COMMUN. STATIST.-THEOR. METH., A5(14), 1349-1368 (1976)

ON SOME ALTERNATIVE PROCEDURES USING RANKS FOR THE ANALYSIS OF EXPERIMENTAL DESIGNS

W. J. Conover

Texas Tech University

R. L. Iman

Sandia Laboratories

Key Words & Phrases: rank transforms; aligned ranks; analysis of variance on ranks; robustness; power.

ABSTRACT

The analysis of data from experimental designs is often hampered by the lack of more than one procedure available for the analysis, especially when that procedure is based on assumptions which do not apply in the situation at hand. In this paper two classes of alternative procedures are discussed and compared. One is the aligned ranks procedure which first standardizes the data by subtracting an appropriate estimate of location, then replaces the data with ranks, and finally uses an appropriate test statistic which has asymptotically a chi-square distribution. The second procedure is the rank transform which first replaces all of the data with the ranks, and then employs the usual parametric methods, but computed on the ranks instead of the data. Some Monte Carlo simulations for a test of interaction in a two way layout with replication enable the robustness and power of

these two methods to be compared with the usual analysis of variance.

1. INTRODUCTION

Objective scientific research relies heavily upon the controlled experiment, and interpretation of the results from controlled experiments often depends on the availability of appropriate statistical methods. Some of the simpler experimental designs may be analyzed by any one of several appropriate methods, both parametric and nonparametric. Some of the more complicated designs may be analyzed only using parametric methods, while other designs have only approximate statistical methods available. Most statistical methods are less than perfect for one reason or another. Nonparametric methods suffer from lack of availability for most of the experimental designs in common usage, and where they are available they appear quite often to be custom made for the situation. Parametric methods on the other hand are available for a wide variety of designs, but their validity depends on a set of assumptions which are seldom met and usually ignored.

Considerable research has been devoted to showing that various parametric procedures are robust (in terms of Type I error) under different conditions. However, it is not as widely realized that robustness in the above sense is only half of the problem and that violation of the assumptions can also result in a loss of power. In these circumstances nonparametric procedures, when available, sometimes have considerably more power and therefore should be used.

Another widespread misconception is that all nonparametric procedures are, ipso facto, robust. Virtually all nonparametric procedures for experimental designs employ asymptotic distributions instead of exact distributions for the test statistics, and virtually all of them make some assumptions about the model which may or may not be satisfied. Therefore both robustness and power should be considered when comparing parametric and nonparametric procedures. The dangerous myth, that nonparametric procedures are always more robust and parametric procedures are always more powerful, has blinded the minds of otherwise reasonable men, and resulted in inferior methods being used to analyze experimental results.

There is a well-established need for alternative procedures which can be used when the parametric assumptions are not met. For those designs that have no alternative procedures presently available, a reasonable approach is to examine the many alternative procedures available for some simple designs, choose those that have especially good properties, and attempt to extend them to include other designs. In this paper we examine two known statistical methods that were developed using that approach, and we compare them with the usual analysis of variance on the basis of robustness and power.

2. SOME NOTATION AND MATHEMATICAL BACKGROUND

Let I_k represent the k × k identity matrix, J_k represent the k × k matrix of ones, and ℓ_k represent the k × 1 vector of ones. We use the direct product notation × defined by

$$A \times B = \begin{bmatrix} a_{11}B & a_{12}B \dots a_{1k}B \\ \dots & \dots \dots \dots \\ a_{r1}B & a_{r2}B \dots a_{rk}B \end{bmatrix}$$

and state the following well-known results (Graybill, 1969).

(i) If $\underline{a}$ is a scalar, then $(aA) \times B = A \times (aB) = a(A \times B)$.

(ii) $(A \times B) \times C = A \times (B \times C)$

(iii) $(A \times B)' = A' \times B'$

(iv) $(A \times B)(F \times G) = (AF) \times (BG)$ if AF and BG are defined.

(v) $(A^{m\times m}+B^{m\times m}) \times C^{n\times n} = (A \times C) + (B \times C)$

(vi) If rank $(A) = r_1$ and rank $(B) = r_2$, then rank $(A \times B) = r_1 r_2$.

(vii) A and B positive (semi)definite $\rightarrow$ $A \times B$ is positive (semi) definite.

Since $(I_a - \frac{1}{a}J_a)$ is idempotent with rank (a-1), it is easy to show, using the above results, that $(I_a - \frac{1}{a} J_a) \times (I_b - \frac{1}{b} J_b)$ is idempotent, with rank (a-1)(b-1).

Let $\underset{\sim}{Y}$ be an $(abn \times 1)$ vector whose elements are Y_{ijk}, $k=1,\ldots,n$; $j=1,\ldots,b$; and $i=1,\ldots,a$; in that order. That is,

$$\underset{\sim}{Y}' = (Y_{111}, Y_{112}, \ldots, Y_{11n}, Y_{121}, \ldots, Y_{abn}).$$

Let $\underset{\sim}{R}(\underset{\sim}{Y})$ be the vector whose elements are the ranks, from 1 to abn, of the elements of $\underset{\sim}{Y}$.

If $\underset{\sim}{Y}$ represents a multivariate random vector with mean vector $\underset{\sim}{\mu}$ and covariance matrix Σ, then $\underset{\sim}{C}\,\underset{\sim}{Y}$ has mean $\underset{\sim}{C}\,\underset{\sim}{\mu}$ and covariance matrix $\underset{\sim}{C}\,\underset{\sim}{\Sigma}\,\underset{\sim}{C}'$, where $\underset{\sim}{C}$ is any $(r \times abn)$ matrix of constants. If $\underset{\sim}{Y}$ is multivariate normal $(\underset{\sim}{\mu}, \underset{\sim}{I})$, then the positive semi definite quadratic forms $\underset{\sim}{Y}'\underset{\sim}{A}\,\underset{\sim}{Y}$ and $\underset{\sim}{Y}'\underset{\sim}{B}\,\underset{\sim}{Y}$ are independent if and only if $\underset{\sim}{A}\,\underset{\sim}{B} = 0$. If $\underset{\sim}{Y}$ is multivariate normal $(\underset{\sim}{\mu}, \underset{\sim}{V})$, then the quadratic form $\underset{\sim}{Y}'\underset{\sim}{A}\,\underset{\sim}{Y}$ has the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2}\underset{\sim}{\mu}'\underset{\sim}{A}\,\underset{\sim}{\mu}$ if and only if $\underset{\sim}{A}\,\underset{\sim}{V}$ is idempotent of rank k.

3. ANALYSIS OF VARIANCE

The name analysis of variance covers a long list of statistical methods, which virtually always incorporate the assumptions of a linear model, and independent identically normally distributed error terms. With these two assumptions, and occasionally additional ones, statistical methods have been developed for many experimental designs, and most of these methods have good properties such as being likelihood ratio tests. Many of these methods are asymptotically nonparametric in the sense that as the sample size gets large the central limit theorem assures that the distribution of the test statistic approaches some prescribed distribution, even though the error terms might be nonnormal. The methods which are asymptotically nonparametric tend to be the same ones which prove to be robust against nonnormality.

For purposes of comparing the various methods, we have selected as our vehicle of comparison the test for interaction in the two way layout with replication. The linear model assumed in the analysis of variance is

$$Y_{ijk} = \mu_{ijk} + E_{ijk} \qquad i=1,2,\ldots,a$$
$$\qquad j=1,2,\ldots,b \quad (3.1)$$
$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \rho_k, \qquad k=1,2,\ldots,n$$

where α_i is the effect of the i^{th} block, β_j is the effect of the j^{th} treatment, γ_{ij} is the block-treatment interaction, and where ρ_k is the effect of the k^{th} replication. The effects, summed over any subscript, total zero, and the E_{ijk}'s are independent, normal $N(0,\sigma^2)$ random variables. The null hypothesis of no interaction is represented by H_0: $\gamma_{ij} = 0$ for all (i,j).

Consider the matrix $\underset{\sim}{N} = \underset{\sim}{M} \times \underset{\sim}{Q}$, where $\underset{\sim}{M} = (I_a - \frac{1}{a} J_a) \times (I_b - \frac{1}{b} J_b)$, and $\underset{\sim}{Q} = \frac{1}{n} \ell_n'$. Note that $\underset{\sim}{M}$ is idempotent of rank $(a-1)(b-1)$ and $\underset{\sim}{Q}\,\underset{\sim}{Q}' = \frac{1}{n}$. The vector $\underset{\sim}{Y}/\sigma$ has the multivariate normal distribution with mean $\underset{\sim}{\mu} = \{\mu_{ijk}/\sigma\}$ and covariance matrix I. Therefore the quadratic form

$$n(\underset{\sim}{N}\,\underset{\sim}{Y})'(\underset{\sim}{N}\,\underset{\sim}{Y})/\sigma^2 = \frac{n}{\sigma^2} \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij.}-Y_{i..}-Y_{.j.}+Y_{...})^2, \quad (3.2)$$

where the dot subscript represents averages, has a chi-square distribution with noncentrality parameter $\lambda = n\underset{\sim}{\mu}'(\underset{\sim}{N}'\underset{\sim}{N})\underset{\sim}{\mu}/2$ and $(a-1)(b-1)$ degrees of freedom if and only if $n\,\underset{\sim}{N}'\underset{\sim}{N}$ is idempotent of rank $(a-1)(b-1)$. But this latter condition is easily shown to be true from the above considerations. Further, λ simplifies to $\lambda = n \sum_{i=1}^{a} \sum_{j=1}^{b} \gamma_{ij}^2/2\sigma^2$ because $\sigma\underset{\sim}{N}\underset{\sim}{\mu} = \{(\mu_{ij.}-\mu_{i..}-\mu_{.j.}+\mu_{...})\} = \{(\gamma_{ij})\}$. The noncentrality parameter equals zero if and only if the null hypothesis is true.

Consider next the matrix $\underset{\sim}{E} = (I_{ab} - \frac{1}{ab} J_{ab}) \times (I_n - \frac{1}{n} J_n)$. Then $\underset{\sim}{E}$ is idempotent of rank $(ab-1)(n-1)$ and the quadratic form

$$(\underset{\sim}{E}\,\underset{\sim}{Y})'(\underset{\sim}{E}\,\underset{\sim}{Y})/\sigma^2 = \frac{1}{\sigma^2} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk}-Y_{ij.}-Y_{..k}+Y_{...})^2 \quad (3.3)$$

has the central chi-square distribution with $(ab-1)(n-1)$ degrees of freedom under the above assumptions on $\mathbf{Y}$. Furthermore, $\mathbf{N}\,\mathbf{E} = 0$ so the two quadratic forms are independent, and the statistic

$$F = \frac{n(\mathbf{N}\,\mathbf{Y})'(\mathbf{N}\,\mathbf{Y})/(a-1)(b-1)}{(\mathbf{E}\,\mathbf{Y})'(\mathbf{E}\,\mathbf{Y})/(ab-1)(n-1)} \tag{3.4}$$

has the F distribution, $(a-1)(b-1)$ and $(ab-1)(n-1)$ degrees of freedom, central F under the null hypothesis and noncentral F under the alternative.

If we relax the assumption of normality on the E_{ijk}, but maintain the assumption of independence, mean zero, variance σ^2, and the linear model (3.1), then the central limit theorem may be applied to $Y_{ij.}$ for "large" n, so that (3.2) still has, asymptotically at least, a noncentral chi-square distribution with the indicated parameters. Also $(\mathbf{E}\,\mathbf{Y})'(\mathbf{E}\,\mathbf{Y})/(ab-1)(n-1)$ represents a consistent estimator of σ^2, so F has as an asymptotic distribution the F distribution with infinite degrees of freedom in the denominator. In this way, usage of the F-distribution to approximate the distribution of F in (3.4) seems justified.

However, relaxation of other assumptions, such as equal variance of the E_{ijk}, or the linear model (3.1), may lead to serious discrepancies between the actual distribution of F and the usual approximation. Even if the true size of the critical region, obtained from F-tables, is close to the nominal level of significance, the power of the test may be substantially less than that of alternative procedures which may be used. Therefore it becomes worthwhile to investigate some of the alternative procedures.

4. ALIGNED RANKS

In searching for alternative methods to use, it seems reasonable to expand upon a method that works well in simpler models. One such method is the Wilcoxon signed-rank test for use with paired observations, or equivalently, with a two-way layout which has two treatments and several (n) blocks. Although not usually presented in this way, the Wilcoxon signed-rank test may be considered as follows. First "align" the observations within each block by subtracting the block mean. Thus the two residuals in block i become $E_{i1} = Y_{i1}-Y_{i.}$ and $E_{i2} = Y_{i2}-Y_{i.}$, where $Y_{i.} = (Y_{i1}+Y_{i2})/2$. Then rank all of the residuals from smallest, rank 1, to largest, rank 2n. The usual paired t statistic, or some monotonic function of it, can be computed on the ranks and compared with either the exact distribution or an approximation based on the asymptotic distribution. This test has desirable properties under both the null and alternative hypotheses (see for example Iman, 1974b, and Conover, 1971). Hodges and Lehmann (1962) suggested using this method of alignment before ranking for other designs. Mehra and Sen (1969) discuss specifically the aligned ranks procedure as a test for interaction in the two way layout with replication. Their procedure is essentially as follows.

First transform the vector $\underset{\sim}{Y}$ of observations to the vector $\underset{\sim}{Z}$ of residuals within replications using $\underset{\sim}{Z} = \underset{\sim}{S}\cdot\underset{\sim}{Y}$ where

$$\underset{\sim}{S} = (I_a - \frac{1}{a} J_a) \times (I_b - \frac{1}{b} J_b) \times I_n, \tag{4.1}$$

obtaining $\underset{\sim}{Z} = \{Z_{ijk}=Y_{ijk}-Y_{i.k}-Y_{.jk}+Y_{..k}\}$. Rank the elements of $\underset{\sim}{Z}$ from smallest, rank 1, to largest, rank abn, and denote the vector of ranks thus obtained by $\underset{\sim}{R}(\underset{\sim}{Z})$. Consider the quadratic form

$$n(\underset{\sim}{N}\cdot\underset{\sim}{R}(\underset{\sim}{Z}))'(\underset{\sim}{N}\cdot\underset{\sim}{R}(\underset{\sim}{Z})) = n \sum_{i=1}^{a} \sum_{j=1}^{b} (R_{ij.}-R_{i..}-R_{.j.}+R_{...})^2 \tag{4.2}$$

as in (3.2), where $\underset{\sim}{N}$ is the same as before. Assume the linear model (3.1) to hold as before, but make no distributional assumptions on the E_{ijk} except that the vectors $(E_{11k},E_{12k},\ldots,E_{abk})$ are mutually independent, k=1,...,n, and have a distribution function which is symmetric in its arguments. Then the vectors of residuals $(Z_{11k},Z_{12k},\ldots,Z_{abk})$ are mutually independent, k=1, ...,n, and their common distribution function is symmetric in its arguments under the null hypothesis. Thus all permutations of the ranks assigned to replication k are equally likely. Mehra and Sen (1969) consider the conditional distribution of (4.2), given the configuration of ranks in rows, columns, and replications, and show that the statistic

$$T = \frac{n(\underset{\sim}{N}.\underset{\sim}{R}(Z))'(\underset{\sim}{N}.\underset{\sim}{R}(Z))}{(\underset{\sim}{S}.\underset{\sim}{R}(\underset{\sim}{Z}))'(\underset{\sim}{S}.\underset{\sim}{R}(\underset{\sim}{Z}))/(n(a-1)(b-1))} \tag{4.3}$$

where $\underset{\sim}{S}$ is given in (4.1), has asymptotically a chi-square distribution with (a-1)(b-1) degrees of freedom, central chi-square under the null hypothesis and noncentral chi-square under the alternative hypothesis. The denominator of T represents the sum of all possible values of $(R_{ijk}-R_{.jk}-R_{i.k}+R_{..k})^2$, divided by the rank of $\underset{\sim}{S}$.

Actually the aligned ranks procedure may be used with any set of scores satisfying rather general requirements, but we are restricting ourselves to ranks as scores in this paper. The statistic (4.3) may be used as an alternative procedure when the distribution of E_{ijk} is nonnormal, and because of the use of ranks we can expect the robustness and the power of this procedure to compare favorably, more often than not, with that of the F-test when the underlying distribution is decidedly nonnormal. Note however that this procedure employs the assumptions of a linear model (3.1) as does the F-test.

5. THE RANK TRANSFORM

The idea of the rank transform is simple. If there is a parametric method available for analysis of the data, but the assumptions of the parametric method are not appropriate for the data, then one merely replaces the data with their ranks, ranking everything together from smallest to largest. Then the parametric method of analysis is applied to the ranks rather than the original data. The idea of replacing the data with the ranks is to transform the original observations into numbers that more nearly satisfy the assumptions of the parametric model, and at the same time retain all of the ordinal information contained in the original data.

In the example used in this paper, the vector of ranks $\underset{\sim}{R}(\underset{\sim}{Y})$ is used instead of $\underset{\sim}{Y}$ in the computation (3.4) of F, to get

$$F_R = \frac{n(\underset{\sim}{N}\ \underset{\sim}{R}(\underset{\sim}{Y}))'(\underset{\sim}{N}\ \underset{\sim}{R}(\underset{\sim}{Y}))/(a-1)(b-1)}{(\underset{\sim}{E}\ \underset{\sim}{R}(\underset{\sim}{Y}))'(\underset{\sim}{E}\ \underset{\sim}{R}(\underset{\sim}{Y}))/(ab-1)(n-1)} \tag{5.1}$$

where the notation is the same as in Section 3.

The justification for the rank transform is not so simple. Like any transformation, it is intended to transform the data into something more resembling normality. For example, it is well known that the average rank $R_{ij.}$ approaches normality quickly as n grows, under rather general conditions, while the same might not be true of the average of the data $Y_{ij.}$, especially with highly skewed populations or in the presence of "outliers". This gives some legitimacy to the numerator of F_R. The denominator of F_R represents a quadratic form which is not sensitive to the presence of interaction. If the covariance matrix of $\underset{\sim}{R}(\underset{\sim}{Y})$ is approximately a constant times the identity matrix, then the quadratic forms in the numerator and the denominator are approximately independent. (This condition is not necessary, but it is sufficient.) Thus the justification for the rank transform is empirical rather than theoretical. We shall see in the next section how well the empirical evidence supports the use of the rank transform in this situation.

Like the aligned ranks procedure, the rank transform procedure represents an extension of a nonparametric method that has proved worthy in a simpler design. In the one way layout, the rank transform procedure is equivalent to the Kruskal-Wallis test (Iman, Quade and Alexander, 1975) and is therefore distribution free. In more complicated designs the null hypothesis does not assure that all arrangements of ranks are equally likely, so the rank transform procedure is no longer distribution free. It is merely a robust procedure that appears to behave remarkably well in the situations in which it has been examined. To be more specific, in Iman (1974a) the rank transform was compared with the analysis of variance on the untransformed data. The design was a two way layout with interaction, so three F statistics were involved; rows, columns, and interaction. Many situations were examined, involving the presence or absence of various effects, small sample sizes, and normal, exponential, and contaminated normal data. In all cases the null distribution of the test statistic behaved as well with the rank transform as without, and the power of the tests was never much worse, but sometimes much better with the rank transform. This is the only study of the rank transform we are aware of.

Until more evidence has been collected, we recommend the following use of the rank transform. In any experimental design, analyze the data using a standard parametric procedure. Then apply the rank transform and use the same procedure on the ranks. Compare the results of the two analyses. If the results are nearly identical, the rank transform has merely confirmed that the original analysis is likely to be accurate. If the results differ considerably using the two methods, the original data should be examined for possible clues such as outliers, to see why the original observations indicate one conclusion while the ordinal information retained after the rank transformation indicates another. Such examination can provide insight leading to the proper interpretation of the experimental results.

6. A MONTE CARLO COMPARISON OF METHODS

In order to compare the robustness and the power of the three methods, we have selected the lognormal distribution for study, since the lognormal model is used as an approximation in many different applied fields. We have chosen three different situations for robustness studies and two situations for power comparisons.

Specifically, let X_{ijk} be independent normal random variables with means

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \rho_k \qquad \begin{array}{l} i=1,\ldots,4 \\ j=1,\ldots,3 \\ k=1,\ldots,5 \end{array}$$

and unit variances, as in the linear model (3.1). Let $Y_{ijk} = \exp\{X_{ijk}\}$ be the observable lognormal variates. Thus the $\{Y_{ijk}\}$

do not satisfy the assumptions of the linear model (3.1) or equal variances. For each of the five situations mentioned above, 1000 sets of data were generated, and F, T, and F_R were computed for each set. Rather than graph the empirical distribution functions vs. the theoretical (or approximate) distribution functions, which are not the same for all three statistics, cumulative plots are made of the observed critical levels (the smallest α for which H_0 could be rejected) vs. the theoretical critical levels which follow the uniform distribution in each case. In this way comparisons of robustness and power are more readily made.

For the first situation, all μ_{ijk} are set equal to zero to see how the test statistics behave under the null distribution in the absence of any nuisance parameters. As Figure 1 indicates all three statistics follow their hypothesized distributions quite closely. Table I gives some numerical comparisons of the three graphs shown in Figure 1. The column headed by D gives the observed value of the Kolmogorov goodness of fit statistic (the maximum vertical deviation of observed from theoretical curves) while the last column gives the exact two sided critical level associated with that observed D, for n = 1000. Thus the closer examination afforded by Table 1, and specifically the last column, shows that the statistic F_R is the only one that agrees well enough with the theoretical distribution to have come in fact from the theoretical distribution. The other statistics have distributions which may be approximated reasonably well by their theoretical distributions however, as indicated by the small values of D.

The second situation introduces nuisance parameters in the form of replication effects ρ_1=-.5, ρ_2=+.5, while the other parameters remain at zero. This is intended to represent a mild form of deviation from the first situation. The results as given in Figure 2 show essentially the same results as in Figure 1, which indicates that all three statistics are relatively robust in the presence of slight nuisance factors. Table II presents some

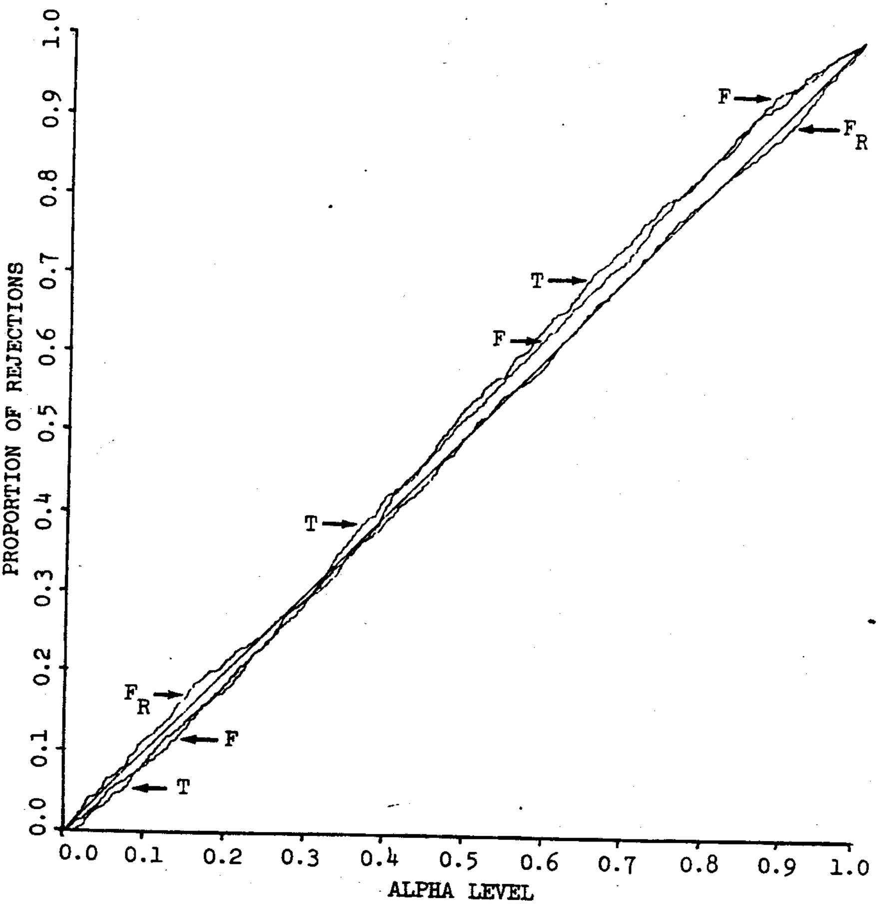

FIG. 1

The Null Distribution of F, T, and F_R When No Effects Are Present

TABLE I

Test Statistic	Proportions of Rejections at			D	$\hat{\alpha}$
	α=.01	.05	.10		
F	.007	.042	.082	.0448	.0350
T	.000	.032	.084	.0490	.0159
F_R	.009	.056	.114	.0237	.6187

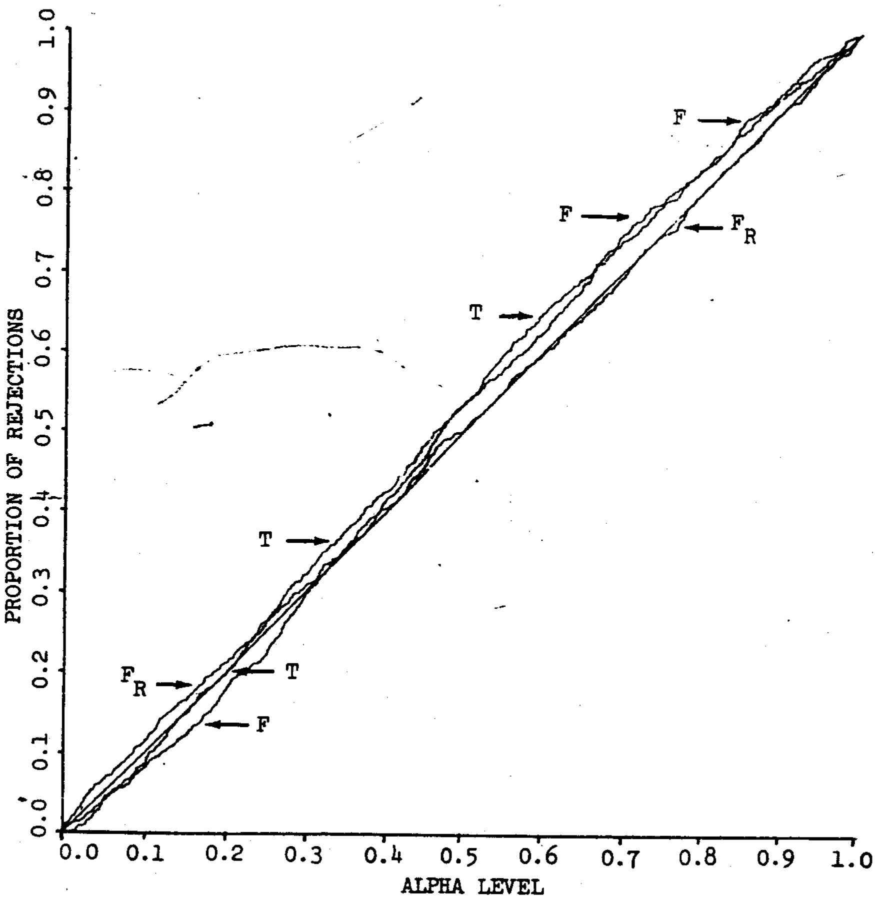

FIG. 2

The Null Distribution of F, T, and F_R in the Presence of Replication Effects Only

TABLE II

Test Statistic	Proportion of Rejections at			D	$\hat{\alpha}$
	α=.01	.05	.10		
F	.010	.040	.082	.0490	.0159
T	.001	.039	.087	.0507	.0113
F_R	.016	.066	.115	.0227	.6717

numerical comparisons. The same comments apply here as applied to Table I.

In the third and final situation for the null hypothesis, severe strain is put on the procedures in the form of block, treatment, and replication factors being present. For this situation we let $\alpha_2=.5$, $\alpha_3=-.5$, $\beta_1=.5$, $\beta_2=-.5$, in addition to $\rho_1=-.5$, $\rho_2=+.5$ as before. The three graphs are presented in Figure 3. While the distribution of the T statistic appears to be affected the most by the nuisance parameters, behavior in the region of small α is still reasonably good. Table III shows the same results as the other two tables; that all three tests are reasonably robust in the presence of nuisance parameters, and that the distribution of the rank transform statistic F_R agrees remarkably well with the theoretical F distribution.

For power studies we first introduce interaction in the form $\gamma_{11} = \gamma_{33} = .5$, $\gamma_{13} = \gamma_{31} = -.5$, with no nuisance parameters present (all α_i, β_j, ρ_k, the remaining γ_{ij}, and μ equal zero). The results are presented in Figure 4 in the form of the three empirical power curves plotted against the level of significance α. Here the rank transform F_R clearly has the most power, while the usual F test and the Mehra-Sen T are essentially equivalent. Some numerical comparisons are given in Table IV.

The final power study incorporates the nuisance parameters used in situation three above, and the interaction effects of the previous power study, to study how the power might be affected by the presence of nuisance parameters. The result shows the power of the rank statistics T and F_R to be relatively unaffected by the nuisance parameters, while the power of the F test on the raw data is diminished somewhat, especially for values of α between .10 and .50.

As a separate item of interest the 60×60 sample covariance matrix of $\underset{\sim}{R}(\underset{\sim}{Y})$ was computed for the 1000 repetitions of the experiment, in each of the five cases. The correlation between ranks was remarkably close to zero and the variances were almost equal,

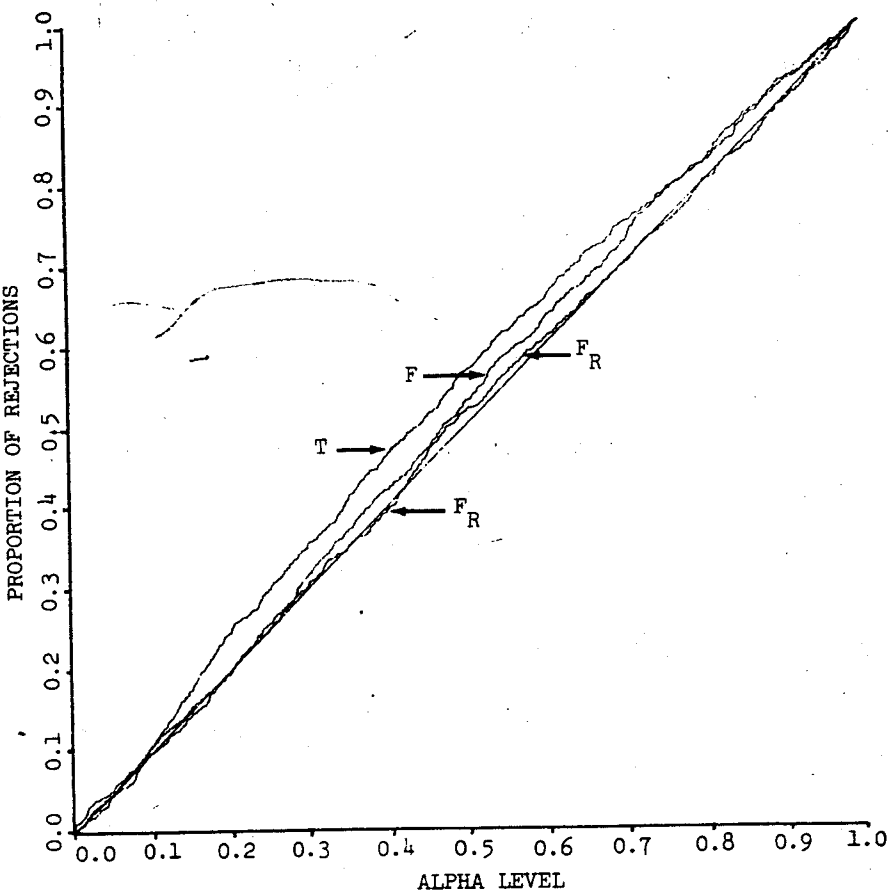

FIG. 3

The Null Distribution of F, T, and F_R in the Presence of Block, Treatment, and Replication Effects

TABLE III

Test Statistic	Proportion of Rejections at			$\underline{D}$	$\underline{\hat{\alpha}}$
	α=.01	.05	.10		
F	.017	.057	.100	.0407	.0708
T	.008	.049	.109	.0704	.000094
F_R	.009	.048	.108	.0219	.7138

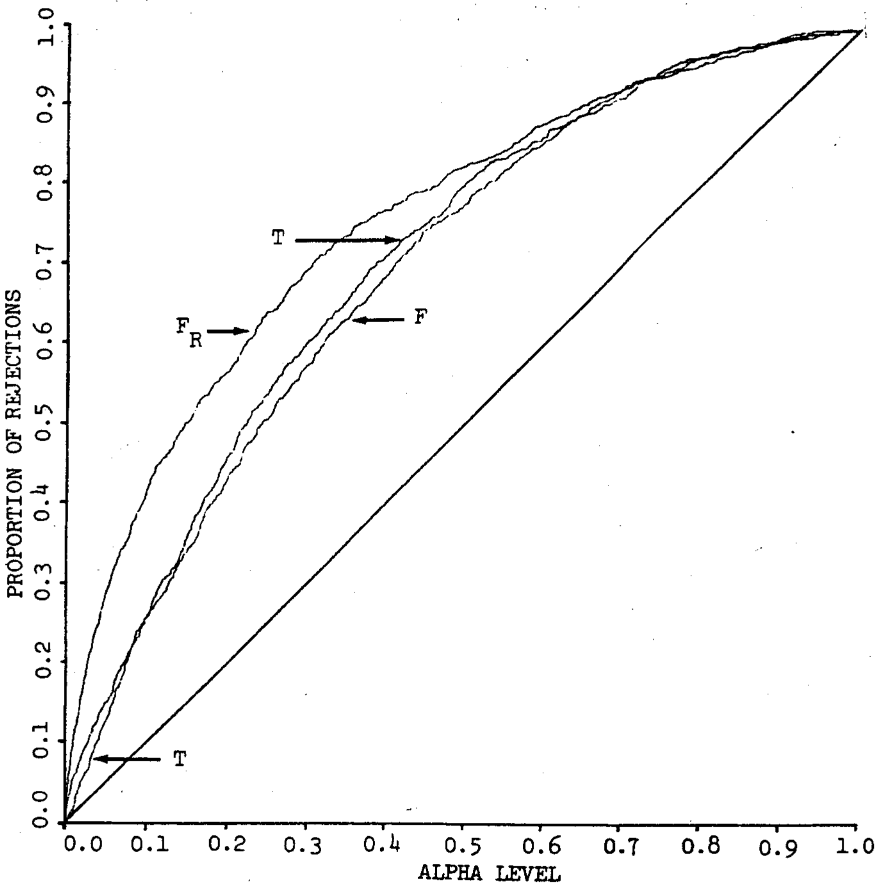

FIG. 4

The Nonnull Distribution of F, T, and F_R When No Effects Are Present Except Interaction

TABLE IV

Test Statistic	Proportion of Rejections at						
	α=.01	.025	.05	.10	.15	.20	.25
F	.055	.096	.150	.258	.344	.433	.506
T	.016	.062	.126	.261	.361	.459	.536
F_R	.101	.188	.290	.418	.502	.564	.639

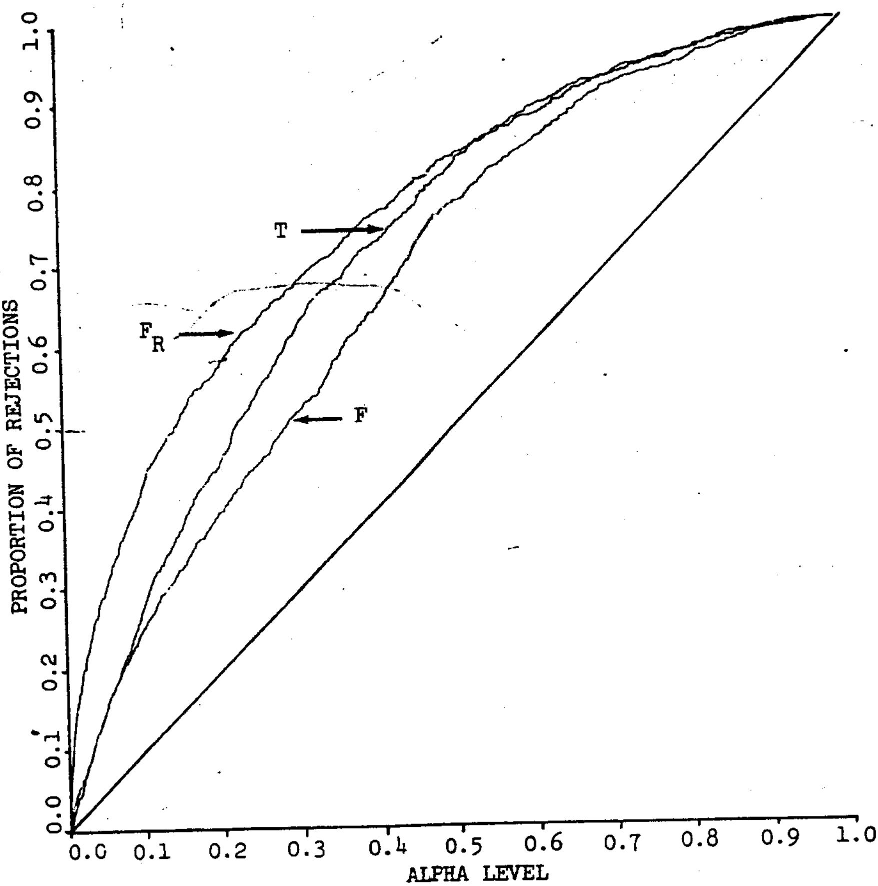

FIG. 5

The Nonnull Distribution of F, T, and F_R When Block, Treatment, Replication and Interaction Effects Are Present

TABLE V

Test Statistic	Proportion of Rejections at						
	α=.01	.025	.05	.10	.15	.20	.25
F	.046	.082	.155	.254	.327	.396	.457
T	.036	.081	.152	.280	.375	.455	.548
F_R	.140	.218	.298	.428	.514	.576	.631

which presented a covariance matrix that resembled a constant times the identity matrix. This offers some explanation for the good behavior of the F statistic with the rank transform.

Another side study involved repeating the above simulation study on normal rather than lognormal data. The results were similar to the above under the null situations (cases 1, 2 and 3), and there was no appreciable difference in power between F and F_R under the alternative hypothesis (cases 4 and 5). The statistic T showed inferior power in this study.

7. DISCUSSION AND SUMMARY

The F-test in the analysis of variance is a powerful and convenient procedure to use when the assumptions of a linear model and normally distributed error terms are reasonable for the data being analyzed. Violation of these assumptions might not affect the robustness of the test, but may seriously decrease the power of the test, as illustrated in this monte carlo study.

Aligned ranks procedures offer an alternative to the analysis of variance in many situations. The exact conditional distribution, given certain configurations of ranks, of the test statistic can be obtained to furnish an exact nonparametric procedure, but the computations involved are too laborious even for high speed computers except in the simplest cases. Therefore an asymptotic approximation is usually employed for the test statistic. Aligned ranks procedures still rely on the assumption of a linear model for their validity. Although aligned ranks tests are available for many situations such as the one illustrated for interaction in a two way layout with replication, they are not available for other situations such as for testing treatment or block effects in the presence of treatment-block interaction. These tests are more general than we indicate in this paper; one may use a wide variety of scores, other than merely ranks, and one may align using medians, midquartiles, or other statistics rather than means as we used here. This allows more flexibility for the experienced practitioner, but also causes considerable confusion on the part of the vast majority of data analysts who prefer a cookbook approach.

The rank transform procedure has the same convenience features of the analysis of variance. For computer analysis, one merely inserts a subroutine that replaces all of the original observations with their ranks, and then employs the usual computer packages for data analysis. Analysis of the data and of the ranks can be examined side by side to see if they agree, indicating validity of the parametric assumptions or if they disagree, indicating that the data should be reexamined for possible violations of the parametric assumptions. The null distribution of the rank transform statistic behaves very well in the situations examined, and the power appears to be greater than that of its competitors in those same situations. The rank transform procedure is not a nonparametric procedure, it merely uses ranks as a means of transforming the data into numbers that more nearly fit the assumptions of the parametric model, and yet retain all of the ordinal information contained in the data.

BIBLIOGRAPHY

Conover, W. J. (1971). Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc.

Graybill, Franklin A. (1969). Introduction to Matrices with Applications in Statistics. Belmont, Ca.: Wadsworth Publ. Co.

Hodges, J. L., Jr., and Lehmann, E. L. (1962). Rank methods for combination of independent experiments in the analysis of variance. Ann. Math. Statist. 33, 482-97.

Iman, Ronald L. (1974a). A power study of a rank transform for the two-way classification model when interaction may be present. Canad. J. Statist. C: Appl. 2 (2), 227-39.

Iman, Ronald L. (1974b). Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed ranks test statistic. Commun. Statist. 3 (8), 795-806.

Iman, Ronald L., Quade, Dana, and Alexander, Douglas A. (1975). Exact probability levels for the Kruskal-Wallis test. Selected Tables in Mathematical Statistics, 3 (Eds., H. L. Harter and D. B. Owen). Providence, R.I.: American Mathematical Society, 329-84.

Mehra, K. L. and Sen, P. K. (1969). On a class of conditionally distribution-free tests for interactions in factorial experiments. Ann. Math. Statist. 40, 658-64.

Refereed by:

Hans K. Ury, Permanente Medical Group, Oakland, California;
and
A member of the Editorial Board of this journal.