Chi Square Problems With Solutions

Understanding Chi Square Problems With Solutions: A Practical Guide

Statistics can often feel like a maze of formulas and assumptions, but the chi-square test stands out as one of the most intuitive and widely used tools for analyzing categorical data. Whether you are a student facing an exam, a researcher testing hypotheses, or a data analyst exploring survey results, encountering **chi square problems with solutions** is inevitable. This article will demystify the chi-square test, walk you through its various applications, and present clear, step-by-step examples so you can confidently tackle any chi-square problem that comes your way.

The chi-square (χ^2) test is a non-parametric statistical method used to determine whether there is a significant association between categorical variables or whether a sample distribution matches an expected distribution. By the end of this guide, you will be able to identify which chi-square test to use, how to set up the hypotheses, calculate the test statistic, and interpret the results. Let's dive into the world of **chi square problems with solutions** and transform a challenging topic into a manageable skill.

What Is the Chi-Square Test?

The chi-square test compares observed frequencies (what you actually collect) with expected frequencies (what you would expect if there were no relationship or pattern). The core formula is:

$\chi^2 = \Sigma ((O – E)^2 / E)$

where O = observed count, E = expected count, and the summation is over all categories or cells. A large chi-square value indicates a large deviation from expectations, often leading to rejection of the null hypothesis.

There are three common types of chi-square tests:

- **Chi-Square Goodness-of-Fit Test** – tests if a sample distribution matches a hypothesized distribution (e.g., equal proportions).
- **Chi-Square Test of Independence** – examines whether two categorical variables are independent (e.g., gender and voting preference).
- **Chi-Square Test of Homogeneity** – checks whether different populations have the same distribution of a categorical variable (e.g., treatment outcomes across different hospitals).

Each type has its own data layout and degrees of freedom calculation, but the underlying logic remains the same. Now, let's explore concrete **chi square problems with solutions** for each type.

Chi-Square Goodness-of-Fit Problem With Solution

Problem 1: Do customers prefer one flavor equally?

A ice cream shop claims that customers have no preference among four flavors: vanilla, chocolate, strawberry, and mint. Over a weekend, they sell 40 vanilla, 25 chocolate, 20 strawberry, and 15 mint. Test at a significance level of $\alpha = 0.05$ whether the sales are equally distributed.

Step 1: State the Hypotheses

- **Null hypothesis (H_0):** Customers prefer all flavors equally (each flavor has 25% of sales).
- **Alternative hypothesis (H_1):** At least one flavor preference is different from the expected proportion.

Step 2: Calculate Expected Frequencies

Total sales = 40 + 25 + 20 + 15 = 100. Under H_0, each flavor expected = 100 / 4 = 25.

Step 3: Compute Chi-Square Statistic

$\chi^2 = \Sigma (O – E)^2 / E$

- Vanilla: $(40 – 25)^2 / 25 = 225 / 25 = 9.0$
- Chocolate: $(25 – 25)^2 / 25 = 0 / 25 = 0.0$
- Strawberry: $(20 – 25)^2 / 25 = 25 / 25 = 1.0$
- Mint: $(15 – 25)^2 / 25 = 100 / 25 = 4.0$

$\chi^2 = 9.0 + 0.0 + 1.0 + 4.0 = 14.0$

Step 4: Determine Degrees of Freedom and Critical Value

df = (number of categories – 1) = 4 – 1 = 3. At $\alpha = 0.05$, the critical value from chi-square distribution table is 7.815.

Step 5: Decision and Interpretation

Since $\chi^2 = 14.0 > 7.815$, we reject H_0. There is sufficient evidence that customer preferences are not equally distributed. This is a classic example of **chi square problems with solutions** involving a single categorical variable.

Chi-Square Test of Independence Problem With Solution

Problem 2: Is there a relationship between education level and exercise frequency?

A survey of 200 adults classified them by education (High School, College, Graduate) and exercise frequency (Regular, Irregular). The observed data is below. Test at $\alpha = 0.05$ whether exercise frequency is independent of education level.

Education / Exercise	Regular	Irregular	Total
High School	20	40	60
College	30	30	60
Graduate	40	40	80
Total	90	110	200

Step 1: State Hypotheses

- **H_0:** Education and exercise frequency are independent.
- **H_1:** They are not independent.

Step 2: Calculate Expected Frequencies

For each cell, E = (row total × column total) / grand total.

- High School & Regular: $(60 \times 90) / 200 = 5400 / 200 = 27$
- High School & Irregular: $(60 \times 110) / 200 = 6600 / 200 = 33$
- College & Regular: $(60 \times 90) / 200 = 27$
- College & Irregular: $(60 \times 110) / 200 = 33$
- Graduate & Regular: $(80 \times 90) / 200 = 7200 / 200 = 36$
- Graduate & Irregular: $(80 \times 110) / 200 = 8800 / 200 = 44$

Step 3: Compute Chi-Square

$\chi^2 = \Sigma (O - E)^2 / E$

- High School/Regular: $(20–27)^2/27 = 49/27 \approx 1.8148$
- High School/Irregular: $(40–33)^2/33 = 49/33 \approx 1.4848$
- College/Regular: $(30–27)^2/27 = 9/27 = 0.3333$
- College/Irregular: $(30–33)^2/33 = 9/33 \approx 0.2727$
- Graduate/Regular: $(40–36)^2/36 = 16/36 \approx 0.4444$
- Graduate/Irregular: $(40–44)^2/44 = 16/44 \approx 0.3636$

Sum = 1.8148 + 1.4848 + 0.3333 + 0.2727 + 0.4444 + 0.3636 = 4.7136

Step 4: Degrees of Freedom and Critical Value

df = (rows – 1) × (columns – 1) = (3 – 1) × (2 – 1) = 2 × 1 = 2. At $\alpha=0.05$, critical value = 5.991.

Step 5: Decision

$\chi^2 = 4.7136 < 5.991$, so we fail to reject H_0. There is insufficient evidence to conclude a relationship between education level and exercise frequency. This independence test is a frequent type of **chi square problems with solutions** in social science research.

Chi-Square Test of Homogeneity Problem With Solution

Problem 3: Are survival rates the same across three hospitals?

A study records whether patients survived a critical surgery in three different hospitals. Data: Hospital A – survived 45, died 5; Hospital B – survived 38, died 12; Hospital C – survived 42, died 8. Test at $\alpha = 0.05$ whether survival distribution is homogeneous (same proportion) across hospitals.

Step 1: Hypotheses

- **H_0:** The proportion of survivors is the same across all three hospitals.
- **H_1:** At least one hospital has a different survival proportion.

Step 2: Build Contingency Table

Rows: hospitals; columns: survived/died.

Hospital	Survived	Died	Row Total
A	45	5	50
B	38	12	50

Hospital	Survived	Died	Row Total
C	42	8	50
Column Total	125	25	150

Step 3: Expected Frequencies

Under homogeneity, overall survival proportion = 125/150 = 0.8333, death = 25/150 = 0.1667.

E for each cell = row total × column proportion. Since each row total is 50:

- Survived for any hospital: 50 × (125/150) = 50 × 0.83333 = 41.6667
- Died for any hospital: 50 × (25/150) = 50 × 0.16667 = 8.3333

So expected values (same for each hospital): Survived = 41.67, Died = 8.33.

Step 4: Compute Chi-Square

$\chi^2 = \Sigma (O - E)^2 / E$

- Hospital A survived: $(45 - 41.67)^2 / 41.67 \approx (11.09) / 41.67 \approx 0.266$
- Hospital A died: $(5 - 8.33)^2 / 8.33 \approx (11.09) / 8.33 \approx 1.331$
- Hospital B survived: $(38 - 41.67)^2 / 41.67 \approx 13.47 / 41.67 \approx 0.323$
- Hospital B died: $(12 - 8.33)^2 / 8.33 \approx 13.47 / 8.33 \approx 1.617$
- Hospital C survived: $(42 - 41.67)^2 / 41.67 \approx 0.11 / 41.67 \approx 0.003$
- Hospital C died: $(8 - 8.33)^2 / 8.33 \approx 0.11 / 8.33 \approx 0.013$

Sum = 0.266 + 1.331 + 0.323 + 1.617 + 0.003 + 0.013 = 3.553

Step 5: Degrees of Freedom and Critical Value

df = (rows – 1) ×