OUTLIERS MATTER IN SURVIVAL ANALYSIS

Gaineford J. Hall
William H. Rogers
Daryl Pregibon

April 1982

P-6761

OUTLIERS MATTER IN SURVIVAL ANALYSIS

Gaineford J. Hall, The Rand Corporation
William H. Rogers, The Rand Corporation
Daryl Pregibon, University of Washington*

1. INTRODUCTION

In recent years regression-type models have been extended beyond the usual least-squares framework. As the scope and frequency of applications have increased, so has apprehension about model validity and stability. Since assumptions are so difficult to check, they are usually not verified. This situation often undermines the credibility of nonlinear regression-type models.

This paper illustrates some diagnostic statistics for the Weibull log-linear regression model. These diagnostic statistics pinpoint extreme data points, quantify the effects of outliers and overly influential leverage points, and provide some data on the validity of the underlying maximum likelihood asymptotics. In five cancer datasets these diagnostics show that single observations can substantially affect inferences. The same observations also have drastic effects on the fitted Cox proportional hazards regression model, and Cox-Snell (1968) residual plots often do not pick out these observations.

Experience with ordinary least-squares regression diagnostics should guide the development of diagnostic tools for regression-like models. For more information about regression diagnostics see Cook (1977) and Belsley, Kuh, and Welsch (1980). In contrast with robust/resistant methods, diagnostics are inexpensive, easy to program on the computer, and straightforward to describe.

*Now at Bell Telephone Laboratories, Murray Hill, NJ.

There are several directions by which diagnostic tools can be extended from linear models to nonlinear models. Only some of these work well. See Pregibon (1981a) and Rogers (1982) for recent work in this area.

Section 2 of this paper describes the familiar Weibull regression model and the similarities between it and ordinary linear regression. Section 3 reviews Cook's distance (Cook (1977)) for the ordinary linear model. Section 4 briefly presents the development of similar regression diagnostics for the Weibull regression model. Section 5 discusses some classical residual plotting techniques for the Weibull regression model. Section 6 applies our techniques to five well-known and previously studied cancer datasets.

2. WEIBULL LOG-LINEAR REGRESSION MODELS

We shall consider those failure time models belonging to the intersection of the class of accelerated failure time models and the class of proportional hazard models. As is well known (Kalbfleisch and Prentice (1980)) this intersection is precisely the class of Weibull (including exponential) log-linear regression models. If T denotes the (random) lifetime of a subject, $Y = \log T$ and if $z = (z_1, \ldots, z_p)$ is the vector of covariates, this implies that

$$Y = \alpha + z\beta + \sigma W, \tag{2.1}$$

where the error term W is extreme value with density $\exp(w - \exp w)$, $-\infty < w < \infty$. Equivalently if the (conditional on z) hazard for T is

$$\lambda(t;z) = \lambda p(\lambda t)^{p-1} e^{z\gamma}, \tag{2.2}$$

the model for the transformed variable $Y = \log T$ is

$$Y = -\log \lambda + z(-\gamma p^{-1}) + p^{-1} W.$$

Thus $\alpha = -\log \lambda$, $\sigma = p^{-1}$, and $\beta = \gamma p^{-1}$. When $p = 1$, this yields the familiar exponential regression model.

The Cox nonparametric proportional hazards regression model postulates the hazard function as

$$\lambda(t;z) = \lambda_0(t) e^{z\gamma} \tag{2.3}$$

where the baseline hazard function $\lambda_0(t)$ is completely arbitrary and unspecified. Identification of parameters in Eqs. (2.1)-(2.3) shows that $\beta = -p^{-1}\gamma$, where γ is as in Eq. (2.3). Empirical comparisons of Weibull and Cox results support this identification (Kalbfleisch and Prentice (1980)). The similarity of model (2.1) and the ordinary regression model makes the Weibull log-linear regression class of models the obvious choice to develop diagnostics for examining the presence of influential and/or outlying cases for users of both the Weibull and Cox models.

3. REVIEW OF ORDINARY LEAST SQUARES REGRESSION DIAGNOSTICS

To motivate the development of influence diagnostics for the Weibull regression model, we briefly review recent developments of regression diagnostics for ordinary regression models. The principal diagnostic of interest here is Cook's distance (1977). See also

Belsley, Kuh, and Welsch (1980), and Cook and Weisberg (1980), Andrews and Pregibon (1978).

These regression diagnostics are used to identify influential extreme points. These are points which unduly affect estimates either because they are outliers in the dependent variable or <u>response space</u> or because they are outliers in the independent variables, or <u>factor space</u>. Regression diagnostics for the linear model are linear functions of the data, and the stability of the estimates due to small perturbations in the data is easily computed.

Let y_i denote the response for the ith case, and let x_i = $(1,x_{i1},\ldots,x_{ip})$ denote the corresponding vector of factors. Then the linear regression model specifies that

$$y_i = x_i\beta + \varepsilon_i,$$

where the error terms ε_i are uncorrelated, with mean 0 and common variance σ^2. Let Y denote the n x 1 vector of responses, X the n x (p + 1) matrix of explanatory (factor) variables, and ε the n x 1 vector of error terms. Then the model is

$$Y = X\,\beta + \varepsilon\ ,$$

and the OLS estimate of β is

$$\hat{\beta} = (X'X)^{-1}\,X'Y.$$

The vector of fitted values is

$$\hat{Y} = X\hat{\beta}$$
$$= X(X'X)^{-1} X'Y$$
$$= HY$$

where H is Tukey's hat matrix. The vector of computed residuals is

$$R = Y - \hat{Y}$$

$$= (I - H)Y.$$

If h_i is the ith diagonal component of H, then since H and I - H are idempotent,

$$Var(y_i) = h_i\sigma^2 ,$$

$$Var(r_i) = (1 - h_i)\sigma^2.$$

Simple diagnostics may be obtained from a traditional residual analysis. The computed residuals r_i and fitted values $\hat{y}_i$ are heteroscedastic and may be "studentized" to achieve more homogeneous residuals. The most common definition of studentized residual is

$$t_i = (y_i - \hat{y}_i)/(\hat{\sigma}\sqrt{1 - h_i})$$
$$= r_i/(\hat{\sigma}\sqrt{1 - h_i}) , \; r_i = y_i - \hat{y}_i,$$

where the estimate $\hat{\sigma}$ of σ is based on the full dataset. Plots of t_i are quite useful in detecting outliers in the response space. However, studentized residuals fail to assess the sensitivity of $\hat{\beta}$ to the position of x_i in factor space. To assess this sensitivity and to detect overly influential (and outlying) cases, Cook (1977) defines a metric D_i to determine the influence of the ith case on the estimate of β. Cook's distance is

$$D_i = [(\hat{\beta} - \hat{\beta}_{(i)})'X'X(\hat{\beta} - \hat{\beta}_{(i)})]/(p'\hat{\sigma}^2)$$

where $p' = p + 1$, and $\hat{\beta}_{(i)}$ is the estimate of β based on the dataset without the ith case. This distance defines a metric to determine the (scaled) magnitude of the shift $\hat{\beta} - \hat{\beta}_{(i)}$ in the estimate of β. Large values of D_i (say > 1) indicate that the ith case is quite influential in the estimate $\hat{\beta}$ of β and may warrant special consideration. Other investigators (Belsley, Kuh, and Welsch (1980)) have defined similar metrics. Cook (1977) also noted that D_i may be written equivalently as

$$D_i = [1/p'] \, t_i^2 [h_i/(1 - h_i)].$$

Note that t_i^2 is a measure of the (relative) magnitude of the response y_i, and $h_i/(1 - h_i)$ is a measure of the distance of x_j from the center of the factor variable configuration.

It is also instructive and useful to examine the influence of the ith case on each component of $\hat{\beta}$. Let

$$\Delta_i\hat{\beta} = (\hat{\beta} - \hat{\beta}_{(i)}), \quad 1 \le i \le n,$$

and $\Delta_i\hat{\beta}_j$ be the jth component of $\Delta_i\hat{\beta}$, for $1 \le j \le p$. Since some (but not all) of the p explanatory variables may be of much greater interest to the analyst (e.g., we may be most interested in treatment effects, or policy variables), it is important to examine the leverage of each case on the individual regression coefficients. As suggested in Cook (1977) and Belsley, Kuh, and Welsch (1980), we may compute $\Delta_i\hat{\beta}_j/se(\hat{\beta}_j)$ and plot this vs. the index number i, for each j.

4. DERIVATION OF WEIBULL LOG-LINEAR REGRESSION DIAGNOSTICS

In this section we will outline the derivation of analogous regression diagnostics for the Weibull model. For further details and a fuller treatment for more general nonlinear models, see Pregibon (1981a) and Rogers (1982).

The data for the Weibull log-linear regression model are of the form

$$(t_i, \delta_i, Z_{i1}, \ldots, Z_{ip}),\ 1 \le i \le n,$$

where t is the failure ($\delta = 1$) or censoring ($\delta = 0$) time to some event, δ is the censoring indicator, and Z is a vector of covariates. In the following discussion, think of $Y = \log T$ as a response, which is somehow related to the covariate vector through a regression-like model.

Consider more generally a parametric model which relates a response y_i to a vector of covariates Z_i, $1 \le i \le n$. Let L_i denote the contribution of the ith case to the likelihood function L, that is, $L = \Pi L_i$. The development here will be quite general, and we assume that $L_i = L(\eta_i; y_i)$ is a function of the ith response y_i and that η_i is a vector relating the unknown parameters to the covariate vector $x_i = (1,Z_i)$. Specifically, we assume that $\theta = (\beta, \lambda)$ is the unknown parameter vector where β always enters the likelihood specification via $x'\beta$. Thus

$\eta_i = (x_i'\beta, \lambda)$, and λ is a nuisance parameter.

For the Weibull model, let $w_i = (y_i - \alpha - Z_i\beta)/\sigma$, $f(w) = \exp(w - e^w)$, and

$$\delta_i = \begin{cases} 0, & \text{if ith case censored} \\ 1, & \text{if not} \end{cases}$$

Then

$$L_i = (\sigma^{-1}\ f(w_i))^{\delta_i}\ s(w_i)^{1-\delta_i}$$

where $s(w) = \sigma^{-1} \int_w^\infty f(u)du$ is the survival function. Here $\lambda = \sigma$.

We have thus specified the likelihood function of the data, and the log-likelihood is

$$\Lambda = \log L = \sum_{i=1}^{n} \log L_i = \sum_{i=1}^{n} \Lambda_i\ .$$

The score vector is

$$S(\theta) = (\partial/\partial\theta)\ \Lambda = \sum_{i=1}^{n} (\partial/\partial\theta)\ \Lambda_i = \sum_{i=1}^{n} S_i$$

and the (observed) information matrix is

$$\begin{aligned} R_2(\theta) &= -\ (\partial/\partial\theta)S(\theta) \\ &= -\sum_{i=1}^{n} (\partial/\partial\theta)S_i(\theta)\ . \end{aligned}$$

Define the accumulated cross-product matrix of scores as

$$R_1(\theta) = \sum_{i=1}^{n} S_iS_i'\ .$$

The score equations are nonlinear in θ, and iteration is required to obtain the maximum likelihood estimate $\hat{\theta}$. This is usually obtained by the Newton-Raphson iteration.

$$\theta_{h+1} = \theta_h + R_{2[h]}^{-1} S_{[h]},\ h = 0, 1, \ldots, *,$$

where $S_{[h]}$ and $R_{2[h]}$ are S and R_2 evaluated at θ_h. At convergence set $\hat{\theta} = \hat{\theta}_*$ and $R_2^{-1} = R_{2[*]}^{-1}$ is generally used as the large sample covariance matrix of $\hat{\theta}$. R_1^{-1} is another frequent choice.

This is typically where the standard maximum likelihood theory ends. It is also where the use of regression diagnostics begins. In particular we want to

1) identify cases not well explained by the model;
2) identify cases which exert undue influence on the fitted model; and
3) quantify the effects of such points on the current fit.

In addition, we wish to add the constraint that these goals be achieved with relatively little computational effort. This is extremely important for nonlinear modeling problems if the methods are to be routinely applied.

In order to develop diagnostics, we require a method to perturb the input (i.e., the data) so that we may observe the effect of such a perturbation on the output (i.e., some aspect of the fitted model). A standard model perturbation (used in linear regression diagnostics and elsewhere) is that of observation deletion. Thus we may examine the effect of the ith observation on the estimate $\hat{\theta}$ by computing $\hat{\theta}_{(i)}$, the estimate of $\hat{\theta}$ based on all the data but the ith case. If this value differs greatly from θ, then the fitted model is unduly sensitive to this data point and possible explanation for this behavior should be sought.

Since the calculation of $\hat{\theta}$ required iteration, so will that of $\hat{\theta}_{(j)}$. This is undesirable as the complete set of diagnostics for i = 1, 2,...,n would be prohibitively expensive and, more importantly, not necessary in most problems. The reason for the latter statement is that we do not require a precise measure of the influence of individual observations. For observations which are not extreme, $\hat{\theta}_{(i)}$ will be close to $\hat{\theta}$ and a one-step iteration will suffice. For extreme points, we are concerned less with exactitude than the direction and relative magnitude of the effect of the perturbation.

These considerations lead one to consider the class of one-step revised estimates

$$\tilde{\theta}_{(i)} = \hat{\theta} + R_{2(i)}^{-1} S_{(i)}$$

where $R_{2(i)}$ and $S_{(i)}$ are the observed information matrix R_2 and the score vector S computed without the ith case. Since $S_{(i)} = S - S_i$ and $S(\hat{\theta}) = 0$, we obtain

$$\tilde{\theta}_{(i)} = \hat{\theta} - R_{2(i)}^{-1} S_i.$$

Thus the shift in the estimate of θ from deleting the ith case is measured as

$$\Delta_i\hat{\theta} = \hat{\theta} - \tilde{\theta}_{(i)} = R_{2(i)}^{-1} S_i.$$

In order to standardize the measure $\Delta_i\hat{\theta}$ to an appropriate scale for assessment of its magnitude, we form the metric

$$F_i = (\Delta_i\hat{\theta})'M(\Delta_i\hat{\theta})$$

where M is some appropriately chosen matrix. Usually, M^{-1} is some estimate of the covariance structure of $\hat{\theta}$, or of $\hat{\theta}_{(i)}$, and may depend on i. For nonlinear problems, choice of M may be determined by computational ease. Additionally, it is desirable to use a matrix which is not too sensitive to departures from the underlying model assumptions.

Under the correct likelihood, $R_1 = R_2$ asymptotically. In general, this equality does not hold up when the model is false. Huber (1965) has shown that one can estimate the covariance matrix of the parameter estimate when the likelihood is not necessarily correct by using

$$\text{Cov }(\hat{\theta}) = R_2^{-1} R_1 R_2^{-1}.$$

The corresponding estimate of the covariance of $\hat{\theta}_{(i)}$ is

$$\text{Cov }(\hat{\theta}_{(i)}) = R_{2(i)}^{-1} R_{1(i)} R_{2(i)}^{-1}.$$

For deletion of the ith case we shall choose

$$M^{-1} = B \text{ Cov }(\hat{\theta}_{(i)}),$$

where B is the dimension of θ.

This choice of M (or M_i) leads to a particularly simple expression for the metric F_i; this is

$$F_i = B^{-1}(\Delta_i\hat{\theta})'(R_{2(i)}^{-1} R_{1(i)} R_{2(i)}^{-1})^{-1} (\Delta_i\hat{\theta})$$

$$= B^{-1}(S_i' R_{2(i)}^{-1})(R_{2(i)}^{-1} R_{1(i)} R_{2(i)}^{-1})^{-1} (R_{2(i)}^{-1} S_i)$$

$$= B^{-1} S_i' R_{1(i)}^{-1} S_i$$

$$= [S_i' R_1^{-1} S_i]/[B(1 - S_i' R_1^{-1} S_i)]$$

since $R_{1(i)} = R_1 - S_i S_i'$. The individual components $S_i' R_1^{-1} S_i$ or F_i can be plotted or printed as desired.[1] The statistic F_i can be interpreted as a confidence curve according to the F distribution with B and n - B - 1 degrees of freedom.

An alternative choice for M leading to another convenient closed-form metric is $M = R_2$. This yields the diagnostic

$$C_i = (\Delta_i\hat{\theta})' R_2(\Delta_i\hat{\theta})$$

$$= [S_i' R_2^{-1} S_i]/[(1 - S_i' R_2^{-1} S_i)^2]$$

We also tried several approximations to $\Delta_i\hat{\theta}$ which did not work as well as we had hoped. For example, start as before with $\Delta_i\hat{\theta} = R_{2(i)} S_i$. Moreover, $R_{2(i)} = R_2 + (\partial/\partial\theta)S_i$. Now if $(\partial/\partial\theta)S_i$ could be written as $(\partial/\partial\theta)S_i = b_i b_i'$ for some vector b_i, and

[1] A practical advantage of this diagnostic is that $0 < S_i' R_1^{-1} S_i \leq 1$.

$$(R - bb')^{-1} = R^{-1} + (R^{-1} bb' R^{-1})/(1 - b'R^{-1}b)$$

for a matrix R and vector b, we obtain for the one-step estimate

$$\tilde{\theta}_{(i)} = \hat{\theta} - [R_2^{-1} + (R_2^{-1} b_i b_i' R_2^{-1})/(1 - b_i' R_2^{-1} b_i)] S_i.$$

In linear models the b_i are easily obtained. In nonlinear models (such as Weibull regression), we may exploit the fact that for regular cases, $E(S_i S_i') = -E(\partial/\partial\theta)S_i$. Thus $R_{2(i)} \approx R_2 - S_i S_i'$, and hence

$$\tilde{\theta}_{(i)} \approx \hat{\theta} - [(R_2^{-1} S_i)/(1 - S_i' R_2^{-1} S_i)]$$

or

$$\Delta_i \hat{\theta} \approx [(R_2^{-1} S_i)/(1 - S_i' R_2^{-1} S_i)]$$

The problem with this diagnostic is the large random factor introduced by S_i. The simpler approximation $R_2 \approx R_{2(i)}$ was much more effective for data analysis. One may expect that the size of large $\Delta\hat{\beta}_{(i)}$ will be underestimated.

There are several other diagnostics of interest which can be expressed using the fundamental building blocks R_1, R_2 and S_i, $1 \leq i \leq n$. For example, since $R_1 = R_2$ asymptotically when the likelihood is correct, we may inquire as to how different these matrices are. Two measures of this difference are the determinant $|R_2^{-1}R_1|$ and the trace $tr(R_2^{-1}R_1)$. Ideally, $|R_2^{-1}R_1|$ would equal one, since if $R_2 =$

R_1 then $|R_2^{-1}R_1| = 1$. Similarly, ideally $tr(R_2^{-1}R_1)$ would equal B, the number of unknown parameters. Pregibon (1982) describes $tr(R_2^{-1}R_1)$ in more detail. See also White (1982) for new work in this area.

5. CLASSICAL RESIDUAL PLOTS

Given the tool kit developed thus far, we could compare Weibull regression diagnostics with no residual analysis. This comparison would be unfair unless we consider other existing options. Most prominent among these are the Cox-Snell residual plots (1968), the variance-stabilized plots of Aitken and Clayton (1980), and the graphical method of Lagakos (1981). (The latter method has not been included in our comparisons since the necessary software was not readily available.)

Let t_i denote the survival time of the ith individual and $\Lambda(t)$ the baseline cumulative hazard function. The ith Cox-Snell residual is

$$\exp\{(\log t_i - z_i\beta)/\sigma\}$$

If the model is correct, the e_i's form a censored sample from the exponential distribution with failure rate 1. To get actual estimates, we replace β and σ by their estimates $\hat{\beta}$ and $\hat{\sigma}$. If the model is appropriate, the e_i's should be similar to a censored exponential sample. The Kaplan-Meier survival curve estimate $\hat{S}(t)$ based on the residuals should, when plotted on a log scale, yield approximately a straight line with slope -1. The variance stabilizing transform Aitken and Clayton (1980)) is

$$\hat{T}(t) = \sin^{-1}\sqrt{\hat{S}(t)}$$

When plotted against $\sin^{-1}(\exp(-e_i/2))$ this should yield a straight line with slope + 1.

Bear in mind that the various methods (Cox-Snell, Lagakos, and ours) may address different aspects of goodness-of-fit. Cox-Snell appears designed to find gross outliers in failure times, where we are searching for influential observations. The regression analogy suggests the two ideas can be quite different. Lagakos appears concerned about the proportionality of the hazards over time, a different issue still.

6. EVALUATION OF FIVE SURVIVAL DATASETS

Let $V_1 = R_1^{-1}$ and $V_2 = R_2^{-1}$. For each of the five cancer datasets we will discuss some (or all) of the following diagnostics:

(1) The metric

$$F_i = S_i' R_1^{-1} S_i/[B(1 - S_i' R_1^{-1} S_i)]$$

(2) The score for location

$$\frac{\partial \Lambda_0}{\partial \beta_i}$$

plotted against the index and the continuous variables.

(3) For each of its β components, a plot of $\Delta_i\hat{\theta}_j \approx R_2^{-1}S_i/se(\theta_j)$ vs. the index number i.

(4) The trace $tr(R_2^{-1}R_1)$ and determinant $|R_2^{-1}R_1|$.

(5) A plot of the Cox-Snell residuals and the Aitken-Clayton transform plot.

6.1. A Leukemia Dataset

We will start by examining the simple leukemia dataset treated in Feigl and Zelen (1965). There were 33 cases (none of which was censored), stratified according to AG type (AG + refers to the presence of Ager rods or granulations). Within AG type, white blood cell counts (WBC) were used as a prognostic factor for survival. To keep matters simple, we shall examine the data for AG = 1. Note that cases 14, 15 and 17 all have WBC's of 100,000.

A Weibull regression model was fit to these 17 cases, with LOGWBC as the sole explanatory variable (making three parameters in all). Table 6.1 shows the estimates for α and β, where α is the intercept, and β is the coefficient of LOGWBC.

Table 6.1

SUMMARY STATISTICS FOR THE FEIGL-ZELEN DATA

Parameter	Coeff. (se)	-case 17
INTERCEPT	8.48 (1.73)	1.26
LOGWBC	-0.482(0.198)	-0.949
n	17	16
Trace	2.05	1.70
Determinant	0.56	0.69
Max F	1.27 (17)	0.24 (2)

Table 6.1 also presents the trace and determinant diagnostics. Note that the trace is 2.05, quite close to 2 (parameters). (In the Weibull model, ln σ is not significantly different from 0, indicating that an exponential model is appropriate.)

Figures 6.1a-d all point to case 17 as the most influential case. Table 6.1 shows the parameter estimates after deletion of case 17. Note that α and β have shifted by more than two (original) standard deviations. Note that the effect of case 17 on β is identified by the diagnostic, but is actually much larger than the diagnostic suggests. Figures 6.1e-h should be compared with Figs. 6.1a-d. The second set of figures demonstrate how well-behaved data appear.

Evidently case 17 is overly influential since the patient lived a comparatively long time (65 days) despite a WBC of 100,000. How should the analyst respond? Given a need to maintain objectivity, we can only throw out case 17 if there are other compelling reasons to do so. However, close scrutiny of this case might reveal an excluded variable, possibly related to treatment or a subsequent remission. Since the effect of LOGWBC is much stronger without case 17, we can be sure of its relation to survival.

6.2. A Myeloma Dataset

Krall (1975) used these data to illustrate a search for factors predicting the survival of myeloma patients. Out of 16 possible prognostic factors, Krall selected a model based on three variables (LOGBUN, HEMOGLOBIN, and CALCIUM). Using a slightly different form of the exponential survival model, Lawless and Singhal (1978) also find these to be the best set of three variables. We applied the regression diagnostics to the Lawless and Singhal exponential model[1], detecting observation 40 to be the most significant case, shifting the observation to its 20 percent confidence ellipsoid. The index plots, Figs. 6.2a-b and 6.2f-h, show that although case 40 is highly influential, so are cases 1 and 2, for certain components.

Figures 6.2c-e plot the score for location ($\partial A_i/\partial \beta_0$) against each of the independent variables. Figure 6.2c suggests that either the variable LOGBUN enters curvilinearly, or there are three outliers. A curvilinear term is significant if fitted.

The Cox-Snell residual plot suggests that the set of cases 44, 48, and 40 is separated from the main body of data. However, these points do not appear to be outliers in either Fig. 6.2i or Fig. 6.2j.

Observation 40 is a multivariate outlier because it has an unusual combination of explanatory variables. The effect on the fitted parameters is noted in Table 6.2. The effects became substantially stronger. For example, HEMOGLOBIN changes from being nonsignificant to definitely significant.

After case 40 is removed, case 2 is the most significant among the remaining data, although case 48 has the largest residual among those remaining. Even though case 2 has a small residual (Fig. 6.2i), it is more influential because of its high value of CALCIUM (at 18, the largest value of CALCIUM in the entire database). Thus cases may be

[1]Since two patients survived only one week (1/4 month), Krall added one month to all survival times. Lawless and Singhal did the same. This is unnecessary for survival models of the type $\lambda(t,z) = \lambda_0(t)\exp(z\beta)$, so we did not add the one month.

Table 6.2

SUMMARY FOR THE MYELOMA EXAMPLE

Variable	Coeff (se)	-40	-40,2	-2	Cox	Cox-40
INTERCEP	5.743 (1.731)	7.172	6.134	4.779		
LOGBUN	-1.559 (0.574)	-1.885	-1.719	-1.506	1.63	1.90
HEMOGLOBIN	0.108 (0.059)	0.161	0.154	0.116	-.125	-.166
CALCIUM	-0.126 (0.137)	-0.290	-0.201	-0.050	.133	.262
n	65	64	63	64		
Trace	5.08	4.14	3.96	4.65		
Det	1.8	1.10	0.91	1.57		
Max F	0.39 (40)	0.17 (2)	0.17(44)	0.34(40)		

influential because of large computed residuals or because of one or more extreme values in the factor space.

Table 6.2 also shows that after deletion of case 40, the trace is much closer to 4 (parameters) and the determinant much closer to 1, as desired. The summary also shows that case 2 is moderately influential, particularly for CALCIUM, because of its value of 18. After removing cases 40 and 2, CALCIUM is again nonsignificant. Apparently case 2 is the sole reason for the presence of CALCIUM in the model.

As Kalbfleisch and Prentice point out, the parameters for the exponential and Cox models are usually very close. Outliers we have found using the parametric Weibull and/or exponential log-linear regression models are also found to be extremely influential in fitting the Cox proportional hazards regression model. For the myeloma data we found that with case 40 the chi-square significance test for HEMOGLOBIN is not significant, but after the outlier is removed this variable becomes highly significant. Our results for the Cox model (with and without observation 40) are shown in the last two columns of Table 6.2. Users of Cox's model should not be complacent merely because this model is termed nonparametric. The mechanism for describing the relationship between survival and the prognostic factors has just as many parameters as the exponential survival regression method.

6.3. An Advanced Head and Neck Carcinoma Dataset

Our third example is a dataset which was kindly made available to us by Dr. David Schoenfeld of the Sidney Farber Cancer Institute. The study was a randomized clinical trial comparing three chemotherapy regimens (labeled A, B, and C). Two stratification variables--performance status (not ambulatory = 2) and stage of disease (three stages, least widespread = 3)--were used for treatment balance. The remaining prognostic variable was age. A Weibull model was fit to the data (Table 6.3). Performance and staging appear to be the only

Table 6.3

SUMMARY FOR THE HEAD AND NECK EXAMPLE

Variable	Coeff (se)	-8	-8,39	Cox	Cox-8
INTERCEPT	6.222 (0.481)	6.422	6.545		
TREAT1	-0.005 (0.139)	-0.056	-0.004	0.016	0.057
TREAT2	-0.122 (0.173)	-0.136	-0.076	0.208	0.217
STAG1	-0.371 (0.165)	-0.291	-0.288	0.471	0.412
STAG2	-0.140 (0.160)	-0.050	-0.096	0.168	0.097
PERF	-0.660 (0.135)	-0.616	-0.586	0.887	0.859
AGE	0.0045(0.008)	-0.0009	-0.0043	-0.0007	0.0035
(LNSIGMA)	-0.322 (0.068)	-0.339	-0.365		
n	202	201	200	202	201
Trace	9.24	8.92	8.66		
Det	2.18	1.68	1.34		
Max F	0.16 (8)	0.13 (39)	0.09 (70)		

important variables. In particular, the treatments appear to have little effect.

From the plots (Fig. 6.3) it appears that there are several highly influential cases, the most important of which is case 8. The five most influential cases are 8, 39, 68, 70, and 138. The first four have performance = 1, all are censored, and all survived a relatively long time.

After deleting case 8 (next column of Table 6.3), staging no longer appears significant. The summaries show that although case 39 is not severely influential, it is still different from the main body of data. The last column records the parameter estimates with cases 8 and 39 deleted. Again, only performance appears essential. The plot of the score against age (Fig. 6.3c) appears to be somewhat heteroscedastic, with more variation on the right. This suggests a nonproportional hazard function, rising faster for young patients than old ones. This effect is significant if tested.

The Cox-Snell plots for this dataset point to a longer right tail than the Weibull allows. Although Huber's robust estimate for the parameter covariance matrix protects the validity of our inferences, this would be a good reason to prefer Cox's model over Weibull regression. Nevertheless, the Weibull model still helps us find influential observations for both models. Observation 8 changes STAG1 from significant to nonsignificant in the Cox model also.

6.4. The Glasser Lung Cancer Dataset

Glasser (1967) examines a set of lung cancer data from a clinical trial. The data consist of the number of days survived after surgery and the age of each patient in two groups of patients with lung cancer. Group 1 patients are those with low vital capacity/predicted vital capacity ratios. Group 2 patients are those with high ratios. Two patients are listed with zero duration. Since the Weibull model requires log-duration, this value was changed to 0.5 (the mean of a uniform (0,1) distribution).

The results obtained from the Weibull regression are presented in Table 6.4. Both TREAT and AGE appear significant. The trace is 4.67, vs. 4 (parameters), and the determinant is 1.65. The index plots "peak" at i = 36, since there are 36 cases in Group 1, and the duration times are in increasing order. The strange appearance of plot 6.4 is an artifact of the unequal sample sizes for the two treatment groups and

Table 6.4

SUMMARY OF RESULTS FOR THE GLASSER LUNG CANCER DATA

Variable	Coeff (se)	-35,36	-1,37	-35,36,1,37
INTERCEPT	8.873 (1.551)	9.676	6.159	9.307
TREAT	0.897 (0.430)	1.330	0.814	1.233
AGE	-0.057 (0.026)	-0.077	-0.051	-0.070
(LNSIGMA)	0.338 (0.120)	0.288	0.256	0.204
n	131	129	129	127
Trace	4.67	4.37		4.38
Det	1.65	1.29		1.30
Max F	0.09 (36)	0.07 (31)	0.09 (36)	0.07 (42)

the sort order of the dataset. The gap in Fig. 6.3c indicates that much information is contained in the identity of the censored observations, as opposed to the actual survival times. The five most influential cases appear to be 36, 35, 1, 37 and 42. Cases 1 and 37 may be expected to be important, since their duration times were actually zero. Cases 35 and 36 are influential because they are two of the oldest in Group 1, yet survived at least 84 days longer than the next longest survivor. These cases are also identified by the Aitken-Clayton version of the Cox-Snell plots.

The results from deleting 36 and 35 are presented in the next two columns of Table 4.1. The TREAT and AGE effects appear stronger. The trace is now 4.37 and the determinant is 1.29.

Since there is some uncertainty about how to handle cases 1 and 37 (with zero duration time), we also checked what would happen without these cases. Cases 35 and 36 are the most influential, with case 42 the next most influential. The last column shows that when four cases are removed, case 42 is still influential. This is because case 42 only survived 5 days, while at a fairly young age (40). Thus, cases of advanced age who survive a long time, or of young age who fail in a short time, may be influential. Note that this "left-hand" observation is found despite a declining hazard function. The Cox-Snell plot cannot do this.

6.5. The VA Lung Cancer Dataset

Prentice (1973) presented a much-analyzed dataset on the survival of advanced VA lung cancer patients originally collected by the Veterans Administration Lung Cancer Study Group. Survival time is predicted by a test treatment (coded as the dummy variable TEST1), Karnofsky performance status (PERFORM), months from diagnosis (TDIS), age in years (AGE), prior therapy (10 = yes, 0 = no), and cell type (SQUAMOUS, ADENO, SMALL, and LARGE). This dataset has been analyzed with both the Weibull regression model and the Cox proportional hazards model.

The results for the Weibull regression, together with our diagnostics, are shown in Table 6.5A. Case 44 appears to be the most influential overall, and it has an appreciable effect on the coefficient of TEST. From the DBETA, case 44 is pulling the coefficient of TEST up by about 0.4 standard errors. However, a balanced view of the influential observations (see Table 6.5B) leaves the inference about TEST intact.

The coefficient of SMALL (relative to LARGE, the omitted cell type), is also strongly affected by several of the outliers. Removing the four biggest outliers makes SMALL strikingly significant, but SQUAMOUS less significant. Cell type remains an important determinant of survival, whether or not outliers are included.

Two of the influential observations (cases 106 and 118) point up possible data or misspecification problems. A Karnofsky performance rating of 10, near death, seems suspiciously low in light of the subsequent long survival. Additional runs were made testing a spline function for PERFORM (by introducing a variable SPLINE = max(0,PERFORM-50)), dropping suspicious case 118. This variable is significant (T = -2.9), indicating that performance is much more important in its lower range. The linear specification of TDIS is problematic, highlighted by case 106. However, TDIS does not appear to matter in any specification.

Table 6.5A

DIAGNOSTICS FOR FULL VA LUNG CANCER DATASET

Model and DBETAS

Variable	Coefficient	se	T	Max |DBETA|	case
INTERCEP	2.6575	0.747	3.56	0.2559	17
PERFORM	0.0301	0.005	6.48	-0.0021	118
TDIS	-0.0004	0.007	-0.06	0.0038	106
AGE	0.0057	0.010	0.57	-0.0046	17
PRIORTH	-0.0046	0.021	-0.22	0.0103	75
SQUAMOUS	0.3986	0.227	1.75	0.0934	75
SMALL	-0.4270	0.237	-1.80	0.0988	44
ADENO	-0.7263	0.188	-3.85	-0.0705	58
TEST	0.2038	0.177	1.31	0.0713	44
(LNSIGMA)	-0.0744	0.069	-1.08	0.0386	44

Largest Influence Observations

Case	DAYS	F	C	Comment
44	392	0.32	1.94	Out of place in the SMALL group.
75	991	0.16	967.12	Long survivor.
118	48	0.13	6.53	PERFORM=10 is lowest.
17	384	0.11	11.13	Very young. Long survivor for SMALL.
106	51	0.07	0.48	TDIS=87 is an outlier.
58	553	0.07	0.43	Long survivor.

Table 6.5B

BEHAVIOR OF COEFFICIENTS WHEN INFLUENTIAL POINTS ARE DELETED

Coefficient	Full	(T)	-44	(T)	-44,75	(T)	-44,75 118,17	(T)
INTERCEP	2.66	(3.6)	2.90	(3.9)	2.72	(3.8)	2.42	(3.7)
PERFORM(/100)	3.01	(6.5)	3.12	(6.9)	3.12	(7.2)	3.39	(8.5)
TDIS(/1000)	-0.44	(-0.1)	2.74	(0.4)	5.76	(0.8)	6.73	(1.0)
AGE(/100)	0.56	(0.6)	0.06	(0.1)	0.36	(0.4)	0.56	(0.7)
PRIORTH(/100)	-0.46	(-0.2)	-0.25	(-0.1)	-1.52	(-0.8)	-0.97	(-0.6)
SQUAMOUS(/10)	3.99	(1.8)	3.88	(1.8)	2.79	(1.3)	2.97	(1.4)
SMALL(/10)	-4.27	(-1.8)	-5.53	(-2.5)	-5.86	(-2.7)	-6.61	(-3.2)
ADENO(/10)	-7.26	(-3.8)	-7.09	(-3.7)	-7.17	(-3.8)	-8.17	(-4.5)
TEST(/10)	2.31	(1.3)	1.34	(0.8)	2.01	(1.3)	1.92	(1.3)
LNSIGMA	-0.07	(-1.1)	-0.14	(-2.1)	-0.16	(-2.4)	-0.21	(-3.0)

Another troublesome aspect of this dataset is the low value for DET, 0.26 for the full dataset. This is the farthest from 1 in the five datasets studied. Figure 6.5 shows a plot of the score against the variable PERFORM. This plot suggests a nonproportional hazard. We carried out a test of proportionality by allowing LNSIGMA to be a function of PERFORM in an expanded likelihood setting. The results were significant.

7. SUMMARY AND CONCLUSIONS

This paper has presented a regression diagnostic for survival analysis (failure) models which is similar to Cook's distance for ordinary linear models and has demonstrated its application on five cancer datasets. We have also compared the diagnostic results with the usual log survival curve plot of the Cox-Snell residuals and the corresponding variance-stabilized plot. The diagnostic was applied to both Weibull and exponential regression models, and in one example the results were compared with those from the Cox proportional hazards model.

The results demonstrated that

- o Generally, but not always, the most influential observations (cases) possess the largest Cox-Snell residuals. A case which has a variable far out in the factor space may be more influential than a case with a large residual. Either kind of outlier can affect inference.
- o The plots of the log survival curve of the residuals and the corresponding variance-stabilized transformation generally do not indicate the importance (influence) of large residuals on estimated parameters.
- o Regression parameter estimates from the Cox proportional hazards model are just as sensitive to influential cases as are fully parametric models.
- o Influential observations often suggest other modelling deficiencies, such as a poorly-specified factor, nonproportional hazards, or an omitted variable. The analyses indicate how broadly-based the conclusions are. We found no evidence to support automatic exclusion of outlying observations.
- o Outlier diagnostics based on linearization (the log-likelihood and its derivatives) work very well. The measure of importance tends to be compressed at its upper end, so the impact of very influential observations tends to be understated.

REFERENCES

Aitken M, Clayton D: The fitting of exponential, Weibull, and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics* 29(2): 156-163 (1980).

Andrews D, Pregibon D: Finding the Outliers that Matter. *J Royal Statist Soc B* 40: 85-94 (1978).

Belsley DA, Kuh E, Welsch RE: *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. J Wiley and Sons (1980).

Cook RD: Detection of influential observations in linear regression. *Technometrics* 19: 15-18 (1977).

Cook RD, Weisberg S: Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics* 22(4): 495-508 (1980).

Cox DR: Regression models and life tables (with discussion). *J R Statist Soc Ser B* 34: 187-220 (1972).

Cox DR, Snell EJ: A general definition of residuals. *J Royal Statist Soc Ser B* 30: 826-838 (1968).

Daniel C, Wood FS: *Fitting Equations to Data* (2nd edition). J Wiley and Sons (1980).

Elandt-Johnson RC, Johnson NL: *Survival Models and Data Analysis*. J Wiley and Sons (1980).

Feigl P, Zelen M: Estimation of exponential survival probabilities with concomitant information. *Biometrics* 21: 826-838 (1965).

Glasser M: Exponential survival with covariates. *Journal of the American Statistical Association* 62: 561-568 (1967).

Greenberg RA, Bayard S, Byar DP: Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit in an exponential survival model. *Biometrics* 30: 601-608 (1974).

Helwig JT, Council KA, Reinhardt PS (ed): *SAS User's Guide, 1979 Edition*. SAS Institute, Inc. (1979).

Huber PJ: The behavior of maximum likelihood estimates under nonstandard conditions. *Fifth Berkeley Symposium*. 221-233 (1965).

Kalbfleisch JD, Prentice RL: The Statistical Analysis of Failure Time Data. J Wiley and Sons (1980).

Kister S, Aroesty J, Rogers WH, Huber C, Willis K, Morrison P, Shangold G, Lincoln T: An analysis of predictor variables for adjuvant treatment of breast cancer. Cancer Chemotherapy and Pharmacology 2: 147-158 (1979).

Krall JM, Uthoff VA, Harley JB: A step-up procedure for selecting variables associated with survival. Biometrics 31: 49-59 (1975).

Lagakos SW: The graphical evaluation of explanatory variables in proportional hazard regression models. Biometrika 68: 93-98 (1981).

Lawless JF, Singhal K: Efficient screening of nonnormal regression models. Biometrics 34: 318-327 (1978).

Pregibon D: Logistic regression diagnostics. Annals of Statistics 9(4): 705-724 (1981a).

Pregibon D: Diagnostics for some nonstandard models with medical applications. Technical Report #48, Department of Biostat, School of Public Health and Community Medicine, University of Washington, (1981b).

Pregibon D: An Alternative Covariance Estimate for Generalized Linear Models. GLIM Newsletter, Issue #6. June 1982. (1982)

Prentice RL: Exponential survivals with censoring and explanatory variables. Biometrika 60: 279-288 (1973).

Rogers WH: A tool kit for regression-like models (1982, forthcoming).

Rogers WH, Hanley J: Weibull regression and hazard estimation. SAS Users' Group Proceedings (February 1982).

Rogers WH, Opfell R, Plotkin D: Prognostic factors with combination chemotherapy for non-oat-cell carcinoma of the lung. Abstract. ASCO (1979).

Tsiatis AA: A large-sample study of Cox's regression model. Annals of Statistics 9: 93-108 (1981).

Weisberg S: Applied Linear Regression. J Wiley and Sons (1980).

White H: Maximum likelihood estimation of misspecified models. Econometrica 50(1): 1-25 (1982).

FIGURE 6.1A: PLOT OF INFLUENCE I AGAINST OBSERVATION INDEX
PLOT OF _FINFLU_*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

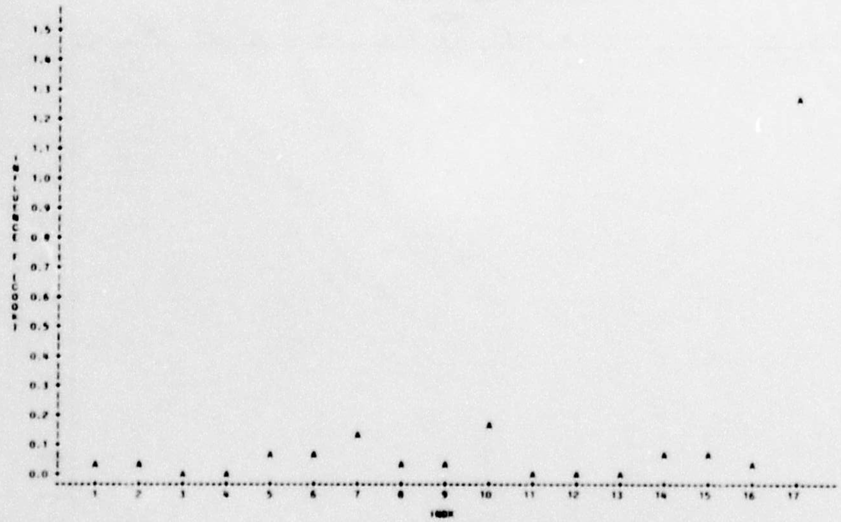


FIGURE 6.1B: PLOT OF SCORE AGAINST OBSERVATION INDEX
PLOT OF RU*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

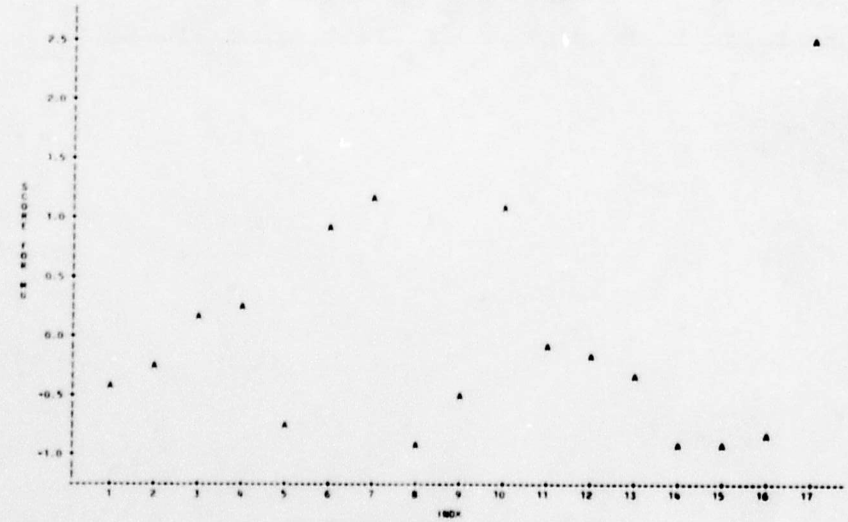


FIGURE 6.1C: PLOT OF SCORE AGAINST VARIABLE LOGWBC
PLOT OF RU*LOGWBC LEGEND: A = 1 OBS, B = 2 OBS, ETC.

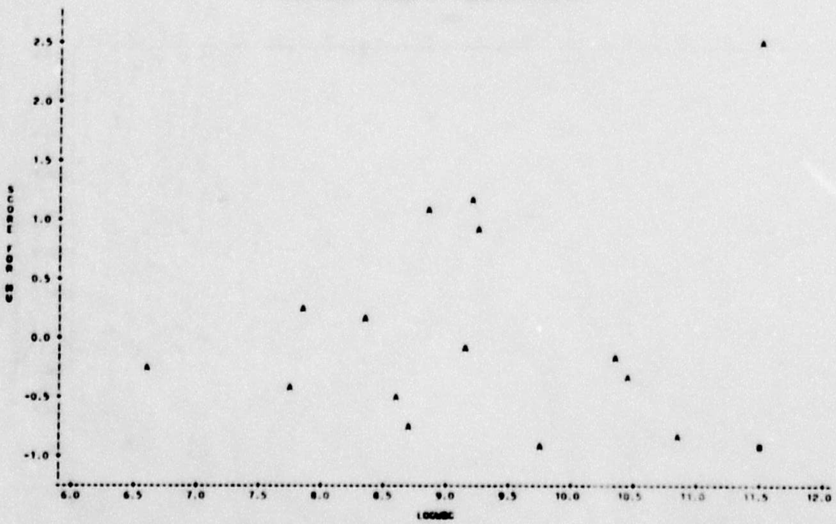


FIGURE 6.1D: STANDARDIZED DBETA FOR LOGWBC AGAINST INDEX
PLOT OF DBETA002*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

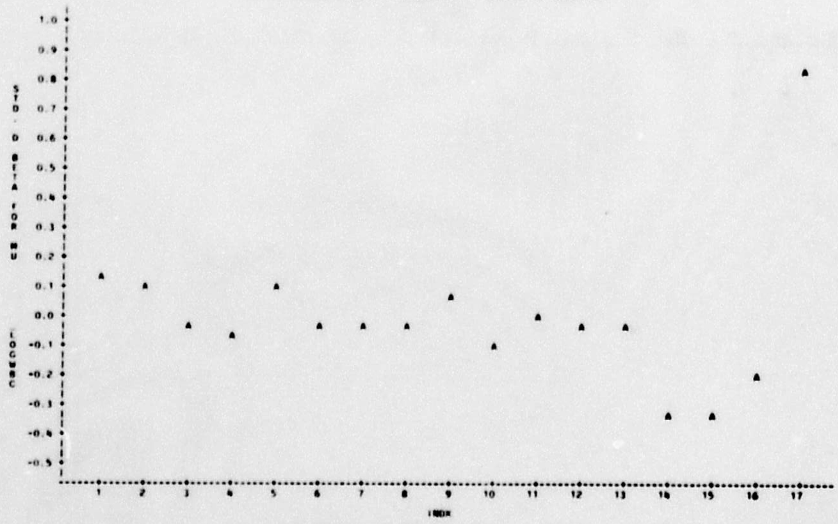


FIGURE 6.1E: PLOT OF INFLUENCE F AGINST OBSERVATION INDEX
OBSERVATION 17 DELETED -- COMPARE WITH FIG. 6.1A

PLOT OF _FINFLU_*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

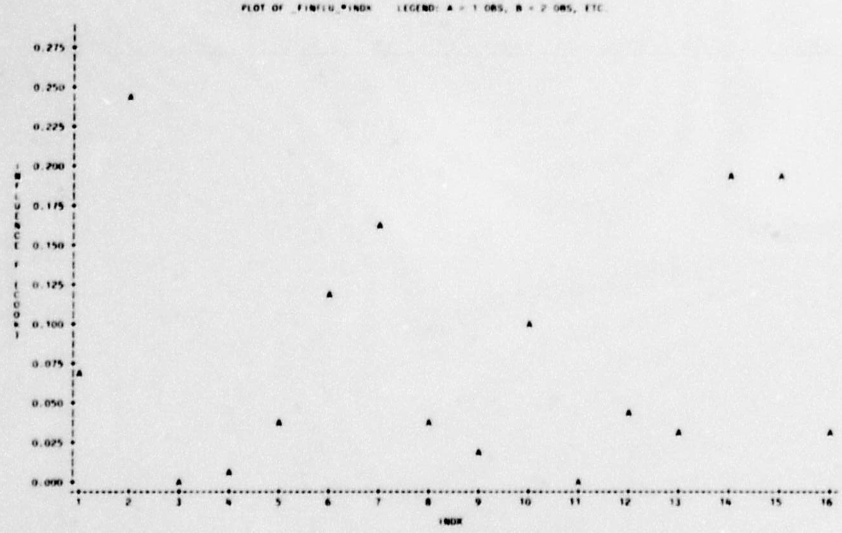


FIGURE 6.1F: PLOT OF SCORE AGAINST OBSERVATION INDEX
OBSERVATION 17 DELETED -- COMPARE WITH FIG. 6.1B

PLOT OF MU*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

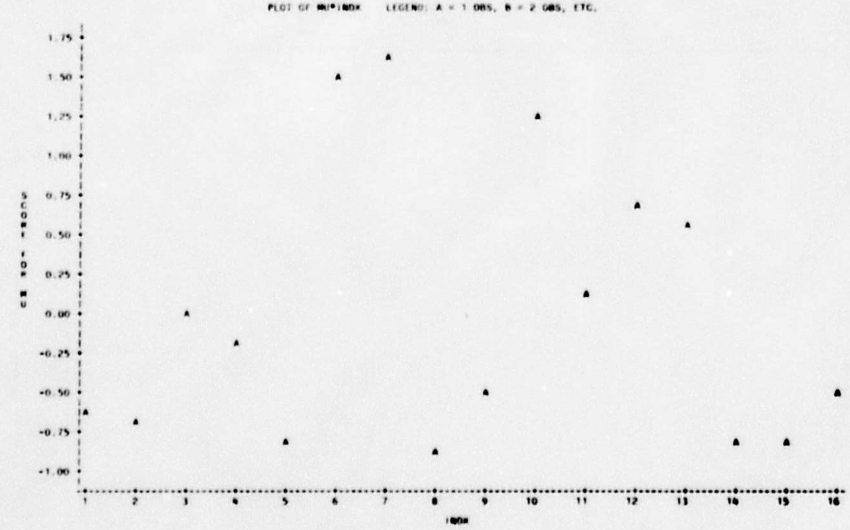


FIGURE 6.1G: PLOT OF SCORE AGAINST VARIABLE LOGWBC
OBSERVATION 17 DELETED -- COMPARE WITH FIG. 6.1C

PLOT OF MU*LOGWBC LEGEND: A = 1 OBS, B = 2 OBS, ETC.

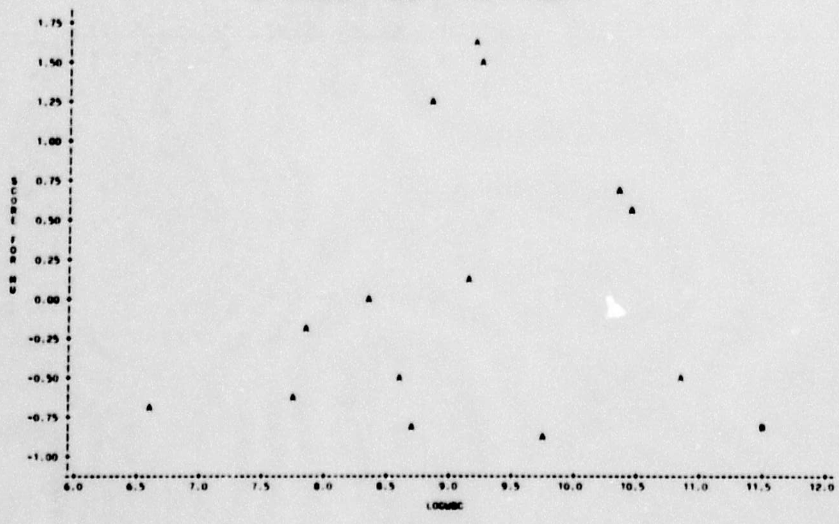


FIGURE 6.1H: STANDARDIZED DBETA FOR LOGWBC AGAINST INDEX
OBSERVATION 17 DELETED -- COMPARE WITH FIG. 6.1D

PLOT OF DBETA002*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

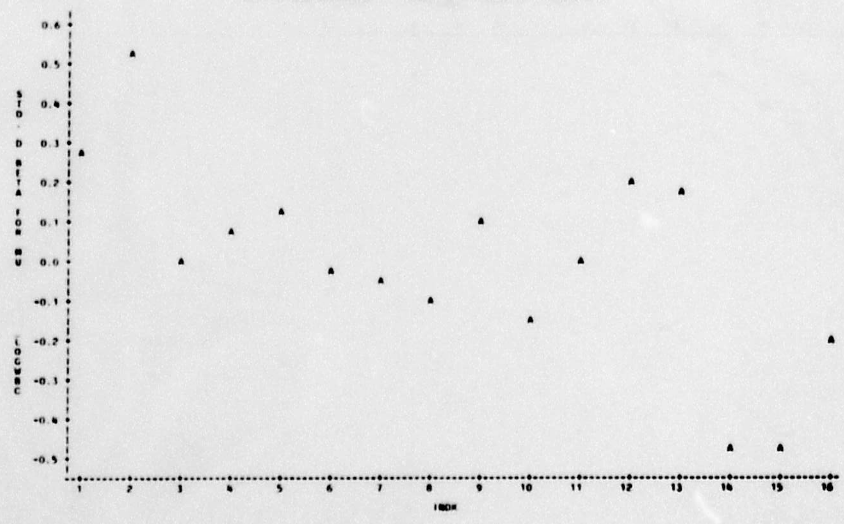


FIGURE 6.2A: PLOT OF INFLUENCE I AGAINST OBSERVATION INDEX

PLOT OF _FINFLU_*INDX   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

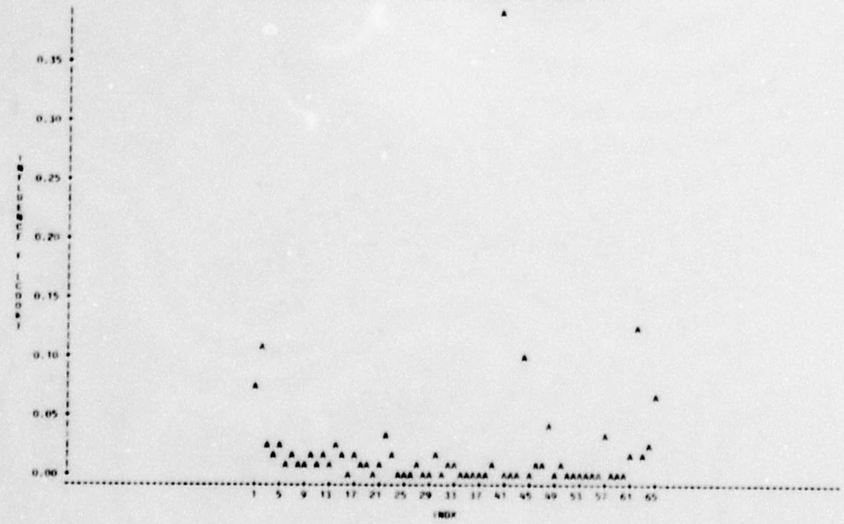


FIGURE 6.2B: PLOT OF SCORE AGAINST OBSERVATION INDEX

PLOT OF MU*INDX   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

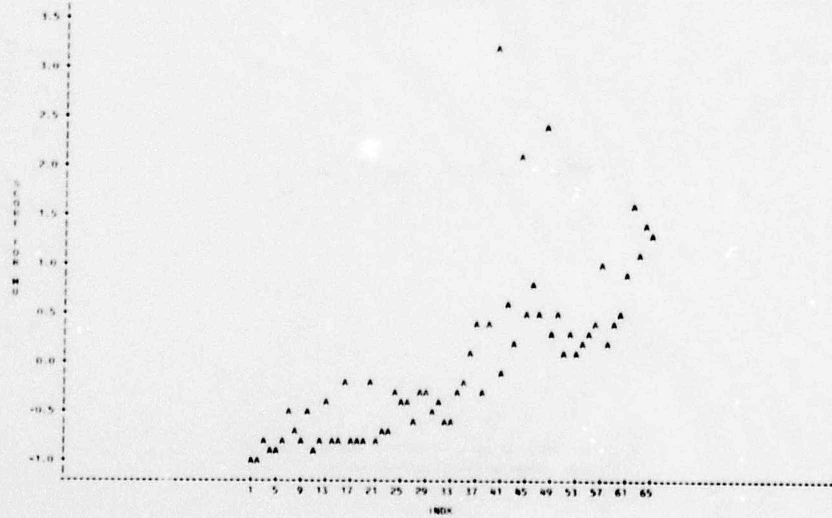


FIGURE 6.2C: PLOT OF SCORE AGAINST VARIABLE LOGBUN

PLOT OF MU*LOGBUN   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

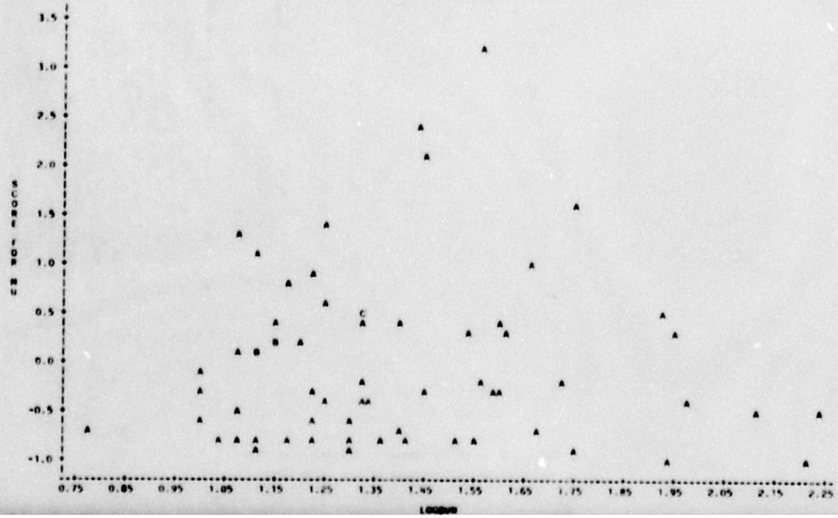


FIGURE 6.2D: PLOT OF SCORE AGAINST VARIABLE HEMOGLOB

PLOT OF MU*HEMOGLOB   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

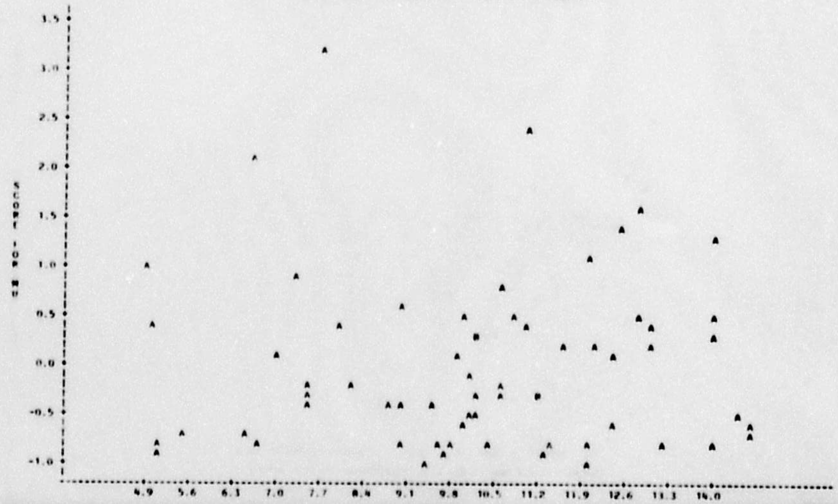


FIGURE 6.2E: PLOT OF SCORE AGAINST VARIABLE CALCIUM

PLOT OF MU*CALCIUM LEGEND: A = 1 OBS, B = 2 OBS, ETC.

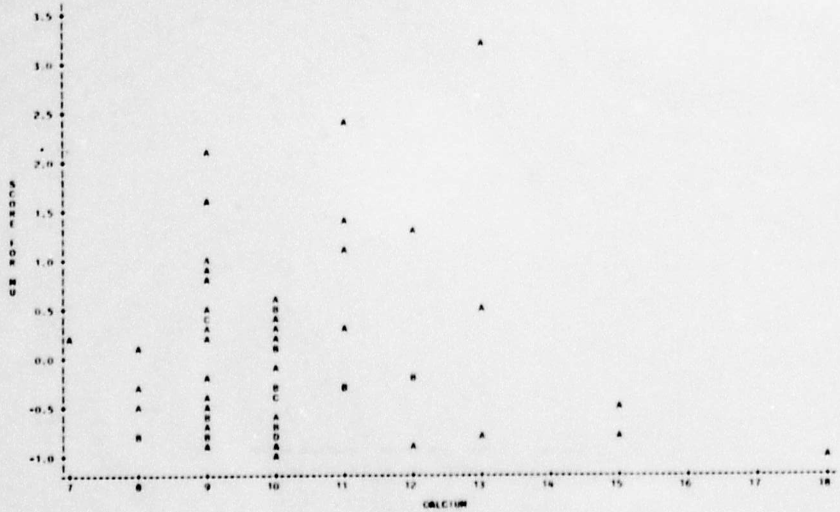


FIGURE 6.2F: STANDARDIZED DBETA FOR LOCBUN AGAINST INDEX

PLOT OF DBETA002*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

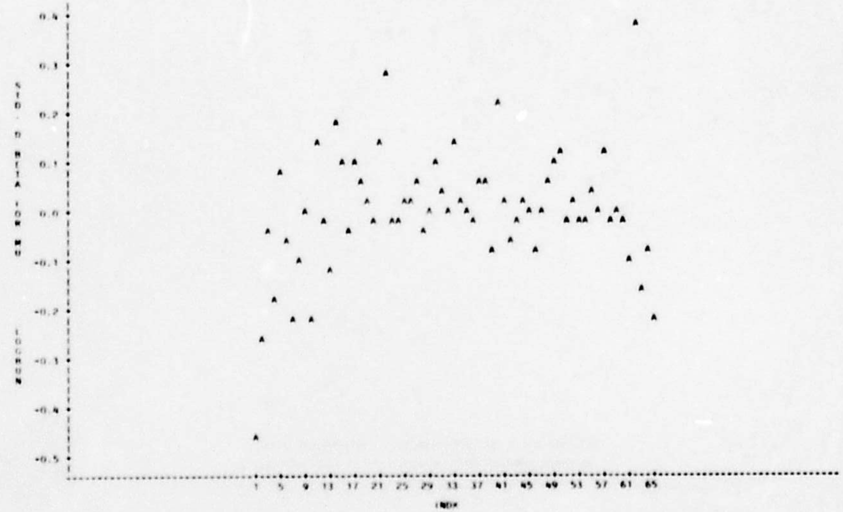


FIGURE 6.2G: STANDARDIZED DBETA FOR HEMOGLOB AGAINST INDEX

PLOT OF DBETA003*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

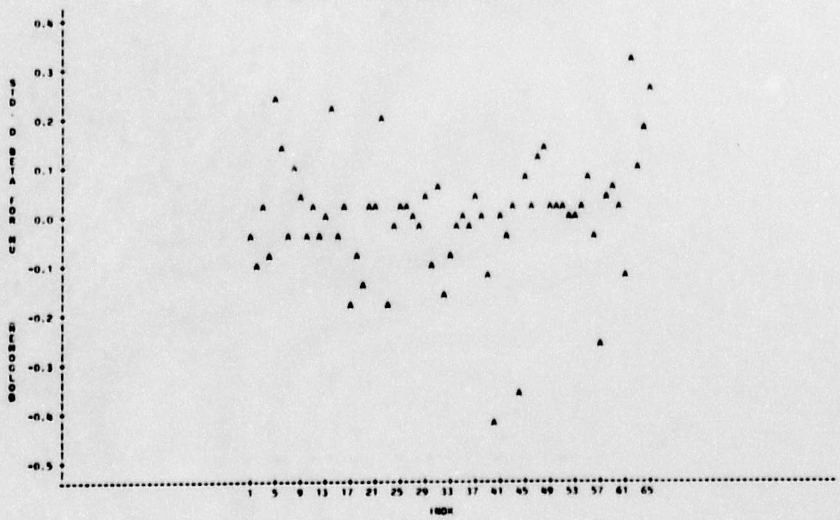


FIGURE 6.2H: STANDARDIZED DBETA FOR CALCIUM AGAINST INDEX

PLOT OF DBETA004*INDX LEGEND: A = 1 OBS, B = 2 OBS, ETC.

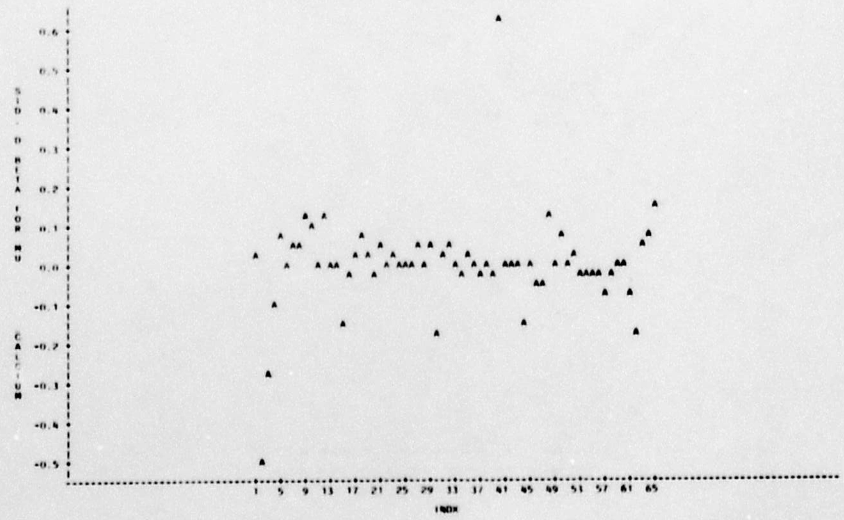


FIGURE 6.21: COX-SNELL RESIDUAL PLOT FOR MYELOMA DATA

PLOT OF LOGKAPM*XTIME   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

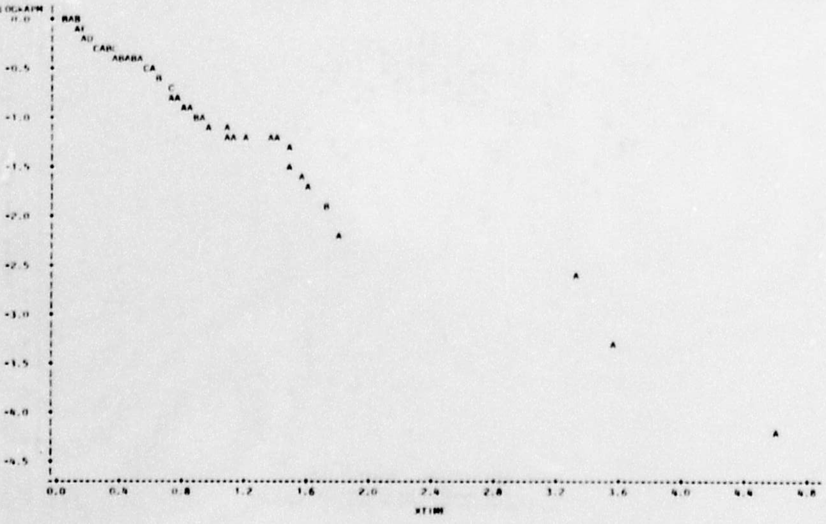


FIGURE 6.22: AITKEN-CLAYTON STABILIZED COX-SNELL RESIDUAL PLOT

PLOT OF AITKENS*AITKENT   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

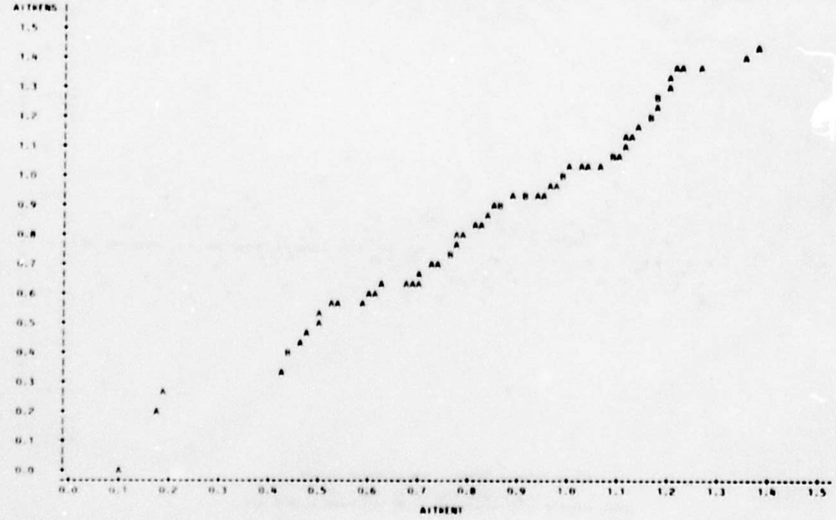
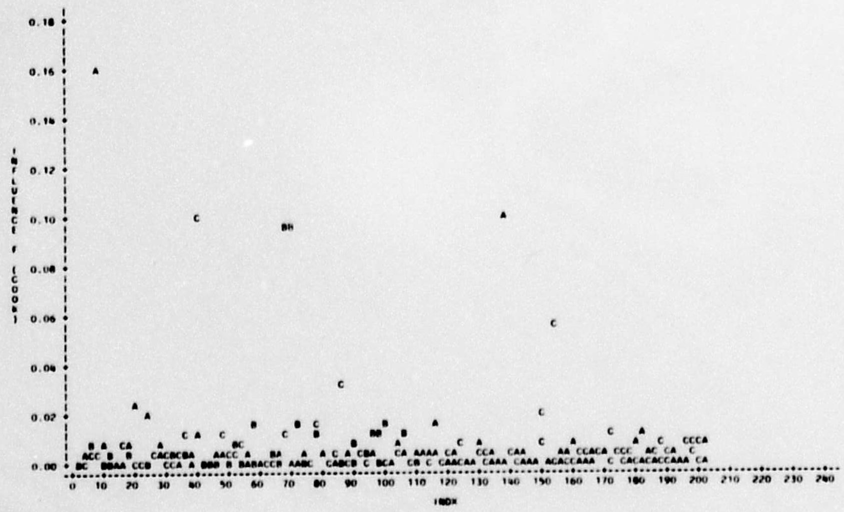


FIGURE 6.3A: PLOT OF INFLUENCE F AGAINST OBSERVATION INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

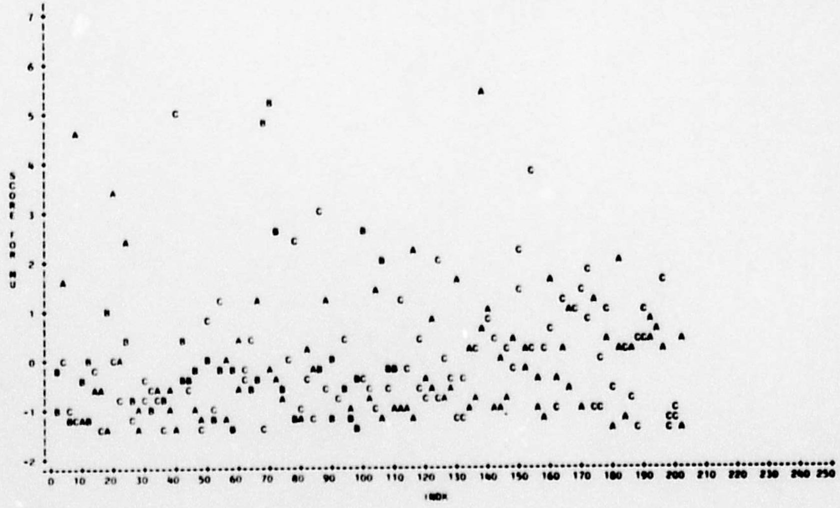
PLOT OF _FINFLU_*INDX SYMBOL IS VALUE OF REG



NOTE: 27 OBS HIDDEN

FIGURE 6.3B: PLOT OF SCORE AGAINST OBSERVATION INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

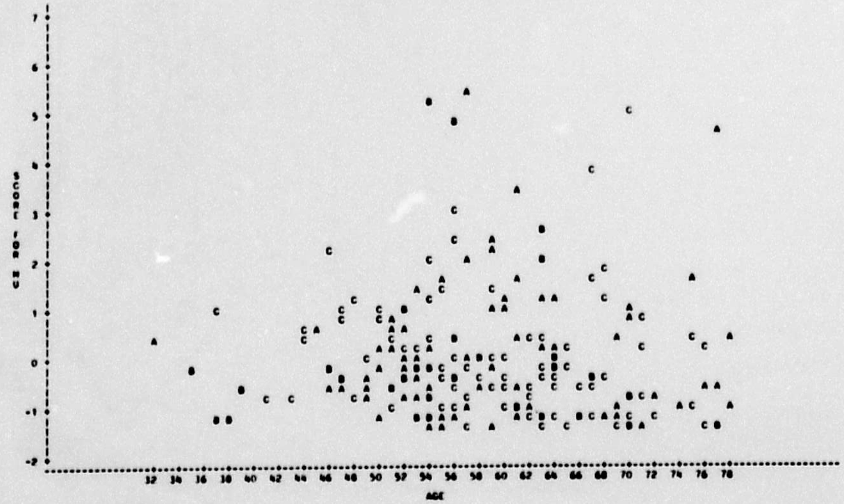
PLOT OF MU*INDX SYMBOL IS VALUE OF REG



NOTE: 4 OBS HIDDEN

FIGURE 6.3C: PLOT OF SCORE AGAINST VARIABLE AGE
PLOT SYMBOL IS THE TREATMENT (A,B,C)

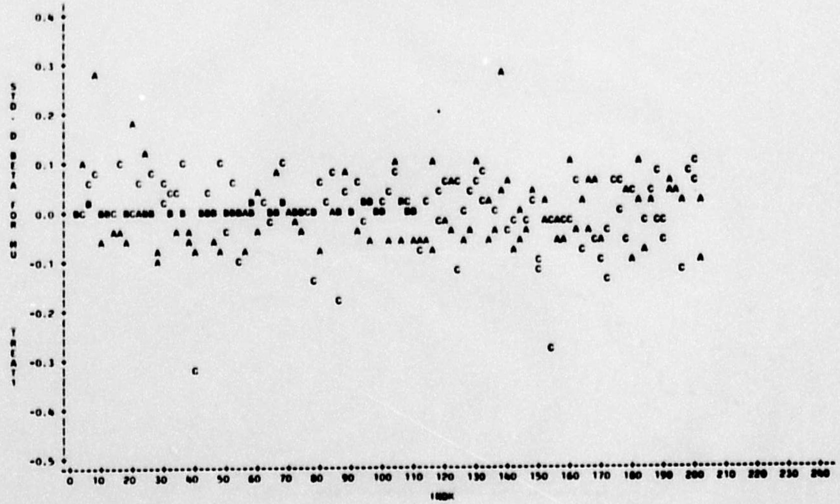
PLOT OF MU*AGE SYMBOL IS VALUE OF REG



NOTE: 34 OBS HIDDEN

FIGURE 6.3D: STANDARDIZED DBETA FOR TREAT1(A VS C) AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

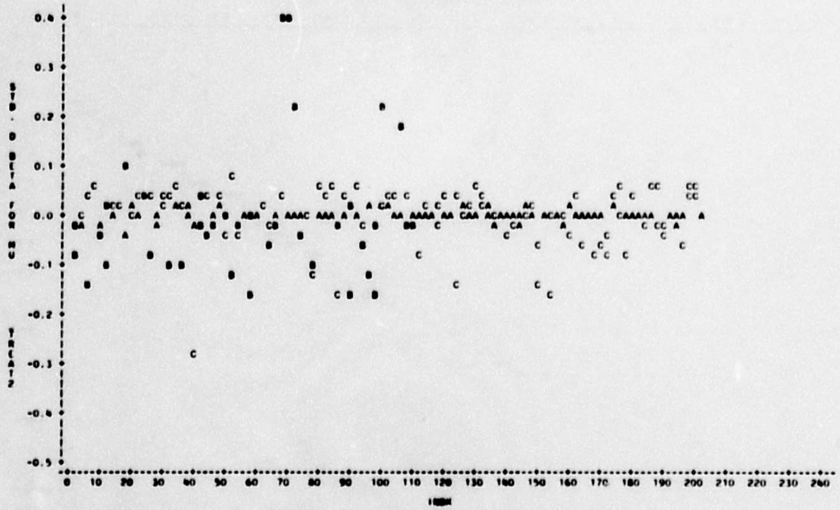
PLOT OF DBETA002*INDX SYMBOL IS VALUE OF REG



NOTE: 12 OBS HIDDEN

FIGURE 6.3E: STANDARDIZED DBETA FOR TREAT2(B VS C) AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

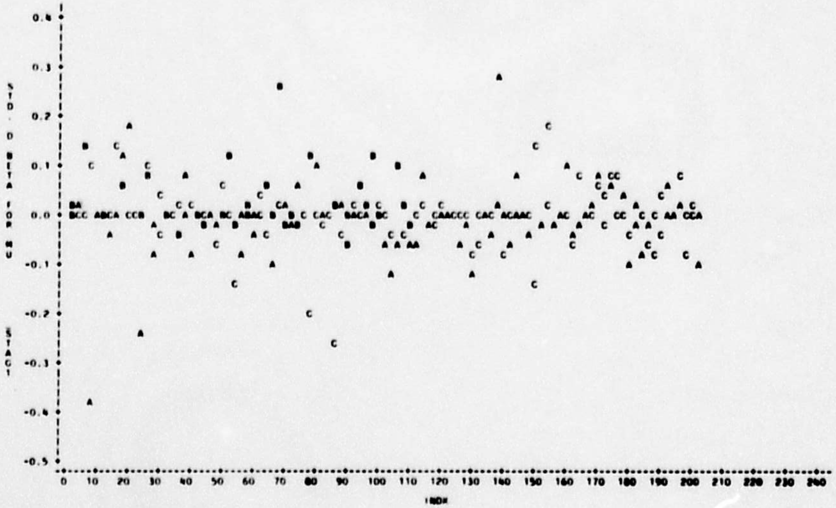
PLOT OF DBETA003*INDX   SYMBOL IS VALUE OF REG



NOTE: 16 OBS HIDDEN

FIGURE 6.3F: STANDARDIZED DBETA FOR STAG1 AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

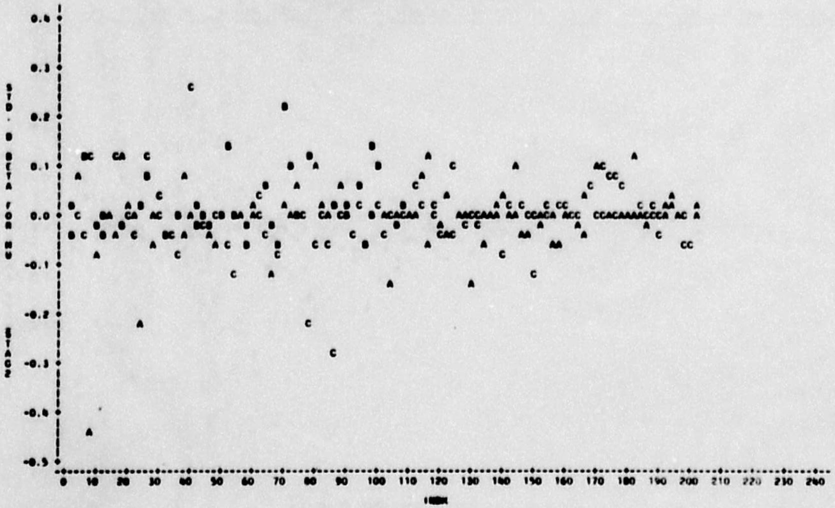
PLOT OF DBETA004*INDX   SYMBOL IS VALUE OF REG



NOTE: 19 OBS HIDDEN

FIGURE 6.3G: STANDARDIZED DBETA FOR STAG2 AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

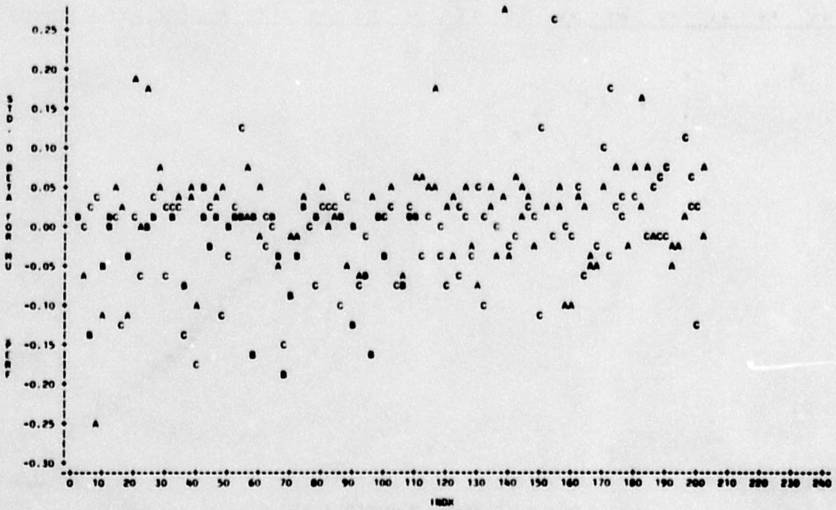
PLOT OF DBETA005*INDX   SYMBOL IS VALUE OF REG



NOTE: 16 OBS HIDDEN

FIGURE 6.3H: STANDARDIZED DBETA FOR PERF AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

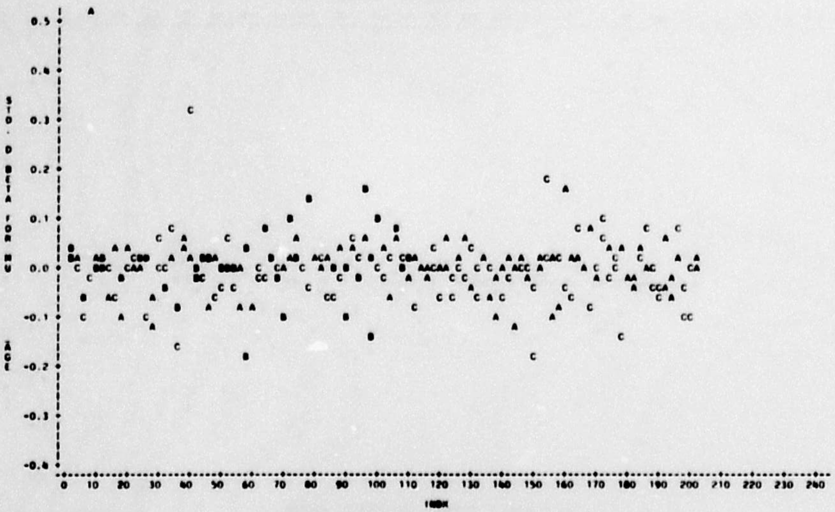
PLOT OF DBETA006*INDX   SYMBOL IS VALUE OF REG



NOTE: 7 OBS HIDDEN

FIGURE 6.3I: STANDARDIZED DBETA FOR AGE AGAINST INDEX
PLOT SYMBOL IS THE TREATMENT (A,B,C)

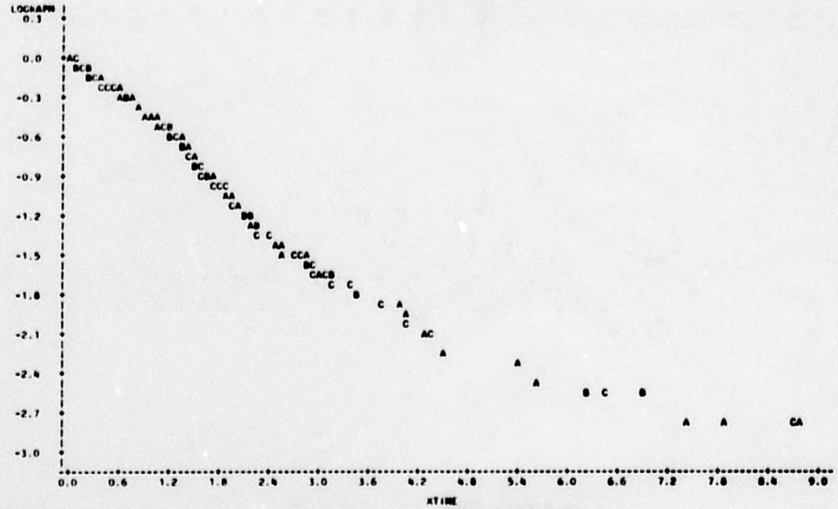
PLOT OF DBETA007*INDX   SYMBOL IS VALUE OF REG



NOTE: 8 OBS HIDDEN

FIGURE 6.3J: COX-SNELL RESIDUAL PLOT FOR HEAD/NECK CANCER DATA
PLOT SYMBOL IS THE TREATMENT (A,B,C)

PLOT OF LOGKAPM*XTIME   SYMBOL IS VALUE OF REG



NOTE: 124 OBS HIDDEN

FIGURE 6.3K: AITKEN-CLAYTON STABILIZED COX-SNELL RESIDUAL PLOT   25
PLOT SYMBOL IS THE TREATMENT (A,B,C)

PLOT OF AITHENS*AITHENT   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

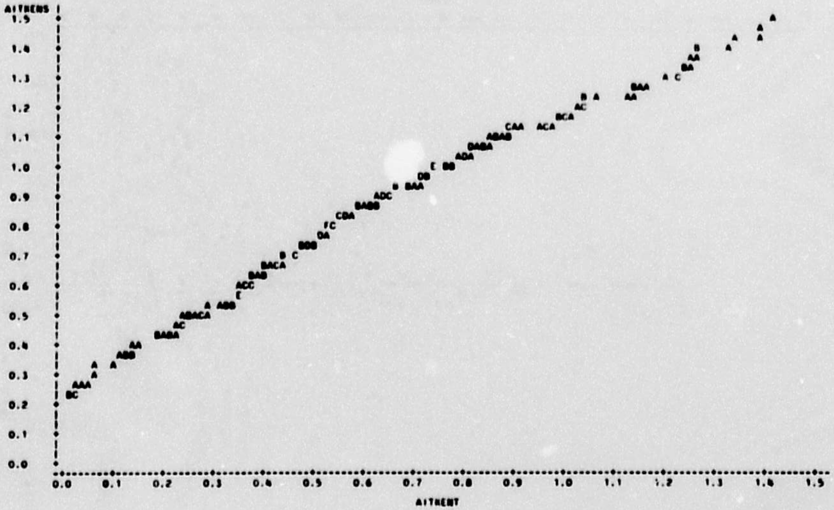
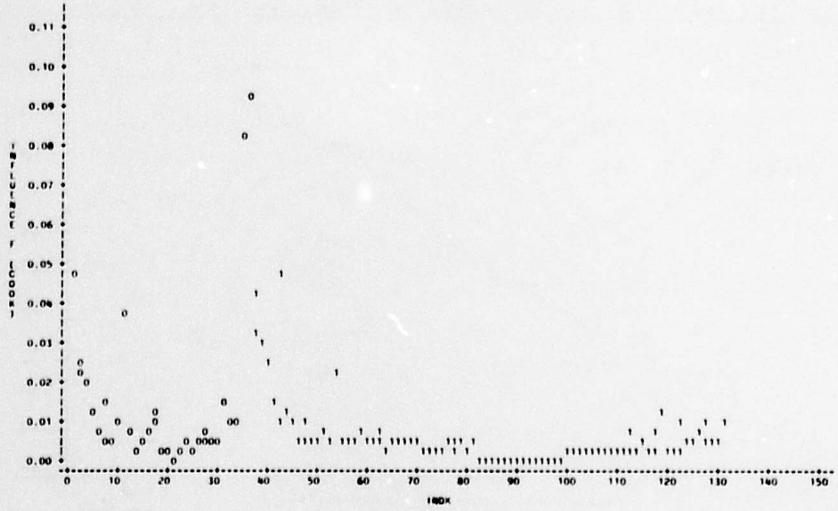


FIGURE 6.4A: PLOT OF INFLUENCE F AGAINST OBSERVATION INDEX
PLOT SYMBOL IS THE TREATMENT DUMMY VARIABLE

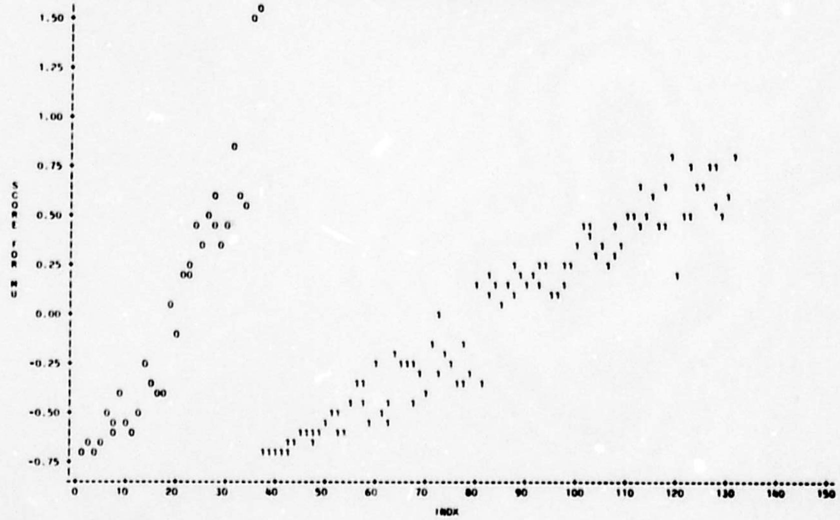
PLOT OF _FINFLU_*INDX   SYMBOL IS VALUE OF TREAT



NOTE: 13 OBS HIDDEN

FIGURE 6.4B: PLOT OF SCORE AGAINST OBSERVATION INDEX

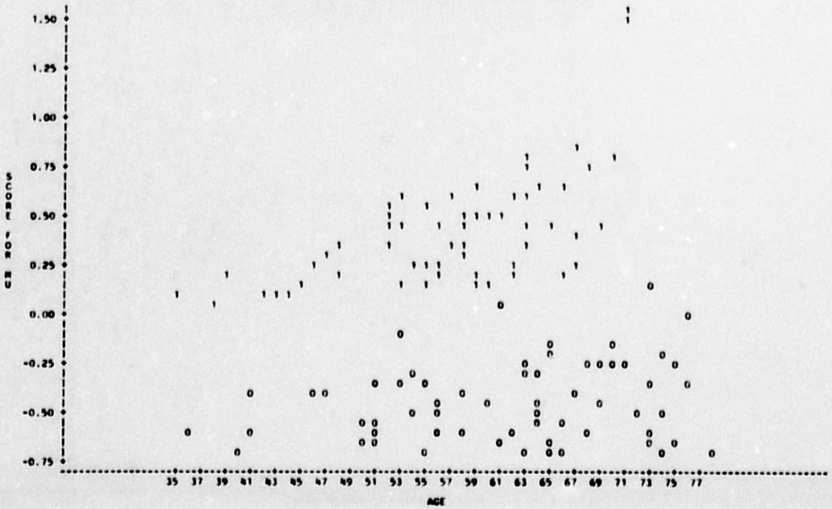
PLOT OF MU*INDX   SYMBOL IS VALUE OF TREAT



NOTE: 5 OBS HIDDEN

FIGURE 6.4C: PLOT OF SCORE AGAINST VARIABLE AGE
PLOT CHARACTER IS CENSORED DATA INDICATOR (1=CENSORED)

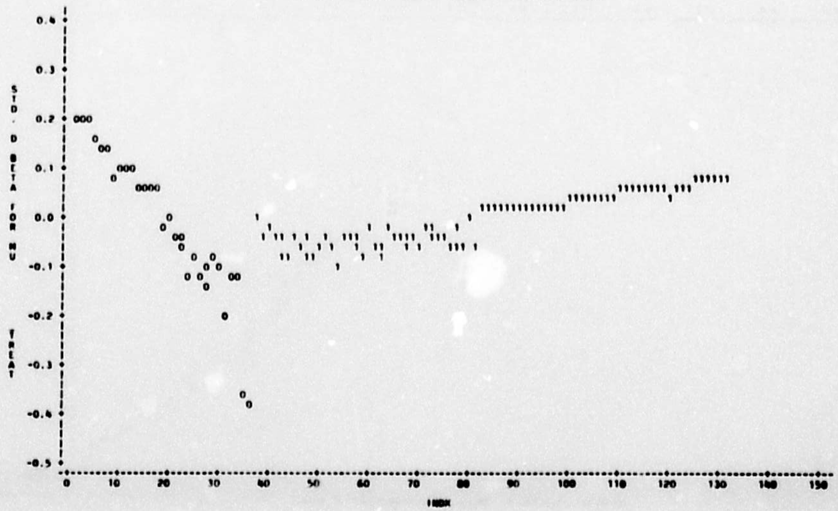
PLOT OF MU*AGE   SYMBOL IS VALUE OF CENSORED



NOTE: 12 OBS HIDDEN

FIGURE 6.4D: STANDARDIZED DBETA FOR TREAT AGAINST INDEX
PLOT CHARACTER IS THE TREATMENT DUMMY VARIABLE

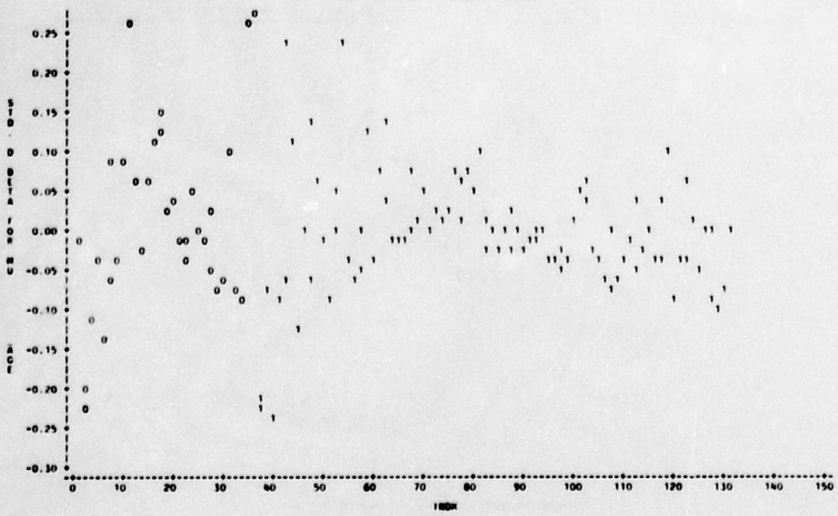
PLOT OF DBETA002*INDX   SYMBOL IS VALUE OF TREAT



NOTE: 17 OBS HIDDEN

FIGURE 6.4E: STANDARDIZED DBETA FOR AGE AGAINST INDEX
PLOT CHARACTER IS THE TREATMENT DUMMY VARIABLE

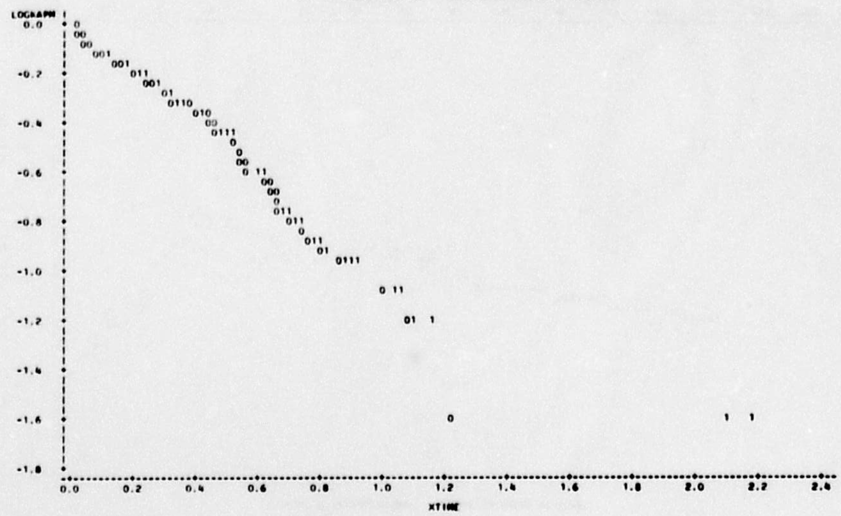
PLOT OF DBETA003*INDX   SYMBOL IS VALUE OF TREAT



NOTE: 3 OBS HIDDEN

FIGURE 6.4F: COX-SNELL RESIDUAL PLOT FOR THE GLASSER DATA
PLOT CHARACTER IS THE CENSORED DATA INDICATOR (1=CENSORED)

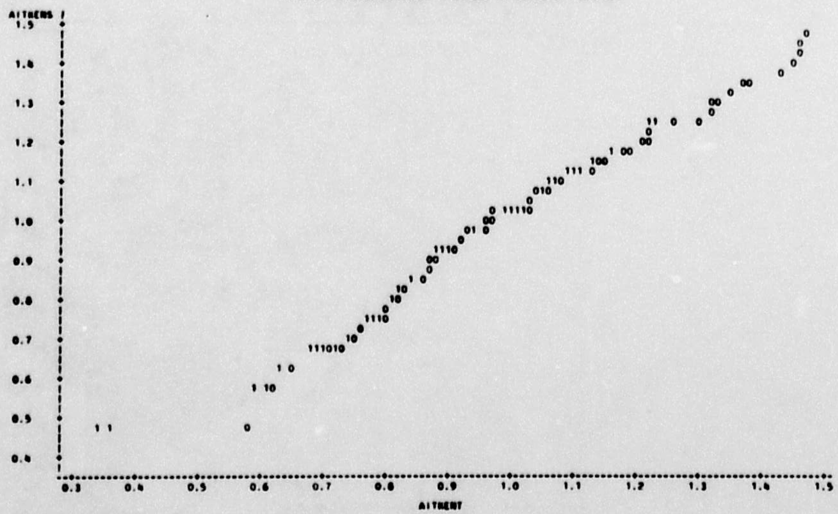
PLOT OF LOGKAPM*XTIME   SYMBOL IS VALUE OF CENSORED



NOTE: 62 OBS HIDDEN

FIGURE 6.4G: AITKEN-CLAYTON STABILIZED COX-SNELL RESIDUAL PLOT
PLOT CHARACTER IS THE CENSORED DATA INDICATOR (1=CENSORED)

PLOT OF AITHERS*AITHERT   SYMBOL IS VALUE OF CENSORED



NOTE: 49 OBS HIDDEN