TRANSFORMATION OF SURVIVAL DATA

Richard A. Johnson

University of Wisconsin

1. Introduction

It is good statistical practice to perform more than one analysis on a given data set. Normal theory methods usually provide one alternative. By employing transformations their domain of application can be greatly extended. Under normal theory, we have the advantage of simple interpretations of linear regression and interactions. Coupled with the relative ease of computation and availability of diagnostic techniques, for complete samples, normal theory has much to recommend it. Except for computational ease, the other advantages are retained in the survival analysis setting.

By transforming survival data, and then applying parametric estimation methods, we obtain an estimated survival curve which may be compared to the Kaplan-Meier (1958) estimate. In a regression setting comparisons could be made with the analysis based on the Cox (1972) proportional hazards model, or the methods of Miller (1976), Buckley and James (1979) and Koul, Susarla and Van Ryzin (1981).

2. Background and Notation

We briefly review the literature that pertains to our extensions. Box and Cox (1964) suggest the family of transformations

(1) $$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda}-1}{\lambda} , & \lambda \neq 0 \\ \ln(x) , & \lambda = 0 \end{cases}$$

for improving the approximation of positive random variables to normality. They tentatively assume that $X^{(\lambda)}$ has a normal distribution for some choice of λ. Under this assumption, $X^{(\lambda)}$ has density

$$x^{\lambda-1} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x^{(\lambda)}-\mu}{\sigma} \right)^2 \right\}$$

and a random sample leads to the log-likelihood

(2) $$\frac{-n}{2} \ln(2\pi) - \frac{1}{2} \sum_{j=1}^{n} \left(\frac{x_j^{(\lambda)} - \mu}{\sigma} \right)^2 + (\lambda-1) \sum_{j=1}^{n} \ln(x_j) \quad .$$

We remark that $X^{(\lambda)}$ cannot be exactly normal, except possibly for $\lambda = 0$, since its support has a finite lower bound.

Hernandez and Johnson (1980) show that, asymptotically, selecting λ to maximize the expression (2) is equivalent to selecting λ to minimize the Kullback-Liebler information number between $g_{\lambda}(z)$, the true pdf of $X^{(\lambda)}$, and a normal distribution $\phi_{\mu,\sigma}(z)$. That is, the information number

(3) $$I[g_{\lambda};\phi_{\mu,\sigma}] = \int g_{\lambda}(z) \ln \left[\frac{g_{\lambda}(z)}{\phi_{\mu,\sigma}(z)} \right] dz$$

is minimized by this choice of λ,μ,σ. Since $g_{\lambda}(z) = g((\lambda z+1)^{\frac{1}{\lambda}})(\lambda z+1)^{\frac{1}{\lambda}-1}$ where $g(\cdot)$ is the pdf of X, the information number can also be expressed as

(4) $$I[g_{\lambda};\phi_{\mu\sigma}] = \int g(x) \ln \left[\frac{g(x)}{\phi_{\mu,\sigma}\left(\frac{x^{\lambda}-1}{\lambda} \right) x^{\lambda-1}} \right] dx \quad .$$

Several examples appear in Hernandez and Johnson (1980).

In Section 3, we treat transformation of survival data. Because the transformation technique has proven especially effective in a regression setting, in Section 6 we extend its domain of application to survival analysis with covariates.

Carroll (1980) and Bickel and Doksum (1981) raise some question about the sampling properties of estimators determined by the Box-Cox procedure. Recent evidence, however, indicates that predictions and tests for significance of regression parameters remain valid (see Carroll and Ruppert (1981)).

3. Survival Analysis Setting

In the survival analysis setting, the times of entry into the study are haphazard or random. We assume the arrival process, for items or persons, is independent of life length. Consequently, we model the life lengths as independent identically distributed random variables with c.d.f. $G(\cdot)$. The time on test for the i^{th} person will be denoted by L_i = (current time) - (entry time). We either observe x_i = life length, or censor the test at L_i. We tentatively assume that some power transformation is normal. The likelihood is then

$$(5)\qquad L(\lambda,\mu,\sigma) = \prod_{i\varepsilon F} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{1}{2\sigma^2}\left(\frac{x_i^{\lambda}-1}{\lambda}-\mu\right)^2} \cdot x_i^{\lambda-1} \prod_{i\notin F}\left[1-\Phi\left(\frac{\frac{L_i^{\lambda}-1}{\lambda}-\mu}{\sigma}\right)\right]$$

where $F=\{i:x_i<L_i\}$ is the set of items that fail during the trial.

The log-likelihood can be maximized numerically over λ,μ and σ. Because of the censoring, there is not even a partial analytic solution as in the complete sample case.

EXAMPLE: [Stanford Heart Transplant Data]

We consider the first $n = 184$ patients reported in Miller and Halpern (1981). Numerical minimization of $-\ell n\ L(\lambda,\mu,\sigma)$ provides the estimates (we replaced the 0 lifelength by .5)

$$\hat{\lambda} = .0042 \ , \quad \hat{\mu} = 6.3706 \ , \quad \hat{\sigma} = 2.4956$$

and

$$-\ell n\ L(\hat{\lambda},\hat{\mu},\hat{\sigma}) = 859.0402 \quad .$$

The log-normal has $\lambda = 0$, $\hat{\mu}(0) = 6.2833$, $\hat{\sigma}(0) = 2.4452$ and $-\ell n\ L(0,\hat{\mu}(0),\hat{\sigma}(0)) =$ 859.044. Because of the near equivalence of the maximized likelihoods, it is just as reasonable to take $\ell n\ X_i$ as approximately normal. In fact, Miller and Halpern (1981) use $\ell n\ X_i$ without explanation. Figure 1 displays the graph of $-\ell n\ L(\lambda,\hat{\mu}(\lambda),\ \hat{\sigma}(\lambda))$ versus λ. If the usual asymptotic theory

$$-2\ \ell n\ [L(\lambda,\hat{\mu}(\lambda),\hat{\sigma}(\lambda))\ /\ L(\hat{\lambda},\hat{\mu},\hat{\sigma})] \text{ approximately } \chi_1^2$$

applies, values of λ in the interval -.09 to .10 should be considered reasonable choices.

It is also possible to estimate the survival function. Proceeding as if $X^{(\hat{\lambda})}$ has a normal distribution with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$, we consider the survival estimate

$$(6) \qquad \hat{S}(x) = \widehat{P[X > x]} = 1 - \Phi\left(\frac{\frac{x^{\hat{\lambda}}-1}{\hat{\lambda}} - \hat{\mu}}{\hat{\sigma}} \right) \quad .$$

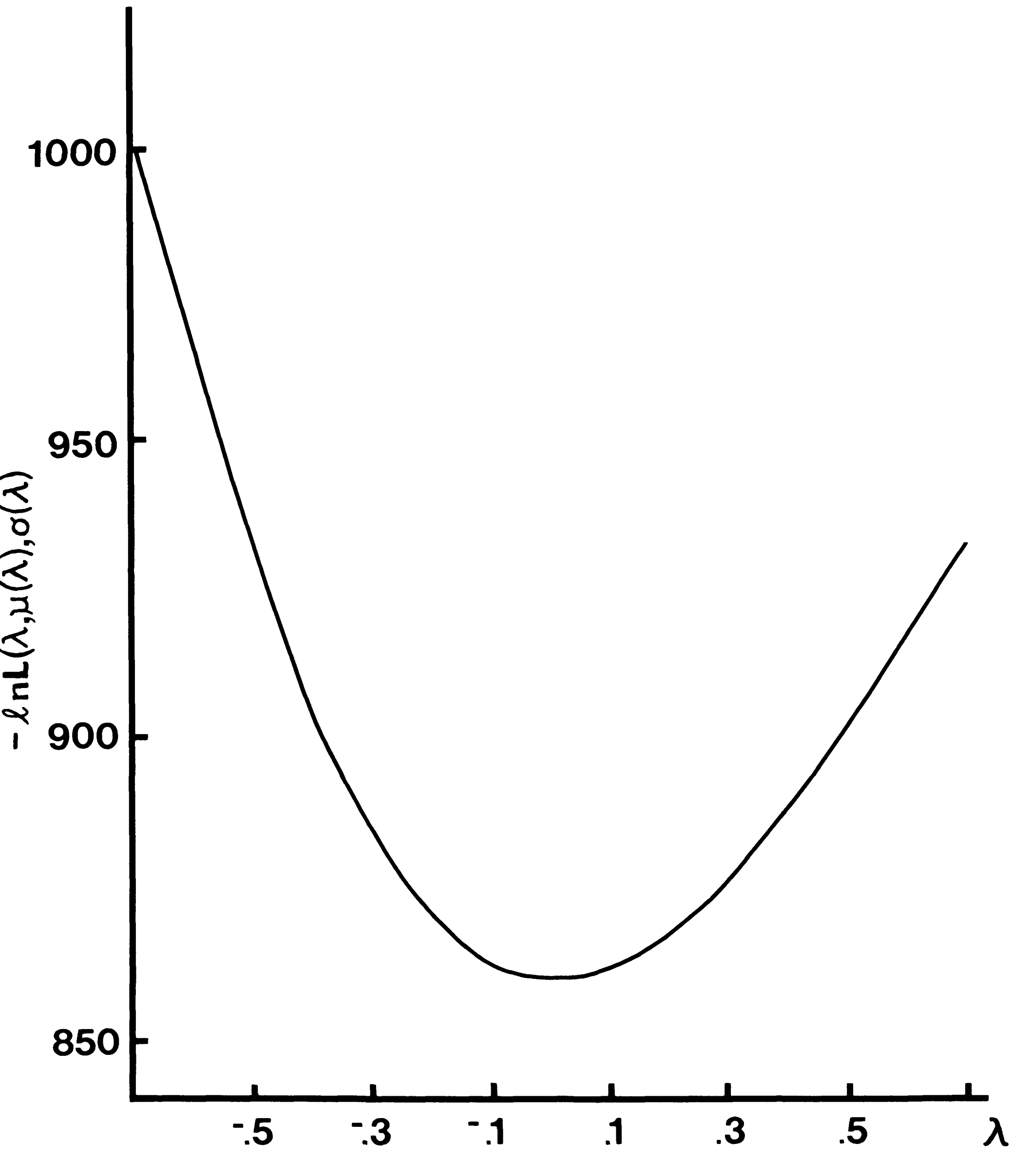

FIGURE 1: The negative of the partially maximized log-likelihood.

Figure 2 displays the estimated survival function (6), along with the Kaplan-Meier estimate, for the heart transplant data.

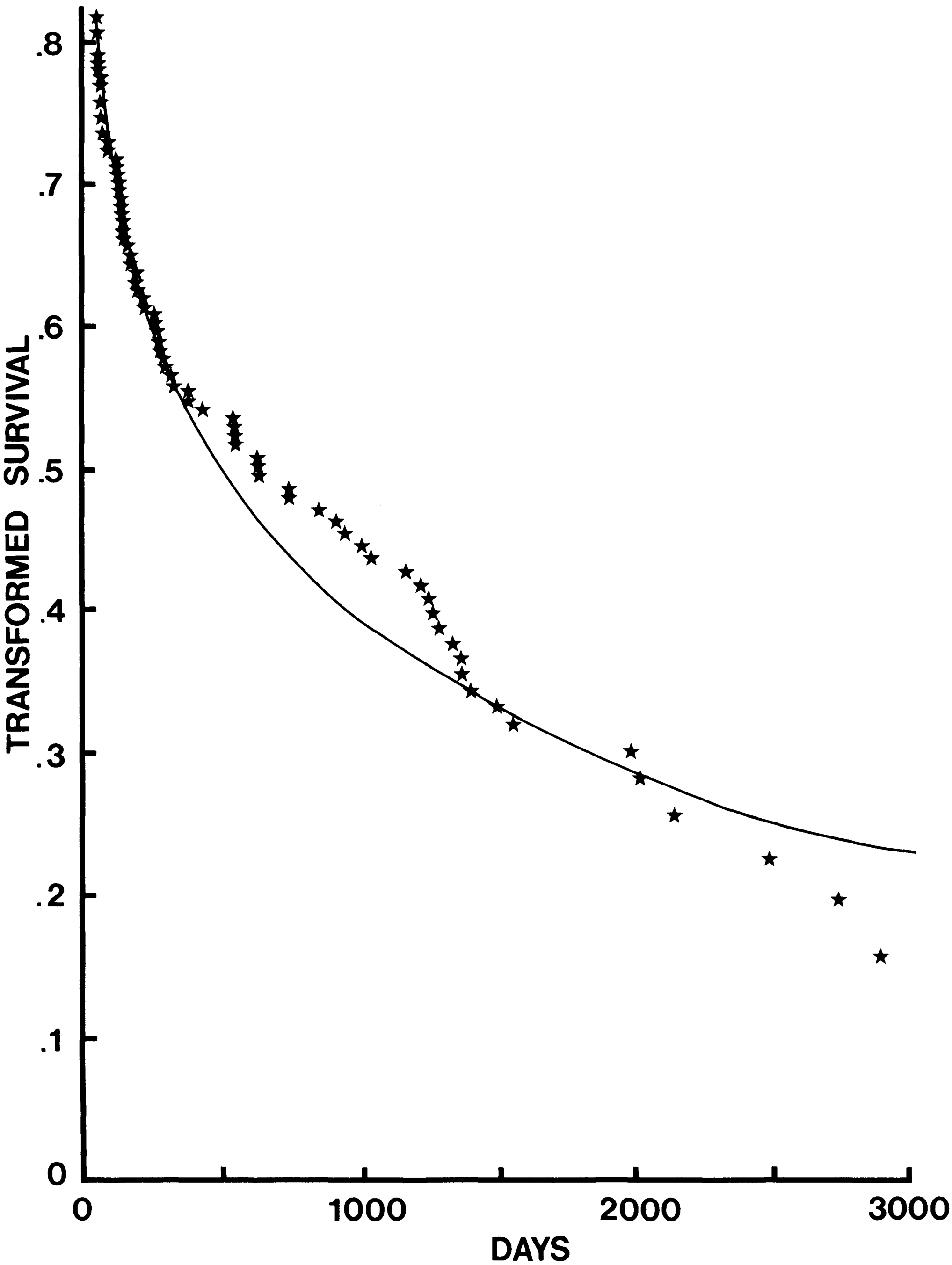

FIGURE 2: Estimated survival functions $\hat{S}$ from (6) and the Kaplan-Meier estimate.

4. Large Sample Properties

For ease of exposition, we establish our results for trials conducted over a fixed time period [0,T] and where items enter only at one of the K fixed times

$$0 = t_1 < t_2 < \cdots < t_K < T \quad .$$

Let n_i = number of items entering at time t_i and $n = \sum_{i=1}^{K} n_i$. Setting $L_i = T - t_i$ and $\eta_i = \ell im(n_i/n)$, we introduce

$$H_{\underset{\sim}{\eta}}(\lambda,\mu,\sigma) = \sum_{i=1}^{K} \eta_i \left\{ [1 - G(L_i)] \, \ell n \left[\frac{1-\Phi\left(\frac{L_i^{(\lambda)} - \mu}{\sigma}\right)}{1 - G(L_i)} \right] \right.$$

(7)

$$\left. + E_g\left(I_{[0,L_i]} \, \ell n \left[\frac{\phi\left(\frac{X_{i1}^{(\lambda)} - \mu}{\sigma}\right) X_{i1}^{\lambda-1}}{g(X_{i1})} \right] \right) \right\} ,$$

which is the negative of the Kullback-Leibler information number between $X^{(\lambda)}$ and some normal distribution obtained by weighting the K censored population numbers by the proportions η_i.

THEOREM 1:

Let $n_i/n \to \eta_i (0 < \eta_i < 1)$ and suppose the following conditions are satisfied:

(i) the parameter space Ω is the compact subset of R^3 defined by $\Omega = \{\theta = (\lambda,\mu,\sigma)' \mid |\mu| \leq M, \; c_1 \leq \sigma \leq c_2, \; a \leq \lambda \leq b$ for some $0 < M, c_1, c_2, b < \infty$ and $-\infty < a < 0\}$,

(ii) the moments $E_g(X^{2a})$ and $E_g(X^{2b})$ are finite,

(iii) $H_{\underset{\sim}{\eta}}(\lambda,\mu,\sigma)$ has a unique global maximum at $(\lambda_0,\mu_0,\sigma_0) = \underset{\sim}{\theta}_0'$.

Then, (1) $(\hat{\lambda},\hat{\mu},\hat{\sigma}) \xrightarrow{a.s.} (\lambda_0,\mu_0,\sigma_0)$ as $n \to \infty$.

Furthermore, if:

(iv) $(\lambda_0,\mu_0,\sigma_0)$ is an interior point of Ω,

(v) both $E_g[X^a \; \ell n(X)]^2$ and $E_g[X^b \; \ell n(X)]^2$ are finite,

(vi) $\nabla H_{\underset{\sim}{\eta}} (\lambda_0,\mu_0,\sigma_0) = \underset{\sim}{0}$,

(vii) $V = \{\nabla^2 H_{\underset{\sim}{\eta}} (\lambda_0,\mu_0,\sigma_0)\}^{-1}$ exists,

then, (2) $\sqrt{n}\,(\hat{\lambda}-\lambda_0,\ \hat{\mu}-\mu_0,\ \hat{\sigma}-\sigma_0)' \xrightarrow{d} N_3(\underset{\sim}{0},VWV')$ as $n \to \infty$, with the elements of $W = (w_{uv})$, $u,v = 1,2,3$, given by

$$\sum_{i=1}^{K} \eta_i \left\{ [1-G(L_i)] \left(\frac{\partial}{\partial\theta_u} \ell n \left[1-\Phi\left(\frac{L_i^{(\lambda)}-\mu}{\sigma}\right)\right]\Bigg|_{\underset{\sim}{\theta}_0} \right) \left(\frac{\partial}{\partial\theta_v} \ell n \left[1-\Phi\left(\frac{L_i^{(\lambda)}-\mu}{\sigma}\right)\right]\Bigg|_{\underset{\sim}{\theta}_0} \right) \right.$$

$$\left. + E_g \left[I_{[0,L_i]}(X) \left(\frac{\partial \, \ell n \, \phi\left(\frac{X^{(\lambda)}-\mu}{\sigma}\right) X^{\lambda-1}}{\partial\theta_u}\Bigg|_{\underset{\sim}{\theta}_0} \right) \left(\frac{\partial \, \ell n \, \phi\left(\frac{X^{(\lambda)}-\mu}{\sigma}\right) X^{\lambda-1}}{\partial\theta_v}\Bigg|_{\underset{\sim}{\theta}_0} \right) \right] \right\} .$$

PROOF: The log-likelihood, divided by the sample size $n = \sum_{i=1}^{K} n_i$, is the sum of the K terms

$$n^{-1} \ell_n = \sum_{i=1}^{K} \frac{n_i}{n} \left\{ \sum_{j=1}^{n_i} \left[-\frac{1}{2}\ell n(2\pi) - \frac{1}{2}\ell n\,\sigma - \frac{1}{2}\left(\frac{x_{ij}^{(\lambda)}-\mu}{\sigma}\right)^2 + (\lambda-1)\ell n(x_{ij}) \right] \right.$$

$$\left. I_{[0,L_i]}(x_{ij}) + m_i \; \ell n \left[1-\Phi\left(\frac{L_i^{(\lambda)}-\mu}{\sigma}\right)\right] \right\}$$

where $m_i = \sum_{j=1}^{n_i} I_{(L_i,\infty]}(x_{ij})$. The term $\left(\frac{x^{(\lambda)} - \mu}{\sigma}\right)^2 I_{[0,L_i]}(x)$

(i) is dominated by a g-integrable function uniformly in $\underset{\sim}{\theta} = (\lambda,\mu,\sigma)' \in \Omega$ and (ii) is equicontinuous in $\underset{\sim}{\theta}$ for fixed x, on the set $S_m = [0, L_i - \frac{1}{m}] \cup [L_i + \frac{1}{m}, m]$. The uniform strong law (see Rubin (1956)) applies.

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \left(\frac{X_{ij}^{(\lambda)} - \mu}{\sigma}\right)^2 I_{[0,L_i]}(X_{ij}) \xrightarrow{\text{a.s.}} E_g\left[\left(\frac{X_{i1}^{(\lambda)} - \mu}{\sigma}\right)^2 I_{[0,L_i]}(X_{i1})\right]$$

for $i = 1,2,\ldots,K$ uniformly in (λ,μ,σ). The strong law of large numbers applied to $\sum_{j=1}^{n_i} \ell n(X_{ij})/n_i$ and $\sum_{j=1}^{n_i} I_{(L_i,\infty]}(X_{ij})/n_i$, for $i = 1,2,\ldots,K$, establishes the almost sure uniform convergence

$$\frac{1}{n} \ell_n \to H_{\underset{\sim}{\eta}}(\lambda,\mu,\sigma) + \sum_{i=1}^{K} \eta_i\{E_g[I_{[0,L_i]} \ell n g(X_{1i})] + [1 - G(L_i)] \ell n[1 - G(L_i)]\} .$$

Moreover, $H_{\underset{\sim}{\eta}}(\lambda,\mu,\sigma)$ is continuous and, by assumption, has a unique maximum at $(\lambda_0,\mu_0,\sigma_0)$. Consequently, $(\hat{\lambda},\hat{\mu},\hat{\sigma}) \to (\lambda_0,\mu_0,\sigma_0)$ almost surely.

The asymptotic normality follows upon expanding the first partial derivatives of $n^{-\frac{1}{2}}\ell_n$ in a Taylor series about $(\lambda_0,\mu_0,\sigma_0)$. Since $(\hat{\lambda},\hat{\mu},\hat{\sigma}) \to (\lambda_0, \mu_0,\sigma_0)$, which is interior to Ω, $\nabla\ell_n(\hat{\lambda},\hat{\mu},\hat{\sigma}) = \underset{\sim}{0}$ for all sufficiently large n. Similar to the treatment of the single sample problem in Guerrero (1979) and Guerrero and Johnson (1979), we can dominate the individual terms in $n^{-1}\nabla^2\ell_n(\lambda,\mu,\sigma)$ to obtain uniform convergence to its expected value $\nabla^2 H_{\underset{\sim}{\eta}}$. In particular, $n^{-1}\nabla^2\ell_n(\lambda_*,\mu_*,\sigma_*)$ converges a.s. to $\nabla^2 H_{\underset{\sim}{\eta}}(\lambda_0,\mu_0,\sigma_0)$ where $(\lambda_*,\mu_*,\sigma_*)$ is any sequence of intermediate values between $(\hat{\lambda},\hat{\mu},\hat{\sigma})$ and $(\lambda_0,\mu_0,\sigma_0)$. Since

$$n^{-\frac{1}{2}} \nabla\ell_n(\lambda_0,\mu_0,\sigma_0) \xrightarrow{\mathcal{L}} N(\underset{\sim}{0},W)$$

the normal convergence for $(\hat{\lambda},\hat{\mu},\hat{\sigma})'$ follows.

5. Checking the Adequacy of the Transformation

The power transformation was selected by maximizing the likelihood (5) obtained under the tentative assumption that some transformation (1) is normal. Although Theorem 1 gives one set of conditions that insure that, asymptotically, the power transformation closest to a normal is selected, this choice may not be good enough. Therefore, it is necessary to check that the transformation

$$x^{(\hat{\lambda})} = \frac{x^{\hat{\lambda}} - 1}{\hat{\lambda}}$$

has achieved near-normality.

In order to obtain diagnostic plots, a censored observation $x_i = L_i$ is assigned the value

$$\hat{x}_i^{(\hat{\lambda})} = E[X^{(\hat{\lambda})} | X > L_i] = \hat{\mu} + \hat{\sigma}\, h\left(\frac{L_i^{(\hat{\lambda})} - \hat{\mu}}{\hat{\sigma}}\right) , \tag{8}$$

where $h(\cdot) = \phi(\cdot) / [1 - \Phi(\cdot)]$ is the hazard rate for the standard normal (see Schmee and Hahn (1979)). That is, the expectation is computed as if $X^{(\hat{\lambda})}$ is normal with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$. Using these estimates, we have

$$x_i^{(\hat{\lambda})} \qquad , \text{ if failure}$$

$$\hat{x}_i^{(\hat{\lambda})} = \hat{\mu} + \hat{\sigma}\, h\left(\frac{L_i^{(\hat{\lambda})} - \hat{\mu}}{\hat{\sigma}}\right) \qquad , \text{ if censored.}$$

These can be ordered and displayed in a normal Q-Q plot.

Figure 3 shows a plot of the transformed heart transplant survival times versus the approximate normal scores $\Phi^{-1}(i/(n+1))$. The predicted values, plotted as open squares, do not seem to conform to the straight line pattern. At this state of development, it is not clear that (8) provides the proper estimates for diagnostic plots. Figure 4 shows the transformed failure times by themselves, plotted against the same scores as in Figure 3. It is these uncensored observations on which the adequacy of the normal approximation should be judged.

In the complete sample situation, goodness-of-fit can be tested using the correlation coefficient calculated from the normal Q-Q plot. Verrill (1981) has recently determined the large sample distribution of the correlation coefficient calculated from data that are right-censored. His results apply to either fixed order statistic censoring or fixed time censoring but not to the staggered entry situation graphed in Figure 3.

6. Survival Analysis Setting with Covariates

When r predictors $\underset{\sim}{z}' = (z_1,\ldots,z_r)$ are available, the tentative assumption becomes

$$X^{(\lambda)} \text{ is distributed } N(\alpha + \underset{\sim}{\beta}'\underset{\sim}{z}, \sigma^2)$$

for some choice of λ. Under this assumption, the likelihood becomes

$$L(\lambda,\alpha,\underset{\sim}{\beta},\sigma) = \prod_{i\varepsilon F} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{1}{2\sigma^2}(x_i^{(\lambda)} - \alpha - \underset{\sim}{\beta}'\underset{\sim}{z}_i)^2} x_i^{\lambda-1}$$

$$\cdot \prod_{i \notin F} \left[1 - \Phi\left(\frac{\frac{L_i^{\lambda}-1}{\lambda} - \alpha - \underset{\sim}{\beta}'\underset{\sim}{z}_i}{\sigma} \right) \right] \quad . \tag{9}$$

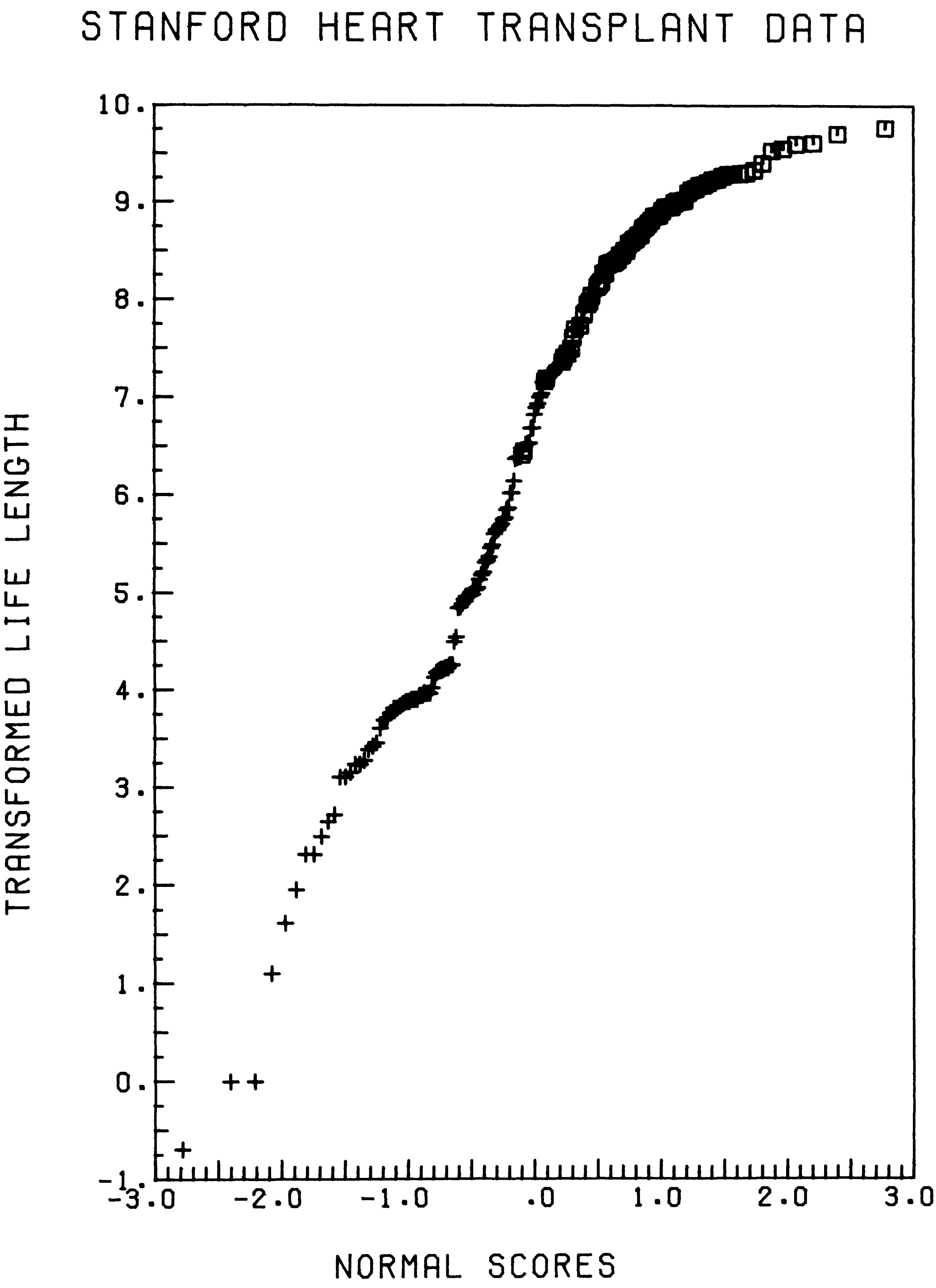

FIGURE 3: Normal scores plot of heart transplant data.

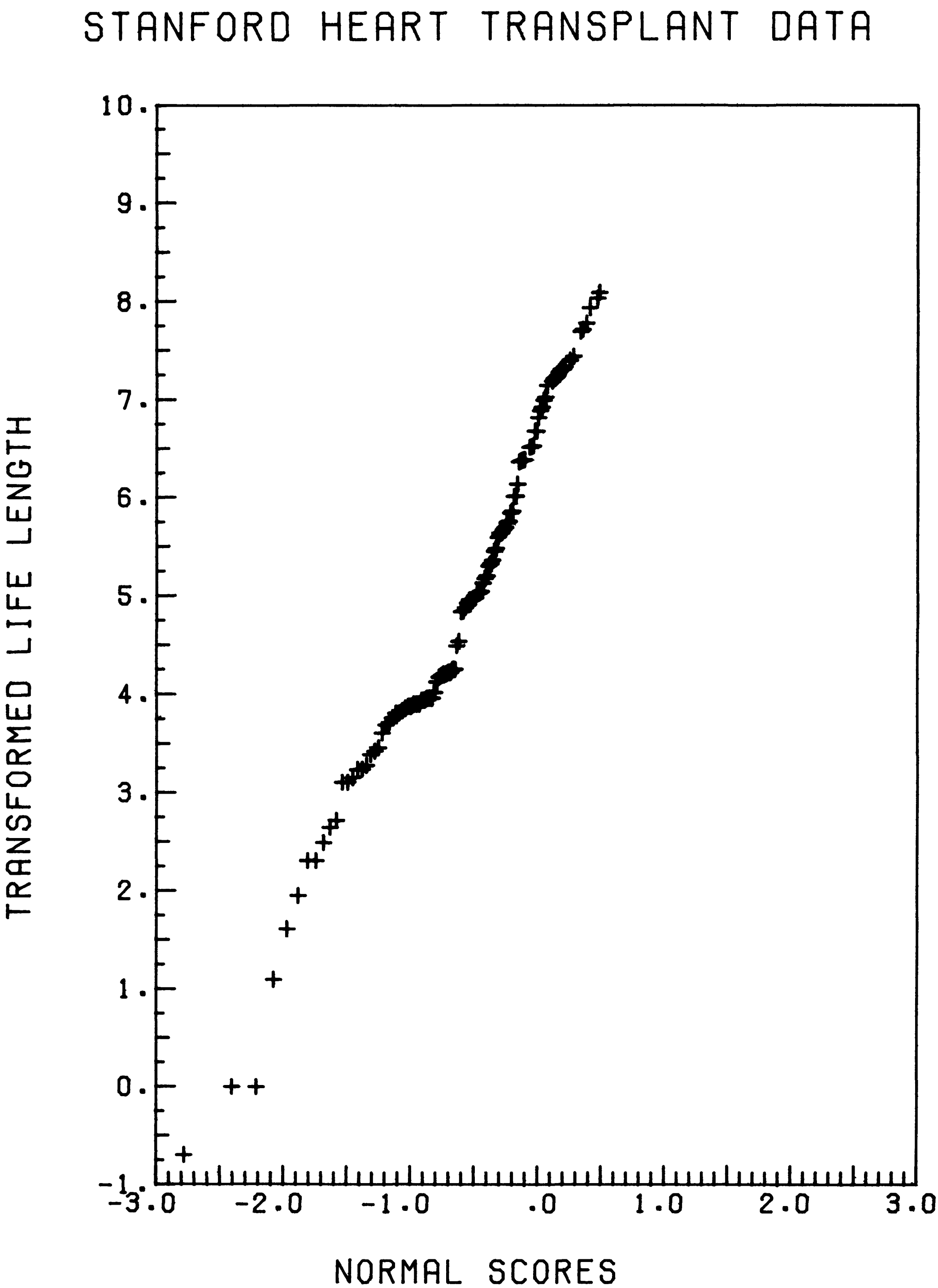

FIGURE 4: Portion of normal scores plot from death times.

Our parametric approach is to maximize (9). Miller (1980) contains a discussion of several alternative methods for formulating the regression model.

EXAMPLE:

We return to the Stanford heart-transplant data and use $z = \text{age}$ as a predictor variable. A computer calculation provides the estimates

$$\hat{\lambda} = .0090 \ , \quad \hat{\alpha} = 7.9339 \ , \quad \hat{\beta} = -.0349 \ , \quad \hat{\sigma} = 2.5490$$

and

$$-\ell n \ L(\hat{\lambda},\hat{\alpha},\hat{\beta},\hat{\sigma}) = 857.3343 \quad .$$

The maximized likelihood for $\lambda = 0$, the log-transformation, is nearly the same. Note also that $-2 \ \ell n[L(\hat{\lambda},\hat{\mu},\hat{\sigma}) \, / \, L(\hat{\lambda},\hat{\alpha},\hat{\beta},\hat{\sigma})]$ is less than χ^2_1 (.05), suggesting that age is not a good predictor.

Regression diagnostics need to be developed to check both the normal assumption for $X^{(\lambda)}$ and the regression equation. To obtain plots, we replace each censored value by its conditional expected value

$$\hat{x}^{(\hat{\lambda})} = \hat{\alpha} + \hat{\underset{\sim}{\beta}}'\underset{\sim}{z} + \hat{\sigma}\, h\left(\frac{x^{(\hat{\lambda})} - \hat{\alpha} - \hat{\underset{\sim}{\beta}}'\underset{\sim}{z}}{\hat{\sigma}} \right) \quad . \tag{10}$$

The residuals, divided by $\hat{\sigma}$, are then

$$\hat{\varepsilon} = \begin{cases} \dfrac{x^{(\hat{\lambda})} - \hat{\alpha} - \hat{\underset{\sim}{\beta}}'\underset{\sim}{z}}{\hat{\sigma}} \ , & \text{if failure} \\[2ex] h\left(\dfrac{x^{(\hat{\lambda})} - \hat{\alpha} - \hat{\underset{\sim}{\beta}}'\underset{\sim}{z}}{\hat{\sigma}} \right) \ , & \text{if censored} \end{cases} \quad . \tag{11}$$

A normal Q-Q plot of $\hat{\varepsilon}$ for the heart transplant data looks almost identical to Figure 3. A plot of $\hat{\varepsilon}$ versus $x^{(\hat{\lambda})}$ (or $\hat{x}^{(\hat{\lambda})}$) is shown as Figure 5 where squares represent the residuals from censored observations. A plot of $\hat{\varepsilon}$ versus patient number is given as Figure 6. Note how the predicted residuals from the later cases in the study form a bounding curve.

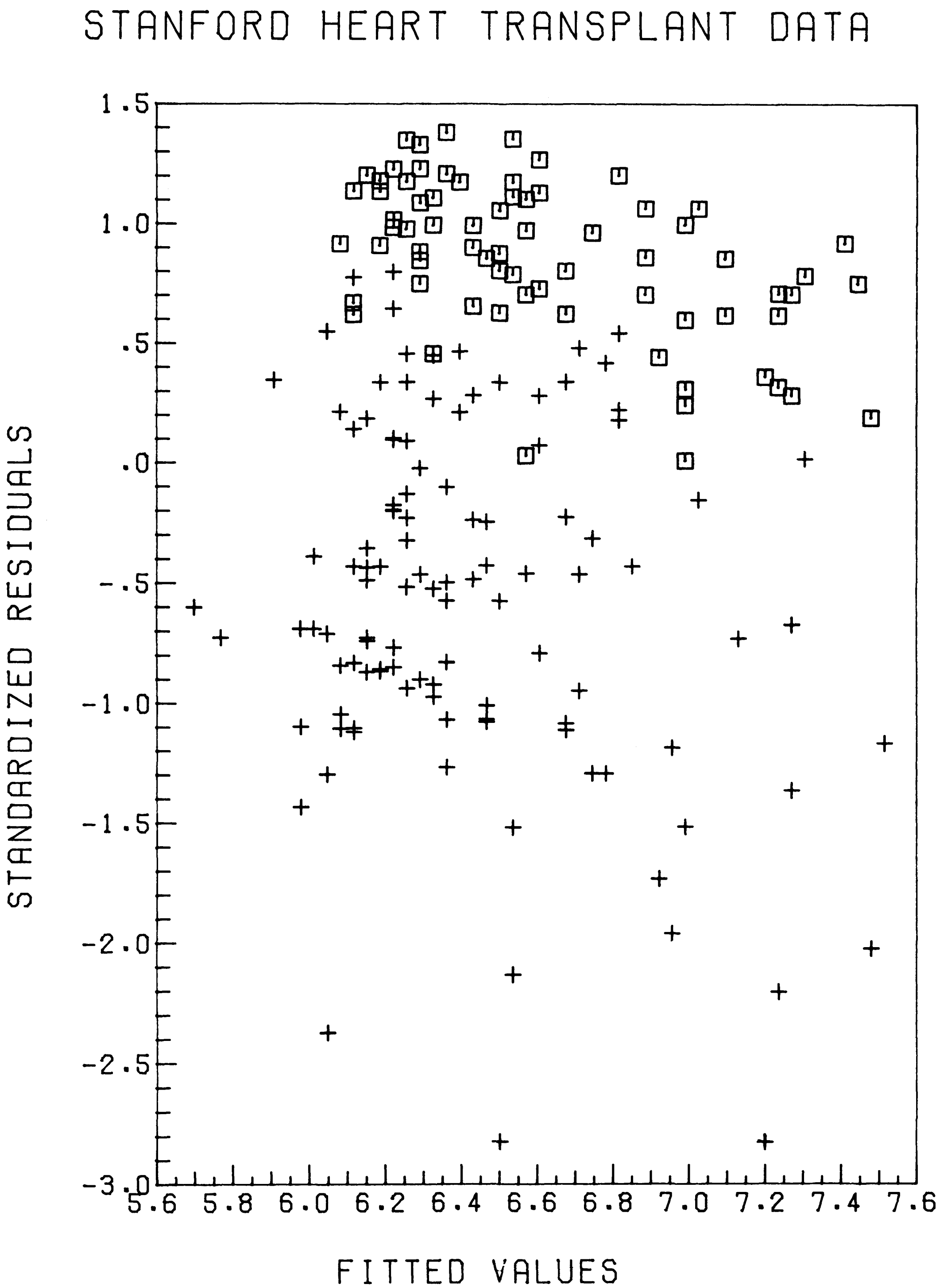

FIGURE 5: Standardized residuals versus fitted values when age is a covariate. ⊡ Censored value.

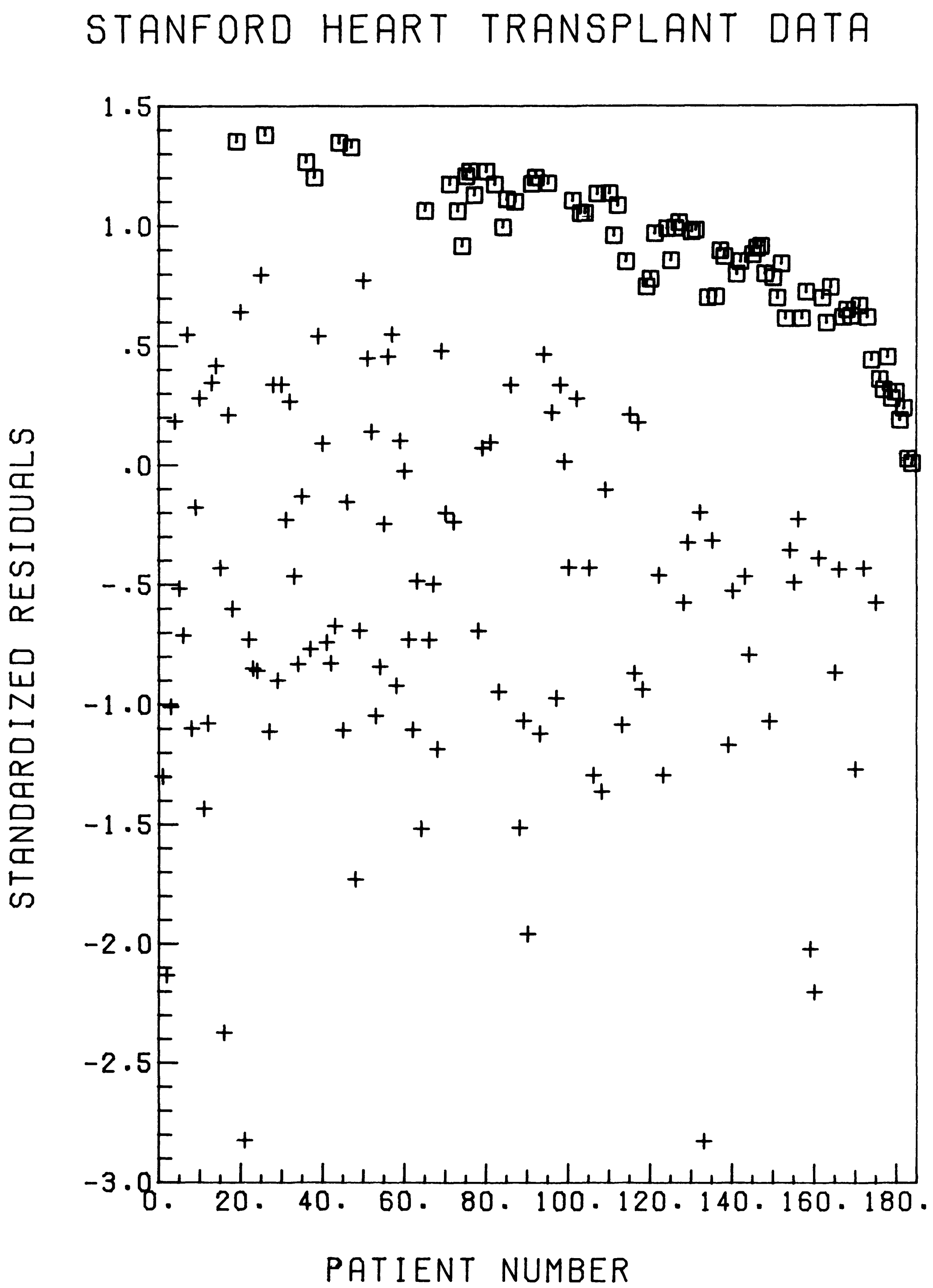

FIGURE 6: Standardized residuals versus patient number.
⊡ Censored value.

ACKNOWLEDGEMENT

This research was sponsored by the Office of Naval Research under Grant No. N00014-78-C-0722.

REFERENCES

Bickel, P.J. and Doksum, K. (1981). An analysis of transformations revisited. Journal of the American Statistical Association, 76, 296-311.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. Journal of the Royal Statistical Society B, 26, 211-243.

Buckley, J. and James, I. (1979). Linear regression with censored data. Biometrika 66, 429-436.

Carroll, R.J. (1980). A robust method for testing transformations to achieve approximate normality. Journal of the Royal Statistical Society B, 42, 71-78.

Carroll, R.J. and Ruppert, D. (1981). On prediction and the power transformation family. Biometrika 68, 609-615.

Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society B, 34, 187-202.

Guerrero, V. (1979). Extensions of the Box-Cox transformation to grouped-data situations. Ph.D. Thesis, Department of Statistics, University of Wisconsin.

Guerrero, V. and Johnson, R.A. (1979). Transformation of grouped or censored data to near normality. Technical Report 542, Department of Statistics, University of Wisconsin-Madison.

Hernandez, F. and Johnson, R.A. (1980). The large-sample behavior of transformations to normality. Journal of the American Statistical Association, 75, 855-861.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. Annals of Statistics 9, 1276-1288.

Miller, R.G. (1981). Survival analysis. John Wiley, New York.

Miller, R.G. (1976). Least squares regression with censored data. Biometrika **63**, 449-464.

Miller, R.G. and Halpern, J. (1981). Regression with censored data. Technical Report No. 66, Division of Biostatistics, Stanford University.

Rubin, H. (1956). Uniform convergence of random functions with applications to statistics. Annals of Mathematical Statistics **27**, 200-203.

Schmee, J. and Hahn, G.J. (1979). A simple method for regression analysis with censored data. Technometrics **21**, 417-432.