Tests for Goodness of Fit

Null hypothesis H_0 : The goodness of fit is appropriate for the given data set

Alternative hypothesis H_1 : The goodness of fit is not appropriate for the given data set

The test statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where k = number of classes

Note: If any of E_i is found less than 5, the corresponding class frequency may be pooled with preceding or succeeding classes such that E_i's of all classes are greater than or equal to 5. It may be noted that the value of k may be determined after pooling the classes.

Problem:

Five coins are tossed 640 times and the following results were obtained.

Number of heads	0	1	2	3	4	5
Frequency	19	99	197	198	105	22

Fit binomial distribution to the above data.

Step 1 : Null hypothesis H_0: Fitting of binomial distribution is appropriate for the given data.

Alternative hypothesis H_1: Fitting of binomial distribution is not appropriate to the given data.

n = number of coins tossed at a time = 5

Let X denote the number of heads (success) in n tosses

N = number of times experiment is repeated = 640

To find mean of the distribution

x	f	fx
0	19	0
1	99	99
2	197	394
3	198	594
4	105	420
5	22	110
Total	640	1617

$$\text{Mean : } \bar{x} = \frac{\sum fx}{\sum f} = \frac{1617}{640} = 2.526$$

The probability mass function of binomial distribution is :

$$p(\mathrm{x}) = {}^{\mathrm{n}}\mathrm{C}_{\mathrm{x}}\, p^{\mathrm{x}}\, q^{\mathrm{n-x}},\ x = 0, 1, \dots, \mathrm{n}$$

Mean of the binomial distribution is $\bar{x} = np$.

Hence, $$\hat{p} = \frac{\bar{x}}{n} = \frac{2.526}{5} \approx 0.5$$

$$\hat{q} = 1 - \hat{p} \approx 0.5$$

$$P(X = 0) = P(0) = 5c_0\, (0.5)^5 = 0.03125$$

The expected frequency at $x = N\,P(x)$

The expected frequency at $x = 0$: $N \times P(0)$

$= 640 \times 0.03125 = 20$

similarly

Number of heads	0	1	2	3	4	5	Total
Expected frequencies	20	100	200	200	100	20	640

The test statistic is computed as under:

Observed frequency (O_i)	Expected frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
19	20	–1	1	0.050
99	100	–1	1	0.010
197	200	–3	9	0.045
198	200	–2	4	0.020
105	100	5	25	0.250
22	20	2	4	0.200
			Total	0.575

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= 0.575$$

Degrees of freedom = $k - 1 = 6 - 1 = 5$

Critical value for *d.f* 5 at 5% level of significance is ----

conclusion:

As the calculated χ_0^2 (=0.575) is less than the critical value $\chi^2_{5,\ 0.05}$, we do not reject the null hypothesis. Hence, the fitting of binomial distribution is appropriate.

problem 2

A packet consists of 100 ball pens. The distribution of the number of defective ball pens in each packet is given below:

x	0	1	2	3	4	5
f	61	14	10	7	5	3

Examine whether Poisson distribution is appropriate for the above data at 5% level of significance.

Null hypothesis H_0: Fitting of Poisson distribution is appropriate for the given data.

Alternative hypothesis H_1: Fitting of Poisson distribution is not appropriate for the given data.

To find the mean of the distribution.(expected freq)

x	f	fx
0	61	0
1	14	14
2	10	20
3	7	21
4	5	20
5	3	15
Total	100	90

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{90}{100} = 0.9$$

Probability mass function of Poisson distribution is:

$$p(x) = \frac{e^{-m} m^x}{x!}; \; x = 0, 1, \ldots \qquad (2.2)$$

In the case of Poisson distribution mean $(m) = \overline{x} = 0.9$.

At $x = 0$, equation (2.2) becomes

$$p(0) = \frac{e^{-m} \; m^0}{0!} = e^m = e^{0.9} = 0.4066 \cdot$$

The expected frequency at x is N P(x)

Therefore, The expected frequency at 0 is

$N \times P\,(0)$

$= 100 \times 0.4066$

$= 40.66$

Table of expected frequency distribution (on rounding to the nearest integer)

x	0	1	2	3	4	5
Expected frequency	41	37	16	5	1	0

The test statistic is computed as under:

Observed frequency (O_i)	Expected frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
61	41	20	400	9.756
14	37	-23	529	14.297
10	16	-6	36	2.250
7	5			
5 } 15	1 } 6	9	81	13.5
3	0			
			Total	51.253

$$\chi_0{}^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= 51.253$$

df = k-1=3

table value (5% level + df=3) is greater

Conclusion:

H0 is rejected

Fitting of poisson distribution is not appropriate for the given data