Yates's continuity correction

Catalina Stefanescu [*], Vance W. Berger [†]
Scott Hershberger [‡]

Yates's correction [17] is used as an approximation in the analysis of 2×1 and 2×2 contingency tables. A 2×2 contingency table shows the frequencies of occurrence of all combinations of the levels of two dichotomous variables, in a sample of size N. A schematic form of such a table is given by the figure below.

Table 1: A 2×2 contingency table.

	Column variable		
Row variable	1	2	Totals
1	A	B	$A+B$
2	C	D	$C+D$
Totals	$A+C$	$B+D$	N

A research question of interest is often whether the variables summarized in a contingency table are independent of each other. The test to determine if this is so depends on which, if any, of the margins are fixed, either by

[*]London Business School. Email: cstefanescu@london.edu
[†]University of Maryland Baltimore County. Email: vance917@comcast.net
[‡]California State University, Department of Psychology. Email: scotth@csulb.edu

design or for the purposes of the analysis. For example, in a randomized trial in which the number of subjects to be randomized to each treatment group has been specified, the row margins would be fixed but the column margins would not (it is customary to use rows for treatments and columns for outcomes). In a matched study, however, in which one might sample 100 cases (smokers, say) and 1000 controls (non–smokers), and then test each of these 1100 subjects for the presence or absence of some exposure that may have predicted their own smoking status (perhaps a parent who smoked), it would be the column margins that are fixed. In a random and unstratified sample, in which each subject sampled is then cross–classified by two attributes (say smoking status and gender), neither margin would be fixed. Finally, in Fisher's famous tea–tasting experiment [15], in which a lady was to guess whether the milk or the tea infusion was first added to the cup by dividing 8 cups into two sets of 4, both the row and the column margins would be fixed by the design. Yet in the first case mentioned, that of a randomized trial with fixed row margins but not fixed column margins, the column margins may be treated as fixed for the purposes of the analysis, so as to ensure exactness [2].

When the row and column margins are fixed, either by design or for the analysis, independence can be tested using Fisher's exact test [4]. This test is based on the hypergeometric distribution and it is computationally intensive, especially in large samples. Therefore, Fisher advocated the use of Pearson's statistic,

$$X^2 = \frac{N(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)},$$

which under the null hypothesis has a χ^2 distribution with one degree of freedom. Yates [17] argued that the χ_1^2 distribution gives only approximate estimates of the discrete probabilities associated with frequency data, and thus the p–values based on Pearson's X^2 statistic will generally underestimate the true p–values. In general, when a statistic takes discrete values $a < b < c$, the p–value corresponding to b is estimated by the tail of the continuous function defined by the point $\frac{a+b}{2}$. Therefore, the tail of the continuous function computed at b will underestimate the p–value. In this context, Yates suggested that X^2 should be corrected for continuity and proposed the corrected test statistic

$$\frac{N(|AD - BC| - \frac{1}{2}N)^2}{(A+B)(C+D)(A+C)(B+D)}.$$

Although Yates's correction is best known for its use in the analysis of 2×2 contingency tables, it is also applicable to the analysis of 2×1 contingency tables. A 2×1 contingency table displays the frequencies of occurrence of two categories in a random sample of size N, drawn from a population in which the proportions of cases within the two categories are p and $1 - p$. The research question is usually whether the observed numbers of cases x and $N - x$ in the two categories have been sampled from a population with some pre–specified value of p. This can be tested using Pearson's statistic,

$$X^2 = \frac{(x - Np)^2}{Np(1-p)},$$

which asymptotically has a χ_1^2 distribution under the null hypothesis. Yates showed that, in this case as well, the use of Pearson's X^2 results in p–values which systematically underestimate the true p–values based on the binomial distribution. Therefore, he suggested the corrected statistic

$$\frac{(|x - Np| - \frac{1}{2})^2}{Np(1-p)}. \tag{1}$$

Kendall and Stuart [8] remarked that Yates's procedure is a special case of a general concept of a continuity correction, while Pearson [11] noted that Yates's correction derives naturally from the Euler–Maclaurin theorem used to approximate binomial and hypergeometric distributions. Subsequently, the use of Yates's correction to Pearson's X^2 has been widely emphasized for the analysis of contingency tables [16]. There are, however, several issues related to Yates's correction and we shall discuss some of these in turn.

Firstly, in the analysis of 2×1 contingency tables, the p–values associated with the corrected statistic (1) tend to overestimate the true p–values in the tails of the distribution and to underestimate them towards the center. This is illustrated in Table 2 which displays the two–tailed p–values in a contingency table with $N = 10$ and $p = 0.5$, obtained with Pearson's X^2 statistic and Yates's correction. The table reports as well the true binomial p–values which are the gold standard. It should also be noted [17] that the p–values obtained with the continuity correction are much less accurate when the binomial probability p is substantially different from 0.5.

Secondly, Yates's correction is appropriate only for one–sided tests, as it is based on a comparison between the observed contingency and the next strongest contingency in the same direction ([7], [9]). For two–sided tests, the statistic involves an overcorrection. Along the same lines, it can be proven analytically that Yates's correction is systematically conservative when carrying out two–sided tests [10].

Thirdly, a more important issue related to Yates's correction is its appli-

Table 2: Binomial distribution for $N = 10$ and $p = 0.5$, and two–tailed p–values. (Adapted from [14])

		p–values		
x	$p(x)$	Pearson	Yates	Binomial
0, 10	0.0010	0.0016	0.0044	0.0020
1, 9	0.0098	0.0114	0.0268	0.0215
2, 8	0.0439	0.0580	0.1138	0.1094
3, 7	0.1172	0.2060	0.3428	0.3437
4, 6	0.2051	0.5270	0.7518	0.7539
5, 5	0.2461	1.0000	1.0000	1.0000

cability to the analysis of contingency tables arising from different research designs. Many researchers have argued that Yates's correction is based upon comparisons among contingency tables with fixed row and column marginal totals, particularly since Yates is specifically concerned with approximating the hypergeometric distribution from Fisher's exact test. However, Yates's method has also been recommended for the analysis of 2×2 contingency tables arising from sampling schemes where one or both sets of marginal totals are free to vary and are thus subject to sampling errors. It should be noted that such sampling schemes are the ones most frequently found in actual research context. While Yates [18] argues along the lines of Fisher's reasoning that the analysis of 2×2 contingency tables should always be performed conditional on the observed marginal totals, this approach is still subject to debate [14]. On the other hand, when the marginal totals are not fixed, Yates's procedure involves an additional overcorrection and the test statistic is conservative. This has been investigated through Monte Carlo simulations ([5], [13]), and confirmed analytically ([3], [7]). In particular, Grizzle [5] notes that for contingency tables with non–fixed marginal totals, Yates's procedure "produces a test that is so conservative as to be almost useless".

Finally, Yates's correction originated as a device of eliminating the discrepancies which arised when approximating the hypergeometric distribution in Fisher's exact test. The approximation using Pearson's X^2 was necessary "for the comparative simplicity of the calculations" ([4], p.99), because the exact analysis of 2×2 contingency tables with the limited computing power available at the time was prohibitive in many cases. This is no longer the case today. Indeed, Agresti [1] notes that Yates's correction is not necessary anymore since current software makes Fisher's exact test computationally feasible even when the sample sizes are large.

References

[1] Agresti, A. (2002) *Categorical Data Analysis.* New York: Wiley.

[2] Berger, V.W. (2000) Pros and cons of permutation tests in clinical trials. *Statistics in Medicine,* **19**, 1319–1328.

[3] Conover, W.J. (1974) Some reasons for not using the Yates continuity correction on 2×2 contingency tables. *Journal of the American Statistical Association,* **69**, 374–376.

[4] Fisher, R.A. (1934) *Statistical Methods for Research Workers.* 5th Edition, Edinburgh: Oliver and Boyd.

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

[6] Haber, M. (1980) A comparison of some continuity corrections for the chi–squared test on 2×2 tables. *Journal of the American Statistical Association,* **75**, 510–515.

[7] Haber, M. (1982) The continuity correction and statistical testing. *International Statistical Review,* **50**, 135–144.

[8] Kendall, M.G. and Stuart, A. (1967) *The Advanced Theory of Statistics.* Vol 2, 2nd Edition, London: Griffin.

[9] Mantel, N. (1976) The continuity correction. *The American Statistician,* **30**, 103–104.

[10] Maxwell, E.A. (1976) Analysis of contingency tables and further reasons for not using Yates correction in 2×2 tables. *Canadian Journal of Statistics,* **4**, 277–290.

[11] Pearson, E.S. (1947) The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika,* **34**, 139–167.

[12] Plackett, R.L. (1964) The continuity correction in 2×2 tables. *Biometrika,* **51**, 327–337.

[13] Richardson, J.T.E. (1990) Variants of chi–square for 2×2 contingency tables. *British Journal of Mathematical and Statistical Psychology,* **43**, 309–326.

[14] Richardson, J.T.E. (1994) The analysis of 2×1 and 2×2 contingency tables: an historical review. *Statistical Methods in Medical Research,* **3**, 107–133.

[15] Salsburg, D. (2002) *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century.* London: Owl Books.

[16] Siegel, S. and Castellan, N.J. (1988) *Nonparametric Statistics for the Behavioural Sciences.* 2nd Edition, New York: McGraw–Hill.

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

[18] Yates, F. (1984) Tests of significance for 2×2 contingency tables. *Journal of the Royal Statistical Society Series A,* **147**, 426–463.