In **Figures 2** to **6** (see **Figure 1** for the legend), we computed the average Type I error rate of the three tests under these five subcategories. The light grey area corresponds to the liberal criterion from Bradley (1978), who regards a departure from the nominal alpha level as acceptable whenever the Type I error rate falls within the interval [0.5 × α; 1.5 × α]. The dark grey area corresponds to the more conservative criterion from which departures from the nominal alpha is considered negligible as long as the Type I error rate falls within the interval [0.9 × α; 1.1 × α].

In **Figures 2** and **3** (cells a, b, and c in **Table 1**), the population variance is equal between all groups, so the homoscedasticity assumption is met. The *F*-test and *F**-test only marginally deviate from the nominal 5%, regardless of the underlying distribution and the *SD*-ratio. The *W*-test also only marginally deviates from the nominal 5%, except under asymmetry (the tests becomes a little more liberal) or extremely heavy tails (the test becomes a bit more conservative), consistently with observations in Harwell et al. (1992). However, deviations don't exceed the liberal criterion of Bradley (1978).

In **Figures 4, 5** and **6** (cells d to i, **Table 1**) the population variance is unequal between groups, so that the homoscedasticity assumption is not met. When sample sizes are equal across groups (**Figure 4**) and when there is a positive correlation between sample sizes and *SDs* (**Figure 5**), the Type I error rate of the *W*-test is closer to the nominal 5% than the Type I error rate of the *F**-test and the *F*-test, the latter which is consistently at the lower limit of the liberal interval suggested by Bradley, in line with Harwell et al. (1992), Glass et al. (1972), Nimon (2012) and Overall et al. (1995). Heteroscedasticity does not impact the Type I error rate of the *W*-test, regardless of the distribution (the order of the distribution shape remains the same in all conditions).

When there is a negative correlation between sample sizes and *SDs* (**Figure 6**), the Type I error rate of the *F**-test is slightly closer of the nominal 5% than the Type I error rate of the *W*-test, for which the distributions (more specifically, the skewness) has a larger impact on the Type I error rate than when there is homoscedasticity. This is consistent with conclusions of Lix et al. (1996) about

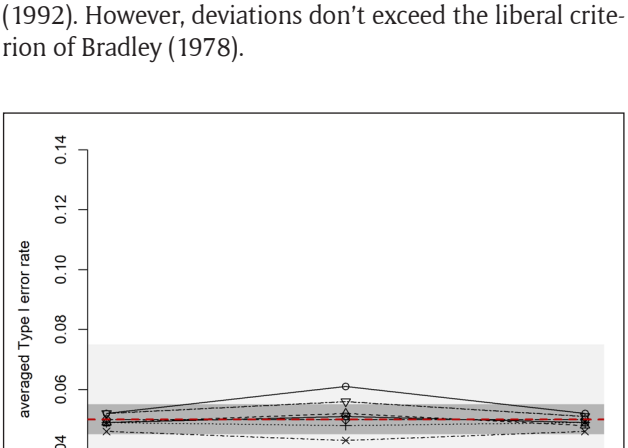

Figure 2: Type I error rate of the F-test, W-test and F*-test when there are equal SDs across groups and equal sample sizes (cell a in Table 1).

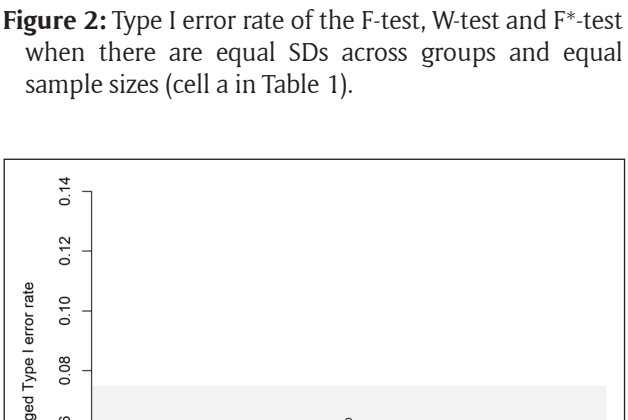

Figure 3: Type I error rate of the F-test, W-test and F*-test when there are equal SDs across groups and unequal sample sizes (cells b and c in Table 1).

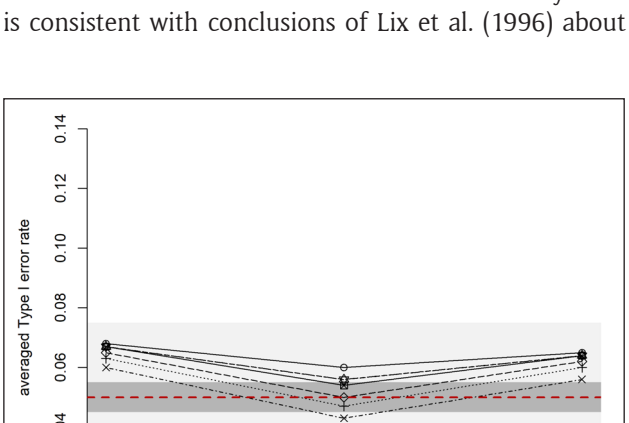

Figure 4: Type I error rate of the F-test, W-test and F*-test when there are unequal SDs across groups and equal sample sizes (cells d and g in Table 1).

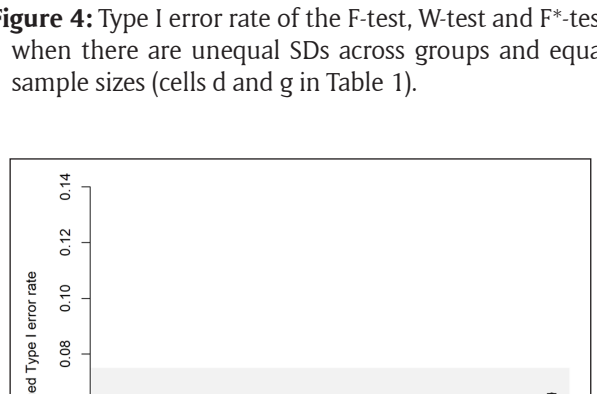

Figure 5: Type I error rate of the F-test, W-test and F*-test when there are unequal SDs across groups, and positive correlation between sample sizes and SDs (cells e and i in Table 1).