**1122 BM508 Assignment 3 Spring 2024 Due 5/14/2024 @ 11:59pm**

* Please specify the software you used for the questions.
* Please write down and submit your answers using Word document and attach the outputs generated from your software.
* Do not simply submit the output from Excel, R, or SPSS 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 and TA know in advance.)

Problem 1 (44%)

A bank company is investigating the risk of late payment for its issued credit cards from two populations: young adults vs old seniors. The observed frequencies showing the number of samples with late payment and the number of samples without late payment for the past six months are as follows:

|  |  |  |
| --- | --- | --- |
|  | Young Adults | Old Seniors |
| Late Payment | 74 | 48 |
| No Late Payment | 606 | 632 |
| Total | 680 | 680 |

1. What are the late payment rates for each group from the sample? (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. What is the relative risk (risk ratio) of late payment for young adults compared to old seniors? (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. What is the odds ratio of late payment for young adults compared to old seniors? (4%) How do we determine which group had higher late payment rate based on your calculation? (4%)

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

此比值比表明，年輕人逾期付款的可能性是年長老年人的 1.6078 倍。

1. Use the Chi-squared test procedure to see if there is a significant difference between the population late payment rates for the two groups. In this practice do not consider continuity correction. Given alpha=0.05. Please provide statistical hypotheses (4%), the calculated value of Chi-square statistic (4%), p-value (4%) and your conclusion. (4%)

H0：年輕人和老年人的逾期付款比例沒有差異 v .s H1：年輕人和老年人的逾期付款比例有差異

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

因為P-value = 0.0136 < 0.05 拒絕虛無假設，有證據證明年輕人和老年人的逾期付款比例有差異。

1. Now, let’s consider continuity correction. Given alpha=0.05, what is the calculated Chi-squared statistic (4%), p-value (4%) and your conclusion (4%)?

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

P-value = 0.0177 < 0.05 ，在應用連續性校正後，年輕人與老年人之間的逾期付款率差異也具有統計顯著性，拒絕虛無假設，有證據證明兩個年齡組之間的逾期付款率存在統計上的顯著差異

Problem 2 (16%)

M&M’s candy milk chocolate candies come in a variety of colors. The over proportions for the colors are 0.2 blue, 0.1 brown, 0.18 green, 0.17 orange, 0.15 red, and 0.2 yellow. In a sampling survey, several bags of M&M’s milk chocolates were opened, and the color counts were obtained in the file *MM5.xlsx*. Use a 0.05 level of significance and the sample data to test the hypothesis that the overall proportions for the colors are as stated above. Please provide the statistical hypotheses (4%), the calculated value of Chi-square statistic (4%), the critical value for the test (4%), and your conclusion. (4%)

H0：顏色比例與預期相同 v.s H1：顏色比例與預期不全相同

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

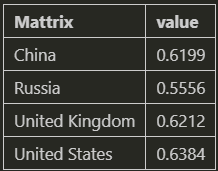
自動產生的描述

因為檢定統計量 = 12.3894 > 臨界值 = 11.0705 所以拒絕虛無假設，有證據證明顏色比例與預期不同

Problem 3 (24%)

Social media is becoming more and more popular around the world. Statista.com provides estimates of the number of social media users in various countries in 2017 as well as the projections for 2022. Assume that the results for surveys in the United Kingdom, China, Russia, and the United States are stored in the file *SocialMedia6.xlsx*.

1. What are the sample proportion of adults using social media for each of the four countries? (8%)



1. Conduct a hypothesis test to determine whether the proportion of adults using social media is equal for all four countries. (4%) What is the critical value given alpha=0.05? (4%) What is the p-value? (4%) Using a .05 level of significance, what is your conclusion? (4%)

H0：四個國家使用社群媒體的成年人比例相同 v.s H1：四個國家使用社群媒體的成年人比例不全相同

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

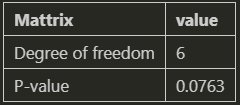
由於 Chi-square 統計量（9.19）超過臨界值（7.81），且 P-value（0.0268）小於 0.05，我們拒絕虛無假設。這表明四個國家（英國、中國、俄羅斯和美國）中成年人使用社交媒體的比例存在統計上的顯著差異

Problem 4 (16%)

A J. D. Power and Associates vehicle quality survey asked new owners a variety of questions about their recently purchased automobile. One question asked for the owner's rating of the vehicle using categorical responses of average, outstanding, and exceptional. Another question asked for the owner's education level with the categorical responses some high school, high school graduate, some college, and college graduate. Assume the sample data in the file *AutoQuality6.xlsx* are for a sample of owners who had recently purchased an automobile.

Use a .05 level of significance and a test of independence to determine if a new owner's vehicle quality rating is independent of the owner's education. Develop the statistical hypotheses. (4%) What is the degree of freedom? (4%) What is the p-value (4%) and what is your conclusion (4%)?

H0：車輛品質評級與車主的教育程度是獨立的v.s H1：車輛品質評級與車主的教育程度非獨立的



P-value = 0.0763 > 0.05，無法拒絕原假設，沒有足夠的證據證明車輛品質評級與車主教育程度之間是獨立的。