The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1449-1458

ESTIMATING REGRESSION COEFFICIENTS BY MINIMIZING THE DISPERSION OF THE RESIDUALS

BY LOUIS A. JAECKEL[1]

Bell Telephone Laboratories, Incorporated

An appealing approach to the problem of estimating the regression coefficients in a linear model is to find those values of the coefficients which make the residuals as small as possible. We give some measures of the dispersion of a set of numbers, and define our estimates as those values of the parameters which minimize the dispersion of the residuals. We consider dispersion measures which are certain linear combinations of the ordered residuals. We show that the estimates derived from them are asymptotically equivalent to estimates recently proposed by Jurečková. In the case of a single parameter, we show that our estimate is a "weighted median" of the pairwise slopes $(Y_j - Y_i)/(c^j - c^i)$.

1. Introduction. An appealing approach to the problem of estimating the regression coefficients in a linear model is to find those values of the coefficients which make the residuals as small as possible. We shall give some measures of the dispersion of a set of numbers, and define our estimates as those values of the parameters which minimize the dispersion of the residuals.

Let $Y_1, Y_2, \cdots, Y_N$ be independent random variables with continuous distribution functions

$$F(y - \alpha^\circ - \beta^\circ c^i) , \qquad i = 1, 2, \cdots, N$$

where the c^i are known K-vectors and α° and the K-vector β° are unknown. We shall consider only the problem of estimating β°; we shall not estimate α°.

Let $D(z)$ be a translation-invariant measure of the dispersion of $z = (z_1, z_2, \cdots, z_N)$; that is, $D(z + b) = D(z)$, where $b = (b, b, \cdots, b)$. Our estimate of β° will be any β which minimizes $D(Y - \beta C)$, where $Y = (Y_1, Y_2, \cdots, Y_N)$ and $C = ((c_{ji}))$ is the $K \times N$ matrix whose columns are the c^i. Clearly, α° connot be estimated by this method. If $D(z)$ is the variance of the z_i, then our estimates become the usual least squares estimates of the β_i. Because of the sensitivity of this method to large deviations, it is of interest to consider alternatives to it which are more "robust". In Section 2 we consider a particular class of dispersion measures: certain linear combinations of the ordered z_i. We then define our estimates and discuss some of their properties. We show in Section 3 that these regression estimates are asymptotically equivalent to the relatively robust estimates proposed by Jurečková (1971), from which it follows that the asymptotic properties of her estimates hold for ours as well. Her estimates are derived by inverting rank tests for hypotheses about the β_i, a generalization of the method

Received December 17, 1971; revised February 29, 1972.

[1] Now at the University of California, Berkeley.

AMS 1970 *subject classifications*. Primary 62G05, 62J05; Secondary 62G35.

Key words and phrases. Estimation, linear model, regression.

of Hodges and Lehmann (1963) and Adichie (1967). Our approach seems more natural, and our estimates seem to be easier to compute.

In Section 4 we consider the case of a single parameter. We show that our estimate of β° is a "weighted median" of pairwise slopes $(Y_j - Y_i)/(c^j - c^i)$. A closely related estimate has been studied by Theil (1950) and Sen (1968).

We shall follow the notation and assumptions of Jurečková (1971) as much as possible.

2. Definition of some dispersion measures and properties of the regression estimates. Let $a_N(k)$, $k = 1, 2, \cdots, N$ be a non-decreasing set of scores, not all equal, satisfying

$$\sum_{k=1}^{N} a_N(k) = 0 . \tag{2.1}$$

Such scores may be generated as in Jurečková (1971), Equation (2.7) or (2.8). An example is given in Corollary 1 of Section 4 below. If the scores are not non-decreasing, the convexity property in Theorem 1 does not hold in general. For any $z = (z_1, z_2, \cdots, z_N)$, let $z_{(1)} \leqq z_{(2)} \leqq \cdots \leqq z_{(N)}$ be the ordered z_i. Define

$$D(z) = \sum_{k=1}^{N} a_N(k) z_{(k)} .$$

By (2.1), $D(z)$ is translation-invariant. Since $D(bz) = bD(z)$ for any nonnegative constant b, statistics of this type have been proposed as scale parameters. See for example Downton (1966) and Chernoff, Gastwirth and Johns (1967). Because $D(z)$ is small when the z_i are close to each other, we shall use it as a measure of dispersion.

For fixed $Y_1, Y_2, \cdots, Y_N$ and for any β, the residuals are

$$z_i = Y_i - \beta c^i \qquad i = 1, 2, \cdots, N . \tag{2.2}$$

The ordered residuals are

$$z_{(k)} = Y_{i(k)} - \beta c^{i(k)} \qquad k = 1, 2, \cdots, N ,$$

where $i(k)$ is the index of the observation giving rise to the kth ordered residual. (If two residuals are equal there is an ambiguity in $i(k)$ but not in $z_{(k)}$.) The dispersion of the residuals, as a function of β, is

$$D(Y - \beta C) = \sum_{k=1}^{N} a_N(k)[Y_{i(k)} - \beta c^{i(k)}] . \tag{2.3}$$

We now derive some properties of $D(Y - \beta C)$.

THEOREM 1. *For fixed Y, $D(Y - \beta C)$ is a nonnegative, continuous, and convex function of β.*

PROOF. Let $z = (z_1, z_2, \cdots, z_N)$ be any N-vector. Let l be the index of the first positive $a_N(k)$. Then

$$D(z) = \sum a_N(k) z_{(k)} = \sum a_N(k)[z_{(k)} - z_{(l)}] \geqq 0 ,$$

since each term in the sum is nonnegative. Hence D is nonnegative, and if the z_i are not all equal, D is positive.

Let $p = (p(1), p(2), \cdots, p(N))$ be any permutation of the indices $1, 2, \cdots, N$. Let P be the set of all such permutations. By Theorem 368 of Hardy, Littlewood and Pólya (1952),

$$D_p = \sum a_N(k) z_{p(k)}$$

is maximized over P by any permutation which arranges the z_i in non-decreasing order. The proof consists of showing that we can move from any permutation to one which orders the z_i by a sequence of transpositions, none of which decreases D_p.

Therefore we can write

$$D(Y - \beta C) = \sum a_N(k) z_{(k)} = \max_{p \in P} \sum a_N(k) z_{p(k)} ,$$

where the z_i are now the residuals defined in (2.2). For each p in P we have

$$\sum a_N(k) z_{p(k)} = \sum a_N(k)[Y_{p(k)} - \beta c^{p(k)}] ,$$

which is a linear function of β, and therefore continuous and convex in β. The maximum of a finite number of such functions is clearly continuous, and is easily seen to be convex (concave upward). The proof is complete.

We remark that the β-space is divided into a finite number of convex polygonal subsets, on each of which $D(Y - \beta C)$ is a linear function of β.

Let E be the $N \times N$ matrix all of whose entries are $1/N$. Let $\bar{C} = CE$. We shall now show that if $C - \bar{C}$ has rank K, then $D(Y - \beta C)$ attains its minimum, and the set of β for which this occurs is bounded. This is an immediate consequence of the following theorem.

THEOREM 2. *If $C - \bar{C}$ has rank K, then for any D_0, the set $\{\beta : D(Y - \beta C) \leqq D_0\}$ is bounded.*

PROOF. We assume without loss of generality that $D_0 > D(Y)$. Let U be any K-vector such that $||U|| = 1$, and consider D restricted to $\beta = bU$, $b \geqq 0$. Let $t_i = Uc^i$, $i = 1, 2, \cdots, N$, and let $T = (t_1, t_2, \cdots, t_N) = UC$. If we write $\bar{T} = (\bar{t}, \bar{t}, \cdots, \bar{t}) = TE = UCE$, then $\bar{T} = U\bar{C}$. Since $C - \bar{C}$ has rank K, $T - \bar{T} = U(C - \bar{C}) \neq 0$. Therefore the t_i are not all equal. We write

$$D(Y - bUC) = D(Y - bT) = \sum a_N(k)[Y_{i(k)} - bt_{i(k)}] .$$

This is just the measure of dispersion for estimating the scalar parameter b. This case is discussed in more detail in Section 4, where it is shown that dD/db is non-decreasing in b and is eventually positive.

It follows that for some $b_U > 0$, $D(Y - b_U UC) > D_0$. Since D is continuous, there is an open set S_U of unit vectors V, containing U, such that $D(Y - b_U VC) > D_0$. Since $D_0 > D(Y)$ and D is convex, we have $D(Y - bVC) > D_0$ for all $b \geqq b_U$ and all V in S_U. So for each unit vector U, we have an open set of unit vectors S_U containing it. Since the unit sphere is compact, a finite number of these sets covers it. Let B be the maximum of the corresponding finite set of b_U. Then for all $b \geqq B$ and all unit vectors V, $D(Y - bVC) > D_0$. The theorem follows.

We now define our estimate.

DEFINITION. Let β_D be any β which minimizes $D(Y - \beta C)$. Then β_D is our estimate of the vector of regression coefficients β°.

Note that β_D may not be unique, so in practice there may be some arbitrariness in its definition. However, by Theorem 3 below, the diameter of the set of possible values of $N^{\frac{1}{2}}\beta_D$ approaches zero asymptotically, so for large N it will not matter much how β_D is chosen. The estimate is invariant in the sense that if β^1 is any K-vector, then $\beta_D(Y + \beta^1 C) = \beta_D(Y) + \beta^1$.

Computation of β_D involves minimizing a convex function of β, a relatively benign type of minimization problem. Since we can evaluate D everywhere, and (as we shall see in the next section) its derivatives almost everywhere, we can apply an iterative method for searching for the minimum. One possibility is the method of steepest descent, which uses first derivatives. Minimization procedures based on second derivatives cannot be used, because the second derivatives of D are identically zero wherever they exist.

3. Asymptotic equivalence to Jurečková's estimates. We define new variables as in Jurečková (1971). Let

$$\Delta = N^{\frac{1}{2}}\beta \qquad \text{and} \qquad x_{ji} = N^{-\frac{1}{2}}c_{ji}\,.$$

We now choose Δ_D to estimate Δ° by minimizing $D(Y - \Delta X)$. We may assume without loss of generality that $\Delta^\circ = 0$. We rewrite (2.3) as

$$\begin{aligned} D(Y - \Delta X) &= \textstyle\sum_{k=1}^{N} a_N(k)[Y_{i(k)} - \Delta x^{i(k)}] \\ &= \textstyle\sum_{i=1}^{N} a_N(R_i^\Delta)[Y_i - \Delta x^i]\,, \end{aligned}$$

where $R_1^\Delta, R_2^\Delta, \cdots, R_N^\Delta$ are the ranks of the residuals $Y_i - \Delta x^i, i = 1, 2, \cdots, N$.

The partial derivatives of $D(Y - \Delta X)$ exist for almost all Δ, and where they exist are

$$\begin{aligned} \frac{\partial}{\partial \Delta_j} D(Y - \Delta X) &= \textstyle\sum a_N(R_i^\Delta)[-x_{ji}] \\ &= -\textstyle\sum a_N(R_i^\Delta)(x_{ji} - \bar{x}_j) \qquad (3.1) \\ &= -S_{Nj}(Y - \Delta X)\,, \qquad j = 1, 2, \cdots, K\,. \end{aligned}$$

Thus they are minus the rank statistics S_{Nj} defined by Equation (2.14) of Jurečková (1971).

We can now see the relationship between our estimate Δ_D and Jurečková's estimate. Her procedure is to choose a Δ to minimize $\sum |S_{Nj}(Y - \Delta X)|$. This has the effect of making all of the $S_{Nj}(Y - \Delta X)$ near zero. Our procedure chooses Δ to minimize $D(Y - \Delta X)$, which means making all of its partial derivatives nearly equal to zero. But the derivatives of D are minus the S_{Nj}, so the two procedures appear to be essentially the same. We shall show that under the assumptions of Jurečková (1971), $\{\Delta_D\}$ is asymptotically equivalent, as defined in (3.3) below, to the sequence $\{\hat{\Delta}_N\}$, also defined below. Jurečková (1971) showed that her estimate is asymptotically equivalent to this same sequence. It follows that her estimate and our estimate are asymptotically equivalent, if we define that concept analogously to (3.3).

The assumptions made in Jurečková (1971) are, briefly: The Y_i are independent, with distribution functions

$$F(y - \alpha^\circ - \Delta^\circ x^i), \qquad i = 1, 2, \cdots, N,$$

where F has finite Fisher information. The entries in the positive definite matrix $\Sigma = [\sigma_{lj}] = [\sigma^1, \sigma^2, \cdots, \sigma^K]$ are $\sigma_{lj} = \lim_{N\to\infty}(x_l - \bar{x}_l)(x_j - \bar{x}_j)'$. Some further technical assumptions are made about the entries in X and the scores $a_N(k)$. The positive constants γ and A^2 are defined as in (2.10) and (2.11) of Jurečková (1971).

Let

$$S_N(Y) = (S_{N1}(Y), S_{N2}(Y), \cdots, S_{NK}(Y)).$$

Since F is assumed continuous, the $S_{Nj}(Y)$ are well defined with probability one. Define the quadratic function

$$Q(\Delta) = \tfrac{1}{2}\gamma\Delta\Sigma\Delta' - \mathrm{S}_N(Y)\Delta + D(Y).$$

The partial derivatives of $Q(\Delta)$ are

$$\frac{\partial Q}{\partial \Delta_j} = \gamma\Delta\sigma^j - S_{Nj}(Y), \qquad j = 1, 2, \cdots, K. \tag{3.2}$$

The unique point at which $Q(\Delta)$ attains its minimum is the solution of the equations

$$S_{Nj}(Y) = \gamma\Delta\sigma^j, \qquad j = 1, 2, \cdots, K.$$

We call this point $\hat{\Delta}_N$. It is the same as the $\hat{\Delta}_N$ defined by (4.5) of Jurečková (1971). It is shown there that $\hat{\Delta}_N$ is asymptotically normal

$$(0, \gamma^{-2}A^2\Sigma^{-1}).$$

Let

$$B_N = \{\Delta : D(Y - \Delta X) \text{ is minimized}\}.$$

By Theorem 2, B_N is nonempty and bounded. Our estimate Δ_D is thus any point of B_N.

DEFINITION. The two sequences of random vectors $\{\Delta_D\}$ and $\{\hat{\Delta}_N\}$ are said to be *asymptotically equivalent* if the distance between corresponding terms converges to zero in probability; that is, if

$$\lim_{N\to\infty} P_0\{\sup_{\Delta\in B_N} ||\Delta - \hat{\Delta}_N|| \geqq r\} = 0 \tag{3.3}$$

for all $r > 0$.

THEOREM 3. *Under the assumptions of Jurečková* (1971), $\{\Delta_D\}$ *and* $\{\hat{\Delta}_N\}$ *are asymptotically equivalent. It follows that* Δ_D *is asymptotically normal*

$$(0, \gamma^{-2}A^2\Sigma^{-1}).$$

We shall show that $D(Y - \Delta X)$ and $Q(\Delta)$ approach each other as N increases; from this we can show that the points where they are minimized approach each other.

LEMMA 1. *Under the assumptions of the theorem,*

$$\lim_{N\to\infty} P_0\{\max_{||\Delta||\leq C} |Q(\Delta) - D(Y - \Delta X)| \geqq \varepsilon\} = 0$$

for all $\varepsilon > 0$ *and* $C > 0$.

PROOF. We observe first that $Q(0) = D(Y)$. By Theorem 3.1 of Jurečková (1971) we have

$$(3.4)\qquad \lim_{N\to\infty} P_0\left\{\sup_{||\Delta||\leq C} |S_{Nj}(Y - \Delta X) - S_{Nj}(Y) + \gamma\Delta\sigma^j| \geqq \frac{\varepsilon}{CK^{\frac{1}{2}}}\right\} = 0$$

for any $\varepsilon > 0$, $C > 0$, and $j = 1, 2, \cdots, K$. Therefore, by (3.1), (3.2), and (3.4),

$$(3.5)\qquad \lim_{N\to\infty} P_0\left\{\sup_{||\Delta||\leq C}\left|\frac{\partial Q(\Delta)}{\partial\Delta_j} - \frac{\partial D(Y - \Delta X)}{\partial\Delta_j}\right| \geqq \frac{\varepsilon}{CK^{\frac{1}{2}}}\right\} = 0$$

for any $\varepsilon > 0$, $C > 0$, and $j = 1, 2, \cdots, K$.

Choose ε and C, and choose $\Delta_0 \neq 0$ such that $||\Delta_0|| \leqq C$. We write the points on the line segment from 0 to Δ_0 as $t\Delta_0$, $0 \leqq t \leqq 1$. Then

$$(3.6)\qquad \frac{d}{dt}[Q(t\Delta_0) - D(Y - t\Delta_0 X)] = \sum_{j=1}^K \Delta_{0j}\left[\frac{\partial}{\partial\Delta_j} Q(t\Delta_0) - \frac{\partial}{\partial\Delta_j} D(Y - t\Delta_0 X)\right].$$

If the events described in (3.5) do not hold for any j, that is, if all of the differences are less than $\varepsilon/CK^{\frac{1}{2}}$, then the absolute value of the sum in (3.6) is less than

$$||\Delta_0||\left(K\frac{\varepsilon^2}{C^2K}\right)^{\frac{1}{2}} \leqq \varepsilon .$$

In this case, since $Q = D$ for $t = 0$ and the derivative of their difference is less than ε, we have $|Q - D| < \varepsilon$ for $t = 1$, or

$$|Q(\Delta_0) - D(Y - \Delta_0 X)| < \varepsilon .$$

If this strict inequality holds for all Δ_0 such that $||\Delta_0|| \leqq C$, then

$$\max_{||\Delta||\leq C} |Q(\Delta) - D(Y - \Delta X)| < \varepsilon ,$$

because $Q - D$ is continuous and hence must attain its maximum on $||\Delta|| \leqq C$. Since by (3.5) this inequality holds with probability approaching one, the lemma is proved.

PROOF OF THE THEOREM. We must show that (3.3) holds for all $r > 0$. It suffices to show that for all $r > 0$ and $p > 0$,

$$(3.7)\qquad P_0\{\sup_{\Delta\in B_N} ||\Delta - \hat{\Delta}_N|| \geqq r\} \leqq p$$

for all sufficiently large N.

Choose $r > 0$ and $p > 0$. Since $\hat{\Delta}_N$ is asymptotically normal, there is a C_0 such that for sufficiently large N,

$$(3.8)\qquad P_0\{||\hat{\Delta}_N|| > C_0\} \leqq p/2 .$$

Let

$$T = \min\{Q(\Delta) : \|\Delta - \hat{\Delta}_N\| = r\} - Q(\hat{\Delta}_N) .$$

Clearly, $T > 0$. By Lemma 1,

$$P_0\{\max_{\|\Delta\| \leqq C_0 + r} |Q(\Delta) - D(Y - \Delta X)| \geqq T/2\} \leqq p/2 \tag{3.9}$$

for sufficiently large N. Therefore, by (3.8) and (3.9), for suffiiciently large N we have with probability at least $1 - p$:

$$\|\hat{\Delta}_N\| \leqq C_0 , \qquad D(Y - \hat{\Delta}_N X) < Q(\hat{\Delta}_N) + T/2 ,$$

and for all Δ such that $\|\Delta - \hat{\Delta}_N\| = r$,

$$\|\Delta\| \leqq C_0 + r$$

and

$$D(Y - \Delta X) > Q(\Delta) - T/2 .$$

It follows that for all Δ such that $\|\Delta - \hat{\Delta}_N\| = r$,

$$\begin{aligned} D(Y - \Delta X) &> Q(\Delta) - T/2 \geqq T + Q(\hat{\Delta}_N) - T/2 \\ &= Q(\hat{\Delta}_N) + T/2 > D(Y - \hat{\Delta}_N X) . \end{aligned}$$

By the convexity of D, we must have

$$D(Y - \Delta X) > D(Y - \hat{\Delta}_N X)$$

for all Δ such that $\|\Delta - \hat{\Delta}_N\| \geqq r$. This implies that any Δ for which $D(Y - \Delta X)$ is minimized must satisfy

$$\|\Delta - \hat{\Delta}_N\| < r .$$

So (3.7) is satisfied, and the theorem is proved.

4. Estimation of a single parameter. When $K = 1$ we have

$$D(Y - \beta C) = \sum a_N(k)(Y_{i(k)} - \beta c^{i(k)}) ,$$

where β and each c^i are numbers rather than vectors. The c^i are assumed to be not all equal. We shall show that if the $a_N(k)$ are symmetric, β_D is a "weighted median" of the pairwise slopes $(Y_j - Y_i)/(c^j - c^i)$. We then consider some special cases.

We assume the Y_i are fixed. We write

$$D'(Y - \beta C) = \frac{dD}{d\beta} = -\sum a_N(k) c^{i(k)} .$$

Since D is convex and D' takes on only a finite number of values, D' is a nondecreasing step function in β. Since, for sufficiently large β, we have

$$Y_j - \beta c^j < Y_i - \beta c^i$$

for all i and j such that $c^j > c^i$, it follows that for very large β,

$$c^{i(1)} \geqq c^{i(2)} \geqq \cdots \geqq c^{i(N)} .$$

By the argument given in the proof of Theorem 1, D' then assumes its maximum value, which is positive. Similarly, for β sufficiently small, the ordering of the c^i is essentially reversed, so that D' assumes its minimum value, which is negative.

The jumps in D' occur where the ordering of the residuals changes. If $c^j > c^i$, then the corresponding residuals cross each other when $\beta = \beta_{ij}$, where

$$Y_i - \beta_{ij}c^i = Y_j - \beta_{ij}c^j \; ; \tag{4.1}$$

that is, when

$$\beta_{ij} = \frac{Y_j - Y_i}{c^j - c^i} \, .$$

We assume the β_{ij} are all distinct. (This event occurs with probability one because F is continuous.) We now find the height of one of these jumps. Suppose, for $\beta = \beta_{ij}$, the two residuals in (4.1) are the kth and $(k + 1)$st among the ordered residuals. Then the increase in D' which occurs when β moves from just below β_{ij} to just above it is

$$\begin{aligned} -\, a_N(k+1)(c^i - c^j) - a_N(k)(c^j - c^i) \\ = [a_N(k+1) - a_N(k)](c^j - c^i) \, . \end{aligned} \tag{4.2}$$

We can now define β_D as a certain quantile of the probability distribution whose distribution function is

$$G(\beta) = \frac{D'(Y - \beta C) - \min D'}{\max D' - \min D'} \, .$$

THEOREM 4. *β_D is a point such that, if the random variable T has distribution function G, then*

$$P(T < \beta_D) \leqq \frac{-\min D'}{\max D' - \min D'}$$

and

$$P(T > \beta_D) \leqq \frac{\max D'}{\max D' - \min D'} \, .$$

If the $a_N(k)$ are symmetric, that is, if $a_N(k) = -a_N(N + 1 - k)$, $k = 1, 2, \cdots, N$, then β_D is a median of G.

PROOF. D is minimized at β_D if $D' \leqq 0$ for $\beta < \beta_D$ and $D' \geqq 0$ for $\beta > \beta_D$. The first assertion then follows from the definition of G. If the $a_N(k)$ are symmetric, $\max D' = -\min D'$, and we have the second assertion.

An interesting special case is given in the following corollary.

COROLLARY 1. *If the $a_N(k)$ are the Wilcoxon scores*

$$a_N(k) = \frac{k}{N+1} - \tfrac{1}{2} \, , \qquad k = 1, 2, \cdots, N \, ,$$

then β_D is a median of the set of pairwise slopes

$$\beta_{ij} = \frac{Y_j - Y_i}{c^j - c^i}$$

for $c^j > c^i$, where each slope is assigned weight proportional to $c^j - c^i$. The asymptotic variance of $N^{\frac{1}{2}}\beta_D$ in this case, if the assumptions of Theorem 3 hold, is

$$\frac{1}{12\sigma^2\{\int f^2(x)\,dx\}^2},$$

where

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum (c^i - \bar{c})^2$$

and f is the density of F.

PROOF. If the $a_N(k)$ are defined as above, the differences between consecutive scores are all equal, so the heights of the jumps in (4.2) depend only on $c^j - c^i$. $G(\beta)$ therefore has jumps at β_{ij} with height proportional to $c^j - c^i$. By Theorem 4, half of the weight (heights of jumps in G) falls on each side of β_D.

The asymptotic variance formula is a consequence of Theorem 3 and the definitions in Jurečková (1971), with $\varphi(u) = u - \frac{1}{2}$.

Theil (1950) proposed that the median of the pairwise slopes, with each assigned the same weight, be used as an estimate of β. The asymptotic behavior of this estimate was studied by Sen (1968). He showed that its asymptotic variance depends on the c^i in a complex way, whereas the asymptotic variance of β_D depends on the c^i only through their variance. In fact, by Corollary 1 and Sen (1968), page 1385, we see that Sen's asymptotic variance is always equal to or greater than that of β_D.

If the c^i are all 0 or 1, we have the two-sample problem of estimating difference in location, and β_D becomes an estimate of that difference. We then have the following result.

COROLLARY 2. *If $c^1 = c^2 = \cdots = c^m = 0$ and $c^{m+1} = c^{m+2} = \cdots = c^N = 1$, and if we renumber the observations so that $Y_1 < Y_2 < \cdots < Y_m$ and $Y_{m+1} < Y_{m+2} < \cdots < Y_N$, then β_D is the quantile given by Theorem 4 of the set of $m(N-m)$ pairwise differences*

$$\beta_{ij} = Y_j - Y_i\,, \qquad i = 1, 2, \cdots, m\,; \quad j = m+1, m+2, \cdots, N\,;$$

where each β_{ij} is assigned weight proportional to

$$a_N(i+j-m) - a_N(i+j-m-1)\,.$$

PROOF. The pairs i and j considered in (4.2) are those for which $c^j > c^i$. Here they are the pairs with $i \leqq m$ and $j \geqq m+1$, for which $c^j - c^i = 1$. When $\beta = \beta_{ij}$ as defined by (4.1), that is, when the ith and jth residuals are equal, then the number of residuals less than $Y_i = Y_j - \beta_{ij}$ is $(i-1) + (j-m-1)$. These two residuals therefore occupy positions $i+j-m-1$ and $i+j-m$ among the ordered residuals. By (4.2) the corollary is proved.

A result analogous to Corollary 2 for the one-sample problem of estimating location appears in Jaeckel (1969).

If in Corollary 2 the $a_N(k)$ are the scores of Corollary 1, then the weights are all equal, and we have the well-known result of Hodges and Lehmann (1963), page 602: The estimate is the median of the pairwise differences $Y_j - Y_i$.

The author is indebted to Dr. Colin L. Mallows for several valuable conversations on this subject.

REFERENCES

ADICHIE, J. N. (1967). Estimates of regression coefficients based on rank tests. *Ann. Math. Statist.* **38** 894–904.

CHERNOFF, H., GASTWIRTH, J. L., and JOHNS, M. V. (1967). Asymptotic distributions of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

DOWNTON, F. (1966). Linear estimates with polynomial coefficients. *Biometrika* **53** 129–141.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1952). *Inequalities* (2nd ed.). Cambridge Univ. Press.

HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

JAECKEL, L. A. (1969). Robust estimates of location. Ph. D. dissertation, University of California, Berkeley.

JUREČKOVÁ, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* **42** 1328–1338.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

THEIL, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I, II, and III. *Nederl. Akad. Wetensch. Proc. Ser. A.* **53** 386–392, 521–525, and 1397–1412.

BELL TELEPHONE LABORATORIES
MURRAY HILL, NEW JERSEY 07974