**Chi-square test of goodness of fit**

**Uniform Distribution**:

A computer scientist has developed an algorithm for generating pseudorandom integers over the interval 0-9. He codes the algorithm and generates 1000 pseudorandom digits. The data are shown in table below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Digit** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| **Frequency** | 94 | 93 | 112 | 101 | 104 | 95 | 100 | 99 | 108 | 94 |

Is there evidence that the random number generator is working correctly?

**Solution**:

**Null and Alternative Hypothesis**

H0: Random number generator is working correctly i.e. frequency of each digit are same or frequencies are according to uniform distribution

H1: Random number generator is not working correctly i.e. frequency of each digit are not same



Cal Chi-square = 3.72

**Critical chi-square**

degrees of freedom = k – 1 – r = 10 – 1 – 0 = 9

Critical value of chi-square from the table

= 16.919

**Statistical Decision**

Do not reject H0

**Conclusion**

The algorithm that generates pseudo random number is working well.

**Fixed proportion: Specified distribution**

A supermarket has observed the following distribution of modes of payment on the express line: The data has been collected over 3 years’ period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mode of payment | Cash | QR | Credit and Debit card | Other |
| Percentage of transactions | 41 % | 24 % | 26 % | 9 % |

In an effort to make express checkout more efficient, the supermarket has just began offering a 5 % discount for cash payment in the express checkout line. The following table lists the frequency distribution of the modes of payment for a sample of 500 express-line customers after the discount went into effect.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mode of payment | Cash | QR | Credit and Debit card | Other | Total |
| No. of customers | 240 | 105 | 108 | 47 | 500 |

Test at the 5 % significance level whether the distribution of modes of payment in the express checkout line changed after the discount went into effect.

Solution:

**Null and Alternative hypothesis**

H0: the distribution of modes of payment in the express checkout line is not changed

H1: the distribution of modes of payment in the express checkout line is changed.

**Calculated chi-square**

|  |  |  |  |
| --- | --- | --- | --- |
| Mode of payment | Observed frequency (Oi) | Expected frequency (Ei) | (Oi – Ei) ^ 2 / Ei |
| Cash | 240 | 41 % of 500 = 205 | 5.9756 |
| QR | 105 | 24 % of 500 = 120 | 1.875 |
| Card | 108 | 26 % of 500 = 130 | 3.7231 |
| Others | 47 | 9 % of 500 = 45 | 0.0889 |
| Total | 500 | 500 | cal = 11.6626 |

Calculated chi-square = 11.6626

**Tabulated chi-square**

Degrees of freedom = k – 1 – r = 4 – 1 – 0 = 3

= 7.815

**Statistical Decision**

Since calculated chi-square = 11.6626 > Critical chi-square= 7.815, we reject H0 at 5 % level of significance.

**Conclusion**

The distribution of modes of payment in the express checkout line is changed.

**Test of goodness of fit for Poisson Distribution**

The following table shows a data set of the number of errors found in a total of n = 85 software products.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. of errors found in a software product | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
| Frequency | 3 | 14 | 20 | 25 | 14 | 6 | 2 | 0 | 1 | n = 85 |

Is it plausible that the number of errors has Poisson distribution with mean λ = 3?

**Solution**:

**Null and Alternative hypothesis**

H0: No. of errors has a Poisson distribution with mean λ = 3

(Frequencies are according to Poisson distribution with parameter λ = 3)

H1: No of errors do not have a Poisson distribution with mean λ = 3

Test Statistic

where,

O = observed frequency

E = Expected frequency according to distribution specified (here poison distribution)

**Calculated chi-square**

|  |  |  |  |
| --- | --- | --- | --- |
| X = No. of errors | observed frequency (Oi) | p (x) = | Expected frequency (Ei)  = n x p(x)  (Sample size x prob) |
| 0 | 3 | 0.04979 | p (0) = 4.23 |
| 1 | 14 | 0.14936 | p (1) = 12.70 |
| 2 | 20 | 0.22404 | p (2) = 19.04 |
| 3 | 25 | 0.22404 | p (3) = 19.04 |
| 4 | 14 | 0.16803 | p (4) = 14.28 |
| 5 | 6 | 0.10082 | p (5) = 8.57 |
| 6 | 2 | 0.05041 | p (6) = 4.28 |
| 7 | 0 | 0.02160 | p (7) = 1.84 |
| 8 | 1 | 0.00810 | P (8) = 0.69 |
| Total | 85 |  |  |

**After grouping**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **No. of errors** | 0-1 | 2 | 3 | 4 | 5 | ≥ 6 | Total |
| **Observed frequency** | 17 | 20 | 25 | 14 | 6 | 3 | 85 |
| **Expected frequency** | 16.93 | 19.04 | 19.04 | 14.28 | 8.57 | 7.14 | 85 |
| **(O-E) ^2 / E** |  |  |  |  |  |  | 5.09 |

Cal chi-square = 5.09

**Critical chi-square**

Degrees of freedom = k – 1 – r = 9 – 1 – (1+3) = 4

From table,

= 9.488

**Statistical Decision**

Since Cal = 5.09 < Critical = 9.488, we do not reject our null hypothesis.

**Conclusion**

The given frequencies are according to Poisson Distribution with parameter λ =3