Math 247: Homework Set 4 **Name:___________________________**

StatCrunch ________________ / 16 points **Non-StatCrunch** ______________/4 points

Self-Assessment: Determine total number of correctly done problems (they don't have to be done correctly the first time…just make sure you find and fix any errors in your work!) and put that score in the last column.

Instructions for submitting homework:

a. **All (and ONLY) StatCrunch problems need to be typed up on a Google Doc,** with the StatCrunch results cut and pasted in for Step 3. See the example below (Problem 8.63).

b. Again, ONLY StatCrunch problems need to be submitted!

c. **I'm working on access so you can include the cover page with the "Assigned Problems" list with the numbers filled in (as usual). More on this later.**

d. When the homework is due (on the day of the final exam), you will send your Google Doc to me by sharing

- Cover pages, with the numbers filled in the assignment (Maybe…I'm working on this
- All of your StatCrunch homework typed up as shown in the example below.

To Share: Select "Share" 

on the upper right corner of Google Docs page, then fill in my email address, pwright@cuesta.edu and click "Send".

EXAMPLE of how you should do the STATCRUNCH homework problems. This is actually one of your homework problems so hopefully there won't be any confusion on how I'd like to have you submit these problems.

This is what your work should look like:

8.63: Treatment for Leukemia

Parking Lot
Treatment Group (received medication)
n1 = 110
x1 = 101
p1-hat = 101/110 =.918 = 91.8%

Placebo Group
n2 = 110
x2 = 88
p2-hat =88/110 = .8 = 80%

Pooled proportion (needed to check conditions): p-hat = (101 + 88)/(110 + 110) = .859 = 85.9%
(85.9% of patients survived overall for both groups)

Step 1: Hypothesize
p = the proportion of ALL leukemia patients who would survive in either situation

H0: p1 = p2 , so **p1 - p2 = 0** The new drug combo and the placebo would yield the same proportion of survivors.

Ha: p1 > p2 , so **p1 - p2 > 0** A higher proportion of patients would survive if given the drug.

Step 2: Prepare
Significance level is .05
Choose test: Two Sample z-Test for Two Population Proportions
Check conditions:

1. Random sample? Assume; Independence WITHIN sample? Assume 1 patient doesn't affect another's patient's outcome WITHIN the groups.
2. Large Sample? Expected number of successes and failures at least 10 in each group?
   Treatment Group (received medication)
   Success: E = n1*p-hat = 110(.8) = 88 yes
   Failure: E = n1*p-hat = 110(.2) = 22 yes

   Placebo Group
   Success: E = n2*p-hat = 110(.8) = 88 yes
   Failure: E = n2*p-hat = 110(.2) = 22 yes

3. Large Population: Yes, there are a lot of people with leukemia
4. Independence BETWEEN groups? Yes, assume no interaction between groups.

Step 3: Compute
StatCrunch printout:

Two sample proportion summary hypothesis test:

p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: Difference in proportions
$H_0 : p_1 - p_2 = 0$
$H_A : p_1 - p_2 > 0$

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	101	110	88	110	0.11818182	0.046914568	2.5190857	0.0059

Step 4: Interpret

P-value = .0059 < .05 = alpha
Reject Ho, Accept Ha

The proportion of leukemia survivors was significantly higher in the group given the drug combo as compared to the placebo group.

Assigned Problems

	Read this section:	Do these problems:	Completed/ Total
1	Section 8.4: Comparing Proportions from Two Samples	8.4 Exercises, page 401: 61, 62, **StatCrunch:** 63, 69 Your work should still include Step 1: setting up the hypotheses, Step 2: choosing the test and checking the conditions, Step 3 “Compute” by using StatCrunch, and Step 4: writing a conclusion (interpret).	/4
2	Section 7.5: Comparing Two Population Proportions with Confidence	7.5 Exercises, page 350: 63, 73 **StatCrunch:** 67, 69, 71 For these 3 problems use StatCrunch to find the Confidence Intervals. Your work should still include reporting the confidence interval, writing a conclusion (interpret).	/5
3	Section 9.1: Sample Means of Random Samples	9.1 Exercises, page 460: 1, 2, 6, 8, 9, 10	/6
4	Section 9.3: Answering Questions about the Mean of a Population (Confidence Intervals)	9.3 Exercises, page 463: · Find t* values: 25, 26 · Confidence Interval interpretation: 19, 20 Find Confidence Intervals · By hand (no StatCrunch!): 29 · **Using StatCrunch:** 21, 22, 35 (Note: Type these small data sets into StatCrunch directly) · Concept Problems: 31, 33, 51	/1 1
5	Section 9.4: Hypothesis Testing for Means	9.4 Exercises, page 465: Hypothesis Tests: 41, 42 (The StatCrunch printouts are given for 41 and 42) **StatCrunch:** 37, 39, 44 You do not have to do any of the compute work by hand, but be sure to do all steps 1,2, 3(StatCrunch), and 4 of the hypothesis test. (Note: Type these small data sets into StatCrunch directly)	/5

6	Section 9.5: Comparing Two Population Means	9.5 Exercises, page 466: 53, 55, 59, 61, 71 **Use StatCrunch:** 63, 65, 69 - (The Excel files for these problems are on the wrightmath.info website.) You do not have to do any of the compute work by hand, but be sure to do all steps 1,2, 3(StatCrunch), and 4 of the hypothesis test.	/8
7	Section 11.1: Multiple Comparisons Read pages 534 – 536 (Stop at "Bonferroni Correction") Section 11.2: The Analysis of Variance	11.2 Exercises, page 563: 15, 16, 17, 19, 20, 21, 22	/7
8	Section 11.3: The ANOVA Test Section 11.4: Post Hoc Procedures	11.3 Exercises, page 565: 25, 26, 27, 28, 29, **StatCrunch:** 31, 35 (data is located on the website under the Math 247, StatCrunch. You do not have to do any of the compute work by hand, but be sure to do all steps 1,2, 3(StatCrunch), and 4 of the hypothesis test. **11.4 Exercises,** page 569: 49, 50	/9

Even Answers

8.62: You will get a smaller P-value when you have a larger sample. The test is more sensitive to small differences between two populations when you have a larger sample. Remember that this is where the question of "statistical significance" vs "clinical significance" comes in!

9.2 They are statistics because they come from a sample

9.6 a. Right-skewed. It cannot be Normal, because in a Normal distribution, 68% of the observations are within one standard deviation of the mean. This is not possible here, because phone calls cannot last negative minutes, so the distribution must be right-skewed.

b. Because the sample mean is unbiased, the mean will be about the same as the population mean: 3.25 minutes.

c. The standard deviation of this distribution, also called the standard error, is $4.2/\sqrt{100} = 0.42$ minute.

9.8 the sampling distribution of means

9.10 a. 60,000 is the expected sample mean, assuming an unbiased sample. The sample mean is an unbiased estimator of the population mean.

b. The standard error is $30{,}000/\sqrt{400} = 1500$.

9.20 a. Only iii. is correct. We are 95% confident that the population mean is between $\$18,546 - \$1398 = \$17,148$ and $\$18,546 + \$1398 = \$19,944$.

b. We cannot reject \$18,000 because it is within the interval.

c. The 95% interval is wider because a greater level of confidence requires a larger t^*.

9.22 *Random Sample*, *Independence*, *Big Population*, and *Large Sample* (*Normal Distribution*) are given.

a. ii. is correct (19.55, 21.00). Both i. and iii. are incorrect.

$$x \pm t^* \frac{s}{\sqrt{n}} = 20.275 \pm 3.182\left(\frac{0.457}{\sqrt{4}}\right) = 10.325 \pm 0.727$$

Inverse Cumulative Distribution Function

```
Student's t distribution with 3 DF
P( X <= x )        x
      0.975  3.18245
```

One-Sample T: Carrot Weight

Variable	N	Mean	StDev	SE Mean	95% CI
Carrot Weight	4	20.275	0.457	0.229	(19.547, 21.003)

b. No, we cannot reject the claim of 20 pounds, because 20 is within the interval.

9.26 a. Use $t^* = 2.797$.

b. It is larger because the sample size is smaller (so the distribution is wider) and also because the level of confidence is greater.

9.40 a. You should not be able to reject 5 pounds, because the confidence interval (4.9 to 5.3) did capture 5 pounds.

b. *Step 1*: H_0: $\mu = 5$, H_a: $\mu \neq 5$

Step 2: One-sample *t*-test: *Random Sample*, *Independence*, and *Big Population* are met. *Large Sample* (*Normal Distribution*) is given, $\alpha = 0.05$

Step 3: $t = 1.99$, p-value = 0.141.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{5.125 - 5}{0.126/\sqrt{4}} = 1.99$$

One-Sample T: Tomato Weight

Test of mu = 5 vs not = 5

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Tomato Weight	4	5.1250	0.1258	0.0629	(4.9248, 5.3252)	1.99	0.141

Step 4: Do not reject H_0. The mean is not significantly different from 5 pounds.

9.44 a. *Step 1*: H_0: $\mu = 128$, H_a: $\mu < 128$, where μ is the population mean weight of 20-year-old women.

Step 2: One-sample *t*-test: *Random Sample*, *Independence*, and *Big Population* are met. *Large Sample* (*Normal Distribution*): $n = 40 > 25$, $\alpha = 0.05$.

Step 3: $t = -2.53$, p-value = 0.008.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{122 - 128}{15/\sqrt{40}} = -2.53$$

One-Sample T

Test of mu = 128 vs < 128

N	Mean	StDev	SE Mean	95% Upper Bound	T	P
40	122.00	15.00	2.37	126.00	-2.53	0.008

Step 4: Reject H_0. The mean for vegetarian women is significantly less than 128.

b. *Step 3*: $t = -4.00$, p-value < 0.001.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{122 - 128}{15/\sqrt{100}} = -4.00$$

One-Sample T

Test of mu = 128 vs < 128

N	Mean	StDev	SE Mean	95% Upper Bound	T	P
100	122.00	15.00	1.50	124.49	-4.00	0.000

Step 4: Reject H_0. The mean for vegetarian women is significantly less than 128.

c. Larger *n*, smaller standard error (narrower and taller sampling distribution) with less area in the tails, as shown by the smaller p-value.

11.16 Group A,B,C would have F = 9.38 and Group G,H,K would have F = 25. How can we tell? Note that the difference in means (the center of the symmetric boxplots) is the same for each grouping. What is different is the amount of variation in each group. Since there is more variation WITHIN the groups in A,B,C, the F-statistic for A,B,C will be smaller

11.20 a. SS class = 3273.8 – 3247.2 = 26.6
b. 26.6/3 = 8.87, which rounded is 8.9.
c. 8.9/30.6 = 0.291, which rounded is 0.29.
d. If MS factor (MS class) is less than MS Error, the *F*-value will be less than 1.

11.22 a. The highest sample mean was for the sophomores, and the lowest mean was for the freshmen.
b. μ is the population mean number of TV hours per week. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$,
H_a: At least population one mean is different from another or class has an effect on TV hours.
c. $F = 0.29$
d. No. There was no random assignment. There could be confounding factors, such as age, hours of work for money, or living situation.

11.26 p-value = 0.833. Do not reject H_0. We do not have enough evidence to conclude that class affects the number of TV hours.

11.27 The pulse rates are not in three independent groups, so the condition of independent groups fails.

11.28 The boxplots show that the data are too skewed for us to use ANOVA with sample sizes under 25. Also, the standard deviations are too different, because the ratio of the largest to the smallest is 10.793/1.165, or about 9.3, which is larger than 2; thus the same-variance condition fails.