

A Test of “Independence”

Chi Square (χ^2)

Statistical Independence

- Observed (Cell) Frequencies
- Expected (Cell) Frequencies

Computing “expected frequencies”.

a	b	a+b
c	d	c+d
a+c	b+d	N

The expected frequency for cell a is $(a+b)*(a+c)/N$. Similarly for all other cells, take the product of the marginal frequencies and divide by N.

How different are the expected and observed frequencies? The answer is provided by χ^2 .

How to compute χ^2 .

1. Compute the difference between the observed and expected frequencies for each cell.
2. Square these differences.
3. Divide the squared differences by the expected frequencies.
4. Sum the results of Step 3.

Examples of “Independence”

Let's return to our Vote and Gender example...

Q's:

Are Gender and Vote independent?

Does the distribution of males and females influence the distribution of vote?

I.

Gender

	M	F
Bush	50 50% 50% 25%	50 50% 50% 25%
Gore	50 50% 50% 25%	50 50% 50% 25%

Cells: Frequency, Row%, Col%, Total %

II.

Gender

	M	F
Bush	25 50% 25% 12%	25 50% 25% 12%
Gore	75 50% 75% 385%	75 50% 75% 38%

Cells: Frequency, Row%, Col%, Total %

III.

Gender

	M	F
Bush	80 33% 80% 27%	160 67% 80% 53%
Gore	20 33% 20% 7%	40 67% 20% 13%

Cells: Frequency, Row%, Col%, Total %

Common Confusions:

Statistical Independence and Independent Variables

Now what's the actual relationship between vote and gender? Our 1992 National Election Survey SETUPS file will tell us.

Is the vote of each respondent independent of his or her gender? What is the evidence?

Chi-Square will give us a summary of the difference between the actual and expected frequencies. Whether this difference is "big enough" to consider important is established by the significance level.

The standard significance level is .05.

For a 2X2 table, a χ^2 of 3.84 or higher provides evidence that the variables are not independent of each other. We're confident that this statement is true 95 times out of 100 tests. It doesn't mean that we're "certain" but we usually feel confident.

The appropriate value of a χ^2 varies with the number of rows and columns in the table.

Another way of saying this is that the appropriate value varies with the number of degrees of freedom. In a 2X2 table there is 1 degree of freedom. (one way of thinking about degrees of freedom is to imagine you know the marginal frequencies but not the cell frequencies. How many cell frequencies would you need to know to fill in the entire table? One (if the table is 2x2). More generally, the number of degrees of freedom in a table is

$$df = (\#rows-1) \times (\#columns -1)$$