Journal of Applied Statistical Science
Volume 14, Number 3/4, pp. 225-235
ISSN 1067-5817


# An Improvement to the Aligned Rank Statistic for Two-Factor Analysis of Variance

***Scott J. Richter*[1]**
Department of Mathematical Sciences, The University of North Carolina at Greensboro, Greensboro, NC 27402
***Mark E. Payton*[2]**
Department of Statistics, Oklahoma State University
Stillwater, OK 74078-1056

## Summary

We modify the nonparametric method of Brunner, Dette and Munk (1997) by applying their method to the aligned ranks of the data. We compare this new approach to three other rank tests: the F-test applied to the ranks of the data, the approximate F-test of Brunner, Dette and Munk (1997) applied to the ranks of the data, and the F-test applied to the aligned ranks of the data. We also compare the new test to parametric F-test and to the approximate F-test using the modified Box-type adjustment of Brunner, Dette and Munk (1997). Simulation studies show that the proposed test alleviates the problem of inflated Type I error rates shown by both the test based on simple ranks and the rank test of Brunner, Dette and Munk (1997). In addition, we show that using the Box-type adjustment alleviates the problem of Type I error rate inflation seen previously for the aligned rank test, without a noticeable loss of power, as long as sample sizes are moderate ( $n \geq 7$ ).

**Key words:** Aligned rank, ANOVA, nonparametric.

## 1 Introduction

There have been many attempts to find robust alternatives to the parametric analysis of variance for investigating the effects of two factors in a completely randomized design. Conover and Iman (1976, 1981) proposed performing the usual parametric analysis on the ranks of the data, and suggested that if the results for the ranks differed significantly from the results of the usual parametric procedure, then the assumptions of the parametric procedure were doubtful and the analysis on the ranks likely to be more appropriate. They examined a two-factor arrangement and presented simulation results supporting the performance of the method, which became known as the rank transform, and it quickly gained widespread acceptance due to the simplicity of implementation. While this approach yields good results for simple cases, for situations involving possible interaction the method has potential flaws. Simulation studies found that for large effect sizes the method could suffer from inflated Type I

[1] E-mail address: sjricht2@uncg.edu
[2] E-mail address: mpayton@okstate.edu (corresponding author)

error rates for tests of interaction (Fligner, 1981; Blair, Sawilowsky and Higgins, 1987; Thompson and Amman, 1990). In fact, when both main effects were present without interaction, the Type I error rate of the test for interaction can approach one as the effect sizes increase.

A variation of the rank transform, where the observations are "aligned" before ranking, has also been proposed (Hodges and Lehmann, 1962; Puri and Sen, 1969; Koch, 1969; Hettmansperger, 1984). Mansouri (1999) investigated asymptotic properties of the aligned rank transform (ART) and outlined multiple comparison procedures. The aligned rank test has been shown to alleviate the major problems with the rank transform of Conover and Iman (Higgins and Tashtoush (1994), Richter and Payton (1999)). In addition, the test has power similar to the F-test for error distributions that are approximately normal, and can have much higher power for certain non-normal error distributions. However, some inflation of nominal Type I error rates have been observed for skewed error distributions.

Asymptotically distribution-free tests for testing nonparametric hypotheses in factorial experiments based on a rank version of the Wald statistic have also been developed (see Akritas 1990; Thompson 1991; Akritas and Arnold 1994; Akritas, Arnold and Brunner 1997). Brunner, Dette and Munk (1997) investigated a modified version of an adjustment by Box (1954) and Paitnik (1949) as a small sample alternative to the rank-based Wald statistic. They found this test to be an improvement over the rank version of the Wald test used by Akritas 1990, Thompson 1991, Akritas and Arnold 1994 and Akritas, Arnold and Brunner 1997, in terms of controlling the nominal Type I error rate, and it compares very favorably in terms of power, sometimes outperforming the rank version of the Wald test. They state this method should not be confused with the rank transform technique of Conover and Iman (1976) since their method adjusts for the effects introduced by ranking the observations. We show, however, that the nonparametric method of Brunner, Dette and Munk (1997), while improving upon the performance of the rank transform test procedure, still suffers from some of the same problems, and thus cannot be recommended.

In this paper we modify the nonparametric method of Brunner, Dette and Munk (1997) by applying their method to the aligned ranks of the data. We compare this new approach to three other rank tests: the F-test applied to the ranks of the data, the approximate F-test of Brunner, Dette and Munk (1997) applied to the ranks of the data, and the F-test applied to the aligned ranks of the data. We also compare the new test to parametric F-test and to the approximate F-test using the modified Box-type adjustment of Brunner, Dette and Munk (1997). We conducted a simulation study to examine the performance of the six tests in question. Each test was used to test for interaction and main effects for data in a two-factor layout, for a variety of effect sizes and combinations and for several error distributions. Standard normal and double exponential error distributions were used as examples of symmetric light-tailed and heavy-tailed distributions, respectively. We also considered three non-symmetric distributions: two contaminated standard normal distributions, one with 10% contamination from a $N(2,1)$ distribution (normal with mild to moderate outliers) and one with 10% contamination from a $N(9,1)$ distribution (normal with extreme outliers), and finally an exponential distribution (skewed, heavy-tailed). Simulation studies show that the proposed test alleviates the problem of inflated Type I error rates shown by both the test based on simple ranks and the rank test of Brunner, Dette and Munk (1997). In addition, we show that using the Box-type adjustment alleviates the problem of Type I error rate inflation seen previously for the aligned rank test, without a noticeable loss of power, as long as sample sizes are moderate ( $n \geq 7$ ).

## 2 Model Investigated

Consider the two-factor model

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} ,$$

where $\mu_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$, and

$\alpha_i$ is the effect of the *i*th level of factor *A*
$\beta_j$ is the effect of the *j*th level of factor *B*
$(\alpha\beta)_{ij}$ is the effect of interaction between the *i*th level of factor *A* and the *j*th level of factor *B*
$\varepsilon_{ijk}$ is the random error associated with the *k*th replicate under the *i*th level of factor *A* and the *j*th level of factor *B*.

The statistic used for all six tests described above is the usual analysis of variance F-ratio statistic. For example, to test for the effect of interaction, one would compute the statistic

$$F = \frac{MSAB}{MSE}$$

where *MSAB* refers to the sum of squares due to interaction divided by the degrees of freedom for interaction, and likewise *MSE* refers to the sum of squares error divided by degrees of freedom for error. For testing the effects of main effects of factors *A* and *B* compute

$$F = \frac{MSA}{MSE}$$

and

$$F = \frac{MSB}{MSE} ,$$

respectively.

For convenience, we will denote the F tests resulting from the six methods in question with the following nomenclature. The parametric F test and the approximate F test of Brunner, Dette and Munk (1997) will be denoted by F and FB, respectively. The F test utilized on the ranked and aligned ranked data will be denoted by FR and FAR, respectively. Finally, FRB and FARB will denote the F tests for the Brunner, Dette and Munk (1997) method applied to the ranks and aligned ranks, respectively.

To develop the approximate F-tests, it will be convenient to formulate the hypotheses and test statistics in terms of matrices. Let $\boldsymbol{\mu} = \{\mu_{ij}\}$ denote the *ab* x 1 vector of population means, $\mathbf{I}_c$ a $c \times c$ identity matrix, and $\mathbf{J}_c$ a $c \times c$ matrix of 1's. Then the hypotheses of no main effects and interaction can be written as

| |
|---|
| $H_0(A): \mathbf{M}_A \boldsymbol{\mu} = 0$ |
| $H_0(B): \mathbf{M}_B \boldsymbol{\mu} = 0$ |
| $H_0(AB): \mathbf{M}_{AB} \boldsymbol{\mu} = 0$ |

where

| |
|---|
| $\mathbf{M}_A = \left( \mathbf{I}_a - \frac{1}{a}\mathbf{J}_a \right) \otimes \frac{1}{b}\mathbf{J}_b$ |
| $\mathbf{M}_B = \frac{1}{a}\mathbf{J}_a \otimes \left( \mathbf{I}_b - \frac{1}{b}\mathbf{J}_b \right)$ |
| $\mathbf{M}_{AB} = \left( \mathbf{I}_a - \frac{1}{a}\mathbf{J}_a \right) \otimes \left( \mathbf{I}_b - \frac{1}{b}\mathbf{J}_b \right).$ |

The symbol $\otimes$ represents the Kronecker product of the matrices. If $\mathbf{Y} = \{Y_{ij}\}$ denotes the vector of observed values, then the sums of squares due to main effects and interaction can be written as

| |
|---|
| $SSA = \mathbf{Y}'\mathbf{M}_a\mathbf{Y}$ |
| $SSB = \mathbf{Y}'\mathbf{M}_b\mathbf{Y}$ |
| $SSAB = \mathbf{Y}'\mathbf{M}_{ab}\mathbf{Y}$ |

and also

$$SSE = \mathbf{Y}'\left( \mathbf{I}_a \otimes \mathbf{I}_b \otimes \left( \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n \right) \right)\mathbf{Y}.$$

Then the test statistics for testing main effects and interaction can be written as

| |
|---|
| $F = \frac{SSA/(a-1)}{MSE}$ |
| $F = \frac{SSB/(b-1)}{MSE}$ |
| $F = \frac{SSAB/[(a-1)(b-1)]}{MSE}$ |

where

$$MSE = \frac{SSE}{ab(n-1)}.$$

These statistics are used for both the F and the FB tests. The FB test adjusts the degrees of freedom associated with the F statistic, so the results will typically not be the same for the two tests.

For the approximate F-test statistics the degrees of freedom are given by

| |
|---|
| $df_{num} = \dfrac{m^2\left[tr\left(\hat{\mathbf{S}}\right)\right]^2}{tr\left[\left(\mathbf{M}\hat{\mathbf{S}}\right)\left(\mathbf{M}\hat{\mathbf{S}}\right)\right]},$ |
| $df_{den} = \dfrac{\left[tr\left(\hat{\mathbf{S}}\right)\right]^2}{tr\left(\hat{\mathbf{S}}^2\Lambda\right)}$ |

where $m$ is the constant diagonal entry of the matrix $\mathbf{M}$, $\hat{\mathbf{S}} = N \cdot diag\left\{\frac{\sigma_1^2}{n_1}, \ldots, \frac{\sigma_{ab}^2}{n_{ab}}\right\}$, $\mathbf{\Lambda} = diag\left\{\frac{1}{n_1 - 1}, \ldots, \frac{1}{n_{ab} - 1}\right\}$, and $N = n_1 + \cdots + n_{ab}$ (Brunner, Dette and Munk 1997).

For the FR and FRB tests, the vector $\mathbf{Y} = \{Y_{ijk}\}$ is replaced by $\mathbf{R} = \{R_{ijk}\}$, where $R_{ijk}$ denotes the rank of observation $Y_{ijk}$.

For tests FAR and FARB, the vector $\mathbf{Y} = \{Y_{ijk}\}$ is replaced by $\mathbf{AR} = \{AR_{ijk}\}$, where $AR_{ij}$ denotes the rank, after alignment, of the observation $Y_{ij.}$, and the observations are aligned as follows:

For testing the main effect of factor $A$, the aligned observation is $AY_{ijk} = Y_{ijk} - \hat{\beta}_j$ where $\hat{\beta}_j$ is estimated using the sample mean for the $j$th level of factor $B$. Likewise for testing the main effect of factor $B$, the aligned observation is $AY_{ijk} = Y_{ijk} - \hat{\alpha}_i$ where $\hat{\alpha}_i$ is estimated using the sample mean for the $i$th level of factor $A$. Finally, for testing interaction, the aligned observation is $AY_{ijk} = Y_{ijk} - \hat{\alpha}_i - \hat{\beta}_j$.

## 3 Simulation Study

The two-way model described above was used to perform the simulations. Two factor level combinations were used (3x6 and 4x3). Power and Type I error rate were examined for tests for interaction and main effects for a variety of effect sizes and combinations and for several error distributions. Standard normal and double exponential error distributions were used as examples of symmetric light-tailed and heavy-tailed distributions, respectively. Three non-symmetric error distributions were considered: contaminated standard normal with 10% contamination from a $N$ (2,1) distribution (normal with mild to moderate outliers); contaminated normal with 10% contamination from a $N$ (9,1) distribution (normal with extreme outliers); exponential (skewed, heavy-tailed). Previous results (Brunner, Dette and Munk(1997); Richter and Payton (2003)) have indicated that the FB test maintains the desired nominal Type I error rate for sample sizes of at least $n = 7$, thus all comparisons used this sample size. All simulation programs were written using SAS IML (SAS Institute, Cary, NC), and all simulations were based on 5000 random samples.

# 4 Results

We examined the ability of each test to detect effects that were present and to maintain a nominal Type I error rate of $\alpha = 0.05$ for effects that were not present. Using simulated results based on 5000 random samples, we can have 99% confidence that the simulated results are within $2.58\sqrt{\frac{0.05*0.95}{5000}} = 0.008$ of the true nominal rate, and thus any simulated result less than 0.04 or greater than 0.06 can be considered "significantly" different from the prescribed nominal rate. Therefore in the following discussion, it can be assumed that any test labeled as "conservative" had a simulated Type I error rate less than 0.04, and any test labeled "liberal" had a simulated Type I error rate greater than 0.06. A sampling of the results follows.

Table 1. Proportion of rejections at $\alpha = 0.05$, standard normal errors with equal variance, based on 5000 samples. 3x6 factorial, A and B effects present ($a_1=a_2=c;a_3= b1=0,b_2-b_6=c$). Equal cell sample size: $n_i$=7.

| Test for: | Method | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| Main Effect A | F | .0472 | .6532 | .9986 | 1 | 1 | 1 | 1 |
| | FB | .0446 | .6458 | .9986 | 1 | 1 | 1 | 1 |
| | FR | .0512 | .6288 | .9964 | 1 | 1 | 1 | 1 |
| | FRB | .05 | .6242 | .9964 | 1 | 1 | 1 | 1 |
| | FAR | .0484 | .63 | .9972 | 1 | 1 | 1 | 1 |
| | FARB | .047 | .625 | .9972 | 1 | 1 | 1 | 1 |
| Main Effect B | F | .049 | .3062 | .9028 | .999 | 1 | 1 | 1 |
| | FB | .0432 | .2846 | .8904 | .9988 | 1 | 1 | 1 |
| | FR | .0488 | .2822 | .8674 | .998 | 1 | 1 | 1 |
| | FRB | .0462 | .2736 | .8674 | .998 | 1 | 1 | 1 |
| | FAR | .0502 | .2828 | .8816 | .9984 | 1 | 1 | 1 |
| | FARB | .0474 | .2712 | .8756 | .9984 | 1 | 1 | 1 |
| Interaction | F | .054 | .054 | .054 | .054 | .054 | .054 | .054 |
| | FB | .0422 | .0422 | .0422 | .0422 | .0422 | .0422 | .0422 |
| | FR | .0496 | .0506 | .0558 | .0818 | .1468 | .2788 | .4472 |
| | FRB | .0438 | .043 | .0436 | .0616 | .1148 | .2254 | .3666 |
| | FAR | .0506 | .0506 | .0506 | .0506 | .0506 | .0506 | .0506 |
| | FARB | .0434 | .0434 | .0434 | .0434 | .0434 | .0434 | .0434 |

Table 1 shows results for a 3x6 layout with A and B main effects present (larger A main effect) and normally distributed errors. All tests have similar power for detecting main effects. Tests F, FB, FAR and FARB maintain Type I error rates near the nominal level of 0.05 for all effect sizes. Tests FR and FRB, however, exhibit inflation of Type I error rates in the test of interaction as the effect size increases. This is the same phenomenon that has been observed for test FR in previous simulation studies, and although the approximate F approach lessens the effect, it does not eliminate the problem.

Table 2. Proportion of rejections at $\alpha = 0.05$, contaminated normal (90% N(0,1), 10% N(9,1)) errors (normal with some extreme outliers), based on 5000 samples. 3x6 factorial, A and B effects present ($a_1$=$a_2$=$c$;$a_3$= $b1$=0,$b_2$-$b_6$=$c$). Equal cell sample size: $n_i$=7.

| Test for: | Method | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| Main Effect A | F | .048 | .13 | .3638 | .6768 | .9026 | .9844 | .9986 |
| | FB | .0368 | .113 | .3456 | .6628 | .892 | .9824 | .9984 |
| | FR | .0528 | .446 | .942 | .9986 | 1 | 1 | 1 |
| | FRB | .0506 | .44 | .9402 | .9986 | 1 | 1 | 1 |
| | FAR | .0512 | .3792 | .9086 | .998 | 1 | 1 | 1 |
| | FARB | .049 | .374 | .9058 | .9976 | 1 | 1 | 1 |
| Main Effect B | F | .0518 | .0762 | .1688 | .3552 | .5814 | .7736 | .911 |
| | FB | .0264 | .0416 | .1196 | .2902 | .5272 | .737 | .8894 |
| | FR | .0538 | .2068 | .6216 | .8834 | .96 | .9802 | .9872 |
| | FRB | .0512 | .1988 | .6062 | .8714 | .9526 | .9732 | .9824 |
| | FAR | .0568 | .1952 | .6174 | .9098 | .9796 | .9918 | .9954 |
| | FARB | .0526 | .1874 | .606 | .9034 | .9758 | .9904 | .9946 |
| Interaction | F | .0514 | .0514 | .0514 | .0514 | .0514 | .0514 | .0514 |
| | FB | .02 | .02 | .02 | .02 | .02 | .02 | .02 |
| | FR | .0494 | .0518 | .053 | .0646 | .0922 | .1302 | .1582 |
| | FRB | .0426 | .0444 | .0436 | .0458 | .0628 | .0802 | .0982 |
| | FAR | .0614 | .0614 | .0614 | .0614 | .0614 | .0614 | .0614 |
| | FARB | .0448 | .0448 | .0448 | .0448 | .0448 | .0448 | .0448 |

Table 2 contains results for the same design except with errors from a contaminated normal CN(9) distribution. Here we observe that all tests have less power to detect main effects, but the power of all the rank tests is much higher than for the F and FB tests. Once again, as the effect sizes increase, we see an inflation of Type I error rates in the test of interaction for the FR and FRB tests. The FAR test has a slightly inflated Type I error rate in the same situation, but the rate is constant and does not become worse as the effect size increases. Note that the Type I error rate of the FARB test shows no inflation, a finding that was consistent for all simulations. In addition, the power of the FARB test for detecting main effects is virtually identical to that of the FAR test. The FB test does not perform well for this skewed error distribution, with lowest overall power and a very conservative Type I error rate (0.02) in the test for interaction.

Table 3. Proportion of rejections at $\alpha = 0.05$, double exponential errors (symmetric, heavy-tailed), based on 5000 samples. 4x3 factorial, A and B effects present ($a_2$=$b_1$= $c$;$a_3$=$b_2$=$c$). Equal cell sample size: $n_i$=7.

| Test for: | Method | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| Main Effect A | F | .0474 | .4384 | .9676 | 1 | 1 | 1 | 1 |
| | FB | .0414 | .4004 | .9586 | 1 | 1 | 1 | 1 |
| | FR | .0546 | .544 | .982 | 1 | 1 | 1 | 1 |
| | FRB | .0504 | .5304 | .9804 | 1 | 1 | 1 | 1 |
| | FAR | .053 | .5608 | .989 | 1 | 1 | 1 | 1 |
| | FARB | .0502 | .5512 | .9888 | 1 | 1 | 1 | 1 |

| Test for: | Method | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| Main Effect B | F | .0438 | .6474 | .9972 | 1 | 1 | 1 | 1 |
| | FB | .0394 | .627 | .996 | 1 | 1 | 1 | 1 |
| | FR | .05 | .7654 | .9988 | 1 | 1 | 1 | 1 |
| | FRB | .0492 | .7588 | .9988 | 1 | 1 | 1 | 1 |
| | FAR | .0486 | .7718 | .9992 | 1 | 1 | 1 | 1 |
| | FARB | .0468 | .7672 | .9992 | 1 | 1 | 1 | 1 |
| Interaction | F | .0494 | .0494 | .0494 | .0494 | .0494 | .0494 | .0494 |
| | FB | .0338 | .0338 | .0338 | .0338 | .0338 | .0338 | .0338 |
| | FR | .0502 | .0522 | .0572 | .0748 | .1082 | .1668 | .2608 |
| | FRB | .0448 | .0428 | .0432 | .0548 | .0778 | .1204 | .194 |
| | FAR | .0508 | .0508 | .0508 | .0508 | .0508 | .0508 | .0508 |
| | FARB | .0464 | .0464 | .0464 | .0464 | .0464 | .0464 | .0464 |

Table 3 shows results for a 4x3 layout, again with A and B main effects present and errors from a double exponential distribution. Here the rank tests again have greater power, although the power advantage is not quite as great as that observed for the skewed error distributions. Once again for tests FR and FRB we see inflation in the Type I error rate for the test of interaction.

Table 4. Proportion of rejections at $\alpha = 0.05$, exponential errors (skewed, heavy-tailed), based on 5000 samples. 4x3 factorial, A*B effect present ($ab_{11}=ab_{33}=c;ab_{13}=ab_{31}=-c$). Equal cell sample size: $n_i$=7.

| Test for: | Method | 0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| Main Effect A | F | .0434 | .0434 | .0434 | .0434 | .0434 | .0434 | .0434 |
| | FB | .0286 | .0286 | .0286 | .0286 | .0286 | .0286 | .0286 |
| | FR | .0472 | .0476 | .0444 | .0444 | .0406 | .0404 | .0392 |
| | FRB | .0442 | .0434 | .0394 | .0378 | .0338 | .0324 | .0334 |
| | FAR | .049 | .0478 | .0476 | .0462 | .0448 | .0444 | .0432 |
| | FARB | .0458 | .044 | .0408 | .039 | .0386 | .0348 | .0338 |
| Main Effect B | F | .0456 | .0456 | .0456 | .0456 | .0456 | .0456 | .0456 |
| | FB | .0338 | .0338 | .0338 | .0338 | .0338 | .0338 | .0338 |
| | FR | .052 | .0536 | .0522 | .0518 | .0474 | .0486 | .0468 |
| | FRB | .0508 | .051 | .0484 | .045 | .0422 | .0442 | .0414 |
| | FAR | .053 | .0532 | .0542 | .0522 | .0484 | .0474 | .0448 |
| | FARB | .051 | .0514 | .049 | .0486 | .0432 | .04 | .0392 |
| Interaction | F | .0394 | .448 | .9676 | 1 | 1 | 1 | 1 |
| | FB | .0218 | .3572 | .943 | .9996 | 1 | 1 | 1 |
| | FR | .0466 | .72 | .9972 | 1 | 1 | 1 | 1 |
| | FRB | .0412 | .699 | .9956 | 1 | 1 | 1 | 1 |
| | FAR | .051 | .6696 | .9956 | 1 | 1 | 1 | 1 |
| | FARB | .0424 | .6352 | .994 | 1 | 1 | 1 | 1 |

We also modeled only one main effect present, only interaction effect present and both main effects and interaction effects present. The results were consistent with those described above. While the FARB test never showed inflated Type I error rates for effects not present in the model, there was an indication of the proportion of rejections getting smaller as the interaction effect size increased, when only interaction was in the model (See Table 4). This did not seem to have any noticeable effect on the power to detect main effects when they were present along with interaction, however (See Table 5).

Table 5. Proportion of rejections at $\alpha = 0.05$, contaminated normal (9) errors, based on 5000 samples. 4x3 factorial, all effects present ($ab_{11}=c; ab_{41}=b1=-c$). Equal cell sample size: $n_i$=7.

| **Test for:** | **Method** | **0** | **0.5** | **1.0** | **1.5** | **2.0** | **2.5** | **3.0** | **4.0** |
|---|---|---|---|---|---|---|---|---|---|
| Main Effect A | F | | .051 | .061 | .0812 | .1116 | .1506 | .2006 | .32 |
| | FB | | .0294 | .0366 | .0502 | .0746 | .1124 | .157 | .2694 |
| | FR | | .0688 | .0914 | .1034 | .1032 | .1036 | .1018 | .109 |
| | FRB | | .0648 | .0854 | .096 | .092 | .0898 | .089 | .0958 |
| | FAR | | .0672 | .1134 | .1944 | .2962 | .3896 | .4826 | .5956 |
| | FARB | | .0628 | .107 | .184 | .2854 | .3762 | .4682 | .582 |
| Main Effect B | F | | .129 | .3786 | .6934 | .8968 | .979 | .9968 | 1 |
| | FB | | .1016 | .3476 | .6716 | .881 | .9768 | .996 | 1 |
| | FR | | .4524 | .9308 | .9948 | .999 | .9996 | .9998 | 1 |
| | FRB | | .4476 | .928 | .9942 | .9988 | .9996 | .9996 | 1 |
| | FAR | | .4014 | .8926 | .9894 | .9988 | .9996 | .9998 | 1 |
| | FARB | | .3938 | .8876 | .9882 | .9982 | .9994 | .9996 | 1 |
| Interaction | F | | .05 | .062 | .089 | .1262 | .1836 | .2592 | .4358 |
| | FB | | .0214 | .027 | .0404 | .0642 | .1034 | .1602 | .3342 |
| | FR | | .0702 | .108 | .126 | .1274 | .1274 | .127 | .1334 |
| | FRB | | .06 | .092 | .1026 | .0996 | .0964 | .0926 | .0946 |
| | FAR | | .0776 | .1348 | .253 | .4156 | .5938 | .7298 | .8786 |
| | FARB | | .0604 | .1138 | .219 | .3798 | .5552 | .7004 | .8608 |

# 5 Example

The following data set (Table 6) for a 3x3 factorial experiment is adapted from Freund and Wilson (1997).

Table 6. Responses for a 3x3 factorial experiment.

| | | FACTOR B | |
|---|---|---|---|
| Levels | 1 | 2 | 3 |
| | 3.6 | 7.1 | 8.2 |
| 2 | 4.5 | 7.3 | 6.0 |
| | 3.9 | 5.0 | 9.8 |
| | 7.7 | 9.8 | 10.3 |

| | | FACTOR B | | |
|---|---|---|---|---|
| FACTOR A | 4 | 7.7 | 10.0 | 12.3 |
| | | 7.7 | 8.6 | 8.6 |
| | | | | |
| | | 6.5 | 8.0 | 6.0 |
| | 6 | 8.5 | 6.5 | 6.9 |
| | | 6.3 | 9.6 | 6.1 |

The corresponding aligned ranks of the data are given in Table 7.

Table 7. Aligned and ranked responses from Table 6.

| | | FACTOR B | | |
|---|---|---|---|---|
| | Levels | 1 | 2 | 3 |
| | | 5 | 20 | 23 |
| | 1 | 11 | 22 | 10 |
| | | 8 | 3 | 27 |
| FACTOR A | | 13 | 15 | 18 |
| | 2 | 13 | 16 | 25 |
| | | 13 | 7 | 4 |
| | | 21 | 17 | 1 |
| | 3 | 26 | 6 | 9 |
| | | 19 | 24 | 2 |

The analysis of variance tests on the raw data, aligned ranks with unadjusted and adjusted p-values, for main effects and interaction, are given below.

Table 8. F-statistics and p-values for the three test procedures. The P-value columns represent the following: "Raw" – ANOVA on raw data; "Unadjusted"- ANOVA on aligned ranks; "Adjusted"- Adjusted ANOVA on aligned ranks.

| | | | **P-value** | |
|---|---|---|---|---|
| **Effect** | **F** | **Raw** | **Unadjusted** | **Adjusted** |
| A | 14.23 | 0.0002 | 0.0001 | 0.0010 |
| B | 6.89 | 0.0060 | 0.0160 | 0.0240 |
| A*B | 3.18 | 0.0384 | 0.040 | 0.0660 |

# 6 Conclusion

We recommend the FARB test. The FAR test performed well also, but has shown a tendency to have slightly inflated Type I error rates for skewed error distributions. The approximate F-test using the modified Box-type adjustment of Brunner, Dette and Munk (1997) eliminates this problem without a noticeable loss of power. The same procedure on the nonaligned ranks cannot be recommended, due to the familiar problem of inflated nominal Type I error rates also associated with the rank transform procedure.

## References


Akritas, M.G. (1990). The rank transform method in some two-factor designs, *Journal of the American Statistical Association*, **85**, 73-78.

Akritas, M.G. and Arnold, S.F. (1994). Fully nonparametric hypotheses for factorial designs I: multivariate repeated-measures designs, *Journal of the American Statistical Association*, **89**, 336-343.

Akritas, M.G., Arnold, S.F. and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs, *Journal of the American Statistical Association*, **92**, 258-265.

Blair, R. C., Sawilowsky, S. S. and Higgins, J. J. (1987). Limitations of the rank transform statistic in tests of interaction, *Communications in Statistics: Computation and Simulation,* **B16**, 1133-1145.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *Annals of Mathematical Statistics,* **25**, 290-302.

Brunner, E., Dette, H. and Munk, A. (1997). Box-type approximations in nonparametric factorial designs, *Journal of the American Statistical Association*, **92**, 1494-1502.

Conover, W.J. and Iman, R.L. (1976). On some alternative procedures using ranks for the analysis of experimental designs, *Communications in Statistics,* **A5**, pp. 1349-1368.

Conover, W.J. and Iman, R.L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics, *The American Statistician*, **35**, 124-129.

Fligner, M. A. (1981). Comment. *The American Statistician,* **35**, 131-132.

Freund, R. J. and Wilson, W. J. (1997). *Statistical Methods*. Academic Press, San Diego, CA.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks.* John Wiley, New York.

Higgins, J. J. and Tashtoush, S. (1994). An aligned rank transform test for interaction, *Nonlinear World*, **1**, 201-211.

Hodges, J. L. and Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance, *Annals of Mathematical Statistics,* **27**, 324-335.

Koch, G. G. (1969). Some aspects of the statistical analysis of split-plot experiments in completely randomized layouts, *Journal of the American Statistical Association,* **64**, 485-506.

Mansouri (1999). Multifactor analysis of variance based on the aligned rank transform technique, *Computational Statistics and Data Analysis*, **29**, 177-189.

Paitnik, P. B. (1949). The noncentral chi-square and F-distributions and their applications, *Biometrika*, **36**, 202-232.

Puri, M. L. and Sen, P. K. (1969). A class of rank order tests for a general linear model, *Annals of Mathematical Statistics,* **40**, 1325-1343.

Richter, S.J. and Payton, M.E. (1999). Nearly exact tests in factorial experiments using the aligned rank transform, *Journal of Applied Statistics*, **26**, 203-217.

Richter, S. J. and Payton, M. E. (2003). Performing two-way analysis of variance under variance heterogeneity, *Journal of Modern Applied Statistical Methods*, in press.

Thompson, G.L. (1991). A unified approach to rank tests for multivariate and repeated measures designs, *Journal of the American Statistical Association*, **86**, 410-419.

Thompson, G. L. and Amman, L. P. (1990). Efficiencies of interblock rank statistics for repeated measures designs, *Journal of the American Statistical Association,* **85**, 519-528.

Toothaker, L. E. and Chang, H. (1980). On the analysis of ranked data derived from completely randomized factorial designs, *Journal of Educational Statistics*, **5**, 169-176.

Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics*, **I**, 80-82.