

# Re: Chi-square for binomial samples

---

*Source:* <http://sci.tech-archive.net/Archive/sci.stat.math/2006-06/msg00078.html>

---

- *From:* Michael McLaughlin <[mmclaughlin1@xxxxxxx](mailto:mmclaughlin1@xxxxxxx)>
  - *Date:* Sat, 03 Jun 2006 13:48:18 -0400
- 

I guess I did not explain the problem well. Let me try again.

Here is dataset A: Binomial variates (successes) coming from 10 samples, each of size 10, with  $p = 0.5$  (hypothetically).

4  
5  
5  
6  
7  
3  
4  
5  
4  
6

The chi-square statistic is 2.6. The significance of this value could be easily determined with a parametric bootstrap without having to worry about sample size, cell frequency being too small, asymptotic assumptions, etc.

Here is dataset B: Binomial variates with observed proportions of success as indicated.

4/11  
6/10  
8/13  
2/5  
2/4  
12/25  
9/15  
3/7  
5/10  
7/15

The ML value for  $p$  is 0.50435 (to five sig. figs.).

I was wondering how a chi-square statistic would be formulated in this case (assuming that is could). That formulation could then be used in an analogous parametric bootstrap to assess its significance, assuming the ML value for  $p$ .

## Re: Chi-square for binomial samples

Note: The log-likelihood values resulting from such a bootstrap do not have much power compared to chi-square, unless there is an outlier, so they make a poor test.

FWIW, my own guess wrt the proper procedure would be either

- a) treat each observation separately, compute a chi-square term for it, then sum these terms, or
- b) pool all variates in dataset B having the same denominator, then proceed as in a).

I was querying this list looking for more options, esp. if one of them was preferred.

David Winsemius wrote:> Michael McLaughlin <mmclaughlin1@xxxxxxx> wrote in

[Xuofg.24929\\$ZW3.3160@dukeread04:](mailto:Xuofg.24929$ZW3.3160@dukeread04:)">news:[Xuofg.24929\\$ZW3.3160@dukeread04:](mailto:Xuofg.24929$ZW3.3160@dukeread04:)

Scenario:

N experimenters take samples from a common population, effectively infinite, BUT these samples are of various sizes. They each look for a given (fixed) attribute of interest and record their frequency of success.

Were their samples all of the same size, then every statistics book ever written would discuss how to compute the associated chi-square statistic.

Most chi-square statistics that I have seen (and there are many) assume variable numbers in each stratum. Can you give us some examples?

Moreover, computing a maximum-likelihood value for  $p$  is still straightforward since every term in the log-likelihood is well-defined.

Question:

Is there an accepted method to compute chi-square under the conditions stated — where  $p$  is assumed known but sample size is variable? Is this even a sensible question?

The chi-square is just the sum of squared deviations from the expected. Usual basis for expected is row-sum  $\times$  column-sum/total. Are you proposing to substitute a known  $p$  (times the row total presumably) for that number? If so, you may want to investigate GLMs with offsets.

## Re: Chi-square for binomial samples

The analogous question could be asked wrt the beta-binomial distribution as well. There, the scenario described is, in fact, the norm. The question could be asked again, a fortiori, wrt the hypergeometric distribution (two parameters nominally fixed).

It sounds as though you are asking for a test of equal proportions among  $k$  multiple samples. Results (successes, failures and associated marginals) would conventionally be displayed in a 2 column  $\times$   $k$  row table with row and column sums. The  $X$ -squared statistic would be compared to chi-squared distribution with  $(k-1)$  degrees of freedom. The row totals do not need to be equal. The usual condition for validity is that the number of cells with counts below 5 is fewer than 20%.

Your added information about "p being known" may be a ringer. What sort of global hypothesis are you thinking of testing when the proportion of successes is already given?