Chapter 9: Hypothesis Tests

9.1 Null and Alternative Hypotheses

A hypothesis test is a process that uses ______________ statistics to test a claim about the value of a _______________________ parameter.

A ___________hypothesis, H_o, is a statistical hypothesis that contains a statement of equality such as $\leq, \geq, =$

The____________________hypothesis, H_a, is the complement of the null hypothesis. It is a statement that must be true if H_o is false and it contains a statement of strict inequality such as $<, >, \neq$.

Example: State the null and alternative hypothesis for each situation.

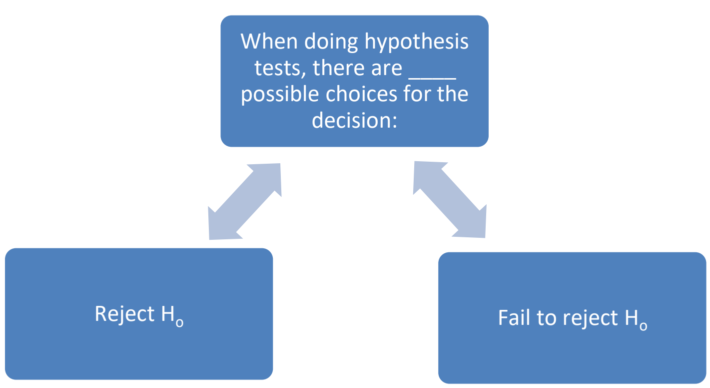
a) The standard deviation for the number of movies an adult watches per year in a theater is at least 1.3.

b) A study claims that National Football League stadiums have an average of 71,212 seats.

Notice the null and alternative hypotheses are claims about the _______________________.

c) The probability that U.S. greenhouse gas emissions come from nitrous oxide is .051. You believe that the probability is actually less than .051.

9.2 Outcomes and the Type I and Type II Errors



Because there are two possible choices for the decision of a hypothesis test, there are two types of errors that can be made when conducting a study:

Type I Error: H_o is true, but you _______________ H_o.

Type II Error: H_o is false, but you ________________________ H_o.

Example: A study shows that at most 15% of Hancock athletes will play for a university

a) State the null and alternative hypothesis.

b) Describe what a Type I error is for this situation.

c) Describe what a Type II error is for this situation.

9.3 Distribution Needed for Hypothesis Testing

Particular distributions are associated with hypothesis testing. (More on this in 9.5)

Perform tests of a **population mean** using a **________________ distribution** or a **___________________________ distribution**. (Remember, use a Student's *t*-distribution when the population **standard deviation** is unknown and the distribution of the sample mean is approximately normal.)

Perform tests of a **population proportion** using a _______________ distribution (usually *n* is large).

9.4 Rare Events, the Sample, Decision, and Conclusion

When doing hypothesis tests, you are using your knowledge about a sample to predict a __________________________________. Since there is a variation in samples, you can sometimes reject H_o when it is true. To account for this variation, we allow for a ___________________________ in the data.

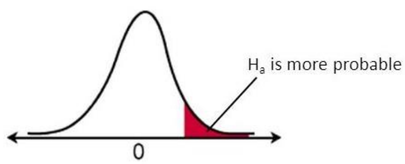
In a hypothesis test, the <u>level of significance</u>, α , is your maximum allowable probability of making a Type ___ error. i.e. P(Type __ Error)= α .

Example: If an innocent person goes to trial 100 independent times and α = .05, then the person will be found guilty 5 times. (There is a 95% chance that the person will be found correctly innocent.)

Note: If a level of significance is not specified, it is assumed to be 5%.

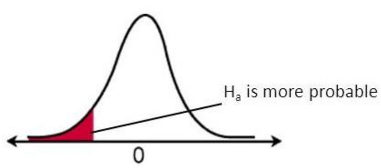
Using p-value method for doing a hypothesis test:

If H_o is true, a <u>P-value</u> of a hypothesis test is the ______________________ of obtaining a sample statistic with a value as extreme as or more extreme than the one determined from the sample data.



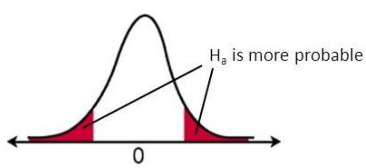
Right-tail test

H_a: $\mu >$ value



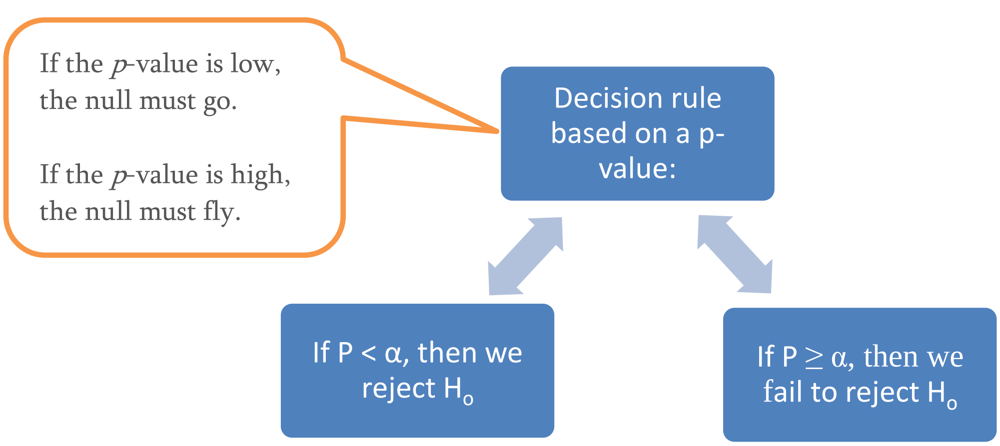
Left-tail test

H_a: $\mu <$ value



Two-tail test

H_a: $\mu \neq$ value



Example: It's a Boy Genetics Labs claim their procedures improve the chances of a boy being born.

H_0:
H_a:

Let's say we use a level of significance of 1% and get a p-value of 0.025.

Interpret the results and state a conclusion in simple terms.

9.5 Full Hypothesis Test Examples

<u>Outline of a Hypothesis Test Using P-values</u>

1. State H_0, H_a, α.
2. Determine what standardized test statistic (which test) to use for the problem. (z, t, χ^2)
3. Verify normality.
4. Find the corresponding P-value using ClassCalc.
5. Decide to reject or fail to reject H_0.
6. Interpret the decision in the context of the original claim.

Hypothesis Testing for the Mean (σ known)

z test for a population mean μ

The z-test can be used when ___________________________________ is known.

Verify normality: either states it is normally distributed or $n \geq 30$

The test statistic is $\overline{x}$. The standardized test statistic is $z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$.

Example: Suppose T.K. says that when he works out, he squats at least 400 pounds, on average. Several of his teammates do not believe him. To persuade his teammates that he is right, T.K. decides to do a hypothesis test. He has a teammate record his squat weight in his next 10 workouts and gets a mean squat weight of 375 pounds. T.K. knows from doing hundreds of squat workouts that the standard deviation for his squat weight is 50 pounds. Assume the distribution of all his squat weights is normal.

Mean Z Test on ClassCalc:

- Stat → tests → z-test
- Enter Stats (μ is the mean used in the hypotheses) or Data
- Choose inequality symbol per H_a

Hypothesis Testing for the Mean (σ unknown)

t-test for a population mean μ

The t-test is used when ___________________________________ is unknown.

Verify normality: either states it is normally distributed or $n \geq 30$

The test statistic is $\overline{x}$. The standardized test statistic is $t = \dfrac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}}$ where d.f.= n – 1.

Example: The government claims the average student loan debt is $39,381. A SRS of 42 students who just graduated from college finds that their average student loan debt is $47,763 with a sample standard deviation of $983. Use $\alpha = .01$ to test the claim.

Mean T Test on ClassCalc:

- Stat → tests → t-test
- Enter Stats (μ is the mean used in the hypotheses) or Data
- Choose inequality symbol per H_a

Example: The following data shows a SRS of paddle times (in minutes) for the members of a kayak team over a three mile course. The team members paddle with Werner paddles. Conduct a hypothesis test at a 5% level of significance to determine if paddling with Werner paddles gives an average time that is faster than 27 minutes. Assume that the data is approximately normal.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13
Werner paddle time	35	24	33	23	27	29	22	21	25	25	30	32	26

Hypothesis Testing for Proportions

z-test for a population proportion p

The z-test can be used when a binomial distribution is given.

To verify normality: $np \geq 5$ and $nq \geq 5$

The test statistic is $\hat{p}$. The standardized test statistic is $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$.

Example: Serena Williams claims if she wins the first point in a match, there is at least a 90% chance she will win the match. A random sample of 25 of her matches finds that this was true for 19 of them. Test her claim at the 5% level.

Proportion Z Test on ClassCalc:

- Stat → tests → 1-prop-z-test
- Enter Stats (P_0 is the proportion used in the hypotheses)
- Choose inequality symbol per H_a

Chapter 10: Hypothesis Testing with Two Samples

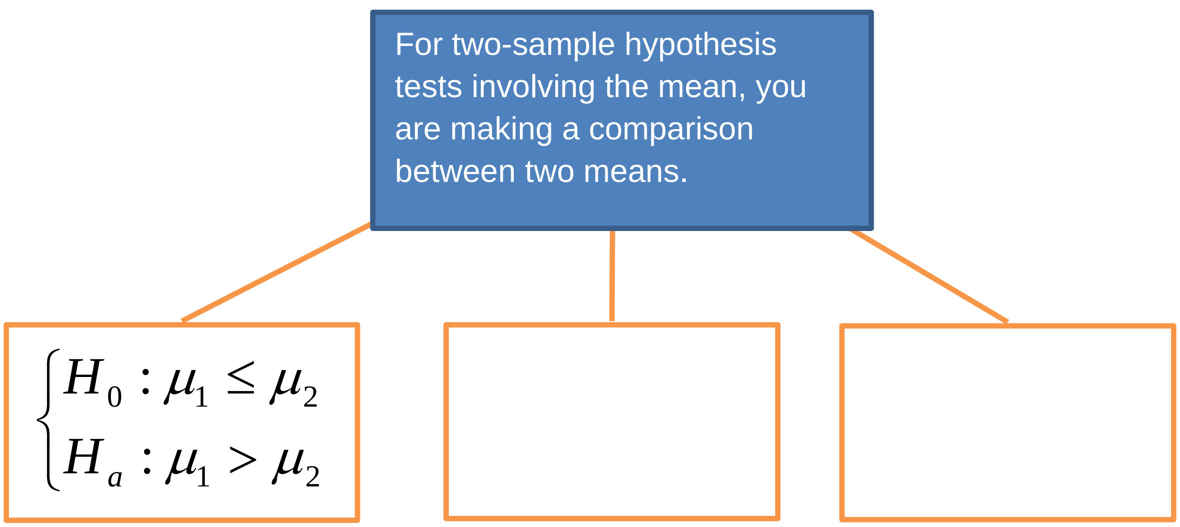
10.1 Two Population Means with Unknown Standard Deviations

Two samples are independent if the sample selected from one population is ________ related to the sample selected from the second population.

Example:

Two samples are ________________ if each member of one sample corresponds to a member of the other sample. Dependent samples are also called **paired samples** or matched samples.

Example:



<u>Two-sample t-test for the difference between means</u>

The two-sample t-test can be used when:

- both samples are random
- both populations are normal or $n_1 \geq 30$ and $n_2 \geq 30$
- both samples are independent
- σ_1**,** σ_2 **are unknown**

The test statistic is $\bar{x}_1 - \bar{x}_2$.

The standardized test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ where d.f.=smlr($n_1 - 1, n_2 - 1$) and $\sigma_1^2 \neq \sigma_2^2$ (Not pooled)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ where d.f.= $n_1 + n_2 - 2$ and $\sigma_1^2 = \sigma_2^2$ (Pooled)

Example: Some people believe there are better hitters overall in the National League than in the American League in the MLB. You took a sample of 42 NL players and found their mean batting average this season was 0.252 with a standard deviation of 0.06. The mean batting average for a sample of 35 AL players this season is 0.247 with a standard deviation of 0.08. Test the claim that the batting averages of the National League are better than that of the American League at a 10% level of significance.

Two sample t test for a difference between means on ClassCalc:

- Stat → tests → 2-samp-t-test
- Enter in stats or data

Example: Dr. Colin Campbell, one of the doctors featured on the documentary Forks Over Knives, claims that the average cost of food for a plant-based diet is significantly less than a diet that includes meat. (Dr. Campbell's patients have diabetes and other severe health problems and are looking for treatment other than surgery, so Dr. Campbell advocates for a plant-based diet.) In a random sample of 32 of his patients who have not yet switched to a plant-based diet, Dr. Campbell found that they spend an average of $425 per month on food and diabetes medicine with a sample standard deviation of $65. In a random sample of 52 of his patients that are on a plant-based diet, the average cost of food per month is $300 with a sample standard deviation of $63. Does the data support Dr. Campbell's claim?

Group Work 10.1: Complete this problem with your group and turn it in on Canvas.

Example: Despite fighting hard for parity, women soccer players believe they are paid a lot less than male soccer players. The following are randomly selected players' salaries for 2022:

NWSL	**MLS**
Sydney Leroux – $70,000	Gonzalo Higuain — $5,793,750
Alex Morgan – $450,000	Spencer Richey – $132,000
Elizabeth Ball – $46,400	Brad Smith – $550,000
Victoria Pickett – $200,000	Javier Hernandez – $6,000,000
	George Acosta – $65,500

Test the women soccer players' claim at a 1% level of significance.

10.2 Two Population Means with Known Standard Deviations

Two-sample z-test for the difference between means

The two-sample z-test can be used when:

- both samples are random
- both populations are normal or $n_1 \geq 30$ and $n_2 \geq 30$,
- both samples are independent
- σ_1, σ_2 **are known**.

The test statistic is $\bar{x}_1 - \bar{x}_2$.

The standardized test statistic is $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$.

Example: A baseball fan wanted to know if there is a difference between the number of games played in a World Series when the American League won the series versus when the National League won the series. From 1922 to 2012, the population standard deviation of games won by the American League was 1.14, and the population standard deviation of games won by the National League was 1.11. Of 39 randomly selected World Series games won by the American League, the mean number of games won was 5.76. The mean number of 47 randomly selected games won by the National League was 5.42.

Two sample z test for a difference between means on ClassCalc:

- Stat → tests → 2-samp-z-test
- Enter in stats or data

10.3 Comparing Two Independent Population Proportions

two-sample z-test for the difference between proportions

The two-sample proportion z test can be used when:

- both samples are randomly selected
- both samples are independent
- both samples are normal. Normality is verified by showing that $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, $n_2\bar{q} \geq 5$.

The test statistic is $\hat{p}_1 - \hat{p}_2$.

The standardized test statistic is $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ where $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$.

Example: A one month study is done on two groups of athletes. 220 of the randomly selected athletes did two workouts a day. The study shows that 182 put on muscle mass. 154 of the randomly selected athletes did one workout a day but got an extra hour of sleep instead of the second workout. 116 of these athletes put on muscle mass. Is there enough evidence to show that the athletes who work out twice a day have a higher chance of putting on more muscle than those who got more sleep?

Two sample z test for a difference between proportions on ClassCalc:

- Stat → test → 2-prop-z-test
- Enter in stats

10.4 Matched or Paired Samples (Dependent Samples)

t-test for the difference between means for paired data

The t-test for paired data can be used when:

- a sample is randomly selected from each population
- both populations are normal or the number of pairs of data is at least 30
- each member of the first sample is paired with a member of the second sample

The test statistic is $\bar{d}$, the average of the differences between data values.

The standardized test statistic is $t = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$ where d.f.= n – 1.

Example: A golf instructor is interested in determining if her new technique for improving players' golf scores is effective. She takes four new students. She records their 18-hole scores before learning the technique and then after having taken her class. She conducts a hypothesis test. Assume a normal distribution.

	Player 1	**Player 2**	**Player 3**	**Player 4**
Mean score before class	83	78	93	87
Mean score after class	80	80	86	86

t test for difference between means for PAIRED data on ClassCalc:

- Make lists A = [enter data] and B = [enter data]
- Set C = B – A
- Stat → test → t-test
- TTest(0, [C])
- Choose inequality based on H_a

Who Has the Higher Average Vertical Jump—

Basketball Players or Football Players?

How could we find out?

Basketball

Name	Vertical Reach	Flag Height	Vertical Jump

Average Vertical Jump:

Football

Name	Vertical Reach	Flag Height	Vertical Jump

Average Vertical Jump:

How confident do we feel in our results? What are some pitfalls of our research?

How could we determine if our sample data is a fluke? In other words, if we repeated our experiment, we would obtain results as extreme or more extreme than what we first got?

We are going to use a **two-sample t test.**

Think back to what you know about a one-sample t test. Why do you think we chose a t test and not a z test?

Hypotheses:

Graph:

P-value:

Conclusion:

What are some problems with our experiment design?

How could we have improved our experiment design?