Comment

T. P. Hettmansperger and James C. Aubuchon

We would like to thank David Draper for an illuminating discussion of some contemporary alternatives to traditional data analysis. Least squares provides one of the great unifying principles of statistical inference. Methodology derived from this principle includes estimation, testing, confidence intervals and multiple comparisons in experimental contexts that range from simple univariate location models through multivariate linear models. This unification is enhanced by the elegant geometry that flows from the L_2 norm underlying least squares and by Gaussian likelihoods that relate this geometry to statistical inference. A benefit (or curse, depending on your point of view) of having a single underlying principle is widespread and extensive implementation of the least squares methods in statistical packages. When combined with modern diagnostic methods, least squares can be the basis of thoughtful and revealing data analysis; see Daniel and Wood (1980) for impressive examples.

Tukey (1960) broke the spell of least squares when he pointed out that statistical methods, optimal for a specific underlying model, may deteriorate rapidly in a neighborhood of the assumed model. This paved the way for the development and acceptance of alternative, robust methods.

By replacing the L_2 norm by alternative norms that generate more robust statistical methods, we retain the unity and elegance provided by an underlying geometry and, at the same time, achieve a more stable and perhaps more powerful set of statistical methods.

The L_1 norm generates robust and resistant methods—the sign test and the sample median in the simple location model. However, the efficiency of the L_1 with respect to the L_2 methods at the Gaussian model is only 64%. If we write $e_i(\boldsymbol{\beta}) = y_i - \mathbf{x}_i'\boldsymbol{\beta}$, $i = 1, \ldots, n$, for the errors in a linear model then inference about $\boldsymbol{\beta}$ can be based on a weighted L_1 norm,

$$D(\boldsymbol{\beta}) = \sum R(|e_i(\boldsymbol{\beta})|)\,|e_i(\boldsymbol{\beta})|, \tag{1}$$

T. P. Hettmansperger is Professor of Statistics, Department of Statistics, Pond Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802. James C. Aubuchon is Senior Statistician, Minitab, Inc., 3081 Enterprise Drive, State College, Pennsylvania 16801.

where $R(|e_i(\boldsymbol{\beta})|)$ is the rank of $|e_i(\boldsymbol{\beta})|$ among $|e_1(\boldsymbol{\beta})|, \ldots, |e_n(\boldsymbol{\beta})|$. A bit of algebra shows that

$$D(\boldsymbol{\beta}) = \sum\sum_{i \le j} \left| \frac{e_i(\boldsymbol{\beta}) + e_j(\boldsymbol{\beta})}{2} \right| + \sum\sum_{i<j} \left| \frac{e_i(\boldsymbol{\beta}) - e_j(\boldsymbol{\beta})}{2} \right|, \tag{2}$$

and further that

$$\sum\sum_{i<j} \left| \frac{e_i(\boldsymbol{\beta}) - e_j(\boldsymbol{\beta})}{2} \right| = \sum \left[R(e_i(\boldsymbol{\beta})) - \frac{n+1}{2} \right] e_i(\boldsymbol{\beta}). \tag{3}$$

Inference can be based on (1) directly. However, for reasons discussed below, we suggest the following strategy. Use the second term in (2) to develop inference methods for the regression parameters in a linear model and then use the resulting residuals and the first term in (2) to develop inference methods for the intercept parameter. As (3) shows, the second term of (2) is simply a linear combination of errors with weights equal to their centered ranks. Hence, rank-based inference for both the vector of regression parameters and the scalar intercept parameter is directly related to the L_1 norm. The resulting estimates and tests enjoy robustness induced by the L_1 norm and, at the same time, achieve a surprising efficiency with respect to L_2 methods of 95.5% for an underlying Gaussian model.

The reason for recommending the two-part strategy over direct application of (1) is that (3) does not depend on the intercept and does not require symmetry of the error distribution to develop the asymptotic theory for tests and estimates. For a given estimate of the vector of regression parameters, minimizing the first term in (2) is the same as applying the Wilcoxon signed-rank procedure to the uncentered residuals, which requires symmetry. When symmetry is in doubt we recommend using (3) along with an L_1 estimate of the intercept.

Draper has presented an interesting exposition of two approaches that generate robust and efficient inferences in the linear model by intervening in the L_2 geometry in different ways. If we apply (3) to the two-sample location model, the resulting estimate of shift is the Hodges-Lehmann estimate. By casting the multiway layouts with several observations per cell as a composite of two-sample shift models, Lehmann developed new estimates of contrasts and thus new tests. Interpretation of tests remains the same as in least squares; only the estimates have been changed to protect the experiments.

The second approach attacks the issue more directly. Namely, replace the L_2 norm by the weighted L_1 norm, and proceed immediately to estimation by minimizing the new distance measure to the model subspaces and to testing by comparing these new distances. The inferential strategy remains the same but the norm (and hence metric) have been changed. The value of the second approach lies in the breadth of application. Models ranging from simple one-sample location through the linear model with AOV designs, regression designs and analysis of covariance designs are handled in a unified way.

The implementation of this second approach requires a fully developed asymptotic distribution theory for the estimates and tests, estimation methods for a ubiquitous scaling parameter ($\theta = \int f^2(x)\,dx$, where $f(x)$ is the density of the error distribution) and the development of algorithms to carry out the required computations.

Most of what is known about the estimation of θ has been mentioned by Draper. We would like to close this discussion with some additional comments and references on the asymptotics and on computation.

In her seminal paper, Jurečková (1971) made rather complicated assumptions about the design matrix in order to develop the asymptotic theory for her regression R-estimates. Unfortunately, in practical problems, there is no way to check whether these assumptions are reasonable. Subsequent authors who built on this work adopted the same assumptions. However, as Heiler and Willers (1979) show, the only necessary assumption on the design matrix is the same as for the asymptotic theory of least squares procedures: Huber's assumption that the diagonal elements of the least squares projection matrix (the leverages) tend to zero as n tends to infinity.

Published work on rank-based methods for linear models typically suggests doing the computations via Newton's method (using the Hessian of the quadratic approximation developed in the asymptotic theory). Osborne (1985) has derived a rather different approach which takes full advantage of the fact that the dispersion is a convex polyhedral function. This approach should be seriously considered by anyone implementing these methods on the computer.

Derivation of confidence and multiple comparison procedures through replacement of the normal theory parameter estimates and estimated error variance by their robust analogues is connected with the Wald test statistic: a quadratic form in the full model estimate of the parameter vector. To develop confidence procedures which would be tied to the drop-in-dispersion test statistic, one would have to find, for example, all values of the parameter vector which could not be rejected by the test. This presents a rather difficult computational problem which, we believe, has not been attacked as yet.

In closing, we would like to reiterate the fact that both approaches described by Draper have been implemented. Fortran routines are available from Draper for the L_1ehmann methods and from J. W. McKean at Western Michigan for the Jaeckel and Hettmansperger-McKean methods, while an experimental rank regression command will be available in Release 6 of Minitab for many computer systems. It is hoped that people will subject these methods to the ultimate test: real data.

ADDITIONAL REFERENCES

HEILER, S. and WILLERS, R. (1979). Asymptotic normality of R-estimates in the linear model. Forschungsbericht N.R. 79/6, Univ. Dortmund.

OSBORNE, M. R. (1985). *Finite Algorithms in Optimization and Data Analysis.* Wiley, New York.

Comment

Peter J. Bickel

I want to use the opportunity of discussing this excellent exposition of rank-based methods in the linear model in part to revive a suggestion I made in a review paper on robust methods generally (Bickel, 1976). This approach may have some computational advantages over the Jaeckel-McKean-Hettmansperger (I would add Jurečková-Koul to the list) approach and relates the methods more closely to classical analysis of variance. The idea is to first fit the full

Peter J. Bickel is Professor of Statistics, University of California, Berkeley, California 94720.