Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

If it is clear that the data are not normally distributed with the same variance then it is possible to perform the nonparametric Kruskal-Wallis test, either alone, or at the same time as 1-way ANOVA to compare the results. Of course, it would be usual to pre-determine which result to accept, otherwise the Bonferroni principle would have to be used.

Example 1

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			

Kruskal-Wallis Nonparametric One Way Analysis of Variance

Test statistic	NDOF	p
23.30	4	0.0001

Clearly the null hypothesis of equal column means and medians must be rejected at the 1% significance level as $p < 0.01$ for both parametric and nonparameteric 1-way ANOVA.

Example 2

The sample sizes need not be identical for 1-way ANOVA, and the next case to be considered is where there are 5 groups of sizes 5, 8, 6, 8, and 8 for weight gain in pounds of pigs from 5 different litters.

23	29	38	30	31
27	25	31	27	33
26	33	28	28	31
19	36	35	22	28
30	32	33	33	30
	28	36	34	24
	30		34	29
	31		32	30

As the sample sizes differ, the data cannot be entered as a matrix this time, and must be entered as individual column vectors, from a project archive, or as a library file which simply holds the locations of individual data files for each of the columns.

So now repeat the above procedure, but this time select to supply a library file and input the test file `anova.TFL` which then reads in data from the test files `column1.tf1`, `column1.tf2`, ..., `column1.tf5`.

1-Way Analysis of Variance: Grand Mean 29.89
Transformation: x (untransformed data)

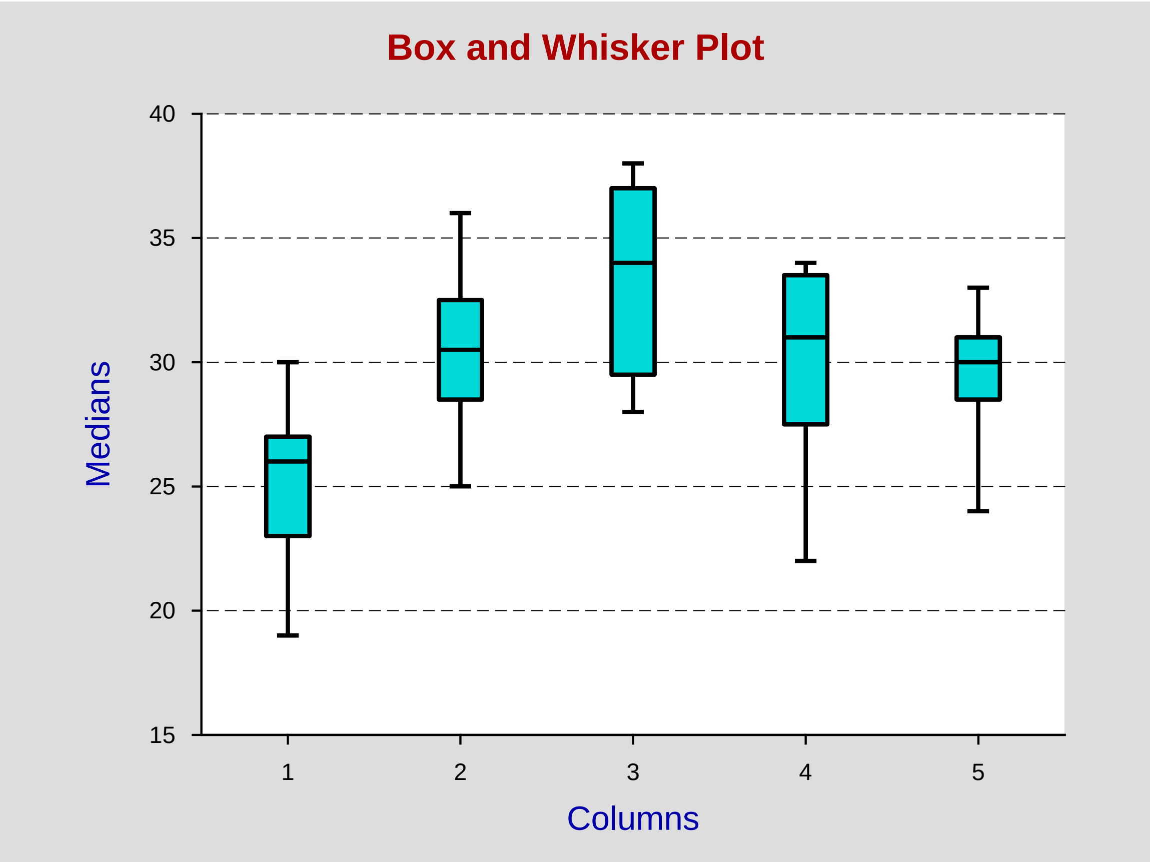
Source	SSQ	NDOF	MSQ	F	p
Between Groups	202.0	4	50.51	3.931	0.0111
Residual	385.5	30	12.85		
Total	587.5	34			

Kruskal-Wallis Nonparametric One Way Analysis of Variance

Test statistic	NDOF	p
10.54	4	0.0323

Clearly the null hypothesis of equal column means and medians must be rejected at the 5% significance level, but not the 1% level, as $p < 0.05$ for both the parametric and nonparameteric 1-way ANOVA.

There is yet another way to display 1-way ANOVA data as illustrated by the next plot.



This is more in keeping with a nonparametric test and displays the ranges and medians as a box and whisker plot, that is: the lowest value, the 25% point, the 50% median point, the 75% point, and the largest value.

The Kruskal-Wallis test

The null hypothesis for standard 1-way ANOVA is

H_0: The groups (i.e., column vectors) are from the same normal distribution,

while for the Kruskal-Wallis analysis of variance by ranks it is the weaker condition

H_0: The groups (i.e., column vectors) are from the same distribution.

The Kruskal-Wallis test is in reality an extension of the Mann-Whitney U test to k independent samples, and it is actually designed to test H_0: the medians are all equal.

The pooled sample is ranked, with tied scores assigned average ranks, then a test statistic H for k groups, each with n_i observations, is calculated as

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$
$$\text{where } n = \sum_{i=1}^{k} n_i$$

and R_i is the sum of the ranks of the n_i observations in group i.

This test is actually a 1-way ANOVA carried out on the ranks of the data. The p value are calculated exactly for small samples, but the fact that H approximately follows a χ^2_{k-1} distribution is used for large samples.

If there are ties, then H is corrected by dividing by λ defined as

$$\lambda = 1 - \frac{\sum_{i=1}^{m}(t_i^3 - t_i)}{n^3 - n}$$

where t_i is the number of tied scores in the ith group of ties, and m is the number of groups of tied ranks.

The test is $3/\pi$ (i.e, 95%) as powerful as the 1-way ANOVA test when the parametric test is justified, but it is more powerful, and should always be used if the assumptions of the linear normal model are not appropriate.

As it is unusual for the sample sizes to be large enough to verify that all the samples are normally distributed and with the same variance, rejection of H_0 in the Kruskal-Wallis test (which is the higher order analogue of the Mann-Whitney U test, just as 1-way ANOVA is the higher analogue of the t test) should always be taken seriously. This is one reason why SimFIT provides the convenient option to perform both 1-way ANOVA and the Kruskal-Wallis test at the same time.