Assignment 4 (Week 7 - 8)

STAT 2601 - Business Statistics (2024 Fall)
School of Mathematics and Statistics, Carleton University

Due Date and Time: Wednesday 13 November 2024, before 10:00 am
Total Marks: 39

Q1: [8]

1. Formulate the null and alternative hypotheses to test the claim. [1]

   Solution: Let p be the population proportion of Canadians who support increased immigration levels.

   - H_0: $p \leq 0.6$ (or $p = 0.6$)[0.5]
   - $H_a: p > 0.6$ [0.5]

2. Calculate the test statistic and p-value. [2.5]

   Solution: The test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

   where

   - $\hat{p} = \dfrac{1980}{3000} = 0.66$ is the sample proportion.[0.5]

   Substitute these values in the test statistic formula:

$$z = \frac{0.66 - 0.60}{\sqrt{\frac{0.60\times(1-0.60)}{3000}}}[0.5] = \frac{0.06}{0.008944} \approx 6.71 \quad [0.5]$$

   Using standard normal distribution table for $z = 6.71$, the p -value is extremely small (much less than 0.01):

$$\text{p -value} = P(Z > 6.71) = P(Z < -6.71) \quad [0.5] \approx 0 \,(\text{or}\, p - value < 0.0002) \quad [0.5]$$

3. Write the decision rule and your decision (reject or fail to reject the null hypothesis) using p-value approach and justify your answer. Explain the managerial decision in the context of public support for immigration in Canada. [3]

   Solution:

   - Decision Rule: If $p - value < \alpha$, we reject H_0, otherwise we fail to reject H_0. [1]
   - Decision: Since the p-value is significantly smaller than the significance level of 0.01, we reject the null hypothesis [1] and conclude that there is strong evidence to support the claim that more than 60% of Canadians favor increased immigration levels [1].

4. What assumptions and conditions are required for the validity of this test? [1.5]

   Solution:

   (i) The sample of 3,000 Canadian residents surveyed is a random sample. [0.5]

   (ii) $np_0 = 3,000 \times 0.6 = 1800 > 5$[0.5], $nq_0 = 3,000 \times 0.4 = 1200 > 5$ [0.5].

Q2: [14]

1. Can we assume that the variances of the purchase amounts for customers using Twitter are equal to those for customers using Facebook? Explain. [1]

   Solution:

- For Twitter:

$$s_T^2 = 15^2 = 225$$

- For Facebook:

$$s_F^2 = 18^2 = 324$$

The ratio of the two variances is $\frac{s_T^2}{s_F^2} = \frac{225}{324} = 0.694$ [0.5]. This value falls between 1/3 and 3, therefore we can assume equal population variances [0.5].

2. Construct and interpret a 90% confidence interval for the difference in average purchase amounts between customers. [6]

Solution: Since we are assuming equal variances, the confidence interval for the difference in means is given by the following formula:

$$[\bar{x}_T - \bar{x}_F] \pm t_{\alpha/2}(n_T + n_F - 2) \cdot s_p\sqrt{\frac{1}{n_T} + \frac{1}{n_F}} \quad [0.5]$$

where:

- $n_T = 17$, the number of customers using Twitter
- $n_F = 15$, the number of customers using Facebook
- degrees of freedom $df = n_T + n_F - 2 = 17 + 15 - 2 = 30$ [0.5]
- $\bar{x}_T - \bar{x}_F = 128 - 117 = 11$, the difference between sample means of two social media platforms [0.5].
- Significant level $\alpha = 0.1$. Thus, $\alpha/2 = 0.05$
- Use the t-Table at $\alpha/2 = 0.05$ and $df = 30$ to get the critical value: $t_{\alpha/2}(n_T + n_F - 2) = t_{0.05,30} \approx 1.645$[1]
- The pooled standard deviation is

$$s_p = \sqrt{\frac{(17-1)\cdot 225 + (15-1)\cdot 324}{17+15-2}} = \sqrt{\frac{3600 + 4536}{30}} \approx \sqrt{271.2} \approx 16.47 \quad [1]$$

- The margin of error is

$$t_{\alpha/2}(n_T + n_F - 2) \cdot s_p\sqrt{\frac{1}{n_T} + \frac{1}{n_F}} = 1.645(16.47)\sqrt{0.0588 + 0.06667} \approx 1.645(16.47)(0.354246) \approx 9.60 \quad [0.5]$$

Thus, the 90% confidence interval is:

$$11 \pm 9.60 \quad \Rightarrow \quad (1.4, 20.6) \quad [0.5]$$

It can be said with 90% confidence [0.5] that the average purchase amounts of customers who engaged with Twitter ads is higher [0.5] than those of who engaged with Facebook by amount lying anywhere between approximately 1.4 and 20.6 dollars. [0.5]

3. Test if there is sufficient evidence to suggest that advertising promotions on Twitter lead to a higher average purchase amount among customers compared to those on Facebook using critical value method at 10% level of significance. Write hypotheses, critical value, decision rule, test statistic, decision with justification and managerial statement in the business context. [4.5]

   Solution:

   - $H_0 : \mu_T = \mu_F$ (or $\mu_T \leq \mu_F$).[0.5]
   - $H_a : \mu_T > \mu_F$.[0.5]

   The test statistic is:

   $$t = \frac{\bar{x}_T - \bar{x}_F - (\mu_T - \mu_F)}{s_p\sqrt{\frac{1}{n_T} + \frac{1}{n_F}}} \quad [0.5]$$

   where

   - μ_T = average purchase amount for Twitter customers
   - μ_F = average purchase amount for Facebook customers
   - The difference in means is given by $\mu_T - \mu_F = 0$.

   Thus,

   $$t = \frac{128 - 117 - 0}{16.47 \cdot \sqrt{0.0588 + 0.06667}} \approx \frac{11}{5.83} \approx 1.885 \quad [0.5]$$

   Critical Value: t critical value for $df = 30$ at $\alpha = 0.10$ is approximately $t_{0.10,30} \approx 1.282$.[0.5]

   Decision Rule: Reject H_0 if $t > 1.282$. [0.5]

   Decision: Since the test statistic 1.885 is greater than the critical value 1.282 [0.5], we reject the null hypothesis based on sample evidence at 10% significance level [0.5] implying that advertising promotions on Twitter lead to a higher average purchase amount compared to Facebook.[0.5]

4. What is the p-value for this test? Write your decision rule and decision using p-value approach. [2.5]

   Solution:

   - $p - value = P(t > 1.885)$[0.5] $\approx .025 < p - value < 0.05$ (using t table)[0.5].
   - Decision Rule: Reject H_0 if $p - value < \alpha$, otherwise we fail to reject H_0.[0.5]
   - Decision: Since $.025 < p - value < 0.05 < \alpha = 0.10$ [0.5], we reject H_0[0.5].

Q3: [9]

1. Formulate the null and alternative hypotheses for this scenario. [1]

   Solution:
   $H_0 : p_1 = p_2$ [0.5]
   $H_a : p_1 \neq p_2$ [0.5]

2. Using a significance level of 0.05, conduct a hypothesis test comparing the two population proportions. Calculate the test statistic and p-value and determine whether to reject or fail to reject the null hypothesis. [8]

   Solution:
   For AliExpress sample:

   - Sample size $(n_1) = 420$
   - Number satisfied $(x_1) = 210$
   - Sample proportion $(\hat{p}_1) = \frac{210}{420} = 0.5$ [1]

   For Temu sample:

   - Sample size $(n_2) = 480$

- Number satisfied $(x_2) = 260$
- Sample proportion $(\hat{p}_2) = \frac{260}{480} \approx 0.5417$ [1]

The test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE(\hat{p}_1 - \hat{p}_2)} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} \quad [1]$$

where $\hat{p}$ is the pooled proportion that is calculated as follows:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{210 + 260}{420 + 480} = \frac{470}{900} \approx 0.5222 \quad [1]$$

The standard error (SE) for the difference in proportions is given by:

$$SE = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.5222 \times (1 - 0.5222)\left(\frac{1}{420} + \frac{1}{480}\right)} \approx \sqrt{0.0011} \approx 0.033 \quad [1]$$

Now substituting back into the test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE(\hat{p}_1 - \hat{p}_2)} = \frac{0.5 - 0.5417}{0.033} \approx -1.26 \quad [1]$$

p-value:

$$p\text{-value } = p(Z < -1.26) + P(Z > 1.26) = 2 \times P(Z < -1.26) \approx 2 \times 0.1038 \approx 0.2076 \quad [1]$$

Since p-value $\approx 0.2076 > 0.05$ [0.5], we fail to reject the null hypothesis and conclude that there is insufficient evidence to suggest a difference in customer satisfaction regarding pricing between AliExpress and Temu.[0.5]

Q4: [8] Use the following two methods **Excel** to solve this question

Excel Instructions

- **Method 1**
  - Create a new column represents the difference between the paired observations (sample1 − sample2).
  - Count the number of pairs, n, using `=COUNT(range)` where the range corresponds to your differences column.
  - Calculate the mean (`=AVERAGE(range)`) and standard deviation (`=STDEV.S(range)`) of the differences.
  - Calculate the Test Statistic, t-statistic, for two dependent means (matched pairs).
  - Calculate the P-Value `=T.DIST.2T(ABS(t-statistic), n-1)`.
- **Method 2**
- **File** > **Option** > **Add-ins** > **Option** > **Manage:** Excel Add-ins - Go > add **Analysis ToolPak**
- **Data** > **Data Analysis** > **t-Test: Paired Two Sample for Means**.
- Input the All-Season MPG values in **Variable 1 Range** and Winter MPG values in **Variable 2 Range**.
- Set **Hypothesized Mean Difference:** 0 and **Alpha:** 0.05.

Use a significance level of 0.05 to conduct a paired t-test and decide whether to reject or fail to reject the null hypothesis that there is no difference in fuel efficiency between cars equipped with winter tires and those with all-season tires. Report all results from both methods.

Solution: The excel output [1 for creating column `Difference`], [3 for results from `method 1`], and [3 for results from `method 2`]

All-Season MPG	Winter MPG	Difference
20	21	-1
22	25	-3
18	20	-2
24	24	0
27	29	-2
28	28	0
24	25	-1
21	20	1
22	21	1
20	22	-2
18	19	-1
26	28	-2

Results from Method 1	
sample size	12
Average of All-Season	22.5
Average of Winter	23.5
Average of difference	-1
Standard devation of All-Season	11.18181818
Standard devation of Winter	12.27272727
Standard devation of Difference	1.279204298
Test Statistic	-2.708012802
P-value	0.020363073

Results from Method 2		
t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	22.5	23.5
Variance	11.18181818	12.27272727
Observations	12	12
Pearson Correlation	0.931240394	
Hypothesized Mean Difference	0	
df	11	
t Stat	-2.708012802	
P(T<=t) one-tail	0.010181537	
t Critical one-tail	1.795884819	
P(T<=t) two-tail	0.020363073	
t Critical two-tail	2.20098516	

- Null Hypothesis (H_0) : There is no difference in fuel efficiency between cars equipped with winter tires and those with all-season tires ($\mu_d = 0$).
- Alternative Hypothesis (H_a) : There is a difference in fuel efficiency ($\mu_d \neq 0$).

Since the p-value is significantly smaller than the significance level of 0.05, we reject the null hypothesis [0.5]. This indicates that there is sufficient evidence to suggest that there is a difference in fuel efficiency between cars equipped with winter tires and those with all-season tires [0.5].