have $\rho_{12} = 1$. A similar argument gives $\rho_{12} = -1$ when $\sigma_S^2 = 0$. Thus, since we required that $-1 < \rho_{12} < 1$, σ_S^2 and σ_D^2 will never equal 0 in our applications.

To extend Pitman's approach to obtain a simultaneous test for the means and variances, consider the expectation of D given S. Since X_1 and X_2 are bivariate normal, it follows that D and S are also bivariate normal. So, noting that $E(S) = \mu_1 + \mu_2$ and $E(D) = \mu_1 - \mu_2$, from any standard mathematical statistics textbook (e.g., Hogg and Craig 1978, p. 118), we have

$$\begin{aligned} E(D \mid S) &= (\mu_1 - \mu_2) + \rho_{SD}(\sigma_D/\sigma_S)(S - (\mu_1 + \mu_2)) \\ &= (\mu_1 - \mu_2) + [(\sigma_1^2 - \sigma_2^2)/\sigma_S^2][S - (\mu_1 + \mu_2)] \\ &= \beta_0 + \beta_1 S, \end{aligned}$$

where

$$\beta_0 = (\mu_1 - \mu_2) - [(\sigma_1^2 - \sigma_2^2)/\sigma_S^2](\mu_1 + \mu_2)$$

and

$$\beta_1 = (\sigma_1^2 - \sigma_2^2)/\sigma_S^2.$$

Now, $\sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ iff $\beta_0 = \beta_1 = 0$, so a simultaneous test of the equivalence of the means and variances of X_1 and X_2 is an F test calculable from standard results from the regression of D on S. The test statistic is

$$F = [(\Sigma d_i^2 - \text{SSE})/2]/[\text{SSE}/(n - 2)],$$

where n is the number of pairs of data observed, Σd_i^2 is the sum of the squares of the n observed differences, and SSE is the residual error sum of squares from the regression of D on S. For an α level test, reject the null hypothesis of equal means and variances if F exceeds the upper α point of an $F(2, n - 2)$ distribution.

[Received May 1988. Revised January 1989.]

REFERENCES

Hogg, R. V., and Craig, A. T. (1978), *Introduction to Mathematical Statistics* (4th ed.), New York: Macmillan.

Pitman, E. G. (1939), "A Note on Normal Correlation," *Biometrika*, 31, 9–12.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods* (7th ed.), Ames: Iowa State University Press.

Using U Statistics to Derive the Asymptotic Distribution of Fisher's Z Statistic

D. L. HAWKINS*

A simple derivation of the asymptotic distribution of Fisher's Z statistic for general bivariate parent distributions F is obtained using U-statistic theory. This method easily reveals that the asymptotic variance of Z generally depends on the correlation ρ and on certain moments of F. It also reveals the particular structure of F that makes the asymptotic variance of Z independent of ρ, and shows that there are many distributions F with this property. The bivariate normal is only one such F.

KEY WORDS: Central limit theorem; Delta method.

1. INTRODUCTION

Let $\mathbf{X}_i^T = (X_{1i}, X_{2i})$, $i = 1, \ldots, n$, be iid random vectors with cdf F, mean $\boldsymbol{\mu}^T = (\mu_1, \mu_2)$, variances $\sigma_1^2 = \text{var}(X_{1i})$, $\sigma_2^2 = \text{var}(X_{2i})$, and correlation ρ. Let r_n be the usual sample correlation coefficient and $Z_n = f(r_n) = \frac{1}{2} \ln((1 + r_n)/(1 - r_n))$ be Fisher's Z statistic. It is well known (see, e.g., Anderson 1984) that when F is bivariate normal (BVN), Z_n is asymptotically normal with mean $f(\rho)$ and variance independent of ρ. What is perhaps less well known is that if F is not BVN, Z_n is still asymptotically normal with the same mean, but the asymptotic variance may depend on ρ and on certain moments of F. This may be seen, for example, in the work of Gayen (1951), who obtained the exact pdf of Z_n for finite n when F is a bivariate Type A Edgeworth distribution.

The purpose of this note is to illustrate how U-statistic theory may be used to simply obtain the limit distribution of Z_n for *any* F with finite fourth moments, in a form that clearly reveals how this distribution depends on F. The derivation of the main result might make a good exercise in a large-sample theory class.

2. DERIVATION OF THE LIMIT DISTRIBUTION

The idea is to express Z_n as a function of a U statistic and then to use the delta method along with the central limit theorem for U statistics. Let

$$\mathbf{A}_n = \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}}_n)(\mathbf{X}_j - \overline{\mathbf{X}}_n)^T = \begin{bmatrix} a_{11}(n) & a_{12}(n) \\ a_{12}(n) & a_{22}(n) \end{bmatrix},$$

where $\overline{\mathbf{X}}_n = 1/n \sum_{j=1}^{n} \mathbf{X}_j$ and $\mathbf{U}_n = (n - 1)^{-1}[a_{11}(n), a_{22}(n), a_{12}(n)]^T$. Then $Z_n = \eta(\mathbf{U}_n)$, where, for $\mathbf{u} = (u_1, u_2, u_3)^T$, $\eta(\mathbf{u}) = f(g(\mathbf{u}))$, $g(\mathbf{u}) = u_3/(u_1 u_2)^{1/2}$, and $f(x) = \frac{1}{2} \ln[(1 + x)/(1 - x)]$.

*D. L. Hawkins is Assistant Professor, Department of Mathematics, University of Texas, Arlington, TX 76019.

Next observe that $\mathbf{U}_n$ is a (vector-valued) U statistic. Let

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

denote the covariance matrix of $\mathbf{X}_j$. Consider the (matrix-valued) kernel $h(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{X}_1\mathbf{X}_1^T - \mathbf{X}_1\mathbf{X}_2^T$. Then $E[h(\mathbf{X}_1, \mathbf{X}_2)] = \mathbf{\Sigma}$, and the symmetrized version of h is $h^*(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2}[h(\mathbf{X}_1, \mathbf{X}_2) + h(\mathbf{X}_2, \mathbf{X}_1)] = \frac{1}{2}(\mathbf{X}_1 - \mathbf{X}_2)(\mathbf{X}_1 - \mathbf{X}_2)^T$. Finally, observe that

$$(n - 1)^{-1}\mathbf{A}_n = \binom{n}{2}^{-1} \sum_{1 \le \beta_1 < \beta_2 \le n} h^*(\mathbf{X}_{\beta_1}, \mathbf{X}_{\beta_2}),$$

so $(n-1)^{-1}\mathbf{A}_n$ is a (matrix-valued) U statistic for $\mathbf{\Sigma}$. If we now let ϕ be the linear map taking a symmetric 2×2 matrix with elements b_{11}, b_{12}, b_{22} to the vector $(b_{11}, b_{22}, b_{12})^T$ in $\mathbf{R}^3$, then it follows that

$$\mathbf{U}_n = \binom{n}{2}^{-1} \sum_{1 \le \beta_1 < \beta_2 \le n} \phi h^*(\mathbf{X}_{\beta_1}, \mathbf{X}_{\beta_2})$$

is a U statistic for the parameter $\phi\mathbf{\Sigma} = (\sigma_1^2, \sigma_2^2, \rho\sigma_1\sigma_2)^T$ corresponding to the symmetric kernel ϕh^*, which has degree 2.

Following standard U-statistic theory, define the projection function, for $\mathbf{x} = (x_1, x_2)^T$, as

$$\mathbf{p}(\mathbf{x}) = E\{\phi h^*(\mathbf{x}, \mathbf{X}_1)\} = \frac{1}{2} E\left\{\begin{bmatrix} (x_1 - X_{11})^2 \\ (x_2 - X_{21})^2 \\ (x_1 - X_{11})(x_2 - X_{21}) \end{bmatrix}\right\}, \quad (2.1)$$

and the projections as $\xi_j = \mathbf{p}(\mathbf{X}_j) - \phi\mathbf{\Sigma}$. The ξ_j's are iid with mean zero and variance $\mathbf{\Gamma}_1 = E(\xi_j \xi_j^T)$ (the existence of which requires finite fourth moments). Then by a generalization of the standard central limit theorem for scalar U statistics (see, e.g., Randles and Wolfe 1979, chap. 3), we have the following:

Lemma. Assume that F has finite fourth moments. Then $n^{1/2}(\mathbf{U}_n - \phi\mathbf{\Sigma}) \xrightarrow{\mathcal{L}} N_3(\mathbf{0}, 4\mathbf{\Gamma}_1)$, as $n \to \infty$.

The lemma and the delta method (see, e.g., Bishop, Fienberg, and Holland 1975, chap. 14) then imply, since $Z_n = \eta(\mathbf{U}_n)$ and η is differentiable, the following:

Theorem. Assume that F has finite fourth moments. Then as $n \to \infty$, $\sqrt{n}[Z_n - \eta(\phi\Sigma)] \xrightarrow{\mathcal{L}} N(0, \tau_F^2)$, where $\tau_F^2 = 4 \cdot \mathbf{\nabla}^T\eta(\phi\Sigma)\mathbf{\Gamma}_1\mathbf{\nabla}\eta(\phi\Sigma)$ and $\mathbf{\nabla}\eta$ denotes the gradient of η.

3. DEPENDENCE OF THE LIMIT DISTRIBUTION ON F

It is clear that the asymptotic mean, $\eta(\phi\Sigma) = f(\rho)$, does not depend on F. We now compute τ_F^2 and find that it varies with F, in general. We have $\mathbf{\nabla}^T\eta(\mathbf{u}) = [1 - g^2(\mathbf{u})]^{-1}[-\frac{1}{2} u_3 u_1^{-3/2} u_2^{-1/2}, -\frac{1}{2} u_3 u_1^{-1/2} u_2^{-3/2}, u_1^{-1/2} u_2^{-1/2}]$, and hence $\mathbf{\nabla}^T\eta(\phi\Sigma) = (1 - \rho^2)^{-1}[-\rho/(2\sigma_1^2), -\rho/(2\sigma_2^2), 1/(\sigma_1\sigma_2)]$. To get $\mathbf{\Gamma}_1$, we need $\mathbf{p}(\mathbf{x})$, which is simplified if we note that Z_n is location and scale invariant, so we may assume that $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$ with no loss in generality. Hence we obtain, for $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{p}(\mathbf{x}) = \frac{1}{2}[x_1^2 + 1, x_2^2 + 1, x_1x_2 + \rho]^T$, so $\xi_j = \mathbf{p}(\mathbf{X}_j) - \phi\mathbf{\Sigma} = \frac{1}{2}[X_{1j}^2 - 1, X_{2j}^2 - 1, X_{1j}X_{2j} - \rho]^T$. Writing $m_{rs} = m_{rs}(F) = E_F(X_{1j}^r X_{2j}^s)$, one obtains

$$\mathbf{\Gamma}_1 = E_F(\xi_j\xi_j^T) = \frac{1}{4}\begin{bmatrix} m_{40} - 1 & m_{22} - 1 & m_{31} - \rho \\ & m_{04} - 1 & m_{13} - \rho \\ \text{symmetric} & & m_{22} - \rho^2 \end{bmatrix}.$$

Combining these results, we obtain

$$\tau_F^2 = (1 - \rho^2)^{-2}\, \tfrac{1}{4}\{(m_{40} + 2m_{22} + m_{04})\rho^2 - 4(m_{31} + m_{13})\rho + 4m_{22}\}. \quad (3.1)$$

This result may be compared with Gayen's (1951) finite-n approximation noted in the introduction, and τ_F^2 will be found to be the limit (as $n \to \infty$) of n times his variance approximation. Gayen's result, however, though valid for all finite n, applies only to the Edgeworth distributions; (3.1) holds for any F with finite fourth moments, but only as $n \to \infty$.

Several features of (3.1) are notable. First, the asymptotic variance τ_F^2 of Z_n depends on F only through the moments m_{rs} appearing in (3.1). Second, generally, τ_F^2 depends on ρ. Third, if F is BVN, we have (see, e.g., Kendall and Stuart 1979, vol. 1, p. 91) $m_{22} = 1 + 2\rho^2$, $m_{40} = m_{04} = 3$, and $m_{31} = m_{13} = 3\rho$, giving $\tau_F^2 = 1$ regardless of ρ, agreeing with the classical result.

From these observations we realize that there are many distributions F—those in the set $\mathcal{F}_N$, say, with moments m_{rs} in (3.1) matching those of the BVN distribution with 0 means and unit variances—under which τ_F^2 equals 1. Such a distribution F may be constructed by the following method, due to Professor George Woodworth. Let W_1 and W_2 be independent random variables with (possibly different) distributions G_1 and G_2 satisfying $E_{G_i}W_i = 0$, $E_{G_i}W_i^2 = 1$, $E_{G_i}W_i^3 = 0$, $E_{G_i}W_i^4 = 3$ $(i = 1, 2)$. Let $\mathbf{\Sigma}$ be defined so that $\phi\mathbf{\Sigma} = (1, 1, \rho)$, and let $X_1 = c_{11}W_1 + c_{12}W_2$ and $X_2 = c_{21}W_1 + c_{22}W_2$, where $c_{11} = c_{22} = [(1 + \rho)^{1/2} + (1 - \rho)^{1/2}]/2$ and $c_{12} = c_{21} = [(1 + \rho)^{1/2} - (1 - \rho)^{1/2}]/2$. Then it may be checked that X_1 and X_2 have the same product moments m_{rs} $(0 \le r + s \le 4)$ as the BVN distribution with mean 0 and covariance $\mathbf{\Sigma}$. In particular, the moments m_{22}, m_{40}, m_{04}, m_{31}, and m_{13} of X_1 and X_2 match those of this BVN distribution, as required. To obtain an explicit example of F we must produce a G_1 and G_2 with the required moment structure. A simple example is obtained if we take $G_1 = G_2$ and let G_1 be a discrete distribution that places masses p_b, p_a, p_a, and p_b on the real numbers $-b$, $-a$, a, and b, where $b > a > 0$. G_1 is obviously symmetric about 0, so it has mean and third moment 0, and the numbers a, b, p_a, and p_b can be chosen to give variance 1 and fourth moment 3 (e.g., taking $a = \frac{1}{2}$, $b = 1.913$, $p_a = .39$, $b_b = .11$ will do this, as follows from setting up and solving obvious equations for the required conditions).

Interesting problems would be to characterize $\mathcal{F}_N$ in a more lucid manner, to devise tests for $F \in \mathcal{F}_N$ versus $F \notin \mathcal{F}_N$, or to determine if there exist $F \notin \mathcal{F}_N$ such that τ_F^2 is independent of F and ρ.

[*Received November 1988. Revised May 1989.*]

REFERENCES

Anderson, T. W. (1984), *An Introduction to Multivariate Analysis* (2nd ed.), New York: John Wiley.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA: MIT Press.

Gayen, A. K. (1951), "The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn From Non-normal Universes," *Biometrika*, 38, 219–247.

Kendall, M. G., and Stuart, A. (1979), *The Advanced Theory of Statistics* (4th ed.), New York: Macmillan.

Randles, R. H., and Wolfe, D. A. (1979), *Introduction to the Theory of Nonparametric Statistics*, New York: John Wiley.

An Empirical Nonlinear Data-Fitting Approach for Transforming Data to Normality

LAWRENCE I-KUEI LIN and EDWARD F. VONESH*

A general method is proposed by which nonnormally distributed data can be transformed to achieve approximate normality. The method uses an empirical nonlinear data-fitting approach and can be applied to a broad class of transformations including the Box–Cox, arcsine, generalized logit, and Weibull-type transformations. It is easy to implement using standard statistical software packages. Several examples are provided.

KEY WORDS: Iteratively reweighted least squares; Nonlinear model; Normal scores; Order statistics.

1. INTRODUCTION

Suppose that $U_1, U_2, \ldots, U_n$ constitute a random sample from a probability density function that is not necessarily normal. We assume there exists an invertible and strictly monotone transformation, $Y_i = g_{\boldsymbol\theta}(U_i)$, such that Y_i is normally distributed with mean μ and variance σ^2. For example, letting $\boldsymbol\theta' = (\lambda, c)$, the power and logarithmic transformations are encompassed by the Box–Cox transformation (Box and Cox 1964):

$$g_{\boldsymbol\theta}(U) = \frac{(U + c)^{\lambda} - 1}{\lambda} \quad \text{if } \lambda \neq 0$$

$$= \log_e (U + c) \quad \text{if } \lambda = 0. \qquad (1.1)$$

When the data have a positive skewness, a common choice for $g_{\boldsymbol\theta}(U)$ is $\log_e(U + c)$; for data with a negative skewness, a reasonable choice for $g_{\boldsymbol\theta}(U)$ is $(U^{\lambda} - 1)/\lambda$ or, more simply, U^{λ}.

Various methods have been proposed for estimating the values of $\boldsymbol\theta' = (\lambda, c)$. Box and Cox (1964) suggested using maximum likelihood (ML) techniques or a Bayesian approach to estimate λ and/or c. All subsequent analyses and inference would then be conditioned on the estimate $\hat{\boldsymbol\theta}$. Carroll and Ruppert (1984) introduced the transformation on both sides of a general nonlinear model and used ML techniques for estimating $\boldsymbol\theta$. Their approach allows $\hat{\boldsymbol\theta}$ to be random rather than fixed with respect to inference on estimated model parameters. Berry (1987) suggested using estimates of c, for the transformation $\log_e(U + c)$, by minimizing, with respect to the residuals $Y - \hat{\mu}$, both skewness and kurtosis. Subsequent analysis would then be based on fixing $c = \hat{c}$.

Our goal is to find a $\boldsymbol\theta$ such that we may transform the data "as close to" normality as possible given any appropriate monotone $g_{\boldsymbol\theta}(U)$. Here, "as close to" reflects the use of nonlinear least squares as the method by which we estimate $\boldsymbol\theta$. In other words, our criterion is to achieve a straight line on a normal probability plot of the residuals. We can use any software that has a nonlinear regression program to do such an evaluation. Our approach has an advantage over the Box–Cox method in that our method does not restrict the form of the transformation to (1.1). Other popular transformations, such as the generalized logit, arcsine, inverse hyperbolic sine, and Weibull, may also be performed. In addition, this method has a natural extension and is easily implemented in applications involving general linear or nonlinear models.

2. METHODS FOR TRANSFORMING DATA

2.1 The Single-Sample Case

Let $U_{(1)}, \ldots, U_{(n)}$ be the order statistics of $U_1, \ldots, U_n$. We can write

$$\sigma^{-1}(Y_{(i)} - \mu) = -\sigma^{-1}\mu + \sigma^{-1}g_{\boldsymbol\theta}(U_{(i)}),$$

where the $Y_{(i)}$ denote the order statistics of Y_i ($i = 1, \ldots, n$). As indicated by Blom (1958) and Tukey (1962), the expected value of the normal score $\sigma^{-1}(Y_{(i)} - \mu)$ can be approximated by the Z score

$$Z_{(i)} = \psi[(r_i - 3/8)/(n + 1/4)], \qquad (2.1)$$

*Lawrence I-kuei Lin and Edward F. Vonesh are Senior Research Statisticians, Statistical Services, Baxter Healthcare Corporation, Round Lake, IL 60073.

The authors express their sincere gratitude to the editor, associate editor, and referee for their valuable comments and criticisms, to Laurene Strauch and Beverly Adelphia for their assistance in preparing the manuscript, and to their fellow statisticians at Baxter Healthcare Corporation for their support and comments.