Detecting Mean-Shift Outliers via Distances

1 INTRODUCTION

Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_I$ be least-squares solutions, and s^2 and s_I^2 the residual mean squares, from the model $\mathbf{Y}_0 = \mathbf{X}_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0$ with and without r observations $\mathbf{Y}_I$ to be assessed for their joint influence. Deletion diagnostics of type $D(\hat{\boldsymbol{\beta}}, \mathbf{M}, c) = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I)'\mathbf{M}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I)/c$ are posed as squared norms for the vector-value sample influence curve $SIC_I = (N-r)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I)$ as in Chapter 3 of Cook and Weisberg (1982). Choices in vogue include $C_I = D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, ks^2)$, $WK_I = D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, ks_I^2)$, $W_I = D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, ks_I^2)$, and $D_I = D(\hat{\boldsymbol{\beta}}, \Sigma^+, rs_I^2)$, due to Cook (1997), Welsch and Kuh (1977), Welsch (1982), and Jensen and Ramirez (1998a). Here N is the full sample size and k the number of elements in $\boldsymbol{\beta}$; $\mathbf{X}$ retains undeleted design points; and Σ^+ is the Moore-Penrose inverse of the dispersion matrix $V(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I) = \Sigma$. Excluding D_I, it remains to determine benchmarks for their proper use. This entails nonstandard distributions, as none of C_I, WK_I, and W_I is properly scaled as a Mahalanobis (1936) distance. Here we draw from distribution theory as set forth in Jensen and Ramirez (1998a), and an algorithm from Ramirez and Jensen (1991), to compute p-values for selected diagnostics in case studies from the literature.

2 BASIC RESULTS

Vectors and matrices appear in bold type, with $\mathbf{M}'$and $\mathbf{M}^{-1}$ as the transpose and inverse; $\mathbf{I}_k$ is the identity of order k; and $O(n)$ denotes the real orthogonal group acting on the Euclidean space $\mathbb{R}^n$. Designate by $\mathcal{L}(\mathbf{Y})$ its law of distribution, by *cdf* its cumulative distribution function, by $N_k(\boldsymbol{\mu}, \Sigma)$ the Gaussian law on $\mathbb{R}^k$ with parameters $(\boldsymbol{\mu}, \Sigma)$, by $\chi^2(\nu)$ the central chi-squared distribution having ν degrees of freedom, and by $F(w; k, \nu, \lambda)$ the Snedecor-Fisher *cdf* with noncentrality λ.

Generalized F distributions are assembled from independent $\{N_1(\omega_i, 1); 1 \leq i \leq r\}$ variates $\mathbf{U}' = [U_1, \cdots, U_r]$, from fixed weights $\{\alpha_1 \geq \cdots \geq \alpha_r > 0\}$, and from $\mathcal{L}(V) = \chi^2(\nu)$ independently of $\mathbf{U}$, as follows. With $T = \alpha_1 U_1^2 + \cdots + \alpha_r U_r^2$ and $W = (T/r)/(V/\nu)$, the *cdf* of W is designated as $F_r(w; \alpha_1, \cdots, \alpha_r, \omega_1, \cdots, \omega_r; \nu)$, reserving $F_r(w; \alpha_1, \ldots, \alpha_r; \nu)$ and $F_r(w; \alpha, \ldots, \alpha; \nu) = F_r(w; \alpha; \nu)$ for the central case when $\boldsymbol{\omega}' = [\omega_1, \cdots, \omega_r] = \mathbf{0}$. These distributions satisfy

$$F_r(w; \alpha_1; \nu) \leq F_r(w; \alpha_1, \ldots, \alpha_r; \nu) \leq F_r(w; \alpha^*; \nu) \ , \tag{1}$$

with α_1 as the maximum and α^* the geometric mean of $\{\alpha_1, \cdots, \alpha_r\}$, and they may be expanded as weighted series of standard F distributions, as shown in Jensen and Ramirez (1991). Since the weights $\{c_i; 0 \leq i \leq \infty\}$ are positive and sum to one, $\sum_{i=\tau+1}^{\infty} c_i$ provides an easy to compute global bound for the truncation error for the *cdf*.

The distributions of WK_I, W_I, and D_I have been characterized in Jensen and Ramirez (1998a); C_I belongs to a separate class owing to dependencies between numerator and denominator. These characterizations in turn stem from a canonical form of the model which we now describe.

Partition $\mathbf{Y}_0$, $\mathbf{X}_0$, and ε_0 conformally as $\mathbf{Y}_0' = [\mathbf{Y}', \mathbf{Y}_I']$, $\mathbf{X}_0' = [\mathbf{X}', \mathbf{Z}']$, and $\varepsilon_0' = [\varepsilon', \varepsilon_I']$, where we suppose that $\mathbf{X}_0$, $\mathbf{X}$, and $\mathbf{Z}$ are full rank of orders $(N \times k)$, $(n \times k)$, and $(r \times k)$, with $k < n < N$, $n + r = N$, and $r \leq k$ for notational convenience. Invoking the theory of singular decompositions, we choose $\mathbf{Q}_1 \in O(n)$, $\mathbf{O}_2 \in O(r)$, and a nonsingular $\mathbf{G}(k \times k)$, such that $\mathbf{Q}_1\mathbf{X}\mathbf{G} = [\mathbf{I}_k, \mathbf{0}']'$ and $\mathbf{Q}_2\mathbf{Z}\mathbf{G} = [\mathbf{D}_\gamma, \mathbf{0}]$, where $\mathbf{D}_\gamma$ is diagonal with elements $\{\gamma_1 \geq \cdots \geq \gamma_r > 0\}$ as the square roots of the eigenvalues of $\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}'$. The eigenvalues $\{\lambda_1 \geq \cdots \geq \lambda_r > 0\}$ of $\mathbf{Z}(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{Z}'$, called the *canonical leverages,* satisfy $\{\lambda_i = \gamma_i^2/(\gamma_i^2+1);\ 1 \leq i \leq r\}$. These operations in turn transform the model $\mathbf{Y}_0 = \mathbf{X}_0\boldsymbol{\beta} + \varepsilon_0$ one-to-one into our canonical form $\mathbf{U} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\eta}$ with $\boldsymbol{\theta} = \mathbf{G}^{-1}\boldsymbol{\beta}$, which we partition as $\boldsymbol{\theta}' = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2']$ with $\boldsymbol{\theta}_1 \in \mathbb{R}^r$ and $\boldsymbol{\theta}_2 \in \mathbb{R}^s$, for $s = k - r$. To model a possible shift of $\boldsymbol{\xi}$ units in $E(\mathbf{Y}_I)$ at design points in $\mathbf{Z}$, we equivalently model a shift of $\boldsymbol{\delta} = \mathbf{Q}_2\boldsymbol{\xi}$ units in the corresponding elements of $\mathbf{U}$. For further details see Jensen and Ramirez (1998a). Basic properties of D_I, and of scaled versions of WK_I and W_I, may be summarized as follows.

Theorem 1 *Suppose that* $\mathcal{L}(\mathbf{Y}_0) = N_N(\mathbf{X}_0\boldsymbol{\beta} + \boldsymbol{\xi}_0, \sigma^2\mathbf{I}_N)$, *with* $\boldsymbol{\xi}_0' = [\mathbf{0}', \boldsymbol{\xi}']$, *and let* $\{\omega_i = \delta_i/(\sigma(\gamma_i^2+1)^{1/2});\ 1 \leq i \leq r\}$.
(i) The cdf of $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, rs_I^2) = (k/r)WK_I$ *is given by* $F_r(w; \gamma_1^2, \cdots, \gamma_r^2, \omega_1, \cdots, \omega_r; n-k)$ *for each* $r \leq k$.
(ii) The cdf of $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, rs_I^2) = (k/r)W_I$ *is given by* $F_r(w; \lambda_1, \cdots, \lambda_r, \omega_1, \cdots, \omega_r; n-k)$ *for each* $r \leq k$.
(iii) The cdf of $D_I = D(\hat{\boldsymbol{\beta}}, \Sigma^+, rs_I^2)$ *is given by* $F_r(w; r, n-k, \lambda(\boldsymbol{\delta}))$ *with* $\lambda(\boldsymbol{\delta}) = \sum_{i=1}^r \delta_i^2/(\sigma^2(\gamma_i^2+1))$.

Proof. See Jensen and Ramirez (1998a). ■

We further remark that the variance-ratio statistic for testing $H_0{:}\,\boldsymbol{\delta} = \mathbf{0}$ against $H_1{:}\,\boldsymbol{\delta} \neq \mathbf{0}$ in canonical form is

$$F_I = D(\hat{\boldsymbol{\theta}}_1, \mathbf{D}_\gamma^{-1}(\mathbf{I}_r + \mathbf{D}_\gamma^2)\mathbf{D}_\gamma^{-1}, rs_I^2). \tag{2}$$

In particular, its distribution is identical to that of $D_I = D(\hat{\boldsymbol{\beta}}, \Sigma^+, rs_I^2)$ as in Theorem 1. Regarding the power of the F test, it is seen from the noncentrality parameter $\lambda(\boldsymbol{\delta}) = \sum_{i=1}^r \delta_i^2/(\sigma^2(\gamma_i^2+1))$ that high leverages tend to mask a given shift $\boldsymbol{\delta}$ from the null hypothesis. See Cook and Weisberg (1982, page 21) for $r = 1$.

For the case $r = 1$, we note that $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, s_i^2)/\gamma_1^2$, $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, s_i^2)/\lambda_1$, and $D(\hat{\boldsymbol{\beta}}, \Sigma^+, s_i^2)$ have identical distributions. Thus the three p-values from Theorem 1 are identical when $r = 1$. Moreover, single-case outliers can be tested using the Studentized deleted residuals $t_i^* = (y_i - \hat{y}_{(i)})/(s_i\sqrt{1 + \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i})$, or the R-Student externally Studentized residuals $t_i = (y_i - \hat{y}_i)/(s_i\sqrt{1 - h_{ii}})$. Here $\widehat{y}_{(i)}$ and $\widehat{y}_i$ denote predicted values using $(\mathbf{Y}, \mathbf{X})$ and $(\mathbf{Y}_0, \mathbf{X}_0)$, respectively, and h_{ii} is the ordinary leverage. Jensen and Ramirez (1998b) showed that the p-values from these two tests are also equal to the p-values from Theorem 1 when $r = 1$.

We turn next to the matter of evaluating p-values numerically for selected diagnostics in case studies from the literature.

3 EXAMPLES: SINGLE-ROW INFLUENCE

We use the data on manpower and work load for U.S. Navy Bachelor Officers' Quarters (BOQ) from Myers (1990). The linear model has $N = 25$ and $k = 8$, and so with $r = 1$, $\nu = N-k-r = 16$. Table I reports the five sites with p-values < 0.10. The analysis of Table I shows, using $D_i(\mathbf{X}_0'\mathbf{X}_0) = D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, s_i^2)$, that the influential sites are 23 and 24, with $p_i < 0.01$. The high values of γ_1^2 correspond to high values for λ_1 since $\lambda_1 = \gamma_1^2/(\gamma_1^2 + 1)$. As noted in Belsley, *et al.* (1980, p.49), high leverage can be viewed as either neutral or beneficial. Since the distribution of $D_i(\mathbf{X}_0'\mathbf{X}_0)$ is scaled by γ_1^2, high values of γ_1^2 do not *necessarily* mean the observation is influential. Theorem 1 shows that it is *not* the magnitude of $D_i(\mathbf{X}_0'\mathbf{X}_0)$ alone that determines an observation of high influence, but rather the magnitude of the ratio of $D_i(\mathbf{X}_0'\mathbf{X}_0)$ to γ_1^2. As $r = 1$, $D_i(\mathbf{X}_0'\mathbf{X}_0)/\gamma_1^2 = D_i(\mathbf{X}'\mathbf{X})/\lambda_1 = D_i(\Sigma^+) = t_i^2$, and all the criteria have identical p-values.

Table I. Single-Row Influence for the BOQ Data from Myers (1990)

Site	γ_1^2	$D_i(\mathbf{X}_0'\mathbf{X}_0)$	λ_1	$D_i(\mathbf{X}'\mathbf{X})$	$D_i(\Sigma^+)$	t_i	s_i^2	p_i
23	85.656	2353.980	.989	27.165	27.482	-5.242	80999	.0001
24	7.076	72.886	.876	9.025	10.300	3.209	133917	.0055
20	.578	4.747	.366	3.008	8.213	-2.866	145464	.0112
15	1.260	7.999	.558	3.539	6.347	-2.519	157611	.0228
21	.076	.319	.070	.297	4.220	2.054	174199	.0567

Table I has been ranked on p_i, the p-values in the last column. This is equivalent to ranking on s_i^2 or on $1/D_i(\Sigma^+) = 1/t_i^2$, as follows.

Theorem 2 *If I_1 and I_2 are two subsets of $\{1, \cdots, N\}$ with r elements, then the following are equivalent, where $Q_I(\hat{\boldsymbol{\beta}}, \mathbf{M}) = D(\hat{\boldsymbol{\beta}}, \mathbf{M}, 1)$.*
(1) $s_{I_1}^2 \leq s_{I_2}^2$,
(2) $Q_{I_1}(\hat{\boldsymbol{\beta}}, \Sigma^+) \geq Q_{I_2}(\hat{\boldsymbol{\beta}}, \Sigma^+)$,
(3) $D_{I_1}(\Sigma^+) \geq D_{I_2}(\Sigma^+)$,
(4) $F_{I_1} \geq F_{I_2}$, and
(5) $p_{I_1} \leq p_{I_2}$.

Proof. Following Jensen and Ramirez (1998a), we partition the residual sum of squares for the full data as

$$(N-k)s^2 = Q_I(\hat{\boldsymbol{\beta}}, \Sigma^+) + (n-k)s_I^2.$$

Since this is fixed in a given experiment, we see that $s_{I_1}^2 \leq s_{I_2}^2$ is equivalent to $Q_{I_1}(\hat{\boldsymbol{\beta}}, \Sigma^+) \geq Q_{I_2}(\hat{\boldsymbol{\beta}}, \Sigma^+)$. The remaining conclusions follow directly. ■

4 EXAMPLES: MULTIPLE-ROW INFLUENCE

We next consider multiple-row influence in the BOQ data for pairs of observations with $I = \{i_1, i_2\}$. Extensions to larger subsets proceed similarly. Table I reports that sites 23 and 24 are outliers with $p < 0.01$. We now seek out, from the 300 pairs of sites, other sites that are jointly influential. We first use the statistic $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, rs_I^2)$ as reported in Table II together with the two nonzero eigenvalues $\{\gamma_1^2, \gamma_2^2\}$ of $\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}'$, and s_I^2. Also shown are the lower (LB) and upper (UB) bounds for the p-values from Equation 1, the estimated p-values p_A (using the average of $\{\gamma_1^2, \gamma_2^2\}$), the condition number $\kappa(\gamma^2) = \gamma_1^2/\gamma_2^2$, the exact p-values, and the number of terms τ used in the series expansion for the generalized F distribution to achieve a global error bound of 10^{-4}. Here we have $r = 2$ and $v = 15$.

Table II: Multiple-Row Influence for BOQ using $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, rs_I^2)$

Sites	D_I	γ_1^2	γ_2^2	s_I^2	LB	p_A	UB	$\kappa(\gamma^2)$	p_I	τ
15,20	10.03	1.263	.577	91312	.0008	.0012	.0044	2.2	.0018	13
20,25	36.53	5.900	.062	106150	.0000	.0007	.0110	95.9	.0031	760
20,21	3.34	.578	.076	107346	.0002	.0016	.0138	7.7	.0043	57
11,15	5.05	1.261	.142	136013	.0008	.0064	.0404	8.9	.0136	67
19,20	2.23	.589	.090	144071	.0020	.0090	.0468	6.6	.0165	48
7,20	2.73	.592	.215	144645	.0052	.0081	.0276	2.88	.0110	18
13,20	2.23	.578	.087	154713	.0018	.0083	.0445	6.6	.0155	49
14,20	2.23	.579	.107	155115	.0027	.0093	.0447	5.4	.0161	39
8,22	23.95	7.109	.366	161945	.0003	.0097	.0619	19.4	.0209	151

In Tables II, III, and IV, we have shown only those pairs of sites that do not contain either site 23 or 24 and with $p_A < 0.01$. Our computer simulations show that p_A is a good estimate of the p-value when κ is small.

Table III: Multiple-Row Influence for BOQ using $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, rs_I^2)$

Sites	D_I	λ_1	λ_2	s_I^2	LB	p_A	UB	$\kappa(\lambda)$	p_I	τ
15,20	5.22	.558	.366	91312	.0009	.0010	.0023	1.3	.0012	8
20,25	5.44	.855	.058	106150	.0000	.0008	.0100	14.8	.0030	113
20,21	2.20	.366	.070	107346	.0004	.0017	.0122	5.2	.0040	37
11,15	2.33	.558	.124	136013	.0029	.0077	.0360	4.5	.0130	32
19,20	1.43	.371	.082	144071	.0040	.0103	.0445	4.5	.0164	32
7,20	1.72	.372	.177	144645	.0084	.0106	.0274	2.1	.0127	12
13,20	1.41	.366	.080	154713	.0038	.0101	.0444	4.6	.0164	32
14,20	1.41	.367	.096	155115	.0055	.0115	.0447	3.8	.0171	26
8,22	2.95	.877	.268	161945	.0115	.0197	.0618	3.3	.0257	22

Note for any pair of sites, that $\kappa(\lambda) = ((\gamma_2^2+1)/(\gamma_1^2+1))\kappa(\gamma^2) < \kappa(\gamma^2)$. Since the number of terms τ used in the series expansion is related to the condition number, we prefer $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, rs_I^2)$ to $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, rs_I^2)$, requiring fewer terms.

Table IV reports the influential pairs of sites found using $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, s_I^2)$. The columns are ranked on p_I which, by Theorem 2, is equivalent to ranking on s_I^2 or $1/D_I(\Sigma^{+})$.

Table IV: Multiple-Row Influence for BOQ using $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, s_I^2)$

Sites	$D_I(\Sigma^{+})$	α_1	α_2	s_I^2	p_I
15,20	11.79	1	1	91312	.0008
20,25	9.09	1	1	106150	.0026
20,21	8.91	1	1	107346	.0028
11,15	5.45	1	1	136013	.0167
19,20	4.72	1	1	144071	.0256
7,20	4.67	1	1	144645	.0264
13,20	3.88	1	1	154713	.0438
14,20	3.85	1	1	155115	.0446
8,22	3.37	1	1	161945	.0617

The diagnostic $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, s_I^2)$ has the advantage of having the easily computed distribution $F(w; r, n-k)$ from Theorem 1.

Cook (1977) considered the data sets of Longley (1967) and Hald (1952). Cook found that the Hald data were well behaved with observation 8 having the largest C_i value. Here $N = 13$ and $k = 5$. We now can add that the p-value for observation 8 is $p_8 = 0.0835$, using the statistics $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, s_i^2)$, $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, s_i^2)$, or $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, s_i^2)$; or using the studentized deleted residuals; or the R-Student t_i; or using the mean shift outlier model. We also add that the only pair of observations that is possibly influential is $I = \{6, 8\}$ with p-values $p_I = 0.0231$, 0.0218, and 0.0197 based on $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, 2s_I^2)$, $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, 2s_I^2)$, and $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, 2s_I^2)$, respectively.

For the Longley data, Cook noted that observations 5 and 16 may be jointly influential. Here $N = 16$ and $k = 7$. We now can add that the individual p-values for these observations are $p_5 = 0.1027$ and $p_{16} = 0.2459$. With $I = \{5, 16\}$, $p_I = 0.1421$, 0.1293, and 0.1235 based on $D(\hat{\boldsymbol{\beta}}, \mathbf{X}_0'\mathbf{X}_0, 2s_I^2)$, $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, 2s_I^2)$, and $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, 2s_I^2)$, respectively, offering only marginal support for their joint influence.

Our recommendation is to screen initially for joint outliers using $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, rs_I^2)$, or equivalently, using F_I (Equation 2). Theorem 2 shows that the ranking based on $D(\hat{\boldsymbol{\beta}}, \Sigma^{+}, rs_I^2)$ is the same as the ranking based on s_I^2. These calculations can be found easily using, for example, Minitab. Finally, we recommend that the p-values be computed for $D(\hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{X}, rs_I^2)$. The Fortran77 program GEN_F is available from the second author.

References

[1] Belsley, D., Kuh, E., and Welsch, R. (1980). *Regression Diagnostics.* John Wiley & Sons, Inc., New York.

[2] Cook, R. (1977). Detecting of influential observations in linear regression models, *Technometrics,* 19, 15-18.

[3] Cook, R. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, London.

[4] Hald, A. (1952). *Statistical Theory with Engineering Applications.* John Wiley & Sons. Inc., New York.

[5] Jensen, D. R. and Ramirez, D. E. (1991). Misspecified T^2 tests. I. Location and scale, *Commun. Statist. -Theory Meth.,* 20, 249-259.

[6] Jensen. D. R. and Ramirez, D. E. (1998a). Some exact properties of Cook's D_I statistic. In *Handbook of Statistics*, vol. 16, Balakrishnan, N. and Rao, C. eds., pp. 387-384, Elsevier Science Publishers, Amsterdam.

[7] Jensen, D. R. and Ramirez, D. E. (1998b). Detecting outliers with Cook's D_I statistic. *Computing Science and Statistics* 29(1), 581-586.

[8] Longley, H, (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user, *J. Amer. Statis. Assoc.*, 62, 819-841.

[9] Mahalanobis, P. C. (1936). On the generalized distance in statistics, *Proc. Nat. Inst. Sci. India,* 2, 49-55.

[10] Myers, R. H. (1990). *Classical and Modern Regression with Applications.* Duxbury Press, Belmont, California.

[11] Ramirez, D. E. and Jensen, D. R. (1991). Misspecified T^2 tests. II. Series expansions, *Commun. Statist. -Simula.* 20, 97-108.

[12] Welsch, R. E. (1982). Influence functions and regression diagnostics. In Launer, R. L. and Siegel, A. F., eds., *Modern Data Analysis.* Academic Press, New York.

[13] Welsch, R. E. and Kuh, E. (1977). Linear regression diagnostics. *Technical Report 923-77.* Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

D. R. Jensen
Department of Statistics
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

D. E. Ramirez
Department of Mathematics
University of Virginia, Charlottesville, VA 22903