A NOTE ON ROBUST

ESTIMATION IN ANALYSIS OF VARIANCE

By

Emil Spjøtvoll
1. INTRODUCTION.

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

\[(1.1) \quad x_{i\alpha} = \xi_i + \bigcup_{i\alpha} = 1,2,\ldots,n_i \quad i = 1,2,\ldots,c,\]

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

\[(1.2) \quad y_{ij} = \text{med} (x_{i\alpha} - x_j)\]

be the median of the $n_i \times n_j$ differences $x_{i\alpha} - x_j$ ($\alpha = 1,2,\ldots,n_i, \beta = 1,2,\ldots,n_j$). 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

\[(1.3) \quad \sum_{i\neq j} (y_{ij} - (\xi_i - \xi_j))^2\]

giving (see [2])

\[(1.4) \quad z_{ij} = y_i - y_j,\]

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 the purpose of reference we state a theorem of Lehmann. Let the sample sizes $n_i$ tend to infinity in such a way that $n_i = \xi^i N(\sum n_i)$. 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

$$V_i = N^{\frac{1}{2}} (Y_{ic} - (\hat{\xi}_i - \hat{\xi}_c))$$

is asymptotically normal with zero mean and covariance matrix

$$\text{Var}(V_i) = (1/12)(1/\xi_i + 1/\xi_c) / (\int f^2(x)dx)^2$$

$$\text{Cov}(V_i, V_j) = (1/12 \xi_c) / (\int f^2(x)dx)^2.$$

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. Weighed Estimates.

Define the $\xi_c$ ($c-1$)-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
where \( a_{ij,kl} \) denote the elements of \( A^{-1} \).

Since from Theorem 1 asymptotically \( E Y_{ij} = \hat{\delta}_i - \hat{\delta}_j \) it seems reasonable to minimize (2.1) even if the \( Y_{ij} \) are not exactly unbiased estimates of \( \hat{\delta}_i - \hat{\delta}_j \) 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

\[
(2.2) \sum w_{ij,kl}(Y_{ij} - (\hat{\delta}_i - \hat{\delta}_j))(Y_{kl} - (\hat{\delta}_k - \hat{\delta}_l)).
\]

Let \( Y_{ij}^* \) denote the minimizing value of \( \hat{\delta}_i - \hat{\delta}_j \) 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.

**THEOREM 2.**

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

\[
\sqrt{N}(Y_{ij}^* - Y_{ij}) \sim 0
\]

**PROOF.** To get a full rank regression problem we introduce the parameters

\[
(2.3) \quad Q_i = \hat{\delta}_i - \hat{\delta}_c \quad i = 1, 2, \ldots, c-1.
\]

Then

\[
(2.4) \quad \hat{\delta}_i - \hat{\delta}_j = \Theta_i - \Theta_j.
\]

Let \( B \) denote the design matrix such that (2.2) can be written
The value of \( \Theta \) minimizing (2.5) is

\[
\hat{\Theta}_N = (B'W_NB)^{-1}B'W_NY.
\]

Define \( Y_1 \) by

\[
Y_1 = (Y_{1c}, \ldots, Y_{c-1}, c).
\]

By Theorem 1(i)

(2.6) \( N^{\frac{3}{2}}(Y - B_1Y_1) \sim 0. \)

We have

(2.7) \( N^{\frac{1}{2}}(\hat{\Theta}_N - Y_1) = \left[ (B'W_NB)^{-1}(B'W_NB)^{-1}B'W_0 \right] N^{\frac{3}{2}}(Y - BY_1) + (B'W_0B)^{-1}BW_0B'W_0N^{\frac{1}{2}}(Y - BY_1). \)

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

\( N^{\frac{3}{2}}(\hat{\Theta}_N - BY_1) \sim 0. \)

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

$$Q = \sum (1/n_i + 1/n_j)^{-1}(Y_{ij} - (\xi_i - \xi_j))^2$$

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 $\xi_i$ it is found that the minimizing values are given by the solutions of the equations

$$\hat{\xi} = \frac{\sum n_i \xi_i}{\sum n_i}$$

for $1 \leq j \leq c - 1$.

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 $\hat{\xi}_i - \hat{\xi}_j$ 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 $\hat{\xi}_i - \hat{\xi}_j$ is better than the $Z_{ij}$. Let $D$ be a subset of the integers $1, 2, \ldots, c$. Suppose that $n_i \rightarrow P_i N$ as $N \rightarrow \infty$ when $i \in D$ while $n_i/N \rightarrow 0$ 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 $N \rightarrow \infty$ such that $n_i \rightarrow P_i N$ when $i=1, 2, \ldots, b$ and $\sum P_i = 1$. Then the asymptotic distribution of the $\hat{\xi}_i - \hat{\xi}_j$ 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

$$Q = \sum_{\max(i,j) \neq b} (1/n_i + 1/n_j)^{-1}(Y_{ij} - (\bar{x}_i - \bar{x}_j))^2$$

(2.10)

$$+ \sum_{\max(i,j) \neq b} (n_i/N)(n_j/N)(n_i/N + n_j/N)^{-1}(Y_{ij} - (\bar{x}_i - \bar{x}_j))^2.$$  

By assumption the last expression on the right hand side of (2.10) tends to zero when $N \to \infty$. Hence

$$Q \sim \sum_{\max(i,j) \neq b} (1/n_i + 1/n_j)^{-1}(Y_{ij} - (\bar{x}_i - \bar{x}_j))^2$$

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 $i,j \neq b$ as for the $Y_{ij}$ with $i,j \neq c$.

In [3] is given an example which shows that the estimate $Z_{12}$ of $\xi - \xi$ 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 $\hat{\xi}_i - \hat{\xi}_j$ do not have this deficiency. If for the same $i$ and $j$ $n_i$ and $n_j$ tends to infinity then $\hat{\xi}_i - \hat{\xi}_j$ is a consistent estimate of $\xi - \xi$.

3. A N A L T E R N A T I V E E S T I M A T E.

Since the estimates $\hat{\xi}_i - \hat{\xi}_j$ of (2.9) is not easily computed unless one have access to an electronic computer, we shall give alternative simpler estimates which also are weighted estimates.

We shall minimize

$$\sum n_i n_j (Y_{ij} - (\bar{x}_i - \bar{x}_j))^2.$$  

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

\[
(3.2) \quad W_{ij} = \sum n_\alpha (Y_{i\alpha} - Y_{j\alpha}) = \bar{Y}_i - \bar{Y}_j
\]

where we have introduced the weighted differences \( \bar{Y}_i = (\sum n_\alpha)^{-1} \sum n_\alpha Y_{i\alpha} \).

\( Y_{i\alpha} \). 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 \( n_1 = n_2 = \ldots = n_c \).

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

**THEOREM 4.** If \( n_1 = n_2 = \ldots = n_c = n \) then the estimates \( Z_{ij} \) are the minimum variance unbiased linear estimates.

**PROOF.** Define \( \sigma^2 \) and \( a \) by

\[
(4.1) \quad \sigma^2 = \text{Var} \ Y_{12}
\]

\[
(4.2) \quad a \sigma^2 = \text{Cov} \ (Y_{12}, Y_{13}).
\]

Note that both \( \sigma^2 \) and \( a \) depend upon \( n \) and \( F \). By symmetry we have

\[
\text{Var} \ Y_{ij} = \sigma^2 \quad i \neq j.
\]

\[
\begin{align*}
\text{Cov} \ (Y_{ij}, Y_{kl}) &= 0 \quad i \neq k, i \neq l, j \neq k, j \neq l.
\text{Cov} \ (Y_{ij}, Y_{il}) &= a \sigma^2 \quad j \neq l, i \neq j, i \neq l.
\text{Cov} \ (Y_{ij}, Y_{kj}) &= a \sigma^2 \quad i \neq k, i \neq j, k \neq j.
\text{Cov} \ (Y_{ij}, Y_{jl}) &= -a \sigma^2 \quad i \neq l, i \neq j, j \neq l,
\text{Cov} \ (Y_{ij}, Y_{ki}) &= -a \sigma^2 \quad k \neq j, i \neq j, k \neq i.
\end{align*}
\]
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 $(1+2(c-2)a)^{-1} G(d)$ where $d = a(1+(c-4)a)^{-1}$. Hence

$$(2.1)$$ is proportional to

$$Q_1 = \sum_{j} \frac{j-1}{2} (Y_{ij} - (\hat{\xi}_i - \hat{\xi}_j))^2$$

$$+ 2d \sum_j \frac{j-1}{2} \sum_{i=1}^{j-1} \sum_{h=1}^{j-1} (Y_{ij} - (\hat{\xi}_i - \hat{\xi}_j))(Y_{jhl} - (\hat{\xi}_h - \hat{\xi}_j))$$

$$+ 2d \sum_j \frac{j-1}{2} \sum_{i=1}^{j-1} \sum_{h=1}^{j-1} (Y_{ij} - (\hat{\xi}_i - \hat{\xi}_j))(Y_{ih} - (\hat{\xi}_h - \hat{\xi}_i)).$$

The part of $Q_1$ involving $\hat{\xi}_\alpha$ can be written

$$\sum (Y_i - (\hat{\xi}_i - \hat{\xi}_\alpha))^2$$

$$+ 2 \sum_j \frac{j-1}{2} \sum_{i=1}^{j-1} \sum_{i=1}^{j-1} (Y_{ij} - (\hat{\xi}_j - \hat{\xi}_\alpha))(Y_{ji} - (\hat{\xi}_i - \hat{\xi}_\alpha))$$

$$+ 2d \sum_j \frac{j-1}{2} \sum_{i=1}^{j-1} \sum_{i=1}^{j-1} (Y_{ij} - (\hat{\xi}_j - \hat{\xi}_\alpha))(Y_{i\alpha} - (\hat{\xi}_i - \hat{\xi}_\alpha)).$$

We find

$$\frac{\partial Q_1}{\partial \hat{\xi}_\alpha} = 2(1+d(c-1))(c\hat{\xi}_\alpha - \sum \hat{\xi}_i - \sum Y_{i\alpha}i).$$

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

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

and hence by (1.4)

$$\hat{\xi}_i - \hat{\xi}_j - c^{-1} \sum Y_{\alpha i} - c^{-1} \sum Y_{\alpha j} = z_{ij}.$$
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 observations are:

<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th></th>
<th>$B_2$</th>
<th></th>
<th>$B_3$</th>
<th></th>
<th>$B_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>61.5</td>
<td>55.0</td>
<td>52.5</td>
<td>42.0</td>
<td>60.3</td>
<td>50.8</td>
<td>56.5</td>
</tr>
<tr>
<td>$A_2$</td>
<td>68.2</td>
<td>42.0</td>
<td>61.8</td>
<td>54.0</td>
<td>51.7</td>
<td>64.7</td>
<td>59.0</td>
</tr>
<tr>
<td>$A_3$</td>
<td>64.0</td>
<td>60.2</td>
<td>49.5</td>
<td>61.0</td>
<td>49.3</td>
<td>61.7</td>
<td>47.2</td>
</tr>
<tr>
<td>$A_4$</td>
<td>65.0</td>
<td>52.7</td>
<td>48.2</td>
<td>48.0</td>
<td>64.0</td>
<td>53.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59.7</td>
<td>39.6</td>
<td>62.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th></th>
<th>$B_2$</th>
<th></th>
<th>$B_3$</th>
<th></th>
<th>$B_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>37.0</td>
<td>56.3</td>
<td>39.7</td>
<td>50.0</td>
<td>59.0</td>
<td>59.5</td>
<td>45.2</td>
</tr>
<tr>
<td>$A_2$</td>
<td>36.3</td>
<td>69.8</td>
<td>46.0</td>
<td>43.8</td>
<td>57.4</td>
<td>52.8</td>
<td>57.0</td>
</tr>
<tr>
<td>$A_3$</td>
<td>68.0</td>
<td>67.0</td>
<td>61.3</td>
<td>54.5</td>
<td>54.0</td>
<td>56.0</td>
<td>61.4</td>
</tr>
<tr>
<td>$A_4$</td>
<td></td>
<td>55.3</td>
<td></td>
<td>47.0</td>
<td></td>
<td>42.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>55.7</td>
<td></td>
<td>54.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Let $\hat{c}_{ij}$ denote the expectation of the variables from $(A_i, B_j)$. In Table 2 are given the estimates of the differences $\hat{c}_{ij} - \hat{c}_{44}$ obtained by the different methods.
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 $\Theta$ do not differ much. The estimates $\Theta$ 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.
It is seen that the estimate of $\hat{\delta}_{11} - \hat{\delta}_{41}$ based on $W$ and $\varnothing$ 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 $\hat{\delta}_{11} - \hat{\delta}_{41}$, while there are $6 + 2$ observations behind the other estimates.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\delta}<em>{11} - \hat{\delta}</em>{41}$</th>
<th>$\hat{\delta}<em>{11} - \hat{\delta}</em>{42}$</th>
<th>$\hat{\delta}<em>{41} - \hat{\delta}</em>{42}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>15.8</td>
<td>18.05</td>
<td>3.1</td>
</tr>
<tr>
<td>$Z$</td>
<td>15.52</td>
<td>18.3</td>
<td>2.78</td>
</tr>
<tr>
<td>$W$</td>
<td>15.66</td>
<td>18.40</td>
<td>2.74</td>
</tr>
<tr>
<td>$\varnothing$</td>
<td>15.61</td>
<td>18.38</td>
<td>2.77</td>
</tr>
<tr>
<td>Classical</td>
<td>14.72</td>
<td>17.78</td>
<td>3.06</td>
</tr>
</tbody>
</table>
REFERENCES

Comparison of the normal scores and Wilcoxon tests.
1 307-318. Univ. of California Press.

Statist. 34 598-691.