112-2 BM508 Mid-Term Exam Spring 2024 Due 4/30/2024 @11:59pm

* Please specify the software you used for the questions. **Python**
* Please write down and submit your answers using Word document and attach the outputs generated from your software.
* Do not simply send an Excel output as your homework without organizing it in the Word document.
* If you are using R, please also submit your codes.
* Late Penalty: 5 points per day!! (If you have any question that may delay your submission, please let me know in advance.)

The file “Park*Spend5*” contains sample data from tourists of an amusement park. The file contains the following data on visits by families to the park: amount spent, size of the family, the distance the family lives from the park (the gate attendee asks for the postal code of each family entering the park), whether or not the family has a park membership (1 = yes, 0 = no), expected satisfaction score at entry (pre score), and the post-visit satisfaction score at the exit (post score). Each family were asked about their expected satisfaction of the visit upon entry. Then a follow-up survey was conducted for their satisfaction after the visit. The satisfaction was measured in a scale from 0-100 and was treated as a continuous variable. At the significance level of 0.05, please analyze the following questions.

Question 1

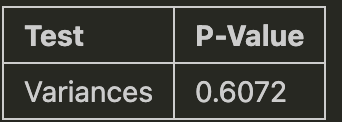
Investigate whether there is a difference in the distance the family lives from the Park between member status. [28%]

1. the mean and standard deviation for each group (8%)

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

1. Is there a difference in variance between the two groups? Formulate the statistical hypotheses. (2%) Conduct the test, (2%) and provide the p-value. (2%) What is your conclusion? (2%)



H0: Variances are equal between the two groups.

H1: Variances are not equal between the two groups.

因為P-Value = 0.6072 >= 0.05 無法拒絕H0，表示變異數沒有顯著差異

1. Is there a difference in mean between two groups? Formulate the statistical hypotheses, conduct the test based on 2), and provide the p-value (6%)

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

𝐻0 : The means are equal between the two groups

𝐻1​ : The means are not equal between the two groups

1. The conclusion from the hypothesis testing for 3) (2%)

因為P-Value = 0.5295 > 0.05 無法拒絕H0，表示平均數沒有顯著差異

1. What is the 95% confidence interval of population mean difference? Provide the result and the interpretation (4%)

一張含有 文字, 字型, 螢幕擷取畫面, 圖形 的圖片

自動產生的描述

區間包括0，表示會員和非會員之間距公園距離的真實平均差異可能為零，意味著沒有顯著差異，在 95% 的信賴水準下，無法保證兩者存在顯著差異。

Question 2

Investigate whether there is a difference in the spending between member status. [28%]

1. the mean and standard deviation for each group (8%)

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

1. Is there a difference in variance between the two groups? Formulate the statistical hypotheses. (2%) Conduct the test, (2%) and provide the p-value. (2%) What is your conclusion? (2%)

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

H0: Variances are equal between the two groups.

H1: Variances are not equal between the two groups.

由於P值 = 0.0031小於0.05拒絕H0，有證據證明會員和非會員之間的支出變異數存在顯著差異

1. Is there a difference in mean between two groups? Formulate the statistical hypotheses, conduct the test based on 2), and provide the p-value (6%)

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

H0: The means are equal between the two groups.

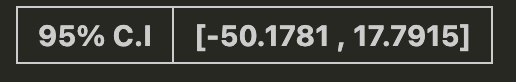
H1: The means are not equal between the two groups.

P-Value = 0.3469

1. The conclusion from the hypothesis testing for 3) (2%)

由於P值大於0.05無法拒絕H0，表示會員和非會員之間的支出平均不存在顯著差異

1. What is the 95% confidence interval of population mean difference? Provide the result and the interpretation (4%)



區間包括0，表示會員和非會員之間支出的真實平均差異可能為零，意味著沒有顯著差異，在 95% 的信賴水準下，無法保證兩者存在顯著差異。

Question 3

The park has been promoting the membership to families living around the park. The goal was to achieve 70% for families residing within 50 kilometers from the park. Based on sampling, investigate whether the goal is achieved. (Note: We are only concerned whether the goal is not met!) [10%]

1. How many families from the data are there with the distance less than 50km and what is the proportion of membership from them? (2%)

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

1. Is the goal achieved? Develop the null and alternative hypotheses. (2%)

Yes

H0: p >= 0.7 v.s H1: p < 0.7

1. Conduct the test and provide the p-value (4%)

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

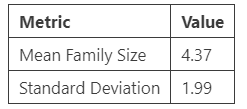
1. Your conclusion (2%)

由於 p 值 0.05，無法拒絕H0，不能斷定公園沒有達到其目標，由於觀察到的會員比例為 62.16%，實際上不足 70%，雖然無法拒絕H0，但公園並未達到目標。

Question 4

Investigate whether the size of family coming to the park is different from the national average of the family size, which is 3.8. Assume the family size is approximately normally distributed. (20%]

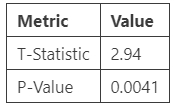
1. Calculate the mean and standard deviation of the group (4%)



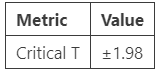
1. Formulate the statistical hypotheses (2%)

H0: μ=3.8 v.s H1: μnot = 3.8

1. Conduct the test and the p-value (4%)



1. Calculate the critical value(s) for the test and provide the rejection rule with the critical value approach. (4%)

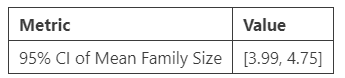


T值為|2.94| >|1.98| 拒絕虛無假設

1. Your conclusion for the hypothesis testing (2%)

當 T 檢定 P 值為 0.0041 時，我們拒絕H0，有證據支持H1公園遊客的家庭規模與全國平均 3.8 人顯著不同。

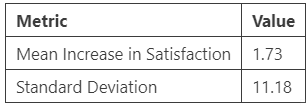
1. Provide the 95% confidence interval of population family size of interest and your interpretation (4%)



Question 5

Investigate whether there is an increase in their satisfaction (post score) from their expectation (pre score). (14%]

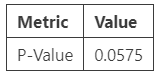
1. What is the mean and the standard deviation for the increase in the satisfaction score? (4pts)



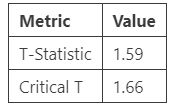
1. Formulate the hypotheses to test for an increase in the mean satisfaction score from their expected score before visiting the park. (2pts)

H0: μ increase ≤ 0 v.s H1: μ increase > 0

1. What is the p-value? (2pts)



1. What is the value of test statistic and the critical value to reject the null hypothesis? (4pts)



1. At the level of significance of 0.05, what is your conclusion? (2pts)

顯著水準為0.05的情況下，我們無法拒絕虛無假設H0，表示沒有足夠的證據證明參觀公園後滿意度得分平均有所提高。