DANA 4800 Homework #4

1. A survey has been conducted to determine what percentage of Canadian who have smoked Marijuana during their teenage years (13-19 years old). A random sample of 1,000 Canadians was drawn, and it is reported that 210 of them did smoke marijuana during their teenage years.

   a) Identify the parameter of interest. **[2 marks]**

      Subjects = Canadians

      Variable = whether or not they have smoked marijuana during their teenage years

      Type = categorical variable

      Answer = Define p as the proportion of all Canadians who smoked marijuana during their teenage years

   b) Estimate the above parameter using a 95% confidence interval. **[2+2 marks]**
      Note: Make sure you understand that checking assumptions is part of this procedure.

$$\text{Confidence Level } 95\% => Z = 1.96$$

$$\bar{p} \pm Z * \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\bar{p} = \frac{210}{1000} = 0.21$$

$$0.21 \pm 1.96 * \frac{\sqrt{0.21(1-0.21)}}{\sqrt{1000}}$$

$$\mathbf{0.21} \pm 0.0252$$

      Conditions checking:

      A. n = 1000, which is bigger than the requirement of 30,
      B. n*p = 1000*0.21 = 210, which is bigger than the requirement of 5,
      C. n*(1-p) = 1000*(1-0.21) = 790, which is bigger than the requirement of 5.

      Since all three conditions are met, the approximation is good.

2. One way to know the efficiency of international airports is to measure the percentage of flights that leave the airport on time on a daily basis**. From a random sample of 100 days**, it was found that an average of 81.5% of flights leave YVR (Vancouver International Airport) on time and a standard deviation of 4.5%. From past experience, it is known that the percentage of flights that leave YVR on time has a Normal distribution.

   a) Identify the subjects of interest. **[1 mark]**

      Subjects = days

   b) Provide a description of the parameter of interest in this study. **[2 marks]**

      Define μ as the average percentage (an average of 81.5%) of all flights that leave YVR on time on a daily basis.

c) Construct a 98% confidence interval of the parameter in part (a). **[3 marks]**

98% Confidence Interval of μ
98% CI means the middle-area is 0.98. That also means we have two 2% on either sides.
Using CA = 0.0099 (closest to 0.01) and look up the Z-table, we get a Z-score of 2.33.

$$\bar{p} \pm Z * s/\sqrt{n}$$

$$0.815 \pm 2.33 * \frac{0.045}{\sqrt{100}}$$

0.815 ±0.0105

1. According to previous study in 2016, 42% of Canadian employees had extended health care coverage from their employers. With organization putting more focus on the employees' physical and mental health, an HR personnel would like to see if the number has gone up since. A recent study found that 180 out of 400 randomly sampled workers have the extended health care coverage from their employer's health care plan.

a) Identify the parameter of interest. **[2 bonus marks]**
Define $\boldsymbol{p}$ as the proportion of all Canadian employees who have extended health care coverage from their employers

b) Set up the null hypothesis and the alternative hypothesis. **[2 bonus marks]**
Ho <=42%
Ha >42%

c) Calculate the test statistic and find the p-value. **[4 bonus marks]**
Note: Make sure you understand checking conditions is part of the procedure.

$$\bar{P} = \frac{180}{400} = 0.45$$

$$\mathrm{T.S.} = \frac{\bar{p} - p0}{\sqrt{\frac{p0(1-p0)}{n}}}$$

$$p - \text{value} = 0.112.[1]$$

Checking's
1) $\boldsymbol{n} = \mathbf{400}$, $\boldsymbol{n} > 30$,
2) $\boldsymbol{n} \times \boldsymbol{p0} = \mathbf{400 \times 0.42 = 168}$, $\boldsymbol{n} \times \boldsymbol{p0} > 5$,
3) $\boldsymbol{n} \times (\mathbf{1} - \boldsymbol{p0}) = \mathbf{400 \times 0.58 = 232}$, $\boldsymbol{n} \times (\mathbf{1} - \boldsymbol{p0}) > 5$.

d) Make an appropriate conclusion in plain English. Use α = 0.01. **[2 bonus marks]**

p-value(11%) is bigger than the 1% level of significance, we do not have enough statistical evidence to reject the null hypothesis and conclude that proportion of all Canadian employees who have the extended health care coverage from their employers is not significantly higher than 42%.

2. At ABC University, the average scholarship examination scores for freshman applications have been 900. Every year, the Associate Dean uses a sample of applications to determine whether the average examination score for the new freshman applications is different from 900. A random sample of 200 applications was drawn this year and the average score was found to be 935 and a standard deviation of 180. Historically, it is known that the exam score has a left-skewed distribution.

   a) Identify the parameter of interest. **[2 marks]**

   Define $\boldsymbol{\mu}$ as the average scholarship examination score of all applications at ABC University

   b) Set up the null hypothesis and the alternative hypothesis. **[2 marks]**

   $\boldsymbol{H}0{:}\boldsymbol{\mu} = \mathbf{900}$
   $\boldsymbol{Ha}{:}\boldsymbol{\mu} \neq \mathbf{900}$

   c) Calculate the test statistic and find the p-value. **[2 marks]**

$$\bar{X} = \mathbf{935},$$
$$\boldsymbol{s} = \mathbf{180}$$
$$\boldsymbol{n} = \mathbf{200}.$$

$$\boldsymbol{T.S.} = \frac{\bar{X} - \mu 0}{s/\sqrt{n}} = \frac{935-900}{180/\sqrt{200}} = +\mathbf{2.75} = 0.9970$$

$$\text{P-value} = \mathbf{2} \times (\mathbf{0.0030}) = \mathbf{0.0060}$$

   d) Note that the p-value is only an approximation. Briefly explain why. Also briefly justify why we can still make a valid conclusion in the following part? **[1+2 marks]**

   The p-value is an approximation because the exam scores are not normally distributed. However, since the sample size is greater than 30, the sampling distribution of the sample mean is **approximately normal (left-skewed)** due to the Central Limit Theorem.

   e) Draw a conclusion so that the Dean could understand. Use 5% significance level. **[2 marks]**

   P-value (0.6%) is less than the 5% level of significance, we have enough statistical evidence to reject the null hypothesis and conclude that the average scholarship examination score of all freshman applications at ABC University is significantly different from 900.