Pearson's Chi Square Test

December 16, 2025

Contents

1 For $2 \times n$ contingency tables

```
data <- matrix(c(50, 20, 30, 40, 10, 40), nrow = 2,
               dimnames = list(Treatment = c("Drug A", "Drug B"),
                               Outcome = c("Cured", "No Change", "Worse")))
data
```

```
##           Outcome
## Treatment Cured No Change Worse
##    Drug A    50        30    10
##    Drug B    20        40    40
```

```
test <- chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 31.848, df = 2, p-value = 1.214e-07
```

```
print(test$stdres)
```

```
##           Outcome
## Treatment     Cured  No Change    Worse
##    Drug A  5.073012 -0.9511897 -4.51523
##    Drug B -5.073012  0.9511897  4.51523
```

Interpreting the p-value:

The distributions across columns are different or unequal or not just due to chance (randomness) since p-value < 0.05.

1.1 Interpreting the Residuals

- Drug A / Cured (+5.1): Positive and large. Drug A is strongly associated with being cured.
- Drug A / Worse (-4.5): Negative and large. Drug A is strongly protected against getting worse.
- Drug B: Shows the exact opposite pattern.

1.2 For small sample sizes

use `simulate.p.value = TRUE` in `chisq.test()`

Primarily when you receive the warning *"Chi-squared approximation may be incorrect"* which is usually triggered by sparse data.

2 For $1 \times n$ contingency tables

When your table has only 1 row (a single variable with counts across multiple categories), you are no longer testing for "independence" between two variables. Instead, you are performing a **Chi-Square Goodness-of-Fit Test**.

You are asking: *"Does the distribution of my single row of data match a specific theoretical distribution (like 50/50)?"*

```r
# Testing for Equal Proportions (Default)
# Data: Drug A outcomes (Cured=60, Not Cured=40)
outcomes <- c(60, 40)

# Test: Are these outcomes split 50/50?
test <- chisq.test(outcomes)

# Null: Probabilities are equal (0.5, 0.5)
# If p < 0.05, the split is significantly different from 50/50.
print(test)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  outcomes
## X-squared = 4, df = 1, p-value = 0.0455
```

```r
# Testing Specific Proportions
# Data: Drug A outcomes (Cured=45, Not Cured=55)
observed <- c(45, 55)

# Null: Cured = 30%, Not Cured = 70%
expected_probs <- c(0.3, 0.7)

test_specific <- chisq.test(x = observed, p = expected_probs)
print(test_specific)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed
## X-squared = 10.714, df = 1, p-value = 0.001063
```