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,

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

and

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

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

$$\left(\frac{(n-1)}{na_1}, \frac{-1}{na_1}, \cdots \frac{-1}{na_1}\right)$$

$$\left(\frac{n-2}{n(a_1+a_2)}, \frac{(n-2)}{n(a_1+a_2)}, \frac{-2}{n(a_1+a_2)}, \cdots, \frac{-2}{n(a_1+a_2)}\right)$$

$$\vdots$$

$$\left(\frac{1}{n(a_1+\ldots+a_{n-1})}, \frac{1}{n(a_1+\ldots+a_{n-1})}, \cdots, \frac{-(n-1)}{n(a_1+\ldots+a_{n-1})}\right).$$

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

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

$$= \sigma^2R^2\,(R^2+1)/C^2,$$

since $\mathrm{var}\,(\hat{\sigma}) = \sigma^2/m'V^{-1}m = \sigma^2/R^2$, 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*

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

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.

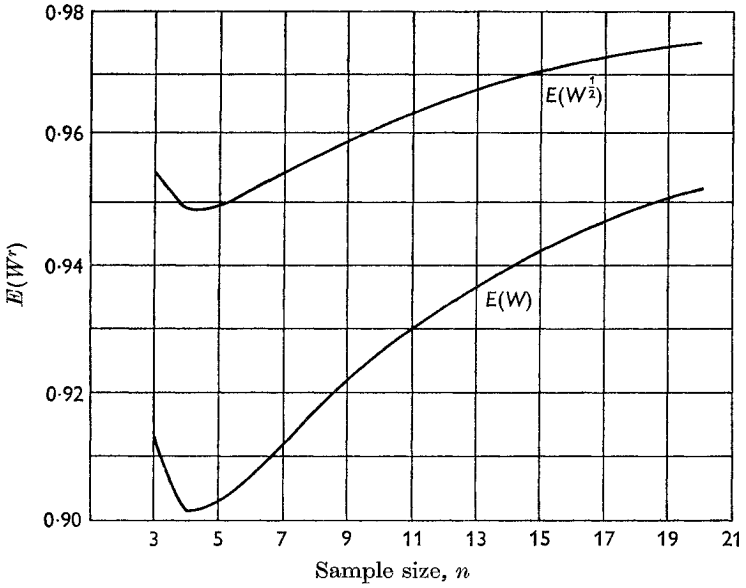

Fig. 1. Moments of W, $E(W^r)$, $n = 3(1)20$, $r = \frac{1}{2}, 1$.

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

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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.