chi square test of independence example problems with answers

Chi square test of independence example problems with answers are vital tools in the field of statistics, helping researchers and analysts understand the relationship between categorical variables. This article will delve into the chi-square test of independence, providing detailed examples and solutions to help clarify this important statistical method.

Understanding the Chi Square Test of Independence

The chi-square test of independence is a statistical method used to determine whether there is a significant association between two categorical variables. It compares the observed frequencies in each category of a contingency table to the frequencies we would expect if the variables were independent.

Key Concepts

1. Null Hypothesis (H0): Assumes that there is no association between the two categorical variables.
2. Alternative Hypothesis (H1): Assumes that there is a significant association between the two variables.
3. Degrees of Freedom (df): Calculated as (number of rows - 1) × (number of columns - 1) in the contingency table.
4. Chi-Square Statistic (χ^2): Measures how expectations compare to actual observed data.

Steps to Conduct a Chi Square Test of Independence

To perform a chi-square test of independence, follow these essential steps:

1. State the hypotheses: Formulate the null and alternative hypotheses.
2. Create a contingency table: Organize the data into a table format.
3. Calculate the expected frequencies: Use the formula:
$$
E = \frac{(Row Total) \times (Column Total)}{Grand Total}
$$
4. Compute the chi-square statistic: Use the formula:
$$
\chi^2 = \sum \frac{(O - E)^2}{E}
$$
where O is the observed frequency and E is the expected frequency.
5. Determine the degrees of freedom: Calculate the degrees of freedom for your table.
6. Find the critical value: Use the chi-square distribution table to find the critical value based on the degrees of freedom and significance level (commonly $\alpha = 0.05$).

7. Make a decision: If the chi-square statistic is greater than the critical value, reject the null hypothesis.

Example Problem 1: Gender and Preference for a Type of Movie

Suppose we want to determine if there is an association between gender (Male, Female) and preference for movie genre (Action, Drama). We survey 100 individuals, and the results are as follows:

	Action	Drama	Total
Male	30	20	50
Female	10	40	50
Total	40	60	100

Solution

1. State the hypotheses:
- H0: There is no association between gender and movie preference.
- H1: There is an association between gender and movie preference.

2. Calculate expected frequencies:
- For Male Action: $E = \frac{50 \times 40}{100} = 20$
- For Male Drama: $E = \frac{50 \times 60}{100} = 30$
- For Female Action: $E = \frac{50 \times 40}{100} = 20$
- For Female Drama: $E = \frac{50 \times 60}{100} = 30$

	Action	Drama	Total
Male	30 (20)	20 (30)	50
Female	10 (20)	40 (30)	50
Total	40	60	100

3. Compute the chi-square statistic:
$$
\chi^2 = \frac{(30-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30}
$$
$$
\chi^2 = \frac{100}{20} + \frac{100}{30} + \frac{100}{20} + \frac{100}{30} = 5 + 3.33 + 5 + 3.33 = 16.66
$$

4. Determine degrees of freedom:
$$
df = (2-1)(2-1) = 1
```
```

\]

5. Find critical value:
For $df = 1$ at $\alpha = 0.05$, the critical value from the chi-square table is 3.841.

6. Make a decision:
Since $16.66 > 3.841$, we reject the null hypothesis.

Conclusion

There is a significant association between gender and movie preference.

Example Problem 2: Smoking and Lung Disease

A health study examines the relationship between smoking (Smoker, Non-Smoker) and the presence of lung disease (Yes, No). The data collected is as follows:

	Lung Disease Yes	Lung Disease No	Total
Smoker	30	70	100
Non-Smoker	10	90	100
Total	40	160	200

Solution

1. State the hypotheses:
- H0: There is no association between smoking and lung disease.
- H1: There is an association between smoking and lung disease.

2. Calculate expected frequencies:
- For Smoker Yes: $E = \frac{100 \times 40}{200} = 20$
- For Smoker No: $E = \frac{100 \times 160}{200} = 80$
- For Non-Smoker Yes: $E = \frac{100 \times 40}{200} = 20$
- For Non-Smoker No: $E = \frac{100 \times 160}{200} = 80$

3. Compute the chi-square statistic:
$$
\chi^2 = \frac{(30-20)^2}{20} + \frac{(70-80)^2}{80} + \frac{(10-20)^2}{20} + \frac{(90-80)^2}{80}
$$
$$
\chi^2 = \frac{100}{20} + \frac{100}{80} + \frac{100}{20} + \frac{100}{80} = 5 + 1.25 + 5 + 1.25 = 12.5
$$

4. Determine degrees of freedom:
\[

df = (2-1)(2-1) = 1
\]

5. Find critical value:
For \(df = 1\) at α = 0.05, the critical value from the chi-square table is 3.841.

6. Make a decision:
Since \(12.5 > 3.841\), we reject the null hypothesis.

Conclusion

There is a significant association between smoking and lung disease.

Conclusion

Understanding the chi-square test of independence through example problems allows researchers to analyze categorical data effectively. By following structured steps and applying statistical techniques, one can draw meaningful conclusions about relationships between variables. The examples provided demonstrate the application of this test in real-world scenarios, enhancing comprehension for those who seek to utilize this statistical method in their studies or professions.

Frequently Asked Questions

What is a chi-square test of independence?

The chi-square test of independence is a statistical method used to determine if there is a significant association between two categorical variables.

How do you set up a chi-square test of independence?

To set up a chi-square test of independence, you need to create a contingency table that displays the frequency counts of the categories of the two variables.

What are the steps to perform a chi-square test of independence?

The steps include: 1) State the null and alternative hypotheses, 2) Create a contingency table, 3) Calculate the expected frequencies, 4) Compute the chi-square statistic, and 5) Determine the p-value and make a conclusion.

Can you provide a simple example problem involving a chi-square test of independence?

Sure! Suppose we have survey data on students' preferences for two different study methods (A and

B) across two majors (Math and Science). We create a contingency table and perform the chi-square test to see if study method preference is independent of major.

What is the formula for calculating 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.

How do you interpret the results of a chi-square test?

If the p-value is less than the significance level (commonly 0.05), you reject the null hypothesis, concluding that there is a significant association between the two variables.

What are common mistakes to avoid when performing a chi-square test of independence?

Common mistakes include having expected frequencies less than 5 in any cell, not checking for independence conditions, and misinterpreting the p-value.

What are the assumptions of the chi-square test of independence?

The main assumptions are that the samples are independent, the data is categorical, and the expected frequency in each cell of the contingency table should be at least 5.

How can software be used to perform a chi-square test of independence?

Statistical software like R, SPSS, or Python libraries can perform chi-square tests by inputting the contingency table and using built-in functions to calculate the test statistic and p-value.

Chi Square Test Of Independence Example Problems With Answers

Chi Square Test Of Independence Example Problems With Answers

Related Articles

- classical conversations foundations guide 5th edition
- christian version of twas the night before christmas
- clarinet basics by paul harris

Back to Home