The Chi Square Test for Counted Data

Katherine Dorfman, UMass Biology Department, 2019

One test for statistical significance applicable to many experiments that count data in categories (e.g., number of cells in particular phases of mitosis) is the Chi (pronounced like *sky* without the *s*) squared (χ^2) test. It tests for whether the values are distributed in the categories as predicted (expected) by chance.

$$\chi^2 = \sum_{i=1}^{m} \frac{(o_i - e_i)^2}{e_i}$$

χ^2 is based on the difference between a series of observed values ($o_1, o_2, \ldots o_m$) and expected values ($e_1, e_2, \ldots e_m$). Bigger deviations from the expected yield a bigger χ^2.

The differences between observed and expected are squared for two reasons. The first is that this makes all the terms positive so the negative ones don't cancel out the others. The second reason is that this accentuates large differences and minimizes small ones.

The squared difference is divided by the expected value to obtain something like the percent deviation, thereby weighting the individual $\frac{(o_i-e_i)^2}{e_i}$ terms so that the relative deviation rather than the absolute deviation is used in calculating the statistic.

The individual $\frac{(o_i-e_i)^2}{e_i}$ terms are summed to give an overall indicator of the difference between observed and expected across all the categories.

χ^2 lets you determine the probability that the deviation from your prediction was due to chance alone.

Example:

For the entering class in 2012, UMass accepted 11,918 out of 18,006 female applicants, and 9552 out of 16320 male applicants[1]. Is this distribution according to chance, or is there some non-random factor at work? Imagine the admissions process taking place by the admissions officials putting all the applications into a giant bag, and pulling out 21,470 (11918 + 9552) without looking. How likely would they be to pull out 11,918 applications from females and 9552 from males?

It helps to look at a 2-way table:

gender	**accepted**	**rejected**	**total applications**
female	11,918	6088	**18,006**
male	9,552	6768	**16,320**
totals	**21,470**	**12,856**	**34,326**

[1] UMass Amherst At a Glance 2012-2013, http://www.umass.edu/oapa/publications/glance/FS_gla_01.pdf, accessed 8/26/13

What would we expect from the grab-bag admissions technique? The applicant pool was 52.5% female (18,006/34,326), so the population of admitted students should also be 52.5% female. Similarly, 62.5% of all applicants were accepted, so we expect 62.5% of the female applicants to be accepted. Here's the two-way table again, with the expected values calculated:

gender	**accepted**	**rejected**	**total applications**
female	$\frac{18006}{34326} \times 21470 = 11{,}262.3$	$\frac{18006}{34326} \times 12856 = 6743.7$	**18,006**
male	$\frac{16320}{34326} \times 21470 = 10{,}207.7$	$\frac{16320}{34326} \times 12856 = 6112.3$	**16,320**
totals	**21,470**	**12,856**	**34,326**

Here's the calculation for χ^2:

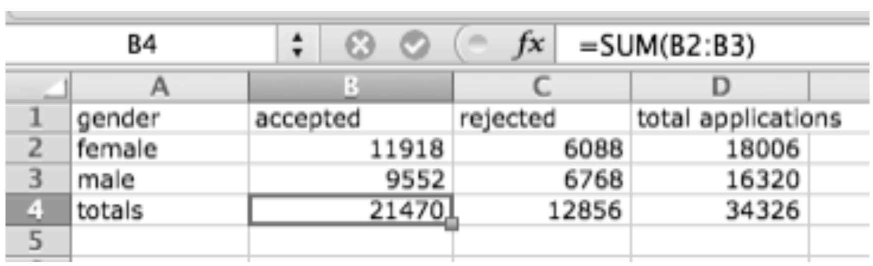
$$\chi^2 = \frac{(11918 - 11262.3)^2}{11262.3} + \frac{(9552 - 10277.7)^2}{10277.7} + \frac{(6088 - 6743.7)^2}{6743.7} + \frac{(6768 - 6112.3)^2}{6112.3} = 214.4$$

The calculated χ^2 by itself has very little meaning. The significance (literally) comes from calculating the probability of getting a χ^2 at least this big by random assignment of individuals to categories. Imagine the admissions committee doing the grab-bag technique over and over and over. Most of the time, a random assignment procedure will produce results similar to the prediction.

Every once in a while, however, the person pulling applications from the bag to accept will grab mostly applications from women, or mostly from men. This would produce a larger χ^2. The question is: how often could we expect to see a χ^2 as big as or bigger than our calculated one, if the applications really were just pulled from a bag?

You can get Excel to do the heaviest of the lifting for you, that is, calculating p from the distribution. However, you do have to get Excel to calculate the expected values for each category. Here's how I did the calculations for the application example:

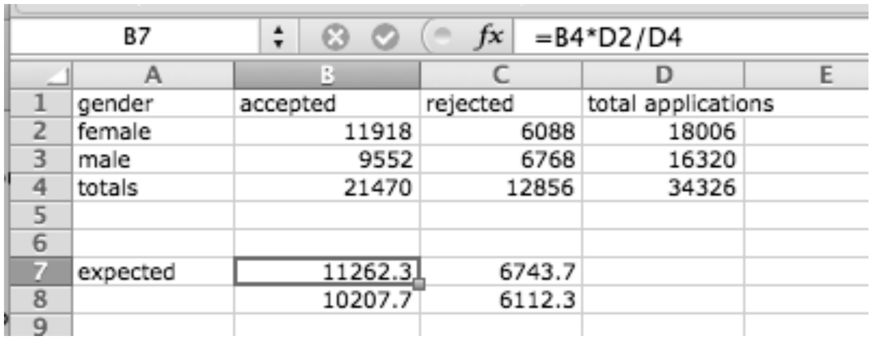
I typed in the raw data from the UMass website, then used formulas to calculate the subtotals and the grand total:



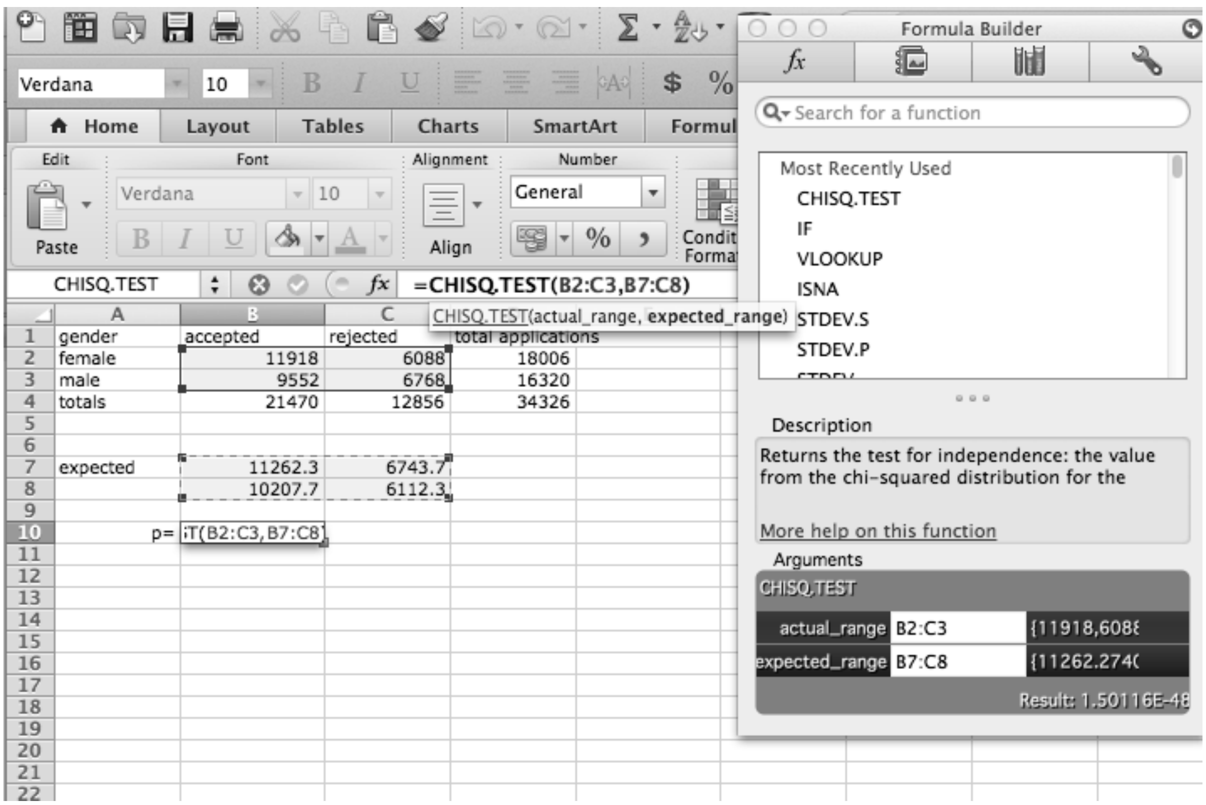
B4 | fx =SUM(B2:B3)

	A	B	C	D
1	gender	accepted	rejected	total applications
2	female	11918	6088	18006
3	male	9552	6768	16320
4	totals	21470	12856	34326
5				

Then I made another table for the expected values, and typed in the appropriate formulas to have Excel do the calculations for me.



I used CHISQ.TEST in the formula builder to calculate the p associated with the χ^2 for this data set. You can see why it matters how you array your observed and expected values. Their relative positions must match.



The result, as Excel calls it, is the probability of getting a χ^2 as big as yours if only chance was sorting the results into categories. In this case, the tiny probability (1.5×10^{-48}) tells us that UMass did *not* accept the same proportion of male and female applicants.

To assess where the discrepancy lies, we have to actually calculate the individual elements that contribute to the χ^2. You can see how I did it, to the right. These calculations show that the biggest contributor to the large χ^2 is the higher than predicted rejection of male applicants.

C14 =(C3-C8)^2/C8

	A	B	C	D
1	gender	accepted	rejected	total applications
2	female	11918	6088	18006
3	male	9552	6768	16320
4	totals	21470	12856	34326
5				
6				
7	expected	11262.3	6743.7	
8		10207.7	6112.3	
9				
10	p=	1.50116E-48		
11				
12				
13	X2=	38.2	63.8	
14		42.1	70.3	214.4

Remember, when you have finished these calculations, you do not have *proof* that chance did or did not cause your results; you simply know how likely it is to get results like yours if chance were the only factor. In the case of the UMass admissions data, we can conclude that there is a difference between the acceptance rates of male and female applicants, which you might describe in one of these ways:

> UMass accepted significantly fewer male applicants than female applicants by the Chi-square test ($p = 1.5 \times 10^{-48}$).
>
> A significantly greater percentage of female applicants (66%) got in to UMass, as compared to male applicants (59%) ($\chi^2 = 214$, $p = 1.5 \times 10^{-48}$).