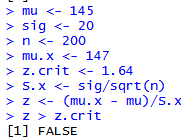
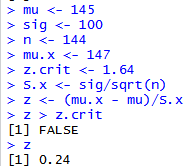
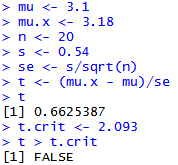
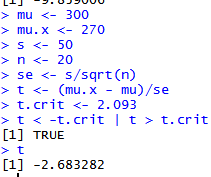
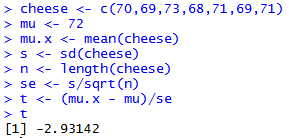
* **Hypothesis testing** = a kind of statistical inference that involves asking a question, collecting data, + then examining what the data tells us about how to proceed.
* In a *formal* hypothesis test, hypotheses *are always statements about the POPULATION*.
* In statistical hypothesis testing, there are always 2 hypotheses.
* The hypothesis to be tested is = **null hypothesis**, **H0**, which states that there is no difference between a hypothesized population mean + a sample mean.
* Test the null hypothesis against an **alternative hypothesis, Ha**, often the hypothesis you believe yourself + includes the outcomes not covered by the null hypothesis.
* We have a medicine being manufactured + each pill is supposed to have 14 mg of active ingredient 🡪 H0 : µ = 14 Ha : µ != 14
* Alternative hypothesis can be supported only by **rejecting the null =** finding a large enough difference between your sample mean + the hypothesized (null) mean that it raises real doubt that the true population mean is what we said.
* **In each hypothesis test, must decide in advance what the magnitude of that difference must be to allow us to reject the null hypothesis.**
* **2-tailed hypothesis tests** 🡪 do not specify whether we believe the true mean to be higher or lower than the hypothesized mean, just believe it must be different.
* In a two-tailed test, reject the null if your sample mean falls in *either* tail of the distribution.
* For this reason, the alpha level (let’s assume .05) is split across the 2 tails into 2 critical regions for alpha = 0.025
* The **z-scores** that designate the start of the critical region = **critical values**
* If the sample mean taken from the population falls w/in these critical/**rejection regions**," conclude there was too much of a difference to have happened by chance + reject the null
* Use a 1-tail hypothesis test when the direction of the results is anticipated or we are only interested in 1 direction of the results (ex: only decide to adopt the textbook if it improved student achievement relative to the old textbook)
* When performing a 1-tail hypothesis test, h(a) utilizes > or < or less than. For example, let’s say we were claiming that the average SAT score of graduating seniors
* 1-tail hypothesis tests also only have 1 critical region b/c we put the entire critical region into just 1 side of the distribution.
* When h(a) is that the sample mean is greater, the critical region is on the right of the distribution + when the sample is smaller, the critical region is on the left side
* Remember some sample are extremes + are going to happen about 5% of the time, since 95% of all sample means fall w/in about 2 SD’s of the mean.
* If we run a hypothesis test + get an extreme sample mean, it won’t look like our hypothesized mean, even if it comes from that distribution 🡪 would be likely to reject the null, *but we would be wrong.*
* Which type of error (Type 1 = FN, Type II = FP) is more serious depends on the specific research situation, but ideally both types of errors should be minimized during analysis.
* The general approach to hypothesis testing focuses on Type I = rejecting null when it may be true.
* **The level of significance/the alpha level = the probability of making a Type I error**.
* At 0.05, the decision to reject the hypothesis may be incorrect 5% of the time.
* Calculating the probability of making a Type II error is not as straightforward
* You should be able to recognize what each type of error looks like in a particular hypothesis test.
* Ex: Testing whether listening to rock music helps improve memory of 30 random objects + assume further it doesn’t.
* Type I error = Say it improves memory, but it actually makes it worse or doesn’t affect it at all
* Type I errors only occur when the null is false.
* Type II error = assume listening to rock *does* improve memory, but we conclude it didn’t, but it really did
* *failing to find a significant difference when one in fact exists.*
* It is also important to realize the chance of making a Type I error is under our direct control.
* Often WE establish an alpha level based on the severity of consequences of making a Type I error
* If consequences are not that serious, we could set an alpha level at 0.10 or 0.20
* In other words, we’re comfortable making a decision where we could falsely reject the null 10 to 20% of the time.
* However, in a field like medical research, we would set the alpha level very low (at 0.001) if there was potential bodily harm to patients.
* **Critical values =** values that indicate the edge of the **critical region** = describe the entire area of values that indicate to reject the null (values not included in the initial claim)
* The **tails** of a test = values outside of critical values = the ends of a distribution = begin at the greatest/least value included in the alternative hypothesis (the critical values).
* Ex: Researcher claims black horses are, on average, more than 30 lbs. heavier than white horses, which average 1100 lbs.
* H(0): weight(b) <= 1400 h(a): weight(b) > 1140
* One-tailed t-test 🡪 right-tailed
* Ex: Package of gum claims flavor lasts more than 39 minutes.
* H(0): flavor <= 39 h(a): flavor > 39
* One-tailed t-test 🡪 right-tailed
* Ex: An ice pack claims to stay cold between 35-65 minutes.
* H(0): 35 > μ ∪ μ > 65 h(a): 35 < μ ∪ μ > 65
* 2-tailed t-test 🡪 right-tailed
* Z-score critical values of a normal distribution are an appropriate resource for conducting tests when *we KNOW the parameters of our population*.
* Z-crit for 95% confidence in a 2-tailed test 🡪 split 5% under the curve into 2.5% **🡪 +/- 1.96**
* Z-crit for a right-tailed test w/ α = 0.01 🡪 99% confidence 🡪 0.9900 🡪 **+2.33**
* Recall that hypothesis testing is a form of **statistical inference**.
* Previously, we **inferred** about a population by calculating a **confidence interval** + estimated a true mean of a population from a sample mean + created a CI provided a **margin of error** for an estimate.
* When conducting a hypothesis test, we are asking ourselves whether the info in the sample is consistent or inconsistent w/ the null hypothesis about the population.
* We follow a series of 4 basic steps:
* 1. State the null and alternative hypotheses.
* 2. Select the appropriate significance level and check the test assumptions.
* 3. Analyze the data and compute the test statistic.
* 4. Interpret the result
* If we reject the null, we are saying the difference between the observed sample mean + the hypothesized population mean is too great to be attributed to chance.
* When we fail to reject the null, we are saying the difference between the observed sample mean and the hypothesized population mean is probable if the null hypothesis is true.
* Essentially, we are willing to *attribute this difference to sampling error*.
* Conducting a Hypothesis Test on One Sample Mean When the Population Parameters are Known
* Although this is rarely the case, we can use our familiar z-statistic to conduct a hypothesis test on a SINGLE sample mean.
* In short, we find the z-statistic of our sample mean in the sampling distribution + determine if that z-score falls w/in the critical region or not.
* *This test is only appropriate when you know the true mean + SD of the population.*
* Ex: school nurse thinks the average height of 7th graders has increased. The average height of a 7th grader 5 years ago = 145 cm w/ SD = 20 cm. She takes a random sample of 200 students + finds average height = 147 cm. Are 7th graders now taller than they were before? Conduct a single-tailed hypothesis test using a .05 significance level to evaluate h(0) and h(a)
* H(0): mu.x <= mu 🡪 mu.x <= 145 h(a) = mu.x > mu
* Alpha = .05 for ONE TAIL therefore z-crit = +1.64
*  🡪 FAIL TO REJECT NULL
* 🡪92.07% chance to get this height
* Ex: Farmer trying out a planting technique to increase yield on pea plants. The average # of pods on 1 plant = 145 pods w/ SD = 100. This year, after trying his new technique, he takes a random sample of 144 plants + finds average # of pods to = 147. He wonders whether or not this is a statistically significant increase w/ alpha = 0.05
* H(0) = mu.x <= mu 🡪 mu.x <= 145 h(a): mu.x > mu 🡪 mu.x > 145
* Alpha = .05 for ONE TAIL therefore z-crit = +1.64
*  🡪 FAIL TO REJECT NULL
* We can also evaluate a hypothesis by asking, “What is the probability of obtaining the value of the test statistic we did if the null hypothesis is true?” 🡪 **the p−value.**
* Ex: farmer wondering if # of pea pods per plant has gone up w/ a new planting technique + finds from a sample of 144 peas an average of 147 pods/plant (compared to previous average of 145).
* To determine the p, we ask what is **P(z >. 24)**? = what is the probability of obtaining a z-score greater than .24 if the null hypothesis is true?
* Using technology, we find this probability to be .49 = indicates a 49% chance that under the null hypothesis, the peas will produce more than 145 pods.
* Alternatively, we can just indicate that p >.05.
* Since we set alpha at .05, we won’t reject if the probability of observing that sample mean is >.05.
* When using technology to conduct a hypothesis test, you will receive a p-value as part of the output.
* p-values = the likelihood of observing that particular sample value if the null hypothesis were true
* *Therefore, if p is smaller than your significance level, you can reject the null hypothesis.*
* A z-test makes for an easy hypothesis test, but most of the time we can’t use it.
* The reality is that most analyses are done when we don’t know what is true about a population.
* Instead, we want to conduct a hypothesis test using only sample + the info it can provide us.
* Back in the early 1900’s, a chemist at a brewery in Ireland discovered that when working w/ very small samples, the distributions of the mean differed significantly from the normal distribution.
* He noticed that as his sample sizes changed, the shape of the distribution changed as well.
* He published his results under the pseudonym ’Student’ + this concept + the distributions for small sample sizes are now known as **Student’s t−distributions**
* This is similar to the normal distribution, except it is more spread out + wider in appearance w/ thicker tails.
* As the # of observations gets larger, the t-distribution shape becomes more + more like the shape of the normal distribution.
* In fact, if we had an infinite number of observations, the t distribution would perfectly match the normal distribution.
* It’s the t-distribution that allows us to test hypotheses when we DON’T know the true population SD
* The differences between the t + normal distribution are more exaggerated w/ fewer data points + therefore fewer **degrees of freedom = #** of samples that have the ’freedom’ to change w/out affecting the sample mean.
* All you really need to know about degrees of freedom is that there is always 1 less degree of freedom than the number of data points:
* When you use a t-distribution for a hypothesis test, there is a different critical value for each dF
* The larger your sample, the closer the critical value gets to the z-score for your alpha level.
* If conducting a 2-tailed hypothesis test on a sample of 25 students w/ df = 24 + α = 0.05, t = ±2.064 b/c there is 0.025 in each tail.
* Conditions for using the t-test
* The t−distribution can be used w/ any statistic having a bell-shaped distribution.
* The CLT states the sampling distribution of a statistic will be close to normal w/ a large enough sample size.
* As a rough estimate, the CLT predicts a roughly normal distribution under any of the following conditions:
* The population distribution is normal; or
* The sampling distribution is *symmetric* + sample size is ≤ 15;
* The sampling distribution is *moderately skewed* + sample size is 16 ≤ n ≤ 30
* The sample size is greater than 30, w/out outliers.
* **When we KNOW the population standard deviation 🡪 use the normal distribution.**
* **When we DON’T know it 🡪 need to use sample standard deviation 🡪 use the t-distribution.**
* We use the Student’s t−distribution in hypothesis testing the same way we use the normal distribution.
* Each row in the t table represents a different t−distribution, each associated w/ a unique dF
* In calculating t, we do **(x(bar) – mu(0))/(s/sqrt(n)**
* x(bar) = sample mean
* mu(0) = population mean under null
* s = sample SD
* s/sqrt(n) = estimated standard error
* n = dF



* Assumptions of the single sample t-test:
* A random sample is used.
* The random sample is made up of independent observations
* The population distribution must be nearly normal, or the size of the sample is large.
* Ex: AS is asked if football players are doing as well academically as the other student athletes. We know from a previous study the average GPA for student athletes = 3.10. After an initiative to help improve GPA of student athletes, AD randomly samples 20 football players + finds average GPA = 3.18 w/ a sample SD = 0.54. Is there a significant improvement? Use a 0.05 significance level.
* H(0): mu.x = mu 🡪 mu.x = 3.10 h(a): mu.x != mu 🡪 mu.x != 3.10
*  🡪 fail to reject null
* average GPA of football players is not significantly different than other student athletes
* difference between sample mean + hypothesized value is not sufficient enough to attribute it to any factor other than sampling error
* Ex: Duracell manufactures batteries the CEO claims will last an average of 300 hours under normal use. A researcher randomly selected 20 batteries from the production line + tested these batteries. The tested batteries had a mean life span of 270 hours w/ a SD = 50 hours. Do we have enough evidence to suggest that the claim of an average lifetime of 300 hours is false?
* H(0): mu.x = mu 🡪 mu.x = 300 h(a): mu.x != mu 🡪 mu.x != 300
*  🡪 Reject null
* average life span of batteries is significantly different than claims
* When a hypothesis is rejected, it is often useful to turn to estimation to try to capture the true value of the population mean.
* i.e. Now that we have rejected the claim of average lifetime of a battery = 300 hours, we want to know how long these batteries do indeed last.
* We will choose a 95% CI, essentially the same degree of confidence we had in our t-test, since **confidence = 1 - α**
* To calculate CI, we need to know 3 things 🡪 mean of our sample, standard error, + the critical value.
* **CI = x(bar) +/- margin of error** where Margin of Error **(ME) = t-crit \* SE(x)**
* Note we are using t-critical (rather than z) b/c we are working only w/ sample data.
* We do not know the true population mean or standard deviation.
* We have to use both our sample mean and our sample standard error to create the CI
* So, for batteries, the Margin of Error = 2.093\*11.18 = **23.4**
* Therefore, our 95% CI = (270 –23.40) to (270 + 23.40) 🡪 **(246.6, 293.4)**
* This shows us the CI for the population mean does NOT include 300.
* This is what we would expect, since we rejected the null hypothesis in our earlier hypothesis test.
* If we were to *interpret* the 95% CI, we’d say: “I am 95% confident the true population mean of battery lifespan is between 246.6 and 293.4 hours.”
* Ex: You have just taken ownership of a pizza shop + previous owner told you you’d save $ if you bought mozzarella cheese in a 4.5 pound slab. Each time you purchase a slab of cheese, you weigh it to ensure that you’re receiving 72 ounces of cheese. The results of 7 random measurements are 70, 69, 73, 68, 71, 69 and 71 ounces. Are these differences due to chance or is the distributor giving you less cheese than you deserve?
* H(0): mu.x = mu 🡪 mu.x = 72 h(a): mu.x != mu 🡪 mu.x != 72



* Would the null hypothesis be rejected at the:
* 10% level? 🡪 t-crit = +/-1.943 🡪 reject
* 5% level? 🡪 t-crit = +/-2.447 🡪 reject
* 1% level? 🡪 t-crit = +/-3.707 🡪 fail to reject