Revista Colombiana de Estadística - Theoretical Statistics
January 2026, volume 49, issue 1, pp. 89–107
http://doi.org/10.15446/rce.v49n1.118682

Asymmetric Chi-square Test and Cohen's w in Contingency Tables

Prueba de chi-cuadrado y w de Cohen asimétricas en tablas de contingencia

Luis D'Angelo[1,2,a]

[1]Master's Program in Medical Molecular Biology, Faculty of Pharmacy and Biochemistry, University of Buenos Aires (UBA), CABA, Argentina

[2]Statistics Department, Biostatistics, National University of Pilar, Provincia de Buenos Aires, Argentina

Abstract

This article presents a new asymmetric version of Cohen's w for analyzing contingency tables. As an extension of this established effect size measure, the proposed index quantifies the effect of one variable on another, providing a valuable complement to null hypothesis significance testing. While specific procedures exist for assessing these directional relationships, they exhibit significant limitations in certain scenarios.

Furthermore, we introduce a normalization process that constrains the coefficient to a $[0, 1]$ range, enhancing interpretability for both researchers and practitioners.

Finally, we present an asymmetric chi-square coefficient that aligns naturally with the proposed effect size, ensuring full conceptual coherence between hypothesis testing and effect size estimation. This coefficient also avoids the interpretability pitfalls that commonly arise when the traditional chi-square test is applied to inherently asymmetric relationships.

Keywords: Asymmetric relationship; Chi-square test; Cohen's w.

Resumen

Este artículo presenta una nueva versión asimétrica de la w de Cohen para analizar tablas de contingencia. Como una extensión de esta medida de tamaño del efecto ya establecida, el índice propuesto cuantifica el efecto de una variable sobre otra, constituyendo un valioso complemento a las pruebas de significación de hipótesis nula. Si bien existen procedimientos específicos para evaluar estas relaciones direccionales, estos presentan limitaciones significativas en ciertos escenarios.

[a]Master's degree. E-mail: luis11dangelo@gmail.com

Además, introducimos un proceso de normalización que restringe el coeficiente al rango $[0, 1]$, mejorando su interpretabilidad tanto para investigadores como para profesionales.

Finalmente, presentamos un coeficiente de chi-cuadrado asimétrico que se alinea naturalmente con el tamaño del efecto propuesto, garantizando una coherencia conceptual plena entre la prueba de hipótesis y la estimación de magnitud del efecto. Este coeficiente también evita los problemas de interpretabilidad que comúnmente surgen cuando la prueba de chi-cuadrado tradicional se aplica a relaciones inherentemente asimétricas.

***Palabras clave*:** Prueba de chi-cuadrado; Relación asimétrica; w de Cohen.

1. Introduction

While the mathematical foundations of the chi-square distribution were laid by multiple authors, it became a cornerstone of modern statistics through the work of English mathematician Karl Pearson (1857–1936). Pearson's application of the distribution, especially in the context of goodness-of-fit tests, culminated in the development of the chi-square test, published in 1900. This test revolutionized statistical hypothesis testing. Later refinements by prominent statisticians such as Yule (1911), Fisher (1925), Yates (1934), and Cochran (1952, 1954) expanded its utility, introducing corrections for small sample sizes and broadening its application in various fields of research.

Building on these foundations, Cohen (1988) provided a pivotal contribution by introducing Cohen's w, a measure of effect size for chi-square tests in contingency tables. His work emphasized the importance of interpreting statistical results not only in terms of significance but also through the lens of practical relevance, thus enriching the applicability of the chi-square test in both theoretical and applied research.

Our study builds upon this historical groundwork to expand the analytical toolkit for two-variable contingency tables by introducing measures specifically designed for asymmetric relationships. We present several key innovations:

1. An asymmetric version of Cohen's w that enables a more nuanced comparison between each category of the independent variable and the marginal distribution of the dependent variable. By accounting for data asymmetry, this measure provides a more accurate representation of the relationship's strength and direction, potentially offering deeper insights into variable dependencies.

2. A normalization process that constrains the coefficient to the $[0, 1]$ range, making results more accessible to researchers and practitioners while facilitating easier interpretation.

3. An asymmetric chi-square coefficient that addresses asymmetries in variable relationships. This measure integrates null hypothesis significance testing with effect size estimation, thereby enhancing the interpretability of the

results and facilitating a more accurate understanding of underlying data associations.

2. Chi-Square Distribution and Tests

Perhaps the simplest way to define the chi-square distribution is to consider it as the sum of squares of independent standard normal random variables, each following $N(0,1)$ with degrees of freedom df:

$$\chi^2_{df} = \sum_{i=1}^{df} Z_i^2 = Z_1^2 + Z_2^2 + Z_3^2 + \cdots + Z_{df}^2. \tag{1}$$

Given that its expected value $E(X) = df$, and its variance $V(X) = 2df$.

2.1. Chi-Square Test (χ^2)

Pearson (1900) developed the chi-square goodness-of-fit test. Given a hypothetical distribution and an empirical distribution, both can be compared using a chi-square test.

2.2. Goodness-of-Fit Test

The *chi-square goodness-of-fit test* evaluates whether empirically observed data conform to a theoretical probability distribution by comparing observed frequencies with expected frequencies derived from the theoretical model. The test statistic is calculated as:

$$\chi^2 = \sum_{i=1}^{r} \frac{(o_i - e_i)^2}{e_i}. \tag{2}$$

The term o_i represents the observed frequency in category i, while e_i refers to the expected frequency under the theoretical distribution. The resulting test statistic is compared against a critical value from the chi-square distribution with degrees of freedom:

$$df = (r - 1), \tag{3}$$

in which r denotes the total number of categories or classes under consideration. This statistical technique, since its original proposal by Pearson, has undergone virtually no changes.

2.3. Independence Test

Pearson (1904) introduced the concept of *contingency* and the *chi-square test* as a hypothesis test for independence between two qualitative variables.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}, \tag{4}$$

where each term represent:

- χ^2: the calculated chi-square statistic, measuring the discrepancy between observed and expected frequencies.
- r: the number of rows in the contingency table.
- c: the number of columns in the contingency table.
- o_{ij}: the observed frequency in the cell at row i and column j.
- e_{ij}: the expected frequency in the cell at row i and column j, calculated under the assumption of independence between the variables.

And e_{ij} are calculated as follows:

$$e_{ij} = \left(\frac{r_i}{n}\right)\left(\frac{c_j}{n}\right) n, \tag{5}$$

with:

- r_i: the total of row i,
- c_j: the total of column j,
- n: the grand total of all observations in the table.

In other words, since we are assuming independence, we expect in cell ij the product of the probabilities, multiplied by n to obtain its corresponding frequency.

Then, the test is conducted according to the chi-square distribution based on its degrees of freedom and significance level (p-value):

$$df = (r-1)(c-1), \tag{6}$$

where:

- r: the number of rows in the contingency table.
- c: the number of columns in the contingency table.

This procedure is designed for analyzing associations between two categorical variables using a single representative sample.

The Pearson chi-square test is mathematically designed to treat variables symmetrically (Agresti, 2002, p. 87). Consequently, its application to an asymmetric hypothesis, where one variable is theorized as dependent and the other independent, is statistically incorrect. The test's structure cannot uphold a required directional specification.

2.4. Homogeneity Test

Pearson's chi-square framework can also be applied to test for homogeneity across groups (Fisher, 1925; Crack, 2018). The key distinction lies in the null hypothesis. As Cramer notes (Cramer, 1946, p. 445), the null hypothesis of homogeneity posits an unknown common distribution p_i for all groups, whether compared by columns or rows (DeGroot, 1988, p. 517).

Despite this conceptual difference, authors consistently emphasize that the test statistic for homogeneity is calculated identically to that for independence (Equations (4), (5), and (6)). Critically, the homogeneity test requires two or more independent samples, unlike the test of independence, which analyzes two variables within a single sample.

To clarify further, in the homogeneity test, finding significant differences in proportions between groups suggests that the distributions are different; however, it does not directly test for an association between the variables. Why doesn't it necessarily imply association? Observed differences in proportions may arise from variability within the groups or from other factors not accounted for in the design. In other words, although a difference exists, the test does not evaluate a direct dependency between the variables (Franke et al., 2012).

3. Effect Size in Contingency Tables

For 2×2 contingency tables, numerous authors, including Rosenthal (1994), Agresti (2002), Sanchez-Meca et al. (2003), Fleiss et al. (2003), and Rita & Komonen (2008), and Borenstein et al. (2009), widely recommend the use of odds ratios as the most appropriate effect size measure. These experts highlight its utility and robustness in quantifying the strength of association between variables, making it a widely accepted and reliable metric across various research fields.

Although the *odds ratio* is highly valuable for quantifying association strength in studies with dichotomous variables, its applicability is limited. When contingency tables cannot be simplified to a 2×2 format, alternative effect size measures become necessary. Commonly used metrics for larger tables include Cramer's V and Cohen's w. Unlike the *odds ratio*, which focuses on comparing two proportions, these measures assess the overall relationship between categorical variables.

However, these effect size indices have been criticized by Fleiss et al. (2003) and by Haddock et al. (1998) as their utilization, based on treating variables as if they were quantitative, sometimes tends to underestimate the true effect size.

The *phi* coefficient (ϕ) is based on the chi-square:

$$phi(\phi) = \sqrt{\frac{\chi^2}{n}}. \tag{7}$$

Cramer's V is simply an extension of *phi* for the case of mxn tables:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}, \tag{8}$$

where $k = \min(r, c)$ is defined as the minimum value between the number of rows (r) or columns (c).

3.1. Cohen's *w*

Cohen's w coefficient is an association measure analogous to the *phi* coefficient and Cramer's V. Although originally defined by Cohen (1988), the following formula represents an adapted version proposed in that work:

$$w = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\frac{o_{ij}}{n} - \frac{mr_i\, mc_j}{n^2}\right)^2}{\frac{mr_i\, mc_j}{n^2}}}, \tag{9}$$

where,

- o_{ij} is the observed frequency in cell ij.
- mr_i is the sum of observed frequencies in row i.
- mc_j is the sum of observed frequencies in column j.
- n is the total number of cases.

The statistic is calculated by summing the squared differences between observed and expected probabilities, normalized by the expected probabilities, across all cells of the contingency table. The final value w is obtained by taking the square root of this sum.

Cohen's w measures the strength of association between categorical variables. A value near 0 suggests a weak association, while values approaching or exceeding 1 indicate a strong association.

Cohen (1988, p. 216) points out that the calculation of w is analogous to that of the chi-square statistic and emphasizes the measure's symmetry.

Another way to calculate w is based on Cramer's V:

$$w = V\ \sqrt{k-1}. \tag{10}$$

Where k is the number of categories of the variable with the fewer categories.

Note that Ben-Shachar and others have recently proposed, for the case of goodness-of-fit tests, a coefficient derived from Cohen's w, which takes the following form (Ben-Shachar et al., 2023; Jané et al., 2024):

$$Fei = \sqrt{\frac{\chi^2}{n\left(\frac{1}{\min(p_e)} - 1\right)}} \tag{11}$$

- χ^2: the chi-square statistic calculated from goodness-of-fit test.
- n: the total sample size.
- $\min(p_e)$: the smallest expected probability among all categories under the null hypothesis.

The *Fei* coefficient, unlike Cohen's w, has the virtue of remaining between the values 0 and 1.

3.2. Confidence Intervals for Cohen's w

The confidence interval of this effect size statistic can be calculated by various procedures.

- Bootstrap method: A resampling technique used to estimate the distribution of a statistic from observed data. This method generates multiple bootstrap samples by randomly selecting observations with replacement. Confidence intervals can be estimated from these samples. This procedure can be carried out in cases of studies with representative samples. Various R libraries are available to perform this procedure. We recommend using *rcompanion*, developed by Mangiafico (2023).
- Permutation method: Another resampling technique that can be adapted to construct confidence intervals for parameters and group differences without distributional assumptions. The method generates multiple samples through random permutations of observations across groups without replacement, allowing estimation of sampling variability and calculation of confidence intervals (Edgington & Onghena, 2007; Good, 2005). This technique is especially useful in cases of experimental designs.

  Calculations can be obtained on the Lock et al. (2021): `https://www.lock5stat.com/StatKey`
- Non-central chi-square parameter method (NCCP): This approach constructs confidence intervals for Cohen's w by estimating the non-centrality parameter (λ) of a chi-squared distribution. The relationship between Cohen's w and the non-central chi-square distribution is given by $\lambda = nw^2$, where n is the sample size and λ is the non-centrality parameter. Under the alternative hypothesis, the chi-square statistic follows a non-central chi-square distribution with $df = (r-1)(c-1)$ degrees of freedom.

  A confidence interval for Cohen's w can be obtained by first determining the confidence interval for the non-centrality parameter λ. This approach preserves the exact relationship between the effect size and the chi-square statistic, ensuring that the resulting interval accurately reflects the theoretical properties of the non-central chi-square distribution. The bounds of λ are calculated numerically from the appropriate quantiles of the non-central chi-square distribution, after which the confidence interval for Cohen's w

is directly derived. Compared to bootstrap or permutation methods, this procedure is generally more reliable and considerably more precise than normal approximations, particularly for small to moderate sample sizes (Algina et al., 2006; Grissom & Kim, 2012; Kelly, 2007; Steiger, 2004).

The NCCP method can be applied in both experimental studies, where group assignment is randomized, and observational studies with representative samples. For practical implementation, we recommend using the *effectsize* package in R (Ben-Shachar et al., 2023) or the Real Statistics Excel add-in by Zaiontz (2024) .

4. Chi-square from Cohen's w

One way of calculating chi-square is given by:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = w^2\, n. \tag{12}$$

Hence, by (9) the chi-square statistic can also be computed as:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\frac{o_{ij}}{n} - \frac{mr_i\, mc_j}{n^2}\right)^2}{\frac{mr_i\, mc_j}{n^2}} n. \tag{13}$$

The degrees of freedom (df) should be calculated as in Equation (6). This formulation provides another approach for computing the chi-square statistic using observed and expected proportions of independence. Please make a note of this method as it will be beneficial for us later on.

5. Asymmetric Cohen's w

Thus far, we have discussed the classic method for calculating Cohen's w, as originally proposed by the author. This conventional formulation treats variables symmetrically, lacking distinction between independent and dependent variables. This approach shares a fundamental limitation with the Pearson chi-square test upon which it is based: the inability to model directional relationships, which often leads to researcher error. This is corroborated by empirical evidence; for instance, Franke et al. (2012) reported that over half of the published studies using chi-square tests misapplied or misinterpreted them, with correct applications observed in only about 44% of cases. Many of these errors stemmed precisely from treating asymmetric hypotheses with symmetric tools.

However, when we clearly define independent and dependent variables, the measures of association should be asymmetric. An asymmetric relationship between variables occurs when one variable is treated as dependent (the outcome to be explained or predicted) and the other as independent (the explanatory, the

factor or predictor variable). In this type of relationship, the association is not reciprocal: changes in the independent variable are analyzed in terms of their effect on the dependent variable, but the reverse is not necessarily true. Notable examples of such measures include Goodman and Kruskal's Lambda (λ) and Tau (τ_a and τ_b) Goodman & Kruskal (1963), as well as Theil's uncertainty coefficient (U). All three can be interpreted as measures of reduction in error or uncertainty, though they differ in focus: Lambda quantifies the proportional reduction of errors when predicting the modal category of the dependent variable, Tau measures the reduction of prediction error across all categories, and U quantifies the reduction in uncertainty when the independent variable is known. While these measures are informative, they have limitations: for instance, Lambda may underestimate association when the dependent variable is highly skewed, Tau tends to underestimate association in the central portion of its range, as it is affected by sparse or highly dispersed category distributions, and U behaves similarly to Tau but can be complex to interpret in practice (Berry et al., 2018).

To address these limitations while providing a standardized and interpretable effect size, this paper introduces a novel application of Cohen's w. Rather than testing for total independence, we examine a scenario where the null hypothesis is defined by the marginal distribution of one variable. For simplicity, we will consider only the columns as the independent (or factor) variable, examining how they relate to a *single sample* while focusing on *asymmetric relationships* between the variables:

$$w_{c,\ \text{Asym}} = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\frac{o_{ij}}{mc_j} - \frac{mr_i}{n}\right)^2}{\frac{mr_i}{n}}} \tag{14}$$

- o_{ij} represents the observed frequency in each cell ij.
- mc_j is the marginal frequency of each column j in the contingency table.
- $\frac{o_{ij}}{mc_j}$ denotes the alternative hypothesis.
- mr_i denotes the marginal frequency of row i, which is the total sum of values in the same row.
- n represents the total number of observations in the contingency table.
- $\frac{mr_i}{n}$ represents the marginal probability of the contingency table in the column direction, corresponding to the null hypothesis.
- r is the number of rows, and c refers to the number of columns in the contingency table, which, in this context, represents the factor variable.

This formula computes the asymmetric Cohen's w by summing, across all cells of the contingency table, the squared differences between each cell's proportion

within its column (j) and the corresponding marginal probability, with each difference normalized by this marginal probability. The final statistic is the square root of this sum.

This asymmetric adaptation retains Cohen's original computational framework while conditioning on the column variable. The interpretation mirrors standard Cohen's w but is applied directionally (for simplicity, this paper focuses only on the column variable as the explanatory factor). A value of 0 indicates no effect of the factor (column) variable on the response (row) variable, while values close to or exceeding 1 indicate a strong effect. Cohen's conventional benchmarks are provided in Table 1: small ($w = 0.10$), medium ($w = 0.30$), and large ($w = 0.50$).

TABLE 1: Equivalents of $w_{c,\text{Asym}}$ and Cramer's V

Effect size	Very small	Small	Medium	Large	Very large
$w^*_{c,\text{Asym}}$	less than 0.1	0.100	0.300	0.500	more than 0.5
Cramer's V ($c = 2$)	less than 0.1	0.100	0.300	0.500	more than 0.5
Cramer's V ($c = 3$)	less than 0.071	0.071	0.212	0.354	more than 0.354
Cramer's V ($c = 4$)	less than 0.058	0.058	0.173	0.289	more than 0.289
Cramer's V ($c = 5$)	less than 0.050	0.050	0.150	0.250	more than 0.250
Cramer's V ($c = 6$)	less than 0.045	0.045	0.134	0.224	more than 0.224

* The equivalence of $w_{c,Asym}$ holds only for 2×2 tables, as indicated by Cohen (1988, p. 221) in the case of the (symmetric) w.

6. Converting Asymmetric Cohen's w into Asymmetric Cramer's V

Effect sizes are generally considered more interpretable when scaled to a $[0, 1]$ range. This normalization can be achieved by exploiting the mathematical relationship between Cramer's V and Cohen's w:

$$\text{Cramer's } V = \frac{w_{\text{Sym}}}{\sqrt{k-1}}, \tag{15}$$

where w_{Sym} denotes the standard Cohen's w and k represents the smaller dimension of the contingency table, i.e., the minimum of the number of rows or columns.

As Cohen (1988, p. 221), the maximum attainable value of w is $\sqrt{k-1}$. Thus, Cramer's V can be understood as a normalization of w to the range $[0, \sqrt{k-1}]$.

However, for the asymmetric Cramer's V, an additional adjustment is necessary to ensure proper scaling to the $[0, 1]$ interval:

$$V_{c,\text{Asym}} = \frac{w_{c,\text{Asym}}}{\sqrt{\frac{1-\left(\frac{mr_i}{n}\right)_{\min}}{\left(\frac{mr_i}{n}\right)_{\min}} \cdot c}} \tag{16}$$

- $w_{c,\text{Asym}}$ is the asymmetric Cohen's w coefficient, treating the columns as the predictor (factor) variable.
- $\left(\frac{mr_i}{n}\right)_{\min}$ is the minimum ratio of the sum of observed frequencies in row i to the total sample size n.
- n is the total number of cases.
- c is the number of columns.

Consequently, the asymmetric Cramer's V is normalized to the $[0, 1]$ interval.

The interpretation follows Cohen's original heuristic conventions, which proposed three approximate cutoff points for 2×2 tables: *small* for values around 0.1, *medium* for values around 0.3, and *large* for values around 0.5. Additional categories (*very small* < 0.1 and *very large* > 0.5) constitute personal adaptations extending this logical framework, consistent with similar scales found in the literature. Furthermore, as the number of columns c in the table increases, the cutoff values are adjusted by dividing the original thresholds by $\sqrt{c-1}$.

7. Asymmetric Chi-square

The relationship between chi-square and Cohen's w –like in Equation (12)– is:

$$\chi^2 = w^2\, n. \tag{17}$$

We now extend this relationship to the asymmetric case:

$$\chi^2_{c,Asym} = w^2_{c,Asym} n. \tag{18}$$

The formal derivation stems directly from the asymmetric Cohen's w (see Equation (14)):

$$\chi^2_{c,Asym} = \left(\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(\frac{o_{ij}}{mc_j} - \frac{mr_i}{n}\right)^2}{\frac{mr_i}{n}} \right) n. \tag{19}$$

Unlike the traditional chi-square test of independence, which evaluates global independence between variables, this asymmetric approach computes the statistic under an alternative null hypothesis based on the conditional distribution of the dependent variable.

The degrees of freedom (df) are obtained as $(r-1)(c-1)$, as indicated in Equation (6). This result follows from the logic of the test: because the statistic compares the row proportions within each column to the overall row marginal distribution, which is estimated from the data. Column totals are treated as fixed. Under this fixed-columns design, each of the c columns contributes $(r-1)$ degrees of freedom, since the observed frequencies within a column must sum to its predetermined total. However, the marginal row proportions $\frac{mr_i}{n}$ are estimated

from the data, which incurs an additional loss of $(r-1)$ degrees of freedom. Consequently, the total degrees of freedom for the asymmetric statistic are given by $c(r-1)-(r-1)=(r-1)(c-1)$.

This approach differs fundamentally from the standard test of homogeneity, which requires independent samples for each subgroup, marginal totals are irrelevant, and tests against an unspecified common population distribution. In contrast, the asymmetric chi-square evaluates whether each column's distribution conforms to the observed marginal distribution of the population. This makes it particularly suitable for assessing the effect of an independent variable on a dependent variable in representative sampling contexts. By explicitly conditioning on the observed population distribution, this measure also helps address the documented misuse of traditional chi-square tests in analyzing variable relationships (see Franke et al., 2012).

8. Confidence Intervals of Asymmetric Cohen's w

As with the symmetric Cohen's w, confidence intervals for the asymmetric coefficient can be obtained through several techniques, including the bootstrap resampling, permutation method, the non-central chi-square parameter procedure, and the delta method. Here, the non-central chi-square approach is proposed as likely the most accurate, although further testing and validation are needed (Grissom & Kim, 2012; Kelly, 2007; Steiger, 2004).

9. A Simulation Study Comparing Association Coefficients and Effect Sizes Across Different Frequency Distributions in Contingency Tables

To illustrate how association coefficients and effect sizes are calculated in contingency tables, we use 3×3 tables with identical total sample sizes but different frequency distributions between main diagonal and off-diagonal cells. This design enables us to examine how association coefficients and effect sizes vary as frequencies progressively shift from off-diagonal to main diagonal positions.

For each pair of values in the "*Diagonal - Remaining cells*" column of Table 2 (e.g., 14–14, 16–13, etc.), a 3×3 contingency table is constructed as follows:

The cells on the main diagonal (positions (1,1), (2,2), and (3,3)) take the value indicated on the left-hand. The off-diagonal cells take the value indicated on the right-hand. Thus, each case produces a table with identical total frequency but a different distribution patterns between diagonal and off-diagonal cells. It should also be noted that, since the variables are nominal, the diagonal structure is arbitrary; any rearrangement of categories preserves the same underlying statistical relationships.

TABLE 2: Comparison of symmetric and asymmetric chi-square for different diagonal-remaining cells in 3×3 tables.

Frequencies Diagonal – Remaining cells	Symmetric chi-square	p symmetric	Asymmetric chi-square	p asymmetric
14–14	0.00	1.000	0.00	1.000
16–13	1.29	0.864	3.86	0.426
18–12	5.14	0.273	15.43	0.004
20–11	11.57	0.021	34.71	0.000
22–10	20.57	0.000	61.71	0.000
24–9	32.14	0.000	96.43	0.000
26–8	46.29	0.000	138.86	0.000
28–7	63.00	0.000	189.00	0.000
30–6	82.29	0.000	246.86	0.000
32–5	104.14	0.000	312.43	0.000
34–4	128.57	0.000	385.71	0.000
36–3	155.57	0.000	466.71	0.000
38–2	185.14	0.000	555.43	0.000
40–1	217.29	0.000	651.86	0.000
42–0	252.00	0.000	756.00	0.000

This analysis considers two scenarios. The first, termed the *symmetric chi-square*, corresponds to the classical test of independence, where C denotes the column variable and R the row variable, evaluating the alternative hypothesis of mutual association between the variables (Formula 4).

The second scenario, termed *asymmetric chi-square*, refers to the hypothesis test in which C is considered the predictor of R, assessing the association in a specific direction and quantifying the dependence of the rows on the columns (Formula 20).

In both scenarios, the chi-square statistic increases as frequencies concentrate in the diagonal cells.

However, this increase is markedly more pronounced in the asymmetric chi-square than in its symmetric counterpart.

Crucially, applying the symmetric test to a hypothesis specifying a directional relationship, where C functions as a predictor of R, would be methodologically inappropriate. In such cases, the asymmetric chi-square should be employed to correctly evaluate the directed dependence between the variables.

Once again, we face two possible scenarios. If variables C and R mutually influence each other, the symmetric Cohen's w is appropriate. However, if the hypothesis posits that C affects R, one of the asymmetric versions should be used.

In this context, we aim to examine the behavior of the new asymmetric Cohen's w (Formula 14), or its equivalent, asymmetric Cramer's V (according to formula 16), in comparison to the previously mentioned Goodman and Kruskal's Lambda and Tau coefficients.

Table 3: Association measures varying diagonal concentrations in 3×3 tables

Frequencies Diag.–Rem. cells	Sym. Cohen's w	Asym. Cohen's w	Goodman & Kruskal's λ	Goodman & Kruskal's τ	Asym. Cramer's V	Interpretation asym. V
14–14	0.00	0.00	0.00	0.00	0.00	null/very small
16–13	0.10	0.17	0.07	0.01	0.07	small
18–12	0.20	0.35	0.14	0.02	0.14	small
20–11	0.30	0.52	0.21	0.05	0.21	medium
22–10	0.40	0.70	0.29	0.08	0.29	medium
24–9	0.51	0.87	0.36	0.13	0.36	large
26–8	0.61	1.05	0.43	0.18	0.43	very large
28–7	0.71	1.22	0.50	0.25	0.50	very large
30–6	0.81	1.40	0.57	0.33	0.57	very large
32–5	0.91	1.57	0.64	0.41	0.64	very large
34–4	1.01	1.75	0.71	0.51	0.71	very large
36–3	1.11	1.92	0.79	0.62	0.79	very large
38–2	1.21	2.10	0.86	0.74	0.86	very large
40–1	1.31	2.27	0.93	0.86	0.93	very large
42–0	1.41	2.45	1.00	1.00	1.00	perfect

We observe that the Lambda coefficients are practically identical, equal in this case due to rounding, whereas the Tau coefficient shows a similar pattern at the extremes but yields considerably lower values in the intermediate range. This is consistent with existing literature, which notes that Tau tends to underestimate association.

In conclusion, the asymmetric Cohen's w, or its equivalent asymmetric Cramer's V, performs better than Tau, but it does not appear to offer a distinct advantage over Lambda in the scenarios examined.

We present a comparison analogous to that in Table 2, now applied to 3×3 tables with a constant total sample size. In this design, a single cell accumulates the frequencies systematically reallocated from all other cells.

The initial configuration (14–14) distributes frequencies uniformly across all cells. In the subsequent scenario (22–13), each of the eight remaining cells contributes one frequency unit (reducing their counts from 14 to 13), which are pooled into the target cell, raising its count from 14 to 22.

This reallocation scheme is repeated across the different scenarios, progressively concentrating frequencies into a single cell and generating increasingly asymmetric distributions.

Once again in this case, we observe that in both approaches, the chi-square coefficient increases as frequencies become concentrated in a single cell.

However, this increase is markedly more pronounced for the asymmetric chi-square than for its symmetric counterpart, consistent with the pattern observed in Table 2.

TABLE 4: Comparison of symmetric and asymmetric chi-square for different one cell-remaining cells in 3×3 tables.

Frequencies One cell – Remaining cells	Symmetric chi-square	p symmetric	Asymmetric chi-square	p asymmetric
14–14	0.00	1.000	0.00	1.000
22–13	1.97	0.742	5.62	0.229
30–12	6.22	0.183	17.63	0.001
38–11	11.34	0.023	33.09	0.000
46–10	16.66	0.002	51.80	0.000
54–9	21.88	0.000	74.74	0.000
62–8	26.84	0.000	103.75	0.000
70–7	31.50	0.000	141.75	0.000
78–6	35.84	0.000	193.54	0.000
86–5	39.87	0.000	267.61	0.000
94–4	43.60	0.000	380.85	0.000
102–3	47.06	0.000	572.51	0.000
110–2	50.26	0.000	960.24	0.000
118–1	53.24	0.000	2132.06	0.000
126–0	56.00	0.000	$> 100\,000\,000$	0.000

As in previous analyses, we consider two distinct scenarios. If variables C and R exhibit mutual dependence, the symmetric Cohen's w is the appropriate measure. Conversely, if the hypothesis posits a directional effect of C on R, one of the asymmetric versions should be applied. As illustrated in Table 3, the two coefficients yield distinct values under these different conditions.

An examination of the behavior of the new asymmetric Cohen's w, and its equivalent, the asymmetric Cramer's V (Formula 16), in comparison with Goodman and Kruskal's Lambda and Tau coefficients reveals distinct patterns. Lambda fails entirely in this scenario, returning values of zero throughout, a limitation consistent with its documented methodological constraints. Tau displays a similar pattern at the distribution extremes but yields substantially lower values across the intermediate range, aligning with existing literature noting its tendency to underestimate association strength.

In conclusion, the asymmetric Cohen's w, and its equivalent asymmetric Cramer's V, not only outperforms Tau but, crucially, remains functional in scenarios where Lambda fails entirely. Thus, while the new measure may behave similarly to Lambda in some contexts, it offers superior robustness, providing meaningful results across a wider range of distributional patterns.

The interpretation of the asymmetric Cramer's V coefficients presented in Tables 3 and 5 follows the guidelines in Table 1, specifically using the reference values for $c = 3$.

TABLE 5: Association measures varying one cell concentrations in 3×3 tables

Frequencies One cell–Rem. cells	Sym. Cohen's w	Asym. Cohen's w	Goodman & Kruskal's λ	Goodman & Kruskal's τ	Asym. Cramer's V	Interpretation asym. V
14–14	0.00	0.00	0.00	0.00	0.00	null/very small
22–13	0.13	0.21	0.00	0.01	0.08	small
30–12	0.22	0.37	0.00	0.03	0.14	small
38–11	0.30	0.51	0.00	0.05	0.18	small
46–10	0.36	0.64	0.00	0.08	0.21	large
54–9	0.42	0.77	0.00	0.11	0.23	large
62–8	0.46	0.91	0.00	0.14	0.25	large
70–7	0.50	1.06	0.00	0.17	0.27	large
78–6	0.53	1.24	0.00	0.19	0.29	large
86–5	0.56	1.46	0.00	0.22	0.31	large
94–4	0.59	1.74	0.00	0.25	0.33	large
102–3	0.61	2.13	0.00	0.27	0.34	large
110–2	0.63	2.76	0.00	0.29	0.36	very large
118–1	0.65	4.11	0.00	0.31	0.37	very large
126–0	0.67	136.62	0.00	0.33	0.38	very large

10. Conclusions

Although association coefficients such as Goodman and Kruskal's Lambda and Tau are widely employed, they present significant limitations in capturing asymmetric relationships in contingency tables. Lambda, which quantifies the proportional reduction in prediction error using modal categories, can produce values of zero despite non-trivial associations, particularly in tables with skewed marginal distributions. Tau, which measures proportional reduction in error based on overall variability, tends to underestimate association strength in the mid-range of its values. Similarly, Theil's uncertainty coefficient (U) behaves comparably to Tau, quantifying uncertainty rather than effect size, and can be difficult to interpret. Consequently, exclusive reliance on these conventional measures may obscure meaningful dependencies or misrepresent the true magnitude of associations.

To overcome these limitations, we develop an asymmetric extension of Cohen's w alongside its corresponding directional hypothesis test. This effect size quantifies the strength of directional influence between nominal-level categorical variables, while the companion test assesses its statistical significance without assuming variable interchangeability. The resulting framework provides a more accurate and interpretable method for analyzing directional dependencies, particularly in scenarios where traditional symmetric measures are inadequate.

The proposed method can be applied in two primary contexts. First, in observational studies, which often involve samples collected at a single time point with only the total sample size controlled, symmetric association coefficients cannot capture directional effects because they inherently distribute associations reciprocally between variables. The asymmetric framework presented here allows researchers

to quantify both the presence and magnitude of directional effects with greater accuracy, thereby enhancing interpretability and enabling more meaningful cross-study comparisons. Results in these cases should not be interpreted as conclusive evidence of causality; hypothesis test outcomes should always be reported alongside effect sizes and confidence intervals.

Second, in experimental designs with controlled group assignment, such as clinical trials or intervention studies where the direction of influence is explicitly defined, asymmetric coefficients provide a robust framework for evaluating effects in the intended direction. These coefficients accurately attribute effects from the manipulated independent variable to the measured outcome variable. In such cases, if the experiment is properly designed and other variables are adequately controlled, both statistical significance and effect size can be attributed to the experimental variable, yielding precise estimates consistent with contemporary best practices in statistical inference.

Furthermore, the framework clarifies the choice of analytical tools: while the classical chi-square test of homogeneity remains appropriate for comparing multiple samples, the asymmetric chi-square test and Cohen's asymmetric w are better suited for analyzing the effect of one nominal variable on another within a single sample.

In summary, this work introduces a refined methodological framework consisting of two key components: 1) asymmetric effect size measures that quantify the strength of directional relationships between nominal variables, and 2) corresponding hypothesis tests that evaluate the statistical significance of these directional effects. Together, these components provide a comprehensive approach for analyzing nominal data within asymmetric or conditional hypotheses. This approach enhances interpretability, supports more robust statistical inference, and offers a versatile solution applicable across diverse research designs.

This framework, centered on the adoption of asymmetric Cohen's w and its associated chi-square test, bridges a critical gap between statistical methodology and substantive theory. It enables a more nuanced analysis in which detecting, quantifying, and communicating effects aligns directly with theoretical predictions of directionality. Consistent with this approach, the reporting of effect sizes alongside their confidence intervals promotes transparency, facilitates cross-study comparison, and ensures that empirical findings are presented with both clarity and statistical rigor.

[Received: February 2025 — Accepted: November 2025]

References

Agresti, J. (2002), *Categorical Data Analysis*, John Wiley & Sons.

Algina, J., Keselman, H. J. & Penfield, R. D. (2006), 'Confidence interval coverage for Cohen's effect size statistic', *Educational and Psychological Measurement* **66**(6), 945–960. Sage Publications.

Ben-Shachar, M. S., Patil, I., Thériault, R., Wiernik, B. M. & Lüdecke, D. (2023), 'Phi, Fei, Fo, Fum: Effect sizes for categorical data that use the chi-squared statistic', *Mathematics* **11**(9).

Berry, K. J., Johnston, J. E. & Mielke, J. P. (2018), *The Measurement of Association. A Permutation Statistical Approach*, Springer.

Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. (2009), *Introduction to Meta-Analysis*, John Wiley & Sons, Ltd.

Cochran, W. G. (1952), 'The χ^2 test of goodness of fit', *Annals of Mathematical Statistics* **23**, 315–345.

Cochran, W. G. (1954), 'Some methods for strengthening the common chi-squared tests', *Biometrics* **10**, 417–451.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates.

Crack, T. F. (2018), 'A note on Karl Pearson's 1900 chi-squared test: Two derivations of the asymptotic distribution, and uses in goodness of fit and contingency tests of independence, and a comparison with the exact sample variance chi-square result'.

Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press.

DeGroot, M. H. (1988), *Probabilidad y estadística*, Addison Wesley.

Edgington, E. S. & Onghena, P. (2007), *Randomization Tests*, Chapman & Hall/CRC.

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

Fleiss, J., Levin, B. & Paik, M. C. (2003), *Statistical Methods for Rates and Proportions*, Wiley.

Franke, T. M., Ho, T. & Christie, C. A. (2012), 'The chi-square test: Often used and more often misinterpreted', *American Journal of Evaluation* pp. 448–458.

Good, P. (2005), *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, Springer.

Goodman, L. A. & Kruskal, W. H. (1963), 'Measures of association for cross classifications iii: Approximate sampling theory', *Journal of the American Statistical Association* **58**(302), 310–364.

Grissom, R. J. & Kim, J. J. (2012), *Effect Sizes for Research: Univariate and Multivariate Applications*, 2 edn, Routledge, New York, NY.

Haddock, C. K., Rindskopf, D. & Shadish, W. R. (1998), 'Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues', *Psychological Methods* **3**(3).

Jané, M. B., Ben-Shachar, M. S., Moreau, D., Steele, J., Qinyu, X., Caldwell, A. R. & Zloteanu, M. (2024), 'Guide to effect sizes and confidence intervals'.

Kelly, K. (2007), 'Confidence intervals for standardized effect sizes: Theory, application, and implementation', *Journal of Statistical Software* **20**(8), 1–24.

Lock, R. H., Lock, P. F., Lock Morgan, K., Lock, E. F. & Lock, D. F. (2021), *Statistics: Unlocking the Power of Data*, Wiley.

Mangiafico, S. (2023), *An R Companion for the Handbook of Biological Statistics*.

Pearson, K. (1900), 'On the probability that two independent distributions of frequency are really samples from the same population', *Biometrika* **8**(1), 250–254.

Pearson, K. (1904), 'On the theory of contingency and its relation to association and normal correlation', *Biometric* pp. 1–34. `https://ia801300.us.archive.org/8/items/cu31924003064833/cu31924003064833.pdf`

Rita, H. & Komonen, A. (2008), 'Odds ratio: an ecologically sound tool to compare proportions', *Ann. Zool. Fennici* **45**(1), 66–72.

Rosenthal, R. (1994), Parametric measures of effect size, *in* H. Cooper & L. V. Hedges, eds, 'The Handbook of Research Synthesis', Russell Sage Foundation, p. 239.

Sanchez-Meca, J., Marín-Martínez, F. & Salvador Chacón-Moscoso, S. (2003), 'Effect-size indices for dichotomized outcomes in meta-analysis', *Psychological Methods* **8**(4).

Steiger, J. H. (2004), 'Beyond the f test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis', *Psychological Methods* **9**(2), 164–182.

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

Yule, G. U. (1911), *An Introduction to the Theory of Statistics*, C. Griffin and company.

Zaiontz, C. (2024), 'Real statistics using excel'. https://real-statistics.com/