Testing of Hypothesis: Large Sample Tests

Lesson 11: Test of Hypothesis for Independence of Attributes

In this test, the qualitative characters or attributes are categorized into two or more categories and the observed data is shown in a table with m rows and n columns. Let us consider two attributes A and B where $A_1, A_2, \ldots\ldots, A_m$ are m categories for attribute A and $B_1, B_2, \ldots\ldots, B_n$ are n categories for attribute B.

A \ B	B_1	B_2	…….	B_n	Row Total
A_1	$f_{A_1B_1}$	$f_{A_1B_2}$	…….	$f_{A_1B_n}$	$\sum f_{A_1B_j}$
A_2	$f_{A_2B_1}$	$f_{A_2B_2}$	…….	$f_{A_2B_n}$	$\sum f_{A_2B_j}$
……..	……	……...	…….	…….	…….
A_m	$f_{A_mB_1}$	$f_{A_mB_2}$	…….	$f_{A_mB_n}$	$\sum f_{A_mB_j}$
Column Total	$\sum f_{A_iB_1}$	$\sum f_{A_iB_2}$	……	$\sum f_{A_iB_n}$	N = total frequency

The $m \times n$ table thus formed is called the contingency table. The row and column totals are called the marginal frequencies and we must have $\sum_{i=1}^{m}\sum_{j=1}^{n} A_iB_j = N$ = the total frequency.

The Steps for testing are:

- On the basis of the contingency table, the expected frequency $f_{e_{ij}}$ for each cell is to be calculated from the given formula

$$f_{e_{ij}} = \frac{\text{row total} \times \text{column total}}{\text{total frequency}}$$

- Formulate null hypothesis H_0 : Attributes are independent i.e, there exists no significant differences between $f_{o_{ij}}$ and $f_{e_{ij}}$
- The test statistic used for testing the null hypothesis is

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}} - - - - - - - (1)$$

  with $(m-1)(n-1)$ degrees of freedom
- The critical value of χ^2 is then located from the χ^2 distribution table for $(m-1)(n-1)$ degrees of freedom
- If the computed χ^2 is more than the tabulated χ^2 at a specific significant level, then H_0 is rejected and the conclusion is that the attributes are not independent. In that case, the attributes are said to be associated.

Problems:

Ex.1. A random sample of 500 students were classified according to the economic condition of their family and also according to the merit. Test whether the two attributes "Merit" and "Economic Condition" are independent at 1% significant level.

Merit	Economic Condition			Total
	Rich	Middle Class	Poor	
Meritorious	42	137	61	240
Non-Meritorious	58	113	89	260
Total	100	250	150	500

Solution:

Based on the contingency table, let us calculate the expected frequency for each cell:

Merit	Economic Condition			Total
	Rich	Middle Class	Poor	
Meritorious	$\frac{240 \times 100}{500} = 48$	$\frac{240 \times 250}{500} = 120$	$\frac{240 \times 150}{500} = 72$	240
Non-Meritorious	$\frac{260 \times 100}{500} = 52$	$\frac{260 \times 250}{500} = 130$	$\frac{260 \times 150}{500} = 78$	260
Total	100	250	150	500

- Let us formulate the null hypothesis H_0 : Attributes are independent

- The test statistic is

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}}$$

$$= \frac{(42-48)^2}{48} + \frac{(137-120)^2}{120} + \frac{(61-72)^2}{72} + \frac{(58-52)^2}{52} + \frac{(113-130)^2}{130} + \frac{(89-78)^2}{78}$$

= 9.3 with $(m-1)(n-1) = (2-1)(3-1)$

= 2 degrees of freedom

- The critical value of χ^2 at 1% significant level for 2 degrees of freedom is given by 9.21 [*refer to the table]

- Since computed χ^2 is more than tabulated χ^2, hence H_0 is rejected

Therefore we conclude that the two attributes "Merit" and "Economic Condition" are not independent, but associated.

Ans.

Ex.2. Test the hypothesis at 5% level of significance that the presence or absence of hypertension is independent of smoking habits from the following data of 180 persons.

Hypertension	Smoking Habits			Total
	Non Smokers	Moderate Smokers	Heavy Smokers	
Present	21	36	30	87
Absent	48	26	19	93
Total	69	62	49	180

Solution:

Based on the contingency table, let us calculate the expected frequency for each cell:

Hypertension	Smoking Habits			Total
	Non Smokers	Moderate Smokers	Heavy Smokers	
Present	$\frac{87 \times 69}{180} = 33.35$	$\frac{87 \times 62}{180} = 29.97$	$\frac{87 \times 49}{180} = 23.68$	87
Absent	$\frac{93 \times 69}{180} = 35.65$	$\frac{93 \times 62}{180} = 32.03$	$\frac{93 \times 49}{180} = 25.32$	93
Total	69	62	49	180

- Let us formulate the null hypothesis H_0 : Attributes are independent

- The test statistic is

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}}$$

$$= \frac{(21-33.35)^2}{33.35} + \frac{(36-29.97)^2}{29.97} + \frac{(30-23.68)^2}{23.68} + \frac{(48-35.65)^2}{35.65} + \frac{(26-32.03)^2}{32.03} + \frac{(19-25.32)^2}{25.32}$$

= 14.46 with $(m-1)(n-1) = (2-1)(3-1)$

= 2 degrees of freedom

- The critical value of χ^2 at 5% significant level for 2 degrees of freedom is given by 5.991 [*refer to the table]

- Since computed χ^2 is more than tabulated χ^2, hence H_0 is rejected

Therefore we conclude that the two attributes "Hypertension" and "Economic Condition" are not independent, but associated.

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