The variance ratio F-test statistic

The Fisher F-test statistic is based on the ratio of two experimentally observed variance, which are squared standard deviations. Therefore, it is useful to test whether two standard deviations s_1 and s_2, calculated from two independent datasets are significantly different in terms of precision from each other.

In this case, the ratio of the estimated variances, F is equal to s_1^2 / s_2^2 where $s_1 > s_2$, and depending on two separate degrees of freedom (n_1-1) for s_1, and (n_2-1) for s_2.

As said, the purpose of carrying out the F-test is to see if one standard deviation is significantly better or different from the other. In fact, it is also fundamental to another important statistical tool, the analysis of variance (ANOVA).

We can apply the F-test on comparing the precision of two analytical methods, the performance of two laboratory analysts or two separate laboratories.

In significance testing, we usually put the null hypothesis Ho as $s_1^2 = s_2^2$. When testing two method against another to see whether the performance is similar, we state the alternative hypothesis Ha as $s_1^2 \neq s_2^2$. This is called a two-tailed test. On the other hand, when testing whether variance s_1^2 is significantly greater (i.e. less precision), the alternative hypothesis Ha is then $s_1^2 > s_2^2$, and vice versa. This is called a one-sided test.

A worked example

A food laboratory used two different methods (Dumas and Kjeldahl) in the determination of protein nitrogen of a similar sample of wheat flour repeatedly. The results obtained are tabulated below:

Trial #	Dumas, %	Kjeldahl, %
1	3.12	3.07
2	3.01	2.92
3	3.05	3.01
4	3.04	3.00
5	3.04	3.02
6	2.98	3.02
7	3.08	2.98
8	3.09	2.92
9		3.01
10		3.05
Mean, %	3.05	3.00
Standard deviation, %	0.0449	0.0490
Variance	0.002013	0.00240

The variance ratio calculated, $F = 0.00240/0.002013 = 1.193$. We may carry out two hypothesis testing by considering alternative hypothesis as either $s_D^2 \neq s_K^2$ or $s_D^2 < s_K^2$ where s_D and s_K are standard deviations of Dumas' and Kjeldahl's methods respectively. So, we have:

$Ho:\ s_D^2 = s_K^2$

$Ha:\ s_D^2 \neq s_K^2$ (2-tailed test, $\alpha=0.025$), or,

$Ha:\ s_D^2 < s_K^2$ (1-tailed test, $\alpha=0.05$)

We now compare the F-value obtained against the F- critical values under these two alternative hypothesis situations. The MS Excel spreadsheet can provide the following results with probability p-values corresponding with the F-value against their respective critical value:

		Two-tailed test
p-Value	0.582	=F.DIST(1.193,9,7,TRUE)
F-critical value	4.823	=F.INV(0.975,9,7)
F-critical value	4.823	=F.INV.RT(0.025,9,7)
		One-tailed test
p-value	0.418	=F.DIST.RT(1.193,9,7)
F-critical value	3.677	=F.INV(0.95,9,7)
F-critical value	3.677	=F.INV.RT(0.05,9,7)

Since the calculated F-value of 1.193 was significantly smaller than the F-critical values of 3.677 (for one-tailed test) and 4.823 (for two-tailed test), we have 95% confidence or commit 5% risk to conclude that the precisions of these two methods for protein N were comparable and there was no significant difference between them.

Take note of the Excel functions used in the spreadsheet of 2013 or 2016 versions to find the critical values of F at an error α=0.05.