The Mathematical Differences Between the *F*-test, *W*-test, and *F**-test

The mathematical differences between the *F*-test, *W*-test and *F**-test can be explained by focusing on how standard deviations are pooled across groups. As shown in (1) the *F* statistic is calculated by dividing the inter-group variance by a pooled error term, where s_j^2 and n_j are respectively the variance estimates and the sample sizes from each independent group, and where *k* is the number of independent groups:

$$F = \frac{\frac{1}{k-1}\sum_{j=1}^{k}\left[n_j\left(\bar{x}_j - \bar{x}_{..}\right)^2\right]}{\frac{1}{N-k}\sum_{j=1}^{k}(n_j - 1)s_j^2} \tag{1}$$

The degrees of freedom in the numerator (2) and in the denominator (3) of the *F*-test are computed as follows:

$$df_n = k - 1 \tag{2}$$

$$df_d = N - k, \tag{3}$$

With $N = \sum_{j=1}^{k} n_j$. As a generalization of the Student's *t*-test, the *F*-test is calculated based on a pooled error term. This implies that all samples are considered as issued from a common population variance (hence the assumption of homoscedasticity). When there is heteroscedasticity, and if the larger variance is associated with the larger sample size, the error term, which is the denominator in (1), is overestimated. The *F*-value is therefore smaller, leading to fewer significant findings than expected, and the *F*-test is too conservative. When the larger variance is associated with the smaller sample size the denominator in (1) is underestimated. The *F*-value is then inflated, which yields more significant results than expected.

The *F** statistic proposed by Brown and Forsythe (1974) is computed as follows:

$$F^* = \frac{\sum_{j=1}^{k}\left[n_j\left(\bar{x}_j - \bar{x}_{..}\right)^2\right]}{\sum_{j=1}^{k}\left[\left(1-\frac{n_j}{N}\right)s_j^2\right]} \tag{4}$$

Where x_j and s_j^2 are respectively the group mean and the group variance, and $\bar{x}_{..}$ is the overall mean. As it can be seen in (4) the numerator of the *F** statistic is equal to the sum of squares between groups (which is equal to the numerator of the *F* statistic when one compares two groups). In the denominator, the variance of each group is weighted by 1 minus the relative frequency of each group. This adjustment implies that the variance associated with the group with the smallest sample size is given more weight compared to the *F*-test. As a result, when the larger variance is associated with the larger sample size, *F** is larger than *F*, because the denominator decreases, leading to more significant findings compared to the *F*-test. On the other hand, when the larger variance is associated with the smaller sample size, *F** is smaller than *F*, because the denominator increases, leading to fewer significant findings compared to the *F*-test. The degrees of freedom in the numerator and in the denominator of *F**-test are computed as follows (with the same principle as the denominator computation of the *F** statistic):

$$df_n = k - 1 \tag{5}$$

$$df_d = \frac{1}{\sum_{j=1}^{k}\left[\frac{\left(\frac{\left(1-\frac{n_j}{N}\right)s_j^2}{\sum_{j=1}^{k}\left[\left(1-\frac{n_j}{N}\right)s_j^2\right]}\right)^2}{n_j - 1}\right]} \tag{6}$$

Formula (7) provides the computation of the *W*-test, or Welch's *F*-test. In the numerator of the *W*-test the squared deviation between group means and the general mean are weighted by $\frac{n_j}{s_j^2}$ instead of n_j (Brown & Forsythe, 1974). As a consequence, for equal sample sizes, the group with the highest variance will have smaller weight (Liu, 2015).

$$W = \frac{\frac{1}{k-1}\sum_{j=1}^{k}\left[w_j\left(\bar{X}_j - \bar{X}'\right)^2\right]}{1 + \frac{2(k-2)}{k^2-1}\sum_{j=1}^{k}\left[\left(\frac{1}{n_j - 1}\right)\left(1 - \frac{w_j}{w}\right)^2\right]} \tag{7}$$

where:

$$w_j = \frac{n_j}{s_j^2}$$

$$w = \sum_{j=1}^{k}\left(\frac{n_j}{s_j^2}\right)$$

$$\bar{X}' = \frac{\sum_{j=1}^{k}\left(w_j \bar{x}_j\right)}{w}$$

The degrees of freedom of the *W*-test are approximated as follows:

$$df_n = k - 1 \tag{8}$$

$$df_d = \frac{k^2 - 1}{3\sum_{j=1}^{k}\left[\frac{\left(1-\frac{w_j}{w}\right)^2}{n_j - 1}\right]} \tag{9}$$

When there are only two groups to compare, the *F**-test and *W*-test are identical (i.e., they have exactly the same statistical value, degrees of freedom and significance). However, when there are more than two groups to compare, the tests differ. In the appendix we illustrate the calculation of all three statistics in detail for a fictional three-group design for educational purposes.