Today:

- Hypothesis testing
- $p$-value and its interpretation
The Final Expression For The Confidence Interval

The confidence interval for a proportion is given by

\[ \hat{p} \pm z^* \cdot SE(\hat{p}) \]

The sample statistic

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.
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: 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: \(H_0: p_{AA \text{ in SD}} = p_{AA \text{ in US}}\)

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

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: \(H_A: p_{AA \text{ in SD}} \neq p_{AA \text{ in US}}\)

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 (before 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:

- Adopt some belief for the moment
- "SD is usually cold, let's wear trousers"
- Operating under this assumption, you collect some data
  - "Temperature $> 23\, ^\circ C$ three weeks in a row"
- If the data supports your belief, you continue in this mindset
- The data might, instead, support an alternative belief
- Discard your old belief in favor of a new one
  - "Let's wear shorts"

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.
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:

1. Adopt some belief for the moment. “SD is usually cold, let’s wear trousers.”
2. Operating under this assumption, you collect some data. “Temperature $> 23^\circ C$ three weeks in a row.”
3. The data might, instead support an alternative belief.
4. Discard your old belief in favor of a new one. “Let’s wear shorts.”

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.

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:

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

$$SE = \sqrt{\frac{pq}{n}} = \sqrt{0.11 \times 0.89 \times 350} \approx 0.0167.$$

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.
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:

\[
\mu_{model} = p_{placebo} = 0.11
\]

\[
SE = \sqrt{\frac{pq}{n}} = \sqrt{0.11 \times 0.89} \approx 0.0167.
\]

**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.

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.
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

Assume $H_0$ is true

Collect data and compute estimate

Compute the $p$-value

Reject $H_0$ in favor of $H_A$

$p$-value > $\alpha$

$p$-value < $\alpha$

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.
Hypothesis Testing Framework

Assume $H_0$ is true

Collect data and compute estimate

Compute the $p$-value

Reject $H_0$ in favor of $H_A$

$p$-value $> \alpha$

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.
Another Example

You read that 5.6% of Americans identify as Asian. You wonder if San Diego is different. After sampling 400 random San Diegans, you find that 17 self-identify as Asian. What do you make of this?
Another Example

You read that 5.6% of Americans identify as Asian. You wonder if San Diego is different. After sampling 400 random San Diegans, you find that 17 self-identify as Asian. What do you make of this?

Let \( p \) be the percentage of Asians in San Diego.
(Always begin by defining your parameter)
You read that 5.6% of Americans identify as Asian. You wonder if San Diego is different. After sampling 400 random San Diegans, you find that 17 self-identify as Asian. What do you make of this?

Let $p$ be the percentage of Asians in San Diego.  
(Always begin by defining your parameter)

$H_0$: $p = 0.056$  
$H_A$: $p \neq 0.056$  
(Choose one-sided vs. two-sided based on what would be interesting to you, not based on what the data suggest)
Another Example

You read that 5.6% of Americans identify as Asian. You wonder if San Diego is different. After sampling 400 random San Diegans, you find that 17 self-identify as Asian. What do you make of this?

Let $p$ be the percentage of Asians in San Diego.
(Always begin by defining your parameter)

$H_0$: $p = 0.056$
$H_A$: $p \neq 0.056$
(Choose one-sided vs. two-sided based on what would be interesting to you, not based on what the data suggest)

We assume $H_0$ is true (to get started). In our particular sample, we get

$$\hat{p} = \frac{17}{400} = 0.0425.$$
By assuming $H_0$, we build a universe where $p = 0.056$, and any samples (of size $n = 400$) are drawn from such a universe.
By assuming $H_0$, we build a universe where $p = 0.056$, and any samples (of size $n = 400$) are drawn from such a universe.

The sampling distribution is approximately Normal if we meet the Independence and 10 Successes/Failures conditions:

- We randomly chose people and 400 is far less than 10% of San Diego’s total population.
- We expect $np = 400 \times 0.056 = 22.4 \geq 10$ successes and $nq = 400 \times 0.944 = 377.6 \geq 10$ failures.
By assuming $H_0$, we build a universe where $p = 0.056$, and any samples (of size $n = 400$) are drawn from such a universe.

The sampling distribution is approximately Normal if we meet the Independence and 10 Successes/Failures conditions:

- We randomly chose people and 400 is far less than 10% of San Diego’s total population.
- We expect $np = 400 \times 0.056 = 22.4 \geq 10$ successes and $nq = 400 \times 0.944 = 377.6 \geq 10$ failures.

Sampling distribution: is (approximately) Normal with parameters

$$
\mu = p = 0.056 \text{ and } SE = \sqrt{\frac{0.056 \times 0.944}{400}} \approx 0.0115.
$$
For a **two-sided alternative**, plot your sample and the symmetrically placed result in the picture (0.0425 and 0.0695).

Shade both tails.
For a **two-sided alternative**, plot your sample and the symmetrically placed result in the picture (0.0425 and 0.0695).

Shade both tails.

Our $p$-value is $2 \times 0.1202 \simeq 0.24 > 0.05$.
Here, we do not reject $H_0$.
Our result is not strange enough for us to abandon $H_0$. 
“Death Postponement”: The theory that people will somehow delay their death until after an important life event (e.g., birthday, wedding of a child, etc...).
“Death Postponement”: The theory that people will somehow delay their death until after an important life event (e.g., birthday, wedding of a child, etc...).

Let $p$ be the percentage of people that die in the three-month window before their birthdays.
“Death Postponement”: The theory that people will somehow delay their death until after an important life event (e.g., birthday, wedding of a child, etc...).

Let $p$ be the percentage of people that die in the three-month window before their birthdays.

$H_0$: Death postponement is nonsense: $p = 1/4$.

$H_A$: Death postponement is real: $p < 1/4$. 

Researchers looked at 747 deaths in Salt Lake City and found 60 deaths occurred in the three-month window before a person’s birthday. (Newsweek, 3/6/1978)
“Death Postponement”: The theory that people will somehow delay their death until after an important life event (e.g., birthday, wedding of a child, etc...).

Let \( p \) be the percentage of people that die in the three-month window before their birthdays.

\( H_0: \) Death postponement is nonsense: \( p = \frac{1}{4} \).
\( H_A: \) Death postponement is real: \( p < \frac{1}{4} \).

Researchers looked at 747 deaths in Salt Lake City and found 60 deaths occurred in the three-month window before a person’s birthday. \((\text{Newsweek, 3/6/1978})\)
To the Sampling Distribution!

Assuming $H_0$, the universe should give us sample from

$$N \left( p, \sqrt{\frac{pq}{n}} \right) \sim N \left( 0.25, \sqrt{\frac{0.25 \times 0.75}{747}} \right)$$

$$= N(0.25, 0.0158)$$

Our data gives $\hat{p} = \frac{60}{747} \sim 0.08$. 
To the Sampling Distribution!

Assuming $H_0$, the universe should give us sample from

$$N \left( p, \sqrt{\frac{pq}{n}} \right) \approx N \left( 0.25, \sqrt{\frac{0.25 \times 0.75}{747}} \right)$$

$$= N(0.25, 0.0158)$$

Our data gives $\hat{p} = \frac{60}{747} \approx 0.08$.

Shading the area to the left of 0.08 gives a $p$-value of $2.67 \cdot 10^{-27} \ll 0.05$. 
To the Sampling Distribution!

Assuming $H_0$, the universe should give us sample from

$$N \left( p, \sqrt{pq/n} \right) \simeq N \left( 0.25, \sqrt{\frac{0.25 \times 0.75}{747}} \right)$$

$$= N(0.25, 0.0158)$$

Our data gives $\hat{p} = \frac{60}{747} \simeq 0.08$.

Shading the area to the left of 0.08 gives a $p$-value of $2.67 \cdot 10^{-27} \ll 0.05$.

We reject $H_0$ in favor of $H_A$. 
Where Does This 0.05 Cutoff Value Come From?

$\alpha = 0.05$ is a historical artifact derived from one sentence in a 1931 book by R.A. Fisher, *The design of Experiments*. He thought that a 1 in 20 event (\(= 5\%\)) might be surprising enough to toss out one's belief system (\(H_0\)) in favor of something else (\(H_A\)).

Some fields have a far more demanding threshold like $\alpha = 0.0000003$. This is usually called the "5 sigma rule": you need to see an event 5 SE's from the assumed mean in order to discard \(H_0\) in favor of \(H_A\).

Examples:

- Particle physics
- Pharmacology
- Aircraft design processes
α = 0.05 is a historical artifact derived from one sentence in a 1931 book by R.A. Fisher, *The design of Experiments*. He thought that a 1 in 20 event (≈ 5%) might be surprising enough to toss out one’s belief system (H₀) in favor of something else (Hₐ).
Where Does This 0.05 Cutoff Value Come From?

$\alpha = 0.05$ is a historical artifact derived from one sentence in a 1931 book by R.A. Fisher, *The design of Experiments*. He thought that a 1 in 20 event (= 5%) might be surprising enough to toss out one’s belief system ($H_0$) in favor of something else ($H_A$).

Some fields have a far more demanding threshold like $\alpha = 0.0000003$. This is usually called the “5 sigma rule”: you need to see an event 5$SE$’s from the assumed mean in order to discard $H_0$ in favor of $H_A$.

Examples:

- Particule physics
- Pharmacology
- Aircraft design processes
$P$ Overload!

- $p$ is the proportion of some trait in a population.
  It is a parameter.

- $\hat{p}$ is the proportion of some trait in a sample.
  It is a statistic.

- $P(A)$ means the probability of some event $A$ occurring.
  It is a probability.

- A $p$-value is a conditional probability:
  It is the probability of getting the value $\hat{p}$ (or something more extreme) in a universe where $p$ is the law of the land. That is,

  $$p\text{-value} = P(\hat{p} \text{ or something more extreme} \mid H_0 \text{ is true}).$$

  It is calculated by finding an area under a sampling distribution curve, whose shape is determined by $H_0$. 
Does Extra-Sensory Perception Exist?

In a 2011 article, Daryl Benn claims to have found evidence for Extra-Sensory Perception (ESP). Participants had to choose which of two curtains on a computer screen had an erotic picture behind it. They were able to do this 829 out of 1560 times. Do these data suggest the ability to perceive erotica beyond what we expect from random chance?
Does Extra-Sensory Perception Exist?

In a 2011 article, Daryl Benn claims to have found evidence for Extra-Sensory Perception (ESP). Participants had to choose which of two curtains on a computer screen had an erotic picture behind it. They were able to do this 829 out of 1560 times. Do these data suggest the ability to perceive erotica beyond what we expect from random chance?

\( H_0 \): ESP does not exist with erotic pictures.

\( H_A \): ESP allows for better-than-random perception of erotic imagery.
Does Extra-Sensory Perception Exist?

In a 2011 article, Daryl Benn claims to have found evidence for Extra-Sensory Perception (ESP). Participants had to choose which of two curtains on a computer screen had an erotic picture behind it. They were able to do this 829 out of 1560 times.

Do these data suggest the ability to perceive erotica beyond what we expect from random chance?

\( H_0: \) ESP does not exist with erotic pictures.
\( H_A: \) ESP allows for better-than-random perception of erotic imagery.

Let \( p \) be the percentage of erotic pictures identified by those claiming to have ESP. We have

\[
H_0: \ p = 0.5 \\
H_A: \ p > 0.5
\]

In this study, \( \hat{p} = \frac{829}{1560} = 0.531 \).
Under $H_0$, we are on the sampling distribution

$$N \left( 0.5, \sqrt{\frac{0.5 \times 0.5}{1560}} \right) \simeq N(0.5, 0.01266).$$
Under $H_0$, we are on the sampling distribution

$$N \left( 0.5, \sqrt{\frac{0.5 \times 0.5}{1560}} \right) \approx N(0.5, 0.01266).$$

These data are strong enough to move to the alternative saying that ESP exists!!

Such a study is part of the field of Parapsychology. For more info on such studies, see a conference of Chris French

Remark: C. French and D. Bem aren't best friends... (link)

Tech approach: Use Minitab/calculator to find the P-value.

With $P = 0.007 < 0.05$, we reject $H_0$!

$$Z = \frac{0.531 - 0.5}{0.01266} \approx 2.45$$

Now use a Z-table to get the same answer.
Under $H_0$, we are on the sampling distribution

$$N \left( 0.5, \sqrt{0.5 \times 0.5 \over 1560} \right) \approx N(0.5, 0.01266).$$

These data are strong enough to move to the alternative saying that ESP exists!!

Tech approach: Use Minitab/calculator to find the P-value.

With $P = 0.007 < 0.05$, we reject $H_0$!

These data are strong enough to move to the alternative saying that ESP exists!!

$$Z = \frac{0.531 - 0.5}{0.01266} \approx 2.45$$

Now use a Z-table to get the same answer.
Under $H_0$, we are on the sampling distribution

$$N \left(0.5, \sqrt{\frac{0.5 \times 0.5}{1560}} \right) \simeq N(0.5, 0.01266).$$

These data are strong enough to move to the alternative saying that ESP exists!!

Such a study is part of the field of Parapsychology. For more info on such studies, see a conference of Chris French

Remark: C. French and D. Bem aren't best friends... (link)

Tech approach: Use Minitab/calculator to find the P-value.

With $P = 0.007 < 0.05$, we reject $H_0$!

These data are strong enough to move to the alternative saying that ESP exists!!

Now use a Z-table to get the same answer.

$$Z = \frac{0.531 - 0.5}{0.01266} \approx 2.45$$