The F-approximation of the Friedman test and the associated Rank Sum multiple comparison test (Conover) 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 \textit{model} statement and the \textit{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 $X_{i1}$, $X_{i2}$, ..., $X_{iK}$, where $i = 1, 2, ..., b$. 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.

<table>
<thead>
<tr>
<th>BLOCKS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_{11}$</td>
<td>$X_{12}$</td>
<td>$X_{13}$</td>
<td>...</td>
<td>$X_{1K}$</td>
</tr>
<tr>
<td>2</td>
<td>$X_{21}$</td>
<td>$X_{22}$</td>
<td>$X_{23}$</td>
<td>...</td>
<td>$X_{2K}$</td>
</tr>
<tr>
<td>3</td>
<td>$X_{31}$</td>
<td>$X_{32}$</td>
<td>$X_{33}$</td>
<td>...</td>
<td>$X_{3K}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>b</td>
<td>$X_{b1}$</td>
<td>$X_{b2}$</td>
<td>$X_{b3}$</td>
<td>...</td>
<td>$X_{bK}$</td>
</tr>
</tbody>
</table>

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

$$R_j = \sum_{i=1}^{b} R(X_{ij})$$

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:

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

where

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

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 $\alpha$.

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

\texttt{Model VARIABLE = BLOCK TREATMENT.}

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

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

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.

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

where

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

and

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

$$= \frac{\left[ \sum_{i=1}^{b} \sum_{j=1}^{K} X_{ij} \right]^2}{\sum_{i=1}^{b} \sum_{j=1}^{K} X_{ij}^2} = \frac{bK(K+1)^2}{4bK}$$


$$\text{MTSS} = \frac{\left[ \sum_{j=1}^{K} \sum_{i=1}^{b} R_j^2 - \text{bK}(K+1)^2 \right]}{K-1}$$

Equation (3) is given by

$$\text{MTSS} = \frac{\left[ \sum_{j=1}^{K} \sum_{i=1}^{b} R_j^2 - \text{bK}(K+1)^2 \right]}{K-1}$$

1146
Now consider the denominator of the ratio (2), i.e., MESS

\[ MESS = \text{Total SS} - \text{Treatment SS} - \text{Block SS} - (5) \]

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

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 \( bK(K-1)-(b-1)(K-1) = (b-1)(K-1) \)

Therefore, equation (5) is given by

\[ MESS = \frac{(b-1)(B_2 - bK(k+1))^2}{A_2^2} \]

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

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

\[ \text{Model Variable = Block Treatment;} \]

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), indicates that the treatment \( i \) and \( j \) are considered different if the following inequality is satisfied.

\[ |R_j - R_i| > t_{1-\alpha} \frac{(2b(A_2-B_2))^{1/2}}{(b-1)(K-1)} \]  

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

\( R_i \) and \( R_j \) are the sums of the ranks for treatments \( j \) and \( i \) and \( t_{1-\alpha/2} \) is the \( 1-\alpha/2 \) quantile of the t-distribution with \((b-1)(K-1)\) degrees of freedom. The value of \( \alpha \) 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.

\[ |\overline{R}_j - \overline{R}_i| > \frac{1}{2} \left( \frac{V(D)}{t_{1-\alpha}} \right)^{1/2} \]  

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

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

First consider

\[ V(D) = \text{VAR}(\overline{R}_j - \overline{R}_i) \]

\[ = \text{VAR}(\overline{R}_j) + \text{VAR}(\overline{R}_i) \]  

\( \overline{R}_j \) and \( \overline{R}_i \) are independent.

\[ = \frac{2\sigma^2}{b} \]

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

\[ \left( \frac{1}{2} \left( \frac{V(D)}{t_{1-\alpha}} \right)^{1/2} \right) \]

\[ = \left( \frac{2(A_2-B_2)}{b(b-1)(K-1)} \right)^{1/2} \]  

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

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

\[ |\overline{R}_j - \overline{R}_i| > t_{1-\alpha} \frac{(2b(A_2-B_2))^{1/2}}{(b-1)(K-1)} \]

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 PDIF;',

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).

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