

Comment

A. H. Welsh

This is a stimulating and timely paper written from a practical data-analytic viewpoint. The introduction contains a careful discussion of strategies for data analysis which should be read critically by all statisticians. The difference in perspective makes this discussion a most useful supplement to the related discussions in Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986). The remainder of the paper provides an easily accessible exposition of two rank-based approaches to analyzing linear models. I would like to discuss several points which arise from the paper.

1. INFERENCE STRATEGIES

Draper's discussion of the widely used but naive "do-nothing" approach to dealing with violations of the classical assumptions highlights the failings of the approach. Although the more sophisticated data-analytic approach offers a potential improvement over the do-nothing approach, it is not as widely appreciated as it should be that the data-analytic approach has serious pitfalls. Essentially, problems arise from the fact that in linear model (and more complicated model) problems, much of the data analysis is based on the residuals from some preliminary parameter estimate (which is usually the least squares estimate) rather than on the observations themselves and any subsequent analysis should take this fact into account. This is particularly true of methods for outlier detection. It is a popular misconception, for example, that we can proceed to delete outliers from the sample and then apply classical least squares techniques to the reduced sample. Ruppert and Carroll (1980) showed that if we delete observations with extreme residuals and apply least squares to the reduced data set, the initial estimator (from which the residuals were calculated) has a persistent effect which does not vanish asymptotically. Consequently, the second-stage estimator can be no more efficient or robust than the initial estimator and the standard errors obtained from standard least squares formulae will be too small so that confidence intervals and tests will be misleading. The problem is that the second stage of the analysis (which is a naive least squares analysis) incorrectly ignores the effects of the preliminary

analysis. The problem is not solved by basing the preliminary analysis on an initial estimator other than least squares. However, the problem can be overcome by constructing a trimmed estimator which takes the preliminary analysis into account (see Welsh, 1987a).

Incidentally, I cannot agree with Draper's view that recent developments in nonparametric regression are a part of robustness work. Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986) have argued convincingly that robustness is concerned with underlying parametric models. Although some extension of the theory away from this strict viewpoint may be possible, without the specification of a precise underlying model, the whole concept of a deviation from such a model loses its foundation. Indeed, in nonparametric regression all data points are treated as equally good and consequently receive equal weight in the analysis.

2. ASYMMETRY

The role of symmetry in the robust analysis of linear models has caused persistent confusion. Draper makes some comments on the role of symmetry which deserve clarification. In most applications of linear models, the important inferential questions involve the slope parameters. The slope parameters are identifiable (and hence can be estimated) when the errors have an asymmetric distribution. *L*- and *M*-estimators, for example, estimate the slopes when the errors have an asymmetric distribution (see Carroll and Welsh (1987) for references) and the same ought to be true of *R*-estimators. If symmetry has any role to play, it is in the estimation of the intercept. However, as Draper notes, in practice we can either use the ordinary median of the residuals (as advocated by Aubuchon and Hettmansperger, 1984b) as the intercept estimator or we can try to determine a more appropriate estimator based on the nature of the problem at hand. Hence, the presence of asymmetry in the errors means that we should think carefully about the choice of intercept estimator; it certainly does not mean that a robust analysis is inappropriate.

3. VARIANCE ESTIMATION

The problem of estimating the variance of *R*-estimators is interesting and important but still seems to have been relatively little studied. A number of questions immediately arise.

A. H. Welsh is Lecturer, Department of Statistics, The Faculties at The Australian National University, G. P. O. Box 4, Canberra A.C.T. 2601, Australia.