Letters to the Editor

CLEMONS, T., AND PAGANO, M. (1999), "ARE BABIES NORMAL?" *THE AMERICAN STATISTICIAN*, 53, 298–302: COMMENT BY STELLMAN

Clemons and Pagano (1999) examined the origin of the initially anomalous distribution of birth weights derived from the Centers for Disease Control's computerized birth certificate files. They surmised that some hospitals record birth weights in grams and others in pounds and ounces, and then showed that this assumption leads to an appropriate choice of bins for tabulating frequency histograms which reduces or eliminates the anomaly together with an unusual digit bias. They have thus elegantly demonstrated that the source of the initial anomaly is an administrative rather than an inherently statistical one.

They showed also (Figure 7) that the re-binned data produce a normal Q-Q plot, except for very light and very heavy babies. They cite studies that tend to explain the deviation from normality at the low end as due to inclusion of pre-term births, and at the high end as due in part to births to diabetic mothers.

I wish to suggest that the deviation at the low end may contain an additional administrative component. Infant mortality (for which low birth weight is a major risk factor) is a closely watched health "indicator," especially in large cities like New York where municipally funded delivery of health care and social service resources may be based partly on neighborhood-specific rates. During 1988–1991, when I served as the New York City Assistant Commissioner of Health with responsibility for birth registration (a state function in most other jurisdictions), after closing the birth registration files for a given year, we searched the medical records of hospitals in high-risk areas for unreported infant deaths. In a year with about 135,000 reported live births, it was not uncommon to find up to 50 unreported infant deaths relative to 1,700 initially reported. Many such unreported deaths were among infants of extremely low birth weight who expired shortly after birth—sometimes in minutes, or following an abnormally short pregnancy. We learned through discussions that some physicians or hospitals assumed that in these instances births, deaths, or both were not legally reportable, while others thought they needed only to file reports of spontaneous terminations of pregnancy. However, the New York City Health Code defines a live birth in part as "expulsion or extraction from its mother of a product of conception, regardless of the duration of the pregnancy, which ... shows evidence of life, such as breathing, beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles ..." (New York City Department of Health 1990).

On the basis of the legal definition, such births must be counted as live births, and should therefore have been counted in the histograms presented by Clemons and Pagano. Our experience suggests that the reporting of many such births is dependent upon the degree of resources which a given birth jurisdiction commits to "extra" surveillance over and above the hospital's legal reporting obligations. Deviation from normality among low-birth weight infants may be due in part to biased reporting of births due to lack of complete understanding of the local health code and reporting requirements on the part of physicians and hospital administrators.

Steven D. STELLMAN
Division of Epidemilogy
American Health Foundation
300 East 42nd Street
New York, NY 10017

REFERENCES

New York City Department of Health (1990), *New York City Health Code, The New York City Charter, and the New York City Administrative Code Relating to Births, Deaths, Terminations of Pregnancy, Funeral Directors, Cemeteries, and Disposals of Human Remains*, New York: author.

HENGARTNER, N. W. (1999), "A NOTE ON MAXIMUM LIKELIHOOD ESTIMATION," *THE AMERICAN STATISTICIAN*, 53, 123–125: COMMENT BY KHATTREE AND SEN

Hengartner (1999) presented an interesting example of the situation where the use of additional information results in a maximum likelihood estimator that is inferior to the one which ignores this additional information. This seemingly counterintuitive phenomenon has been observed earlier in other contexts, as well. For example, Birch and Robertson (1983) demonstrated that, while estimating the variance σ^2 of a Normal population with known mean μ based on a random sample $X_1, X_2, \ldots, X_n$, the maximum likelihood estimator $\hat{\sigma}^2 = 1/n \sum_{i=1}^{n}(X_i - \mu)^2$ is inferior to the plug-in estimator $\tilde{\sigma}^2 = 1/n \sum_{i=1}^{n}(X_i - \overline{X})^2$ (MLE when μ is unknown), obtained by replacing μ by $\overline{X}$, with respect to the mean-squared error (MSE) criterion. The difference between this case and the one presented by Hengartner (1999) is that, while in the former the ignored information is that of a parameter of the original population, the latter deals with the issue of ignoring available information arising from an additional random sample from a population with a mean shifted with respect to the original population.

We submit the following comments to point out that the specific findings by the author in this case are largely dependent on the chosen criterion for comparison. In what follows, we explore the same situation the author has investigated, but on the basis of a probabilistic criterion based on the notion of Pitman measure of closeness (PC). Given two estimators, T_1 and T_2, of a parameter θ, under PC criterion, T_1 is said to be better than T_2, if

$$\Pr\left[|T_1 - \theta| < |T_2 - \theta|\right] > \frac{1}{2}. \tag{1}$$

Following the notation of Hengartner (1999), let $X_1, \ldots, X_n$, and $Y_1, \ldots, Y_m$ be independent random samples from $\mathrm{Bin}(1, p)$ and $\mathrm{Bin}(1, q)$, respectively, with $q \leq p$. In order to estimate p, the estimators considered by the author are

$$\hat{p} = n^{-1} \sum_{i=1}^{n} X_i,$$

and

$$p = \begin{cases} \hat{p}, & \text{if} \quad n^{-1}\sum_{i=1}^{n} X_i \geq m^{-1}\sum_{j=1}^{m} Y_j \\ \left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j\right) \Big/ (n+m), & \\ \quad \text{if} \quad n^{-1}\sum_{i=1}^{n} X_i < m^{-1}\sum_{j=1}^{m} Y_j. & \end{cases}$$

In the present context, the roles of T_1 and T_2 in (1) will be assumed by $\hat{p}$ and $\tilde{p}$, respectively, and θ is our parameter of interest p. Note now that the definition of $\tilde{p}$ entails that

$$\Pr\left[|\hat{p} - p| = |\tilde{p} - p|\right] = \Pr\left[n^{-1}\sum_{i=1}^{n} X_i \geq m^{-1}\sum_{j=1}^{m} Y_j\right] > 0.$$

Following Nayak's (1990) suggestion in such cases, we should replace (1) by a direct comparison of $\mathrm{PC}(\hat{p}, \tilde{p}; p) = \Pr[|\hat{p} - p| \leq |\tilde{p} - p|$ and $\mathrm{PC}(\tilde{p}, \hat{p}; p) = \Pr[|\tilde{p} - p| \leq |\hat{p} - p|]$ in order to conclude the superiority of one estimator over the other. In other words, $\hat{p}$ will be better than $\tilde{p}$ in PC sense if

$$\mathrm{PC}(\hat{p}, \tilde{p}; p) > \mathrm{PC}(\tilde{p}, \hat{p}; p). \tag{2}$$

Preference for $\tilde{p}$ over $\hat{p}$ is established by the reverse inequality.

Since $\hat{p} \leq \tilde{p}$,

$$\begin{aligned} \mathrm{PC}(\hat{p}, \tilde{p}; p) &= \Pr[(\hat{p} - \tilde{p})(\tilde{p} + \hat{p} - 2p) \leq 0] \\ &= \Pr[\hat{p} - \tilde{p} = 0] + \Pr[\tilde{p} + \hat{p} - 2p \geq 0] \end{aligned}$$

and similar considerations hold for $\mathrm{PC}(\tilde{p}, \hat{p}; p)$. Thus, it suffices to compare the probabilities $\Pr[\tilde{p} + \hat{p} - 2p \geq 0]$ and $\Pr[\hat{p} + \tilde{p} - 2p \leq 0]$. It is easily seen that, when $p > 1/2$

$$\Pr[\tilde{p} + \hat{p} - 2p \geq 0] = \Pr\left[\frac{\tilde{p} + \hat{p}}{2} \geq p\right] \geq \Pr(\hat{p} \geq p) > \frac{1}{2}, \tag{3}$$

since in this case distribution of $\hat{p}$ is negatively skewed. Thus, for $p > 1/2$, $\hat{p}$ is preferred to $\tilde{p}$ in PC sense. This preference region is in stark contrast to the MSE preference region obtained by the author. It may also be noted that unlike the MSE criterion, the PC preference region here does not depend on m and n.

It may be interesting to establish a sufficient condition under which $\tilde{p}$ is preferred over $\hat{p}$ in the sense of (2). A calculation similar to (3) shows that one range of p over which $\tilde{p}$ dominates $\hat{p}$ is given by the restriction: $p - \text{median}(\tilde{p}) > 0$, which is a highly complex, nonlinear function of both p and q. It is thus imperative that dominance of one estimator over the other depends on the chosen comparison criterion and in many cases—including the present one—may yield completely contradictory results under different criteria. It may be pointed out that similar comparison problems under additional information and under PC criterion have been investigated by Khattree (1992), Gupta and Khattree (1993, 1994), and Bose, Datta, and Ghosh (1993).

Ravindra KHATTREE
Ananda SEN

Department of Mathematics and Statistics
Oakland University
Rochester, MI 48309-4485

REFERENCES

Bose, S., Datta, G. S., and Ghosh, M. (1993), "Pitman's Measure of Closeness for Symmetric Stable Distributions," *Statistics and Probability Letters*, 17, 245–251.

Birch, J. J., and Robertson, T. (1983), "A Classroom Note on the Sample Variance and the Second Moment," *American Mathematical Monthly*, 90, 703–705.

Gupta, R. D., and Khattree, R. (1993), "A Comparison of Some Estimators of the Variance Covariance Matrix When the Population Mean is Known," *Statistics*, 24, 331–345.

——— (1994), "Some Comments on Maximum Likelihood Estimators of Scale Parameter or its Square," *Journal of Indian Statistical Association*, 32, 123–131.

Khattree, R. (1992), "Comparing Estimators for Population Vaxiance Using Pitman Nearness," *The American Statistician*, 46, 214–217.

Nayak, T. K. (1990), "Estimation of Location and Scale Parameters Using Generalized Pitman Nearness Criterion," *Journal of Statistical Planning and Inference*, 24, 259–268.

DAWSON, K. S., GENNINGS, C., AND CARTER, W. H. (1997), "TWO GRAPHICAL TECHNIQUES USEFUL IN DETECTING CORRELATION STRUCTURE IN REPEATED MEASURES DATA," *THE AMERICAN STATISTICIAN*, 51, 275–283: COMMENT BY KESELMAN, ALGINA, AND KOWALCHUK

In the article by Dawson, Gennings, and Carter (1997) graphical procedures were presented intended to help researchers identify the correct covariance structure of their data in order to arrive at better tests of the fixed-effects in mixed model analyses of repeated measures data with SAS' (SAS Institute 1996) PROC MIXED program. That is, one of the newer approaches to the analysis of repeated measurements is based on a mixed model approach (see Littell, Milliken, Stroup, and Wolfinger 1996). The potential benefit of this approach is that it allows a user to model the covariance structure of the data rather than presuming a certain type of structure as is the case with the traditional univariate and multivariate test statistics. Parsimoniously modeling the covariance structure of the data should result in more efficient estimates of the fixed-effects parameters of the model and consequently more powerful tests of the repeated measures effects.

The mixed approach, and specifically PROC MIXED, allows users to fit various covariance structures to the data. For example, some of the structures that can be fit with PROC MIXED are: (a) compound symmetric (CS); (b) unstructured (UN); (c) spherical; (d) first-order autoregressive (AR); and (e) random coefficients (RC). The spherical structure is assumed by the traditional univariate F tests in SAS's GLM program, while the UN structure is assumed by GLM's multivariate tests of the repeated measures effects. AR and RC structures more appropriately reflect that measurement occasions that are closer in time are more highly correlated than those farther apart in time. In addition, PROC MIXED allows users to specify, separately and jointly, between-subjects and within-subjects heterogeneity. It is suggested that users first determine the appropriate covariance structure prior to conducting tests of significance for the repeated measures effects (see Littell et al. 1996). Hence, Dawson et al. (1997) suggested that researchers use their graphical procedures (i.e., draftman's display and parallel axis display) in conjunction with the Akaike (1974) information criterion and/or Schwarz (1978) Bayesian criterion values in order to select the correct covariance structure (see also Littell et al.).

Unfortunately, the research by Keselman, Algina, Kowalchuk, and Wolfinger (1998, 1999) indicates that the default F tests that SAS uses to test the within-subjects effects can be moderately biased in certain cases.

In particular, they compared the mixed model approach, the multivariate Welch–James nonpooled test enumerated by Keselman, Carriere, and Lix (1993), and the corrected df test (improved general approximation test) due to Huynh (1978). The tests were compared for unbalanced nonspherical repeated measures designs containing one between-subjects and one within-subjects variable when the assumptions of covariance homogeneity and multivariate normality were violated separately and jointly. Specifically in a 3×4 design where the data were generated so that the sphericity parameter (ϵ) equaled .75, they varied the: (1) covariance structure (UN, ARH, and RCH, where H designates between-subjects heterogeneity); (2) equality (1:1:1)/inequality (1:3:5) of the between-subjects covariance matrices; (3) equality/inequality of the group sizes (unequal group sizes cases were equal to: (a) $8, 10, 12$ and $6, 10, 14\,(N = 30)$, (b) $12, 15, 18$ and $9, 15, 21\,(N = 45)$, and (c) $16, 20, 24$ and $12, 20, 28\,(N = 60)$); (4) type of pairing of covariance matrices and group sizes (positive/negative); and (5) distributional form of the data (multivariate normal/lognormal).

Their results indicated that the default tests available through PROC MIXED typically were conservative or liberal when unequal covariance matrices were paired in either a positive or negative way with unequal group sizes. In particular, the rates of error were depressed or inflated when the PROC MIXED tests were based on either the best Akaike or Schwarz criteria. As well, the rates were liberal when the *correct* covariance structure was used with the tests of the repeated measures effects.

Thus, whether a covariance structure is selected with graphical methods and/or with the Akaike (1974) and Schwarz (1978) criteria will not alter the fact that the default F tests available through PROC MIXED can be biased under certain conditions. Accordingly, any presumed power benefits must be discounted when the procedure is prone to excessive rates of Type I error. The liberal nature of many of the PROC MIXED tests may be due to the fact that the default F approximation is based on the adjusted residual df. We are currently investigating more conservative approximations.

On the other hand, the tests enumerated by Keselman et al. (1993), and Huynh (1978), were generally able to control their rates of Type I error even when asssumptions were jointly violated. The Welch–James test, however, required a larger sample size to achieve robustness. Based on the results reported by Keselman et al. (1998, 1999), Keselman et al. (1993), and Algina and Keselman (1998) we recommend the Welch–James test for analyzing effects in repeated measures designs. The Welch–James test typically will not only provide a robust test of repeated measures effects but as well will generally provide a more powerful test of nonnull effects compared to Huynh's (1978) improved general approximation test, in repeated measures designs. Indeed, Algina and Keselman found—when Type I error rates were controlled—power differences in favor of Welch–James as large as 60 percentage points! However, if sample sizes are smaller than values recommended to ensure robustness, we suggest users adopt the improved general approximation test. (The second author will provide upon request SAS/IML programs for obtaining numerical results.)

Finally, in addition to power benefits, Lix and Keselman (1995) presented a SAS/IML program that enables users to compute between-subjects and within-subjects Welch–James tests for omnibus as well as subeffect tests (e.g., contrasts among the repeated measures main and/or interaction means); users need only input the data, sample sizes, and a contrast matrix or vector to obtain numerical results.

H. J. KESELMAN
Department of Psychology
The University of Manitoba
Winnipeg, Manitoba
Canada R3T 2N2

James ALGINA
Department of Educational Psychology
University of Florida
Gainesville, FL 32611

Rhonda K. KOWALCHUK
Department of Educational Psychology
University of Wisconsin
Milwaukee, WI 53201

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Algina, J., and Keselman, H. J. (1998), "A Power Comparison of the Welch–James and Improved General Approximation Tests in the Split-Plot Design," *Journal of Educational and Behavioral Statistics*, 23, 152–169.

Dawson, K. S., Gennings, C., and Carter, W. H. (1997), "Two Graphical Techniques Useful in Detecting Correlation Structure in Repeated Measures Data," *The American Statistician*, 51, 275–283.

Huynh, H. (1978), "Some Approximate Tests for Repeated Measurement Designs," *Psychometrika*, 43, 161–175.

Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998), "A Comparison of Two Approaches for Selecting Covariance Structures in the Analysis of Repeated Measurements," *Communications in Statistics, Simulation, and Computation*, 27, 591–604.

——— (1999), "A Comparison of Recent Approaches to the Analysis of Repeated Measurements," *British Journal of Mathematical and Statistical Psychology*, 52, 63–78.

Keselman, H. J., Carriere, K. C., and Lix, L. M. (1993), "Testing Repeated Measures Hypotheses When Covariance Matrices are Heterogeneous," *Journal of Educational Statistics*, 18, 305–319.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D., (1996), *SAS System for Mixed Models*, Cary, NC: SAS Institute.

Lix, L. M., and Keselman, H. J. (1995), "Approximate Degrees of Freedom Tests: A Unified Perspective on Testing for Mean Equality," *Psychological Bulletin*, 117, 547–560.

SAS Institute (1996), *SAS/STAT Software: Changes and Enhancements Through Release 6.11*, Cary, NC: SAS Institute Inc.

HAMMER, H. (1997), "SYMBOLIC EXCLUSION IN STATISTICAL LITERATURE: THE IMPACT OF GENDERED LANGUAGE," *THE AMERICAN STATISTICIAN*, 51, 13–19: COMMENT BY HAMER AND JONES

We agree with Hammer (1997) that the use of male pronouns when referring to statisticians or scientists has become inappropriate. Thus, it is fair for her to cite examples of such usage in recent statistics books to sensitize today's authors to this issue.

Hammer errs, however, in citing "Thurstone (1965, xi)" as an example of "statisticians who chose to carry on in the tradition of excluding women . . . (d)uring the late 1960s and early 1970s." L. L. Thurstone was a psychologist rather than a statistician; he died in 1955. The sole citation to Thurstone in Hammer's reference list is to Thurstone (1947), *Multiple-Factor Analysis*. On page xi of his preface, Thurstone did refer to "the scientist," as "he." At the time Thurstone wrote, the pronoun "he" often was used to represent a person of unspecified gender. Today, most authors are more sensitive because they recognize the value of choosing gender-free pronouns. Unfortunately, when writing in English in the singular case, we still need the somewhat awkward "he or she."

The ease with which men are able to work for and with women, while not a perfect indicator of a freedom from sexist attitudes, nonetheless may be suggestive. About Thurstone, we note the following: He reached the mandatory retirement age of 65 at the University of Chicago in 1952. He considered offers from the University of California, Berkeley, the University of Washington, and the University of North Carolina (UNC). He moved with his wife to UNC because, of the three opportunities, this was the only one to offer professorial appointments both to him and to his wife, Thelma Gwinn Thurstone. He was recruited by the chair of the psychology department, Dorothy Adkins. His recruitment was strongly supported by Gertrude Cox, Director of the Institute of Statistics, which included faculty members from a department of mathematical statistics at UNC and an applied statistics department at North Carolina State University (NCSU). Thurstone's coauthor in a number of undertakings was his wife, for whose scholarship he expressed great respect.

In her critique of the use of gendered pronouns, it would have been better for Hammer to focus exclusively on current literature. It is particularly unfortunate that Thurstone (1947) was mistakenly characterized as a more recent publication.

Robert M. HAMER
Department of Psychiatry
UMDNJ Robert Wood Johnson Medical School
Piscataway, NJ 08854

Lyle V. JONES
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3270

REFERENCES

Hammer, H. (1997), "Symbolic Exclusion in Statistical Literature: The Impact of Gendered Language," *The American Statistician*, 51, 13–19.

Thurstone, L. L. (1947), *Multiple Factor-Analysis*, Chicago: The University of Chicago Press.

DRISCOLL, M. F. (1999), "AN IMPROVED RESULT RELATING QUADRATIC FORMS AND CHI-SQUARE DISTRIBUTIONS," *THE AMERICAN STATISTICIAN*, 53, 273–275: COMMENT BY HARVILLE AND RELPY

Having (like Professor Driscoll) made repeated attempts at devising proofs (for the standard results on the distribution of quadratic forms) that will be palatable to students, I have the following reactions.

1. *Sufficiency*. Sufficiency can be proved rather simply as follows. Let $\mathbf{A}$ represent a $p \times p$ symmetric matrix of rank r, and let $\mathbf{x} \sim N_p(\mu, \mathbf{V})$, where μ is a $p \times 1$ vector and $\mathbf{V}$ a $p \times p$ positive definite matrix. Observe, also, that there exists a nonsingular matrix $\mathbf{U}$ such that $\mathbf{V} = \mathbf{U}\mathbf{U}'$. Now, suppose that $\mathbf{AV}$ is idempotent, in which case

$$\mathbf{A} = \mathbf{AVV}^{-1} = \mathbf{AVAVV}^{-1} = \mathbf{AVA};$$

implying (since $\mathbf{AVA} = \mathbf{A}'\mathbf{VA}$) that $\mathbf{A}$ is nonnegative definite and hence that $\mathbf{U}'\mathbf{AU}$ is nonnegative definite, so that there exists a $p \times r$ matrix $\mathbf{R}$ such that $\mathbf{U}'\mathbf{AU} = \mathbf{RR}'$. Furthermore, let $\mathbf{L} = \mathbf{R}'\mathbf{U}^{-1}$, and define $\mathbf{z} = \mathbf{Lx}$. Then, observing that $\mathbf{A} = \mathbf{L}'\mathbf{L}$ and further that

$$\mathbf{LVL}' = (\mathbf{LL}')^{-1}\mathbf{LAVAL}'(\mathbf{LL}')^{-1} = (\mathbf{LL}')^{-1}\mathbf{LAL}'(\mathbf{LL}')^{-1} = \mathbf{I},$$

we find that $\mathbf{x}'\mathbf{Ax} = \mathbf{z}'\mathbf{z}$ and $\mathbf{z} \sim N_r(\mathbf{L}\mu, \mathbf{I})$. Assuming that we have defined the chi-square distribution to be the distribution of a sum of independent unit-variance normal random variables, which is perhaps the most common definition and the one I favor, or alternatively that we have derived the moment generating function (MGF) of such a sum—this is a rather elementary derivation—and have defined the chi-square distribution to be the distribution with that MGF, the proof is complete upon observing that $\mathbf{z}'\mathbf{z} \sim \chi^2(r, \gamma)$ (noncentral chi-square with degrees of freedom r and noncentrality parameter γ), where $\gamma = (\mathbf{L}\mu)'\mathbf{L}\mu = \mu'\mathbf{A}\mu$.

This proof is similar to one given by Hocking (1985, p. 28), except it avoids use of the spectral decomposition of $\mathbf{U}'\mathbf{AU}$—the very existence of the spectral decomposition is a deep result. It has significant advantages over the proof given by Driscoll in that it does not require knowledge of the general formula for the MGF of a quadratic form or of results on the eigenvalues of idempotent matrices; and depending on how we have defined the noncentral chi-square distribution, it avoids using the rather deep result that identity of MGF's implies identity of distributions.

2. *Necessity*. An appealing alternative to Driscoll's proof can be obtained by examining the behavior of the MGF at points away from the origin. Define $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{U}$ as in the proof of sufficiency, and observe that there exists a matrix $\mathbf{P}$ with r [= rank($\mathbf{A}$)] orthonormal columns and an $r \times r$ diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, with $\lambda_1 \geq \cdots \geq \lambda_r$, such that $\mathbf{U}'\mathbf{AU} = \mathbf{P}\Lambda\mathbf{P}'$ (the spectral decomposition of $\mathbf{U}'\mathbf{AU}$). Furthermore, let $\mathbf{K} = \mathbf{P}'\mathbf{U}^{-1}$, and define $\mathbf{z} = (z_1, \ldots, z_r)' = \mathbf{Kx}$ and $\omega = (\omega_1, \ldots, \omega_r)' = \mathbf{K}\mu$. Then, $\mathbf{x}'\mathbf{Ax} = \mathbf{z}'\Lambda\mathbf{z} = \sum_{i=1}^r \lambda_i z_i^2$, and $\mathbf{z} \sim N(\omega, \mathbf{I})$. And, letting I_1 represent the interval (ℓ, u), where $\ell = -\infty$ or $\ell = 1/(2\lambda_r)$, depending on whether $\lambda_r > 0$ or $\lambda_r < 0$, and similarly where $u = \infty$ or $u = 1/(2\lambda_1)$, depending on whether $\lambda_1 < 0$ or $\lambda_1 > 0$, the MGF, say $m_1(\bullet)$, of the distribution of $\sum_{i=1}^r \lambda_i z_i^2$ (and hence of $\mathbf{x}'\mathbf{Ax}$) is expressible as

$$m_1(t) = \left[\prod_{i=1}^r (1 - 2t\lambda_i)^{-1/2}\right] \times \exp\left\{-(1/2)\sum_{i=1}^r \omega_i^2 \left[1 - (1 - 2t\lambda_i)^{-1}\right]\right\} \qquad (t \in I_1),$$

as can easily be verified from first principles.

Now, suppose that $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2(d, \gamma)$ for some d and γ. And, letting I_2 represent the interval $(-\infty, 1/2)$, recall that the MGF, say $m_2(\bullet)$, of $\chi^2(d, \gamma)$ is expressible as

$$m_2(t) = (1-2t)^{-d/2} \exp\{-(1/2)\gamma[1-(1-2t)^{-1}]\} \quad (t \in I_2).$$

Setting $[m_2(t)]^{-2} = [m_1(t)]^{-2}$ gives (for $t \in I_1 \cap I_2$)

$$(1-2t)^d \exp\{\gamma[1-(1-2t)^{-1}]\} = \left[\prod_{i=1}^{r}(1-2t\lambda_i)\right] \exp\left\{\sum_{i=1}^{r} \omega_i^2[1-(1-2t\lambda_i)^{-1}]\right\}. \quad (1)$$

Let c represent the number of elements in the set $\{\lambda_1, \ldots, \lambda_r\}$ that equal one. The remainder of the proof consists of using equality (1) to show that $c = r$—implying (since $\lambda_1, \ldots, \lambda_r$ are the nonzero eigenvalues of $\mathbf{U}'\mathbf{A}\mathbf{U}$) that $\mathbf{U}'\mathbf{A}\mathbf{U}$ is idempotent and hence that $(\mathbf{A}\mathbf{V})^2 = (\mathbf{U}')^{-1}(\mathbf{U}'\mathbf{A}\mathbf{U})^2\mathbf{U}' = (\mathbf{U}')^{-1}\mathbf{U}'\mathbf{A}\mathbf{U}\mathbf{U}' = \mathbf{A}\mathbf{V}$ (i.e., $\mathbf{A}\mathbf{V}$ is idempotent). We have that $0 < \lambda_i \leq 1$ (for $i = 1, \ldots, r$)—since if $\lambda_i < 0$ for some i, the limit (as $t \downarrow \ell$) of the right side of equality (1) is 0 while that of the left side is nonzero; and similarly if $\lambda_i > 1$ for some i (in which case $u < 1/2$), the limit (as $t \uparrow u$) of the right side is 0 while that of the left side is nonzero—with the implication that $I_1 \cap I_2 = I_2 = (-\infty, 1/2)$. Furthermore, $d = c$ and $\gamma = \sum_{i=1}^{c} \omega_i^2$. To see this, observe that (for $t < 1/2$)

$$(1-2t)^{d-c} \exp\left\{\left(\gamma - \sum_{i=1}^{c}\omega_i^2\right)[1-(1-2t)^{-1}]\right\} = \left[\prod_{i=c+1}^{r}(1-2t\lambda_i)\right] \exp\left\{\sum_{i=c+1}^{r} \omega_i^2[1-(1-2t\lambda_i)^{-1}]\right\},$$

(where if $c = r$, the right side of this equality is to be interpreted as 1) and that if $d \neq c$ or $\gamma \neq \sum_{i=1}^{c} \omega_i^2$, the limit (as $t \uparrow 1/2$) of the left side of this equality is 0 or infinity, while the limit of the r.h.s. is nonzero and finite. And, upon setting $d = c$ and $\gamma = \sum_{i=1}^{c} \omega_i^2$, we obtain (for $t < 1/2$) the equality

$$1 = \left[\prod_{i=c+1}^{r}(1-2t\lambda_i)\right] \exp\left\{\sum_{i=c+1}^{r} \omega_i^2[1-(1-2t\lambda_i)^{-1}]\right\},$$

from which it follows that $c = r$—since if $c < r$ the limit (as $t \uparrow 1/2$) of the right side of this equality is less than 1.

Not only is this proof relatively straightforward, it does not assume familiarity with cumulants or with the general formula for the MGF of the distribution of a quadratic form.

3. *Laha's lemma*. In previous (coauthored) articles, Driscoll has discussed a "lemma" (on polynomials) stated (but not proved) by Laha (1956, p. 791), and has considered its use in establishing the necessity of the necessary and sufficient conditions for the independence of quadratic forms. As he indicated in his 1999 article, this lemma can also be used to considerable advantage in establishing the necessity of the conditions under which a quadratic form has a noncentral chi-square distribution. In particular, if we make use of Laha's lemma, we have as an almost immediate consequence of equality (1) that

$$(1-2t)^d = \prod_{i=1}^{r}(1-2t\lambda_i). \quad (2)$$

This equality holds for all t in the nondegenerate interval $I_1 \cap I_2$ and hence (since both sides are polynomials) for all t. Then, upon observing that $\lambda_1 = \cdots = \lambda_r = 1$ [since if $\lambda_i \neq 1$ for some i, the right side of equality (2) has a root at $t = 1/(2\lambda_i)$, while the left side does not], the proof is essentially complete.

Should we use Laha's lemma or not? That (in Driscoll's words) "this is a deep result whose proof is difficult" is a compelling reason to avoid its use. However, a case can also be made for taking advantage of this lemma. Harville and Kempthorne (1997) gave a proof of the one-variable version of the lemma—which is all that is really needed in its application to the distribution of quadratic forms—that while somewhat tedious and not especially intuitive, is accessible even to those with no knowledge of the theory of functions of complex variables. Moreover, if the lemma is to be invoked in establishing results on independence (as is commonly done), there would seem to be less reason to resist its use in establishing results on "chisquareness."

4. *Terminology*. Driscoll adheres to Johnson, Kotz, and Balakrishnan's (1995, p. 433) usage of the term noncentrality parameter, and seems to regard Searle (1971, p. 49, and 1987, p. 228) as something of a maverick in asserting that he "defined the noncentrality parameter as one-half the noncentrality parameter used by others." The reality is that Searle has considerable company in his usage of this term; company that includes Christensen (1996, p. 406), Hocking (1985, p. 22), and Myers and Milton (1991, p. 58). Presumably, the dual usage arose and persists because which usage is most convenient depends on the circumstances. In any case, it has caused more confusion than can possibly be justified by any conceivable advantage. I suggest that we settle on a single usage. I do not think it much matters which one, but for definiteness I suggest that we follow Driscoll and conform to Johnson, Kotz, and Balakrishnan's usage.

David A. HARVILLE
Mathematical Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598–0218

REFERENCES

Christensen, R. (1996), *Plane Answers to Complex Questions: The Theory of Linear Models* (2nd ed.), New York: Springer-Verlag.

Harville, D. A., and Kempthorne, O. (1997), "An Alternative Way to Establish the Necessity Part of the Classical Result on the Statistical Independence of Quadratic Forms," *Linear Algebra and its Applications*, 264, 205–215.

Hocking, R. R. (1985), *The Analysis of Linear Models*, Monterey, CA: Brooks/Cole.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions* (2nd ed.), New York: Wiley.

Laha, R. G. (1956), "On the Stochastic Independence of Two Second-Degree Polynomial Statistics in Normally Distributed Variates," *Annals of Mathematical Statistics*, 27, 790–796.

Myers, R. H., and Milton, J. S. (1991), *A First Course in the Theory of Linear Statistical Models*, Boston: PWS-KENT.

Searle, S. R. (1971), *Linear Models*, New York: Wiley.

——— (1987), *Linear Models for Unbalanced Data*, New York: Wiley.

RESPONSE

I appreciate Professor Harville's pertinent and thorough comments. I would agree that a given presentation of results about independence and distribution of quadratic forms in normal variates should be based either on Laha's lemma or on the cumulant arguments that I have advanced (with others), or on the approach that Professor Harville has demonstrated here.

Michael F. DRISCOLL
Department of Mathematics
Arizona State University
Tempe, AZ 85287-1804

AHMAD, I. (1996), "A CLASS OF MANN–WHITNEY–WILCOXON TYPE STATISTICS," *THE AMERICAN STATISTICIAN*, 50, 324–327: COMMENT BY ADAMS, ADAMS, CHANG, ETZEL, KUO, MONTEMAYOR, AND SCHUCANY; AND REPLY

As a team project, our Spring 1997 class in nonparametric theory read with interest the note by Ahmad (1996). We recognized that the consistency class for the Mann–Whitney–Wilcoxon (MWW) test is succinctly specified in terms of the functional $\int F(x)dG(x)$. The notion of comparing the maxima of subsets is an attractive one, because their distributions are powers of the respective cdfs. Thus, it is a reasonable extension to consider $\delta^{r,s} = \int F^r(x)dG^s(x)$. It seems to be a nice idea to investigate the existence of tests that are more efficient than $r = s = 1$.

However, the article should not equate the von Mises statistic (from plugging in the empiricals for $F(x)$ and $G(y)$) with the corresponding U statistic. These V statistics are generally not equal to the related U statistics unless the degree(s) of the kernel is (are) 1 (Sen and Singer 1993, pp. 210–211. This appeared to be for ease of motivation, but not a necessary step in the author's development. It is quite satisfactory and pedagogically sound to go directly to the U statistic (2.4) as the "empirical" version.

There are some minor errors in the details in the examples. Theorem 2.1 should actually require $m/(m+n) \to \lambda$. The author suppressed the factor of $\lambda(1-\lambda)$ in all of the expressions for PAE(r,s) in Section 2.2. This is inconsequential provided that the ARE comparisons are to be made only among various values of r and s. In the normal example the familiar $3/\pi$ is the ARE relative to the Student t and not PAE(1,1) as stated.

There are some pleasant surprises in the results. In particular the MWW is best in this large collection of tests for all of the two-sample examples except the uniform. In the numerical example the author evidently again "modified" the statistic to agree with the usual MWW counts, $W = 48$. We were unable to confirm the value of 49.5. It was disconcerting that the pattern of superiority of the MWW did not appear to hold true for the logistic in the one-sample case. Indeed, the *correct* expression for the derivative there is $2r/(r+1)(r+2)$. Squaring these and dividing by the null variance, the correct PAE is *not* increasing in r and the first few values are 1/3, 1/3, .3316, and .3203. Therefore, the patterns are in fact the same for the corresponding one-sample examples.

Bryan E. ADAMS

Amber G. ADAMS

Hui-Min CHANG

Carol ETZEL

Jo-Kang KUO

Jorge MONTEMAYOR

William R. SCHUCANY

Statistical Science Department
Southern Methodist University
PO Box 750332
Dallas, TX 75275

REFERENCES

Sen and Singer (1993), *Large Sample Methods in Statistics*, New York: Chapman Hall.

RESPONSE

The comments of Professor Schucany and his students center around three points. First, yes, a U statistic and a corresponding V statistic need not be identical. To answer their comment here, Equation (2.4) has some missing parts. It should read

$$\delta_{m,n}^{(r,s)} = \int_{-\infty}^{\infty} F_m^r(x) dG_n^s(x) = (m^r n^s)^{-1} \sum_{i_1=1}^{m} \cdots \sum_{i_r=1}^{m} \sum_{j_1=1}^{n} \cdots \sum_{j_s=1}^{n} I\left(\max\left(X_{i_1},\ldots,X_{i_r}\right) < \max\left(Y_{j_1},\ldots,Y_{j_s}\right)\right),$$

which is asymptotically equivalent to the following U statistic

$$U_{m,n}^{(r,s)} = \left[\binom{m}{r}\binom{n}{s}\right]^{-1} \sum_{C} I\left(\max\left(X_{i_1},\ldots,X_{i_r}\right) < \max\left(Y_{j_1},\ldots,Y_{j_s}\right)\right), \qquad (2.6)$$

Second, in the statement of Theorem 2.1, the ratio (m/n) should read $(m/(m+n))$, this is just a misprint.

Third, they are correct in their derivation of the PAE for the logistic distributions, but their last two numeric values should be .3 and .2667 instead of .3316 and .3203.