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.

(largest contemplated) linear model Ω by Jaeckel's method, say using (3.15). Having obtained the resulting $\mathbf{e}(\hat{\beta}_{JW})$ residuals $\mathbf{e}(\hat{\beta}_{JW})$ and an estimate of $\hat{\sigma}_R$, form pseudo values for the data as follows,

$$Y_i^* = [X\hat{\beta}_{JW}]_i + \hat{\sigma}_R\left[R_i(\hat{\beta}_{JW}) - \frac{(N+1)}{2}\right], \quad i = 1, \cdots, n.$$

The heuristics of Bickel (1976) can I believe be rigorized in this case also to conclude that if we now act as if the Y_i^* were the data, apply ordinary least squares methods in fitting subhypotheses and then calculate the usual F statistics we are asymptotically right in the sense that the asymptotic null distribution and power functions of these statistics agree with the χ^2 approximations to the corresponding Hettmansperger-McKean statistics. We expect more. For instance, application of Tukey's method of multiple comparisons to the pseudo observations should have the same efficiency (say in terms of length of the intervals) with respect to the method applied to the original observations as the Wilcoxon test has to the t test.

Of course the asymptotic χ^2 approximations here too will be inadequate as Draper points out. However, one might hope that the same empirical observations made by Draper continue to hold, viz., using the classical degrees of freedom for F works adequately.

Let me add a caution. As Draper points out what is done here guarantees robustness only against heavy tails. In particular, sensitivity to high leverage points among the $[X\beta]_i$ is not affected. Nor is sensitivity to heteroscedasticity, dependence, transformation of the Y scale, etc. Perhaps the pseudo values could be used as a first step in procedures where the second step fitting method would address these departures and of course, one would then iterate.

It is worth noting that the scope of the methods discussed by Draper has recently been enlarged by Tsiatis (1986) to handle the case of right censoring of the Y_i. It's not clear what happens to the pseudo value-based procedures in this context.

Finally, it is worth remembering that the scope of purely rank-based procedures is much greater than what is suggested by the Kruskal-Wallis, Friedman-Tukey tests. In particular, ranks not rank of residual procedures are appropriate when one considers transformation models of the form

$$h(Y_i) = [X\beta]_i + e_i \quad i = 1, \cdots, n$$

where the e_i are assumed to come from some parametric family but h is an unknown monotone transformation. See Doksum (1987) and Bickel (1986) for example.

I congratulate David Draper on this clear insightful presentation.

ADDITIONAL REFERENCES

BICKEL, P. J. (1976). Another look at robustness: A review of reviews and some new developments (with discussion). *Scand. J. Statist.* **3** 145–168.

BICKEL, P. J. (1986). Efficient testing in a class of transformation model. Papers on semiparametric models at the ISI Centenary Session, C.W.I., Amsterdam.

DOKSUM, K. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15** 325–345.

TSIATIS, A. (1986). Estimating regression parameters using linear rank statistics for censored data. Technical Report, School of Public Health, Harvard Univ.

Comment

R. Douglas Martin

Dr. Draper has provided a very nice exposition and review of two rank-based robust methods for fixed effects ANOVA problems. In so doing, he concentrates on (i) the formal structure of the methods and (ii) robust inference based on rank-based analogues of the classical test statististics, where robust inference is taken to mean robustness of validity and efficiency. Given the author's commitment to focus on the R-estimate approach, I would only wish that he had given some emphasis to examples, and in so doing revealed the exploratory data-analytic use of the methods. As far as the focus on rank-based methods goes, I have a pragmatically motivated reservation based on a concern I share with Draper, namely, robust methods are not widely available in the major statistical packages.

As Draper points out, R-estimates comprise just one of three major classes of robust estimates, with L-estimates and M-estimates being the other two, and

R. Douglas Martin is Professor of Statistics, Department of Statistics, University of Washington, Seattle, Washington 98195.