

# ROBUST ESTIMATION IN ANALYSIS OF VARIANCE<sup>1</sup>

BY E. L. LEHMANN

*University of California, Berkeley*

**1. Summary.** In linear models with several observations per cell, estimates of all contrasts are given whose small and large sample behaviour is analogous to that of the estimate of a shift parameter proposed in [2]. In particular, the asymptotic efficiency of these estimates relative to the standard least squares estimates, as the number of observations in each cell gets large, is shown to be the same as the Pitman efficiency of the Wilcoxon test relative to the  $t$ -test.

**2. A compatible set of difference estimates.** Consider the  $c$ -sample model, in which the observations are

$$(2.1) \quad X_{i\alpha} = \xi_i + U_{i\alpha} \quad (\alpha = 1, \dots, n_i; i = 1, \dots, c)$$

where the variables  $U_{i\alpha}$  are independently distributed with distributions  $F_i$  having a common median. Let

$$(2.2) \quad Y_{ij} = \text{med} (X_{i\alpha} - X_{j\beta})$$

be the median of the  $n_{ij}$  differences  $X_{i\alpha} - X_{j\beta}$  ( $\alpha = 1, \dots, n_i; \beta = 1, \dots, n_j$ ). It was shown in [2] that the estimate  $Y_{ij}$  of  $\xi_i - \xi_j$  has more robust efficiency than, and hence may be preferable to, the standard estimate  $T_{ij} = X_{i\cdot} - X_{j\cdot}$ , where  $X_{i\cdot} = \sum X_{i\alpha}/n_i$ .

Unfortunately, unlike the estimates  $T_{ij}$ , the estimates  $Y_{ij}$  are incompatible in the sense (see [4]) that they do not satisfy the linear relations satisfied by the differences they estimate. This leads to ambiguities. To estimate, for example,  $\xi_3 - \xi_1$ , we will get a different answer if we estimate this difference directly by  $Y_{31}$  or if we take the sum of the estimates  $Y_{32}$  and  $Y_{21}$  of  $\xi_3 - \xi_2$  and  $\xi_2 - \xi_1$ . This is particularly awkward in the estimation of more general contrasts, where there is no longer a single "natural" estimate. Thus, for estimating  $(\xi_3 + \xi_4) - (\xi_1 + \xi_2)$  one will get the estimate  $Y_{31} + Y_{42}$  when writing the contrast as  $(\xi_3 - \xi_1) + (\xi_4 - \xi_2)$ , but  $Y_{41} + Y_{32}$  when writing it as  $(\xi_4 - \xi_1) + (\xi_3 - \xi_2)$ .

To avoid this difficulty, we shall replace the *raw* estimates  $Y_{ij}$  by *adjusted* estimates  $Z_{ij}$  which have the form  $\hat{\xi}_i - \hat{\xi}_j$  or equivalently which satisfy the set of restrictions  $Z_{ij} + Z_{jk} + Z_{ki} = 0$  for all  $i, j, k$ .

If we suppose that any serious effect of gross errors has been eliminated by replacing the estimates  $T_{ij}$  by  $Y_{ij}$ , it is natural to adjust these latter estimates by minimizing the sum of squares

$$(2.3) \quad \sum_{i \neq j} [Y_{ij} - (\xi_i - \xi_j)]^2.$$

Received December 20, 1962.

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

If  $Y_{ii}$  is defined to be zero for all  $i$  and  $\hat{\xi}_i - \hat{\xi}_j$  denote the minimizing values of (2.3), differentiation of (2.3) gives  $\sum_{j=i}^c (Y_{ij} - \hat{\xi}_i + \hat{\xi}_j) = 0$  and hence  $\hat{\xi}_i - (1/c) \sum_{j=1}^c \hat{\xi}_j = Y_{i.}$ , so that

$$(2.4) \quad Z_{ij} = \hat{\xi}_i - \hat{\xi}_j = Y_{i.} - Y_{j.}$$

We shall in the next sections show that the estimates  $Z_{ij}$  have essentially the same small- and large-sample properties that had been proved for the estimates  $Y_{ij}$  in [2]. They are affected however by a peculiarity which may be considered a disadvantage: The estimate of  $\xi_i - \xi_j$  depends not only on the observations from the  $i$ th and  $j$ th sample but also on those of the other samples, which are quite unrelated to  $\xi_i$  and  $\xi_j$ .

Suppose for example that  $c = 3$  and that  $x_{11} = -4, x_{12} = 0, x_{13} = 6; x_{21} = -10, x_{22} = 3, x_{23} = 7; x_{31} = -7, x_{32} = 0, x_{33} = 8$ . Then the raw estimates of  $\xi_3 - \xi_1, \xi_3 - \xi_2$  and  $\xi_2 - \xi_1$  are 0, 1 and 1 respectively. Since the data suggest that  $\xi_3$  is larger than  $\xi_2$  and  $\xi_2$  is larger than  $\xi_1$ , this is taken into account in arriving at the adjusted estimate for  $\xi_3 - \xi_1$ . Instead of estimating this difference to be zero, the adjusted estimate is  $\frac{1}{3}[2 \cdot 0 + 1 + 1] = \frac{2}{3}$ . As will be seen later, the size of the influence of the other samples becomes small as the sample sizes increase, tending in fact to zero in probability at a rate  $o_p(N^{-1})$  where  $N$  is the total sample size.

While the adjustment of the estimates does not affect their asymptotic efficiency, we unfortunately have no results concerning the effect on small-sample efficiency. It should be pointed out however that the adjustment may result in a serious loss of efficiency when the sample sizes are widely disparate. This may be illustrated on the case  $c = 3$ . The variance of the raw estimate  $Y_{12}$  of  $\xi_1 - \xi_2$  tends to zero as  $n_1$  and  $n_2$  tend to infinity. This is however not true of the variance of  $Z_{12} = \frac{1}{3}(2Y_{12} + Y_{13} - Y_{23})$  unless  $n_3$  also tends to infinity.

**3. The estimation of contrasts.** A contrast is a linear function of the  $\xi$ 's, say  $\sum c_i \xi_i$ , with  $\sum c_i = 0$ . Since this is a function only of the differences of the  $\xi$ 's, it can be written in the form

$$(3.1) \quad \theta = \sum \sum d_{ij}(\xi_i - \xi_j).$$

As estimate of the contrast (3.1) we propose

$$(3.2) \quad \hat{\theta} = \sum \sum d_{ij}Z_{ij} = \sum \sum d_{ij}(Y_{i.} - Y_{j.}).$$

The representation (3.1) is not unique; however the estimate (3.2) is independent of the representation. For if  $\sum \sum d_{ij}(\xi_i - \xi_j) = \sum \sum e_{ij}(\xi_i - \xi_j)$  for all vectors  $(\xi_1, \dots, \xi_c)$ , then obviously

$$\sum \sum d_{ij}(Y_{i.} - Y_{j.}) = \sum \sum e_{ij}(Y_{i.} - Y_{j.})$$

for all vectors  $(Y_{1.}, \dots, Y_{c.})$ .

As an example, consider a two-way layout with several observations per cell. Changing the notation of (2.1), let the observable random variables be  $X_{ij\alpha} =$

$\xi_{ij} + U_{ij\alpha}$  where the  $U$ 's are independent and identically distributed. In the usual decomposition into main effects and interactions, let  $\xi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$  where

$$(3.3) \quad \begin{aligned} \mu &= \xi_{..}, & \alpha_i &= \xi_{i.} - \xi_{..}, & \beta_j &= \xi_{.j} - \xi_{..}, \\ \gamma_{ij} &= \xi_{ij} - \xi_{i.} - \xi_{.j} + \xi_{..} \end{aligned}$$

If we denote the raw estimate of  $\xi_{ij} - \xi_{kl}$  by  $V_{ijkl}$ , so that  $V_{ijkl}$  is the median of all differences  $X_{ij\alpha} - X_{kl\beta}$  between observations in the  $ij$ th and the  $kl$ th cell, then the adjusted estimate of  $\xi_{ij} - \xi_{kl}$  is by (2.4),  $W_{ijkl} = \hat{\xi}_{ij} - \hat{\xi}_{kl} = V_{ij..} - V_{kl..}$ . This leads to the following estimates for the contrasts  $\alpha_i$ ,  $\beta_j$  and  $\gamma_{ij}$ :

$$(3.4) \quad \begin{aligned} \hat{\alpha}_i &= V_{i...} - V_{....}, & \hat{\beta}_j &= V_{.j..} - V_{....}, \\ \hat{\gamma}_{ij} &= V_{ij..} - V_{i...} - V_{.j.} + V_{....}, \end{aligned}$$

where actually  $V_{....} = 0$ .

As an illustration consider the data given by Brownlee (*Statistical Theory and Methodology in Science and Engineering*, Wiley, New York, 1960, p. 379) on the per cent reduction in blood sugar a certain time after injection of insulin into rabbits, which are shown in the table below. There are two factors involved: the dose at three levels, and the preparation of insulin at two levels,  $A$  and  $B$ . The number of observations (rabbits) per cell is four.

Dose	Preparation A			Preparation B		
	2.29	3.63	5.75	2.29	3.63	5.75
	17	64	62	33	41	56
	21	49	72	37	64	62
	49	34	61	40	34	57
	54	63	91	16	64	72

The following table shows three estimates for the difference of any two cell means  $\xi_{ij} - \xi_{kl}$ : The raw estimate  $V_{ijkl} = \text{med}(X_{ij\alpha} - X_{kl\beta})$ , the adjusted estimate  $W_{ijkl} = V_{ij..} - V_{kl..}$ , and the classical estimate  $X_{ij.} - X_{kl.}$ .

	$\xi_{11} - \xi_{12}$	$\xi_{11} - \xi_{13}$	$\xi_{11} - \xi_{21}$	$\xi_{11} - \xi_{22}$	$\xi_{11} - \xi_{23}$	$\xi_{12} - \xi_{13}$	$\xi_{12} - \xi_{21}$
V	-14.5	-40.5	7	-15	-29	-18	23.5
W	-17.33	-36.16	4.42	-15.91	-27.00	-18.83	21.75
Classical	-17.25	-36.25	3.75	-15.5	-26.5	-19	21

  

	$\xi_{12} - \xi_{22}$	$\xi_{12} - \xi_{23}$	$\xi_{13} - \xi_{21}$	$\xi_{13} - \xi_{22}$	$\xi_{13} - \xi_{23}$	$\xi_{21} - \xi_{22}$	$\xi_{21} - \xi_{23}$	$\xi_{22} - \xi_{23}$
V	0	-8	37	24	5.5	-24	-27	-11.5
W	1.42	-9.67	40.58	20.25	9.16	-20.33	-31.42	-11.09
Classical	1.75	-0.25	40	20.75	9.75	-19.25	-30.25	-11

At least in this example, adjustment has moved the estimates closer to the classical ones. This is perhaps not surprising since the adjustment puts the

estimates onto the same linear surface on which the classical estimates are lying.

From this table we obtain the estimates (3.4) of the  $\alpha$ 's,  $\beta$ 's and  $\gamma$ 's, which are shown below together with the corresponding classical estimates

	$\alpha_1 = -\alpha_2$	$\beta_1$	$\beta_2$	$\beta_3$	$\gamma_{11} = -\gamma_{21}$	$\gamma_{21} = -\gamma_{22}$	$\gamma_{13} = -\gamma_{23}$
Estimates (3.4)	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

Returning now to the general case, the following theorem gives two conditions under which the estimate  $\hat{\theta}$  of any contrast  $\theta$  is symmetrically distributed about the true value of the contrast, and hence in particular is unbiased.

**THEOREM 1.** *For any contrast  $\theta$  given by (3.1), the distribution of the estimate  $\hat{\theta}$  given by (3.2) is symmetric about  $\theta$  if either*

- (i) *the distributions  $F_i$  defined in relation to (2.1) are symmetric or*
- (ii) *all sample sizes  $n_1, \dots, n_c$  are equal.*

**PROOF.**

(i) Let  $X, Y, Z$  denote the matrices  $(X_{i\alpha}), (Y_{ij}), (Z_{ij})$  respectively and let  $Y = f(X)$  and  $Z = g(Y)$ . Then it follows from 2(i) of [2] that  $f(-X) = -f(X)$  and from Equation (2.4) of the present paper that  $g(-Y) = -g(Y)$ . Hence we have  $g[f(-X)] = -g[f(X)]$  and therefore

$$(3.5) \quad \hat{\theta}(-X) = -\hat{\theta}(X).$$

The result now follows from (3.5) as did Corollary (i) to Theorem 2 in [2] from the Equation (8.1) of that paper, with  $\Psi$  here being the rectangular distribution.

(ii) The proof of (ii) is exactly analogous, with Equation (8.3) of [2] in place of (8.1).

**4. Asymptotic distribution and efficiency.** In the remainder of the paper we shall assume that the variables  $U_{i\alpha}$  have a common distribution  $F$ . To obtain the asymptotic distribution of the adjusted estimates  $Z_{ij}$ , we begin by determining the asymptotic distribution of the raw estimates  $Y_{ij}$ . This is given by the following theorem, where the sample sizes  $n_i$  are assumed to tend to infinity in such a way that  $n_i = \rho_i N, N \rightarrow \infty$ .

**THEOREM 2.**

- (i) *The joint distribution of  $(V_1, \dots, V_{c-1})$  where*

$$(4.1) \quad V_i = N^{1/2}[Y_{ic} - (\xi_i - \xi_c)]$$

*is asymptotically normal with zero mean and covariance matrix*

$$(4.2) \quad \begin{aligned} \text{Var}(V_i) &= (1/12)(1/\rho_i + 1/\rho_c) / \left( \int f^2(x) dx \right)^2 \\ \text{Cov}(V_i, V_j) &= (1/12\rho_c) / \left( \int f^2(x) dx \right)^2. \end{aligned}$$

*Here the density  $f$  of  $F$  is assumed to satisfy the regularity conditions of Lemma 3(a) of [1].*

(ii) For any  $i$  and  $j$

$$(4.3) \quad N^{\frac{1}{2}}Y_{ij} \sim N^{\frac{1}{2}}(Y_{ic} - Y_{jc})$$

where  $\sim$  indicates that the difference of the two sides tends to zero in probability.

The proof of (i) rests on the following lemma.

LEMMA 1. Suppose the variables  $X_{i\alpha}$  have the distribution specified in connection with (2.1) with fixed  $F$  but a sequence of means  $(\xi_1, \dots, \xi_c) = (\xi_1^{(N)}, \dots, \xi_c^{(N)})$  satisfying  $\xi_i^{(N)} - \xi_c^{(N)} = -a_i/N^{\frac{1}{2}}$ . Let  $h_{ij}$  be defined by

$$(4.4) \quad n_{ij}h_{ij} = \text{Number of pairs } (\alpha, \beta) \text{ such that } X_{i\alpha} < X_{j\beta}$$

and let  $\mu_{ij} = E(h_{ij}) = P(X_{i\alpha} < X_{j\beta})$  where actually both  $h_{ij}$  and  $\mu_{ij}$  depend on  $N$ . Then the variables  $(W_1, \dots, W_{c-1})$  given by

$$(4.5) \quad W_i = N^{\frac{1}{2}}(h_{ic} - \mu_{ic}) \quad i = 1, \dots, c - 1$$

have a joint asymptotic normal distribution as  $N \rightarrow \infty$ , with zero mean and covariance matrix

$$(4.6) \quad \text{Var}(W_i) = [1/\rho_i + 1/\rho_c]/12; \quad \text{Cov}(W_i, W_j) = 1/12\rho_c.$$

The proof of this lemma is given in the appendix.

PROOF OF THEOREM 2(i). By (9.1) of [2],

$$\lim P\{N^{\frac{1}{2}}[Y_{ic} - (\xi_i - \xi_c)] \leq a_i \text{ for all } i\} = \lim P_N\{N^{\frac{1}{2}}(h_{ic} - \frac{1}{2}) \leq 0 \text{ for all } i\}$$

where  $P_N$  indicates that the probability is computed for a sequence of means satisfying (4.4). Since under the assumed regularity conditions

$$N^{\frac{1}{2}}(\mu_{ic} - \frac{1}{2}) \rightarrow -a_i \int f^2(x) dx \quad \text{as } N \rightarrow \infty$$

(see for example Lemma 3(a) of [1]), it follows that

$$\begin{aligned} \lim P\{N^{\frac{1}{2}}[Y_{ic} - (\xi_i - \xi_c)] \leq a_i \text{ for all } i\} \\ = \lim P_N\{N^{\frac{1}{2}}(h_{ic} - \mu_{ic}) \leq a_i \int f^2(x) dx \text{ for all } i\}. \end{aligned}$$

By Lemma 1, this is equal to  $Q(a_1, \dots, a_{c-1})$  where  $Q$  is the  $(c - 1)$ -dimensional multivariate normal distribution with zero mean and covariance matrix (4.2), as was to be proved.

Part (ii) of Theorem 2 can also be proved through  $U$ -statistics. Alternatively it is a simple consequence of the following lemma which was communicated to me by Professor L. LeCam.

LEMMA 2. (LeCam). Let  $(A_1^{(N)}, A_2^{(N)}, A_3^{(N)})$  be a sequence of random vectors such that for any  $1 \leq i < j \leq 3$ , the pair  $(A_i^{(N)}, A_j^{(N)})$  converges in law to a bivariate distribution with mean  $(\eta_i, \eta_j)$  and covariance matrix  $\begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{pmatrix}$ . If the  $\sigma$ 's satisfy

$$(4.7) \quad \sum_i \sigma_{ii} + 2 \sum_{i < j} \sigma_{ij} = 0,$$

then  $\sum_{i=1}^3 A_i^{(N)} \rightarrow \sum_{i=1}^3 \eta_i$  in probability.

PROOF. Since each sequence  $\{A_i^{(N)}\}$  has a non-degenerate limit distribution, there exists a sequence  $\{N_k\}$  such that the distribution of  $(A_1^{(N_k)}, A_2^{(N_k)}, A_3^{(N_k)})$  converges completely, say that  $(A_1^{(N_k)}, A_2^{(N_k)}, A_3^{(N_k)})$  tends in law to  $(A_1, A_2, A_3)$ . The covariance matrix of this limiting distribution by assumption satisfies (4.7) and hence the variance of  $\sum A_i$  is

$$\text{Var}(\sum A_i) = \sum_i \sigma_{ii} + 2\sum\sum_{i<j} \sigma_{ij} = 0,$$

so that  $\sum A_i = \sum \eta_i$  with probability one. Since this argument applies to every converging subsequence, it follows that  $\sum A_i^{(N)} \rightarrow 0$  in probability, as was to be proved.

PROOF OF THEOREM 2(ii). Let

$$A_1^{(N)} = N^{\frac{1}{2}}[Y_{ij} - (\xi_i - \xi_j)]; \quad A_2^{(N)} = N^{\frac{1}{2}}[Y_{jc} - (\xi_j - \xi_c)];$$

$$A_3^{(N)} = N^{\frac{1}{2}}[Y_{ci} - (\xi_c - \xi_i)].$$

Since each pair of the  $Y$ 's have one subscript in common, and using the fact that  $Y_{ji} = -Y_{ij}$ , it follows from part (i) of Theorem 2 that the assumptions of Lemma 2 are satisfied with

$$\sigma_{11} = \frac{1}{12} \left( \frac{1}{\rho_i} + \frac{1}{\rho_j} \right) / \left( \int f^2(x) dx \right)^2, \quad \sigma_{22} = \frac{1}{12} \left( \frac{1}{\rho_j} + \frac{1}{\rho_c} \right) / \left( \int f^2(x) dx \right)^2,$$

$$\sigma_{33} = \frac{1}{12} \left( \frac{1}{\rho_i} + \frac{1}{\rho_c} \right) / \left( \int f^2(x) dx \right)^2, \quad \sigma_{12} = -\frac{1}{12\rho_j} / \left( \int f^2(x) dx \right)^2,$$

$$\sigma_{13} = -\frac{1}{12\rho_i} / \left( \int f^2(x) dx \right)^2, \quad \sigma_{23} = -\frac{1}{12\rho_c} / \left( \int f^2(x) dx \right)^2,$$

so that (4.6) holds. Hence  $N^{\frac{1}{2}}(Y_{ij} + Y_{jc} + Y_{ci})$  tends to zero in probability, which is equivalent to (4.3).

The following corollary shows the asymptotic distribution of the adjusted estimate  $Z_{ij}$  to be the same as that of the raw estimate  $Y_{ij}$  given by Theorem 2; the corollary also establishes the statement at the end of Section 2, that the adjustments tend to become small as  $N \rightarrow \infty$ .

COROLLARY. *The difference  $N^{\frac{1}{2}}(Z_{ij} - Y_{ij})$  tends to zero in probability for all  $i, j$ .*

PROOF. By definition,  $Z_{ij} = \sum_{\alpha} (Y_{i\alpha} - Y_{j\alpha})/c$ . Using the fact that by Theorem 2(ii),  $N^{\frac{1}{2}}Y_{j\alpha} \sim N^{\frac{1}{2}}(Y_{jc} - Y_{\alpha c})$  and  $N^{\frac{1}{2}}Y_{i\alpha} \sim N^{\frac{1}{2}}(Y_{ic} - Y_{\alpha c})$ , we see that  $N^{\frac{1}{2}}Z_{ij} \sim N^{\frac{1}{2}}(Y_{ic} - Y_{jc}) \sim N^{\frac{1}{2}}Y_{ij}$ .

We are now in a position to prove that the asymptotic efficiency of the estimate  $\hat{\theta}$  of a contrast  $\theta$  given by (3.2) and (3.1), relative to the classical least squares estimate of  $\theta$  is the same as the Pitman efficiency of the Wilcoxon test relative to the  $t$ -test.

THEOREM 3. *The asymptotic efficiency of the estimate  $\hat{\theta} = \sum \sum d_{ij}Z_{ij}$  of  $\theta = \sum \sum d_{ij}(\xi_i - \xi_j)$  relative to the estimate  $\sum \sum d_{ij}(X_i - X_j)$  is*

$$(4.8) \quad e = 12\sigma^2[\int f^2(x) dx]^2$$

where  $\sigma^2 = \text{Var}(X_{i\alpha})$ .

PROOF. Let  $T_{ij} = X_i - X_j$  and let  $V'_i = N^{1/2}[T_{ic} - (\xi_i - \xi_c)]$ . Then the variables  $(V'_1, \dots, V'_{c-1})$  have an asymptotic normal distribution with zero mean and covariance matrix

$$\text{Var}(V'_i) = \sigma^2(1/\rho_i + 1/\rho_c), \quad \text{Cov}(V_i, V_j) = \sigma^2/\rho_c.$$

Furthermore,  $T_{ij} = T_{ic} - T_{jc}$ . Comparing this with the asymptotic distribution of the  $Z_{ij}$ , which is given by Theorem 2 with  $Y_{ij}$  replaced by  $Z_{ij}$  throughout, the result follows.

So far, the means  $\xi_1, \dots, \xi_c$  were assumed to be completely unrestricted. However, in more general linear models, they may be required to satisfy certain linear restrictions. These may then affect the classical estimate. However, whenever this estimate is a linear function of the differences  $X_i - X_j$ , then the estimate obtained from it by replacing  $X_i - X_j$  by  $Y_i - Y_j$  has asymptotic relative efficiency (4.1). This is an immediate consequence of Theorem 2.

As an example suppose that  $E(X_{i\alpha}) = \xi_i$  ( $i = 1, \dots, 4$ ) where  $\xi_1 = \mu + \gamma$ ,  $\xi_2 = \mu + \delta$ ,  $\xi_3 = \mu + \gamma + \delta$ ,  $\xi_4 = \mu - \gamma - \delta$ . Then the classical estimate of  $\gamma$  is

$$\tilde{\gamma} = (10X_1 - 8X_2 + 3X_3 - 5X_4)/18,$$

which is a linear function of the differences  $X_i - X_j$  since the sum of the coefficients is zero. Replacing  $X_i - X_j$  by  $Y_i - Y_j$ , we obtain the estimate

$$\begin{aligned} \hat{\gamma} &= (10Y_1 - 8Y_2 + 3Y_3 - 5Y_4)/18 \\ &= (18Y_{12} + 7Y_{13} + 15Y_{14} - 11Y_{23} - 3Y_{24} + 8Y_{34})/72 \end{aligned}$$

and the asymptotic efficiency of  $\hat{\gamma}$  relative to  $\tilde{\gamma}$  is given by (4.8).

**5. An alternative approach.** There exists another class of estimates of the parameters in a linear model, not restricted to contrasts and not suffering from the peculiarity discussed at the end of Section 2. However, these estimates appear to have the efficiency of those treated in the main body of this paper only when the underlying distribution  $F$  is symmetric.

Consider once more the  $c$ -sample model (2.1). In [2] it is shown that the estimate

$$(5.1) \quad \xi_i^* = \text{med} [(X_{i\alpha} + X_{i\beta})/2]$$

has asymptotic efficiency (4.8) relative to the standard estimate  $X_i$  as  $n_i \rightarrow \infty$  provided the distribution  $F$  of the  $X$ 's is symmetric. Hence the estimate  $\sum c_i \xi_i^*$  of  $\sum c_i \xi_i$  has asymptotic efficiency (4.8) relative to the standard estimate  $\sum c_i X_i$  when  $F$  is symmetric.

As an example in which the method of Sections 2-4 is not applicable consider a weighing design, in which the sum and the difference of two objects  $A$  and  $B$  is each weighed  $n$  times, so that  $E(X_{1\alpha}) = \xi + \eta$ ,  $E(X_{2\alpha}) = \eta - \xi$  for  $\alpha = 1, \dots, n$ , and suppose that the distribution  $F$  of the  $X$ 's is symmetric. Then  $\text{med} [(X_{1\alpha} + X_{1\beta})/2]$  and  $\text{med} [(X_{2\alpha} + X_{2\beta})/2]$  are independent estimates of  $\xi + \eta$  and  $\eta - \xi$  with asymptotic efficiency (4.8) relative to  $X_1$  and  $X_2$ . Hence

$$\frac{1}{2}\{\text{med} [(X_{1\alpha} + X_{1\beta})/2] + \text{med} [(X_{2\alpha} + X_{2\beta})/2]\}$$

and

$$\frac{1}{2}\{\text{med} [(X_{1\alpha} + X_{1\beta})/2] - \text{med} [(X_{2\alpha} + X_{2\beta})/2]\}$$

are estimates of  $\eta$  and  $\xi$  having asymptotic efficiency (4.8) relative to the standard estimates  $\frac{1}{2}(X_{1.} + X_{2.})$  and  $\frac{1}{2}(X_{1.} - X_{2.})$ .

**6. Appendix: generalized  $U$ -statistics.** The following theorem which is a slight extension of Hoeffding's theorem on  $U$ -statistics [3], is known to many workers in nonparametric inference. It is stated here solely for the purpose of convenient reference; the proof exactly parallels Hoeffding's argument. (Part (i) for  $s = 2$  is stated in Fraser: *Nonparametric Methods in Statistics*, p. 229.)

(i) Let  $\varphi(x_1, \dots, x_{m_1}; y_1, \dots, y_{m_2}; z_1, \dots, z_{m_3}; \dots)$  be a function symmetric in the  $x$ 's, in the  $y$ 's, in the  $z$ 's,  $\dots$ . Let  $X_1, \dots, X_{n_1}; Y_1, \dots, Y_{n_2}; Z_1, \dots, Z_{n_3}; \dots$  be samples of sizes  $n_i = \rho_i N$  from distributions  $F, G, H, \dots$  and consider the generalized  $U$ -statistic

$$U = \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \binom{n_3}{m_3}^{-1} \sum \varphi(X_{\alpha_1}, \dots, X_{\alpha_{m_1}}; Y_{\beta_1}, \dots, Y_{\beta_{m_2}}; Z_{\gamma_1}, \dots, Z_{\gamma_{m_3}}; \dots)$$

where the summation extends over all

$$\binom{n_1}{m_1} \binom{n_2}{m_2} \binom{n_3}{m_3} \dots$$

sets of subscripts satisfying  $1 \leq \alpha_1 < \alpha_2 < \dots < \alpha_{m_1} \leq n_1; 1 \leq \beta_1 < \beta_2 < \dots < \beta_{m_2} \leq n_2; \dots$ . Let  $E(U) = \theta(F, G, H, \dots)$ . Then as  $N \rightarrow \infty, Z = N^{1/2}(U - \theta)$  tends in law to  $N(0, \sigma^2)$  with

$$\sigma^2 = (m_1^2/\rho_1)\zeta_{10\dots 0} + (m_2^2/\rho_2)\zeta_{010\dots 0} + \dots + (m_s^2/\rho_s)\zeta_{00\dots 1}.$$

Here  $\zeta_{c_1, \dots, c_s}$  is the covariance of

$$\varphi(X_1, \dots, X_{c_1}, X_{c_1+1}, \dots, X_{m_1}; Y_1, \dots, Y_{c_2}, Y_{c_2+1}, \dots, Y_{m_2}; \dots)$$

with  $\varphi(X_1, \dots, X_{c_1}, X'_{c_1+1}, \dots, X'_{m_1}; Y_1, \dots, Y_{c_2}, Y'_{c_2+1}, \dots, Y'_{m_2}; \dots)$  where  $X_1, \dots, X_{m_1}, X'_{c_1+1}, \dots, X'_{m_1}$  are independently distributed according to  $F; Y_1, \dots, Y_{m_2}, Y'_{c_2+1}, \dots, Y'_{m_2}$  according to  $G$ ; etc.

(ii) Let  $U^{(i)}$  ( $i = 1, \dots, r$ ) be  $r$  generalized  $U$ -statistics defined in terms of functions  $\varphi^{(i)}$ , where the numbers  $m_1, m_2, \dots$  may depend on  $i$  but not the sample sizes  $n_1, n_2, \dots$ . The joint distribution of the variables  $Z^{(1)} = N^{1/2}(U^{(1)} - \theta^{(1)})$ ,  $\dots, Z^{(r)} = N^{1/2}(U^{(r)} - \theta^{(r)})$  is asymptotically normal, with mean zero and covariance matrix  $\Sigma = (\sigma_{ij})$  where

$$\sigma_{ij} = (m_1^{(i)} m_1^{(j)} / \rho_1) \zeta_{10\dots 0}^{(i,j)} + \dots + (m_s^{(i)} m_s^{(j)} / \rho_s) \zeta_{00\dots 1}^{(i,j)}.$$

Here  $\zeta_{c_1, \dots, c_s}^{(i,j)}$  is the covariance of

$$\varphi^{(i)}(X_1, \dots, X_{c_1}, X_{c_1+1}, \dots, X_{m_1}; Y_1, \dots, Y_{c_2}, Y_{c_2+1}, \dots, Y_{m_2}; \dots)$$



with

$$\varphi^{(j)}(X_1, \dots, X_{c_1}, X'_{c_1+1}, \dots, X'_{m_1}; Y_1, \dots, Y_{c_2}, Y'_{c_2+1}, \dots, Y'_{m_2}; \dots).$$

(iii) Suppose that the distributions  $F, G, H, \dots$  depend on  $N$ . Let  $\theta_N^{(s)} = E_N[U^{(s)}]$  and let  $\Sigma_N$  denote the covariance matrix  $\Sigma$  computed for  $(F_N, G_N, H_N, \dots)$ . Then if  $\Sigma_N$  tends to a non-singular covariance matrix  $\Sigma$  as  $N \rightarrow \infty$ , it remains true that

$$N^{\frac{1}{2}}(U^{(1)} - \theta_N^{(1)}), \dots, N^{\frac{1}{2}}(U^{(r)} - \theta_N^{(r)})$$

tend in law to a normal distribution with zero mean and covariance matrix  $\Sigma$ .

As an application, we now prove Lemma 1. It will be convenient to use the present notation rather than that of Section 4.

For simplicity of notation consider the case of  $c = 3$  samples  $X_1, \dots, X_{n_1}; Y_1, \dots, Y_{n_2}; Z_1, \dots, Z_{n_3}$ . Let

$$\varphi^{(1)}(x_1, y_1, z_1) = \begin{cases} 1 & \text{if } x_1 < z_1 \\ 0 & \text{otherwise} \end{cases}$$

$$\varphi^{(2)}(x_1, y_1, z_1) = \begin{cases} 1 & \text{if } y_1 < z_1 \\ 0 & \text{otherwise} \end{cases}$$

and consider a sequence of distributions

$$F_N(x) = F(x - a_1N^{-\frac{1}{2}}), \quad G_N(y) = F(y - a_2N^{-\frac{1}{2}}), \quad H_N(z) = F(z)$$

for the variables  $X_\alpha, Y_\beta$  and  $Z_\gamma$ . Then  $n_1n_3U^{(1)} =$  Number of pairs  $(\alpha, \gamma)$  such that  $X_\alpha < Z_\gamma$  and  $n_2n_3U^{(2)} =$  Number of pairs  $(\beta, \gamma)$  such that  $Y_\beta < Z_\gamma$ .

To apply (iii), it is only necessary to compute the limiting covariance matrix  $\Sigma$ .

Now

$$\begin{aligned} \zeta_{100}^{(1)} &= E\varphi^{(1)}(X_1, Y_1, Z_1)\varphi^{(1)}(X_1, Y'_1, Z'_1) - [P(X_1 < Z_1)]^2 \\ &= P\{X_1 < Z_1, Z'_1\} - [P(X_1 < Z_1)]^2 \rightarrow \frac{1}{3} - \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

Similarly  $\zeta_{010}^{(1)} \rightarrow \frac{1}{12}$  while  $\zeta_{001}^{(1)} = 0$ . Thus  $\sigma_{11} = (1/\rho_1 + 1/\rho_3)/12$ . Exactly similarly,  $\sigma_{22} = (1/\rho_2 + 1/\rho_3)/12$ . Finally,

$$\zeta_{100}^{(1,2)} = P\{X_1 < Z_1 \text{ and } Y'_1 < Z'_1\} - P(X_1 < Z_1)P(Y_1 < Z_1) = 0 = \zeta_{010}^{(1,2)}$$

while

$$\zeta_{001}^{(1,2)} = P\{X_1 < Z_1 \text{ and } Y'_1 < Z'_1\} - P(X_1 < Z_1)P(Y_1 < Z_1) \rightarrow \frac{1}{12}$$

so that  $\sigma_{12} = 1/12\rho_3$ . This verifies the covariance matrix (4.6) and completes the proof of Lemma 1.

REFERENCES

[1] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1 307-318.

- [2] HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598-611.
- [3] Hoeffding, Wassily (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293-325.
- [4] LEHMANN, E. L. (1957). A theory of some multiple decision problems, I. *Ann. Math. Statist.* **28** 1-25.