Hypothesis Testing Notes

Inferential statistics is the process of employing hypothesis testing to random samples drawn from unknown population distribution to determine a targeted characteristic. The central limit theorem provides the foundation upon which it is possible to have confidence in the value of the point estimates subjected to hypothesis testing.

Steps for a Hypothesis Test

1. State the null hypothesis.
2. State the alternative hypothesis.
3. Determine the critical value (based upon chosen type of distribution (normal or Student's t)
4. Determine if this will be a 1-tail or 2-tailed test.
5. Calculate the test statistic.
6. Draw a conclusion (inference) from the results.

The process of hypothesis testing begins with the framing of two hypotheses from the research question – the null hypothesis and the alternative hypothesis.

H_0: The **Null Hypothesis** is an opposing statement to your proposed research question. The process of hypothesis testing is actually null hypothesis testing such that the results will allow you to either *reject the null hypothesis* or *fail to reject the null hypothesis*. You will never use the results of a hypothesis to accept the null hypothesis.

H_a: The **Alternative Hypothesis** is a direct statement drawn from your research question and it is basically the contradictory statement to the null hypothesis (H_0). When you perform the hypothesis test, the result will be used to either *reject the null hypothesis* or *fail to reject the null hypothesis*. When you reject the null hypothesis, this means that there is sufficient evidence to support the alternative hypothesis (H_a). If the results of the test require you to *fail to reject the null hypothesis*, this means that there is insufficient evidence to support the alternative hypothesis (H_a).

In this course, you've been asked to develop a research question that falls into one of the following cases:

(1) Comparing a sample mean to a proposed population parameter value.

Examples;

Is the mean length of female alligators in the Everglades greater than 200 cm?

Is the mean length of female alligators in the Everglades less than 200 cm?

(2) Compare the mean of one sample to the mean of another sample.

Examples:

Does the mean length of female alligators in the Everglades differ from the mean length of female alligators in Louisiana?

Comparing the sample mean to a proposed population parameter value.

Let's say that you've created a research question that fits into case (1) above. Here are a few examples of how you might format the null and alternative hypotheses using the examples:

Is the mean length of female alligators in the Everglades greater than 200 cm?

H_0: *The mean length of female alligators in the Everglades is less than or equal to 200 cm.*

H_a: *The mean length of female alligators in the Everglades is greater than 200 cm.*

In the example above, we are testing the sample mean against the value of 200 cm. Symbolically, we can represent these hypotheses as follows:

H_0: $\mu \leq 200$

H_a: $\mu > 200$

In the null hypothesis we always include the *or equal to* when we use an inequality. The following table shows the required symbolic relations for the alternative hypothesis compared to how the null hypothesis is formatted:

H_o	H_a
equal (=)	not equal (≠), or greater than (>), or less than (<)
greater than or equal to (≥)	less than (<)
less than or equal to (≤)	greater than (>)

Consider the following example:

Research Question: *Is the mean nose length of Tasmanian spotted-aardvarks greater than 9.2 inches?*

Null Hypothesis wording: *The mean nose length of Tasmanian spotted-aardvarks is less than or equal to 9.2 inches.*

Null Hypothesis symbolically: H_0: $\mu \leq 9.2$

Alternative Hypothesis wording: *The mean nose length of Tasmaian spotted-aardvarks is greater than 9.2 inches.*

Alternative Hypothesis symbolically: H_0: $\mu > 9.2$

This kind of hypothesis test will involve the use of a 1-sample t-test that determines the statistical significance of the relationship between the sample mean and the proposed population parameter.

Note: For samples greater than or equal to 30, you can assume a normal distribution and use z-score calculations for your hypothesis tests. However, since the t-distribution approximates the normal curve for sample sizes over 30, we will only explore the application of t-tests below.

Comparing Sample Means (Independent Means)

Let's say that you have the following research question in which the mean of one sample is compared to the mean of another sample.

Does the mean weight of Texas White-tailed deer in 2000 differ from the mean weight of Texas White-tailed deer in 2020?

Here's how you would format the null hypothesis:

H_0: *The mean weight of Texas White-tailed deer in 2000 does not differ from the mean weight of Texas White-tailed deer in 2020.*

H_0: $\mu_1 = \mu_2$

Here's how you would format the alternative hypothesis:

H_a: *The mean weight of Texas White-tailed deer in 2000 differs from the mean weight of Texas White-tailed deer in 2020.*

H_a: $\mu_1 \neq \mu_2$

This type of question is generally known as a comparison of the means. The statistical test employed is a 2-sample t-test.

Another approach to determining the relationship between two sample means is to test the differences between the matched or paired variables. In this case, you would be attempting to determine if there is a greater than zero difference between the samples. The hypotheses would be established as follows:

H_0: $\mu_d = 0$

H_a: $\mu_d > 0$

The Level of Significance

When a researcher conducts a study, the level of statistical significance is typically determined before the hypothesis test is performed. The level of significance, α, the probability of rejecting the null hypothesis when it is true. For example, a significance level of .05 indicates a 5% risk of rejecting a correct null hypothesis. While it is possible to select any value for the level of significance, researchers will select the lowest percentage that they feel will be acceptable for their research. The most common level of significance is α=.05 which also implies a 95% confidence level.

The P-value

The results of a hypothesis test will always report a p-value. The p-value represents the probability associated with the test statistics assuming that the null hypothesis is valid. In practice, the p-value is used as a decision value when compared to the selected level of significance α to determine if the null hypothesis should be rejected. A small p-value will mean that there is strong evidence to support the alternative hypothesis.

If p−value < α, then *reject the null hypothesis*. This means that there is sufficient evidence to support the alternative hypothesis.

If p−value ≥ α, then *fail to reject the null hypothesis*. This result means that there is not sufficient evidence to support the alternative hypothesis.

When you conduct hypothesis testing, there are only four outcomes. The table below summarizes these outcomes:

	Null is True	**Null is False**
Reject	*Type I Error (α)*	*OK*
Fail to Reject	*OK*	*Type II Error (β)*

As the table shows, a Type I Error occurs when the null hypothesis is rejected when it is actually true. The probability of this occurring is represented by α. When the null is false and we fail to reject it, this is known as a Type II Error. The probability of a Type II error is represented by β. It is worth noting that α and β should be as small as possible due to the fact that represents the probability of making these errors, however they are never zero. A low β results in a higher power for the hypothesis test represented by: $power = 1 - \beta$. In general, the power of the hypothesis test can be increased by increasing the sample size.

Example 1: The nap times of a random sample of 25 Nifflers was observed with a mean of 22.1 minutes and a standard deviation of 5.3 minutes. Is the mean nap time of Nifflers less than 25.4 minutes? Use a level of significance of $\alpha = .05$.

H_o: $\mu \geq 25.4$ H_a: $\mu < 25.4$

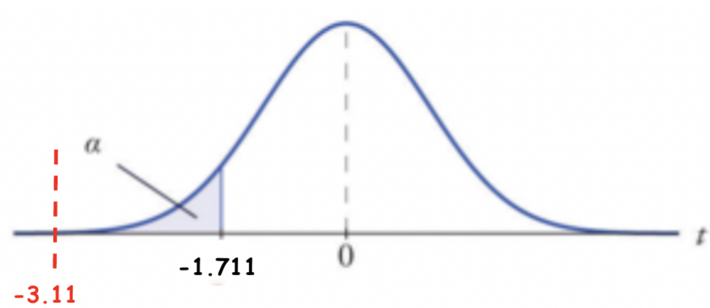
n = 25; df = 25 - 1 = 24 $\alpha = .05$

t-critical = -1.711 found using https://goodcalculators.com/student-t-value-calculator/

Use https://www.socscistatistics.com/pvalues/tdistribution.aspx to find the p-value from the t-score and degrees of freedom.

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{22.1-25.4}{\frac{5.3}{\sqrt{25}}} = -3.11$$

p-value: $p(t \geq -3.11) = .002386$



Results: Since the p-value < α (.002386 < .05), we will *reject the null hypothesis*.

Inference Drawn: *Based upon the data collected, it appears that the mean nap time for Nifflers is less than 25.4 minutes.*

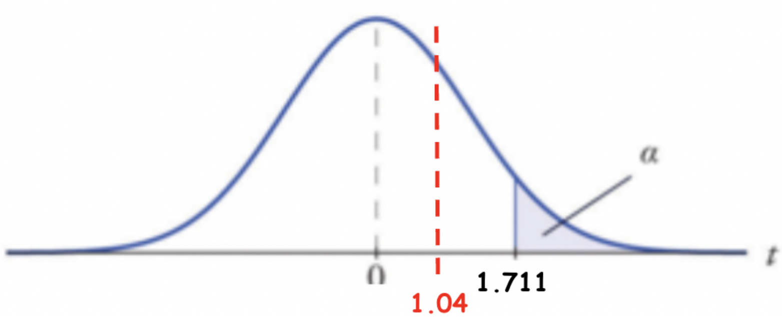
Example 5: From a random sample of 25 peanut farms in Virginia, it was determined that the mean number of peanuts per acre was 3120 with a standard deviation of 578. Is the mean number of peanuts per acre for all farms in Virginia greater than 3000 ($\alpha = .05$)?

H_o: $\mu \leq 3000$ H_a: $\mu > 3000$

n = 25; df = 25 - 1 = 24 $\alpha = .05$

t-critical = 1.711 found using https://goodcalculators.com/student-t-value-calculator/

Our t-value will be:



$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{3120 - 3000}{\frac{578}{\sqrt{25}}} = 1.04$$

p-value: $p(t \leq 1.04) = 0.1543$

Results: Since the p-value > α (.1543 > .05), we will *fail to reject the null hypothesis*.

Inference Drawn: *Based upon the data collected, it appears there is not sufficient evidence to suggest that the mean* number of peanuts per acre for all farms in Virginia is greater than 3000.