Biostatistics

Chi-square test

What is it?

Types of its use?

Inferential statistics

Mohamed abu libdah

Statistician and Data Analyst at the
NCI - Cairo University - Egypt

introduction

The origin of this test goes back to the famous statistical psychologist Karl Pearson (1900).

The chi-square test depends on the number of qualitative variables studied. If we have one qualitative variable studied, then the appropriate test is the goodness of fit test or the homogeneity test. If we have two variables Of the two types studied, the appropriate test is the independence test. The x2 test is used to calculate the significance of the differences in frequencies or numerical data that can be converted into frequencies. The chi-square test aims to determine whether the observed or observed frequencies (experimental) differ from the expected frequencies (theoretical).) For reasons related to coincidence factors or to fundamental factors. There are three types of chi-square test , which are as follows:

1. **Chi-Square Goodness of Fit Test**
2. **Chi-Square Test of Independence**
3. **Chi-Square Test of Homogeneity**

- **The data must be qualitative (nominal or ordinal) or quantitative data (interval or relative) that Is converted into qualitative data.**
- **The data must be in the form of frequencies.**
- **The total number of observed frequencies should not be less than 20 occurrences, and preferably more than 40 repetitions .**
- **The total expected repetitions In any cell of the classification should not be less than 5 repetitions In 80% of the cells. If this Is not achieved, it is possible in some cases to resort to merging cells as a solution to this problem, provided that the merging Is in accordance with justified considerations so that the cells After merging, It is meaningful.**
- **The frequencies must be Independent of each other, and this means that the frequency resulting from the measurement of any individual from the sample does not affect the measurement of the frequencies of the rest of the sample members.**

- In the event that the degree of freedom Is equal to 1, that Is, in the case of an association table of Figure (2 x 2), and the expected frequencies in one of the cells are less than 10. The modified x2 test (Yates correction) Is applied by subtracting 0.5 from the value

To calculate chi-square we use the following formula:

where:

Fo: observed frequencies

Fe: expected frequencies.

$$\chi^2 = \sum \frac{(fo - fe)}{fe}$$

Chi-square test for goodness of fit:

It is a one-sample test, and is used to calculate the significance of differences for data of one qualitative variable when they are in the form of frequencies, and the frequencies are represented in an association table consisting of one row representing the frequencies of the sample band two or more columns to represent the categories of the qualitative variable, i.e. In the form of a simple table (1×2) or (1×3)...etc., it is used to show the extent to which the frequencies of the categories of the qualitative variable specified in the sample, which are called the observed frequencies, match the expected frequencies in the studied population, and the meaning of This is because the sample drawn belongs to the population under study, so it is called a test of goodness of fit or goodness of fit. If the sample is representative of the population in its repetitions and identical with it; The test value is usually equal to zero, and this value increases to become more than zero whenever there are differences between the observed frequencies of the sample and the expected frequencies of the population to which it belongs.

To calculate the chi-square for goodness of fit, we use the following formula:

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

$$Df = C - 1$$

Where:

fo: observed frequencies

fe: expected frequencies.

Df: degree of freedom.

C: number of columns (number of categories of the variable).

Example:

Here is the following table, which represents the opinions of a sample of mobile phone network users about the level of quality of services provided to them by the mobile phone company.

The level of quality of services provided	Bad	Acceptable	Excellent
Customer reviews (frequency)	15	20	25

Required:

Test the hypothesis that there is a correspondence (no differences) in the opinions of mobile phone network users about the quality of the level of services provided to them by the company.

The solution:

- **Ask the problem:**
  Are there differences between mobile phone network users in their opinions? About their level of use of the mobile phone network and about the quality of services provided to them through the company.
- **Formulating hypotheses:**

- **Null hypothesis:**
  There are differences between mobile phone network users in their opinions about their level of use of the mobile phone network and about the quality of services provided to them through the company.
- **Alternative hypothesis:**
  There is no difference between mobile phone network users in their opinions about the level of their use of the mobile phone network and the quality of services provided to them through the company.
- **Choosing the appropriate statistical method:**

Since we have one qualitative variable, which is the level of mobile network usage measured using a single sample, and the usage conditions are: data in the form of frequencies, independence of frequencies, the appropriate statistical method is the chi-square test of goodness of fit.

- **Performing calculations:**

1. **Calculating the expected frequency**

The expected frequency is calculated from the following formula:

Total repetitions / number of frequency

$$fe = \frac{25 + 20 + 15}{3} = 20$$

2. **Calculating the chi-square**

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

$$\chi^2 = \frac{(25-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(15-20)^2}{20}$$

$$\chi^2 = \frac{25}{20} + \frac{0}{20} + \frac{25}{20}$$

$$\chi^2 = 1.25 + 0 + 1.25$$

$$\boxed{\chi^2 = 2.5}$$

3. **Find the degree of freedom Df**

$$Df = C - 1$$
$$Df = 3 - 1$$
$$\boxed{Df = 2}$$

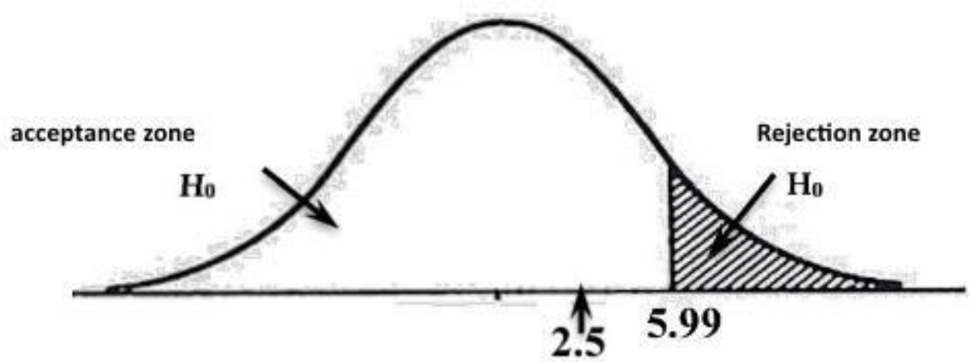
4. **Determine the tabular chi-square value (x2) :**

5.99 χ_T^2 at the degree of freedom 2 and at the significance level of 0.05

5. **Comparison and decision making:**

Since the calculated chi-square value (χ_C^2) which is equal to 2.5, is less than the tabular chi-square value $\chi_T^2$2), which equal to 5.99, we accept The null hypothesis.

The following graph shows this:



,

Interpretation of the decision: We are confident with a confidence rate of 95% that there is a match (no differences) between customers of the mobile phone network in their opinions about the level of quality provided to them by the company, with an error rate of 5%, and at a degree of freedom of 2 = Df.

Chi-square table

df	0.1	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	**5.99**	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.95
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
31	41.42	44.99	48.23	52.19	55.00
32	42.58	46.19	49.48	53.49	56.33
33	43.75	47.40	50.73	54.78	57.65
34	44.90	48.60	51.97	56.06	58.96
35	46.06	49.80	53.20	57.34	60.27
40	51.81	55.76	59.34	63.69	66.77
45	57.51	61.66	65.41	69.96	73.17
50	63.17	67.50	71.42	76.15	79.49
60	74.40	79.08	83.30	88.38	91.95
70	85.53	90.53	95.02	100.43	104.21
80	96.58	101.88	106.63	112.33	116.32
90	107.57	113.15	118.14	124.12	128.30
100	118.50	124.34	129.56	135.81	140.17

Mohamed abu libdah

Statistician and Data Analyst at the NCI - Cairo University - Egypt