



Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.

<https://simfit.org.uk>

<https://simfit.silverfrost.com>

It is frequently necessary to check for homogeneity of variance given n samples, that is to perform a statistical test to see if it reasonable to assume they have a common population variance. In particular, analysis of variance (ANOVA) is based on the assumption that all samples are normally distributed with the same variance.

Now it is often stated that ANOVA procedures are relatively insensitive to small departures from normality, but are much more affected by differences in variances between groups so, for that reason, variance-stabilizing transformations are frequently resorted to. Variance homogeneity tests are best done interactively on the data set, so that the effect of transformations on variance stabilizations can be judged before proceeding to ANOVA. SIMFIT provides the facility to read in samples as individual data sets with possibly differing sample sizes, or as a matrix if all samples have the same size, but then to test for homogeneity of variance under all conditions of variance stabilizing transformations in order to check that the transformation selected has succeeded.

The next table illustrates analysis of data in the test file `anova1.tff1` for homogeneity of variance, using the Bartlett test, and also the Levene test.

Homogeneity of variance test 1: Bartlett

Transformation	x (untransformed data)
B	0.69006
C	1.0800
B/C	0.63895
Number of degrees of freedom	4
$P(\chi^2 \geq B/C)$	0.9586
Upper tail 1% point	13.277
Upper tail 5% point	9.4877

Homogeneity of variance test 2: Levene (median)

Transformation	x (untransformed data)
W	0.18458
Degrees of freedom 1	4
Degrees of freedom 2	25
$P(F \geq W)$	0.9442
Upper tail 1% point	4.1774
Upper tail 5% point	2.7587

In order to interpret these results it is necessary to understand the assumptions involved and the statistics that are calculated, so such issues will now be discussed.

The Bartlett test

With just two normal samples the F test is recommended, and this can be performed routinely in SIMFIT as part of the t test procedure, which is actually equivalent to 1-way ANOVA when there are only two samples. Where there are n normal samples the Bartlett test is recommended, and this is just the same as the F test when there are two samples of the same size. If there are k groups, with sample size n_i , $\nu_i = n_i - 1$, and sample

variances s_i^2 , then the pooled variance estimate s_p^2 , and parameters B and C can be calculated as follows,

$$s_p^2 = \frac{\sum_{i=1}^k v_i s_i^2}{\sum_{i=1}^k v_i}$$

$$B = \log(s_p^2) \sum_{i=1}^k v_i - \sum_{i=1}^k v_i \log(s_i^2)$$

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{v_i} - \frac{1}{\sum_{i=1}^k v_i} \right).$$

To test homogeneity of variance, the Bartlett test statistic B/C is approximately chi-square distributed with $k - 1$ degrees of freedom.

The Levene test

When normality cannot be assumed, the Levene test can be performed. If the total number of observations is $N = \sum_{i=1}^k n_i$, then the test statistic W is defined as

$$W = \frac{(N - k) \sum_{i=1}^k n_i (Z_{i.} - Z_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - Z_{i.})^2},$$

where $Z_{..}$ is the mean of all Z_{ij} , and $Z_{i.}$ is the group mean of the Z_{ij} . If Y_{ij} is observation j in group i the definitions are

$$Z_{ij} = |Y_{ij} - Y_{i.}|$$

$$Z_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}$$

$$Z_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij},$$

but note that there are several ways to define $Y_{i.}$. Usually this is taken to be the median of group i , but if there are long tails in the distribution as with the Cauchy distribution, the the trimmed mean can be used. The group mean can also be used if the data are similar to a normal distribution. To test variance homogeneity, the Levene test statistic W is approximately F distributed with $k - 1$ and $N - k$ degrees of freedom.

The above table illustrates that the null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

cannot be rejected for the data in test file `anova1.tff1`.