

The F-ratio when the Homogeneity of Variance assumption is broken

In Box 8.3 I explained a bit about the Brown-Forsythe F -ratio, which corrects for violations of the homogeneity of variance assumption. To recap, we discovered that:

$$F_{BF} = \frac{SS_M}{SS_{R_{BF}}} = \frac{SS_M}{\sum s_k^2 \left(1 - \frac{n_k}{N}\right)}$$

So, for the Viagra data, SS_M is the same as before (20.135), and the equation becomes:

$$\begin{aligned} F_{BF} &= \frac{20.135}{s_{group1}^2 \left(1 - \frac{n_{group1}}{N}\right) + s_{group2}^2 \left(1 - \frac{n_{group2}}{N}\right) + s_{group3}^2 \left(1 - \frac{n_{group3}}{N}\right)} \\ &= \frac{20.135}{1.70 \left(1 - \frac{5}{15}\right) + 1.70 \left(1 - \frac{5}{15}\right) + 2.50 \left(1 - \frac{5}{15}\right)} \\ &= \frac{20.135}{3.933} \\ &= 5.119 \end{aligned}$$

The model sum of squares stay the same, but the error sum of squares are given by:

$$\begin{aligned} \frac{1}{df_R} &= \sum \frac{\left(\frac{s_k^2 \left(1 - \frac{n_k}{N}\right)}{SS_{R_{BF}}} \right)^2}{(n_k - 1)} \\ &= \frac{\left(\frac{1.133}{3.933}\right)^2}{5-1} + \frac{\left(\frac{1.133}{3.933}\right)^2}{5-1} + \frac{\left(\frac{1.667}{3.933}\right)^2}{5-1} \\ &= 0.0864 \\ df_R &= \frac{1}{0.0864} \\ &= 11.57 \end{aligned}$$

The Welch (1951) F -ratio is somewhat more complicated (and hence why it's stuck in an appendix). First we have to work out a weight that is based on the sample size, n_k , and variance, s_k^2 , for a particular group:

$$w_k = \frac{n_k}{s_k^2}$$

We also need to use a grand mean based on a weighted mean for each group. So we take the mean of each group, \bar{x}_k , and times it by its weight, w_k , do this for each group and add them up, then divide this total by the sum of weights:

$$\bar{x}_{grand}^{Welch} = \frac{\sum w_k \bar{x}_k}{\sum w_k}$$

The easiest way to do this is in table form:

Group	Variance s^2	Sample size, n_k	Weight w_k	Mean \bar{x}_k	$w_k \times \bar{x}_k$
Placebo	1.70	5	2.941	2.2	6.4702
Low Dose	1.70	5	2.941	3.2	9.4112
High Dose	2.50	5	2.000	5.0	10.000
			$\Sigma = 7.882$		$\Sigma = 25.8814$

So, we get:

$$\bar{x}_{grand}^{Welch} = \frac{25.8814}{7.882} = 3.284$$

Think back, to equation 8.5, the model sum of squares was:

$$SS_M = \sum n_k (\bar{x}_k - \bar{x}_{grand})^2$$

In Welch's F this is adjusted to incorporate the weighting and the adjusted grand mean:

$$SS_M^{Welch} = \sum w_k (\bar{x}_k - \bar{x}_{grand}^{Welch})^2$$

And to create a mean squares we divide by the degrees of freedom, $k - 1$:

$$\begin{aligned} MS_M^{Welch} &= \frac{\sum w_k (\bar{x}_k - \bar{x}_{grand}^{Welch})^2}{K - 1} \\ &= \frac{2.941(2.2 - 3.284)^2 + 2.941(3.2 - 3.284)^2 + 2(5 - 3.284)^2}{2} \\ &= 4.683 \end{aligned}$$

We now have to work out a term called lambda, which is based again on the weights:

$$\Lambda = \frac{3 \sum \left(\frac{1 - \frac{w_k}{\sum w_k}}{n_k - 1} \right)^2}{K^2 - 1}$$

This looks horrendous, but is just based on the sample size in each group, the weight for each group (and the sum of all weights), and the total number of groups, K . For the Viagra data this gives us:

$$\begin{aligned} \Lambda &= \frac{3 \left(\frac{\left(1 - \frac{2.941}{7.882}\right)^2}{5-1} + \frac{\left(1 - \frac{2.941}{7.882}\right)^2}{5-1} + \frac{\left(1 - \frac{2}{7.882}\right)^2}{5-1} \right)}{3^2 - 1} \\ &= \frac{3(0.098 + 0.098 + 0.139)}{8} \\ &= 0.126 \end{aligned}$$

The F ratio is then given by:

$$F_W = \frac{MS_M^{Welch}}{1 + \frac{2\Lambda(K-2)}{3}}$$

Where, K , is the total number of groups. So, for the Viagra data we get:

$$\begin{aligned} F_W &= \frac{4.683}{1 + \frac{(2 \times 0.126)(3-2)}{3}} \\ &= \frac{9.336}{1.084} \\ &= 4.32 \end{aligned}$$

As with the Brown-Forsythe F , the model degrees of freedom stay the same at $K-1$ (in this case 2), but the residual degrees of freedom, df_R , are $1/\Lambda$ (in this case, $1/0.126 = 7.94$).

This document is distributed on the CD-ROM of:

Field, A. P. (2004). *Discovering Statistics Using SPSS (2nd Edition)*. London: Sage.

It should be referenced accordingly, and is copyright protected with the rest of the book.