The Korean Journal of Applied Statistics (2010)
23(4), 777–781

The Eccentric Properties of the Chi-Squared Test with Yates' Continuity Correction in Extremely Unbalanced 2×2 Contingency Tables

Seung-Ho Kang[1] · Tae-Hyuk Kwon[2]

[1]Department of Applied Statistics, Yonsei University
[2]Department of Applied Statistics, Yonsei University

(Received June 2010; accepted July 2010)

Abstract

Yates' continuity correction of the chi-squared test for testing the homogeneity of two binomial proportions in 2×2 contingency tables is developed to lower the value of the test statistic slightly. The effect of continuity correction is expected to decrease as the sample size increases. However, in extremely unbalanced 2×2 contingency tables, we find some cases where the effect of continuity correction is eccentric and is larger than expected. In such cases, we conclude that the chi-squared test with continuity correction should not be employed as a test statistic in both asymptotic tests and exact tests.

Keywords: Exact test, type I error, homogeneity, binomial distribution.

1. Introduction

Testing the homogeneity of two binomial proportions is one of the most fundamental problems in modern statistics and studies of extremely unbalanced 2×2 contingency tables are scant. We often encounter extremely unbalanced 2×2 contingency tables in many application fields when we would like to compare the proportions of interest between a very large population and a small population.

The chi-squared test with Yates' continuity correction seems to still be popular and is introduced in many books (for example, Rosner, 2000; Indrayan, 2008). Therefore, it is important to understand the characteristics of the chi-squared test with Yates' continuity correction. This paper shows that the chi-squared test with Yates' continuity correction has unexpected eccentric properties in extremely unbalanced 2×2 contingency tables. Specifically, the effect of continuity correction is expected to be small for large samples. Therefore, it is believed that the use of continuity correction is not an issue in large samples. However, in Sections 3 and 4, we provide some cases in which continuity correction makes quite a significant between the values of the chi-squared test with and without correction. We investigate why such phenomena occur. In Section 5, we discuss the most appropriate statistical tests for analyzing extremely unbalanced 2×2 contingency tables.

[1]Corresponding author: Professor, Department of Applied Statistics, Yonsei University, 262 Seongsanno, SeoDaeMun-Gu, Seoul 120-749, Korea. E-mail: seungho@yonsei.ac.kr

Table 2.1. A 2×2 contingency table

	group 1	group 2	Totals
response	X_{11}	X_{12}	R_1
no response	X_{21}	X_{22}	R_2
Totals	n_1	n_2	n

2. Notation and Review

Let X_{11} and X_{12} be two independent binomial random variables, $X_{1k} \sim B(n_k, p_k)$, $k = 1, 2$. Then we have the following 2×2 contingency table:

where $R_j = X_{j1} + X_{j2}$, for $j = 1, 2$, $n = n_1 + n_2$. We wish to test $H_0 : p_1 = p_2 = p$ versus $H_1 : p_1 \neq p_2$ where $0 < p < 1$.

Ostensibly, the chi-squared test has been the most popular method for testing the homogeneity of two binomial proportions, if the sample size is neither too small nor extremely unbalanced. The chi-squared test compares the observed frequency with the expected frequency in each category in the contingency table under the null hypothesis. It is well-known that the chi-squared test statistic approximately follows the chi-square distribution with one degree of freedom under the null hypothesis. The chi-squared test rejects H_0 if $X^2 > \chi^2_{1,\alpha}$, where $\chi^2_{1,\alpha}$ is the upper α-quintile of the chi-square distribution with one degree of freedom and

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(X_{ij} - E_{ij})^2}{E_{ij}}, \qquad E_{ij} = \frac{R_i n_j}{n}, \quad \text{for } i, j = 1, 2.$$

The chi-squared test is valid when all expected cell frequencies (E_{ij}) are reasonably large. Fisher (1925) argued that the chi-squared test should be used only when $\min(E_{ij}) > 5$. Many textbooks adopt this guideline (for example, Rosner, 2000).

In order to improve this continuous approximation to a discrete distribution, a continuity correction is sometimes used. Although there are many different types of continuity correction, the following Yates' continuity correction seems to be the most popular (Yates, 1934; Martin Andres and Silva Mato, 1994; Rosner, 2000)

$$X_c^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(|X_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}.$$

Continuity correction decreases the value of the test statistic and it increases the corresponding p-value. The correction makes the test more conservative and less likely to reject the null hypothesis. This improves the approximation in some cases but can make the test overly conservative in other cases. There have been debates regarding the use of continuity correction. Grizzle (1967) and Conover (1968, 1974) recommend that the correction for continuity not be applied. They give as their reason an apparent lowering of the actual significance level when the correction is used. A lowered significance level results in a reduction in power. Haviland (1990) also argued against use of Yates' continuity correction. Mantel and Greenhouse (1968) illustrate the inappropriateness of Grizzle's analyses and refute his argument against the use of the correction. It is not the present purpose of this paper to explore the debate.

Table 3.1 displays four 2×2 contingency tables whose values of the chi-squared test are similar.

As expected, Table 3.1 shows that the differences between X^2 and X_c^2 decrease as the sample size

Table 2.2. The effect of continuity correction as the sample size increases

n_1	n_2	X_{11}	X_{12}	X^2	X_c^2	$X^2 - X_c^2$
30	30	10	4	3.35	2.33	1.02
100	100	30	20	2.67	2.16	0.51
1000	1000	300	265	3.02	2.85	0.17
10000	10000	3000	2890	2.91	2.86	0.05

Table 3.1. The value of X_c^2 is much smaller than that of X^2

	group 1	group 2	Totals
response	10	5200	5210
no response	996	994001	994997
Totals	1006	999201	1000207

increases. In Sections 3 and 4, we will show that there are two categories in which the differences between X^2 and X_c^2 are greater than expected, although the total sample size is large.

3. The Cases in Which the Values of X_c^2 are Much Smaller than Those of X^2

Let us consider the following hypothetical 2×2 contingency table.

The values of X^2 and X_c^2 in Table 3.1 are 4.35 and 3.48, respectively. Considering the total sample size $(1,000,207)$, the difference between the values of X^2 and X_c^2 $(0.87 = 4.35 - 3.48)$ is larger than expected. When $n_1 : n_2 \sim 1 : 1000$ and $n \sim 10^6$, we find many hypothetical extremely unbalanced 2×2 contingency tables whose expected cell frequencies are all greater than five, but there are significantly large differences between the values of X^2 and X_c^2. Table 3.1 is just one of them.

The reason for this phenomenon is that we subtract 0.5 in the numerator of X_c^2. When the value of $|X_{ij} - E_{ij}|$ is large, subtracting 0.5 just makes the value of X_c^2 slightly smaller. Especially, when the sample size is large, the differences between the values of X^2 and X_c^2 become very small. However, in an extremely unbalanced table such as Table 3.1 above, the value of $|X_{ij} - E_{ij}|$ is not large in cell (1,1) and it causes a discrepancy between the values of X^2 and X_c^2. Particularly, in Table 3.1,

$$|X_{11} - E_{11}| = 4.8, \qquad \frac{(X_{11} - E_{11})^2}{E_{11}} = 4.3, \qquad \frac{(|X_{11} - E_{11}| - 0.5)^2}{E_{11}} = 3.5.$$

We may employ either asymptotic tests or exact tests to analyze Table 3.1. Since all expected cell frequencies are greater than five ($E_{11} = 5.2$, $E_{12} = 5204.8$, $E_{21} = 1000.8$ and $E_{22} = 993996.2$), the chi-square approximation seems to be well justified. In asymptotic tests, the p-values of X^2 and X_c^2 are 0.037 and 0.062, respectively, and produce opposite results at a 5% nominal level. Exact tests may be employed as alternatives. However, the values of $X^2(= 4.35)$ and $X_c^2(= 3.48)$ do not change in exact tests and only exact distributions are used instead of the asymptotic distribution to compute the p-value. So continuity correction makes the chi-squared test very conservative in both asymptotic tests and exact tests. Therefore, we conclude that Yates' continuity correction should not be used in both asymptotic tests and exact tests to analyze Table 3.1. In Section 5, we will discuss which method is the most appropriate for analyzing data in Table 3.1.

4. The Cases in Which the Values of X_c^2 are Much Larger than Those of X^2

Kang *et al.* (2006) found that the size of the chi-squared test with continuity correction is much greater than the nominal level in extremely unbalanced 2×2 contingency tables. This result

Table 4.1. The value of X_c^2 is larger than that of X^2

	group 1	group 2	Totals
response	0	3	3
no response	1000	99397	100397
Totals	1000	99400	100400

motivates the search for extremely unbalanced 2×2 contingency tables in which the chi-squared test with continuity correction produces huge values. We found many examples without difficulty and the following table is just one example.

The values of X^2 and X_c^2 are 0.03 and 7.47, respectively. Since two sample proportions are almost identical ($X_{11}/n_1 = 0$ and $X_{12}/n_2 = 3.02 \times 10^{-5}$), the result of X_c^2 is absurd.

The reason that X_c^2 produces an absurd result can be explained as follows. Since we have $X_{11} = 0$ and $E_{11} = 3 \times 1000/100400 = 0.03$,

$$\frac{(X_{11} - E_{11})^2}{E_{11}} = E_{11} = 0.03, \qquad \frac{(|X_{11} - E_{11}| - 0.5)^2}{E_{11}} = \frac{(E_{11} - 0.5)^2}{E_{11}} = 7.40.$$

That is, when $X_{11} = 0$, 0.5 in the numerator of X_c^2 inflates the value of X_c^2.

Note that some expected cell frequencies are smaller than five ($E_{11} = 0.03$ and $E_{12} = 2.97$). Therefore, the value of X_c^2 will not be compared with the 95% percentile of the chi-square distribution with one degree of freedom. However, exact tests can be used as alternatives. For data in Table 4.1, if exact tests are employed with X_c^2 as a test statistic, the result is misleading.

5. Discussion

The question of which statistical tests are the most appropriate for analyzing Tables 3.1 and 4.1 could be interesting for future research. In Table 3.1, since all expected cell frequencies are greater than five, the chi-square approximation seems to be valid. However, Kang *et al.* (2006) showed that the chi-squared test with and without continuity correction may have inflated type I error in an extremely unbalanced 2×2 contingency table when the common success probability is very small. Specifically, the inflation of type I error rates depends on both the sample ratio (n_2/n_1) and the common success probability (p). Since the common success probability is unknown, it is not easy to ensure that the type I error rates are well controlled under the nominal level. Therefore, we think that exact tests are safer ways to analyze Table 3.1 than asymptotic tests. In addition, since some expected cell frequencies are very small in Table 4.1, exact tests are the only feasible methods to analyze that table.

It takes less than one second to conduct a conditional exact chi-squared test for Tables 3.1 and 4.1 with StatXact on a personal computer. However, the unconditional exact chi-squared test is not feasible with StatXact because the total sample size is too large. Recently, Kang and Ahn (2008) showed that the conditional exact chi-squared test is more powerful than the unconditional exact chi-squared test in extremely unbalanced 2×2 contingency tables when the sample ratio is greater than 20. Martin Andres and Silva Mato (1994) and Martin Andres *et al.* (1998) investigated the optimal unconditional exact test when the sample size is not large. It is of interest to find a test statistic in exact tests which produce the greatest power in extremely unbalanced 2×2 contingency tables. The chi-squared test with continuity correction should not be employed as a test statistic in exact tests for extremely unbalanced 2×2 contingency tables.

Acknowledgements

This work was supported by grant No. R01-2007-000-20713-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

References

Conover, W. J. (1968). Uses and abuses of the continuity correction, *Biometrics*, **24**, 1028.

Conover, W. J. (1974). Some reasons for not using the Yates' continuity correction on 2×2 contingency tables (With comments), *Journal of the American Statistical Association*, **69**, 374–382.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, Edinburgh, Oliver and Boyd.

Grizzle, J. E. (1967). Continuity correction in the χ^2 test for 2×2 tables, *American Statistician*, **21**, 28–32.

Haviland, M. G. (1990). Yates's correction for continuity and the analysis of 2×2 contingency tables, *Statistics in Medicine*, **9**, 363–367.

Indrayan, A. (2008). *Medical biostatistics*, Chapman & Hall/CRC, Boca Raton.

Kang, S. H. and Ahn, C. (2008). Tests for the homogeneity of two binomial proportions in extremely unbalanced 2×2 contingency tables, *Statistics in Medicine*, **27**, 2524–2535.

Kang, S. H., Lee, Y. and Park, E. S. (2006). The sizes of the three popular asymptotic tests for testing homogeneity of two binomial proportions, *Computational Statistics and Data Analysis*, **51**, 710–722.

Mantel, N. and Greenhouse, S. W. (1968). What is the continuity correction?, *American Statistician*, **22**, 27–30.

Martin Andres, A. and Silva Mato, A. (1994). Choosing the optimal unconditioned test for comparing two independent proportions, *Computational Statistics and Data Analysis*, **17**, 555–574.

Martin Andres, A., Sanchez Quevedo, M. J. and Silva Mato, A. (1998). Fisher's mid-p-value arrangement in 2×2 comparative trials, *Computational Statistics and Data Analysis*, **29**, 107–115.

Rosner, B. (2000). *Fundamentals of Biostatistics*, third edition, PWS-Kent, New York.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test, *Journal of the Royal Statistical Society Supplement*, **1**, 217–235.