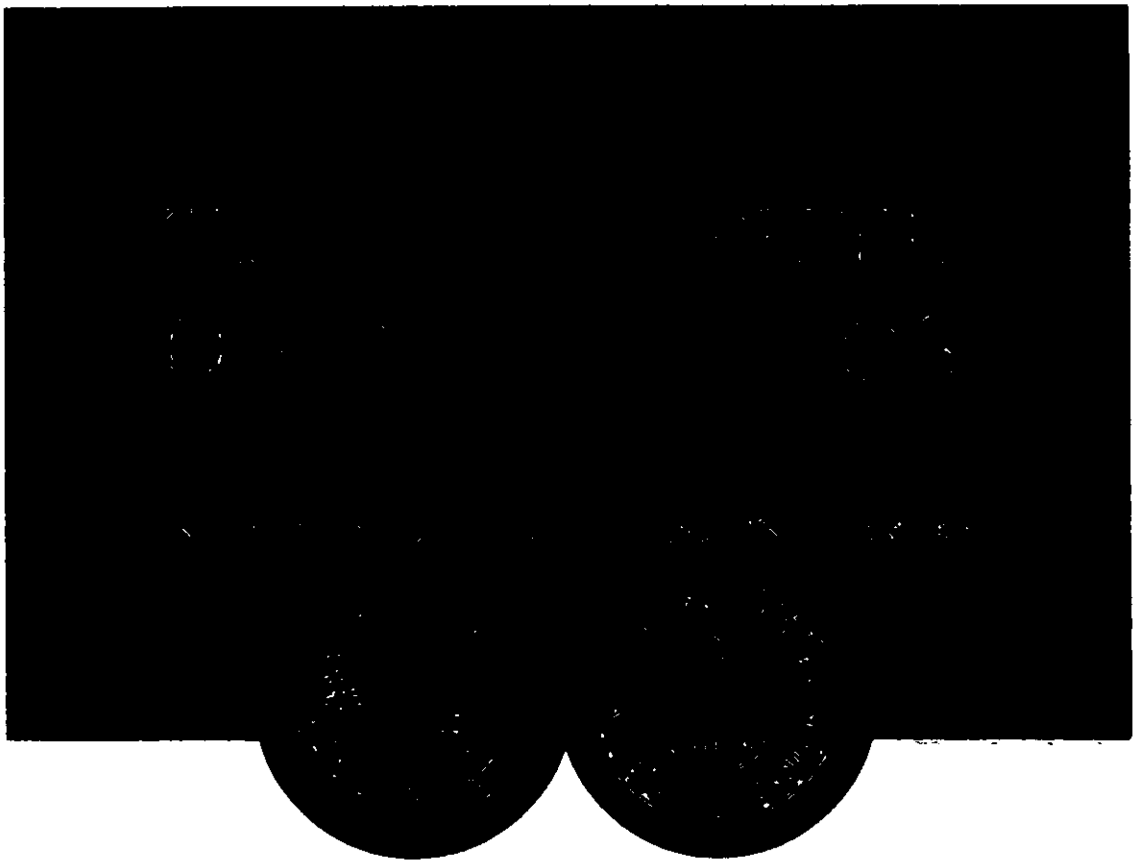

ON ESTIMATING A SLOPE AND INTERCEPT
IN A NONPARAMETRIC STATISTICS COURSE

E. Jacquelin Dietz

Institute of Statistics Mimeo Series No. 1689 R
North Carolina State University
Raleigh, NC 27695

July 1987

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

On Estimating a Slope and Intercept in a Nonparametric Statistics Course

E. Jacquelin Dietz

Department of Statistics, Box 8203

North Carolina State University

Raleigh, NC 27695

E. Jacquelin Dietz is Associate Professor in the Department of Statistics, Box 8203, North Carolina State University, Raleigh, NC 27695. The author thanks Dennis D. Boos for many very helpful discussions and the referees for thorough reviews of earlier drafts.

ABSTRACT

Various estimators of slope, intercept, and mean response in the simple linear regression problem are compared in terms of unbiasedness, efficiency, breakdown, and mean squared error. Theil's estimator of slope and two intercept estimators based on Theil's estimator are recommended for inclusion in nonparametrics courses as robust, efficient, and easy-to-calculate alternatives to least-squares.

KEY WORDS: Linear regression; Robust estimation; Point estimation; Breakdown; Theil estimator.

1. INTRODUCTION

Few teachers of an introductory statistics course would think of omitting the topic of simple linear regression. Yet that topic is either omitted entirely or covered incompletely in most popular nonparametric statistics texts. Many such books include a distribution-free test or confidence interval for the slope of the regression line, but omit any discussion of the intercept. Those books that do discuss the intercept offer a variety of estimators without providing any basis for choosing among them. The problem of estimating the mean response at a given x is also neglected in these texts.

In this article, I review various point estimators for the slope, intercept, and mean response in the simple linear regression problem. None of the estimators are new, but many are not well known. I illustrate the calculation of the estimators with an example. After considering unbiasedness, efficiency, breakdown, and mean squared error (MSE) of these estimators, I make the following recommendations: Theil's estimator of slope is robust, easy to compute, and competitive in terms of MSE with alternative slope estimators. For symmetric error distributions, the median of pairwise averages of residuals based on Theil's estimator provides an attractive estimator of the intercept. If the errors are asymmetric or heavily contaminated with outliers, the median of residuals based on Theil's estimator is preferable.

2. THE ESTIMATORS

I assume the model $Y_i = \alpha + \beta x_i + e_i$, $i = 1,2,\ldots,n$, where α and β are unknown parameters, $x_1 \leq x_2 \leq \ldots \leq x_n$ are known constants (not all equal), and the e_i's are independent and identically distributed continuous random variables with mean zero.

2.1 Slope Estimators

The estimators of β considered in this paper can all be viewed as functions of the N sample slopes

$$S_{ij} = (Y_j - Y_i)/(x_j - x_i) \ , \ i < j \ , \ x_i \neq x_j .$$

If the x_i's are all distinct, $N = \binom{n}{2}$.

The least-squares estimator of β, $\hat{\beta}_{LS}$, is a weighted average of the S_{ij}'s. Specifically,

$$\hat{\beta}_{LS} = \sum_{i<j} w_{ij} S_{ij} / \sum_{i<j} w_{ij} \ ,$$

where $w_{ij} = (x_j - x_i)^2$.

Randles and Wolfe (1979, Problem 3.1.6) suggest as an estimator of β the unweighted average of the S_{ij}'s,

$$\hat{\beta}_A = \sum_{i<j} S_{ij}/N \ .$$

The estimator of β mentioned most frequently in textbooks on nonparametric statistics is the Theil (1950) estimator, $\hat{\beta}_M$, the median of the S_{ij}'s.

Sievers (1978) and Scholz (1978) generalize the Theil estimator by assigning a weight of w_{ij} to each S_{ij} ($w_{ij} = 0$ for $x_i = x_j$). Their estimator of β is the median of the probability distribution obtained by assigning probability $w_{ij} / \sum_{i<j} w_{ij}$ to S_{ij}. The Theil estimator corresponds to $w_{ij} = 1$ for all $i < j$, $x_i \neq x_j$. Other weights considered by Sievers and Scholz are w_{ij} = j–i and w_{ij} = $x_j - x_i$ (recommended by Sievers). The estimators corresponding to these two sets of weights will be referred to as $\hat{\beta}_{W1}$ and $\hat{\beta}_{W2}$, respectively.

These estimators of β are summarized in Table 1.

(Insert Table 1 here).

2.2 Intercept Estimators

The estimators of α considered in this paper can be divided into two groups, those that are functions of the N sample intercepts

$$A_{ij} = (x_jY_i - x_iY_j)/(x_j - x_i)\ ,\ i < j,\ x_i \neq x_j\ ,$$

and those based on residuals associated with a particular slope estimate.

The least-squares estimator of α, $\hat{\alpha}_{LS}$, is a weighted average of the A_{ij}'s. Specifically,

$$\hat{\alpha}_{LS} = \sum_{i<j} w_{ij}A_{ij} / \sum_{i<j} w_{ij}\quad ,$$

where $w_{ij} = (x_j - x_i)^2$.

Randles and Wolfe (1979, Problem 3.1.6) suggest as an estimator of α the unweighted average of the A_{ij}'s ,

$$\hat{\alpha}_A = \sum_{i<j} A_{ij}/N.$$

The median of the A_{ij}'s, $\hat{\alpha}_M$, provides an estimator of α analogous to Theil's estimator of β. Maritz (1979) shows that the test statistic corresponding to this estimator is not distribution-free. He considers instead tests and estimators based on the median of a subset of the A_{ij}'s. (Maritz considers only the case of distinct x's.) Since the emphasis here is on point estimation, I use the median of all the A_{ij}'s.

These estimators of α, all functions of the A_{ij}'s, are summarized in Table 1.

Another class of intercept estimators is obtained by taking the median or median of pairwise averages of $Y_i - \hat{\beta}x_i$, $i = 1,2,\ldots,n$, where $\hat{\beta}$ is some estimator of β. Bhattacharyya (1968) considers the median of pairwise averages of the $(Y_i - \hat{\beta}x_i)$'s, where $\hat{\beta}$ is either the Theil estimator $\hat{\beta}_M$ or the weighted median estimator $\hat{\beta}_{W1}$ with weights j–i. (In Bhattacharyya's paper, $x_i = i$ for $i = 1,2,\ldots,n$.) Hettmansperger (1984) considers both the median and median of pairwise averages of the $(Y_i - \hat{\beta}_{W2}\, x_i)$'s. Members of this class of estimators are listed and named in Table 2.

(Insert Table 2 here)

Finally, Conover (1980, p. 267) proposes the estimator

$$\hat{\alpha}_C = Y_{.50} - \hat{\beta}_M x_{.50} ,$$

where $Y_{.50}$ and $x_{.50}$ are the sample medians of the Y's and x's, respectively. Note the analogy with the least-squares estimator, which can be written as $\hat{\alpha}_{LS} = \bar{Y} - \hat{\beta}_{LS}\bar{x}$, where $\bar{Y}$ and $\bar{x}$ are the sample means of the Y's and x's, respectively.

2.3 Estimators of Mean Response

The mean response at a given x value, $E(Y) = \alpha + \beta\, x$, is often a more interesting parameter than the intercept α. (Of course, if $x = 0$, then $E(Y) = \alpha$.) The estimators of E(Y) considered here are of the form $\hat{\alpha} + \hat{\beta}\, x$, where each $\hat{\alpha}$ in Tables 1 and 2 is associated with the $\hat{\beta}$ given in the same row of the table, and $\hat{\alpha}_C$ is associated with $\hat{\beta}_M$.

There are, of course, many other estimators of slope and intercept which could have been considered here. In fact, Simon (1982) compares more than forty estimators of slope, all functions of the S_{ij}'s. I have tried to select estimators that are simple to understand and easy to compute.

3. UNBIASEDNESS AND SYMMETRY

In this section, I consider unbiasedness and symmetry properties possessed by the estimators. Since $E(S_{ij}) = \beta$, it follows that $\hat{\beta}_{LS}$ and $\hat{\beta}_A$ are unbiased estimators of β. Sen (1968, sec. 5) shows that the distribution of $\hat{\beta}_M$ is symmetric about β, and Theorem 5 of Sievers (1978) implies that $\hat{\beta}_{W1}$ and $\hat{\beta}_{W2}$ are asymptotically unbiased, under certain conditions on the x's.

At the beginning of Section 2, I assumed that the e_i's have mean zero. Then $E(A_{ij}) = \alpha$, which implies that $\hat{\alpha}_{LS}$ and $\hat{\alpha}_A$ are unbiased estimators of α. The alternative assumption that the e_i's have median zero leads to a reparameterization more appropriate for most of the other intercept estimators. The simulation study of Dietz (1986) suggests that under this alternative parameterization, $\hat{\alpha}_M$, $\hat{\alpha}_{1,M}$, $\hat{\alpha}_{1,W1}$, $\hat{\alpha}_{1,W2}$, and $\hat{\alpha}_C$ are median unbiased estimators of α (that is, each has a distribution with median α). The simulation study also suggests that $\hat{\alpha}_{2,M}$, $\hat{\alpha}_{2,W1}$, and $\hat{\alpha}_{2,W2}$ are estimators of α plus the pseudomedian of e_i (Hollander and Wolfe 1973, p. 458). For symmetric error distributions, these various parameterizations coincide and all of the intercept estimators estimate the same parameter.

If the distribution of the e_i's is symmetric about zero, every estimator of α or β in Tables 1 and 2 has a distribution which is symmetric about the value of the corresponding parameter. Symmetry of $\hat{\alpha}$ and $\hat{\beta}$ implies symmetry of $\hat{\alpha} + \hat{\beta}$ x about $\alpha + \beta$ x. The distribution of $\hat{\alpha}_C$ is symmetric about α if $\beta = 0$; however, it does not seem possible to demonstrate any unbiasedness property for $\hat{\alpha}_C$ for general β.

4. EFFICIENCY

The asymptotic relative efficiency of one estimator with respect to another is defined as the ratio of the asymptotic variances of the two estimators. Let $\sigma^2 = \mathrm{Var}(e_i)$ and let f be the density function of the e_i's. Then $e_S = 4\sigma^2 f^2(0)$ and $e_W = 12\sigma^2[\int f^2(x)dx]^2$ are the well-known efficiencies of the sample median and the sample median of pairwise averages, respectively, with respect to the sample mean.

Sen (1968) shows that, under suitable conditions on the x's, the efficiency of the Theil estimator $\hat{\beta}_M$ with respect to $\hat{\beta}_{LS}$ is $\rho^2 e_W$, where $\rho^2 = \lim_{n \to \infty} \rho_n^2$, and ρ_n is the product moment correlation coefficient between $(x_1, x_2, \ldots, x_n)$ and $(1, 2, \ldots, n)$. For equally spaced x's, ρ equals one, and the efficiency of $\hat{\beta}_M$ achieves its maximum value of e_W.

Scholz (1978) shows that weighted median estimators of β (e.g., $\hat{\beta}_{W1}$, $\hat{\beta}_{W2}$, and $\hat{\beta}_M$) cannot have efficiency exceeding e_W relative to $\hat{\beta}_{LS}$. The estimator $\hat{\beta}_{W2}$ achieves the optimal efficiency of e_W; the efficiency of $\hat{\beta}_{W1}$ equals that of $\hat{\beta}_M$ (Sievers 1978). (Note that both Sievers (1978, Example 3) and Lehmann (1975, p. 313) erroneously attribute to Bhattacharyya (1968) the claim that weights $w_{ij} = j-i$ and $w_{ij} = 1$ lead to the same estimator of β. The resulting estimators, $\hat{\beta}_{W1}$ and $\hat{\beta}_M$, are equivalent in terms of asymptotic efficiency, but are not identical. In fact, Bhattacharyya proposes $\hat{\beta}_{W1}$ and $\hat{\beta}_M$ as alternative estimators.)

Direct computation shows that for distinct x's, Var $(\hat{\beta}_A)$ and Var $(\hat{\alpha}_A)$ increase without bound as the smallest distance between any two x values decreases to zero. As a consequence, $\hat{\beta}_A$ and $\hat{\alpha}_A$ can have very low efficiency relative to other estimators. (In fairness to Randles and Wolfe (1979), they do not recommend these estimators; they arise merely as answers to a textbook exercise.)

Incomplete information is available concerning the asymptotic relative efficiencies of the intercept estimators. Assume for the remainder of this section that f is symmetric, so that the intercept parameter is well-defined.

Hettmansperger (1984, p. 250) shows that the efficiency of $\hat{\alpha}_{2,W2}$ relative to $\hat{\alpha}_{LS}$ is e_W, the efficiency of the median of pairwise averages. He also gives results (Hettmansperger 1984, p. 251) that imply that the efficiency of $\hat{\alpha}_{1,W2}$ relative to $\hat{\alpha}_{LS}$ is given by

$$\frac{1 + (\mu^2/\sigma_x^2)}{e_S^{-1} + (\mu^2/\sigma_x^2)e_W^{-1}} \quad , \qquad (1)$$

where $\mu = \lim_{n \to \infty} \bar{x}$ and $\sigma_x^2 = \lim_{n \to \infty} \sum_{i=1}^{n} (x_i - \bar{x})^2/n$. If the x's are centered so that $\mu = 0$, then expression (1) reduces to e_S, the efficiency of the sample median. On the other hand, for large μ^2/σ_x^2, expression (1) is dominated by e_W. Note that in this latter case, we are less likely to be interested in making inferences about α, since a very large value of μ^2/σ_x^2 implies that zero is outside the range of the x values.

Maritz (1979) gives an efficiency result for an estimator related to $\hat{\alpha}_M$, but based on a certain subset of the A_{ij}'s. For equally spaced x values and normal errors, the efficiency of his estimator with respect to $\hat{\alpha}_{LS}$ is $2/\pi$, the value of e_S for normal f.

Note that two nonparametric estimators $\hat{\beta}_i$ and $\hat{\beta}_j$ (or $\hat{\alpha}_i$ and $\hat{\alpha}_j$) can be compared using

$$ARE(\hat{\beta}_i, \hat{\beta}_j) = ARE(\hat{\beta}_i, \hat{\beta}_{LS})/ARE(\hat{\beta}_j, \hat{\beta}_{LS}),$$

where $ARE(\hat{\beta}_i, \hat{\beta}_j)$ is the asymptotic relative efficiency of $\hat{\beta}_i$ with respect to $\hat{\beta}_j$ (Randles and Wolfe 1979, Problem 5.2.3).

5. BREAKDOWN

Donoho and Huber (1983) define the following notion of the breakdown point of an estimator. Suppose a fixed sample of size n is corrupted by replacing an arbitrary subset of size m from the sample by arbitrary values. Then the corrupted sample contains a fraction $\varepsilon = m/n$ of contaminated values. The breakdown point ε^* of an estimator is the smallest value of ε for which the estimator, when applied to the corrupted sample, can take values arbitrarily far from the value of the estimator for the uncorrupted sample. Donoho and Huber discuss two types of corruption in the simple linear regression problem: corruption in only the Y's and corruption in both the x's and Y's.

For either type of corruption, $\hat{\alpha}_{LS}$, $\hat{\beta}_{LS}$, $\hat{\alpha}_A$, and $\hat{\beta}_A$ have breakdown point $1/n$; that is, a single bad observation can cause these estimators to behave arbitrarily badly.

The Theil estimator $\hat{\beta}_M$ breaks down if and only if at least half of the S_{ij} are contaminated. The resulting breakdown point equals that of the median of pairwise averages, that is, approximately $1 - 2^{-1/2} \doteq .293$ for large n (Donoho and Huber

1983). The same reasoning implies that $\hat{\alpha}_M$ has this same breakdown point.

The estimators of α that depend on $\hat{\beta}_M$, namely $\hat{\alpha}_{1,M}$, $\hat{\alpha}_{2,M}$, and $\hat{\alpha}_C$, also have breakdown point $\varepsilon^* \approx .293$, because of $\hat{\beta}_M$. (However, if $x_{.50} = 0$, then $\hat{\alpha}_C$ has breakdown point .5. This would be the case if we "median-center" the x values.)

The weighted median estimator $\hat{\beta}_{W1}$ with weights j-i breaks down if and only if the sum of weights corresponding to contaminated S_{ij}'s exceeds half the sum of all the weights. This leads to a breakdown point of approximately $1 - 2^{-1/3} = .206$ for large n (Scholz 1978). The estimators of α that depend on $\hat{\beta}_{W1}$, namely $\hat{\alpha}_{1,W1}$ and $\hat{\alpha}_{2,W1}$, share the breakdown point of $\hat{\beta}_{W1}$.

The breakdown point of $\hat{\beta}_{W2}$, the weighted median estimator with weights $x_j - x_i$, depends on the type of corruption. For contamination in the x's as well as the Y's, the breakdown point is 1/n. For contamination in the Y's only, the breakdown point depends on the placement of the x's. The breakdown point can be as low as 1/n, if one bad Y value corresponds to a very influential x value. Specifically, if the sum of weights associated with one observation exceeds half the sum of all the weights, the associated Y value has an arbitrarily large effect on $\hat{\beta}_{W2}$. The estimators of α that depend on $\hat{\beta}_{W2}$, namely $\hat{\alpha}_{1,W2}$ and $\hat{\alpha}_{2,W2}$, share the breakdown point of $\hat{\beta}_{W2}$ in a given situation.

The breakdown point of an estimator $\hat{\alpha} + \hat{\beta}x$ of a mean response is the smaller of the breakdown points of $\hat{\alpha}$ and $\hat{\beta}$.

6. MEAN SQUARED ERROR

In this section, I summarize the main conclusions of a simulation study in which I estimated and compared the mean squared errors (MSE's) of the various slope and intercept estimators (Dietz 1986). In that study, five hundred samples were generated for each combination of n = 20 and 40, three x designs, and nine error distributions. The value of each estimator of α and β defined in Section 2 was computed for each sample. For n = 20, and the eight symmetric error distributions, the value of each estimator of E(Y) was also computed for each of two x values, one involving interpolation, the other extrapolation. (Note that E(Y) is not the natural parameter to estimate for an asymmetric error distribution.) The 500 values of an estimator thus obtained were used to estimate the MSE of that estimator for that n, x design, and error distribution.

The x designs consisted of the expected order statistics from samples of size n from the uniform, normal, and (approximately) the double exponential distributions. Each set of x's was standardized so that $\sum_{i=1}^{n} x_i = 0$ and $\sum_{i=1}^{n} x_i^2 = 1$. The resulting sets of x's vary in the extent to which information is concentrated around x = 0. Nine error distributions were considered -- the standard normal, six contaminated normal distributions, the heavy-tailed t distribution with three degrees of freedom, and the asymmetric lognormal distribution. See Dietz (1986) for details concerning computation and the generation of the random variates.

For estimating slope, the main findings of the simulation study were:

1. The MSE of $\hat{\beta}_A$ is very large, especially for the double exponential x design and heavily contaminated error distributions.

2. The least-squares estimator, $\hat{\beta}_{LS}$, has smaller MSE than any other estimator for normal errors. However, for the other error distributions, the MSE of $\hat{\beta}_{LS}$ is exceeded only by that of $\hat{\beta}_A$.

3. The estimator $\hat{\beta}_{W2}$ performs well until the errors are heavily contaminated at which point $\hat{\beta}_M$ is better. The MSE of $\hat{\beta}_{W1}$ is usually between those of $\hat{\beta}_M$ and $\hat{\beta}_{W2}$.

For estimating mean response (the intercept is a special case), the main results for symmetric error distributions were:

1. The MSE of $\hat{\alpha}_A + \hat{\beta}_A$ x is very large compared to those of the other estimators.

2. The estimator $\hat{\alpha}_M + \hat{\beta}_M$ x is also non-competitive in terms of MSE; its MSE is usually one of the two or three largest.

3. The least-squares estimator, $\hat{\alpha}_{LS} + \hat{\beta}_{LS}$ x, has smaller MSE than any other estimator for normal errors. For other error distributions, its MSE is usually exceeded only by that of $\hat{\alpha}_A + \hat{\beta}_A$ x.

4. For most situations considered, the estimators based on the $\hat{\alpha}_1$'s ($\hat{\alpha}_{1,M}$, $\hat{\alpha}_{1,W1}$, $\hat{\alpha}_{1,W2}$) do not differ among themselves in MSE; the same is true of the estimators based on the $\hat{\alpha}_2$'s ($\hat{\alpha}_{2,M}$, $\hat{\alpha}_{2,W1}$, $\hat{\alpha}_{2,W2}$). The $\hat{\alpha}_2$ estimators have smaller MSE than any other estimators, except for normal errors, where the least-squares estimator is preferable, and the most heavily contaminated normal error distribution, where the $\hat{\alpha}_1$ estimators and $\hat{\alpha}_C + \hat{\beta}_M$ x are preferable.

5. When estimation of the mean response requires extrapolation, the choice of $\hat{\beta}$ estimator can be more important than the choice between the $\hat{\alpha}_1$ and $\hat{\alpha}_2$ groups of estimators. That is, it makes little difference whether the median or median of pairwise averages is used to estimate E(Y); the choice of $\hat{\beta}$ used to form the residuals is more important. The estimators based on $\hat{\beta}_{W2}$ perform well until the errors are heavily contaminated at which point estimators based on $\hat{\beta}_M$ are better.

See Dietz (1986) for simulation results for the asymmetric lognormal error distribution.

7. EXAMPLE

McEntee and Mair (1978) studied the relationship between cerebrospinal fluid concentration of certain brain metabolites and the extent of memory impairment in nine patients suffering from Korsakoff's syndrome. Memory function was measured by IQ–MQ, where IQ is the full-scale intelligence quotient derived from the Wechsler Adult Intelligence Scale and MQ is the memory quotient derived from the Wechsler Memory Scale. The larger IQ–MQ, the more severe the memory impairment. McEntee and Mair computed Pearson correlation coefficients between IQ–MQ and the concentration of each of four brain metabolites. They also fit the least-squares line for IQ–MQ and MHPG, the one metabolite for which a significant correlation with IQ–MQ was found.

Table 3 shows the values of IQ and MQ and the concentration of the brain metabolite homovanillic acid (HVA) for each of the nine patients. Figure 1 shows a scatterplot of HVA versus IQ–MQ. Table 4 shows the ordered values of the sample intercepts A_{ij} and the ordered values of the sample slopes S_{ij} for these two variables. Also shown are the weights j-i and $x_j - x_i$ corresponding to each S_{ij}. The weighted median estimates of β are given by appropriate averages of S_{ij}'s: $\hat{\beta}_{W1} = (-.059 + .067)/2$ and $\hat{\beta}_{W2} = (.154 + .185)/2$. The values of the various estimates of slope and intercept are shown in Table 5.

(Insert Tables 3, 4, 5 and Figure 1 here.)

Certain estimated lines from Section 2.3 are displayed in Figure 1. Because each $\hat{\alpha}_2$ estimate is equal to or nearly equal to the corresponding $\hat{\alpha}_1$ estimate, and the value of $\hat{\alpha}_C$ is very similar to $\hat{\alpha}_{1,M} = \hat{\alpha}_{2,M}$, certain lines are omitted from Figure 1.

The outlying observation for Patient 9 has a large effect on the lines $y = \hat{\alpha}_{LS} + \hat{\beta}_{LS}\,x$ and $y = \hat{\alpha}_{1,W2} + \hat{\beta}_{W2}\,x$. Note that the sum of the weights $x_j - x_i$ associated with Patient 9 is 361, more than half the sum of all the weights (654). Thus the Y value for Patient 9 has an arbitrarily large effect on $\hat{\beta}_{W2}$, and therefore on $\hat{\alpha}_{1,W2}$. (See Section 5 on breakdown.) Note also that the value of $\hat{\alpha}_M$ seems unreasonably large.

Although the lines based on the three $\hat{\alpha}_1$ estimators look fairly different from each other in Figure 1, near the center of the data they yield very similar estimates of E(Y). In fact, at x = 31, the estimates of E(Y) corresponding to $\hat{\alpha}_{1,M}$, $\hat{\alpha}_{1,W1}$, and $\hat{\alpha}_{1,W2}$ are all equal to 27.00.

8. DISCUSSION AND RECOMMENDATIONS

Given all of this information, what robust alternatives to the least-squares estimators should we teach in our introductory nonparametrics courses?

For estimating slope, we can eliminate $\hat{\beta}_A$ from consideration, leaving the weighted median estimators $\hat{\beta}_M$, $\hat{\beta}_{W1}$, and $\hat{\beta}_{W2}$. In that order, the asymptotic relative efficiencies of these estimators with respect to $\hat{\beta}_{LS}$ are $\rho^2 e_W$, $\rho^2 e_W$, and e_W; their breakdown points are, approximately, .293, .206, and as low as 1/n. Thus, $\hat{\beta}_{W2}$ can have higher efficiency, but is less robust, than $\hat{\beta}_M$ and $\hat{\beta}_{W1}$.

Although sometimes less efficient, the Theil estimator is considerably easier to compute than $\hat{\beta}_{W1}$ or $\hat{\beta}_{W2}$. An exact confidence interval related to $\hat{\beta}_M$ can be found easily using tables for Kendall's tau (Hollander and Wolfe 1973). Confidence intervals corresponding to $\hat{\beta}_{W1}$ or $\hat{\beta}_{W2}$ are more difficult to compute and are based on large sample approximations (Sievers 1978). Thus, I recommend $\hat{\beta}_M$ as a robust, easy-to-compute slope estimator whose MSE is often smaller and never much larger than that of competing estimators.

Of the estimators of mean response, we can eliminate those based on $\hat{\alpha}_A$ and $\hat{\alpha}_M$ from consideration, leaving those based on the $\hat{\alpha}_1$'s, $\hat{\alpha}_2$'s, and $\hat{\alpha}_C$.

If the mean response to be estimated is well within the range of the x values, the estimators within the $\hat{\alpha}_1$ group or within the $\hat{\alpha}_2$ group are very similar in MSE. Then there is no motivation for choosing the estimators based on the computationally difficult $\hat{\beta}_{W1}$ or $\hat{\beta}_{W2}$ over those based on $\hat{\beta}_M$. The MSE of $\hat{\alpha}_{2,M} + \hat{\beta}_M x$ is smaller than that of $\hat{\alpha}_{1,M} + \hat{\beta}_M x$ or $\hat{\alpha}_C + \hat{\beta}_M x$ for all symmetric error distributions considered in the simulation study except the most heavily contaminated normal distribution. Thus, for symmetric errors with light or moderate contamination, I recommend the estimator based on $\hat{\alpha}_{2,M}$, the median of pairwise averages of residuals based on $\hat{\beta}_M$. For heavily contaminated errors, estimators based on $\hat{\alpha}_{1,M}$ or $\hat{\alpha}_C$ are preferable.

For asymmetric error distributions, most of the estimators considered here do not estimate the mean response, but rather the median or pseudomedian of the response. The pseudomedian, estimated by $\hat{\alpha}_{2,M}$, is a parameter of dubious interest; thus in this situation, I recommend the estimators based on $\hat{\alpha}_{1,M}$ or $\hat{\alpha}_C$.

The estimator $\hat{\alpha}_C$ is robust and easy to compute; however, it has unknown bias and efficiency properties.

In conclusion, I will suggest the following in my nonparametrics course from now on: To estimate slope in the linear regression problem, use $\hat{\beta}_M$. The choice of an intercept estimator depends on the assumptions you are willing to make about the error distribution. If you are willing to assume symmetry, use $\hat{\alpha}_{2,M}$; if not, or if the errors are very heavily contaminated, $\hat{\alpha}_{1,M}$ is a better choice.

REFERENCES

Bhattacharyya, G.K. (1968), "Robust Estimates of Linear Trend in Multivariate Time Series," *Annals of the Institute of Statistical Mathematics*, 20, 299-310.

Conover, W.J. (1980), *Practical Nonparametric Statistics* (2nd ed.), New York: John Wiley.

Dietz, E.J. (1986), "A Comparison of Robust Estimators in Simple Linear Regression," Institute of Statistics Mimeograph Series No. 1690R, North Carolina State University, Dept. of Statistics.

Donoho, D.L., and Huber, P.J. (1983), "The Notion of Breakdown Point," in *A Festschrift for Erich L. Lehmann*, eds. P.J. Bickel, K.A. Doksum, and J.L. Hodges, Jr., Belmont, Calif.: Wadsworth.

Hettmansperger, T.P. (1984), *Statistical Inference Based on Ranks*, New York: John Wiley.

Hollander, M., and Wolfe, D.A. (1973), *Nonparametric Statistical Methods*, New York: John Wiley.

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

Maritz, J.S. (1979), "On Theil's Method in Distribution-Free Regression," *Australian Journal of Statistics*, 21, 30-35.

McEntee, W.J., and Mair, R.G. (1978), "Memory Impairment in Korsakoff's Psychosis: A Correlation with Brain Noradrenergic Activity," *Science*, 202, 905-907.

Randles, R.H., and Wolfe, D.A. (1979), *Introduction to the Theory of Nonparametric Statistics*, New York: John Wiley.

Scholz, F.W. (1978), "Weighted Median Regression Estimates," *The Annals of Statistics*, 6, 603-609.

Sen, P.K. (1968), "Estimates of the Regression Coefficient Based on Kendall's Tau," *Journal of the American Statistical Association*, 63, 1379-1389.

Sievers, G.L. (1978), "Weighted Rank Statistics for Simple Linear Regression," *Journal of the American Statistical Association*, 73, 628-631.

Simon, S.D. (1982), "A Class of Non-Iterative Slope Estimators in Linear Regression," unpublished Ph.D. dissertation, University of Iowa.

Theil, H. (1950), "A Rank-Invariant Method of Linear and Polynomial Regression Analysis," *Indagationes Mathematicae*, 12, 85-91.

Table 1. Estimators of β and α Based on Sample Slopes and Intercepts

Description	$\hat{\beta}$	$\hat{\alpha}$
Least-squares	$\hat{\beta}_{LS}$	$\hat{\alpha}_{LS}$
Unweighted average	$\hat{\beta}_{A}$	$\hat{\alpha}_{A}$
Median	$\hat{\beta}_{M}$	$\hat{\alpha}_{M}$
Weighted median, weights j-i	$\hat{\beta}_{W1}$	
Weighted median, weights x_j-x_i	$\hat{\beta}_{W2}$	

Table 2. Estimators of α Based on $Y_i-\hat{\beta}x_i$

$\hat{\beta}$ Used	Median	Median of Pairwise Averages
$\hat{\beta}_M$	$\hat{\alpha}_{1,M}$	$\hat{\alpha}_{2,M}$
$\hat{\beta}_{W1}$	$\hat{\alpha}_{1,W1}$	$\hat{\alpha}_{2,W1}$
$\hat{\beta}_{W2}$	$\hat{\alpha}_{1,W2}$	$\hat{\alpha}_{2,W2}$

Table 3. HVA Concentrations (Nanograms per Milliliter), Full-Scale Intelligence Quotients (IQ), and Memory Quotients (MQ) for Nine Patients with Korsakoff's Syndrome

Patient*	HVA	IQ	MQ
1	21	89	60
2	23	90	59
3	25	122	102
4	25	87	64
5	26	89	61
6	31	106	79
7	40	104	80
8	48	106	80
9	75	127	88

*Patients have been renumbered according to increasing value of HVA.

Source: McEntee and Mair (1978).

Table 4. Ordered Sample Intercepts A_{ij} and Ordered Sample Slopes S_{ij} with Corresponding Weights for HVA and IQ-MQ

	Ordered A_{ij}	Ordered S_{ij}	j-i	x_j-x_i
	-180.000	-5.500	1	2
	-102.000	-4.000	2	2
	-9.167	-2.250	2	4
	2.889*	-1.500	3	4
	6.333	-1.000	3	3
	6.857*	-.500	4	8
	8.000	-.412	5	17
	10.500*	-.333	1	9
	13.333	-.286	2	14
	13.478	-.263	6	19
	14.000	-.200	5	10
	15.000*	-.200	4	5
	18.545*	-.200	6	25
	19.739	-.200	1	5
	21.333	-.111	7	27
	22.163*	-.091	3	22
	25.111*	-.059	2	17
	27.462*	.067	3	15
	28.824	.130	4	23
	30.364	.154*	7	52
	31.333	.185*	8	54
	33.200	.224*	4	49
	33.200	.250	1	8
	33.200	.261	5	23
	34.526	.267	4	15
	35.429	.273*	3	44
	35.600	.320*	5	50
	37.333	.380*	6	50
	40.471	.429*	2	35
	42.500	.481*	1	27
	54.000	.667	2	6
	60.500	1.000	1	2
	76.250	1.167	3	6
	123.000	5.000	1	1
	157.500	8.000	2	1
Sum			119	654
Mean	23.452	.061		

*A_{ij}'s and S_{ij}'s associated with Patient 9.

Table 5. Estimates of Slope and Intercept for HVA and IQ-MQ

$\hat{\beta}_{LS} = .1964$	$\hat{\alpha}_{LS} = 20.592$	$\hat{\alpha}_{1,M} = 24.933$
$\hat{\beta}_{A} = .0614$	$\hat{\alpha}_{A} = 23.452$	$\hat{\alpha}_{1,W1} = 26.878$
$\hat{\beta}_{M} = .0667$	$\hat{\alpha}_{M} = 27.462$	$\hat{\alpha}_{1,W2} = 21.745$
$\hat{\beta}_{W1} = .0039$	$\hat{\alpha}_{C} = 25.267$	$\hat{\alpha}_{2,M} = 24.933$
$\hat{\beta}_{W2} = .1695$		$\hat{\alpha}_{2,W1} = 26.878$
		$\hat{\alpha}_{2,W2} = 21.652$

Figure title

Figure 1. Scatterplot of HVA versus IQ-MQ with Fitted Regression Lines. Notice the outlying point for Patient 9.

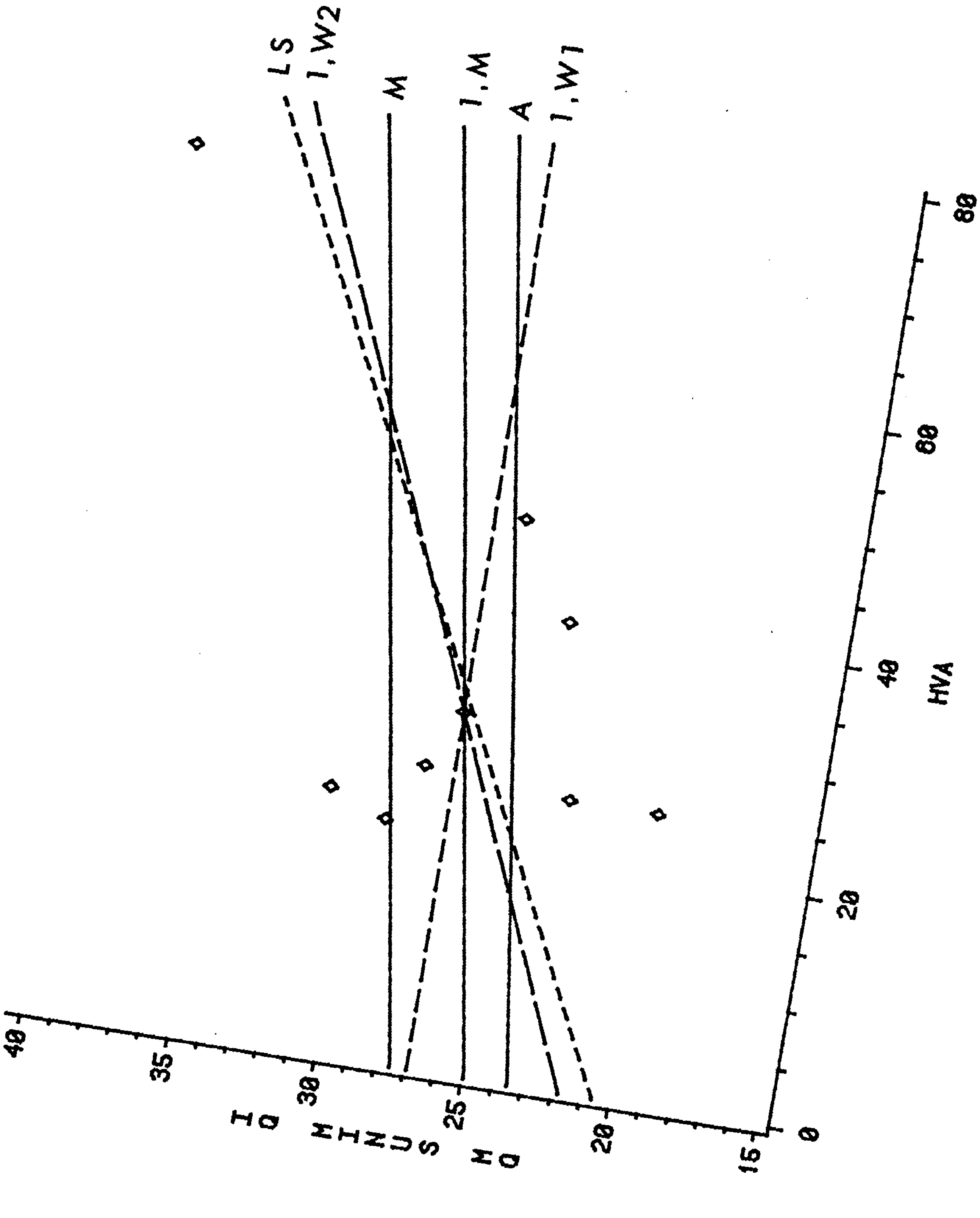

Figure 1