A significant minimization of Pearson's χ^2 statistics in 2x2 contingency tables: preliminary results for small samples

Nicola Serra Ph.D.

(1) Department of Pediatrics, School of Medicine and Surgery, University Federico II of Naples, Italy - E-mail: nicola.serra@unina.it

CORRESPONDING AUTHOR: *Nicola Serra Ph.D.- Department of Pediatrics, School of Medicine and Surgery, University Federico II of Naples, Italy E-mail: nicola.serra@unina.it*

DOI: 10.2427/12949
Accepted on September 5, 2018

ABSTRACT

The Pearson's chi-square test or χ^2 test represents a nonparametric test more used in Medicine, Biology and Social Sciences, but it introduces some error for 2x2 contingency tables, therefore Yates introduces a continuity correction. This correction produces a very conservative result of χ^2 statistics with overestimation of p-value and consequently a type II error is very likely. The goal of this paper is to define, with a statistical approach, a significant minimization of Pearson's χ^2 statistics for small data sample, based on the concept of the arithmetic mean, that could be a possible efficient statistic for reducing the type II error in the calculation of p-value.

Key words: Pearson's χ^2 statistics; Yates's continuity correction; 2x2 contingency table; minimized Pearson's χ^2 statistics; Monte Carlo simulation; Fisher's exact test

INTRODUCTION

Frequently in Biomedicine and Social Sciences, the researchers applied both parametric and nonparametric statistical tests, to define the best diagnostic tool or therapy. The nonparametric test commonly used, is the Pearson's chi-square test or χ^2 test. This test is based on the calculation of Pearson's χ^2 statistics from a sample of a population characterized by two o more dichotomous variables. For two dichotomous variables, it is possible to define a 2x2 contingency table with the frequencies of occurrence of all combinations of their levels, considering a sample size equal to *N*. Table 1 shows a generic scheme of 2x2 contingency table.

The most convenient formula to compute Pearson's χ^2 statistic is [1]:

$$\chi_P^2 = \frac{(ad-bc)^2 N}{(a+b)(c+d)(a+c)(b+d)} = \frac{(ad-bc)^2 N}{r_1 r_2 c_1 c_2} \quad (1)$$

where r_1, r_2, c_1 and c_2, i.e. the totals across rows and columns are generally called marginal totals. Using the χ^2 distribution to interpret Pearson's χ^2 statistic requires one to assume that the discrete probability of observed binomial frequencies of 2x2 contingency table, can be approximated by the continuous χ^2 distribution. This assumption is not entirely correct and introduces some error. To reduce the error in approximation, Yates F.

TABLE 1. 2x2 contingency table form.

	Column variable		
Row variable	*State 1*	*State 2*	*Row totals*
State 1	*a*	*b*	$a + b = r_1$
State 2	*c*	*d*	$c + d = r_2$
Column totals	$a + c = c_1$	$b + d = c_2$	$N = a + b + c + d$

[2] suggested a correction for continuity that adjusts the formula for Pearson's χ^2 test by subtracting the value 0.5, from the difference between each observed value and its expected value for 2x2 contingency table. This correction reduces the χ^2 value obtained and consequently increases its p-value. The formula to compute Yates's χ^2 statistics in a 2x2 contingency table is [1]:

$$\chi^2_{Yates} = \frac{\left(|ad-bc|-\frac{1}{2}N\right)^2 N}{(a+b)(c+d)(a+c)(b+d)} = \frac{\left(|ad-bc|-\frac{1}{2}N\right)^2 N}{r_1 r_2 c_1 c_2} \qquad (2)$$

The effect of Yates's correction is to prevent underestimation of statistical significance for small data samples. This formula is chiefly used when at least one cell of the table has an expected count lower than 5. Unfortunately, Yates's correction may tend to overcorrect of p-value; this can implicate an overly conservative result, i.e., the null hypothesis (H_0) is accepted when it should be rejected (type II error), as reported by several other authors, [3-8].

The continuity correction of original Pearson's χ^2 statistics for 2x2 contingency tables, could be done considering a minimization of Pearson's χ^2 statistics. In this paper we propose a significant minimization of Pearson's χ^2 statistics obtained with the numerical and statistical approach and based on the concept of the arithmetic mean, considering only small samples in this step. This study represents the first step to define an optimum continuity correction of Pearson's χ^2 statistics for 2x2 contingency tables.

METHODS

The approach to defining a minimization of Pearson's χ^2 statistics, is based on the observation that the denominator r_1 r_2 c_1 c_2 of (1), can be interpreted as a geometric mean (G):

$$G = \left(\sqrt[4]{r_1 r_2 c_1 c_2}\right) \Rightarrow G^4 = r_1 r_2 c_1 c_2 \qquad (3)$$

It is noted that the geometric mean is less or equal to arithmetic mean ($\bar{X}$), [9]

$$G \le \bar{X} \Rightarrow G^4 \le \bar{X}^4$$

therefore we can write,

$$\chi^2 = \frac{(ad-bc)^2 N}{r_1 r_2 c_1 c_2} = \frac{(ad-bc)^2 N}{G^4} \ge \frac{(ad-bc)^2 N}{\bar{X}^4} = \frac{(ad-bc)^2 N}{\left(\frac{r_1+r_2+c_1+c_2}{4}\right)^4} = \frac{(ad-bc)^2 N}{\left(\frac{N}{2}\right)^4} = \frac{16}{N^3}(ad-bc)^2 = \chi^2_{\min} \qquad (4)$$

where: $r_1 + r_2 = c_1 + c_2 = N$.

Finally, data were expressed as mean ± Standard Deviation (SD) or percentage. Multicomparison test among χ^2_P, χ^2_{Yates} and χ^2_{min} was performed with analysis of variance for repeated measures (rANOVA). If rANOVA was significant (p-value<0.05), the Bonferroni correction of p-value was used in the pairwise comparison [10]. The Binomial sign test was performed to test the difference between two paired proportions [9]. A value of p-value<0.05 was considered statistically significant.

RESULTS

600,000 2x2 contingency tables were simulated randomly with Monte Carlo method [11], and only 9547 were univocal cases and therefore considered. For every contingency table we computed Pearson's χ^2 (χ^2_P), Yates's χ^2 (χ^2_{Yates}) and minimized Pearson's χ^2 statistic (χ^2_{min}), assigning to *a*, *b*, *c* and *d*, the random numbers generated from uniform pseudorandom number distribution, under condition that the sample size was an arbitrary number less or equal to 25 ($N \le 25 = a + b + c + d$) and under hypotheses of continuity correction. Therefore the simulated distributions of χ^2_P, χ^2_{Yates} and χ^2_{min} statistic were defined and their characteristics were shown in Table 2.

Notably, by χ^2 distribution with one degree of freedom and a significant level equal to 0.05 (α=0.05), the correspondent values of χ^2_P, χ^2_{Yates} and χ^2_{min} for significant tests, are higher than χ^2_α = 3.84 (table of the values of χ^2 distribution). The chi-square distribution is a distinct case of the gamma distribution. It is one of the most widely used probability distributions in inferential statistics [11]. Therefore under this condition, we compared χ^2_P, χ^2_{Yates} and χ^2_{min} for both significant and no significant tests. In Table 3, we reported our results.

By Table 3, the χ^2_P test was significant in 24.40%

TABLE 2. Characteristics of simulated distributions of χ^2_P, χ^2_{Yates} and χ^2_{min}

Parameters	χ^2_P	χ^2_{min}	χ^2_{Yates}
Sample size	9547	9547	9547
Lowest value	0.00	0.00	0.00
Highest value	17.63	17.57	14.43
Arithmetic mean	2.60	2.01	1.53
95% CI for the Arithmetic mean	[2.53; 2.66]	[1.95; 2.06]	[1.48;1.58]
Median	1.27	0.84	0.38
95% CI for the median	[1.22; 1.34]	[0.80;0.89]	[0.35; 0.41]
Variance	10.23	7.49	5.57
Standard deviation	3.20	2.74	2.36
Standard error of the mean	0.033	0.028	0.024

TABLE 3. Test results, for χ^2_P, χ^2_{Yates} and χ^2_{min}, considering a significant level α<0.05.

	Pearson's χ^2 statistic		
Yates's χ^2 statistic	*Significant test*	*No significant test*	**Minimum's χ^2 statistic**
Significant test	13.01% (1242/9547)	0.00%(0/9547)	*Significant test*
Significant test	0.79% (75/9547)	0.00%(0/9547)	*No significant test*
No significant test	4.96% (474/9547)	0.00%(0/9547)	*Significant test*
No significant test	5.38%(538/9547)	75.60%(7218/9547)	*No significant test*

(2329/9547), χ^2_{Yates} test in 13.79% (1317/9547), and χ^2_{min} test in 17.97% (1716/9547). Particularly considering only significant tests of χ^2_P, it resulted that χ^2_{Yates} test was significant in 56.55% (1317/2329) and χ^2_{min} test in 73.68% (1716/2329). Therefore, considering only significant tests of χ^2_P statistic, it resulted $\chi^2_{min} > \chi^2_{Yates}$ (73.68% vs. 56.55%, p-value < 0.0001), i.e., there was, on a significant test of χ^2_P, a considerable presence of significant cases of χ^2_{min} in comparison to χ^2_{Yates}, considering small data samples, as Figure 1 showed.

Additional investigations were considered to verify that χ^2_{min} is an excellent significant continuity correction, in comparison to χ^2_{Yates}.

For this scope, we performed both qualitative and quantitative analysis. By qualitative analysis (Table 4), we compared χ^2_P, χ^2_{Yates} and χ^2_{min} values for every simulated contingency table.

The observed proportions by simulations, expressed as a percentage, were subsequetly compared to hypothetical pre-specified percentages [12]. These represented significant cut-off rates individuated with optimized values of significant levels and power tests, according to Machin et al. [12]. In other terms, we found by Table 4, a very high probability that for small data sample, the inequalities: $\chi^2_P \geq \chi^2_{min}$, $\chi^2_P \geq \chi^2_{Yates}$, and $\chi^2_{min} \geq \chi^2_{Yates}$ are verified with percentages upper to 99.80%, 80%, and 70% respectively.

By quantitative analysis, we tested with repeated measures analysis of variances (rANOVA test), if between the simulated distributions of χ^2_P, χ^2_{Yates} and χ^2_{min}, there were significant differences. It resulted that there was a substantial difference among χ^2_P, χ^2_{Yates} and χ^2_{min} distribution (p-value<0.0001). Subsequently, since the rANOVA test was significant (p-value < 0.05), we performed the pairwise comparisons, where the different measurements were compared to each other, using Bonferroni correction, as suggested by Bland M., [11]. In Table 5 we report the post hoc rANOVA test results and power test (1-β) [12].

Finally to confirm our results and verify the efficacy of χ^2_{min} test, a comparison among Fisher's exact test (FET), Yates, and χ^2_{min} test was performed. The Fisher's exact test is an alternative to the χ^2_P test, based on the calculation of marginal probabilities [13,14]. Particularly the Fisher's exact test, actually is considered the "gold test", used when χ^2_P test is not appropriate, i.e., if the sample size is small and the expected values in any of the cells of a 2x2 contingency table are below 5 [4,9,11,13]

For this step, the significant tests obtained with χ^2_{Yates} and χ^2_{min} were all cases with χ^2_{Yates} and $\chi^2_{min} > \chi^2_{\alpha} = 3.84$. Conversely, the significant tests with FET were obtained considering a two-tail p-value <0.05 [13]. Therefore under these conditions, we compared the three methods for both significant and no significant simulated cases, as showed in Table 6.

By Table 6, FET was significant in 17.72% (1692/9547), χ^2_{Yates} test in 13.80% (1317/9547) and χ^2_{min} test in 17.97% (1716/9547). Particularly considering only significant tests obtained by FET, it resulted that χ^2_{Yates} test was significant in 77.84% (1317/1692) and χ^2_{min} test in 88.59% (1499/1692).

FIGURE 1. Values of χ^2_P, χ^2_{Yates} and χ^2_{min} statistic, for significant simulated cases of χ^2_P

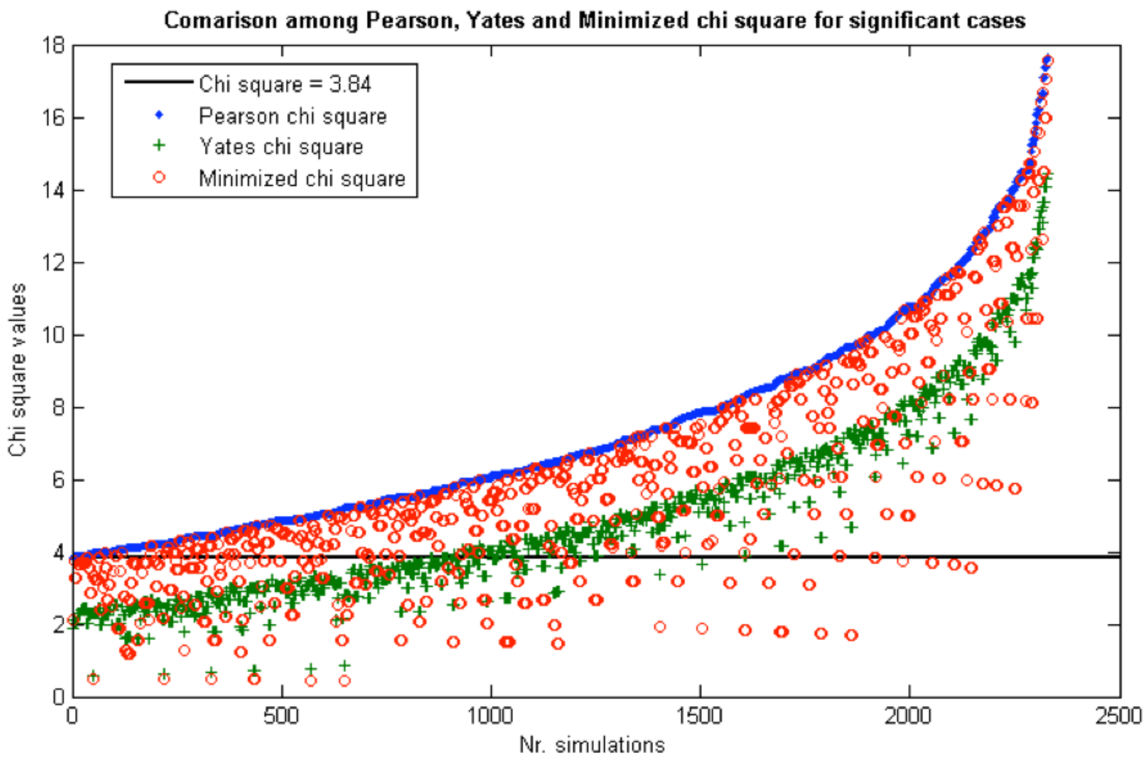

TABLE 4. Results and percentages on simulation data, among χ^2_P, χ^2_{Yates} and χ^2_{min} distributions and cut-off rates with significant levels and power tests.

Results	Proportion (%)	*Hypothesis* (H$_1$)	*Significant level*	*Power test*
$\chi^2_P \geq \chi^2_{min}$	100% (9547/9547)	100% ≥ **99.80 %** * (Z)	α < 0.0001 %	1-β>99.999%
$\chi^2_P \geq \chi^2_{Yates}$	86.78% (8285/9547)	86.78% ≥ **80.00%** * (Z)	α < 0.0001 %	1-β>99.999%
$\chi^2_{min} \geq \chi^2_{Yates}$	79.18% (7559/9547)	79.18% ≥ **70.00%*** (Z)	α < 0.0001 %	1-β>99.999%

* *= significant test; Z= Z-test; cut-off percentage = percentage in bold*

TABLE 5. Pairwise comparisons test, after the significant multicomparison rANOVA test (p-value < 0.0001) among distributions of χ^2_P, χ^2_{Yates} and χ^2_{min} with power test.

Hypothesis H$_1$	Mean difference	Std. error	*p-value*	95% C.I.	*Significant level*	*Power test*
$\chi^2_P \geq \chi^2_{min}$	0.59	0.010	p<0.00001*	0.57-0.62	α < 0.0001 %	1-β>99.99%
$\chi^2_P \geq \chi^2_{Yates}$	1.07	0.001	p<0.00001*	1.04-1.09	α < 0.0001 %	1-β>99.99%
$\chi^2_{min} \geq \chi^2_{Yates}$	0.48	0.007	p<0.00001*	0.46-0.49	α < 0.0001 %	1-β>99.99%

p-value was computed with t-Student test with Bonferroni correction for multiple comparison;
* *= significant test; p =p-value*

Therefore in this case, with χ^2_{min} statistic there is a considerable increase about 10.75% in comparison to χ^2_{Yates}, (88.59% vs. 77.84%, p-value < 0.0001), with a probability greater to 99.99%. Instead, no significant differences there were between the percentages of significant tests obtained with FET and χ^2_{min} test (17.72% vs. 17.97%, p-value > 0.05). In other words for small samples, χ^2_{min} is less conservative in comparison to χ^2_{Yates} statistic, confirming the previous results.

DISCUSSION

The χ^2 statistic defined, is a minimization of Pearson's χ^2 statistics, significant statistically, efficient and very simple

TABLE 6. Significant and no significant number of tests performed with Fisher's exact test, χ^2_{Yates} and χ^2_{min} statistic, considering a significant level α=0.05 and one degree of freedom.

	Fisher's exact test (FET)		
Yates's χ^2 statistic	*Significant test*	*No significant test*	**Minimum's χ^2 statistic**
Significant test	13.01% (1242/9547)	0.00%(0/9547)	*Significant test*
Significant test	0.79% (75/9547)	0.00%(0/9547)	*No significant test*
No significant test	2.59% (257/9547)	2.27%(217/9547)	*Significant test*
No significant test	1.24%(118/9547)	80.00%(7638/9547)	*No significant test*

to compute (4). By statistical tests, we observed that for small samples, the minimized Pearson's χ^2 statistic in 2x2 contingency table, reduces the Pearson's χ^2 value, but it provides to active correction in comparison Yates's χ^2 value. In other words, it seems to be more conservative compared with Pearson's χ^2 statistics, reducing the type I error and less cautious in comparison to Yates's χ^2 statistics, reducing the type II error. Consequently, it increases the power of the test.

These results were confirmed, comparing Fisher's exact test with Yates, and χ^2_{min} test. Considering only significant tests, we observed no significant difference between FET and Minimized chi-square test (17.72% vs. 17.97%, p-value > 0.05) while a significant difference was found between FET and Yates chi-square test (17.72% vs. 13.80%, p-value < 0.0001). Notably, there was 2.3% (217/9547) of cases that were significant with FET but not with χ^2_{min} test. In these cases, we observed that the χ^2_{min} test fail if in contingency table there was in any of the cells an expected value less than 2.

Vice versa in 2.03% (193/9547) of cases, we had a significant χ^2_{min} test, but not FET. In these cases, no evidence was observed from the data. In both cases, more frequently were found, for a statistical analysis that fails between χ^2_{min} and FET, values of χ^2_{min} and p-value near to significant level (χ^2_{α} = 3.85 and α=0.05 respectively). We conclude that for small samples, the performance of Fisher's exact test and χ^2_{min} test were statistically equal. Therefore, based on the results obtained here, the χ^2_{min} statistic, could be used to define an efficient χ^2_{min} distribution, to minimize the continuity correction problem.

The objective of further developments will be oriented, to define a minimum χ^2 distribution, to improve the approximation of the discrete probability of observed binomial frequencies in 2x2 contingency tables with χ^2 distribution.

Competing interests

There are no competing interests for this study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Yates F, (1984). Tests of Significance for 2 × 2 Contingency Tables. Journal of the Royal Statistical Society. Series A (General), Vol. 147, No. 3 (1984), pp. 426-463. DOI: 10.2307/2981577
2. Yates F, (1934). Contingency tables involving small numbers and the c2 test. Journal of the Royal Statistical Society, Suppl.1, 217–235. DOI: 10.2307/2983604
3. Camilli G and Hopkins KD, (1978). Applicability of chi-square to 2 x 2 contingency tables with small expected cell frequencies. Psychological Bulletin, 85, 163-167. DOI: 10.1037/0033-2909.85.1.163
4. Campbell I (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Statistics in Medicine 26:3661-3675. DOI: 10.1002/sim.2832
5. Haber M, (1982), The continuity correction and statistical testing. International Statistical Review, 50, 135–144. DOI: 10.2307/1402597
6. Maxwell EA (1976) Analysis of contingency tables and further reasons for not using Yates correction in 2 × 2 tables. Canadian Journal of Statistics, 4, 277–290. DOI: 10.2307/3315141
7. Richardson JTE, (1990). Variants of chi–square for 2x2 contingency tables. British Journal of Mathematical and Statistical Psychology, 43, 309–326. DOI: 10.1111/j.2044-8317.1990.tb00943.x
8. Richardson JTE (2011). The analysis of 2 x 2 contingency tables - Yet again. Statistics in Medicine 30:890. DOI: 10.1002/sim.4116
9. Sheskin DJ (2004) Handbook of parametric and nonparametric statistical procedures. 3rd ed. Boca Raton: Chapman & Hall /CRC
10. Bland, M. (2015). An introduction to medical statistics. Oxford University Press (UK).
11. Rubinstein, R. Y., & Kroese, D. P. (2011). Simulation and the Monte Carlo method (Vol. 707). John Wiley & Sons. DOI: 10.1002/9781118631980
12. Machin D, Campbell MJ, Tan SB, Tan SH (2009) Sample size tables for clinical studies. 3rd ed. Chichester: Wiley-Blackwell.
13. Upton, G. J. (1992). Fisher's exact test. Journal of the Royal Statistical Society. Series A (Statistics in Society), 395-402. DOI: 10.2307/2982890
14. Fisher, R.A. Statistical Methods for Research Workers; Oliver and Boyd, Edinburgh, UK, 1934.