RANK-BASED ROBUST INFERENCE IN REGRESSION MODELS WITH SEVERAL OBSERVATIONS PER CELL

BY

DAVID C. DRAPER*

TECHNICAL REPORT NO. 18
FEBRUARY 1983

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY

*RESEARCH PARTIALLY SUPPORTED BY
NATIONAL SCIENCE FOUNDATION GRANTS
MCS77-01665 AND MCS79-03716

RANK-BASED ROBUST INFERENCE IN REGRESSION MODELS WITH SEVERAL OBSERVATIONS PER CELL[†]

by

David C. Draper[‡]

Department of Statistics
University of California, Berkeley

ABSTRACT

An extension of the Hodges-Lehmann rank-based robust estimation method to regression models with several observations per cell is proposed and investigated. Asymptotically distribution-free tests and confidence procedures are introduced, and it is shown that the new estimator has asymptotic efficiency 1 relative to the best rank-based robust estimator to date (that of Jaeckel), meaning that it shares with Jaeckel's estimator excellent efficiency properties relative to the classical least-squares estimator.

AMS Subject Classification: Primary 62G05, 62J05; Secondary 62F35, 62E20

Key words and phrases: robust estimation, regression, rank-based, Hodges-Lehmann

---

[†]Prepared with partial support of National Science Foundation Grants MCS77-01665 and MCS79-03716.

[‡]Now with the Department of Statistics, University of Chicago.

1. Introduction

Consider the linear model with several observations per cell, which may be written

$$Y_{ij} = \mu_i + e_{ij}, \qquad \left\{ \begin{array}{l} i = 1, \ldots ,I \\ j = 1, \ldots ,n_i \end{array} \quad \sum_{i=1}^{I} n_i = N \right\} , \tag{1}$$

in which the e_{ij} are i.i.d. continuous random variables with density f satisfying

$$\theta = \int_{-\infty}^{\infty} f^2(x)dx < \infty \quad \text{and} \quad \sigma^2 = \text{Var}(e_{ij}) < \infty . \tag{2}$$

Here μ_i is a measure of centering for the ith of I total cells, Y_{ij} is the jth of the n_i observations in cell i, and N is the total number of observations. Suppose the μ_i, instead of being functions of one or more qualitative factors as in the analysis of variance, are thought to depend on the I levels of $p \geq 1$ quantitative variables x_k:

$$\mu_i = \beta_0 + \sum_{k=1}^{p} x_{ik}\beta_k , \tag{3}$$

in which the x_{ik} are known regression constants (the fixed-effects model).

There are several ways in which this model arises in practice:

1) When the independent variables x_k are conceptually continuous but are made discrete by the measuring process;

2) When higher accuracy is desired at some values of the independent variables than at others, and the design clusters at those values;
3) When a pure error estimate is desired; and
4) When optimal design considerations call for replication at specified values.

The classical least-squares estimator is derived under the assumption that the specific error density f is the normal, and it possesses various well-known optimality properties in that model. When the data contain outliers or gross errors, or more generally come from a distribution with tails heavier than those of the normal, it is also well known that the classical procedures can lose their optimality and perform poorly both in terms of efficiency and of sensitivity to extreme observations. One method of dealing with this which has gained some acceptance is the data-analytic approach, in which graphical and numerical methods are employed to detect the ways in which the data do not meet the classical assumptions and the data are modified appropriately (through transformations, for example) before the classical techniques are applied. Another approach involves the use of robust methods which are appropriate under broader assumptions and so may be utilized directly with the original data. The application of one type of robust methods, those based on ranks, to problems of estimation and inference in the regression model with several observations per cell is investigated here.

There is an extensive literature devoted to rank-based regression estimates in the case $n_i = 1$, with principal developments by Theil (1950), Mood (1950), Adichie (1967), Sen (1968), and Jaeckel (1972). Of these methods the best in terms of efficiency is Jaeckel's, which will be discussed in detail below. All of these methods in the case $p = 1$ rely on the following idea: each choice of two points (x,y) with different x-values provides an estimate of the slope parameter β_1, and some estimator of location applied to the set of all such pairwise slopes should yield a reasonable composite estimate of β_1. Asymptotic work on these methods involves letting both I and N become large and requires some growth conditions on the regression constants x_{ik}. Most of these estimators can be applied to the model (1), but none of them take account of the special character of models with several observations per cell. The estimator developed below is different from the previous rank-based estimates in two ways -- its form is based on taking pairs of cells rather than pairs of points, and its asymptotic properties are found by letting the $n_i \to \infty$ while holding I constant, simplifying conditions on the x_{ik}. Other asymptotic schemes, for example letting both I and N grow but still requiring that $I/N \to 0$ as in Huber (1981), are not pursued here.

The above continuity assumption on the e_{ij} and consequently on the Y_{ij} is made to avoid technical complications involving ties in the ranking of the data. When ties are present in linear models data they are often due (as in case 1) above) to the measuring process having made a conceptually continuous variable discrete, and in such situations, provided the size of the roundoff is not large, the methods to be discussed below may be applied with little harm in acting as if the rounding had not occurred (cf. Lehmann (1975)). The finiteness of θ and σ^2 are needed because division by $1/\theta$ and σ^2 play a role in what follows; these conditions place little practical restriction on the use of the methods.

Rank-based estimates of the regression parameters in the model (1) can be found by extending the Hodges-Lehmann method of robust estimation to this setting. The main idea can perhaps best be seen in the context of the sub-model $p = 1$; extensions of these concepts to the case $p > 1$ are considered in section 3 below.

2. One Independent Variable

It is convenient to change the notation for the parameters and the values of the independent variable in this case: suppose that

$$\mu_i = \alpha + \beta x_i \quad , \tag{4}$$

where the x_i are known constants, all distinct. One sensible approach to the estimation of α and β is to estimate β first, by $\bar{\beta}$, say, and then to estimate α from the residuals $Y_{ij} - \bar{\beta}x_i$.

An intuitively appealing estimate of β can be found with no additional assumptions about the errors e_{ij} by considering differences of cell centers μ_i. Note that for $i \neq j$

$$\mu_i - \mu_j = \beta(x_i - x_j) \quad \text{and} \quad \beta = (\mu_i - \mu_j)/(x_i - x_j) \quad , \tag{5}$$

so that each pair of cells can give rise to a separate estimate of β through estimation of $\mu_i - \mu_j$, and then one composite estimate of β can be found by combining these separate estimates in some way.

The simple Hodges-Lehmann estimate of the $\mu_i - \mu_j$,

$$D_{ij} = \text{med}\left\{Y_{ik} - Y_{j\ell} : k = 1, \ldots, n_i; \ \ell = 1, \ldots, n_j\right\} , \tag{6}$$

the median of the set of all pairwise differences among the observations in cells i and j, is unsatisfactory, since these estimates do not satisfy the linearity constraints which the $\mu_i - \mu_j$ themselves do:

$$(\mu_i - \mu_j) + (\mu_j - \mu_k) = (\mu_i - \mu_k) , \tag{7}$$

but

$$D_{ij} + D_{jk} \neq D_{ik} . \tag{8}$$

This makes them unsuitable as a basis for linear inference about the $\mu_i = \alpha + \beta x_i$ and thus about α and β. Lehmann (1963a), when considering applying the Hodges-Lehmann estimation method to analysis of variance models with several observations per cell, proposed adjusting these raw Hodges-Lehmann estimates and estimating $\mu_i - \mu_j$ by

$$W'_{ij} = \bar{D}'_i - \bar{D}'_j , \tag{9}$$

where

$$\bar{D}_i' = I^{-1} \sum_{i=1}^{I} D_{ik} . \tag{10}$$

The linearity problem is thus removed, at the cost of offending intuition by using observations in cells other than i and j to help in the estimation of $\mu_i - \mu_j$. Lehmann pointed out, however, that the size of the influence of cells other than i and j on the estimator of $\mu_i - \mu_j$ tends to 0 in probability as the sample sizes increase. A different drawback of this estimation method was noticed by Spjøtvoll (1968) -- cells with unequal numbers of observations get equal weight in the calculation of the $\bar{D}_i'$. Spjøtvoll suggested several ways of remedying the situation, the simplest of which is to use

$$W_{ij} = \bar{D}_i - \bar{D}_j , \qquad \bar{D}_i = N^{-1} \sum_{k=1}^{I} n_k D_{ik} . \tag{11}$$

This is the form of Hodges-Lehmann estimation which is used in what follows. Thus a Hodges-Lehmann-type estimate of β based on cells i and j is simply

$$\hat{\beta}_{ij} = W_{ij}/(x_i - x_j) . \tag{12}$$

Several reasonable methods of combining these separate estimates come to mind, including taking their weighted average or weighted median or even calculating a weighted one-sample Hodges-Lehmann estimate based on them. If one is looking for a good compromise between robustness and efficiency, however, perhaps the best choice is a weighted average, since a good degree of resistance has already been built into the $\hat{\beta}_{ij}$ themselves and any further such measures might serve only to decrease the efficiency. The composite estimate studied here takes the form

$$\bar{\beta}^* = \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \gamma_{ij} \hat{\beta}_{ij} , \tag{13}$$

in which the weights γ_{ij} can be chosen to minimize the asymptotic variance of the composite estimator $\bar{\beta}^*$ while satisfying a constraint resulting in asymptotic unbiasedness of the estimator.

Perhaps the simplest realistic asymptotic theory is developed by letting the $n_i \to \infty$ in such a way that

$$n_i/N \to \rho_i \in (0,1) \quad \text{for all} \quad i = 1, \ldots ,I . \tag{14}$$

Under these conditions the asymptotic behavior of the D_{ij}, the statistics that $\bar{\beta}^*$ is based on, was worked out by Lehmann (1963a); he found that the rank of the asymptotic covariance matrix of the D_{ij} is only $(I-1)$, even though there are $\binom{I}{2}$ distinct pairs of cells i and j with $i \neq j$. This is because, despite the fact that for finite n_i the D_{ij} do not satisfy linearity constraints of the form

$$D_{ij} + D_{jk} = D_{ik} \tag{15}$$

discussed above, such constraints are satisfied asymptotically (see (24) below), so that in the limit knowing D_{12} and D_{13}, for example, makes

D_{23} redundant.

Therefore, it is enough to characterize the joint asymptotic behavior of the proper choice of $(I-1)$ of the D_{ij}, for example by choosing a reference cell, cell I, say, and working with the cell pairs $(1,I)$, $(2,I)$, ..., $(I-1,I)$. Lehmann defined for $j = 1, \ldots, I-1$ the quantities

$$V_j = N^{1/2}\,[D_{jI} - (\mu_j - \mu_I)] \tag{16}$$

and showed that

$$(V_1, \ldots, V_{I-1}) \xrightarrow{D} N(0, \Sigma) \;, \tag{17}$$

with Σ having diagonal entries for $j = 1, \ldots, I-1$

$$\lim_{N \to \infty} \operatorname{var}(V_j) = \sigma_R^2 (1/\rho_j + 1/\rho_I) \tag{18}$$

and off-diagonal entries for $1 \le i \ne j \le I-1$

$$\lim_{N \to \infty} \operatorname{cov}(V_i, V_j) = \sigma_R^2 (1/\rho_I) \;, \tag{19}$$

in which

$$\sigma_R^2 = 1/12\theta^2 = 1/12(\textstyle\int f^2)^2 \;. \tag{20}$$

σ_R^2 is the rank-analogue of the underlying error variance σ^2 in the linear model; their ratio

$$e_{R,C}(f) = \sigma^2/\sigma_R^2 = 12\sigma^2(\int f^2)^2 \tag{21}$$

is the asymptotic efficiency of Wilcoxon-type rank-based robust methods relative to their classical counterparts, including the Wilcoxon one- and two-sample tests and the Hodges-Lehmann one- and two-sample estimates based on them. Table 1 gives some values of σ_R^2 and this efficiency for various distributions.

Table 1. σ_R^2 and asymptotic relative efficiency of Wilcoxon-type rank-based methods to the corresponding classical methods.

Distribution f	σ^2	σ_R^2	$e_{R,C}(f) = \sigma^2/\sigma_R^2$
Standard normal	1.0	1.047	0.9549
Standard logistic	3.290	3.0	1.097
χ^2 with 8 degrees of freedom	16.0	13.65	1.172
Skewed mixed normal ($\lambda = 0.75$, $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 1.9$, $\sigma_2 = 2$)	2.426	1.817	1.335
($\lambda = 0.82$, $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 1.9$, $\sigma_2 = 3.5$)	3.558	1.714	2.076
t with 3 degrees of freedom	3.0	1.579	1.900
Any f	—	—	≥ 0.864

The skewed mixed normal distribution referred to in Table 1 has cumulative distribution function (cdf)

$$F(x) = \lambda\Phi[(x-\mu_1)/\sigma_1] + (1-\lambda)\Phi[(x-\mu_2)/\sigma_2] \quad , \qquad (21.1)$$

where Φ is the standard normal cdf.

Returning to the estimation of β, the fact that the asymptotic covariance matrix of the D_{ij} has rank only $(I-1)$ means that it is sensible in estimating β to base the estimate on only $(I-1)$ pairs of cells. Consider then instead of $\bar{\beta}^*$ the estimate

$$\bar{\beta} = \sum_{j=1}^{I-1} \gamma_{jI}\, \hat{\beta}_{jI} \quad . \qquad (22)$$

Expressing this estimate in terms of the D_{ij} gives

$$\bar{\beta} = \sum_{j=1}^{I-1} [\gamma_{jI}/(x_j - x_I)] \left[N^{-1} \sum_{k=1}^{I} n_k (D_{jk} - D_{Ik}) \right] \quad . \qquad (23)$$

Now Lehmann also showed that for all j and k,

$$N^{1/2} D_{jk} = N^{1/2} (D_{jI} - D_{kI}) + o_p(1) \quad , \qquad (24)$$

where $o_p(1)$ as usual signifies a term which goes to 0 in probability as the $n_i \to \infty$. From this, since $D_{ji} = -D_{ij}$,

$$N^{1/2} (D_{jk} - D_{Ik}) = N^{1/2} D_{jI} + o_p(1) \quad , \qquad (25)$$

and

$$N^{1/2}\,\bar{\beta} = \sum_{j=1}^{I-1} [\gamma_{jI}/(x_j - x_I)]\,(N^{1/2} D_{jI}) + o_p(1)\ . \tag{26}$$

Written in terms of Lehmann's V_j above, this becomes

$$N^{1/2}\,\bar{\beta} = \sum_{j=1}^{I} [\gamma_{jI}/(x_j - x_I)]\,V_j + N^{1/2}\sum_{j=1}^{I-1} \gamma_{jI}[(\mu_j - \mu_I)/(x_j - x_I)] + o_p(1). \tag{27}$$

But from (5)

$$(\mu_j - \mu_I)/(x_j - x_I) = \beta\ , \tag{28}$$

so, in order to get asymptotic unbiasedness of $\bar{\beta}$ for β, the reasonable condition

$$\sum_{j=1}^{I-1} \gamma_{jI} = 1 \tag{29}$$

needs to be imposed. When this is done, (27) becomes

$$N^{1/2}(\bar{\beta} - \beta) = \sum_{j=1}^{I-1} [\gamma_{jI}/(x_j - x_I)]V_j + o_p(1)\ . \tag{30}$$

This, combined with Lehmann's result (17), then gives

$$N^{1/2}(\bar{\beta} - \beta) \xrightarrow{D} N(0, v_{\gamma}^2)\ , \tag{31}$$

with

$$v_{\underset{\sim}{\gamma}}^2 = \underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c} , \tag{32}$$

in which $\underset{\sim}{c}$ has entries for $j = 1, \ldots, I-1$

$$c_j = \gamma_{jI}/(x_j - x_I) . \tag{33}$$

The problem thus reduces to that of minimizing $v_{\underset{\sim}{\gamma}}^2$ as a function of $\underset{\sim}{\gamma} = (\gamma_{1I}, \ldots, \gamma_{I-1,I})$, subject to the constraint

$$\underset{\sim}{1}'\underset{\sim}{\gamma} = \underset{\sim}{a}'\underset{\sim}{c} = 1 , \tag{34}$$

where $\underset{\sim}{1}$ is a vector of 1's and for $j = 1, \ldots, I-1$

$$a_j = x_j - x_I . \tag{35}$$

This is a standard problem which can be solved, for example, by a straightforward application of Lagrange multipliers; the solution is expressed in the following well-known (cf. Rao (1973) section 1f)

<u>Lemma 1</u>. Suppose $S = (S_1, \ldots, S_k) \xrightarrow{D} N(0,\underset{\sim}{\Sigma})$ with rank $(\underset{\sim}{\Sigma}) = k$, and let $T = \underset{\sim}{\lambda}'\underset{\sim}{S} = \sum_{i=1}^{k} \lambda_i S_i$ so that $T \xrightarrow{D} N(0,v^2(\lambda))$ with $v^2(\underset{\sim}{\lambda}) = \underset{\sim}{\lambda}'\underset{\sim}{\Sigma}\underset{\sim}{\lambda}$. The $\underset{\sim}{\lambda}$ which minimizes $v^2(\underset{\sim}{\lambda})$ subject to the constraint $\underset{\sim}{a}'\underset{\sim}{\lambda} = 1$ is

$$\underset{\sim}{\lambda}_{opt} = \Sigma^{-1}\, \underset{\sim}{a}/\underset{\sim}{a}'\, \Sigma^{-1}\, \underset{\sim}{a} \quad , \tag{36}$$

and the resulting minimum variance is $v^2(\underset{\sim}{\lambda}_{opt}) = 1/a'\, \Sigma^{-1}\, \underset{\sim}{a}$.

From this the solution to the problem above is evidently

$$\underset{\sim}{\lambda}_{opt} = \begin{bmatrix} (x_1 - x_I) & & 0 \\ & \ddots & \\ 0 & & (x_{I-1} - x_I) \end{bmatrix} \underset{\sim}{c}_{opt} \quad , \tag{37}$$

where

$$\underset{\sim}{c}_{opt} = \Sigma^{-1}\, \underset{\sim}{a}/\underset{\sim}{a}'\, \Sigma^{-1}\, \underset{\sim}{a} \quad , \tag{38}$$

and the minimizing variance is

$$v^2_{\underset{\sim}{\Upsilon}_{opt}} = 1/\underset{\sim}{a}'\, \Sigma^{-1}\, \underset{\sim}{a} \quad . \tag{39}$$

Now Σ^{-1} has diagonal elements for $j = 1, \ldots, I - 1$,

$$(\Sigma^{-1})_{jj} = \left(\rho_j \sum_{i \neq j} \rho_i \right) \Big/ \sigma_R^2 \quad , \tag{40}$$

and off-diagonal elements for $1 \leq i \neq j \leq I - 1$

$$(\Sigma^{-1})_{ij} = -\rho_i \rho_j / \sigma_R^2 \quad . \tag{41}$$

After simplification the optimal weights work out to

$$(\gamma_{jI})^*_{opt} = \rho_j(x_j - x_I) \sum_{k=1}^{I} \rho_k(x_j - x_k)/\sigma_x^2 \quad , \tag{42}$$

and the minimum variance comes out

$$v^2_{\underline{\gamma}_{opt}} = \sigma_R^2/\sigma_x^2 \quad . \tag{43}$$

Here

$$\sigma_x^2 = \sum_{k=1}^{I} \rho_k(x_k - \bar{x})^2 \tag{44}$$

is the limiting variance of the x-values,

$$\bar{x} = \sum_{i=1}^{I} \rho_i x_i = \lim_{N \to \infty} \bar{x}_N \tag{45}$$

is their limiting mean, and $\bar{x}_N$ is the mean of the N x-values,

$$\bar{x}_N = N^{-1} \sum_{i=1}^{I} n_i x_i \quad . \tag{46}$$

In practice one of course would substitute n_i/N for ρ_i and use the weights

$$(\gamma_{jI})_{opt} = \left[n_j(x_j - x_I) \sum_{k=1}^{I} n_k(x_j - x_k) \right] \Big/ \left[N \sum_{k=1}^{I} n_k(x_k - \bar{x}_N)^2 \right] \quad . \tag{47}$$

Note that these weights need not all be in the range $(0,1)$; they need only sum to 1.

How does the asymptotic variance (43) compare to that for previous rank-based robust slope estimates? The best estimator in this class to date was developed independently by Adichie (1967) and Jaeckel (1972), although Adichie did not realize that his estimator had a closed-form expression and Jaeckel did not recognize his estimator to be the solution of Adichie's iterative procedure. Rewriting the model (1,4) in more general form as

$$Y_i = \alpha + \beta x_i + e_i \ , \qquad i = 1, \ldots, N \ , \tag{48}$$

the Adichie-Jaeckel slope estimate is a weighted median of the set

$$\left\{ (Y_j - Y_i)/(x_j - x_i), \quad (i,j) \ni x_i \neq x_j \right\} \tag{49}$$

of all pairwise slopes, in which the weights are proportional to the absolute distance $|x_i - x_j|$ between the independent variable values. Adichie and Jaeckel both showed that the asymptotic efficiency of this estimator relative to the classical least-squares estimator is given by (21), and their asymptotic variance expressions, when specialized to the model (1,4), coincide with that (43) of the above estimator $\bar{\beta}$ (22) with optimal weights (47). Thus asymptotically the estimator (22) and Jaeckel's estimator are equally accurate.

It is interesting to contrast the above rank-based estimator (22) with Jaeckel's in terms of their method of estimation, since both are in a sense generalizations of the two-sample Hodges-Lehmann estimator. The idea behind the method described here is to compute many robust estimates of β first, by working with pairwise slopes, and then take a nonrobust (weighted) average of the separate estimates; Jaeckel's procedure in effect does the same thing but in the reverse order, by constructing many nonrobust pairwise estimates of β and then taking a robust average (weighted median) of them. It is not intuitively obvious that the two operations being performed should commute in such a way that the results have the same asymptotic behavior, but it is seen that in this case they do.

The above results are summmarized in the following

Proposition 1. In the model (1) specialized to (4) above, let the $n_i \to \infty$ in such a way that

$$n_i/N \to \rho_i \in (0,1) \quad \text{for all} \quad i = 1, \dots, I \ . \tag{50}$$

Then with weights $\underset{\sim}{\gamma} = (\gamma_{1I}, \dots, \gamma_{I-1,I})$ given by (47), the estimator

$$\bar{\beta} = \sum_{j=1}^{I-1} [\gamma_{jI}/(x_j - x_I)] \, W_{jI} \ , \tag{51}$$

in which the W_{ij} are as in (11), is asymptotically normal with (asymptotic) mean β and variance σ_R^2/σ_x^2, in which σ_x^2 is given by (44) and $\sigma_R^2 = 1/12\theta^2$ with $\theta = \int f^2$. This is the smallest asymptotic variance attainable by an asymptotically unbiased estimator of the form (51). The asymptotic efficiency of $\bar{\beta}$ relative to the classical least-squares estimator is σ^2/σ_R^2, where σ^2 is the variance of the error density f.

In addition to working out the asymptotic distribution of the D_{ij}, Lehmann (1963a) also showed that D_{ij} is symmetrically distributed about $\mu_i - \mu_j$ if either the underlying error density f is symmetric about 0 or the sample sizes n_i are all equal. Since the weights used in constructing $\bar{\beta}$ sum to 1, this implies that under either of these conditions $\bar{\beta}$ is unbiased for β.

Note that when the classical least-squares estimator of β is expressed in the same form as $\bar{\beta}$,

$$\hat{\beta}_C = \sum_{j=1}^{I-1} [\lambda_{jI}/(x_j - x_I)]\, T_{jI} \quad , \tag{52}$$

where

$$T_{ij} = Y_{i\cdot} - Y_{j\cdot} \tag{53}$$

in which $Y_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$, the weights λ_{ij} coincide with the optimal weights (47) for $\bar{\beta}$. This is due to the fact that the statistics W_{ij} (11) and T_{ij} (53) have the same asymptotic covariance structure up to a multiplicative constant, a point noted by Lehmann (1963b). This relationship could be used to provide a somewhat shorter but perhaps less illuminating proof of Proposition 1, when combined with the fact that the classical least-squares estimator is uniform minimum variance unbiased (UMVU) for β.

The above proposition can serve as the basis for large-sample confidence procedures and tests for the slope parameter β, in combination with an estimator of σ_R^2. The proposition argues that for large N

$$\bar{\beta} \;\dot{\sim}\; N\left(\beta,\ \sigma_R^2/N\sigma_x^2\right) \quad , \tag{54}$$

so with $\hat{\sigma}_R^2$ as any consistent estimate of σ_R^2 an approximate $100(1-\lambda)\%$ distribution-free confidence interval for β is

$$\bar{\beta} \;\pm\; \Phi^{-1}\,(1-\lambda/2)\;\hat{\sigma}_R/N^{1/2}\sigma_x^* \quad , \tag{55}$$

where

$$\sigma_x^* \;=\; \left[N^{-1}\sum_{i=1}^{I} n_i(x_i - \bar{x}_N)^2\right]^{1/2} \quad . \tag{56}$$

Critical regions for asymptotically distribution-free tests concerning β are constructed analogously.

The estimation of σ_R^2 in linear models with several observations per cell is investigated by Draper (1982). Several estimators of σ_R^2 based on the lengths of distribution-free confidence intervals associated with the one- and two-sample Hodges-Lehmann estimators are proposed and shown to have good small-samples properties. The best of these, the so-called Lehmann-type two-sample estimator with bias correction based on pairs of cells, has the form

$$\left[\hat{\sigma}_R^2(\alpha)\right]_2 \;=\; \left[\sum_{i=1}^{I-1}\sum_{j=i+1}^{I} n_{ij}\;T_{ij}(\alpha)\right]^2\Big/12 \quad , \tag{57}$$

in which (for $I \geq 3$)

$$\eta_{ij} = 2[(I-1)(n_i + n_j) - N]/[N(I-1)(I-2)] \quad (58)$$

(for $I = 2$ η_{12} is of course 1), and

$$T_{ij}(\alpha) = \left[3n_i n_j \Big/ (n_i + n_j)\right]^{1/2} L_{ij}(\alpha) \Big/ t^{-1}_{n_i+n_j-2}(1-\alpha/2) , \quad (59)$$

where $L_{ij}(\alpha)$ is the length of the $100(1-\alpha)\%$ confidence interval for $(\mu_i - \mu_j)$ based on the Wilcoxon rank-sum statistic (cf. Lehmann (1963c)) and t_k^{-1} is the inverse t cdf with k degrees of freedom. From simulation work focusing on the bias of $[\hat{\sigma}_R^2(\alpha)]_2$ the best choice for the confidence level in general appears to be $100(1-\alpha)\% \doteq 50\%$. See Draper (1982) for other estimators of σ_R^2 in this context and a discussion of their relative merits.

Small-sample refinements of the above large-sample approach to inference about β, including obtaining confidence coefficients and critical regions from a distribution other than the normal, could be investigated with a simulation study.

A natural way to estimate the intercept parameter in the model (1,4) once the slope β has been estimated by $\bar{\beta}$ is to regard α as the center of the distribution of the random variables $e^*_{ij} = Y_{ij} - \beta x_i$ and to apply some estimate of location to the residuals

$$\hat{e}^*_{ij} = Y_{ij} - \bar{\beta} x_i \quad . \quad (60)$$

In order that α be identifiable, an assumption on the manner in which the errors e_{ij} are centered at 0 is necessary; for example, if $E(e_{ij}) = 0$ is assumed then α becomes the expectation of the e^*_{ij}. With the further assumption of symmetry of the error density about 0, one reasonable robust estimator of location to apply to the $\hat{e}^*_{ij}$ is the one-sample Hodges-Lehmann estimate. Renumbering the residuals from 1 to N as $\hat{e}^*_i$, a Hodges-Lehmann-type estimator of α is then

$$\hat{\alpha} = \text{med}\left\{(\hat{e}^*_i + \hat{e}^*_j)/2, \quad 1 \le i \le j \le N\right\} \quad . \tag{61}$$

3. <u>More than One Independent Variable</u>

Here are sketches of extensions of the above ideas to the case $p > 1$ and to the analysis of covariance.

Working again with the model (1),

$$Y_{ij} = \mu_i + e_{ij} \,, \qquad \left\{ \begin{matrix} i = 1, \ldots, I & \sum_{i=1}^{I} n_i = N \\ j = 1, \ldots, n_I & \end{matrix} \right\} , \tag{62}$$

with, for $p > 1$,

$$\mu_i = \beta_0 + \sum_{k=1}^{p} x_{ik}\beta_k \quad , \tag{63}$$

the approach is, as in the case $p = 1$, to estimate $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)$ first, with $\hat{\underset{\sim}{\beta}}$, say, and then to estimate β_0 from the residuals $Y_{ij} - \sum_{k=1}^{p} x_{ik}\hat{\beta}_k$.

To obtain an intuitively reasonable estimate of $\underline{\beta}$ using the ideas of section 2, note that for $i_1 \neq i_2$,

$$\mu_{i_1} - \mu_{i_2} = \sum_{k=1}^{p} \left(x_{i_1 k} - x_{i_2 k} \right) \beta_k \quad , \tag{64}$$

one equation in the p unknowns β_k. The method used in section 2 for $p = 1$ was first to solve this one equation for β_1, and then to estimate $\mu_{i_1} - \mu_{i_2}$; here it is necessary first to generate and solve $p > 1$ linearly independent equations in the β_k. Choose $p + 1$ integers (cells) $i_1, \ldots ,i$ from $\{1, \ldots ,I\}$; there are $\binom{I}{p+1} \equiv M$ such choices. Number these $m = 1, \ldots ,M$ and, for each such choice, form the system of equations

$$\left\{ \begin{array}{rcl} \mu_{i_1} - \mu_{i_2} & = & \sum_{k=1}^{p} \left(x_{i_1 k} - x_{i_2 k} \right) \beta_k \\ & \vdots & \\ \mu_{i_1} - \mu_{i_{p+1}} & = & \sum_{k=1}^{p} \left(x_{i_1 k} - x_{i_{p+1} k} \right) \beta_k \end{array} \right\} \quad , \tag{65}$$

or in matrix terms

$$X_m \underline{\beta} = \underline{\mu}_m \quad , \tag{66}$$

where

$$
{}_p(x_m)_p = \begin{bmatrix} \left(x_{i_1 1} - x_{i_2 1}\right) & \cdots & \left(x_{i_1 p} - x_{i_2 p}\right) \\ \vdots & \ddots & \vdots \\ \left(x_{i_1 1} - x_{i_{p+1} 1}\right) & \cdots & \left(x_{i_1 p} - x_{i_{p+1} p}\right) \end{bmatrix} \qquad (67)
$$

and

$$
\underline{\mu}_m' = \left(\mu_{i_1} - \mu_{i_2}, \ldots, \mu_{i_1} - \mu_{i_{p+1}}\right) \quad . \qquad (68)
$$

Provided the cells $i_1, \ldots, i_{p+1}$ have been chosen in such a way that x_m is full rank p, the solution of (66) for $\underline{\beta}$ is of course

$$
\underline{\beta} = x_m^{-1} \underline{\mu}_m \quad , \qquad (69)
$$

true for each $m = 1, \ldots, M'$, where M' is M minus the number S of choices of $(p+1)$ cells which result in singular x matrices. ($p+1$ points do not always determine a p-dimensional hyperplane.) Thus as in section 2 each of the M' choices of $(p+1)$ cells can generate a separate estimate of $\underline{\beta}$ through estimation of $\underline{\mu}_m$, and then a composite estimate of $\underline{\beta}$ can be constructed by combining the separate estimates.

The number S of inadequate choices of $(p+1)$ cells depends on the design configuration and is typically quite a bit smaller than $\binom{I}{p+1}$. For example, for the design illustrated in Figure 1 below, of the $M = \binom{8}{3} = 56$ choices of $(p+1) = 3$ cells only the $S = 4$ choices determining the indicated lines fail to span the $x_1 - x_2$ plane. (The analogous

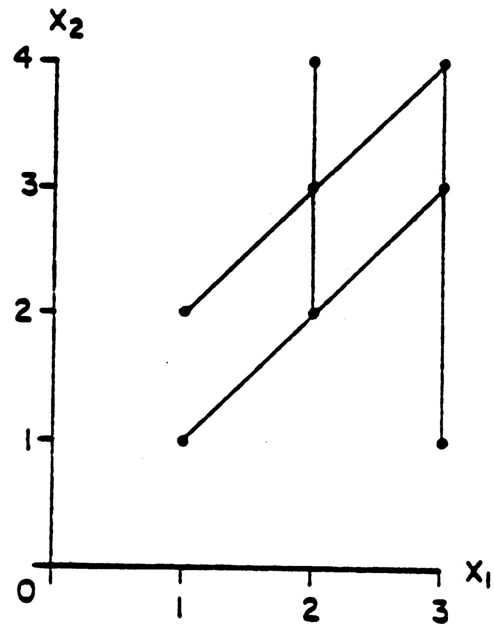

Figure 1. A multiple regression design with $p = 2$, $I = 8$.

problem in the case $p = 1$ would have manifested itself as an attempt to estimate the slope of a vertical line, and the problem could not arise in that case since the values of the independent variable in the different cells were all distinct.)

As in section 2 a good robust estimate of $\underline{\mu}_m$ is already available, namely

$$\hat{\underline{\mu}}_m = \begin{bmatrix} W_{i_1 i_2} \\ \vdots \\ W_{i_1 i_{p+1}} \end{bmatrix} , \tag{70}$$

where W_{ij} is as in (11). Thus a Hodges-Lehmann-type estimator of $\underset{\sim}{\beta}$ based on choice m of (p + 1) of the cells is

$$\hat{\underset{\sim}{\beta}}_m = x_m^{-1} \hat{\underset{\sim}{\mu}}_m \quad \text{for each} \quad m = 1, \ldots, M' \quad . \tag{71}$$

As in section 2, perhaps the most efficient way to combine these separate estimates is through a weighted average. Let $\hat{B}$ be the $(p \times M')$ matrix whose columns are the $\hat{\underset{\sim}{\beta}}_m$; then the analogue of the weighted average in section 2 is

$$\hat{\underset{\sim}{\beta}} = \underset{p\,M'}{\hat{B}} \, \underset{\sim}{\gamma}_1 \quad , \tag{72}$$

where $\underset{\sim}{\gamma}$ is a vector of weights chosen to minimize some function of the asymptotic covariance matrix of $\hat{\underset{\sim}{\beta}}$ while preserving asymptotic unbiasedness. As before, since the rank of the asymptotic covariance matrix of the W_{ij} is only $(I - 1)$, a greatly reduced subset of the $\hat{\underset{\sim}{\beta}}_m$ suffices in the construction of the composite estimate. The choice of which of the $\hat{\underset{\sim}{\beta}}_m$ to use and the determination of the optimal weights $\underset{\sim}{\gamma}$ can then be carried out in a manner analogous to that in section 2.

Note that for $p > 1$ the matrix inversion in (71) is the analogue of division by $x_i - x_j$ in (12) in the case $p = 1$. This approach thus seems less attractive for large p than for small since the matrix inversions involved may prove relatively costly.

The ideas of section 2 also extend naturally to the analysis of covariance, either by using dummy variables to express the qualitative factors in regression coding and proceeding as above or by preserving the mixed ANOVA-regression notation and doing the obvious thing. For example, in the analysis of covariance model

$$Y_{ij} = \eta_{ij} + e_{ij} \quad , \qquad \left\{ \begin{array}{l} i = 1, \ldots, I \\ j = 1, \ldots, n_i \end{array} \right\} \quad , \tag{73}$$

where

$$\eta_{ij} = \beta_i + \gamma z_{ij} \quad , \tag{74}$$

for each choice of i and $j_1 \neq j_2$ one obtains

$$\left(\eta_{ij_1} - \eta_{ij_2}\right) \Big/ \left(z_{ij_1} - z_{ij_2}\right) = \gamma \quad , \tag{75}$$

so separately for each level of the ANOVA factor the problem reduces to that studied above in section 2 and a composite estimate of γ can then be formed from the separate estimates.

It should in principle be feasible to derive exact confidence regions and tests which might be preferable to the asymptotic ones described in section 2 by using in some way the information provided by the spread in the individual regression estimates (12) and (71), in the manner of Sen (1968) and others. This approach has not been pursued here.

The estimation of β_0 in the case of more than one independent variable can of course be done in the same way as was outlined in section 2, by calculating the one-sample Hodges-Lehmann estimate applied to the residual vector

$$\hat{e}^{*}_{ij} = Y_{ij} - \sum_{k=1}^{p} x_{ik}\hat{\beta}_k \quad . \tag{76}$$

This would, as before, require the further assumption of symmetry of the error density f about 0.

Comparisons of the small-sample behavior of these new regression estimates relative to Jaeckel's and to the classical estimates are worthwhile because, while the new estimator and Jaeckel's have the same asymptotic performance, one or the other may have an advantage in small-sample bias, standard error, or computational efficiency; and both the new estimator and Jaeckel's may exhibit better or worse small-sample properties relative to the classical estimates in practice than those indicated by the asymptotic results. This will be the subject of future investigation.

REFERENCES

Adichie, J. N. (1967). "Estimates of regression parameters based on rank tests," Annals of Mathematical Statistics, Vol. 38, pp. 894-904.

Draper, David C. (1982). "Rank-based robust analysis of linear models. II. Estimation of the asymptotic variance of rank-based estimators and associated statistics," technical report, Department of Statistics, University of California/Berkeley.

Huber, Peter J. (1981). Robust Statistics. New York: Wiley.

Jaeckel, Louis A. (1972). "Estimating regression coefficients by minimizing the dispersion of the residuals," Annals of Mathematical Statistics, Vol. 43, pp. 1449-1458.

Lehmann, E. L. (1963a). "Robust estimation in analysis of variance," Annals of Mathematical Statistics, Vol. 34, pp. 957-966.

Lehmann, E. L. (1963b). "Asymptotically nonparametric inference: an alternative approach to linear models," Annals of Mathematical Statistics, Vol. 34, pp. 1494-1506.

Lehmann, E. L. (1963c). "Nonparametric confidence intervals for a shift parameter," Annals of Mathematical Statistics, Vol. 34, pp. 1507-1512.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Mood, A. M. (1950). Introduction to the Theory of Statistics. New York: Wiley.

Rao, C. Radhakrishna (1973). Linear Statistical Inference and Its Applications, second edition. New York: Wiley.

Sen, P. K. (1968). "Estimates of the regression coefficient based on Kendall's Tau," Journal of the American Statistical Association, Vol. 63, pp. 1379-1389.

Spjøtvoll, Emil (1968). "A note on robust estimation in analysis of variance," Annals of Mathematical Statistics, Vol. 39, pp. 1486-1492.

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