substantial and has greatly increased my awareness of the structure of regression problems, particularly with regard to the role of individual and groups of observations. However, for progress beyond linear models and a more complete understanding of past results, ad hoc reasoning no longer seems sufficient. Competing goals must be carefully weighed and influence measures must be formulated with a broader base. Likelihood is the foundation for many analyses and in the long term we should strive for methods that directly reflect the difference between the full sample likelihood and the likelihood obtained after deletion. From a Bayesian perspective, the pioneering work of Johnson and Geisser (1982, 1983, 1985) is relevant.

Broadening the concept of influence to include more than the deletion of observations is a second direction that may prove fruitful. Deletion can be viewed as just one of many ways of perturbing a problem formulation to assess influence. Minor modifications of the values of a selected explanatory variable in linear or nonlinear regression, for example, can uncover relevant structure in the data that would not normally be detected by deletion, and lead to fresh interpretations of certain patterns in added variable plots. These and related issues are addressed in Cook (1986).

ADDITIONAL REFERENCES

COOK, R. D. (1975). Detection of influential observations in linear regression. Technical Report 256, School of Statistics, Univ. Minnesota.

COOK, R. D. (1977b). Letter to the Editor. *Technometrics* **19** 348.

COOK, R. D. (1979). Influential observations in linear regression. *J. Amer. Statist. Assoc.* **74** 169–174.

COOK, R. D. (1982). Discussion of Dr. Atkinson's paper. *J. Roy. Statist. Soc. Ser. B* **44** 28.

COOK, R. D. (1986). Assessment of local influence (with discussion). To appear in *J. Roy. Statist. Soc. Ser. B.*

COOK, R. D., PENA, D. and WEISBERG, S. (1984). The likelihood displacement: a unifying principle for influence measures. MRC Technical Summary Report 2751, Univ. Wisconsin, Madison.

COOK, R. D. and WANG, P. C. (1983). Transformations and influential cases in regression. *Technometrics* **25** 337–343.

FREEDMAN, D. A. and NAVIDI, W. C. (1986). Regression models for adjusting the 1980 census (with discussion). *Statist. Sci.* **1** 3–39.

JOHNSON, W. and GEISSER, S. (1982). Assessing the predictive influence of observations. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds.) 343–358. North-Holland, Amsterdam.

JOHNSON, W. and GEISSER, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *J. Amer. Statist. Assoc.* **78** 137–144.

JOHNSON, W. and GEISSER, S. (1985). Estimative influence measures for the multivariate general linear model. *J. Statist. Plann. Inference* **11** 33–56.

Comment: Aspects of Diagnostic Regression Analysis

A. C. Atkinson

1. INTRODUCTION

The rapidity of acceptance of the group of techniques known as regression diagnostics is remarkable. The methods are already included in many regression packages and there are at least three books devoted to the subject. The emphasis of each book is distinct. Belsley, Kuh, and Welsch (1980) are primarily concerned with applications in economics; Cook and Weisberg (1982) are the most mathematical of the three; Atkinson (1985) includes much material on transformations. In addition, an introduction is given by Weisberg (1985, Chapters 5 and 6). Now we have the present review article by Chatterjee and Hadi. In my comments I shall go beyond the area defined by their title, to describe several recent developments which reflect important aspects of diagnostic regression analysis. An example of the use of these methods is given in Section 5.

A. C. Atkinson is Professor of Statistics, Department of Mathematics, Imperial College, Queen's Gate, London SW7 2BZ, United Kingdom.

Diagnostic procedures are essentially concerned with the detection of disagreements between the model and the data to which it is fitted. As Chatterjee and Hadi suggest, the variety of such procedures can be bewildering. There are, however, some underlying ideas which provide a framework for comparisons. A succinct summary of principles is given by Weisberg (1983). Among other aspects he stresses: 1) the relationship with score tests for parameterized departures from assumptions, 2) the importance of graphical methods, and 3) influence analysis, that is calculation of the effect of individual observations on inferences drawn from the data.

2. GENERALIZATIONS

Chatterjee and Hadi's discussion is almost entirely concerned with the normal theory linear model. Pregibon (1981) gives the extension of diagnostic methods to generalized linear models, although his detailed discussion and examples concentrate on the analysis of binary data. Chapter 12 of McCullagh and Nelder (1983), Model Checking, also describes the extension

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.* www.jstor.org

of these ideas to generalized linear models, including explicit discussion of score tests, and gives an example with gamma errors. The example in Cook and Weisberg (1982, Section 5) is again of logistic regression.

A difficulty in the diagnostic approach to the analysis of binary data is that the residuals can take only one of two values, depending on whether a "success" or a "failure" is observed. Landwehr, Pregibon, and Shoemaker (1984) discuss graphical and diagnostic methods for the analysis of binary data, some of which use smoothing techniques. Wang (1985) provides added variable plots for generalized linear models. Jørgensen (1983) indicates how the results of Pregibon (1981) could be used in an extended class of generalized linear and nonlinear models.

3. TRANSFORMATIONS

Data are often better analyzed after a transformation of the response. Let this parametric transformation be $y(\lambda)$. In the family analyzed by Box and Cox (1964)

$$y(\lambda) = (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) \tag{1}$$

where $\dot{y}$ is the geometric mean of the observations. In many examples, use of such a transformation leads to a simple additive model, approximate normality of errors, and a reconciliation of apparent outliers with the body of the data. Several examples are given by Atkinson (1985, Chapter 6).

Conclusions about a transformation can, however, be strongly influenced by one or a few observations. Some form of influence analysis is therefore required. One development leads to procedures analogous to the added variable plot for the addition of an extra carrier in multiple regression. The hope of the analysis is that, for some suitable λ, the observations will follow the normal theory model

$$z(\lambda) = X\beta + \varepsilon. \tag{2}$$

Expansion of the transformed response (2) in a Taylor series about the null value λ_0 yields the approximation

$$z(\lambda) \cong z(\lambda_0) + (\lambda - \lambda_0)w(\lambda_0)$$

with

$$w(\lambda_0) = \partial z(\lambda)/\partial\lambda \,|_{\lambda=\lambda_0}. \tag{3}$$

Box (1980) called $w(\lambda_0)$ a constructed variable. The approximate linear model is found by substitution in (3) to be

$$z(\lambda_0) = X\beta - (\lambda - \lambda_0)w(\lambda_0) + \varepsilon. \tag{4}$$

Results such as equation (49) of Chatterjee and Hadi can be used to provide an approximate score test for the transformation, which is the t test for regression on the constructed variable in (4) (Atkinson, 1973).

The added variable plot of residual $z(\lambda_0)$ against residual $w(\lambda_0)$ provides an indication of the influence of individual observations on the evidence for a transformation. However, the plot can fail if points of high leverage are present. Because leverage points give rise to relatively small residuals, such points may make a small contribution to the plot, even if they are important for the transformation. Cook and Wang (1983) give an example and describe alternative diagnostic plots. Atkinson (1986a) provides a generalization of their method based on writing the approximate score statistic for the transformation as a function of sums of squares and products of residual $z(\lambda_0)$ and $w(\lambda_0)$. Standard formulae for the effect of deletion on the residual sum of squares (Chatterjee and Hadi, equation 14) can therefore be adapted to estimate the effect on the score statistic of the deletion of individual observations.

The idea of a constructed variable is not confined to transformations. Pregibon (1980) develops the method for a parameterized link function in a generalized linear model. Atkinson (1985, page 240) gives an added variable plot for such a constructed variable. An alternative graphical procedure would be again to calculate the effect of deletion on the score test.

The fit of a model to data, as measured by the residual sum of squares, can often be improved by transformation of the response, deletion of an observation, or addition of an extra carrier. Deletion of the ith observation is equivalent to regression on a variable in which all values except the ith are zero. Expression of transformation of the response as regression on a constructed variable provides a formal link between the three possibilities. However, the physical implications are quite distinct.

4. MASKING AND ROBUST REGRESSION

The methods described so far are, like those given by Chatterjee and Hadi, concerned with the deletion of a single observation. The extension to deletion of several observations at once is algebraically straightforward, but a computational nightmare, as there is a combinatorial explosion of possibilities to be explored. Unfortunately, sequential deletion of observation cannot be relied on to reveal multiple outliers. Some examples of this phenomenon, known as masking, are described by Chatterjee and Hadi in their Section 8.

One method of revealing masking relies on a combination of robust regression and diagnostic methods. The robust method is least median of squares regression described by Rousseeuw (1984). For the parameter value b let the residual $r_i = y_i - x_i^T b$. Then two criteria for the choice of b are:

Least sum of squares regression: $\min_b \sum r_i^2$

Least median of squares regression: $\min_b \operatorname{median}_i r_i^2$.

The purpose of least median of squares regression in the presence of outliers is to fit a line to the "good" observations while revealing the "bad" observations as outliers. The proportion of outlying observations can be large, but must be less than the asymptotic limit of one-half (Hampel et al., 1986, page 330).

For a regression model with p carriers, the least median of squares estimates of the parameters are calculated by repeated sampling of elemental sets of p observations until a stable pattern of residuals emerges. The results of Rousseeuw (1984) and Atkinson (1986b) indicate that the method works excellently as an exploratory tool. However, the estimates of the parameters of the linear model have poor properties and a second, confirmatory, stage is required.

Rousseeuw (1984) uses the least median of squares estimate as a starting point for robust regression with M estimators. Atkinson (1986b) uses extensions of some of the regression diagnostics described by Chatterjee and Hadi.

Inspection of an index plot of the least median of squares residuals often suggests the exploratory deletion of several observations. Let there be m of these. The deletion residuals r_i^* for the remaining $n - m$ observations are given by

$$r_i^* = r_i/\{s_{(i)}\sqrt{(1 - h_i)}\}, \tag{5}$$

where r_i is the least squares residual. The deletion residual is Chatterjee and Hadi's t_i^* in a different notation. In particular, we use the standard notation h_i for the diagonal elements of the $(n - m) \times (n - m)$ hat matrix $H = X(X^TX)^{-1}X^T$.

The variance of prediction at a point x_i for a case not included in the fitted model is

$$\text{var}(\hat{y}_i) = \sigma^2\{1 + x_i^T(X^TX)^{-1}x_i\} = \sigma^2(1 + d_i).$$

Agreement between $\hat{y}_i$ and the observed y_i is tested by the prediction residual

$$r_i^o = r_i/\{s\sqrt{(1 + d_i)}\}. \tag{6}$$

This t test is the analogue of the deletion residual (5), but with a change of sign in the denominator.

The modified Cook statistic (Chatterjee and Hadi, equation 38) measures the effect of deletion of the ith observation on the parameter estimates. The effect of addition of an observation on the parameter estimates is found from the analogous quantity

$$C_i^o = \left\{\frac{n - p - m}{p}\,\frac{d_i}{1 + d_i}\right\}^{1/2} |r_i^o|. \tag{7}$$

Atkinson (1986b) gives examples of the use of plots of these quantities to check the deletion results of the exploratory robust analysis.

5. AN EXAMPLE

Table 1 gives the record time for 35 hill races, together with the distance in miles and the climb in feet. The data, taken from the 1984 fixture list of the Scottish Hill Runners Association, have been simplified by omission of a third explanatory variable, the time of year. Analysis of the reduced data provides a good illustration of many of the points discussed in the two previous sections.

One way to start the analysis is to fit a first order model in distance and climb with record time as response. Half normal plots of the deletion residuals r_i^* and of the modified Cook statistics C_i provide a check for outliers and influential observations. Interpretation of these plots is aided by the use of simulation envelopes (Atkinson, 1981; Dempster et al., 1984). The plot of r_i^*, not shown here, indicates that observations 7 and 18 are outliers. If these two observations are deleted, the half normal plots of both r_i^* and of C_i show observation 33 also lying outside the simulation envelope. When these three observations are deleted, the half normal plot of r_i^*, Figure 1, clearly reveals three appreciable outliers. One conclusion is that when all observations are included in the analysis, the outlying nature of observation 33 is masked by observations 7 and 18.

A safer approach to the identification of outliers in the presence of masking is to start with least median of squares regression. An arbitrary number of 1000 elemental sets was sampled. Figure 2 is an index plot of the residuals from the best fit, standardized by a robust estimate of scale. The observations giving the five largest absolute residuals are, in decreasing size of residual, 7, 18, 11, 33, and 35. The best fit was obtained from the 845th elemental set, for which the estimate of σ^2 was 6.21. The next smallest estimate was 6.42, obtained after 79 sets, for which a similar pattern of residuals was obtained.

Figure 2 is typical of the pattern which occurs in the presence of leverage points. If the leverage points are not included in the optimum elemental set, small fluctuations in the fitted line can cause large residuals at remote points in the space of the carriers. However, the confirmatory application of diagnostic methods leads to discrimination between outliers and well behaved leverage points. In this case half normal plots in which all five observations are deleted show that observations 11 and 35 agree with the bulk of the data. The conclusion is again reached that observations 7, 18, and 33 are, in some way, different from the rest of the data.

Although Figure 1 shows three clear outliers, there are also some smaller residuals which are slightly too large. This may be an indication that a transformation of the response is needed. We now briefly look at the

TABLE 1
Record times for Scottish Hill races

Observation number	Name of race	x_1: distance	x_2: climb	y: record time
		miles	*ft*	*hr.min.sec*
1	Greenmantle New Year Dash	2.5	650	16.05
2	Carnethy "5" Hill Race	6	2500	48.21
3	Craig Dunain Hill Race	6	900	33.39
4	Ben Rha Hill Race	7.5	800	45.36
5	Ben Lomond Hill Race	8	3070	62.16
6	Goatfell Hill Race	8	2866	73.13
7	Bens of Jura Fell Race	16	7500	3.24.37
8	Cairnpapple Hill Race (Veterans only)	6	800	36.22
9	Scolty Hill Race	5	800	29.45
10	Traprain Law Race	6	650	39.45
11	Lairig Ghru Fun Run	28	2100	3.12.40
12	Dollar Hill Race	5	2000	43.03
13	Lomonds of Fife Hill Race	9.5	2200	65.00
14	Cairn Table Hill Race	6	500	44.08
15	Eildon Two Hills Race	4.5	1500	26.56
16	Cairngorm Hill Race	10	3000	1.12.15
17	Seven Hills of Edinburgh Race	14	2200	1.38.25
18	Knock Hill Race	3	350	1.18.39
19	Black Hill Race	4.5	1000	17.25
20	Creag Beag Hill Race	5.5	600	32.34
21	Kildoon Hill Race	3	300	15.57
22	Meall Ant-Suidhe Hill Race	3.5	1500	27.54
23	Half Ben Nevis	6	2200	47.39
24	Cow Hill Race	2	900	17.56
25	North Berwick Law Race	3	600	18.41
26	Creag Dubh Hill Race	4	2000	26.13
27	Burnswark Hill Race	6	800	34.26
28	Largo Law Race	5	950	28.34
29	Criffel Hill Race	6.5	1750	50.30
30	Achmony Hill Race	5	500	20.57
31	Ben Nevis Race	10	4400	1.25.35
32	Knockfarrel Hill Race	6	600	32.23
33	Two Breweries Fell Race	18	5200	2.50.15
34	Cockleroi Hill Race	4.5	850	28.06
35	Moffat Chase	20	5000	2.39.50

effect of outliers on the evidence for the power transformation (1). For the null hypothesis of no transformation, that is $\lambda_0 = 1$, the constructed variable is

$$w(1) = y\{\log(y/\dot{y}) - 1\},$$

provided the model includes a constant.

The asymptotically standard normal approximate score statistic for the power transformation for all 35 observations has the value -4.11, the negative sign indicating a value of $\hat{\lambda}$ less than 1. In fact the maximum likelihood estimator is 0.55, so a square root transformation of y would seem to be indicated. However, the index plot of the value of the statistic as each observation in turn is deleted (Figure 3) shows that observations 7 and 18 have appreciable influence. If these two observations, which we already believe to be outlying, are deleted, the absolute value of the statistic decreases to -2.91. The index plot analogous to Figure 3 now shows observation 33 to be highly influential. If this observation is also deleted, the test statistic for the remaining 32 observations equals -1.91, with the deletion estimates oscillating around this nonsignificant value. The analysis of transformations thus reveals the importance of the same three outlying observations as did the analysis of the untransformed data. If these observations had not been detected, there would have appeared to be appreciable evidence for the square root transformation.

The purpose of this summary analysis is to illustrate the use of some diagnostic techniques. Further understanding of the data can be obtained from scatter plots of y against the explanatory variables. The stronger univariate relationship is with x_1, distance, and the plot reveals observations 7 and 18 as outliers. Observation 33 is not clearly outlying, whereas observation 11 appears strongly outlying on the plot against climb. This is because it comes from a race with low climb for its distance. As the preceding analysis has shown,

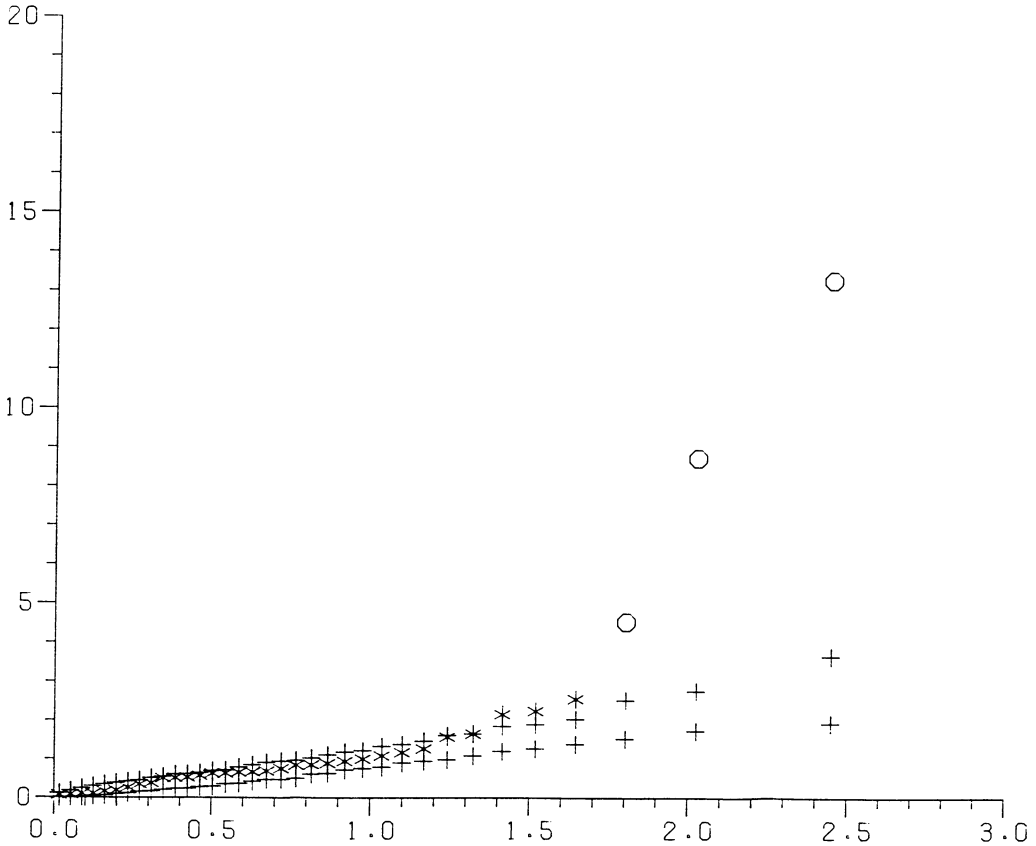

FIG. 1. *Hill race data: half normal plot of deletion residuals* r_i^*: O deleted observations 7, 18, and 33.

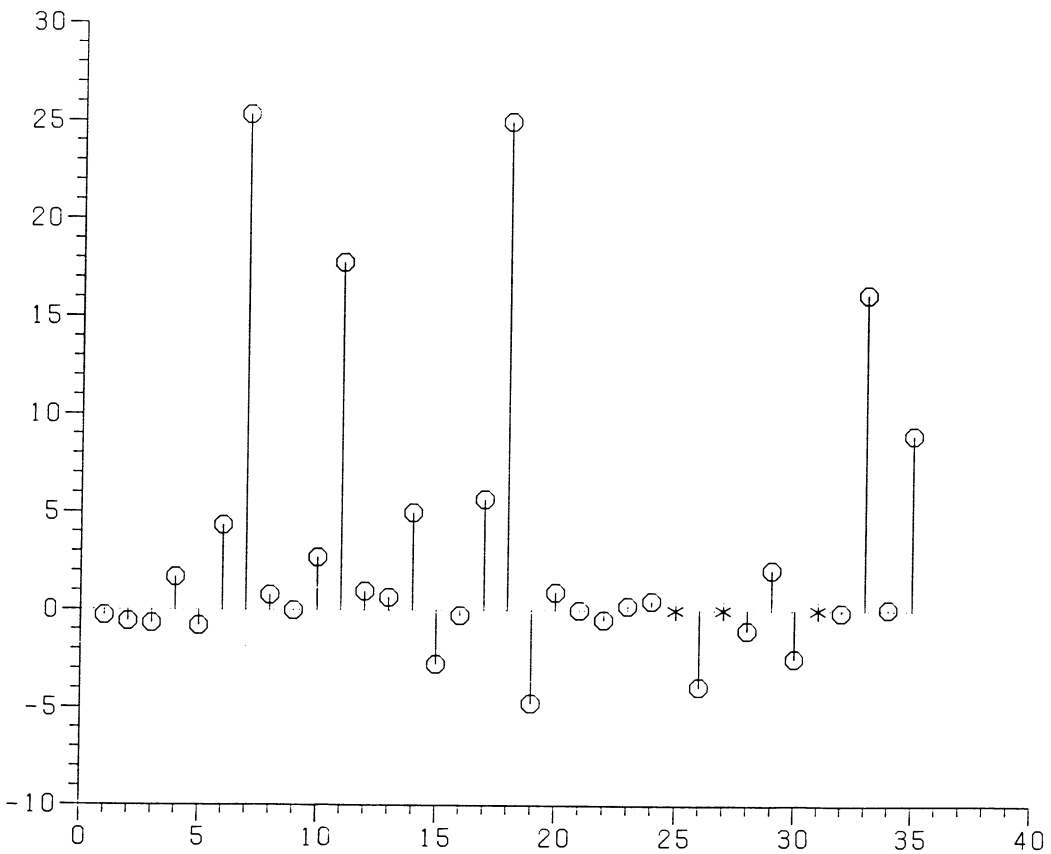

FIG. 2. *Hill race data: index plot of standardized least median of squares residuals.*

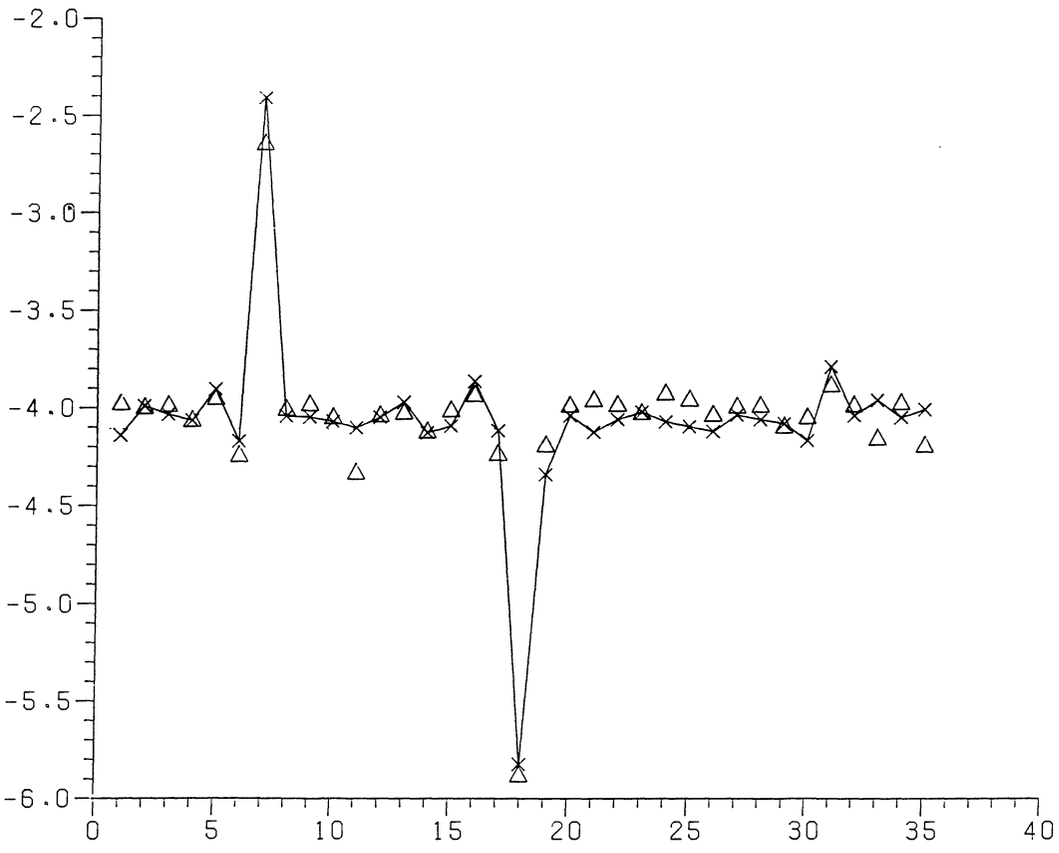

FIG. 3. *Hill race data: index plot of deletion versions of approximate score statistic for testing parametric transformation of the response: Δ, observation i deleted: ×, effect of deletion estimated.*

this leverage point is in agreement with the rest of the data. Further analysis would include checking the data in Table 1 both against the original fixture list and against the lists for other years. Another extension would be to the use of extreme value distributions, which are appropriate for modeling record times.

ACKNOWLEDGMENT

I am grateful to Geoff Cohen of the University of Edinburgh for introducing me to these data.

ADDITIONAL REFERENCES

ATKINSON, A. C. (1973). Testing transformations to normality. *J. Roy. Statist. Soc. Ser. B* **35** 473–479.

ATKINSON, A. C. (1985). *Plots, Transformations, and Regression.* University Press, Oxford.

ATKINSON, A. C. (1986a). Diagnostic tests for transformations. *Technometrics* **28** 29–37.

ATKINSON, A. C. (1986b). Masking unmasked. *Biometrika* **73** (in press).

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–246.

COOK, R. D. and WANG, P. C. (1983). Transformations and influential cases in regression. *Technometrics* **25** 337–343.

DEMPSTER, A. P., SELWYN, M. R., PATEL, C. M. and ROTH, A. J. (1984). Statistical and computational aspects of mixed model analysis. *Appl. Statist.* **33** 203–214.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics.* Wiley, New York.

JØRGENSEN, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika* **70** 19–28.

LANDWEHR, J. L., PREGIBON, D. and SHOEMAKER, A. C. (1984). Graphical methods for assessing logistic regression models (with discussion). *J. Amer. Statist. Assoc.* **79** 61–83.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

PREGIBON, D. (1980). Goodness of link tests for generalized linear models. *Appl. Statist.* **29** 15–24.

PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9** 705–724.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

WANG, P. C. (1985). Adding a variable in generalized linear models. *Technometrics* **27** 273–276.

WEISBERG, S. (1983). Some principles for regression diagnostics and influence analysis. *Technometrics* **25** 240–244.

WEISBERG, S. (1985). *Applied Linear Regression*, 2nd ed. Wiley, New York.