Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

An Approximate Analysis of Variance Test for Normality

S. S. Shapiro & R. S. Francia

To cite this article: S. S. Shapiro & R. S. Francia (1972) An Approximate Analysis of Variance Test for Normality, Journal of the American Statistical Association, 67:337, 215-216

To link to this article: http://dx.doi.org/10.1080/01621459.1972.10481232

© Journal of the American Statistical Association
March 1972, Volume 67, Number 337
Theory and Methods Section

An Approximate Analysis of Variance Test for Normality

S. S. SHAPIRO and R. S. FRANCIA*

This article presents a modification of the Shapiro-Wilk W statistic for testing normality which can be used with large samples. Shapiro and Wilk gave coefficients and percentage points for sample sizes up to 50. These coefficients required obtaining an approximation to the covariance matrix of the normal order statistics. The proposed test uses coefficients which depend only on the expected values of the normal order statistics which are generally available. Results of an empirical sampling study to compare the sensitivity of the test statistic to the W test statistic are briefly discussed.

1. THE W TEST FOR NORMALITY

The analysis of variance test statistic W developed in [5] for testing complete samples for normality is derived as follows:

Let $m'=(m_1, m_2, \cdots, m_n)$ denote the vector of expected values of standard normal order statistics, and let $V=(v_{ij})$ be the corresponding $n\times n$ covariance matrix. That is, if $x_1\leq x_2\leq \cdots \leq x_n$ denotes an ordered random sample of size n from a standard normal distribution ($\mu=0$, $\sigma=1$), then

$$E(x_i) = m_i(i = 1, 2, \cdots, n), \tag{1.1}$$

and

$$\operatorname{Cov}(x_i, x_j) = v_{ij}(i, j = 1, 2, \cdots, n). \tag{1.2}$$

Let $y'=(y_1, y_2, \cdots, y_n)$ denote a vector of ordered random observations. It is desired to derive a test for the composite hypothesis that $\{y_i\}$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2. If $\{y_i\}$ is a normal sample, then y_i may be expressed as

$$y_i = \mu + \sigma x_i(i = 1, 2, \cdots, n). \tag{1.3}$$

The W test statistic is then defined as

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{1.4}$$

where

$$a' = (a_1, a_2, \cdots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{\frac{1}{2}}}\cdot \tag{1.5}$$

To date, however, the elements of V are given only for sample sizes up to 20 in [4]. Approximations were developed in [5] for calculating coefficients for using the W test up to sample size 50.

2. AN APPROXIMATE W′ TEST FOR NORMALITY

For large samples, the observations $\{y_i\}$ may be treated as if they were independent and as suggested in [1], the identity matrix I can be substituted for V^{-1} in the estimation of the slope of the regression line. Consequently, an approximate W' test for normality may be defined as

$$W' = \frac{(\sum_{i=1}^n b_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{2.1}$$

where

$$b' = (b_1, b_2, \cdots, b_n) = \frac{m'}{(m'm)^{\frac{1}{2}}}\cdot \tag{2.2}$$

Values of m are given in [2] for sample sizes $n=2(1)100(25)300(50)400$.

The null distribution of W' was approximated by an empirical sampling study. One thousand normal random sample were generated for each sample size $n=35$, 50, 51(2)99, with the use of IBM subroutine GAUSS [3]. For each sample size, the 1,000 values of W' were calculated by equation (2.1), and the quantities $W'_j(p)$ were determined from the corresponding ordered W_j' using the relationship $p_j=j/n-1$,

$$j = 10,50(50)200(300)800(50)950,990.$$

The table gives the empirical percentage points of W' for the above values of n and j.

3. COMPARISON OF THE SENSITIVITIES OF THE W AND W′ TEST STATISTICS

An empirical sampling study was conducted to compare the sensitivities of the W and W' test statistics in detecting non-normality. One hundred samples of size $n=35$ and 50 were generated randomly from 37 alternative distributions. The alternative distributions came from 12 different families and represented a wide variety of shapes including skewed and symmetric as well as

* S. S. Shapiro is director, Scientific Services, and R. S. Francia is statistician, both with Programming Methods, Inc., 1301 Sixth Avenue, New York, N. Y. 10010. The authors are indebted to Mrs. Ruth Kurts and Mr. Tom Brady who programmed the routines for generating the statistical tables for the power study and the W' statistic, and to the editor and referees for their helpful comments.

EMPIRICAL PERCENTAGE POINTS OF THE APPROXIMATE W′ TEST

n	P .01	.05	.10	.15	.20	.50	.80	.85	.90	.95	.99
35	0.919	0.943	0.952	0.956	0.964	0.976	0.982	0.985	0.987	0.989	0.992
50	.935	.953	.963	.968	.971	.981	.987	.988	.990	.991	.994
51	0.935	0.954	0.964	0.968	0.971	0.981	0.988	0.989	0.990	0.992	0.994
53	.938	.957	.964	.969	.972	.982	.988	.989	.990	.992	.994
55	.940	.958	.965	.971	.973	.983	.988	.990	.991	.992	.994
57	.944	.961	.966	.971	.974	.983	.989	.990	.991	.992	.994
59	.945	.962	.967	.972	.975	.983	.989	.990	.991	.992	.994
61	0.947	0.963	0.968	0.973	0.975	0.984	0.990	0.990	0.991	0.992	0.994
63	.947	.964	.970	.973	.976	.984	.990	.991	.992	.993	.994
65	.948	.965	.971	.974	.976	.985	.990	.991	.992	.993	.995
67	.950	.966	.971	.974	.977	.985	.990	.991	.992	.993	.995
69	.951	.966	.972	.976	.978	.986	.990	.991	.992	.993	.995
71	0.953	0.967	0.972	0.976	0.978	0.986	0.990	0.991	0.992	0.994	0.995
73	.956	.968	.973	.976	.979	.986	.991	.992	.993	.994	.995
75	.956	.969	.973	.976	.979	.986	.991	.992	.993	.994	.995
77	.957	.969	.974	.977	.980	.987	.991	.992	.993	.994	.996
79	.957	.970	.975	.978	.980	.987	.991	.992	.993	.994	.996
81	0.958	0.970	0.975	0.979	0.981	0.987	0.992	0.992	0.993	0.994	0.996
83	.960	.971	.976	.979	.981	.988	.992	.992	.993	.994	.996
85	.961	.972	.977	.980	.981	.988	.992	.992	.993	.994	.996
87	.961	.972	.977	.980	.982	.988	.992	.993	.994	.994	.996
89	.961	.972	.977	.981	.982	.988	.992	.993	.994	.995	.996
91	0.962	0.973	0.978	0.981	0.983	0.989	0.992	0.993	0.994	0.995	0.996
93	.963	.973	.979	.981	.983	.989	.992	.993	.994	.995	.996
95	.965	.974	.979	.981	.983	.989	.993	.993	.994	.995	.996
97	.965	.975	.979	.982	.984	.989	.993	.993	.994	.995	.996
99	.967	.976	.980	.982	.984	.989	.993	.994	.994	.995	.996

continuous and discrete distributions. The W and the W' tests were performed on each sample. Using a 10 percent confidence level, a count was made of the number of times, out of a possible 100, that the W and W' tests rejected the hypothesis that the sample came from a normal population.

In general, the W' test appeared to be more sensitive than the W test when the alternative distribution was continuous and symmetric with high fourth moment (as compared to the normal distribution), when it was near normal and when it was discrete and skewed. The two tests appeared to be equivalent for alternative distributions which were continuous and asymmetric with high fourth moment and discrete and symmetric. The W test is superior to the W' test for other alternative distributions. Overall, the differentials in the power were small.

An earlier article [6] reported that for many alternative distributions, the W test has power as good or better than the following test procedures: $\sqrt{b_1}$, b_2, Kolmogorov-Smirnov, Cramer-Von Mises, Weighted Cramer-Von Mises, Durbin, chi-squared and David's u. Thus the W' test would also have this property.

4. CONCLUDING REMARKS

An approximate and simplified version of the W test was developed by estimating the slope of a regression line by simple least squares instead of the more precise procedure using generalized least squares. The operational constants for the W' test are easier to obtain and it is not limited to sample sizes less than 51. The sampling studies undertaken seem to indicate that the sensitivities of the W test and the approximate W' test were similar and any differences depended on the characteristics of the alternative distributions. Consequently, one was either less powerful, equal or more powerful than the other, depending on the nature of the alternative distribution. The use of the W' procedure for sample sizes greater than 50 is recommended and the extension to sample sizes over 100 can readily be accomplished. However, one must be careful in such extrapolations. In [7] and [8] it was shown that the chi-squared goodness-of-fit test had zero asymptotic efficiency relative to the Kolmogorov-Smirnov test while the results of [6] indicated that for sample size 50 for many alternative distributions the chi-squared procedure was the more powerful. This clearly cannot hold true as the sample size increases and a crossover point will eventually be reached.

[Received June 1970. Revised June 1971]

REFERENCES

[1] Gupta, A. K., "Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample," *Biometrika*, 39 (December 1952), 260–73.

[2] Harter, H. L., "Expected Values of Normal Order Statistics," *Biometrika*, 48, Nos. 1 and 2 (June 1961), 151–65.

[3] IBM Application Program, Document No. H20-0205-3, IBM Corporation, Poughkeepsie, N.Y., 77.

[4] Sarhan, A.E., and Greenberg, B. G., "Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples Part I," *Annals of Mathematical Statistics*, 27 (June 1956), 427–51.

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

[6] ———, Wilk, M. B., and Chen, H. J., "A Comparative Study of Various Tests for Normality," *Journal of the American Statistical Association*, 63 (December 1968), 1343–72.

[7] Wald, A., and Wolfowitz, J., "Confidence Limits for Continuous Distribution Functions," *Annals of Mathematical Statistics*, 10 (March 1939), 105–18.

[8] ——— and Wolfowitz, J., "Notes on Confidence Limits for Continuous Distribution Functions," *Annals of Mathematical Statistics*, 12 (March 1941), 118–19.