



Published in final edited form as:

Commun Stat Theory Methods. 1997 January 1; 26(3): . doi:10.1080/03610927708831932.

THE DISTRIBUTION OF COOK'S D STATISTIC

Keith E. Muller and

Dept. of Biostatistics, CB#7400 University of North Carolina Chapel Hill, North Carolina, 27599

Mario Chen Mok

Dept. of Biostatistics, CB#7400 University of North Carolina Chapel Hill, North Carolina, 27599

Abstract

Cook (1977) proposed a diagnostic to quantify the impact of deleting an observation on the estimated regression coefficients of a General Linear Univariate Model (GLUM). Simulations of models with Gaussian response and predictors demonstrate that his suggestion of comparing the diagnostic to the median of the F for overall regression captures an erratically varying proportion of the values.

We describe the exact distribution of Cook's statistic for a GLUM with Gaussian predictors and response. We also present computational forms, simple approximations, and asymptotic results. A simulation supports the accuracy of the results. The methods allow accurate evaluation of a single value or the maximum value from a regression analysis. The approximations work well for a single value, but less well for the maximum. In contrast, the cut-point suggested by Cook provides widely varying tail probabilities. As with all diagnostics, the data analyst must use scientific judgment in deciding how to treat highlighted observations.

Keywords

regression diagnostics; influence; residual analysis

1. INTRODUCTION

1.1 Motivation

A wide variety of applications in the medical, social, and physical sciences use regression models with continuous predictors. Often the predictors may plausibly be assumed to follow a multivariate Gaussian distribution. For example, a paleontologist may wish to model total skeleton length of fossils of a particular species, as a function of sizes for a limited number of bones. Many diagnostics have been suggested to aid in evaluating the validity of such models.

Most research in regression diagnostics has centered on the impact of deleting a single observation, with many different measures suggested. Cook (1977) recommended evaluating the standardized shift in the vector of estimated regression coefficients. He suggested comparing the statistic to the median of the F statistic for the test of all coefficients equal to zero. Such highlighted observations merit further examination in terms of their credibility and also their implications for validity of the model assumptions.

Belsley, Kuh, and Welsch (1980, p28) and Cook and Weisberg (1982, p114) discussed two alternatives for judging diagnostic statistics. *Internal scaling* involves judging a value with respect to the distribution in the sample at hand. *External scaling* involves judging a value with respect to the distribution that might occur over repeated samples. Both principles have merit in data analysis.

A standard approach for a diagnostic with known sampling distribution, such as studentized residuals, involves three steps. First, highlight observations by reference to the sampling distribution. Second, investigate the highlighted observations' values and role in the analysis. Third, decide on the disposition of the observation, in light of all knowledge about the data. Possible actions include doing nothing, correcting a discovered error, or deleting an impossible value.

Data analysts first encountering p -values for regression diagnostics may hope to use them for automatic elimination of observations. Sophisticated analysts use the reference distributions to provide a common metric for the three-step process (highlight, investigate, decide). Kleinbaum, Kupper, and Muller (1988, p201), in their introductory regression book, summarized their discussion of diagnostics by stating: "One should be cautioned that deleting the most deviant observations will in all cases slightly improve, and sometimes substantially improve, the fit of the model. One must be careful not to data snoop simply in order to polish the fit of the model by discarding troublesome data points."

Although conceptually attractive to some observers, Cook's statistic has not elicited universal enthusiasm. For example, Obenchain (1977) suggested ignoring the statistic and concentrating on its two components, the residual and the leverage. The difficulty in using the statistic stems from uncertainty as to what cut-point to use for highlighting troublesome observations. Our experience led us to the belief that the statistic flags only values already highlighted by residual analysis. Unpublished simulations (Chen Mok, 1993) confirmed the impression.

The ability to compute quantiles for Cook's statistic based on Gaussian predictors, described in §2, provides an accurate metric for the statistic and hence allows the diagnostic to consistently highlight values worthy of further examination. The new results in this paper also imply a framework and approach for describing distributions and other properties of other diagnostics.

1.2 Related Earlier Work

Nearly all current regression texts consider regression diagnostics in some detail. Excellent book-length treatments include, in chronological order, Belsley, Kuh and Welsch (1980), Cook and Weisberg (1982), Atkinson (1985), and Chatterjee and Hadi (1986).

We consider two versions of the General Linear Univariate Model (GLUM) with iid Gaussian errors. For each observational unit the predictors will be assumed to be either a set of fixed values or to follow a multivariate Gaussian distribution. Sampson (1974) described the setting with fixed predictors as the conditional model, and the setting with Gaussian predictors as the unconditional model. As detailed in §2, the distribution and interpretation of Cook's statistic depend directly on the distribution of the predictors. See Jensen and Ramirez (1996, 1997) for the distribution of Cook's statistic for fixed predictors.

2. DISTRIBUTION THEORY

2.1 Notation and Definitions

In this section we present many standard results for regression diagnostics. Rather than cite a single source for each result, we recommend that the reader consult any of the book-length treatments just cited. LaMotte (1994) provided a "Rosetta Stone" for translating among the many names used for residuals.

A number of standard distributions must be considered. In general, indicate the cumulative distribution function (CDF) of the random variable U , which depends on parameters α_1

through a_k , as $F_U(t; a_1 \dots a_k)$, with density $f_U(t; a_1 \dots a_k)$ and p th quantile $F_U^{-1}(p; \alpha_1 \dots \alpha_k)$. For notational convenience write the CDF of $U|V = v$ as $F_{U|V}(t; a_1 \dots a_k)$. Resolution of conflict between random variable and matrix notation and the random or fixed nature of a variable will be specified when not obvious from context. Let $N(\mu, \sigma)$ indicate a multivariate Gaussian vector, with mean μ , non-singular covariance Σ , and CDF $\Phi(t; \mu, \Sigma)$. Most results in this paper involve χ^2 , F , or β random variables (Johnson and Kotz, Chapter 17, 1970a; Chapters 24 and 26, 1970b). Let $\chi^2(\nu)$ indicate a central χ^2 random variable on ν degrees of freedom, and let $F(\nu_1, \nu_2)$ indicate a central F random variable on ν_1 and ν_2 degrees of freedom. Similarly let $\beta(\kappa_1, \cdot)$ indicate a β random variable, with support $(0, 1)$.

Most results for regression diagnostics concern fixed predictors, and hence the conditional model described by Sampson (1974). In particular, consider

$$y = X\beta + e. \quad (2.1)$$

$N \times 1$ $(N \times q)(q \times 1)$ $N \times 1$

Let y_i indicate the i th row of y , X_i the i th row of X , and e_i the i th row of e . Here X contains fixed values, known conditionally on having designated the sampling units, β contains fixed unknown values, and $F_{e|X}(t) = \Phi(t; \mu, \Sigma)$. Assume throughout that $N \gg q$ and that X has full rank of q . Let $\nu = (N - q)$ indicate the error degrees of freedom. Indicate the usual estimators as

$$\hat{\beta} = (X'X)^{-1}X'y, \quad (2.2)$$

$$\hat{\sigma}^2 = y'(I - H)y/\nu. \quad (2.3)$$

Define

$$H = X(X'X)^{-1}X', \quad (2.4)$$

the *hat matrix* because $\hat{y} = Hy$ (Hoaglin and Welsch, 1978). Let h_i indicate the i th diagonal element of H , the *leverage* for the i th observation:

$$h_i = X_i(X'X)^{-1}X_i'. \quad (2.5)$$

Refer to

$$\hat{e} = (y - \hat{y}) \quad (2.6)$$

as the vector of residuals. Note that

$$F_{\hat{e}|X}(t) = \Phi \left[t; 0, \sigma^2 (I - H) \right]. \quad (2.7)$$

In turn define the i th squared *standardized residual* as

$$R_i^2 = \frac{\hat{e}_i^2}{\hat{\sigma}^2 (1 - h_i)}. \quad (2.8)$$

Belsley, Kuh and Welsch (1980), Cook and Weisberg (1982) and Atkinson (1985) reviewed the algebra of deletion and properties of residuals. Let $(-i)$ indicate deletion of the i th observation and index the N statistics generated by doing so. Let $X_{(-i)}$ indicate the $(N - 1) \times q$ matrix created by deleting the i th row, with corresponding leverage

$h_{(-i)} = X_i \left(X'_{(-i)} X_{(-i)} \right)^{-1} X'_i$. The process creates sets of N estimates of β , $\{\hat{\beta}_{(-i)}\}$, predicted values, $\{\hat{y}_{(-i)} = X_i \hat{\beta}_{(-i)}\}$, residuals, $\{\hat{e}_{(-i)} = y_i - \hat{y}_{(-i)}\}$, and variance estimates, $\{\hat{\sigma}^2_{(-i)}\}$. The resulting squared and standardized residual, the *studentized residual*, equals

$$\begin{aligned} R^2_{(-i)} &= \frac{\hat{e}^2_{(-i)}}{\hat{\sigma}^2_{(-i)}(1-h_{(-i)})} \\ &= \frac{\hat{e}^2_i}{\hat{\sigma}^2_{(-i)}(1-h_i)} \quad (2.9) \\ &= R^2_i \cdot \left(\frac{\nu-1}{\nu-R^2_i} \right), \end{aligned}$$

with

$$F_{R^2_{(-i)}|X}(t) = F_F(t; 1, \nu - 1). \quad (2.10)$$

Cook's statistic measures the standardized shift in predicted values and the shift in $\hat{\beta}$ due to deleting the i th observation:

$$D_i = \frac{(\hat{y}_{(-i)} - \hat{y})' (\hat{y}_{(-i)} - \hat{y})}{q \cdot \hat{\sigma}^2} = \frac{(\hat{\beta}_{(-i)} - \hat{\beta})' (X'X) (\hat{\beta}_{(-i)} - \hat{\beta})}{q \cdot \hat{\sigma}^2}. \quad (2.11)$$

Furthermore

$$\begin{aligned} D_i &= R^2_i \cdot \frac{h_i}{q(1-h_i)} \quad (2.12) \\ &= R^2_i \cdot C_i. \end{aligned}$$

Finding d such that $Pr\{D_i > d\} = \alpha$ would provide a metric for Cook's statistic. This idea motivates the current work. The results also provide a test of whether a particular D_i arose from the distribution of D_i implied by the GLUM assumptions. As highlighted in §1.1 and §4.3, the latter interpretation has more risks than benefits in practical use for the diagnostic setting.

2.2 The Distribution of Cook's Statistic for Fixed Predictors

For fixed predictors C_i does not vary randomly. Hence, conditional on X ,

$$\begin{aligned} D_i &= C_i \cdot R^2_i \\ &= C_i \cdot \nu \cdot \beta [1/2, (\nu - 1) / 2] \quad (2.13) \\ &= C'_i \cdot \beta [1/2, (\nu - 1) / 2]. \end{aligned}$$

Usually if $i \neq i'$ then $C'_i \neq C'_{i'}$. The value of C'_i does not vary randomly with fixed predictors, but does vary with the i th leverage, h_i , and hence typically varies across sampling units.

In order to provide a metric for judging Cook's statistic it would seem natural to eliminate the heterogeneity between sampling units which occurs with fixed predictors. However,

doing so eliminates the variability due to C_i and makes D_i a simple multiple of R_i^2 , with no distinct information. At least with predictor values assigned by the experimenter, Obenchain's (1977) preference for considering the leverages and residuals separately seems appealing. See Jensen and Ramirez (1996, 1997) for a thorough treatment of fixed predictors.

2.3 The Distribution of Cook's Statistic for Gaussian Predictors

Theorem—Let $a_0 = [q(N-1)]^{-1}$, $a_1 = (q-1)N[q\nu(N-1)]^{-1}$, and $t_0 = \max(a_0, d/\nu)$. For $d > 0$ and Gaussian predictors

$$\Pr\{D_i \leq d\} = 1 - \int_{t_0}^{\infty} \Pr\left\{\beta\left(\frac{1}{2}, \frac{\nu-1}{2}\right) > \frac{d}{\nu t}\right\} f_{C_i}(t) dt, \quad (2.14)$$

with corresponding density

$$f_D(d) = \int_{t_0}^{\infty} f_{\beta}\left(\frac{d}{\nu t}; \frac{1}{2}, \frac{\nu-1}{2}\right) (\nu t)^{-1} f_{C_i}(t) dt. \quad (2.15)$$

Here

$$f_{C_i}(t) = \begin{cases} 0 & t < a_0 \\ f_F\left[(t-a_0)a_1^{-1}; q-1, \nu\right] a_1^{-1} & a_0 \leq t \end{cases}. \quad (2.16)$$

Lemma 1—(Weisberg, 1985, p114) Conditional on knowing X (fixed X)

$$R_i^2 = \nu \cdot \beta[1/2, (\nu-1)/2]. \quad (2.17)$$

Lemma 2—A leverage value from a model containing an intercept and $(q-1)$ multivariate Gaussian predictors, with each row iid, equals a one-to-one function of an F random variable.

Proof—Belsley, Kuh, and Welsch (p66, 1980) proved that

$$F_i = \frac{(h_i - 1/N) / (q-1)}{(1-h_i) / \nu} = F(q-1, \nu). \quad (2.18)$$

Solving their result for h_i yields

$$h_i = \frac{F_i(q-1) / \nu + 1/N}{1 + F_i(q-1) / \nu}. \quad (2.19)$$

Lemma 3—With Gaussian predictors, $C_i = a_0 + a_1 F_i$, so that

$$\begin{aligned} \Pr\{C_i \leq t\} &= \Pr\{a_0 + a_1 F_i \leq t\} \\ &= \Pr\{F_i \leq (t - a_0) / a_1\}, \end{aligned} \quad (2.20)$$

and

$$f_{C_i}(t) = \begin{cases} 0 & t < a_0 \\ f_F[(t - a_0)/a_1; q - 1, \nu] a_1^{-1} & a_0 \leq t \end{cases} \quad (2.21)$$

Proof—For Gaussian predictors the expression in (2.19) for h_i allows stating

$$C_i = \frac{F_i(q-1)/\nu+1/N}{q(1-1/N)} = a_0 + a_1 F_i. \quad (2.22)$$

Lemma 4—Let $X_* = XT$, with T a full rank $q \times q$ matrix of constants. Note that $T^{-t} = (T')^{-1} = (T^{-1})'$. Then H does not vary due to this transformation of the predictors.

Proof—Observe that

$$\begin{aligned} H = X(X'X)^{-1}X' &= XT \left[T^{-1}(X'X)^{-1}T^{-t} \right] T'X' \\ &= XT(T'X'XT)^{-1}T'X' \\ &= X_*(X'_*X_*)^{-1}X'_*. \end{aligned} \quad (2.23)$$

Corollary 4.1— H does not vary due to the covariance matrix of iid random predictors.

Proof—Let $\Sigma_X = FF'$ indicate a factoring of the $(q - 1) \times (q - 1)$ covariance matrix of a row of random predictors, assumed full rank. Choosing

$$T = \begin{bmatrix} 1 & 0 \\ 0 & F^{-t} \end{bmatrix} \quad (2.24)$$

corresponds to considering a new model with predictors $X_* = XT$. The model contains an intercept and $q - 1$ random predictors, with $\Sigma_{X_*} = I$.

Corollary 4.2— $h_i, \hat{e}_i, \hat{\sigma}^2, R_i, C_i$ and D_i do not vary due to full rank transformation of the predictors or the covariance matrix of random predictors.

Proof—Each quantity depends on X only through elements of H .

Lemma 5—With Gaussian predictors $F_{R_{(-i)}^2|h_i}(t) = F_{R_{(-i)}^2|X}(t)$.

Proof—Consider $R_{(-i)}^2$ in terms of three pieces: $(1 - h_i), \hat{\sigma}_{(-i)}^2$ and \hat{e}_i^2 .

- i. Obviously $(1 - h_i)$ depends on X only through h_i .
- ii. Conditional on $X, \hat{\sigma}_{(-i)}^2 (\nu - 1) / \sigma^2 = \chi^2 (\nu - 1)$, and does not depend on X .
- iii. $F_{\hat{e}_i^2|X}(t) = \Phi \left[0, (1 - h_i) \sigma^2 \right]$ and therefore $F_{\hat{e}_i^2|X}(t) = F_{\hat{e}_i^2|h_i}(t)$.

- iv. Conditional on X , by the nature of deletion \hat{e}_i^2 and $\hat{\sigma}_{(-i)}^2$ are statistically independent (LaMotte, 1994, example 1) and $F_{\hat{e}_i/\hat{\sigma}_{(-i)}^2|X}(t) = F_{\hat{e}_i|X}(t) F_{1/\hat{\sigma}_{(-i)}^2|X}(t)$.
- v. Combining i) through iv) completes the proof

Corollary 5.1—With Gaussian predictors $F_{R_i^2|h_i}(t) = F_{R_i^2|X}(t)$.

Proof—Use the last line of (2.9) to write $R_i^2 = \nu \left[(\nu - 1) / R_{(-i)}^2 + 1 \right]^{-1}$. Hence R_i^2 depends on X only through $R_{(-i)}^2$ which depends on X only through h_i .

Corollary 5.2—With Gaussian predictors $F_{R_i^2|C_i}(t) = F_{R_i^2|X}(t)$.

Proof— $C_i = h_i/[q(1 - h_i)]$ and hence depends on X only through h_i .

Proof of the Theorem—Use the law of total probability to state

$$Pr\{D_i > d\} = \int_{a_0}^{\infty} Pr\left\{ \left(R_i^2 | C_i = t \right) > d/t \right\} f_{C_i}(t) dt. \quad (2.25)$$

Equation (2.17) describes the distribution function of R_i^2 conditional on X , which equals the distribution of R_i^2 conditional on C_i , by Corollary 5.2. Combining the distribution in (2.17) with (2.25) allows concluding that

$$Pr\{D_i > d\} = \begin{cases} 1 & d \leq 0 \\ \int_{a_0}^{\infty} Pr\left\{ \beta\left(\frac{1}{2}, \frac{\nu-1}{2}\right) > \frac{d}{vt} \right\} f_{C_i}(t) dt & 0 < d < a_0\nu \\ \int_{d/\nu}^{\infty} Pr\left\{ \beta\left(\frac{1}{2}, \frac{\nu-1}{2}\right) > \frac{d}{vt} \right\} f_{C_i}(t) dt & a_0\nu < d \end{cases} \quad (2.26)$$

Note that $t_0 = \max(a_0, d/\nu)$ and simplify. Finding the density requires differentiating each form in (2.26) separately, and recognizing that the lower limit depends on d . The two apparently distinct forms reduce to a single one upon noting that $f_{\beta}[1/2, (\nu - 1)/2] = 0$.

2.4 Computational Forms for Numerical Integration

Although tantalizing in form, the integral for the CDF of D_i does not allow closed form integration. Numerical integration allows accurate and convenient computation of $Pr\{D_i > d\}$. Both functions in the integral require careful consideration in order to produce a form amenable to computation. Among various forms considered, the ones used here provide the simplest proofs and least computational time for any level of accuracy, except perhaps for small values of $Pr\{D_i > d\}$. Interest usually centers on large values of $Pr\{D_i > d\}$.

Two distinct representations create a finite region of integration, which greatly simplifies numerical integration. First express the density of C_i in terms of an F . If $u = (t - a_0)/a_1$, so that $t = a_1u + a_0$ and $u_0 = (t_0 - a_0)/a_1$ then

$$Pr\{D_i > d\} = \int_{u_0}^{\infty} Pr\left\{ \beta\left[\frac{1}{2}, \frac{(\nu-1)}{2}\right] > \frac{d}{\nu(a_1u+a_0)} \right\} f_F(u; q-1, \nu) du, \quad (2.27)$$

or equivalently

$$Pr \{D_i > d\} = \int_{u_0}^{\infty} Pr \left\{ F(1, \nu - 1) > \frac{\nu - 1}{(a_1 u + a_0) \nu / d - 1} \right\} f_F(u; q - 1, \nu) du. \quad (2.28)$$

The relationship of F and β random variables allows creating a finite region of integration. If $z = (q - 1)u[\nu + (q - 1)u]^{-1}$ then $u = \nu(q - 1)^{-1}z(1 - z)^{-1}$ and $Z_0 = (q - 1)u_0[\nu + (q - 1)u_0]^{-1}$. Also let

$$s(z) = \frac{\nu - 1}{[a_1 \nu (q - 1)^{-1} z(1 - z)^{-1} + a_0] \nu / d - 1}. \quad (2.29)$$

With this transformation

$$Pr \{D_i > d\} = \int_{z_0}^1 Pr \{F(1, \nu - 1) > s(z)\} f_{\beta}\left(z; \frac{q - 1}{2}, \frac{\nu}{2}\right) dz. \quad (2.30)$$

A second useful representation results from applying the transformation $w = u/(1 + u)$ to the integral in (2.28). With $w_0 = u_0/(1 + u_0)$ and

$$h(w) = \frac{\nu - 1}{[a_1 w(1 - w)^{-1} + a_0] \nu / d - 1} \quad (2.31)$$

it follows that

$$Pr \{D_i > d\} = \int_{w_0}^1 Pr \{F(1, \nu - 1) > h(w)\} f_F\left[\frac{w}{(1 - w)}; q - 1, \nu\right] \frac{1}{(1 - w)^2} dw. \quad (2.32)$$

2.5 Approximations

Equation (2.27) allows recognizing that $Pr\{D_i > d\}$ equals the expected value of a function of a random variable whenever $t_0 = a_0$. For fixed q $\lim_{N \rightarrow \infty} d/\nu = \lim_{N \rightarrow \infty} a_0 = 0$. Consequently the expected value interpretation holds, at least asymptotically, in all cases. The accuracy of a series based on treating the integral as an expected value depends both on the remainder term and on any discrepancy due to $d/\nu > a_0$.

Creating a two term Taylor's series approximation for (2.30) involves noting that $\varepsilon\beta(q - 1)/2, \nu/2 = (q - 1)/(\nu + q - 1)$. Ignoring any discrepancy due to $d/\nu > a_0$ yields

$$Pr \{D_i > d\} \approx Pr \left\{ F(1, \nu - 1) > \frac{\nu - 1}{[a_1 + a_0] \nu / d - 1} \right\}. \quad (2.33)$$

Applying a series expansion for an F random variable, using (2.27) or (2.28), requires $\nu > 2k$ to insure finite k th moment. If $\nu > 2$ then $\varepsilon F[(q - 1), \nu] = \nu/(\nu - 2)$ and, ignoring discrepancy due to $d/\nu > a_0$, a two term series equals

$$Pr \{D_i \leq d\} \approx Pr \left\{ F(1, \nu - 1) \leq \frac{\nu - 1}{[a_1 \nu / (\nu - 2) + a_0] \nu / d - 1} \right\}. \quad (2.34)$$

For $\nu = 2$, a one term F based expansion about the number 1 yields

$$Pr \{D_i \leq d\} \approx Pr \left\{ F(1, \nu - 1) \leq \frac{\nu - 1}{[a_1 + a_0] \nu / d - 1} \right\}, \quad (2.35)$$

which corresponds to the two term expansion for the β representation (in 2.34). The approximate probability of (2.35) will never be greater than that of (2.34).

The probability approximations imply approximations for quantiles of D_i :

$$\tilde{d}_p = [a_1 \cdot m + a_0] \nu \left[1 + \frac{\nu - 1}{F_F^{-1}(p; 1, \nu - 1)} \right]^{-1}. \quad (2.36)$$

Here $m = 1$ for (2.34) and (2.36), or $m = \nu/(\nu - 2)$ for (2.35). Assigning m the value of the median, $F_F^{-1}(.50; q - 1, \nu)$, or mode, $\nu(q - 3)/[(q - 1)(\nu + 2)]$, for $q > 3$, also provides a one term approximation.

One convenient form for creating a long series arises from (2.28):

$$\begin{aligned} Pr \{D_i > d\} &= \int_{u_0}^{\infty} Pr \left\{ F(1, \nu - 1) > \frac{\nu - 1}{(a_1 u + a_0) \nu / d - 1} \right\} f_F(u; q - 1, \nu) du \\ &= \int_{u_0}^{\infty} Pr \left\{ F(\nu - 1, 1) \leq \frac{(a_1 u + a_0) \nu / d - 1}{\nu - 1} \right\} f_F(u; q - 1, \nu) du \\ &= \int_{u_0}^{\infty} \{ F(\nu - 1, 1) \leq c_1 u + c_0 \} f_F(u; q - 1, \nu) du \\ &= \int_{u_0}^{\infty} P(u) f_F(u; q - 1, \nu) du. \end{aligned} \quad (2.37)$$

In turn

$$\begin{aligned} P^{(0)}(u) &= \int_0^{c_1 u + c_0} f_F(s; \nu - 1, 1) ds \\ P^{(1)}(u) &= c_1 f_F(c_1 u + c_0; \nu - 1, 1) \\ &\vdots \\ P^{(k)}(u) &= c_1^k f_F^{(k)}(c_1 u + c_0; \nu - 1, 1). \end{aligned} \quad (2.38)$$

2.6 Large Sample Properties

The behavior of D_i in large samples merits separate consideration. The results have both analytic and computational value. Rather than study D_i directly, consider $D_{i*} = \nu D_i$. Then

$$\begin{aligned} Pr \{D_{i*} > d_*\} &= Pr \{ \nu \cdot D_i > d_* \} \\ &= Pr \{ D_i > d_* / \nu \} \\ &= Pr \{ D_i > d \}, \end{aligned} \quad (2.39)$$

with $d = d_* / \nu$. Using (2.28) the distribution function for D_{i*} may be expressed as

$$Pr \{D_{i*} > d_*\} = \int_{u_{0*}}^{\infty} Pr \{ F(1, \nu - 1) > s_*(d_*/\nu, u) \} f_F(u; q - 1, \nu) du, \quad (2.40)$$

with

$$s_*(d_*/\nu, u) = \frac{(\nu - 1) / \nu}{\left[u \left(\frac{q-1}{q} \right) \left(\frac{N}{N-1} \right) + q^{-1} \left(\frac{\nu}{N-1} \right) \right] / d_* - 1 / \nu}, \quad (2.41)$$

$$u_{0*} = [t_0(d_*/\nu) - a_0] / a_1, \text{ and } t_0(d_*/\nu) = \max(a_0, d_*/\nu^2).$$

Consider D_{i*} as $N \rightarrow \infty$. In that case

$$\lim_{N \rightarrow \infty} s_* \left(\frac{d_*}{\nu}, u \right) = \frac{1}{\left[u \left(\frac{q-1}{q} \right) + q^{-1} \right] / d_*} = \frac{d_* q}{[u(q-1)+1]}. \quad (2.42)$$

That $\lim_{N \rightarrow \infty} a_0 = 0$ and $\lim_{N \rightarrow \infty} d_*/\nu^2 = 0$ combine to imply $\lim_{N \rightarrow \infty} u_{0*} = 0$. Therefore

$$\lim_{N \rightarrow \infty} Pr \{D_{i*} > d_*\} = \int_0^\infty Pr \left\{ \chi^2(1) > \frac{d_* q}{[u(q-1)+1]} \right\} (q-1) f_{\chi^2} [(q-1)u; q-1] du. \quad (2.43)$$

Let $w = (q-1)u$, so that $dw = (q-1)du$. Then

$$\lim_{N \rightarrow \infty} Pr \{D_{i*} > d_*\} = \int_0^\infty Pr \left\{ \chi^2(1) > \frac{d_* q}{w+1} \right\} f_{\chi^2}(w; q-1) dw. \quad (2.44)$$

A Taylor's series about $\varepsilon W = (q-1)$ yields the two term approximation

$$Pr \{D_{i*} > d_*\} \approx Pr \left\{ \chi^2(1) > d_* \right\}. \quad (2.45)$$

Also, with $d = d_*/\nu$, for large N

$$Pr \{D_i \leq d\} \approx Pr \left\{ \chi^2(1) \leq \nu \cdot d \right\}, \quad (2.46)$$

with corresponding quantile approximation

$$\tilde{d}_p \approx F_{\chi^2}^{-1}(p; 1) / \nu. \quad (2.47)$$

The F based approximation in (2.36) provides more accuracy, except in large samples. Additional terms are required for the approximation to vary with q .

Three conclusions follow. First, as N increases D_i converges to a degenerate random variable with all mass at zero. Second, D_{i*} converges to a non-degenerate random variable. Third, calculations of quantiles in terms of D_{i*} can greatly reduce numerical difficulties with large samples.

2.7 The Maximum of N Values of Cook's Statistic

Fitting a linear model leads to considering N values of D_i . The non-independence of the set of D_i makes an analytic description of their joint distribution unclear, and computing associated probabilities rather onerous. Despite that, ignoring the multiple testing problem would lead to spuriously rejecting valid data. A Bonferroni correction provides the simplest strategy.

A multiple-testing correction for D_i with fixed predictors reduces to consideration of the same issue for residuals. Cook and Prescott (1981) examined the accuracy of a Bonferroni correction in evaluating N residuals, for fixed X . They provided useful lower bounds, based on residual correlations, to complement the Bonferroni upper bound. The accuracy of the Bonferroni correction decreases as correlations among the residuals increase. Experimental designs with purposeful confounding can create extremely high correlations among some pairs of residuals. Recall that, given X , the residuals have covariance matrix $(I_N - H)\sigma^2$. The Bonferroni correction seems more likely to be universally applicable with Gaussian

predictors. As described in §2.3, for the study of $\{D_i\}$ the covariance matrix for each row of Gaussian predictors may be assumed to be I_{q-1} . Hence the expected correlation for any pair of residuals should be modest and asymptotically zero. The excellent performance of the Bonferroni correction with independent events of small probability promises good accuracy here.

3. NUMERICAL EVALUATIONS

3.1 Exact Probability and Quantile Computations

All exact probabilities reported in this paper were computed by applying Simpson's rule to equation (2.32). All calculations were expressed in terms of the variable $D_{i*} = v \cdot D_i$ in order to provide better numerical accuracy for large sample cases. All exact quantiles were computed via a bisection algorithm (Thisted, 1988, p169) applied to equation (2.32), in terms of D_{i*} . An approximate quantile from equation (2.35) provided the starting value. Equation (2.34) or (2.35) provided starting values. Properties of the function were exploited to refine the code, merely to speed convergence. See Kennedy and Gentle (1980) or Thisted (1988) for a descriptions of Simpson's rule, as well as general discussions of numerical integration, the use of transformations to finite regions such as the one used here, and function inversion algorithms.

3.2 A Simulation

A small simulation study was conducted in order to verify the accuracy of the computational strategy detailed at the end of §2.4, and to assess the accuracy of a Bonferroni correction in evaluating N values of D_i . Assumptions followed those in §2: $y = X\beta + e$ holds, with $\{e_i\}$ iid Gaussian, β fixed and unknown, $\{X_i\}$ iid multivariate Gaussian and independent of $\{e_i\}$. In such cases, finding d such that $Pr\{D_i > d\} = \alpha$ depends only on N , q , and α . The value of d provides a test of whether a particular D_i arose from the hypothesized distribution.

All data were generated under the stated assumptions. Empirical size of the test of D_i was tabulated for each replicate. Two factors were varied in a factorial design: $N \in \{25, 50, 100\}$ and $q \in \{2, 4, 8\}$. For each replicate the first D_i was tested at $\alpha \in \{.01, .05\}$ and the largest D_i was tested at $\alpha \in \{.01/N, .05/N\}$. The pseudo-random generation of data, under valid assumptions, insures that the first D_i represents a pseudo-randomly selected value. In contrast, the distribution of the largest D_i depends on the remaining $N - 1$ values.

Let $Z = [e \ G]$ indicate an $N \times q$ matrix, with $\text{row}_i(Z) = N(\mathbf{0}, I)$. For each combination of N and q a total of 20,000 replicates of Z were created in SAS IML[®], using the function NORMAL. Next $X = [1 \ G]$, $y = X\beta + e$, and $\{D_i\}$ were computed for each, with $\beta = 0_q$ (which implies $y = e$). The first D_i , the largest D_i , N and q were stored for each replicate.

Table I summarizes the empirical size for the tests of D_i with Gaussian predictors, as a function of N , q , and α . The formulas derived in §2 provided accurate probabilities for the simulations of a single value. Furthermore the Bonferroni approximation was quite accurate.

3.3 Comparisons of Approximations

Table II contains probabilities of D_i exceeding $F_F^{-1}(.50; q - 1, \nu)$, and N times the probabilities. Test size systematically and rapidly decreases with N . Ideally a cut-point allows consistent interpretation across regression analyses. The median, or any other quantile of $F(q - 1, \nu)$, does not allow such consistency.

The approximate quantile in equation (2.36) also provides a cut-point requiring only one evaluation of $F_F^{-1}(\cdot; \cdot, \cdot)$. Such values were computed for N and q as in Table II, with target

test sizes of .01 and .05. Equation (2.32) was integrated with Simpson's rule to compute exact probabilities of exceeding the approximate quantiles. In order to approximately evaluate a Bonferroni correction, the same process was followed for target test sizes of $.01/N$ and $.05/N$, with the additional step of multiplying the probabilities by N . As can be seen in Table III, the exact test size ranges from .052 to .058 for a target α of .05, and from .014 to .026 for a target α of .01. The results corresponding to a Bonferroni correction (in the right half of Table III) involve smaller tail probabilities and were much less accurate. An overall target α of .01 gave approximate test sizes ranging from .073 to .398, while a target α of .05 gave approximate test sizes ranging from .200 to .709 (for the conditions examined). Accuracy improves with increasing sample size and number of predictors.

Table IV provides exact critical values of νD_i for $\alpha \in \{.01/N, .05/N\}$ and a range of N and q . Quantiles in Table IV were computed by a simple bisection algorithm (Thisted, 1988, p 169) applied to equation (2.32). An approximate quantile from equation (2.35) provided the starting value. Algorithmic stability across a large range of sample sizes required using $d_* = \nu \times d$.

Different rows in Table IV have different patterns. The range reflects a varying distance from a boundary condition. The studentized residual embedded in D_i requires $N - q - 1 > 0$. The critical value of D_i or $\nu \cdot D_i$ may be taken to be infinity for $N - 1 > q$. The table covers $q \leq 80$. Rows with $N \leq 80$ include the boundary and show a marked upturn in rightmost value. Rows with $N \geq 200$ have no entries near the boundary, and hence display a monotone pattern.

3.4 Comments on Algorithms

Computing and verifying the results in this paper led us to program and evaluate both formulations described in §2.4. Equation (2.32) needed less computation for a fixed accuracy. The advantage of the transformation in (2.32) arises from the shape of the function as both N and q get large. The good performance of the transformation reflects the nature of the random variable $W = U/(1 + U)$, with U following an F distribution. Even though both numerator and denominator degrees of freedom increase, the distribution function of W does not degenerate to a point mass. In contrast, the other two formulations involve convergence (as sample size increases) to degenerate random variables and hence degenerate functions. Two difficulties with (2.32) should be noted. First, $q = 2$ creates a singularity at zero, which often represent the end-point of the interval of integration. Second, extremely large values of d (corresponding to values far beyond those in Table IV) may increase the computational burden.

The care required to insure reasonable numerical performance across a wide range of conditions should not be surprising, given the random variables involved. Kennedy and Gentle (§5.5 and .5.6, 1980) discussed the difficulties in computing F and β probabilities and quantiles. They concluded that no single approach works with all parameter combinations. Thisted (§S.2.2, 1988) provided related material.

4. DISCUSSION

4.1 The Role of Sample Size in Regression Diagnostics

Considering $D_{i*} = \nu D_i$ rather than D_i creates a computational advantage. The two alternatives also reflect two mutually exclusive behaviors for regression diagnostics. For a fixed amount of deviance, distinctions among observations shrink as sample size increases for h_i and D_i (both converge to zero). In contrast, a fixed amount of deviance yields an

interpretation essentially constant as sample size increases for \hat{e}_i , and R_i . In order to emphasize the distinction, compare h_i to

$$h_{i*} = \nu h_i / q = X_i (X'X/\nu)^{-1} X_i' / q = (X_i - 0) (X'X/\nu)^{-1} (X_i - 0)' / q. \quad (4.1)$$

Obviously h_{i*} exhibits the second type of behavior, across N and q . Note that h_{i*} corresponds to the Mahalanobis distance from the origin.

Both types of behavior have merit. Statistics of the first type better reflect the impact of a single observation on the total analysis. With the first type of statistic, the misleading effect of a single observation eventually drowns in rising sample size. Statistics of the second type highlight a given deviant observation, no matter what the sample size. The second type's consistent range of values across sample sizes simplifies interpretation. For example, no matter what the sample size, $R_i = 7$ would demand further attention to the observation.

Sample size also plays a familiar role in the interpretation of D_{i*} and D_i . As always, one must distinguish between statistical "significance" (a small p -value, reflecting rarity of the value) and scientific importance (a difference of consequence in practice). In the present context, some data analysts judge importance by the size of an estimated regression

coefficient, $\hat{\beta}_j$. Others consider the standardized version, $\hat{\beta}_j / \hat{\sigma}(\hat{\beta}_j)$, the corresponding semi-partial correlation coefficient, $r(Y, X_j | \{X_1 \dots X_{j-1}, X_{j+1} \dots X_{q-1}\})$, or the corresponding sums of squares. A regression coefficient may significantly differ from zero but have no practical

importance. Recall that $D_i = (\hat{\beta}_{(-i)} - \hat{\beta})' (X'X)^{-1} (\hat{\beta}_{(-i)} - \hat{\beta}) / [q \cdot \hat{\sigma}^2]$, and captures the *shift* in (standardized) regression coefficients. Hence to judge the importance of an observation highlighted by D_i one should examine the *shift* in $\hat{\beta}$, sum of squares, or multiple correlation, due to deleting the observation.

4.2 Open Questions and Potential Applications

D_i represents one example of a closely related set of diagnostic statistics, including DFFITS _{i} and DFBETAS _{$j(i)$} (Cook and Weisberg, 1982), and a modification of Cook's D_i (Atkinson, 1985). The approach presented here for computing probabilities and quantiles appears to allow similar computations for at least some of the related statistics.

The distribution of the predictors and the purpose of the analysis strongly affects the interpretation of the diagnostics. The results cover only two situations, all fixed or all Gaussian variable predictors plus the special fixed predictor, the intercept. Random but non-Gaussian predictors were also not considered here. Do the results provide an approximation whose quality improves as ν increases? Are the results robust with respect to the form of the distribution?

Another generalization involves models with both fixed and Gaussian predictors. Consider, for example, ANCOVA models, which contain one or more fixed effects, and one or more Gaussian predictors. The fixed parameters for a single fixed factor or for any factorial design can be expressed, without loss of generality, as a set of G cell means, with corresponding columns in X of full rank (G). The statistical independence between rows of data allows the likelihood to be separated into G components. The theory in §1-4 treats the special case of $G = 1$. Consequently the new results apply to each of the G sets of data, considered separately. However, such an analysis would only identify influence within a group, not with respect to all observations in the analysis. Considering all observations simultaneously appears to require additional theoretical results.

More general models may have fixed parameters not expressible as a full-rank cell mean coding (such as a fixed-block design), and/or contain interactions between fixed and Gaussian or Gaussian and Gaussian predictors. Again new theoretical results appear necessary.

Two or more observations may mask the influence of each other. Consequently some research on diagnostics has focused on the impact of deleting two or more observations. Although very appealing, they usually create substantially greater analytic and computational difficulty. Cook and Weisberg (1982) discussed generalizing D_i in this fashion. The theory described here does not accommodate their generalization in any straightforward fashion. The more general result seems worth pursuing. See Jensen and Ramirez (1996, 1997) for the fixed predictor case. Furthermore generalizing the results stated here to multivariate regression (two or more responses) also has merit.

Computational algorithms for probabilities and quantiles of D_i deserve more attention. A series representation would likely provide the best solution, although an even better behaved function for integration might suffice.

4.3 Abusing the Results for Data Analysis

The new results should never be used for automatically discarding an observation. Some widely cited meteorology illustrate the danger of automatic deletion. In order to help process the flood of data from U. S. weather satellites, automatic outlier detection and rejection was applied as part of data formatting and reduction (Kenward, 1988). British scientists (Farman, Gardiner, and Shanklin, 1985), using ground station data, reported a dramatic downward trend across time in ozone levels over the Antarctic. U. S. NASA scientists confirmed the infamous “hole” in the ozone by re-examining their accumulated satellite data, with automatic outlier detection disabled.

4.4 Using the Results for Data Analysis

As discussed in § 1.1, using any diagnostic involves a three step process: 1) highlight bothersome values, 2) investigate the highlighted values, and 3) decide on a disposition, using scientific principles. Whenever the Gaussian predictors assumption seems reasonable, we recommend using the probability and quantile computations for νD_i to highlight observations worthy of investigation. As indicated, explicitly compute and compare $\hat{\beta}$ and $\hat{\beta}_{(-i)}$. As discussed in §4.1, examining the shift in sum of squares or correlation also has appeal. We believe the results presented here provide a useful metric for D_i and valuable insight into its nature and performance.

Acknowledgments

Muller’s work supported in part by NCI grant P01 CA47 982-04, NIH grant M01 RR000-46-33, and NIEHS grant N01-ES-35356. The authors gratefully acknowledge comments on earlier drafts by anonymous reviewers.

BIBLIOGRAPHY

- Atkinson, AC. Plots, Transformations, and Regression. Clarendon Press; Oxford: 1985.
- Belsley, DA.; Kuh, E.; Welsch, RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley; New York: 1980.
- Chatterjee, Sand Hadi AS. Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*. 1986; 1:379–416.
- Chen Mok, M. Evaluating Cook’s D Statistic in Theory and Practice: A Simulation Study. Department of Biostatistics, University of North Carolina; Chapel Hill: 1993. Unpublished Master’s Paper

- Cook RD. Detection of Influential Observations in Linear Regression. *Technometrics*. 1977; 19:15–18.
- Cook RD, Prescott P. On the Accuracy of Bonferroni Significance Levels for Detecting Outliers in Linear Models. *Technometrics*. 1981; 23:59–63.
- Cook, RD.; Weisberg, S. *Residuals and Influence in Regression*. Chapman and Hall; New York: 1982.
- Farman JC, Gardiner BG, Shanklin JD. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*. 1985; 315:207–210.
- Hoaglin DC, Welsch RE. The Hat Matrix in Regression and ANOVA. *American Statistician*. 1978; 32:17–22.
- Jensen, DR.; Ramirez, DE. Computing the CDF of Cook's D_j Statistic. In: Prat, A.; Ripoll, E., editors. *Proceedings of the 12th Symposium in Computational Statistics*; Barcelona, Spain. Instituto de Estadística de Catalunya; 1996. p. 65-66.
- Jensen, DR.; Ramirez, DE. Some exact properties of Cook's D_j . In: Rao, CR.; Balakrishnan, N., editors. *Handbook of Statistics-16: Order Statistics and Their Applications*. North-Holland; Amsterdam: 1997. in press
- Johnson, NL.; Kotz, S. *Continuous Univariate Distributions - 1*. Wiley; New York: 1970a.
- Johnson, NL.; Kotz, S. *Continuous Univariate Distributions - 2*. Houghton Mifflin; Boston: 1970b.
- Kennedy, WJ., Jr.; Gentle, JE. *Statistical Computing*. Marcel Dekker; New York: 1980.
- Kenward M. Surprise, Surprise. *New Scientist*. 1988; 117(1606):16.
- Kleinbaum, DG.; Kupper, LL.; Muller, KE. *Applied Regression Analysis and Other Multivariable Methods*. Second Edition. Duxbury Press; Boston: 1988.
- LaMotte LR. A Note on the Role of Independence in t Statistics Constructed From Linear Statistics in Regression Models. *American Statistician*. 1994; 48:238–240.
- Obenchain RL. Letter to the Editor. *Technometrics*. 1977; 19:348–349.
- Sampson AR. A Tale of Two Regressions. *Journal of the American Statistical Association*. 1974; 69:682–689.
- Thisted, RA. *Elements of Statistical Computing*. Chapman and Hall; New York: 1988.
- Weisberg, S. *Applied Linear Regression*. Wiley; New York: 1985.

TABLE I

Empirical Test Size, $\hat{\alpha}$, for Single and Largest D_i with 20,000 Replications, and standard error of .0007 ($\alpha = .01$) or .0015 ($\alpha = .05$)

q	N	Single		Largest	
		$\alpha = .01$	$.05$	$.01/N$	$.05/N$
2	25	.010	.051	.010	.048
	50	.011	.049	.010	.049
	100	.009	.050	.011	.051
4	25	.010	.049	.010	.047
	50	.011	.050	.011	.050
	100	.011	.051	.010	.051
8	25	.010	.047	.010	.049
	50	.010	.052	.010	.048
	100	.011	.050	.009	.051

TABLE II

Probability of D_i Exceeding F Median as a Function of Sample Size (N) and Number of Gaussian Predictors ($q - 1$)

N	q			
	2	4	8	16
	$Pr\{D_i > F_F^{-1}(.50; q - 1, \nu)\}$			
25	$7.02 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$	$1.19 \cdot 10^{-2}$
50	$7.46 \cdot 10^{-4}$	$3.47 \cdot 10^{-5}$	$6.02 \cdot 10^{-6}$	$5.61 \cdot 10^{-5}$
100	$3.59 \cdot 10^{-5}$	$1.85 \cdot 10^{-7}$	$3.10 \cdot 10^{-9}$	$1.43 \cdot 10^{-10}$
200	$5.55 \cdot 10^{-7}$	$1.20 \cdot 10^{-10}$	$<1 \cdot 10^{-14}$	$<1 \cdot 10^{-14}$
	$N \cdot Pr\{D_i > F_F^{-1}(.50; q - 1, \nu)\}$			
25	$1.76 \cdot 10^{-1}$	$3.82 \cdot 10^{-2}$	$3.46 \cdot 10^{-2}$	$2.98 \cdot 10^{-1}$
50	$3.73 \cdot 10^{-2}$	$1.73 \cdot 10^{-3}$	$3.01 \cdot 10^{-4}$	$2.81 \cdot 10^{-4}$
100	$3.59 \cdot 10^{-3}$	$1.85 \cdot 10^{-5}$	$3.10 \cdot 10^{-7}$	$1.43 \cdot 10^{-8}$
200	$1.11 \cdot 10^{-4}$	$2.41 \cdot 10^{-8}$	$1.39 \cdot 10^{-1}$	$<1 \cdot 10^{-14}$

TABLE III

Probability of D_i Exceeding F -Based Approximate Quantile as a Function of Sample Size (N) and Number of Gaussian Predictors ($q - 1$)

N	q							
	$Pr\{D_i < \tilde{d}.01\}^I$				$N \cdot Pr\{D_i > \tilde{d}.01\}^I$			
25	.022	.023	.022	.026	.182	.166	.149	.296
50	.020	.020	.018	.016	.212	.168	.113	.085
100	.019	.019	.017	.014	.298	.203	.118	.069
200	.019	.019	.016	.014	.398	.268	.142	.073
N	$Pr\{D_i > \tilde{d}.05\}^I$				$N \cdot Pr\{D_i > \tilde{d}.05/N\}^I$			
	25	.054	.058	.057	.058	.295	.286	.260
50	.053	.057	.055	.054	.369	.323	.240	.192
100	.053	.056	.055	.053	.498	.405	.270	.181
200	.053	.056	.055	.053	.709	.539	.330	.200

$$I\tilde{d}_p = [a_1 \cdot m + a_0] \nu \left[1 + (\nu - 1) / F_F^{-1}(p; 1, \nu - 1) \right]^{-1}, \text{ with } m = \nu(\nu - 2) \text{ if } \nu > 2 \text{ and } m = 1 \text{ otherwise.}$$

TABLE IV

Exact Critical Values of $\nu \cdot D_i$ for $\alpha \in \{.01/N, .05/N\}$

$q-1=1$	2	3	4	5	6	7	8	9	10	15	20	40	80	
$\alpha = .05/N$														
10	15.2	16.6	18.2	21.3	28.4	50.2	192.9	
15	15.6	15.6	15.5	15.6	15.9	16.7	18.0	20.3	24.8	34.7	.	.	.	
20	16.4	16.0	15.5	15.1	15.0	14.9	15.0	15.3	15.7	16.4	40.0	.	.	
25	17.2	16.6	15.9	15.3	15.0	14.7	14.6	14.5	14.5	14.6	16.9	44.6	.	
30	17.9	17.1	16.3	15.7	15.2	14.8	14.6	14.4	14.3	14.2	14.6	17.5	.	
40	19.2	18.2	17.2	16.4	15.8	15.4	15.0	14.7	14.5	14.4	13.9	14.1	.	
60	21.3	19.9	18.6	17.7	17.0	16.4	16.0	15.7	15.3	15.1	14.3	13.9	14.5	
80	23.0	21.2	19.8	18.7	17.9	17.3	16.8	16.4	16.1	15.7	14.9	14.3	13.8	
100	24.3	22.3	20.7	19.6	18.7	18.0	17.5	17.1	16.7	16.4	15.4	14.8	13.9	15.6
200	28.8	26.1	24.1	22.5	21.4	20.5	20.0	19.3	18.0	18.4	17.2	16.5	15.2	14.6
400	33.9	30.2	27.6	25.8	24.4	23.5	22.4	22.1	21.3	20.5	19.4	18.5	16.8	16.0
800	40.2	34.0	32.1	29.3	28.2	26.7	25.8	24.4	24.3	23.3	21.8	20.3	18.8	17.5
$\alpha = .01/N$														
10	28.7	31.1	35.1	44.1	66.8	50.5	964.1	
15	26.9	26.1	25.7	25.8	26.7	28.5	31.8	37.8	50.1	80.7	.	.	.	
20	27.2	25.7	24.2	23.6	23.2	23.1	23.3	23.9	24.9	26.5	92.1	.	.	
25	27.9	25.8	24.3	23.2	22.5	22.0	21.7	21.6	21.6	21.8	27.0	102.3	.	
30	28.7	26.4	24.4	23.3	22.4	21.8	21.3	21.0	20.7	20.6	21.5	27.7	.	
40	30.2	27.4	25.4	23.9	22.8	22.2	21.4	20.9	20.5	20.1	19.4	19.7	.	
60	32.3	29.3	26.6	25.2	24.2	23.0	22.2	21.6	21.2	20.7	19.4	18.9	19.8	
80	34.4	31.1	28.1	26.4	25.0	23.9	23.3	22.4	21.8	21.3	19.8	19.2	17.8	
100	36.1	32.6	29.2	27.3	25.8	24.4	24.2	23.3	22.2	22.1	20.2	19.2	18.0	20.7
200	41.2	37.3	34.2	31.3	29.4	28.4	26.9	25.8	25.6	24.5	22.4	21.3	19.3	18.6
400	49.4	45.0	37.6	35.3	34.1	31.0	31.0	29.3	28.2	28.2	25.6	23.3	21.2	20.1
800	68.4	57.7	52.6	40.6	36.9	36.9	33.6	33.6	30.5	30.5	27.7	25.2	22.9	22.9