Testing of Hypothesis: Large Sample Tests

Lesson 10: Test of Hypothesis for Goodness of Fit

This test is used to decide whether the observations are in good agreement with a hypothetical distribution. The observed frequencies $\left(f_{o_i}\right)$ of different classes are compared with the expected frequencies $\left(f_{e_i}\right)$ by the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{o_i} - f_{e_i}\right)^2}{f_{e_i}} - - - - - - - (1)$$

such that $\sum_{i=1}^{k} f_{o_i} = \sum_{i=1}^{k} f_{e_i} = N$ = total frequency, k = number of classes or observations in the frequency distribution. This is called the Pearsonian χ^2 or goodness of fit χ^2.

This test statistic follows the χ^2 distribution with (k - m) degrees of freedom under the assumption that the null hypothesis

H_0: Data is in agreement with the hypothetical population

is true, here m is the number of constraints. If the calculated value of χ^2 is more than the tabulated value of χ^2 at a given level of significance, then the null hypothesis H_0 is rejected.

Note:

(i) If $\chi^2 = 0$, then f_{e_i} and f_{o_i} agree exactly

(ii) If $\chi^2 > 0$ and small, then f_{e_i} and f_{o_i} are close to each other indicating a good fit

(iii) If $\chi^2 > 0$ and large, then f_{e_i} and f_{o_i} differs considerably indicating a poor fit

Conditions for Validity of χ^2 Test:

(i) Large sample size $(n \geq 50)$

(ii) $4 \leq \mathrm{k} \leq 16$

(iii) If the expected frequency $f_{e_i} < 5$ for any item, then some data from the neighbourhood can be put together so that f_{e_i} becomes ≥ 5

Problems:

Ex.1. A dice was thrown 60 times with the attached results.

Face	1	2	3	4	5	6
Frequency	6	10	8	13	11	12

Use χ^2 test to examine if the data is consistent with the hypothesis that the dice is unbiased at 1% significant level.

Solution:

Let us consider the null hypothesis, H_0: The dice is unbiased

Then the probability of each upturned face = 1/6 and since the total frequency N = 60.

$\therefore$ The expected frequency (f_{e_i}) of each face = $60 \times 1/6 = 10$

$\boldsymbol{f_{o_i}}$	6	10	8	13	11	12
$\boldsymbol{f_{e_i}}$	10	10	10	10	10	10
$\boldsymbol{(f_{o_i} - f_{e_i})^2}$	16	0	4	9	1	4
$\boldsymbol{\frac{(f_{o_i} - f_{e_i})^2}{f_{e_i}}}$	1.6	0	0.4	0.9	0.1	0.4

Then

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{o_i} - f_{e_i}\right)^2}{f_{e_i}} = 3.4$$

Here, k = number of classes = 6 and m = number of constraints = 1 (The total frequency 60 is the only constraint here)

∴ Degrees of freedom = $k - m$ = 5. The critical value of χ^2 at 1% significant level for 5 degrees of freedom = 15.086 (refer to the χ^2 table)

As computed $\chi^2 <$ tabulated χ^2, therefore H_0 is accepted and we conclude that the data is consistent with the hypothesis that the dice is unbiased. In other words, the data is a good fit for the distribution.

Ans.

Ex.2. The following table indicates (a) the observed frequencies of a given distribution and (b) the frequencies of the normal distribution with same mean, standard deviation and the total frequency as in (a).

a	1	5	20	28	42	22	15	5	2
b	1	6	18	25	40	25	18	6	1

Apply the χ^2 test of goodness of fit (5% significant level).

Solution:

Let us consider the null hypothesis, H_0: The observed data is in agreement with the normal distribution.

As we know that the expected frequencies cannot be less than 5, let us club the frequencies of 1st and 2nd class and similarly that of the last two classes together to obtain the modified table as:

$\boldsymbol{f_{o_i}}$	6	20	28	42	22	15	7
$\boldsymbol{f_{e_i}}$	7	18	25	40	25	18	7
$\boldsymbol{(f_{o_i} - f_{e_i})^2}$	1	4	9	4	9	9	0
$\boldsymbol{\frac{(f_{o_i} - f_{e_i})^2}{f_{e_i}}}$	1/7	4/18	9/25	4/40	9/25	9/18	0

Then

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{o_i} - f_{e_i})^2}{f_{e_i}} = 1.685$$

Here, k = number of classes = 7 and m = number of constraints = 3 (same mean, standard deviation and total frequency in both distributions)

∴ Degrees of freedom = $k - m$ = 4. The critical value of χ^2 at 5% significant level for 4 degrees of freedom = 9.488 (refer to the χ^2 table)

As computed $\chi^2 <$ tabulated χ^2, therefore H_0 is accepted and we conclude that the data is in agreement with the normal distribution and therefore the fit is good.

Ans.

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892