Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality

Author(s): Steve Verrill and Richard A. Johnson

Source: *Journal of the American Statistical Association*, Vol. 83, No. 404 (Dec., 1988), pp. 1192-1197
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2290156

Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality

STEVE VERRILL and RICHARD A. JOHNSON*

Plotting order statistics versus some variant of the normal scores is a standard graphical technique for assessing the assumption of normality. To obtain an objective evaluation of the normal assumption, it is customary to calculate the correlation coefficient associated with this plot. The Shapiro–Francia statistic is the square of the correlation between the observed order statistics and the expected values of standard normal order statistics, whereas the Shapiro–Wilk statistic also involves the covariances of the standard normal order statistics. In a wide variety of applications, an investigation of the plausibility of the normal (or lognormal) model is needed when the observations on strength or life length are right-censored. The plotting procedure still applies if the observations are censored at a fixed order statistic or a fixed time. Here, the corresponding distribution theory for some modified versions of the Shapiro–Wilk correlation statistic is investigated. Because the asymptotic theory used in this article shows a surprisingly slow rate of convergence even for complete samples, a table of critical values based on a Monte Carlo study is provided. Results from an empirical power study are also presented. Finally, large-sample critical values are obtained and compared with the Monte Carlo values.

KEY WORDS: Asymptotic distributions; Modified Shapiro–Wilk statistics; Normal probability plot; Small-sample power.

1. INTRODUCTION

In numerous applications, including lumber testing and component life-length analysis, it is crucial to evaluate alternative parametric models on the basis of right-censored data. Because the most widely applied check for normality requires calculating some version of the Shapiro–Wilk statistic, we focus on the sampling behavior of its extension to Type I and Type II censored data. Our primary contributions, described in Sections 4 and 5, concern the determination of the limit distribution and its percentiles. Computer calculations reveal a surprisingly slow rate of convergence to the asymptotic distribution. Consequently, in Section 2 we provide approximate critical values based on extensive Monte Carlo studies. These are augmented with a power study that includes the most popular competitors of the censored-data Shapiro–Wilk statistic.

The Shapiro–Wilk (1965) test-of-normality statistic, W, is essentially a ratio of two estimates of scale. In particular,

$$W = (\mathbf{m}'\mathbf{V}^{-1}\mathbf{X})^2/(\mathbf{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m}) \sum_{i=1}^{n} (X_{in} - \overline{X})^2, \quad (1.1)$$

where X_{in} is the ith order statistic from some sample, and $\mathbf{m}$ is the expectation vector and $\mathbf{V}$ is the covariance matrix of the order statistics from a sample of n standard normal random variables. Shapiro, Wilk, and Chen (1968) and Stephens (1974) performed Monte Carlo studies that established that the Shapiro–Wilk statistic yields a powerful omnibus test of normality.

Shapiro and Francia (1972) introduced the modified W statistic

$$W' = (\mathbf{m}'\mathbf{X})^2/(\mathbf{m}'\mathbf{m}) \sum_{i=1}^{n} (X_{in} - \overline{X})^2, \quad (1.2)$$

which is easier to calculate for large samples. Their Monte Carlo power studies indicated that W' has power properties similar to those of W. Sarkadi (1975) established the consistency of the test of normality based on W'.

Filliben (1975) and Ryan and Joiner (1973) noted that the Shapiro–Francia statistic could be written as

$$W' = \frac{[\sum_{i=1}^{n} (X_{in} - \overline{X})(m_{in} - \overline{m})]^2}{\sum_{i=1}^{n} (X_{in} - \overline{X})^2 \sum_{i=1}^{n} (m_{in} - \overline{m})^2}. \quad (1.3)$$

Thus W' is the square of the correlation coefficient associated with a normal probability plot in which the sample order statistics are plotted against the expected values of the standard normal order statistics.

In an effort to reduce computation costs even further, various authors (e.g., Filliben 1975; Ryan and Joiner 1973; Weisberg and Bingham 1975) suggested replacing the m_{in} by other normal scores. For example, replacing m_{in} with $H_{in} \equiv H(i/(n + 1))$, where H is the inverse of the standard normal distribution function, we obtain the statistic

$$r(\mathbf{X}, \mathbf{H}) = \frac{\sum_{i=1}^{n} (X_{in} - \overline{X})(H_{in} - \overline{H})}{(\sum_{i=1}^{n} (X_{in} - \overline{X})^2 \sum_{i=1}^{n} (H_{in} - \overline{H})^2)^{1/2}}. \quad (1.4)$$

DeWet and Venter (1972, 1973) obtained the asymptotic distribution of $r(\mathbf{X}, \mathbf{H})$. Leslie, Stephens, and Fotopoulos (1986) showed that $r^2(\mathbf{X}, \mathbf{H})$ and the original Shapiro–Wilk statistic are asymptotically equivalent. Verrill and Johnson (1987a) demonstrated that $r(\mathbf{X}, \mathbf{H})$ and the cor-

* Steve Verrill is Statistics and Probability Group Leader, Computing and Mathematics Research Division, Lawrence Livermore National Laboratory, Livermore, CA 94550. Richard A. Johnson is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. Verrill's work was supported by U.S. Department of Energy Contract W-7405-Eng-48. Johnson's research was supported by U.S. Office of Naval Research Grant N00014-78-C-0722.

Journal of the American Statistical Association
December 1988, Vol. 83, No. 404, Theory and Methods

relation statistics based on the standard versions of the normal scores (including the square root of W') are asymptotically equivalent.

In this article our primary results concern the extension of DeWet and Venter's (1972, 1973) ideas to the censored-data situation.

Consider data that are censored so that only the first k order statistics $X_{1n}, \ldots, X_{kn}$ are available. With censored data, graphical methods continue to play a dominant role in the assessment of normality. The normal-scores plot of X_{in} versus H_{in} $(i \le k)$ is judged for departures from linearity. One objective measure of straightness is given by the value of the correlation coefficient. This correlation coefficient has an interpretation as the cosine of the angle between $\mathbf{X}_{(k)} = [X_{1n} - \overline{X}_k, \ldots, X_{kn} - \overline{X}_k]'$ and $\mathbf{H}_c = [H_{1n} - \overline{H}_k, \ldots, H_{kn} - \overline{H}_k]'$, and also as the accompanying degree of fit to the order statistics by their projection on the centered scores $\mathbf{H}_c$.

Although Ryan and Joiner (1973) mentioned that $r(\mathbf{X}, \mathbf{H})$ has an obvious and immediate extension to censored data, it remained for Smith and Bain (1976) to give the explicit form. To set notation, we write the Type II censored-data correlation statistic as

$$r(\mathbf{X}, \mathbf{H}, \delta) = \frac{\sum_{i=1}^{[n\delta]} (X_{in} - \overline{X}_{n,\delta})(H_{in} - \overline{H}_{n,\delta})}{(\sum_{i=1}^{[n\delta]} (X_{in} - \overline{X}_{n,\delta})^2 \sum_{i=1}^{[n\delta]} (H_{in} - \overline{H}_{n,\delta})^2)^{1/2}}, \tag{1.5}$$

where $0 < \delta < 1$, $[\ ]$ is the greatest integer function, $\overline{H}_{n,\delta} = \sum_{i=1}^{[n\delta]} H_{in}/[n\delta]$, and $\overline{X}_{n,\delta} = \sum_{i=1}^{[n\delta]} X_{in}/[n\delta]$. The Type I censored-data version has the same form [see Eq. (4.4)], except that the upper limit of summation and divisor for the mean is replaced by the random number of order statistics observed.

2. SMALL-SAMPLE PROPERTIES

Smith and Bain (1976) did a small-scale simulation study of the critical values for $1 - r^2(\mathbf{X}, \mathbf{H}, \delta)$. Gerlach (1980) reported some critical values for the correlation coefficient, $r(\mathbf{X}, \mathbf{m}, \delta)$, that uses the normal scores m_{in} (the square root of the censored-data version of W'). Stephens (1986) presented a table of critical values for $n(1 - r^2(\mathbf{X}, \mathbf{m}, \delta))$.

In our study at each sample size 5,000 samples were generated and the $[5{,}000(\alpha)]$th-smallest correlation was taken as the estimate of the (100α)th percentile, η_α, of $r(\mathbf{X}, \mathbf{H}, \delta)$. The scores H_{in} were calculated using the International Mathematical and Statistical Libraries (IMSL) routine MDNRIS. We also obtained the estimated critical values for both $r(\mathbf{X}, \mathbf{m}, \delta)$ and the square root of the censored-data Shapiro–Wilk statistic. The latter statistic, given by

$$r(\mathbf{X}, \mathbf{a}, \delta) = \frac{\sum_{i=1}^{[n\delta]} (X_{in} - \overline{X}_{n,\delta})(a_{in} - \bar{a}_{n,\delta})}{(\sum_{i=1}^{[n\delta]} (X_{in} - \overline{X}_{n,\delta})^2 \sum_{i=1}^{[n\delta]} (a_{in} - \bar{a}_{n,\delta})^2)^{1/2}}, \tag{2.1}$$

involves the first $[n\delta]$ entries of $\mathbf{a} = [a_{1n}, \ldots, a_{nn}]' = \mathbf{V}^{-1}\mathbf{m}$. We calculated the expected values of the order statistics by numerical integration and the entries of $\mathbf{V}$ using the Davis and Stephens (1978) algorithm.

Remark 1. On a Cray 1 running at scalar speeds it took .016, .029, and .056 minutes to calculate $\mathbf{a}$ for $n = 50$, 100, and 200. Approximately 35% of this time was used to calculate $\mathbf{m}$. Since Cray 1 scalar speeds are becoming available on much less expensive machines, it is reasonable that users may soon want to calculate the Shapiro–Wilk statistic even for sample sizes greater than 50.

The results of the simulation, covering $\alpha = .10$, .05, and .01, are presented in Tables 1 and 2. [A version of Tables 1 and 2 that covers $\delta = .1(.1)1.0$ and twice as many values of n appears in Verrill and Johnson (1987b).] Because of the relatively poor power performance of $r(\mathbf{X}, \mathbf{H}, \delta)$ (except in the symmetric, long-tailed case; see Sec. 3), we only report the Monte Carlo critical values of $r(\mathbf{X}, \mathbf{m}, \delta)$ and $r(\mathbf{X}, \mathbf{a}, \delta)$. Critical values for $r(\mathbf{X}, \mathbf{H}, \delta)$ are available in Verrill (1981) and Verrill and Johnson (1987b).

Notice how the critical values for the square root of the censored-data Shapiro–Francia correlation statistic and the square root of the censored-data Shapiro–Wilk statistic approach a common value as the sample size increases. In Section 4 we establish the asymptotic equivalence of these statistics.

Table 1. Nominal Critical Values: Square Root of the Shapiro–Francia Statistic: $r(\mathbf{X}, \mathbf{m}, \delta)$

	δ				
n	.2	.4	.6	.8	1.0
20	**.89085**	**.92061**	**.93992**	**.95375**	**.96052**
	.86464	.89905	.92641	.94407	.95140
	.77801	*.84473*	*.89574*	*.91664*	*.92476*
30	**.90544**	**.93900**	**.95588**	**.96559**	**.97039**
	.88206	.92431	.94512	.95704	.96359
	.83040	*.88869*	*.91559*	*.93782*	*.94504*
40	**.91876**	**.94922**	**.96410**	**.97291**	**.97700**
	.89538	.93701	.95586	.96696	.97217
	.84912	*.90721*	*.93531*	*.95311*	*.95772*
60	**.93502**	**.96200**	**.97359**	**.98075**	**.98377**
	.92081	.95080	.96691	.97618	.98010
	.88454	*.92799*	*.95154*	*.96310*	*.97171*
80	**.94766**	**.97028**	**.98004**	**.98547**	**.98709**
	.93427	.96247	.97494	.98190	.98459
	.90535	*.94059*	*.96123*	*.97234*	*.97802*
100	**.95404**	**.97446**	**.98273**	**.98790**	**.98949**
	.94160	.96709	.97854	.98484	.98728
	.91061	*.94717*	*.96603*	*.97736*	*.98245*
150	**.96583**	**.98135**	**.98782**	**.99133**	**.99258**
	.95612	.97618	.98439	.98927	.99100
	.93116	*.96318*	*.97609*	*.98406*	*.98742*
250	**.97634**	**.98796**	**.99217**	**.99465**	**.99538**
	.97077	.98493	.99016	.99328	.99448
	.95228	*.97514*	*.98458*	*.98991*	*.99262*
500	**.98614**	**.99301**	**.99572**	**.99709**	**.99754**
	.98249	.99142	.99474	.99652	.99709
	.97363	*.98735*	*.99214*	*.99507*	*.99599*

NOTE: The values have been estimated from 5,000 trials: 10% (bold), 5% (Roman), and 1% (italic).

Table 2. Nominal Critical Values: Square Root of the Shapiro–Wilk Statistic: $r(\mathbf{X}, \mathbf{a}, \delta)$

	δ				
n	.2	.4	.6	.8	1.0
20	**.85317**	**.90858**	**.93370**	**.95081**	**.95985**
	.82128	.88556	.92026	.94137	.95166
	.72803	*.82952*	*.88777*	*.91800*	*.93121*
30	**.87101**	**.92608**	**.94975**	**.96307**	**.96919**
	.83697	.91023	.94034	.95514	.96287
	.77513	*.87689*	*.91586*	*.93588*	*.94957*
40	**.87974**	**.93321**	**.95617**	**.96873**	**.97454**
	.85172	.91989	.94805	.96277	.97040
	.79750	*.89325*	*.93114*	*.94991*	*.96048*
60	**.92464**	**.95847**	**.97202**	**.98019**	**.98334**
	.90773	.94934	.96606	.97580	.98014
	.87398	*.92882*	*.95245*	*.96538*	*.97327*
80	**.93909**	**.96712**	**.97847**	**.98463**	**.98684**
	.92616	.95936	.97387	.98182	.98445
	.89644	*.94519*	*.96279*	*.97330*	*.97909*
100	**.94885**	**.97218**	**.98175**	**.98717**	**.98915**
	.93678	.96584	.97778	.98455	.98723
	.91022	*.95150*	*.96910*	*.97872*	*.98341*
150	**.96127**	**.97987**	**.98725**	**.99109**	**.99237**
	.95324	.97550	.98429	.98922	.99116
	.93473	*.96551*	*.97768*	*.98482*	*.98796*
250	**.97418**	**.98691**	**.99180**	**.99436**	**.99526**
	.96844	.98428	.99001	.99325	.99446
	.95564	*.97657*	*.98557*	*.99037*	*.99282*
500	**.98522**	**.99275**	**.99556**	**.99700**	**.99747**
	.98174	.99125	.99463	.99647	.99707
	.97509	*.98772*	*.99265*	*.99511*	*.99609*

NOTE: The values have been estimated from 5,000 trials: 10% (bold), 5% (Roman), and 1% (italic).

Using the critical values from Table 1, the true level is some $\alpha_{\hat{c}_\delta}$. Let $\hat{R}_{[5,000\alpha]}$ be the estimated (100α)th percentile based on 5,000 trials. From the distribution of order statistics, we then have

$$\Pr[\alpha_L < \alpha_{\hat{c}_\delta} \le \alpha_U] = \Pr[\eta_{\alpha_L} < \hat{R}_{[5,000\alpha]} \le \eta_{\alpha_U}]$$
$$= \sum_{j=[5,000\alpha]}^{5,000} \binom{5,000}{j} (\alpha_U)^j(1-\alpha_U)^{5,000-j}$$
$$- \sum_{j=[5,000\alpha]}^{5,000} \binom{5,000}{j} (\alpha_L)^j(1-\alpha_L)^{5,000-j}.$$

We evaluated these binomial probabilities via the IMSL routine MDBIN. [Alternatively, one could evaluate them by using the incomplete beta distribution; see Abramowitz and Stegun (1964, sec. 6.6).] We varied α_L and α_U to obtain the following 99% intervals for the true level: (.006, .014) for $\alpha = .01$, (.042, .058) for $\alpha = .05$, and (.089, .111) for $\alpha = .10$.

If the tabled values were rounded to three decimal places, then $\hat{R}_{[5,000\alpha]}$ would not only be the sample (100α)th percentile, but it would also be numerous other percentiles because of ties in a very dense list of values. For instance, in the complete-sample case with $n = 90$ there were 128 ties at the estimated 10th percentile of the Shapiro–Francia statistic. This could change the level, up or down, by more than .02. When four decimals were reported, the number of ties decreased to 19. The situation is worse for larger sample sizes, so we used five decimal places. (In the complete-sample case with $n = 500$ and five decimal places reported, there were only nine ties at the estimated 10th percentile of the Shapiro–Francia statistic.)

3. POWER STUDY

Several competitors of $r(\mathbf{X}, \mathbf{H}, \delta)$ already exist. Pettitt (1976) provided natural extensions of the Cramer–Von Mises and Anderson–Darling statistics; these were further discussed by Pettitt and Stephens (1976). Mihalko and Moore (1980) proposed a censored-data version of the chi-squared goodness-of-fit statistic and studied its asymptotic properties.

To gain some understanding of the merits of these censored-data goodness-of-fit procedures relative to the correlation statistics, we performed a Monte Carlo power study. The statistics compared were (a) Cramer–Von Mises: $w2 = \sum_{i=1}^{[n\delta]} (\hat{\tau}_{in} - (i - \frac{1}{2})/n)^2$ and $\hat{\tau}_{in} = \Phi((X_{in} - \hat{\mu})/\hat{\sigma})$, where $\hat{\mu}$ and $\hat{\sigma}$ are the Gupta (1952) censored-data estimates; (b) Anderson–Darling: $a2 = \sum_{i=1}^{[n\delta]} [(2i-1)/n][\ln(1 - \hat{\tau}_{in}) - \ln(\hat{\tau}_{in})] - 2\sum_{i=1}^{[n\delta]} \ln(1 - \hat{\tau}_{in})$; (c) Shapiro–Wilk: $r(\mathbf{X}, \mathbf{a}, \boldsymbol{\delta})$; (d) Shapiro–Francia: $r(\mathbf{X}, \mathbf{m}, \delta)$; (e) DeWet–Venter: $r(\mathbf{X}, \mathbf{H}, \delta)$; and (f) chi-square: chisq, for which the cells were data-dependent and the number of cells equaled $\max[\{[n\delta/5] + 1, 2\}]$ (see Mihalko and Moore 1980).

For the choice of alternative distributions, we tried to cover a range of cases. These were selected from a larger power study and were intended to be representative of (a) symmetric and short-tailed, (b) symmetric and long-tailed, (c) skewed and short-tailed, and (d) skewed and long-tailed distributions.

First, the nominal $\alpha = .10$ critical values were estimated for each statistic, from 5,000 trials at $\delta = .1\ (.1)\ 1.0$ and for sample sizes $n = 20$, 40, and 70. Power was then estimated from 1,000 trials.

From the estimated power curves (see Verrill and Johnson 1987b), we conclude the following:

1. The Shapiro–Wilk statistic did very well, except against symmetric long-tailed alternatives. This finding conforms to known results for the complete-sample case.
2. Although they were seldom the most powerful tests, on the whole the Anderson–Darling and Cramer–Von Mises tests also performed quite well. This finding is in accord with the results of the power trials of Stephens (1974). Given that they do not require the calculation of an order statistic covariance matrix, the $a2$ and $w2$ tests are less computer-intensive than the Shapiro–Wilk test. This fact, together with their superior power properties in the symmetric, long-tailed case, leads us to regard them as serious competitors to the correlation coefficient tests.
3. The Shapiro–Francia statistic was roughly comparable in power with the Anderson–Darling and Cramer–Von Mises statistics. The DeWet–Venter statistic performed extremely well in the symmetric, long-tailed case

and quite poorly in the other cases [see Ryan and Joiner (1973) for a discussion of the differences in the power behavior of the DeWet–Venter and the Shapiro–Wilk statistics]. The performance of the chi-square test was erratic, but generally poor.

4. Power is not always monotone in the censoring proportion δ. For alternatives such as the uniform, increased censoring can improve power. This feature was also noticed by Koziol and Petkau (1984) in their study of asymptotic efficiency. Presumably, the nonmonotone power can be attributed to the fact that we are dealing with omnibus tests. A test derived especially for testing a normal versus a uniform should have power that is monotone in δ.

5. On the other hand, censoring can have a drastic *negative* effect on power. For example, for a sample of size 40, the uncensored Shapiro–Wilk and Anderson–Darling tests have high power (.88 and .75, respectively) against a triangular distribution with density $2x$ on [0, 1]. Nevertheless, for even moderate censoring, $\delta \leq .8$, the power of all six tests is essentially reduced to level .1.

4. THE LIMIT DISTRIBUTION

With increasing sample size, even Monte Carlo determination of critical values can become infeasible. For instance, in one of our applications more than 5,000 interarrival times needed to be modeled for input to a simulation. Extending the method of DeWet and Venter (1972, 1973), who established the asymptotic theory for the complete-sample case, we obtain the censored sample results.

To obtain a limit distribution, $2[n\delta](1 - r(\mathbf{X}, \mathbf{H}, \delta))$ needs to be centered by $a_{n,\delta}/t^2_{n,\delta}$, where

$$t^2_{n,\delta} = \sum_{i=1}^{[n\delta]} (H_{in} - \overline{H}_{n,\delta})^2/[n\delta]. \tag{4.1}$$

The constant $a_{n,\delta}$, defined in Verrill and Johnson (1983), involves $H'(i/(n+1))$, the standard normal pdf and $t^2_{n,\delta}$. Further, $t^2_{n,\delta}$ can be shown to have limit

$$K_{3,\delta} = \frac{1}{\delta}\int_0^\delta H^2(x)\,dx - \left(\frac{1}{\delta}\int_0^\delta H(x)\,dx\right)^2. \tag{4.2}$$

A crucial role in the limit is played by truncated versions of the Hermite polynomials. In particular, let

$$\begin{aligned} g_m(x) & \\ &= 0, \qquad x = 0, \\ &= \frac{1}{(2^m m!)^{1/2}} h_m\left(\frac{H(x)}{2^{1/2}}\right), \quad x \in (0, \delta], \quad m = 0, 1, \ldots, \\ &= g_m(\delta), \qquad x \in (\delta, 1], \end{aligned} \tag{4.3}$$

where h_m is the mth Hermite polynomial (see Rainville 1960). Note that each $g_m(x)$ should be written as $g_{m,\delta}(x)$, but for convenience we drop the δ subscript.

We now present the asymptotic limit distribution. Let the constants $a_{n,\delta}$ and $J_{uv,\delta}$ ($u \geq v = 0, 1, 2$) be defined as in Verrill and Johnson (1983) and Johnson and Verrill (1988).

Theorem 1. For normal random variables, $2[n\delta](1 - r(\mathbf{X}, \mathbf{H}, \delta)) - a_{n,\delta}/t^2_{n,\delta} \xrightarrow{D} Y_\delta/K_{3,\delta}$, where

$$\begin{aligned} Y_\delta = \sum_{m=3}^{\infty} \frac{1}{m}(W_m^2 - E(W_m^2)) &+ J_{22,\delta}(W_2^2 - E(W_2^2)) \\ &+ J_{11,\delta}(W_1^2 - E(W_1^2)) + J_{00,\delta}(W_0^2 - E(W_0^2)) \\ &+ J_{21,\delta}(W_2W_1 - E(W_2W_1)) + J_{20,\delta}(W_2W_0 - E(W_2W_0)) \\ &+ J_{10,\delta}(W_1W_0 - E(W_1W_0)); \end{aligned}$$

for every M, the random variables $W_0, W_1, \ldots, W_M$ have a joint multivariate normal distribution with expectation vector $\mathbf{0}$ and covariance matrix $\Sigma_M = (\sigma_{ij})$ ($i, j = 0, 1, \ldots, M$), where $\sigma_{ij} = \int_0^1 g_i(x)g_j(x)\,dx$ and the g_m are defined in (4.3).

Remark 2. Note that the $g_m(x)$ and hence the W_i depend on δ. The case of no censoring corresponds to $\delta = 1$. Then, the orthogonality of the Hermite polynomials implies that the covariance terms $\sigma_{ij} = 0$ ($i \neq j$). The censoring introduces correlations among the terms of the series for Y_δ. Also, all of the $J_{uv,\delta}$ equal 0 when there is no censoring. Thus our representation reduces to the complete-sample expression given by DeWet and Venter (1972, 1973).

Remark 3. The limit distribution also applies to the squared correlation statistic. Replace $2[n\delta](1 - r(\mathbf{X}, \mathbf{H}, \delta))$ with $[n\delta](1 - r^2(\mathbf{X}, \mathbf{H}, \delta))$.

Theorem 2. Let $r(\mathbf{X}, \mathbf{a}, \delta)$, the square root of the censored-data Shapiro–Wilk statistic, be given by (2.1). For normal random variables, $n(r(\mathbf{X}, \mathbf{a}, \delta) - r(\mathbf{X}, \mathbf{H}, \delta)) \to 0$ in probability, so $r(\mathbf{X}, \mathbf{a}, \delta)$ and $r(\mathbf{X}, \mathbf{H}, \delta)$ have the same limit distribution.

Proof. Set $\mathbf{H} = [H_{1n}, \ldots, H_{nn}]$. The conclusion follows directly from the work of Leslie, Stephens, and Fotopoulos (1986) and Verrill and Johnson (1987a, theorem 3.1). In the notation of that theorem, the scores associated with the Shapiro–Wilk statistic, $r(\mathbf{X}, \mathbf{a}, \delta)$ are given by $\mathbf{b} = \frac{1}{2}\mathbf{V}^{-1}\mathbf{m}$. Since

$$\begin{aligned} \|\mathbf{b} - \mathbf{H}\| = \|\tfrac{1}{2}\mathbf{V}^{-1}\mathbf{m} - \mathbf{H}\| &\leq \|\tfrac{1}{2}\mathbf{V}^{-1}\mathbf{m} - \mathbf{m}\| + \|\mathbf{m} - \mathbf{H}\| \\ &\leq C_1/[\log(n)]^{1/2} + C_2/[\log(n)]^{1/2}, \end{aligned}$$

$\lim_{n\to\infty} \log\log(n) \sum_{i=1}^n (b_{in} - H_{in})^2 = 0$, which establishes Theorem 2. Here, the C_1 bound is in Leslie, Stephens, and Fotopoulos [1986, eq. (1)], and the C_2 bound is given in Verrill and Johnson (1987a, lemmas 3.2, 3.3).

Remark 4. Verrill and Johnson (1987a) showed that the same limit distribution applies if $H(i/(n+1))$ is replaced with other approximate scores. In particular, it applies if we use the scores m_{in} (the Shapiro–Francia statistic), the Filliben median scores, or the Weisberg–Bingham scores.

Remark 5. The same limit result is obtained for Type I censored data and the correlation statistic

$$r_{N(\omega)} = \frac{\sum_{i=1}^{N(\omega)} (X_{in} - \overline{X}_{N(\omega)})(H_{in} - \overline{H}_{N(\omega)})}{(\sum_{i=1}^{N(\omega)} (X_{in} - \overline{X}_{N(\omega)})^2 \sum_{i=1}^{N(\omega)} (H_{in} - \overline{H}_{N(\omega)})^2)^{1/2}}, \tag{4.4}$$

where $\overline{H}_{N(\omega)} = \sum_{i=1}^{N(\omega)} H_{in}/N(\omega)$, $\overline{X}_{N(\omega)} = \sum_{i=1}^{N(\omega)} X_{in}/N(\omega)$, and $N(\omega)$ is equal to the number of $X_{in}(\omega) \leq L$ for a specified L. In this case, the δ in Y_δ equals $\Phi((L - \mu)/\sigma)$. For the limit distribution approximation, the upper limits in $a_{n,\delta}$ and $t_{n,\delta}^2$ need to be changed to $N(\omega)$. See Verrill and Johnson (1987a) and Verrill (1981) for details on the equivalence of the limits under Type I and Type II censoring.

5. LARGE–SAMPLE APPROXIMATIONS TO THE CRITICAL VALUES OF $r(\mathbf{X}, \mathbf{H}, \delta)$

Our primary motivation for obtaining the limit distribution was that its percentiles could be used for critical values when sample size is large and extensive Monte Carlo results are not available. They also provide a check for the Monte Carlo studies. These critical values are obtained in the following manner.

Let $Y_{\delta,M}$ denote the series for Y_δ that is truncated after the term $M^{-1}(W_M^2 - E(W_M^2))$. Since the critical values of $Y_{\delta,M}$ converge to those of Y_δ, we may approximate the critical values for Y_δ by those of $Y_{\delta,M}$ for large M. The random part of $Y_{\delta,M}$ has characteristic function

$$\int_{R^{M+1}} \exp(it\mathbf{w}'\mathbf{B}\mathbf{w}) \frac{1}{\left[(2\pi)^{(M+1)/2}\left|\sum_M\right|^{1/2}\right]} \exp(-\tfrac{1}{2}\mathbf{w}' \textstyle\sum_M^{-1} \mathbf{w})\, d\mathbf{w},$$

where $\mathbf{w}' = (w_0, w_1, \ldots, w_M)$ and $\mathbf{B} = \text{diag}(\mathbf{A}, \frac{1}{3}, \frac{1}{4}, \ldots, 1/M)$, with $a_{ii} = J_{ii,\delta}$, $a_{ik} = J_{ik,\delta}/2$ $(i \neq k = 0, 1, 2)$. Make the change of variable $\mathbf{z} = \mathbf{C}^{-1}\mathbf{w}$, where $\mathbf{C}' \sum_M^{-1} \mathbf{C} = \mathbf{I}$, and write $\mathbf{Z}'\mathbf{C}'\mathbf{B}\mathbf{C}\mathbf{Z} = \lambda_{0M}(\mathbf{Z}'\mathbf{e}_{0M})^2 + \cdots + \lambda_{MM}(\mathbf{Z}'\mathbf{e}_{MM})^2$, where the λ_{mM} are the eigenvalues of $\mathbf{C}'\mathbf{B}\mathbf{C}$ and the $\mathbf{e}_{mM}$ are the eigenvectors. Since the $\mathbf{Z}'\mathbf{e}_{mM}$ are independent standard normal, the integral is recognized as the characteristic function of $\sum_{m=0}^{M} \lambda_{mM}\chi_{mM}^2$, where the χ_{mM}^2 are distributed as independent chi-squares with 1 df.

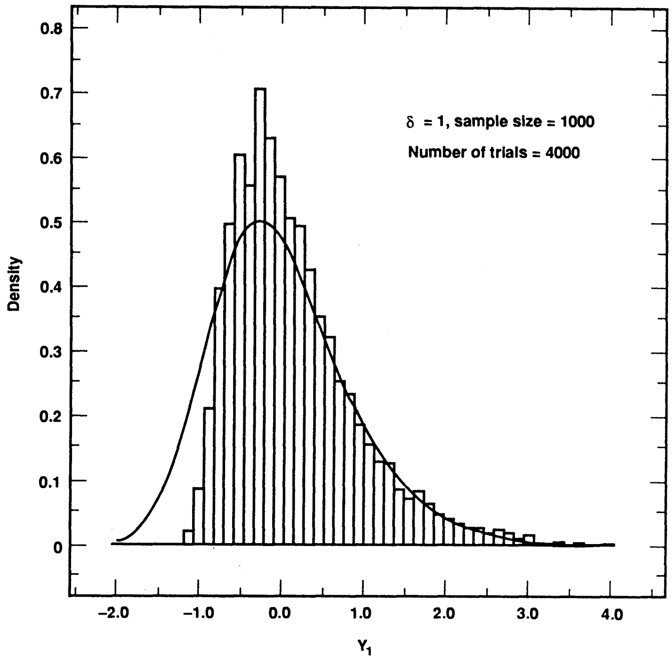

Figure 1. Comparison of the Monte Carlo Results to the Limit Distribution.

Table 3. Critical Values of $r_{n,\delta}$ Obtained From Asymptotic Theory: 10% (bold) and 5% (Roman)

	δ				
n	*.2*	*.4*	*.6*	*.8*	*1.0*
50	**.823705**	**.924432**	**.955979**	**.971312**	**.977455**
	.770705	.903632	.944762	.964550	.972865
100	**.907840**	**.960569**	**.977077**	**.985114**	**.988193**
	.881340	.950169	.971468	.981733	.985898
150	**.937006**	**.973053**	**.984353**	**.989862**	**.991919**
	.919340	.966119	.980614	.987607	.990389
200	**.951911**	**.979431**	**.988069**	**.992282**	**.993828**
	.938661	.974231	.985265	.990591	.992681
500	**.979641**	**.991306**	**.994977**	**.996768**	**.997392**
	.974341	.989226	.993855	.996092	.996933
1,000	**.989376**	**.995475**	**.997395**	**.998331**	**.998646**
	.986726	.994435	.996834	.997993	.998416
2,000	**.994462**	**.997650**	**.998652**	**.999140**	**.999299**
	.993137	.997130	.998372	.998971	.999184
4,000	**.997119**	**.998783**	**.999305**	**.999558**	**.999638**
	.996456	.998523	.999164	.999474	.999581
5,000	**.997666**	**.999015**	**.999438**	**.999644**	**.999708**
	.997136	.998807	.999326	.999576	.999662

The eigenvalues of $\mathbf{C}'\mathbf{B}\mathbf{C}$ are the same as those of $\mathbf{B}\sum_M$ and were obtained by solving $\mathbf{B}^{-1}\mathbf{x} = \lambda^{-1}\sum_M \mathbf{x}$. We experienced some numerical difficulties when computing the eigenvalues and eigenvectors. Using single precision on a Cray (a 64-bit machine) and the Numerical Algorithms Group (NAG) routine Fϕ2ADF, we could only obtain 15–20 terms when δ was near .7. We overcame the numerical singularities and obtained 300 terms by using double precision and NAG routine Fϕ2BJF. After calculating the eigenvalues λ_{mM}, we applied an algorithm of Davies (1980) to obtain the critical values of $Y_{\delta,M}$. [See Verrill and Johnson (1987b) for the results.]

Next, let $c_\delta(\alpha)$ denote the approximation to the level-α critical value of Y_δ. Since $2[n\delta](1 - r(\mathbf{X}, \mathbf{H}, \delta)) - a_{n,\delta}/t_{n,\delta}^2 \xrightarrow{D} Y_\delta/K_{3,\delta}$, the large-sample approximation to the level-α critical value of $r(\mathbf{X}, \mathbf{H}, \delta)$ is

$$1 - \frac{c_\delta(\alpha)/K_{3,\delta} + a_{n,\delta}/t_{n,\delta}^2}{2[n\delta]}. \tag{5.1}$$

One feature of the large-sample approximation merits emphasis. We have found that convergence to the limit distribution is slow. Even for the complete-sample case, which is well established and widely used, the agreement between the simulated distribution of $2[n\delta](1 - r(\mathbf{X}, \mathbf{H}, \delta)) - a_{n,\delta}/t_{n,\delta}^2$ and Y_1 is not particularly good. (See Fig. 1 for $n = 1{,}000$.) Fortunately, the approximation to the upper tail of Y_δ and the critical values of $r(\mathbf{X}, \mathbf{H}, \delta)$ is quite good for moderately large samples.

In Table 3, we provide values of (5.1) for $\alpha = .10, .05$, and several large sample sizes n. Notice how the critical values increase with δ and n. Table 4 gives the estimated levels when the asymptotic critical values (5.1) are used. The number of trials at each sample size (2,000–10,000) are indicated in parentheses. For a fixed sample size, the quality of the approximation deteriorates as δ decreases. Our Monte Carlo studies of the critical values suggest that

Table 4. Estimated Level for Nominal α = .05 and .10

n^*	$\delta = .2$		$\delta = .4$		$\delta = .6$		$\delta = .8$		$\delta = 1.0$	
	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$	$\alpha = .10$
50 (10)	.000	.005	.010	.027	.022	.048	.035	.064	.044	.087
100 (8)	.005	.019	.021	.042	.031	.060	.040	.074	.052	.098
200 (4)	.016	.034	.030	.054	.039	.070	.046	.087	.060	.110
500 (4)	.024	.049	.036	.071	.045	.083	.047	.093	.060	.108
1,000 (4)	.028	.061	.038	.080	.046	.083	.049	.090	.061	.110
2,000 (2)	.037	.069	.038	.084	.046	.084	.048	.094	.057	.108
5,000 (2)	.050	.081	.052	.094	.058	.104	.066	.101	.056	.110

* Number of trials, in thousands, given in parentheses.

the large-sample approximations are fairly good for $[n\delta] \geq 100$, unless there is heavy censoring, where $[n\delta] \geq 500$ may be required.

[*Received July 1986. Revised March 1988.*]

REFERENCES

Abramowitz, M., and Stegun, I. A. (1964), *Handbook of Mathematical Functions*, Washington, DC: National Bureau of Standards.

Davies, R. B. (1980), "The Distribution of a Linear Combination of χ^2 Random Variables," *Applied Statistics*, 38, 52–73.

Davis, C. S., and Stephens, M. A. (1978), "Approximating the Covariance Matrix of Normal Order Statistics," *Applied Statistics*, 27, 206–212.

DeWet, T., and Venter, J. H. (1972), "Asymptotic Distributions of Certain Test Criteria of Normality," *South African Statistical Journal*, 6, 135–149.

——— (1973), "Asymptotic Distributions for Quadratic Forms With Applications to Tests of Fit," *The Annals of Statistics*, 1, 380–387.

Filliben, J. J. (1975), "The Probability Plot Correlation Coefficient Test for Normality," *Technometrics*, 17, 111–117.

Gerlach, B. (1980), "A Correlation-Type Goodness-of-Fit Test for Normality With Censored Sampling," *Mathematische Operationsforschung und Statistik* (Series Statistics), 11, 207–218.

Gupta, A. K. (1952), "Estimation of the Mean and Standard Deviation of a Normal Population From a Censored Sample," *Biometrika*, 39, 260–273.

Johnson, R., and Verrill, S. (1988), "The Large Sample Distribution of the Shapiro–Wilk Statistic and Its Variations Under Type I and Type II Right Censoring," unpublished manuscript.

Koziol, J., and Petkau, A. J. (1984), "Relative Efficiencies of Goodness-of-Fit Procedures With Truncated Data," *Canadian Journal of Statistics*, 12, 107–117.

Leslie, J. R., Stephens, M. A., and Fotopoulos, S. (1986), "Asymptotic Distribution of the Shapiro–Wilk *W* for Testing for Normality," *The Annals of Statistics*, 14, 1497–1506.

Mihalko, D. P., and Moore, D. S. (1980), "Chi-Square Tests of Fit for Type II Censored Data," *The Annals of Statistics*, 8, 625–644.

Pettitt, A. N. (1976), "Cramer–Von Mises Statistics for Testing Normality With Censored Samples," *Biometrika*, 63, 475–481.

Pettitt, A. N., and Stephens, M. A. (1976), "Modified Cramer–Von Mises Statistics for Censored Data," *Biometrika*, 63, 291–298.

Rainville, E. D. (1960), *Special Functions*, New York: Chelsea Publishing.

Ryan, T., and Joiner, B. (1973), "Normal Probability Plots and Tests for Normality," technical report, Pennsylvania State University, Dept. of Statistics.

Sarkadi, K. (1975), "The Consistency of the Shapiro–Francia Test," *Biometrika*, 62, 445–450.

Shapiro, S. S., and Francia, R. A. (1972), "An Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association*, 67, 215–216.

Shapiro, S. S., and Wilk, M. B. (1965), "Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, 591–611.

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968), "A Comparative Study of Various Tests for Normality," *Journal of the American Statistical Association*, 63, 1343–1372.

Smith, R. M., and Bain, L. J. (1976), "Correlation Type Goodness-of-Fit Statistics With Censored Sampling," *Communications in Statistics, Part A—Theory and Methods*, 5, 119–132.

Stephens, M. A. (1974), "EDF Statistics for Goodness-of-Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730–737.

——— (1986), "Tests Based on Regression and Correlation," in *Goodness-of-Fit Techniques*, eds. R. D'Agostino and M. A. Stephens, New York: Marcel Dekker, pp. 195–233.

Verrill, S. P. (1981), "Some Asymptotic Results Concerning Censored Data Versions of the Shapiro–Wilk Goodness of Fit Test," unpublished Ph.D. dissertation, University of Wisconsin–Madison, Dept. of Statistics.

Verrill, S. P., and Johnson, R. A. (1983), "The Asymptotic Distributions of Censored Data Versions of the Shapiro–Wilk Test of Normality Statistic," Technical Report 702, University of Wisconsin–Madison, Dept. of Statistics.

——— (1987a), "The Asymptotic Equivalence of Some Modified Shapiro–Wilk Statistics—Complete and Censored Sample Cases," *The Annals of Statistics*, 15, 413–419.

——— (1987b), "A Large Sample Approximation to the Distribution of a Modified Shapiro–Wilk Statistic Under Right Censoring," UCRL-97603 preprint, Lawrence Livermore National Laboratory.

Weisberg, S., and Bingham, C. (1975), "An Approximate Analysis of Variance Test for Non-normality Suitable for Machine Calculation," *Technometrics*, 17, 133–134.