

# Effect-size formulas

## Purpose

`testflow` reports one primary effect-size estimate with each workflow when an effect size is defined. The formulas below describe the implemented estimators. They are written to match the package code, including the direction of signed effects.

## Cohen's d

### One-sample Cohen's d

For a numeric sample  $x_1, \dots, x_n$  compared with reference value  $\mu_0$ , `test_one_sample()` reports:

$$d = \frac{\bar{x} - \mu_0}{s_x}$$

where  $\bar{x}$  is the sample mean and  $s_x$  is the sample standard deviation, both computed after removing missing values.

### Independent-groups Cohen's d

For two independent groups  $x$  and  $y$ , `test_two_groups()` reports Cohen's d on the Student and Welch t-test branches:

$$d = \frac{\bar{x} - \bar{y}}{s_p}$$

with pooled standard deviation:

$$s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}$$

The sign follows the group order used internally by the workflow: the first group minus the second group.

### Paired-sample Cohen's dz

For paired measurements, `test_paired()` computes each paired difference as:

$$d_i = after_i - before_i$$

and reports Cohen's  $d_z$ :

$$d_z = \frac{\bar{d}}{s_d}$$

where  $\bar{d}$  and  $s_d$  are the mean and standard deviation of the paired differences.

## ANOVA-style effect sizes

### Eta squared

For one-way and factorial ANOVA workflows, `testflow` reports eta squared for the selected ANOVA term:

$$\eta_j^2 = \frac{SS_j}{\sum SS}$$

where  $SS_j$  is the term sum of squares and  $\sum SS$  is the sum of all ANOVA-table sums of squares, including residual variation.

For repeated-measures ANOVA, the implemented repeated-time eta squared is:

$$\eta_{time}^2 = \frac{SS_{time}}{SS_{total}}$$

with:

$$SS_{total} = \sum_i (y_i - \bar{y})^2$$

and:

$$SS_{time} = \sum_t n_t (\bar{y}_t - \bar{y})^2$$

where  $n_t$  is the number of observations at time  $t$ ,  $\bar{y}_t$  is the mean at time  $t$ , and  $\bar{y}$  is the grand mean.

### Kruskal-Wallis epsilon squared

For the non-parametric branch of `test_groups()`, the reported effect is:

$$\epsilon^2 = \frac{H - k + 1}{n - k}$$

where  $H$  is the Kruskal-Wallis statistic,  $k$  is the number of groups, and  $n$  is the number of complete observations.

## Categorical and rank-based effect sizes

### Cramer's V

For categorical association workflows, `testflow` reports:

$$V = \sqrt{\frac{\chi^2}{n(\min(r, c) - 1)}}$$

where  $\chi^2$  is the Pearson chi-square statistic,  $n$  is the table total,  $r$  is the number of rows, and  $c$  is the number of columns.

## Rank-biserial correlation

For the Wilcoxon branch of `test_two_groups()`, the reported rank-biserial correlation is:

$$r_{rb} = 1 - \frac{2W}{n_1 n_2}$$

where  $W$  is the `stats::wilcox.test()` statistic for the first group after R's conversion to the Mann-Whitney  $U$  scale, and  $n_1$  and  $n_2$  are the non-missing group sizes. The sign follows the same first-group reference.

## Kendall's W

For Friedman repeated numeric workflows, `testflow` reports:

$$W = \frac{\chi_F^2}{n(k-1)}$$

where  $\chi_F^2$  is the Friedman statistic,  $n$  is the number of complete subjects, and  $k$  is the number of repeated measurements.

For Cochran Q repeated categorical workflows, the implemented analogous effect is:

$$W = \frac{Q}{n(k-1)}$$

where  $Q$  is the Cochran Q statistic.

## Other reported workflow summaries

Some workflows expose a scalar summary in the effect-size field when a standard Cohen-style effect size is not used:

- `test_proportion()` reports the observed success proportion:

$$\hat{p} = \frac{x}{n}$$

- `test_paired_categorical()` reports the number of discordant pairs:

$$b + c$$

- `test_multinomial()` reports the chi-square goodness-of-fit statistic:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- `test_outliers()` reports the number of rows flagged by the selected outlier rule.

## Magnitude labels

Magnitude labels are descriptive thresholds used consistently by the package:

- Cohen's  $d$ : negligible if  $|d| < 0.2$ , small if  $|d| < 0.5$ , moderate if  $|d| < 0.8$ , otherwise large.
- Eta squared and epsilon squared: negligible if the estimate is  $< 0.01$ , small if  $< 0.06$ , moderate if  $< 0.14$ , otherwise large.
- Cramer's V, rank-biserial absolute value, and Kendall's W: negligible if  $< 0.1$ , small if  $< 0.3$ , moderate if  $< 0.5$ , otherwise large.