An Analysis of Variance Test for Normality (Complete Samples)
Author(s): S. S. Shapiro and M. B. Wilk
Source: *Biometrika*, Vol. 52, No. 3/4 (Dec., 1965), pp. 591-611
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2333709

Biometrika (1965), **52**, 3 *and* 4, *p.* 591
With 5 *text-figures*
Printed in Great Britain

An analysis of variance test for normality (complete samples)†

BY S. S. SHAPIRO AND M. B. WILK
General Electric Co. and Bell Telephone Laboratories, Inc.

1. INTRODUCTION

The main intent of this paper is to introduce a new statistical procedure for testing a complete sample for normality. The test statistic is obtained by dividing the square of an appropriate linear combination of the sample order statistics by the usual symmetric estimate of variance. This ratio is both scale and origin invariant and hence the statistic is appropriate for a test of the composite hypothesis of normality.

Testing for distributional assumptions in general and for normality in particular has been a major area of continuing statistical research—both theoretically and practically. A possible cause of such sustained interest is that many statistical procedures have been derived based on particular distributional assumptions—especially that of normality. Although in many cases the techniques are more robust than the assumptions underlying them, still a knowledge that the underlying assumption is incorrect may temper the use and application of the methods. Moreover, the study of a body of data with the stimulus of a distributional test may encourage consideration of, for example, normalizing transformations and the use of alternate methods such as distribution-free techniques, as well as detection of gross peculiarities such as outliers or errors.

The test procedure developed in this paper is defined and some of its analytical properties described in §2. Operational information and tables useful in employing the test are detailed in §3 (which may be read independently of the rest of the paper). Some examples are given in §4. Section 5 consists of an extract from an empirical sampling study of the comparison of the effectiveness of various alternative tests. Discussion and concluding remarks are given in §6.

2. THE W TEST FOR NORMALITY (COMPLETE SAMPLES)

2·1. *Motivation and early work*

This study was initiated, in part, in an attempt to summarize formally certain indications of probability plots. In particular, could one condense departures from statistical linearity of probability plots into one or a few 'degrees of freedom' in the manner of the application of analysis of variance in regression analysis?

In a probability plot, one can consider the regression of the ordered observations on the expected values of the order statistics from a standardized version of the hypothesized distribution—the plot tending to be linear if the hypothesis is true. Hence a possible method of testing the distributional assumption is by means of an analysis of variance type procedure. Using generalized least squares (the ordered variates are correlated) linear and higher-order models can be fitted and an F-type ratio used to evaluate the adequacy of the linear fit.

† Part of this research was supported by the Office of Naval Research while both authors were at Rutgers University.

This approach was investigated in preliminary work. While some promising results were obtained, the procedure is subject to the serious shortcoming that the selection of the higher-order model is, practically speaking, arbitrary. However, research is continuing along these lines.

Another analysis of variance viewpoint which has been investigated by the present authors is to compare the squared slope of the probability plot regression line, which under the normality hypothesis is an estimate of the population variance multiplied by a constant, with the residual mean square about the regression line, which is another estimate of the variance. This procedure can be used with incomplete samples and has been described elsewhere (Shapiro & Wilk, 1965*b*).

As an alternative to the above, for complete samples, the squared slope may be compared with the usual symmetric sample sum of squares about the mean which is independent of the ordering and easily computable. It is this last statistic that is discussed in the remainder of this paper.

2·2. *Derivation of the W statistic*

Let $m' = (m_1, m_2, \ldots, m_n)$ 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 $x_1 \leqslant x_2 \leqslant \ldots x_n$ denotes an ordered random sample of size n from a normal distribution with mean 0 and variance 1, then

$$E(x)_i = m_i \quad (i = 1, 2, \ldots, n),$$

and

$$\operatorname{cov}(x_i, x_j) = v_{ij} \quad (i, j = 1, 2, \ldots, n).$$

Let $y' = (y_1, \ldots, y_n)$ denote a vector of ordered random observations. The objective is to derive a test for the hypothesis that this is a sample from a normal distribution with unknown mean μ and unknown variance σ^2.

Clearly, if the $\{y_i\}$ are a normal sample then y_i may be expressed as

$$y_i = \mu + \sigma x_i \quad (i = 1, 2, \ldots, n).$$

It follows from the generalized least-squares theorem (Aitken, 1935; Lloyd, 1952) that the best linear unbiased estimates of μ and σ are those quantities that minimize the quadratic form $(y - \mu 1 - \sigma m)' V^{-1} (y - \mu 1 - \sigma m)$, where $1' = (1, 1, \ldots, 1)$. These estimates are, respectively,

$$\hat{\mu} = \frac{m' V^{-1} (m1' - 1m') V^{-1} y}{1' V^{-1} 1 m' V^{-1} m - (1' V^{-1} m)^2}$$

and

$$\hat{\sigma} = \frac{1' V^{-1} (1m' - m1') V^{-1} y}{1' V^{-1} 1 m' V^{-1} m - (1' V^{-1} m)^2}.$$

For symmetric distributions, $1' V^{-1} m = 0$, and hence

$$\hat{\mu} = \frac{1}{n} \sum_{1}^{n} y_i = \bar{y}, \quad \text{and} \quad \hat{\sigma} = \frac{m' V^{-1} y}{m' V^{-1} m}.$$

Let

$$S^2 = \sum_{1}^{n} (y_i - \bar{y})^2$$

denote the usual symmetric unbiased estimate of $(n-1)\sigma^2$.

The W test statistic for normality is defined by

$$W = \frac{R^4 \hat{\sigma}^2}{C^2 S^2} = \frac{b^2}{S^2} = \frac{(a'y)^2}{S^2} = \left(\sum_{i=1}^{n} a_i y_i\right)^2 \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

where
$$R^2 = m'V^{-1}m,$$
$$C^2 = m'V^{-1}V^{-1}m,$$
$$a' = (a_1, \ldots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{\frac{1}{2}}}$$
and
$$b = R^2\hat{\sigma}/C.$$

Thus, b is, up to the normalizing constant C, the best linear unbiased estimate of the slope of a linear regression of the ordered observations, y_i, on the expected values, m_i, of the standard normal order statistics. The constant C is so defined that the linear coefficients are normalized.

It may be noted that if one is indeed sampling from a normal population then the numerator, b^2, and denominator, S^2, of W are both, up to a constant, estimating the same quantity, namely σ^2. For non-normal populations, these quantities would not in general be estimating the same thing. Heuristic considerations augmented by some fairly extensive empirical sampling results (Shapiro & Wilk, 1964*a*) using populations with a wide range of $\sqrt{\beta_1}$ and β_2 values, suggest that the mean values of W for non-null distributions tends to shift to the left of that for the null case. Further it appears that the variance of the null distribution of W tends to be smaller than that of the non-null distribution. It is likely that this is due to the positive correlation between the numerator and denominator for a normal population being greater than that for non-normal populations.

Note that the coefficients $\{a_i\}$ are just the normalized 'best linear unbiased' coefficients tabulated in Sarhan & Greenberg (1956).

2·3. *Some analytical properties of W*

LEMMA 1. *W is scale and origin invariant*

Proof. This follows from the fact that for normal (more generally symmetric) distributions,
$$-a_i = a_{n-i+1}$$

COROLLARY 1. *W has a distribution which depends only on the sample size n, for samples from a normal distribution.*

COROLLARY 2. *W is statistically independent of S^2 and of $\bar{y}$, for samples from a normal distribution.*

Proof. This follows from the fact that $\bar{y}$ and S^2 are sufficient for μ and σ^2 (Hogg & Craig, 1956).

COROLLARY 3. $EW^r = Eb^{2r}/ES^{2r}$, *for any r.*

LEMMA 2. *The maximum value of W is* 1.

Proof. Assume $\bar{y} = 0$ since W is origin invariant by Lemma 1. Hence
$$W = [\sum_i a_i y_i]^2 / \sum_i y_i^2.$$
Since
$$(\sum_i a_i y_i)^2 \leqslant \sum_i a_i^2 \sum_i y_i^2 = \sum_i y_i^2,$$
because $\sum_i a_i^2 = a'a = 1$, by definition, then W is bounded by 1. This maximum is in fact achieved when $y_i = \eta a_i$, for arbitrary η.

LEMMA 3. *The minimum value of W is* $na_1^2/(n-1)$.

Proof.† (Due to C. L. Mallows.) Since W is scale and origin invariant, it suffices to consider the maximization of $\sum_{i=1}^{n} y_i^2$ subject to the constraints $\Sigma y_i = 0$, $\Sigma a_i y_i = 1$. Since this is a convex region and Σy_i^2 is a convex function, the maximum of the latter must occur at one of the $(n-1)$ vertices of the region. These are

$$\left(\frac{(n-1)}{na_1}, \frac{-1}{na_1}, \dots \frac{-1}{na_1}\right)$$
$$\left(\frac{n-2}{n(a_1+a_2)}, \frac{(n-2)}{n(a_1+a_2)}, \frac{-2}{n(a_1+a_2)}, \dots, \frac{-2}{n(a_1+a_2)}\right)$$
$$\vdots$$
$$\left(\frac{1}{n(a_1+\dots+a_{n-1})}, \frac{1}{n(a_1+\dots+a_{n-1})}, \dots, \frac{-(n-1)}{n(a_1+\dots+a_{n-1})}\right).$$

It can now be checked numerically, for the values of the specific coefficients $\{a_i\}$, that the maximum of $\sum_{i=1}^{n} y_1^2$ occurs at the first of these points and the corresponding minimum value of W is as given in the Lemma.

Lemma 4. *The half and first moments of W are given by*

$$EW^{\frac{1}{2}} = \frac{R^2\,\Gamma\{\frac{1}{2}(n-1)\}}{C\Gamma(\frac{1}{2}n)\sqrt{2}}$$

and

$$EW = \frac{R^2(R^2+1)}{C^2(n-1)},$$

where $R^2 = m'V^{-1}m$, and $C^2 = m'V^{-1}V^{-1}m$.

Proof. Using Corollary 3 of Lemma 1,

$$EW^{\frac{1}{2}} = Eb/ES \quad \text{and} \quad EW = Eb^2/ES^2.$$

Now,

$$ES = \sigma\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)\Big/\Gamma\left(\frac{n-1}{2}\right) \quad \text{and} \quad ES^2 = (n-1)\,\sigma^2.$$

From the general least squares theorem (see e.g. Kendall & Stuart, vol. II (1961)),

$$Eb = \frac{R^2}{C}E\hat{\sigma} = \frac{R^2}{C}\sigma$$

and

$$Eb^2 = \frac{R^4}{C^2}E\hat{\sigma}^2 = \frac{R^4}{C^2}\{\operatorname{var}(\hat{\sigma}) + (E\hat{\sigma})^2\}$$
$$= \sigma^2R^2(R^2+1)/C^2,$$

since $\operatorname{var}(\hat{\sigma}) = \sigma^2/m'V^{-1}m = \sigma^2/R^2$, and hence the results of the lemma follow.

Values of these moments are shown in Fig. 1 for sample sizes $n = 3(1)\,20$.

Lemma 5. *A joint distribution involving W is defined by*

$$h(W, \theta_2, \dots, \theta_{n-2}) = KW^{-\frac{1}{2}}(1-W)^{\frac{1}{2}(n-4)}\cos^{n-4}\theta_2 \dots \cos\theta_{n-3},$$

over a region T on which the θ_i's and W are not independent, and where K is a constant.

† Lemma 3 was conjectured intuitively and verified by certain numerical studies. Subsequently the above proof was given by C. L. Mallows.

Proof. Consider an orthogonal transformation B such that $y = Bu$, where

$$u_1 = \sum_{i=1}^{n} y_i/\sqrt{n} \quad \text{and} \quad u_2 = \sum_{i=1}^{n} a_i y_i = b.$$

The ordered y_i's are distributed as

$$n!\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}n} \exp\left\{-\tfrac{1}{2}\sum_i\left(\frac{y_i-\mu}{\sigma}\right)^2\right\} \quad (-\infty < y_1 < \ldots < y_n < \infty).$$

After integrating out, u_1, the joint density for $u_2, \ldots, u_n$ is

$$K^* \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n} u_i^2\right\}$$

over the appropriate region T^*. Changing to polar co-ordinates such that

$$u_2 = \rho \sin\theta_1, \text{ etc},$$

and then integrating over ρ, yields the joint density of $\theta_1, \ldots, \theta_{n-2}$ as

$$K^{**} \cos^{n-3}\theta_1 \cos^{n-4}\theta_2 \ldots \cos\theta_{n-3},$$

over some region T^{**}.

From these various transformations

$$W = \frac{b^2}{S^2} = \frac{u_2^2}{\sum_{i=1}^{n} u_i^2} = \frac{\rho^2 \sin^2\theta_1}{\rho^2} = \sin^2\theta_1,$$

from which the lemma follows. The θ_i's and W are not independent, they are restricted in the sample space T.

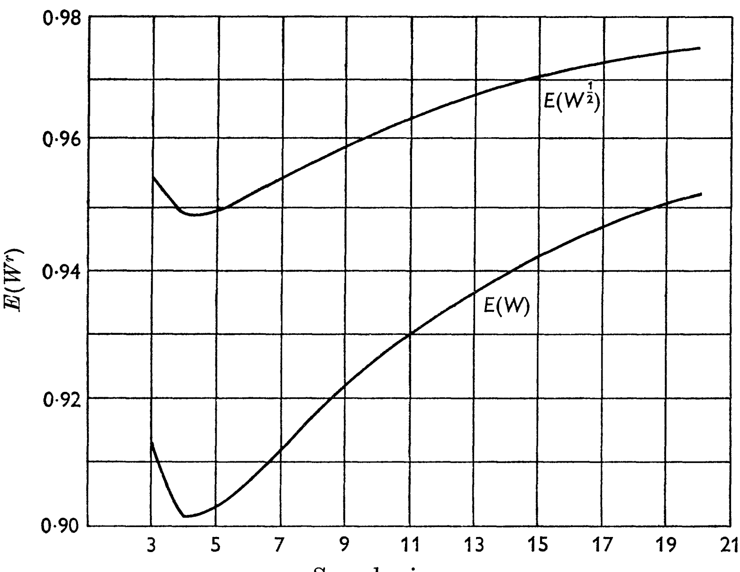

Fig. 1. Moments of W, $E(W^r)$, $n = 3(1)20$, $r = \frac{1}{2}, 1$.

Corollary 4. *For $n = 3$, the density of W is*

$$\frac{3}{\pi}(1-W)^{-\frac{1}{2}} W^{-\frac{1}{2}}, \quad \tfrac{3}{4} \leqslant W \leqslant 1.$$

Note that for $n = 3$, the W statistic is equivalent (up to a constant multiplier) to the statistic (range/standard deviation) advanced by David, Hartley & Pearson (1954) and the result of the corollary is essentially given by Pearson & Stephens (1964).

It has not been possible, for general n, to integrate out of the θ_i's of Lemma 5 to obtain an explicit form for the distribution of W. However, explicit results have also been given for $n = 4$, Shapiro (1964).

2·4. *Approximations associated with the W test*

The $\{a_i\}$ used in the W statistic are defined by

$$a_i = \sum_{j=1}^{n} m_j v^{ij}/C \quad (j = 1, 2, \ldots, n),$$

where m_j, v_{ij} and C have been defined in § 2·2. To determine the a_i directly it appears necessary to know both the vector of means m and the covariance matrix V. However, to date, the elements of V are known only up to samples of size 20 (Sarhan & Greenberg, 1956). Various approximations are presented in the remainder of this section to enable the use of W for samples larger than 20.

By definition,

$$a = \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{\frac{1}{2}}} = \frac{m' V^{-1}}{C}$$

is such that $a'a = 1$. Let $a^* = m' V^{-1}$, then $C^2 = a^{*\prime} a^*$. Suggested approximations are

$$\hat{a}_i^* = 2m_i \quad (i = 2, 3, \ldots, n-1),$$

and

$$\hat{a}_1^2 = \hat{a}_n^2 = \begin{cases} \dfrac{\Gamma(\frac{1}{2}n)}{\sqrt{2}\,\Gamma\{\frac{1}{2}(n+1)\}} & (n \leqslant 20), \\[2ex] \dfrac{\Gamma\{\frac{1}{2}(n+1)\}}{\sqrt{2}\,\Gamma(\frac{1}{2}n+1)} & (n > 20). \end{cases}$$

A comparison of a_i^* (the exact values) and $\hat{a}_i^*$ for various values of $i \neq 1$ and $n = 5$, 10, 15, 20 is given in Table 1. (Note $a_i = -a_{n-i+1}$.) It will be seen that the approximation is generally in error by less than 1%, particularly as n increases. This encourages one to trust the use of this approximation for $n > 20$. Necessary values of the m_i for this approximation are available in Harter (1961).

Table 1. *Comparison of* $|a_i^*|$ *and* $|\hat{a}_i^*| = |2m_i|$, *for selected values of* $i(\neq 1)$ *and* n

n	$i =$	2	3	4	5	8	10
5	Exact	1·014	0·0	—	—	—	—
	Approx.	0·990	0·0	—	—	—	—
10	Exact	2·035	1·324	0·757	0·247	—	—
	Approx.	2·003	1·312	0·752	0·245	—	—
15	Exact	2·530	1·909	1·437	1·036	0·0	—
	Approx.	2·496	1·895	1·430	1·031	0·0	—
20	Exact	2·849	2·277	1·850	1·496	0·631	0·124
	Approx.	2·815	2·262	1·842	1·491	0·630	0·124

A comparison of a_1^2 and $\hat{a}_1^2$ for $n = 6(1)\,20$ is given in Table 2. While the errors of this approximation are quite small for $n \leqslant 20$, the approximation and true values appear to cross over at $n = 19$. Further comparisons with other approximations, discussed below, suggested the changed formulation of $\hat{a}_1^2$ for $n > 20$ given above.

Table 2. *Comparison of a_1^2 and $\hat{a}_1^2$*

n	Exact	Approximate	n	Exact	Approximate
6	0·414	0·426	13	0·287	0·283
7	·388	·392	14	·276	·272
8	·366	·365	15	·265	·261
9	·347	·343	16	·256	·254
10	·329	·324	17	·247	·245
11	·314	·308	18	·239	·237
12	·300	·295	19	·231	·231
			20	·224	·226

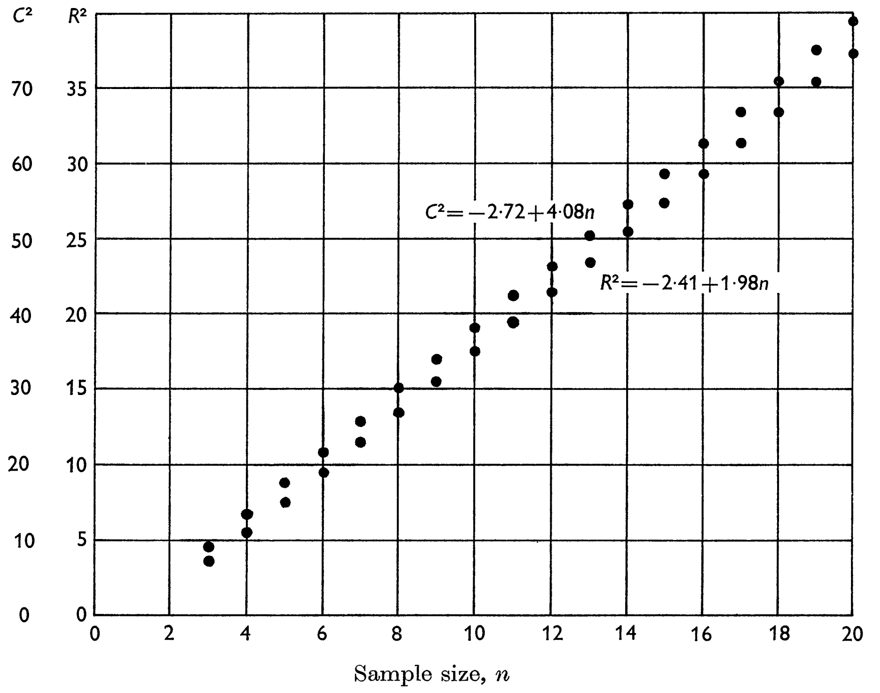

Fig. 2. Plot of $C^2 = m'V^{-1}V^{-1}m$ and $R^2 = m'V^{-1}m$ as functions of the sample size n.

What is required for the W test are the normalized coefficients $\{a_i\}$. Thus $\hat{a}_1^2$ is directly usable but the $\hat{a}_i^*$ $(i = 2, \ldots, n-1)$, must be normalized by division by $C = (m'V^{-1}V^{-1}m)^{\frac{1}{2}}$.

A plot of the values of C^2 and of $R^2 = m'V^{-1}m$ as a function of n is given in Fig. 2. The linearity of these may be summarized by the following least-squares equations:

$$C^2 = -2{\cdot}722 + 4{\cdot}083n,$$

which gave a regression mean square of 7331·6 and a residual mean square of 0·0186, and

$$R^2 = -2{\cdot}411 + 1{\cdot}981n,$$

with a regression mean square of 1725·7 and a residual mean square of 0·0016.

These results encourage the use of the extrapolated equations to estimate C^2 and R^2 for higher values of n.

A comparison can now be made between values of C^2 from the extrapolation equation and from $\sum_1^n \hat{a}_i^{*2}$, using

$$\hat{a}_1^{*2} = \frac{\hat{a}_1^2}{1-2\hat{a}_1^2} \sum_2^{n-1} \hat{a}_i^{*2}.$$

For the case $n = 30$, these give values of 119·77 and 120·47, respectively. This concordance of the independent approximations increases faith in both.

Plackett (1958) has suggested approximations for the elements of the vector a and R^2. While his approximations are valid for a wide range of distributions and can be used with censored samples, they are more complex, for the normal case, than those suggested above. For the normal case his approximations are

$$\tilde{a}_j^* = nm_j[F(m_{j+1}) - F(m_{j-1})] \quad (j = 2, 3, \ldots, n-1),$$

$$\tilde{a}_j^* = n\left\{\frac{m_j f(m_j)^2}{F(m_j)} + m_j^2 f(m_j) - f(m_j) + m_j[F(m_{j+1}) - F(m_j)]\right\} \quad (j = 1),$$

where $F(m_j)$ = cumulative distribution evaluated at m_j,

$f(m_j)$ = density function evaluated at m_j,

and $$\tilde{a}_1^* = -\tilde{a}_n^*.$$

Plackett's approximation to R^2 is

$$\tilde{R}^2 = 2\left\{\frac{m_1^2 f(m_1)^2}{F(m_1)} + m_1^3 f(m_1) + m_1 f(m_1) - 2F(m_1) + 1\right\}.$$

Plackett's $\tilde{a}_i^*$ approximations and the present $\hat{a}_i^*$ approximations are compared with the exact values, for sample size 20, in Table 3. In addition a consistency comparison of the two approximations is given for sample size 30. Plackett's result for a_1 ($n = 20$) was the only case where his approximation was closer to the true value than the simpler approximations suggested above. The differences in the two approximations for a_1 were negligible, being less than 0·5 %. Both methods give good approximations, being off no more than three units in the second decimal place. The comparison of the two methods for $n = 30$ shows good agreement, most of the differences being in the third decimal place. The largest discrepancy occurred for $i = 2$; the estimates differed by six units in the second decimal place, an error of less than 2 %.

The two methods of approximating R^2 were compared for $n = 20$. Plackett's method gave a value of 36·09, the method suggested above gave a value of 37·21 and the true value was 37·26.

The good practical agreement of these two approximations encourages the belief that there is little risk in reasonable extrapolations for $n > 20$. The values of constants, for $n > 20$, given in §3 below, were estimated from the simple approximations and extrapolations described above.

As a further internal check the values of a_n, a_{n-1} and a_{n-4} were plotted as a function of n for $n = 3(1)50$. The plots are shown in Fig. 3 which is seen to be quite smooth for each of the three curves at the value $n = 20$. Since values for $n \leqslant 20$ are 'exact' the smooth transition lends credence to the approximations for $n > 20$.

Table 3. *Comparison of approximate values of* $a^* = m'V^{-1}$

n	i	Present approx.	Exact	Plackett
20	1	−4·223	−4·2013	−4·215
	2	−2·815	−2·8494	−2·764
	3	−2·262	−2·2765	−2·237
	4	−1·842	−1·8502	−1·820
	5	−1·491	−1·4960	−1·476
	6	−1·181	−1·1841	−1·169
	7	−0·897	−0·8990	−0·887
	8	−0·630	−0·6314	−0·622
	9	−0·374	−0·3784	−0·370
	10	−0·124	−0·1243	−0·123
30	1	−4·655	—	−4·671
	2	−3·231	—	−3·170
	3	−2·730	—	−2·768
	4	−2·357	—	−2·369
	5	−2·052	—	−2·013
	6	−1·789	—	−1·760
	7	−1·553	—	−1·528
	8	−1·338	—	−1·334
	9	−1·137	—	−1·132
	10	−0·947	—	−0·941
	11	−0·765	—	−0·759
	12	−0·589	—	−0·582
	13	−0·418	—	−0·413
	14	−0·249	—	−0·249
	15	−0·083	—	−0·082

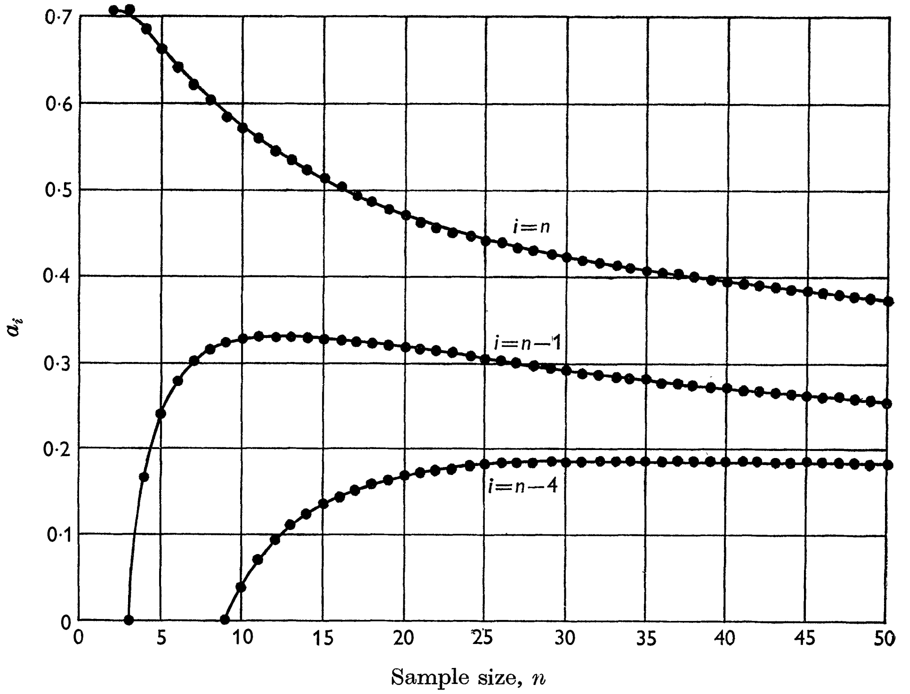

Fig. 3. a_i plotted as a function of sample size, $n = 2(1)50$, for $i = n, n-1, n-4$ $(n > 8)$.

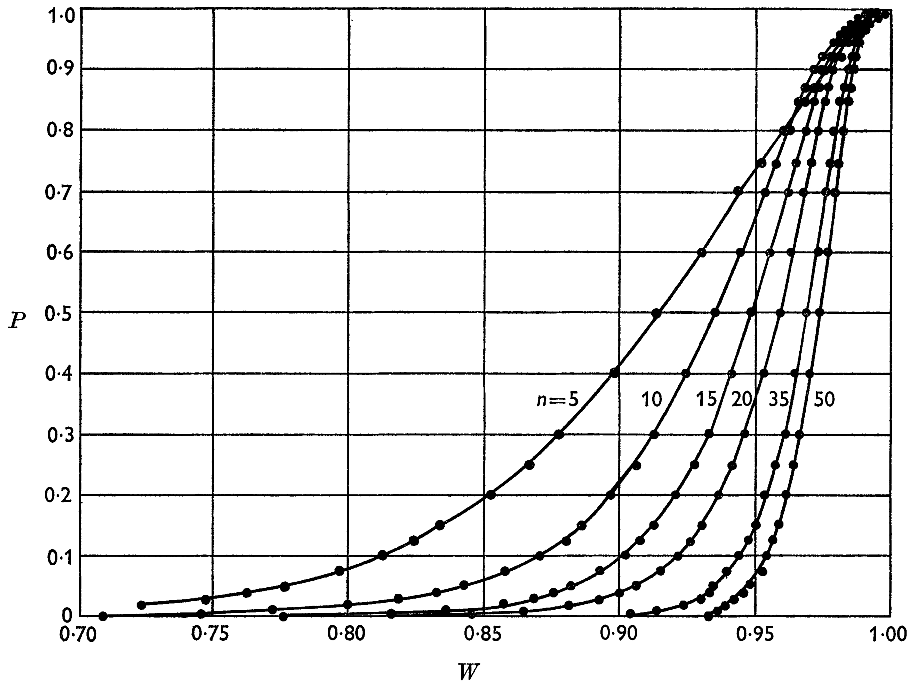

Fig. 4. Empirical c.d.f. of W for $n = 5, 10, 15, 20, 35, 50$.

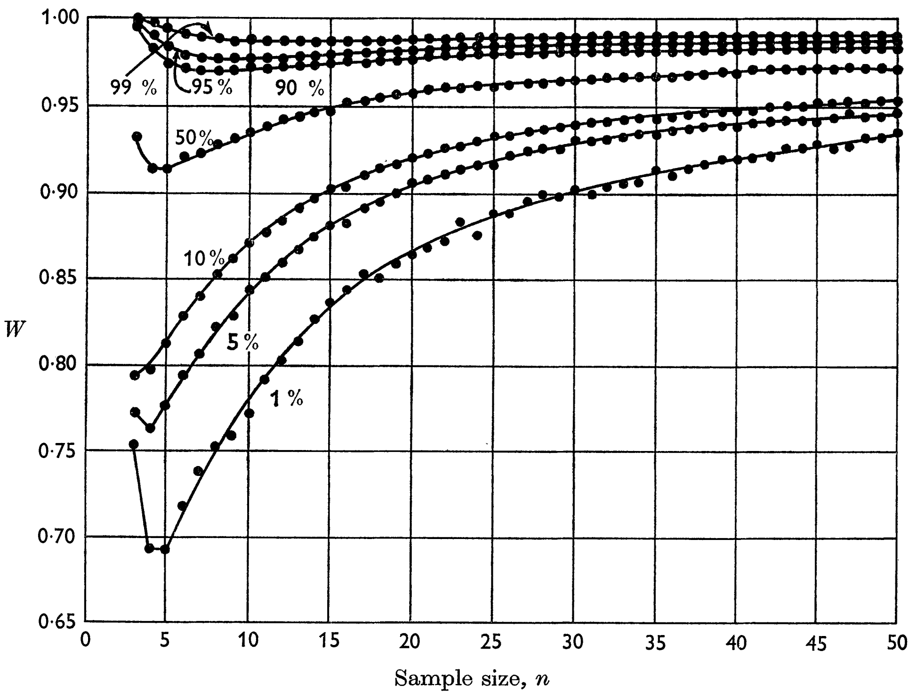

Fig. 5. Selected empirical percentage points of W, $n = 3(1)50$.

Table 4. *Some theoretical moments (μ_i) and Monte Carlo moments ($\hat{\mu}_i$) of W*

n	$\mu_{\frac{1}{2}}$	$\hat{\mu}_{\frac{1}{2}}$	μ_1	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3/\hat{\mu}_2^{\frac{3}{2}}$	$\hat{\mu}_4/\hat{\mu}_2^2$
3	0·9549	0·9547	0·9135	0·9130	0·005698	−0·5930	2·3748
4	·9486	·9489	·9012	·9019	·005166	− ·8944	3·7231
5	·9494	·9491	·9026	·9021	·004491	− ·8176	7·8126
6	0·9521	0·9525	0·9072	0·9082	0·003390	−1·1790	5·4295
7	·9547	·9545	·9123	·9120	·002995	−1·3229	6·4104
8	·9574	·9575	·9174	·9175	·002470	−1·3841	7·1092
9	·9600	·9596	·9221	·9215	·002293	−1·5987	8·4482
10	·9622	·9620	·9264	·9260	·001972	−1·6655	9·2812
11	0·9643	0·9639	0·9303	0·9295	0·001717	−1·7494	11·0547
12	·9661	·9661	·9337	·9338	·001483	−1·7744	11·9185
13	·9678	·9678	·9369	·9369	·001316	−1·7581	13·0769
14	·9692	·9693	·9398	·9399	·001168	−1·9025	14·0568
15	·9706	·9705	·9424	·9422	·001023	−1·8876	16·7383
16	0·9718	0·9717	0·9447	0·9445	0·000964	−1·7968	17·6669
17	·9730	·9730	·9470	·9470	·000823	−1·9468	22·1972
18	·9741	·9741	·9491	·9492	·000810	−2·1391	24·7776
19	·9750	·9750	·9508	·9509	·000711	−2·1305	29·7333
20	·9757	·9760	·9523	·9527	·000651	−2·2761	32·5906
21	—	0·9771	—	0·9549	0·000594	−2·2827	36·0382
22	—	·9776	—	·9558	·000568	−2·3984	44·5617
23	—	·9782	—	·9570	·000504	−2·1862	40·7507
24	—	·9787	—	·9579	·000504	−2·3517	43·4926
25	—	·9789	—	·9584	·000458	−2·3448	46·3318
26	—	0·9796	—	0·9598	0·000421	−2·4978	58·9446
27	—	·9801	—	·9607	·000404	−2·5903	60·5200
28	—	·9805	—	·9615	·000382	−2·6964	64·1702
29	—	·9810	—	·9624	·000369	−2·6090	68·9591
30	—	·9811	—	·9626	·000344	−2·7288	71·7714
31	—	0·9816	—	0·9636	0·000336	−2·7997	77·4744
32	—	·9819	—	·9642	·000326	−2·6900	76·8384
33	—	·9823	—	·9650	·000308	−3·0181	93·2496
34	—	·9825	—	·9654	·000293	−3·0166	100·4419
35	—	·9827	—	·9658	·000268	−2·8574	108·5077
36	—	0·9829	—	0·9662	0·000264	−2·7965	91·7985
37	—	·9833	—	·9670	·000253	−3·1566	120·0005
38	—	·9837	—	·9677	·000235	−3·0679	118·2513
39	—	·9837	—	·9678	·000239	−3·3283	134·3110
40	—	·9839	—	·9682	·000229	−3·1719	136·4787
41	—	0·9840	—	0·9684	0·000227	−3·0740	129·9604
42	—	·9844	—	·9691	·000212	−3·2885	136·3814
43	—	·9846	—	·9694	·000196	−3·2646	151·7350
44	—	·9846	—	·9695	·000193	−3·0803	140·2724
45	—	·9849	—	·9701	·000192	−3·1645	137·2297
46	—	0·9850	—	0·9703	0·000184	−3·3742	176·0635
47	—	·9854	—	·9710	·000170	−3·3353	179·2792
48	—	·9853	—	·9708	·000179	−3·2972	173·6601
49	—	·9855	—	·9712	·000165	−3·2810	183·9433
50	—	·9855	—	·9714	·000154	−3·3240	212·4279

2·5. *Approximation to the distribution of W*

The complexity in the domain of the joint distribution of W and the angles $\{\theta_i\}$ in Lemma 5 necessitates consideration of an approximation to the null distribution of W. Since only the first and second moments of normal order statistics are, practically, available, it follows that only the one-half and first moments of W are known. Hence a technique such as the Cornish–Fisher expansion cannot be used.

In the circumstance it seemed both appropriate and efficient to employ empirical sampling to obtain an approximation for the null distribution.

Accordingly, normal random samples were obtained from the Rand Tables (Rand Corp. (1955)). Repeated values of W were computed for $n = 3(1)50$ and the empirical percentage points determined for each value of n. The number of samples, m, employed was as follows:

$$\text{for}\quad n = 3(1)\,20, \qquad m = 5000,$$
$$n = 21(1)\,50, \quad m = \left[\frac{100{,}000}{n}\right].$$

Fig. 4 gives the empirical C.D.F.'s for values of $n = 5$, 10, 15, 20, 35, 50. Fig. 5 gives a plot of the 1, 5, 10, 50, 90, 95, and 99 empirical percentage points of W for $n = 3(1)50$.

A check on the adequacy of the sampling study is given by comparing the empirical one-half and the first moments of the sample with the corresponding theoretical moments of W for $n = 3(1)20$. This comparison is given in Table 4, which provides additional assurance of the adequacy of the sampling study. Also in Table 4 are given the sample variance and the standardized third and fourth moments for $n = 3(1)50$.

After some preliminary investigation, the S_B system of curves suggested by Johnson (1949) was selected as a basis for smoothing the empirical null W distribution. Details of this procedure and its results are given in Shapiro & Wilk (1965*a*). The tables of percentage points of W given in §3 are based on these smoothed sampling results.

3. Summary of operational information

The objective of this section is to bring together all the tables and descriptions needed to execute the W test for normality. This section may be employed independently of notational or other information from other sections.

The object of the W test is to provide an index or test statistic to evaluate the supposed normality of a complete sample. The statistic has been shown to be an effective measure of normality even for small samples ($n < 20$) against a wide spectrum of non-normal alternatives (see §5 below and Shapiro & Wilk (1964*a*)).

The W statistic is scale and origin invariant and hence supplies a test of the composite null hypothesis of normality.

To compute the value of W, given a complete random sample of size n, $x_1, x_2, \ldots, x_n$, one proceeds as follows:

(i) Order the observations to obtain an ordered sample $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$.

(ii) Compute
$$S^2 = \sum_1^n (y_i - \bar{y})^2 = \sum_1^n (x_i - \bar{x})^2.$$

(iii) (*a*) If n is even, $n = 2k$, compute

$$b = \sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i),$$

where the values of a_{n-i+1} are given in Table 5.

(*b*) If n is odd, $n = 2k+1$, the computation is just as in (iii) (*a*), since $a_{k+1} = 0$ when $n = 2k+1$. Thus one finds

$$b = a_n(y_n - y_1) + \ldots + a_{k+2}(y_{k+2} - y_k),$$

where the value of y_{k+1}, the sample median, does not enter the computation of b.

(iv) Compute $W = b^2/S^2$.

(v) 1, 2, 5, 10, 50, 90, 95, 98 and 99 % points of the distribution of W are given in Table 6. *Small values of W are significant*, i.e. indicate non-normality.

(vi) A more precise significance level may be associated with an observed W value by using the approximation detailed in Shapiro & Wilk (1965*a*).

Table 5. *Coefficients* $\{a_{n-i+1}\}$ *for the W test for normality, for* $n = 2(1)50$.

i \ n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0·7071	0·7071	0·6872	0·6646	0·6431	0·6233	0·6052	0·5888	0·5739
2	—	·0000	·1677	·2413	·2806	·3031	·3164	·3244	·3291
3	—	—	—	·0000	·0875	·1401	·1743	·1976	·2141
4	—	—	—	—	—	·0000	·0561	·0947	·1224
5	—	—	—	—	—	—	—	·0000	·0399

i \ n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0·5601	0·5475	0·5359	0·5251	0·5150	0·5056	0·4968	0·4886	0·4808	0·4734
2	·3315	·3325	·3325	·3318	·3306	·3290	·3273	·3253	·3232	·3211
3	·2260	·2347	·2412	·2460	·2495	·2521	·2540	·2553	·2561	·2565
4	·1429	·1586	·1707	·1802	·1878	·1939	·1988	·2027	·2059	·2085
5	·0695	·0922	·1099	·1240	·1353	·1447	·1524	·1587	·1641	·1686
6	0·0000	0·0303	0·0539	0·0727	0·0880	0·1005	0·1109	0·1197	0·1271	0·1334
7	—	—	·0000	·0240	·0433	·0593	·0725	·0837	·0932	·1013
8	—	—	—	—	·0000	·0196	·0359	·0496	·0612	·0711
9	—	—	—	—	—	—	·0000	·0163	·0303	·0422
10	—	—	—	—	—	—	—	—	·0000	·0140

i \ n	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1	0·4643	0·4590	0·4542	0·4493	0·4450	0·4407	0·4366	0·4328	0·4291	0·4254
2	·3185	·3156	·3126	·3098	·3069	·3043	·3018	·2992	·2968	·2944
3	·2578	·2571	·2563	·2554	·2543	·2533	·2522	·2510	·2499	·2487
4	·2119	·2131	·2139	·2145	·2148	·2151	·2152	·2151	·2150	·2148
5	·1736	·1764	·1787	·1807	·1822	·1836	·1848	·1857	·1864	·1870
6	0·1399	0·1443	0·1480	0·1512	0·1539	0·1563	0·1584	0·1601	0·1616	0·1630
7	·1092	·1150	·1201	·1245	·1283	·1316	·1346	·1372	·1395	·1415
8	·0804	·0878	·0941	·0997	·1046	·1089	·1128	·1162	·1192	·1219
9	·0530	·0618	·0696	·0764	·0823	·0876	·0923	·0965	·1002	·1036
10	·0263	·0368	·0459	·0539	·0610	·0672	·0728	·0778	·0822	·0862
11	0·0000	0·0122	0·0228	0·0321	0·0403	0·0476	0·0540	0·0598	0·0650	0·0697
12	—	—	·0000	·0107	·0200	·0284	·0358	·0424	·0483	·0537
13	—	—	—	—	·0000	·0094	·0178	·0253	·0320	·0381
14	—	—	—	—	—	—	·0000	·0084	·0159	·0227
15	—	—	—	—	—	—	—	—	·0000	·0076

Table 5. *Coefficients* $\{a_{n-i+1}\}$ *for the W test for normality, for* $n = 2(1)50$ (*cont.*)

i \ n	31	32	33	34	35	36	37	38	39	40
1	0·4220	0·4188	0·4156	0·4127	0·4096	0·4068	0·4040	0·4015	0·3989	0·3964
2	·2921	·2898	·2876	·2854	·2834	·2813	·2794	·2774	·2755	·2737
3	·2475	·2463	·2451	·2439	·2427	·2415	·2403	·2391	·2380	·2368
4	·2145	·2141	·2137	·2132	·2127	·2121	·2116	·2110	·2104	·2098
5	·1874	·1878	·1880	·1882	·1883	·1883	·1883	·1881	·1880	·1878
6	0·1641	0·1651	0·1660	0·1667	0·1673	0·1678	0·1683	0·1686	0·1689	0·1691
7	·1433	·1449	·1463	·1475	·1487	·1496	·1505	·1513	·1520	·1526
8	·1243	·1265	·1284	·1301	·1317	·1331	·1344	·1356	·1366	·1376
9	·1066	·1093	·1118	·1140	·1160	·1179	·1196	·1211	·1225	·1237
10	·0899	·0931	·0961	·0988	·1013	·1036	·1056	·1075	·1092	·1108
11	0·0739	0·0777	0·0812	0·0844	0·0873	0·0900	0·0924	0·0947	0·0967	0·0986
12	·0585	·0629	·0669	·0706	·0739	·0770	·0798	·0824	·0848	·0870
13	·0435	·0485	·0530	·0572	·0610	·0645	·0677	·0706	·0733	·0759
14	·0289	·0344	·0395	·0441	·0484	·0523	·0559	·0592	·0622	·0651
15	·0144	·0206	·0262	·0314	·0361	·0404	·0444	·0481	·0515	·0546
16	0·0000	0·0068	0·0131	0·0187	0·0239	0·0287	0·0331	0·0372	0·0409	0·0444
17	—	—	·0000	·0062	·0119	·0172	·0220	·0264	·0305	·0343
18	—	—	—	—	·0000	·0057	·0110	·0158	·0203	·0244
19	—	—	—	—	—	—	·0000	·0053	·0101	·0146
20	—	—	—	—	—	—	—	—	·0000	·0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0·3940	0·3917	0·3894	0·3872	0·3850	0·3830	0·3808	0·3789	0·3770	0·3751
2	·2719	·2701	·2684	·2667	·2651	·2635	·2620	·2604	·2589	·2574
3	·2357	·2345	·2334	·2323	·2313	·2302	·2291	·2281	·2271	·2260
4	·2091	·2085	·2078	·2072	·2065	·2058	·2052	·2045	·2038	·2032
5	·1876	·1874	·1871	·1868	·1865	·1862	·1859	·1855	·1851	·1847
6	0·1693	0·1694	0·1695	0·1695	0·1695	0·1695	0·1695	0·1693	0·1692	0·1691
7	·1531	·1535	·1539	·1542	·1545	·1548	·1550	·1551	·1553	·1554
8	·1384	·1392	·1398	·1405	·1410	·1415	·1420	·1423	·1427	·1430
9	·1249	·1259	·1269	·1278	·1286	·1293	·1300	·1306	·1312	·1317
10	·1123	·1136	·1149	·1160	·1170	·1180	·1189	·1197	·1205	·1212
11	0·1004	0·1020	0·1035	0·1049	0·1062	0·1073	0·1085	0·1095	0·1105	0·1113
12	·0891	·0909	·0927	·0943	·0959	·0972	·0986	·0998	·1010	·1020
13	·0782	·0804	·0824	·0842	·0860	·0876	·0892	·0906	·0919	·0932
14	·0677	·0701	·0724	·0745	·0765	·0783	·0801	·0817	·0832	·0846
15	·0575	·0602	·0628	·0651	·0673	·0694	·0713	·0731	·0748	·0764
16	0·0476	0·0506	0·0534	0·0560	0·0584	0·0607	0·0628	0·0648	0·0667	0·0685
17	·0379	·0411	·0442	·0471	·0497	·0522	·0546	·0568	·0588	·0608
18	·0283	·0318	·0352	·0383	·0412	·0439	·0465	·0489	·0511	·0532
19	·0188	·0227	·0263	·0296	·0328	·0357	·0385	·0411	·0436	·0459
20	·0094	·0136	·0175	·0211	·0245	·0277	·0307	·0335	·0361	·0386
21	0·0000	0·0045	0·0087	0·0126	0·0163	0·0197	0·0229	0·0259	0·0288	0·0314
22		—	·0000	·0042	·0081	·0118	·0153	·0185	·0215	·0244
23	—	—	—	—	·0000	·0039	·0076	·0111	·0143	·0174
24	—	—	—	—	—	—	·0000	·0037	·0071	·0104
25	—	—	—	—	—	—	—	—	·0000	·0035

Table 6. *Percentage points of the W test* for $n = 3(1)50$*

n	Level 0·01	0·02	0·05	0·10	0·50	0·90	0·95	0·98	0·99
3	0·753	0·756	0·767	0·789	0·959	0·998	0·999	1·000	1·000
4	·687	·707	·748	·792	·935	·987	·992	·996	·997
5	·686	·715	·762	·806	·927	·979	·986	·991	·993
6	0·713	0·743	0·788	0·826	0·927	0·974	0·981	0·986	0·989
7	·730	·760	·803	·838	·928	·972	·979	·985	·988
8	·749	·778	·818	·851	·932	·972	·978	·984	·987
9	·764	·791	·829	·859	·935	·972	·978	·984	·986
10	·781	·806	·842	·869	·938	·972	·978	·983	·986
11	0·792	0·817	0·850	0·876	0·940	0·973	0·979	0·984	0·986
12	·805	·828	·859	·883	·943	·973	·979	·984	·986
13	·814	·837	·866	·889	·945	·974	·979	·984	·986
14	·825	·846	·874	·895	·947	·975	·980	·984	·986
15	·835	·855	·881	·901	·950	·975	·980	·984	·987
16	0·844	0·863	0·887	0·906	0·952	0·976	0·981	0·985	0·987
17	·851	·869	·892	·910	·954	·977	·981	·985	·987
18	·858	·874	·897	·914	·956	·978	·982	·986	·988
19	·863	·879	·901	·917	·957	·978	·982	·986	·988
20	·868	·884	·905	·920	·959	·979	·983	·986	·988
21	0·873	0·888	0·908	0·923	0·960	0·980	0·983	0·987	0·989
22	·878	·892	·911	·926	·961	·980	·984	·987	·989
23	·881	·895	·914	·928	·962	·981	·984	·987	·989
24	·884	·898	·916	·930	·963	·981	·984	·987	·989
25	·888	·901	·918	·931	·964	·981	·985	·988	·989
26	0·891	0·904	0·920	0·933	0·965	0·982	0·985	0·988	0·989
27	·894	·906	·923	·935	·965	·982	·985	·988	·990
28	·896	·908	·924	·936	·966	·982	·985	·988	·990
29	·898	·910	·926	·937	·966	·982	·985	·988	·990
30	·900	·912	·927	·939	·967	·983	·985	·988	·900
31	0·902	0·914	0·929	0·940	0·967	0·983	0·986	0·988	0·990
32	·904	·915	·930	·941	·968	·983	·986	·988	·990
33	·906	·917	·931	·942	·968	·983	·986	·989	·990
34	·908	·919	·933	·943	·969	·983	·986	·989	·990
35	·910	·920	·934	·944	·969	·984	·986	·989	·990
36	0·912	0·922	0·935	0·945	0·970	0·984	0·986	0·989	0·990
37	·914	·924	·936	·946	·970	·984	·987	·989	·990
38	·916	·925	·938	·947	·971	·984	·987	·989	·990
39	·917	·927	·939	·948	·971	·984	·987	·989	·991
40	·919	·928	·940	·949	·972	·985	·987	·989	·991
41	0·920	0·929	0·941	0·950	0·972	0·985	0·987	0·989	0·991
42	·922	·930	·942	·951	·972	·985	·987	·989	·991
43	·923	·932	·943	·951	·973	·985	·987	·990	·991
44	·924	·933	·944	·952	·973	·985	·987	·990	·991
45	·926	·934	·945	·953	·973	·985	·988	·990	·991
46	0·927	0·935	0·945	0·953	0·974	0·985	0·988	0·990	0·991
47	·928	·936	·946	·954	·974	·985	·988	·990	·991
48	·929	·937	·947	·954	·974	·985	·988	·990	·991
49	·929	·937	·947	·955	·974	·985	·988	·990	·991
50	·930	·938	·947	·955	·974	·985	·988	·990	·991

* Based on fitted Johnson (1949) S_B approximation, see Shapiro & Wilk (1965*a*) for details.

To illustrate the process, suppose a sample of 7 observations were obtained, namely $x_1 = 6$, $x_2 = 1$, $x_3 = -4$, $x_4 = 8$, $x_5 = -2$, $x_6 = 5$, $x_7 = 0$.

(i) Ordering, one obtains

$$y_1 = -4, \quad y_2 = -2, \quad y_3 = 0, \quad y_4 = 1, \quad y_5 = 5, \quad y_6 = 6, \quad y_7 = 8.$$

(ii) $S^2 = \Sigma y_i^2 - \frac{1}{7}(\Sigma y_i)^2 = 146 - 28 = 118.$

(iii) From Table 5, under $n = 7$, one obtains

$$a_7 = 0{\cdot}6233, \quad a_6 = 0{\cdot}3031, \quad a_5 = 0{\cdot}1401, \quad a_4 = 0{\cdot}0000.$$

Thus $$b = 0{\cdot}6233(8+4) + 0{\cdot}3031(6+2) + 0{\cdot}1401(5-0) = 10{\cdot}6049.$$

(iv) $W = (10{\cdot}6049)^2/118 = 0{\cdot}9530.$

(v) Referring to Table 6, one finds the value of W to be substantially larger than the tabulated 50 % point, which is 0·928. Thus there is no evidence, from the W test, of non-normality of this sample.

4. EXAMPLES

Example 1. Snedecor (1946, p. 175), makes a test of normality for the following sample of weights in pounds of 11 men: 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, 236.

The W statistic is found to be 0·79 which is just below the 1 % point of the null distribution. This agrees with Snedecor's approximate application of the $\sqrt{b_1}$ statistic test.

Example 2. Kendall (1948, p. 194) gives an extract of 200 'random sampling numbers' from the Kendall–Babington Smith, *Tracts for Computers No.* 24. These were totalled, as number pairs, in groups of 10 to give the following sample of size 10: 303, 338, 406, 457, 461, 469, 474, 489, 515, 583.

The W statistic in this case has the value 0·9430, which is just above the 50 % point of the null distribution.

Example 3. Davies *et al.* (1956) give an example of a 2^5 experiment on effects of five factors on yields of penicillin. The 5-factor interaction is confounded between 2 blocks. Omitting the confounded effect the *ordered* effects are:

C	0·0958	ABC	0·0002
BC	·0333	CD	−0·0026
ACDE	·0293	B	−0·0036
BCE	·0246	BD	−0·0042
ACD	·0206	BCD	−0·0113
ABCE	·0194	ABE	−0·0139
DE	·0191	ABD	−0·0211
BE	·0182	AC	−0·0333
BDE	·0173	AD	−0·0341
ADE	·0132	ACE	−0·0363
BCDE	·0102	ABCD	−0·0363
ABDE	·0084	AB	−0·0402
CDE	·0077	CE	−0·0582
D	·0058	A	−0·1184
AE	·0016	E	−0·1398

In their analysis of variance, Davies *et al.* pool the 3- and 4-factor interactions for an error term. They do not find the pooled 2-factor interaction mean square to be significant but note that CE is significant at the 5 % point on a standard F-test. However, on the basis of a Bartlett test, they find that the significance of CE does *not* reach the 5 % level.

The overall statistical configuration of the 30 unconfounded effects may be evaluated against a background of a null hypothesis that these are a sample of size 30 from a normal population. Computing the W statistic for this hypothesis one finds a value of 0·8812, which is substantially below the tabulated 1 % point for the null distribution.

One may now ask whether the sample of size 25 remaining after removal of the 5 main effects terms has a normal configuration. The corresponding value of W is 0·9326, which is above the 10 % point of the null distribution.

To investigate further whether the 2-factor interactions taken alone may have a non-normal configuration due to one or more 2-factor interactions which are statistically 'too large', the W statistic may be computed for the ten 2-factor effects. This gives

$$W = 0{\cdot}9465,$$

which is well above the 50 % point, for $n = 10$.

Similarly, the 15 combined 3 and 4-factor interactions may be examined from the same point of view. The W value is 0·9088, which is just above the 10 % value of the null distribution.

Thus this analysis, combined with an inspection of the ordered contrasts, would suggest that the A, C and E main effects are real, while the remaining effects may be regarded as a random normal sample. This analysis does not indicate any reason to suspect a real CE effect based only on the statistical evidence.

The partitioning employed in this latter analysis is of course valid since the criteria employed are independent of the observations *per se*.

In the situation of this example, the sign of the contrasts is of course arbitrary and hence their distributional configuration should be evaluated on the basis of the absolute values, as in half-normal plotting (see Daniel, 1959). Thus, the above procedure had better be carried out using a half-normal version of the W test if that were available.

5. Comparison with other tests for normality

To evaluate the W procedure relative to other tests for normality an empirical sampling investigation of comparative properties was conducted, using a range of populations and sample sizes. The results of this study are given in Shapiro & Wilk (1964*a*), only a brief extract is included in the present paper.

The null distribution used for the study of the W test was determined as described above. For all other statistics, except the χ^2 goodness of fit, the null distribution employed was determined empirically from 500 samples. For the χ^2 test, standard χ^2 table values were used. The power results for all procedures and alternate distributions were derived from 200 samples.

Empirical sampling results were used to define null distribution percentage points for a combination of convenience and extensiveness in the more exhaustive study of which the results quoted here are an extract. More exact values have been published by various authors for some of these null percentage points. Clearly one employing the Kolmogorov–Smirnov procedure, for example, as a statistical method would be well advised to employ the most accurate null distribution information available. However, the present power results are intended only for indicative interest rather than as a definitive description of a procedure, and uncertainties or errors of several percent do not materially influence the comparative assessment.

Table 7 gives results on the power of a 5 % test for samples of size 20 for each of nine test procedures and for fifteen non-normal populations. The tests shown in Table 7 are: W; chi-squared goodness of fit (χ^2); standardized 3rd and 4th moments, $\sqrt{b_1}$ and b_2; Kolmogorov–Smirnov (KS) (Kolmogorov, 1933); Cramér–Von Mises (CVM) (Cramér, 1928); a weighted, by $F/(1-F)$, Cramér–Von Mises (WCVM), where F is the cumulative distribution function (Anderson & Darling, 1954); Durbin's version of the Kolmogorov–Smirnov procedure (D) (Durbin, 1961); range/standard deviation (u) (David *et al.* 1954).

Table 7. *Empirical power for* 5 % *tests for selected alternative distributions; samples all of size* 20

Population title	$\sqrt{\beta_1}$	β_2	W	χ^2	$\sqrt{b_1}$	b_2	KS	CVM	WCVM	D	u
$\chi^2(1)$	2·83	15·0	0·98	0·94	0·89	0·53	0·44	0·44	0·54	0·87	0·10
χ^2 (2)	2·00	9·0	·84	·33	·74	·34	·27	·23	·27	·42	·08
χ^2 (4)	1·41	6·0	·50	·13	·49	·27	·18	·13	·16	·15	·06
χ^2 (10)	0·89	4·2	·29	·07	·29	·19	·11	·10	·11	·07	·06
Non-cent. χ^2	0·73	3·7	·59	·10	·50	·20	·19	·16	·18	·20	·10
Log normal	6·19	113·9	·93	·95	·89	·58	·44	·48	·62	·82	·06
Cauchy	—	—	·88	·41	·77	·81	·45	·55	·98	·85	·56
Uniform	0	1·8	·23	·11	·00	·29	·13	·09	·10	·08	·38
Logistic	0	4·2	·08	·06	·12	·06	·06	·03	·05	·05	·07
Beta (2, 1)	−0·57	2·4	·35	·08	·08	·13	·08	·10	·12	·12	·23
La Place	0	6·0	·25	·17	·25	·27	·07	·07	·29	·16	·19
Poisson (1)	1·00	4·0	·99	1·00	·26	·11	·55	·22	·31	1·00	·35
Binomial, (4, 0·5)	0	2·5	·71	1·00	·02	·03	·38	·15	·17	1·00	·20
*T(5, 2·4)	0·79	2·2	·55	·14	·24	·20	·23	·20	·22	—	—
*T(10, 3·1)	0·97	2·8	·89	·32	·51	·24	·32	·30	·30	—	—

* Variates from this distribution $T(a, \lambda)$ are defined by $y = aR^\lambda - (1-R)^\lambda$, where R is uniform (0, 1) (Hastings, Mosteller, Tukey & Winsor, 1947). Also note that (*a*) the non-central χ^2 distribution has degrees of freedom 16, non-centrality parameter 1; (*b*) the beta distribution has $p = 2$, $q = 1$ in standard notation; (*c*) the Poisson distribution has expectation 1.

In using the non-scale and non-origin invariant tests the mean and variance of the hypothesized normal was taken to agree with the known mean and variance of the alternative distribution. For the Cauchy the mode and intrinsic accuracy were used.

The results of Table 7 indicate that the W test is comparatively quite sensitive to a wide range of non-normality, even with samples as small as $n = 20$. It seems to be especially sensitive to asymmetry, long-tailedness and to some degree to short-tailedness.

The χ^2 procedure shows good power against the highly skewed distributions and reasonable sensitivity to very long-tailedness.

The $\sqrt{b_1}$ test is quite sensitive to most forms of skewness. The b_2 statistic can usefully augment $\sqrt{b_1}$ in certain circumstances. The high power of $\sqrt{b_1}$ for the Cauchy alternative is probably due to the fact that, though the Cauchy is symmetric, small samples from it will often be asymmetric because of the very long-tailedness of the distribution.

The KS test has similar properties to that of the CVM procedure, with a few exceptions. In general the WCVM test has higher power than KS or CVM, especially in the case of long-tailed alternatives, such as the Cauchy, for which WCVM had the highest power of all the statistics examined.

The use of Durbin's procedure improves the KS sensitivity only in the case of highly skewed and discrete alternatives. Against the Cauchy, the D test responds, like $\sqrt{b_1}$, to the asymmetry of small samples.

The u test gives good results against the uniform alternative and this is representative of its properties for short-tailed symmetric alternatives.

The χ^2 test has the disadvantages that the number and character of class intervals used is arbitrary, that all information concerning sign and trend of discrepancies is ignored and that, for small samples, the number of cells must be very small. These factors might explain some of the lapses of power for χ^2 indicated in Table 7. Note that for almost all cases the power of W is higher than that of χ^2.

As expected, the $\sqrt{b_1}$ test is in general insensitive in the case of symmetric alternatives as illustrated by the uniform distribution. Note that for all cases, except the logistic, $\sqrt{b_1}$ power is dominated by that of the W test.

Table 8. *The effect of mis-specification of parameters*

(n = 20, 5 % test, assumed parameters are $\mu = 0$, $\sigma = 1$)

Actual parameters			Sample	Tests				
μ	σ	μ/σ	size	KS	CM	WCVM	D	χ^2
0·00	1·2	0·00	20	0·06	0·08	0·18	0·09	0·07
·00	1·3	·00	20	·12	·12	·29	·10	·09
·15	1·0	·15	20	·05	·08	·10	·03	·04
·18	1·2	·15	20	·08	·16	·24	·11	·12
·195	1·3	·15	20	·07	·12	·31	·12	·10
·30	1·0	·30	20	·14	·26	·31	·07	·11
·36	1·2	·30	20	·21	·34	·46	·16	·21
·39	1·3	·30	20	·21	·38	·55	·19	·26

The b_2 test is not sensitive to asymmetry. Its performance was inferior to that of W except in the cases of the Cauchy, uniform, logistic and Laplace for which its performance was equivalent to that of W.

Both the KS and CVM tests have quite inferior power properties. With sporadic exception in the case of very long-tailedness this is true also of the WCVM procedure. The D procedure does improve on the KS test but still ends up with power properties which are not as good as other test statistics, with the exceptions of the discrete alternatives. (In addition, the D test is laborious for hand computation.)

The u statistic shows very poor sensitivity against even highly skewed and very long-tailed distributions. For example, in the case of the $\chi^2(1)$ alternative, the u test has power of 10 % while even the KS test has a power of 44 % and that for W is 98 %. While the u test shows interesting sensitivity for uniform-like departures from normality, it would seem that the types of non-normality that it is usually important to identify are those of asymmetry and of long-tailedness and outliers.

The reader is referred to David *et al.* (1954, pp. 488–90) for a comparison of the power of the b_2, u and Geary's (1935) 'a' (mean deviation/standard deviation) tests in detecting departure from normality in symmetrical populations. Using a Monte Carlo technique, they found that Geary's statistic (which was not considered here) was possibly more effective than either b_2 or u in detecting long-tailedness.

The test statistics considered above can be put into two classes. Those which are valid for composite hypotheses and those which are valid for simple hypotheses. For the simple hypotheses procedures, such as χ^2, KS, CVM, WCVM and D, the parameters of the null distribution must be pre-specified. A study was made of the effect of small errors of specification on the test performance. Some of the results of this study are given in Table 8. The apparent power in the cases of mis-specification is comparable to that attained for these procedures against non-normal alternatives. For example, for $\mu/\sigma = 0{\cdot}3$, WCVM has apparent power of between 0·31 and 0·55 while its power against $\chi^2(2)$ is only 0·27.

6. Discussion and concluding remarks

6·1. *Evaluation of test*

As a test for the normality of complete samples, the W statistic has several good features—namely, that it may be used as a test of the composite hypothesis, that is very simple to compute once the table of linear coefficients is available and that the test is quite sensitive against a wide range of alternatives even for small samples ($n < 20$). The statistic is responsive to the nature of the overall configuration of the sample as compared with the configuration of expected values of normal order statistics.

A drawback of the W test is that for large sample sizes it may prove awkward to tabulate or approximate the necessary values of the multipliers in the numerator of the statistic. Also, it may be difficult for large sample sizes to determine percentage points of its distribution.

The W test had its inception in the framework of probability plotting. The formal use of the (one-dimensional) test statistic as a methodological tool in evaluating the normality of a sample is visualized by the authors as a supplement to normal probability plotting and not as a substitute for it.

6·2. *Extensions*

It has been remarked earlier in the paper that a modification of the present W statistic may be defined so as to be usable with incomplete samples. Work on this modified W^* statistic will be reported elsewhere (Shapiro & Wilk, 1965*b*).

The general viewpoint which underlies the construction of the W and W^* tests for normality can be applied to derive tests for other distributional assumptions, e.g. that a sample is uniform or exponential. Research on the construction of such statistics, including necessary tables of constants and percentage points of null distributions, and on their statistical value against various alternative distributions is in process (Shapiro & Wilk, 1964*b*). These statistics may be constructed so as to be scale and origin invariant and thus can be used for tests of composite hypothesis.

It may be noted that many of the results of §2·3 apply to any symmetric distribution.

The W statistic for normality is sensitive to outliers, either one-sided or two-sided. Hence it may be employed as part of an inferential procedure in the analysis of experimental data as suggested in Example 3 of §4.

The authors are indebted to Mrs M. H. Becker and Mrs H. Chen for their assistance in various phases of the computational aspects of the paper. Thanks are due to the editor and referees for various editorial and other suggestions.

References

Aitken, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edin.* **55**, 42–8.

Anderson, T. W. & Darling, D. A. (1954). A test for goodness of fit. *J. Amer. Statist. Ass.* **49**, 765–9.

Cramér, H. (1928). On the composition of elementary errors. *Skand. Aktuar.* **11**, 141–80.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two level experiments. *Technometrics*, **1**, 311–42.

David, H. A., Hartley, H. O. & Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika*, **41**, 482–3.

Davies, O. L. (ed). (1956). *Design and Analysis of Industrial Experiments*. London: Oliver and Boyd.

Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika*, **48**, 41–55.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, **27**, 310–32.

Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika*, **48**, 151–65.

Hastings, C., Mosteller, F., Tukey, J. & Winsor, C. (1947). Low moments for small samples: a comparative study of order statistics. *Ann. Math. Statist.* **18**, 413–26.

Hogg, R. V. & Craig, A. J. (1956). Sufficient statistics in elementary distribution theory. *Sankyā*, **17**, 209–16.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–76.

Kendall, M. G. (1948). *The Advanced Theory of Statistics*, **1**. London: C. Griffin and Co.

Kendall, M. G. & Stuart, A. (1961). *The Advanced Theory of Statistics*, **2**. London: C. Griffin and Co.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di une legge di distribuzione. *G. Ist. Att.-uari*, 83–91.

Lloyd, E. H. (1952). Least squares estimation of location and scale parameters using order statistics. *Biometrika*, **39**, 88–95.

Pearson, E. S. & Stephens, M. A. (1964). The ratio of range to standard deviation in the same normal sample. *Biometrika*, **51**, 484–7.

Plackett, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist.* **29**, 131–42.

Rand Corporation (1955). *A Million Random Digits with* 100,000 *Normal Deviates*.

Sarhan, A. E. & Greenberg, B. G. (1956). Estimation of location and scale parameters by order statistics from singly and double censored samples. Part I. *Ann. Math. Statist.* **27**, 427–51.

Shapiro, S. S. (1964). An analysis of variance test for normality (complete samples). Unpublished Ph.D. Thesis, Rutgers—The State University.

Shapiro, S. S. & Wilk, M. B. (1964*a*). A comparative study of various tests for normality. Unpublished manuscript. Invited paper, Amer. Stat. Assoc, Annual Meeting, Chicago, Illinois, December 1964.

Shapiro, S. S. & Wilk, M. B. (1964*b*). Tests for the exponential and uniform distributions. (Unpublished manuscript.)

Shapiro, S. S. & Wilk, M. B. (1965*a*). Testing the normality of several samples. (Unpublished manuscript).

Shapiro, S. S. & Wilk, M. B. (1965*b*). An analysis of variance test for normality (incomplete samples). (Unpublished manuscript.)

Snedecor, G. W. (1946). *Statistical Methods*, 4th edition. Ames: Iowa State College Press.