Appendix 1: The Mathematical Development of the F-test, W-test, and F^*-test: Numerical Example

Descriptive statistics are presented in Table A1. The raw data are available here : https://github.com/mdelacre/W-ANOVA/tree/master/Functions (see "practical example.R"). The dependent variable is a score that can vary from 0 to 40. The independent variable is a three-level factor A (levels = A_1, A_2 and A_3).

Table A1. *Summary of the data of the fictive case*

	A1	A2	A3
n_j	41	21	31
$\bar{X}$	24	23	27
S^2	81.75	10.075	38.40

The overall mean (i.e. the mean of the global dataset) is a weighted mean of the sample means :

$$\bar{X}_{..} = \frac{(41 \times 24) + (21 \times 23) + (31 \times 27)}{41 + 21 + 31} = \frac{2304}{93} \approx 24.77$$

The F-test statistic and degrees of freedom are computed by applying equation (1) and related degrees of freedom :

$$F = \frac{\frac{1}{3-1}[41 \times (24 - \frac{2304}{93})^2 + 21 \times (23 - \frac{2304}{93})^2 + 31 \times (27 - \frac{2304}{93})^2]}{\frac{1}{93-3}[(41-1) \times 81.75 + (21-1) \times 10.075 + (31-1) \times 38.4]} \approx 2.377$$

$$df_n = 3 - 1 = 2$$

$$df_d = 93 - 3 = 90$$

The F^*-test statistic and his degrees of freedom are computed by applying equation (2) and related degrees of freedom :

$$F^* = \frac{41 \times (24 - \frac{2304}{93})^2 + 21 \times (23 - \frac{2304}{93})^2 + 31 \times (27 - \frac{2304}{93})^2}{(1 - \frac{41}{93}) \times 81.75 + (1 - \frac{21}{93}) \times 10.075 + (1 - \frac{31}{93}) \times 38.4} \approx 3.088$$

$$df_n = 3 - 1 = 2$$

$$df_d = \frac{1}{\frac{\left(\frac{(1-\frac{41}{93}) \times 81.75}{\sum_{j=1}^{k}(1-\frac{n_j}{N})S_j^2}\right)^2}{41-1} + \frac{\left(\frac{(1-\frac{21}{93}) \times 10.075}{\sum_{j=1}^{k}(1-\frac{n_j}{N})S_j^2}\right)^2}{21-1} + \frac{\left(\frac{(1-\frac{31}{93}) \times 38.4}{\sum_{j=1}^{k}(1-\frac{n_j}{N})S_j^2}\right)^2}{31-1}} \approx 81.149$$

where $\sum_{j=1}^{k}(1 - \frac{n_j}{N}) \times S_j^2 \approx 79.11$

Finally, the W-test and his degrees of freedom are computed by applying equation (3) and related degrees of freedom :

$$W = \frac{\frac{1}{3-1}[\frac{41}{81.75}(24-\bar{X}')^2 + \frac{21}{10.075}(23-\bar{X}')^2 + \frac{31}{38.4}(27-\bar{X}')^2]}{\frac{2(3-2)}{3^2-1}[(\frac{1}{41-1})(1-\frac{\frac{41}{81.75}}{w})^2 + (\frac{1}{21-1})(1-\frac{\frac{21}{10.075}}{w})^2 + (\frac{1}{31-1})(1-\frac{\frac{31}{38.4}}{w})^2] + 1} \approx 4.606$$

where: $w = \sum_{j=1}^{k} w_j \approx 3.39$ and $\bar{X}' = \frac{\sum_{j=1}^{k}(w_j \bar{X}_j)}{w} \approx 24.1$

$$df_n = 3 - 1$$

$$df_d = \frac{3^2-1}{3[\frac{(1-\frac{w_j}{w})^2}{41-1} + \frac{(1-\frac{w_j}{w})^2}{21-1} + \frac{(1-\frac{w_j}{w})^2}{31-1}]} \approx 59.32$$

One should notice that in this example, the biggest sample size has the biggest variance. As previously mentioned, it means that the F-test will be too conservative, because the F value decreases. The F^*-test will also be a little too conservative, even if the test is less affected than the F-test. As a consequence: $W > F^* > F$.

Appendix 2: Justification for the choice of distributions in simulations

The set of simulations described in the article was repeated for 7 distributions. We used R commands to generate data from different distributions:

- *k* normal distributions (Figure A1): in order to assess the Type I error rate and power of the different tests under the assumption of normality, data were generated by means of the function "rnorm" (from the package "stats"; "R: The Normal Distribution," 2016).
- *k* double exponential distributions (Figure A2): In order to assess the impact of high kurtosis on the Type I error rate and power of all tests, data were generated by means of the function "rdoublex" (from the package "smoothmest"; "R: The double exponential (Laplace) distribution," 2012).
- *k* mixed normal distributions (Figure A3): In order to assess the impact of extremely high kurtosis on the Type I error rate and power of all tests, regardless of variance, data were generated by means of the function "rmixnorm" (from the package "bda"; Wang & Wang, 2015).
- *k* normal right skewed distributions (Figure A4): In order to assess the impact of moderate skewness on the Type I error rate and power, data were generated by means of the function "rsnorm" (from the package "fGarch"; "R: Skew Normal Distribution," 2017). The normal skewed distribution was chosen because it is the only skewed distribution where the standard deviation ratio can vary without having an impact on skewness.
- *k*-1 normal left skewed distributions (Figure A5) and 1 normal right skewed distribution (Figure A2.4): In order to assess the impact of unequal shapes, in terms of skewness, on the Type I error rate and power, when data have moderate skewness, data were generated by means of the functions "rsnorm" (from the package "fGarch"; "R: Skew Normal Distribution," 2017).
- *k*-1 chi-squared distributions with two degrees of freedom (See Figure A6), and one normal rigt skewed distribution (Figure A2.4): In order to assess the impact of high asymetry on the Type I error rate an power, k-1 distributions were generated by means of the functions "rchisq" ("R: The (non-central) Chi-squared Distribution," 2016). The last distribution was generated by means of "rsnorm" in order to follow a normal right skewed distribution with a mean of 2 (from the package "fGarch"; "R: Skew Normal Distribution," 2017). Because the chi-squared is non-negative, it is not possible to generate chi-squared where population SD= 1, 4 or 8 and population mean is the same than the chi-squared with two degrees of freedom. However, we wanted to assess the impact of different SD-ratio on Type I error rate. For these reasons, the last distribution was generated by means of "rsnorm" in order to follow a normal skewed distribution with positive skewness of +0.99 and mean = 2 (from the package "fGarch"; "R: Skew Normal Distribution," 2017).
- *k*-1 chi-squared distributions with two degrees of freedom (See Figure A6), and one normal left skewed distribution (Figure A5): In order to assess the impact of unequal shapes, in terms of skewness, on Type I error rate and power when distributions have extreme skewness, k-1 distributions were generated by means of the functions "rchisq" ("R: The (non-central) Chi-squared Distribution," 2016). The last distribution was generated by means of "rsnorm" in order to follow a normal right skewed distribution with a mean of 2 (from the package "fGarch"; "R: Skew Normal Distribution," 2017)

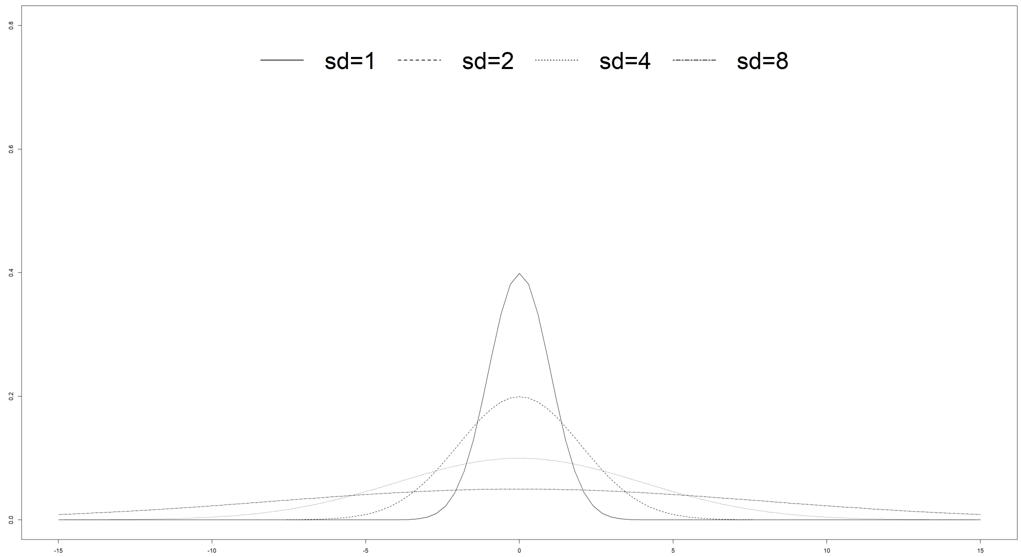

Figure 1: centered normal probability density function, as a function of the population SD

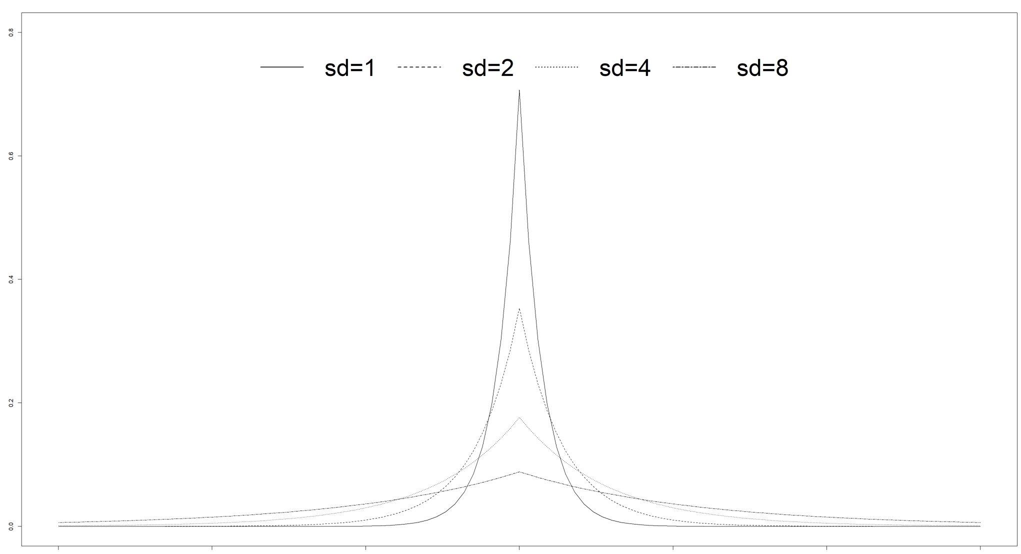

Figure 2: centered double exponential probability density function, as a function of the population SD

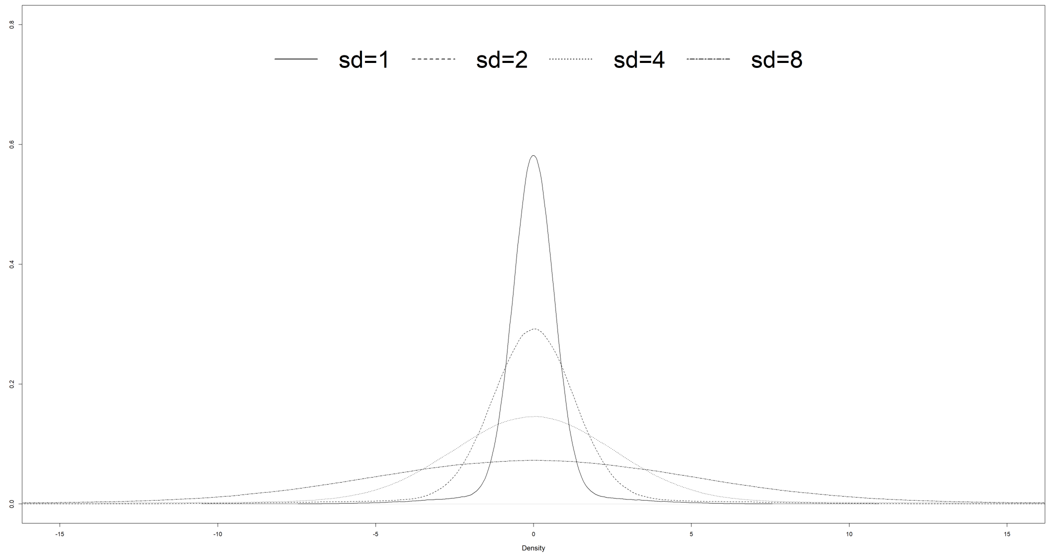

Figure 3: centered mixed normal probability density function, as a function of the population SD

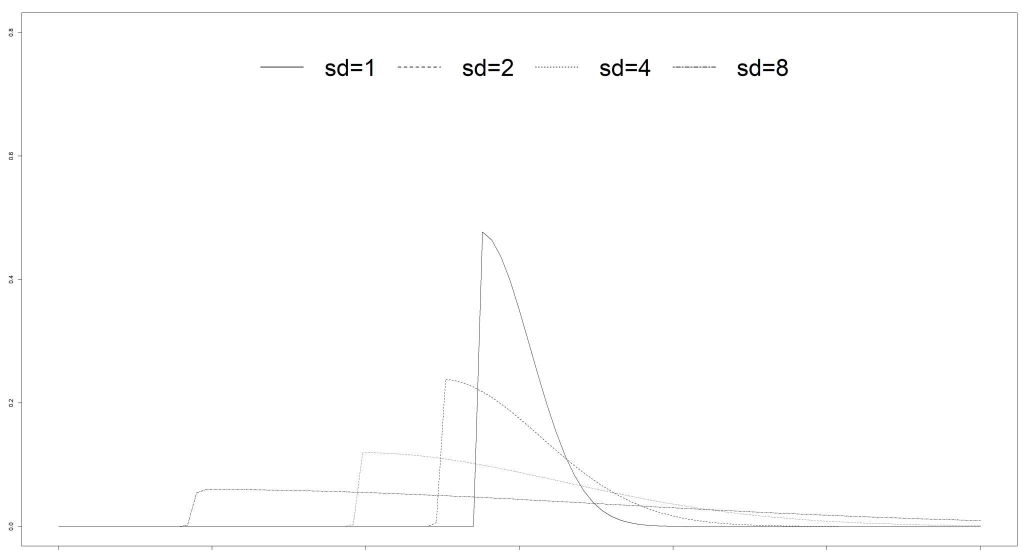

Figure 4: centered normal right skewed probability density function, as a function of the population SD

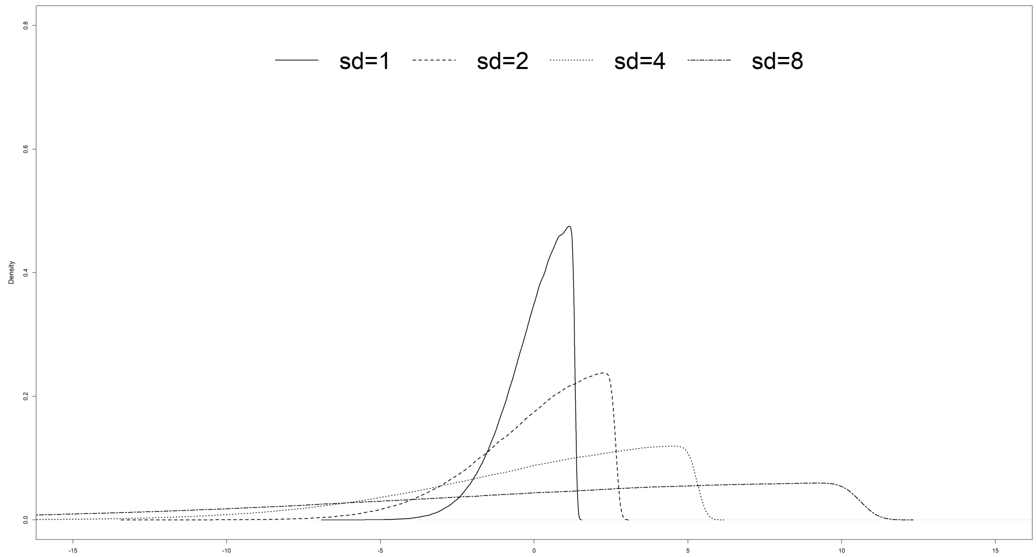

Figure 5: centered normal left skewed probability density function, as a function of the population SD

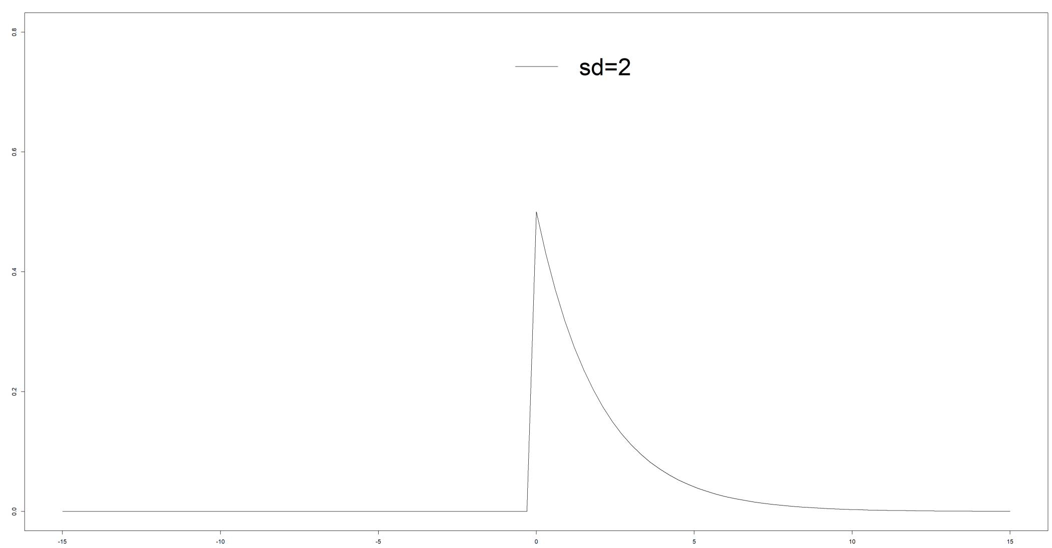

Figure 6: chi-squared with 2 degrees of freedom probability density function, when SD=2