# ST2132 Mathematical Statistics
# Assignment 5
# Due date: 5 April 2024 23:59 on Canvas
# Please submit in pdf format.

Please work on the assignment by yourself only. We follow a strict rule on academic honesty.

1. We would like to investigate whether there is a relationship between weather conditions and the incidence of violent crime. We classified 1361 homocides according to the season, resulting in the following data:

| Winter | Spring | Summer | Fall |
|---|---|---|---|
| 340 | 395 | 358 | 268 |

Test the null hypothesis of equal proportions using a significance level of $\alpha = 0.01$.

2. We would like to investigate the incidence of concussions among athletes. Samples of soccer players, non-soccer athletes, and non-athletes are collected.

| | # of Concussions | | | |
|---|---|---|---|---|
| | **0** | **1** | **2** | **≥ 3** |
| **Soccer** | 45 | 25 | 11 | 10 |
| **N-S Athletes** | 68 | 15 | 8 | 5 |
| **Non-athletes** | 45 | 5 | 3 | 0 |

To see whether these three types of individuals have the same distribution, conduct a chi-square test of homogeneity at $\alpha = 5\%$. State (i) the degree of freedom, (ii) the expected frequency count of each cell and (iii) the conclusion of the test.

3. (Does Microsoft stock price really follow a lognormal distribution?) In finance literature, one popular model for stock price is the lognormal distribution. Suppose that we have the daily stock price of a stock, say $S_1, S_2, \ldots, S_n$. We can then compute the so-called log-return of the stock price $R_1, R_2, \ldots, R_{n-1}$, where

$$R_i = \ln\left(\frac{S_{i+1}}{S_i}\right), \quad i = 1, \ldots, n-1.$$

On Canvas, there is a dataset "MSFT.csv" that contains the daily stock price of Microsoft from March 14 1986 to March 12 2024, with $n = 9575$. Use your favourite computing language to answer this question, and attach your code at the end of your assignment.

(a) Now, we conduct the chi-square goodness-of-fit test, with the following $H_0$:

$$H_0 : R_i \overset{i.i.d.}{\sim} N(0, 0.02^2).$$

Fill in the following table:

| Categories | Observed | Expected |
|---|---|---|
| $< -0.001$ | | |
| $[-0.001, -0.0004)$ | | |
| $[-0.0004, 0)$ | | |
| $[0, 0.0004)$ | | |
| $[0.0004, 0.001)$ | | |
| $\geq 0.001$ | | |

Do we reject the null hypothesis at $\alpha = 5\%$? What is the degree of freedom that you use?

(b) Conduct the chi-square goodness-of-fit test, with the following $H_0$:

$$H_0 : R_i \text{ is normally distributed.}$$

Estimate the two parameters by their maximum likelihood estimates and fill in the following table:

| Categories | Observed | Expected |
|---|---|---|
| $< -0.001$ | | |
| $[-0.001, -0.0004)$ | | |
| $[-0.0004, 0)$ | | |
| $[0, 0.0004)$ | | |
| $[0.0004, 0.001)$ | | |
| $\geq 0.001$ | | |

Do we reject the null hypothesis at $\alpha = 5\%$? What is the degree of freedom that you use?