A0235422N Xiao Zhehao

1. $H_0: p_{spring} = p_{summer} = p_{autumn} = p_{winter}$ H_1: H_0 not true. at $\alpha = 0.01$

Expected no. of homicides $= 1361/4 = 340.25$ d.o.f $= 4-1 = 3$ $q_3 = \sum \frac{(O_i - E_i)^2}{E_i} = 25.078 > 11.34 = \chi^2_{0.01}(3)$. Hence, we reject H_0 that the homicides are of equal proportion.

2. i) The d.o.f. $= (3-1)(4-1) = 6$

iii) $\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 22.03 > 12.59 = \chi^2_{0.05}(6)$ Hence, we reject H_0 at $\alpha = 0.05$ that the distributions are homogenous

ii) The observed table is

# of Concussions	0	1	2	≥3	Row Total
Soccer Athletes	45	25	11	10	91
Non Soccer Athletes	68	15	8	5	96
Non Athletes	45	5	3	0	53
Col Total	158	45	22	15	240

The expected table is $E_{ij} = \frac{c_j}{240} \cdot r_i$

# of Concussions	0	1	2	≥3
Soccer Athletes	59.91	17.06	8.34	5.69
Non Soccer Athletes	63.2	18	8.8	6
Non Athletes	34.89	9.93	4.86	3.31

3. a) d.o.f = num categories $-1 = 5$ $\chi^2 = 614.473 > 11.07 = \chi^2_{0.05}(5)$ Hence, we reject H_0 at $\alpha = 5\%$ that $R_i \sim N(0, 0.02)$

Categories	Observed	Expected
< -0.001	4451	4715.161
$[-0.001, -0.0004)$	141	117.476
$[-0.0004, 0)$	60	78.363
$[0, 0.0004)$	293	78.363
$[0.0004, 0.001)$	135	117.476
≥ 0.001	4742	4715.161

b) d.o.f. = num categories $-1-$ num est params $= 6-1-2 = 3$ $\hat{\mu}_{mle} = \bar{x} = \frac{1}{n}\sum r_i = 0.000839$ $\hat{\sigma}_{mle} = \sqrt{\frac{1}{n}\sum(r_i - \hat{\mu}_{mle})^2} = 0.0211$

Categories	Observed	Expected
< -0.001	4451	4569.986
$[-0.001, -0.0004)$	141	111.129
$[-0.0004, 0)$	60	74.195
$[0, 0.0004)$	293	74.251
$[0.0004, 0.001)$	135	111.423
≥ 0.001	4742	4881.017

$\chi^2 = 667.247 > 7.815 = \chi^2_{0.05}(5)$ Hence, we reject H_0 at $\alpha = 5\%$ that R_i is Normally distributed

ST2132 Mathematical Statistics
Assignment 5
Due date: 11 April 2025 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.

Generative AI policy: The use of generative AI tools (e.g., ChatGPT, Gemini, etc.) is strictly prohibited for structural or conceptual questions in this course. However, these tools may be used to assist with coding tasks, provided their use is clearly documented.

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 10 2025, with $n = 9823$. 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		
$[-0.001, -0.0004)$		
$[-0.0004, 0)$		
$[0, 0.0004)$		
$[0.0004, 0.001)$		
≥ 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		
$[-0.001, -0.0004)$		
$[-0.0004, 0)$		
$[0, 0.0004)$		
$[0.0004, 0.001)$		
≥ 0.001		

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