

Comparing Frequencies using Chi-square

Chi-square or χ^2 : pronounced “*ki squared*”

- Used to test whether the frequency of nominal data fit a certain pattern (**goodness of fit**) or whether two variables have a dependency (**test for independence**).
- Can be used to test whether data are normally distributed and for homogeneity of proportions
- Frequency of each nominal category is computed and compared to an expected frequency.

Goodness of Fit:

- Need to know expected pattern of frequencies.
- If not known assume equal distribution among all categories.

Assumptions:

- data are from a random sample
- expected frequency for each category must be 5 or more

Examples:

- test for product / procedure preference (each is assumed equally likely to be selected)
- test for “fairness” of coin, die, roulette wheel (expect each outcome equally)
- test for expected frequency distribution (need theoretically expected pattern)

Goodness of Fit Test

Step 1

H_0 : data fit the expected pattern

H_1 : data do not fit expected pattern

Step 2

Find critical value from χ^2 table. Test is always a one-tailed right test with $n-1$ degrees of freedom, where n is number of categories.

Step 3

Compute test value from:
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = observed freq. E = expected frequency

Step 4

Make decision. If $\chi^2 >$ critical value reject H_0 .

Step 5

Summarize the results.

E.g., There is (not) enough evidence to accept/reject the claim that there is a preference for _____.

E.g., Coin is fair / unfair

 Die is fair / "loaded"

 Wheel is fair / flawed

Test for Independence

Step 1

H_0 : two variables are independent

H_1 : two variables are dependent

Step 2

Find critical value from χ^2 table. Test is always one-tailed right with $(n_{\text{row}} - 1)(n_{\text{col.}} - 1)$ degrees of freedom, where n_{row} and $n_{\text{col.}}$ are the number of categories of each variable. These correspond to the number of rows and columns in the contingency table.

Step 3

Create the contingency table to derive the expected values (see next page).

Compute test value from:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = observed freq. E = expected frequency

Step 4

Make decision. If $\chi^2 >$ critical value reject H_0 .

Step 5

Summarize the results.

E.g.: Getting a cold is dependent upon whether you took a cold vaccine.

- Smoking and lung disease are dependent.
- Is a cure dependent upon placebo vs. drug.

Contingency Table

First, enter observed (O) scores and compute **row** and **column** totals.

	Col.1	Col.2	Col.3	totals
Row 1	25	10	5	40
Row 2	10	20	5	35
Row 3	5	30	15	50
totals	40	60	25	125* / 125*

* Notice sum of row and column totals must equal.

Second, compute expected (E) values based of **row** and **column** totals.

	Col.1	Col.2	Col.3
Row 1	$40 \times 40 / 125 = E_{11}$	$40 \times 60 / 125 = E_{12}$	$40 \times 25 / 125 = E_{13}$
Row 2	$35 \times 40 / 125 = E_{21}$	$35 \times 60 / 125 = E_{22}$	$35 \times 25 / 125 = E_{23}$
Row 3	$50 \times 40 / 125 = E_{31}$	$50 \times 60 / 125 = E_{32}$	$50 \times 25 / 125 = E_{33}$

Finally, compute the test value:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$