2 x 2 Contingency Table: Fisher's Exact Test

Padam Singh, PhD, *Gurgaon, India*

ABSTRACT

While dealing with comparison of new drug with the standard treatment one is faced with the problem of small numbers specially in phase-1 trials or pilot studies. The analysis of data in such situations requires special treatment. This paper presents the issues involved in the test of significance for the association of attributes for small samples. It is recommended to use Fisher's Exact Test in such situations.

Keywords: Attributes, Association, Chi- square test, Correction for continuity, Exact test

Background

In medical research, a new drug/ intervention is generally compared with standard treatment and outcomes assessed in terms of cured or not cured, and occurrence and non-occurrence of side effects etc. These variables are generally dichotomous i.e. Involves only two possible values. The data is generally displayed in a 2 x 2 contingency table that 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 below.

Table 1: A 2 × 2 contingency table.

Outcome			
	Cured	**Not-Cured**	**Total**
New Drug	a	b	a+b
Standard Drug	c	d	c+d
Total	a+c	b+d	n

In the above contingency table it is customary to use rows for treatments and columns for outcomes. The margins in the contingency table are fixed, either by 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. On the other hand in a retrospective study, where one might sample cases

From: Medanta - The Medicity, Sector 38, Gurgaon, India (P.S.)

Corresponding Author: Dr. Padam Singh, PhD
Medanta - The Medicity, Sector 38, Gurgaon, Haryana-122001, India
Ph: +91-124-4141414 | Fax: +91-124-4834111
Email: dr.padamsingh2013@gmail.com

(diseased persons) and controls (non–diseased), and then record the presence or absence of some exposure, it would be the column margins that are fixed. However, in a random sample, in which each subject sampled is cross–classified by two attributes none of the margins would be fixed.

A research question of interest is whether the attributes in the contingency table are associated or independent of each other. The null hypothesis would be that there is no association or there is no difference in proportions. Two methods of analysis are available for this set-up.

i) Proportion tests based on Gaussian distribution

ii) Chi-square tests based on contingency table

For large samples both tests are equivalent and give identical results. However for small samples the analysis requires special treatment. This paper highlights the issues dealing with small numbers in 2 X 2 contingency table and suggests the appropriate procedure to be used with suitable illustrations.

Chi-square tests

Karl Pearson (1900) suggested use of test statistic as under:

$$\chi^2 = \frac{n\,(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

This test statistic under the null hypothesis follows a Chi-square (χ^2) distribution with one degree of freedom.

Following graph presents a comparative picture of Chi square distribution with 1, 5 and 10 degrees of freedom. Evidently the shape of Chi-square distribution with one degree of freedom is distinctly different with those of higher degrees of freedom and hence demands special

treatment in the statistical analysis.

Chi Square Distribution
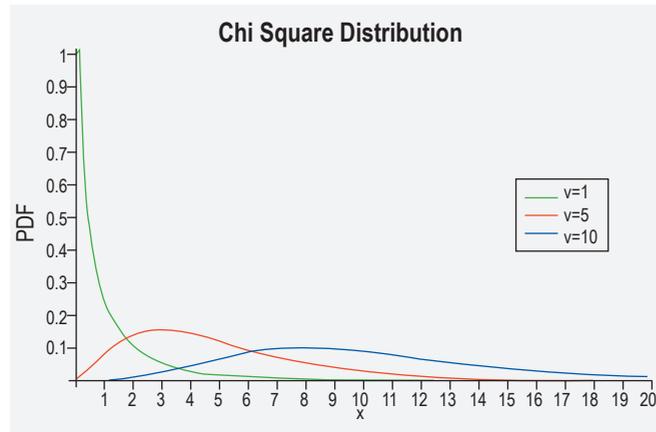
Consider the following examples:

Example 1:

	Cured	Not-Cured	Total
New Drug	8	1	9
Standard Drug	4	5	9
Total	12	6	18

Chi-Square = 4.000,   DF = 1,   P-Value = 0.046

The resulting p-value, 0.046, from the x^2 test indicates that there is a statistically significant difference i.e. the cure rate is higher for new drug as compared to the standard drug (at α = 0.05 level).

Example 2:

	Side effects	No Side-effects	Total
Drug A	8	14	22
Drug B	1	3	4
Total	9	17	26

Chi-Square = 0.1931, DF = 1,   P-Value = 0.6603

The resulting p-value 0.6603 indicates that the difference in the proportion of side effects in two drugs is not statistically significant.

In the following sections, we shall examine whether the interpretation will undergo any change if other methods are used for analysis.

Yates's continuity correction

Yates (1934) argued that in case of small samples the χ^2 distribution gives only approximate estimates of the discrete probabilities associated with frequency data, and thus the p–values based on Pearson's χ^2 statistic will generally underestimate the true p–values. In view of this, Yates suggested that χ^2 should be corrected for continuity and proposed the corrected test statistic as under

$$\chi^2 = \frac{n(|ad - bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

For the example 1 considered above the Chi-square value with continuity correction works out as 2.25. The p value for this is 0.13361440. Contrary to the usual Chi-square test the results indicate no statistically significant difference. For example 2 the Chi-square value is 0.0173 and for this the p value works out as 0.0.8953. Thus the results after Yates correction for continuity are far different from the results of usual Chi-square tests.

It is important to mention that the Yates' continuity correction is designed to make the approximation better. A comparison of binomial distribution (n =12 Π=0.5) with the approximating normal distribution is presented to illustrate the need for continuity correction for small 'n' When either nΠ or n-nΠ (Π being the binomial probability) are below 10, the accuracy of hypothesis test on the normal approximation can be improved by the introduction of a continuity correction. The continuity correction adjusts the numerator of the test statistic so that there is a closer fit. This is illustrated for observing 9 or more events out of 12 with probability, Π=0.5 in Table-2 below.

Table-2: Comparisons of the different methods of probability of observing 9 or more events, when **n**=12 and **Π**=0.5.

Probability of observing 9 or more events, when **n**=12 and **Π**=0.5.	
Calculated using binomial probability	
9 Events	$220 \times 0.5^{12} = 0.0537$
10 Events	$66 \times 0.5^{12} = 0.0161$
11 Events	$12 \times 0.5^{12} = 0.0029$
12 Events	$1 \times 0.5^{12} = 0.0002$
Total of 9+events (exact probability)	**0.0729**
Using approximating normal distribution:	
Based on area above 9	0.0418
With continuity correction, based on area above 8.5	**0.0749**

Source: *Kirkwood BR, Sterne JAC. Medical Statistics, 2nd edition. Oxford: Blackwell Science Ltd; 2003, Page 146*

The above table shows that incorporating a continuity correction and calculating the area under the normal

curve above 8.5 gives a close approximation to the exact binomial probability of observing 9 events or more. On contrast the area of the normal curve above 9 is not a good approximation.

The same is presented graphically as under:
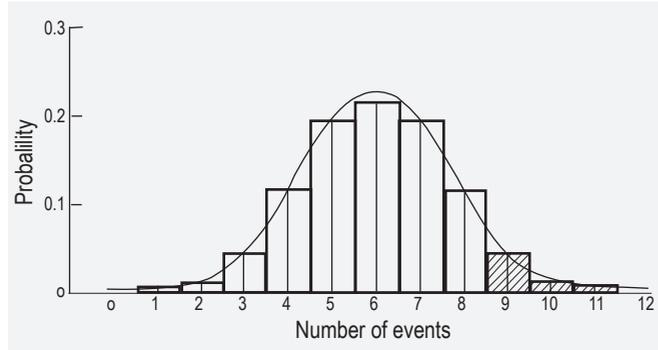
There are several issues related to Yates's correction.

1. The p–values obtained with the continuity correction are much less accurate when the binomial probability **Π** is substantially away from 0.5.

2. Yates's correction is more 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. For two–sided tests, the statistic involves an over correction. Yates's correction is systematically conservative when carrying out two–sided tests.

Fisher's Exact Test

When the row and column margins in 2 X 2 contingency table are fixed, either by design or for the analysis, independence of attributes can be tested using Fisher's exact test. This test is based on the hyper-geometric distribution. The 2 X 2 contingency table can be presented as under:

Outcome			
	Cured	**Not-Cured**	**Total**
New Drug	n_{11}	n_{12}	$n_{1\cdot}$
Standard Drug	n_{21}	n_{22}	$n_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	n

The computation of probability for the contingency table with given cell frequencies using hyper-geometric distribution is as under

$$P\{n_{ij}\} = \frac{n_{1\cdot}!\; n_{2\cdot}!\; n_{\cdot 1}!\; n_{\cdot 2}!}{n\,!\; n_{11}\,!\; n_{12}\,!\; n_{21}\,!\; n_{22}\,!}$$

Computation of p values using Fisher's Exact Test

To start with for given marginal totals all possible two way tables are generated which has desired marginal totals. Thereafter, using the smallest marginal total the table for each ordered pair of integers with that sum is created. The 2 X 2 contingency table for each of these ordered pairs are then completed. The probability for each contingency table is computed using hyper-geometric distribution.

For a one tailed test n_{11} is compared with its expected value computed as

$$\frac{\text{Corresponding row total X Corresponding column total}}{\text{Grand Total}}$$

If n_{11} exceeds its expected value then p value is the sum of the probabilities of all n_{11} more than equal to the expected value. Alternatively, if n_{11} is less than its expected value then p value is the sum of the probabilities of all n_{11} less than equal to the expected value.

The computation of p value for the two examples is explained as under:

Illustration 1: Following table presents the expected frequencies along with the observed frequencies for example 1

Table 3: Comparison of observed and expected frequencies

		Cured	Not-Cured	Total
New Drug	Observed frequencies	8	1	9
	Expected frequencies	6.00	3.00	9.00
Standard Drug	Observed frequencies	4	5	9
	Expected frequencies	6.00	3.00	9.00
All		12	6	18

It is observed that marginal totals are 9, 9 and 12, 6

Fisher's exact test considers all possible cell combinations that would result in the marginal frequencies (namely 9, 9 and 12, 6).

The hyper-geometric probability distribution is used to compute the probability of the observed results. In addition to the table of Example 1, other tables that will be consistent with the marginal frequencies of 9, 9 and 12, 6, along with their associated probabilities, are shown in Table 4.

Table 4: Computation of probability for different 2 X 2

tables corresponding to desired marginal totals.

Table	Hyper Geometric Computation	Associated Probability
9 0 3 6	9!*9!*12!*6! 18!*9!*0!*3!*6!	0.004524887
7 2 5 4	9!*9!*12!*6! 18!*7!*2!*5!*4!	0.244343891
6 3 6 3	9!*9!*12!*6! 18!*6!*3!*6!*3!	0.380090498
5 4 7 2	9!*9!*12!*6! 18!*5!*4!*7!*2!	0.244343891
4 5 8 1	9!*9!*12!*6! 18!*4!*5!*8!*1!	0.061085973
3 6 9 0	9!*9!*12!*6! 18!*3!*6!*9!*0!	0.004524887

Source: When To Use Fisher's Exact Test by Keith M. Bower, M.S.

To compute Fisher's exact test results, look at the tables with probabilities less than or equal to the probability of the observed results (0.061085972). They are highlighted with an *. Add these probabilities together, along with the probability of the observed results, to obtain the p-value for the test. This resulting p-value is 0.13122. Therefore, one would fail to reject the null hypothesis of independence at $\alpha = 0.05$ level. This also contradicts the results from the χ^2 test and incidentally the p-value is close to that obtained using Yates correction.

Illustration 2: Following table presents the expected frequencies along with the observed frequencies for example 2

Table 5: Comparison of observed and expected frequencies

		Side-effects	No Side-effects	Total
Drug A	Observed frequencies	8	14	22
	Expected frequencies	7.6	14.4	9.00
Drug B	Observed frequencies	1	3	22
	Expected frequencies	1.4	2.6	9.00
All		9	17	4

Assuming that the marginal totals are fixed, all possible two way tables which have those marginal totals are worked out. Using the smallest marginal total and the tables for each ordered pair of integers with that sum are created. Thereafter the 2x2 contingency tables for each of these ordered pairs are completed.

Since $n_{2.} = 4$ is the smallest marginal total, we look for the following ordered pairs(n_{21}, n_{22}): (0, 4), (1, 3), (2, 2), (3, 1) and (4,0).

For each table the probability is computed as under:

Table 6: Computation of probability for different 2 X 2 tables of Example 2

Table	Hyper Geometric Computation	Associated Probability
9 13 0 4	22!*4!*9!*17! 26!*9!*13!*0!*4!	0.159197
8 14 1 3	22!*4!*9!*17! 26!*8!*14!*1!*3!	0.409365
7 15 2 2	22!*4!*9!*17! 26!*7!*15!*2!*2!	0.327492
6 16 3 1	22!*4!*9!*17! 26!*6!*16!*3!*1!	0.095518
5 17 4 0	22!*4!*9!*17! 26!*5!*17!*4!*0!	0.008428

n_{11}	$P\{n_{ij}\}$
9	0.159197
8	0.409365
7	0.327492
6	0.095518
5	0.008428

Since the expected value is 7.6, which is less than the observed frequency 8 the P-value is .159197 + .409365 = .568562. Although the inference for Example 2 remains the same for test using χ^2 test, Yates correction and Fisher's Exact Test, the p-value is distinctly different.

Conclusion

The table below summarizes the comparison of p values for the two examples using different tests of significance

Tests of significance	Example 1	Example 2
Chi square	0.0460	0.6603
Yates Correction	0.1336	0.8953
Fisher's Exact test	0.1312	0.5685

The highlights of conclusion are as under:

- χ^2 test without Yates's correction yields much lower p values.

- The p–values obtained using Yates's correction tends to overestimate the true p–values.

- The Fisher's test is exact because it uses the hyper-geometric distribution.

- The results using Yates correction at times provide p-values close to the Fisher's Exact test but it is not the case always.

Importantly, the approximation using Pearson's χ^2 was necessary for the comparative simplicity of the calculations because the exact analysis of 2 X 2 contingency tables with the limited computing power available at the time was prohibitive in many cases. This is no longer the case. Yates's correction is not necessary anymore since current software makes Fisher's exact test computationally feasible. In view of this it is recommended to use Fisher's Exact test for 2 X 2 contingency tables involving small numbers.

References

1. 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.

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

3. Yates, F. (1934) Contingency tables involving small numbers and the χ2 test. Journal of the Royal Statistical Society, Suppl.1, 217–235.

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

5. Kirkwood BR, Sterne JAC. Medical Statistics, 2nd edition. Oxford: Blackwell Science Ltd; 2003