Carleton University – School of Mathematics and Statistics
STAT 2507 and BIT 2000 – Assignment #5 – Winter 2020

INSTRUCTIONS:

1. You must print this assignment and write your answers in the space provided by either printing this assignment or writing directly on this PDF with a tablet. If you print the assignment, you must use A4 or letter size paper. DO NOT CHANGE THE EXISTING SPACING OF THE QUESTION.
2. Assignments are to be uploaded to the course website on cuLearn as a single PDF file by **Wednesday, April 8 at 8:35am**. No late assignments will be accepted. No other file types will be accepted. Technical issues are not an excuse so don't wait until the deadline to submit.
3. The document should in the proper orientation so as not to require rotation. The document must be of sufficient resolution and writing must be legible. Pictures of your work are NOT an acceptable substitute for a scan.
4. You must show and explain all of your work. This includes explicitly defining random variables where necessary, writing out any formula that you use, and explaining your reasoning where applicable. No credit will be given for answers without justification.
5. **This assignment is intended to represent your individual knowledge. It is not a group assignment.**
6. When you save your PDF file, save it with the format: LastName.StudentNumber.A5.pdf .
7. Failure to follow these instructions will result in a grade of zero.

Name: Connor Raymond Stewart
Student ID: 101041125
Lab Section: STAT-2507-H4

A market research study is designed to determine if more than 25% of the residents of a large city would be a new type of cereal. A random sample of 600 people is selected and is given the new cereal to try. The 600 people are then asked if they would buy this cereal and 168 responded they would. At $\alpha = 0.05$, is there enough evidence to conclude that more than 25% of the residents of this city would buy the new type of cereal? Use the *p*-value method.

n = Sample Size = 600
X = # of Successes = 168
α = Significance level = 0.05

Claim: Proportion is greater then 25% (0.25)

The Claim is either the null hypothesis or the alternative hypothesis. The null hypothesis states the value is equal to the proportion mentioned in the Claim. The alternative hypothesis is that the proportion is greater then 25%

$H_0: p = 0.25$

$H_a: p > 0.25$

Since Ha is $p > 0.25$ & Ha has $>$, the test is Right-tailed

Sample proportion:

$$\hat{p} = \frac{X}{n} = \frac{168}{600} = 0.28$$

Test-Statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.28 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{600}}} = \frac{0.03}{\sqrt{\frac{1}{3200}}} = \frac{6\sqrt{2}}{5} \doteq 1.6971 \doteq 1.70$$

Since $n \geq 30$: $\sigma \doteq s$ & we can use the Z-Score test by CLT, $\hat{p} \sim N(p, \frac{pq}{n})$

To find the p-value:

we know:

Test is Right-tailed, $\hat{p}$, p, & Z

$$p = P(Z > 1.7) = 1 - P(Z \leq 1.7) = 1 - 0.9554 = \boxed{0.0446}$$

If the p-value is smaller then the significance level, Ho is rejected:

$\boxed{0.0446 < 0.05} \rightarrow$ Reject Ho, thus the null hypothesis can be rejected at the 5% level

So, we have sufficient evidence to support the Claim that $\hat{p} = 0.28$ is a significant increase from $p = 0.25$ at the 5% level.

$\therefore p > 0.25$ meaning we can conclude that more then 25% of the residents of this city would buy the new type of cereal.

A company that manufactures pavers used in residential landscaping. The company guarantees that the paver average weight is no more than 10kg. However, recently the company has received many complaints that the pavers are heavier than expected. In an effort to address the customers complaints, the quality control manager selected a random sample of 40 pavers. The average and standard deviation obtained from this sample are 10.3kg and 3.1 kg, respectively. At $\alpha = 0.10$, can it be concluded the pavers average weight is higher than 10kg? Use the critical value/rejection point method.

n = Sample Size = 40
$\bar{x}$ = Sample Mean = 10.3
S = Sample Standard deviation = 3.1
α = Significance level = 0.10

Claim: Mean is greater then 10 kg

The Claim is either the null hypothesis or the alternative hypothesis:
The null hypothesis States that the weight mean is the Same as the value Claimed (10kg).
The alternative hypothesis is the opposite; that the mean's greater then 10kg.

$H_0: \mu = 10$ kg
$H_a: \mu > 10$ kg

Knowns:
Since H_a is $\mu > 10$, the test's Right-tailed

Sampling dist.:

μ = mean = Claimed mean = 10 Kg

$\frac{\sigma}{\sqrt{n}}$ = Standard deviation (Sample)

Since $n > 30$ (the Sample's large), we can Approximate the pop. SD as the Sample SD;
$\sigma \doteq S$, we can also use Z-Score test by the CLT, $\bar{X} \sim N(\mu, \frac{\sigma^2}{40})$

So, the z-value is:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \doteq \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{10.3 - 10}{3.1/\sqrt{40}} \doteq 0.6121 = \boxed{0.61}$$

To find the rejection region:

$1 - P = \alpha = 0.1$ Such that $\boxed{Z_\alpha \doteq 1.28}$

thus any Z-value not within the rejection region 1.28 fails to reject H_0

Check Z:

$Z < Z_\alpha$ as $0.61 < 1.28$, thus $Z = 0.61$ dosn't fall in the rejection region of $Z_\alpha = 1.28$ w/ $\alpha = 0.10$, thus the $\boxed{\text{null hypothesis Cannot be rejected}}$ at the 10% level.

∴ we lack Sufficient evidence to Support the Claim that the Pavers average weight is higher than 10Kg.

A new hip replacement procedure is being evaluated at a big research hospital. This new procedure is less expensive; however, it might lead to a longer recovery time. A study was conducted to compare this new procedure with the current procedure. A randomly sample of 15 patients were operated on with the current procedure, and another sample of 12 patients were operated with the new procedure. The average and standard deviation of the recovery time for the sample that was treated with the current procedure are 35 days and 7.5 days, respectively. The average and standard deviation for the recovery time for the sample that was operated with the new procedure 39 days and 6.2 days, respectively. Is there enough evidence to conclude that, on average, patients who undergo a hip replacement using the new procedure need more time to recover? Test at $\alpha = 0.01$. Use the critical value/rejection point method.

n_1 = Sample Size (Old procedure) = 15
n_2 = Sample Size (new) = 12
$\bar{X}_1$ = Sample mean (old) = 35
$\bar{X}_2$ = Sample mean (new) = 39
S_1 = Sample SD (old) = 7.5
S_2 = Sample SD (new) = 6.2
α = Significance Level = 0.01

Knowns:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Claim: Patients who undergo the new procedure need more time to recover on average.

- The Claim is either the null hypothesis or the alternative hypothesis
  - the null hypothesis States that the time to recover's the Same for both Procedures.
  - the alternative hypothesis States that the time to recover is longer for the new Procedure.

$\left.\begin{array}{l} H_0: \bar{X}_1 - \bar{X}_2 = 0 \\ H_a: \bar{X}_1 - \bar{X}_2 < 0 \end{array}\right\}$ Since H_a is $<$, the test's left-tail

D_0 = Specified difference to test = 0
$n_1 < 30$ & $n_2 < 30$ So:
we may use the t-test | $\bar{X}_1 - \bar{X}_2 = 0$ in H_0 So $D_0 = 0$

So, the t-value is:

$$t = \frac{35 - 39 - 0}{\sqrt{\frac{7.5^2}{15} + \frac{6.2^2}{12}}} = -\frac{4}{\sqrt{\frac{1043}{150}}} \doteq \boxed{-1.5169}$$

To find the rejection Region:

$df = n_1 + n_2 - 2 = 15 + 12 - 2 = 25$

$t_{0.01/2,\,25} = t_{0.005/25} = 2.787$

Check t:

Since $|t| < t_{0.005,25}$ as $|-1.5169| < 2.787$ we know that at $\alpha = 0.01$ $\boxed{\text{we lack enough evidence to reject } H_0.}$

$\therefore$ We lack Sufficient evidence to Support the Claim that the Patients who undergo a hip replacment using the new procedure need more time to recover.

4. A researcher wants to test the claim that women who are exposed glycol ethers are more likely to have miscarriages than those who are not. Among the 40 women who are exposed glycol ethers that he randomly selected, 15 had miscarriages. Whereas, among the randomly selected 780 women who are not exposed glycol ethers, 130 had miscarriages. Is there enough evidence to conclude that the claim is correct? Use $\alpha = 0.01$. Use the p-value method.

n_1 = Sample Size (exposed) = 40
x_1 = # of miscarriages (exposed) = 15
n_2 = Sample Size (Not-exposed) = 780
x_2 = # of miscarriages (not-exposed) = 130
α = Significance Level = 0.01

Known:

$\hat{p} = \frac{x}{n}$, $Z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, $D_0 = 0$ ($p_1 - p_2 = 0$ means 0 difference)

Claim: Women who are exposed to glycol ethers are more likely to have miscarriages than those who are not.
- The claim is either the null hypothesis or the alternative hypothesis.
  - The null hypothesis states that women who are exposed to glycol ethers have the same rate of miscarrages as those who are not.
  - The alternative hypothesis states that women who are exposed to glycol ethers have a higher rate of miscarrages.

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 > 0$ ∴ right-tailed

find the proportions:

$\hat{p}_1 = \frac{x_1}{n_1} = \frac{15}{40} = 0.375$ = proportion exposed in sample

$\hat{p}_2 = \frac{x_2}{n_2} = \frac{130}{780} \doteq 0.1667$ = proportion not exposed in sample

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{15+130}{40+780} = \frac{29}{164}$

Since $n_1 > 30$ & $n_2 > 30$:
$\sigma \doteq s$ & we can use the Z-score test by the CLT
$(\hat{p}_1 - \hat{p}_2) \sim N\left(p_1 - p_2, \frac{p_1 q_1}{40} + \frac{p_2 q_2}{780}\right)$

$$Z = \frac{0.375 - 0.1667 - 0}{\sqrt{(29/164)(1 - 29/164)}\sqrt{\frac{1}{40}+\frac{1}{780}}} \doteq \boxed{3.3677} \doteq \boxed{3.37}$$

find the probability:

We Know:
test is right-tailed, $\hat{p}_1, \hat{p}_2, \hat{p}$, & Z

$P = P(Z > 3.3677) = 1 - P(Z \leq 3.3677) \doteq 1 - P(Z \leq 3.37) \doteq 1 - 0.9996 = \boxed{0.0004}$

If the P-value is smaller then the significance level, H_0 is rejected:

$0.0004 < 0.01 \rightarrow$ Reject H_0, thus $\boxed{\text{the null hypothesis can be rejected}}$ at the 1% level.

∴ $p_1 - p_2 > 0$ meaning we can conclude that women exposed to glycol ethers have a higher rate of miscarrages

5. Do left-hand and right-hand average reaction times differ? To test this, 12 right-handed subjects were randomly selected and the reaction times of their right and left hands, in thousands of seconds, were recorded. The following table contains the results.

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Right	189	96	112	166	115	131	169	158	110	102	182	157
Left	222	168	188	205	195	168	178	169	138	186	152	218
left − right	33	72	76	39	80	37	9	11	28	84	−30	61

At $\alpha = 0.10$, test the claim that there is a difference in the average reaction times of the left and right hands. Use the p-value method.

$n = 12 =$ Sample Size
$\alpha = 0.10 =$ Significance level
$\sum_{i=1}^{n} d_i = 500$
$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{500}{12} = \frac{125}{3}$
$\sum_{i=1}^{n} d_i^2 = 34002$

Claim: there's a difference in the average reaction time for left & right hands.
- The null hypothesis states there's no difference in reaction time
- The alternative hypothesis is that the average reaction times of those w/ left & right hands is different.

$H_0: \mu_d = 0$ So $\mu_d = 0$
$H_a: \mu_d \neq 0$ ∴ two-tailed Since H_a is $\neq$

Standard deviation of Difference:

$$S_d^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} d_i^2 - \frac{(\sum_{i=1}^{n} d_i)^2}{n}\right]$$

$$= \frac{1}{12-1}\left[34002 - \frac{500^2}{12}\right] = \frac{1}{11}\left[13168\tfrac{2}{3}\right] = \frac{39506}{33}$$

$$S_d = \sqrt{S_d^2} = \sqrt{\frac{39506}{33}} \doteq \underline{34.5999}$$

We Know:

Since $n < 30$:
We must use the t-test
$df = n - 1 = 12 - 1 = 11$
$\alpha = 0.10$

~~$t_{\alpha/2, df} = t_{0.10/2, 11} = t_{0.05,11} = 1.796$ $\boxed{1.80}$~~

Also:

$$t = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \frac{\left(\frac{125}{3}\right)}{(34.5999/\sqrt{12})} \doteq 4.1716 \doteq \boxed{4.17}$$

$t \leq -4.17$ & $t \geq -4.17$

$t = 4.17$ w/ $df = 11$
↳ this t value does not fall under any range of crit. values
It's larger then $t_{0.005,11} = 3.106$
$t > 2(3.106)$ So $P < 0.005$

Since $P < \alpha$, $\boxed{\text{We Can reject the null hypothesis } H_0 \text{ at } \alpha = 0.10}$

∴ We have enough evidence to Conclude there's a difference in the average reaction times of the left & right hands.

6. A fish-processing company is concerned about the shelf life of its new cat food. A random sample in Halifax and another sample in Dartmouth were examined. The results are summarized below.

	Halifax	Dartmouth
Sample Size	35	40
Sample Mean (months)	13	11
Sample Standard Deviation	1.44	1.52

Is there a difference in the average shelf life between the two locations? Conduct a hypothesis test 0.05 significance level. Use the critical value/rejection point method.

n_1 = Sample Size (Halifax) = 35
n_2 = Sample Size (Dartmouth) = 40
$\bar{X}_1$ = Sample mean (Halifax) = 13
$\bar{X}_2$ = Sample mean (Dartmouth) = 11
S_1 = Sample Standard deviation (Halifax) = 1.44
S_2 = Sample Standard deviation (Dartmouth) = 1.52
α = Significence level = 0.05

Claim: there's a difference in the average Shelf life between the two locations.
- The null States there's no difference
- The alternative States there's a difference

$H_0: \mu_1 - \mu_2 = 0 \therefore D_0 = 0$
$H_a: \mu_1 - \mu_2 \neq 0 \therefore$ two-tailed Since H_a is $\neq$

We Know:

n_1 & n_2 are greater then 30, So:

$\sigma \doteq s$ & we can use the Z-score test by CLT; $(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{35}, \frac{\sigma_2^2}{40}\right)$

$D_0 = 0 \rightarrow$ Specified difference to test

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} =$$

So, the Z-value's:

$$Z = \frac{13 - 11 - 0}{\sqrt{\frac{1.44^2}{35} + \frac{1.52^2}{40}}} = \frac{2}{\sqrt{\frac{5119}{43750}}} \doteq \boxed{5.8469} \doteq \boxed{5.85}$$

To find the rejection Region:

$1 - P = \alpha = 0.05$ Such that $Z_\alpha = 1.645$ (Z-Score between 1.64 & 1.65)

thus any Z-value not within the rejection region 1.645 fail to reject Ho

Check Z:

$Z_\alpha < Z$ as $1.645 < 5.85$ thus $Z = 5.85$ falls within the rejection region of $Z_\alpha = 1.645$ W/ $\alpha = 0.05$, thus the null hypothesis Can be rejected at the 5% level.

$\therefore$ We have Sufficient evidence to Support the Claim that the average Shelf life between the two locations is different.

A group of university students are interested in comparing the average age of cars owned by students and the average age of cars owned by faculty. They randomly selected 25 cars that are own by students and 20 cars that are owned by faculty. The average age and standard deviation obtained from the students' cars are 6.78 years and 5.21 years, respectively. The sample of faculty cars produced a mean and a standard deviation of 5.86 years, and 2.72.

A. Construct and interpret a 90% confidence interval for the difference between the average age of students' cars and average age of faculty cars.

B. At $\alpha = 0.05$, is there enough average to conclude that on average faculty cars are newer than students' cars? Use the p-value method.

$n_1 = \#$ Student Cars $= 25$
$n_2 = \#$ factuly Cars $= 20$ } $n < 30$ So use t-test

Claim: faculty cars have a lower mean age then Student Cars
- The null States theres no differ.
- The alternative States there's a diff.

$H_0: \overline{X_1} - \overline{X_2} = 0$
$H_a: \overline{X_1} - \overline{X_2} > 0$

$\overline{X_1}$ = average age of Students Cars $= 6.78$
$\overline{X_2}$ = average age of faculty Cars $= 5.86$
S_1 = Sample Standard deviation of Student Cars age $= 5.21$
S_2 = Sample Standard deviation of faculty Cars age $= 2.72$

A) $df = \min(25-1, 20-1) = 19$

$\frac{S_1}{S_2} = \frac{5.21^2}{2.72^2} \doteq 3.6689 > 3$ So $\sigma_1^2 \neq \sigma_2^2$, use s_1^2 & s_2^2

$t_{0.1/2,43} = t_{0.05,19} = 1.721$

to find the Confidence interval: $\left(E = t_{\alpha/2} \cdot S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{X_1}-\overline{X_2} \pm E\right)$

$(\overline{X_1}-\overline{X_2}) \pm t_{0.05,19}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} = (6.78-5.86) \pm (1.729)\sqrt{\frac{5.21^2}{25}+\frac{2.72^2}{20}} \doteq$

$0.92 \pm 2.0861 \Rightarrow \boxed{(-1.1661, 3.0061)}$

thus, the Confidence interval of 90% Ranges from the values Shown for the difference between Student & faculty Car ages.

B) α = Significance level $= 0.05$

$S = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(25-1)(5.21)^2+(20-1)(2.72)^2}{25+20-2}} \doteq 4.29177$

$t = \frac{\overline{X_1}-\overline{X_2}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{6.78-5.86}{4.29177\sqrt{1/25+1/20}} \doteq \underline{0.7145}$

$t = 0.7145$ & $df = 19$ thus: $t \leq -0.7145$ & $t \geq 0.7145$

this is greater then $t_{0.100,19} = 1.328$ So $P > 0.1$

$0.1 > 0.05$ thus, P falls outside the rejection range & $\boxed{\text{the null hypothesis } H_0 \text{ Cannot be rejected}}$

∴ We Can Say that at $\alpha = 0.05$ we dont have enough evidence to Conclude faculty Cars are newer on average then Student Cars.