Comparing Proportions

Last modified: 2026-09-29 00:05:17 (PDT)

This page reviews tests for comparing groups on a categorical outcome: the chi-square test and Fisher's exact test for contingency tables. It uses the chi-square reference distribution defined on the Statistical Inference page. This page is adapted from Vittinghoff et al. (2012), Chapter 3.

The HERS data

The examples on this page use the HERS data, which the Comparing Means page describes. The `rmb` R package includes the dataset; `haven::as_factor()` converts its Stata value labels to factors:

```
hers <- rmb::hers |> haven::as_factor()
```

Comparing two groups: categorical outcomes

Contingency tables

> **Example 0.1** (Exercise by treatment group in HERS)**.** Table 1 cross-tabulates regular exercise at baseline by treatment group, as a contingency table with row percentages.

Table 1: Exercise by treatment group in HERS

```
hers |>
  gtsummary::tbl_cross(
    row = exercise,
    col = HT,
    label = list(
      exercise ~ "Exercises regularly",
      HT ~ "Treatment group"
    ),
    percent = "row"
  )
```

	Treatment group		
	placebo	hormone therapy	Total
Exercises regularly			
no	853 (50%)	842 (50%)	1,695 (100%)
yes	530 (50%)	538 (50%)	1,068 (100%)
Total	1,383 (50%)	1,380 (50%)	2,763 (100%)

The chi-square test

Definition 0.1 (Expected count under independence)**.** Let a contingency table have r rows and c columns, with observed count O_{ij} in row i and column j, row totals R_i, column totals C_j, and grand total n. The **expected count** in cell (i, j) under independence is

$$E_{ij} \stackrel{\text{def}}{=} \frac{R_i \, C_j}{n}.$$

Example 0.2 (Expected count in a 2 x 2 table)**.** In a table with $n = 100$, first-row total $R_1 = 30$, and first-column total $C_1 = 40$, the expected count in cell $(1, 1)$ is $E_{11} = 30 \cdot 40/100 = 12$.

Definition 0.2 (Pearson's chi-square test of independence)**.** With observed counts O_{ij} and expected counts E_{ij} in an $r \times c$ contingency table, **Pearson's chi-square test** of the null hypothesis that the row and column variables are independent uses the statistic

$$X^2 \stackrel{\text{def}}{=} \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Its p-value is $\Pr(W \geq X^2)$, where W has the $\chi^2_{(r-1)(c-1)}$ distribution (chi-square distribution).

Theorem 0.1 (Large-sample null distribution of the chi-square statistic)**.** *Let n observations be sampled independently and classified by two categorical variables, and let the two variables be independent. Then as $n \to \infty$, the distribution of X^2 (Definition 0.2) converges to the $\chi^2_{(r-1)(c-1)}$ distribution (Hogg et al. 2019, sec. 9.2, p. 440).*

The chi-square approximation is poor when some expected counts are small. A common rule of thumb asks for every E_{ij} to be at least 5.

Example 0.3 (Chi-square test of exercise by treatment group in HERS)**.** The expected counts and statistic of Definition 0.2 for Table 1:

```
observed <- table(hers$exercise, hers$HT)
expected <- outer(rowSums(observed), colSums(observed)) / sum(observed)
x_squared <- sum((observed - expected)^2 / expected)
expected
#>     placebo hormone therapy
#> no   848.42          846.58
#> yes  534.58          533.42
c(
  x_squared = x_squared,
  p_value = pchisq(x_squared, df = 1, lower.tail = FALSE)
)
#> x_squared   p_value
#>  0.128054  0.720458
```

Every expected count is large, so the χ^2_1 approximation of Theorem 0.1 is reasonable. `chisq.test()` with `correct = FALSE` reports the same values:

```
chisq.test(hers$exercise, hers$HT, correct = FALSE)
#>
#>  Pearson's Chi-squared test
#>
#> data:  hers$exercise and hers$HT
#> X-squared = 0.1281, df = 1, p-value = 0.72
```

The p-value is large: the data give no evidence that exercise depends on treatment group, as randomization would lead us to expect.

For a 2×2 table, `chisq.test()` applies Yates' continuity correction by default, which subtracts 0.5 from each $|O_{ij} - E_{ij}|$ before squaring, and so gives a smaller statistic than Definition 0.2. `correct = FALSE` turns the correction off.

Fisher's exact test

Definition 0.3 (Fisher's exact test)**.** Take a 2×2 contingency table with cells a, b, c, and d as in the contingency table definition, and hold its row and column totals fixed. Under independence of the row and column variables, the probability that the top-left cell equals x is the hypergeometric probability

$$p(x) \stackrel{\text{def}}{=} \frac{\binom{a+b}{x}\binom{c+d}{a+c-x}}{\binom{n}{a+c}}.$$

Fisher's exact test of independence has p-value

$$\sum_{x\,:\,p(x) \le p(a)} p(x),$$

the total probability of the tables with the same totals that are no more probable than the observed table.

The p-value is exact: it comes from the null distribution itself, not from a large-sample approximation, so the test is valid even when expected counts are small, and it is often used for 2×2 tables in which some expected count is below 5 (Theorem 0.1). Other two-sided versions exist; this one is the version that R's `fisher.test()` computes.

Example 0.4 (Fisher's exact test of exercise by treatment group in HERS)**.** The p-value of Definition 0.3 for Table 1, computed from the hypergeometric probabilities with `dhyper()`:

```r
a <- observed[1, 1]
row1 <- sum(observed[1, ])
row2 <- sum(observed[2, ])
col1 <- sum(observed[, 1])
x <- max(0, col1 - row2):min(row1, col1)
p_x <- dhyper(x, m = row1, n = row2, k = col1)
p_a <- dhyper(a, m = row1, n = row2, k = col1)
sum(p_x[p_x <= p_a * (1 + 1e-7)])
#> [1] 0.72528
```

The tolerance `1e-7` keeps tables whose probability equals $p(a)$ up to rounding error, as `fisher.test()` does. `fisher.test()` reports the same p-value:

```r
fisher.test(hers$exercise, hers$HT)
#>
#>  Fisher's Exact Test for Count Data
#>
#> data:  hers$exercise and hers$HT
#> p-value = 0.725
#> alternative hypothesis: true odds ratio is not equal to 1
#> 95 percent confidence interval:
#>  0.879664 1.202192
#> sample estimates:
#> odds ratio
#>    1.02836
```

With counts this large, the exact p-value is close to the chi-square p-value of Example 0.3.

Measures of association for 2×2 tables

Tests of independence say whether two binary variables are associated, but not how strongly. Risk differences, risk ratios, and odds ratios measure the strength of the association; see Odds Ratios and Relative Risks.

References

Hogg, Robert V., Elliot A. Tanis, and Dale L. Zimmerman. 2019. *Probability and Statistical Inference*. Tenth edition. Pearson.

Vittinghoff, Eric, David V Glidden, Stephen C Shiboski, and Charles E McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. Springer. https://doi.org/10.1007/978-1-4614-1353-0.