Statistical Research Report
Institute of Mathematics
University of Oslo

No. 6
September 1967

A NOTE ON ROBUST ESTIMATION IN ANALYSIS OF VARIANCE

By

Emil Spjøtvoll

1. INTRODUCTION.

Consider the c-sample model, in which the observations are

(1.1) $$X_{i\alpha} = \xi_i + U_{i\alpha} \qquad \alpha = 1,2,\dots,n_i, \quad i = 1,2,\dots,c,$$

where the variables $U_{i\alpha}$ are independently distributed with cumulative distribution function F. Let

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

be the median of the $n_i \; n_j$ differences $X_{i\alpha} - X_{j\beta}$ ($\alpha = 1,2,\dots,n_i$, $\beta = 1,2,\dots,n_j$). It has been shown by the Hodges and Lehmann [2] that the estimate Y_{ij} of $\xi_i - \xi_j$ has more robust efficiency than the standard estimate $T_{ij} = X_{i.} - X_{j.}$ where $X_{i.} = \sum X_{i\alpha} / n_i$.

The estimates Y_{ij} do not satisfy the linear relations satisfied by the differences they estimate. To remedy this, the raw estimates Y_{ij} were by Lehmann [3] replaced by adjusted estimates Z_{ij} of the form $\hat{\xi}_i - \hat{\xi}_j$. This was done by minimizing the sum of squares

(1.3) $$\sum_{i \neq j} (Y_{ij} - (\xi_i - \xi_j))^2$$

giving (see [2])

(1.4) $$Z_{ij} = Y_{i.} - Y_{j.}$$

where $Y_{i.} = (1/c) \sum Y_{ij}$ and where Y_{ii} is defined to be zero for all i.

The purpose of this note is to argue that in the sum of squares (1.3) there should be used weights according to the number of observations on which the different Y_{ij} are based.

For purpose of reference we state a theorem of Lehmann. Let the sample sizes n_i tend to infinity in such a way that $n_i = \rho_i N (N = \sum n_i)$. Then we have the following theorem (Theorem 2 of [3]).

THEOREM 1.

(i) The joint distribution of $(V_1, V_2, \ldots, V_{c-1})$ where $V_i = N^{\frac{1}{2}} (Y_{ic} - (\xi_i - \xi_c))$ is asymptotically normal with zero mean and covariance matrix

$$\mathrm{Var}\,(V_i) = (1/12)(1/\rho_i + 1/\rho_c) \,/\, (\int f^2(x)dx)^2$$
$$\mathrm{Cov}(V_i, V_j) = (1/12\rho c)/(\int f^2(x)dx)^2 .$$

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

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

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

2. WEIGTED ESTIMATES.

Define the $\frac{1}{2}c$ (c-1)- component vector $Y' = [Y_{12}, Y_{13}, \ldots, Y_{c-1,c}]$. Denote the covariance matrix of Y by A. Suppose that $E\,Y_{ij} = \xi_i - \xi_j$ (conditions under which this holds or approximately holds are given in [2]) for all i and j . To estimate the differences $\xi_i - \xi_j$ can then be treated as an ordinary regression problem. The minimum variance unbiased linear estimates of the $\xi_i - \xi_j$ are obtained by minimizing

(2.1) $\sum a^{ij,kl}(Y_{ij} - (\xi_i - \xi_j))(Y_{kl} - (\xi_k - \xi_l))$

where $a^{ij,kl}$ denote the elements of A^{-1}.

Since from Theorem 1 asymptotically $E\,Y_{ij} = \xi_i - \xi_j$ it seems reasonable to minimize (2.1) even if the Y_{ij} are not exactly unbiased estimates of $\xi_i - \xi_j$ for finite N.

Unfortunately the elements of A are unknown. But suppose we use an arbitrary matrix W with elements $w_{ij,kl}$ such that we shall minimize

(2.2) $\sum w_{ij,kl}(Y_{ij} - (\xi_i - \xi_j))(Y_{kl} - (\xi_k - \xi_l))$.

Let Y^*_{ij} denote the minimizing value of $\xi_i - \xi_j$ in (2.2). We shall study the asymptotic distribution of the Y^*_{ij}. We shall allow the matrix W to vary with the number of observations, and use the notation $W(n_1, n_2, \ldots, n_c) = W_N$. Let the n_i tend to infinity as in Theorem 1.

<u>T H E O R E M 2.</u>

For any sequence $\{W_N\}$ of matrices of rank $\geq$ c-1 converging to a matrix W_0 of rank $\geq$ c-1 , asymptotically for any i and j

$$\sqrt{N}\,(Y^*_{ij} - Y_{ij}) \sim 0$$

<u>P R O O F.</u> To get a full rank regression problem we introduce the parameters

(2.3) $\theta_i = \xi_i - \xi_c \qquad i = 1,2,\ldots,c-1.$

Then

(2.4) $\xi_i - \xi_j = \theta_i - \theta_j$.

Let B denote the design matrix such that (2.2) can be written

(2.5) $(Y - B\,\Theta)'\ W_N\ (Y - B\,\Theta)$

where $\Theta' = (\Theta_1, \Theta_2, \ldots, \Theta_{c-1})$

The value of Θ minimizing (2.5) is

$$\hat{\Theta}_N = (B'\,W_N\,B)^{-1}\,B'\,W_N\,Y.$$

Define Y_1 by $Y_1' = (Y_{1c}, \ldots, Y_{c-1,c})$.

By Theorem 1(i)

(2.6) $N^{\frac{1}{2}}\,(Y - B\,Y_1) \sim 0.$

We have

(2.7) $N^{\frac{1}{2}}(\hat{\Theta}_N - Y_1) = \left[(B'W_NB)^{-1}B'W_N - (B'W_oB)^{-1}B'W_o\right] N^{\frac{1}{2}}(Y - BY_1) + (B'W_oB)^{-1}BW_o$ $B'W_o\ N^{\frac{1}{2}}\ (Y - BY_1)$.

By (2.6)and the continuity of the second function and the uniform convergence in any closed interval of the first function on the right hand side of (2.7) it follows that

$$N^{\frac{1}{2}}(\hat{\Theta}_N - BY_1) \sim 0.$$

Hence $N^{\frac{1}{2}}(Y^*_{ic} - Y_{ic}) \sim 0$ for any i. By Theorem 1(ii) and the fact that $Y^*_{ij} = Y^*_{ic} - Y^*_{jc}$ it follows that $N^{\frac{1}{2}}(Y^*_{ij} - Y_{ij}) \sim 0$ for any i and j. The theorem is proved.

It is seen from the above theorem that the asymptotic distribution of the estimates does not depend on the matrices W_N. Hence the asymptotic distribution will be the same as for the best linear estimates. (The solution of (2.1)). In particular this is true for the estimates Z_{ij} given by Lehmann.

But , of course, the best unbiased linear estimates will give better estimates than the Z_{ij} for finite N. Since A is unknown we cannot find the former. By Theorem 1(i) the asymptotic value of A is known, but it is singular and cannot be used in (2.1).

We now propose to use the asymptotic variances of the Y_{ij} as weights i.e. we want to minimize

$$(2.8) \quad Q = \sum (1/n_i + 1/n_j)^{-1}(Y_{ij} - (\xi_i - \xi_j))^2 \text{ with respect to } \xi_i - \xi_j .$$

We introduce (2.3) and (2.4) in (2.8). After derivation of (2.8) with respect to the θ_i it is found that the minimizing values are given by the solutions of the equations

$$(2.9) \quad \hat{\theta}_\alpha \Big(\sum_{i \neq \alpha} \frac{n_i}{n_i + n_\alpha}\Big) - \sum_{i \neq \alpha, c} \frac{n_i}{n_i + n_\alpha} \hat{\theta}_i = \sum \frac{n_i}{n_i + n_i} Y_{\alpha i}$$

$$\alpha = 1,2,\ldots,c-1.$$

It does not seem easy to find an explicit algebraic solution of (2.9), though for each specific set of the n_i we can solve (2.9), if necessary with the aid of an electronic computer.

It follows from Theorem 2 that the asymptotic distribution of the $\hat{\theta}_i - \hat{\theta}_j$ is equal to the asymptotic distribution of the estimates Z_{ij} and hence the same is true regarding asymptotic efficiencies.

We now proceed to prove that in some respects the estimates $\hat{\theta}_i - \hat{\theta}_j$ is better than the Z_{ij}. Let D be a subset of the integers 1,2,...,c. Suppose that $n_i \to \rho_i N$ as $N \to \infty$ when $i \in D$ while $n_i/N \to 0$ when $i \notin D$. We shall study the asymptotic distribution of the estimates in this case. Without loss of generality we may assume $D = \{1,2,\ldots,b\}$ for some $b < c$.

THEOREM 3.

Suppose that $N \to \infty$ such that $n_i \to \rho_i N$ when $i=1,2,\ldots,b$ $(\sum_1^b \rho_i = 1)$. Then the asymptotic distribution ofthe $\hat{\theta}_i - \hat{\theta}_j$ of (2.9) for $i,j \leq b$ is equal to the asymptotic distribution of the Y_{ij} in Theorem 1 when c is replaced by b.

P R O O F. Q can be written

$$Q = \sum_{\max(i,j)\leq b} (1/n_i + 1/n_j)^{-1}(Y_{ij}-(\xi_i-\xi_j))^2$$

(2.1o)

$$+ \sum_{\max(i,j)>b} (n_i/N)(n_j/N)(n_i/N + n_j/N)^{-1} (Y_{ij} -(\xi_i - \xi_j))^2.$$

By assumption the last expression on the right hand side of (2.1o) tends to zero when $N \to \infty$. Hence

$$Q \sim \sum_{\max(i,j)\leq b} (1/n_i+1/n_j)^{-1}(Y_{ij}-(\xi_i-\xi_j))^2$$

which is of the form (2.8) with c replaced by b. The theorem now easily follows since the same results holds for the Y_{ij} with $i,j \leq b$ as for the Y_{ij} with $i,j \leq c$.

In [3] is given an example which shows that the estimate Z_{12} of $\xi_1 - \xi_2$ is not consistent when n_1 and n_2 tends to infinity unless also n_3 tends to infinity (c=3). Theorem 3 proves that the new estimates $\hat{\theta}_i - \hat{\theta}_j$ do not have this definciency. If for the same i and j n_i and n_j tends to infinity then $\hat{\theta}_i - \hat{\theta}_j$ is a consistent estimate of $\xi_i - \xi_j$.

3. AN ALTERNATIVE ESTIMATE.

Since the estimates $\theta_i - \theta_j$ of (2.9) is not easily computed unless one have access to an electronic computor, we shall give alternative simpler estimates which also are weighted estimates.

We shall minimize

$$(3.1) \quad \sum n_i n_j (Y_{ij} -(\xi_i - \xi_j))^2 .$$

By differentiation it is easily found that the values of $\hat{\xi}_i - \hat{\xi}_j$ minimizing (3.1) is

(3.2) $W_{ij} = \sum n_\alpha (Y_{i\alpha} - Y_{j\alpha}) = \tilde{Y}_i - \tilde{Y}_j$

where we have introdused the weighted differences $\tilde{Y}_i = (\sum n_\alpha)^{-1} \sum n_\alpha Y_{i\alpha}$. Compare (1.4).

It follows from Theorem 2 that the estimates W_{ij} have the same asymptotic properties as the Z_{ij} and $\hat{\theta}_i - \hat{\theta}_j$. Furthermore it is easily seen that Theorem 3 holds for the W_{ij}.

4. THE CASE $n_1 = n_2 = \dots = n_c$.

When $n_1 = n_2 = \dots = n_c$ both $\hat{\theta}_i - \hat{\theta}_j$ and W_{ij} reduce to Z_{ij}. Further we have

THEOREM 4. If $n_1 = n_2 = \dots = n_c = n$ then the estimates Z_{ij} are the minimum variance unbiased linear estimates.

PROOF. Define σ^2 and a by

(4.1)
$$\sigma^2 = \text{Var } Y_{12}$$
$$a\sigma^2 = \text{Cov } (Y_{12}, Y_{13}).$$

Note that both σ^2 and a depend upon n and F. By symmetry we have

(4.2)
$$\begin{aligned}
&\text{Var } Y_{ij} &&= \sigma^2 && i \neq j. \\
&\text{Cov } (Y_{ij}, Y_{kl}) &&= 0 && i \neq k,\ i \neq l,\ j \neq k,\ j \neq l. \\
&\text{Cov } (Y_{ij}, Y_{il}) &&= a\sigma^2 && j \neq l,\ i \neq j,\ i \neq l. \\
&\text{Cov } (Y_{ij}, Y_{kj}) &&= a\sigma^2 && i \neq k,\ i \neq j,\ k \neq j. \\
&\text{Cov } (Y_{ij}, Y_{jl}) &&= -a\sigma^2 && i \neq l,\ i \neq j,\ j \neq l, \\
&\text{Cov } (Y_{ij}, Y_{ki}) &&= -a\sigma^2 && k \neq j,\ i \neq j,\ k \neq i
\end{aligned}$$

Let the covariance matrix of Y given by (4.2) be $G(a)\sigma^2$. It can be verified that the inverse of $G(a)$ is $(1+2(c-2)da)^{-1}\,G(d)$ where $d = -\,a(1+(c-4)a)^{-1}$. Hence

(2.1) is proportional to

$$Q_1 = \sum_j \sum_{i=1}^{j-1} (Y_{ij} - (\xi_i - \xi_j))^2$$

$$+ 2d \sum_j \sum_{i=1}^{j-1} \sum_{h=i+1}^{j-1} (Y_{ij} - (\xi_i - \xi_j))(Y_{hj} - (\xi_h - \xi_j))$$

$$+ 2d \sum_j \sum_{i=1}^{j-1} \sum_{h=1}^{j-1} (Y_{ij} - (\xi_i - \xi_j))(Y_{ih} - (\xi_i - \xi_h)) .$$

The part of Q_1 involving ξ_α can be written

$$\sum (Y_i - (\xi_i - \xi_\alpha))^2$$

$$+ 2d \sum_j \sum_{i \neq \alpha} (Y_{j\alpha} - (\xi_j - \xi_\alpha))(Y_{ji} - (\xi_j - \xi_i))$$

$$+ 2d \sum_j \sum_{i=1}^{j-1} (Y_{j\alpha} - (\xi_j - \xi_\alpha))(Y_{i\alpha} - (\xi_i - \xi_\alpha)) .$$

We find

$$\frac{\partial Q_1}{\partial \xi_\alpha} = 2\,(1+d(c-1))(c\,\xi_\alpha - \sum \xi_i - \sum_i Y_{\alpha i}).$$

Hence the minimizing values $\hat{\xi}_i$ satisfy

$$\hat{\xi}_\alpha = c^{-1} \sum \hat{\xi}_i + c^{-1} \sum_i Y_{\alpha i}$$

and hence by (1.4)

$$\hat{\xi}_i - \hat{\xi}_j - c^{-1} \sum_i Y_{\alpha i} - c^{-1} \sum_j Y_{\alpha j} = Z_{ij} .$$

5. An example.

In this section the estimates are compared on an example taken from Scheffé:" The Analysis of Variance". p. 140. It is a two-way layout with factors genotype of the foster mother and that of the litter. The observations are weights (average) of the litter. Let A_1, A_2, A_3, A_4 and B_1, B_2, B_3, B_4 denote the different genotypes of the foster mother and the litter respectively. The observatoons are:

Table 1.

B_1				B_2			
A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
61.5	55.0	52.5	42.0	60.3	50.8	56.5	51.3
68.2	42.0	61.8	54.0	51.7	64.7	59.0	40.5
64.0	60.2	49.5	61.0	49.3	61.7	47.2	
65.0		52.7	48.2	48.0	64.0	53 0	
59.7			39.6		62.0		

B_3				B_4			
A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
37.0	56.3	39.7	50.0	59.0	59.5	45.2	44.8
36.3	69.8	46.0	43.8	57.4	52.8	57.0	51.5
68.0	67.0	61.3	54.5	54.0	56.0	61.4	53.0
		55.3		47.0			42.0
		55.7					54.0

Let ξ_{ij} denote the expectation of the variables from (A_i, B_j). In Table 2 are given the estimates of the differences $\hat{\xi}_{ij}-\hat{\xi}_{44}$ obtained by the different methods.

Table 2 .

	$\xi_{11}-\xi_{44}$	$\xi_{21}-\xi_{44}$	$\xi_{31}-\xi_{44}$	$\xi_{41}-\xi_{44}$	$\xi_{12}-\xi_{44}$	$\xi_{22}-\xi_{44}$	$\xi_{32}-\xi_{44}$	$\xi_{42}-\xi_{44}$
Y	14.2	3.5	6.1	0.0	3.9	10.7	5.0	- 2.2
Z	14.35	3.86	4.65	0.14	3.14	12.0	4.82	- 2.87
W	14.29	3.88	4.57	0.06	3.01	11.95	4.83	- 2.88
Θ	14.31	3.87	4.60	0.09	3.07	11.98	4.82	- 2.89
Classical	14.62	3.34	5.07	- 0.10	3.27	11.58	4.87	- 3.16

	$\xi_{13}-\xi_{44}$	$\xi_{23}-\xi_{44}$	$\xi_{33}-\xi_{44}$	$\xi_{43}-\xi_{44}$	$\xi_{14}-\xi_{44}$	$\xi_{24}-\xi_{44}$	$\xi_{34}-\xi_{44}$
Y	- 7.8	15.5	2.7	0.5	5.0	6.5	5.5
Z	- 7.50	15.48	3.11	0.31	5.19	6.68	5.74
W	- 7.84	15.51	3.12	0.33	5.21	6.68	5.73
Θ	- 7.66	15.49	3.12	0.31	5.21	6.67	5.74
Classical	- 1.96	15.31	2.54	0.37	5.29	7.04	5.47

From Table 2 the estimates of any $\xi_{ij} - \xi_{kl}$ can be found. It is seen that in this example the estimates Z, W and Θ do not differ much. The estimates Θ tend to lie between Z and W. In the example the sample sizes vary from 2 to 6 while c=16. The results seem to indicate that for such a small variation of sample sizes relative to the value of c, the weighted estimates will not much change the estimates Z.

To see the effect for smaller c when the variation from 2 to 6 of sample sizes seems more important, we select the factor combinations (A_1,B_1), (A_4,B_1) and (A_4,B_2). Then we have c=3 and sample sizes 6,6 and 2. The estimates are given in Table 3.

Table 3.

	$\xi_{11} - \xi_{41}$	$\xi_{11} - \xi_{42}$	$\xi_{41} - \xi_{42}$
Y	15.8	18,05	3.1
Z	15.52	18.3	2.78
W	15.66	18.40	2.74
θ	15.61	18.38	2.77
Classical	14.72	17.78	3.06

It is seen that the estimate of $\xi_{11} - \xi_{41}$ based on W and θ are closer to the original estimate Y than the estimate based on Z. This is as should be expected since there are 6 + 6 observations behind the estimate of the difference $\xi_{11} - \xi_{41}$, while there are 6 + 2 observations behind the other estimates.

R E F E R E N C E S

[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. Univ. of California Press.

[2] HODGES, J.L.,JR. and LEHMANN, E.L. (1963).
Estimation of location based on ranks. Ann.Math. Statist. 34 598-691.

[3] LEHMANN, E.L. (1963). Robust estimation in analysis of variance. Ann. Math. Statist. 34 957-966.