

An aligned rank transform test for interaction

James J. Higgins¹ and Suleiman Tashtoush²

¹Kansas State University, Manhattan, KS, USA

²Mu'tah University, Mu'tah Al-Karak, Jordan

Received March 3, 1993

Abstract. A solution is proposed for the important problem of testing for interaction in factorial experiments when Gaussian assumptions are violated. The proposed rank test can be implemented with existing statistical packages and provides a fix-up for the flawed rank transform procedure. Simulation results suggest that the test is valid for the small and moderate sample sizes typically found in practice when error distributions are symmetric or moderately skewed. The procedure has advantages over standard analysis of variance in the presence of outliers or when error distributions are heavy tailed.

Key words: Factorial experiments, split-plot repeated measures design, robust interference

AMS classification scheme numbers: 62G10, 62G35

1. Introduction

Analysis of variance is a powerful tool for exploring causal relationships among variables in multifactor experiments. The simplest mathematical models for responses observed in such experiments are additive. That is, the aggregate effect of the factors being studied is a sum of the effects of the individual factors. More complex models include nonadditive effects, or interactions. An important statistical problem is that of testing for the presence of interactions. Such tests can reveal complex relationships among the factors not adequately explained by the additive model.

It is well-known that statistical procedures such as analysis of variance, which are optimal under the assumption of a Gaussian probability model, can be sensitive to the violations of the Gaussian assumption or to the presence of outliers. Such procedures involve sample means which are notoriously ill-behaved when outliers are present or when distributions of observations are heavy tailed. The

consequences of this can be particularly devastating when using analysis of variance to test for interactions in multifactor experiments. Spurious effects due to violations of assumptions can mask themselves as interaction effects leading researchers to postulate more complicated structures than warranted by the underlying model.

Rank tests have long been valued as alternatives to Gaussian theory methods when Gaussian assumptions are violated. While such tests have long been available for single factor experiments or for multifactor experiments with additive effects, the development of corresponding procedures for multifactor experiments with interactions has lagged behind. This is due in part to computational complexities of various proposed methods and in part to lack of consensus on which methods should be developed. Sen (1968) proposed a class of rank order tests in two-way layouts which solve in the two-way setting the problem posed here but which have never been computationally implemented in the standard statistical packages. Draper (1988) reviewed rank-based robust methods which could also be used in this context, but again such methods have lagged in implementation.

Among the rank based procedures for multifactor experiments presently in use, the one that appears to have gained the most popularity is the rank transform method proposed by Iman and Conover (1981). In this method, the analysis is carried out by replacing original observations with their respective ranks, computing parametric tests on these ranks (typically an analysis of variance), and referring the values of the test statistics so obtained to the usual table of critical values. Its popularity comes from the fact that it can be implemented in most of the standard statistical packages by simply ranking data, which most packages do, and applying standard methods to the ranked data. For instance, this is a procedure recommended by IMSL (1987) for two and three factor experiments.

Theoretical results (Iman, Hora, and Conover, 1984) suggest that the rank transform procedure provides asymptotically valid tests for analyzing experiments when additive effects are present. Simulation studies (Conover and Iman, 1976, and Iman, 1976) carried out in a connection with a 4×3 factorial design seem to show that the rank transform procedure is also valid for analyzing experiments with interaction. These authors have concluded that the rank transform tests are competitive with Gaussian theory tests under Gaussian assumptions and may enjoy considerable power advantages over such tests when observations are taken from non-Gaussian distributions.

Unfortunately, the evidence for the validity of the rank transform procedure applied to factorial experiments with interaction comes from rather incomplete simulation studies. Blair, Higgins, and Sawilosky (1987) were among the first to raise questions about the limitations of this methodology. Subsequent work by Sawilosky, Blair, and Higgins, (1989), Akritas (1990, 1991), and Thompson (1991) have further shed light on the potential undesirable behavior of the rank transform method when nonadditive effects are present. The underlying reason for the flaw is simple. When nonlinear transformations such as the rank transform

are made on a set of variables, the transformed data may or may not exist. There is no justification for the use of rank tests in factorial experiments where it is appropriate.

The procedure proposed here is set forth by Sen (1968). Applied by Higgins (1989) in a factorial experiment, the procedure is a split-plot repeated measures design on our recommended two-way factorial treatment combinations. Recommended procedures for applications to other

2. The aligned rank transform

Alignment is a procedure for removing the effect of "nuisance" factors. For instance, to remove the effect of a factor (Sen, 1975). In the aligned rank transform, the data are then analyzed using the aligned rank transform for a balanced two-way factorial design.

2.1 The aligned rank transform

The mathematical

where $i = 1, \dots, r$ variables with means μ_i will be referred to as the interaction effect. The adjustment factor is subtracted from the data so that the adjusted means are α_i , β_j , and $(\alpha\beta)_{ij}$. The adjustment factor is $\sum_i(\alpha\beta)_{ij} = \sum_j(\alpha\beta)_{ij}$.

are made on a set of data, interaction structures that exist in the original data may or may not exist in the transformed data. That is, interaction may exist in the transformed data but not in the original data, or vice versa. As a result, there is no justification for the general applicability of the rank transform procedure in factorial experiments with interaction although there may be special cases where it is appropriate.

The procedure proposed here combines the notion of alignment of the data as set forth by Sen (1968) with the rank transform method of Iman and Conover. Applied by Higgins, Tashtoush, and Blair (1990) in the context of agricultural experimentation, the procedure is quite general and can be applied to balanced factorial experiments of all dimensions in completely random designs and in split-plot repeated measures experimental designs. Section 2 of the paper focuses on our recommended procedure in completely randomized designs with two-way factorial treatment structures. Section 3 investigates modifications of the recommended procedure using robust estimates of location. Section 4 discusses applications to other designs.

2. The aligned rank transform in two-way completely random designs

Alignment is a procedure applied in multiparameter models for removing the effect of "nuisance" parameters when testing for the effects of parameters of interest. For instance, in a randomized complete block design data may be aligned to remove the effect of blocks in testing for the effect of treatments (Lehmann, 1975). In the aligned rank transform procedure, data are aligned, ranked, and then analyzed using an appropriate parametric procedure. In this section, the aligned rank transformation procedure is defined and its properties investigated for a balanced two-way completely randomized design.

2.1 The aligned rank transform procedure

The mathematical model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where $i = 1, \dots, r$, $j = 1, \dots, c$, $k = 1, \dots, n$, and the ϵ_{ijk} 's are i.i.d. random variables with mean 0 and common standard deviation σ . The α_i 's and β_j 's will be referred to as row effects and column effects, respectively, and $(\alpha\beta)_{ij}$ is the interaction effect. To align data in testing for interaction, an adjustment factor is subtracted from the Y_{ijk} 's to remove the row effects and column effects so that the resulting data will depend only on the $(\alpha\beta)_{ij}$'s and the ϵ_{ijk} 's. The adjustment factors proposed here are based on the usual estimates of μ , α_i , β_j , and $(\alpha\beta)_{ij}$ under the following restrictions: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$. These estimates are: $\hat{\mu} = \bar{Y}_{...}$, $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$,

Table 2. Maximum power advantages attained by AOV and aligned rank procedures CRD with 4×3 treatment structure test for interaction, $\alpha = .05$, 1000 simulated data sets $E(Y_{11k}) = E(Y_{33k}) = c\sigma$; $E(Y_{13k}) = E(Y_{31k}) = -c\sigma$; $E(Y_{ijk}) = 0$, other

		Number of observations per cell		
		2	5	10
Error distribution				
Gaussian	AR	.012	.000	.001
	AOV	.047	.035	.026
uniform	AR	.017	.000	.002
	AOV	.052	.070	.088
$tdf = 3$	AR	.044	.144	.193
	AOV	.000	.000	.000
exponential	AR	.068	.176	.323
	AOV	.011	.000	.000

to expect of rank tests. For light tailed, symmetric distributions, the analysis of variance procedure has a modest power advantage over the aligned rank transform procedure, generally less than .10 at the maximum. On the other hand, for heavy tailed distributions or skewed distributions, the aligned rank transform procedure is superior to analysis of variance, and the power advantages can be substantial with maximum power advantages often in the .15 to .30 range.

Typical results obtained in the simulation study are summarized in Table 2. Power functions for the analysis of variance and the aligned rank procedures were simulated in a 4×3 factorial treatment structure with alternatives of the form

$$E(Y_{11k}) = E(Y_{33k}) = c\sigma, \quad E(Y_{13k}) = E(Y_{31k}) = -c\sigma, \\ E(Y_{ijk}) = 0 \quad \text{otherwise,}$$

where c ranged from 0 to a value sufficiently large for the simulated power functions to be greater .99. The simulations were based on 1000 repetitions of the sampling experiment with $\alpha = .05$. Since it happened in some cases that the simulated power functions crossed each other (either due to sampling error in the estimates of power or due to real differences), the maximum power difference attained by each power function over the other was obtained. These results are reported in Table 2. For the Gaussian and uniform distributions, the analysis of variance procedure never had a power advantage greater than .09. On the other hand, the rank transform procedure enjoyed much larger power advantages over the analysis of variance procedure when samples were selected from the exponential distribution and the Student's t distribution with three degrees of freedom. In one case the maximum advantage exceeded .30.

Table 3. Simulated type I error rates for aligned rank transform procedure robust estimates of location CRD with 4×3 factorial treatment structure test for interaction, $\alpha = .05$, 1000 simulated data sets

Error distribution	Number of observations per cell			
	median		trimmed mean	
	5	10	5	10
Gaussian	.078	.061	.076	.064
<i>tdf</i> = 3	.069	.065	.069	.056
exponential	.091	.075	.080	.078

3. Alignment based on robust estimates of location

We investigated alternative methods of alignment based on robust estimates of location of cell means. In the alignment procedure it is possible to use any estimate of location in place of a sample cell mean as follows. Let m_{ij} denote any robust estimate of the $E(Y_{ijk})$. The aligned data then take the form

$$AB_{ijk} = Y_{ijk} - \bar{m}_{i.} - \bar{m}_{.j} + \bar{m}_{...}$$

The robust estimates considered in this study were the median and the 40% trimmed mean (20% trimmed from each side). When error distributions are symmetric and have finite mean, the mean, median, and trimmed mean all estimate the mean of the distribution from which samples are taken. However, this is not the case when the error distribution is skewed. Thus alignment using robust estimates for skewed error distributions may not be appropriate. Indeed investigation of Type I error rates showed the robust procedures to give unacceptable inflation in Type I error rates for the exponential distribution. Thus, we do not recommend the use of robust estimates except in cases of symmetric error distributions. While, for the symmetric error distributions, the Type I errors tended to be somewhat inflated especially for the smaller sample sizes, they appear to be close enough to the nominal values for practical purposes. Simulated Type I error rates are shown in Table 3.

The power advantages of tests based on alignment with robust estimates of location over those based on alignment with means is modest, with the maximum advantage being generally less than .07. Given the slight inflation in Type I error rate for the robust alignment compared to the alignment with means, the real advantage appears to be even less than this. Typical results are shown in Table 4 for two symmetric distribution, the Gaussian and Student's t with 3 degrees of freedom.

Table 4. Max. p with 4 by 3 fact data sets $E(Y_{ijk})$

error distribution
Gaussian
tdf = 3

Table 5. Type I error rates for treatment structure from standard Gaussian data sets

TEST
AOV
AR (mean)
AR (median)
AR (trimmed mean)

Another issue in the analysis of variance is to investigate this effect in a design with no interaction in a Gaussian distribution. When simulating outliers, the Type I error rates do not show significant inflation for spurious interaction rates above the nominal level.

As the simulation results show, the Type I error rates are immune from the effect of outliers, with error rates ranging from .05 to .10. Type I error rates for $\mu = 6$ are in the range .10 to .15, with no significant inflation for alignment with robust estimates, somewhat smaller

robust estimates of interaction, $\alpha = .05$,

per cell trimmed mean
10
.064
.056
.078

robust estimates of interaction possible to use any vs. Let m_{ij} denote the form

median and the 40% trimmed mean all estimates taken. However, alignment using appropriate. Indeed, estimates to give unacceptably high Type I error rates. Thus, in cases of symmetry, the Type I error rates for smaller sample sizes, for practical purposes.

robust estimates of interaction with the maximum inflation in Type I error rates. In fact, the real situation is shown in Table 4 with 3 degrees of

Table 4. Max. power advantages of robust alignment over alignment with means, CRD with 4 by 3 factorial treatment structure, test for interaction, $\alpha = .05$, 1000 simulated data sets $E(Y_{11k}) = E(Y_{33k}) = c\sigma$; $E(Y_{13k}) = E(Y_{31k}) = -c\sigma$; $E(Y_{ijk}) = 0$, other

	Number of observations per cell			
	median		trimmed mean	
error distribution	5	10	5	10
Gaussian	.035	.051	.027	.047
$tdf = 3$.062	.043	.062	.032

Table 5. Type I error rates for AOV and aligned rank tests with outliers 4×3 factorial treatment structure, two outliers in cell 1,1 with mean $\mu = 2, 4, 6$, other observations from standard Gaussian distribution five observations per cell, $\alpha = .05$, 1000 simulated data sets

TEST	μ		
	2	4	6
AOV	.103	.169	.278
AR (mean)	.086	.114	.129
AR (median)	.110	.113	.119
AR (trimmed mean)	.092	.087	.100

Another issue of importance is the possibility that outliers may cause the analysis of variance test to show interaction when no interaction exists. To investigate this effect, five observations per cell were generated in a 4×3 factorial design with no interaction. All observations were selected from a standard Gaussian distribution except for two observations in cell 1,1 which were generated from a Gaussian distribution with mean μ , $\mu = 2, 4, 6$, and variance 1, thus simulating outliers in one cell. What one would hope is that the tests would not show significant interactions due to the outlier effects. The extent to which spurious interactions were detected was measured by the inflation in Type I error rate above the nominal .05 rate.

As the simulated Type I error rates in Table 5 show, none of the tests were immune from the effects of outliers. For the mean $\mu = 2$, tests had Type I error rates ranging from .086 to .110. As the mean of the outliers increased, the Type I error rate for analysis of variance continued to increase to a value of .278 for $\mu = 6$. However, the Type I error rates for the aligned tests stabilized in the range .100 to .129. For the larger values of μ , there was somewhat more inflation for alignment with means while alignment with trimmed means had somewhat smaller inflation.

There appears to be some slight advantage in aligning with robust estimates in terms of power and in terms of controlling the effects of outliers when observations have symmetric error distributions. However, it is the process of ranking after alignment that gives the aligned rank transform tests their desirable properties. Given the ease with which alignment with means can be done, we would generally recommend this procedure.

4. Other designs

4.1 Split-plot or repeated measures designs

The model for the split-plot or repeated measures design is given by

$$Y_{ijk} = \mu + \alpha_i + \delta_{ik} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where $i = 1, \dots, r$, $j = 1, \dots, c$, $k = 1, \dots, n$, the δ_{ik} 's are i.i.d. random variables with mean 0 standard deviation σ_δ , the ϵ_{ijk} 's are i.i.d. random variables with mean 0 and standard deviation σ_ϵ , and the δ_{ik} 's are independent of the ϵ_{ijk} 's. The α_i 's are the whole-plot or between subjects effects and the δ_{ik} 's and the whole-plot or between subjects errors. The β_j 's are the subplot or within subject effects, and the ϵ_{ijk} 's are the subplot effects or within subject error. The term $(\alpha\beta)_{ij}$ is the interaction term.

There are two methods of alignment that may be used. A naive approach would be to use the same alignment as in the two-way completely random design (the CRD method). This method is generally not desirable since it does not remove the effect of the whole-plot error. In the method proposed below, the effect of the whole-plot error is removed in the alignment process. This generally leads to a more powerful test especially when the whole-plot error variance is large in relation to subplot error variance as is often the case. To test for interactions, whole-plot effects and whole-plot errors are removed by subtracting the whole-plot means $\bar{Y}_{i.k}$'s from the observations. Then the subplot effects are subtracted. Finally, the aligned data are ranked and the usual split-plot analysis of variance is performed on aligned ranks. As before, sums of squares for main effects are ignored while the F ratio for the interaction effect is referred to the F distribution with appropriate degrees of freedom. The aligned data for testing for interactions is given by

$$\begin{aligned} AB_{ijk} \text{ (split-plot)} &= Y_{ijk} - \bar{Y}_{i.k} - \hat{\beta}_j \\ &= Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{.j.} + \bar{Y} \dots \end{aligned}$$

Table 6 shows simulated power for the analysis of variance tests and the aligned rank transform tests using both the split-plot alignment and the naive alignment for completely random designs. Clearly we see the advantages of the

Table 6. Simulate treatment structur
 $-E(Y_{41k}) = 1.5$,
 data sets

Whole plot std. d
 Error distribution
 Gaussian
 uniform
 $tdf = 3$
 exponential

AOV = Analysis
 SP/AR = split-plc
 CR/AR = comple
 ** AOV and SP/A

aligned rank tran
 when error distr
 advantages over
 smaller, but as v
 lose power.

4.2 Higher orde

Interactions in e
 using the aligned
 the alignment for

$$AB_{ijk} = Y_{ijk}$$

For three-way int
 is given schemat

$$ABC_{ij}$$

The pattern for n

Table 6. Simulated power for aligned rank procedures for split-plot designs 4×3 factorial treatment structure test for interaction, $\alpha = .05$, subplot std. dev. $\sigma = 1$ $E(Y_{11k}) = -E(Y_{41k}) = 1.5$, $E(Y_{ijk}) = 0$ for other i and j , 5 observations per cell, 1000 simulated data sets

	Test statistic				
	AOV	SP/AR	CR/AR		
Whole plot std. dev.	**	**	1	2	4
Error distribution					
Gaussian	.732	.716	.695	.653	.636
uniform	.732	.703	.664	.666	.664
$tdf = 3$.374	.512	.514	.436	.373
exponential	.756	.824	.812	.781	.745

AOV = Analysis of variance test

SP/AR = split-plot aligned rank test

CR/AR = completely random aligned rank test

** AOV and SP/AR not affected by size of whole-plot std. dev.

aligned rank transform method with split-plot alignment in testing for interaction when error distributions are heavy tailed. The naive alignment also has power advantages over the analysis of variance when whole-plot error variances are smaller, but as whole-plot error variances get larger the naive alignment may lose power.

4.2 Higher order interactions in completely random designs

Interactions in experiments involving higher order interactions can be tested using the aligned rank transform procedure. To see the pattern, let us re-express the alignment for the two-way case schematically as follows:

$$AB_{ijk} = Y_{ijk} - (\text{sum of one-way means involving } i, j) + \text{overall mean.}$$

For three-way interactions in experiments involving three factors, the alignment is given schematically as

$$\begin{aligned} ABC_{ijkl} = & Y_{ijkl} - (\text{sum of 2-way means involving } i, j, k) \\ & + (\text{sum of 1-way means involving } i, j, k) \\ & - \text{overall mean.} \end{aligned}$$

The pattern for more than three factors is apparent.

Tests are carried out by ranking the aligned data and applying n -way analysis of variance to the ranks, ignoring all tests so obtained except for the test of interaction of interest.

5. Summary

Although the aligned rank transform tests are not distribution free, they are robust with respect to the underlying error distribution, and they have many of the desirable power properties of the common nonparametric tests. The methodology has wide applicability and may be carried out with standard analysis of variance programs after data are aligned and ranked which themselves are procedures that may be carried out with ease with many popular statistical packages. Moreover, the tests do not have the same potential for giving misleading results as the ordinary rank transform tests when applied to multifactor experiments with interaction.

References

1. AKRITAS, M. G. (1990). The Rank Transform Method in Some Two-Factor Designs. *Journal of the American Stat. Assoc.* **85**, p. 73–78.
2. AKRITAS, M. G. (1991). Limitations of the Rank Transform Procedure: A Study of Repeated Measures Designs, Part I. *Journal of the American Stat. Assoc.* **86**, p. 457–460.
3. BLAIR, R. C., HIGGINS, J. J., SAWILOSKY, S. S. (1987). Limitations of the Rank Transform Statistic in Tests for Interaction. *Communications in Stat.- Simulation* **16** (4), p. 1133–1145.
4. DRAPER, David (1988). Rank-Based Robust Analysis of Linear Models. I. Exposition and Review. *Statistical Science*. **3** (2), p. 239–271.
5. CONOVER, W. J., IMAN, R. L. (1976). On Some Alternative Procedures Using Ranks for the Analysis of Experimental Designs. *Communications in Stat.* **A5**, p. 1349–1368.
6. HIGGINS, J. J., BLAIR, R. C., TASHTOUSH, S. (1990). The Aligned Rank Transform Procedure. *Proceedings of the 1990 Conference on Applied Statistics in Agriculture*. p. 185–195.
7. IMAN, R. L. (1974). A Power Study of a Rank Transform for the TwoWay Classification when Interactions May Be Present. *Canadian Journal of Stat.* **2**, p. 227–239.
8. IMAN, R. L., CONOVER, W. J. (1981). Rank Transformation as a Bridge Between Parametric and Nonparametric Statistics. *American Statistician* **35**, p. 131–132.
9. IMAN, R. L., HORA, S. C., CONOVER, W. J. (1984). Comparison of Asymptotically Distribution-Free Procedures for the Analysis of Complete Blocks. *Journal of the American Stat. Assoc.* **79**, p. 674–685.
10. International Mathematical and Statistical Library (1987). IMSL Library Interface (April, Version 10.0/LIBR-10G-10.0), Houston.
11. LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based On Ranks*. Holden-Day, Inc. San Francisco.

12. SAWILOSKY, S.
Error and Pov
Journal of Ed
13. SEN, P. K. (19
Annals of Ma
14. THOMPSON, G
peated Measu:

12. SAWILOSKY, S. S. BLAIR, R. C., HIGGINS, J. J. (1989). An Investigation of the Type I Error and Power Properties of the Rank Transform Procedure in Factorial ANOVA. *Journal of Educational Stat.* **14**, p. 225–268.
13. SEN, P. K. (1968). On a Class of Aligned Rank Order Tests in Two-Way Layouts. *Annals of Mathematical Statistics.* **39**, p. 1115–1124.
14. THOMPSON, G. L. (1991). A Unified Approach to Rank Tests for Multivariate Repeated Measures Designs. *Journal of the American Stat. Assoc.* **86**, p. 410–419.

g n -way analysis
it for the test of

e, they are robust
ave many of the
The methodology
alysis of variance
e procedures that
s. Moreover, the
ts as the ordinary
with interaction.

vo-Factor Designs.

edure: A Study of
Assoc. **86**, p. 457–

of the Rank Trans-
Simulation **16** (4),

odels. I. Exposition

dures Using Ranks
Stat. **A5**, p. 1349–

d Rank Transform
tics in Agriculture.

TwoWay Classifi-
at. **2**, p. 227–239.

a Bridge Between
i, p. 131–132.

of Asymptotically
ks. *Journal of the*

. Library Interface

Based On Ranks.