Math 11
Calculus-Based Introductory Probability and Statistics

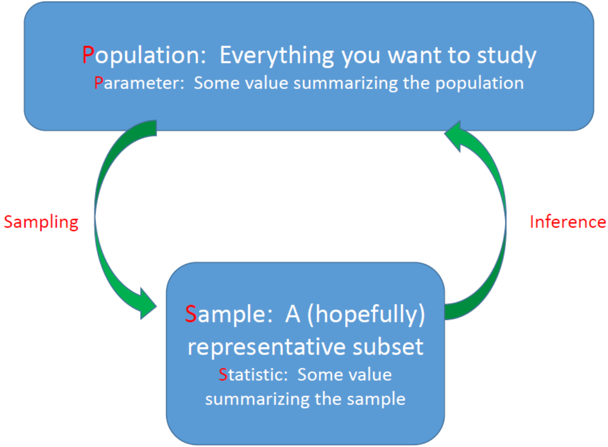
Eddie Aamari
S.E.W. Assistant Professor

`eaamari@ucsd.edu`
`math.ucsd.edu/~eaamari/`
AP&M 5880A

Today:

- Inference of a proportion
- Confidence intervals
- Hypothesis testing

Bringing Probability and Statistics Together: Populations and Samples



Statistical inference is the attempt to say something about the population parameter given a particular sample statistic (i.e. point estimate).

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.

This is **true**: we found $\hat{p} = \dfrac{130}{200} = 65\%$.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.
This is **true**: we found $\hat{p} = \dfrac{130}{200} = 65\%$.

Claim 2: 65% of UCSD students are Facebook users.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.
This is **true**: we found $\hat{p} = \dfrac{130}{200} = 65\%$.

Claim 2: 65% of UCSD students are Facebook users.
This is **false**: the population parameter may not match the sample statistic.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.
This is **true**: we found $\hat{p} = \dfrac{130}{200} = 65\%$.

Claim 2: 65% of UCSD students are Facebook users.
This is **false**: the population parameter may not match the sample statistic.

Claim 3: About 65% of UCSD students use Facebook.

Why We Need Inference

You draw a SRS of 200 UCSD students and ask if they have a Facebook account. 130 say they do.

Claim 1: 65% of our sample are Facebook users.
This is **true**: we found $\hat{p} = \frac{130}{200} = 65\%$.

Claim 2: 65% of UCSD students are Facebook users.
This is **false**: the population parameter may not match the sample statistic.

Claim 3: About 65% of UCSD students use Facebook.
This is **vague**, and we need to learn how to do better. The language "about" is not precise enough for statisticians.

The Key To Inference: The Sampling Distribution

Population: San Diego Population
Parameter: p_{LGBT} (3.9%)

The goal is to find this percentage.
Pretend we don't know it.

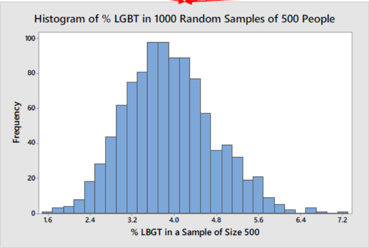
Sample 1: 500 people randomly selected
Statistic: $\boxed{\hat{p}_{\text{LGBT}}}$ in sample 1

Sample 2
$\boxed{\hat{p}_{\text{LGBT}}}$ in sample 2

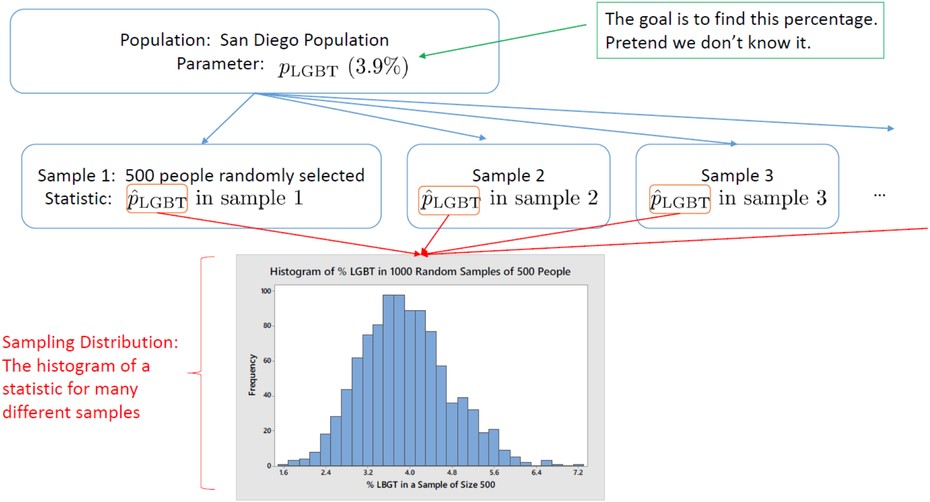
Sample 3
$\boxed{\hat{p}_{\text{LGBT}}}$ in sample 3

...

Sampling Distribution: The histogram of a statistic for many different samples

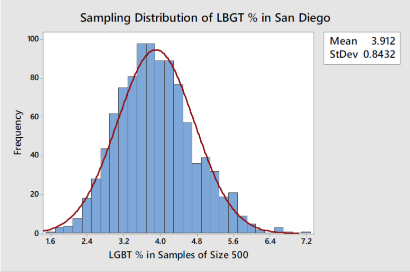


The Key To Inference: The Sampling Distribution

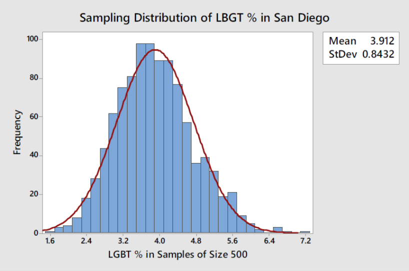


From the Central Limit Theorem, the sampling distribution is (almost) a Normal distribution.

But Which Normal Model?



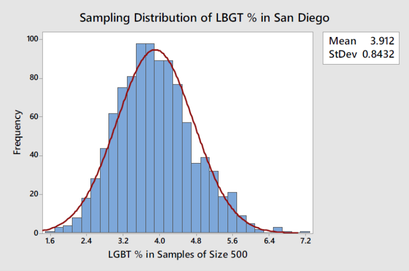
But Which Normal Model?



Two facts from statistics:

$$Center_{model} = p_{LGBT} \qquad SE_{model} = \sqrt{\frac{p_{LGBT}q_{LGBT}}{n}}$$

But Which Normal Model?



Two facts from statistics:

$$Center_{model} = p_{LGBT} \qquad SE_{model} = \sqrt{\frac{p_{LGBT}q_{LGBT}}{n}}$$

Notice that the above formulas use the population **parameters**.
In general, they are NOT known, and this is why you are drawing a sample in the first place.

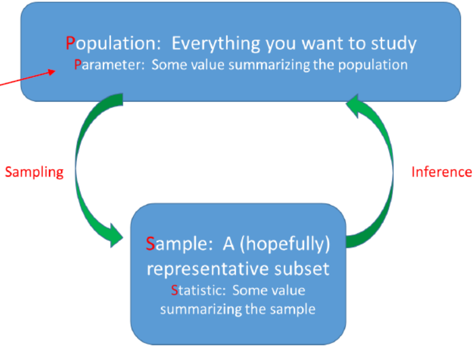
How Could The Sampling Distribution Help When...?

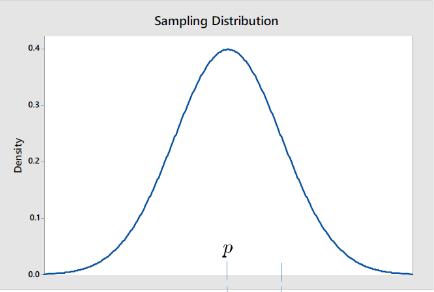
In real life, we don't know the population parameter. So what good is the sampling distribution?

We learned:

$$\text{center}_{\text{model}} = p_{\text{LGBT}}$$

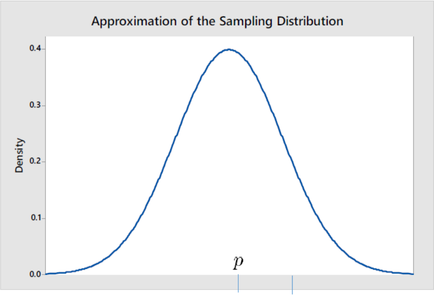
$$\text{spread}_{\text{model}} = \frac{\sqrt{p_{\text{LGBT}} \cdot q_{\text{LGBT}}}}{\sqrt{n}}$$





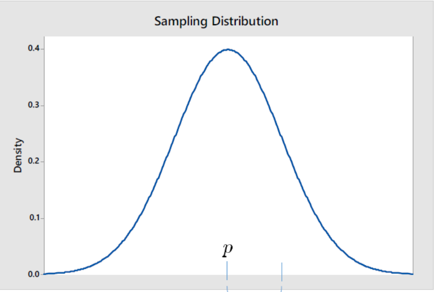
$$SD = \sqrt{\frac{pq}{n}}$$

This model can help us say something more precise, but it is built upon the very information (p) we are trying to find. So, it will be of no use.



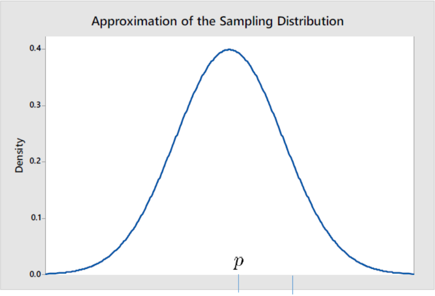
$$SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

We can approximate the sampling distribution by calculating the standard error. This is just the SD formula with the sample statistic (p hat) instead of the population parameter (p).



$$SD = \sqrt{\frac{pq}{n}}$$

This model can help us say something more precise, but it is built upon the very information (*p*) we are trying to find. So, it will be of no use.



$$SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

We can approximate the sampling distribution by calculating the standard error. This is just the SD formula with the sample statistic (p hat) instead of the population parameter (p).

Important Note: Many people do not distinguish between these two worlds and just use SE to mean "the standard deviation of the sampling distribution" (for either the theoretical model or the approximation).

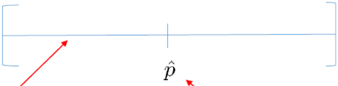
OK. Now What?

OK. Now What?

The sampling distribution helps us create a **confidence interval** (CI), a range of values around a point estimate that convey our uncertainty about the population parameter (as well as a range of plausible values for it).

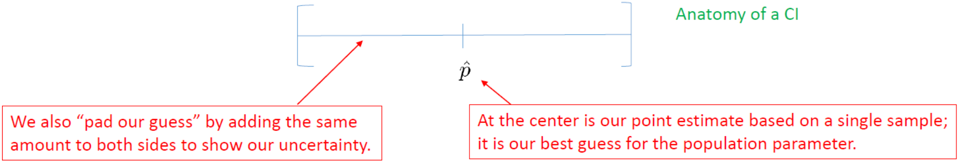
OK. Now What?

The sampling distribution helps us create a **confidence interval** (CI), a range of values around a point estimate that convey our uncertainty about the population parameter (as well as a range of plausible values for it).



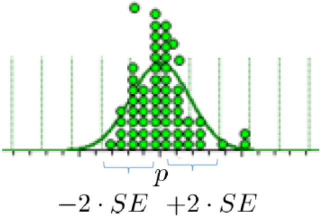
OK. Now What?

The sampling distribution helps us create a **confidence interval** (CI), a range of values around a point estimate that convey our uncertainty about the population parameter (as well as a range of plausible values for it).

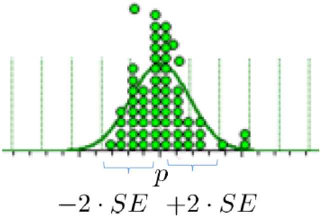


The amount you pad your answer by (i.e., the width of the CI) is determined by how sure you want to be that the interval will contain the true population parameter.
Metaphors: Fishing net, criminal capture radius.

Recall that for ANY normal curve (e.g. the sampling distribution), about 95% of all values fall within 2 standard deviations (SE's) from the center.

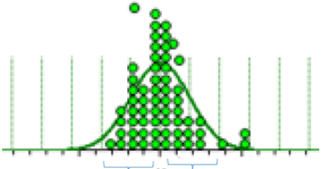


Recall that for ANY normal curve (e.g. the sampling distribution), about 95% of all values fall within 2 standard deviations (SE's) from the center.



So 95% of all point estimates (**green dots**) are within $\pm 2 \times SE$ of p.

Recall that for ANY normal curve (e.g. the sampling distribution), about 95% of all values fall within 2 standard deviations (SE's) from the center.



So 95% of all point estimates (**green dots**) are within $\pm 2 \times SE$ of p.

Said differently, if you stand at a green dot and reach out a distance of $2 \times SE$, 95% of the time your will include p.

Stand at p. Reach out about 2 SEs. You will grab about 95% of sample means (green dots, or values of $\hat{p}$).

equivalent ideas

Stand at a green dot. Reach out about 2 SEs. If you did this at every green dot, about 95% of those reaches include p.

Our First CI (And a Review of the Main Ideas)

1. You want to know the proportion of some trait in a population. You draw a random sample and calculate the proportion in the sample.

Our First CI (And a Review of the Main Ideas)

1. You want to know the proportion of some trait in a population. You draw a random sample and calculate the proportion in the sample.

130 people in our sample of 200 UCSD students use Facebook.
So $\hat{p} = 65\%$.

Our First CI (And a Review of the Main Ideas)

1. You want to know the proportion of some trait in a population. You draw a random sample and calculate the proportion in the sample.

130 people in our sample of 200 UCSD students use Facebook.
So $\hat{p} = 65\%$.

2. You want to infer from the sample back to the population, so you decide on a confidence level C (C is usually 90%, 95% or 99%).

Our First CI (And a Review of the Main Ideas)

1. You want to know the proportion of some trait in a population. You draw a random sample and calculate the proportion in the sample.

130 people in our sample of 200 UCSD students use Facebook.
So $\hat{p} = 65\%$.

2. You want to infer from the sample back to the population, so you decide on a confidence level C (C is usually 90%, 95% or 99%).

When I infer back to the population of **all** UCSD students, I will have to give a range of possible FB percents. I want to be 95% sure that my range has the true population value p.

3. This confidence level dictactes how many SE's you must reach out from $\hat{p}$ to ensure you've grabbed the population parameter p (about $C\%$ of the time).

3. This confidence level dictactes how many SE's you must reach out from $\hat{p}$ to ensure you've grabbed the population parameter p (about $C\%$ of the time).

For a Normal curve, 95% of the area in the middle is withing 2 SE's of the mean. So we must reach 2 SE's each way to grab p with 95% confidence.

3. This confidence level dictactes how many SE's you must reach out from $\hat{p}$ to ensure you've grabbed the population parameter p (about $C\%$ of the time).

For a Normal curve, 95% of the area in the middle is withing 2 SE's of the mean. So we must reach 2 SE's each way to grab p with 95% confidence.

4. Build a confidence interval using the decisions from the previous steps.

3. This confidence level dictactes how many SE's you must reach out from $\hat{p}$ to ensure you've grabbed the population parameter p (about $C\%$ of the time).

For a Normal curve, 95% of the area in the middle is withing 2 SE's of the mean. So we must reach 2 SE's each way to grab p with 95% confidence.

4. Build a confidence interval using the decisions from the previous steps.

In our problem, $SE = \sqrt{\frac{65 \times 35}{200}} \simeq 3.37\%$.

For 95% confidence, we find:

$$\hat{p} \pm 1.96 \times SE = 65\% \pm 1.96 \times 3.37\%$$

Recall that the 68-95-99.7% Rule was only an approximation. From now on, use 1.96 instead of 2 for 95% confidence (see later slides to see how this number is precisely found).

3. This confidence level dictactes how many SE's you must reach out from $\hat{p}$ to ensure you've grabbed the population parameter p (about $C\%$ of the time).

For a Normal curve, 95% of the area in the middle is withing 2 SE's of the mean. So we must reach 2 SE's each way to grab p with 95% confidence.

4. Build a confidence interval using the decisions from the previous steps.

In our problem, $SE = \sqrt{\dfrac{65 \times 35}{200}} \simeq 3.37\%$.

For 95% confidence, we find:

$$\hat{p} \pm 1.96 \times SE = 65\% \pm 1.96 \times 3.37\%$$

Recall that the 68-95-99.7% Rule was only an approximation. From now on, use 1.96 instead of 2 for 95% confidence (see later slides to see how this number is precisely found).

Our confidence interval is $CI = (58.4\%, 71.6\%)$.

5. Write a sentence that summarizes your findings. Be very careful with your language.

5. Write a sentence that summarizes your findings. Be very careful with your language.

We are 95% confident that the percentage of Facebook users among **all** UCSD students is in $CI = (58.26\%, 71.74\%)$.

5. Write a sentence that summarizes your findings. Be very careful with your language.

We are 95% confident that the percentage of Facebook users among **all** UCSD students is in $CI = (58.26\%, 71.74\%)$.

Note: In general, the safest, clearest language to use is:

We are ($C\%$) confident that the (population parameter) is in (CI).

What Does It Really Mean to Be Confident About An Interval?

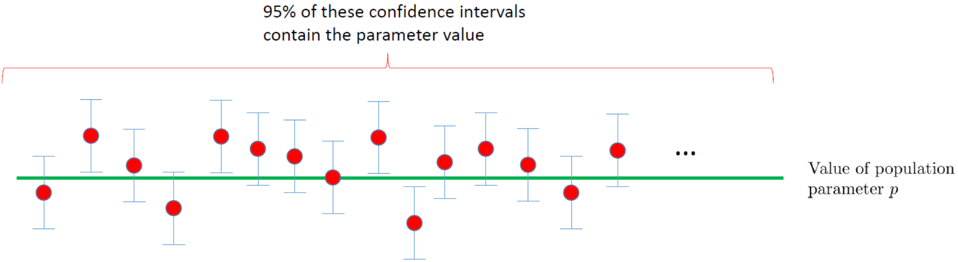
The phrase "95% confident" technically mean this:

If you drew many, many samples, and for each one, you found $\hat{p}$ and built a confidence interval by reaching out $\pm 1.96SE$, then the true population parameter would be in about 95% **of these intervals**.

What Does It Really Mean to Be Confident About An Interval?

The phrase "95% confident" technically mean this:

If you drew many, many samples, and for each one, you found $\hat{p}$ and built a confidence interval by reaching out $\pm 1.96SE$, then the true population parameter would be in about 95% **of these intervals**.



Changing the Confidence Level: Finding the “Critical Value”

New goal:
Construct an 80% CI for the percent of all UCSD on Facebook.

Changing the Confidence Level: Finding the "Critical Value"

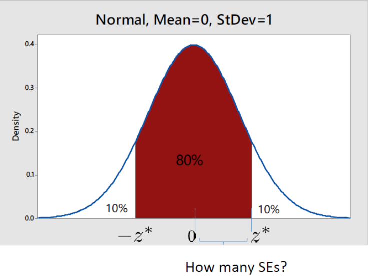
New goal:
Construct an 80% CI for the percent of all UCSD on Facebook.

How many SE's do we need to go in each direction to get 80% of the area under a Normal curve?

Changing the Confidence Level: Finding the "Critical Value"

New goal:
Construct an 80% CI for the percent of all UCSD on Facebook.

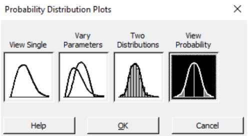
How many SE's do we need to go in each direction to get 80% of the area under a Normal curve?



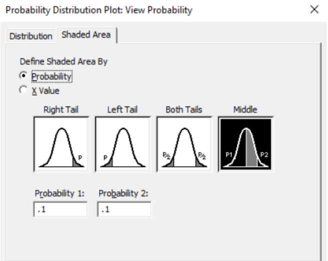
This is like a "reverse" area problem: Instead of calculating the area under the curve up to some z-score (or between z-scores), we need to know the z-score that has 0.8 area between $-z^*$ and z^*.

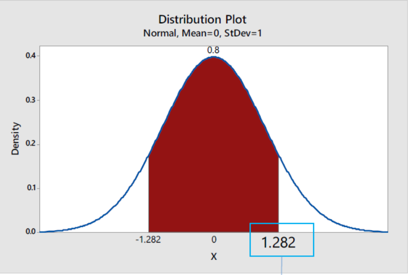
Minitab Can Help with This!

Go to Graph » Probability Distribution Plot » View Probability

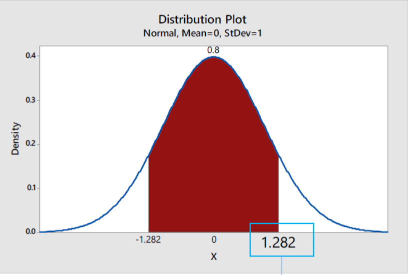


Use the standard Normal curve (mean 0, SD 1), and select the Shaded Area tab. Choose the Probability selector and the "Middle" option. Type in the probabilities (areas under the curve) for the non-shaded parts on the left and right (here, both are 0.1 to get an area of 0.8 in the shaded zone).



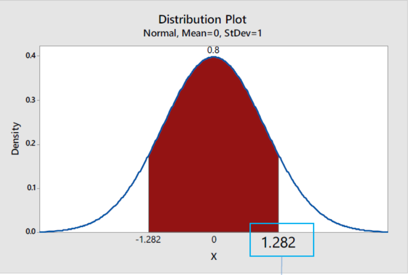


This is z^*, the "critical value". It is HOW MANY SEs we must reach out to grab 80% of the dots (sample proportions).



This is z^*, the "critical value". It is HOW MANY SEs we must reach out to grab 80% of the dots (sample proportions).

Common Confidence Levels in Real World	z^* for that Level of Confidence
90%	1.645
95%	1.960 (notice this isn't 2 because the 68-95-99.7 rule is just an approximation; use 1.96 not 2)
99%	2.576

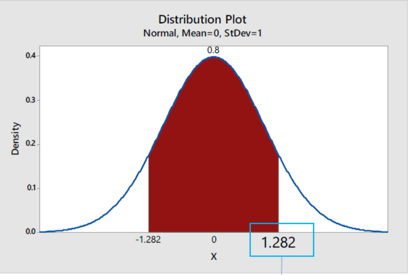


This is z^*, the "critical value". It is HOW MANY SEs we must reach out to grab 80% of the dots (sample proportions).

Common Confidence Levels in Real World	z^* for that Level of Confidence
90%	1.645
95%	1.960 (notice this isn't 2 because the 68-95-99.7 rule is just an approximation; use 1.96 not 2)
99%	2.576

In general, the confidence interval formula is $\hat{p} \pm z^* \times SE$.

We get $65\% \pm 1.282 \times 3.37\% = (60.08\%, 69.32\%)$.



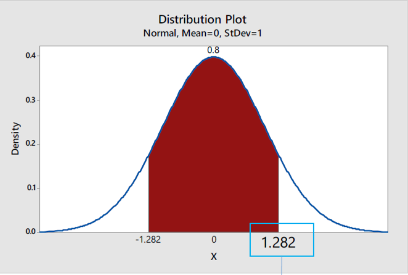
This is z^*, the "critical value". It is HOW MANY SEs we must reach out to grab 80% of the dots (sample proportions).

Common Confidence Levels in Real World	z^* for that Level of Confidence
90%	1.645
95%	1.960 (notice this isn't 2 because the 68-95-99.7 rule is just an approximation; use 1.96 not 2)
99%	2.576

In general, the confidence interval formula is $\hat{p} \pm z^* \times SE$.

We get $65\% \pm 1.282 \times 3.37\% = (60.08\%, 69.32\%)$.

Notice: This is smaller than the 95% CI. Why?



This is z^*, the "critical value". It is HOW MANY SEs we must reach out to grab 80% of the dots (sample proportions).

Common Confidence Levels in Real World	z^* for that Level of Confidence
90%	1.645
95%	1.960 (notice this isn't 2 because the 68-95-99.7 rule is just an approximation; use 1.96 not 2)
99%	2.576

In general, the confidence interval formula is $\hat{p} \pm z^* \times SE$.

We get $65\% \pm 1.282 \times 3.37\% = (60.08\%, 69.32\%)$.

Notice: This is smaller than the 95% CI. Why?
The more confident you want the real population parameter to be in the CI's, the wider the CI's!

The Final Expression For The Confidence Interval

The confidence interval for a proportion is given by

$$\hat{p} \pm \underbrace{z^* \cdot SE(\hat{p})}_{}$$

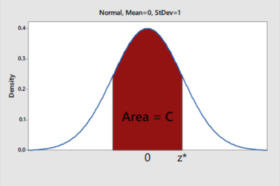
The sample statistic

Margin of Error:
How much you
go up and down
from p hat

The "standard error":
The (approximation of the) standard deviation of the sampling distribution.

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The critical value: If you want a confidence level of $C\%$, this is the z-score z^* on a standard normal curve so that the area under the curve between $-z^*$ and z^* is equal to C.



Practice

Which things **increase** the margin of error in a study?

1. Drawing a larger sample
2. Drawing a smaller sample
3. Demanding a higher level of confidence
4. Demanding a lower level of confidence

Practice

Which things **increase** the margin of error in a study?

1. Drawing a larger sample
2. Drawing a smaller sample
3. Demanding a higher level of confidence
4. Demanding a lower level of confidence

Answer: 2. and 3.

Practice

You work for a polling company exploring an upcoming election. Your boss demands a very, very small confidence interval. Which are good options for getting a small confidence interval?

1. Switch from a 95% level of confidence to a 20% level of confidence. This shrinks the confidence interval dramatically
2. Increase your sample size from 1,000 to 100,000. This makes the SE much smaller, decreases the margin of error, and creates a small confidence interval
3. Both are good ideas
4. Both are bad ideas

Practice

You work for a polling company exploring an upcoming election. Your boss demands a very, very small confidence interval. Which are good options for getting a small confidence interval?

1. Switch from a 95% level of confidence to a 20% level of confidence. This shrinks the confidence interval dramatically
2. Increase your sample size from 1,000 to 100,000. This makes the SE much smaller, decreases the margin of error, and creates a small confidence interval
3. Both are good ideas
4. Both are bad ideas

Answer: 3.

Practice

You work for a polling company exploring an upcoming election. Your boss demands a very, very small confidence interval. Which are good options for getting a small confidence interval?

1. Switch from a 95% level of confidence to a 20% level of confidence. This shrinks the confidence interval dramatically
2. Increase your sample size from 1,000 to 100,000. This makes the SE much smaller, decreases the margin of error, and creates a small confidence interval
3. Both are good ideas
4. Both are bad ideas

Answer: 3.
Warning: 1. decreasing the confidence level and 2. increasing the sample size both have the CI shrink, but they also make the whole process dirty/messy:

- A 20% level of confidence yields a very unreliable CI
- A sample size of 100,000 might be very costly to collect.

In practice, these are actually artificial solutions to the problem.

Practice

You survey 200 random UCSD students and find that 43% love Twitter. The margin of error for your study is 7% given that you demanded a 95% confidence level.
Which of the below statements are correct, responsible ways to report your findings?

1. The % of Twitter lovers at UCSD is 43%
2. The % of Twitter lovers at UCSD is about 43%
3. The % of Twitter lovers at UCSD is in the interval $(36\%, 50\%)$
4. If we drew many, many samples of size 200, 95% of the confidence intervals would contain the % of UCSD students who love Twitter
5. I am 95% confident that the true % of students at UCSD who love Twitter is in $(36\%, 50\%)$
6. 95% of the time, the true % of students at UCSD who love Twitter is in $(36\%, 50\%)$

Practice

You survey 200 random UCSD students and find that 43% love Twitter. The margin of error for your study is 7% given that you demanded a 95% confidence level.
Which of the below statements are correct, responsible ways to report your findings?

1. The % of Twitter lovers at UCSD is 43%
2. The % of Twitter lovers at UCSD is about 43%
3. The % of Twitter lovers at UCSD is in the interval $(36\%, 50\%)$
4. If we drew many, many samples of size 200, 95% of the confidence intervals would contain the % of UCSD students who love Twitter
5. I am 95% confident that the true % of students at UCSD who love Twitter is in $(36\%, 50\%)$
6. 95% of the time, the true % of students at UCSD who love Twitter is in $(36\%, 50\%)$

Answer: 4. and 5.
2. is false because it is too vague.
6. is wrong because the true % of Twitter lovers is either in the CI, or outside the CI. (100% or 0%)

Always Check Your Conditions!

To be able to approximate the sampling distribution of $\hat{p}$ by a Normal Curve, you need:

Independence: The surveyed persons must be picked randomly, and you must not sample more than 10% of the total population.

10 Successes/10 Failures Conditions: Technically, this means that you need $n\hat{p} \geq 10$ and $n\hat{q} = n(1 - \hat{p}) \geq 10$.

Why Exactly Are Bigger Samples Better?

Whether we are studying the proportion in a sample, or the mean of a sample, there is some variability in the answer we get. As we learnt, the amount of variation is given by

$$SE_{model} = \frac{\sqrt{pq}}{\sqrt{n}} \text{ for proportions,}$$

$$SE_{model} = \frac{\sigma}{\sqrt{n}} \text{ for means.}$$

So, each time you quadruple n, you cut the variability in half (because the square root on n in the denominator).

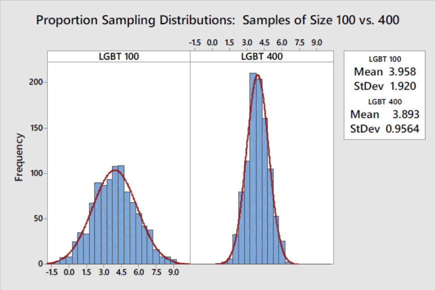
Why Exactly Are Bigger Samples Better?

Whether we are studying the proportion in a sample, or the mean of a sample, there is some variability in the answer we get. As we learnt, the amount of variation is given by

$$SE_{model} = \frac{\sqrt{pq}}{\sqrt{n}} \text{ for proportions,}$$

$$SE_{model} = \frac{\sigma}{\sqrt{n}} \text{ for means.}$$

So, each time you quadruple n, you cut the variability in half (because the square root on n in the denominator).



Imagine taking many samples of 100 San Diegans and finding the LGBT proportion. Imagine taking many samples of 400 San Diegans and finding the LGBT proportion.

With larger samples, there is less variation in the sampling distribution, so you are more confident that the answer you get is closer to the "true" percentage in the population!

This will also cut the width of CI's by half since $CI = \hat{p} \pm z^* SE$.

Choosing a Sample Size

Your boss thinks sampling 200 students was too few. If she wants to maintain a 95% level of confidence, how many students must be chosen in a new sample if the goal is to get a 2% margin of error (ME)?

Choosing a Sample Size

Your boss thinks sampling 200 students was too few. If she wants to maintain a 95% level of confidence, how many students must be chosen in a new sample if the goal is to get a 2% margin of error (ME)?

We know that $ME = z^* SE(\hat{p})$.

Choosing a Sample Size

Your boss thinks sampling 200 students was too few. If she wants to maintain a 95% level of confidence, how many students must be chosen in a new sample if the goal is to get a 2% margin of error (ME)?

We know that $ME = z^* SE(\hat{p})$.

Demanding a 95% confidence level requires $z^* = 1.96$ (not 2!)

Choosing a Sample Size

Your boss thinks sampling 200 students was too few. If she wants to maintain a 95% level of confidence, how many students must be chosen in a new sample if the goal is to get a 2% margin of error (ME)?

We know that $ME = z^* SE(\hat{p})$.

Demanding a 95% confidence level requires $z^* = 1.96$ (not 2!)

We use $\hat{p} = 65\%$ from our first study to guide our new sampling:

$$2\% = 1.96 \times \sqrt{\frac{65 \times 35}{n}} \qquad \Longrightarrow \qquad \sqrt{n} = \frac{1.96 \times \sqrt{65 \times 35}}{2},$$

Choosing a Sample Size

Your boss thinks sampling 200 students was too few. If she wants to maintain a 95% level of confidence, how many students must be chosen in a new sample if the goal is to get a 2% margin of error (ME)?

We know that $ME = z^* SE(\hat{p})$.

Demanding a 95% confidence level requires $z^* = 1.96$ (not 2!)

We use $\hat{p} = 65\%$ from our first study to guide our new sampling:

$$2\% = 1.96 \times \sqrt{\frac{65 \times 35}{n}} \qquad \Longrightarrow \qquad \sqrt{n} = \frac{1.96 \times \sqrt{65 \times 35}}{2},$$

so $n \simeq 2185$ students.

Am I (Really) Different?

Sometimes you draw a sample and calculate a proportion not just to find the proportion, but to see if it is different than you expected.

Am I (Really) Different?

Sometimes you draw a sample and calculate a proportion not just to find the proportion, but to see if it is different than you expected.

You contact 500 random San Diegans to learn if the percentage of Asian-American is **different** than the national average.

Am I (Really) Different?

Sometimes you draw a sample and calculate a proportion not just to find the proportion, but to see if it is different than you expected.

You contact 500 random San Diegans to learn if the percentage of Asian-American is **different** than the national average.

You wonder if giving 200 random freshmen a "How to Succeed in College" course will **decrease** the proportion that dropout as compared to the general student population.

Am I (Really) Different?

Sometimes you draw a sample and calculate a proportion not just to find the proportion, but to see if it is different than you expected.

You contact 500 random San Diegans to learn if the percentage of Asian-American is **different** than the national average.

You wonder if giving 200 random freshmen a "How to Succeed in College" course will **decrease** the proportion that dropout as compared to the general student population.

You administer a new drug to 350 heartburn patients and see what percentage report an **improvement** in symptoms versus a placebo.

Hypothesis Testing

Hypothesis: A claim that may or may not be true.

Hypothesis Testing

Hypothesis: A claim that may or may not be true.

Asian American: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Hypothesis Testing

Hypothesis: A claim that may or may not be true.

Asian American: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Success Course: $p_{\text{Drop out with course}} < p_{\text{Drop out without course}}$

Hypothesis Testing

Hypothesis: A claim that may or may not be true.

Asian American: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Success Course: $p_{\text{Drop out with course}} < p_{\text{Drop out without course}}$

Drug: $p_{\text{Symptom relief with drug}} > p_{\text{Symptom relief with placebo}}$

Hypothesis Testing

Hypothesis: A claim that may or may not be true.

Asian American: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Success Course: $p_{\text{Drop out with course}} < p_{\text{Drop out without course}}$

Drug: $p_{\text{Symptom relief with drug}} > p_{\text{Symptom relief with placebo}}$

Suppose we draw a sample and find $\hat{p}_{\text{new drug}} = 0.14$.
If we are told $p_{\text{placebo}} = 0.11$, how do we decide if the difference we see is sampling variability or suggestive evidence of a real difference?

Sample Variation or True Effect?

Hypothesis testing is the rigorous way statisticians have devised to sort out how confident we can be that sampling variation is not the cause.

Sample Variation or True Effect?

Hypothesis testing is the rigorous way statisticians have devised to sort out how confident we can be that sampling variation is not the cause.

Step 1: Write down a **null hypothesis** (H_0).
This is a statement that says nothing interesting is happening. It (almost) always uses an equal sign. It uses population parameters.

Sample Variation or True Effect?

Hypothesis testing is the rigorous way statisticians have devised to sort out how confident we can be that sampling variation is not the cause.

Step 1: Write down a **null hypothesis** (H_0).
This is a statement that says nothing interesting is happening. It (almost) always uses an equal sign. It uses population parameters.

Asian American: H_0: $p_{\text{AA in SD}} = p_{\text{AA in US}}$

Drug: H_0: $p_{\text{Symptom relief with drug}} = p_{\text{Symptom relief with placebo}}$

Step 2: Write down an **alternative hypothesis** (H_A).
This is what you suspect might be true and is what you hope to show.

Step 2: Write down an **alternative hypothesis** (H_A).
This is what you suspect might be true and is what you hope to show.

Asian American: H_A: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Drug: H_A: $p_{\text{Symptom relief with drug}} > p_{\text{Symptom relief with placebo}}$

Step 2: Write down an **alternative hypothesis** (H_A).
This is what you suspect might be true and is what you hope to show.

Asian American: H_A: $p_{\text{AA in SD}} \neq p_{\text{AA in US}}$

Drug: H_A: $p_{\text{Symptom relief with drug}} > p_{\text{Symptom relief with placebo}}$

Two types of alternative hypotheses:

- A **one-sided alternative hypothesis** will use a $>$ or $<$ sign. You are hoping your percentage is on a certain side of the comparison percentage.
- A **two-sided alternative hypothesis** will use a $\neq$. You are just wondering if your percentage is different than the comparison percentage.

The kind of alternative hypothesis you use simply depends on what you are guessing/hoping might be true (<u>before</u> any data are collected).

Mimicking real-life

How do we decide between H_0 and H_A?
Answer: How we often decide between beliefs in real life:

Mimicking real-life

How do we decide between H_0 and H_A?
Answer: How we often decide between beliefs in real life:

If the data supports your belief,
you continue in this mindset

Adopt some belief
for the moment
"SD is usually cold,
let's wear trousers"

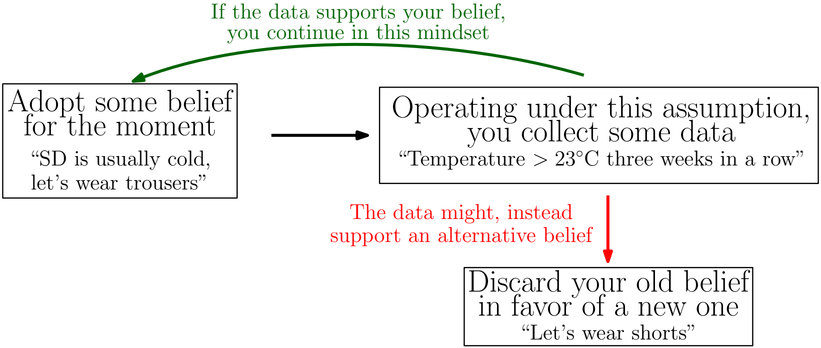
Operating under this assumption,
you collect some data
"Temperature > 23°C three weeks in a row"

The data might, instead
support an alternative belief

Discard your old belief
in favor of a new one
"Let's wear shorts"

Mimicking real-life

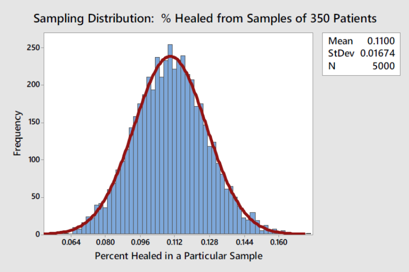
How do we decide between H_0 and H_A?
Answer: How we often decide between beliefs in real life:



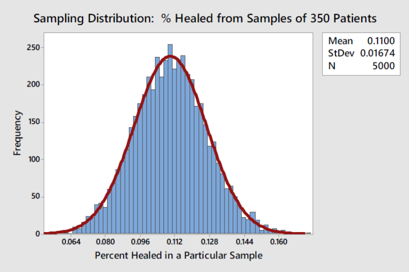
Notice that you are comparing the data from your life against some belief that you hold temporarily (here, wearing trousers). Perhaps the data support it, perhaps they support movement to an alternative.

Step 3: Draw a sample and consider it assuming H_0 is true.

Step 3: Draw a sample and consider it assuming H_0 is true.
Say, in a sample of 350 taking the new drug, 14% show improvement.



Step 3: Draw a sample and consider it assuming H_0 is true.
Say, in a sample of 350 taking the new drug, 14% show improvement.



The universe where our drug is the same as a placebo (H_0) would have a sampling distribution centered at the placebo's healing percentage (11%), with a standard error we can easily calculate:

$$\mu_{model} = p_{placebo} = 0.11$$

$$SE = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.11 \times .89}{350}} \simeq 0.0167.$$

Notice: in the universe where our drug is no different than a placebo, it is possible to get healing percentages around 14% just from random chance.

How Does The Sampling Distribution Help?

Use the normal model to calculate the probability of getting the observed percentage (14%) or anything more extreme.

How Does The Sampling Distribution Help?

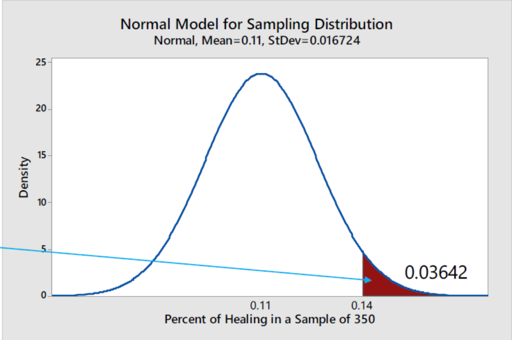
Use the normal model to calculate the probability of getting the observed percentage (14%) or anything more extreme.
(We must use the language "14% or more extreme" because $P(X = 14\%) = 0$ since we have a continuous distribution)

How Does The Sampling Distribution Help?

Use the normal model to calculate the probability of getting the observed percentage (14%) or anything more extreme.
(We must use the language "14% or more extreme" because $P(X = 14\%) = 0$ since we have a continuous distribution)

The value we get is called a **P-value**. It is a probability: the chance of seeing our result (14%) or something more extreme if our universe is "H_0: The drug works just as well as a placebo".

Our sample is among the top 3.6% biggest percentages the sampling distribution would give us. That's strange...



Step 4: Decide what you wish to say about the null hypothesis given the p-value.

Step 4: Decide what you wish to say about the null hypothesis given the p-value.

Two Possible Choices:

1) Reject the null hypothesis.You do this when your p-value (here, 0.036) is quite small; many scientific journals suggest you do this when the p-value is below 0.05 ("cutoff" or "significance level"). The observed value (14%) seems really out of place in your universe (here, a drug = placebo 11% universe).

2) Do not reject the null hypothesis. Do this when your p-value isn't particularly small.
The observed value isn't that out of place in your universe.

Step 4: Decide what you wish to say about the null hypothesis given the p-value.

Two Possible Choices:

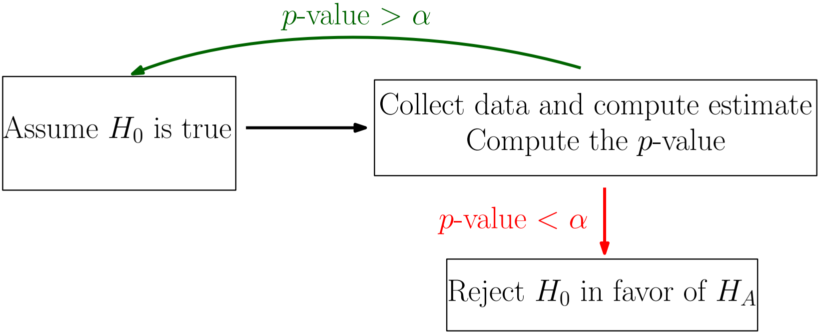
1) Reject the null hypothesis.You do this when your p-value (here, 0.036) is quite small; many scientific journals suggest you do this when the p-value is below 0.05 ("cutoff" or "significance level"). The observed value (14%) seems really out of place in your universe (here, a drug = placebo 11% universe).

2) Do not reject the null hypothesis. Do this when your p-value isn't particularly small.
The observed value isn't that out of place in your universe.

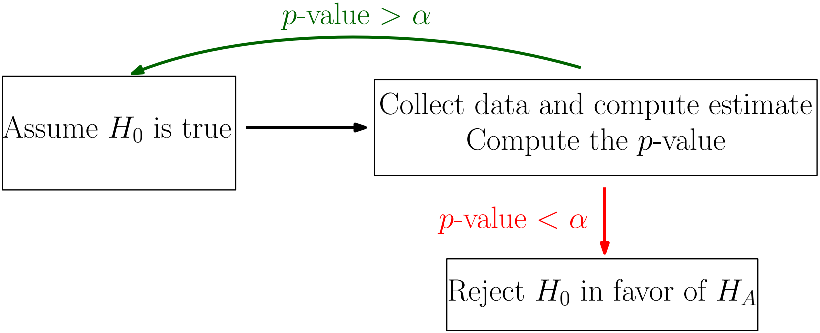
In our drug example, we get a p-value of 0.036. If the drug really is no more effective than a placebo, then only 3.6 samples in 100 would give us this result (or something more extreme). As such, we reject the null hypothesis:

There is good evidence the drug is more effective than the placebo.

Hypothesis Testing Framework



Hypothesis Testing Framework



Note that our data **do not** prove the null is true, nor that the alternative is true.

The data simply suggests which we should adopt moving forward.