EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- Recap on Formulating Hypotheses
  - And draw the picture!
- Level of Significance
- Choosing a Statistical Test
- Find the Critical Values (“Cut Offs”)
- Calculate the Test Statistic
- State a Conclusion
- Hypothesis Testing in R

Recap Hypothesis Formulation

- Hypotheses always comes in pairs
  - The Null and the Alternative
- Hypotheses are statements about the population, so you only use μ in the notation
- Hypotheses may be formulated to test either:
  - Two Tailed Test
    - No direction, only a difference of some sort
  - One Tailed Test
    - There is an assertion of direction, like greater or less than

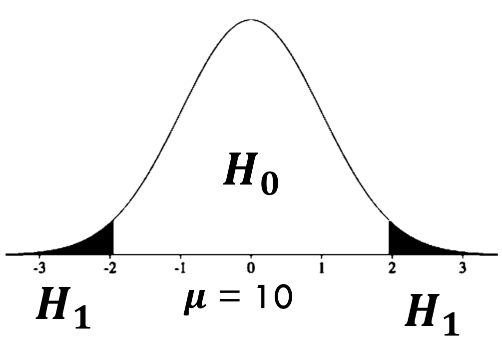
Recap Hypothesis Formulation

Two Tailed

Different Than

- H_0: There is no difference in this school's performance on the STAR test compared to national average (10).
- H_1: There IS a difference in this school's performance on the STAR test compared to national average (10).

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

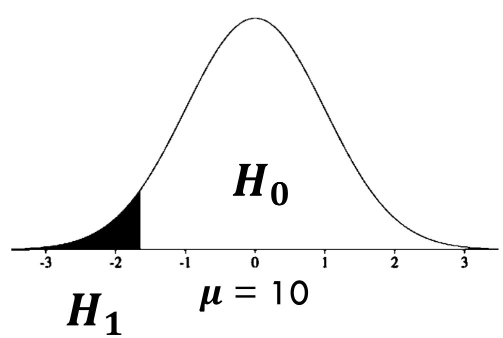


One Tailed (Left)

Less Than

- H_0: This school's performance on the STAR test is greater than or equal to the national average (10).
- H_1: This school's performance on the STAR test is less the national average (10).

$$H_0: \mu \geq 10$$
$$H_1: \mu < 10$$

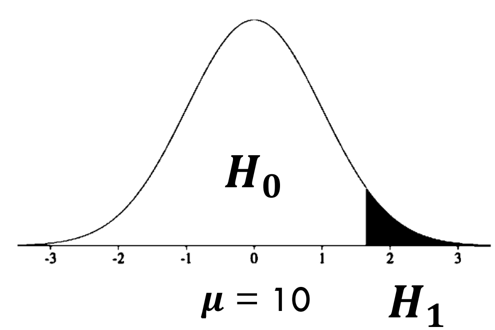


One Tailed (Right)

Greater Than

- H_0: This school's performance on the STAR test is less than or equal to the national average (10).
- H_1: This school's performance on the STAR test is greater the national average (10).

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$



Significance Test: Step by Step

A significance test about a hypothesis has six steps.

1. State the null and alternative hypotheses
   - And draw the picture!
2. Select a level of significance (α)
3. Choose a statistical test
4. Find the critical values
5. Compute the test statistic and the p-value
6. Formulate a decision (reject or fail to reject the null hypothesis)

Selecting a Level of Significance α

Step 2: Select a Level of Significance α

- Remember our 95% (or 90%, 99%) confidence level for our confidence intervals? What did that mean, again?
  - 95% or .95 = the proportion of the time that we capture the "truth" (the population parameter)
    - 95 out of 100 times we will capture it
    - 19 out of 20 times we will capture it
  - 5% or .05 = the proportion of the time that we are willing to FAIL to capture the "truth"
    - 5 out of 100 times we will FAIL to capture the truth
    - 1 out of 20 times we will FAIL to capture the truth

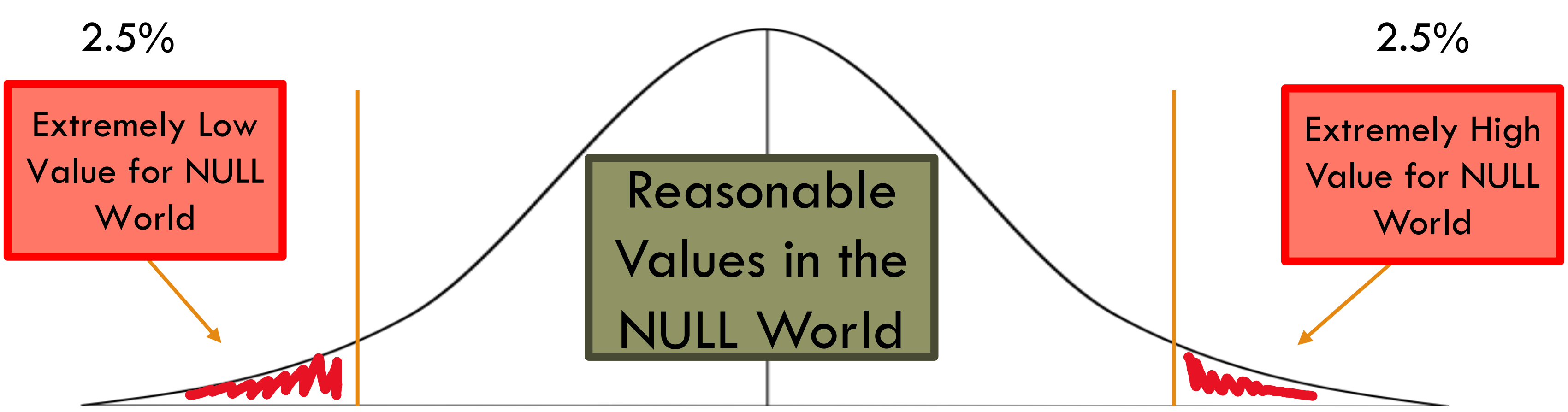
The compliment to "confidence level" is the "level of significance."

Step 2: Select a Level of Significance α

- "How much of a risk are you willing to take about being wrong if you reject the null hypothesis?"
  - With a level of significant at .05, you risk being wrong 1 in 20 times or 5% of the time.
- Common Levels of Significance:
  - 5% (.05) and 1% (.01)
    - More strict levels of significance (.01) for high stakes outcomes
      - Ex. Medical testing, infrastructure safety
    - Sometimes you see 10% (.10), but it's usually not accepted
  - Significance Levels are denoted at α (alpha)
    - Ex. $\alpha = .05$
      - Same as 95% confidence level (1- α = Confidence Level)
      - Same as .05 significance level

Step 2: Select a Level of Significance α

- In hypothesis testing, we always assume the NULL hypothesis is true, then we test how likely it is we would see this sample mean $\bar{x}$ if the null is true.
- Alpha level (α): Under the sampling distribution of H_0, we decide what proportion (ex. 5%) of sample means are too *extreme*, "unlikely enough."

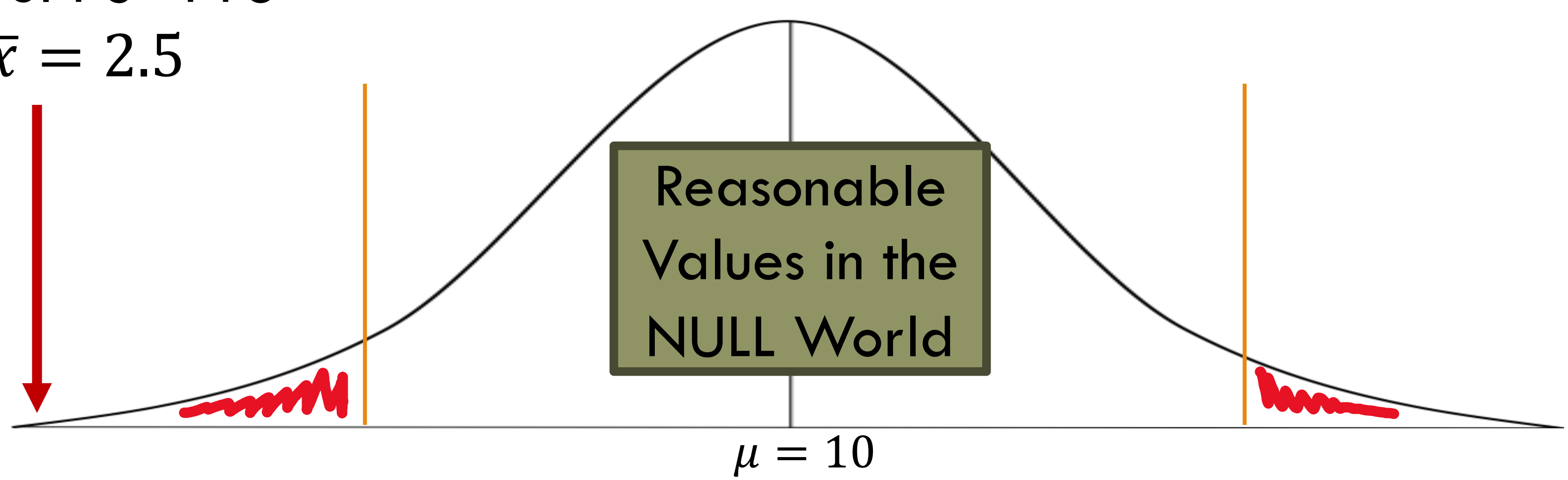


Step 2: Select a Level of Significance α

- When you select a level of significance, and by extension your confidence level, you are defining the boundaries for what will be considered "unlikely enough" or "extreme enough" for us to reject the assumption we live in the NULL world.
  - Imagine one school had an average STAR test score that was way down here… Do you think everything is fine and it's just a fluke, they just happened to have such a low average, or would you reject the idea that everything is ok (the null) and maybe want to investigate further?

School PS -118

$\bar{x} = 2.5$



How Unlikely is "Unlikely Enough"?

You want to know whether getting tutoring in statistics will have an effect on your class grade (in GPA grading). The average grade in a statistics course is 3.0.

$$H_0\colon \mu = 3.0$$
$$H_1\colon \mu \neq 3.0$$

H_0: The average GPA for tutored students is 3.0.

H_1: The average GPA for tutored students is not 3.0.

(Two-Tailed because no assertion of direction was made.)

Remember, even if the tutoring has no effect, the mean for the tutored group might not be exactly 3.0… What is 3.0 enough for us to conclude tutors has no effect?

What if your sample had an average GPA of 3.4? Do we reject H_0?

What if your sample had an average GPA of 2.9? Do we reject H_0?

What are the chances?

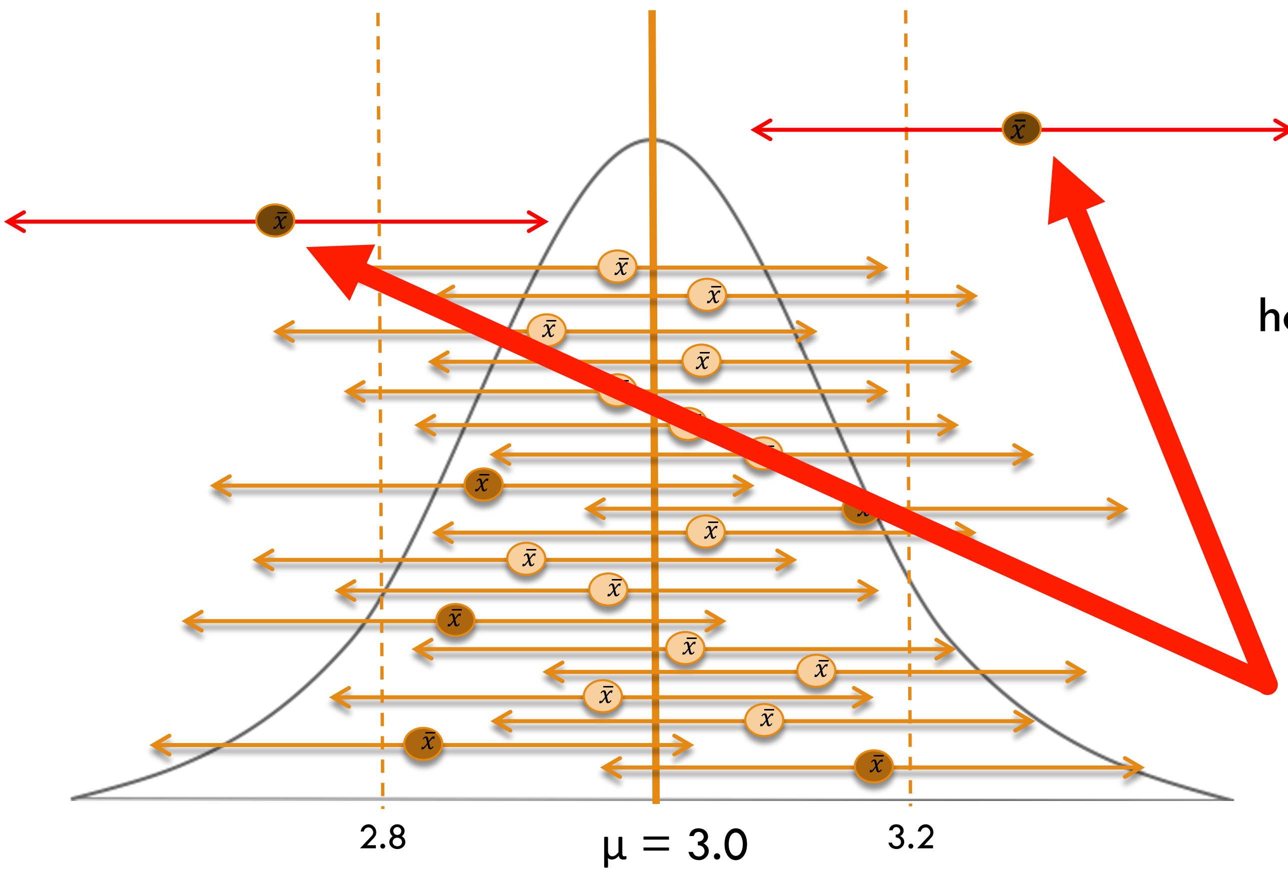
- Let's say the mean for the sample you collected was $\bar{x} = 3.4$ How likely is it that this observed sample value is….
  - An extreme value of the NULL distribution?
    - We would expect to happen by random chance ~5% of the time

OR

  - A value from a truly different ALTERNATIVE distribution?
    - A reasonable mean value for a different alternative distribution of people
      - Ex. A different group of people
        - Tutored vs Non-Tutored

What are the chances?

- Let's assume the NULL is true.
  - H_0: The average GPA for tutored students is 3.0.



If the null were true, we would expect our sample means to be within a certain range (ex. 2.8-3.2*)

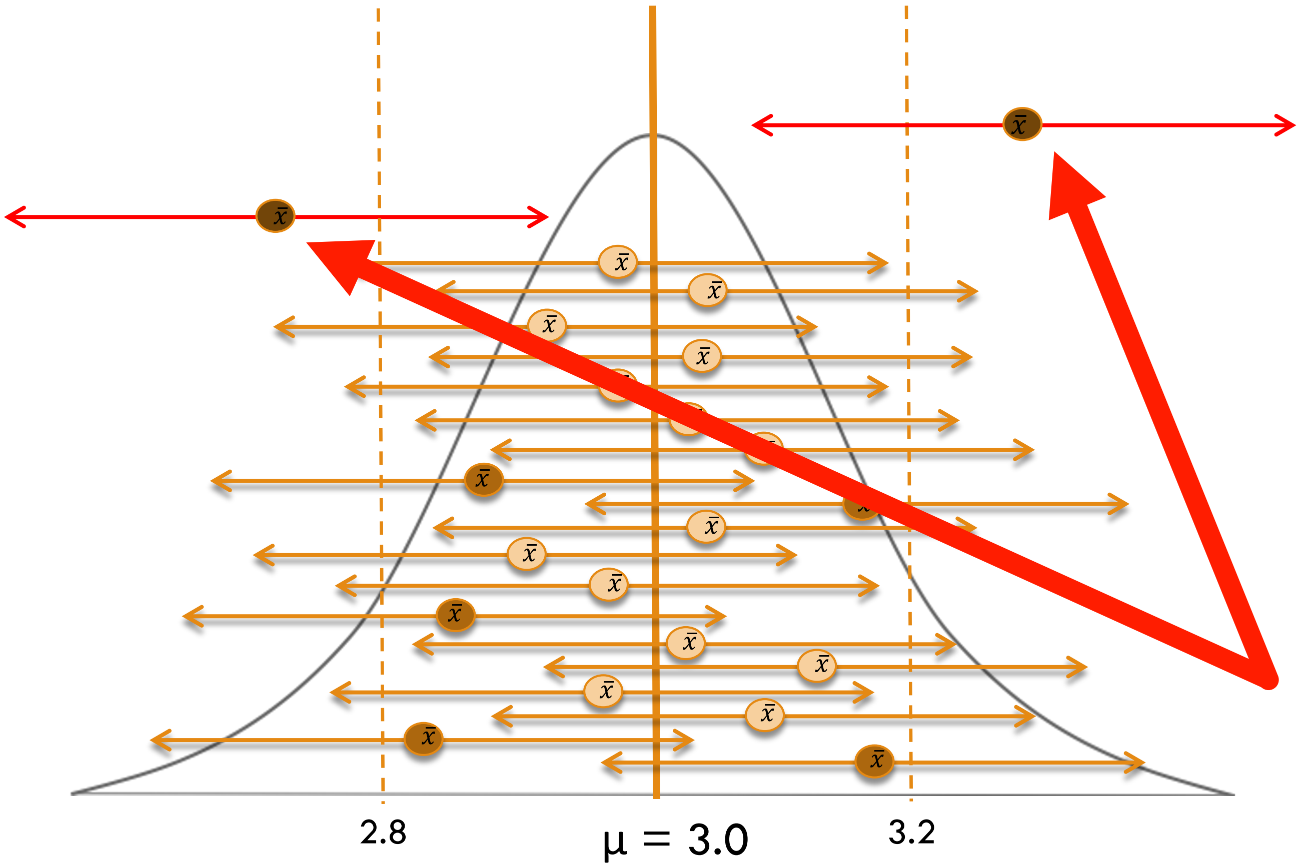
If this were the case, how likely is it that I get a sample that is less than 2.8 or greater than 3.2?

What are the chances we get one of these?

*bounds are made up

What are the chances?

- Let's assume the NULL is true.
  - H_0: The average GPA for tutored students is 3.0.

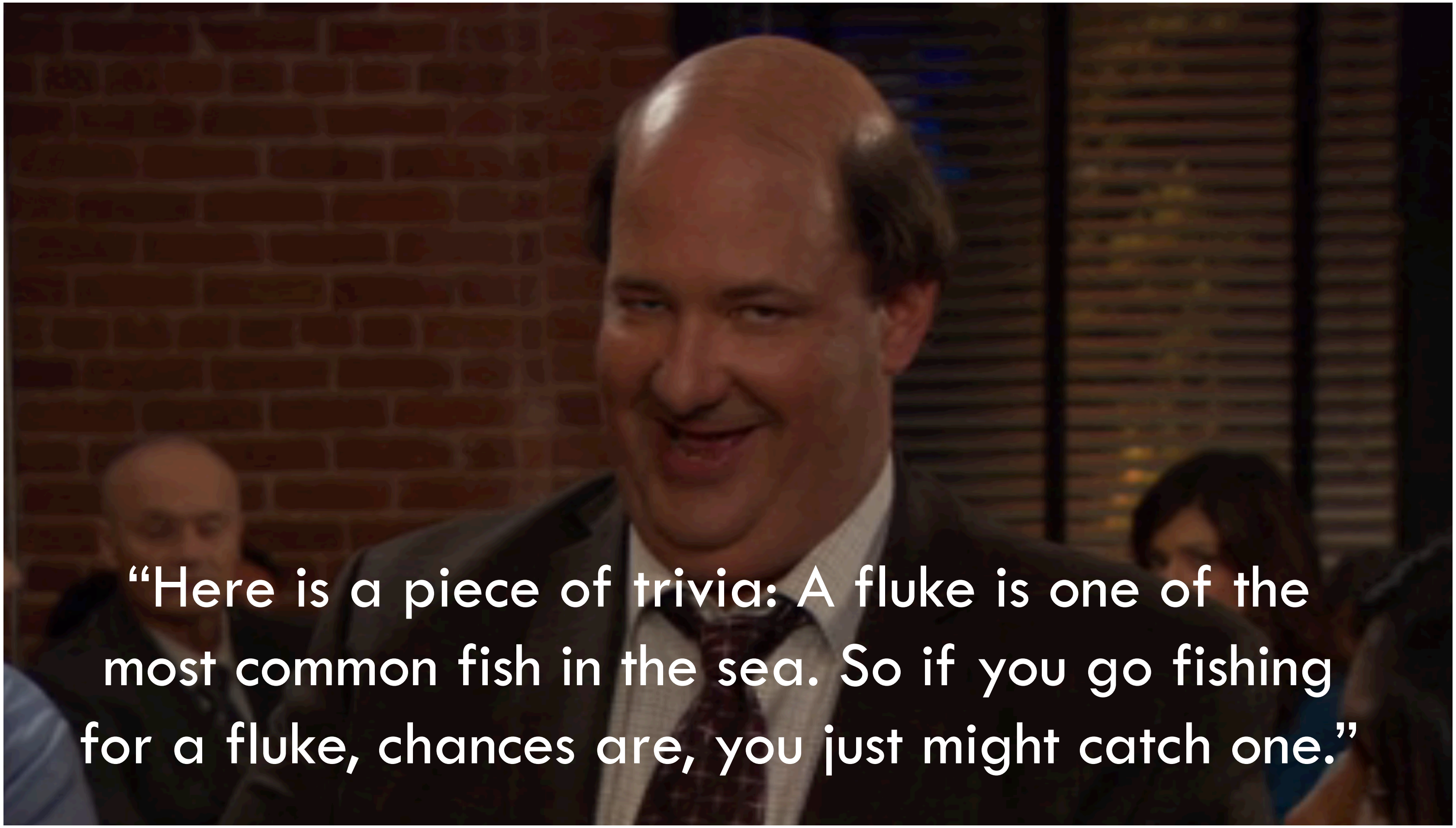


What are the chances we get one of these?

If the Null were TRUE and we set our level of significance to .05, the chances of getting a sample mean that is lower than 2.8 or higher than 3.2 is **5%**. This would be a fluke under the null hypothesis.

*bounds are made up

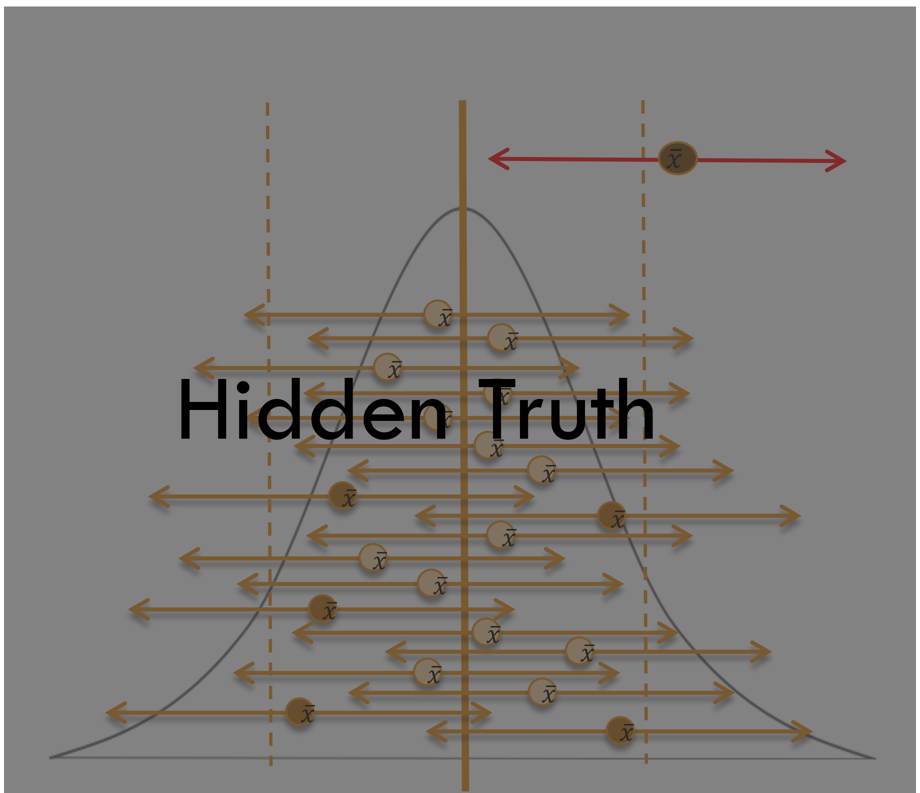
Wisdom from Kevin Malone



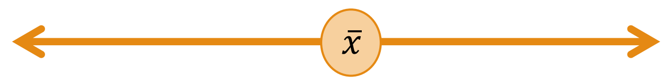
Trivia, The Office: Season 8, Episode 11

What are the chances?

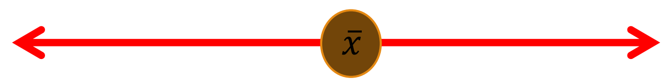
- Unfortunately, in the real world, we only take one sample…
  - How do we know if our sample is one of the 95% that is “correct” or one of those 5% of extremes that are “incorrect”?



We only see this…



And have no idea if it is really one of these…

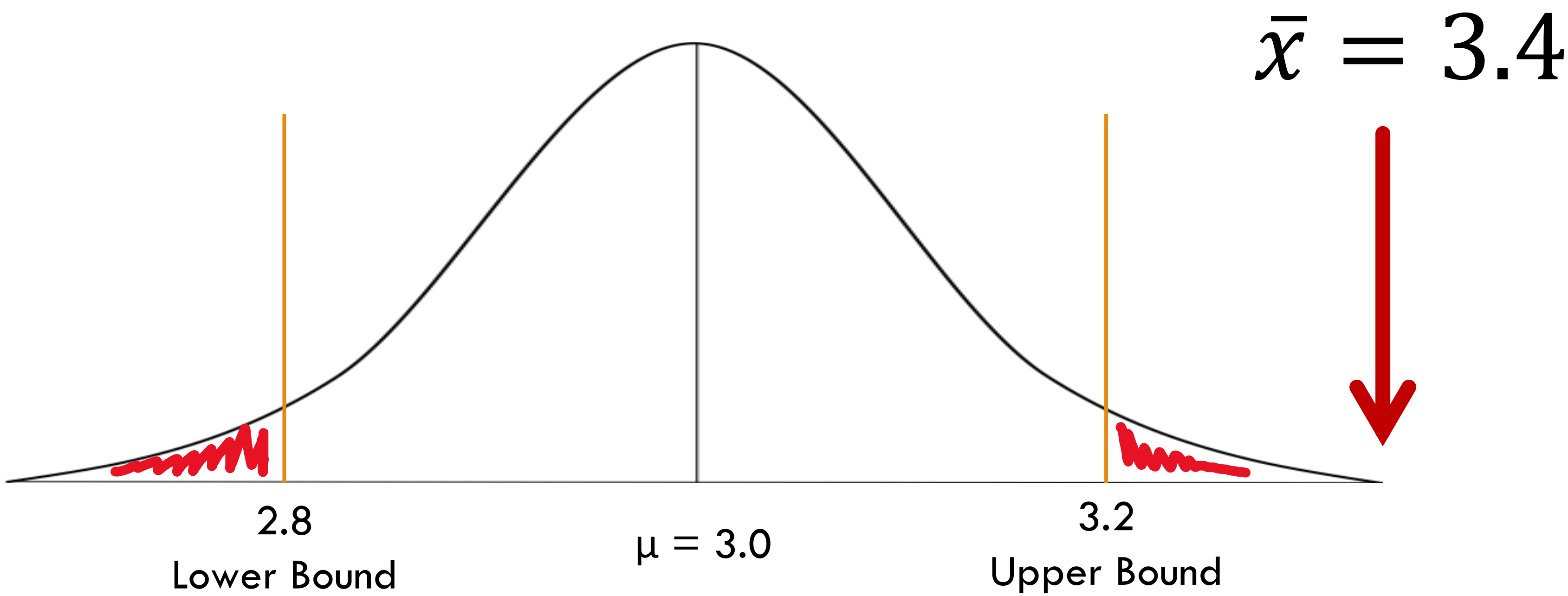


Sadly, for the most part,
we can never know for sure…

This is why we have to acknowledge our p-value (the probability of randomly getting one of those extreme cases if the NULL were true)

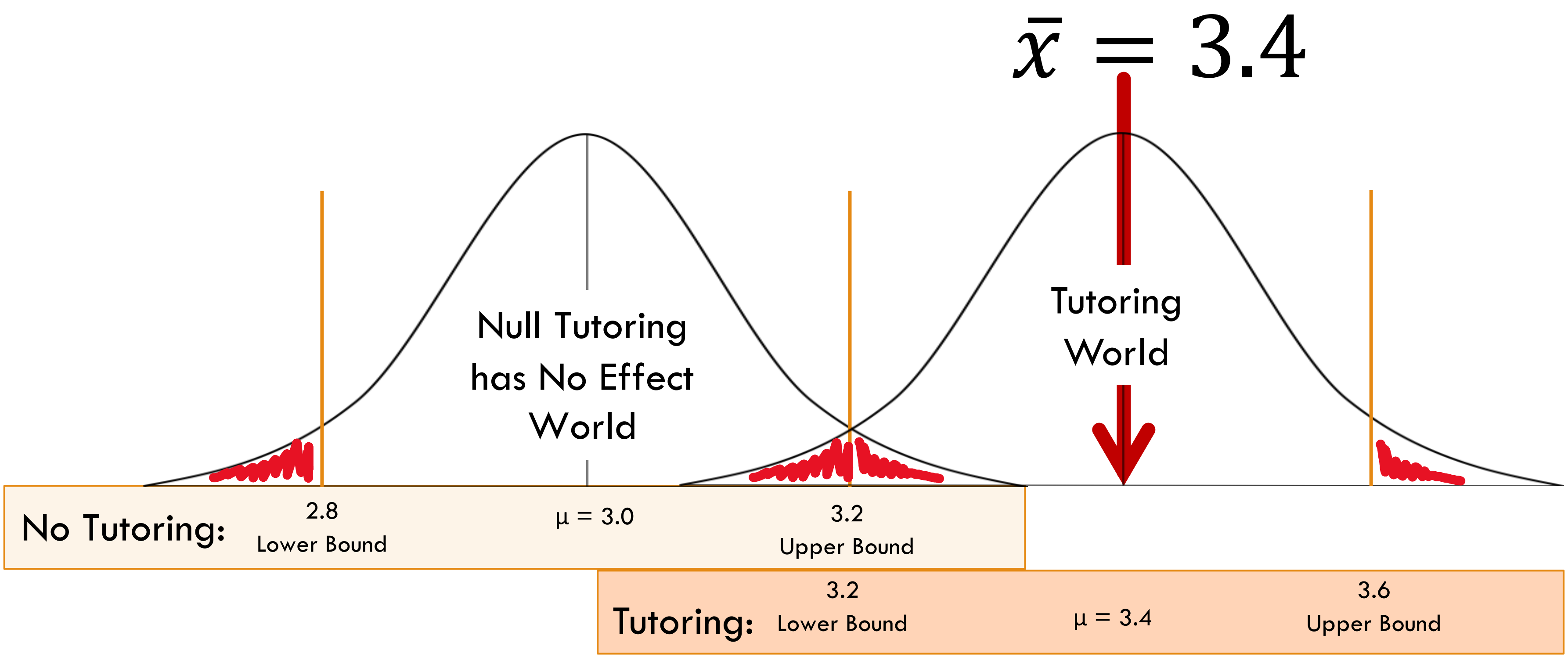
Maybe It's Extreme Null Value

- Maybe I just took a sample of people who would have done well without tutoring anyway, maybe I sampled some overachievers
- It could just be one of those extreme cases we expect to get 1 in 20 times (if $\alpha = .05$)
  - 5% chance



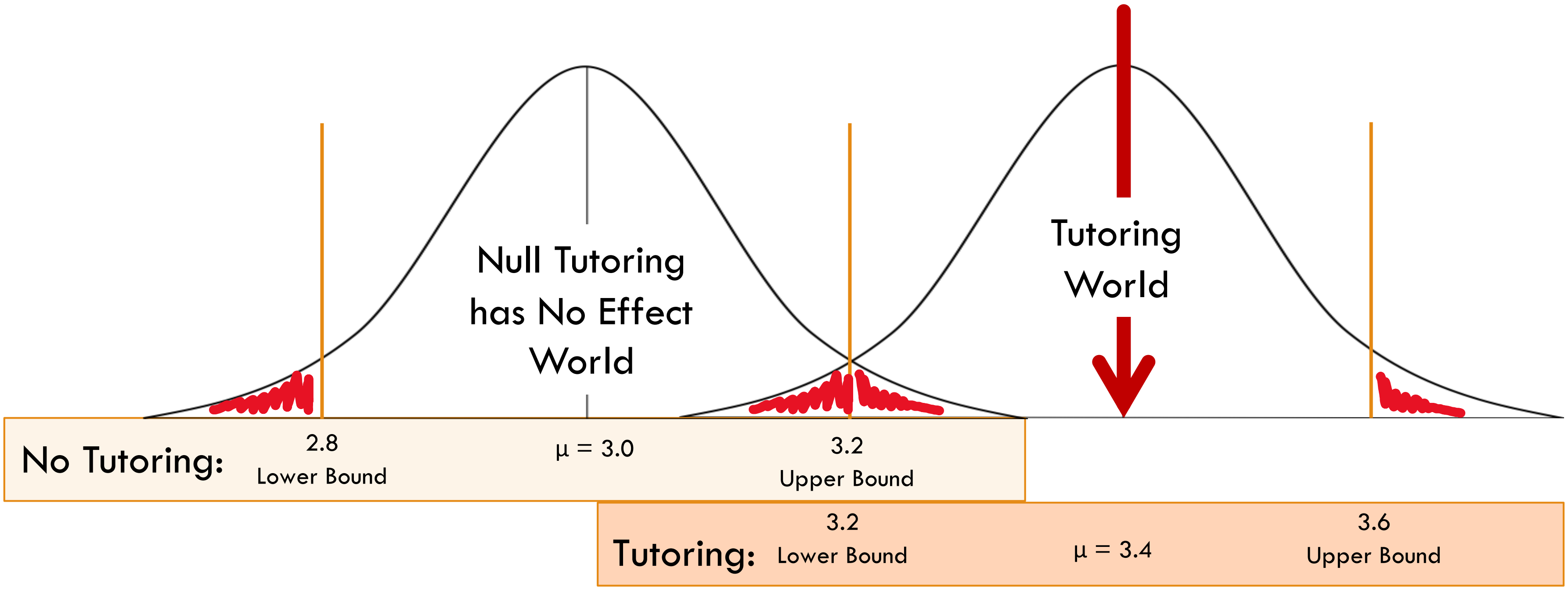
Or Maybe It's From Somewhere Else…

- Maybe there is a different population… The population of those that got tutoring, and their average is higher than non-tutored.
- What is more likely, that you got one of those rare 5% extremely high means from the NULL (tutoring has no effect) world or that you got a reasonable mean for an ALTERNATIVE world?



The Essence of Hypothesis Testing

- There a many different types of hypothesis testing, but this example highlights the essence of trying to determine if a certain result is a fluke or some sort of "real" difference all based on how likely something is.



Choosing a Statistical Test

Step 3: Choose a Statistical Test

- There are many different types of statistical tests out there. Which one you choose depends on:
  - 1) Your research question
  - 2) The data you have
    - Nominal data?
    - Quantitative data?
- Ideally, if you were a funded researcher formulating a study, you'd ask your question then determine how to collect the data you need
  - But sometimes you are stuck with a data set (maybe you didn't collect it) and will have to work with it

Step 3: Choose a Statistical Test

- "What kind of test should you run given the data?"
  - Here are just a few… We will be covering these statistical tests for the rest of the class.

Statistic	Outcome[1] Variable	Variable 2	Language	Notes
One-sample *t*-test	Interval/ratio	-	Comparison of one group's mean with one value	Only 1 group, compare mean with a known reference mean
Independent Samples *t*-test	Interval/ratio	Nominal (2 groups)	Group comparison	Two groups must be independent
Related samples *t*-test	Interval/ratio	Nominal (2 groups)	Group comparison	Two groups are related – either pre & post or matched pairs
ANOVA	Interval/ratio	Nominal (≥ 2 groups)	Group comparison	Independent groups
Pearson's Correlation	Interval/Ratio	Interval/Ratio	Relationship	
Regression	Interval/Ratio	Any	Prediction	Simple regression
Chi-Squared χ^2	Nominal	Nominal	Group comparison or relationship	

Step 3: Choose a Statistical Test

- From the statistical test you choose, you will calculate a "test statistic." This is the thing you are going to compare to a certain cut off to then make a decision about whether to reject or fail to reject the null hypothesis.
- A test statistic is a standardized number that you compute from a sample (i.e. it will be in a standardized language like z or t)

$$z_{stat} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

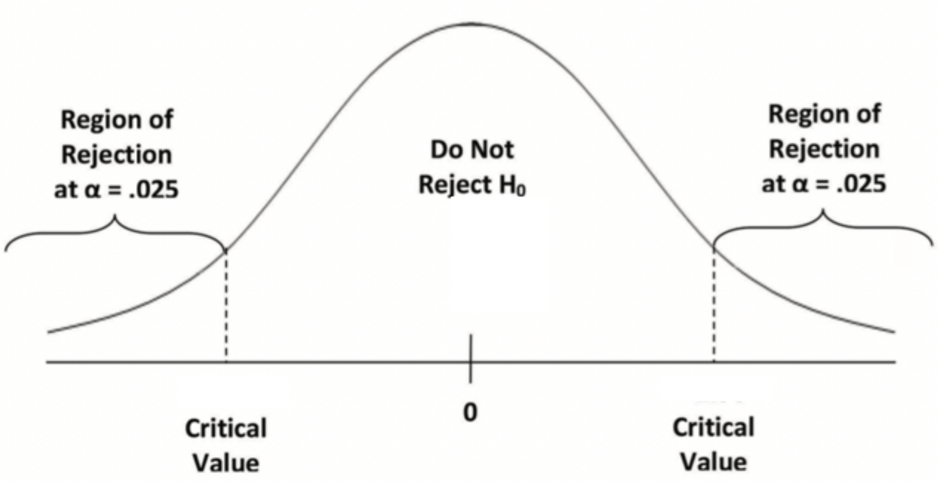
$$t_{stat} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Finding the Critical Value(s)

Step 4: Find the Critical Value(s)

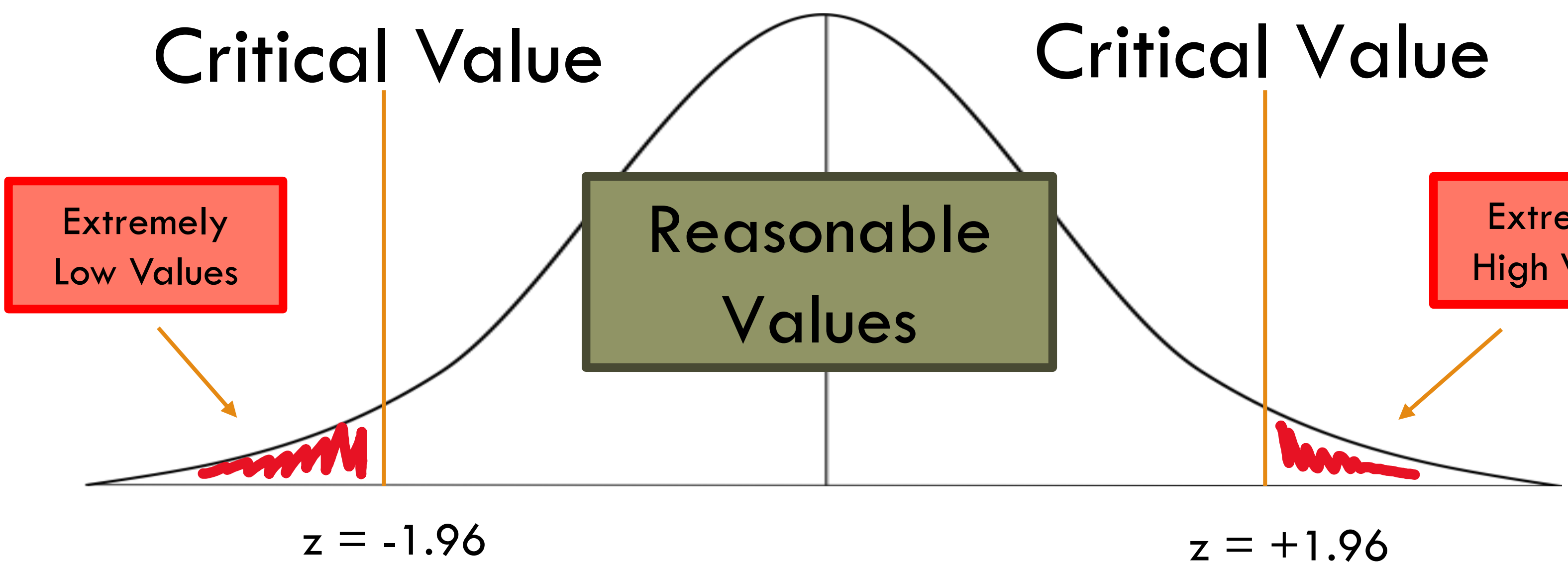
- The critical value is your "cut off" value that you will use to determine if a test statistic is "extreme enough" or "unlikely enough" to reject the null.
- Critical values come from standardized table (depending on what data you have, z for known σ, t for unknown σ)

Two-Tailed Test at a .05 Level of Significance



Step 4: Find the Critical Value(s)

- The critical value tells you the value associated with a certain level of significance (ex. .05)



Critical t-values depend on the degrees of freedom, ex. (n-1)

Standardized Cut-Offs Example

Job Promotion Example from Previous PPT

- A random sample of 25 college graduates revealed that they worked an average of 6 years at a job before being promoted, with a standard deviation of 1.3 years. Compute and interpret a 99% confidence interval for the mean number of years worked at a job before being promoted.

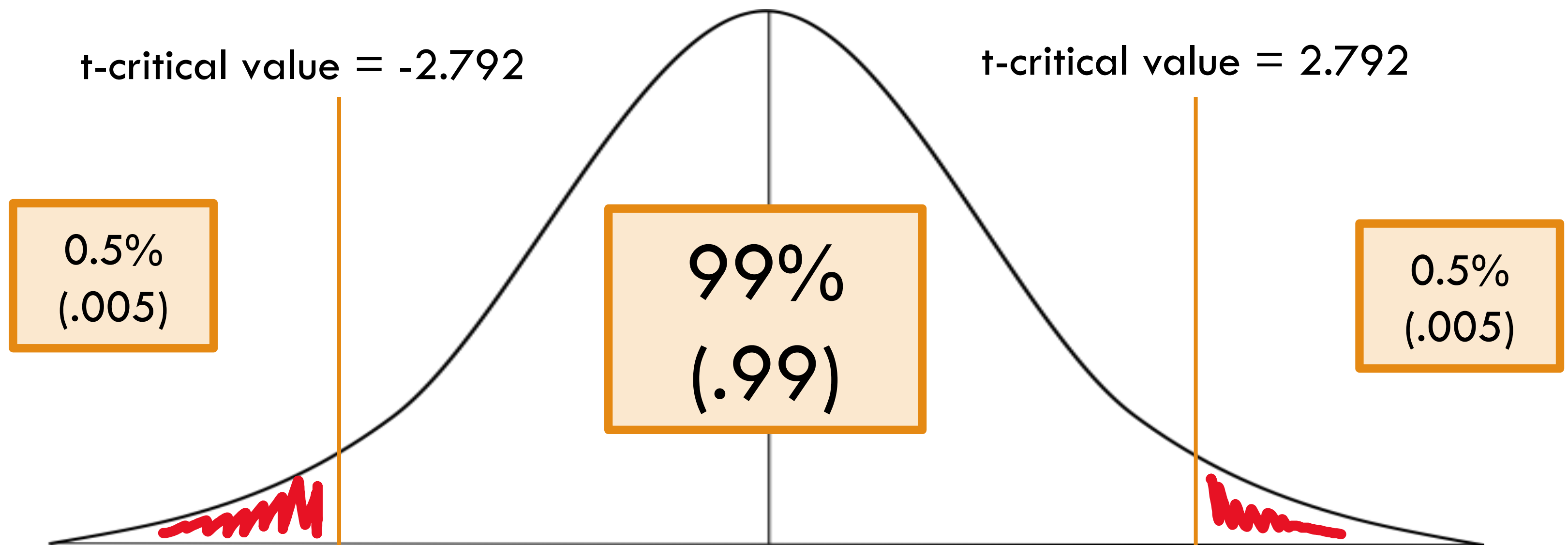
$$\bar{x} = 6, \quad s = 1.3$$
$$n = 25 \quad \rightarrow \quad df = 24$$
$$Level\ of\ Significance = 99\%\ or\ .01$$

How high/low would your t-statistic (or observed mean) have to be to reject the null hypothesis?

Standardized Cut-Offs Example

Job Promotion Example from Previous PPT

How high or low would your t-statistic have to be to reject the null hypothesis?



Standard Scale: t-critical value = -2.792 (Lower Bound) | t = 0 (Standardized Mean) | t-critical value = +2.792 (Upper Bound)

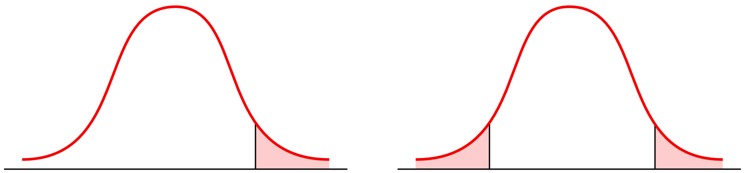
Original Scale: Lower Bound = 5.27 (Original Scale) | $\bar{x}$ = 6 (Original Scale Mean) | Lower Bound = 6.73 (Original Scale)

t-table

We need to determine critical values based on a few things…

- Degrees of Freedom
- Level of Significance
- One or Two Tail Test

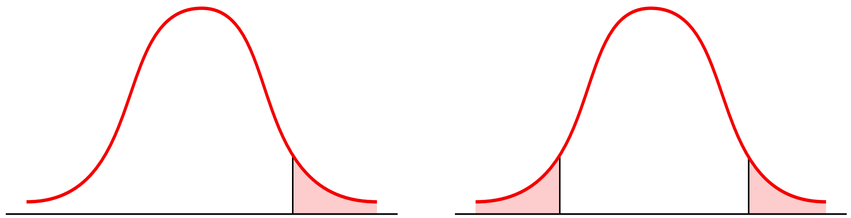
Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

The t-table, Again

- We've looked at the t-table before when we were working on creating confidence intervals. Then we only looked at the "Proportion in Two Tail Combined" row, which distributes the alpha level you want on two sides, ex. if $\alpha = .05$, then the proportion on either side is .025 ($.05/2 = .025$) or 2.5%
- If we were running a two tailed test, we use the same row as before, but if we were running a *one tailed test*, then we look at the row above where the .05 is lumped together on one side.



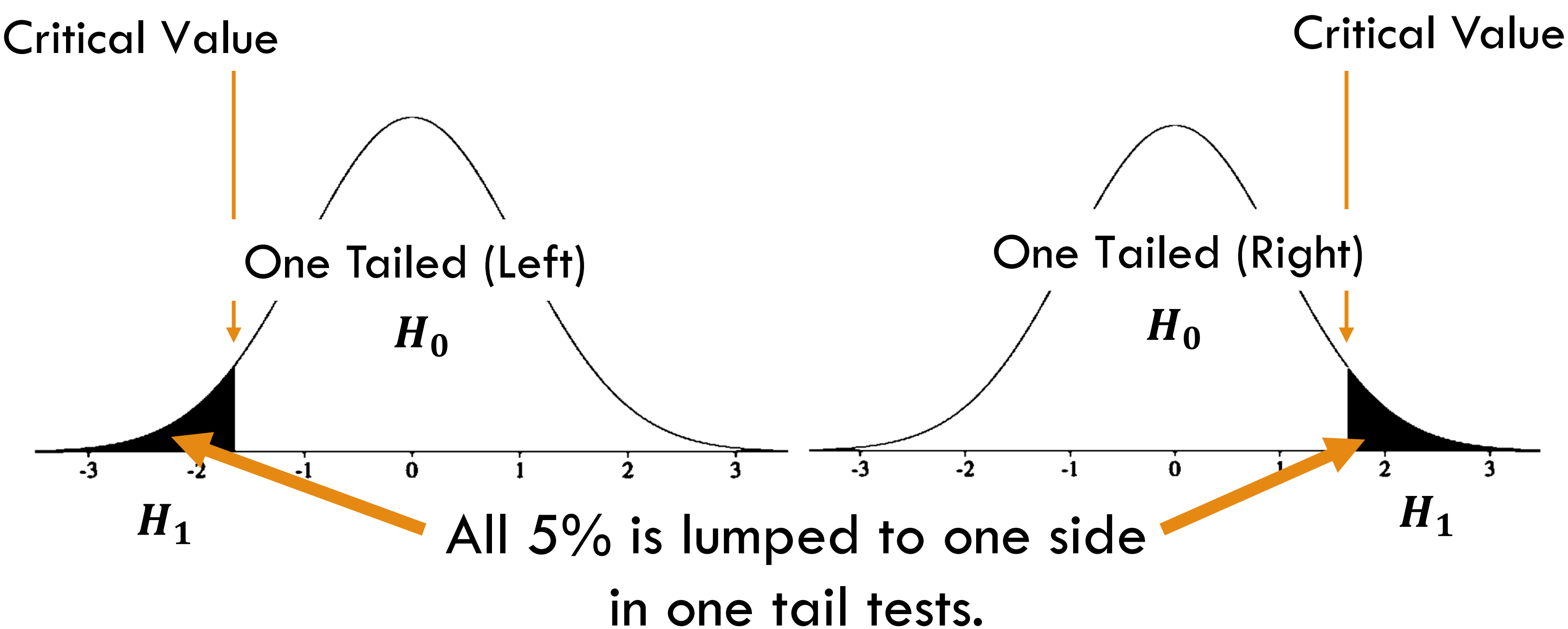
One tail (either right or left)

Two tails combined

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925

Step 4: Find the Critical Value(s)

- Same procedure but if we are running a one sided test, we only need one critical value because we are only looking in one direction (greater than or less than).



Try it.

Find the critical values for each scenario (assume that $\boldsymbol{\sigma}$ **is unknown**, so we will use a t-table).

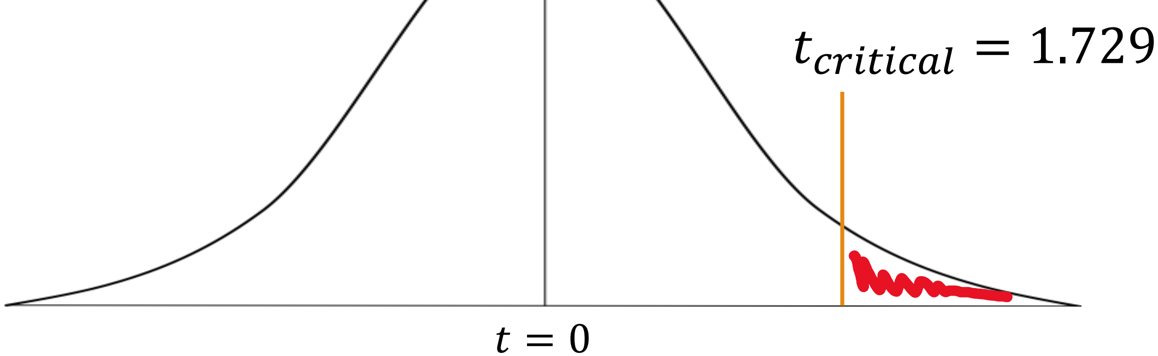
- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer. $(\mathrm{n} = 20, \alpha = .05, \bar{x} = 13, s = 1.3)$
- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire. $(\mathrm{n} = 25, \alpha = .05, \bar{x} = 39, s = 6.9)$
- 3) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week. Test if there is an effect using the sample mean of 4 and sample size of 12 to test this. $(\mathrm{n} = 12, \alpha = .05, \bar{x} = 4, s = 2.5)$

Try it: Emergency Room Wait Time

- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer. $(\text{n} = 20, \alpha = .05)$

$$\boldsymbol{H_0}: \mu \leq 10$$
$$\boldsymbol{H_1}: \mu > 10$$
$$\alpha = .05$$
$$\text{df} = 19$$

One Tailed Test (Right)

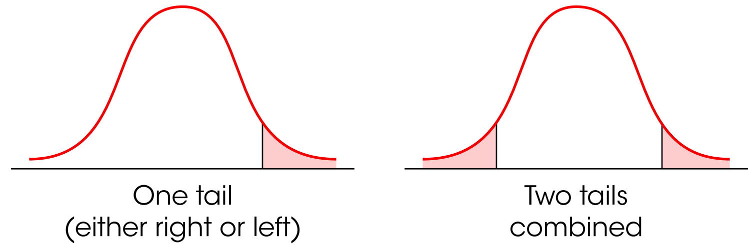


t-table

- One Tailed Test (Right)
  - $\alpha = .05$
  - $\text{df} = 19$

$$t_{critical} = 1.729$$

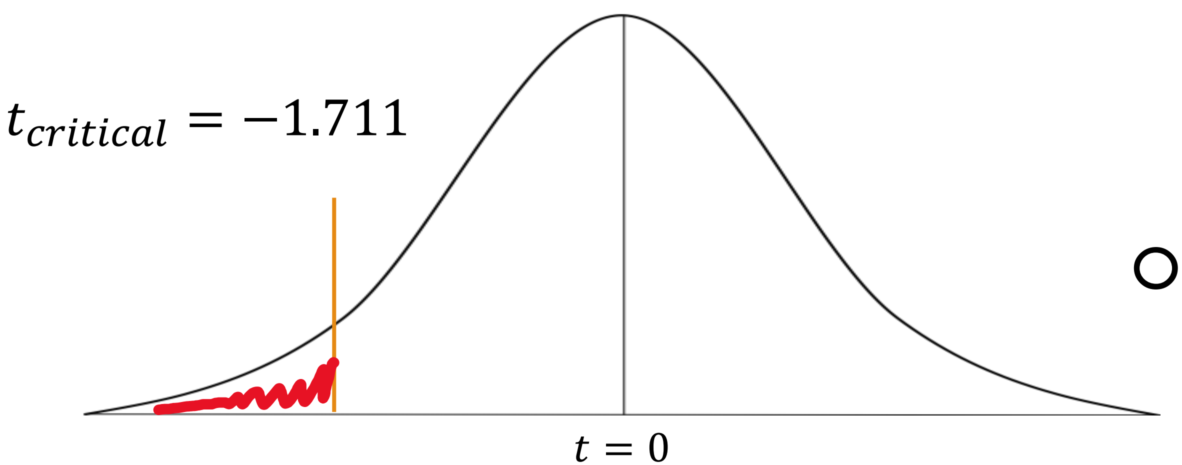
Table entries are values of t corresponding to proportions in one tail or in two tails combined.



	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	$\boxed{1.729}$	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Try it: Football and Self-Esteem

- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire. $(\mathrm{n} = 25, \alpha = .05)$



$\boldsymbol{H_0}: \mu \geq 40$
$\boldsymbol{H_1}: \mu < 40$

$\alpha = .05$

$\mathrm{df} = 24$

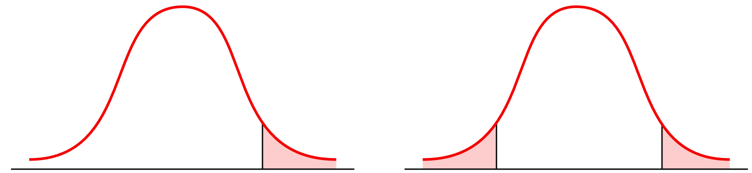
One Tailed Test (Left)

t-table

- One Tailed Test (Left)
  - $\alpha = .05$
  - $\text{df} = 24$

$$t_{critical} = -1.711$$

Table entries are values of t corresponding to proportions in one tail or in two tails combined.



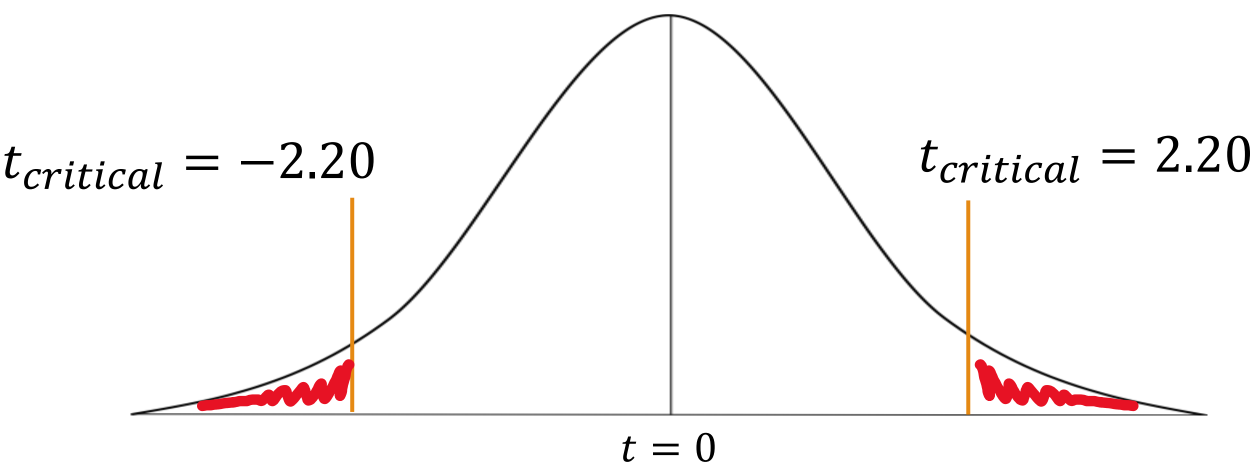
One tail (either right or left)

Two tails combined

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	$\boxed{1.711}$	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Try it: Behavior Modification and Sodas

- 3) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week.
  - $(\mathrm{n} = 12, \alpha = .05)\ t_{crit} = \pm 2.20$



$\boldsymbol{H_0}: \mu = 8$
$\boldsymbol{H_1}: \mu \neq 8$

$\alpha = .05$

$\mathrm{df} = 11$

Two Tailed Test

Compute the Test Statistic

Step 5: Compute the Test Statistic (or Determine P-Value)

- If our test statistic falls past our critical value (falls in the **critical region**), then we reject H_0.
  - It we reject the null, we can then say the p value was less than our set alpha value
    - It states that though we could incorrectly be rejecting the null, the chances of that are $p < .05$

$$z_{stat} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad t_{stat} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

APA Notation: t(df) = t-statistic value, p < or > α level.
Example: $t(21) = 2.46$, $p < .05$

Try it.

Find the test statistic for each scenario. Does it fall in the critical region?

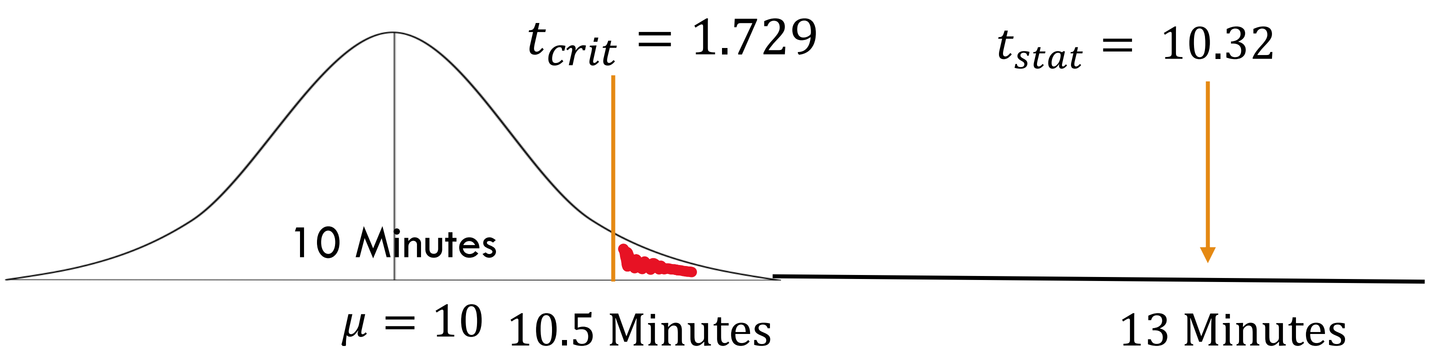
- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer. ($\mathrm{n} = 20, \alpha = .05, \bar{x} = 13, s = 1.3$)
- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire. ($\mathrm{n} = 25, \alpha = .05, \bar{x} = 39, s = 6.9$)
- 3) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week. Test if there is an effect using the sample mean of 4 and sample size of 12 to test this. ($\mathrm{n} = 12, \alpha = .05, \bar{x} = 4, s = 2.5$)

Try it: Emergency Room Wait Time

- 1) An emergency room advertises a wait time of 10 minutes, but you believe that it is longer.
  - $\text{n} = 20, \alpha = .05, \bar{x} = 13, s = 1.3, t_{crit} = 1.729$

Our t_{stat} is further than t_{crit} $(10.32 > 1.729)$, so we reject H_0.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{13 - 10}{1.3/\sqrt{20}} \approx 10.32$$

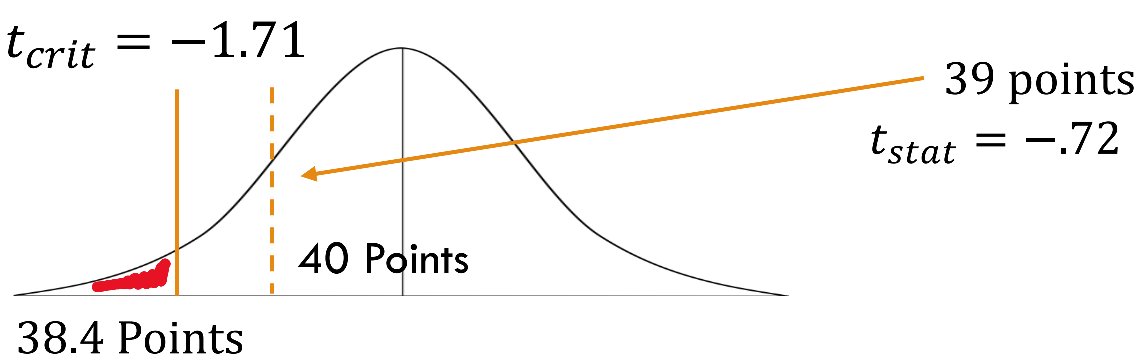


Try it: Football and Self-Esteem

- 2) A psychologist believes that watching 9 or more hours of football a week reduces men's self esteem. The population of men score an average of 40 points on a self-esteem questionnaire.
  - $\mathrm{n} = 25, \alpha = .05, \bar{x} = 39, s = 6.9, t_{crit} = -1.71$

Our t_{stat} is not extreme enough (t_{stat} is not further than t_{crit}), so we fail to reject H_0.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{39 - 40}{6.9/\sqrt{25}} \approx -.72$$



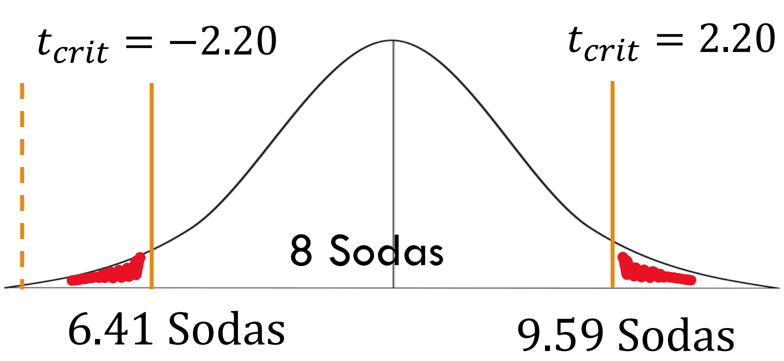
Try it: Behavior Modification and Sodas

- 3) It is suspected behavior modification will have an effect on the average number of sodas a person drinks (8) in a week. Test if there is an effect using the sample mean of 4 and sample size of 12 to test this.
  - $(\mathrm{n} = 12, \alpha = .05, \bar{x} = 4, s = 2.5)$ $t_{crit} = 2.20$

Our is t_{stat} past the t_{crit}, so we reject H_0.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{4 - 8}{2.5/\sqrt{12}} \approx -5.54$$

$t_{stat} = -5.54$
4 Sodas



P Values

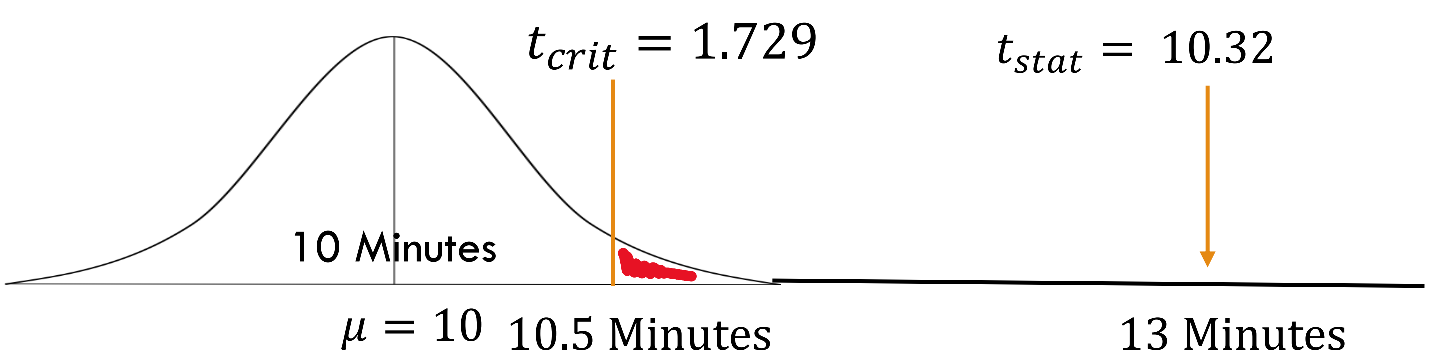
- A **p-value** is the probability of getting the calculated test statistic or a more extreme number <u>if the null hypothesis was true</u>.
    - Remember, we are always working under the assumption that the null is true.
  - The smaller the p-values, the greater the likelihood that the null hypothesis is false
- In practice, statistical software like R tells us the exact p-value, but the decision using critical values to determine if the probability of a fluke is less than ex. .05 will result in the same conclusion (i.e. reject or fail to reject).
  - This is because t_{stat} corresponds to p, and t_{crit} corresponds to α
    - You can find the probability by looking up the t-statistic value in the table and find the probability on top, though usually we just say:

$$p < \alpha \rightarrow \text{Reject } H_0$$
$$p \geq \alpha \rightarrow \text{Fail to reject } H_0$$

State a Conclusion

Step 6: Draw Your Conclusions

- When we "Reject the Null Hypothesis," we is saying that the probability of seeing this results (this certain sample mean $\bar{x}$), if the Null were true, is low enough (ex. less than 5% chance) that we can reject the idea that we live in the null world. We can conclude there is a statistically significant difference.



Statistical Significance

- Something is statistically significant if we are able to reject the null hypothesis.
  - "The results were statistically significant at the .05 level."
- When we "Fail to Reject the Null Hypothesis," the results are not statistically significant. The $\bar{x}$ we observe is close enough to the Null that there is no statistically significant difference. If the p-value is greater than ex. .05, we are saying that the probability of seeing this $\bar{x}$ if the Null is true are higher than 5%.
  - "The results were not statically significant at the .05 level."

Step 6: Draw Your Conclusions

- After all the steps, you will come to a conclusion whether to "reject" or "fail to reject the null."
- The APA notation (and the notation you need to write to get full credit) is as follows:

$$t(19) = 10.32, p < .05$$

Degrees of Freedom

t-statistic
(The one you calculate.)

< for "Reject"
> for "Fail to Reject"

The level you set your α to

Step 6: Draw Your Conclusions

- Technically, the "real" APA notation gives the actual p-value rather than the $>$ or $<$ symbol because statistical software like R will give you a p-value

$$t(19) = 10.32,\ p = 0.00$$

Degrees of Freedom

t-statistic
(The one you calculate.)

Actual p-value
(There are some numbers after a lot of zeros, but we don't need to write it all out, 1.580561e-09.)

Up Next…

- We know now that p tells up the probability of making a mistake…
- There are two main types of mistakes:
  - Type I Error
  - Type II Error

Hypothesis Testing in R

Critical Values in R

- Instead of using a table, you can use the "qt( )" function to find the critical t-value in R.
  - You need to specify the level of confidence you want (ex. 95%) and the degrees of freedom

```{r}
################################
###### Critical t Values #######
###### One & Two Tailed ########
################################

##########################################
##### TWO TAILED Tests Critical t-values ######
##########################################

# 95% Critical t-values for 95% Confidence, significance = 0.05
two_tail_crit_t_95 <- qt(p = c(.025, .975), df = )
# 99% Critical t-values for 99% Confidence, significance = 0.01
two_tail_crit_t_99 <- qt(p = c(.005, .995), df = )

##########################################
##### ONE TAILED Tests Critical t-value ######
##########################################

#### RIGHT TAIL TEST ####
# If you are testing the hypothesis something is GREATER than
#
# 95% Critical t-values for 95% Confidence, significance = 0.05
right_tail_crit_t_95 <- qt(p = .95, df = )
# 99% Critical t-values for 99% Confidence, significance = 0.01
right_tail_crit_t_99 <- qt(p = .99, df = )

#### LEFT TAIL TEST ####
# If you are testing the hypothesis something is LESS than
#
# 95% Critical t-values for 95% Confidence, significance = 0.05
left_tail_crit_t_95 <- qt(p = .05, df = )
# 99% Critical t-values for 99% Confidence, significance = 0.01
left_tail_crit_t_99 <- qt(p = .01, df = )
```

Emergency Room Wait Time (Right Tail)

```r
#########################################
##### t-test One sample, RIGHT Tail #####
#########################################

# First fill in the information that is know: Emergency Room Wait Time

sample_mean <- 13
pop_mean <- 10
sd <- 1.3
n <- 20
df <- n-1

# Find the critical values: 95% Critical t-values for 95% Confidence, significance = 0.05
right_tail_crit_t_95 <- qt(p = .95, df = df) # critical t = 1.729

# Calculate the t-statistic
t_stat_one_sample <- (sample_mean - pop_mean)/(sd/sqrt(n)) # t-statistic = 10.32

# Our t-statistic (10.32) is greater than the critcal t value (1.729)
# We can reject the null hypothesis.
```

Football and Self-Esteem (Left Tail)

```
########################################
##### t-test One sample, LEFT Tail #####
########################################

# First fill in the information that is know: Football and Self-Esteem

sample_mean <- 39
pop_mean <- 40
sd <- 6.9
n <- 25
df <- n-1

# Find the critical values: 95% Critical t-values for 95% Confidence, significance = 0.05
left_tail_crit_t_95 <- qt(p = .05, df = df) # critical t = -1.71

# Calculate the t-statistic
t_stat_one_sample <- (sample_mean - pop_mean)/(sd/sqrt(n)) # t-statistic = -0.72

# Our t-statistic (-0.72) is not LESS than the critcal t value (-1.71)
# We FAIL reject the null hypothesis.
```

Behavior Modification and Sodas (Two Tails)

```
######################################
##### t-test One sample, TWO Tailed #####
######################################

# First fill in the information that is know: Behavior Modification and Sodas

sample_mean <- 4
pop_mean <- 8
sd <- 2.5
n <- 12
df <- n-1

# Find the critical values: 95% Critical t-values for 95% Confidence, significance = 0.05
two_tail_crit_t_95 <- qt(p = c(.025, .975), df = df) # critical ts = -2.20, 2.20

# Calculate the t-statistic
t_stat_one_sample <- (sample_mean - pop_mean)/(sd/sqrt(n)) # t-statistic = -5.54

# Our t-statistic (-5.54) is further out than the critcal t values (-2.20, 2.20)
# We reject the null hypothesis.
```

Looking Up Exact p-values in R

```r
########################################
##### t-test One sample, TWO Tailed #####
############### p-values ################
########################################

# With R, we can also look up the exact p-value using "pt()"
# The "p" in "pt()" is asking for the p value from a t distribution
# the "q" in "pt()" is asking you to put in the t-statistic you calculated above
# We also need to use the "abs()" function to take the absolute value of the t-statistic,
# then add a "-" sign in front... I know it's strange, but it works!

# Example of a RIGHT Tail t-test
# Emergency Room Wait Time: t-statistic (10.32), df = 19
# p = 0.00000
pt(q = -abs(10.32), df = 19)

# Example of a LEFT Tail t-test
# Football and Self-Esteem: t-statistic (-0.72), df = 24
# p = 0.24
pt(q = -abs(-0.72), df = 24)

# Example of a TWO Tail t-test
# Behavior Modification and Sodas: t-statistic (-5.54), df = 11
# p = 0.0001
pt(q = -abs(-5.54), df = 11, lower.tail = T)*2
```

Instead of just checking if the t-static is greater than the critical value, you could look up the exact p-value in R