We will use samples to find evidence that an assumption about the population is *not* correct.

General Education Statistics
Class Notes
Introduction to Hypothesis Testing (Section 10.1)

In a criminal trial, the defendant is assumed to be innocent. Evidence is gathered by the prosecutor to prove “beyond a reasonable doubt” that the defendant is, in fact, guilty. If not enough evidence is presented, the defendant is found to be “not guilty”. Notice she is never declared “innocent”, just not guilty. Here, we do nearly the same with sample evidence to test if something is true of the population.

Suppose a friend bets you that he can roll a six-sided die four times and get a 6 each and every time. Would you pay him $10 if he could? Well, the probability that he could do that is

$$
\begin{aligned}
&P(\text{four 6's in a row}) \\
&= P(6 \textit{ and } 6 \textit{ and } 6 \textit{ and } 6) \\
&= \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} \\
&= \frac{1}{1296} \\
&\approx .00077
\end{aligned}
$$

If he managed this feat, what might you conclude? Is he a true friend? Or a shyster? A bamboozler? A dirty rotten fiend?

Or maybe he’s just lucky? How can we tell?

Now, you may never know if he cheated you or not, but you can say that, statistically speaking, the result is highly unlikely.
Either your friend is a cheat (and the probability of his die turning up a 6 is not really 1/6) or he beat the odds and managed a very, very unlikely event. After you punch him, consider this.

We make an assumption about reality (like the probability of his die turning up a 6 is 1/6). We look at sample evidence (like him performing this cool trick) to see if it contradicts our assumption. If you feel the evidence contradicts the assumption, then maybe it’s chance or maybe the assumption is wrong (and he has a loaded die; now punch him again).

Some definitions…

Definition: A **hypothesis** is a statement regarding a characteristic of one or more populations. In this chapter, we look at hypotheses regarding a single population parameter.

Definition: Hypothesis testing is a procedure, based on sample evidence and probability, used to test a hypothesis.

Steps in Hypothesis Testing:

1. Make a statement regarding the nature of the population.
2. Collect evidence (sample data) to test the statement.
3. Analyze the data to assess the plausibility of the statement.

This assumes we do *not* know the population information. If we do, there is *no* need to sample.

Definitions: The **null hypothesis**, denoted H_0 ("H-naught"), is a statement to be tested. The null hypothesis is a statement of no change, no effect, or no difference and *is assumed true until evidence indicates otherwise.*

The **alternative hypothesis**, denoted H_1 ("H-sub one"), is a statement that we are trying to find evidence to support.

Three Different Hypothesis Tests:

In this chapter, there are three ways to set up the null and alternative hypotheses.

1. Equal versus *not* equal hypothesis (**two-tailed test**)

   H_0: parameter $=$ some value

   H_1: parameter $\neq$ some value

The null hypothesis is assumed to be true until we have evidence it is *not*. We seek evidence that supports the alternative hypothesis.

2. Equal versus less than (**left-tailed test**)

   H_0: parameter $=$ some value

   H_1: parameter $<$ some value

3. Equal versus greater than (**right-tailed test**)

   H_0: parameter $=$ some value

   H_1: parameter $>$ some value

Notice each starts off with the same null hypothesis.

expl 1: Determine the null and alternative hypotheses. Label them as H_0 and H_1. Use the appropriate abbreviations for the population parameters (μ , σ , or p). Tell whether the test is a two-tailed, left-tailed, or right-tailed test.

a.) Federal law requires that a jar of peanut butter that is labeled 32 ounces contains at least 32 ounces. A consumer advocate suspects that a certain company is shorting their customers.

b.) A bottling plant bottles jars of peanut butter in 32 ounce jars. A quality-control engineer is concerned that there might be too little or too much peanut butter in each jar.

The test needed may differ based on the researcher's point of view.

c.) The standard deviation in the pressure required to open a certain valve is known to be $\sigma = 0.8$ psi. Due to changes in the manufacturing process, the quality-control manager feels that the pressure has been embiggened (increased).

d.) A Gallup poll in 2008 showed that 80% of Americans felt satisfied with their personal lives. A researcher in 2018 wonders if this percentage has changed, but she does not have a guess if it has decreased or increased.

Remember, because we will be using sample data, we cannot say *for sure* that our conclusion about the population is absolutely, positively correct. However, using good statistics can tell us if the null hypothesis is supported or not.

Type I and Type II Errors:

Since there is room for making an incorrect conclusion, we need to talk about the probability of making certain errors. Consider the four possible outcomes for a hypothesis test.

Four Outcomes from Hypothesis Testing:

1. Reject the null hypothesis when the alternative hypothesis is true.
   This decision would be correct.

2. Do *not* reject the null hypothesis when the null hypothesis is true.
   This decision would be correct.

false positive

3. Reject the null hypothesis when the null hypothesis is true.
   This decision would be incorrect. This type of error is called a **Type I error.**

false negative

4. Do *not* reject the null hypothesis when the alternative hypothesis is true.
   This decision would be incorrect. This type of error is called a **Type II error.**

To keep these outcomes straight, here they are in table form.

		(Unknown) Reality	
		H_0 is true	**H_1 is true**
Conclusion	**Do Not Reject H_0**	Correct conclusion	Type II error
	Reject H_0	Type I error	Correct conclusion

expl 2: Think back to the example of a criminal trial. The null hypothesis is that the defendant is innocent. The alternative hypothesis is that she is guilty (*not* innocent). Which conclusion is a Type I error and which is a Type II error?

a.) The defendant is really guilty but is *not* convicted.

We either reject H_0 and convict; or do *not* reject H_0 and do *not* convict.

b.) The defendant is really innocent but is convicted.

expl 3: Consider these prior scenarios with hypothesis tests in place. Explain what would Type I and Type II errors look like.

a.) Federal law requires that a jar of peanut butter that is labeled 32 ounces contains at least 32 ounces. A consumer advocate suspects that a certain company is shorting their customers.

H_0: $\mu = 32$ ounces

H_1: $\mu < 32$ ounces

Type I error: Reject the null hypothesis when it's true.

Type II error: Do *not* reject the null hypothesis when it's *not* true.

Type I error:

Type II error:

b.) A Gallup poll in 2008 showed that 80% of Americans felt satisfied with their personal lives. A researcher in 2018 wonders if this percentage has changed, but she does not have a guess if it has decreased or increased.

H_0: $p = .80$

H_1: $p \neq .80$

Type I error:

Type II error:

The Probability of Type I and II Errors:

We never know if our conclusion is correct (because we do *not* take a census). However, we can find the probabilities of making Type I and Type II errors. We will use the following notation.

$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when it's true})$

$\beta = P(\text{Type II error}) = P(\textit{not} \text{ rejecting } H_0 \text{ when } H_1 \text{ is true})$

Greek letters "alpha" and "beta"

Definition: The **level of significance**, $\boldsymbol{\alpha}$, is the probability of making a Type I error. The level of significance is chosen by the researcher before the sample data is collected.

The connection between α and β:

The choice of the level of significance depends on the consequences of making a Type I error. If the consequences are severe, the level of significance should be small, like α = 0.01. Otherwise, we will use α = 0.05 or α = 0.10. With regard to criminal trials, we do *not* want to send innocent people to jail, so we try to make that probability (Type I error) low.

However, reducing the probability of making a Type I error increases the probability of making a Type II error. So if we do *not* want innocent people sent to jail, we accept that some guilty will go free.

Stating Conclusions in Hypothesis Tests:

Just as a defendant is *not* proclaimed innocent (they are *not guilty*), we do *not* whole-heartedly *accept* H_0.

If we end up *not* rejecting the null hypothesis, we will say something like "there is *not* sufficient evidence to conclude that the population parameter is as it's stated in the alternative hypothesis".

If we end up rejecting the null hypothesis, we will say something like "there is sufficient evidence to conclude that the population parameter is as it's stated in the alternative hypothesis".

Definition: Statistically significant: When observed results are *unlikely* under the assumption that the null hypothesis is true, we say the result is **statistically significant**. The sample result was so far from the assumed parameter's value that the difference is *significant* enough for us to say the assumption (about the population's parameter) was likely wrong. When results are found to be statistically significant, **we *reject* the null hypothesis**.

Optional worksheet: Statistically Significant:
We will practice rewording statistical significance using the context of studies.

expl 4: For each scenario, write out H_0 and H_1. Then state the conclusion appropriately.

a.) A researcher wants to test if the percentage of Americans who have tried and like fried Twinkies is more than 20%. They sample Americans to perform a hypothesis test. They do *not* reject the null hypothesis.

b.) According to the *CTIA – The Wireless Association*, the mean monthly revenue per cell phone was $48.79 in 2014. A researcher suspects that the mean monthly revenue per cell phone is different today. They sample and perform a hypothesis test. They reject the null hypothesis.