Math 247: Test 4 (Wright, Fall 2019) **Name:**______________________________

__________/100 pts

I encourage you to work with other students in the class but the final work you hand in must be your own. Your answers must match every step of your work; otherwise, you may lose most or all of the points for the problem. You may consult with tutors but do not ask them to solve the problems for you!

- *This exam is due at the beginning of class on Tuesday, 12/3/19. Be sure that all answers are written in your own words; i.e., do not write verbatim the same answer as another student.*
- *Your work should be typed except for calculations and graphs, which can be hand-written.*
- *Include printouts for all StatCrunch work. Please label the StatCrunch printouts with corresponding problem number and letter from the test.*

Scoring will be based on organization of your work, accuracy, and thoughtful, well-written answers using complete sentences!

1. (10 pts) The following is the taken from the "Annual Drinking Water Quality Report", from 2004, for the town of Brookston, Indiana. In the final report there was this statement:

   "I'm pleased to report that our drinking water is safe and meets federal and state requirements."

   Here are some of the actual test results: Violation "Y/N" means Yes/No. Yes = violation because the water exceeds the safety standard; No = no violation because the water is under the MCL.

   MCL is the maximum contaminant level, the highest level of a contaminant that is allowed in drinking water.

   Beta/photon emitters and alpha emitters refers to radioactivity detected in the water.

   The Level Detected values are confidence intervals in the Margin of Error format.

Contaminant	Violation Y/N	Level Detected	Unit measurement	MCL
Beta/photon emitters BPE	N	2.1 ± 3.2	mrem/yr	4
Alpha emitters	N	0 ± 1.6	pCi/l	15

BPE CI

$\bar{x} \pm M$

2.1 ± 3.2

$\Rightarrow (-1.1, 5.3)$

Captures lots of values over 4!!

$\Rightarrow$ over MCL

One of these contaminant violation results should actually be a "yes" instead of a "no." Which one is it and why? Include a discussion of what the numbers in the confidence interval tell you about the sample mean (that came from water samples) and what the inference is for the population (the entire water supply).

The "Beta/photon emitters" contaminant should be Yes (as in "Yes, this contaminant could exceed safe levels (MCL)."

The CI is 2.1 ± 3.2 so the water samples had an average contaminant level of 2.1 mrem/yr. This is below the MCL of 4. But the margin of error is ± 3.2, giving a CI of $(-1.1, 5.3)$, indicating there's a chance the mean for the entire water supply could exceed the MCL. So the evidence suggests there's a distinct possibility the water is NOT safe.

2. (20 pts) Research conducted in 2015 showed that 35% of Cuesta students had travelled outside the US. A recent survey in 2019 showed that out of the 100 randomly sampled students, 48 have travelled outside the US.

4 (a) By hand, construct a 95% confidence interval for the proportion of all Cuesta students who have traveled out of the US. For full credit, clearly show all of your work.

CI: Estimate ± Margin of Error

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.45 \pm 1.96 \sqrt{\frac{.45(1-.45)}{100}}$$

$$\boxed{.45 \pm .098}$$

$$\boxed{(.352, .548)}$$

Parking Lot

$P_{2015} = .35$

Sample (2019)

$n = 100$

$x = 45$ successes

$\hat{p} = \frac{45}{100} = .45$

95% confidence

$z^* = 1.96$

4 (b) Find the confidence interval using StatCrunch. Print out and attach your work to this test.

3 (c) Interpret the confidence interval the context of the problem.

We are 95% confident that the true proportion of all Cuesta students who have traveled outside the U.S. is between 35.2% and 54.8%.

4 (d) What would the hypotheses be if you wanted to test whether there has been a significant change in the proportion of Cuesta students who have traveled outside of the US since 2015?

$H_0: p = .35$ The proportion of students who have traveled outside the U.S. is still 35%

$H_a: p \neq .35$ The proportion has changed since 2015 (so is no longer 35%)

5 (e) Would you reject or fail to reject the null hypothesis above at the .05 significance level, based on the confidence interval you found in part (a)? Explain your answer.

Based on the confidence interval we would reject the null (at the .05 level of significance) since the CI did not capture the null ($p = .35$). Moreover, the CI tells us there has been a significant increase in the proportion of students who have traveled outside the U.S. (again, at the .05 level of significance).

3. (40 pts) A student doing a research project on whether the Academic Success Center helps students took a random sample of 30 students in the Math Lab and found they had a mean GPA of 3.26, with a standard deviation of 0.81. The average GPA of all Questa College students is 2.93.

(a) Test whether the students in the Math Lab have, on average, significantly different GPA's than general Cuesta students. Include all 4 steps of the hypothesis test. Write the hypotheses with symbols and with words. Do all compute work by hand, up to finding the P-value. Use StatCrunch to confirm your results and to find the P-value. Draw a well-labeled curve that illustrates the sampling distribution, the sample mean, the t-value, and the P-value. Include a printout of the StatCrunch work.

Parking Lot
$n = 30$
$\bar{x} = 3.26$
$s = 0.81$
$\mu_{Cuesta} = 2.93$

1. Hypothesize $H_0: \mu = 2.93$ (The mean GPA of Math Lab students is the same as all Cuesta)

$H_a: \mu \neq 2.93$ (The GPA of Math Lab students, on average, is different)

2. Plan and Prepare Use $\alpha = .05$ and 1 Sample t-Test

Conditions: 1. Random Sample? Yes, stated
Independent observations? Assume
2. Large sample? Yes, $n = 30 \geq 25$

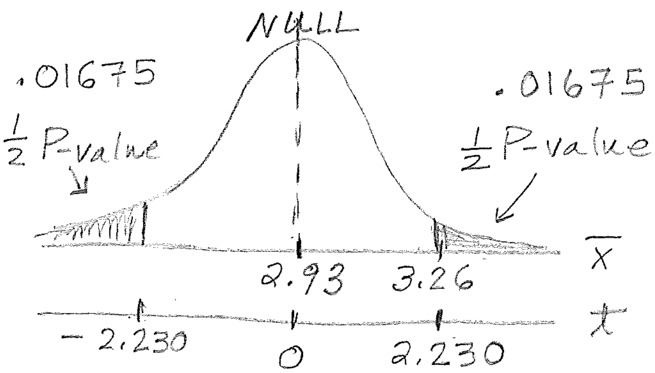
3. Compute

$SE = \frac{s}{\sqrt{n}} = \frac{.81}{\sqrt{30}} \approx .148$

Test Stat $t = \frac{\bar{x} - \mu_0}{SE} = \frac{3.26 - 2.93}{.148} = 2.230$

$df = n - 1 = 29$

P-value = .0335



4. Interpret: (a) P-value = .0335 < .05 = α
Reject H_0, accept H_a

(b) We have convincing evidence that the mean GPA of students who use the Math Lab is significantly different from the mean GPA of all Cuesta students.

(b) Referring to the hypothesis test in part (a), if we were to change the test to see whether the mean GPA of students in the Math Lab is significantly higher than from the general Cuesta student mean GPA,

- What would the new Alternative Hypothesis be? $H_a: \mu > 2.93$
- What would the new P-value be? P-value = .01675
- Would changing the Alternative Hypothesis change the outcome of the hypothesis test? Explain.

No, both P-values are less than the .05 level of significance, but the smaller P-value for $\mu > 2.93$ gives the impression the result is more highly significant, even though the evidence is the same!

- Which test is stricter (harder to get a significant result)? One-Tailed Test Two-Tailed Test

#3 (continued)

(c) By hand, construct a 95% confidence interval for the mean GPA of all students who use the Math Lab.

CI: Estimate ± Margin of Error

$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

$3.26 \pm 2.045\left(\frac{.81}{\sqrt{30}}\right)$

$\boxed{3.26 \pm .302}$

$\boxed{(2.958, 3.562)}$

$\bar{x} = 3.26$
$s = .81$
$n = 30$
$t^* = 2.045$
for 95% confidence and df = 29

(d) Use StatCrunch to find the CI. Include the printout with your test. ✓

(e) Interpret the confidence interval in the context of the problem.

We are 95% Confident that the average GPA for all students who use the Math Lab is between 2.958 and 3.562.

(f) Does the CI support your conclusion to the hypothesis test in part(a)? Explain.

Yes, the CI does not capture the null of $\mu = 2.93$ so it tells us to reject the null (at the .05 level of significance to match the 95% confidence level.)

(g) Does the CI support your conclusion to the hypothesis test in part (b)? Explain.

Yes, same result.

(h) In order to relate a confidence interval to a hypothesis test, the test must be TWO - Tailed.

(i) Can we conclude that using the Math Lab *causes* students to have higher GPA's? Explain why or why not.

No! This was an observational study! A huge confounder is motivation (also confidence). Students who are more motivated (or confident) will seek out help, like the Math Lab, and will also work harder in their classes. Also, time is another huge confounder, as in having time to seek help and also to study (hence higher GPA).

4. (30 points) Dr. Dean Ornish conducted a five-year study (https://jamanetwork.com/journals/jama/fullarticle/188274) in which he showed that coronary artery disease (CAD) is reversible. Patients with advanced CAD were randomized (randomly assigned) to an intensive lifestyle change group (low-fat vegetarian diet, moderate aerobic exercise, stress-management training, group support) or to a usual-care control group. After 5 years, the control group of 20 patients had an average of 2.25 cardiac "events" per patient (heart attack, hospitalization, by-pass surgery, cardiac-caused death), with standard deviation of 1.15. The experimental group of 28 patients had an average of .89 cardiac events per patient with standard deviation .38.

(a) Perform a hypothesis test to see whether there is a significant difference in the mean number of cardiac events between the two groups. Include all 4 steps of the hypothesis test. Write the hypotheses with symbols and with words. Use StatCrunch to do the compute step but draw a well-labeled curve by hand that illustrates the sampling distribution of the difference in means, the observed difference, the Test Statistic (t-value), and the P-value. Include a printout of the StatCrunch work. Use alpha = .05

Parking Lot

Sample 1 (Lifestyle)	Sample 2 (Control)
$n_1 = 28$	$n_2 = 20$
$\bar{X}_1 = .89$	$\bar{X}_2 = 2.25$
$S_1 = .38$	$S_2 = 1.15$

1. Hypothesize

$H_0: \mu_{Life} - \mu_{Control} = 0$ → There is no difference in the average number of cardiac events. Lifestyle changes make no difference, on average.

$H_a: \mu_{Life} - \mu_{control} \neq 0$ → There IS a difference and lifestyle changes caused it!

Note: You could reverse the order of the hypotheses and sample numbers.

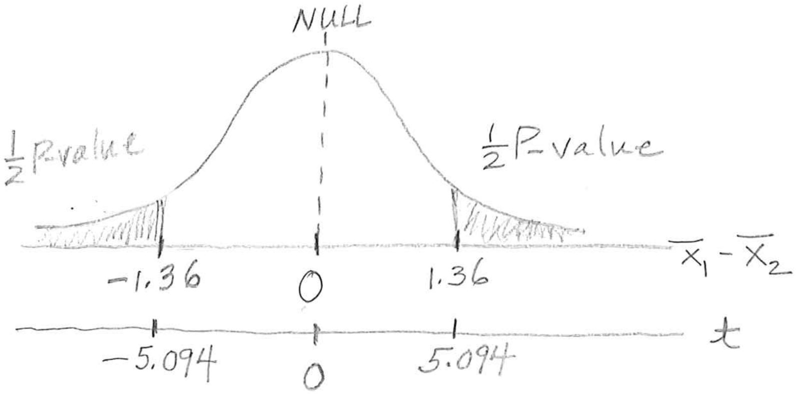
2. Plan and Prepare: Use $\alpha = .05$ and the 2-sample t-Test.

1. Random sample? unknown, but sample WAS randomized } assume
   Independence WITHIN samples? Again, unknown.
2. Independence BETWEEN samples? Assume no connection between groups.
3. Large samples? $n_1 = 28 \geq 25$ yes
   $n_2 = 20 < 25$ no

Because n_2 is a small sample we have to assume the distribution of cardiac events is normal.

3. Compute

Observed difference:
$\bar{X}_1 - \bar{X}_2 = -1.36$
Test statistic: $t = -5.094$
P-value < .0001



4. Interpret

(a) $P < .0001 < .05 = \alpha$

(b) We have strong evidence that the lifestyle changes cause a significant difference in the mean number of cardiac events as compared to the usual care control for patients with CAD, over a 5-year period.

#4 (continued)

(b) Does the result from the hypothesis test above tell us whether there were significantly *fewer* cardiac events in the Lifestyle Change group? Explain.

Technically speaking, no. The alternative hypothesis was just that there is a significant change in the mean number of cardiac events due to lifestyle intervention.
However, if we consider the evidence, it's clear there was a reduction of cardiac events, on average, in this group, and a one-tailed test would yield an even smaller P-value.

(c) Find the 95% confidence interval for the difference in means using StatCrunch (include printout)

CI: (−1.914, −.806) OR (.806, 1.914) if you switch the order of μ's in the hypotheses.

(d) Explain how you can tell whether or not to reject the null hypothesis by just looking at the CI.

Since the CI did not capture ZERO, we can see that there IS a significant difference, which means we would reject the null.

(e) Interpret the confidence interval in the context of the study. Include what the confidence interval tells us about which group had significantly fewer cardiac events, on average.

We are 95% confident that the difference in mean cardiac events between the Lifestyle group and the Control group is between .8 and 1.9 events.

What this means is that implementing the lifestyle change program for ALL CAD patients could reduce cardiac events, on average, by .8 to 1.9 events, because the lifestyle group had significantly fewer events.

Note: This result (real data!) assumes the conditions for using the 2-sample t-test are fully satisfied which probably isn't entirely true. Still, this is good news for people with CAD since the lifestyle changes have no side effects!

#2 part b)

One sample proportion summary confidence interval:

p : Proportion of successes
Method: Standard-Wald

95% confidence interval results:

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	45	100	0.45	0.049749372	0.35249302	0.54750698

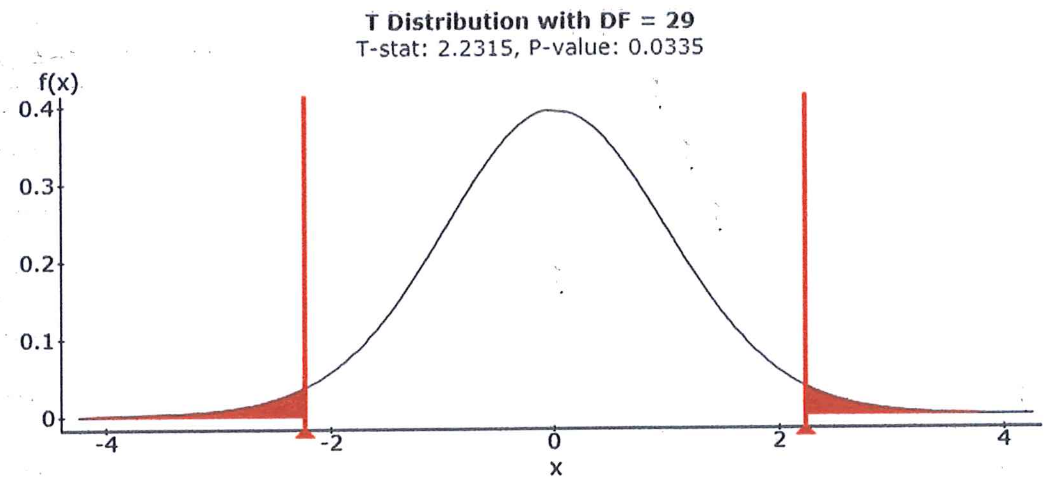
#3 part(a)

One sample T summary hypothesis test:

μ : Mean of population
$H_0 : \mu = 2.93$
$H_A : \mu \neq 2.93$

Hypothesis test results:

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	3.26	0.14788509	29	2.2314623	0.0335



#3 part(d)

One sample T summary confidence interval:

μ : Mean of population
95% confidence interval results:

Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	3.26	0.14788509	29	2.957541	3.562459

#4 part(a)

Two sample T summary hypothesis test:

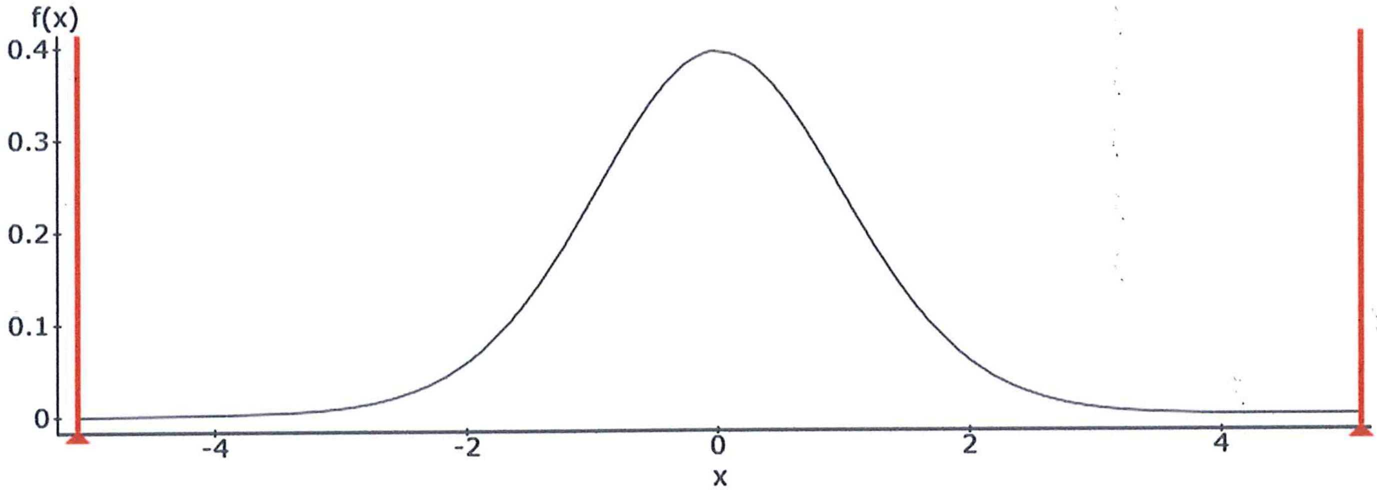
μ_1 : Mean of Population 1
μ_2 : Mean of Population 2
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Note: If you use the T-distribution calculator on StatCrunch, you can find the exact P-value = .0004198

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-1.36	0.26698716	21.985116	-5.0938779	<0.0001

T Distribution with DF = 21.985116
T-stat: -5.0939, P-value: <0.0001



#4 part(c)

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-1.36	0.26698716	21.985116	-1.9137192	-0.80628079