Chapter 9: Inferences from Two Samples

Section 9.1: Two Proportions Stat 50

Introduction Scenario:
Think about whether you plan to vote in the next presidential election in 2020. Let's say I split the responses into two groups: Females and Males
Notice: There are **TWO** proportions here: $p_1 =$ $p_2 =$
QUESTION: Is there a statistically significant difference?
We don't care what the proportions actually are, we care about whether they're the __________ or not.

** In this section (9.1), we will only deal with independent samples with ____________________ variables.

The Logic

If $p_1 = p_2$ then $p_1 - p_2 =$ ______.

So we will first collect ________________________________ and find ______ and ________

Then we will see if $\hat{p}_1 - \hat{p}_2$ is anywhere close to _______.

Recall the Logic: Say $H_0: p_1 = p_2$ and $H_1: p_1 < p_2$. If we get…

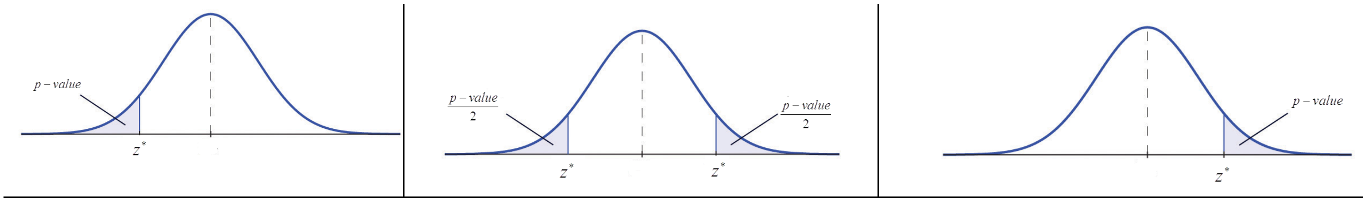
- $\hat{p}_1 = \hat{p}_2$ ___________________ H_0
- $\hat{p}_1 > \hat{p}_2$ ___________________ H_0
- $\hat{p}_1 < \hat{p}_2$ by "a little" ___________________ H_0
- $\hat{p}_1 < \hat{p}_2$ by "a lot" then we ___________________ H_0

Steps for Hypothesis Test Regarding the Difference between p_1 and p_2		
Pre-Step: *Check Requirements* ● The samples are simple random and _____________	● At least five success $n\hat{p} \geq 5$ and five failures $n\hat{q} \geq 5$ for each sample.	
Step 1: *Hypotheses*		
$\boldsymbol{H_0: p_1 = p_2}$ $\boldsymbol{H_1: p_1 \neq p_2}$	$\boldsymbol{H_0: p_1 = p_2}$ $\boldsymbol{H_1: p_1 < p_2}$	$\boldsymbol{H_0: p_1 = p_2}$ $\boldsymbol{H_1: p_1 > p_2}$
Step 2: *Level of Significance* If it's not given, then use ________________. Choice depends on seriousness of making Type ___ error.		
Step 3: *Test Statistic* $z^* = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}}$ where the point estimate is $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$		

Step 4: *Find a Critical Value or P-Value*

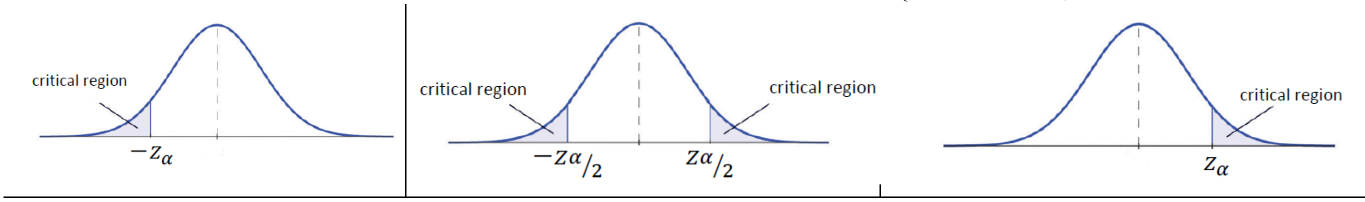
P-VALUE METHOD **DECISION**

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } P\text{-value} \le \alpha \\ \textit{Fail to Reject } H_0 \sim \text{ if } P\text{-value} > \alpha \end{cases}$



CRITICAL VALUE METHOD **DECISION**

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } z^* \text{ lies in the critical region} \\ \textit{Fail to Reject } H_0 \sim \text{ if } z^* \text{ doesn't lie in the critical region} \end{cases}$



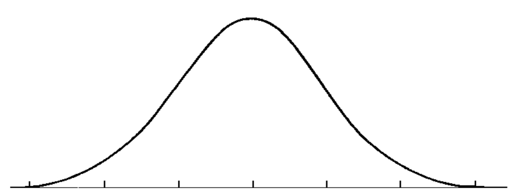
Step 5: Write a *CONCLUSION* either rejecting or failing to reject H_0

Ex: An insurance company is concerned that men are more likely to speed than women. In a sample of 500 randomly selected women, 27 have been ticketed for speeding in the last year. In a sample of 250 randomly selected men, 26 have been ticketed for speeding in the last year. Use a 0.05 significance level to test the insurance company's claim that the percentage of women ticketed for speeding is less than the percentage of men.

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region



Decision about Null Hypothesis

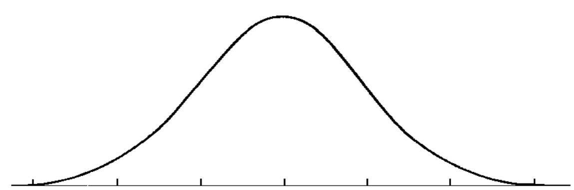
Conclusion

Ex: In clinical trials of the anti-inflammatory drug Inflaminex, adult and adolescent allergy patients were randomly divided into two groups. Some patients received 500mcg of Inflaminex, while some patients received a placebo. Of the 2103 patients who received Inflaminex, 520 reported bloody noses as a side effect. Of the 1671 patients who received the placebo, 368 reported bloody noses as a side effect. Is there significant evidence to conclude that the proportion of Inflaminex users who experienced bloody noses as a side effect is greater than the proportion of the placebo group at the $\alpha = 0.01$ level of significance?

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region



Decision about Null Hypothesis

Conclusion

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ 2-PropZTest

(b) Enter
$$\begin{cases} x_1 \ / \ x_2 \ = \text{ number of success in sample \#1 / \#2} \\ n_1 \ / \ n_2 \ = \text{ size of sample \#1 / \#2} \\ p_1 \ \sim \text{ alternative hypothesis} \end{cases}$$

Confidence Interval for the Difference of Two Population Proportions

Alternative Forms: $\left(\hat{p}_1 - \hat{p}_2\right) - E < p_1 - p_2 < \left(\hat{p}_1 - \hat{p}_2\right) + E$ or $\left(\hat{p}_1 - \hat{p}_2\right) \pm E$

where the margin of error is given by $E = z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}$

Graphing Calculator (TI-83 or 84)

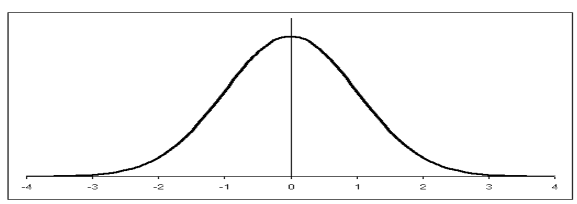
Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ 2-PropZInt

(b) Enter
$$\begin{cases} x_1 \;/\; x_2 \;=\; \text{number of successes in sample \#1 / \#2} \\ n_1 \;/\; n_2 \;=\; \text{size of sample \#1 / \#2} \\ C-level \;=\; \text{confidence level} \end{cases}$$

Ex: A study was conducted to test the effectiveness of a sweetener called xylitol in preventing ear infections in preschool children. In a randomized experiment, 159 preschool children took five daily doses of xylitol, and 46 of these children got an ear infection during the three months of the study. Meanwhile, 165 children took five daily doses of placebo syrup, and 68 of these children got an ear infection during the study. Construct a 90% confidence interval for the difference in the proportion of children that got ear infections for the control group and the xylitol group.

Find the point estimate (difference between sample proportions)

Determine critical value $z_{\alpha/2}$



Find margin of error

Construct confidence interval

Does it appear that the sweetener is effective at reducing ear infections?

Steps for a Hypothesis Test When Applied to testing μ_1 and μ_2

Pre-Step: *Check Requirements*

- The samples are ________________ and randomly obtained
- The values of the population standard deviation σ_1 and σ_2 are not known and unequal.
- Populations are normally distributed **OR** $n_1 >$ _____ **and** $n_2 >$ _____

Step 1: *Hypotheses*

$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 < \mu_2$	$H_1: \mu_1 > \mu_2$

Step 2: *Level of Significance*

If it's not given, then use ________________. Choice depends on seriousness of making Type ___ error.

Step 3: *Test Statistic*

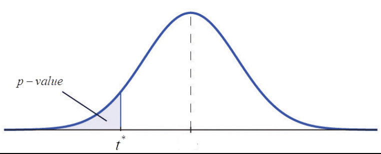
$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

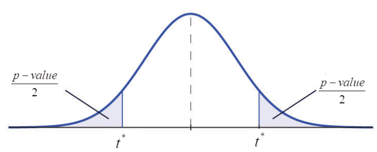
Step 4: *Find a Critical Value or P-Value*

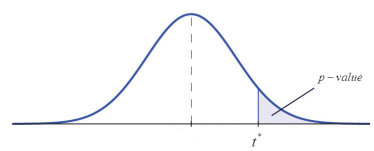
P-VALUE METHOD

DECISION

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } \textit{P-value} \le \alpha \\ \textit{Fail to Reject } H_0 \sim \text{ if } \textit{P-value} > \alpha \end{cases}$



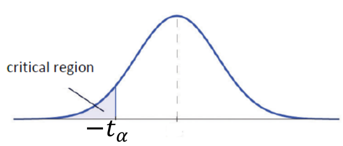


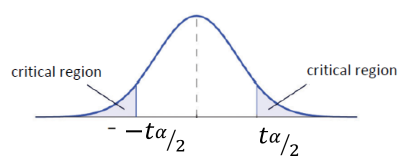


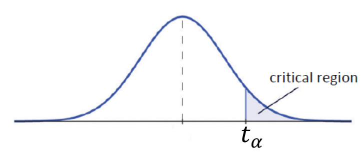
CRITICAL VALUE METHOD

DECISION

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } t^* \text{ lies in the critical region} \\ \textit{Fail to Reject } H_0 \sim \text{ if } t^* \text{ doesn't lie in the critical region} \end{cases}$







Step 5: Write a *CONCLUSION* either rejecting or failing to reject H_0

Graphing Calculator (TI-83 or 84) Instructions

Instructions: (a) STAT ⇒ TESTS ⇒ 2-SampTTest

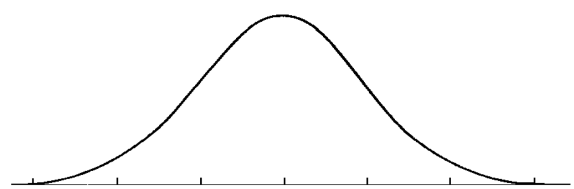
(b) Enter
- $\bar{x}_1$ / $\bar{x}_2$ = mean of sample #1 / #2
- s_1 / s_2 = standard deviation of sample #1 / #2
- n_1 / n_2 = size of sample #1 / #2
- μ_1 ~ alternative hypothesis (*not pooled*)

Ex: A study of zinc-deficient mothers was conducted to determine whether zinc supplements during pregnancy results in babies with increased weights at birth. 294 expectant mothers were given a zinc supplement, and the mean birth weight was 3214 grams with a standard deviation of 669 g. There were 286 expectant mothers who were given a placebo, and the mean weight was 3088 g with a standard deviation of 728 grams. Using a 0.01 significance level, is there sufficient evidence to support the claim that a zinc supplement does result in increased birth weights?

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region



Decision about Null Hypothesis

Conclusion

Ex: A professor at a large community college wanted to determine whether there is a difference in the means of final exam scores between students who were allowed to text in class and those who weren't. She believed that the mean of the final exam scores for the texting class would be lower than that of the non-texting class. Was the professor correct? She randomly selected 30 final exam scores from each group, and they are listed below.

Texting class
class

67.6	41.2	85.3	55.9	82.4	91.2	73.5	94.1	64.7	64.7
70.6	38.2	61.8	88.2	70.6	58.8	91.2	73.5	82.4	35.5
94.1	88.2	64.7	55.9	88.2	97.1	85.3	61.8	79.4	79.4

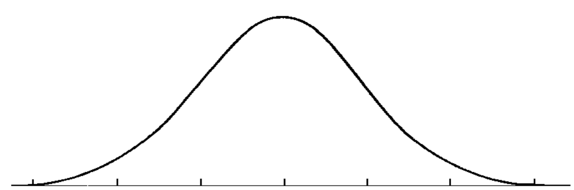
Non-texting class
class

77.9	95.3	81.2	74.1	98.8	88.2	85.9	92.9	87.1	88.2
69.4	57.6	69.4	67.1	97.6	85.9	88.2	91.8	78.8	71.8
98.8	61.2	92.9	90.6	97.6	100	95.3	83.5	92.9	89.4

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region



Decision about Null Hypothesis

Conclusion

Confidence Interval for the Difference of Two Population Means

Alternative Forms: $\left(\bar{x}_1 - \bar{x}_2\right) - E < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + E$ or $\left(\bar{x}_1 - \bar{x}_2\right) \pm E$

where the margin of error is given by $E = t_{\alpha/2} \cdot \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Graphing Calculator (TI-83 or 84) Instructions

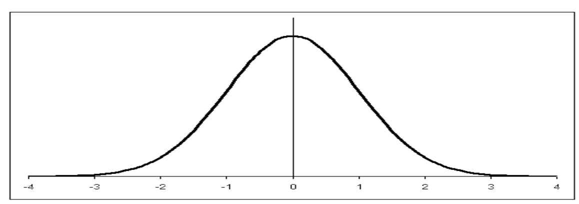
Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ 2-SampTInt

(b) Enter
- $\bar{x}_1 \,/\, \bar{x}_2$ = mean of sample #1 / #2
- $s_1 \,/\, s_2$ = standard deviation of sample #1 / #2
- $n_1 \,/\, n_2$ = size of sample #1 / #2
- $C-level$ = confidence level (*not pooled*)

Ex: The Gallup Organization wanted to investigate the time that American men and women spend hanging out with their friends. A random sample of 700 men surveyed spent a mean time of 10 hrs per week with their friends with a standard deviation of 1.9 hours. On the other hand, 740 women surveyed spent a mean time of 7.5 hours with a standard deviation of 1.6 hours. Construct a 95% confidence interval estimate for the difference between the corresponding population means.

Find the point estimate (difference between sample means)

Determine critical value $t_{\alpha/2}$



Find margin of error

Construct confidence interval

Does it appear that there is a difference between men and women?

Def Two sets of observations are *paired* if each observation in one set has a special correspondence or connection with exactly one observation in the other set.

Ex: State if the samples are dependent or independent and if the variable is qualitative or quantitative.

1) Among competing acne medications, does one perform better than the other? To answer this question, researchers applied Medication A to one part of the subject's face and Medication B to a different part of the subject's face to determine the proportion of subjects whose acne cleared up for each medication. The part of the face that received Medication A was randomly determined.

Sample 1 =

Sample 2 =

Variable =

2) A researcher wishes to determine the effects of alcohol on people's reaction time. She randomly divides 100 people 21 years or older into two groups. Group 1 is asked to drink 3 ounces of alcohol, while group 2 drinks a placebo. Thirty minutes later they measure their reaction time.

Sample 1 =

Sample 2 =

Variable =

3) A statistician wants to compare the treatment of female and male actors based on their ages. She looks at the ages of the females and males who won best actor/actress in the last five years of the Oscars.

Sample 1 =

Sample 2 =

Variable =

TURN TWO SAMPLES INTO ONE SAMPLE:

Consider an experiment where a researcher throws a stick towards someone and first asks them to catch it with their dominant hand then again with their non-dominant hand. The times below showed how long it took several individuals to react to the toss.

Student	Dominant Hand, X_i	Nondominant Hand, Y_i	Difference, d_i
1	0.177	0.179	0.177 − 0.179 = −0.002
2	0.210	0.202	0.210 − 0.202 = 0.008
3	0.186	0.208	−0.022
4	0.189	0.184	0.005
5	0.198	0.215	−0.017
6	0.194	0.193	0.001
7	0.160	0.194	−0.034
8	0.163	0.160	0.003
9	0.166	0.209	−0.043
10	0.152	0.164	−0.012
11	0.190	0.210	−0.020
12	0.172	0.197	−0.025
			$\sum d_i = -0.158$

These are **Dependent Samples** because:

Create one **New Sample of *Differences*** then run a ________________________.

Symbols:

True mean of the differences =

Sample standard deviation of the differences =

Sample size =

Recall the logic: Say $H_0: \mu_d = 0$ and $H_1: \mu_d < 0$. If we get…

- $\bar{d} = 0$ we ____________________ H_0
- $\bar{d} > 0$ we ____________________ H_0
- $\bar{d} < 0$ by "a little", we ____________________ H_0
- $\bar{d} < 0$ by "a lot", we ____________________ H_0

Steps for Hypothesis Test when Applied to testing μ_d

Pre-Step: *Check Requirements*	***Step 1:*** *Hypotheses*	***Step 2:*** *Level of Significance*
• Samples are dependent and simple random • Differences are **normal** or $\boldsymbol{n > 30}$	$H_0: \mu_d = 0$ $H_1: \mu_d < 0$ or $\mu_d > 0$ or $\mu_d \neq 0$	

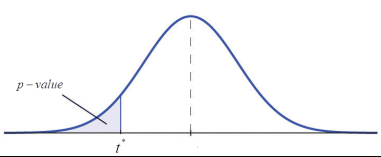
Step 3: *Test Statistic* ***(note: it can be negative!)***

Find a z-score, t-value, X^2 value or F-value

$$t^* = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

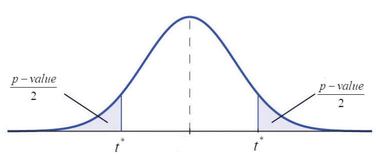
Step 4: *Find a Critical Value or P-Value*

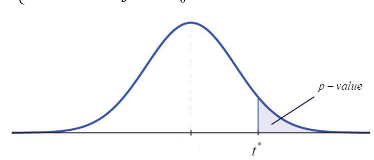
P-VALUE METHOD



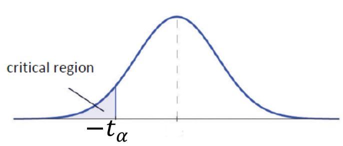
DECISION

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } \textit{P-value} \le \alpha \\ \textit{Fail to Reject } H_0 \sim \text{ if } \textit{P-value} > \alpha \end{cases}$



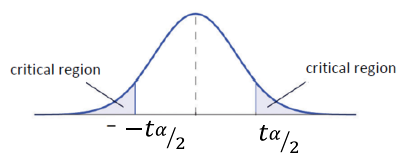


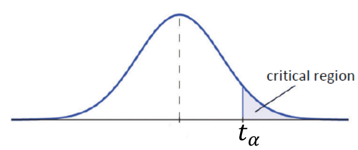
CRITICAL VALUE METHOD



DECISION

$\begin{cases} \textit{Reject } H_0 \sim \text{ if } t^* \text{ lies in the critical region} \\ \textit{Fail to Reject } H_0 \sim \text{ if } t^* \text{ doesn't lie in the critical region} \end{cases}$





Step 5: Write a *CONCLUSION* either rejecting or failing to reject H_0

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ T-Test

(b) Enter
$$\begin{cases} \mu_0 = \text{population mean stated in } H_0 \ (\mu_d) \\ s = \text{sample standard deviation } (s_d) \\ \bar{x} = \text{sample mean } (\bar{d}) \\ n = \text{sample size} \\ \mu \sim \text{alternative hypothesis} \end{cases}$$

Ex: It is a commonly held belief that Crossovers are safer than small cars. If a Crossover and small car are in a collision, does the Crossover sustain less damage (as suggested by the cost of repair)? Consumer Reports crashed Crossovers into small cars, with the Crossover moving 15 miles per hour and the front of the Crossover crashing into the rear of the small car. The data is normally distributed. Below are the repair costs:

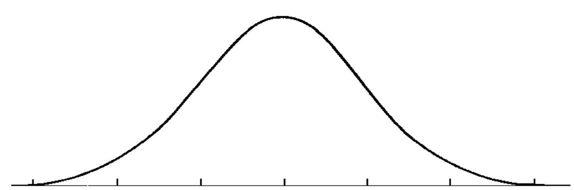
Crossover into Car	Small Car Damage	Crossover Damage	
Lexus RX-350 into Honda Insight	1274	1721	
Nissan Pathfinder into Hyundai Elantra	2327	1434	
Toyota RAV4 into Kia Forte	3223	850	
Jeep Cherokee into Kia Niro Hybrid	2058	2329	
Ford Explorer into Toyota Camry	3095	1415	
Honda CR-V into Ford Focus	3386	1470	
Chevrolet Equinox into Nissan Sentra	4560	2884	

Do the sample data suggest that Crossovers are safer? Use the level of significance $\alpha = 0.01$.

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region



Decision about Null Hypothesis

Conclusion

Confidence Interval for the Mean Difference from Dependent Samples

Alternative Forms: $\bar{d} - E < \mu_d < \bar{d} + E$ or $\bar{d} \pm E$

where the margin of error is given by $E = t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$

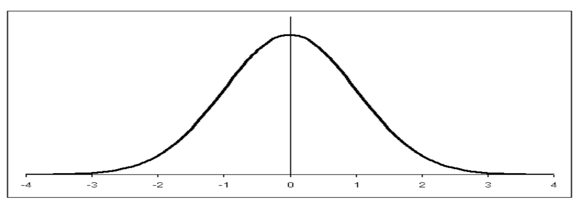
Ex: A company claims that its 12-week special exercise program significantly reduces weight. A random sample of eight persons was selected, and the following table gives the weights (in lbs) of those eight persons before and after the program.

	Weight (in pounds)							
Before	180	195	177	221	208	199	148	230
After	185	187	171	214	208	194	150	227

Construct a 90% confidence interval for the mean before-after differences.

Find point estimate (sample mean)

Determine critical value $t_{\alpha/2}$



Find margin of error

Construct confidence interval

Does it appear that the weight loss program is effective?