Review Article

Chi-square test applications

Parisa Ataee Dizaji [1] , Fatemeh Heidary [2] and Reza Gharebaghi [2]

[1] *Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran*
[2] *IVORC Academic Foundation (International Virtual Ophthalmic Research Center), Texas, United States*

ABSTRACT

Background: The chi-squared (χ^2) test is a fundamental non-parametric statistical method. It is widely employed in clinical, epidemiological, and biomedical research, including ophthalmology and optometry. It is useful for testing hypotheses regarding independence between categorical variables and for assessing the goodness-of-fit of observed categorical frequencies to a specified expected distribution. In this review, we present a thorough examination of the statistical principles and clinical relevance of the χ^2 test, focusing on its application in vision science and related research domains.

Methods: We outline the conceptual framework and methodological steps for conducting the χ^2 test, emphasizing two common applications considered in this review: the goodness-of-fit test and the test of independence. We discuss key assumptions, such as the independence of observations, use of frequency data, and minimum expected cell counts in detail. Moreover, we explain the process of calculating degrees of freedom (*df*) and interpreting results based on critical values from the χ^2 distribution. Additionally, appropriate measures of effect size, i.e., Phi (φ) for 2 × 2 tables and Cramer's V for larger tables, for assessing association strength, are introduced. To contextualize its clinical relevance, we present four examples from ophthalmology and optometry.

Results: In the first example, the association between vision impairment (VI) category and sex was examined using a 6 × 2 contingency table. Using unrounded expected frequencies, the Pearson χ^2 statistic was 5.31 with 5 *df*; however, two expected cell frequencies were < 1, indicating that an exact procedure is preferable for formal inference. Cramer's V was 0.097, indicating a negligible association. The second example examined the association between age category and first-year persistence with antiglaucoma therapy. Here, $\chi^2 = 6.18$ ($df = 2$, $P = 0.045$), indicating a statistically significant association, although Cramer's V was negligible (0.054). In the third example, a 2 × 2 table was used to examine the association between sex and the type of anti-vascular endothelial growth factor injection (aflibercept or ranibizumab). This yielded $\chi^2 = 0.027$ ($df = 1$, $P = 0.870$) and $\varphi = 0.017$, indicating no statistically significant evidence of association and a negligible effect. In the goodness-of-fit example, the observed contact-lens distribution differed statistically from the pooled historical reference proportions used in the analysis ($\chi^2 = 3.99$, $df = 1$, $P = 0.046$).

Conclusions: The χ^2 test is a widely used tool for analyzing categorical data, enabling clinicians and researchers to assess potential associations between variables. However, its reliability depends on its proper application, including verification of assumptions and appropriate interpretation of effect sizes, along with consideration of statistical significance. In clinical disciplines, such as ophthalmology or optometry, understanding and utilizing the χ^2 test enhances research rigor and the validity of research findings, facilitating better-informed decisions in patient care and in program development.

KEYWORDS

chi-square test, categorical data, contingency tables, goodness-of-fit, effect size, ophthalmology, optometry

Correspondences: Parisa Ataee Dizaji, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. Email: p.ataie.593@gmail.com. ORCID iD: https://orcid.org/0000-0001-8664-1412.

How to cite this article: Ataee Dizaji P, Heidary F, Gharebaghi R. Chi-square test applications. Med Hypothesis Discov Innov Optom. 2025 Winter; 6(4): 150-159. DOI: https://doi.org/10.51329/mehdioptometry234.

Received: 13 June 2025; Accepted: 19 December 2025

INTRODUCTION

The chi-square (χ^2) test is a widely used, non-parametric statistical method for analyzing categorical data [1]. In this review, we focus on two common applications: evaluating the goodness-of-fit between observed and theoretical (expected) distributions and testing independence between two categorical variables in a contingency table [2–5]. The Pearson χ^2 test of homogeneity uses the same test statistic but addresses equality of categorical distributions across independently sampled groups and is not discussed separately here [4].

Categorical data are a fundamental data type encountered in experimental and clinical sciences. These data represent the counts of observations that are classified into qualitative categories and are often arranged in contingency tables to facilitate analysis. Such tables are conventionally structured as r × c matrices, where r denotes the number of rows and c represents the number of columns [6, 7], corresponding to the levels of the variables under study. In particular, the Pearson χ^2 test is widely used for testing associations within such tables [8, 9]. Under simple sampling with independent observations, the Pearson χ^2 statistic is equivalent to the score test statistic for independence in an r × c contingency table [9].

Originally introduced by Karl Pearson in 1900 to assess the goodness-of-fit of observed frequency distributions, the χ^2 test was subsequently extended, in 1904, for assessing the independence of categorical variables arranged in contingency tables [4]. Since then, the χ^2 family of tests has become integral to hypothesis testing in biomedical, psychological, and social sciences [4, 10]. Among the strengths of this test are its simplicity and broad applicability, but the validity of its usual asymptotic approximation depends on adequately large expected cell frequencies rather than on the total sample size alone [1, 11].

In this review, we aim to provide a comprehensive overview of the statistical foundations and practical applications of the χ^2 test, with a particular focus on its use in vision research. Using examples drawn from the published literature, we illustrate the relevance and utility of χ^2 methodologies in analyzing categorical data in clinical settings. The definitions and key concepts pertinent to the application of χ^2 tests in this review are presented in Table 1.

METHODS

This article is a narrative methodological review rather than a systematic review. References and clinical examples were selected purposively to illustrate established χ^2 concepts and applications in vision research; no systematic literature search or evidence synthesis was undertaken.

Assumptions of the X^2 Test

Although the χ^2 test is a non-parametric procedure, its validity and interpretation depend on assumptions concerning the sampling design, independence of observations, and expected cell frequencies. The key considerations are as follows.

1. Sampling and Scope of Inference: Observations should arise from a sampling or allocation process appropriate to the intended inference, and each observation should contribute independently to the table. Random sampling supports population-level generalizability; when convenience, clinical, or registry samples are used, the scope of inference should be interpreted accordingly [1, 14].

2. Frequency Data: The values entered into the cells of the contingency table must represent raw frequencies or counts of cases, and not percentages, proportions, or any other transformed data [1, 15].

3. Variable Types: For the χ^2 test of independence, both variables under investigation must be categorical and may be nominal or ordinal [1, 2, 15]. For the goodness-of-fit test, the analyzed variable must be categorical. For ordinal variables, the Pearson χ^2 test assesses general association without using the natural ordering of the categories; if an ordered or linear trend is the primary hypothesis, an appropriate trend test should be considered [2, 7, 16, 17].

4. Independence of Observations: All observations must be independent of one another. The χ^2 test is not appropriate if the groups being compared are related or matched (e.g., paired samples); in such cases, an alternative test should be used [2, 15].

5. Expected Frequencies: At least 80% of the cells in the contingency table should have expected frequencies of ≥ 5, and no cell should have an expected frequency < 1 [1, 11, 15].

Table 1. Definitions of key terms relevant to the chi-square (χ^2) test

Term	Definition
Categorical data	Data in which observations are classified into discrete, mutually exclusive categories, typically representing qualitative attributes. Analysis involves counting the number of observations within each category [6, 7].
Observed frequency	The actual number of cases or observations recorded in each cell of a contingency table [1, 6, 7].
Expected frequency	The theoretically calculated number of cases that would be expected in each cell of a contingency table if the null hypothesis (e.g., independence of variables) were true [6, 7, 12].
Degrees of freedom	For a Pearson χ^2 test of independence in an r × c contingency table, degrees of freedom (df) = (r − 1) × (c − 1), where r and c are the numbers of row and column categories, respectively [11]. For a goodness-of-fit test with k categories and fully prespecified expected probabilities, $df = k - 1$; when p independent parameters defining those probabilities are estimated from the same data under the usual regularity conditions, df is typically $k - 1 - p$ [13].

Alternatives to the X^2 Test When Assumptions Are Violated

For 2 × 2 contingency tables, when expected cell frequencies are too small for the usual χ^2 approximation to be reliable, Fisher's exact test provides an exact alternative for assessing association [11, 12]. In particular, concern arises when more than 20% of cells have expected frequencies < 5 or any expected cell frequency is < 1 [1, 11]. Fisher's exact test is classified as an exact test, in contrast to the χ^2 test, which is based on approximations [11, 12]. For contingency tables larger than 2 × 2 with sparse expected frequencies, exact methods such as the Fisher–Freeman–Halton test, with Monte Carlo estimation when necessary, should be considered [18, 19]. The usual asymptotic likelihood-ratio χ^2 test is also based on an asymptotic approximation and should not be regarded as an automatic solution to sparse expected counts; exact likelihood-ratio procedures are separate methods [18, 19].

Software Tools Supporting the X^2 Test

Various statistical software packages and programming environments support χ^2 analyses, including IBM SPSS Statistics, R, and Python with appropriate statistical libraries [20–22].

Types of X^2 Tests

As stated earlier, this review focuses on two common applications of the Pearson χ^2 statistic [15, 23, 24]:

1. To evaluate whether the observed distribution of a categorical variable conforms to a specified expected distribution (goodness-of-fit test) [15, 23].
2. To test whether two categorical variables are independent in a contingency table (test of independence) [15, 23].

The closely related χ^2 test of homogeneity compares categorical distributions across independently sampled groups but is not discussed separately here [4].

1. X^2 Goodness-of-Fit Test

The χ^2 goodness-of-fit test is used to assess whether the observed distribution (frequencies) of a single categorical variable conforms to an expected distribution [4, 5, 23, 24]. This type of test involves only one variable with multiple levels.

The Pearson χ^2 test is widely used in this context. When the expected probabilities for k categories are fully specified independently of the observed sample, the test statistic has an approximate χ^2 distribution with $k - 1$ degrees of freedom (*df*) as the sample size becomes large [13]. If p independent parameters defining those probabilities are estimated from the same data under the usual regularity conditions (e.g., by maximum likelihood), the *df* are typically reduced to $k - 1 - p$ [13].

- H_0 (null hypothesis): The population category probabilities equal the specified expected probabilities [4, 23].
- H_1 (alternative hypothesis): At least one population category probability differs from its specified expected probability [4, 23].

2. X^2 Test of Independence

The χ^2 test of independence is used to assess whether two categorical variables are statistically associated. Specifically, it examines whether the distribution of one categorical variable differs across the levels of the other [1, 4, 23].

The χ^2 statistic is calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \quad (1),$$

where O_i and E_i are the observed and expected frequencies, respectively, in cell i, and K is the total number of cells included in the calculation

The test involves calculating the difference between the observed and expected values in each cell, squaring the differences to remove negative values, and dividing this value by the expected frequency to normalize the contribution of each cell. Summing these normalized values yields the test statistic, which is then compared to a critical value from the χ^2 distribution with *df* equal to (r − 1) × (c − 1), where r is the number of rows and c is the number of columns [4, 17, 23, 25].

- H_0: The two categorical variables will be independent [4, 23].
- H_1: The two categorical variables will be associated [4, 23].

The χ^2 test statistic is interpreted as follows [26]: A sufficiently large discrepancy between the observed and expected frequencies produces a large χ^2 statistic and may lead to rejection of the null hypothesis [26]. Conversely, a small χ^2 statistic provides less evidence against the null hypothesis [26]; failure to reject H_0 should not be interpreted as proof that H_0 is true [27].

The expected frequency of each cell in the contingency table is calculated by using the following formula:

$$E_{ij} = \frac{n_{i+}\, n_{+j}}{N} \quad (2),$$

where E_{ij} is the expected frequency in row i, column j, n_{i+} is the corresponding row total, n_{+j} is the corresponding column total, and N is the grand total number of observations [12, 26].

Strength of Association in X^2 Tests

When using contingency tables to analyze categorical variables, not only should statistical significance be assessed, but the strength of association should also be quantified. To this end, several effect size measures can be used, including the Cramer's V and Phi (φ) coefficient [1, 28, 29].

Cramer's V statistic is preferred for contingency tables larger than 2 × 2. This is interpreted similarly to a correlation coefficient, which ranges from 0 (complete independence) to 1 (complete dependence). Cramer's V remains interpretable regardless of table dimensions, and provides a standardized measure of the effect size that accounts for table size [1, 28, 30]. Cramer's V is calculated as follows [1, 28, 29]:

$$V=\sqrt{\frac{\chi^2}{N\times\min\ (r-1,\ c-1)}} \qquad (3),$$

where χ^2 is the Pearson χ^2 statistic, N is the total number of observations, and min (r – 1, c − 1) is the smaller of the row- and column-dimension terms.

For 2 × 2 contingency tables, the φ coefficient is commonly used to assess effect size. It is analogous to the Pearson correlation coefficient and provides a measure of the degree of association between two binary variables. Similar to Cramer's V, it is interpreted as a correlation coefficient, ranging from 0 (complete independence) to 1 (complete dependence). When φ is calculated from χ^2 as in Equation (4), it represents the non-negative magnitude of association between two binary variables, equivalent to the absolute magnitude of their Pearson correlation. This magnitude is unchanged when the rows and columns of a 2 × 2 table are exchanged; for larger contingency tables, Cramer's V is preferred [28–30].

For 2 × 2 tables, the φ coefficient is computed to indicate the effect size by using the following formula [28, 29]:

$$\varphi=\sqrt{\frac{\chi^2}{N}} \qquad (4)$$

Interpretation of Effect Size

The strength of the association can be interpreted as follows, based on the calculated effect size [28]:

- Effect size = 0 to < 0.10 → Negligible association
- Effect size = 0.1 to < 0.2 → Weak association
- Effect size = 0.20 to < 0.4 → Moderate association
- Effect size = 0.40 to < 0.60 → Relatively strong association
- Effect size = 0.60 to < 0.80 → Strong association
- Effect size = 0.80 to < 1.00 → Very strong association

Beyond statistical significance, these thresholds provide a descriptive guide to the magnitude of association; practical or clinical significance should still be judged in the context of the research question [1, 10, 28, 31].

Power of the X^2 Test

The statistical power of the χ^2 test depends on multiple factors, including the effect size, sample size, *df*, and the significance level (α) [1, 10, 28]. Adequate power is important for detecting associations of a specified magnitude when they exist and for reducing the probability of a Type II error.

Determining an appropriate sample size is a critical component of the design of a study. Power analysis techniques allow calculation of the minimum number of observations required to achieve a desired level of power for a given effect size and significance threshold [4, 10]. When power is insufficient, studies may fail to detect meaningful differences, leading to misleading conclusions [1].

A central element in the power analysis of categorical data is Cohen's w, a standardized effect size specific to χ^2 tests [10, 27, 32]. Cohen proposed conventional thresholds for small (w = 0.10), medium (w = 0.30), and large (w = 0.50) effect sizes. These benchmarks are widely used as conventional values for planning χ^2 analyses when a more context-specific effect size is unavailable [10].

RESULTS and DISCUSSION

Stepwise Application of the X^2 Test of Independence in Examples 1–3

To demonstrate the practical application of the χ^2 test of independence, we present three examples from vision research. These examples are didactic secondary calculations based on aggregate counts reported in the cited publications. The observed frequencies were obtained from those publications, whereas the expected frequencies, χ^2 statistics, *P*-values, and effect-size measures were independently calculated for this review; no individual-level data were reanalyzed.

Example 1: Association Between Vision Impairment and Sex

Vision impairment (VI) is a significant global public health issue, with an uneven prevalence across World Health Organization regions. Identifying the causes and impacts of VI is essential for targeting interventions, planning effective programs, and prioritizing resources [33]. In this example, we examined whether sex (a nominal variable) and VI, based on the International Classification of Diseases (an ordinal variable), were associated. This relationship was statistically tested using the χ^2 test of independence.

- H_0: VI and sex are independent.
- H_1: VI and sex are associated.

Table 2. Observed and expected frequencies in the category of vision impairment based on distance vision in the better eye with the best possible lens correction in patients referred to a low-vision rehabilitation clinic over 7 years

VI Category	Observed Frequency			Expected Frequency		
	Men	Women	Total	Men	Women	Total
1	42	32	74	44.113	29.887	74
2	171	108	279	166.317	112.683	279
3	55	31	86	51.266	34.734	86
4	46	45	91	54.247	36.753	91
5	23	13	36	21.460	14.540	36
6	1	0	1	0.596	0.404	1
Total	**338**	**229**	**567**	**338.000**	**229.000**	**567**

Note: Comparison between men and women. Categories 1–6 are based on the International Classification of Diseases-11 for mortality and morbidity statistics. Category 1, mild vision impairment; Category 2, moderate vision impairment; Category 3, severe vision impairment; Categories 4–6, blindness with better eye vision > +1.30 logMAR to no light perception. Observed frequencies were obtained from the source study [33]. Expected frequencies were calculated in this review from the marginal totals and are shown to three decimal places. Unrounded expected frequencies were used for all subsequent calculations.

Table 3. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
42	44.113	0.101
32	29.887	0.149
171	166.317	0.132
108	112.683	0.195
55	51.266	0.272
31	34.734	0.401
46	54.247	1.254
45	36.753	1.850
23	21.460	0.110
13	14.540	0.163
1	0.596	0.274
0	0.404	0.404
Total χ^2		**5.31**

Step 1: Present the Observed and Expected Frequencies

In Table 2, the observed and expected frequencies (calculated using Equation 2) for each category are presented, based on a dataset originating from patients with low vision who attended a rehabilitation clinic at a tertiary referral center [33].

Step 2: Calculate the Degrees of Freedom

The *df* for the χ^2 test in the contingency table was calculated using the following formula: *df* = (number of rows − 1) × (number of columns − 1). In this example, the table consists of six rows (VI categories) and two columns (sex categories) and excludes the totals. Therefore: *df* = (6 − 1) × (2 − 1) = 5.

Step 3: Compute the X^2 Statistic

Using Equation (1) and the unrounded expected frequencies shown in Table 3, the Pearson χ^2 statistic was 5.31, with *df* = 5.

Step 4: Determine Statistical Significance

At α = 0.05, the asymptotic critical value of the χ^2 distribution for 5 *df* is 11.07 [23]. However, two expected cell frequencies in Category 6 were < 1 (0.596 and 0.404), indicating that the expected-frequency criterion for the usual Pearson χ^2 approximation was not satisfied. Therefore, the Pearson χ^2 statistic is presented for illustrative purposes, whereas an exact test for the 6 × 2 contingency table, such as the Fisher–Freeman–Halton test with Monte Carlo estimation when necessary, is preferable for formal inference. A Monte Carlo estimate for the Fisher–Freeman–Halton exact test did not provide evidence of an association between VI category and sex ($P \approx 0.375$).

Step 5: Evaluate the Strength of Association

Given that the contingency table was larger than 2 × 2, Cramer's V was used to describe the magnitude of association. Using the corrected Equation (3), $V = \sqrt{\frac{5.306}{567 \times 1}} \approx 0.097$, indicating a negligible association according to the effect-size classification used in this review.

Example 2: Association Between Age and First-Year Persistence with Antiglaucoma Therapy

Age, treatment regimen, and adverse effects have been associated with persistence or discontinuation of antiglaucoma therapy. Maintaining long-term persistence with therapy is important for reducing the risk of disease progression and irreversible vision loss [34]. In this context, persistence was defined as the proportion of patients who continued taking any antiglaucoma medication, regardless of any changes in the specific drug, by 1 year after treatment initiation [34].

In this example, we evaluated whether persistence with antiglaucoma medication in the first year (a nominal variable) was associated with age-group (an ordinal variable). This relationship was tested using a χ^2 test for independence.

- H_0: Persistence rate and age-group are independent.
- H_1: Persistence rate and age-group are associated.

Table 4. Observed and expected frequencies by age group and first-year persistence with antiglaucoma drug use

Age-group (years)	Observed Frequency			Expected Frequency		
	Persistent	Non-persistent	Total	Persistent	Non-persistent	Total
18–44	64	9	73	67.093	5.907	73
45–64	420	48	468	430.131	37.869	468
≥65	1481	116	1597	1467.776	129.224	1597
Total	**1965**	**173**	**2138**	**1965.000**	**173.000**	**2138**

Note: Observed frequencies were obtained from the source study [34]; Expected frequencies were calculated from the marginal totals and are shown to three decimal places; unrounded expected frequencies were used for all subsequent calculations.

Table 5. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
64	67.093	0.143
420	430.131	0.239
1481	1467.776	0.119
9	5.907	1.620
48	37.869	2.710
116	129.224	1.353
Total χ^2		**6.184**

Step 1: Present the Observed and Expected Frequencies

Table 4 presents the observed and expected frequencies (calculated using Equation 2) for each category. The data involve patients categorized by age (18–44 years, 45–64 years, and 65 years or older) and their persistence with antiglaucoma therapy during the first year of follow-up [34].

Step 2: Calculate the Degrees of Freedom

The *df* for this contingency table are computed as follows:

df = (number of rows − 1) × (number of columns − 1)

In this example, the table consists of three age groups (rows) and two persistence categories (columns), excluding the totals. Thus: *df*= (3 − 1) × (2 − 1) = 2

Step 3: Compute the X^2 Statistic

Applying Equation (1) to the unrounded expected frequencies, the Pearson χ^2 statistic was 6.18, with *df* = 2 (*P* = 0.045) (Table 5):

Step 4: Determine Statistical Significance

At $\alpha = 0.05$, the critical value for 2 *df* is 5.99 [23]. Since the computed χ^2 statistic of 6.18 exceeds 5.99 (*P* = 0.045), the null hypothesis is rejected. Thus, age group and first-year therapy persistence were statistically significantly associated in this sample.

Step 5: Evaluate the Strength of Association

As the contingency table was 3 × 2, Cramer's V was used to quantify the magnitude of association. Using the corrected Equation (3), $V = \sqrt{\frac{6.184}{[2138 \times 1]}} \approx 0.054$, indicating a negligible association according to the classification used in this review. Thus, although the association was statistically significant, its magnitude was negligible.

Example 3: Association Between Sex and Intravitreal Injection Type for Treating Diabetic Macular Edema (DME)

Vascular endothelial growth factor (VEGF) plays a central role in retinal barrier disruption and DME development. Although laser photocoagulation was historically an important treatment for DME, intravitreal anti-VEGF agents such as aflibercept and ranibizumab are now widely used [35].

Table 6. Observed and expected frequencies in sex categories with intravitreal injections of aflibercept or ranibizumab

Intravitreal injections	Observed Frequency			Expected Frequency		
	Male	Female	Total	Male	Female	Total
Aflibercept	14	39	53	14.354	38.646	53
Ranibizumab	12	31	43	11.646	31.354	43
Total	**26**	**70**	**96**	**26.000**	**70.000**	**96**

Note: Observed frequencies were obtained from the source study [35]; Expected frequencies were calculated from the marginal totals and are shown to three decimal places; unrounded expected frequencies were used for all subsequent calculations.

Table 7. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
14	14.354	0.009
12	11.646	0.011
39	38.646	0.003
31	31.354	0.004
Total χ^2		**0.027**

In this example, we examined whether the type of intravitreal injection (aflibercept vs. ranibizumab) was associated with patient sex; both variables were nominal. This potential association was tested using the χ^2 test for independence.

- H_0: The type of injection (aflibercept or ranibizumab) and sex are independent.
- H_1: The type of injection and sex are associated.

Step 1: Observed and Expected Frequencies

Table 6 presents the observed and expected frequencies, where the expected values were calculated using formula (2), based on data collected from patients who received intravitreal injections of aflibercept or ranibizumab, categorized by sex [35].

Step 2: Calculate the Degrees of Freedom

The *df* for a contingency table were computed as:

df = (number of rows − 1) × (number of columns − 1)

In this 2 × 2 table (two sexes and two injection types), the *df* was: $df = (2 - 1) \times (2 - 1) = 1$

Step 3: Compute the X² Statistic

Using Equation (1) and the unrounded expected frequencies, the Pearson χ^2 statistic was 0.027, with $df = 1$ ($P = 0.870$) (Table 7): The source article reported $P = 0.8$ for the sex comparison [35]; direct recalculation of the uncorrected Pearson χ^2 test from the published aggregate counts yielded $P = 0.870$. This difference does not alter the inferential conclusion.

Step 4: Determine Statistical Significance

At $\alpha = 0.05$, the critical value for a χ^2 distribution with 1 *df* is 3.84 [23]. In this example, $\chi^2 = 0.027$ ($P = 0.870$), which is below the critical value. Therefore, the null hypothesis was not rejected, and no statistically significant evidence of an association between sex and injection type was detected in this sample.

Step 5: Evaluate the Strength of Association

Given that the contingency table was 2 × 2, the φ coefficient was used to quantify effect size. Using Equation (4),

$\Phi = \sqrt{\frac{0.0268}{96}} \approx 0.017$, indicating a negligible association according to the classification used in this review.

Stepwise Application of the X² Goodness-of-Fit Test in Example 4

To demonstrate the practical application of the χ^2 goodness-of-fit test, we present the following example:

Example 4: Goodness-of-Fit Test for Contact Lens Prescribing Patterns

A historical international survey conducted during 2007–2011 reported 12 230 rigid and 100 670 soft contact-lens fits, corresponding to approximately 10.8% rigid and 89.2% soft fittings [36]. In this example, these pooled historical proportions are used as an illustrative reference distribution for the contact-lens prescribing data reported from a university clinic in Trinidad and Tobago [37] (Table 8). For this didactic goodness-of-fit example, the pooled historical proportions were treated as fixed reference probabilities; a formal comparison of two empirical samples would instead be framed as a test of homogeneity or another appropriate two-sample comparison.

Hypotheses:

- H_0: The distribution of soft and rigid contact-lens fittings at the university clinic follows the pooled historical proportions reported in the source study [36].
- H_1: The distribution of soft and rigid contact-lens fittings at the university clinic differs from the pooled historical proportions reported in the source study [36].

To perform the χ^2 goodness-of-fit test, expected proportions were derived from the pooled counts reported in the source study [36]: 100 670 soft and 12 230 rigid fittings among 112 900 total fittings. For the clinic sample of $N = 243$, the expected frequencies were therefore:

Soft contact lenses: $(100\,670 / 112\,900) \times 243 = 216.677$

Rigid contact lenses: $(12\,230 / 112\,900) \times 243 = 26.323$.

The *df* were determined using the formula: *df* = number of categories − 1 = 2 − 1 = 1

Using the observed frequencies and the unrounded expected frequencies derived from the source study [36] (Table 9), the χ^2 statistic was 3.99, with $df = 1$ ($P = 0.046$).

To interpret the result, the calculated χ^2 statistic was compared with the critical value at $\alpha = 0.05$. For 1 *df*, the critical value is 3.84, based on the χ^2 distribution table [23]. Since $\chi^2 = 3.99$ exceeds the critical value of 3.84 ($P = 0.046$), the null hypothesis is rejected at $\alpha = 0.05$. Thus, the observed clinic distribution differed statistically from the pooled historical proportions used as the reference distribution. Because the source study [36] demonstrated substantial between-country variation in rigid-lens fitting rates, this result should be interpreted as an illustrative comparison with a historical pooled benchmark rather than as evidence of departure from a single fixed global distribution.

Table 8. Observed frequencies for the type of contact lenses used

Prescribing rates of contact lenses	Observed Frequency
Soft CL	207
RGP CL	36
Total	243

Abbreviations: CL, contact lens; RGP, rigid gas permeable. Note: Observed frequencies were obtained from the source study [37].

Table 9. Summarizing the data for calculating the χ^2 statistic

Prescribing rates of CLs	Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
Soft CL	207	216.677	0.432
RGP CL	36	26.323	3.557
Total χ^2			**3.989**

Abbreviations: CL, contact lens; RGP, rigid gas permeable. Note: Expected frequencies were calculated from the exact pooled soft- and rigid-lens counts reported in the source study [36].

CONCLUSIONS

The χ^2 test is a widely used, non-parametric statistical method for assessing associations between categorical variables and, in goodness-of-fit applications, for comparing observed categorical frequencies with a specified expected distribution. Its strengths lie in its simplicity and applicability across a range of disciplines, including epidemiology, clinical research, and public health. When applied correctly, the χ^2 test provides valuable insights into patterns of association that might otherwise be missed in categorical data.

This review outlines the step-by-step procedures for conducting the χ^2 tests of independence and goodness-of-fit, including hypothesis formulation, *df* calculation, test statistic interpretation, and effect size evaluation, using φ and Cramer's V. Through practical examples from ophthalmology and optometry, including the association between age group and glaucoma-therapy persistence, sex and anti-VEGF injection type, VI category and sex, and a goodness-of-fit comparison of contact-lens prescribing patterns, we demonstrated how χ^2 methods can be applied to clinical and vision-research questions. In the goodness-of-fit example, the observed contact-lens distribution differed statistically from the pooled historical reference proportions ($\chi^2 = 3.99$, $df = 1$, $P = 0.046$), leading to rejection of the specified null hypothesis.

Understanding the assumptions and limitations of the χ^2 test is essential for clinicians and researchers, including for optometrists and ophthalmologists. These considerations include ensuring adequate expected cell frequencies, independence of observations, and a sampling design appropriate to the intended inference. Misinterpretation or misuse of the test may lead to incorrect inferences, compromising the validity of the study results.

Moreover, we emphasize that complementing statistical significance with appropriate measures of effect size, such as Cramer's V or the φ coefficient, provides a more nuanced understanding of the strength of associations. This is particularly important when statistically significant results do not translate into clinically meaningful effects.

The χ^2 test is a foundational tool for categorical data analysis. Its correct application can support evidence-based interpretation and decision-making by eye-care professionals and clinical researchers and contribute meaningfully to the advancement of knowledge in vision science.

ETHICAL DECLARATIONS

Ethical approval: Not required.

Conflict of interests: The second and third authors of this article are staff members of the IVORC and its affiliated journals. However, in accordance with journal policy [38], articles submitted by the board, staff, or editor-in-chief undergo peer review to maintain impartiality. These submissions are managed by a different editorial board member with no conflicts of interest, ensuring that those with potential conflicts are excluded from the process. Furthermore, the editor overseeing the review process remains blinded to the reviewers' identities.

FUNDING

None.

ACKNOWLEDGMENTS

None.

REFERENCES

1. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb). 2013;23(2):143-9. doi: 10.11613/bm.2013.018. PMID: 23894860; PMCID: PMC3900058.
2. Schober P, Vetter TR. Chi-square Tests in Medical Research. Anesth Analg. 2019 Nov;129(5):1193. doi: 10.1213/ANE.0000000000004410. PMID: 31613806.
3. Bewick V, Cheek L, Ball J. Statistics review 8: Qualitative data - tests of association. Crit Care. 2004 Feb;8(1):46-53. doi: 10.1186/cc2428. Epub 2003 Dec 30. PMID: 14975045; PMCID: PMC420070.
4. Franke TM, Ho T, Christie CA. The chi-square test: Often used and more often misinterpreted. American journal of evaluation. 2012 Sep;33(3):448-58. doi: 10.1177/1098214011426594.
5. Fisher MJ, Marshall AP, Mitchell M. Testing differences in proportions. Aust Crit Care. 2011 May;24(2):133-8. doi: 10.1016/j.aucc.2011.01.005. Epub 2011 May 4. PMID: 21536451.

6. Xu B, Feng X, Burdine RD. Categorical data analysis in experimental biology. Dev Biol. 2010 Dec 1;348(1):3-11. doi: 10.1016/j.ydbio.2010.08.018. Epub 2010 Sep 6. PMID: 20826130; PMCID: PMC3021327.
7. Hazra A, Gogtay N. Biostatistics Series Module 4: Comparing Groups - Categorical Variables. Indian J Dermatol. 2016 Jul-Aug;61(4):385-92. doi: 10.4103/0019-5154.185700. PMID: 27512183; PMCID: PMC4966396.
8. Scott M, Flaherty D, Currall J. Statistics: dealing with categorical data. J Small Anim Pract. 2013 Jan;54(1):3-8. doi: 10.1111/j.1748-5827.2012.01298.x. Epub 2012 Nov 28. PMID: 23190085.
9. Lipsitz SR, Fitzmaurice GM, Sinha D, Hevelone N, Giovannucci E, Hu JC. Testing for independence in J×K contingency tables with complex sample survey data. Biometrics. 2015 Sep;71(3):832-40. doi: 10.1111/biom.12297. Epub 2015 Mar 11. PMID: 25762089; PMCID: PMC4567525.
10. Abdul Rahman H, Noraidi AA, Hj Khalid AN, Mohamad-Adam AZ, Zahari NH, Tuming NE. Practical guide to calculate sample size for chi-square test in biomedical research. BMC Med Res Methodol. 2025 May 26;25(1):144. doi: 10.1186/s12874-025-02584-4. PMID: 40419970; PMCID: PMC12107878.
11. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod. 2017 May;42(2):152-155. doi: 10.5395/rde.2017.42.2.152. Epub 2017 Mar 30. PMID: 28503482; PMCID: PMC5426219.
12. Nowacki A. Chi-square and Fisher's exact tests (From the "Biostatistics and Epidemiology Lecture Series, Part 1"). Cleve Clin J Med. 2017 Sep;84(9 Suppl 2):e20-e25. doi: 10.3949/ccjm.84.s2.04. PMID: 28937359.
13. Chang S, Li D, Qi Y. Pearson's goodness-of-fit tests for sparse distributions. J Appl Stat. 2021 Dec 30;50(5):1078-1093. doi: 10.1080/02664763.2021.2017413. PMID: 37009596; PMCID: PMC10062227.
14. Whitley E, Ball J. Statistics review 2: samples and populations. Crit Care. 2002 Apr;6(2):143-8. doi: 10.1186/cc1473. Epub 2002 Feb 7. Erratum in: Crit Care. 2003 Feb;7(1):8. PMID: 11983040; PMCID: PMC137296.
15. Egbuchulem KI. THE KARL PEARSON'S CHI-SQUARE: A MEDICAL RESEARCH LIBERO, AND A VERSATILE TEST STATISTIC: AN EDITORIAL. Ann Ib Postgrad Med. 2024 Aug 30;22(2):5-8. PMID: 40007702; PMCID: PMC11848378.
16. Kotronoulas G, Miguel S, Dowling M, Fernández-Ortega P, Colomer-Lahiguera S, Bağçivan G, Pape E, Drury A, Semple C, Dieperink KB, Papadopoulou C. An Overview of the Fundamentals of Data Management, Analysis, and Interpretation in Quantitative Research. Semin Oncol Nurs. 2023 Apr;39(2):151398. doi: 10.1016/j.soncn.2023.151398. Epub 2023 Mar 2. PMID: 36868925.
17. Warner P. Quantifying association in ordinal data. J Fam Plann Reprod Health Care. 2010 Apr;36(2):83-5. doi: 10.1783/147118910791069187. PMID: 20406551.
18. Oster RA. An examination of statistical software packages for categorical data analysis using exact methods. The American Statistician. 2002 Aug 1;56(3):235-46. doi: 10.1198/00031300274.
19. Oster RA. An examination of statistical software packages for categorical data analysis using exact methods—Part II. The American Statistician. 2003 Aug 1;57(3):201-13. doi: 10.1198/0003130031928.
20. Deo V, Ranganathan P. Statistical tools and packages for data collection, management, and analysis - A brief guide for health and biomedical researchers. Perspect Clin Res. 2024 Oct-Dec;15(4):209-212. doi: 10.4103/picr.picr_160_24. Epub 2024 Oct 9. PMID: 39583916; PMCID: PMC11584161.
21. Shahrul AI, Syed Mohamed AMF. A Comparative Evaluation of Statistical Product and Service Solutions (SPSS) and ChatGPT-4 in Statistical Analyses. Cureus. 2024 Oct 28;16(10):e72581. doi: 10.7759/cureus.72581. PMID: 39610603; PMCID: PMC11602406.
22. Abbasnasab Sardareh S, Brown GT, Denny P. Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. Teaching Statistics. 2021 Jul;43:S157-72. doi: 10.1111/test.12274.
23. Rana R, Singhal R. Chi-square test and its application in hypothesis testing. Journal of the practice of cardiovascular sciences. 2015 Jan 1;1(1):69-71. doi: 10.4103/2395-5414.157577.
24. Pulkstenis E, Robinson TJ. Goodness-of-fit tests for ordinal response regression models. Stat Med. 2004 Mar 30;23(6):999-1014. doi: 10.1002/sim.1659. PMID: 15027085.
25. Howell DC. Chi-square test: analysis of contingency tables. In International encyclopedia of statistical science 2011 (pp. 250-252). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-04898-2_174.
26. Pandis N. The chi-square test. Am J Orthod Dentofacial Orthop. 2016 Nov;150(5):898-899. doi: 10.1016/j.ajodo.2016.08.009. PMID: 27871717.
27. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof. 2021;18:17. doi: 10.3352/jeehp.2021.18.17. Epub 2021 Jul 30. PMID: 34325496; PMCID: PMC8441096.
28. Sapra RL, Saluja S. Understanding statistical association and correlation. Current Medicine Research and Practice. 2021 Jan 1;11(1):31-8. doi: 10.4103/cmrp.cmrp_62_20.
29. Ben-Shachar MS, Patil I, Thériault R, Wiernik BM, Lüdecke D. Phi, Fei, Fo, Fum: effect sizes for categorical data that use the chi-squared statistic. Mathematics. 2023 Apr 22;11(9):1982. doi: 10.3390/math11091982.
30. Lee DK. Alternatives to P value: confidence interval and effect size. Korean J Anesthesiol. 2016 Dec;69(6):555-562. doi: 10.4097/kjae.2016.69.6.555. Epub 2016 Oct 25. PMID: 27924194; PMCID: PMC5133225.
31. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. J Grad Med Educ. 2012 Sep;4(3):279-82. doi: 10.4300/JGME-D-12-00156.1. PMID: 23997866; PMCID: PMC3444174.

32. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021 Feb 15;31(1):010502. doi: 10.11613/BM.2021.010502. Epub 2020 Dec 15. PMID: 33380887; PMCID: PMC7745163.
33. Riazi A, Gharebaghi R and Heidary F. Seven-year experience in a low vision rehabilitation clinic at a tertiary referral center. Medical hypothesis, discovery & innovation in optometry. 2022;3(4):128–35. doi: 10.51329/mehdioptometry161.
34. Menino J, Camacho P, Coelho A. Persistence with medical glaucoma therapy in newly diagnosed patients. Med Hypothesis Discov Innov Ophthalmol. 2024 Aug 14;13(2):63-69. doi: 10.51329/mehdiophthal1495. PMID: 39206081; PMCID: PMC11347956.
35. Ali Ali MA, Hegazy HS, Abdelkhalek Elsayed MO, Tharwat E, Mansour MN, Hassanein M, Ezzeldin ER, GadElkareem AM, Abd Ellateef EM, Elsayed AA, Elabd IH, Abd Rbu MH, Amer RS, Gabbar AGAE, Mahmoud H, Abdelhameed HM, Abdelkader AME. Aflibercept or ranibizumab for diabetic macular edema. Med Hypothesis Discov Innov Ophthalmol. 2024 Jul 1;13(1):16-26. doi: 10.51329/mehdiophthal1490. PMID: 38978826; PMCID: PMC11227664.
36. Efron N, Morgan PB, Woods CA; International Contact Lens Prescribing Survey Consortium. International survey of rigid contact lens fitting. Optom Vis Sci. 2013 Feb;90(2):113-8. doi: 10.1097/OPX.0b013e31827cd8be. PMID: 23262991.
37. Ezinne NE, Ekemiri KK, Harbajan GN, Crooks AC, Douglas D, Ilechie AA, Mashige KP. Contact Lens Prescribing Patterns in a University Clinic in Trinidad and Tobago. Vision (Basel). 2022 Sep 2;6(3):55. doi: 10.3390/vision6030055. PMID: 36136748; PMCID: PMC9503470.
38. Medical hypothesis, discovery & innovation in optometry. Available at: https://mehdijournal.com/index.php/mehdioptometry/JournalPolicies (Accessed: December 10, 2025).