

Re: Categorical Data Question

Source: <http://sci.tech-archive.net/Archive/sci.stat.edu/2006-08/msg00113.html>

- *From:* Bruce Weaver <bweaver@xxxxxxxxxxxxx>
 - *Date:* Wed, 30 Aug 2006 17:36:13 -0400
-

Richard Ulrich wrote:

---- snip ----

Then how should I test goodness of fit? Should I use chi-square or G-square?

I think you are suggesting "Pearson" versus "Likelihood" tests on the contingency table. They are not usually the best choice.

I also took G-squared to mean the likelihood ratio chi-square. But "contingency table" suggests a chi-square test of association, whereas the OP wants a goodness of fit test, I think. Although it may not be done as often, one can compute a likelihood ratio chi-square for a goodness of fit test too.

$$G^2 = 2 * \sum [O * \ln(O/E)]$$

where O = observed count and E = expected count.

I've never seen anything that compares the properties of G² and Pearson's chi-square for goodness of fit tests. The closest thing I have is this passage from Agresti (1990) comparing them for contingency tables.

----- Beginning of excerpt -----

It is not simple to describe the sample size needed for the chi-squared distribution to approximate well the exact distributions of X² and G² [or L²]. For a fixed number of cells, X² usually converges more quickly than G². The chi-squared approximation is usually poor for G² when n/IJ < 5 [where n = the grand total and IJ = rc = the number of cells in the table]. When I or J [i.e., r or c] is large, it can be decent for X² for n/IJ as small as 1, if the table does not contain both very small and moderately large expected frequencies. (Agresti, 1990, p. 49)

----- End of excerpt -----

X² = Pearson's chi-square test statistic

G² = likelihood ratio chi-square test statistic (sometimes called L²)

Re: Categorical Data Question

Bruce Weaver
bweaver@xxxxxxxxxxxx
www.angelfire.com/wv/bwhomedir
.