Lecture 10 - Inference for One Mean

Sta 102

May 27, 2016

Colin Rundel & Mine Çetinkaya-Rundel

Confidence Intervals in the Real World

A small problem

Lets assume we are interested in understanding the blood pressure of high school athletes, we collect a large sample (n=200) from the population and record each student's blood pressure in mmHg. Here we do not know either the population mean (μ) or variance (σ^2).

We want to construct a 95% confidence interval based on the observed sample average, which we do by calculating:

$$\begin{aligned} CI_{95\%} &= \bar{X} \pm Z^\star \; SE \\ &= \bar{X} \pm 1.96 \; \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Missing σ

When working with real samples the population standard deviation (σ) is almost never known, we address this by plugging in the sample standard deviation when calculating the standard error. However, when we do this it changes the sampling distribution.

- We estimate the standard error using the sample standard deviation, this adds uncertainty to inference process.

Missing σ

When working with real samples the population standard deviation (σ) is almost never known, we address this by plugging in the sample standard deviation when calculating the standard error. However, when we do this it changes the sampling distribution.

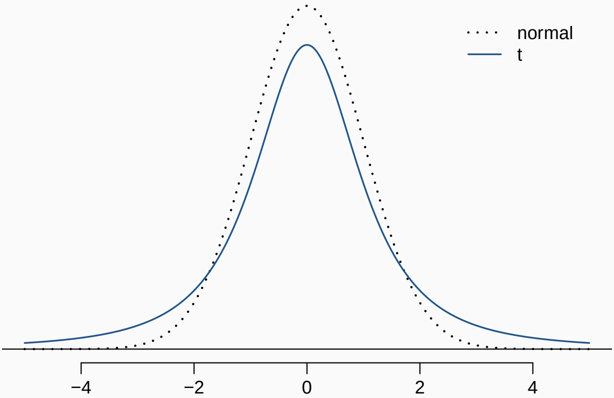
- We estimate the standard error using the sample standard deviation, this adds uncertainty to inference process.
- Our new sampling distribution is still symmetric and roughly bell shaped, but its tails are *thicker* than the normal distribution.

Missing σ

When working with real samples the population standard deviation (σ) is almost never known, we address this by plugging in the sample standard deviation when calculating the standard error. However, when we do this it changes the sampling distribution.

- We estimate the standard error using the sample standard deviation, this adds uncertainty to inference process.
- Our new sampling distribution is still symmetric and roughly bell shaped, but its tails are *thicker* than the normal distribution.
- Observations are more likely to fall beyond two SDs from the mean than with the normal distribution.

t distribution

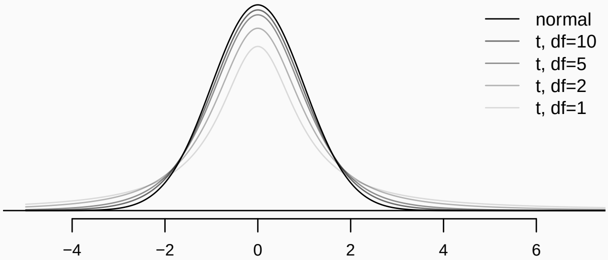


First described by William Gosset ...

- Oxford Graduate with a degree in Chemistry and Mathematics
- Hired by the Guinness Brewery in 1899
- Spent 1906 - 1907 studying with Karl Pearson
- Published "The probable error of a mean" in 1908 under the pseudonym "A. Student"
- Much of his work was promoted by R.A. Fisher

Properties of the t distribution

- it is centered at zero*, like the standard normal (Z) distribution.
- it has a single parameter, *df* (*degrees of freedom*), which determines the thickness of the tails.



- as *df* increases the t distribution converges to the Z distribution.

Finding probabilities

As before we can find any probability we are interested by knowing how to calculate the area under the tail of the t distribution. For example, if we want to know $P(T_{df=19} > 1.16)$ then we can use:

Finding probabilities

As before we can find any probability we are interested by knowing how to calculate the area under the tail of the t distribution. For example, if we want to know $P(T_{df=19} > 1.16)$ then we can use:

- R:

```
1-pt(1.16,df=19)

## [1] 0.1302092
```

Finding probabilities

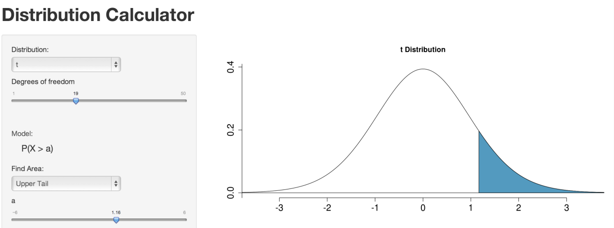
As before we can find any probability we are interested by knowing how to calculate the area under the tail of the t distribution. For example, if we want to know $P(T_{df=19} > 1.16)$ then we can use:

- R:

```
1-pt(1.16,df=19)

## [1] 0.1302092
```

- App: (`https://gallery.shinyapps.io/dist_calc/`):



Finding Probabilities - t table

Locate the T value on the appropriate df row, obtain the probability from the corresponding column heading (one or two tail).

one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
⋮	⋮	⋮	⋮	⋮	
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
⋮	⋮	⋮	⋮	⋮	
400	1.28	1.65	1.97	2.34	2.59
500	1.28	1.65	1.96	2.33	2.59
∞	1.28	1.64	1.96	2.33	2.58

Finding probabilities - upper tail

Using the table below find:

$$P(T_{df=19} > 1.16)$$

	one tail	0.100	0.050	0.025	0.010	0.005
	two tails	0.200	0.100	0.050	0.020	0.010
df	17	1.33	1.74	2.11	2.57	2.90
	18	1.33	1.73	2.10	2.55	2.88
	19	1.33	1.73	2.09	2.54	2.86
	20	1.33	1.72	2.09	2.53	2.85

Finding probabilities - upper tail

Using the table below find:

$$P(T_{df=19} < -2)$$

	one tail	0.100	0.050	0.025	0.010	0.005
	two tails	0.200	0.100	0.050	0.020	0.010
df	17	1.33	1.74	2.11	2.57	2.90
	18	1.33	1.73	2.10	2.55	2.88
	19	1.33	1.73	2.09	2.54	2.86
	20	1.33	1.72	2.09	2.53	2.85

Finding probabilities - two tails

Using the table below find:

$$P(T_{df=19} < -1.5 \text{ or } T_{df=19} > 1.5)$$

	one tail	0.100	0.050	0.025	0.010	0.005
	two tails	0.200	0.100	0.050	0.020	0.010
df	17	1.33	1.74	2.11	2.57	2.90
	18	1.33	1.73	2.10	2.55	2.88
	19	1.33	1.73	2.09	2.54	2.86
	20	1.33	1.72	2.09	2.53	2.85

CLT vs. t

From the Central Limit Distribution we have,

$$\bar{X} \sim N(\mu, \sigma^2/n)$$
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Since σ is unknown we must use s which results in the following

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{df=n-1}$$

Implications of t distribution for Confidence intervals

Confidence intervals are always of the form

$$\text{point estimate} \pm CV \times SE$$

Implications of t distribution for Confidence intervals

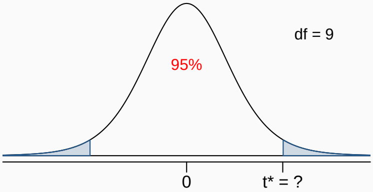
Confidence intervals are always of the form

$$\text{point estimate} \pm CV \times SE$$

If our point estimate is a sample mean and σ is unknown, then our sample mean follows a t distribution (and not a Z distribution), the critical value is then given by $t^\star_{df}$ (as opposed to a $Z^\star$) and the SE is $s/\sqrt{n}$ (and not $\sigma/\sqrt{n}$).

$$\bar{X} \pm t^\star_{df} \times \frac{s}{\sqrt{n}}$$

Finding the critical t ($t^\star$)



$n = 10$, $df = 10 - 1 = 9$

$t^\star$ is at the intersection of row $df = 9$ and two tails column 0.05.

one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17

Constructing a CI

We would like to calculate a 95% confidence interval for the average rental price of an apartment in Durham. We sample craigslist and find

$$\begin{aligned}\text{Rent} = \{&625, 733, 895, 929, 775, 1349, 599, 749, 1020, 799,\\ &705, 665, 1282, 1143, 1209, 500, 1495, 1076, 975, 879\}\end{aligned}$$

Constructing a CI

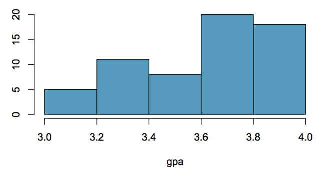
We would like to calculate a 95% confidence interval for the average rental price of an apartment in Durham. We sample craigslist and find

$$\begin{aligned} \text{Rent} = \{&625, 733, 895, 929, 775, 1349, 599, 749, 1020, 799, \\ &705, 665, 1282, 1143, 1209, 500, 1495, 1076, 975, 879\} \end{aligned}$$

$$\bar{X} = 920.1 \qquad s = 271 \qquad n = 20 \qquad SE = s/\sqrt{n} = 60.6$$

Confidence Intervals as Inference

In 2001 the average GPA of students at Duke University was 3.37. Last semester 63 introductory statistics students reported their GPA on an in class survey. The mean was 3.58, and the standard deviation 0.53. A histogram of the data is shown below.



Assuming that this sample is random and representative of all Duke students, do these data provide convincing evidence that the average GPA of Duke students has changed over the last decade and a half?

Example - Z vs t

Your friend has collected some data as part of a summer REU - they collected tadpoles from a local different stream and measured their lengths. From the stream they were able collect 50 tadpoles which had an average length 2.3 cm and a standard deviation of 0.2 cm.

They argue that since it is well known that the distribution of tadpole lengths is normal they should be able to use the Z distribution when constructing their confidence intervals for the average lengths. Are they correct? If not, how serious a mistake are they making? (Construct the CIs both ways for both steams and compare)

Recap: Inference using CIs for sample means

If σ is unknown, then $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ has a t distribution with $df = n - 1$ when the CLT holds.

Recap: Inference using CIs for sample means

If σ is unknown, then $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ has a t distribution with $df = n - 1$ when the CLT holds.

Conditions (same as CLT):

- independence of observations (often verified by a random sample, and if sampling without replacement, $n < 10\%$ of population)
- sample size is large or population not overly skewed or heavy/light tailed

Recap: Inference using CIs for sample means

If σ is unknown, then $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ has a t distribution with $df = n - 1$ when the CLT holds.

Conditions (same as CLT):

- independence of observations (often verified by a random sample, and if sampling without replacement, $n < 10\%$ of population)
- sample size is large or population not overly skewed or heavy/light tailed

Confidence interval:

$$\bar{X} \pm t^{\star}_{df} \frac{s}{\sqrt{n}}, \text{ where } df = n - 1$$

Hypothesis Tests for one mean

Null Value Hypothesis Testing framework

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

Null Value Hypothesis Testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We develop an *alternative hypothesis (H_A)* that represents our research question (what we're testing for). It should be mutually exclusive to H_0.

Null Value Hypothesis Testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We develop an *alternative hypothesis (H_A)* that represents our research question (what we're testing for). It should be mutually exclusive to H_0.
- We conduct the hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods.

Null Value Hypothesis Testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We develop an *alternative hypothesis (H_A)* that represents our research question (what we're testing for). It should be mutually exclusive to H_0.
- We conduct the hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods.
- We examine how likely our data (or something more extreme) is under this assumption, and use that as evidence against the null hypothesis (and hence for the alternative).

Example - Grade inflation?

In 2001 the average GPA of students at Duke University was 3.37. Last semester Duke students in a Stats class were surveyed and ask for their current GPA. This survey had 63 respondents and yielded an average GPA of 3.56 with a standard deviation of 0.31.

Assuming that this sample is random and representative of all Duke students, do these data provide convincing evidence that the average GPA of Duke students has *changed* over the last decade?

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.
- There may be two explanations why our sample mean is higher than the average GPA from 2001.
  - The true population mean has changed.
  - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

- There may be two explanations why our sample mean is higher than the average GPA from 2001.
  - The true population mean has changed.
  - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

- We start with the assumption that nothing has changed.

$$H_0 : \mu = 3.37$$

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

- There may be two explanations why our sample mean is higher than the average GPA from 2001.
  - The true population mean has changed.
  - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

- We start with the assumption that nothing has changed.

$$H_0 : \mu = 3.37$$

- We test the claim that average GPA has changed.

$$H_A : \mu \neq 3.37$$

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject* H_0.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject H_0*.
- *Large* p-value ($> \alpha$) we claim it is likely to observe these data if the null hypothesis were true, and therefore *do not reject H_0*.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject H_0*.
- *Large* p-value ($> \alpha$) we claim it is likely to observe these data if the null hypothesis were true, and therefore *do not reject H_0*.
- We never accept H_0 since we're not in the business of trying to prove it. We just want to know if the data provide *convincing* evidence against H_0.

What is a p-value

What is a p-value:

- The probability of the observed data (sample statistic) or something more extreme in favor of the null hypothesis given the null hypothesis is true.
- Indirect evidence against H_0.

What a p-value *isn't*:

- A p-value is not the probably H_0 is true
- A p-value is not the probably H_A is false

Conditions for inference - GPA

Back to the GPA example, in order to perform inference on these data we need to use the CLT, and therefore we need to check the conditions:

Conditions for inference - GPA

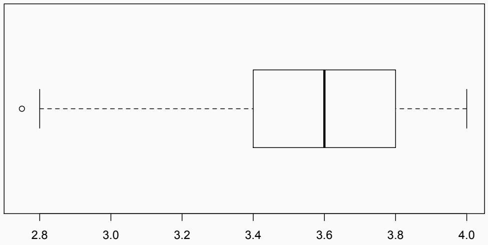
Back to the GPA example, in order to perform inference on these data we need to use the CLT, and therefore we need to check the conditions:

1. *Independence:*
    - We have already assumed this sample is random.
    - Assume sampling without replacement, but $63 < 10\%$ of all current Duke students.

    $\Rightarrow$ it appears reasonable to assume that GPA of one student in this sample is independent of another.

Conditions for inference - GPA

2. *Sample size / skew:* The distribution appears to be slightly left skewed (but not extremely) and $n = 63$ so we will assume that the sampling distribution of the sample means should be nearly normal by the CLT.



Calculating the p-value

p-value - probability of observing data at least as favorable to H_A as our current data set, if in fact H_0 is true (the true population mean $\mu = 3.37$).

In this case because we are not making any claims about GPAs going up or down, we need to consider GPA changes in both directions. E.g. a sample average GPA of 3.18 is just as much in favor of H_A as a sample average GPA of 3.56.

Drawing a Conclusion / Inference

$$\text{p-value} = 4.2 \times 10^{-6}$$

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.

Drawing a Conclusion / Inference

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.
- Since the p-value is *small* (lower than 5%) we *reject H_0*.

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.
- Since the p-value is *small* (lower than 5%) we *reject H_0*.
- Claim - the data provide convincing evidence that Duke students' average GPA has changed since 2001. E.g. the difference between the null value of a 3.37 GPA and observed sample mean of 3.56 GPA is *not due to chance / sampling variability*.

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the 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 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the 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 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the 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 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Are the conditions for inference met?

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the 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 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Are the conditions for inference met?

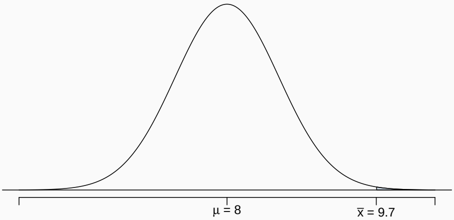
Yes - Independence ✓, Nearly Normal ✓

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 was true (the true population mean is 8).

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 was true (the true population mean is 8).



College Applications - Making a decision

$$p - value < 0.005$$

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely that a sample mean of 9.7 or more schools is likely to happen simply by chance.

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely 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.

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely 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 on average to more than 8 schools.

What about a confidence interval?

We can also assess this claim using a confidence interval.

What about a confidence interval?

We can also assess this claim using a confidence interval.

$$\bar{X} = 9.7 \qquad s^2 = 7^2 \qquad n = 206$$

What about a confidence interval?

We can also assess this claim using a confidence interval.

$$\bar{X} = 9.7 \qquad s^2 = 7^2 \qquad n = 206$$

We construct a 95% confidence interval using
$t^*_{df=205} \approx t^*_{df=200} = 1.97$,

Recap: Null Value Hypothesis Testing

Regardless of the sample statistic of interest, all null value hypothesis testing takes exactly the same form.

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

Recap: Null Value Hypothesis Testing - Sample Means

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

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Check assumptions and conditions
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq \textit{null value}$
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad\qquad T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

4. Make a decision, and interpret it in context of the research question
   - If p-value $< \alpha$, reject H_0
   - If p-value $> \alpha$, do not reject H_0