The familiar F tests associated with tests of main effects in the analysis of complete block designs rely on a linear model with a normal error term, and are based on calculations that adjust for the presence of main effects. When the necessary assumptions are satisfied, the parametric F test is the preferred procedure to use. However, the necessary assumptions frequently cannot be satisfied, and analysts turn to methods employing ranks. In this paper recent literature will be reviewed and examples presented, to provide guidance on the selection of proper procedures for the analysis of complete blocks. The material presented in this paper is excerpted from a new text in preparation by Hora and Iman (1988a) that presents rank based methods for performing statistical analyses.

The concept of matched pairs extends quite naturally to blocks of data that contain three or more related measurements, such as measurements made on the same subject with each measurement made under a different experimental condition or treatment. Thus, the primary factor of interest is the "treatments", while a secondary factor, the "blocks" are homogeneous values grouped together. Some examples are:

1. Animals from a liter are randomly assigned to different feeding regimens. The liters are the blocks and the feeding regimens are the experimental treatments.

2. Five types of ultraviolet inhibitors are tested with each of eight different formulations of house paint. The inhibitors are the treatments and the types of paint are the blocks.

3. Each of ten leadership traits are ranked in order of importance from 1 (most important) to 10 (least important) by each of seven managers. Inferences are to be made about leadership traits. Each manager's evaluations forms a block.

4. Four football handicappers make point spread predictions on professional football games during the season. The absolute difference between the predicted spread and the actual spread is used to measure the handicappers' accuracy. The handicappers are the treatments and each game is a block.

5. Four shapes of liquid dishwashing soap bottles are each evaluated for ease of use on a scale from 1 to 5 by each member of a panel of eleven consumers. The shapes of bottles are treatments and each consumer is a block.

The experimental design that leads to the blocks and treatments data layout is usually called a block design. A commonly used block design is the randomized complete block (RCB) design. Randomized refers to the random assignment of experimental units within a block to the various treatments. This is done to avoid introducing bias in the assignment of units to treatments. The term complete, in the phrase complete blocks, refers to assignment of the experimental units in each block to each of the treatments. Each treatment and block combination is represented in a complete block design.

Each of the five examples presented above represents a complete design. Example 1 is a classic example of a randomized complete block experiment as members of the liter are randomly assigned to each feeding regimen. Example 2 may also be a randomized complete block design if, for instance, one gallon of each formulation of paint is randomly assigned to each of the ultraviolet inhibitors. However, in example 3 the block is formed by having one individual rank the ten traits with respect to one another, so the treatments are not randomized within the block. In example 4, the treatments again are not randomized, but unlike example 3 where one individual ranked each treatment, each handicapper makes his or her predictions independent of the other handicappers. Example 5 is similar to example 3 in that one individual provides the observational unit for each treatment, but unlike example 3 where the treatments were compared against each other to obtain the rankings from 1 to 10, the bottles are each evaluated on a scale from 1 to 5, so two bottles could easily have the same value assigned.

The previous examples are not meant to exhaust all possible situations that may appear under the guise of a complete block design, but rather serve to illustrate that complete block designs can appear in many forms. Some designs have the observational units randomly
assigned to the treatments and some do
not. The terminology, complete blocks,
may be used when there is at least one
observational unit assigned to each
treatment within each block. If each
treatment-block combination has the same
number of observational units the design
is said to be proportional. Frequently,
balanced designs have only one observa­
tional unit for each treatment-block
combination. Many times the same physi­
ical unit appears under each of the
treatments in one block such as in
examples 3 to 5. For instance, in
example 4 each of four handicappers
(treatments) handicap the same game
(physical unit) within each block.

STATISTICAL MODELS FOR COMPLETE BLOCKS

EXPERIMENTS

Model I. The model most frequently
associated with the analysis of complete
blocks is the additive, or linear, model
in which the block effects are repre­
sented by shifts in the mean response
within each block. This model contains
two sets of shift terms, one set to
account for the treatment differences
and one set to account for the block
differences. For k levels of the ex­
perimental treatments and b blocks, we
have

\[ Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \]  

where \( \mu \) is an overall mean, \( \alpha_i \) is the
shift due to block effects and \( \beta_j \) is the
shift due to treatment effects. The
error term \( \epsilon_{ij} \) captures random fluctua­
tions in the jth observation in the ith
block that has no systematic behavior.
The error terms are assumed to be inde­
pendently and identically distributed
and have an expected value of zero.

Since Model I contains both an over­
all mean \( \mu \) and two types of differential
effects, the \( \alpha_i \)'s and the \( \beta_j \)'s, it is
possible to add a fixed amount to \( \mu \) and
subtract the same amount from either all
of the \( \alpha_i \)'s or all of the \( \beta_j \)'s without
changing the mean of any \( Y_{ij} \). This
indeterminacy can be removed by adding
the constraint that the sum of the block
differential effects is zero and the sum
of the treatment differential effects is
zero. With the added constraint for the
\( \alpha_i \)'s

\[ (1/b) \sum_{i=1}^{b} E(Y_{ij}) = \mu + \alpha_i \]  

(3)

The constraints in (2) and (3) together
imply

\[ (1/k) \sum_{j=1}^{k} \sum_{i=1}^{b} E(Y_{ij}) = \mu \]  

(4)

so that the two type of differential
effects can be easily isolated by
subtraction of sums such as subtracting
the left side of (4) from both (2) and
(3).

An important property of the dif­
terential effects model is that the
magnitude of any \( \alpha_i \) or \( \beta_j \) is a function
of the other levels of the factor or
blocks chosen to be included in the
experiment. These quantities are pro­
perly interpreted as relative values
rather than absolute values. For
example, if the average m.p.g. for Fords
is compared to the average m.p.g. for
Hondas the differential effect for Fords
may be negative, while if Fords are
compared with Cadillacs the differential
effect for Fords may be positive.

Model II. A difficulty with Model I
is that the block effects must be of a
very specific type, that is linear
shifts. A more general formulation of
the problem can be obtained through the
distribution functions of the observa­
tions. Denoting the distribution func­
tion of \( Y_{ij} \) by \( F_{ij}(y) \), we have for \( i = 
1, \ldots, b \) and \( j = 1, \ldots, k \)

\[ F_{ij}(y) = F_{i}(y - \beta_j) = P(Y_{ij} \leq y) \]  

(5)

Model II requires that the distribu­
tions within a block be shift transla­
tions of each other, but between block
distributions are arbitrary. In terms
of Model I, this means a relaxation of
the i.i.d. requirement on the errors.
Therefore, Model II may be viewed as a
generalization of Model I, in that
whenever Model I holds, Model II holds.

HYPOTHESES FOR COMPLETE BLOCKS

The primary hypothesis that is
examined in the complete blocks design
is the absence of treatment effects.
For Example 5 the hypothesis implies
that there are no systematic preferences
among consumers for any particular
bottle design. That is, all bottles, on
the average, are equally preferred.
This hypothesis is expressed as

\[ H_0: \beta_1 = \beta_2 = \ldots = \beta_k. \]  

(6)

The alternative hypothesis is that

\[ \beta_1 \neq \beta_2 \neq \ldots \neq \beta_k. \]
HI: \( \beta_j = \beta_{j'} \) for at least one pair \( j, j' \) whenever the alternative is true, at least two of the treatment levels affect the response differently.

In addition to the primary hypothesis, it is sometimes, but not always, interesting to determine whether systematic differences among the blocks exist. In Example 4 given earlier, one might also wish to determine whether the paint types differ. The secondary null hypothesis is then

\[
H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_b.
\]

The alternative to the secondary null hypothesis is that

\[
H_1: \alpha_i' = \alpha_{i'}
\]

for at least one pair \( i, i' \).

The alternative hypothesis for both types of differential effects, treatment and block effects, are nondirectional in that there is no ordering of the differential effects implied by the alternative hypothesis.

### THE PARAMETRIC PROCEDURE (THE F TEST)

The parametric, or normal theory, approach to the analysis of complete block designs is the familiar F test and relies on computing F-ratio statistics under the assumptions of Model I and the added assumption of normal error terms.

### FRIEDMAN'S TEST

The Friedman procedure tests for treatment effects after first removing block effects through the utilization of a within block ranking scheme. Since there are \( k \) treatment levels, there are \( k \) values in each block and these values will be ranked from 1 to \( k \). Denoting the rank of the \( j \)th value in the \( i \)th block by \( R_{ij} \), a common form of Friedman’s statistic when there are no ties is given by

\[
T_F = \frac{12}{bk(k+1)} \sum_{j=1}^{k} \sum_{i=1}^{b} (R_{ij} - \bar{b}(k+1)/2)^2
\]

The statistic \( T_F \) has a null distribution that depends upon the number of treatment levels and the number of blocks, but not on the within block distributions that are given in Model II. It is only necessary to assume that the within block distributions are continuous so that ties among data values within a block are precluded. It is not even necessary to assume that the random variables within a block are independent. It is only necessary to assume that there is independence between blocks and random variables within each block possess a property called exchangeability.

### Exchangeability:

A set of \( k \) random variables are exchangeable if

\[
P(X_1 < x_1, \ldots, X_k < x_k) = P(X_1 < x_1, \ldots, X_j < x_j)
\]

for all possible reorderings of the \( k \) random variables within the probability expression and all possible values of \( x_1, \ldots, x_k \).

Exchangeability is a symmetry condition that often plays a role in rank statistics similar to the role played by the i.i.d. assumption in Model I for normal theory statistics. The i.i.d. condition is a special case of the exchangeability condition, but exchangeability is a much weaker assumption.

### THE RANK TRANSFORMATION VERSION OF FRIEDMAN'S STATISTIC

Conover and Iman (1981) show that the usual F statistic for the complete blocks design calculated on within block ranks produces a test comparable to Friedman’s test but with the added advantage of a built in correction for ties. This is an important consideration because whenever ties are present, the tables of the exact distribution for Friedman’s statistic are inappropriate. The monotone relation between \( T_F \) and \( F_F \) given in Conover and Iman can be written as

\[
F_F = (b-1)T_F/[b(k-1) - T_F].
\]

The statistic \( F_F \) is compared against quantiles of the \( F \) distribution with \( k-1 \) and \( (b-1)(k-1) \) degrees of freedom. In the case of identical rankings within each of the blocks the statistic \( F_F \) is undefined as \( b(k+1) - T = 0 \) in (11). In this case the p-value associated with the rejection of the null hypothesis is \((1/k!)^{b-1}\). Further, it is shown in Iman and Davenport (1980) that the F statistic more closely approximates the null distribution of Friedman’s statistic than does the usual chi-squared approximation. It should be noted that for \( b=2 \) the Friedman test is equivalent to Spearman’s rho, and for \( k=2 \) it is equivalent to the sign test, in which case it is usually better to use the Wilcoxon signed ranks test.

### THE RANK TRANSFORMATION METHOD USING AN OVERALL RANKING

Next, we consider the rank transformation procedure that is based upon an overall ranking (RI-1) of the data,
rather than a ranking within blocks. The motivation for using an overall ranking is that no interblock information is lost such as occurs in tests based on within block ranks.

Application of the RT-1 procedure requires the data to be replaced with an overall ranking. Thus, the ranks will be from 1 to kb. As suggested in Conover and Iman (1981), the usual normal theory F statistic is calculated on the ranks and compared to the F distribution with k-1 and (k-1)(b-1) degrees of freedom. The resulting p-value is used to interpret the test outcome in the standard manner.

The rank transformation test based upon an overall ranking of the data can be applied without making any assumptions more stringent than those required by Friedman's test (see Hora and Conover, 1984 and Iman, Hora and Conover, 1984).

The RT-1 procedure is demonstrated on a set of data given by Berry (1987) to show the impact of outliers. The data consist of plasma digoxin levels measured on six different days for each of 15 patients after they were administered digoxin. These data are given in Figure 1. As Berry points out, these data contain two obvious outliers of 143 and 121. The F-test on the original data for Days gives \( F_t = 2.32 \) with a p-value of .0523, while the F-test on Subjects gives \( F_b = 1.62 \) with a p-value of .0957. Hence, at the .05 level neither of these effects would be considered significant.

The AOV for the RT-1 analysis of these data gives \( F_t = 5.19 \) and \( F_b = 2.72 \) with respective p-values of .0004 and .0030 indicating that both main effects are significant. Berry presents an analysis of the original data where the two outliers are replaced by substituting estimates for the outliers. The analysis on the adjusted data yields p-values of .0009 and .0019 while the corresponding RT-1 analysis is basically unchanged with p-values of .00009 and .001. This example demonstrates the robustness of the RT-1 analysis with respect to outliers.

**THE ALIGNED RANKS TEST**

An alternative to Friedman's procedure, the normal theory (linear model) procedure, and the overall ranking procedure is the aligned ranks procedure given in Hodges and Lehmann (1962), Mehra and Sen (1969), and Lehmann (1975). The aligned ranks procedure incorporates some of the features of the linear model approach with an overall ranking.

Implementation of the aligned ranks procedure is accomplished by aligning the observations within the ith block by subtracting from each observation within the ith block a centrality estimate such as the sample mean or sample median of the observations in the ith block. The purpose of the alignment is to remove block effects. The aligned, or centered, observations are replaced by overall ranks and the usual linear model-normal theory statistic is calculated on the ranks of the aligned observations.

Given a set of observations \( Y_{ij} \) such that \( Y_{i1}, \ldots, Y_{ik} \) are within the same block, but measured under each of \( k \) different experimental treatments, the aligned observations are given by \( W_{ij} = Y_{ij} - \bar{Y}_i \), where \( \bar{Y}_i \) is the centering statistic. There are many possible choices for the centering statistic, however the block mean is the most commonly used.

The alignment, or centering, of the observations within a block can be accomplished using any symmetric function of the observations within a block such that \( W_{ij} \) is invariant to a change in location of the block observations. By symmetric, we mean that the statistic, \( \bar{Y}_i \), remains unchanged under permutations of the subscripts \( j \) among the \( i = 1, \ldots, i_k \). More simply, the statistic is the same regardless of the order of the observations within a block. The aligned observations, \( W_{ij} \), are invariant
to a change in location of the block’s observations if adding a quantity, say c, to every observation in the block leaves $W_{ij}$ unchanged.

It is simple to show that both the block mean and median possess the desired characteristics and can be used as alignment statistics. Other robust measures of location such as Winsorized and trimmed means can also be used. We will confine our calculations and comparisons to aligned rank statistics based upon the mean, exclusively.

**CALCULATING THE ALIGNED RANKS STATISTIC FOR COMPLETE BLOCKS**

Let $Y_{ij}$ be the observation in the $i$th block under the $j$th treatment. Then

$$T_a = b \sum_{j=1}^{k} (R_{.j} - \bar{R})^2 / s_a^2$$

(12)

where

$$R_{.j} = \sum_{i=1}^{b} R(W_{ij}) / b,$$

$$\bar{R} = \sum_{j=1}^{k} R_{.j} / k = (bk + 1) / 2,$$

$$W_{ij} = Y_{ij} - Y_{.i},$$

$$s_a^2 = \sum_{i=1}^{b} \sum_{j=1}^{k} [R(W_{ij}) - R_{.i}]^2 / [b(k-1)],$$

and

$$R_{.i} = \sum_{j=1}^{k} R(W_{ij}) / k.$$

The statistic $T_a$ is compared to the quantiles of the chi-squared distribution with $k-1$ degrees of freedom. Notice that the denominator of $T_a$ is calculated using deviations from the average block rank, rather than deviations from a mean rank corrected for both block and treatment effects.

The null distribution of the aligned ranks test statistic depends upon the underlying distribution of the data. Because of this dependence, the aligned ranks statistic is not distribution-free under the null hypothesis. A procedure can be developed that produces a conditionally distribution free test, the condition being that certain ranks have appeared in certain blocks.

**SUMMARY OF THE STATISTICS USED TO ANALYZE AN ADHESIVES EXPERIMENT**

Four different statistics: the $F$ test, $T_F$ (Friedman’s test), $R$ (RT-I transform of the data), and $T_a$ (the aligned ranks test) have been presented for the analysis of complete block experiments. These four statistics are now demonstrated on a simple set of adhesives data presented in Figure 2.

**ADHESIVE**

<table>
<thead>
<tr>
<th>SURFACE</th>
<th>ADHESIVE A</th>
<th>ADHESIVE B</th>
<th>ADHESIVE C</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>410</td>
<td>565</td>
<td>512</td>
</tr>
<tr>
<td>II</td>
<td>389</td>
<td>512</td>
<td>530</td>
</tr>
<tr>
<td>III</td>
<td>345</td>
<td>431</td>
<td>411</td>
</tr>
</tbody>
</table>

Figure 2. Breaking Strengths of Three Adhesives Applied to Three Surfaces

Additionally, the chi-squared and $F$ distributions have been used to approximate the exact distribution of Friedman’s test, and the chi-squared distribution serves as approximation to the exact distribution of the aligned ranks test (for which tables of the exact distribution do not exist). The summary of the p-values for these various procedures for testing for differences in the adhesives is as follows.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Test</td>
<td>.0083</td>
</tr>
<tr>
<td>Friedman’s Test</td>
<td>.1944</td>
</tr>
<tr>
<td>Exact Distribution</td>
<td>.0970</td>
</tr>
<tr>
<td>Chi-squared Approximation</td>
<td>.0494</td>
</tr>
<tr>
<td>F Approximation</td>
<td></td>
</tr>
<tr>
<td>Rank Transformation (RT-1)</td>
<td>.0124</td>
</tr>
<tr>
<td>Aligned Ranks Test</td>
<td>.0833</td>
</tr>
<tr>
<td>Exact Distribution</td>
<td></td>
</tr>
<tr>
<td>Chi-squared Approximation</td>
<td>.1045</td>
</tr>
</tbody>
</table>

With so many different p-values possible, which one(s) should be reported by the analyst? First of all, the estimated p-values obtained from approximations based on the chi-squared and $F$ distributions will both approach the p-value of the exact distribution as $b$ and $k$ increase. The problem here is that $b$ and $k$ are both very small, making it very difficult to approximate the discrete exact distribution with a continuous distribution. The Friedman test and the aligned ranks test are both designed to remove block effects and their efficiencies, along with that of the rank transformation test are considered later in the paper. The next example illustrates how the three rank tests behave in the presence of very strong block effects.
THE BEHAVIOR OF THE STATISTICS $T_F$, $F_R$, AND $T_A$
IN THE PRESENCE OF VERY STRONG BLOCK EFFECTS

The Friedman and aligned rank tests are both designed to remove block effects before testing for treatment effects (or by a second analysis removing treatment effects before testing for block effects). The Friedman test removes block effects by assigning the ranks 1 to $k$ within each block, thus making the mean rank in each block equal to $(k+1)/2$. The aligned rank tests makes each block mean zero by subtracting the block mean from each observation within the block. On the other hand, the statistic $F_R$ is calculated on $R_T - 1$ ranks which does not, in and of itself, remove block effects. The next example illustrates how very strong block effects can impact an analysis.

A set of data presented by Damaser, Shor and Orne (1963) and analyzed in Lehmann (1975) consists of measurements of skin potential (adjusted for the initial level) in millivolts. The measurements were made on each of eight patients under hypnosis when emotions of fear, happiness, depression, and calmness were requested (in a random order) from each of eight patients. The data are presented in Figure 3 along with their Friedman ranks.

<table>
<thead>
<tr>
<th>Sub.</th>
<th>Fear</th>
<th>Happy</th>
<th>Depress</th>
<th>Calm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.1(4)</td>
<td>22.7(3)</td>
<td>22.5(1)</td>
<td>22.6(2)</td>
</tr>
<tr>
<td>2</td>
<td>57.6(4)</td>
<td>53.2(2)</td>
<td>53.7(3)</td>
<td>53.7(1)</td>
</tr>
<tr>
<td>3</td>
<td>10.5(3)</td>
<td>9.7(2)</td>
<td>10.8(4)</td>
<td>8.3(1)</td>
</tr>
<tr>
<td>4</td>
<td>23.6(4)</td>
<td>19.6(1)</td>
<td>21.1(2)</td>
<td>21.6(3)</td>
</tr>
<tr>
<td>5</td>
<td>11.9(1)</td>
<td>13.8(4)</td>
<td>13.7(3)</td>
<td>13.3(2)</td>
</tr>
<tr>
<td>6</td>
<td>54.6(4)</td>
<td>47.1(3)</td>
<td>39.2(2)</td>
<td>37.0(1)</td>
</tr>
<tr>
<td>7</td>
<td>21.0(4)</td>
<td>13.6(1)</td>
<td>13.7(2)</td>
<td>14.8(3)</td>
</tr>
<tr>
<td>8</td>
<td>20.3(3)</td>
<td>23.6(4)</td>
<td>16.3(2)</td>
<td>14.8(1)</td>
</tr>
</tbody>
</table>

Figure 3. Skin Potential (adjusted for initial level) in millivolts from each of Eight Subjects During Hypnosis (within block ranks in parentheses)

An analysis of variance for main effects on the original data gives a $p$-value of .0345 for treatments and an $F$ statistic for blocks of 102.86 with 7 and 21 degrees of freedom. The $p$-value associated with the $F$ statistic is on the order of $10^{-14}$. Hence, the block effect is quite strong for these data. However, the assumption of homogeneity of variance within blocks required by the normal theory $F$ test is violated (Bartlett’s test for homogeneity of variance gives a $p$-value of .0001), so the results of the analysis on the original data are meaningless since the computed $F$ statistics do not have an $F$ distribution.

The Friedman test on Emotions gives a value of $T_F = 6.450$ with 3 degrees of freedom with a $p$-value of .0917 (the exact permutation $p$-value is .0911), indicating no treatment effect. The corresponding Friedman test on block effects gives $T_F = 26.09$ with 7 degrees of freedom and a $p$-value of .0005, indicating a very strong block effect.

The block aligned observations are shown in Figure 5 with their overall (aligned) ranks given in Figure 6. The statistic $T_A = 22558.5/2644.75 = 8.53$ is compared with the chi-squared distribution with 3 degrees of freedom which gives a $p$-value of .0362, indicating a significant treatment effect.

Figure 4. Skin Potential Measurements Within Each Block Versus Emotion

One way of observing the strength of the block effect that does not rely on assumptions associated with hypothesis testing is simply by plotting the magnitude of the observations within each block versus the emotion as shown in Figure 4. This figure shows the observations made on subjects 2 and 6 dominate all others with subject 3 below all others. Note also the low variability of the measurements on subject 1 and the high variability on subject 6.

An analysis of variance for the RT-1 analysis gives a $p$-value for treatments of .1639, indicating no treatment effects, in agreement with the Friedman test. The $F$ ratio for block effects is 40.18 with 7 and 21 degrees of freedom and a $p$-value on the order of $10^{-10}$. The block effects are more strongly identified than with the Friedman test.

The analysis of variance for the RT-1 analysis gives a $p$-value for treatments of .1639, indicating no treatment effects, in agreement with the Friedman test. The $F$ ratio for block effects is 40.18 with 7 and 21 degrees of freedom and a $p$-value on the order of $10^{-10}$. The block effects are more strongly identified than with the Friedman test.

The block aligned observations are shown in Figure 5 with their overall (aligned) ranks given in Figure 6. The statistic $T_A = 22558.5/2644.75 = 8.53$ is compared with the chi-squared distribution with 3 degrees of freedom which gives a $p$-value of .0362, indicating a significant treatment effect.
At this point the reader may think that the normal theory F test adjusted out the block and treatment effects and has already established that the treatments are significantly different. The problem with this line of reasoning has already been covered — namely if the assumptions for the normal theory F test are violated, the F statistics do not follow an F distribution. Thus, the conclusions may be right or may be wrong, there is no way of knowing for sure without some additional analyses. Next we consider the robustness and asymptotic relative efficiencies of the various procedures presented in this paper for analyzing complete blocks.

ROBUSTNESS OF METHODS FOR COMPLETE BLOCKS

Robustness under the null hypothesis refers to the ability of a test statistic to retain its null distribution regardless of the parent distributions of the data. Of the procedures we have considered here (Friedman’s test, the rank transformation test, the aligned ranks test, and the normal theory F test) Friedman’s test possesses the greatest robustness under the null hypothesis. The only assumptions that are required are that Model II of (5) holds and that the data are continuous so that no observations within a block will be tied. When these assumptions are met, the exact distribution of Friedman’s statistic under the null hypothesis can be tabulated without reference to the underlying distribution of the data. Thus, Friedman’s test is distribution free.

In practice, however, tables of the exact distribution are available only for a limited selection of block and treatment numbers. Thus, more often than not, asymptotic approximations are used rather than exact tables. This means that under most circumstances the test procedure is approximate rather than exact.

The other two rank procedures, the overall ranking procedure and the aligned ranks procedure are only asymptotically distribution free. That is, the small sample distribution of the test statistic under the null hypothesis and Model II of (5) depends upon the underlying parent distribution of the data, and thus the distribution can only be approximated. Fortunately, the approximation for the overall ranking and aligned ranks statistics are very good even when sample sizes are quite small.

The null distribution of the aligned ranks statistic is somewhat more susceptible to distribution assumptions since it is a quasi-mean statistic in that the block means are used to adjust the data prior to ranking.

ASYMPTOTIC RELATIVE EFFICIENCIES OF THE METHODS FOR COMPLETE BLOCKS

The second type of robustness examined is power robustness — how well a procedure maintains its power characteristics across distributions. The discussion is presented in the form of the asymptotic relative efficiencies (AREs) of the various procedures.

The AREs of the rank transformation procedure relative to the normal theory test, the aligned ranks test, and Friedman’s test have been evaluated by Hora and Iman (1988b). Several within block densities and several methods of selecting block location and scale changes across densities were examined in that study. The within block densities employed include normal, Student, and double exponential densities as given in the following six cases.

1. Normal Densities Within Blocks with Normal Block Shifts.
2. Normal Densities Within Blocks with Uniform Block Shifts.
3. Normal Densities Within Blocks with Block Scale Changes.
4. Double Exponential Densities Within Blocks with Normal Block Shifts.
5. Double Exponential Densities within Blocks and Uniform Block Shifts.

6. Student Densities Within Blocks with Normal Block Shifts.

CONCLUSIONS

The choice of a test procedure for block data is not clear cut. Instead, the performance of the candidate procedures vary in a complex manner as the properties of the data change. Under some circumstances the investigator would be advised to use more than one procedure and then compare the results. If they are confirmatory, then no further action is needed. But, if the results are conflicting, then the data should be examined to diagnose the reason for the conflict (heavy tails, extreme block effects, outliers, small number of treatments, etc.)

In summary, we make some statements about each of the tests presented in this paper.

Normal Theory F test

The normal theory F test performs well whenever the data are normal or the errors have a short-tailed distribution. The gain in efficiency over the rank tests is not large, however, under these circumstances.

Friedman's test

The most noticeable deficiency of Friedman's test is the loss of power when the number of treatments is small. It is usually a poor choice when there are only three or four treatment levels.

Rank transformation procedure

The rank transformation procedure is the only procedure with properties affected by the magnitude of block shifts. When large block shifts are present, the power of this procedure is reduced. Otherwise, it performs well in almost all situations.

Aligned ranks procedure

The aligned ranks procedure is a quasi-mean procedure and is somewhat susceptible to loss of power when heavy-tailed distributions are introduced. Thus, the aligned ranks procedure may not be a good choice in the presence of outliers.

How, then, should one proceed to analyze a complete blocks data set? We suggest that a preliminary assessment be made with regard to the following factors:

1. Examine the distribution of deviations with particular regard to the heaviness of tails. The Wilks-Shapiro test and the Lilliefors plot for normality provide good tools for detecting this situation.

2. Look for the presence or absence of one or several extreme observations.

3. Examine the magnitude of block effects. This can be accomplished by a priori by considering the physical process that has generated the data, by plots such as in Figure 4, or by examining the differences among block means standardized by the square root of the usual estimate of the error variance. When the standard values have absolute values that exceed one unit, then the rank transformation test may lose power.

4. Consider the number of treatment effects. If it is small, Friedman's test should not be considered as a primary test.

From these four steps, one should gain insight into those tests that will perform well. If nothing is found to disqualify tests, then it may be that any of the four tests will work well. Whenever possible, it is best that at least two of the tests be performed.
REFERENCES


