chi square practice problems

Chi square practice problems are essential tools for students and professionals alike who seek to understand the application of the chi-square test in statistics. These problems not only help in grasping theoretical concepts but also enhance problem-solving skills necessary for real-world data analysis. In this article, we will explore the fundamentals of chi-square tests, provide a series of practice problems, and offer detailed solutions to solidify your understanding.

Understanding Chi-Square Tests

The chi-square test is a statistical method used to determine if there is a significant association between categorical variables. It compares the observed frequencies of events to the expected frequencies under the null hypothesis, which states that there is no association between the variables.

There are two primary types of chi-square tests:

- **Chi-Square Test of Independence:** This test assesses whether two categorical variables are independent of one another.
- **Chi-Square Goodness of Fit Test:** This test determines if a sample distribution matches an expected distribution.

The Formula

The formula for the chi-square statistic (χ^2) is:

$$
\chi^2 = \sum \frac{(O - E)^2}{E}
$$

Where:
- O = Observed frequency
- E = Expected frequency

The sum is taken over all categories.

Chi-Square Practice Problems

To help you master the chi-square test, we will present several practice problems, each followed by a thorough solution.

Problem 1: Chi-Square Test of Independence

A researcher wants to investigate whether there is an association between gender and preference for a type of fruit. The following data was collected:

	Apples	Bananas	Oranges	Total
Males	30	10	20	60
Females	20	25	15	60
Total	50	35	35	120

1. State the null and alternative hypotheses.
2. Calculate the expected frequencies.
3. Compute the chi-square statistic.
4. Determine whether to reject or fail to reject the null hypothesis at a significance level of 0.05.

Solution to Problem 1

1. Hypotheses:
- Null Hypothesis (H0): There is no association between gender and fruit preference.
- Alternative Hypothesis (H1): There is an association between gender and fruit preference.

2. Expected Frequencies:
To find the expected frequency for each cell, use the formula:

$$
E = \frac{(Row \ Total) \times (Column \ Total)}{Grand \ Total}
$$

- For Males and Apples:
$$
E = \frac{60 \times 50}{120} = 25
$$
- For Males and Bananas:
$$
E = \frac{60 \times 35}{120} = 17.5
$$
- For Males and Oranges:

$$
E = \frac{60 \times 35}{120} = 17.5
$$
- For Females and Apples:
$$
E = \frac{60 \times 50}{120} = 25
$$
- For Females and Bananas:
$$
E = \frac{60 \times 35}{120} = 17.5
$$
- For Females and Oranges:
$$
E = \frac{60 \times 35}{120} = 17.5
$$

The expected frequency table is:

	Apples	Bananas	Oranges	Total
Males	25	17.5	17.5	60
Females	25	17.5	17.5	60
Total	50	35	35	120

3. Chi-Square Statistic:
$$
\chi^2 = \sum \frac{(O - E)^2}{E}
$$
- For Males and Apples:
$$
\frac{(30 - 25)^2}{25} = \frac{25}{25} = 1
$$
- For Males and Bananas:
$$
\frac{(10 - 17.5)^2}{17.5} = \frac{56.25}{17.5} \approx 3.21
$$
- For Males and Oranges:
$$
\frac{(20 - 17.5)^2}{17.5} = \frac{6.25}{17.5} \approx 0.36
$$
- For Females and Apples:
$$
\frac{(20 - 25)^2}{25} = \frac{25}{25} = 1
$$
- For Females and Bananas:
$$
\frac{(25 - 17.5)^2}{17.5} = \frac{56.25}{17.5} \approx 3.21
$$
- For Females and Oranges:
$$
\frac{(15 - 17.5)^2}{17.5} = \frac{6.25}{17.5} \approx 0.36
$$

\]

Summing these values gives:
\[
\chi^2 = 1 + 3.21 + 0.36 + 1 + 3.21 + 0.36 \approx 9.14
\]

4. Decision:
The degrees of freedom (df) = (rows - 1)(columns - 1) = (2 - 1)(3 - 1) = 2. Looking up the critical value for χ^2 at df = 2 and $\alpha = 0.05$ gives approximately 5.991. Since 9.14 > 5.991, we reject the null hypothesis. There is significant evidence to suggest an association between gender and fruit preference.

Problem 2: Chi-Square Goodness of Fit Test

A researcher believes that a die is fair and would like to test this assumption. She rolls the die 60 times and records the following results:

Face	Frequency
1	10
2	8
3	12
4	15
5	7
6	8

1. State the null and alternative hypotheses.
2. Calculate the expected frequencies.
3. Compute the chi-square statistic.
4. Determine whether to reject or fail to reject the null hypothesis at a significance level of 0.05.

Solution to Problem 2

1. Hypotheses:
- Null Hypothesis (H0): The die is fair (all faces are equally likely).
- Alternative Hypothesis (H1): The die is not fair.

2. Expected Frequencies:
For a fair die, each face should appear equally often. The expected frequency for each face is:
\[
E = \frac{Total \ Rolls}{Number \ of \ Faces} = \frac{60}{6} = 10
\]

3. Chi-Square Statistic:
$$
\chi^2 = \sum \frac{(O - E)^2}{E}
$$
- For Face 1:
$$
\frac{(10 - 10)^2}{10} = 0
$$
- For Face 2:
$$
\frac{(8 - 10)^2}{10} = \frac{4}{10} = 0.4
$$
- For Face 3:
$$
\frac{(12 - 10)^2}{10} = \frac{4}{10} = 0.4
$$
- For Face 4:
$$
\frac{(15 - 10)^2}{10} = \frac{25}{10} = 2.5
$$
- For Face 5:
$$
\frac{(7 - 10)^2}{10} = \frac{9}{10} = 0.9
$$
- For Face 6:
$$
\frac{(8 - 10)^2}{10} = \frac{4}{10} = 0.4
$$

Summing these values gives:
$$
\chi^2 = 0 + 0.4 + 0.4 + 2.5 + 0.9 + 0.4 = 4.6
$$
\

Frequently Asked Questions

What is a chi-square test and when is it used?

A chi-square test is a statistical method used to determine if there is a significant association between categorical variables. It is commonly used in hypothesis testing to compare observed data with expected data under the null hypothesis.

How do you calculate the chi-square statistic?

The chi-square statistic is calculated using the formula: $\chi^2 = \Sigma((O - E)^2 / E)$, where O is the observed frequency and E is the expected frequency for each category.

What are the assumptions of the chi-square test?

The main assumptions include: the data should consist of independent observations, the variables should be categorical, and the expected frequency in each category should be at least 5.

What does a high chi-square value indicate?

A high chi-square value indicates a greater difference between observed and expected frequencies, suggesting that there may be a significant association between the variables being tested.

What is the difference between a chi-square test of independence and a chi-square goodness of fit test?

A chi-square test of independence assesses whether two categorical variables are independent, while a chi-square goodness of fit test determines if a sample distribution matches a specified distribution.

How do you interpret the p-value in a chi-square test?

The p-value indicates the probability of observing the data if the null hypothesis is true. A low p-value (typically less than 0.05) suggests that the null hypothesis can be rejected, indicating a significant association.

Can you perform a chi-square test with small sample sizes?

While it is possible to perform a chi-square test with small samples, it is not recommended unless the expected frequencies in each category are sufficiently high (at least 5). In such cases, other tests like Fisher's Exact Test may be more appropriate.

What is the role of degrees of freedom in a chi-square test?

Degrees of freedom in a chi-square test are calculated as (number of categories - 1) for goodness of fit tests, or (rows - 1) (columns - 1) for tests of independence. They determine the critical value for the chi-square distribution.

What are common applications of chi-square tests?

Common applications include market research, genetics, survey analysis, and any field where relationships between categorical variables need to be analyzed.

How do you perform a chi-square test using software like R or Python?

In R, you can use the 'chisq.test()' function for both goodness of fit and independence tests. In Python, the 'scipy.stats.chi2_contingency()' function is used for independence tests, and 'scipy.stats.chisquare()' for goodness of fit.

Chi Square Practice Problems