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 the independence of categorical data or the goodness-of-fit of the observed data to the expected distributions within contingency tables. 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 its two primary forms: 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.

Results: In the first example, the association between vision impairment (VI) and sex was examined using a 2 × 6 contingency table. The χ^2 statistic was 4.37 with 5 df ($P > 0.05$), indicating no statistically significant association. Cramer's V was 0.04, suggesting a very weak effect. The second example tested the association between age category and first-year persistence with antiglaucoma therapy. Here, $\chi^2 = 5.93$ (df = 2, $P > 0.05$), also showing no significant association, Cramer's V was weak (0.04). In the third example, a 2 × 2 table was used to analyze the association between sex and the type of anti-vascular endothelial growth factor injection (aflibercept or ranibizumab) used. This yielded a $\chi^2 = 0.214$ (df = 1, $P > 0.05$) and $\varphi = 0.05$, again indicating no statistically significant association and a weak effect. In a goodness-of-fit test assessing the pattern of contact lens usage, the χ^2 exceeded the critical threshold, indicating a significant deviation between the observed and expected frequencies, leading to rejection of the null hypothesis.

Conclusions: The χ^2 test is a robust tool for analyzing categorical data, enabling clinicians and researchers to identify potential relationships 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, non-parametric statistics, data analyses, optometry, ocular surgery, sample sizes, statistical data interpretations

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 (or distribution-free statistic) for analyzing categorical data [1]. It is primarily employed in two contexts: to evaluate the goodness-of-fit between observed and theoretical (expected) distributions and to test for independence between two categorical variables in a contingency table [2–5].

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 remains the cornerstone method for testing associations within such tables [8, 9], and its test statistic corresponds to the score for evaluating independence in r × c contingency tables [8, 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 its validity rests on the assumption of sufficiently large sample sizes [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

Assumptions of the X^2 Test

Although the χ^2 test is a non-parametric procedure, it relies on certain assumptions about the data, as is the case in parametric tests. Among these is the assumption that the data were obtained through random sampling [1, 13]. The key assumptions underlying the valid application of the χ^2 test are as follows.

1. Random Sampling: Each individual or observation must appear in the table only once and must have been selected randomly from the population of interest [1, 13].

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, 14].

3. Variable Types: Both variables under investigation should ideally be nominal, implying that the categories are qualitatively distinct, with no inherent or natural ordering [2, 14]. However, the χ^2 test may also be applied to ordinal–ordinal or nominal–ordinal variable combinations, as long as the analysis aims to test general associations, rather than measuring correlation strength or detecting a linear trend [1, 15, 16].

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, 14].

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 of zero [2, 14].

Alternatives to the X^2 Test When Assumptions Are Violated

For 2 × 2 contingency tables, if the assumptions of the χ^2 test (e.g., minimum expected cell counts) are not met, the assumptions underlying the χ^2 approximation are violated. When more than 20% of the cells in a contingency table have expected frequencies < 5, the χ^2 test becomes unreliable. In such cases, Fisher's exact test provides a more accurate and appropriate alternative for assessing the associations between variables [11, 12]. 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 larger contingency tables (greater than 2 × 2) with small expected frequencies, alternative tests, such as the likelihood-ratio χ^2 test or the Fisher–Freeman–Halton Exact Test (possibly with Monte Carlo simulation), are recommended to ensure valid statistical inference [17, 18].

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 χ^2 test in an $r \times c$ contingency table, the degrees of freedom are calculated as $(r-1) \times (c-1)$, where r is the number of levels (categories) of the variable presented in rows and c is the number of levels of the variable presented in columns in the contingency table [12].

Software Tools Supporting the Χ² Test

Various statistical software packages support the application of the χ^2 test, including widely used tools, such as IBM SPSS (IBM Corp., Armonk, New York, USA), Python (Python Software Foundation, Lafayette Boulevard, Fredericksburg, VA, USA), and R (R Foundation for Statistical Computing, Vienna, Austria). These platforms offer flexible and accessible options for conducting χ^2 analyses across diverse research settings [19–21].

Types of Χ² Tests

As stated earlier, the χ^2 test is a non-parametric statistical method primarily used for two purposes [14, 22, 23]:

1. To evaluate the extent to which the observed data distribution fits an expected, theoretical distribution (i.e., goodness-of-fit test) [14, 22, 23].
2. To test the hypothesis of no association between two or more groups, populations, or classification criteria (i.e., to assess independence between two categorical variables) [14, 22, 23].

1. Χ² 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, 22, 23]. This type of test involves only one variable with multiple levels.

The Pearson χ^2 test is widely used in this context. When the number of categories (k) is fixed, the test statistic follows an approximate χ^2 distribution with $k - 1$ degrees of freedom (df) when the sample size (n) becomes large [24].

- H_0 (null hypothesis): The observed frequencies will be consistent with the expected frequencies (i.e., the sample data will follow a specified distribution) [4, 22].
- H_1 (alternative hypothesis): The observed frequencies will not be consistent with the expected frequencies (i.e., the sample data will not follow the specified distribution) [4, 22].

2. Χ² Test of Independence

The χ^2 test of independence is a two-dimensional test that is used to determine whether two categorical variables are associated. Specifically, it examines whether the proportions across the levels of one variable differ significantly depending on the levels of the other [1, 25].

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

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

where O_1, O_2, ..., O_n are the observed frequencies, E_1, E_2, ..., E_n are the expected frequencies, and n is the number of cells in the contingency table.

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 $(r - 1) \times (c - 1)$, where r is the number of rows and c is the number of columns [4, 16, 22, 26].

- H_0: The two categorical variables will be independent [4, 22].
- H_1: The two categorical variables will be associated [4, 22].

The χ^2 test statistic is interpreted as follows [27]: A large difference between the observed and expected frequencies produces a large χ^2 value, leading to rejection of the null hypothesis [27]. On the other hand, a small difference yields a small χ^2 value, supporting the null hypothesis [27].

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

$$E_{ij} = \frac{(i^{th}\text{ row total})(j^{th}\text{column total})}{\text{grand total}} = \frac{(n_{i+})(n_{+j})}{n} \quad (2),$$

where, E_{ij} represents the expected frequency in the cell located in the i^{th} row and j^{th} column under the assumption of independence between the row and column variables. This expected value is obtained by multiplying the row total (i^{th}) and column total (j^{th}) for the respective cell and dividing this product by the overall sample size (grand total; n) [12, 27].

Strength of Association in Χ² 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, 11].

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, 29]. Cramer's V is calculated as follows [1, 28, 30]:

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{n.df}} \quad (3),$$

where χ^2 is the χ^2 statistic, n is the total number of observations, and df is the degrees of freedom, calculated as $(r - 1) \times (c - 1)$, with r and c being the number of rows and columns, respectively.

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). Although the Pearson correlation can be negative, the φ coefficient is always reported as a non-negative, absolute value [28, 29]. Notably, the value of the φ coefficient can vary when the rows and columns are exchanged and is only suitable for 2 × 2 tables [1, 11].

For 2 × 2 tables, the φ coefficient is computed to indicate the effect size by using the following formula [11, 28, 30]:

$$\varphi=\sqrt{\frac{\chi^2}{n}} \quad (4)$$

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 mere statistical significance, these thresholds assist in understanding the practical significance of a finding [1, 10, 28, 31].

Power of the X² 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 essential for detecting true associations when they exist and for avoiding Type II errors.

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, 32, 33]. Cohen proposed conventional thresholds for small (w = 0.10), medium (w = 0.30), and large (w = 0.50) effect sizes. These benchmarks have become the standard for determining the sample size requirements for categorical data analysis [10].

RESULTS and DISSCUSSION

Stepwise Application of the X² Test of Independence in Examples 1–3

To demonstrate the practical application of the χ^2 test of independence, we present three examples from the field of ophthalmology.

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 [34]. 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

Category of VI	Observed Frequency			Expected Frequency		
	Men	Women	Total	Men	Women	Total
Category 1	42	32	74	44	30	74
Category 2	171	108	279	166	113	279
Category 3	55	31	86	51	35	86
Category 4	46	45	91	54	37	91
Category 5	23	13	36	22	14	36
Category 6	1	0	1	1	0	1
Total	338	229	567	338	229	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. Expected frequencies have been rounded and are presented as whole numbers for ease of display. However, these values are typically decimal numbers, as they are derived from relative frequencies.

Table 3. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
42	44	0.09
32	30	0.13
171	166	0.15
108	113	0.22
55	51	0.31
31	35	0.45
46	54	1.18
45	37	1.73
23	22	0.04
13	14	0.07
1	1	0
0	0	0

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 in Tehran, Iran, between March 20, 2012, and March 20, 2019 [34].

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 X² Statistic

Using the χ^2 formula (1), we calculated the following (Table 3):

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 4.37, \text{with df} = 5$$

Step 4: Determine Statistical Significance

To evaluate whether the null hypothesis could be rejected, we compared the calculated χ^2 statistic value with the critical value from the χ^2 distribution table [35]. At $\alpha = 0.05$, the critical value for 5 df is 11.07.

Since the calculated χ^2 statistic value of 4.37 is less than 11.07, we cannot reject the null hypothesis. Thus, VI and sex are not statistically significantly associated in this dataset.

Step 5: Evaluate the Strength of Association

Given that the contingency table was larger than 2 × 2, the appropriate effect size measure was Cramer's V, which was calculated using Equation (3):

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{n.df}} = \sqrt{\frac{4.37}{567 \times 5}} \approx 0.04$$

This suggests a negligible association between VI and sex. The low Cramer's V value agrees with the χ^2 test outcome and further supports the conclusion that these two variables are not significantly associated.

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

Monotherapy, age, and side effects are significant risk factors for discontinuation of antiglaucoma treatment. Maintaining long-term persistence with therapy is essential to slow disease progression and to prevent irreversible vision loss. 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 [36].

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.

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 [36].

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

Table 4. Observed and expected frequencies in the category of 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	6	73
45–64	420	48	468	430	38	468
≥65	1481	116	1597	1468	129	1597
Total	1965	173	2138	1965	173	2138

Note: Expected frequencies have been rounded and are presented as whole numbers for ease of display. However, these values are typically decimal numbers, as they are derived from relative frequencies.

Table 5. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
64	67	0.14
420	430	0.23
1481	1468	0.12
9	6	1.5
48	38	2.63
116	129	1.31

Step 3: Compute the X^2 Statistic

Applying formula (1), the χ^2 statistic is calculated as follows (Table 5):

$\chi^2 = 5.93$, with df = 2

Step 4: Determine Statistical Significance

To determine whether the null hypothesis could be rejected, we compared the calculated χ^2 value with the critical values from the χ^2 distribution table. At $\alpha = 0.05$, the critical value for 2 df was 5.99 [35].

Since the computed χ^2 statistic value of 5.93 is less than the critical value of 5.99, the null hypothesis cannot be rejected. This implies that age and first-year therapy persistence were not associated in this population.

Step 5: Evaluate the Strength of Association

As the contingency table was 3 × 2, Cramer's V was appropriate for measuring effect size, which was calculated using Equation (3):

Cramer's V=$\sqrt{\frac{\chi^2}{n.df}} = \sqrt{\frac{5.93}{2138\times2}} \approx 0.04$

The value of Cramer's V indicated a very weak association between age and persistence. The small effect size was consistent with the results of the χ^2 test and reinforced the conclusion that age is not significantly associated with first-year therapy persistence in this sample.

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 has traditionally been the standard treatment for DME, use of intravitreal anti-VEGF injections, such as aflibercept and ranibizumab, has recently become more widespread and has surpassed the use of laser therapy in many cases [37].

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	15	38	53
Ranibizumab	12	31	43	11	32	43
Total	26	70	96	26	70	96

Note: Expected frequencies have been rounded and are presented as whole numbers for ease of display. However, these values are typically decimal numbers, as they are derived from relative frequencies.

Table 7. Summarizing data for χ^2 statistic calculation

Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
14	15	0.067
12	11	0.090
39	38	0.026
31	32	0.031

In this study, we investigated whether the type of intravitreal injection (aflibercept vs. ranibizumab) (a nominal variable) was associated with patient sex (a nominal variable). 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 [37].

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) × (2 − 1) = 1

Step 3: Compute the X² Statistic

Using formula (1), the χ^2 statistic is calculated as follows (Table 7):

$\chi^2 = 0.214$, with df = 1

Step 4: Determine Statistical Significance

At $\alpha = 0.05$, the critical value for a χ^2 distribution with 1 df is 3.84 [35]. If the calculated statistic exceeds this value, the null hypothesis would be rejected. However, in this example, $\chi^2 = 0.214$. Because the χ^2 statistic value was smaller than the critical value of 3.84, the null hypothesis could not be rejected. This suggests that sex and the type of intravitreal injection administered was not associated in this population.

Step 5: Evaluate the Strength of Association

Given that the contingency table was 2 × 2, the appropriate effect size measure is the φ coefficient, which was calculated using formula (4):

$$\Phi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{0.214}{96}} \approx 0.05$$

This φ value indicates a small effect size, which was consistent with a negligible association, reinforcing the conclusion that sex and the choice of injection (aflibercept or ranibizumab) were not associated [37].

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

Globally, rigid gas-permeable (RGP) lenses account for approximately 10% of all contact lens fittings, whereas soft lenses make up over 90% of contact lens prescriptions [38, 39]. In this example, the prescription patterns of contact lenses at a university clinic in Trinidad and Tobago were assessed to determine whether the observed distribution was in line with global expectations [38] (Table 8).

Hypotheses:

- H_0: The observed distribution of contact lens types at a university clinic in Trinidad and Tobago matches the expected global distribution.
- H_1: The observed distribution of contact lens types at a university clinic in Trinidad and Tobago differs from the expected global distribution.

To perform the χ^2 goodness-of-fit test, expected frequencies were calculated based on the total sample size (n = 243).

Soft contact lenses: 0.90 × 243 = 219

RGP contact lenses: 0.10 × 243 = 24

The df were determined using the formula:

df = number of categories – 1 = 2 – 1 = 1

Using the observed and expected frequencies (Table 9), the χ^2 statistic was computed as:

$\chi^2 = 6.66$, with df = 1

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 [35].

Since the χ^2 statistic = 6.66 is larger than 3.84, the null hypothesis is rejected. This indicates a statistically significant difference between the observed and expected distributions of contact lens types.

The χ^2 goodness-of-fit test results suggest that the prescription pattern at this university clinic deviates significantly from global trends. This finding may reflect the local clinical preferences, patient demographics, and/or access to different lenses [38].

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

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

Table 9. Summarizing the data for calculating the χ² statistic

Prescribing rates of CLs	Observed Frequency	Expected Frequency	$\frac{(O_i - E_i)^2}{E_i}$
Soft CL	207	219	0.66
RGP CL	36	24	6.00

Abbreviations: CL, contact lens; RGP, rigid gas permeable

CONCLUSIONS

The χ² test is a widely used, non-parametric statistical method for assessing associations between two categorical variables. Its strengths lie in its simplicity and applicability across a range of disciplines, including epidemiology, clinical research, and public health. When applied correctly, the χ² 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 χ² 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, such as treatment persistence in glaucoma, sex-based differences in anti-VEGF therapy, sex distribution of VI, and contact lens prescription patterns, we demonstrated how this test can be applied to real-world clinical questions. In our goodness-of-fit test example, assessing the pattern of contact lens usage, the χ² statistic exceeded the critical threshold, indicating a significant deviation between the observed and expected frequencies, and leading to the rejection of the null hypothesis.

Understanding the assumptions and limitations of the χ² test is essential for clinicians and researchers, including for optometrists and ophthalmologists. These considerations include ensuring a sufficient sample size, adequate expected cell counts, and the independence of observations. 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 χ² test is a foundational tool for categorical data analysis. Its correct application can empower eye-care professionals and clinical researchers to make evidence-based decisions, improve patient outcomes, and meaningfully contribute 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 [40], 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. Sharkey AM, Siddiqui N, Downey K, Ye XY, Guevara J, Carvalho JCA. Comparison of Intermittent Intravenous Boluses of Phenylephrine and Norepinephrine to Prevent and Treat Spinal-Induced Hypotension in Cesarean Deliveries: Randomized Controlled Trial. Anesth Analg. 2019 Nov;129(5):1312-1318. doi: 10.1213/ANE.0000000000003704. PMID: 30113395.
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: 10.1177/109821401142659.
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. Bergh D. Chi-Squared Test of Fit and Sample Size-A Comparison between a Random Sample Approach and a Chi-Square Value Adjustment Method. J Appl Meas. 2015;16(2):204-17. PMID: 26075668.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. Weidener L, Fischer M. Artificial Intelligence in Medicine: Cross-Sectional Study Among Medical Students on Application, Education, and Ethical Aspects. JMIR Med Educ. 2024 Jan 5;10:e51247. doi: 10.2196/51247. PMID: 38180787; PMCID: PMC10799276.
26. 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.
27. 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.
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. 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.
30. 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.
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. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. Medical hypothesis, discovery & innovation in optometry. Available at: https://mehdijournal.com/index.php/mehdioptometry/JournalPolicies (Accessed: December 10, 2025).