1. The table below shows the number of accidents per week on a highway in France over a year. It is thought that the data follows a Poisson model with a mean of 2.

| Accidents | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Observed | 5 | 13 | 14 | 8 | 6 | 3 | 2 | 1 |

Conduct a Chi-squared goodness of fit test at the 5% level of significance, to test whether the data does follow a Poisson model.

2. A model is being developed such that the model will follow a geometric distribution. The model should be distributed $X \sim Geo\left(\frac{1}{4}\right)$.

a) Find the mean and variance of the model.

b) The model is developed and tested by doing 100 trials. The following lists the results observed by the 100 trials.

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Freq. | 26 | 24 | 19 | 10 | 10 | 8 | 3 |

Make a table of expected frequencies based on the distribution $X \sim Geo\left(\frac{1}{4}\right)$.

c) The data is to be tested using a chi-squared goodness of fit test to see if the model is geometric. State suitable hypothesis for this test.

d) Calculate the chi-squared test statistic, $p$.

e) Clearly state your degrees of freedom, write down the critical chi-squared value from your tables at the 10% level of significance.

f) Make a conclusion based on your results for b) to e).

3. Chelski football club had some bad injuries to their goalkeepers in the 2006-7 season, but their manager Mr. Morrison claims that the team will perform well whoever the goalkeeper is. Chelski's 60 matches and their results are shown below in the table. A chi squared test is to be carried to test for independence at the 5% level of confidence.

| | **Result** | | | |
|---|---|---|---|---|
| **Goalkeeper** | **Win** | **Draw** | **Lose** | **Total** |
| **Goalkeeper A** | 8 | 4 | 4 | 16 |
| **Goalkeeper B** | 4 | 7 | 3 | 14 |
| **Goalkeeper C** | 20 | 5 | 5 | 30 |
| **Total** | 32 | 16 | 12 | 60 |

a) Write down the null and alternate hypotheses.

b) Calculate a table of expected values for the for the observed data above.

c) Evaluate the chi-squared test statistic, $p$.

d) Clearly stating the degrees of freedom, write down the critical value from the chi-squared table. Hence, write a conclusion to the test.

4. The Wine producers of California believe that the type of wine (White, Red, or Rose) is independent of the gender of the purchaser. However, they wish to test this and sample 1000 bottles of wine purchased across California and the results are below.

| | **Type of wine** | | | |
|---|---|---|---|---|
| **Gender** | **White** | **Red** | **Rose** | **Total** |
| **Female** | 260 | 100 | 40 | 400 |
| **Male** | 150 | 400 | 50 | 600 |
| **Total** | 410 | 500 | 90 | 1000 |

Carry out a test at the 1% level of confidence to check if the type of wine is independent of the gender of the purchaser.

5. Test results in the mathematics examination have, in the past, been normally distributed with a mean of 50% and a variance of 196%.

In 2006 a new type of examination was sat and the results for 500 students were recorded as follows:

| Score (%) | $0 < s \le 20$ | $20 < s \le 34$ | $34 < s \le 50$ | $50 < s \le 64$ | $64 < s \le 80$ | $80 < s \le 100$ |
|---|---|---|---|---|---|---|
| Frequency | 10 | 31 | 179 | 225 | 42 | 13 |

Carry out a chi-squared goodness of fit test at the 5% level to test if the new test fits the model of the old test.

6. A set of dice are thought to be biased. A selection of 10 dice are taken from the set and each one is rolled 36 times, and the number of sixes in the 36 trials is recorded. A chi-squared test for goodness of fit is to be carried out.

a) Write down suitable hypothesis for a test.

b) Write down the number of sixes expected for each of the 10 trials.

c) Below are the results of the test. Use this table and your answer from b) to calculate the chi-squared test statistic, *p*.

| Trial | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 6's | 8 | 9 | 6 | 7 | 6 | 10 | 12 | 7 | 5 | 8 |

d) State your degrees of freedom and write down the critical value from your chi-squared table for the 5% level of confidence.

e) State your conclusion.

7. In a High School there are 4 English teachers, who each teach a class of 20 students. The results for the 5 end of year examination are shown below with a letter grade attached to each.

| | Grade | | | | | |
|---|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | **E** | **Total** |
| **Ms. P** | 0 | 10 | 40 | 30 | 20 | 100 |
| **Ms. Q** | 40 | 30 | 20 | 10 | 0 | 100 |
| **Ms. R** | 20 | 20 | 20 | 20 | 20 | 100 |
| **Ms. S** | 10 | 35 | 25 | 10 | 10 | 100 |
| **Total** | 70 | 95 | 115 | 70 | 50 | 400 |

Test at the 1% level of significance that the grade of the student is independent of the teacher.

8. The percentage of a radioactive substance present in a person's blood is measured at various distances away from a nuclear power station.

The observed frequencies are shown in the table below.

| | **0 – 10 km** | **11 – 20 km** | **21 – 30 km** | **31 – 40 km** | **Totals** |
|---|---|---|---|---|---|
| **0% - 0.5%** | 75 | 30 | 44 | 54 | 220 |
| **0.5% - 1.0%** | 80 | 40 | 26 | 52 | 180 |
| **1.0% - 1.5%** | 50 | 15 | 10 | 24 | 100 |
| **Totals** | 205 | 85 | 80 | 130 | 500 |

The table of expected values is then calculated and written below.

| | **0 – 10 km** | **11 – 20 km** | **21 – 30 km** | **31 – 40 km** | **Totals** |
|---|---|---|---|---|---|
| **0% - 0.5%** | $p$ | 37.4 | 35.2 | $s$ | 220 |
| **0.5% - 1.0%** | 73.8 | $q$ | 28.8 | 46.8 | 180 |
| **1.0% - 1.5%** | 41 | $r$ | 16 | 26 | 100 |
| **Totals** | 205 | 85 | 80 | 130 | 500 |

a) Find the values of $p$, $q$, $r$, and $s$ in the table above.

b) A $\chi^2$ test is to be done to test whether the distance a person lives from the power station is independent of the amount of radioactive substance in their blood. State the alternate hypothesis.

c) Explain why the degrees of freedom are 6.

d) Find the $\chi^2$ value at the 10% level of significance from the Chi-squared tables.

e) The test $\chi^2$ value is calculated as 12.40. Explain clearly what your conclusion is how you arrived at this conclusion.

9. A variable is thought to follow a continuous probability density function that is defined by the model,

$$f(z) = \begin{cases} \frac{6}{5}\left(x^2 - x\right), 1 \leq x \leq 2. \end{cases}$$

The function is split into 5 intervals:

$I_1 = [1, 1.4[$
$I_2 = [1.4, 1.6[$
$I_3 = [1.6, 1.8[$
$I_4 = [1.8, 1.9[$
$I_5 = [1.9, 2]$

An experiment is carried out on the variable and a record of 60 results is shown below.

| $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ |
|---|---|---|---|---|
| 5 | 8 | 11 | 14 | 22 |

a) Calculate the expected values for each interval.

b) At the 5% level of significance does the model hold true for this set of data?

**Answers**

1. $H_0$: the new test results are distributed to fit the model $X \sim Po(2)$
$H_1$: the data does not fit this Poisson distribution

| Accidents | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Expected | 7.04 | 14.07 | 14.07 | 9.38 | 4.69 | 1.88 | 0.63 | 0.18 |

But,

| Accidents | 0 | 1 | 2 | 3 | 4 - 7 |
|---|---|---|---|---|---|
| Expected | 7.04 | 14.07 | 14.07 | 9.38 | 7.43 |

$p$ = 3.76 (the test statistic)

$v=4$, 5% level from table $\chi^2 = 9.488$

Do not reject $H_0$, that is at the 5% level of significance the Poisson model fits the data observed.

2. a) $\mu = 4, \sigma^2 = 12$

b) Expected table:-

| Trial | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Freq. | 25 | 18.75 | 14.06 | 10.54 | 7.91 | 5.93 | 4.45 |

| Trial | 1 | 2 | 3 | 4 | 5 | 6-7 |
|---|---|---|---|---|---|---|
| Freq. | 25 | 18.75 | 14.06 | 10.54 | 7.91 | 10.38 |

c) $H_0$: the new test results are distributed to fit the model $X \sim Geo\left(\frac{1}{4}\right)$.

$H_1$: the data does not fit this geometric distribution.

d) $p=3.87$

e) $v=5$, $\chi^2 = 9.236$

f) Do not reject $H_0$, that is that based on this data the model fits the geometric distribution stated above.

3. a) $H_0$: Chelski's results are independent of the goalkeeper playing.

$H_1$: Chelski's results are not independent of the goalkeeper playing.

b)

| | Result | | |
|---|---|---|---|
| **Goalkeeper** | **Win** | **Draw** | **Lose** |
| **Goalkeeper A** | 8.53 | 4.27 | 3.2 |
| **Goalkeeper B** | 7.47 | 3.7 | 2.8 |
| **Goalkeeper C** | 16 | 8 | 6 |

| | Result | |
|---|---|---|
| **Goalkeeper** | **Win** | **Draw / Lose** |
| **Goalkeeper A** | 8.53 | 7.47 |
| **Goalkeeper B** | 7.47 | 6.5 |
| **Goalkeeper C** | 16 | 14 |

c) $p = 5.71$

d) $v = 2$, $\chi^2 = 5.991$

e) Do not reject $H_0$, Chelski's results are independent of the goalkeeper used.

4. $H_0$: The wine purchased is independent of the gender of the purchaser.
   $H_1$: The wine purchased is not independent of the gender of the purchaser.

| | **Type of wine** | | |
|---|---|---|---|
| **Gender** | **White** | **Red** | **Rose** |
| **Female** | 260 | 100 | 40 |
| **Male** | 150 | 400 | 50 |

$p = 177.73$

$v = 2,\ \chi^2 = 9.210$

Reject $H_0$ in favour of $H_1$, the wine purchased will be dependent of the gender of the purchaser.

5. $H_0$: the new test results are normally distributed with $X \sim N(50, 14^2)$
   $H_1$: the data does not fit this normal distribution.

Table of expected values:

| Score (%) | $0 < s \leq 20$ | $20 < s \leq 34$ | $34 < s \leq 50$ | $50 < s \leq 64$ | $64 < s \leq 80$ | $80 < s \leq 100$ |
|---|---|---|---|---|---|---|
| Frequency | 8 | 55 | 187 | 171 | 71 | 8 |

$p = 43.285$

$v = 5,\ \chi^2 = 11.07$

Reject $H_0$ in favour of $H_1$. The results for 2006 do not fit the normal distribution stated.

6. a) $H_0$: the dice are not biased and fit the distribution $X \sim B\left(36, \frac{1}{6}\right)$.

   $H_1$: the dice are biased and do not fit the distribution $X \sim B\left(36, \frac{1}{6}\right)$.

   b) 6

   c) $p = 12.02$

   d) $v = 9$, $\chi^2 = 16.919$

   e) Do not reject $H_0$, the dice are not biased.

7. $H_0$: The grade achieved is independent of the teacher.
   $H_1$: The grade achieved is not independent of the teacher.

   Expected table:-

| | Grade | | | | |
|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | **E** |
| **Ms. P** | 17.5 | 23.75 | 28.75 | 17.5 | 12.5 |
| **Ms. Q** | 17.5 | 23.75 | 28.75 | 17.5 | 12.5 |
| **Ms. R** | 17.5 | 23.75 | 28.75 | 17.5 | 12.5 |
| **Ms. S** | 17.5 | 23.75 | 28.75 | 17.5 | 12.5 |

$P = 111.16$

$v = 12$, $\chi^2 = 26.217$

Reject $H_0$, at this level of significance there is suffiecient evidence to suggest that the grade achieved is dependent on the teacher.

8. a) $p = 90.2$ $q = 30.6$ $r = 17$ $s = 57.2$

b) $H_1$: The amount of radioactivity in the blood is not independent of the distance from the power station.

c) $v = (4-1)(3-1) = 6$

d) $v = 6$, $\chi^2 = 10.64$

f) Do not reject $H_0$, the amount of radioactivity is independent of the distance lived from the power station.

9. a)

| $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ |
|---|---|---|---|---|
| 7.32 | 10.80 | 17.22 | 11.28 | 13.62 |

b) $p = 9.54$, $v = 4$, $\chi^2 = 9.488$

Reject $H_0$ in favour of $H_1$, the data does not fit the function given in the question.