

An F-test for the 21st Century

No one wants to use 100-year old techniques!

Donald J. Wheeler

A recent question from a statistician in Germany led me to the realization that the F-test of Analysis of Variance fame is in serious need of an update.

WHAT THE F-RATIO DOES

The F-ratio, created by Sir Ronald Fisher around 1925, is a generalization of Student's t-test for comparing two averages. With the F-ratio a set of k averages can be compared to see if any differ from the others. Say we have k experimental conditions and we collect n observations for some response variable at each of these k conditions. Do the experimental conditions affect the response variable? If so, then we would expect the k averages to differ. If not, then the k averages should all be about the same.

The F-ratio works by comparing two quantities computed from the k subgroups of size n . The first quantity, the mean square between (*MSB*), is a measure of variation that depends solely upon the k subgroup averages. The second quantity, the mean square within (*MSW*) is a measure of variation based solely upon the variation within each of the k subgroups of size n . These quantities are found in the traditional ANOVA table of Figure 1:

Source	Sum of Squares	D. F.	Mean Squares
Between	SSB	$k - 1$	MSB
Within	SSW	$k(n - 1)$	MSW
Total	SST	$nk - 1$	

Figure 1: Basic ANOVA Quantities

$$F\text{-ratio} = \frac{\text{Mean Square Between}}{\text{Mean Square Within}}$$

When the k experimental conditions affect the response variable the k subgroup averages will differ by more than routine variation, and these differences will inflate the mean square between. When the k experimental conditions have no effect upon the response variable the k subgroup averages will only differ by routine variation and the mean square between will characterize this routine variation. Thus, the mean square between is the component that is affected by signals within the data.

However, regardless of the effect of the experimental conditions upon the response variable, the mean square within will reflect the routine variation within the subgroups. Thus, the mean square within will always characterize the noise in the data.

So, when the data contain signals, the averages will tend to be different, the mean square between will be inflated relative to the mean square within, which will inflate the F-ratio. When

the data contain no signals, the averages will all be about the same, the mean square between will be similar to the mean square within, and the F-ratio will be near 1.00. Thus, the F-ratio is in effect a signal-to-noise ratio. The numerator contains the signal component while the denominator contains the noise component.

ANOTHER OLD TECHNIQUE

While Sir Ronald Fisher was using the F-ratio to separate signals from noise, Walter Shewhart was creating a different technique to do the same thing. Shewhart's process behavior chart plots the subgroup averages against a set of limits centered on the grand average. These limits, known as three-sigma limits, serve the same function as the denominator in Fisher's F-ratio—they filter out the noise of routine variation. And, like the mean square within, Shewhart's limits are based on the average within-subgroup variation.

So, a little less than 100 years ago, both Fisher and Shewhart filtered out the noise using the within-subgroup variation. While their computations differed, their techniques were built on the same foundation. For nearly a century these techniques have proven themselves over and over.

OUR BRAVE NEW WORLD

The separation of signals and noise was the foundation of Twentieth Century statistics. Now we are in the 21st Century and everything is done electronically. To use today's software you must select various options to do your analysis, and to judge by the software, virtually anything goes. Thus, we are in the age of do-it-yourself statistics where we are no longer constrained by the highbrow statistical theorems of yesterday. We get to choose our own analysis, cafeteria style, from the options provided by our software.

In light of this, the following option does for Fisher's ANOVA what various software options have *already* done for Shewhart's process behavior charts. (It would be a shame for ANOVA to fall behind.) This option is necessary to keep pace with the evolving nature of 21st Century statistics. To set up this option we will need to add a value to the ANOVA table. While this value is one that everyone is already familiar with, it is one that is never included in the traditional ANOVA table. The ANOVA name for this value is the mean square total (*MST*):

Source	Sum of Squares	D. F.	Mean Squares
Between	SSB	$k - 1$	MSB
Within	SSW	$k(n - 1)$	MSW
Total	SST	$nk - 1$	MST

Figure 2: ANOVA for a Brave New World

The *MST* is more commonly known as the variance statistic, s^2 . It is the value we find when we put all nk data in the calculator or spreadsheet, compute the global standard deviation statistic, and square it. Thus, in spite of its name, it is a value we all meet in our introductory statistics class. Given a traditional ANOVA table you may easily find the *MST* by dividing the sum of squares total (*SST*) by the total degrees of freedom ($nk - 1$).

As do-it-yourself statistics evolve, having access to the mean square total will soon result in a new software option—the computation of a brave new F-ratio where the mean square between

will still be the signal component but the mean square total will be used as the noise component:

$$A \text{ Brave New } F\text{-ratio} = \frac{\text{Mean Square Between}}{\text{Mean Square Total}}$$

The fact that there is no rationale for computing this ratio, or that it has no justification in theory or practice will not matter. By providing this option for ANOVA the software will simply be doing *exactly the same thing* for ANOVA as it already does for SPC when it gives you the option of using the “long-term” variation to compute limits for a process behavior chart. So, you can install this option to compute a brave new F-ratio for ANOVA, or equivalently, you can use existing options to compute limits for a process behavior chart using the long-term global standard deviation. Either way you can turn every day into April Fool’s day.

The brave new F-ratio above and the use of long-term variation for three-sigma limits are two sides of the same coin. With do-it-yourself statistics it does not matter that both sides of this coin are complete nonsense. After all, this is the age of alternate facts.

