Week-03-L-04

Hypothesis Testing

Non-parametric testing

Dr. Amandeep Singh
Imagineering Laboratory
Indian Institute of Technology Kanpur

Chi-Square Test

Chi-Square Test of Independence:

Problem Statement

Suppose we want to know whether or not gender is associated with food preference. We take a simple random sample of 500 people and survey them on their food preference. The following table shows the results of the survey:

	Rice	*Wheat*	*Other*
Male	*120*	*90*	*40*
Female	*110*	*95*	*45*

H_0: Association between Gender status and Choice of food

H_0: Gender = Food

H_a: Gender ≠ Food

Chi-Square Test

Chi-Square Test of Independence:

Problem Statement

Suppose we want to know whether or not gender is associated with food preference. We take a simple random sample of 500 people and survey them on their food preference. The following table shows the results of the survey:

	Rice	*Wheat*	*Other*	*Total*
Male	120	90	40	250
Female	110	95	45	250
Total	230	185	85	500

Chi-Square Test

- *Perform a Chi-Square test of independence to determine if gender is associated with food preference.*

Step 1: Define the hypotheses

- We will perform the Chi-Square test of independence using the following hypotheses:
  H_0: Gender and food preference are independent.
  H_1: Gender and food preference are not independent.

Step 2: Calculate the expected values.

- Next, we will calculate the expected values for each cell in the contingency table using the following formula:
  Expected value = (row sum * column sum) / table sum.

  - For example, the expected value for Male Rice eater is: (230*250) / 500 = 115.
  - We can repeat this formula to obtain the expected value for each cell in the table:

Chi-Square Test

Conducting the analysis

O: observed value

	Rice	*Wheat*	*Other*	*Total*
Male	*120*	*90*	*40*	*250*
Female	*110*	*95*	*45*	*250*
Total	*230*	*185*	*85*	*500*

E: expected value

	Rice	*Wheat*	*Other*	*Total*
Male	*115*	*92.5*	*42.5*	*250*
Female	*115*	*92.5*	*42.5*	*250*
Total	*230*	*185*	*85*	*500*

$$\frac{120 - 115}{115}$$

Chi-Square Test

Step 3: Calculate $(O-E)^2 / E$ for each cell in the table.

- Next we will calculate $(O-E)^2 / E$ for each cell in the table where:
  - O: observed value
  - E: expected value

For example, Male Rice Eater would have a value of: $(120-115)^2 / 115 = 0.2174$

- We can repeat this formula for each cell in the table:

Chi-Square Test

	Rice	Wheat	Other	Total
Male	120	90	40	250
Female	110	95	45	250
Total	230	185	85	500

1 Obs.

2. Exp

	Rice	Wheat	Other	Total
Male	115	92.5	42.5	250
Female	115	92.5	42.5	250
Total	230	185	85	500

3

	Rice	Wheat	Other	Total
Male	0.2174	0.0676	0.1471	0.432
Female	0.2174	0.0676	0.1471	0.432
Total	0.4348	0.1352	0.294	0.864

χ^2

Chi-Square Test

Step 4: Calculate the test statistic X^2 and the corresponding p-value.

- The test statistic X^2 is simply the sum of the values in the last table.
- The p-value that corresponds to the test statistic.

- The test statistic X^2 turns out to be 0.8640 and the corresponding p-value is 0.649198.

	Rice	*Wheat*	*Other*	*Total*
Male	*0.2174*	*0.0676*	*0.1471*	***0.432***
Female	*0.2174*	*0.0676*	*0.1471*	***0.432***
Total	***0.4348***	***0.1352***	***0.294***	***0.864***

$$X^2 = 0.864$$
$$p = 0.649198$$

D.o.F

$(R-1)(C-1)$

$= (2-1)(3-1)$

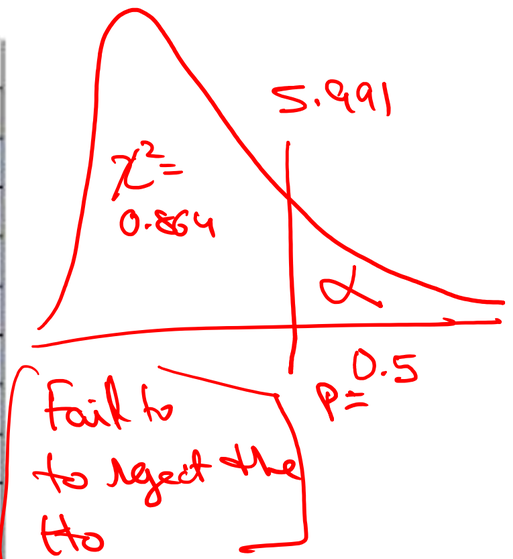
$= 1 \times 2$

$= 2$

Chi-Square Test

Step 4: Calculate the test statistic X^2 and the corresponding p-value.

	P										
DF	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	.0004	.00016	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.55	10.828
2	0.01	0.0506	3.219	4.605	5.991	7.378	7.824	9.21	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.86	16.924	18.467
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086	16.75	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.69	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.18	11.03	13.362	15.507	17.535	18.168	20.09	21.955	24.352	26.124
9	1.735	2.7	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.92	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.3	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.79
18	6.265	8.231	22.76	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.9	27.204	30.144	32.852	33.687	36.191	38.582	41.61	43.82
20	7.434	9.591	25.038	28.412	31.41	34.17	35.02	37.566	39.997	43.072	45.315



$$X^2 = 0.864$$
$$p = 0.649198$$

Chi-Square Test

Step 5: Conclusion

- Since this p-value is not less than 0.05, we fail to reject the null hypothesis.

- This means we do not have sufficient evidence to say that there is an association between gender and food preference.

Source: www.shutterstock.com

Thank you