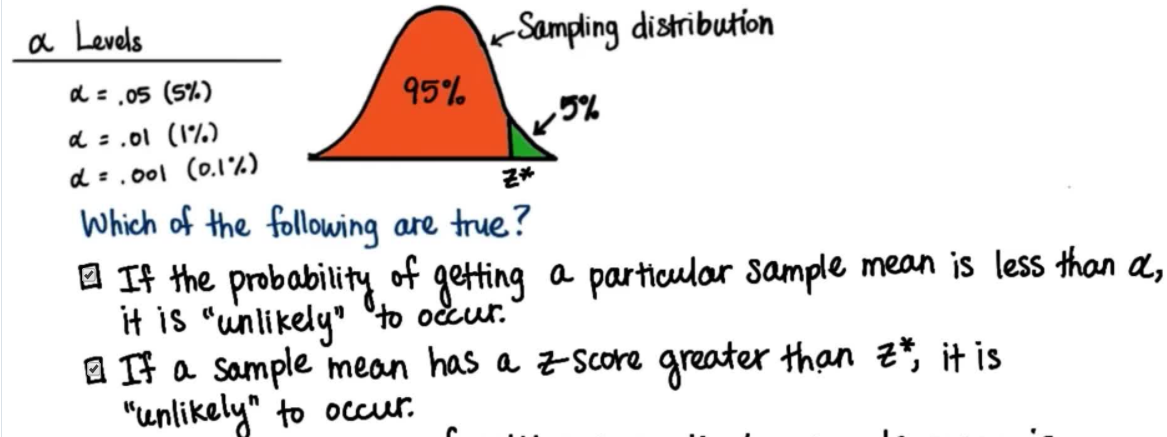
***Udacity Data Analyst Track***

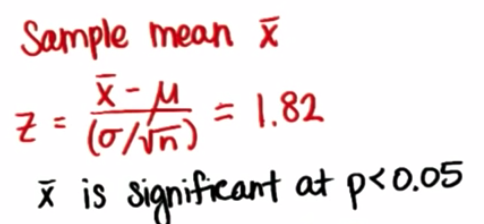
**I. Into to Inferential Stats**

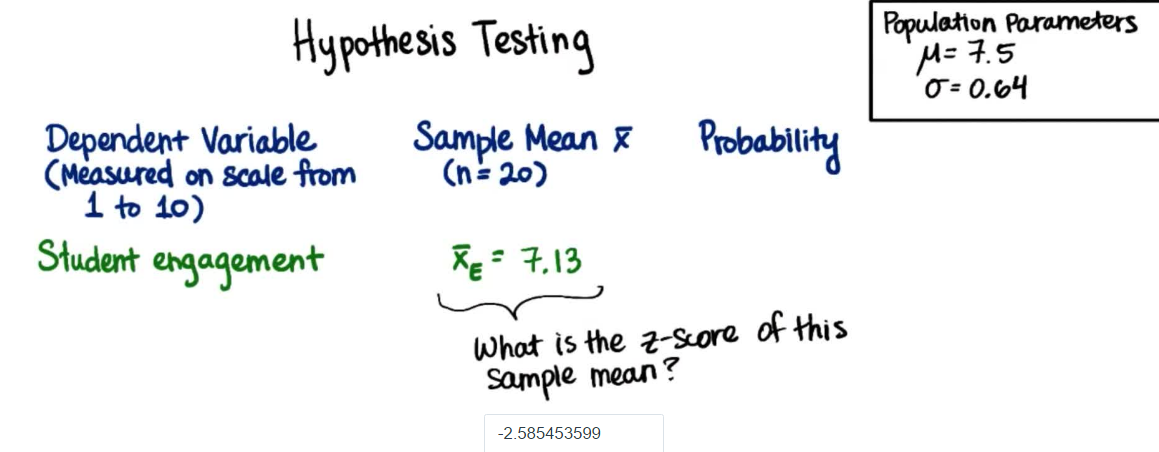
3. Hypothesis Testing

* **A hypothesis test** = used to test a claim about how an observation may be different from the *known* population parameter.
* **Alpha level (a)** =helps us determine the **critical region** of a distribution.
* **Levels of Likelihood/Alpha Levels**
* If probability of getting a sample mean is < 0.05, 0.01, 0.001 (5%, 1%, 0.1%) 🡺 UNLIKELY

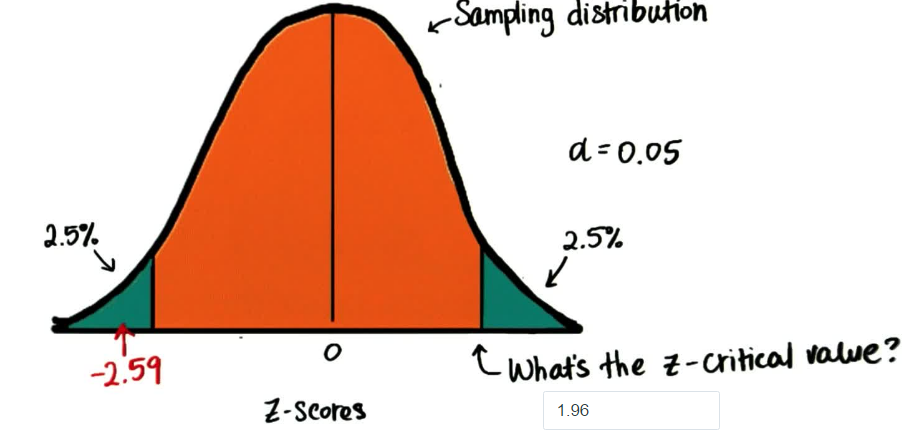


* So, if probability of selecting a sample mean is < alpha ( p < a ), it falls in the **critical region**, which is cut off from the rest of the distribution by the z-score of the critical region = **z-critical value**
* If z-score of a sample mean > z-critical = evident the sample statistics are different from the regular/untreated population
* Alpha = 0.05 🡪 5% 🡪 z-crit = 1.64 0.01 🡪 1% 🡪 z-crit = 2.32 0.001 🡪 0.1% 🡪 3.080.025 🡪 2.5% 🡪 z-crit = 1.96
* Those sample means that fall in any of the critical regions are those that are unlikely to have happened by chance
* sample statistics are very different to population parameters so there’s strong evidence of an effect from some intervention

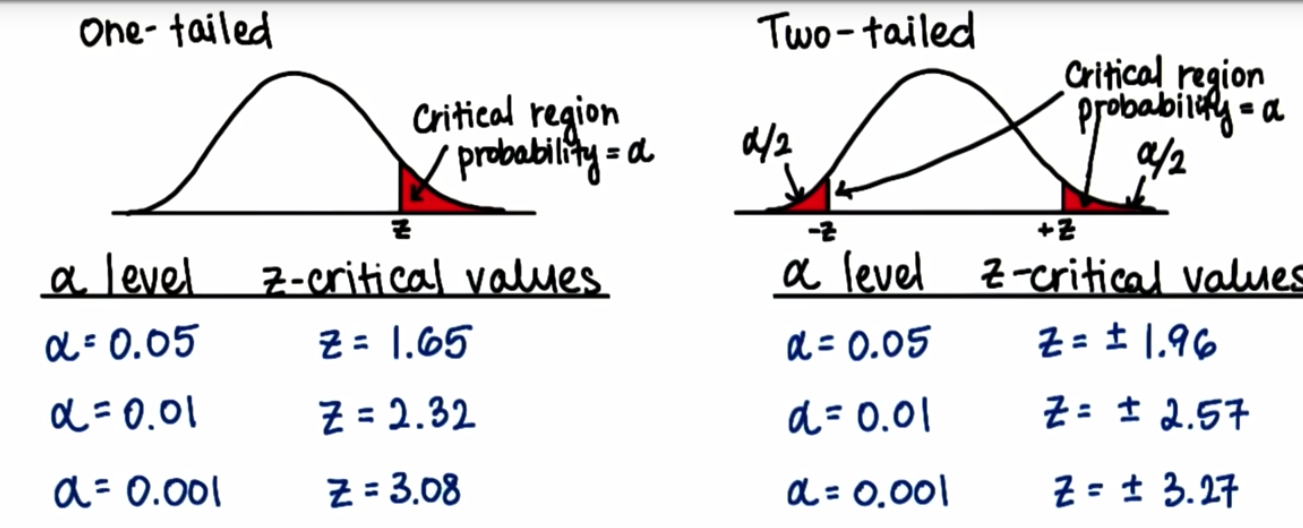




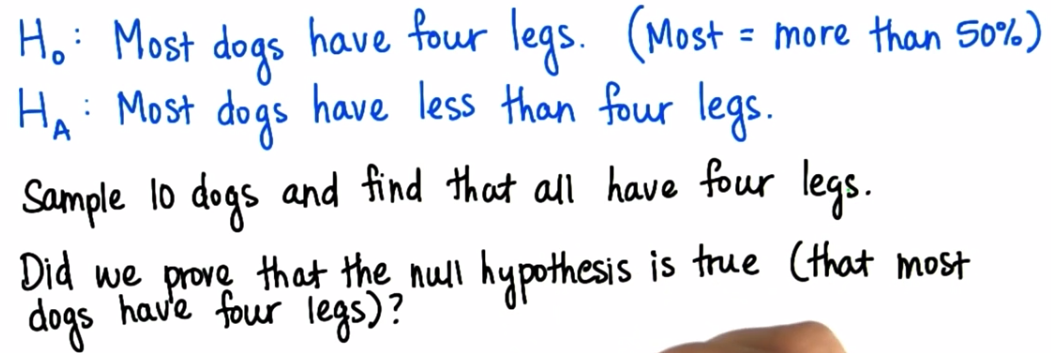
* Alpha split in 2 for TWO critical regions = **Two-Tailed test**
* 0.05 🡪 2.5%/0.025🡪 z-crit = 1.96 0.01 🡪 ½%/0.005 🡪 z-crit = 2.575 0.001 🡪 0.05%/0.0005 🡪 3.32



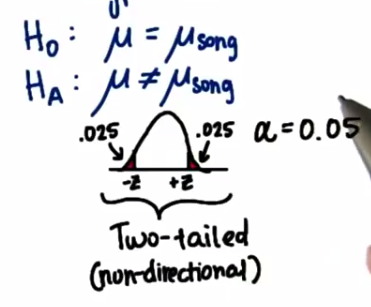
* So, we can see that getting a sample mean of 7.13 in this distribution is unlikely (falls w/in critical region), so there is evidence that singing made students *less* engaged, so a mean engagement score of 7.13 is significant w/ alpha = 0.05



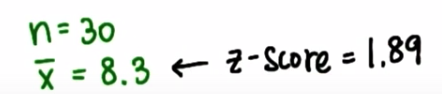
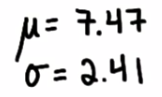
* Sample means in critical regions = not by chance = due to intervention (usually w/ alpha = 0.05)
* See same z-values as calculating CI’s, just used in a different manner
* 1-tailed test 🡪 could be on either side of plot (either tail)
* **Null Hypothesis** = always an equality = the claim we’re trying to provide evidence *against* 🡪 little to no difference between current population parameters and post-intervention population parameters
* Cannot prove true, can only evidence that is isn’t
* H0 : µ0 = µ
* H0 : µ0 >= µ
* H0 : µ0 <= µ
* **Alternative Hypothesis =** theresult we are checking *against the claim* 🡪 always some kind inequality 🡪 some significant difference between current population parameters and post-intervention population parameters
* Ha : µa <> µ 🡪 Ha : µa > µ or Ha : µa < µ

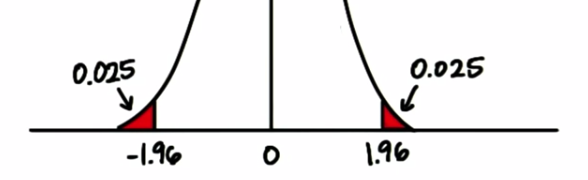


* Did NOT prove H(0) was true, but we have evidence it’s true, so based on our sample we **fail to reject the null**
* If we sampled 10 dogs and 6 have 3 legs, this IS evidence that we CAN reject the null (most dogs in sample have < 4 legs)
* Ex: In survey of course engagement and learning
* not a very clear definition of either (someone only views 1 engagement but got too busy = low engagement, but high learning
* There are arbitrary reasons for responses to the survey so CI’s are not useful here, where we don’t have concrete operational definitions (no REAL meaning to scores)
* Possible H(0)’s 🡪 a song for the lesson does NOT make learners more engaged, does NOT change how much learners are engaged, results in SAME level of engagement
* Possible H(A)’s 🡪 a song for the lesson WILL make learners more engaged, DOES change how much learners are engaged, results in LESS-engaged learners
* 1-tailed tests 🡪 sample mean < population mean, sample mean > population mean
* Predicting DIRECTION of treatment effect (song increases engagement)
* 2-tailed test 🡪 sample mean <> population mean (=/= population mean 🡪 EITHER)
* Do NOT predict direction of treatment effect (song changes engagement)
* More conservative test 🡪 less likely to reject Null
* \*\*Exception 🡪 comparing new treatment w/ established treatment 🡪 only care if new treatment is BETTER, don’t care if worse 🡪 1-tailed

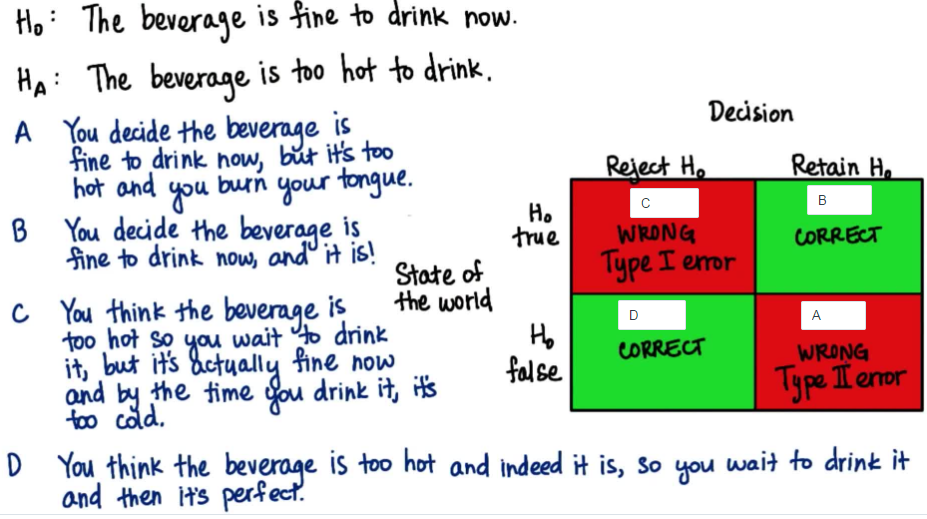


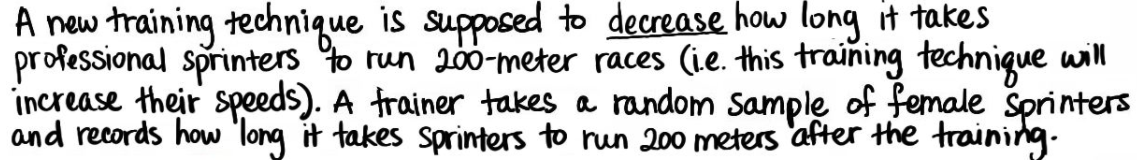
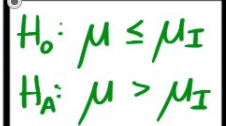
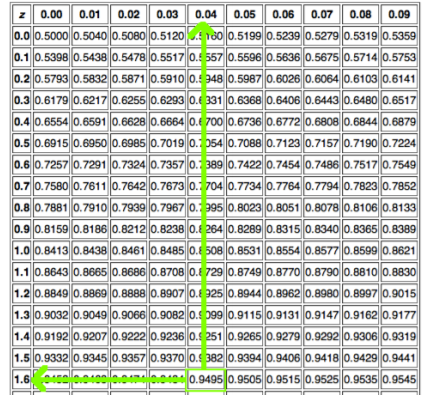
* Reject this null when sample mean is w/in critical region OR is less/greater than z-critical OR probability of getting the sample mean is less than the alpha level





* Fail to reject NULL 🡪no evidence song affected engagement
* Increase sample size to 50 🡪 z-score = 2+ 🡪 reject NULL w/ p < 0.05 (reject b/c probability *p* of getting this mean value with this sample size is very low 🡪 evidence song affected engagement
* This kind of statistical analysis is prone to be misinterpretation
* Possible some who viewed the song were *already* more engaged, but we found the probability of getting that mean score was 0.007, but it’s still possible it was a randomly-obtained mean
* **Statistical Decision Errors**
* Reject Null when it’s true 🡪 **Type I Error**
* Accept Null when it’s false 🡪 **Type II Error**



* So stats is always open to misinterpretations, data only gets us so far. What matters is HOW we get the data (random sample, large enough, etc.)
* The alpha level is smaller for smaller **critical regions** = define unlikely values if Null = TRUE
* When z-score is large, do not automatically accept Null
* To retain the Null is to say we believe the intervention had no effect based on our sample
* If an effect exists, it’s more likely to be detected if n or if pop SD is large
* 
* 
* Ex: A towns census from 2001 reported average age of people living there = 32.3 w/ SD = 2.1 years.
* The town takes a sample of 25 people + finds average age = 38.4 years.
* Test the claim that the average age of people in the town has increased w/ a level = 0.05
* 1st define our hypotheses:
* H0 : µ0 = 32.3 years
* Ha : µ0 > 32.3 years
* ID the important info
* x¯ = 38.4 s = 2.1 n = 25 SE = 2.1/sqrt(25) = 0.42
* Last piece of important info we need is our **critical value**:
* Finding Z-critical value 🡪 look up as close as we can to 95% = **Z-crit =** 1.64
* 
* Once we have all our important info we can now find our **test statistic**:
* Z-score = obs – mean / SE 🡪 38.4 - 32.3 / 0.42 = **14.5238**
* Since our z-score *is much bigger than our z-crit* 🡪 REJECT the claim/null) that the average age of people was 32.3 years
* **Type I Error** 🡪 reject the null when null hypothesis is *actually true* (FP)
* Probability of committing a Type I error = alpha level (a)
* **Type II Error** 🡪 fail to reject the null when it is actually false (FN)
* Probability of committing a Type II error = alpha level (b)
* Ex: An insurance company is reviewing its current policy rates. When originally setting rates, they believed the average claim amount = $1,800. They’re concerned the true mean is actually higher than this, b/c they could potentially lose a lot of money. They randomly select 40 claims (n), + calculate a sample mean = $1,950. Assuming the SD of claims = $500, set a = 0.05 + test to see if the insurance company should be concerned.
* H0 : µ0 = $1,800
* Ha : µ0 > $1,800
* x¯ = $1,950 s = $500 n = 40 SE = 500/sqrt(40) = 79.0569415
* Z-crit = 1.64
* Z-score = obs – mean / SE 🡪 1,950 – 1800 / 79.0569415 = **1.897366596**
* Since our z-score is bigger than our z-crit 🡪 REJECT the claim/null)
* Ex: Explain a type I and type II error in context of the problem. Which is worse?