MODULE 4 LIVE LECTURE

INFERENTIAL STATISTICS AND ANALYTICS

INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

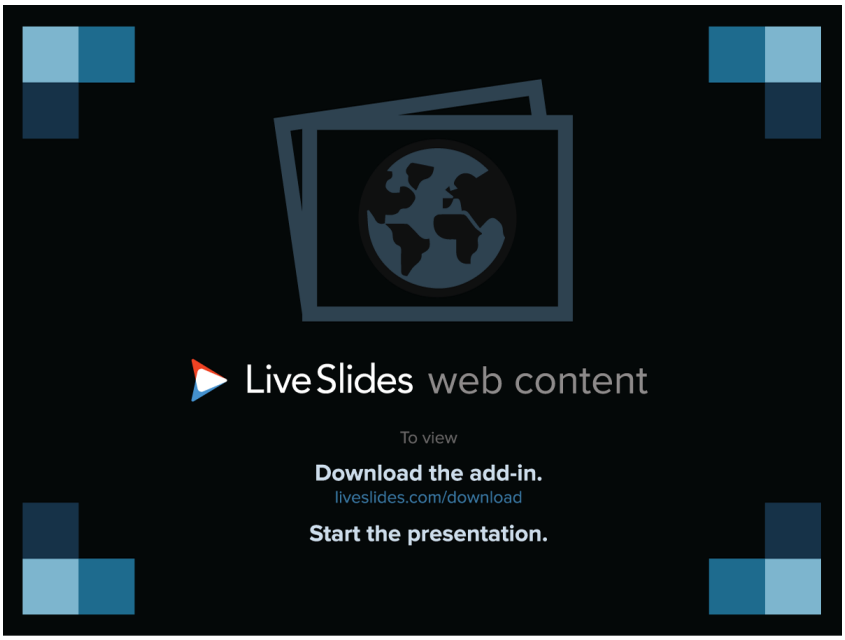
- Module 4 Inferences from two samples
- Module 4 homework
- Quiz 4
- Course project phase 4
- Module 4 Live classroom Grading
- Summary

INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences from two samples***
- Module 4 homework
- Quiz 4
- Course project phase 4
- Module 4 Live classroom Grading
- Summary

INFERENCES FROM TWO SAMPLES

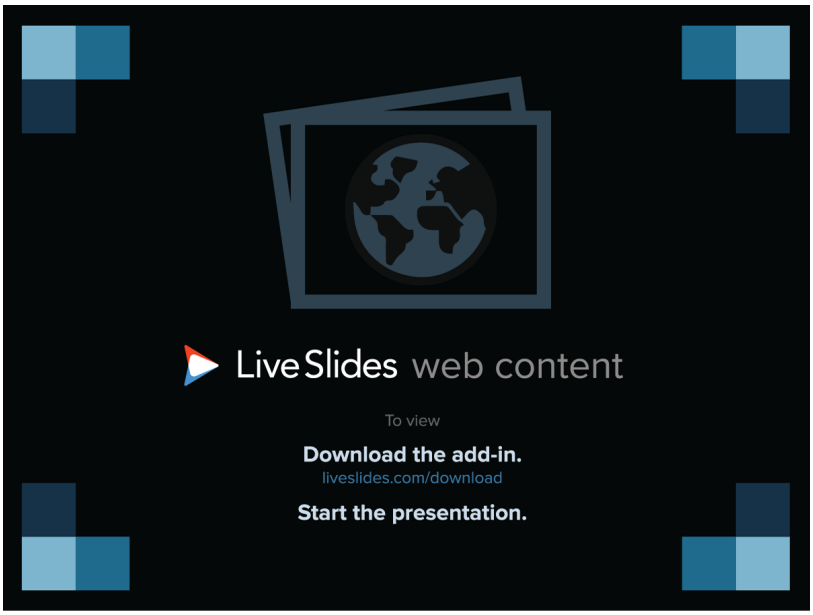
- ***Inferences from Two Independent Samples: Two Proportions***
- If the confidence interval for the difference between two population proportions ***includes 0***, then it is possible that the difference between the two population proportions equals 0.That is there is no significant difference between the two population proportions.
- The confidence interval ***does not include 0***, which means that the two population proportions are not equal.



https://www.youtube.com/watch?v=OIYkOiQX3fk

- ***Inferences from Two dependent Samples***

- Testing hypotheses and constructing confidence intervals involving *the mean of the differences of the values* from two dependent populations.



https://www.youtube.com/watch?v=TMRBGFlj0RM

INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences from two samples***
- ***Module 4 homework***
- Quiz 4
- Course project phase 4
- Module 4 Live classroom Grading
- Summary

MODULE 4 HOMEWORK

1.In a clinical trial, 401,974 adults were randomly assigned to two groups. The treatment group consisted of 201,229 adults given a vaccine and the other 200,745 adults were given a placebo.Among the adults in the treatment group, 33 adults developed the disease and among the placebo group, 115 adults developed the disease. Based on the information presented, answer the following questions:

$n_1 = 201229 \quad n_2 = 200745$
$x_1 = 33 \quad x_2 = 115$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{33}{201229}$$
$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{115}{200745}$$

$\hat{q}_1 = 1 - \hat{p}_1$
$\hat{q}_2 = 1 - \hat{p}_2$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$
$\bar{q} = 1 - \bar{p}$

2.In a clinical trial, 401,974 adults were randomly assigned to two groups. The treatment group consisted of 201,229 adults given a vaccine and the other 200,745 adults were given a placebo.Among the adults in the treatment group, 33 adults developed the disease and among the placebo group, 115 adults developed the disease.The doctors' claim that the rate for the group receiving the vaccine is less than the group receiving the placebo. Answer the following questions:

a. identify the symbolic null and alternative hypothesis.

$H_0 : p_1 = p_2$
$H_1 : p_1 < p_2$

c. Assume that we want to use a 0.05 significance level to test the claim that $p_1 < p_2$. If we want to test that claim by using a confidence interval, what confidence level should we use?
1-5%-5%

2.In a clinical trial, 401,974 adults were randomly assigned to two groups. The treatment group consisted of 201,229 adults given a vaccine and the other 200,745 adults were given a placebo.Among the adults in the treatment group, 33 adults developed the disease and among the placebo group, 115 adults developed the disease.The doctors' claim that the rate for the group receiving the vaccine is less than the group receiving the placebo. Answer the following questions:

$H_0 : p_1 = p_2$

$H_1 : p_1 < p_2$

d. If we test the original claim, we get the confidence interval -0.000508 < p_1 - p_2 < −0.000309 , what does this confidence interval suggest about the claim? Since the limits do not contain 0, there appears to be a significant difference between the vaccine group and the placebo group.

3. Determine whether the samples are independent or dependent.

a.The effectiveness of a new headache medicine is tested by measuring the amount of time before the headache is cured for patients who use the medicine and another group of patients who use a placebo drug. Explain.

b.The effectiveness of a headache medicine is tested by measuring the intensity of a headache in patients before and after drug treatment. The data consist of before and after intensities for each patient. Explain.

Two samples are independent if the sample values selected from one population are unrelated to or paired or matched with the sample values from the second population.

Two samples are dependent if the sample values are paired. (That is, each pair of sample values consists of two measurements from the same subject [such as before/after data], or each pair of sample values consists of matched pairs [such as husband/wife data], where the matching is based on some inherent relationship.)

MODULE 4 HOMEWORK

4.Use the traditional method of hypothesis testing to test the given claim about the means of two populations. Assume that two dependent samples have been randomly selected from normally distributed populations. Using a 0.01 level of significance, test the claim that the tutoring has an effect on the math scores.
Five students took a math test before and after tutoring. Their scores were as follows.

Student	A	B	C	D	E			
Before	71	66	67	77	75			
After	75	75	65	80	87			
d	-4	-9	2	-3	-12	d_bar=-5.2	Sd=5.45	n=5

- a.

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d > 0$$

d = individual difference between the two values in a single matched pair

μ_d = mean value of the differences d for the population of all matched pairs of data (this is equal to 0 since the null hypothesis assumes the mean differences equals 0)

$\bar{d}$ = mean value of the differences d for the paired sample data

s_d = standard deviation of the differences d for the paired sample data

n = number of pairs of sample data

- b.Find the value of the test statistic, t.

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

where $\mu_d = 0$ and degrees of freedom = $n - 1$

- c. Find the critical value(s) and state decision about null hypothesis.

Using a 0.01 level of significance, test the claim that the tutoring has an effect on the math scores.

Since this is a right-tailed test.
df=5-1=4
Use the t table to find the area to the right of t .

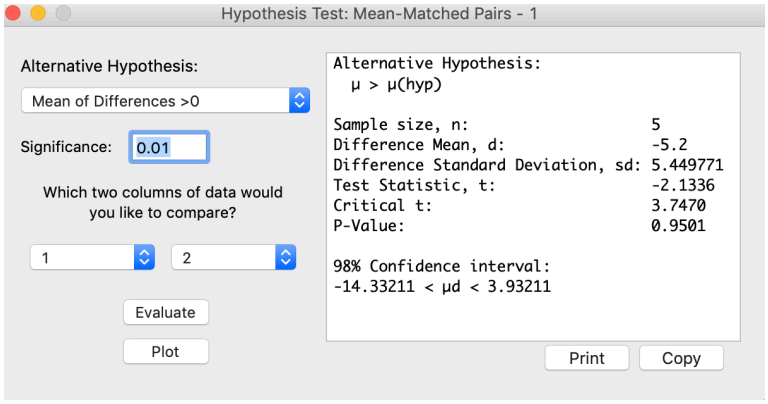
The P-value is area to the right .
If P value < 0.01, reject the null hypothesis.
If P value > 0.01, do not reject the null hypothesis. There is not sufficient evidence to support the claim that tutoring has an effect on math scores.

MODULE 4 HOMEWORK

4. Use the traditional method of hypothesis testing to test the given claim about the means of two populations. Assume that two dependent samples have been randomly selected from normally distributed populations. Using a 0.01 level of significance, test the claim that the tutoring has an effect on the math scores.
Five students took a math test before and after tutoring. Their scores were as follows.

Student	A	B	C	D	E			
Before	71	66	67	77	75			
After	75	75	65	80	87			
d	-4	-9	2	-3	-12	d_bar=-5.2	Sd=5.45	n=5

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d > 0$$



p value > 0.01, we fail to reject the null hypothesis.There is not sufficient evidence to support the claim that the tutoring has an effect on math scores.

MODULE 4 HOMEWORK

5. A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to construct a 99% confidence interval for $\mu_1 - \mu_2$ where μ_1 and μ_2 represent the mean for the treatment group and the control group respectively.

Treatment Group	**Control Group**
$n_1 = 85$	$n_2 = 75$
sample mean=189.1	sample mean = 203.7
$s_1 = 38.7$	$s_2 = 39.2$

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

The following requirements must be met to test a claim about two independent population means:

1. σ_1 and σ_2 are unknown and no assumption is made about the equality of σ_1 and σ_2.
2. The two samples are independent.
3. Both samples are simple random samples.
4. Either or both of these conditions are satisfied: The two sample sizes are both large (over 30) or both samples come from populations having normal distributions.

5. Hypothesis Test for Two Means: Independent Samples

6. Test Statistic: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ where $\mu_1 - \mu_2$ is often assumed to be 0

Degrees of freedom: df = smaller of $n_1 - 1$ or $n_2 - 1$.

Confidence Interval Estimate of $\mu_1 - \mu_2$ Independent Sample

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E \quad \text{where} \quad E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

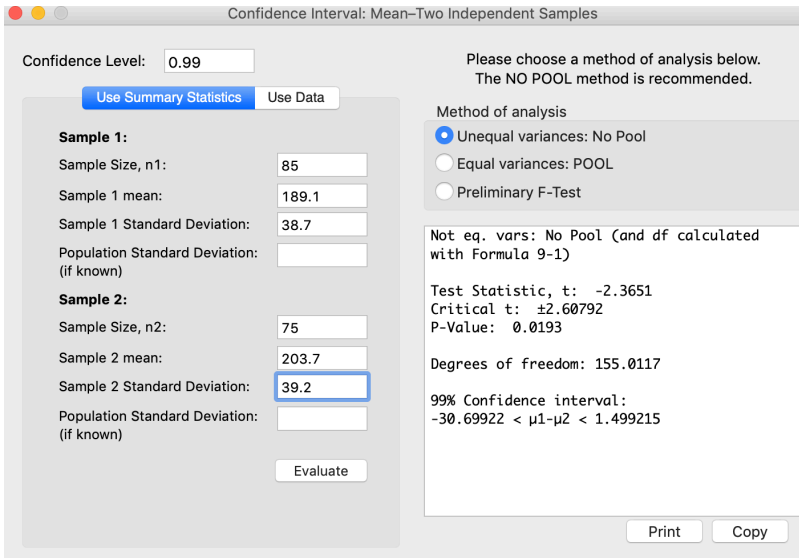
MODULE 4 HOMEWORK

5. A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to construct a 99% confidence interval for $\mu_1 - \mu_2$ where μ_1 and μ_2 represent the mean for the treatment group and the control group respectively. Show all steps of the hypothesis test and all calculations. Use the following t-distribution table - http://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm

Treatment Group
$n_1 = 85$
sample mean=189.1
$s_1 = 38.7$

Control Group
$n_2 = 75$
sample mean = 203.7
$s_2 = 39.2$

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



INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences with two samples***
- ***Module 4 homework***
- ***Quiz 4***
- Course project phase 4
- Module 4 Live classroom Grading
- Summary

QUIZ 4

1) A paint manufacturer made a modification to a paint to speed up its drying time. Independent simple random samples of 11 cans of type A (the original paint) and 9 cans of type B (the modified paint) were selected and applied to similar surfaces. The drying times, in hours, were recorded.The summary statistics are as follows.

Type A	Type B
x1 = 76.3 hrs	x2 = 65.1 hrs
s1 = 4.5 hrs	s2 = 5.1 hrs
n1 = 11	n2 = 9

The following 98% confidence interval was obtained for μ1 - μ2, the difference between the mean drying time for paint cans of type A and the mean drying time for paint cans of type B:

4.90 hrs < μ1 - μ2 < 17.50 hrs

What does the confidence interval suggest about the population means?

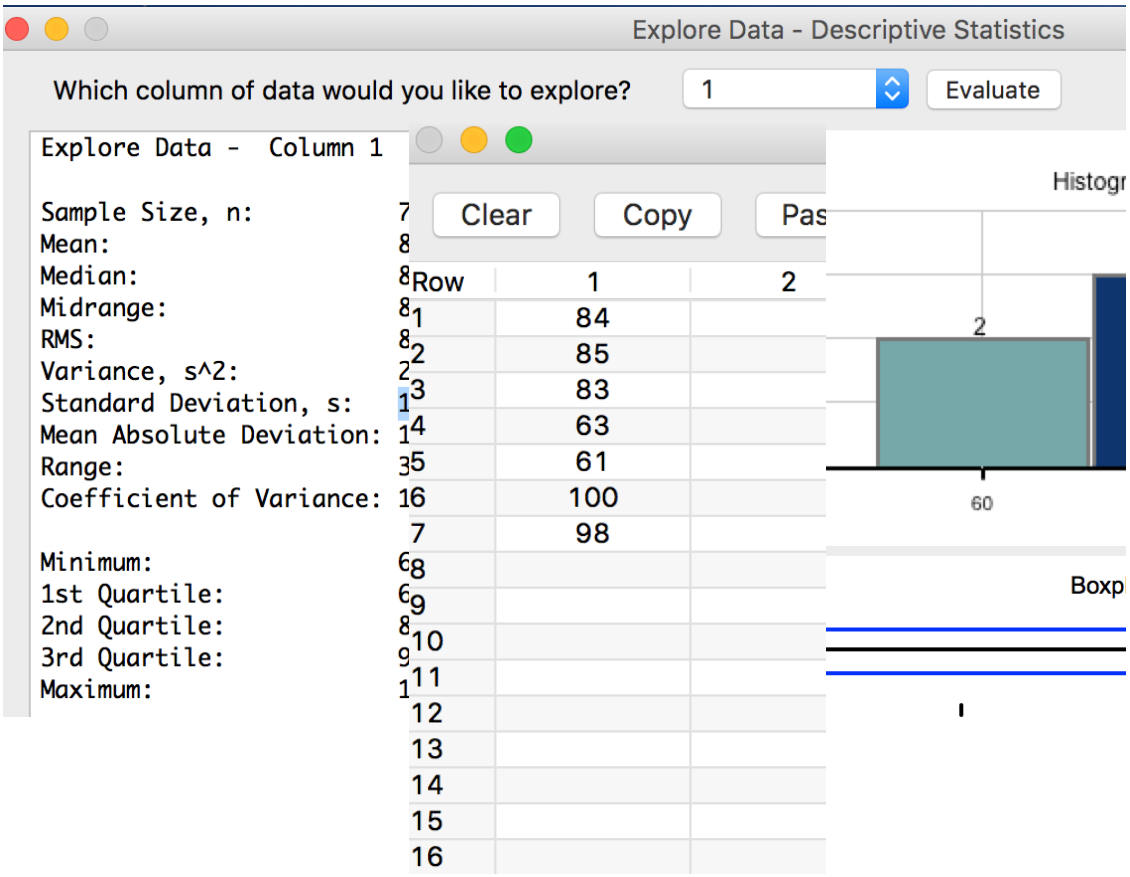
If μ1 - μ2 includes only positive values ––-> the mean drying time for paint type A > the mean drying time for paint type B.

If μ1 - μ2 includes only negative values ––-> the mean drying time for paint type A < the mean drying time for paint type B.

If μ1 - μ2 includes 0 ––-> That is there is no significant difference between the mean drying time for paint type A and the mean drying time for paint type B.

QUIZ 4

2)Consider the set of differences, denoted with d, between two dependent sets: 84, 85, 83, 63, 61, 100, 98. Find the sample standard deviation sd and round it to the nearest tenth.

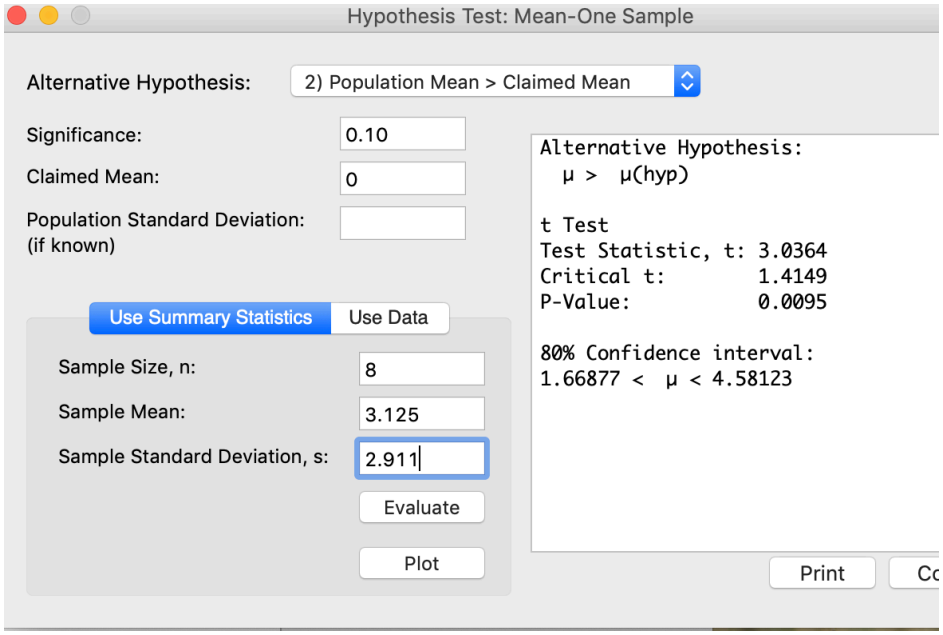


QUIZ 4

3)We wish to compare the means of two populations using paired observations. Suppose that $d = 3.125$, $S_d = 2.911$, and $n = 8$, and that you wish to test the following hypothesis at the 10% level of significance:

$H_0: \mu_d = 0$ against $H_1: \mu_d > 0$

Determine the decision criterion for rejecting the null hypothesis in the given hypothesis test. Reject H_0 if test statistic is greater than 1.415.

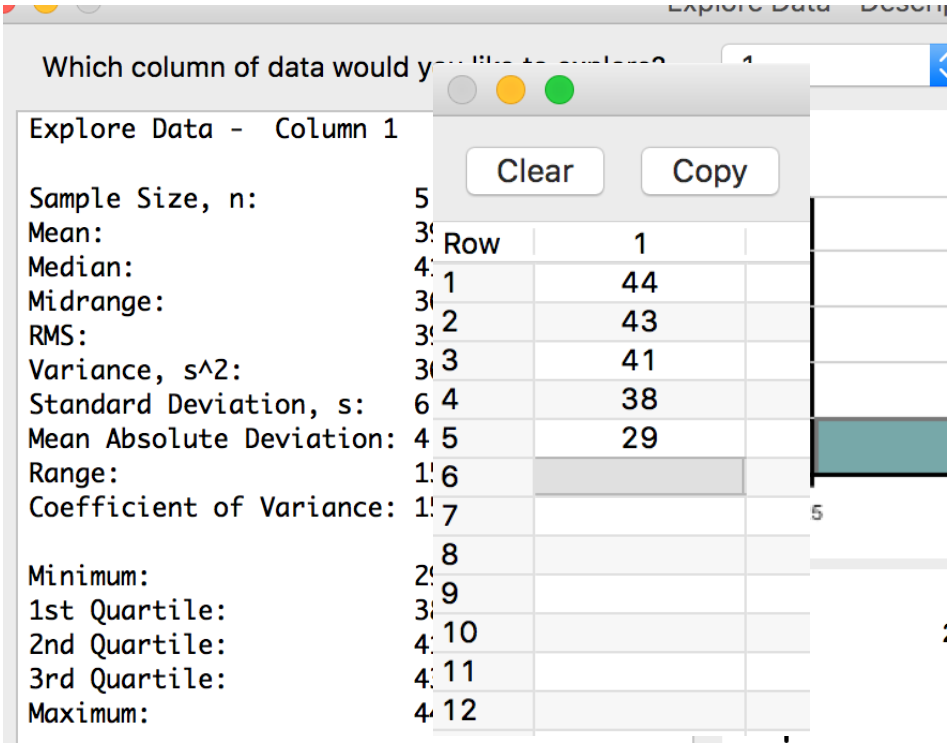


QUIZ 4

4)The two data sets are dependent. Find mean d-bar to the nearest tenth.

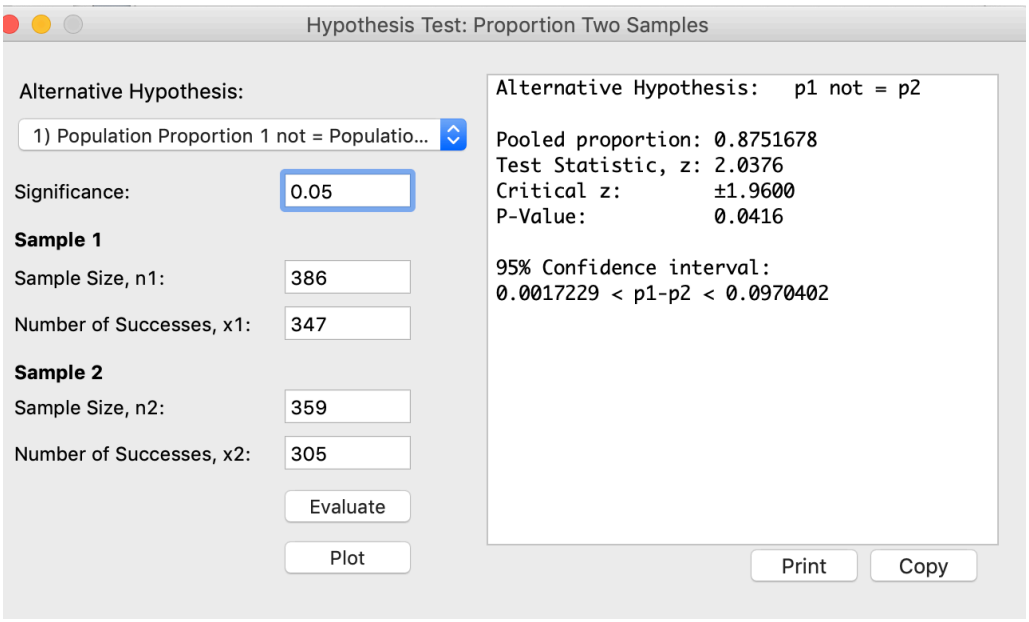
A	69	66	61	63	51
B	25	23	20	25	22
d=A-B	44	43	41	38	29

—-> mean d-bar?



QUIZ 4

5) Randomly selected subjects were asked if they agreed with the statement "It is morally wrong for lawyers to defend guilty people." Among the 386 women surveyed, 347 agreed with the statement. Among the 359 men surveyed, 305 agreed with the statement. When testing the claim that $p_1 = p_2$, a test statistic of $z = 2.04$ is obtained. What is the p-value from this test statistic?



QUIZ 4

6) Randomly selected subjects were asked if they agreed with the statement "It is morally wrong for lawyers to defend guilty people." Among the 386 women surveyed, 347 agreed with the statement. Among the 359 men surveyed, 305 agreed with the statement. When testing the claim that $p_1 = p_2$, a test statistic of z = 2.04 is obtained. What is the value of $\bar{p}$?

$x_1 = 347$ $\quad$ $x_2 = 305$

$n_1 = 386$ $\quad$ $n_2 = 359$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

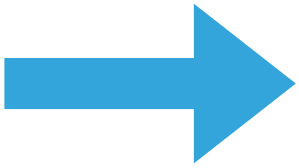
QUIZ 4

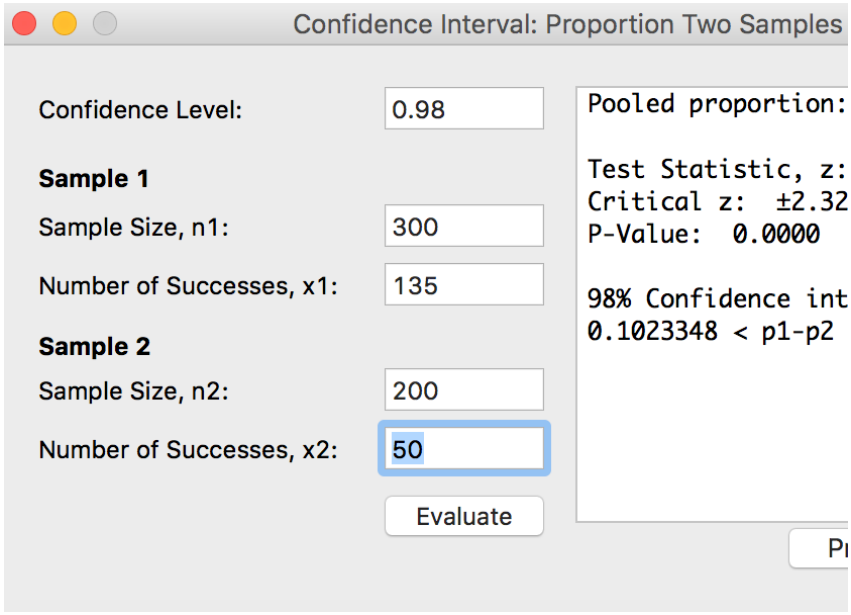
7) Construct the indicated confidence interval for the difference between population proportions p1 - p2. Assume that **the samples are independent** and that they have been randomly selected. In a random sample of 300 women, 45% favored stricter gun control legislation. In a random sample of 200 men, 25% favored stricter gun control legislation. Construct a 98% confidence interval for the difference between **the population proportions** p1 - p2.

Sample 1: n1=300 p1=45% —> x1=300 x45%=135

Sample 2: n2=200 p2=25% —> x2=200 x25%=50

Construct a 98% confidence interval for the difference between the population proportions p1-p2





QUIZ 4

8) Which distribution is used to test the claim that the standard deviation of the ages of Minnesota voters is equal to the standard deviation of voters in Illinois?

$$F = \frac{s_1^2}{s_2^2}$$ where s_1^2 is the larger of the two sample variances

QUIZ 4

9)We are interested in testing the claim that the heights of men and the heights of women have different variances. Given that s = 6.60 for women and s = 6.02 for men, find the value of the F-test statistic.

$$F = \frac{s_1^2}{s_2^2}$$ where s_1^2 is the larger of the two sample variances

$$F = \frac{s_1^2}{s_2^2} = \frac{6.6^2}{6.02^2}$$

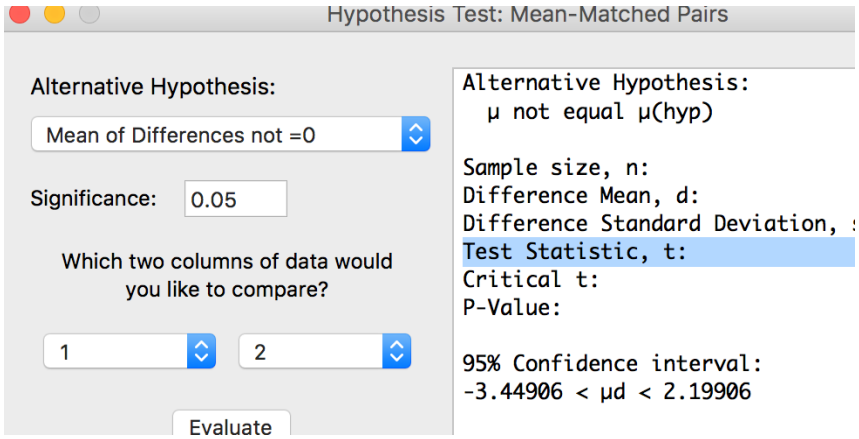
QUIZ 4

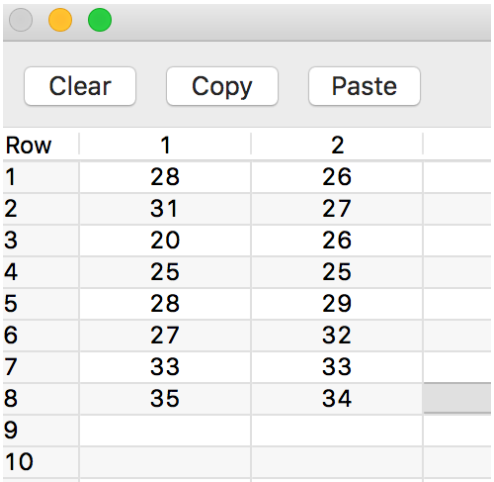
10)Assume that you want to test the claim that the paired sample data come from a population for which the mean difference is μd = 0. Compute the value of the t test statistic. Round intermediate calculations to four decimal places as needed and final answers to three decimal places as needed.

x	28	31	20	25	28	27	33	35
y	26	27	26	25	29	32	33	34

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

where $\mu_d = 0$ and degrees of freedom = $n - 1$





INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences with two samples***
- ***Module 4 homework***
- ***Quiz 4***
- ***Course project phase 4***
- Module 4 Live classroom Grading
- Summary

COURSE PROJECT PHASE 4

- For Phase 4 of your course project, you will want to review grade feedback from your Phase 1, Phase 2, and Phase 3 submissions to make any necessary corrections.
- Once you have made your corrections, you will compile your information from Phase 1, Phase 2, Phase 3 and your final conclusion into one submission and submit this as your rough draft for Phase 4 of the course project.

INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences from two samples***
- ***Module 4 homework***
- ***Quiz 4***
- ***Course project phase 4***
- ***Module 4 Live classroom Grading***
- Summary

MODULE 4 LIVE CLASSROOM GRADING

The live classroom session archive as a URL will be added after the session has ended. The following is how you will receive your points for the module 4 live classroom session:

Question:**What is the difference between dependent and independent samples?**

Please go to module 4 live classroom, enter your response.

You have until midnight CST on Sunday to confirm that you have viewed the live classroom session archive.

INFERENTIAL STATISTICS AND ANALYTICS–MODULE 4 LIVE LECTURE

- ***Module 4 Inferences from two samples***
- ***Module 4 homework***
- ***Quiz 4***
- ***Course project phase 4***
- ***Module 4 Live classroom Grading***
- ***Summary***

SUMMARY

- If the confidence interval for the difference between two population proportions ***includes 0***, then it is possible that the difference between the two population proportions equals 0.That is there is no significant difference between the two population proportions.
- The confidence interval ***does not include 0***, which means that the two population proportions are not equal.
- Module 4 Question:**What is the difference between dependent and independent samples?**