4. Test for goodness of fit :

From the sample data we fit a curve or a distribution (binomial, exponential, poisson or normal). We set the null hypothesis as the given sample is from a specified population.

Let O_i and E_i be the observed and expected frequencies such that $\Sigma O_i = \Sigma E_i = N$ and the test-statistic is,

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

with $v = n-1$ d.o.f.

Note :

1. In the case of fitting a Binomial distribution d.o.f $= n-1$, for poisson distribution. $v = n-2$ and for normal distribution $v = n-3$.

2. If $\chi^2 = 0$, all observed and expected frequencies coincide.

Problems : (165)

(1) Fit a binomial distribution for the following data and also test the goodness of fit.

x	0	1	2	3	4	5	6	Total
f	5	18	28	12	7	6	4	80

Solution :

Null hypothesis : H_0 : Binomial fit is a good fit to the data.

Alternative hypothesis : H_1 : Binomial fit is not a good fit to the data.

Calculation of expected frequencies :

x	0	1	2	3	4	5	6	Total
f	5	18	28	12	7	6	4	80
fx	0	18	56	36	28	30	24	192

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

$$\bar{x} = \frac{192}{80} = 2.4$$

$$\boxed{\bar{x} = 2.4}$$

For Binomial distribution, mean $= np$

i.e., $\bar{x} = np \Rightarrow np = 2.4$

$6p = 2.4$

$p = 0.4$

$q = 0.6$

The expected frequencies are calculated using the formula, $N \times n c_r \, p^r q^{n-r}$

where $N = 80$, $n = 6$, $p = 0.4$, $q = 0.6$ &

$r = 0, 1, 2, \dots 6$

$E(0) = 80 \times 6c_0 (0.4)^0 (0.6)^6 = 3.73$

$E(1) = 80 \times 6c_1 (0.4)^1 (0.6)^5 = 14.93$

$E(2) = 80 \times 6c_2 (0.4)^2 (0.6)^4 = 24.88$

$E(3) = 80 \times 6c_3 (0.4)^3 (0.6)^3 = 22.12$

$E(4) = 80 \times 6c_4 (0.4)^4 (0.6)^2 = 11.06$

$E(5) = 80 \times 6c_5 (0.4)^5 (0.6) = 2.95$

$E(6) = 80 \times 6c_6 (0.4)^6 (0.6)^0 = 0.33$

Hence the observed and expected frequencies are given by,

O_i	5	18	28	12	7	6	4
E_i	4	15	25	22	11	3	0

The first class is combined with the second and the last two classes are combined with the last but second class in order to make the expected frequency in each class greater than or equal to 10. Thus, after regrouping, we have,

O_i	23	28	12	17
E_i	19	25	22	14

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
23	19	4	16	0.842
28	25	3	9	0.36
12	22	−10	100	4.4545
17	14	3	9	0.643
	80			6.39

Test-Statistic :

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

$$= 6.39$$

$$\boxed{\chi^2 = 6.39}$$

Table Value :

At $\alpha = 5\%$ LOS, $v = n - k = 4 - 2 = 2$ d.o.f, the table value of χ^2 is given by,

$$\boxed{\chi^2_\alpha = 5.99}$$

Decision :

Since $\chi^2 > \chi^2_\alpha$, H_0 is rejected. ∴ The binomial fit is not a good fit for the data.

(2) Fit a poisson distribution for the following distribution and also test the goodness of fit.

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400

Solution :

Null hypothesis : H_0: Poisson fit is a good fit to the data.

Alternative hypothesis : H_1 : poisson fit is not a good fit to the data.

Calculation of expected frequencies :

(167)

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400
fx	0	156	138	81	20	5	400

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{400}{400} = 1$$

For poisson distribution, mean $= \lambda$

i.e., $\bar{x} = \lambda \Rightarrow \lambda = 1$

The expected frequencies are given by,

$\dfrac{N e^{-\lambda} \lambda^{r}}{r!}$ where $N = 400$, $\lambda = 1$, $r = 0, 1, \ldots 5$

$$E(0) = 400 \times e^{-1} = 147.2$$

$$E(1) = \frac{400 \times e^{-1}}{1!} = 147.2$$

$$E(2) = \frac{400 \times e^{-1}}{2!} = 73.58$$

$$E(3) = \frac{400 \times e^{-1}}{3!} = 24.53$$

$$E(4) = \frac{400 \times e^{-1}}{4!} = 6.13$$

$$E(5) = \frac{400 \times e^{-1}}{5!} = 1.23$$

Thus,

O_i	:	142	156	69	27	5	1
E_i	:	147	147	74	25	6	1

Let us combine the last 3 groups into one, so that $E_i \geq 10$.

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
142	147	-5	25	0.17
156	147	9	81	0.55
69	74	-5	25	0.34
33	32	1	1	0.03
	400			1.09

Test Statistic:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\boxed{\chi^2 = 1.09}$$

Table value:

At $\alpha = 5\%$ LOS, $v = n - 2 = 4 - 2 = 2$ d.o.f, the value of χ^2 in tables is,

$$\boxed{\chi^2_\alpha = 5.99}$$

Decision: Since $\chi^2 < \chi^2_\alpha$, H_0 is accepted $\therefore$ Poisson fit is a good fit to the data.

Problems :

① Find if there is any association between extravagance in fathers and extravagance in sons from the following data.

	Extravagant father	Miserly father
Extravagant son	327 ^{a}	741 ^{b}
Miserly Son	545 ^{c}	234 ^{d}

Determine the coefficient of association also.

Solution :

Null hypothesis : H_0 : The attributes are independent.

Alternative hypothesis : H_1 : The attributes are dependent.

Test-Statistic :

For a 2×2 Contingency table,

$$\chi^2 = \frac{(ad-bc)^2 (a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \frac{[(327)(234)-(545)(741)]^2 \times (327+234+545+234)}{(875)(975)(1068)(779)}$$

$$\boxed{\chi^2 = 230.24}$$

Table value :

At $\alpha = 5\%$ Los , $v = (p-1)(q-1)$

$= (2-1)(2-1) = 1$ d.o.f , the table value of χ^2 is given by,

$$\boxed{\chi^2_\alpha = 3.841}$$

Decision :

Since $\chi^2 > \chi^2_\alpha$, H_0 is rejected. $\therefore$ The attributes are dependend. Hence there is an association between extravagance in fathers and extravagance in sons.

Coefficient of attributes :

$$\text{Coefficient of association} = \frac{ad - bc}{ad + bc}$$

$$= \frac{327 \times 234 - 741 \times 545}{327 \times 234 + 741 \times 545}$$

$$= \frac{-327330}{480363}$$

$$\boxed{Q = -0.6814}$$