

HOW PRISM 8.0 COMPUTES MULTIPLE COMPARISONS TESTS FOLLOWING ANOVA (ONE-WAY, TWO-WAY AND THREE-WAY) OR CORRESPONDING MIXED-EFFECTS MODEL

Prism 8.0 can perform many kinds of multiple comparisons testing. All are standard, so there should be no surprises in this document, which reviews exactly how Prism does the calculations.

This document was created September 2018. Updates will be posted to:
<http://www.graphpad.com/faq/viewfaq.cfm?faq=1688>

CONTENTS

How the family is defined?	2
One-way ANOVA.....	2
Two-way ANOVA	2
Three-way ANOVA.....	5
How the appropriate error term is defined (DF(Error), MS(Error) or SE(Diff))?	5
Sphericity.....	5
Mixed-effects model (all ANOVAs)	6
One-way ANOVA.....	7
Two-way ANOVA	7
Three-way ANOVA.....	10
Multiple comparisons tests.....	15
Correct for multiple comparisons using statistical hypothesis testing.....	16
Bonferroni test	16
Tukey test.....	16
Dunnet test	16

Newman-Keuls test	2
.....	17
Holm-Sidak test	17
Correct for multiple comparisons by controlling the False Discovery Rate.....	18
Original method of Benjamini and Hochberg.....	18
Two-stage step-up method of Benjamini, Krieger and Yekutieli.....	19
Corrected method of Benjamini & Yekutieli.....	19
Don't correct for multiple comparisons. Each comparison stands alone	19
Fisher's LSD (Least Significant Difference) test	19
Read more	19

HOW THE FAMILY IS DEFINED?

ONE-WAY ANOVA

Compare each column mean with every other column mean

Number of families = 1

Number of comparisons within family = $N * (N - 1)/2$, where N is the number of data set columns

Compare each column mean with the control column mean

Number of families = 1

Number of comparisons within family = $(N - 1)$, where N is the number of data set columns.

Compare selected column means

Number of families = 1

User selects precisely which comparisons to make, so specifies the number of comparisons.

TWO-WAY ANOVA

Compare column means (main column effect)

Number of families = 1

Compare each column mean with every other column mean

Number of comparisons within family = $N * (N - 1)/2$, where N is the number of levels of column factor.

Compare each column mean with the control column mean

Number of comparisons within family = $(N - 1)$, where N is the number of levels of column factor.

Compare row means (main row effect)

Number of families = 1

Compare each row mean with every other row mean

Number of comparisons within family = $N * (N - 1)/2$, where N is the number of levels of row factor.

Compare each row mean with the control row mean

Number of comparisons within family = $(N - 1)$, where N is the number of levels of row factor.

Within each row, compare columns (simple effect within rows)

The user can choose to define all the comparisons to be one family, or to create one family per row.

One family for all comparisons (conservative)

Number of families = 1

Compare each cell mean with every other cell mean on that row

Number of comparisons within family = $M * (N * (N - 1)/2)$, where M is the number of levels of row factor and N is the number of levels of column factor.

Compare each cell mean with the control cell mean on that row

Number of comparisons within each family = $M * (N - 1)$, where M is the number of levels of row factor and N is the number of levels of column factor.

One family per row (recommended)

Number of families = number of levels of row factor.

Compare each cell mean with every other cell mean on that row

Number of comparisons within each family = $N * (N - 1)/2$, where N is the number of levels of column factor.

Compare each cell mean with the control cell mean on that row

Number of comparisons within each family = $(N - 1)$, where N is the number of levels of column factor.

Within each column, compare rows (simple effect within columns)

The user can choose to define all the comparisons to be one family, or to create one family per column.

One family for all comparisons (conservative)

Number of families = 1

Compare each cell mean with every other cell mean on that column

Number of comparisons within family = $M * (N * (N - 1) / 2)$, where M is the number of levels of column factor and N is the number of levels of row factor.

Compare each cell mean with the control cell mean on that column

Number of comparisons within each family = $M * (N - 1)$, where M is the number of levels of column factor and N is the number of levels of row factor.

One family per column (recommended)

Number of families = number of levels of column factor.

Compare each cell mean with every other cell mean on that column

Number of comparisons within each family = $N * (N - 1) / 2$, where N is the number of levels of row factor.

Compare each cell mean with the control cell mean on that column

Number of comparisons within each family = $(N - 1)$, where N is the number of levels of row factor.

Compare each cell mean with the other cell mean in that row

Number of families = 1

Number of comparisons within family = the number of levels of row factor.

Compare each cell mean with the other cell mean in that column

Number of families = 1

Number of comparisons within family = the number of levels of column factor.

Compare all cell means regardless of rows and columns

Number of families = 1

Compare each cell mean with every other cell mean

Number of comparisons within family = $N * (N - 1) / 2$, where N is the number of levels of row factor multiplied by the number of levels of column factor.

Compare each cell mean with the control cell mean

Number of comparisons within family = $(N - 1)$, where N is the number of levels of row factor multiplied by the number of levels of column factor.

THREE-WAY ANOVA

Compare each cell mean with every other cell mean

Number of families = 1

Number of comparisons within family = $N * (N - 1)/2$, where N is the number of levels of row factor multiplied by the number of levels of column factor.

Compare the control mean (A1) with the other cell means

Number of families = 1

Number of comparisons within family = $(N - 1)$, where N is the number of cells.

Compare means of cells that differ by only one factor

Number of families = 1

Number of comparisons within family = $3 * N * M * K - A - B - C$, where N is the number of levels of row factor, M is the number of levels of (AB vs CD) factor and K is the number of levels of (AC vs BD) factor (12 for 2 x 2 x 2 design).

Compare each cell mean in row 1 with the cell mean just below

Number of families = 1

Number of comparisons within family = number of columns multiplied by the number of levels of row factor (4 for 2 x 2 x 2 design).

Compare each row mean with every other row mean

Number of families = 1

Number of comparisons within family = $N * (N - 1)/2$, where N is the number of levels of row factor.

Compare each row mean with the control row mean

Number of families = 1

Number of comparisons within family = $(N - 1)$, where N is the number of levels of row factor.

HOW THE APPROPRIATE ERROR TERM IS DEFINED (DF(ERROR), MS(ERROR) OR SE(DIFF))?

ASSUME SPHERICITY (EQUAL VARIABILITY OF DIFFERENCES)

IF YOU ASSUME SPHERICITY

The multiple comparisons tests performed by Prism use the mean square error for all comparisons. This is a pooled value that assess variability in all the groups. If you assume that variability really is the same in all groups (with any differences due to chance) this gives you more power.

IF YOU DO NOT ASSUME SPHERICITY

If you check the option to not assume sphericity, Prism does two things differently.

- It applies the Geisser-Greenhouse correction when computing the P values for the main effect.
- It computes the multiple comparisons differently. For each comparison of two groups, it uses only the data in those two groups, so has less power than the first method if the assumption of sphericity is true, especially with small samples. This makes sense when scatter increases with time, so later treatments give a more variable response than earlier treatments. See (Maxwell & Delaney, Second edition), pages 552-555.
- When two groups are matched, it is identical to a paired t test and has fewer degrees of freedom $DF(\text{Error}) = N - 1$, where N is the number of paired subjects (e.g. rows in one-way RM ANOVA). When there are no missing values in both samples, $N = N_1 = N_2$ where N_1 and N_2 are sample sizes

$SD_{Pooled} = \sqrt{s^2/DF(\text{Error})}$, where

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_{1i} - y_{2i} - \text{Mean diff})^2$$

and

$$SEDifference = SD_{Pooled} \cdot (1/\sqrt{N_1} + 1/\sqrt{N_2})$$

If one element in some pair (y_{1i} , y_{2i}) is missing, the pair is excluded.

- When two groups are unmatched, it is identical to an unpaired t test, i.e. $DF(\text{Error}) = N_1 + N_2 - 2$, where N_1 and N_2 are sample sizes

$SD_{Pooled} = \sqrt{\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{DF(\text{Error})}}$, where s_1^2 and s_2^2 are sample variances, i.e. $s_k^2 = \frac{1}{N_k-1} \sum_{i=1}^{N_k} (y_{ki} - \text{Mean}_k)^2$

and

$$SEDifference = SD_{Pooled} \cdot (1/\sqrt{N_1} + 1/\sqrt{N_2})$$

We assume that sphericity assumption holds true for all models described below.

MIXED-EFFECTS MODEL (ALL ANOVAS)

For multiple comparisons, the predicted mean difference is computed as

$$\text{Predicted mean diff} = \mathbf{L} \cdot \boldsymbol{\beta}, \quad (1)$$

and the standard error of the difference is computed as

$$SEDifference = \sqrt{\mathbf{L} \cdot \mathbf{C} \cdot \mathbf{L}'}, \quad (2)$$

where \mathbf{L} is a vector/matrix of contrasts, $\boldsymbol{\beta}$ is an estimated vector of fixed-effects parameters and \mathbf{C} is a variance-covariance matrix at best-fit values. $MS(\text{Error})$ is not calculated for mixed-effects models.

If variance of some random effect is zero and simpler mixed-effects model is fitted, DF of that term is considered a 0. If variances of all random effect are zero, then the simpler mixed-effects model is the corresponding ordinary ANOVA.

ONE-WAY ANOVA

ORDINARY (NOT REPEATED MEASURES)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

REPEATED MEASURES

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Yes, it is the same equation as for ordinary (not repeated measures) ANOVA, but the MS(residual) is computed differently with repeated measures.

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual})$$

TWO-WAY ANOVA

ORDINARY (NOT REPEATED MEASURES)

For all kind of comparisons (main effects, simple effects) the MS(Error) and DF(Error) are defined as follows:

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

REPEATED MEASURES (MATCHED VALUES ARE SPREAD ACROSS A ROW)

COMPARE COLUMN MEANS (MAIN COLUMN EFFECT)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual})$$

WITHIN EACH ROW, COMPARE COLUMNS (SIMPLE EFFECT WITHIN ROWS)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual})$$

COMPARE ROW MEANS (MAIN ROW EFFECT)

$$MS(\text{Error}) = MS(\text{Subjects})$$

$$DF(\text{Error}) = DF(\text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Subjects})$$

WITHIN EACH COLUMN, COMPARE ROWS (SIMPLE EFFECT WITHIN COLUMNS)

$$MS(\text{Error}) = (SS(\text{Subjects}) + SS(\text{Residuals})) / (DF(\text{Subjects}) + DF(\text{Residuals}))$$

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Subjects})$$

REPEATED MEASURES (MATCHED VALUES ARE STACKED INTO A SUBCOLUMN)

COMPARE COLUMN MEANS (MAIN COLUMN EFFECT)

$$MS(\text{Error}) = MS(\text{Subjects})$$

$$DF(\text{Error}) = DF(\text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Subjects})$$

WITHIN EACH ROW, COMPARE COLUMNS (SIMPLE EFFECT WITHIN ROWS)

$$MS(\text{Error}) = (SS(\text{Subjects}) + SS(\text{Residuals})) / (DF(\text{Subjects}) + DF(\text{Residuals}))$$

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Subjects})$$

COMPARE ROW MEANS (MAIN ROW EFFECT)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual})$$

WITHIN EACH COLUMN, COMPARE ROWS (SIMPLE EFFECT WITHIN COLUMNS)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual})$$

REPEATED MEASURES (MATCHED VALUES ARE BOTH STACKED AND SPREAD ACROSS A ROW)

COMPARE COLUMN MEANS (MAIN COLUMN EFFECT)

$$MS(\text{Error}) = MS(\text{Column Factor} \times \text{Subjects})$$

$$DF(\text{Error}) = DF(\text{Column Factor} \times \text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Column Factor} \times \text{Subjects})$$

WITHIN EACH ROW, COMPARE COLUMNS (SIMPLE EFFECT WITHIN ROW)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Column Factor} \times \text{Subjects})$$

COMPARE ROW MEANS (MAIN ROW EFFECT)

$$MS(\text{Error}) = MS(\text{Row Factor} \times \text{Subjects})$$

$$DF(\text{Error}) = DF(\text{Row Factor} \times \text{Subjects})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Row Factor} \times \text{Subjects})$$

WITHIN EACH COLUMN, COMPARE ROWS (SIMPLE EFFECT WITHIN COLUMN)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

Mixed-effects model

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Row Factor} \times \text{Subjects})$$

COMPARE CELL MEANS

$$MS(\text{Error}) = MS(\text{Column Factor} \times \text{Row Factor} \times \text{Subjects}) = MS(\text{Residuals})$$

$$DF(\text{Error}) = DF(\text{Column Factor} \times \text{Row Factor} \times \text{Subjects}) = DF(\text{Residuals})$$

Mixed-effects model

If compared cells are from the same row

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Column Factor} \times \text{Subjects})$$

If compared cells are from the same column

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Row Factor} \times \text{Subjects})$$

If compared cells are from different columns and rows

$$DF(\text{Error}) = DF(\text{Residual}) + DF(\text{Column Factor} \times \text{Subjects}) + DF(\text{Row Factor} \times \text{Subjects})$$

THREE-WAY ANOVA

MS(Residual), MS(Subject), MS(Effect x Subject) are values from ANOVA table while Variance(Residual), Variance(Subject), Variance(Effect x Subject) are values from Random effects of Mixed-effects analysis.

In general, MS(Residual) = Variance(Residual), but Variance(Subject) and Variance(Effect x Subject) are point estimates that are not equal to MS(Subject) and MS(Effect x Subject) respectively.

ORDINARY (NOT REPEATED MEASURES)

For all kind of comparisons (main effects, simple effects) the MS(Error) and DF(Error) are defined as follows:

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

REPEATED MEASURES (MATCHING BY FACTOR: AB VS CD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched, e.g. A1 vs. C1, B1 vs. D1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. A2, A1 vs. B1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject})$$

$$DF(\text{Error}) = DF(\text{Subject})$$

REPEATED MEASURES (MATCHING BY FACTOR: AC VS BD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched, e.g. A1 vs. B1, C1 vs. D1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. A2, A1 vs. C1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject})$$

$$DF(\text{Error}) = DF(\text{Subject})$$

REPEATED MEASURES (MATCHING BY FACTOR: ROW FACTOR), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched, e.g. A1 vs. A2, B1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. B1, A1 vs. C2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Residual})$$

REPEATED MEASURES (MATCHING BY FACTORS: AB VS CD & AC VS BD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched by the factor AB vs CD, e.g. A1 vs. C1, B2 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Residual})$$

If compared cells are matched by the factor AC vs BD, e.g. A1 vs. B1, C2 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are matched by the interaction (AB vs CD) x (AC vs BD), e.g. A1 vs. D1, B2 vs. C2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. A2, A1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject})$$

$$DF(\text{Error}) = DF(\text{Subject})$$

REPEATED MEASURES (MATCHING BY FACTORS: ROW FACTOR & AB VS CD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched by Row factor, e.g. A1 vs. A2, B1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor}) + DF(\text{Residual})$$

If compared cells are matched by the factor AB vs CD, e.g. A1 vs. C1, B2 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Residual})$$

If compared cells are matched by the interaction Row factor x (AB vs CD), e.g. A1 vs. C2, B1 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. B1, A1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject} \times \text{Row factor})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor})$$

REPEATED MEASURES (MATCHING BY FACTORS: ROW FACTOR & AC VS BD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched by the factor Row factor, e.g. A1 vs. A2, B1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor}) + DF(\text{Residual})$$

If compared cells are matched by the factor AC vs BD, e.g. A1 vs. B1, C2 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are matched by the interaction Row factor x (AC vs BD), e.g. A1 vs. B2, C1 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are unmatched, e.g. A1 vs. C1, A1 vs. D2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject}) + \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject}) + DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Residual})$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject} \times \text{Row factor})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor})$$

REPEATED MEASURES (MATCHING BY FACTORS: ROW FACTOR & AB VS CD & AC VS BD), INCLUDING MIXED-EFFECTS MODEL

COMPARE CELL MEANS (EACH CELL MEAN WITH EVERY OTHER CELL MEAN, THE CONTROL MEAN (A1) WITH THE OTHER CELL MEANS, DIFFER BY ONLY ONE FACTOR, EACH CELL MEAN IN ROW 1 WITH THE CELL MEAN JUST BELOW)

If compared cells are matched by the factor AB vs CD, e.g. A1 vs. C1, B1 vs. D1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + DF(\text{Residual})$$

If compared cells are matched by the factor AC vs BD, e.g. A1 vs. B1, C1 vs. D1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are matched by the factor Row factor, e.g. A1 vs. A2, B1 vs. B2, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times \text{Row factor}) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are matched by the interaction (AB vs CD) x (AC vs BD), e.g. A1 vs. D1, B1 vs. C1, etc.

$$MS(\text{Error}) = \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual})$$

$$DF(\text{Error}) = DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual})$$

If compared cells are matched by the interaction Row factor x (AB vs CD), e.g. A1 vs. C2, B1 vs. D2, etc.

$$\begin{aligned} MS(\text{Error}) &= \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times \text{Row factor}) \\ &\quad + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) \\ &\quad + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual}) \end{aligned}$$

$$\begin{aligned} DF(\text{Error}) &= DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) \\ &\quad + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual}) \end{aligned}$$

If compared cells are matched by the interaction Row factor \times (AC vs BD), e.g. A1 vs. B2, C1 vs. D2, etc.

$$\begin{aligned} MS(\text{Error}) &= \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor}) \\ &\quad + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) \\ &\quad + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual}) \end{aligned}$$

$$\begin{aligned} DF(\text{Error}) &= DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor}) + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) \\ &\quad + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual}) \end{aligned}$$

If compared cells are matched by the interaction Row factor \times (AB vs CD) \times (AC vs BD), e.g. A1 vs. D2, B1 vs. C2, etc.

$$\begin{aligned} MS(\text{Error}) &= \text{Variance}(\text{Subject} \times (\text{AB vs CD})) + \text{Variance}(\text{Subject} \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor}) \\ &\quad + \text{Variance}(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) \\ &\quad + \text{Variance}(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + \text{Variance}(\text{Residual}) \end{aligned}$$

$$\begin{aligned} DF(\text{Error}) &= DF(\text{Subject} \times (\text{AB vs CD})) + DF(\text{Subject} \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor}) \\ &\quad + DF(\text{Subject} \times (\text{AB vs CD}) \times (\text{AC vs BD})) + DF(\text{Subject} \times \text{Row factor} \times (\text{AB vs CD})) \\ &\quad + DF(\text{Subject} \times \text{Row factor} \times (\text{AC vs BD})) + DF(\text{Residual}) \end{aligned}$$

COMPARE ROWS (EACH ROW MEAN WITH EVERY OTHER ROW MEAN, EACH ROW MEAN WITH THE CONTROL ROW MEAN)

$$MS(\text{Error}) = MS(\text{Subject} \times \text{Row factor})$$

$$DF(\text{Error}) = DF(\text{Subject} \times \text{Row factor})$$

MULTIPLE COMPARISONS TESTS

Definitions:

Alpha = Significance level, set by user, often to 0.05.

K = Number of pairwise comparisons within family (defined above).

M = Number of different means within each family.

MS(Error) = Appropriate mean square error. Defined above.

DF(Error) = Number of degrees of freedom used in computing the MS(error), defined above.

N_1 and N_2 are calculated as follows:

Comparing column/row means

N_1 = number of all values within column/row 1

N_2 = number of all values within column/row 2

Compare cell means

N_1 = number of values within cell 1

N_2 = number of values within cell 2

Mean diff = $Mean_1 - Mean_2$ after ANOVA model or as in (1) after mixed-effects model.

$$SEDifference = \sqrt{MS(Error) * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \text{ after ANOVA model or as in (2) after mixed-effects model.}$$

Statistical functions: *PFromQTukey*, *QFromPTukey*, *PFromQDunnett*, *QFromPDunnett*, *PFromT*, and *TFromP* convert from statistical ratios to probabilities and from probabilities to ratios.

Adjusted vs. not adjusted P values, refers to adjustment for multiple comparisons. See:

<http://www.graphpad.com/faq/viewfaq.cfm?faq=1518>

CORRECT FOR MULTIPLE COMPARISONS USING STATISTICAL HYPOTHESIS TESTING

BONFERRONI TEST

$$t = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue(adjusted) = PFromT(t, DF(Error)) * K$$

$$CIDifference = |Mean\ diff| \pm TFromP\left(\frac{Alpha}{K}, DF(Error)\right) * SEDifference$$

ŠIDÁK TEST

$$t = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue(adjusted) = PFromT(t, DF(Error)) * (1 - (1 - Alpha)^{1/K})$$

$$CIDifference = |Mean\ diff| \pm TFromP\left((1 - (1 - Alpha)^{1/K}), DF(Error)\right) * SEDifference$$

TUKEY TEST

$$q = \frac{|Mean\ diff|}{\frac{1}{\sqrt{2}}SEDifference}$$

$$PValue(adjusted) = PFromQTukey(q, DF, M)$$

$$CIDifference = |Mean\ diff| \pm QFromPTukey(Alpha, DF, M) * SEDifference$$

DUNNET TEST

$$q = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue(adjusted) = PFromQDunnet(q, DF, M)$$

$$CIDifference = |Mean\ diff| \pm QFromPDunnet(Alpha, DF, M) * SEDifference$$

Note that for historical reasons, the standard error of the difference is defined differently for the Dunnett and Tukey tests. The Dunnett test definition lacks the square root of 2 term.

$$q = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue(adjusted) = PFromQDunnett(q, DF, M)$$

$$CIDifference = |Mean\ diff| \pm QFromPDunnett(Alpha, DF, M) * SEDifference$$

NEWMAN-KEULS TEST

Let's assume that family consists of 5 samples with the following means: M1 ... M5. Let's assume that after ranking (sorting in decreasing order) means are ordered as follows: M2, M5, M1, M4, M3.

Then the pairs are considered in the following order: largest vs. smallest, largest vs. next smallest ...largest vs. next largest, next largest vs. smallest ... next smallest vs. smallest.

In our example the order of pairs being considered looks like: (M3 vs. M2), (M3 vs. M5), (M3 vs. M1), (M3 vs. M4), (M4 vs. M2), (M4 vs. M5), (M4 vs. M1), (M1 vs. M2), (M1 vs. M5), (M5 vs. M2).

For each pair let's designate B as the number of Means within this range (e.g. for second pair (M3, M5) P = 4).

After that the pairwise comparisons are calculated as follows (in the order defined above):

$$q = \frac{|Mean\ diff|}{\frac{1}{\sqrt{2}}SEDifference}$$

$$PValue(adjusted) = PFromQTukey(q, DF, B)$$

The calculations are the same as Tukey test, with the one exception. The Tukey test is based on M (number of all means within family), while the Newman-Keuls test is based on B (the number of means within this range).

Another important rule is that if no difference is found between two means then it's concluded that no difference exists between any means enclosed by those two, and such differences are not tested for.

HOLM-SIDAK TEST

First for each pair within family unpaired t test is calculated as follows:

$$t = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue = PFromT(t, DF(Error))$$

Then p values (two-tailed) are sorted in increasing order and statistical significance is calculated by the following algorithm.

- 1) Start with the smallest P value. Set K equal to the number of comparisons.
- 2) Compare the P value to this value: $(1 - \text{Alpha})^{\frac{1}{K}}$
- 3) If the P value is less than or equal to the computed value, that difference is considered to be statistically significant. Go back to step 2, decrease the value of K by 1, and test the next largest P value .
- 4) If the P value is greater than or equal to the computed value, then that difference is considered to not be statistically significant. Don't test any more P values. All the rest are considered to be not statistically significant

CORRECT FOR MULTIPLE COMPARISONS BY CONTROLLING THE FALSE DISCOVERY RATE

Prism offers a choice of three algorithms for controlling the FDR. The three algorithms all work in a similar manner.

1. Calculate a P value for each comparison.
2. Rank the P values from low to high.
3. Start with the largest P value.
4. Compute a threshold for the largest P value. That threshold value depends on the number of P values being looked at. For the method of Benjamini, Krieger and Yekutieli, the threshold also depends on the estimate of the number of true null hypotheses provided by the method.
5. If the P value is less than the threshold, then all P values are flagged as discoveries, and you are done. Otherwise, continue.
6. Go to the second largest P value.
7. Compute a threshold for the second highest P value. This threshold will be smaller than the threshold for the largest P value. Computing the threshold (see below) depends on the rank of the P value, the number of P values and (for the method of Benjamini, Krieger and Yekutieli) the estimate of the number of true null hypotheses (computed by the method; nothing for you to think about).
8. If the P value is less than the threshold, that P value and all smaller ones are flagged as discoveries and you are done. Otherwise continue.
9. Go to the next P lower value.
10. Compute a threshold for this rank. It will be smaller than the previous threshold.
11. If the P value is less than the threshold, that P value and all smaller ones are flagged as discoveries and you are done. Otherwise repeat steps 9-10 until done.

The difference between the three methods is how they compute the threshold values.

Definitions:

Q is the desired false discovery rate (as a percentage)

N is the number of P values in the set

Ntrue is the number of the null hypotheses estimated to be true (part of the second method below)

Define q to equal Q/100. This converts the value you enter as a percentage into a fraction.

ORIGINAL METHOD OF BENJAMINI AND HOCHBERG

Threshold for smallest P value = q/n

Threshold for largest P value = q

See (Benjamini & Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, 1995)

TWO-STAGE STEP-UP METHOD OF BENJAMINI, KRIEGER AND YEKUTIELI

Threshold for smallest P value = $q/[(1+q)N_{true}]$

Threshold for largest P value = $[q/(1+q)] * (N / N_{true})$

See (Benjamini, Krieger, & Yekutieli, Adaptive linear step-up procedures that control the false discovery rate, 2006)

CORRECTED METHOD OF BENJAMINI & YEKUTIELI

Threshold for smallest P value = $q/[N * (1 + 1/2 + 1/3 + \dots + 1/N)]$

Threshold for largest P value = $q / (1 + 1/2 + 1/3 + \dots + 1/N)$

See (Benjamini & Yekutieli, The control of the false discovery rate in multiple testing under dependency, 2001)

DON'T CORRECT FOR MULTIPLE COMPARISONS. EACH COMPARISON STANDS ALONE

FISHER'S LSD (LEAST SIGNIFICANT DIFFERENCE) TEST

$$t = \frac{|Mean\ diff|}{SEDifference}$$

$$PValue(NotAdjusted) = PFromT(t, DF(Error))$$

$$CIDifference = |Mean\ diff| \pm TFromP(Alpha, DF(Error)) * SEDifference$$

Prism does the *unprotected* Fisher's LSD test, which means the individual comparisons are made without regard to the overall ANOVA results. In contrast, the *protected* test only does individual comparisons if the overall P value is less than a preset threshold.

READ MORE

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* , 289-300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* , 1165-1188.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93 , 491-507.

Maxwell, S. E., & Delaney, H. D. (Second edition). *Designing Experiments and Analyzing Data*. Lawrence Erlbaum, ISBN: 0-8058-3718-3.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer Verlag New York, LLC.

Difference between q and t ratios: <http://www.graphpad.com/faq/viewfaq.cfm?faq=540>

How to compute the critical values of q for Tukey and Dunnett test:

<http://www.graphpad.com/faq/viewfaq.cfm?faq=1517>

How to compute critical values of the t distribution: <http://www.graphpad.com/faq/viewfaq.cfm?faq=1431>

The post test for trend: <http://www.graphpad.com/faq/viewfaq.cfm?faq=1544>

Multiple comparisons after repeated measures ANOVA, with or without assuming sphericity:

<http://www.graphpad.com/faq/viewfaq.cfm?faq=1609>

Dunn's multiple comparison test following the nonparametric Kruskal-Wallis ANOVA:

<http://www.graphpad.com/faq/viewfaq.cfm?faq=1156>

All the methods are explained in (Maxwell & Delaney, Second edition)