Lab 4 (Week 7 - 8)

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

Date: November 4 - 8, 2024

Q1: Cannabis Use for Youth (age of 16 - 19)

In 2020 and 2021, two surveys were conducted to assess the use of cannabis in youth age groups between 16 and less than 19. In 2020, a random sample of 500 was taken and it was found that 220 were cannabis users. In 2021, a random sample of 300 was taken and it was found that 110 were Cannabis users. Several studies on the legalization of cannabis claim that the legalization act prevents the age group younger than 19 (i.e., minimum age to buy, use, possess or grow recreational cannabis) from accessing Cannabis at the legal retail store. Do you think the proportion of cannabis users in youth (16 and below 19) has significantly dropped from 2020 to 2021?

Perform all calculations in the **Q1** worksheet of the excel file **Lab 4**.

(a) Write the hypotheses defining p_1 and p_2.

(b) Perform the test of hypothesis in EXCEL at .05 level of significance (calculate $\hat{p}_1$, $\hat{p}_2$, $\hat{q}_1$, $\hat{q}_2$, $\hat{p}$, $\hat{q}$, z).

(c) Calculate the critical value in EXCEL and make decision using critical value approach.

(d) Calculate the p-value in EXCEL and make decision using p-value approach.

(e) Calculate standard error, margin of error, and finally 95% confidence interval in EXCEL for the difference between the young population proportion of cannabis users between 2020 and 2021.

Solution:

(a) **Hypotheses:**

$$H_0 : p_1 \leq p_2 \quad \text{or} \quad p_1 = p_2$$
$$H_A : p_1 > p_2$$

Here, p_1 and p_2 represent the young (between 16 and less than 19) population proportion of cannabis users in 2020 and 2021.

(b) (i) $\hat{p}_1 = \frac{x_1}{n_1}, \quad = \frac{\textbf{B2}}{\textbf{D2}} = 0.44$ (**CELL: E2**).

(ii) $\hat{p}_2 = \frac{x_2}{n_2}, \quad = \frac{\textbf{B3}}{\textbf{D3}} \approx 0.37$ (**CELL: E3**).

(iii) $\hat{q}_1 = 1 - \hat{p}_1, \quad =$ **1 - E2** $= 0.56$ (**CELL: F2**).

(iv) $\hat{q}_2 = 1 - \hat{p}_2, \quad =$ **1 - E3** $= 0.63$ (**CELL: F3**).

(v) $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \quad = \frac{\textbf{B2 + B3}}{\textbf{D2 + D3}} = 0.41$ (**CELL: B7**).

(vi) $\hat{q} = 1 - \hat{p}, \quad =$ **1 - B7** ≈ 0.59 (**CELL: B8**).

(vii) $z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} \quad = \frac{\textbf{E2 - E3}}{\sqrt{\textbf{B7*B8*}(\frac{1}{\textbf{D2}} + \frac{1}{\textbf{D3}})}} \approx 2.04$ (**CELL: B9**).

(c) Critical Value:
= **NORM.S.INV(0.95)** ≈ 1.645 (**CELL: B10**).

Decision: Since test statistic, $z = 2.04 > z_{cv} = 1.645$, we may reject H_0.

(d) p-value:

= **(1 - NORM.S.DIST(B9, TRUE))** ≈ 0.02 (**CELL: B11**).

Decision: Since p-value, $= 0.02 < \alpha = 0.05$, we may reject H_0.

(e) Confidence Interval:

(i) Standard Error, $\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}, \quad = \sqrt{\frac{\textbf{E2*F2}}{\textbf{D2}} + \frac{\textbf{E3*F3}}{\textbf{D3}}} \approx 0.04$ (**CELL: B12**).

(ii) Margin of Error, = **1.96 * B12** ≈ 0.07 (**CELL: B13**).

(iii) Lower Confidence Limit, = **(E2 - E3) - B13** ≈ 0.004 (**CELL: B14**).

(iv) Upper Confidence Limit, = **(E2 - E3) + B13** ≈ 0.14 (**CELL: B15**).

Q2: MBA Salary

MBA Salary worksheet of the excel file **Lab 4** contains the two random samples of annual salaries, in thousands of dollars, earned by individuals who graduated with MBAs in 2015 and 2016 from a business school in Canada. We would like to determine whether the average of salaries for 2015 MBA graduates is higher than that of for 2016 MBA graduates.

(a) Use EXCEL to create a side-by-side box plot for 2015 and 2016 graduates. Can we assume population annual salaries of MBA graduates in 2015 and 2016 are normally distributed with equal variances? Justify.

(b) Write the hypotheses defining μ_1, μ_2.

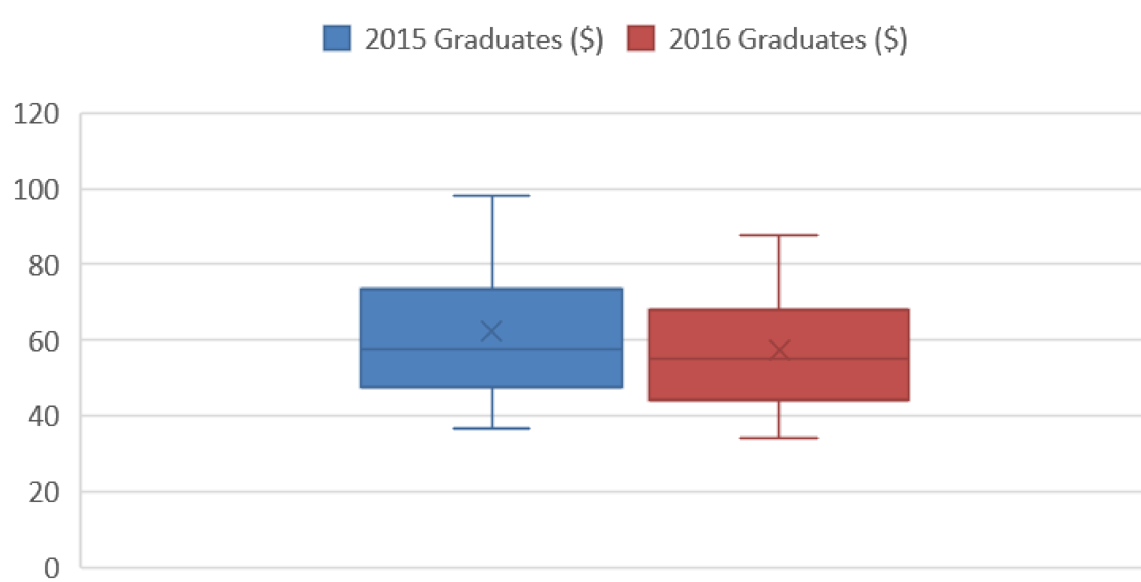
(c) Based on the assumption made in step (a), perform an appropriate parametric test (i.e., t) in EXCEL at 5% level of significance. Make the decision using both critical value and p-value approach.

(d) Calculate 95% confidence interval for the difference between the average salaries of 2015 and 2016 graduates.

Solution:

(a) Box Plot: The sampled annual salaries of 2015 and 2016 MBA graduates are fairly symmetric, so it is reasonable to assume that the population annual salaries of 2015 and 2016 graduates are normally distributed with equal variances.

[EXCEL: Select the data (A1:B16) > Insert > Recommended Charts > All Charts > Box and Whisker > OK > (+) (Chart Elements) > Legends]



(b) **Hypotheses:**

$$H_0 : \mu_1 \leq \mu_2 \quad \text{or} \quad \mu_1 = \mu_2$$
$$H_A : \mu_1 > \mu_2$$

Here, μ_1 and μ_2 represent the average annual salaries of 2015 and 2016 MBA graduates.

(c) Test of μ_1, μ_2 (equal variances):

[EXCEL: Data > Data Analysis > t-Test: Twos-Sample Assuming Equal Variances > Variable 1 Range ($A$1:$A$15) > Variable 2 Range ($B$1:$B$16) > Hypothesized Mean Difference (0) > Labels (✓) > OK]

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*2015 Graduates ($)*	*2016 Graduates ($)*
4	Mean	62.3	57.41333333
5	Variance	359.9015385	256.2840952
6	Observations	14	15
7	Pooled Variance	306.1739753	
8	Hypothesized Mean Difference	0	
9	df	27	
10	t Stat	0.751517724	
11	P(T<=t) one-tail	0.229421103	
12	t Critical one-tail	1.703288446	
13	P(T<=t) two-tail	0.458842205	
14	t Critical two-tail	2.051830516	
15			
16			
17	LCL	-8.455150129	
18	UCL	18.22848346	

Decision (Critical Value Approach): Since $t_{cal} = 0.75 < t_{cv} = 1.70$, we fail to reject the null hypothesis based on sample evidence.

Decision (p-value Approach): Since $p - \text{value} = 0.23 > \alpha = 0.05$, we fail to reject the null hypothesis based on sample evidence.

(d) 95% Confidence Interval for $\mu_1 - \mu_2$:

(i) Lower Confidence Limit, $= (B4 - C4) - B14 * \sqrt{B7 * (\frac{1}{B6} + \frac{1}{C6})} \approx -8.46.$

(i) Upper Confidence Limit, $= (B4 - C4) + B14 * \sqrt{B7 * (\frac{1}{B6} + \frac{1}{C6})} \approx 18.23.$