

PERFORMING THE FRIEDMAN TEST AND THE ASSOCIATED MULTIPLE COMPARISON TEST USING PROC GLM

David Ipe, Syntex (Research)

The F-approximation of the Friedman test and the associated Rank Sum multiple comparison test (Conover)¹ can be performed using the procedure 'PROC GLM'. This procedure is applied after ranking the data within each block from lowest to highest value using the procedure 'PROC RANK'. In this paper an overview of the F-approximation of the Friedman test and the associated rank sum multiple comparison test will be first presented. It will then be shown that by appropriately specifying the 'model' statement and the 'LS Means' statement in the procedure 'PROC GLM' one can arrive at exactly the same results.

Consider a data set of b mutually independent k -dimensional random variables $X_{i1}, X_{i2}, \dots, X_{iK}$ where $i = 1, 2, \dots, b$. For example, in a pharmacological study a human subject represents a block, and K represents the number of treatments administered to each subject over K weeks. X_{ij} is in block i and is associated with the j^{th} treatment. The b blocks can be arranged in K columns where each column corresponds to a treatment.

BLOCKS	TREATMENT				
	1	2	3	...	K
1	X_{11}	X_{12}	X_{13}	...	X_{1K}
2	X_{21}	X_{22}	X_{23}	...	X_{2K}
3	X_{31}	X_{32}	X_{33}	...	X_{3K}
...
...
b	X_{b1}	X_{b2}	X_{b3}	...	X_{bK}

Rank the variables $X_{i1}, X_{i2}, \dots, X_{iK}$, within each block from the lowest (1) to the highest (K) value such that $R(X_{i1}), R(X_{i2}), \dots, R(X_{iK})$ are the corresponding ranks within the i^{th} block, where $i = 1, 2, \dots, b$. Then the sum of the ranks for treatment j is given by

$$R_j = \sum_{i=1}^b R(X_{ij}) \quad j = 1, 2, \dots, K.$$

The assumptions behind developing the test statistics are as follows:

1. K experimental units must be assigned to K treatments at random.
2. Units in different blocks are separate and non interacting (i.e., the b K -dimensional random variables are mutually independent).
3. Within each block the observations may be ranked according to some criterion of interest.

The hypotheses tested are:

- H_0 : All treatments have identical effects.
 H_1 : At least one treatment tends to yield larger observed values than any other treatments.

The test statistic is:

$$T_2 = \frac{(b-1) [B_2 - \frac{bK(K+1)}{4}]^2}{A_2 - B_2} \quad - (1)$$

Where $A_2 = \sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2$, $B_2 = \frac{1}{b} \sum_{j=1}^K R_j^2$

and R_j is the sum of the ranks for treatment j .

If T_2 exceeds the $1-\alpha$ quantile of the F-distribution with $K_1:K-1$ and $K_2 = (b-1)(K-1)$ degrees of freedom then the null hypothesis H_0 is rejected at the level α .

If the usual Randomized Block Design (RBD) describes the experiment, then the model statement in the procedure 'PROC GLM' is given by

Model VARIABLE = BLOCK TREATMENT.

Under the RBD assumption, the F-statistics for treatment differences is given by

$$F = \frac{\text{Mean treatment sum of squares (MTSS)}}{\text{Mean error sum of squares (MESS)}} \quad - (2)$$

The goal is to show that T_2 is algebraically equivalent to F . To attain this goal first consider the numerator of the ratio (2), i.e. MTSS.

$$\text{MTSS} = \frac{\text{Treatment sum of squares (TSS)}}{\text{Number of treatments} - 1} \quad - (3)$$

Where $\text{TSS} = \frac{1}{b} \sum_{j=1}^K R_j^2 - \text{Correction factor (CF)}$

and $\text{CF} = \frac{(\text{Sum of all observations})^2}{\text{Total number of observations}}$

$$= \frac{[b \frac{K(K+1)}{2}]^2}{bK} = \frac{bK(K+1)^2}{4}$$

Equation (3) is given by

$$\text{MTSS} = \frac{\frac{1}{b} \sum_{j=1}^K R_j^2 - \frac{bK(K+1)^2}{4}}{K-1} \quad - (4)$$

Now consider the denomination of the ratio (2), i.e., MESS

$$\text{MESS} = \frac{\text{Total SS} - \text{Treatment SS} - \text{Block SS}}{bK - (K-1) - (b-1) - 1} \quad (5)$$

$$\text{Where Total SS} = \sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2 - CF,$$

$$\text{Treatment SS} = \frac{1}{b} \sum_{j=1}^K R_j^2 - CF,$$

$$CF = \frac{bK(K+1)^2}{4},$$

Block SS = 0 (∵ The same set of ranks were assigned within each block)

and $bK - (K-1) - (b-1) - 1 = (b-1)(K-1)$

Therefore, equation (5) is given by

$$\text{MESS} = \frac{\sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2 - \frac{1}{b} \sum_{j=1}^K R_j^2}{(b-1)(K-1)} \quad (6)$$

Making use of equation (4) and (6), Equation (2) can be written as

$$F = \frac{\text{MTSS}}{\text{MESS}} = \frac{(b-1)[B_2 - \frac{bK(K+1)^2}{4}]}{A_2 - B_2}$$

$$\text{Where } A_2 = \sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2, \quad B_2 = \frac{1}{b} \sum_{j=1}^K R_j^2$$

Which is identical to T_2 .

This shows that F is algebraically equivalent to T_2 . Therefore, in order to arrive at the same results obtained by the statistics T_2 , the model statement in procedure 'PROC GLM' should be specified as

Model Variable = Block Treatment;

It is to be noted that the factor 'block' must be specified in the model statement, even though no source of variation is accounted for by the factor 'block'. This adjusts for the degrees of freedom.

Now, the associated multiple comparison test (Conover)¹, indicates that the treatment i and j are considered different if the following inequality is satisfied.

$$|\bar{R}_j - \bar{R}_i| > t_{1-\frac{\alpha}{2}} \left(\frac{2b(A_2 - B_2)}{(b-1)(K-1)} \right)^{\frac{1}{2}} \quad (7)$$

$$\text{Where } A_2 = \sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2, \quad B_2 = \frac{1}{b} \sum_{j=1}^K R_j^2$$

R_j and R_i are the sums of the ranks for treatments j and i and $t_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of the t-distribution with $(b-1)(K-1)$ degrees of freedom. The value of α is the same as the one used in the Friedman Test.

The 'least squares means' statement with the PDIF option in the procedure 'PROC GLM' requires that the following inequality be satisfied for treatments i and j to be considered different.

$$|\bar{R}_j - \bar{R}_i| > t_{1-\frac{\alpha}{2}} [V(D)]^{\frac{1}{2}} \quad (8)$$

Where \bar{R}_j and \bar{R}_i are the means of the ranks for treatments j and i, $V(D)$ is the variance of D and $D = \bar{R}_j - \bar{R}_i$.

It will now be shown that the expressions given by (7) and (8) are identical.

First consider

$$\begin{aligned} V(D) &= \text{VAR}(\bar{R}_j - \bar{R}_i) \\ &= \text{VAR}(\bar{R}_j) + \text{VAR}(\bar{R}_i) \quad \because \bar{R}_j \text{ and } \bar{R}_i \text{ are independent.} \\ &= \frac{2\sigma^2}{b} \end{aligned}$$

Where the best estimate for σ^2 is the mean square error obtained from ANOVA and is given in equation (6).

$$\therefore [V(D)]^{\frac{1}{2}} = \left(\frac{2}{b} \frac{(A_2 - B_2)}{(b-1)(K-1)} \right)^{\frac{1}{2}} \quad (9)$$

$$\text{Where } A_2 = \sum_{i=1}^b \sum_{j=1}^K [R(X_{ij})]^2, \quad B_2 = \frac{1}{b} \sum_{j=1}^K R_j^2$$

The result obtained after substituting equation (9) in expression (8) is given by

$$\begin{aligned} \left| \frac{R_j}{b} - \frac{R_i}{b} \right| &> t_{1-\frac{\alpha}{2}} \left(\frac{2}{b} \frac{(A_2 - B_2)}{(b-1)(K-1)} \right)^{\frac{1}{2}} \\ \therefore |\bar{R}_j - \bar{R}_i| &> t_{1-\frac{\alpha}{2}} \left(\frac{2b(A_2 - B_2)}{(b-1)(K-1)} \right)^{\frac{1}{2}} \end{aligned}$$

This is identical to expression (7). This implies that both the expressions [(7) and (9)] will give identical results.

Thus it has been shown that by specifying the 'model' statement as

```
'Model Variable = Block Treatment;'
```

and the 'LS Means' statement as

```
'LS Means Variable PDIFF;'
```

in the procedure 'PROC GLM' one arrives at exactly the same results produced by the F-approximation of the Friedman test and the associated rank sum multiple comparison test (Conover).¹

The complete SAS® code to perform the F-approximation of the Friedman test and the associated multiple comparison is given as

```
PROC RANK Data = RAW Out = RDATA;  
BY BLOCK;  
VAR Y;  
RANKS R;
```

```
PROC GLM DATA = RDATA;  
CLASS BLOCK TREAT;  
MODEL R = BLOCK TREAT;  
LSMEANS TREAT/PDIFF;
```

SAS is the registered trademark of the SAS Institute Inc., Cary, NC, USA.

Reference

1. Conover, W.J. Practical Nonparametric Statistics, Second Edition. New York: John Wiley and Sons.

David Ipe
Biostatistician
Syntex Research
3401 Hillview Avenue
Palo Alto, California 94304
415-476-3571