Sociedad de Estadística e Investigación Operativa
Test (2005) *Vol.* **14**, *No.* 1, pp. 99–128

On the Validity Condition of the Chi-Squared Test in 2×2 Tables

A. Martín Andrés*
Bioestadística, Facultad de Medicina
Universidad de Granada, Spain

M. J. Sánchez Quevedo
Estadística e I.O.
Universidad de Cádiz, Spain

J. M. Tapia García and A. Silva–Mato
Bioestadística, Facultad de Medicina
Universidad de Granada, Spain

Abstract

A 2×2 contingency table is usually analyzed by using the chi-squared asymptotic test, with Yates' continuity correction ($c = n/2$, where n is the total size of sample). This correction is the correct one when the chi-squared test is an approximation to Fisher's exact (conditional) test. When the chi-squared test is used as an approximation to Barnard's exact (unconditional) test for comparing two independent proportions (two samples of size n_1 and n_2), or for contrasting independence (one sample with a size of n), the correction c is different ($c = 1$ if $n_1 \neq n_2$ or $c = 2$ if $n_1 = n_2$ in the first case; $c = 0.5$ in the second). Whatever the case, it is traditional to affirm that the asymptotic test is valid when $E > 5$, where E is the minimum expected quantity. Today it is recognized that this condition is too general and may not be appropriate. In the case of Yates' correction, Martín Andrés and Herranz Tejedor (2000) proved that the validity condition must be of the type $E > E^\star$ –where $E^\star$ is a known function depending on the marginals of the table– and that checking the validity of the asymptotic test is equivalent to checking the asymmetry of the base statistic (a hypergeometric random variable). In the present article the authors prove that this argument is valid for the other two continuity corrections, and moreover, that the value $E^\star$ is obtained for all three cases. Given that the function $E^\star$ reaches an absolute maximum, it can be affirmed that the three chi-squared tests referred to are valid when $E > 19.2, 14.9$ or 6.2 (or $E > 8.1, 7.7$ or 3.9 if $n \leq 500$) respectively for the three previous models (although for the first model and the two-tailed test $E > 0$ is sufficient).

*Correspondence to: A. Martín Andrés. Bioestadística, Facultad de Medicina, Universidad de Granada, Spain. E-mail: amartina@ugr.es

This research was supported by the Dirección General de Investigación (Spain), Grant BFM2000-1472.

Received: April 2003; Accepted: November 2003

Key Words: Asymmetry, Barnard's exact test, chi-squared test, conditional test, continuity correction, Fisher's exact test, minimum expected quantity, unconditional test, validity conditions, 2×2 tables.

AMS subject classification: 62F03, 62F04.

1 Introduction

For a 2×2 table it is usual to refer to a presentation of the data as in Table 1. The aim of the experiment is to contrast the null hypothesis H: "the characteristics A and B are independent" with an alternative hypothesis K (one–tailed or two–tailed). But the data may have been obtained from three different types of sampling, and this produces three different statistical models and three different p-values:

i) Under Model III, the values a_i and n_i are previously fixed, the only random variable (r.v. in the following) for the problem is x_1 (for example) and its distribution under H is the hyper-geometric $P(x_1) = C(n_1, x_1) \times C(n_2, x_2)/C(n, a_1)$. Therefore, the exact p-value will be

$$P_{III}(x_1) = \sum_{T(i) \geq T(x_1)} P(i),$$

where T is a direction-sensitive statistic. This yields the well-known Fisher's exact conditional test (Fisher, 1935).

ii) Under Model II, values n_i are previously fixed, the r.v.'s for the problem are x_1 and x_2 (for example) –two binomial independent r.v.'s with parameters p_1 and p_2 – and its joint distribution under H ($p_1 = p_2 = p$) is $P(x_1, x_2) = C(n_1, x_1) \times C(n_2, x_2) p^{a_1} (1-p)^{a_2}$, where p is a nuisance parameter. Therefore, the exact p-value will be

$$P_{II}(x_1, x_2) = \max_{0<p<1} \left\{ \sum_{T(i,j) \geq T(x_1, x_2)} P(i, j) \right\},$$

where T is a direction-sensitive statistic. This gives rise to Barnard's exact unconditional test (Barnard, 1947).

iii) Under Model I, only value n is previously fixed, the r.v.'s for the problem are x_1, y_1 and x_2 (for example) –a multinomial distribution with parameters p_{11}, p_{12}, p_{21} and p_{22} – and its distribution under H ($p_{ij} = p_{i\bullet} \times p_{\bullet j}$) will be $P(x_1, y_1, x_2) = C(n; x_1, x_2, y_1, y_2) p_{\bullet 1}^{a_1} (1 - p_{\bullet 1})^{a_2} p_{1\bullet}^{n_1} (1 - p_{1\bullet})^{n_2}$, where $p_{\bullet 1}$ and $p_{1\bullet}$ are two nuisance parameters. Therefore, the exact p-value will be

$$P_I(x_1, y_1, x_2) = \max_{0<p_{\bullet 1}, p_{1\bullet}<1} \left\{ \sum_{T(i,j,k) \geq T(x_1,y_1,x_2)} P(i,j,k) \right\},$$

where T is a direction-sensitive statistic. This gives rise to Barnard's exact unconditional test (Barnard, 1947).

Table 1: Presentation of results in the form of a 2×2 table

		Characteristic A		
		YES	NO	
Characteristic B	YES	x_1	y_1	n_1
	NO	x_2	y_2	n_2
		a_1	a_2	n

For statisticians who favor conditioning, the three models are all solved the same way: value $P_{III}(x_1)$. For those statisticians who do not use conditioning, each model has its own solution (as indicated above). It is not for us to argue here the merits of one solution or the other (see Yates' discussion (Yates, 1984), and Martín's review (Martín Andrés, 1991)), but in the following we shall adopt the unconditional approach.

The exact p-values P_{III}, P_{II} and P_I (where $P_{III} > P_{II} > P_I$ generally) are progressively more difficult to calculate (in that order). In particular, it is at present not possible to determine the last two (especially P_I) when the marginals are moderately large. The value of P_{III} is offered by many statistical programs; the value of P_{II} by StatXact; free programs for obtaining values P_I and P_{II} may be obtained at the web page `http://www.ugr.es/~bioest/software.htm`.

In any event it is customary to solve the problem in an approximate fashion –whether for pedagogic reasons, for convenience or because of the impossibility of calculating it– by using the chi-squared test (Greenwood and Nikulin, 1996) with the appropriate continuity correction (c.c. in the following). But, because the chi-squared test is an asymptotical test, it will be subject to certain validity conditions (v.c. in the following). It is usual to require that the minimum expected quantity $E = \min(a_1, a_2) \times \min(n_1, n_2)/n$ be sufficiently large ($E > E^\star$), where $E^\star$ is an fixed number yet to be determined. Determining the value of $E^\star$ has, generally speaking, not been the result of exhaustive study, but rather the result of an opinion of the various authors based on some partial results. For example, the value of $E^\star$ would be: *i)* 5 for Fisher (1925), and 5 or 2 for Cochran (1954) in Model III; *ii)* 5 for Grizzle (1967), 2 for Sachs (1986), and 2 for Scott (1999) in Model II; and *iii)* 1 or 2 for Camilli and Hopkins (1978) and Camilli and Hopkins (1979), small (without being specified) for Larntz (1978) and Richardson (1990), and 2 for Scott (1999) in Model I. However, Martín Andrés and Herranz Tejedor (1997), Martín Andrés and Silva Mato (1996) and Martín Andrés and Tapia García (2004), for Models III, II and I respectively, found that the value $E^\star$ is a function of the marginals of the table, and that the condition $E > 5$ may be very liberal or very conservative. Finally, and for the case of Model III, Martín Andrés and Herranz Tejedor (2000) found that affirming that $E > E^\star$ is equivalent to affirming that the asymmetry of the hypergeometric r.v. x_1 is bounded, and this allowed them to determine the form of the function $E^\star$. The aim of this study is to repeat their theoretical-practical analysis for the case of Models I and II, and also modify the results already known for the case of Model III.

2 Criteria for determining the experimental values of $\mathbf{E}^\star$

2.1 Validity of the asymptotic test

Given an experimental table, this will yield an exact p–value (P_E) and an asymptotic p–value (P_A). The aim is that $P_A \simeq P_E$, and so some discrepancy between the two will have to be allowed. Following Cochran (1954) we can say that the asymptotic test is valid ("does not fail") when

$$|P_A - P_E| \leq \delta P_E \tag{2.1}$$

where $\delta = 0.2$ for $P_E = 5\%$ (implying that, for a table where $P_E = 5\%$, the asymptotic test is valid if $4\% \leq P_A \leq 6\%$) and $\delta = 0.5$ for $P_E = 1\%$.

Martín Andrés and Herranz Tejedor (2000) adopted the same values for δ and interpolated them in the case of $1\% < P_E < 5\%$. However, these values for δ yield some excessively large values for $E^\star$ in the case of MODEL III, occasioned by the criterion $4\% \leq P_A \leq 6\%$ being excessively strict. (In the case of MODEL II the values for $E^\star$ are even larger: the data may be requested from the authors).

Alternatively, the criterion may be modified by accepting that a P_A which verifies $3\% \leq P_A \leq 7\%$ is an acceptable estimation of $P_E = 5\%$ (yielding a $\delta = 0.4$). Although the aim of this study is to obtain values of $E^\star$ for $P_E = 5\%$ (because $E^\star$ also varies with the target error), it is difficult to obtained tables with this exact value of P_E. This means having to define the values δ for the other cases, although here one must be less strict (because they are not the object of this study). This is why, in the following, the specified values δ are considered in Table 5 (which serves as a summary of this article). It is also why, in the following, MODEL III is analyzed again (although this was already done, with other δ's, by Martín Andrés and Herranz Tejedor (2000)).

2.2 The chi–squared test to be used

In expression (2.1) reference is made to the asymptotic p–value P_A obtained using the chi-squared statistic. The uncorrected version of this is $\chi^2_U = T(x_1y_2 - x_2y_1)^2/\{a_1a_2n_1n_2\}$, where $T = n$ or $T = n - 1$ according to whichever author is using it. However, the suitability of performing a c.c. is well-known, though it is less well-known that the c.c. in question depends on whichever model is assumed. Thus:

i) In the case of MODEL III, the traditional c.c. is Yates', and this yields the statistic χ^2_Y in Table 5. In the case of the one-tailed test, Martín Andrés and Herranz Tejedor (1997) proved that better results are obtained by using Conover's c.c. (Conover, 1974), which yields the statistic χ^2_{III} in Table 5. Hence, the asymptotic p-value P_A will be $P_{AIII} = P(\chi_{III})$, where $P(\chi) = 1 - F(\chi)$ and $F(\bullet)$ is the typical normal distribution function.

For the two-tailed test, Martín Andrés et al. (1992) proved that the optimal test is χ^2_Y taking the precaution suggested by Mantel (1974): the two-tailed p-value is the sum of the p-values for each tail. This means calculating the table x'_1 for the other tail giving a value $\chi^2_{Y'}$ that is as extreme or more than χ^2_Y, that is:

$$x'_1 = [2E_{11} - x_1], \text{ where } E_{11} = \frac{a_1 n_1}{n} \tag{2.2}$$

in which $[x]$ refers to the rounding of x in the sense of moving away from the value of E_{11}. So, the asymptotic p–value P_A with two tails will be $P_{AIII} = P(\chi_Y) + P(\chi_{Y'})$, as shown in Table 5.

ii) In the case of Model II, Martín Andrés et al. (2002) proved that the most appropriate chi-squared statistic is χ^2_{II} in Table 5; the asymptotic p–values are $P_{AII} = P(\chi_{II})$ for one–tailed test and $2P(\chi_{II})$ for two–tailed test.

iii) Finally, in the case of Model I the most suitable chi-squared statistic (Martín Andrés and Tapia García, 2004) is a modification of Pirie and Hamdan (1972): χ^2_I in Table 5. The p–values will now be $P_{AI} = P(\chi_I)$ or $2P(\chi_I)$ for one- or two–tailed tests respectively.

The asymptotic tests described are optimal for errors α between 1% and 10%, which are the most common errors and those to which we shall refer exclusively in the following.

2.3 The exact test to be used

In expression (2.1) reference is made to the exact p–value P_E: this will be P_{III}, P_{II} or P_I (see Introduction) in terms of the model used. Unfortunately, the value P_E in each model is not unique, but depends on the ordering criterion which defines the statistic T. It seems logical, in each model, to choose statistic T which yields a test which is generally more powerful (because in this type of problem there is no UMP test), and so:

i) In the case of Model III choose $T(x_1) = \chi^2_Y$, because Martín Andrés and Herranz Tejedor (1995) proved that it was one of the most powerful criteria and moreover, the ordering is related to that of the asymptotic statistic. Now, as below, for T to be direction–sensitive it is necessary to distinguish the cases $x_1 > E_{11}$ and $x_1 < E_{11}$. For the present Fisher's exact test, this implies that $P_{III}(x_1) = \sum_{i \geq x_1} P(i)$ for the one-tailed test (positive association), and $P_{III}(x_1) = \sum_{i \geq x_1} P(i) + \sum_{i \leq x'_1} P(i)$ for the two–tailed test (if $x_1 > E_{11}$).

ii) In the case of Model II choose Barnard's order T (1947) because of the results in Martín Andrés et al. (1998). It is complicated to describe and, as it is not an objective in the present case, we refer the reader to the article mentioned.

iii) In the case of Model I, the optimal order T is also Barnard's Barnard's (1947), but the time needed to calculate it prevents it being used for our purposes here. The order based on the statistic χ^2_Y –Shuster (1992)– is more accessible, and yields the generally most powerful of all the tests based on ordering criteria of the chi–squared type (Martín Andrés and Tapia García, 1999).

A summary of the foregoing is also given in Table 5.

2.4 Generation of the sample space

In order to identify which cell the quantity E corresponds to, let us agree that the sample will be reordered so that E is the expected quantity for the cell $x_1 : E = a_1 n_1/n$. Therefore, for fixed values of (n_1, n), all the possible tables are generated in the following way:

i) In the case of Model III, if it is agreed that $a_1 = \min(a_1, a_2, n_1, n_2)$ and $n_1 = \min(n_1, n_2)$, then (Martín Andrés and Herranz Tejedor, 2000): $0 \leq a_1 \leq n_1, 0 \leq x_1 \leq a_1$.

ii) In the case of Model II, if it is agreed that $a_1 = \min(a_1, a_2)$ and $n_1 = \min(n_1, n_2)$, then: $0 \leq a_1 \leq [n/2]^-$, $0 \leq x_1 \leq \min(a_1, n_1)$, where $[x]^-$ refers to the rounding of x.

iii) In the case of Model I act in the same way as for Model III.

2.5 Determining the experimental value of E*

The desired v.c. is of the type $E > E^\star$, that is, and given that $E = a_1 n_1/n$:

$$a_1 > \frac{E^\star n}{n_1} = E^\star(K+1) = a_1^\star, \text{ where } K = \frac{n_2}{n_1} \geq 1 \tag{2.3}$$

Therefore, for each pair (n_1, n) –or equivalent (n_1, K)– the aim is to determine a value $a_1^\star$ such that for $a_1 > a_1^\star$ the asymptotic test will always be valid (that is, it verifies expression (2.1)). Once the value is obtained $a_1^\star$, then $E^\star = a_1^\star/(K+1)$, and so the classic v.c. will be:

$$E > E^\star = \frac{a_1^\star}{K+1} \tag{2.4}$$

In order to determine $a_1^\star$ proceed in the following way. Once the pair (n_1, K) is fixed:

1. Fix the value of a_1, beginning at its maximum value possible (n_1 in Models I and III; $[n/2]^-$ in Model II).

2. For the triplet (a_1, n_1, n), obtain all the tables from $x_1 = 0$ to x_1 = "the maximum possible" (a_1 in Models I and III; $\min(a_1, n_1)$ in Model II), calculate the value of P_A those tables and retain only those where $1\% \leq P_A \leq 10\%$.

3. Calculate the value P_E for those tables and check whether expression (2.4) is verified. If (2.4) is always verified, then the test is valid for that value of a_1, so return a_1 to step **2** after reducing a_1 by one unit. If not, the asymptotic test will fail in a table containing the present value a_1, which is the value $a_1^\star$ wanted (the a_1 of the first failure).

The above procedure requires the values of interest –(n_1, K)– to be fixed. These values can be large in the case of Model III (where the computational intensity is less), but they should be smaller in the case of Model II, and even more so in Model I. Tables 2, 3 and 4 present the results obtained for each case. The data show that: **a)** For constant n_1 (K), the values of $E^\star$ increase generally with K (n_1); **b)** For $K = 1$, $E^\star$ is constant in Models II and III, but increases in Model I; **c)** Compared to the v.c.'s for the two–tailed test, the v.c.'s for the one-tailed test are very similar in Model I, weaker in Model II and stronger in Model III; **d)** Something similar to what we have described above occurs with the values of $a_1^\star$, but one must bear in mind that $a_1^\star$ cannot grow indefinitely while K increases (because of, necessity, $a_1^\star \leq n_1$ or $[n/2]^-$); **e)** It must be emphasized that, for the selected values of δ, Fisher's two–tailed exact test is always valid.

Table 2: Real minimum values (a_1^* and E^*) for the smallest marginal (a_1) and for the minimum expected quantity (E) for which the χ^2_{III} test is valid (as an approximation to Fisher's exact test) and conservative predictions for the value E^*. It is assumed that $K = n_2/n_1 \geq 1$

		One Tail					**One Tail**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	n_1	K	a_1^*	E^*	$\hat{E}^*$
40	1.0	2	1.0	2.5	150	1.0	2	1.0	3.0
40	1.1	2	1.0	2.4	150	1.1	2	1.0	2.9
40	1.2	2	0.9	2.3	150	1.2	2	0.9	2.7
40	1.3	2	0.9	2.6	150	1.3	2	0.9	3.0
40	1.4	2	0.8	2.9	150	1.4	2	0.8	3.3
40	1.5	2	0.8	2.8	150	1.5	2	0.8	3.6
40	1.6	2	0.8	3.1	150	1.6	2	0.8	3.8
40	1.7	2	0.7	3.3	150	1.7	2	0.7	4.1
40	1.8	2	0.7	3.6	150	1.8	2	0.7	4.3
40	1.9	2	0.7	3.4	150	1.9	12	4.1	4.5
40	2.0	2	0.7	3.7	150	2.0	13	4.3	4.7
40	2.5	15	4.3	4.6	150	2.5	16	4.6	6.0
40	2.6	15	4.2	4.7	150	2.6	16	4.4	6.1
40	2.7	16	4.3	4.6	150	2.7	17	4.6	6.2
40	2.8	16	4.2	4.7	150	2.8	18	4.7	6.6
40	2.9	17	4.4	4.9	150	2.9	25	6.4	6.7
40	3.0	17	4.3	5.0	150	3.0	19	4.8	6.8
40	3.5	20	4.4	5.6	150	3.5	29	6.4	7.6
40	4.0	23	4.6	5.8	150	4.0	33	6.6	8.2
40	5.0	38	6.3	6.3	150	5.0	41	6.8	9.2
60	1.0	2	1.0	2.5	200	1.0	2	1.0	3.0
60	1.1	2	1.0	2.4	200	1.1	2	1.0	2.9
60	1.2	2	0.9	2.7	200	1.2	2	0.9	2.7
60	1.3	2	0.9	2.6	200	1.3	2	0.9	3.0
60	1.4	2	0.8	2.9	200	1.4	2	0.8	3.3
60	1.5	2	0.8	3.2	200	1.5	2	0.8	3.6
60	1.6	2	0.8	3.5	200	1.6	2	0.8	3.8
60	1.7	2	0.7	3.7	200	1.7	2	0.7	4.1
60	1.8	2	0.7	3.6	200	1.8	12	4.3	4.3
60	1.9	12	4.1	3.8	200	1.9	12	4.1	4.5
60	2.0	2	0.7	4.0	200	2.0	13	4.3	4.7
60	2.5	15	4.3	5.1	200	2.5	16	4.6	6.0
60	2.6	16	4.4	5.3	200	2.6	23	6.4	6.4
60	2.7	16	4.3	5.4	200	2.7	17	4.6	6.5
60	2.8	17	4.5	5.5	200	2.8	18	4.7	6.6
60	2.9	17	4.4	5.6	200	2.9	25	6.4	6.9
60	3.0	18	4.5	5.8	200	3.0	26	6.5	7.0
60	3.5	21	4.7	6.2	200	3.5	30	6.7	7.8
60	4.0	32	6.4	6.6	200	4.0	34	6.8	8.6

(*Continued on next page*)

(*Table 2. Continued from previous page*)

		One Tail					**One Tail**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	n_1	K	a_1^*	E^*	$\hat{E}^*$
60	5.0	39	6.5	7.3	200	5.0	51	8.5	9.7
100	1.0	2	1.0	2.5	250	1.0	2	1.0	3.0
100	1.1	2	1.0	2.9	250	1.1	2	1.0	2.9
100	1.2	2	0.9	2.7	250	1.2	2	0.9	2.7
100	1.3	2	0.9	3.0	250	1.3	2	0.9	3.0
100	1.4	2	0.8	2.9	250	1.4	2	0.8	3.3
100	1.5	2	0.8	3.2	250	1.5	2	0.8	3.6
100	1.6	2	0.8	3.5	250	1.6	2	0.8	3.8
100	1.7	2	0.7	3.7	250	1.7	2	0.7	4.1
100	1.8	2	0.7	3.9	250	1.8	12	4.3	4.3
100	1.9	12	4.1	4.1	250	1.9	12	4.1	4.5
100	2.0	13	4.3	4.3	250	2.0	13	4.3	5.0
100	2.5	16	4.6	5.4	250	2.5	16	4.6	6.0
100	2.6	16	4.4	5.8	250	2.6	23	6.4	6.4
100	2.7	17	4.6	5.9	250	2.7	17	4.6	6.5
100	2.8	17	4.5	6.1	250	2.8	18	4.7	6.8
100	2.9	18	4.6	6.2	250	2.9	25	6.4	6.9
100	3.0	19	4.8	6.5	250	3.0	26	6.5	7.3
100	3.5	29	6.4	7.1	250	3.5	30	6.7	8.0
100	4.0	33	6.6	7.6	250	4.0	42	8.4	8.8
100	5.0	41	6.8	8.5	250	5.0	51	8.5	9.8

Note: The two-tailed test is always valid.

Table 3: Real minimum values (a_1^* and E^*) for the smallest marginal (a_1) and for the minimum expected quantity (E) from which the χ^2_{II} test is valid (as an approximation to Barnard's exact test in Model II) and conservatives predictions ($\hat{E}^*$) for value E^*. It is assumed that $K = n_2/n_1 \geq 1$

		One Tail			**Two Tails**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
20	1.0	1	0.5	2.5	1	0.5	3.5
20	1.1	2	1.0	2.4	1	0.5	3.3
20	1.2	1	0.5	2.7	5	2.3	3.6
20	1.3	1	0.4	2.6	1	0.4	3.5
20	1.4	4	1.7	2.5	1	0.4	3.3
20	1.5	4	1.6	2.8	1	0.4	3.6
20	1.6	6	2.3	2.7	1	0.4	3.5
20	1.7	4	1.5	2.6	1	0.4	3.3
20	1.8	4	1.4	2.9	1	0.4	3.6
20	1.9	4	1.4	2.8	1	0.3	3.4
20	2.0	4	1.3	3.0	11	3.7	3.7

(*Continued on next page*)

(*Table 3. Continued from previous page*)

		One Tail			Two Tails		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
20	2.5	11	3.1	3.1	1	0.3	3.7
20	2.6	1	0.3	3.3	13	3.6	3.9
20	2.7	12	3.2	3.2	14	3.8	3.8
20	2.8	3	0.8	3.4	7	1.8	3.9
20	2.9	13	3.3	3.3	1	0.3	3.8
20	3.0	13	3.3	3.5	16	4.0	4.0
20	3.5	15	3.3	3.8	17	3.8	4.0
20	4.0	18	3.6	3.8	21	4.2	4.2
20	5.0	22	3.7	4.0	26	4.3	4.3
30	1.0	1	0.5	3.0	1	0.5	4.5
30	1.1	2	1.0	2.9	1	0.5	4.3
30	1.2	7	3.2	3.2	5	2.3	4.1
30	1.3	2	0.9	3.0	1	0.4	4.3
30	1.4	4	1.7	2.9	1	0.4	4.2
30	1.5	4	1.6	3.2	1	0.4	4.0
30	1.6	4	1.5	3.1	1	0.4	4.2
30	1.7	4	1.5	3.3	1	0.4	4.1
30	1.8	4	1.4	3.2	1	0.4	4.3
30	1.9	4	1.4	3.4	1	0.3	4.1
30	2.0	5	1.7	3.7	11	3.7	4.3
30	2.5	12	3.4	4.0	1	0.3	4.6
30	2.6	12	3.3	3.9	15	4.2	4.7
30	2.7	13	3.5	4.1	14	3.8	4.6
30	2.8	13	3.4	4.2	16	4.2	4.7
30	2.9	14	3.6	4.1	16	4.1	4.6
30	3.0	14	3.5	4.3	17	4.3	4.8
30	3.5	16	3.6	4.4	6	1.3	4.9
30	4.0	19	3.8	4.6	22	4.4	5.0
30	5.0	23	3.8	5.0	28	4.7	5.3
40	1.0	2	1.0	3.5	1	0.5	5.0
40	1.1	2	1.0	3.3	1	0.5	4.8
40	1.2	2	0.9	3.2	1	0.5	4.5
40	1.3	7	3.0	3.0	4	1.7	4.8
40	1.4	4	1.7	3.3	1	0.4	4.6
40	1.5	4	1.6	3.2	1	0.4	4.8
40	1.6	8	3.1	3.5	1	0.4	4.6
40	1.7	4	1.5	3.7	1	0.4	4.8
40	1.8	9	3.2	3.6	1	0.4	4.6
40	1.9	10	3.4	3.8	1	0.3	4.8
40	2.0	10	3.3	4.0	11	3.7	5.0
40	2.5	12	3.4	4.3	1	0.3	5.1
40	2.6	12	3.3	4.4	15	4.2	5.3
40	2.7	13	3.5	4.6	15	4.1	5.1

(*Continued on next page*)

(*Table 3. Continued from previous page*)

		One Tail			**Two Tails**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
40	2.8	13	3.4	4.5	16	4.2	5.3
40	2.9	14	3.6	4.6	17	4.4	5.4
40	3.0	14	3.5	4.8	17	4.3	5.3
40	3.5	17	3.8	4.9	20	4.4	5.6
40	4.0	19	3.8	5.2	23	4.6	5.8
40	5.0	34	5.7	5.7	29	4.8	6.0
50	1.0	2	1.0	3.5	1	0.5	5.5
50	1.1	2	1.0	3.3	1	0.5	5.2
50	1.2	2	0.9	3.2	1	0.5	5.0
50	1.3	1	0.4	3.5	1	0.4	5.2
50	1.4	2	0.8	3.3	5	2.1	5.0
50	1.5	6	2.4	3.6	1	0.4	5.2
50	1.6	8	3.1	3.8	1	0.4	5.0
50	1.7	4	1.5	3.7	1	0.4	5.2
50	1.8	9	3.2	3.9	1	0.4	5.0
50	1.9	12	4.1	4.1	1	0.3	5.2
50	2.0	10	3.3	4.0	1	0.3	5.3
50	2.5	12	3.4	4.6	15	4.3	5.4
50	2.6	13	3.6	4.7	1	0.3	5.6
50	2.7	13	3.5	4.9	15	4.1	5.7
50	2.8	14	3.7	4.7	17	4.5	5.8
50	2.9	14	3.6	4.9	17	4.4	5.6
50	3.0	15	3.8	5.0	18	4.5	5.8
50	3.5	17	3.8	5.3	21	4.7	6.0
50	4.0	28	5.6	5.6	24	4.8	6.2
50	5.0	34	5.7	6.0	29	4.8	6.5
60	1.0	2	1.0	3.5	1	0.5	5.5
60	1.1	2	1.0	3.3	1	0.5	5.2
60	1.2	2	0.9	3.6	1	0.5	5.5
60	1.3	2	0.9	3.5	1	0.4	5.2
60	1.4	4	1.7	3.8	1	0.4	5.4
60	1.5	4	1.6	3.6	1	0.4	5.2
60	1.6	4	1.5	3.8	10	3.8	5.4
60	1.7	4	1.5	4.1	10	3.7	5.2
60	1.8	6	2.1	3.9	1	0.4	5.4
60	1.9	4	1.4	4.1	1	0.3	5.5
60	2.0	10	3.3	4.3	1	0.3	5.7
60	2.5	12	3.4	4.9	15	4.3	6.0
60	2.6	13	3.6	5.0	16	4.4	5.8
60	2.7	13	3.5	5.1	16	4.3	5.9
60	2.8	14	3.7	5.0	17	4.5	6.1
60	2.9	14	3.6	5.1	17	4.4	6.2
60	3.0	15	3.8	5.3	18	4.5	6.3

(*Continued on next page*)

(*Table 3. Continued from previous page*)

		One Tail			Two Tails		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
60	3.5	17	3.8	5.6	21	4.7	6.4
60	4.0	28	5.6	6.0	24	4.8	6.6
60	5.0	35	5.8	6.5	40	6.7	7.0
70	1.0	2	1.0	4.0	1	0.5	6.0
70	1.1	2	1.0	3.8	1	0.5	5.7
70	1.2	2	0.9	3.6	5	2.3	5.5
70	1.3	7	3.0	3.5	1	0.4	5.7
70	1.4	4	1.7	3.8	1	0.4	5.4
70	1.5	4	1.6	4.0	1	0.4	5.6
70	1.6	4	1.5	3.8	10	3.8	5.4
70	1.7	4	1.5	4.1	1	0.4	5.6
70	1.8	10	3.6	4.3	11	3.9	5.7
70	1.9	4	1.4	4.1	1	0.3	5.9
70	2.0	10	3.3	4.3	1	0.3	5.7
70	2.5	12	3.4	5.1	15	4.3	6.0
70	2.6	13	3.6	5.0	16	4.4	6.1
70	2.7	13	3.5	5.1	16	4.3	6.2
70	2.8	14	3.7	5.3	17	4.5	6.3
70	2.9	14	3.6	5.4	18	4.6	6.4
70	3.0	15	3.8	5.5	18	4.5	6.5
70	3.5	17	3.8	6.0	21	4.7	6.7
70	4.0	29	5.8	6.2	32	6.4	7.0
70	5.0	35	5.8	6.7	41	6.8	7.3
80	1.0	2	1.0	4.0	1	0.5	6.0
80	1.1	2	1.0	3.8	1	0.5	5.7
80	1.2	2	0.9	3.6	5	2.3	5.9
80	1.3	2	0.9	3.9	4	1.7	5.7
80	1.4	4	1.7	3.8	1	0.4	5.8
80	1.5	6	2.4	4.0	1	0.4	5.6
80	1.6	4	1.5	3.8	10	3.8	5.8
80	1.7	4	1.5	4.1	1	0.4	5.9
80	1.8	6	2.1	4.3	11	3.9	5.7
80	1.9	4	1.4	4.5	1	0.3	5.9
80	2.0	10	3.3	4.7	1	0.3	6.0
80	2.5	12	3.4	5.1	15	4.3	6.3
80	2.6	13	3.6	5.3	16	4.4	6.4
80	2.7	13	3.5	5.4	16	4.3	6.5
80	2.8	14	3.7	5.5	17	4.5	6.6
80	2.9	15	3.8	5.6	18	4.6	6.7
80	3.0	15	3.8	5.8	18	4.5	6.8
80	3.5	18	4.0	6.0	21	4.7	6.9
80	4.0	29	5.8	6.4	33	6.6	7.2
80	5.0	36	6.0	7.0	40	6.7	7.7

(*Continued on next page*)

(*Table 3. Continued from previous page*)

		One Tail			**Two Tails**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
90	1.0	2	1.0	4.0	1	0.5	6.0
90	1.1	2	1.0	3.8	1	0.5	6.2
90	1.2	2	0.9	3.6	5	2.3	5.9
90	1.3	2	0.9	3.9	4	1.7	5.7
90	1.4	4	1.7	3.8	1	0.4	5.8
90	1.5	6	2.4	4.0	1	0.4	6.0
90	1.6	6	2.3	4.2	10	3.8	5.8
90	1.7	4	1.5	4.1	1	0.4	5.9
90	1.8	6	2.1	4.3	11	3.9	6.1
90	1.9	9	3.1	4.5	1	0.3	6.2
90	2.0	10	3.3	4.7	1	0.3	6.0
90	2.5	13	3.7	5.1	15	4.3	6.6
90	2.6	13	3.6	5.3	16	4.4	6.7
90	2.7	14	3.8	5.4	17	4.6	6.8
90	2.8	14	3.7	5.5	17	4.5	6.8
90	2.9	15	3.8	5.6	18	4.6	6.7
90	3.0	15	3.8	5.8	18	4.5	6.8
90	3.5	26	5.8	6.2	22	4.9	7.1
90	4.0	29	5.8	6.6	33	6.6	7.4
90	5.0	36	6.0	7.2	41	6.8	7.8
100	1.0	2	1.0	4.0	1	0.5	6.5
100	1.1	2	1.0	3.8	1	0.5	6.2
100	1.2	2	0.9	3.6	1	0.5	5.9
100	1.3	2	0.9	3.9	4	1.7	6.1
100	1.4	4	1.7	3.8	1	0.4	5.8
100	1.5	6	2.4	4.0	1	0.4	6.0
100	1.6	4	1.5	4.2	10	3.8	6.2
100	1.7	4	1.5	4.4	1	0.4	5.9
100	1.8	4	1.4	4.3	5	1.8	6.1
100	1.9	6	2.1	4.5	1	0.3	6.2
100	2.0	10	3.3	4.7	1	0.3	6.3
100	2.5	13	3.7	5.4	15	4.3	6.6
100	2.6	13	3.6	5.6	16	4.4	6.7
100	2.7	14	3.8	5.7	17	4.6	6.8
100	2.8	14	3.7	5.8	17	4.5	6.8
100	2.9	15	3.8	5.9	18	4.6	6.9
100	3.0	15	3.8	6.0	19	4.8	7.0
100	3.5	26	5.8	6.4	30	6.7	7.3
100	4.0	29	5.8	6.8	34	6.8	7.6
100	5.0	36	6.0	7.3	42	7.0	8.0

Table 4: Real minimum values (a_1^* and E^*) for the smallest marginal (a_1) and for the minimum expected quantity (E) for which the χ_I^2 test is valid (as an approximation for the exact test of MODEL I) and conservative predictions ($\hat{E}^*$) for value E^*. It is assumed that $K = n_2/n_1 \geq 1$

		One Tail			**Two Tails**		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
10	1.0	2	1.0	1.5	2	1.0	1.5
10	1.1	2	1.0	1.4	2	1.0	1.4
10	1.2	2	0.9	1.4	2	0.9	1.4
10	1.3	3	1.3	1.3	2	0.9	1.3
10	1.4	2	0.8	1.3	3	1.3	1.3
10	1.5	3	1.2	1.2	2	0.8	1.2
10	1.7	3	1.1	1.5	2	0.7	1.5
10	1.8	3	1.1	1.4	2	0.7	1.4
10	1.9	2	0.7	1.4	4	1.4	1.4
10	2.0	4	1.3	1.3	2	0.7	1.3
10	2.4	3	0.9	1.5	2	0.6	1.5
10	3.0	5	1.3	1.5	5	1.3	1.8
10	3.5	6	1.3	1.6	6	1.3	1.8
10	4.0	8	1.6	1.6	6	1.2	1.8
10	4.5	8	1.5	1.5	7	1.3	1.8
10	5.0	9	1.5	1.5	8	1.3	1.7
12	1.0	2	1.0	1.5	3	1.5	1.5
12	1.5	3	1.2	1.6	2	0.8	1.6
12	2.0	4	1.3	1.3	4	1.3	1.7
12	2.5	5	1.4	1.4	4	1.1	1.7
12	3.0	5	1.3	1.5	5	1.3	1.8
12	3.5	7	1.6	1.6	9	2.0	2.0
12	4.0	8	1.6	1.6	7	1.4	2.0
15	1.0	2	1.0	1.5	3	1.5	1.5
15	1.2	3	1.4	1.8	3	1.4	1.8
15	1.4	3	1.3	1.7	4	1.7	1.7
15	1.6	4	1.5	1.5	5	1.9	1.9
15	1.8	4	1.4	1.8	4	1.4	1.8
15	2.0	4	1.3	1.7	4	1.3	1.7
15	2.2	5	1.6	1.6	4	1.3	1.9
15	2.4	5	1.5	1.8	6	1.8	1.8
15	2.6	5	1.4	1.7	7	1.9	1.9
15	2.8	6	1.6	1.8	8	2.1	2.1
15	3.0	6	1.5	1.8	5	1.3	2.0
20	1.0	3	1.5	2.0	4	2.0	2.0
20	1.1	3	1.4	1.9	4	1.9	1.9
20	1.2	3	1.4	1.8	4	1.8	1.8
20	1.3	4	1.7	1.7	4	1.7	1.7
20	1.4	4	1.7	1.7	5	2.1	2.1

(*Continued on next page*)

(*Table 4. Continued from previous page*)

		One Tail			Two Tails		
n_1	K	a_1^*	E^*	$\hat{E}^*$	a_1^*	E^*	$\hat{E}^*$
20	1.5	4	1.6	2.0	4	1.6	2.0
20	1.7	4	1.5	1.9	5	1.9	1.9
20	1.8	5	1.8	1.8	5	1.8	2.1
20	1.9	5	1.7	1.7	6	2.1	2.1
20	2.0	5	1.7	2.0	6	2.0	2.0
25	1.0	3	1.5	2.0	4	2.0	2.0
25	1.2	5	2.3	2.3	4	1.8	2.3
25	1.4	4	1.7	2.1	5	2.1	2.1
30	1.0	5	2.5	2.5	4	2.0	2.5

As can be seen, the experimental data show that $E^\star$ is not a constant, but rather that it is a function of the number of the tails of the test, of n_1 and of K (and also of the target error α, but in this case we are restricting ourselves to the usual errors). What is this function?

3 Determining the function $\mathbf{E^*(n_1, K)}$: theoretical-experimental validity conditions

3.1 Condition of bounded asymmetry and form of the function $\mathbf{E^*}$

Martín Andrés and Herranz Tejedor (2000) argued that, as the chi-squared test is based on approximating a hypergeometric distribution (Model III) along a normal distribution, such a approximation will behave badly when the square of the asymmetry coefficient of the first (β_{III}) is excessively large, that is, when $\beta_{III} > B$, where $B > 0$ is a constant which remains to be determined. A similar argument can be applied to Models I and II. The value β_{III} is that of the hypergeometric distribution; the values of β_I and β_{II} are determined in Annexes I and II respectively. In the three cases the form of β is similar:

$$\beta = \frac{(n_2 - n_1)^2}{n_1 n_2} \times \frac{(a_2 - a_1)^2}{a_1 a_2} \times \frac{1}{C} \tag{3.1}$$

where

$$C = \begin{cases} (n-2)^2/(n-1) & \text{in MODEL III} \\ n^2/(n-1) & \text{in MODEL II} \\ n^4/(n-2)^2(n-1) & \text{in MODEL I} \end{cases} \tag{3.2}$$

and also:

$$\beta = \frac{f(a_1) \times f(n_1)}{C} \text{ where } f(x) = \frac{(n-2x)^2}{x(n-x)}$$

However Martín Andrés and Herranz Tejedor (2000) argued that the condition $\beta > B$ should somewhat be modified, since for $K = 1$, $\beta = 0 < B$ is true, which is contrary to the experimental fact that in $K = 1$ the test is not always valid. The proposal of these authors was that the condition should be:

$$\{f(n_1) + A\} \times f(a_1)/C \geq B, \tag{3.3}$$

where A is a constant which must be determined. This is equivalent to requiring that

$$y \leq a + bx, \text{ where } y = \frac{C}{f(a_1)}, \; x = f(n_1) = \frac{(K-1)^2}{K}, \; a = \frac{A}{B}, \; b = \frac{1}{B} \tag{3.4}$$

For each fixed value of x (that is, for each fixed value of K), equality $y = a+bx$ will be reached at a value $y^* = a+bx$ which will be $y^* = C/f(a_1^*)$. Thus one achieves the validity of the chi-squared test (from the perspective of β) when $y > y^*$ or, equivalently (and from the perspective of the previous section), when $a_1 > a_1^*$, so linking both criteria. By working out a_1^* for the equality $y^* = C/f(a_1^*)$, and substituting in expression (2.4), it can be deduced that the chi-squared test will be valid when:

$$E \; > \; E^*(n_1, K) = \frac{1}{K+1}\left[\frac{n}{2}\left\{1 - \sqrt{\frac{C}{C+4y^*}}\right\}\right]^- \tag{3.5}$$

and this determines the form of the function $E^*(n_1, K)$. Note that in practice $C \simeq n$, according to expression (3.2).

3.2 Estimation of the constants a and b. Specific validity condition

Martín Andrés and Herranz Tejedor (2000) argued that, in order to determine the values for a and b, one had to obtain a line $y^* = a + bx$ such that it leaves below all the points (x, y^*) obtained from Table 2, where $x = (K-1)^2 \big/ K$ and $y^* = C/f\,(a_1^*)$, as was stated previously. The same can be said with regard to Tables 3 and 4. Figures 1, 2 and 3 present the scatter plot and the selected line for the data in Tables 2, 3 and 4 respectively. The equations for the lines are included in Table 5. From these values of y^*, and expression (3.5), the predictions $\hat{E}^*$ which are included in Tables 2, 3 and 4 can be determined. Condition $E > \hat{E}^*$ is thus an estimated v.c. but, as it is conservative (because of the manner in which it was obtained), it can be assumed that it will be generally valid (for any value of n_1 and K). Martín Andrés and Herranz Tejedor (2000) found, with many examples, that this assumption is correct. From equations $y^* = a + bx$ for each model, and from expression (3.5), it can be observed that:

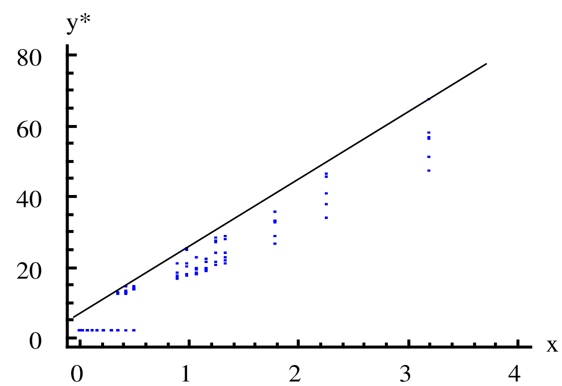

Figure 1: *Scatter plot for (x, y^*) of* Model *III (one tail) –Table 2– and the conservative line that fits them.*

1. In Model I (II) the initial v.c. (defined by constant a) for one and two-tailed tests are very similar (different) and, as K increases, the v.c. become more and more different (similar), as indicated by constant b.

2. The v.c.'s are stronger in Model II and in I, as the values of a and b in the first are greater than those of the second. In the one-tailed test, the v.c.'s are stronger in Models I and II than in Model III for small values of K (because $a_I, a_{II} > a_{III}$), but weaker for large values of K (because $b_I, b_{II} < b_{III}$).

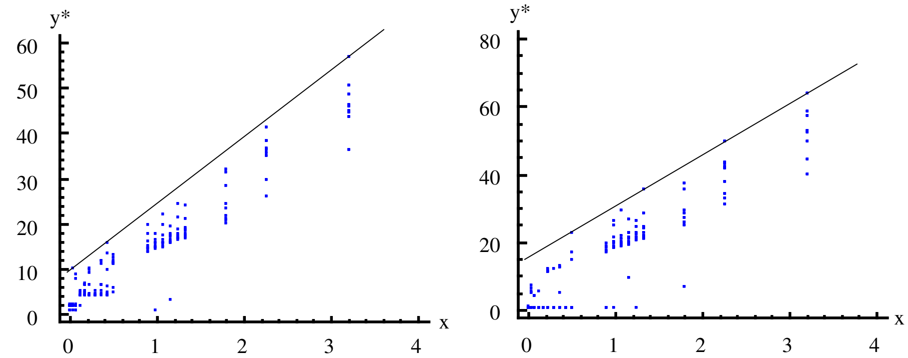

Figure 2: Scatter plot for (x, y^*) of Model *II –Table 3– and the conservative line that fits them. One tail (left) and two tails (right).*

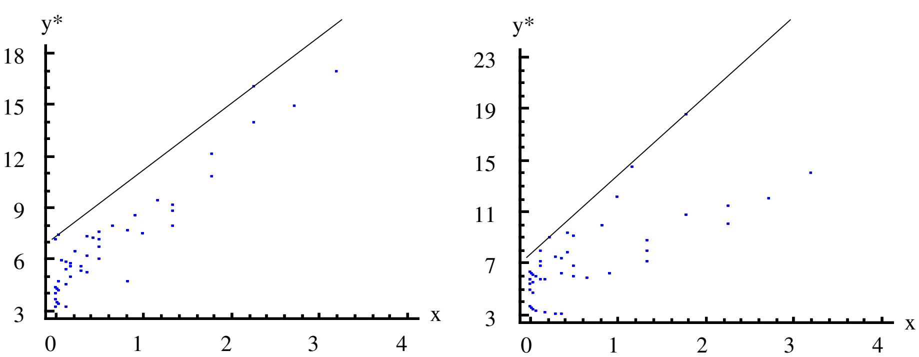

Figure 3: Scatter plot for (x, y^*) of Model *I –Table 4– and the conservative line that fits them. One tail (left) and two tails (right).*

For all the following, and where there is no danger of confusion, the values $\hat{E}^*$ will be noted by E^*.

Through (3.4) the values of a and b allow one to obtain the values of A and B, so that a v.c. equivalent to $E > E^*$ is the one which will be obtained by denying expression (3.3). By substituting values, this implies that the chi-squared test is valid when:

$$\frac{\left\{b(n_2 - n_1)^2 + an_1n_2\right\}(a_2 - a_1)^2}{a_1a_2n_1n_2C} < 1 \tag{3.6}$$

Consequently, given any experimental table, the chi-squared will be valid if expression (3.5) can be verified or, equivalently, if (3.6) can be verified. This v.c. should be properly understood. Because it proceeds from a conservative fit of the data, the statement "the chi-squared test is valid" is true, but the opposite statement is not (even though the v.c.'s are not verified, the chi-squared test can still be valid).

In Section 2.2 it is stated that the statistic χ^2_{III} performs better, as a one-tailed test, than the classic χ^2_Y. From the present perspective this can be checked by determining the line y^* for χ^2_Y. The data (which may be obtained on request from the authors) indicate that now $y^* = 7.785 + 20.627x$, where a and b are greater to those of the case χ^2_{III} (see Table 5).

3.3 More generic validity conditions

Martín Andrés and Herranz Tejedor (2000) studied the specific form of the function $E^*(n_1, K)$, but some of the demonstrations are based on the specific values of a and b they obtained for Model III and for their particular selection of δ. A generic study is given in Annex III. Because $E^*(n_1, K)$ increases in n_1 for each fixed value of K, its maximum will be reached at $E^*(K) = \lim_{n_1 \to \infty} E^*(n_1, K)$; hence the condition:

$$E > E^*(K) = \frac{\left[a + b\frac{(K-1)^2}{K}\right]^-}{K+1} \tag{3.7}$$

is a conservative v.c. and therefore valid.

It seems appropriate to determine a more general v.c. which implies a single value for E^* (which is why it will be a very conservative v.c.). In Annex III it is proved that $\max_{n_1,K} E^*(n_1, K) = b$ for the present values of a and b, which implies the v.c. modified in Table 5. Note the great disparity in terms of the model: in Model I the v.c.'s are very weak ($E > 6.2$ could be a general v.c.); in Model II they are very strong ($E > 14.9$ could be a general v.c.); in Model III they are extreme: very strong for a one-tailed test ($E > 19.2$ could be a general v.c.) or null for the two-tailed test ($E > 0$).

As Martín Andrés and Herranz Tejedor (2000) pointed out, there are other possible ways of providing v.c.:

(A) Expression (3.3) implies that the chi-squared test is valid when $\beta < B - Af(a_1)/C$; as this cannot occur if $B - Af(a_1)/C < 0$, then the chi-squared test will not be valid in such circumstances. Condition $B - Af(a_1)/C < 0$ occurs approximately when

$$a_1 \leq \frac{n}{2}\left(1 - \sqrt{\frac{n}{4a+n}}\right) \to a \text{ when } n \to \infty.$$

This implies (approximately) that the imbalanced tables where $a_1 \leq [a]^-$ can never be analyzed by chi-square.

(B) When all the marginals are balanced ($a_i = n_i = n/2$, then $f(a_1) = f(n_1) = 0$ and it is true that $\beta = 0 < B$: the chi-squared test is always valid.

(C) When $K = 1$, expression (3.7) indicates that the v.c. is $E > [a]^-/2$, which is a less strong v.c. that the universal v.c. of $E > b$.

The principal results described in this section are summarized in Table 5 (in it a v.c. generic is also included for the very habitual case that $n \leq 500$).

4 Conclusions

Given the difficulty of calculating the exact p-value for a 2×2 table under Models I and II (and for pedagogic-practical reasons in Model III), it is advisable to use an asymptotic test which allows one to obtain a p-value (P_A) near the exact p-value P_E. In all three cases it is usual to use a chi-squared test for this purpose, but each model requires a different continuity correction (see Table 5).

The asymptotic test is not always valid. Its validity condition (v.c.) depends on the minimum marginal imbalance (in Models I and III) or the sample imbalance (in Model II) -called K in the article- and on the size of the said marginal (called n_1 in the article). This v.c. adopts the generic form of $E > E^*(n_1, K)$ in the three models, but the function depends on some constants, a and b, which vary with the model and number of tails of the test. The physical sense of these constants is indicated in Table 5, which also gives various rules for v.c.'s.

A universal v.c. that is especially easy to use (but very conservative) is $E > b$ ($E > [a]^-/2$ when $K = 1$). The v.c. $E > b$ is very near the classic $E > 5$ in Model I, but draws much farther away from it in the other two models.

Table 5: Validity conditions for the chi-squared test in Models *I, II and III (notation of Table 1)*

Concept	**MODEL I (1)**	**MODEL II (1)**	**MODEL III (1)(2)**
Exact P_E value for the method:	Unconditional		Conditional
	Order χ_Y^2	Barnard's order	Order χ_Y^2
Asymptotic test used:	$\chi_I^2 = \frac{(\lvert x_1y_2 - x_2y_1\rvert - 0.5)^2}{a_1a_2n_1n_2}(n-1)$	$\chi_{II}^2 = \frac{(\lvert x_1y_2 - x_2y_1\rvert - c)^2}{a_1a_2n_1n_2}(n-1)$ where $\begin{cases} c=2 & \text{if } n_1 = n_2 \\ c=1 & \text{if } n_1 \neq n_2 \end{cases}$	$\chi_{III}^2 = \chi_Y^2 + \frac{n^3}{4a_1a_2n_1n_2}$ where $\chi_Y^2 = \frac{(\lvert x_1y_2 - x_2y_1\rvert - n/2)^2}{a_1a_2n_1n_2}n$
Asymptotic P_A value (3)	$P(\chi_I)$ $2 \times P(\chi_I)$	$P(\chi_{II})$ $2 \times P(\chi_{II})$	$P(\chi_{III})$ $(P(\chi_Y))$ $P(\chi_Y) + P(\chi_{Y'})$ **(4)**
P_A is acceptable if:	$\lvert P_A - P_E\rvert \leq \delta P_E, \quad \text{where} \quad \delta = \begin{cases} 1, & \text{if } P_E \leq 1\% \\ 1.15 - 15 \times P_E, & \text{if } 1\% < P_E \leq 5\% \\ 0.5 - 2 \times P_E, & \text{if } 5\% < P_E \leq 10\% \\ 0.3, & \text{if } P_E > 10\% \end{cases}$		
$y^* = a + bx$ $x = (K-1)^2/K$	$7.318 + 3.899x$ $7.612 + 6.120x$	$9.743 + 14.786x$ $16.315 + 14.889x$	$6.486 + 19.111x$ $(7.785 + 20.627x)$ The asymptotic test is always valid
Conditions	$K = n_2/n_1 \geq 1$, $\lvert n_2 - n_1\rvert \leq \lvert a_2 - a_1\rvert$	$K = n_2/n_1 \geq 1$	$K = n_2/n_1 \geq 1$, $\lvert n_2 - n_1\rvert \leq \lvert a_2 - a_1\rvert$

(*Continued on next page*)

(Table 5. Continued from previous page)

Concept	MODEL I (1)	MODEL II (1)	MODEL III (1)(2)
$v.c.$ $(n_1; K)$	$E > E^* = \frac{1}{K+1}\left[\frac{n}{2}\left(1-\sqrt{\frac{C}{C+4y^*}}\right)\right]^-$ where $[z]^-$ = the whole down of z		
$C(\simeq n)$	$n^4/(n-2)^2(n-1)$	$n^2/(n-1)$	$(n-2)^2/(n-1)$
$v.c.(K)$	$E > [y^*]^-/(K+1) = \left[a + b(K-1)^2/K\right]^-/(K+1)$		
$v.c.$(if $K = 1$) $E > [a]^-/2$	$E > 3.5$ $E > 3.5$	$E > 4.5$ $E > 8$	$E > 3$ (3.5) $E > 0$
General $v.c.$ $E > b$ [if $n \leq 500$]	$E > 3.9$ [3.0] $E > 6.2$ [3.9]	$E > 14.8$ [7.2] $E > 14.9$ [7.7]	$E > 19.2$ [8.1] (20.7) $E > 0$ [0]
If $a_1{=}a_2{=}n_1{=}n_2$	The asymptotic test is always valid		
If n **is large:**	The test is not valid if $a_1 \leq [a]^-$		
	$a_1 \leq 7$ $a_1 \leq 7$	$a_1 \leq 9$ $a_1 \leq 16$	$a_1 \leq 6$ (7) ———

NOTES

(1) When there are two lines of results in one cell, the 1st (2nd) refers to the one-(two-)tailed test.
(2) The data in brackets refer to the case where the one-tailed test is carried out using the classic statistic χ^2_Y.
(3) $P(\chi) = 1 - F(\chi)$, where $F(\cdot)$ is the distribution function of the typical normal distribution.
(4) $\chi^2_{Y'}$ refers to the value of χ^2_Y in the table where $x'_1 = [2E - x_1]$ and $[\cdot]$ refers to the rounding of x in the sense of moving away from the value of E = "minimum expected quantity" = $\min(a_1, a_2) \times \min(n_1, n_2)/n$.

Annex I: Asymmetry coefficient in MODEL I

In order to facilitate the present demonstration, let us agree to call (x_1, x_2, x_3, x_4) the values (x_1, y_1, x_2, y_2) in Table 1, and (p_1, p_2, p_3, p_4) the probabilities. Let $p = p_1 + p_3$ and $q = p_1 + p_2$. Pirie and Hamdan (1972) proved that, in order to contrast independence $(H : p_1 p_4 = p_2 p_3)$ the appropriate statistic is $U = x_1 x_4 - x_2 x_3$ where $E(U) = n(n-1)(p_1 p_4 - p_2 p_3)$, $E(U|H) = 0$ and $V(U|H) = E\left(U^2|H\right) = n^2 (n-1) p_1 p_4 = \sigma^2$.

The moment generating function of the multinomial distribution is

$$\psi^n = \varphi(t_1, t_2, t_3, t_4) = \left(\sum_{i=1}^{4} p_i e^{t_i} \right)^n ,$$

so that any non-central moment can be obtained as follows:

$$\mu'_{r_1,r_2,r_3,r_4} = E\left(\prod_{i=1}^{4} x_i^{r_i} \right) = \left[\frac{\partial^{\sum r_i} \psi^n}{\prod \partial t_i^{r_i}} \right]_{t_1=t_2=t_3=t_4=0.} \tag{I.1}$$

For example:

$$\mu'_{1,0,0,1} = \left[\frac{\partial^2 \psi^n}{\partial t_1 \partial t_4} \right]_{t_1=t_2=t_3=t_4=0} = \left[\frac{\partial}{\partial t_4} \left\{ n \psi^{n-1} p_1 e^{t_1} \right\} \right]_{t_1=t_2=t_3=t_4=0}$$

$$= \left[n p_1 e^{t_1} \left\{ (n-1) \psi^{n-2} p_4 e^{t_4} \right\} \right]_{t_1=t_2=t_3=t_4=0} = n(n-1) p_1 p_4$$

given that in $t_i = 0\,(\forall i)$, we have $\psi = 1$ because $\sum p_i = 1$. By proceeding in similar fashion, one can obtain all the non-central moments necessary: $\mu'_{1,0,0,1}$, $\mu'_{0,1,1,0}$, $\mu'_{1,1,1,1}$, $\mu'_{2,0,0,2}$, $\mu'_{0,2,2,0}$, $\mu'_{2,1,1,2}$, $\mu'_{1,2,2,1}$, $\mu'_{3,0,0,3}$ and $\mu'_{0,3,3,0}$ (the first five were obtained by Pirie and Hamdan (1972), a different way).

Let μ_r (μ'_r) the central (non-central) moments of order r of the statistic U. By definition $\beta_I = \mu_3^2 / \mu_2^3$ is the square of the coefficient of asymmetry of U. As β_I will be obtained under H, then, using expression (I.1),

$$\beta_1 = \frac{{\mu'_3}^2}{{\mu'_2}^3} \frac{E^2\left(U^3|H\right)}{E^3\left(U^2|H\right)} = \frac{E^2\left(U^3|H\right)}{\sigma^6} \tag{I.2}$$

so that only $E\left(U^3|H\right)$ is left unknown. As

$$E\left(U^3\right) = E\left\{ (x_1 x_4 - x_2 x_3)^3 \right\} = \left(\mu'_{3,0,0,3} - \mu'_{0,3,3,0} \right) + 3\left(\mu'_{1,2,2,1} - \mu'_{2,1,1,2} \right),$$

then, by evaluating these differences under the condition $p_1p_4 = p_2p_3$, bearing in mind that $\sum p_i = 1$ and putting everything in terms of p_1 and p_4, one is left with

$$E\left(U^3|H\right) = n^2\,(n-1)\,(n-2)\,p_1p_4\,(2p_1 + 2p_4 - 1)\,.$$

Hence, through expression (I.2), $\beta_I = \left[(n-2)^2\,(2p_1+2p_2-1)^2\right]\Big/\sigma^2$. But, under H, $p_1 = pq$ and $p_4 = (1-p)\,(1-q)$. By substituting in β_I one obtains:

$$\beta_I \;=\; \frac{(n-2)^2\,(1-2p)^2\,(1-2q)^2}{\sigma^2} \tag{I.3}$$

where p and q are two unknown parameters.

In order to estimate β_I, its numerator and denominator will be estimated separately. For the former, it is sufficient to substitute p and q by $\hat{p} = a_1/n$ and $\hat{q} = n_1/n$ (its estimators of maximum likelihood). For the latter it is logical to substitute σ^2 by the value $\hat{\sigma}^2$ used for obtaining the test. Since $\chi^2_I = (U-0.5)^2/V(U|H)$, and $V\,(U\,|H\,) = \sigma^2$ was estimated by Richardson (1990) by $\hat{\sigma}^2 = a_1a_2n_1n_2/(n-1)$, then:

$$\hat{\beta}_I \;=\; \frac{(n-2)^2(n-1)}{n^4} \times \frac{(a_2-a_1)^2(n_2-n_1)^2}{a_1a_2n_1n_2} \tag{I.4}$$

the expression used in this article. The value $\hat{\sigma}^2$ was obtained by Richardson (1990) by substituting $p\,(1-p)$ and $q\,(1-q)$ by their unbiased estimators $\hat{p}(1-\hat{p})n/(n-1)$ and $\hat{q}(1-\hat{q})n/(n-1)$.

For the following, we return to the classic notation of the article.

Annex II: The asymmetry coefficient in MODEL II

For Model II, $x_i \sim B\,(n_i, p_i)$, so that, under $H: p_1 = p_2\,(=p)$, $E\,(\hat{p}_1 - \hat{p}_2) = 0$ and $V\,(\hat{p}_1 - \hat{p}_2) = p\,(1-p)\,n/(n_1n_2)$, where $\hat{p}_i = x_i/n_i$. As $p\,(1-p)$ is unknown, it is substituted by its unbiased estimator

$$\frac{\hat{p}(1-\hat{p})n}{n-1} = \frac{a_1a_2}{n(n-1)},$$

and so $(\hat{p}_1 - \hat{p}_2)^2 \Big/ V\,(\hat{p}_1 - \hat{p}_2) = \chi^2_U$ with factor $(n-1)$, which is the basis of statistic χ^2_{II}. Consequently the statistic whose asymmetry should be calculated is $\hat{p}_1 - \hat{p}_2$.

The first three central moments of a binomial r.v. $x \sim B(n,p)$ are $\mu_1 = 0$, $\mu_2 = np(1-p)$ and $\mu_3 = np(1-p)(1-2p)$. The same central moments of the r.v. x/n will be $\bar{\mu}_i = \mu_r/n^r$. As in Model II, under H, $x_i \sim B(n_i, p_i)$, then the first three central moments of $\hat{p}_i$ will be $\bar{\mu}_{1,i} = 0$, $\bar{\mu}_{2,i} = p(1-p)/n_i$ and $\bar{\mu}_{3,i} = p(1-p)(1-2p)/n_i^2$.

As $\hat{p}_1$ and $\hat{p}_2$ are independents, then the first three central moments $\bar{\bar{\mu}}_i$ of $\hat{p}_1 - \hat{p}_2$ are

$$\begin{aligned}
\bar{\bar{\mu}}_1 &= E(\hat{p}_1 - \hat{p}_2) = 0 \\
\bar{\bar{\mu}}_2 &= V(\hat{p}_1 - \hat{p}_2) = \frac{np(1-p)}{n_1 n_2} \\
\bar{\bar{\mu}}_3 &= E\left\{(\hat{p}_1 - p) - (\hat{p}_2 - p)\right\}^3 = \sum_{h=0}^{3} (-1)^{3-h} C(3,h)\, \bar{\mu}_{h,1} \bar{\mu}_{3-h,2} \\
&= \frac{p(1-p)(1-2p)n(n_2 - n_1)}{n_1^2 n_2^2}.
\end{aligned}$$

As the square of the asymmetry coefficient of $\hat{p}_1 - \hat{p}_2$ is $\beta_{II} = \bar{\bar{\mu}}_3^2/\bar{\bar{\mu}}_2^3$, then $\beta_{II} = (1-2p)^2 (n_2 - n_1)^2 \Big/ \{n_1 n_2 np(1-p)\}$. Finally, and as in Annex I, in order to estimate β_{II}, p is substituted by $\hat{p} = a_1/n$ in the numerator, and $p(1-p)$ by $a_1 a_2/\{n(n-1)\}$ in the denominator (the latter in accordance with the way the statistic was obtained χ_U^2). Therefore the expression used in this article is:

$$\hat{\beta}_{II} \;=\; \frac{n-1}{n^2} \times \frac{(a_2 - a_1)^2 (n_2 - n_1)^2}{a_1 a_2 n_1 n_2}$$

Annex III: Study of the function $\mathbf{E}^*(\mathbf{n_1}, \mathbf{K})$

The function $E^*(n_1, K)$ is given by expression (3.5). In order to simplify the reasoning, and without loss of generality, let us leave out the rounding down (that is, the symbol $[\cdot]^-$). Therefore:

$$E^*(n_1, K) = \frac{n_1}{2}\left(1 - \sqrt{\frac{C}{4y^* + C}}\right) \tag{III.1}$$

In order to study the evolution of E^* in terms of n_1, because $C \simeq n$ in the three models, let us make $C = (K+1)\, n_1$ and note $D = 4y^*/(K+1)$.

So, $E^* > (n_1/2)\left[1-\{n_1/(D+n_1)\}^{0.5}\right]$. By multiplying and dividing by the conjugate the result is $E^* \simeq (D/2)\, z\big/\{1+z^{0.5}\}$, where $z = n_1/(D+n_1)$. As $D > 0$ is independent from n_1, the sign of $\mathrm{d}E^*/\mathrm{d}n_1$ is the same as that of $\mathrm{d}f/\mathrm{d}n_1$, where $f = z\big/\{1+z^{0.5}\}$. But $df/dn_1 = (\partial f/\partial z)\times(\partial f/\partial n_1)$ and, since both partial derivatives are positive, one can deduce that $\mathrm{d}E^*/\mathrm{d}n_1 > 0$ and so E^* increases in n_1 (the increase is not strict if one takes into account the symbol $[\cdot]^-$ which was left out). What has been said above implies that, for each fixed value of K, the maximum of $E^*(n_1, K)$ is reached in $\lim_{n_1\to\infty} E^*(n_1,K) = E^*(K)$. By multiplying and dividing (III.1) by its conjugate, the result is $E^*(K) = [y^*]^-\big/(K+1)$ because C is order n_1. Hence $E > E^*(K)$ is a valid (but conservative) v.c.

In order to obtain a universal v.c., one must calculate the maximum of $E^*(K)$. By again leaving out $[\cdot]^-$ –which is irrelevant for the present purpose– the result is $\mathrm{d}E^*/\mathrm{d}K = f(K)\,K^{-2}(K+1)^{-2}$, where $f(K) = (3-r)\,K^2 - 2K - 1$ and $r = a/b$, and so the sign of $\mathrm{d}E^*/\mathrm{d}K$ is that of $f(K)$. The extreme value $E^*(K)$, if it exists, is reached when $f(K) = 0$, that is, in $K_1,\ K_2 = \left\{1 \pm (4-r)^{0.5}\right\}\big/(3-r)$. If the values of K_i exist, then $f(K) = (3-r)(K-K_1)(K-K_2)$, from which one can deduce that:

i. If $r \geq 3 \Rightarrow f(K) < 0 \Rightarrow E^*(K)$ decreases in $K \Rightarrow$ its maximum is reached in $K = 1$ and it is $[a]^-/2$.

ii. If $r > 3 \Rightarrow K_1 < 0$ (which is not a licit value) and $K_2 > 1 \Rightarrow$ in $K = K_1$ there is a maximum and in $K = K_2$ there is a minimum $\Rightarrow$ the maximum of $E^*(K)$ is the maximum of the values $E^*(K=1)$ $= [a]^-/2$ and $E^*(K=\infty) = b$.

From this it can be stated that a universal v.c. is:

$$\begin{cases} If\ r = a/b \geq 3:\ E > [a]^-/2 \\ If\ r = a/b < 3:\ E > \max\ \{[a]^-/2,\ b\} \end{cases}$$

because in the first case $E^*(n_1, K)$ reaches an absolute maximum in $K = 1$, and in the second it reaches an absolute minimum in $K = K_2 = \left\{1+(4-r)^{0.5}\right\}\big/(3-r)$. This last statement is valid for large values of n (because of the approximation $C \simeq n$).

For the data of the article, $r < 3$ always. This implies, on one hand, that $E^*(n_1, K)$ reaches an absolute maximum in b, because in all three cases

$b > [a]^- /2$. On the other hand, $E^*(n_1, K)$ reaches an absolute minimum in the neighboring of $K = K_2 \simeq 2.2$ (1.5), 1.2 (1.4) and 1.1 (0) for the one- (two-) tailed tests in MODELS I, II and III respectively. This can be confirmed with the values $\hat{E}^*$ of Tables 4, 3 and 2 respectively.

References

BARNARD, G. A. (1947). Significance test 2×2 tables. *Biometrika*, 34:123–138.

CAMILLI, G. and HOPKINS, K. D. (1978). Applicability of chi-square to 2 × 2 contingency tables with small expected cell frequencies. *Psycho–logical Bulletin*, 85:163–167.

CAMILLI, G. and HOPKINS, K. D. (1979). Testing for association in 2 × 2 contingency tables with very small samples sizes. *Psychological Bulletin*, 86(5):1011–1014.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10:417–451.

CONOVER, W. J. (1974). Some reasons for not using the Yates' continuity corrections on 2 × 2 contingency tables. *Journal of the American Statistical Association*, 69:374–376.

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

FISHER, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society. Series A*, 98:39–54.

GREENWOOD, P. E. and NIKULIN, M. S. (1996). *A Guide to Chi-Squared Testing*. Wiley, New York.

GRIZZLE, J. E. (1967). Continuity correction in the χ^2 test for 2×2 tables. *The American Statistician*, 21(4):28–32.

LARNTZ, K. (1978). Small-sample comparisons of exact levels for chi-square goodness-of-fit statistics. *Journal of the American Statistical Association*, 73:253–263.

MANTEL, N. (1974). Comment and a suggestion. *Journal of the American Statistical Association*, 69:378–380.

MARTÍN ANDRÉS, A. (1991). A review of classic non-asymptotic me–thods for comparing two proportions by means of independent samples. *Communications in Statistics. Simulation and Computation*, 20(2&3):551–583.

MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (1995). Is Fisher's exact test very conservative? *Computational Statistics and Data Analysis*, 19:579–591.

MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (1997). On condition for validity of the approximations to Fisher's exact test. *Biometrical Journal*, 39:935–954.

MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (2000). On the minimum expected quantity for the chi-square test in 2×2 tables. *Journal of Applied Statistics*, 27(7):807–820.

MARTÍN ANDRÉS, A., HERRANZ TEJEDOR, I., and LUNA DEL CASTILLO, J. D. (1992). Optimal correction for continuity in the chi-square test in 2×2 tables (conditioned method). *Communications in Statistics. Simulation and Computation*, 21(4):1077–1101.

MARTÍN ANDRÉS, A., SÁNCHEZ 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(1):107–115.

MARTÍN ANDRÉS, A., SÁNCHEZ QUEVEDO, M. J., and SILVA MATO, A. (2002). Asymptotical tests in 2×2 comparative trials (unconditional approach). *Computational Statistics and Data Analysis*, 40(2):339–354.

MARTÍN ANDRÉS, A. and SILVA MATO, A. (1996). Optimal correction for continuity and conditions for validity in the unconditional chi-square test. *Computational Statistics and Data Analysis*, 21:609–626. Erratum in 26:235 (1997).

MARTÍN ANDRÉS, A. and TAPIA GARCÍA, J. M. (1999). Optimal unconditional test in 2×2 multinomial trials. *Computational Statistics and Data Analysis*, 31(3):311–321.

Martín Andrés, A. and Tapia García, J. M. (2004). Optimal unconditional asymptotic test in 2×2 multinomial trials. *Communications in Statistics. Simulation and Computation*, 33(1):83–97.

Pirie, W. R. and Hamdan, M. A. (1972). Some revised continuity corrections for discrete distributions. *Biometrics*, 28:693–701.

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

Sachs, L. (1986). Alternatives to the chi-square test of homogeneity in 2×2 tables and to Fisher's exact test. *Biometrical Journal*, 28(8):975–979.

Scott, W. F. (1999). On 2×2 contingency tables. *The Mathematical Scientist*, 24(2):148–153.

Shuster, J. J. (1992). Exact unconditional tables for significance testing in the 2×2 multinomial trial. *Statistics in Medicine*, 11(7):913–922. Correction in 11:1619 (1992).

Yates, F. (1984). Test of significance for 2×2 contingency tables. *Journal of the Royal Statistical Society. Series A*, 147(3):426–463.