Two important applications of the chi-square test are: test for homogeneity and test for independence. They basically differ in the sampling schemes, the ways in which the data were collected.

1. Test of Homogeneity

For example, suppose we're interested in comparing the proportions of high school freshmen and high school seniors falling into various driving categories, perhaps, those who don't drive at all, those who drive unsafely, and those who drive safely. We randomly select 100 freshmen and 100 seniors and then observe into which of the three driving categories each student falls.

So, we want to test wether two or more (say I) multinomial distributions are equal. The sampling scheme involves taking two or more random (and therefore independent) samples with fixed sizes in advance and observing into which of the K categories each observation falls. So, row totals are fixed.

Is the acceptance rate the same for males and females across four different schools (business, engineering, liberal arts and science)?

2. Test of Independence

Now, suppose 395 people are randomly selected, and are cross-classified into one of eight cells, depending into which age category they fall (4 categories) and whether or not they support legalizing marijuana. The sampling scheme involves taking one random sample of size n, with n fixed in advance, and then cross-classifying each subject into one and only one of the mutually exclusive and exhaustive cells. Note that, in this case, both the row totals and column totals are random, it is only the total number n sampled that is fixed in advance.

Is the desire to ride a bicycle independent of age? Is the belief in an afterlife independent of age or independent of gender?

Comments on these tests:

1. When n is large and the model is true, χ^2 and G^2 (likelihood ratio test statistic) tend to be approximately equal. For large samples, the results of the χ^2 and G^2 tests will be essentially the same.

2. An old-fashioned rule of thumb is that the χ^2 approximation works well provided that n is large enough to have $E_j = n\pi_j \geq 5$ for every j. Nowadays, most agree that we can have $E_j < 5$ for some of the cells (say, 20 % of them). Some of the E_j can be as small as 2, but none of them should fall below 1. If this happens, then the χ^2 approximation isn't appropriate, and the test results are not reliable.

3. In practice, it's a good idea to compute both χ^2 and G^2 to see if they lead to similar results. If the resulting p-values are close, then we can be fairly confident that the large-sample approximation is working well.