The ANOVA 1-way analysis tests for equality between the means in a number of samples but, if the null hypothesis is rejected, it does not indicate for which samples the means are significantly different. In fact, for k samples there are $k(k-1)/2$ possible pairwise comparisons so alternative techniques have been developed. The SimFit ANOVA procedure provides the option to use the Tukey Q post-ANOVA test for this multiple comparison purpose.

Open the SimFit main menu, select the [Statistics] option, choose 1-way-ANOVA, indicate that untransformed data are to be used, then analyze the test file provided which is a data matrix contained in `anova.tf1`. This particular data set is for six replicate estimates for strontium concentrations (mg/ml) in five different locations, and it is wished to test if there are significant differences between the mean levels as listed in the last row.

	28.2	39.6	46.3	41.0	56.3
	33.2	40.8	42.1	44.1	54.1
	36.4	37.9	43.5	46.4	59.4
	34.6	37.1	48.8	40.2	62.7
	29.1	43.6	43.7	38.6	60.0
	31.0	42.4	40.1	36.3	57.3
Means	32.1	40.2	44.1	41.1	58.3

The results are as follows.

1-Way Analysis of Variance: Grand Mean 43.16
Transformation: x (untransformed data)

Source	SSQ	NDOF	MSQ	F	p
Between Groups	2193	4	548.4	56.15	0.0000
Residual	244.1	25	9.765		
Total	24383	29			

Clearly the null hypothesis of equal column means must be rejected at the 1% significance level as $p < 0.01$ so the following results were obtained for the Tukey test.

Tukey Q-test with 5 means and 10 comparisons
5% point = 4.189, 1% point = 5.125

Column	Column	Q	p	5%	1%	n_B	n_A
5	1	20.55	0.0001	*	*	6	6
5	2	14.16	0.0001	*	*	6	6
5	4	13.48	0.0001	*	*	6	6
5	3	11.14	0.0001	*	*	6	6
3	1	9.406	0.0001	*	*	6	6
3	2	3.018	0.2377	NS	NS	6	6
3	4	[[2.338	0.4792]]	No-Test	No-Test	6	6
4	1	7.068	0.0005	*	*	6	6
4	2	[[0.6793	0.9885]]	No-Test	No-Test	6	6
2	1	6.388	0.0013	*	*	6	6

[5%] and/or [[1%]] No-Test results given for reference only

In this table, columns where means differ significantly are indicated by *, columns where means are not significantly different are indicated by NS, and columns that were not tested are indicated by No-test with hypothetical p values in square brackets.

Note that, for the Tukey test, the means are ranked and columns with means between those of extreme columns that differ significantly are not tested, according to the protocol that is recommended for this test. This involves a systematic procedure where the largest mean is compared to the smallest, then the largest mean is compared with the second largest, and so on. If no difference is found between two means then it is concluded that no difference exists between any means enclosed by these two, and so no testing is done.

The test statistic Q for comparing columns A and B with means $\bar{y}_A$, $\bar{y}_B$ and sample sizes n_A, n_B is

$$Q = \frac{\bar{y}_B - \bar{y}_A}{SE}$$

$$\text{where } SE = \sqrt{\frac{s^2}{n}}, \text{ if } n = n_A = n_B$$

$$SE = \sqrt{\frac{s^2}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}, \text{ if } n_A \neq n_B$$

$$s^2 = \text{ error } MS$$

and the significance level for Q is calculated as a studentized range.

Evidently, for these data, we reach the conclusion that

- data in column 5 differs significantly from data in columns 1, 2, 3, and 4,

- data in column 3 differs significantly from data in column 1, and

- data in column 1 differs significantly from data in columns 2 and 4.

This conclusion is also fairly obvious from the plot of means with error bars representing the 95% confidence limits as follows.

Data and Error-Bars