Statistical Analysis Using Chi-Square Test

March 7, 2025

1 Introduction

The chi-square test is a statistical method used to assess the differences between observed and expected categorical data distributions. This report outlines the results of chi-square tests applied to two datasets. The first test analyzes the distribution of zodiac signs among artists, while the second test compares calculated chi-square values to a critical chi-square threshold.

2 Chi-Square Test 1: Zodiac Signs and Artists

The first chi-square test was conducted to determine if the observed distribution of zodiac signs for artists significantly differs from an expected even distribution. This test helps assess whether certain zodiac signs are overrepresented or underrepresented among artists.

2.1 Data

The dataset contains 12 zodiac sign categories with the following observed values:

- Observed values: [12, 15, 14, 13, 10, 14, 16, 13, 11, 12, 14, 14]
- Expected values: Even distribution with 13 occurrences for each zodiac sign

2.2 Python Code

The chi-square test was implemented using Python's `scipy.stats.chisquare` function to compare the observed and expected frequencies. The code is provided below:

```
import numpy as np
from scipy.stats import chisquare

# Observed and expected values
```

```
observed = np.array([12, 15, 14, 13, 10, 14, 16, 13, 11, 12, 14, 14])
expected = np.full(12, 13)

# Perform chi-square test
chi2_stat, p_val = chisquare(f_obs=observed, f_exp=expected)
print(f"Chi-square Statistic: {chi2_stat:.4f}, p-value: {p_val:.4f}")
```

2.3 Results

The chi-square test results were as follows:

- Chi-square statistic: 6.5547
- p-value: 0.8339

Since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis. This indicates that there is no significant difference between the observed distribution of zodiac signs among artists and the expected even distribution.

3 Chi-Square Test 2: Critical Value Comparison

In the second chi-square test, two calculated chi-square values were compared against a critical chi-square value. The goal was to determine whether the observed values significantly differed from expected values based on this comparison.

3.1 Data

The chi-square values and the degrees of freedom for this test are as follows:

- Chi-square value 1: 12.94
- Chi-square value 2: 1.53
- Degrees of freedom: 5
- Significance level: 0.05

3.2 Python Code

To determine the critical chi-square value, the following Python code was used:

```
from scipy.stats import chi2

# Calculate critical chi-square value
critical_value = chi2.ppf(0.95, df=5)
print(f"Critical Chi-square value: {critical_value:.4f}")
```

3.3 Results

The critical chi-square value was calculated as:

- Critical chi-square value (at 5 degrees of freedom and 0.05 significance level): 11.07

The results were interpreted as follows:

- Chi-square value 1 (12.94) is greater than the critical value (11.07), leading to the rejection of the null hypothesis. This indicates a significant difference between the observed and expected distributions.
- Chi-square value 2 (1.53) is less than the critical value (11.07), so the null hypothesis was not rejected, indicating no significant difference.

4 Conclusion

The chi-square tests conducted in this report provided insights into the distribution of categorical data. In the first test, we found no significant difference between the observed and expected distributions of zodiac signs among artists. In the second test, one chi-square value was found to be significantly different from the critical value, while the other was not.

These results demonstrate the utility of chi-square tests for analyzing categorical data and identifying statistically significant deviations from expected distributions.