Lianne Hourigan
H170365C
Simulation and Modelling
Assignment 2

Q. Injuries per month | Frequency of Occurrence

0	35
1	40
2	13
3	6
4	4
5	1
6	1

a. Apply the chi-square test to these data to test the hypothesis that the underlying distribution is Poisson. Use level of significance $\alpha = 0.05$. [6]

Defining hypothesis
H_0: The distribution is Poisson.
H_a: The distribution is not Poisson.

For Poisson Distribution

$$P_i = \frac{e^{-\alpha}\alpha^x}{x!} \qquad \alpha = \bar{x} = \frac{\sum f_i m_i}{n}$$

$$\therefore \alpha = \frac{0\times 35 + 1\times 40 + 2\times 13 + 6\times 3 + 4\times 4 + 1\times 5 + 1\times 6}{100}$$

$$\therefore \alpha = 1.11$$

x_i	O_i	P_i	$E_i = h * P_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
0	35	0.3296	32.96	4.1616	0.126
1	40	0.3658	36.58	11.6964	0.320
2	13	0.2030	20.30	53.2900	2.625
3	6	0.0751	7.51	2.2801	1
4	4	0.0209	2.09	3.6481	0.333
5	1	0.0046	0.46	0.2916	1
6	1	0.0010	0.10	0.8100	1
Total	100	1.0000	100		$\chi_0^2 = 3.400$

Estimate the value of α

$\therefore s = 1$

$\therefore d.f = k - s - 1$

$= 4 - 1 - 1$

$= 2$

$\therefore \chi^2_{0.05,2} = 5.99$

$\chi_0^2 = 3.404 < \chi^2_{0.05,2} = 5.99$

$\therefore$ Ho is accepted

given that data is poisson distribution.

b) Apply the chi-square test to these data to test the hypothesis that the distribution is Poisson with mean 1.0. Again let $\alpha = 0.05$. [6]

Defining Hypothesis:

H_0: data fits to poisson distribution

H_1: data does not fit to poisson distribution

$$P_i = \frac{e^{-\alpha} \alpha^x}{x!} \qquad \alpha_i = 1 \text{ (given)}$$

x_i	O_i	P_i	$E_i = n \times P_i$	$(O_i - E_i)^2 / E_i$
0	35	0.3679	36.79	0.087
1	40	0.3679	36.79	0.280
2	13	0.1839	18.39	1.580
3	6	0.0613	6.13	1
4	4	0.0153	1.53	1.963
5	1	0.0031	0.31	1
6	1	0.0006	0.06	1
Total	100	1.0	100	$x_0^2 = 3.91$

To find a:

we do not estimate value of α

$\therefore s = 0$

$\therefore d.f = k - s - 1$

$= 4 - 0 - 1$

$= 3$

$\because x^2_{0.05,3} = 7.81$ $\quad \therefore H_0$ is accepted

$\because x_0^2 = 3.91 < x^2_{0.05,3} = 7.81$ given data fits to poisson distrib

C (iii) What are the differences between parts (a) and (b) and when might each case arise? [3]

In part (a) the mean is not provided hence had to be calculated and in part (b) the mean was provided.

In (a) we had to subtract k by 2 (k-1-1) since we are calculating two values α and the last value of β so as to get the degree of freedom where as in (b) we subtract one from k (k-1) since only one value was missing, to obtain the degree of freedom.

At the end we got different degrees of freedom.