<u>Section 9: Hypothesis Tests Based on a Single Sample</u>

<u>General Setup of a Hypothesis Test</u>

Statistical Hypotheses – statements about population parameters

<u>Examples</u>
- Mean weight of adult males is greater than 160 ($\mu > 160$)
- Proportion of students with a 4.0 GPA is less than .01 ($\mu < .01$)

In statistics, we test one hypothesis against another

Alternative Hypothesis – the hypothesis that we want to prove, denoted H_a (sometimes denoted as H_1)

Null Hypothesis – the hypothesis formed that contradicts H_a, denoted H_0

In a hypothesis test we actually test the null hypothesis. We try and prove the null hypothesis false with a high level of confidence. After taking the sample, we must make one of the following decisions:
1. Reject H_0 (and believeH_a)
2. Fail to Reject H_0 (there was not sufficient evidence to reject H_0 and conclude H_a)

There are two types of error that can occur in a hypothesis test.

Type I Error – the decision is Reject H_0 when H_0 is true

Type II Error – the decision is Fail to Reject H_0 when H_0 is false

Level of Significance – the probability with which we are willing to risk a type I error, denoted α (alpha), $P(Reject\ H_0|H_0\ is\ true) = \alpha$

In a hypothesis test α represents the same quantity as it did in a confidence interval. If $\alpha = .05$ and we can conclude H_a, then we are 95% confident in our conclusion (the quantity $1 - \alpha$ is the **Level of Confidence**)

The probability of making a type II error is denoted β (beta),
$P(Fail\ to\ Reject\ H_0|H_0\ is\ false) = \beta$

Power – the quantity $1 - \beta$ which represents the probability of rejecting H_0 when in fact it is false (or proving H_a when it is true)

Power is difficult to calculate and is beyond what we will do in this class. However, I do want you to understand the relationship between the two errors in hypothesis testing along with how they relate to power and level of confidence.

The table below identifies when the correct decision is made along with the types of error. Also, the probability of each decision is included in parentheses.

Decision from the Hypothesis Test	**Decision with a perfect census (actual truth about H_0)**	
	H_0 is false	H_0 is true
Reject H_0	Correct Decision (Power, $1-\beta$)	Type I Error (Level of Significance, α)
Fail to Reject H_0	Type II Error (β)	Correct Decision (Confidence Level, $1-\alpha$)

In inferential statistics we obviously always want a small amount of error. For hypothesis testing we desire α and β to be small.

Decreasing α increases β if all other quantities are held constant (it is always the case in inferential statistics that if you decrease one type of error and hold everything else constant, you will increase the other type of error)

Increasing the sample size is the way to ensure that both α and β are small

Test Statistic – the statistic we compute to make the decision in a hypothesis test (sampling distribution of test statistic given that H_0 is true must be known or well approximated – we have already covered these so they will look familiar)

Critical Region (Rejection Region) – the values of the test statistic such that we reject H_0 and conclude that H_a is true (the critical region depends H_a and α)

Critical value – the endpoint of the critical region

To perform a hypothesis test we will typically use the following steps

1) State H_a and H_0
2) Specify critical region
3) Compute test statistic
4) Make decision
5) Interpret the results

<u>Test of Hypothesis for μ with known σ</u>

When conducting a hypothesis test for μ with known σ, the sampling distribution is a standard normal distribution which means that our test statistic is a *Z*. The test statistic should look familiar. It is given below.

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (This is our test statistic when conducting a test for μ with known σ)

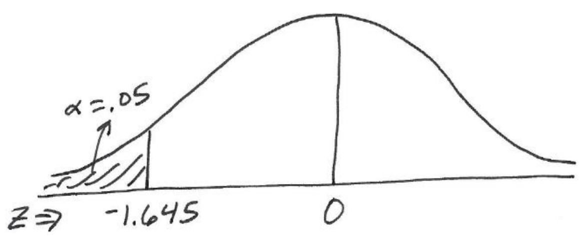
Example

It is claimed that the mean score for elementary education majors on a test of mathematical competency is less than 35. Test the claim at $\alpha = .05$. Assume $n = 165$, $\bar{x} = 32.91$, and $\sigma = 9.5$.

It may be beneficial to look at the steps of a hypothesis test given above. The steps will be number accordingly.

1) $H_a: \mu < 35$
   $H_0: \mu \geq 35$

2) This is a Z since we are doing a hypothesis test for μ with known σ. The critical region depends on H_a (since we are testing <, the critical region is to the left) and α (the area is α). I recommend you draw a picture of the critical region.



Reject H_0 if $Z \leq -1.645$

3) To calculate the test statistic, just plug in the appropriate values. For the parameter, plug in what you are comparing the parameter to (here we will plug in $\mu = 35$).

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{32.91-35}{9.5/\sqrt{165}} = -2.83$$

4) Reject H_0 (since $-2.83 \leq -1.645$, the test statistic is in the critical region)

5) Conclude with 95% confidence that the mean score for elementary education majors on the test of mathematical competency is less than 35.

The p-value Approach to a Test of Hypothesis

The p-value approach is an alternative way to make the decision in a hypothesis test. This approach is commonly used when using software to complete a hypothesis test since the p-value will be reported in the analysis. We will focus on the traditional approach of a hypothesis test, but you are responsible for the p-value approach in the case of estimating μ with known σ. Always use the traditional approach unless specifically asked to give the p-value. If you take additional coursework in statistics, you will likely use software so it is important to know what the p-value represents.

The choice of α in a hypothesis test is subjective. You might choose $\alpha = .05$ and I might choose $\alpha = .01$. The smaller α is, the smaller the rejection region. Thus, the smaller α is the harder it is to Reject H_0.

p-value – the smallest value of α such that H_0 would have been rejected (it is called the p-value since it's a probability value, the probability that corresponds to the test statistic)

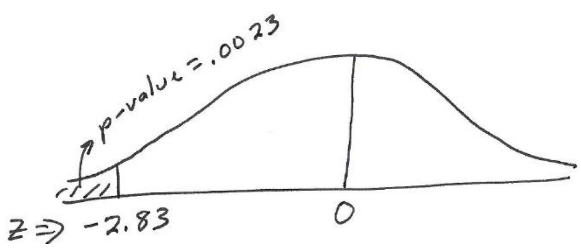
Rules for making decision based on a p-value
If p-value $\leq \alpha$, Reject H_0
If p-value $> \alpha$, Fail to reject H_0

To perform a hypothesis test using the p-value approach
1) State H_a and H_0
2) Compute test statistic
3) Identify p-value
4) Make decision
5) Interpret the results

<u>Example</u>
Go back to the mathematical competency of elementary education majors.

We already stated the hypotheses and calculated the test statistic,
$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{32.91-35}{9.5/\sqrt{165}} = -2.83$$

To calculate the p-value, use we must use the z-table and find the area to left of the test statistic. We find the area to the left since we are testing less than. Picture of this calculation is below.



Thus, p-value $= P(z \leq -2.83) = .5 - .4977 = .0023$

Reject H_0 (since $.0023 \leq .05$, the p-value $\leq \alpha$)

Notice this is the same decision as with the other method as it will always be. If you compare the picture here to the picture using the traditional method, you can see how this works. In the traditional approach, we are comparing the Z values and in the p-value approach we are comparing the areas. Notice if the test statistic is in the critical region then the p-value will be less than or equal to α. Therefore, the outcome will be the same using the two methods.

Test of Hypothesis for μ with unknown σ

The same issue arises in hypothesis tests for μ as was discussed in confidence intervals. The test statistic $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ includes σ so it must be estimated by s yielding $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ as our test statistic. Notice the only change in the formula is σ being replaced by s which means the standard normal distribution becomes a t-distribution. This is the more realistic case since we will not know the population standard deviation when estimating the population mean.

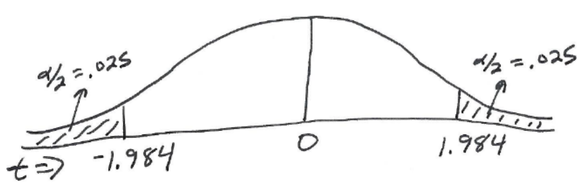
$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ (This is our test statistic when conducting a test for μ with unknown σ)

Example

A consumer protection agency wants to prove that packages of Post Grape Nuts have an average weight that is not 24 oz. Test this assuming $\alpha = .05$, $n = 100$, $\bar{x} = 23.94$, and $s = .13$.

1) $H_a: \mu \neq 24$
   $H_0: \mu = 24$

2) This is a t since we are doing a hypothesis test for μ with unknown σ. The critical region depends on H_a (since we are testing $\neq$, the critical region is divided in half and put in both tails) and α (the area is α). When the exact degrees of freedom are not available, pick the closest option. Remember that degrees of freedom is $n - 1$.



   Reject H_0 if $t \leq -1.984$ or if $t \geq 1.984$

3) $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{23.94-24}{.13/\sqrt{100}} = -4.615$

4) Reject H_0 (since $-4.615 \leq -1.984$, the test statistic is in the critical region)

5) Conclude with 95% confidence that packages of Post Grape Nuts have a population mean weight not equal to 24 oz.

Give a 95% confidence interval using the Post Grape Nuts information.

$E = t\left(\frac{s}{\sqrt{n}}\right) = 1.984\left(\frac{.13}{\sqrt{100}}\right) = .026$
$\bar{x} \pm E$
$(23.94 - .026, 23.94 + .026)$
$(23.914, 23.966)$

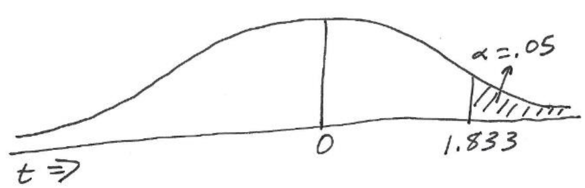
We are 95% confident that the population mean weight of Post Grape Nuts packages is between 23.914 and 23.966 oz.

Notice the consistency between the confidence interval and hypothesis test
If α is the same, then the error in a two tailed hypothesis test is set up the same as it is in a confidence interval. Therefore, in these cases the conclusions will always be consistent. Notice that if we are confident that μ is between 23.914 and 23.966, then we are also confident that $\mu \neq 24$. Obviously, 24 is not in the interval. In these cases you can actually tell the conclusion of the hypothesis test by looking at the confidence interval.

<u>Example</u>
It is desired to prove that the population mean weight of metal components produced by a process is greater than 4.5 oz. Assume $\alpha = .05$, $n = 10$, $\bar{x} = 4.59$, and $s = .504$.

1) $H_a: \mu > 4.5$
   $H_0: \mu \leq 4.5$

2) This is a *t* since we are doing a hypothesis test for μ with unknown σ. The critical region depends on H_a (since we are testing >, the critical region is to the right) and α (the area is α).



   Reject H_0 if $t \geq 1.833$

3) $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.59-4.5}{.504/\sqrt{10}} = .565$

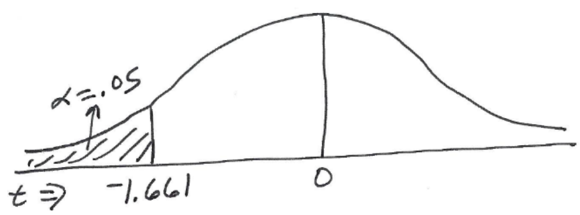
4) Fail to Reject H_0 (since $.565 < 1.833$, the test statistic is not in the critical region)

5) There is not sufficient evidence to prove that the population mean weight of the metal components is greater than 4.5 oz.

Example
A marketing consultant wants to prove that the population mean household income in Lexington is less than \$50,000. Assume $\alpha = .05$, $n = 100$, $\bar{x} = 40{,}571$, and $s = 8316$.

1) $H_a: \mu < 50{,}000$
   $H_0: \mu \geq 50{,}000$

2) This is a *t* since we are doing a hypothesis test for μ with unknown σ. The critical region depends on H_a (since we are testing <, the critical region is to the left) and α (the area is α).



Reject H_0 if $t \leq -1.661$

3) $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{40{,}571-50{,}000}{8316/\sqrt{100}} = -11.338$

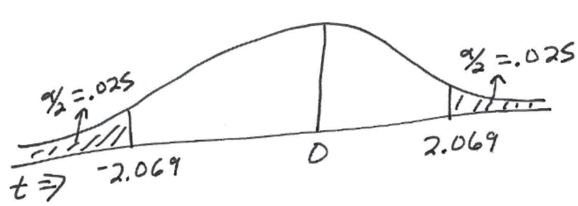
4) Reject H_0 (since $-11.338 \leq -1.661$, the test statistic is in the critical region)

5) Conclude with 95% confidence that the mean household income in Lexington is less than \$50,000.

Example
A farmer's pigs have experienced an average weight gain of 200 lbs over a fixed time. He is experimenting with a new feeding technique and wants to know if the average weight gain will change in either direction. Assume $\alpha = .05$, $n = 24$, $\bar{x} = 197.2$, and $s = 9.8$.

1) $H_a: \mu \neq 200$
   $H_0: \mu = 200$

2) This is a *t* since we are doing a hypothesis test for μ with unknown σ. The critical region depends on H_a (since we are testing $\neq$, the critical region is divided in half and put in both tails) and α (the area is α).



Reject H_0 if $t \leq -2.069$ or if $t \geq 2.069$

3) $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{197.2-200}{9.8/\sqrt{24}} = -1.400$

4) Fail to Reject H_0 (since $-2.069 < -1.400 < 2.069$, the test statistic is not in the critical region)

5) There is not sufficient evidence to prove the average weight gain of the pigs is different from 200 lbs using the new feeding technique.

Test of Hypothesis for *p* (when *n* is large)

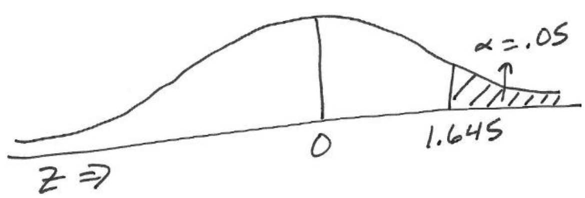
When conducting a hypothesis test for *p* when *n* is large (at least 30), the sampling distribution is a standard normal distribution which means that our test statistic is a *Z*. The test statistic should look familiar. It is given below.

$Z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}}$ (This is our test statistic when conducting a test for *p*)

Example
The current treatment for a type of cancer produces remission 20% of the time. An investigator wishes to prove that a new method is better. Suppose 26 of 100 patients go into remission using the new method. Assume $\alpha = .05$.

1) $H_a: p > .20$
   $H_0: p \leq .20$

2) This is a *Z* since we are doing a hypothesis test for *p* with large *n*. The critical region depends on H_a (since we are testing >, the critical region is to the right) and α (the area is α).



Reject H_0 if $Z \geq 1.645$

3) $\hat{p} = \frac{26}{100} = .26$
   $Z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.26-.20}{\sqrt{\frac{(.20)(.80)}{100}}} = 1.5$

4) Fail to reject H_0 (since $1.5 < 1.645$, the test statistic is not in the critical region)

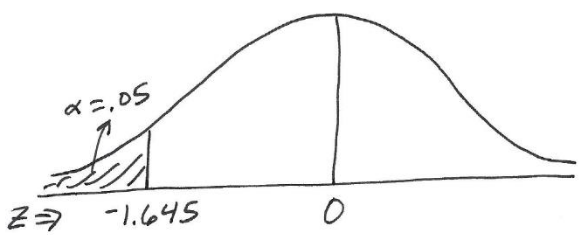
5) There is not sufficient evidence to conclude the new method is better.

Example

Do less than 50% of people prefer Murray's Vanilla Wafer's when compared to other brands? Suppose 42 of the 250 chose Murray's. Assume $\alpha = .05$.

1) $H_a: p < .50$
   $H_0: p \geq .50$

2) This is a Z since we are doing a hypothesis test for p with large n. The critical region depends on H_a (since we are testing <, the critical region is to the left) and α (the area is α).

   Reject H_0 if $Z \leq -1.645$

3) $\hat{p} = \frac{42}{250} = .168$

   $Z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.168-.50}{\sqrt{\frac{(.50)(.50)}{250}}} = -10.499$

4) Reject H_0 (since $-10.499 \leq -1.645$, the test statistic is in the critical region)

5) Conclude with 95% confidence that less than 50% of people prefer Murray's Vanilla Wafer's when compared to other brands.