

# ASYMPTOTICALLY NONPARAMETRIC INFERENCE: AN ALTERNATIVE APPROACH TO LINEAR MODELS<sup>1</sup>

BY E. L. LEHMANN<sup>2</sup>

*University of California, Berkeley*

**1. Summary.** As an alternative to the standard analysis of variance, procedures are developed for linear models with several observations per cell which are asymptotically distribution-free and whose asymptotic efficiency relative to the standard procedures is the same as that of the Wilcoxon test relative to Student's  $t$ -test. Specific procedures discussed are (i) tests of linear hypotheses, (ii) confidence intervals for any contrast, (iii) simultaneous intervals for all contrasts.

**2. Introduction.** One of the basic problems of statistics is the two-sample location problem, which arises in the comparison of two populations. Let  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$  be two samples and suppose that

$$(2.1) \quad P\{X_i \leq x\} = F(x) \quad \text{and} \quad P\{Y_j \leq y\} = F(y - \Delta).$$

The classical estimate of  $\Delta$  is of course  $\bar{Y} - \bar{X}$ , while the hypothesis  $H : \Delta = 0$  was tested during the 19th century by referring the test statistic  $t = (\bar{Y} - \bar{X}) / (1/m + 1/n)^{1/2} S$  to the normal distribution to get an approximate significance level which, for large samples and  $F$  with finite second moment, is justified by the central limit theorem.

A new approach to testing—small sample theory—was initiated by Student (1908) who conjectured the exact form of the distribution of  $t$  for the case that  $F$  is normal. His conjecture was proved by R. A. Fisher (1923) who went on to develop the exact distributions of a number of other test statistics under the assumption of normality. (An earlier such derivation had been that of Helmert (1876) of the  $\chi^2$ -distribution of  $S^2$ ).

Attention was thus focused on the possibility of *exact tests*, i.e., tests whose significance level could be exactly computed under suitable assumptions, for example the assumption of normality. This aspect was further emphasized through the development in the 1930's by Neyman and Pearson [18] of a theory of optimum tests. Tests were sought which are most powerful in a clearly defined parametric model within the class of all similar tests, i.e., tests whose significance level is exactly equal to the specified value independent of any nuisance parameters of the model.

The emphasis on exactness raised the question what to do when the form of

---

Received January 29, 1963.

<sup>1</sup> This research was done while the author was a Professor of the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California.

<sup>2</sup> I am grateful to Professor John W. Tukey for a critical reading of the manuscript and several helpful suggestions.

the distribution could not be assumed to be known, and hence no particular parametric model could be postulated. In response to this need, nonparametric tests were developed, such as the tests proposed by Wilcoxon (1945) (see also [13]), whose levels do not depend on the underlying distribution and are therefore exact for a broadly nonparametric model, for example model (2.1) where  $F$  is only required to be continuous.

Such nonparametric tests were regarded rather skeptically at first. Since they use only the order relationship between the observations, it seemed obvious that these tests discard so much of the information contained in the data that there must result a considerable loss of power. However, in this case intuition proved unreliable. Work by Pitman (1949) and others showed for example that, asymptotically, the Wilcoxon test is about 95.5% efficient in the case of a normal distribution<sup>3</sup>; that it is more efficient than the  $t$ -test if the normal distribution is seriously disturbed by gross errors; and that its asymptotic efficiency relative to the  $t$ -test is never (i.e., for no  $F$ ) less than 86.4% but can be arbitrarily high and even infinite.

The Wilcoxon test is thus seen to be superior to the  $t$ -test in two respects.

(i) Its significance level is more robust in that it is exactly equal to the stated value for all continuous distributions  $F$ . (It has this value also in much more general models, which require no assumption of independence of the variables or identity of their distribution.)

(ii) Its power is more robust, particularly against distributions with heavy tails.

Of these two advantages, the second is perhaps the more important.

If it is only a question of testing the hypothesis  $\Delta = 0$  in model (2.1), a nonparametric test such as the Wilcoxon or Normal Scores test would nearly always seem preferable to the  $t$ -test. The objection is however frequently raised that occasions for such isolated tests are rare; that the nonparametric tests are not adaptable to point estimation, multiple decision problems, etc.; and that they are not suitable for a flexible use in linear models such as is provided by the standard analysis of variance.

It is the purpose of the present paper to describe procedures possessing this flexibility and enjoying the above property (ii), of nonparametric methods although unfortunately not property (i). The approach is here developed only for linear models in which there are several observations per cell. The significance level, confidence level, or other probabilities to be controlled, have their nominal values asymptotically as the numbers of observations in all cells become large; the procedures are therefore asymptotically distribution-free. They are in this sense comparable to the corresponding classical procedures. While the latter are "exact" under the assumption of normality, this assumption is of course never exactly fulfilled and in particular is likely to be disturbed by gross errors. In

---

<sup>3</sup> The exact values obtained by Klotz (1963) for the one-sample problem in the normal case for sample sizes 5 to 10 are even slightly higher.

practice, the classical procedures therefore also have the correct levels only asymptotically.

For some special linear hypotheses, nonparametric tests have been proposed by Friedman (1937), Durbin (1951), Kruskal and Wallis (1952), Benard and Van Elteren (1953), Hodges and Lehmann (1962), and others. This type of test seems however not applicable to many of the standard problems such as, for example, that of testing for the absence of interactions in a two-way layout. A general class of nonparametric procedures, applicable to a variety of linear hypotheses, has been proposed by Mood (1950) and Brown and Mood (1951). In the particular case of the  $c$ -sample problem, the asymptotic efficiency of the Brown-Mood median test relative to the classical  $F$ -test has been shown by Andrews (1954) to be the same as that of the sign test relative to Student's  $t$ -test in the one-sample problem. (For other efficiency results concerning the Brown-Mood tests see Bhapkar (1963)). In contrast, the procedures proposed here have the same asymptotic efficiency relative to the classical procedures as the Wilcoxon test has relative to the  $t$ -test.

**3. Linear models with several observations per cell.** Let the observable random variables be  $X_{i\alpha}$ , and suppose they are of the form

$$(3.1) \quad X_{i\alpha} = \xi_i + U_{i\alpha} \quad (\alpha = 1, \dots, n_i; i = 1, \dots, c)$$

where the  $U$ 's are independently distributed with common distribution  $F$  having density  $f$ , and the  $\xi$ 's are unknown constants. Let

$$(3.2) \quad Y_{ij} = \text{med}(X_{i\alpha} - X_{j\beta})$$

be the median of the  $n_i n_j$  differences  $X_{i\alpha} - X_{j\beta}$  ( $\alpha = 1, \dots, n_i; \beta = 1, \dots, n_j$ ), and let

$$(3.3) \quad Z_{ij} = Y_{i.} - Y_{j.},$$

where  $Y_{i.} = \sum Y_{ij}/c$ . It was shown in [15] that for  $n_i = \rho_i N$  and  $N$  tending to infinity, the variables  $N^{\frac{1}{2}}[Z_{ic} - (\xi_i - \xi_c)]$  have a limiting normal distribution with zero means and covariance matrix  $\Sigma = (\sigma_{ij})$  where

$$(3.4) \quad \begin{aligned} \sigma_{ii} &= \frac{1}{12} \left( \frac{1}{\rho_i} + \frac{1}{\rho_c} \right) / \left[ \int f^2(x) dx \right]^2 \\ \sigma_{ij} &= \frac{1}{12\rho_c} / \left( \int f^2(x) dx \right)^2 \end{aligned} \quad \text{for all } i, j < c.$$

Also for any  $i$  and  $j$

$$(3.5) \quad N^{\frac{1}{2}}Z_{ij} = N^{\frac{1}{2}}(Z_{ic} - Z_{jc}).$$

Consider now instead the statistics  $T_{ij} = X_{i.} - X_{j.}$ . The variables  $N^{\frac{1}{2}}[T_{ic} - (\xi_i - \xi_c)]$  are asymptotically normally distributed with zero means

and covariance matrix  $\Sigma'$ , which is related to  $\Sigma$  through

$$(3.6) \quad \Sigma' = k^2 \Sigma$$

with  $k^2 = 12\sigma^2[\int f^2(x) dx]^2$ . Also of course,  $N^{\frac{1}{2}}T_{ij} = N^{\frac{1}{2}}(T_{ic} - T_{jc})$ .

These facts may be summarized in

**THEOREM 1.** *The matrices with elements  $kN^{\frac{1}{2}}[Z_{ij} - (\xi_i - \xi_j)]$  and  $N^{\frac{1}{2}}[T_{ij} - (\xi_i - \xi_j)]$  respectively, have the same limiting distribution.*

The parallel notation in the definitions of the statistics  $Z_{ij}$  and  $T_{ij}$ , frequently makes it easy to write down the estimates, tests, etc. based on the  $Z$ 's simply by substituting in the corresponding classical procedures. However, in one respect the parallel notation is deceptive. The averaging indicated by a dot in  $X_{i.}$  is over the  $n_i$  observations in the  $i$ th cell; the corresponding average in  $Y_{i.}$ , on the other hand, is taken over the  $c$  cells. This difference sometimes leads to certain dissimilarities in the notation as for instance in the example of Section 4 below.

A principal purpose in introducing the estimates  $Y_{ij}$  is to provide protection against the possibility of gross errors. It should be pointed out in this connection that for  $n_i = n_j = 2$ ,  $Y_{ij}$  coincides with the usual estimate  $T_{ij} = X_{i.} - X_{j.}$  and hence does not provide such protection. More generally, little protection is afforded by  $Y_{ij}$  unless at least both  $n_i \geq 3$  and  $n_j \geq 3$ .

Let us now see how the estimates  $Z_{ij}$  of  $(\xi_i - \xi_j)$  can be used to obtain large sample tests of linear hypotheses concerning contrasts in the model (3.1). Consider first the classical test. If  $F$  were normal with known  $\sigma$ , attention could be restricted to the statistics  $X_{i.} = \xi_i + U_{i.}$ . A linear hypothesis  $H$  specifies that the vector  $(\xi_1, \dots, \xi_c)$  of means of  $(X_{1.}, \dots, X_{c.})$  lies in a  $(c - r)$ -dimensional linear subspace  $\Pi_\omega$  of  $c$ -space. The test is based on the statistic

$$(3.7) \quad \sum n_i (X_{i.} - \bar{\xi}_i)^2 / \sigma^2$$

where  $(\bar{\xi}_1, \dots, \bar{\xi}_c)$  is the projection of  $(X_{1.}, \dots, X_{c.})$  on  $\Pi_\omega$ .

Under  $H$ , the statistic (3.7) has a  $\chi^2$ -distribution with  $r$  degrees of freedom if  $F$  is normal, and has this as limiting distribution for any  $F$  with finite variance. Let

$$(3.8) \quad X_{i.} - \bar{\xi}_i = \sum_{j=1}^{c-1} a_{ij} (X_{j.} - X_{c.}),$$

so that (3.7) may be written as

$$(3.9) \quad \sum n_i [\sum a_{ij} (X_{j.} - X_{c.})]^2 / \sigma^2.$$

From (3.8) it follows that

$$(3.10) \quad \sum a_{ij} (\xi_c - \xi_j) = 0 \quad \text{for all } (\xi_1, \dots, \xi_c) \in \Pi_\omega,$$

and hence from Theorem 1 that the statistics  $N^{\frac{1}{2}} \sum a_{ij} T_{jc}$  ( $i = 1, \dots, c - 1$ ) under  $H$  have the same joint limiting distribution as the statistics  $kN^{\frac{1}{2}} \sum a_{ij} Z_{jc}$ . Since (3.9) has a  $\chi^2$ -distribution with  $r$  degrees of freedom under  $H$ , we have

therefore shown that the limiting distribution under  $H$  of

$$(3.11) \quad k^2 \sum n_i (\sum a_{ij} Z_{ij})^2 / \sigma^2 = 12 \left( \int f^2(x) dx \right)^2 \sum n_i (\sum a_{ij} Z_{ij})^2$$

is  $\chi^2$  with  $r$  degrees of freedom. The statistic  $\sum n_i (\sum a_{ij} Z_{ij})^2$  can be obtained very simply from the classical statistic  $\sum n_i (X_i - \bar{\xi}_i)^2$ . The latter is a quadratic form in the variables  $(X_1, \dots, X_c)$  and it is only necessary to replace these by the variables  $(Y_1, \dots, Y_c)$ , as is illustrated in Section 4 below.

It has been tacitly assumed in the above argument that  $\sigma^2 < \infty$  and

$$(3.12) \quad \int f^2(x) dx < \infty.$$

Actually, the latter assumption alone is enough to insure the result; for it follows from (3.4) and (3.5) that the joint limiting distribution of the variables  $kN^{\frac{1}{2}} \cdot [Z_{ij} - (\xi_i - \xi_j)]$  is the same as the distribution of  $N^{\frac{1}{2}}[(X'_{i\alpha} - X'_{j\alpha}) - (\xi_i - \xi_j)]$  where the variables  $X'_{i\alpha}$  are independently normally distributed with mean  $\xi_i$  and unit variance. Assumption (3.12) clearly holds whenever the density  $f$  is bounded.

The quantity (3.11) is not yet a usable test statistic since it depends on the unknown quantity  $\int f^2(x) dx$ . The following are two ways out of this difficulty.

(i) Let  $T_N$  be any consistent estimate of  $1/\int f^2(x) dx$  as the sample sizes  $n_i$  tend to infinity. Then the test statistic

$$(3.13) \quad W = 12 \sum n_i (\sum a_{ij} Z_{ij})^2 / T_N^2$$

under  $H$  has a limiting  $\chi^2$ -distribution with  $r$  degrees of freedom. The rejection region  $W > C$ , where  $C$  is the  $100\alpha$  upper percentage point of the  $\chi^2$ -distribution with  $r$  degrees of freedom, therefore provides a large-sample level  $\alpha$  test of  $H$ .

Two such consistent estimates are provided in the paper following the present one. If  $T^{(i)}$  denotes the estimate (15) of that paper based on the observations from the  $i$ th cell, one may take for  $T_N$  the average  $\sum T^{(i)}/c$ . Alternatively if  $T^{(i,j)}$  is the estimate (9) of the succeeding paper based on the observations from the  $i$ th and  $j$ th cell, one may take for  $T_N$  the average of the  $T^{(i,j)}$ 's.

(ii) In many cases, it is possible to obtain a quadratic form, say  $Q'$ , which is asymptotically independent of  $Q = \sum n_i (\sum a_{ij} Z_{ij})^2$ , and which after multiplication with  $12(\int f^2(x) dx)^2$  is asymptotically distributed as  $\chi^2$  with, say,  $r'$  degrees of freedom, regardless of whether the hypothesis being tested is true or not. The statistic

$$(3.14) \quad W' = Q/r \div Q'/r'$$

under  $H$  then has a limiting  $F$  distribution with  $r$  and  $r'$  degrees of freedom and hence provides a large-sample test of  $H$ .

Such quadratic forms will be available whenever all the contrasts in a subspace of the full contrast space may be assumed to be zero. If, for example, in a two-way layout it is assumed that the interactions are zero, the quadratic form

that would have been appropriate for testing the hypothesis that the interactions are zero could be used for  $Q'$ .

Other such quadratic forms can be obtained by dividing the observations in each cell into two groups (say, at random, with the sizes of the subgroups equal or differing by one). For the  $i$ th cell one may then set up the spurious hypothesis  $H_i$  that there is no difference in location between the distributions of the observations in the two subgroups. Let the variables in the two subgroups of the  $i$ th cell be denoted by  $X'_{ij}$  ( $j = 1, \dots, n'_i$ ) and  $X''_{ik}$  ( $k = 1, \dots, n''_i$ ) with  $n'_i + n''_i = n_i$ . Then the test statistic appropriate for  $H_i$  is

$$(3.15) \quad Q_i = n_i Y_i'^2 / [(1/n'_i) + (1/n''_i)],$$

where  $Y_i' = \text{med}(Y'_{ij} - Y''_{ik})$ . On multiplication with  $12(\int f^2(x) dx)^2$ ,  $Q_i$  has a limiting  $\chi^2$ -distribution with one degree of freedom. Furthermore,  $Q_1, \dots, Q_c$  are asymptotically independent of each other, and of  $Q$  and of  $Q'$ . If the sample sizes  $n_i$  are sufficiently large for this method to be applicable ( $\geq 6$  to make the half-group size  $\geq 3$ , and preferable  $\geq 8$  so that the  $\chi^2$ -approximation is based on half-group sizes  $\geq 4$ ), then  $Q_1 + \dots + Q_c$  provides a quadratic form with  $c$  degrees of freedom, which may be added to the form  $Q'$  suggested above or be used in its place if  $\xi_1, \dots, \xi_c$  are completely unknown.

The above suggestion of dividing the observations in each cell into two groups was based on the assumption, corresponding to typical practice, that the number of observations per cell is not very large. Of course, if many observations are available for a cell, they can be divided into more than two subgroups, with a corresponding gain in degrees of freedom for the resulting quadrating form. In fact, suppose that the observations in the  $i$ th cell are divided into  $k_i + 1$  subgroups, and let  $Q_i$  be the quadratic form appropriate for testing the hypothesis  $H_i$  that there is no difference in location between the distributions of the observations in these subgroups. Then the same asymptotic results (as the numbers of observations in all subgroups become large) hold as in the case of two subgroups with the only exception that the limiting  $\chi^2$ -distribution of  $Q_i$  now has  $k_i$  degrees of freedom, and that of  $\sum Q_i$  therefore  $\sum k_i$  degrees of freedom.

The approach of this section also leads to large-sample confidence sets for any contrast or set of contrasts. Consider for example a single contrast

$$(3.16) \quad \theta = \sum c_i \xi_i = \sum \sum d_{ij} (\xi_i - \xi_j), \quad (\sum c_i = 0).$$

The standard acceptance region of the hypothesis  $\theta = \theta_0$  (for known  $\sigma$ ) is  $|\sum \sum d_{ij} (X_{i.} - X_{j.}) - \theta_0| / \sigma (\sum c_i^2 / n_i)^{\frac{1}{2}} \leq C$  and the associated confidence statement

$$(3.17) \quad |\sum \sum d_{ij} [(X_{i.} - X_{j.}) - (\xi_i - \xi_j)]| \leq (C\sigma/N^{\frac{1}{2}}) (\sum c_i^2 / \rho_i)^{\frac{1}{2}}$$

has limiting probability  $1 - \alpha$ . By Theorem 1, the statement

$$(12N)^{\frac{1}{2}} \int f^2(x) dx |\sum \sum d_{ij} [(Y_{i.} - Y_{j.}) - (\xi_i - \xi_j)]| \leq C (\sum c_i^2 / \rho_i)^{\frac{1}{2}}$$

also has limiting probability  $1 - \alpha$ . Substituting the consistent estimate  $T_N$  of  $1/\int f^2(x) dx$ , we therefore get the asymptotic confidence intervals at level  $1 - \alpha$

$$(3.18) \quad \begin{aligned} \sum \sum d_{ij}(Y_{i.} - Y_{j.}) - [C/(12N/T_N^2)]^{\frac{1}{2}}[\sum (c_i^2/\rho_i)]^{\frac{1}{2}} &\leq \theta \\ &\leq \sum \sum d_{ij}(Y_{i.} - Y_{j.}) + [C/(12N/T_N^2)]^{\frac{1}{2}}[\sum (c_i^2/\rho_i)]^{\frac{1}{2}}. \end{aligned}$$

Alternatively, let  $Q'$  be a quadratic form which is asymptotically independent of  $\sum \sum d_{ij}(Y_{i.} - Y_{j.})$  and which on multiplication by  $12(\int f^2(x) dx)^2$  is asymptotically distributed as  $\chi^2$  with  $r'$  degrees of freedom. Then

$$N^{\frac{1}{2}}[\sum \sum d_{ij}(Y_{i.} - Y_{j.}) - \theta_0]/(\sum c_i^2/\rho_i)^{\frac{1}{2}} \div (Q'/r')^{\frac{1}{2}}$$

is asymptotically distributed as Student's  $t$  with  $r'$  degrees of freedom. Hence if  $C$  is the critical value for a two-sided  $t$ -test at level  $\alpha$ ,

$$(3.19) \quad \sum \sum d_{ij}(Y_{i.} - Y_{j.}) \pm C(\sum c_i^2/\rho_i)^{\frac{1}{2}}(Q')^{\frac{1}{2}}/r'N^{\frac{1}{2}}$$

will provide another set of asymptotic confidence intervals for  $\theta$  at level  $1 - \alpha$ .

In exactly the same manner, one can obtain simultaneous approximate confidence statements based on the  $Y$ 's for all contrasts, analogous to Scheffé's  $S$ -method. For let  $C'$  be determined so that the limiting probability is  $1 - \alpha$  that (3.18) holds for all matrices  $(d_{ij})$  when  $C$  is replaced by  $C'$ . Then it follows as before that the limiting probability is  $1 - \alpha$ , that (3.18) will hold for all matrices  $(d_{ij})$ , and hence for all contrasts  $\theta$ , when  $C$  is replaced by  $C'$ . The same remark applies of course to the intervals (3.19).

**4. An example.** As an illustration, consider a two-way layout with  $n$  observations per cell. With a slight change of notation, let the observable variables be denoted by

$$(4.1) \quad X_{ij\alpha} = \xi_{ij} + U_{ij\alpha} \quad (i = 1, \dots, a; j = 1, \dots, b; \alpha = 1, \dots, n),$$

where the  $U$ 's are independently and identically distributed. In the usual decomposition into main effects and interactions, let

$$(4.2) \quad \xi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad (\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0)$$

and let  $Y_{ijk} = \text{med}(X_{kl\beta} - X_{ij\alpha})$  be the median of all differences between observations in the  $kl$ th and  $ij$ th cell. Then the adjusted estimates of the  $\alpha$ 's,  $\beta$ 's and  $\gamma$ 's are

$$(4.3) \quad \begin{aligned} \hat{\alpha}_i &= Y_{i\dots} - Y_{\dots}, & \hat{\beta}_j &= Y_{.j\dots} - Y_{\dots} \\ \hat{\gamma}_{ij} &= Y_{ij\dots} - Y_{i\dots} - Y_{.j\dots} + Y_{\dots}, \end{aligned}$$

where actually  $Y_{\dots} = 0$ . Let  $T_N$  be the estimate of  $1/\int f^2(x) dx$  which is discussed in the preceding section. Using method (i) above, we can then base a test of the hypothesis  $H: \alpha_1 = \dots = \alpha_a = 0$  on the statistic  $W = 12bn \sum (Y_{i\dots} - Y_{\dots})^2/T_N^2$  which under  $H$  has asymptotically a  $\chi^2$ -distribution with  $a - 1$  degrees of freedom. If  $C$  is the upper 100  $\alpha$ -percentage point of this

distribution, the rejection region  $W \geq C$  will provide an asymptotically distribution-free test of  $H$ .

Corresponding confidence sets of the  $\alpha$ 's can be based on the acceptance region  $12bn \sum (Y_{i\dots} - Y_{\dots} - \alpha_i^0)^2 / T_N^2 \leq C$  of the hypothesis  $\alpha_1 = \alpha_1^0, \dots, \alpha_a = \alpha_a^0$ . The resulting confidence sets for  $(\alpha_1, \dots, \alpha_a)$  are the intersections of the hyperspheres

$$(4.4) \quad \sum [\alpha_i - (Y_{i\dots} - Y_{\dots})]^2 \leq CT_N^2 / 12bn$$

with the hyperplanes

$$(4.5) \quad \sum \alpha_i = 0.$$

In an exactly analogous manner we obtain tests of hypotheses specifying the  $\beta$ 's or  $\gamma$ 's, or confidence sets for these parameters. The sums of squares on which the test statistics are based, are the terms of a decomposition, exactly analogous to the usual one in the analysis of variance of a two-way layout, namely

$$(4.6) \quad \sum \sum (Y_{ij\dots} - Y_{\dots})^2 = \sum \sum (Y_{ij\dots} - Y_{i\dots} - Y_{.j\dots} + Y_{\dots})^2 + b \sum (Y_{i\dots} - Y_{\dots})^2 + a \sum (Y_{.j\dots} - Y_{\dots})^2.$$

Note that this is the decomposition corresponding to that of analysis of variance when the variance is known, since we take the  $Y_{ij\dots}$  as our basic variables, the averaging over the cells being required for robustness and forming the basis of the large-sample theory.

Instead of (i), we may also use method (ii) above, where however the only available quadratic form for the denominator of (3.14) is  $Q_1 + \dots + Q_c$  with  $Q_i$  defined by (3.15). The changes which this modification requires in the tests and confidence procedures are obvious.

The situation is somewhat different if one is willing to assume the  $\gamma$ 's to be zero. The estimates  $\hat{\alpha}_i$  and  $\hat{\beta}_j$  are then still given by (4.3). However, for the quadratic form  $Q'$  one may now take

$$(4.7) \quad Q' = n \sum \sum \hat{\gamma}_{ij}^2 = n \sum \sum (Y_{ij\dots} - Y_{i\dots} - Y_{.j\dots} + Y_{\dots})^2$$

with  $(a - 1)(b - 1)$  degrees of freedom. A test of the hypothesis  $H: \alpha_1 = \dots = \alpha_a = 0$  is then given by the rejection region

$$(4.8) \quad Q / (a - 1) \div Q' / (a - 1)(b - 1) > C$$

where  $Q = bn \sum (Y_{i\dots} - Y_{\dots})^2$  and  $Q'$  is given by (4.7), and where  $C$  is the 100% upper percentage point of the  $F$ -distribution with  $(a - 1)$  and  $(a - 1)(b - 1)$  degrees of freedom.

The analogy between the test proposed here and the corresponding classical test becomes even closer if instead of with Model I we are concerned with a Model II analysis. Suppose that  $\xi_{ij}$ , instead of by (4.2), is given by

$$(4.9) \quad \xi_{ij} = \mu + A_i + B_j$$



where the  $A$ 's and  $B$ 's are jointly independent random variables, independent of the  $U$ 's, the  $A$ 's having common distribution  $G_A$  and the  $B$ 's common distribution  $G_B$ , say. The test (4.8) is then clearly still (asymptotically) valid for testing the hypothesis that  $G_A$  is concentrated on a single point, and it is exactly analogous to the corresponding classical test.

As an illustration of the above procedures, consider the two-way layout from Brownlee (1960) p. 379 with  $a = 2, b = 3$  and  $n = 4$ , used as an example in [15]. Suppose that the interactions can be assumed to be zero and that we wish to test the hypothesis  $\alpha_1 = \alpha_2 = 0$  in the model given by (4.1) and (4.2), or the hypothesis that  $G_A$  is concentrated on a single point in model (4.9). In [15] the following estimates were obtained for the  $\alpha$ 's,  $\beta$ 's and  $\gamma$ 's.

	$\alpha_1 = -\alpha_2$	$\beta_1$	$\beta_2$	$\beta_3$	$\gamma_{11} = -\gamma_{21}$	$\gamma_{12} = -\gamma_{22}$	$\gamma_{13} = -\gamma_{23}$
Estimates (4.3)	2.50	-17.54	1.29	16.25	-.29	-1.79	2.08
Classical estimates	2.54	-17.17	1.08	16.08	-.67	-1.67	2.33

We thus get the following decomposition of  $\sum \sum (Y_{ij..} - Y_{i...})^2$  and  $\sum \sum (X_{ij.} - X_{i...})^2$ .

	$bn \sum \hat{\alpha}_i^2 / (a - 1)$	$an \sum \hat{\beta}_j^2 / (b - 1)$	$n \sum \sum \hat{\gamma}_{ij}^2 / (a - 1)(b - 1)$
Decomposition (4.5)	150.00	2293.51	30.46
Classical Analysis of Variance	154.84	2218.17	34.67

In this example, the number of observations per cell is too small for use of the quadratic forms (3.15). Using therefore the test (4.8), we find the significance probability to be .16. For the corresponding classical test in model (4.9), the significance probability is .17; on the other hand, for the classical test in model (4.1), where in the denominator one can use the residual sum of squares, we find the significance probability to be .39.

**5. Asymptotic efficiency.** Let us now determine the asymptotic efficiency of the test  $W > C$  with  $W$  given by (3.13). Suppose that the classical statistic (3.7) is based on  $n'_i = \rho_i N'$  ( $i = 1, \dots, c$ ) observations  $X'_{i\alpha}$ , while (3.13) and (3.14) continue to be based on  $n_i = \rho_i N$  observations  $X_{i\alpha}$ , both the  $X'_{i\alpha}$  and  $X_{i\alpha}$  being given by (3.1). Then if  $T'_{ij} = X'_i - X'_j$ , it follows from Theorem 1 that the matrices with elements

$$(5.1) \quad kN^{\frac{1}{2}}[Z_{ij} - (\xi_i - \xi_j)] \quad \text{and} \quad (N')^{\frac{1}{2}}[T'_{ij} - (\xi_i - \xi_j)]$$

respectively, have the same limiting distribution.

Consider now the linear hypothesis  $H$  of Section 3. By (3.10) the null distribution of the test criteria (3.13) and (3.7) may be computed for  $\xi_1 = \dots =$

$\xi_c = 0$ , so that under  $H$  the matrices with elements  $kN^{\frac{1}{2}}Z_{ij}$  and  $(N')^{\frac{1}{2}}T'_{ij}$  have the same limiting distribution. As usual, we shall compare the power of the two tests against a sequence of alternatives  $K_N$  whose difference from  $H$  is of the order  $1/N^{\frac{1}{2}}$ . Because of (3.10), we may assume without loss of generality that the vectors  $(\xi_1, \dots, \xi_c) = (\xi_1^{(N)}, \dots, \xi_c^{(N)})$  of  $K_N$  satisfy

$$(5.2) \quad \xi_i^{(N)} - \xi_j^{(N)} = (b_i - b_j)/N^{\frac{1}{2}} + o(1/N^{\frac{1}{2}}).$$

If  $Z_{ij}^* = Z_{ij} - (\xi_i - \xi_j)$  and  $T_{ij}^* = T'_{ij} - (\xi_i - \xi_j)$ , we then have  $kN^{\frac{1}{2}}Z_{ij} = kN^{\frac{1}{2}}Z_{ij}^* + k(b_i - b_j) + o(1)$  and  $(N')^{\frac{1}{2}}T'_{ij} = (N')^{\frac{1}{2}}T_{ij}^* + (N'/N)^{\frac{1}{2}}(b_i - b_j) + (N')^{\frac{1}{2}}o(1/N^{\frac{1}{2}})$ . It follows from the asymptotic equivalence of the matrices (5.1), that the matrices with elements

$$(5.3) \quad kN^{\frac{1}{2}}Z_{ij} \quad \text{and} \quad (N')^{\frac{1}{2}}T'_{ij}$$

have the same limit distribution under  $K_N$  provided  $N' = g(N)$  is determined in such a way that

$$(5.4) \quad N'/N \rightarrow k^2 \quad \text{as} \quad N \rightarrow \infty.$$

If (5.4) holds, the two test statistics therefore have the same limit distribution under  $K_N$  and hence have the same limiting power against these alternatives. This shows that  $k^2$  is the Pitman efficiency of the test based on (3.13) relative to the classical test based on (3.7), or to the corresponding test with  $\sigma$  replaced by  $S$ . This is of course also the asymptotic efficiency of the test based on (3.14), provided each cell is divided into sufficiently many subgroups so that the number of degrees of freedom  $r'$  of  $Q'$  tends to infinity with the sample sizes.

To discuss the efficiency of the confidence intervals (3.18) we may use the relationship between the power of a test and the probability of false parameter values being covered by the associated confidence sets. From this relationship it follows that  $k^2$  is also the asymptotic efficiency of the confidence intervals (3.18) relative to the intervals (3.17) in the following sense. The intervals (3.18) based on  $N'$  observations have the same limiting probability of covering values  $\theta^{(N)}$  of the contrast differing from the true value by a term of order  $1/N^{\frac{1}{2}}$ , i.e., of the form  $\theta^{(N)} = \theta + \Delta/N^{\frac{1}{2}} + o(1/N^{\frac{1}{2}})$ , provided (5.4) holds. Again, the result is unchanged if in (3.17) the standard deviation  $\sigma$  is replaced by a consistent estimate  $S$ .

Some authors prefer to measure the accuracy of confidence intervals by their length rather than by the probability of covering false values. (For a discussion of this issue see Pratt [20], [21] and Madansky [16].) The length of the intervals (3.18) based on  $N$  observations is

$$(5.5) \quad L = 2C(\sum c_i^2/\rho_i)^{\frac{1}{2}}/(12NT_N^2)^{\frac{1}{2}}$$

while that of the intervals (3.17), based on  $N'$  observations and with  $\sigma$  replaced by  $S$ , is

$$(5.6) \quad L' = 2C(\sum c_i^2/\rho_i)^{\frac{1}{2}}S/(N')^{\frac{1}{2}}.$$

The ratio  $L'/L$  tends in probability to  $k(N/N')^{\frac{1}{2}}$ . Thus, if  $N' = N$ ,  $k^2$  is the efficiency of the intervals (3.18) to the intervals (3.17) in the sense that it is the limit of the ratio of the squares of the lengths of the intervals. Alternatively, it is the efficiency in the sense that if the intervals are based on  $k \sum \rho_i N$  and  $\sum \rho_i N$  observations respectively, then the limiting ratio of the length is 1.

A third approach to the comparison of two sets of confidence intervals can be based on a definition of accuracy suggested by Wolfowitz (1950). If  $(\theta_L, \theta_U)$  are confidence intervals for a parameter  $\theta$ , Wolfowitz proposed as a measure of accuracy the quantity

$$(5.7) \quad W(\theta; \theta_L, \theta_U) = a(\theta_L - \theta)^2 + b(\theta_U - \theta)^2.$$

Let us denote by  $(\theta_{LX}, \theta_{UX})$  and  $(\theta_{LY}, \theta_{UY})$  the confidence intervals for  $\theta$  defined by (3.17) and (3.18) respectively. Then  $N^{\frac{1}{2}}(\theta_{LX} - \theta)$  has a limiting normal distribution with mean and variance  $\gamma = C\sigma(\sum c_i^2/\rho_i)^{\frac{1}{2}}$  and  $\tau^2 = \sigma^2 \sum c_i^2/\rho_i$  and the difference  $D = N^{\frac{1}{2}}[(\theta_{UX} - \theta) - (\theta_{LX} - \theta)]$  tends to  $\lambda = 2C\sigma(\sum c_i^2/\rho_i)^{\frac{1}{2}}$ . (If in (3.18)  $\sigma$  is replaced by  $S$ , the difference  $D$  tends to  $\lambda$  in probability.) By Theorem 1,  $N^{\frac{1}{2}}(\theta_{LY} - \theta)$  then has a normal limiting distribution with mean  $\gamma/k$  and variance  $\tau^2/k^2$ , and the difference  $N^{\frac{1}{2}}[(\theta_{UY} - \theta) - (\theta_{LY} - \theta)]$  tends in probability to  $\lambda/k$ . Suppose now that the intervals  $(\theta_{LX}, \theta_{UX})$  are based on  $N'$  observations where  $N'/N \rightarrow k^2$  as  $N \rightarrow \infty$ . Then it is easily seen that the quantities  $NW(\theta; \theta_{LX}, \theta_{UX})$  and  $NW(\theta; \theta_{LY}, \theta_{UY})$  have the same limit distribution, so that again  $k^2$  appears as a reasonable measure of asymptotic efficiency. The same result clearly holds for more general measures of accuracy.

In a similar sense, the comparison of the two methods for simultaneous estimation of several contrasts also leads to efficiency  $k^2$ . As an illustration consider the confidence spheres for the effects  $(\alpha_1, \dots, \alpha_c)$  discussed in Section 4. If accuracy is measured by the probability of covering false values, it follows of course from the corresponding result for tests that the asymptotic efficiency of the spheres given by (4.4) and (4.5) relative to the classical spheres is  $k^2$ .

An alternative comparison is in terms of the volumes of the spheres. The spheres defined by (4.4) and (4.5) are centered at the point  $(Y_{1\dots} - Y_{\dots}, \dots, Y_{c\dots} - Y_{\dots})$  and the square of their radius is  $R^2 = C/12bnT_N^2$ . On the other hand, the standard confidence spheres are centered on the point  $(X_{1..} - X_{\dots}, \dots, X_{c..} - X_{\dots})$  and the square of their radius is  $R'^2 = CS^2/bn'$  if the number of observations per cell is  $n'$ . It follows that if  $n'$  is determined so that  $n'/n \rightarrow k^2$  as  $n \rightarrow \infty$ , the ratio of the volumes of the two spheres tends to 1.

In analogy to (5.7), one might measure the accuracy of the confidence spheres by

$$(5.8) \quad W = a[\text{min squared distance of } (\alpha_1, \dots, \alpha_c) \text{ from the sphere}] \\ + b[\text{max squared distance of } (\alpha_1, \dots, \alpha_c) \text{ from the sphere}].$$

Then  $NW$  is a function of  $N^{\frac{1}{2}}R$  and  $N^{\frac{1}{2}}(\hat{\alpha} - \alpha)$  where  $R$  is the radius and  $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_c)$  the center of the sphere. From this it is easily seen as before that

$NW$  has the same limit distribution for the spheres defined by (4.5) and (4.6) and based on  $N$  observations, as for the standard spheres based on  $N'$  observations when  $N'/N \rightarrow k^2$ .

Finally, for Scheffé's method of judging all contrasts the comparison may be made, as suggested in [12], in terms of the lengths of the corresponding intervals. Since these are given for all contrasts by (5.5) and (5.6), the remarks made in this connection for a single contrast also apply to the present case.

Throughout the present paper, the procedures have been based on the two-sample estimates of contrasts discussed in Sections 2 to 4 of [15]. Instead, they can of course be based on the one-sample estimates of Section 5 of [15]. All results will then hold with the obvious modifications, subject only to the restriction that the distribution  $F$  of the variables  $U_{i\alpha}$  must be symmetric.

## REFERENCES

- [1] ANDREWS, FRED C. (1954). Asymptotic behavior of some rank tests for analysis of variance. *Ann. Math. Statist.* **25** 724-736.
- [2] BENARD, A. and VAN ELTEREN, PH. (1953). A generalization of the method of  $m$  rankings. *Indag. Math.* **15** 358-369.
- [3] BHAPKAR, V. P. (1963). The asymptotic power and efficiency of Mood's test for two-way classification. *J. Indian Statist. Assoc.* **1** 24-31.
- [4] BROWN, G. W. and MOOD, A. M. (1951). On median tests for linear hypotheses. *Proc. Second Berkeley Symp. Math. Statist. and Prob.* **II** 159-166.
- [5] BROWNLIE, K. A. (1960). *Statistical Theory and Methodology in Science and Engineering*. Wiley, New York.
- [6] DURBIN, J. (1951). Incomplete blocks in ranking experiments. *British J. Psych.* **4** 85-90.
- [7] FISHER, R. A. (1923). Note on Dr. Burnside's recent paper on error of observation. *Proc. Cambridge Philos. Soc.* **21** 655-658.
- [8] FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.* **32** 675-699.
- [9] HELMERT, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers direkter Beobachtungen gleicher Genauigkeit. *Astr. Nachr.* **88** 113-132.
- [10] HODGES, J. L., JR. and LEHMANN, E. L. (1956). The efficiency of some nonparametric competitors of the  $t$ -test. *Ann. Math. Statist.* **27** 324-335.
- [11] 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** 482-497.
- [12] KLOTZ, JEROME (1963). Small sample power and efficiency for the one sample Wilcoxon and normal scores test. *Ann. Math. Statist.* **34** 624-632.
- [13] KRUSKAL, W. H. (1957). Historical notes on the Wilcoxon unpaired two-sample test. *J. Amer. Statist. Assoc.* **52** 356-360.
- [14] KRUSKAL, WILLIAM H., and WALLIS, ALLEN W. (1952). Use of ranks on one-criterion variance analysis. *J. Amer. Statist. Assoc.* **47** 583-621.
- [15] LEHMANN, E. L. (1963). Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957-966.
- [16] MADANSKY, ALBERT (1962). More on length of confidence intervals. *J. Amer. Statist. Assoc.* **57** 586-589.
- [17] MOOD, ALEXANDER McFARLANE (1950). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

- [18] NEYMAN, J. and PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. Ser. A* **231** 289–337.
- [19] PITMAN, E. J. G. (1949). Lecture notes on nonparametric statistical inference. (Unpublished.)
- [20] PRATT, JOHN W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549–567.
- [21] PRATT, JOHN W. (1963). Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* **34** 574–586.
- [22] SCHEFFÉ, HENRY (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40** 87–104.
- [23] STUDENT (1908). On the probable error of a mean. *Biometrika* **6** 1–25.
- [24] WILCOXON, FRANK (1945). Individual comparisons by ranking methods. *Biometrics* **1** 80–83.
- [25] WOLFOWITZ, J. (1950). Minimax estimates of the mean of a normal distribution with known variance. *Ann. Math. Statist.* **21** 218–230.