# The Probability Plot Correlation Coefficient Test

# for Normality

**James J. Filliben**

*National Bureau of Standards*
*Statistical Engineering Laboratory*
*U.S. Department of Commerce*
*Washington, D.C. 20234*

This paper introduces the normal probability plot correlation coefficient as a test statistic in complete samples for the composite hypothesis of normality. The proposed test statistic is conceptually simple, is computationally convenient, and is readily extendible to testing non-normal distributional hypotheses. An empirical power study shows that the normal probability plot correlation coefficient compares favorably with 7 other normal test statistics. Percent points are tabulated for $n = 3(1)50(5)100$.

KEY WORDS

Probability plot
Correlation coefficient
Normal distribution
Tests of distributional hypotheses
Statistical methods
Order statistics
Medians

## 1. INTRODUCTION

Historically, the standard third moment $\sqrt{b_1}$ and the standard fourth moment $b_2$ (see Fisher (1929, 1930), Pearson (1930, 1931, 1963, 1965), and Williams (1935)) were among the first statistics to be used for testing the composite (location and scale unspecified) hypothesis of normality. An additional normal test statistic was provided by Geary (1935, 1936, 1947) who introduced the ratio $a$ of the mean deviation to the standard deviation. David, Hartley, and Pearson (1954) considered the ratio $u$ of the range to the standard deviation as a test statistic for normality; Pearson and Stephens (1964) extended the results for $u$.

Shapiro and Wilk (1965, 1968) introduced the $W$ statistic which is essentially the squared ratio of the best linear unbiased estimator for scale to the standard deviation. The $W$ statistic is computable exactly up to sample size $n = 20$, and approximations exist which are valid up to $n = 50$. $W$ was shown (see Shapiro, Wilk, and Chen (1968) for details) to be a generally superior (in terms of power) omnibus test of normality. Shapiro and Francia (1972) considered a modification $W'$ of the Shapiro–

Wilk statistic which is appropriate for larger sample sizes. The modification consists of weighting each ordered observation by its expected value for which good approximations exist (see, e.g., Harter (1961)) even for $n > 50$. $W'$ proved to be nearly as powerful as $W$ for these larger sample sizes. D'Agostino (1971, 1973) considered a test statistic $D$ which is, up to a constant, the ratio of Downton's (1966) estimator for scale to the standard deviation; the $D$ statistic essentially imposes a linear weighting scheme on the ordered observations. It was shown that the $D$ test was also comparable to $W$ in terms of power for larger sample sizes.

The purpose of this paper is to introduce a new test statistic for the composite hypothesis of normality: the normal probability plot correlation coefficient $r$. The $r$ test statistic is 1) conceptually easy to understand in that it combines two fundamentally simple concepts: the probability plot and the correlation coefficient; 2) computationally simple due to the fact that all operational coefficients are internally computable—none need be stored; and 3) readily extendible as a distributional test statistic for non-normal hypotheses.

## 2. BACKGROUND AND MOTIVATION

In much the same way as the Shapiro and Wilk (1965) study, this present study was undertaken in an attempt to summarize and formalize the distributional information contained in normal probability plots. A normal probability plot, as here used, is defined as a plot of the $i$th order statistic $X_i$ versus some measure of location $\text{loc}(X_i)$ of the $i$th order statistic from a standardized normal distribution. Heretofore, the most frequently used measure of

111

location for the $i$th order statistic has been the order statistic mean $(\text{loc}(X_i) = E(X_i) = E_i$ , say). In general, however, order statistics means have three undesirable properties: 1) The integration technique for computing $E_i$ varies drastically from distribution to distribution—no uniform technique exists for generating the $E_i$ for all distributions; 2) For some distributions (e.g., normal), order statistic means are difficult or time-consuming to compute and so must be stored or approximated; and 3) For other distributions (e.g., Cauchy), order statistic means may not always be defined. All three of these drawbacks are avoided in general by choosing to measure the location of the $i$th order statistic by its median $(\text{loc}(X_i) = \text{med}(X_i) = M_i$ , say) rather than by its mean. Doing so, we thus have as our starting point for our normal test statistic, a probability plot consisting of the ordered observations $X_i$ versus the order statistic medians $M_i$ from a normal $N(0, 1)$ distribution.

Now if the sample was, in fact, generated from the hypothesized (location and scale unspecified) normal distribution, then the plot of $X_i$ versus $M_i$ will be approximately linear. We now note that in general the product moment correlation coefficient is a simple and straightforward measure of linearity between any two variables, and so in particular is an obvious choice to measure the linearity of a normal probability plot.

The proposed test statistic, the normal probability plot correlation coefficient $r$, is thus defined as the product moment correlation coefficient between the ordered observations $X_i$ and the order statistic medians $M_i$ from a normal $N(0, 1)$ distribution:

$$r = \text{Corr}(X, M) \tag{1}$$

$$= \frac{\sum (X_i - \bar{X})(M_i - \bar{M})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (M_i - \bar{M})^2}} \tag{2}$$

In summary, the rationale behind the $r$ statistic is that underlying normality will tend to yield near-linear normal probability plots which in turn will be reflected by near-unity values of the probability plot correlation coefficient $r$.

### 3. PROPERTIES OF $r$

*Lemma 1:* Since $M_i = -M_{n-i+1}$ and $\bar{M} = 0$, then $r$ may be written as $r = \Sigma M_i X_i / (c_n S)$ where $c_n = \sqrt{\Sigma M_i^2}$ and $S = \sqrt{\Sigma (X_i - \bar{X})^2}$.

*Lemma 2:* A computationally optimal form for $r$ is $r = \Sigma_i M_i (X_i - X_{n-i+1}) / (2\Sigma_i M_i^2 \cdot \Sigma_j (X_j - \bar{X})^2)^{1/2}$ where $i = [n/2] + 1, \cdots, n$ and $j = 1, 2, \cdots, n$.

*Lemma 3:* $r$ is location and scale invariant and statistically independent of $\bar{X}$ and $S$. The invariance follows from the corresponding property of the product moment correlation coefficient and the

independence follows from a sufficiency theorem (Kendall and Stuart (1961), Vol. 2, pages 219–220).

*Corollary:* The distribution of $r$ depends only on the sample size $n$ and the standardized normal cumulative distribution function $\Phi(x)$. $r$ does not depend on the values of the (assumedly) unknown location and scale parameters $\mu$ and $\sigma$ and hence is appropriate for testing the composite hypothesis of normality.

*Lemma 4:* $0 < (M_n/c_n) \sqrt{(n/(n-1))} \le r \le 1$.

The non-negativity follows from Lemma 2 by noting that all of the components are non-negative; the minimum value of $r$ follows from an argument analogous to that of Shapiro and Wilk (1965, p. 594).

*Lemma 5:* The mean and variance of $r$ are given by

$$E(r) = (\Sigma M_i E_i \Gamma((n-1)/2) / (c_n \Gamma(n/2) \sqrt{2})$$

and

$$\text{Var}(r)$$

$$= \{\Sigma\Sigma M_i M_j (\sigma_{ij} + E_i E_j) / ((n-1)c_n^2)\} - E(r)^2$$

where $E_i$ denotes the $i$th normal $N(0, 1)$ order statistic mean and $\sigma_{ij}$ denotes the covariance of the $i$th and $j$th order statistics from a normal $N(0, 1)$ distribution. These follow from the fact that $E(r^k) = E((\Sigma M_i X_i)^k) / (c_n^k E(S^k))$.

The above lemmas and corollary are of course in reference to the *normal* probability plot correlation coefficient. If one considers a *generalized* probability plot correlation coefficient whereby one tests the hypothesis $H_0 : D = D_0$ , then after all references to normal are replaced by references to $D_0$ , one finds that all 5 lemmas and corollary still continue to hold with but one exception: $r$ is no longer independent of $X$ and $S$.

### 4. PERCENT POINTS OF $r$

Empirical sampling was employed to obtain an approximation to the distribution of $r$ under normality. For each sample size $n$ $(n = 3(1)50(5)100)$ normal random samples were generated on the Univac 1108 at the National Bureau of Standards using the Rand normal deviates (1955) as input. $N = 10^5$ samples were generated for $n \le 10$, and $N = [10^5/n]$ samples were generated for the specified sample sizes $n > 10$. Table 1 gives the smoothed percent points resulting from the empirical sampling for $p = .005, .01, .025, .05, .10, .25, .50, .75, .90, .95, .975, .99,$ and $.995$. Also included in Table 1 as an indication of the length of the lower tail of $r$ is the analytical minimum of $r$ (the 0 percent point) as given in Lemma 4.

Checks on the accuracy of the empirical sampling were provided by comparing the empirical versus

theoretical mean and standard deviation of $r$ for each sample size $n$, and also by comparing empirical versus theoretical percent points of $r$ for $n = 3$.

## 5. POWER OF $r$

To assess the power of $r$ relative to the other 7 test statistics for normality included in this paper

($\sqrt{b_1}$, $b_2$, $a$, $u$, $W$, $W'$, and $D$) an empirical sampling study was conducted. For sample size $n = 20$, 51 non-normal alternative distributions (see Table 2) were considered; for sample size $n = 50$, a representative subset of 25 non-normal alternatives were considered. For both sample sizes and for each of the alternative distributions, $N = 200$ samples were

TABLE 1—*Percent points of the normal probability plot correlation coefficient* $r$

| n | .000 | .005 | .01 | .025 | .05 | .10 | .25 | .50 | .75 | .90 | .95 | .975 | .99 | .995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | .866 | .867 | .869 | .872 | .879 | .891 | .924 | .966 | .991 | .999 | 1.000 | 1.000 | 1.000 | 1.000 |
| 4 | .784 | .813 | .822 | .845 | .868 | .894 | .931 | .958 | .979 | .992 | .996 | .998 | .999 | 1.000 |
| 5 | .726 | .803 | .822 | .855 | .879 | .902 | .935 | .960 | .977 | .988 | .992 | .995 | .997 | .998 |
| 6 | .683 | .818 | .835 | .868 | .890 | .911 | .940 | .962 | .977 | .986 | .990 | .993 | .996 | .997 |
| 7 | .648 | .828 | .847 | .876 | .899 | .916 | .944 | .965 | .978 | .986 | .990 | .992 | .995 | .996 |
| 8 | .619 | .841 | .859 | .886 | .905 | .924 | .948 | .967 | .979 | .986 | .990 | .992 | .995 | .996 |
| 9 | .595 | .851 | .868 | .893 | .912 | .929 | .951 | .968 | .980 | .987 | .990 | .992 | .994 | .995 |
| 10 | .574 | .860 | .876 | .900 | .917 | .934 | .954 | .970 | .981 | .987 | .990 | .992 | .994 | .995 |
| 11 | .556 | .868 | .883 | .906 | .922 | .938 | .957 | .972 | .982 | .988 | .990 | .992 | .994 | .995 |
| 12 | .539 | .875 | .889 | .912 | .926 | .941 | .959 | .973 | .982 | .988 | .990 | .992 | .994 | .995 |
| 13 | .525 | .882 | .895 | .917 | .931 | .944 | .962 | .975 | .983 | .988 | .991 | .993 | .994 | .995 |
| 14 | .512 | .888 | .901 | .921 | .934 | .947 | .964 | .976 | .984 | .989 | .991 | .993 | .994 | .995 |
| 15 | .500 | .894 | .907 | .925 | .937 | .950 | .965 | .977 | .984 | .989 | .991 | .993 | .994 | .995 |
| 16 | .489 | .899 | .912 | .928 | .940 | .952 | .967 | .978 | .985 | .989 | .991 | .993 | .994 | .995 |
| 17 | .478 | .903 | .916 | .931 | .942 | .954 | .968 | .979 | .986 | .990 | .992 | .993 | .994 | .995 |
| 18 | .469 | .907 | .919 | .934 | .945 | .956 | .969 | .979 | .986 | .990 | .992 | .993 | .995 | .995 |
| 19 | .460 | .909 | .923 | .937 | .947 | .958 | .971 | .980 | .987 | .990 | .992 | .993 | .995 | .995 |
| 20 | .452 | .912 | .925 | .939 | .950 | .960 | .972 | .981 | .987 | .991 | .992 | .994 | .995 | .995 |
| 21 | .445 | .914 | .928 | .942 | .952 | .961 | .973 | .981 | .987 | .991 | .993 | .994 | .995 | .996 |
| 22 | .437 | .918 | .930 | .944 | .954 | .962 | .974 | .982 | .988 | .991 | .993 | .994 | .995 | .996 |
| 23 | .431 | .922 | .933 | .947 | .955 | .964 | .975 | .983 | .988 | .991 | .993 | .994 | .995 | .996 |
| 24 | .424 | .926 | .936 | .949 | .957 | .965 | .975 | .983 | .988 | .992 | .993 | .994 | .995 | .996 |
| 25 | .418 | .928 | .937 | .950 | .958 | .966 | .976 | .984 | .989 | .992 | .993 | .994 | .995 | .996 |
| 26 | .412 | .930 | .939 | .952 | .959 | .967 | .977 | .984 | .989 | .992 | .993 | .994 | .995 | .996 |
| 27 | .407 | .932 | .941 | .953 | .960 | .968 | .977 | .984 | .989 | .992 | .994 | .995 | .995 | .996 |
| 28 | .402 | .934 | .943 | .955 | .962 | .969 | .978 | .985 | .990 | .992 | .994 | .995 | .995 | .996 |
| 29 | .397 | .937 | .945 | .956 | .962 | .969 | .979 | .985 | .990 | .992 | .994 | .995 | .995 | .996 |
| 30 | .392 | .938 | .947 | .957 | .964 | .970 | .979 | .986 | .990 | .993 | .994 | .995 | .996 | .996 |
| 31 | .388 | .939 | .948 | .958 | .965 | .971 | .980 | .986 | .990 | .993 | .994 | .995 | .996 | .996 |
| 32 | .383 | .939 | .949 | .959 | .966 | .972 | .980 | .986 | .990 | .993 | .994 | .995 | .996 | .996 |
| 33 | .379 | .940 | .950 | .960 | .967 | .973 | .981 | .987 | .991 | .993 | .994 | .995 | .996 | .996 |
| 34 | .375 | .941 | .951 | .960 | .967 | .973 | .981 | .987 | .991 | .993 | .994 | .995 | .996 | .996 |
| 35 | .371 | .943 | .952 | .961 | .968 | .974 | .982 | .987 | .991 | .993 | .995 | .995 | .996 | .997 |
| 36 | .367 | .945 | .953 | .962 | .968 | .974 | .982 | .987 | .991 | .994 | .995 | .996 | .996 | .997 |
| 37 | .364 | .947 | .955 | .962 | .969 | .975 | .982 | .988 | .991 | .994 | .995 | .996 | .996 | .997 |
| 38 | .360 | .948 | .956 | .964 | .970 | .975 | .983 | .988 | .992 | .994 | .995 | .996 | .996 | .997 |
| 39 | .357 | .949 | .957 | .965 | .971 | .976 | .983 | .988 | .992 | .994 | .995 | .996 | .996 | .997 |
| 40 | .354 | .949 | .958 | .966 | .972 | .977 | .983 | .988 | .992 | .994 | .995 | .996 | .996 | .997 |
| 41 | .351 | .950 | .958 | .967 | .972 | .977 | .984 | .989 | .992 | .994 | .995 | .996 | .996 | .997 |
| 42 | .348 | .951 | .959 | .967 | .973 | .978 | .984 | .989 | .992 | .994 | .995 | .996 | .997 | .997 |
| 43 | .345 | .953 | .959 | .967 | .973 | .978 | .984 | .989 | .992 | .994 | .995 | .996 | .997 | .997 |
| 44 | .342 | .954 | .960 | .968 | .973 | .978 | .984 | .989 | .992 | .994 | .995 | .996 | .997 | .997 |
| 45 | .339 | .955 | .961 | .969 | .974 | .978 | .985 | .989 | .993 | .994 | .995 | .996 | .997 | .997 |
| 46 | .336 | .956 | .962 | .969 | .974 | .979 | .985 | .990 | .993 | .995 | .995 | .996 | .997 | .997 |
| 47 | .334 | .956 | .963 | .970 | .974 | .979 | .985 | .990 | .993 | .995 | .995 | .996 | .997 | .997 |
| 48 | .331 | .957 | .963 | .970 | .975 | .980 | .985 | .990 | .993 | .995 | .996 | .996 | .997 | .997 |
| 49 | .329 | .957 | .964 | .971 | .975 | .980 | .986 | .990 | .993 | .995 | .996 | .996 | .997 | .997 |
| 50 | .326 | .959 | .965 | .972 | .977 | .981 | .986 | .990 | .993 | .995 | .996 | .996 | .997 | .997 |
| 55 | .315 | .962 | .967 | .974 | .978 | .982 | .987 | .991 | .994 | .995 | .996 | .997 | .997 | .997 |
| 60 | .305 | .965 | .970 | .976 | .980 | .983 | .988 | .991 | .994 | .995 | .996 | .997 | .997 | .998 |
| 65 | .296 | .967 | .972 | .977 | .981 | .984 | .989 | .992 | .994 | .996 | .996 | .997 | .997 | .998 |
| 70 | .288 | .969 | .974 | .978 | .982 | .985 | .989 | .993 | .995 | .996 | .997 | .997 | .998 | .998 |
| 75 | .281 | .971 | .975 | .979 | .983 | .986 | .990 | .993 | .995 | .996 | .997 | .997 | .998 | .998 |
| 80 | .274 | .973 | .976 | .980 | .984 | .987 | .991 | .993 | .995 | .996 | .997 | .997 | .998 | .998 |
| 85 | .268 | .974 | .977 | .981 | .985 | .987 | .991 | .994 | .995 | .997 | .997 | .997 | .998 | .998 |
| 90 | .263 | .976 | .978 | .982 | .985 | .988 | .991 | .994 | .996 | .997 | .997 | .998 | .998 | .998 |
| 95 | .257 | .977 | .979 | .983 | .986 | .989 | .992 | .994 | .996 | .997 | .997 | .998 | .998 | .998 |
| 100 | .252 | .979 | .981 | .984 | .987 | .989 | .992 | .994 | .996 | .997 | .998 | .998 | .998 | .998 |

generated. The Rand Tables (1955) were again used as a source of the required normal and uniform random numbers. Table 2 summarizes the results of the power study.

For symmetric shorter-tailed alternatives, it is seen that $r$ has below-average power; for symmetric longer-tailed alternatives, $r$ has above-average power; for skewed alternatives, $r$ again has above-average power. Relative to the generally acknowledged omnibus normal test statistic $W$, $r$ is seen to be poorer than $W$ for symmetric shorter-tailed alternatives, marginally better than $W$ for symmetric longer-tailed alternatives, and marginally poorer than $W$ for skewed alternatives. Asymptotically, of course, $r$ and $W$ have equivalent power.

We note in passing that the similarity between the power curves of $r$ and the Shapiro–Francia $W'$ is, no doubt, due to the fact that $W'$, upon close scrutiny, may be construed to be a squared normal probability plot correlation coefficient with the order statistic means (rather than $r$'s order statistic medians) used as the order statistic's "typical value". We are thus justified in concluding that nothing has been lost in this the normal case in choosing to use order statistic medians (rather than means) in defining our probability plot correlation coefficient $r$; and yet of course, much will be gained in the extendibility of the normal probability plot correlation coefficient $r$ to non-normal cases due to the existence and computational difficulties inherent in

TABLE 2—*Empirical 5% level power (in %) of normal test statistics for* $n = 20, 50,$ *and* $100$

Symmetric alternatives shorter-tailed than normal (ordered by $\beta_2$)

| Distribution | Skewness Measures $\sqrt{\beta_1}$ | $\rho$ | Tail Length Measures $\beta_2$ | $\tau_2$ | Power for n = 20 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 50 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 100 r |
|---|---|---|---|---|---|---|---|
| Arcsine | 0 | .5 | 1.500 | .035 | 1 69 56 90 70 42 2 41 | 1 99 96 * * 98 13 98 | * |
| Johnson bounded: JSB(0,.5) | 0 | .5 | 1.627 | .136 | 1 59 45 74 47 23 8 19 | 0 98 86 * 98 84 40 80 | * |
| Tukey λ(1.5) | 0 | .5 | 1.753 | .150 | 0 35 22 43 20 7 11 6 | 0 96 67 98 93 54 58 45 | 99 |
| Tukey λ(1.25) | 0 | .5 | 1.761 | .150 | 0 35 22 42 20 7 12 5 | | |
| Uniform | 0 | .5 | 1.800 | .167 | 0 30 20 37 15 4 13 4 | 0 91 59 96 86 41 60 39 | 95 |
| Johnson bounded: JSB(0,.7071) | 0 | .5 | 1.87 | .238 | 1 31 22 32 17 6 13 5 | | |
| Tukey λ(.75) | 0 | .5 | 1.890 | .214 | 0 20 15 26 10 3 11 2 | 0 74 49 87 71 25 54 21 | 72 |
| Tukey λ(.5) | 0 | .5 | 2.082 | .302 | 1 13 9 11 4 1 7 1 | | |
| Anglit | 0 | .5 | 2.194 | .331 | 1 10 6 8 2 1 2 1 | 0 34 17 33 22 4 28 4 | 9 |
| Triangular | 0 | .5 | 2.400 | .352 | 1 4 4 6 1 1 2 1 | 0 13 6 14 8 1 14 1 | 1 |
| Tukey λ(.25) | 0 | .5 | 2.539 | .429 | 2 4 5 5 1 1 3 1 | | |
| Johnson bounded: JSB(0,2) | 0 | .5 | 2.631 | .448 | 3 5 3 5 4 4 7 4 | | |

Symmetric alternatives longer-tailed than normal (ordered by $\tau_2$)

| Distribution | Skewness Measures $\sqrt{\beta_1}$ | $\rho$ | Tail Length Measures $\beta_2$ | $\tau_2$ | Power for n = 20 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 50 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 100 r |
|---|---|---|---|---|---|---|---|
| Johnson unbounded: JSU(0,10) | 0 | .5 | 3.041 | .504 | 6 6 4 6 6 7 8 7 | | |
| Logistic | 0 | .5 | 4.200 | .579 | 13 7 9 7 7 12 9 12 | 19 21 22 16 11 23 20 22 | 29 |
| Johnson unbounded: JSU(0,2) | 0 | .5 | 4.508 | .590 | 15 13 11 9 11 14 13 15 | | |
| LaPlace | 0 | .5 | 6.000 | .617 | 24 23 27 20 26 33 28 33 | 38 49 66 37 43 59 62 60 | 80 |
| Tukey λ9-.25) | 0 | .5 | ----- | .719 | 35 36 38 27 35 42 40 42 | | |
| Johnson unbounded: JSU(0,1) | 0 | .5 | 36.188 | .728 | 35 37 39 30 38 44 43 45 | 56 71 79 61 67 78 78 78 | 95 |
| Student t_2 | 0 | .5 | ----- | .810 | 52 53 57 43 54 59 55 60 | | |
| Tukey λ(-.5) | 0 | .5 | ----- | .827 | 60 63 65 49 66 71 68 71 | 75 91 95 86 88 93 94 93 | * |
| Johnson unbounded: JSU(0,.5) | 0 | .5 | 4x10^6 | .889 | 71 85 86 65 85 89 88 89 | | |
| Tukey λ(-.75) | 0 | .5 | ----- | .899 | 74 78 82 61 83 86 84 87 | | |
| Johnson unbounded: JSU(0,.4) | 0 | .5 | 4x10^10 | .929 | 80 90 94 71 93 94 93 94 | | |
| Cauchy | 0 | .5 | ----- | .941 | 79 84 89 63 89 91 90 92 | 92 99 * 98 99 99 * 99 | * |
| Tukey λ(-1) | 0 | .5 | ----- | .942 | 81 86 91 64 91 93 93 93 | | |
| Tukey λ(-1.25) | 0 | .5 | ----- | .967 | 85 91 95 66 95 96 96 96 | | |
| Tukey λ(-1.5) | 0 | .5 | ----- | .982 | 88 95 96 65 97 98 98 98 | 96 * * * * * * * | * |

Skewed alternatives (ordered by power of W)

| Distribution | Skewness Measures $\sqrt{\beta_1}$ | $\rho$ | Tail Length Measures $\beta_2$ | $\tau_2$ | Power for n = 20 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 50 $\sqrt{b_1}$ $b_2$ a u W W' D r | Power for n = 100 r |
|---|---|---|---|---|---|---|---|
| Weibull(4) | -.087 | .520 | 2.748 | .463 | 2 4 3 4 3 2 3 2 | | |
| Weibull(3) | .168 | .527 | 2.729 | .454 | 2 3 4 4 3 2 3 2 | 4 7 4 8 8 4 6 3 | 1 |
| Weibull(2) | .631 | .618 | 3.245 | .464 | 13 5 2 5 13 12 8 11 | | |
| Weibull(10) | -.638 | .392 | 3.570 | .507 | 18 11 5 5 14 15 9 16 | 42 18 9 7 35 34 18 32 | 50 |
| Extreme Value Type 1 | 1.140 | .664 | 5.400 | .554 | 29 16 9 3 32 31 17 31 | 68 28 18 8 66 63 35 63 | 93 |
| Skewed λ(1.5,.5) | .497 | .655 | 2.209 | .268 | 12 15 13 24 33 18 6 18 | 23 36 34 72 94 70 7 69 | 99 |
| Half-normal | .995 | .709 | 3.869 | .467 | 25 13 10 13 40 32 16 32 | 70 27 14 21 94 85 31 82 | 99 |
| Extreme Value Type 2(10) | 1.910 | .718 | 10.979 | .609 | 49 28 15 3 49 48 30 48 | | |
| Power Lognormal: PLN(.5) | 1.750 | .727 | 8.898 | .590 | 50 24 15 7 56 54 31 54 | 89 51 34 17 94 92 66 91 | 99 |
| Extreme Value Type 2(5) | 3.535 | .768 | 48.092 | .667 | 64 39 24 4 67 65 46 65 | 96 68 49 21 96 96 82 96 | * |
| Exponential | 2.000 | .818 | 9.000 | .579 | 71 32 17 7 86 82 50 82 | 99 60 36 16 * 99 89 99 | * |
| Pareto(100) | 2.062 | .820 | 9.504 | .585 | 73 34 17 6 88 83 51 83 | | |
| Pareto(10) | 2.811 | .844 | 17.829 | .638 | 79 47 26 5 90 88 63 88 | * 76 49 17 * * 93 * | * |
| Extreme Value Type 2(2) | ----- | .882 | ----- | .817 | 89 66 52 5 92 91 82 91 | | |
| Pareto(5) | 4.648 | .868 | 73.800 | .693 | 85 55 32 4 94 93 72 93 | | |
| Lognormal | 6.185 | .877 | 113.936 | .728 | 86 56 40 5 96 94 71 94 | 99 90 73 33 * * 98 * | * |
| Chi-Square(1) | 2.828 | .910 | 15.000 | .631 | 88 51 27 11 97 94 81 94 | * 83 61 21 * * 99 * | * |
| Pareto(2) | ----- | .924 | ----- | .827 | 96 70 56 3 98 98 90 98 | | |
| Extreme Value Type 2(1) | ----- | .970 | ----- | .942 | 99 89 75 1 99 99 98 99 | * * 98 67 * * * * | * |
| Weibull(.5) | 6.619 | .965 | 87.720 | .759 | 99 79 61 5 * * 97 * | | |
| Pareto(1) | ----- | .975 | ----- | .942 | 99 89 76 2 * * 98 * | | |
| Power Lognormal: PLN(3) | 7x10^15 | .997 | 4x10^15 | .954 | * 97 89 2 * * * * | | |
| Extreme Value Type 2(.5) | ----- | .9987 | ----- | .994 | * 98 96 1 * * * * | | |
| Weibull(.2) | 190.113 | .9998 | 2x10^5 | .948 | * 98 96 2 * * * * | | |

The skewness measure $\rho = (G(.975) - G(.5))/(G(.975) - G(.025))$ where $G(p)$ is the percent point function of the distribution.
The tail length measure $\tau_2 = (1 - 1/\tau_1)^{.57854}$ where $\tau_1 = (G(.9975) - G(.0025))/(G(.975) - G(.025))$.
The Tukey $\lambda$ distribution is defined by the percent point function $G(p) = (p^\lambda - (1 - p)^\lambda)/\lambda$.
The Skewed $\lambda$ distribution (with parameters $\lambda_1$ and $\lambda_2$) has percent point function $G(p) = (p^{\lambda_1}/\lambda_1) - ((1 - p)^{\lambda_2}/\lambda_2)$.
The Power Lognormal distribution (with parameter $\gamma$) has ppf $G(p) = \exp(\gamma\Phi^{-1}(p))$ where $\Phi^{-1}(p)$ is the normal $N(0, 1)$ ppf.
– – – – – undefined.
* 100.

order statistic means and absent in order statistic medians as outlined previously in section 2. More detailed discussion regarding $r$, $W'$, and $W$ will be found in section 6.

Although a complete power study of all 8 normal test statistics was not conducted for sample size $n = 100$, the empirical power curve for $r$ itself was computed (see the last column of Table 2). For this larger sample size it is seen that, on an absolute scale, $r$ has good power for even the symmetric shorter-tailed alternatives (triangular and anglit distributions excepted), and has excellent power for the symmetric longer-tailed and the skewed alternative distributions.

## 6. COMPARISON OF $r$, $W$, AND $W'$

The purpose of this section is to point out the similarities and the differences between the normal probability plot correlation coefficient $r$, the Shapiro–Wilk statistic $W$, and the Shapiro–Francia $W'$ modification of $W$. First of all, in regard to origin, all three were developed in an attempt to summarize and quantify the information existent in a normal probability plot. Secondly, if one ignores the squaring involved in the definition of $W$ and $W'$, then it is seen that there is a mathematical similarity between the three statistics—all reducing to a ratio of two statistics—each of which is sensitive to scale. The numerator of the ratio is a linear combination of the ordered observations; the denominator of the ratio is essentially the sample standard deviation. A third point of similarity is in the power curves for the test for normality—this similarity being especially pronounced for the larger sample sizes. Asymptotically, this similarity becomes an equivalence.

In regard to differences, the most obvious is one of formulation. As stated explicitly in Shapiro and Wilk (1965) and Shapiro and Francia (1972) respectively, both $W$ and $W'$ rely on the equivalence (under the null hypothesis of normality) of two measures of variation—the squared slope of the normal probability plot regression line and the residual mean square of the regression line—as a basis for their test. In contrast, $r$ concentrates on the simplest and most obvious characteristic of a normal probability plot—its linearity—as its basis. Since the linearity of a probability plot is that particular aspect which data analysts naturally look for in examining probability plots, its quantification per se seems eminently reasonable. Further, in measuring this linearity, $r$ incorporates yet another element of simplicity—the product moment correlation coefficient. It is the author's contention that this directness and simplicity in the definition of $r$ is a distinct asset over $W$ and $W'$.

The second (and most notable) dissimilarity

between $r$, $W$, and $W'$ is in the coefficients (order statistic weightings) utilized in the numerators of the three statistics. The coefficients of $W$ are essentially the best linear unbiased estimator (BLUE) of the scale parameter for a normal distribution; the coefficients of $W'$ are the normal order statistic means; the coefficients of $r$ are the normal order statistic medians. The numerical distinction between the three weighting systems is insignificant for the normal case; however, the practical distinction between the three systems is critically important in terms of defining a logically extendible non-normal distributional test statistic. A logical extension of $W$ would involve the computation of the BLUE for scale for the non-normal distribution $D$ under consideration; such a BLUE computation is, however, difficult and tedious for most distributions $D$ and most sample sizes $n$. For $W'$, a logical extension necessitates the computation of order statistic means for the distribution $D$ under test; for most distributions $D$, however, this is not easy for the reasons enumerated in section 2. And finally for $r$, its logical extension requires the computation of order statistic medians for the distribution $D$ at hand; but as discussed also in section 2, this is in fact easily done—requiring only a transformation of uniform order statistic medians via the distribution $D$'s percent point function. Thus of the three normal test statistics, a distinguishing feature is that only $r$ is readily extendible per se to testing non-normal distributional hypotheses.

## 7. CALCULATION OF NORMAL ORDER STATISTIC MEDIANS $M_i$

The coefficients $M_i$ required in the calculation of the $r$ statistic are the order statistic medians from a normal $N(0, 1)$ distribution. These normal order statistic medians $M_i$ are exactly related to the order statistic medians $m_i$ from a uniform distribution on $[0, 1]$ by $M_i = \Phi^{-1}(m_i)$ where $\Phi^{-1}(p)$ is the normal percent point function (see Filliben (1969)). High-accuracy algorithms for $\Phi^{-1}(p)$ do exist and are easily implementable on the computer; one such algorithm is that of Abramowitz and Stegun (1964, page 933, formula 26.2.23) which has an accuracy of $(4.5)10^{-4}$.

With respect to the calculation of the uniform order statistic medians $m_i$ on the computer, the following points are relevant: 1) $m_i = 1 - m_{n-i+1}$ for all $i$ and $n$; 2) $m_n = .5^{(1/n)}$ for all $n$; 3) Although closed form expressions for $m_i$ ($i \neq n$ and $i \neq 1$) do not exist, readily programmable algorithms do exist which can, e.g., provide accuracy to 7 (or more) decimal places; and 4) 7 decimal place accuracy is rarely needed nor justified in typical day-to-day data analysis situations; the following algorithm for

$m_i$ is very fast and yet sufficiently accurate (maximum error $\leq .0003$ for all $i$ and $n$) so as to be recommended in practice:

$$m_i = \begin{cases} 1 - m_n & i = 1 \\ (i - .3175)/(n + .365) & i = 2, 3, \cdots, n - 1 \\ .5^{(1/n)} & i = n \end{cases} \quad (3)$$

The author arrived at the above approximational form by taking Blom's (1958) general approximation form for order statistic means, surmised that the same general form would probably be valid for order statistic medians, and applied it specifically to the uniform distribution thereby obtaining the approximation $m_i = (i - \gamma)/(n - 2\gamma + 1)$. This resulting form was further supported by the fact that for the uniform distribution, the $i$th order statistic mean is given exactly by $i/(n + 1)$, the $i$th order statistic mode is given exactly by $(i-1)/(n-1)$, and so the intermediate $i$th order statistic median might reasonably be approximated by $(i - \gamma)/(n - 2\gamma + 1)$ for some appropriate $\gamma$. Experimentation with various values of $\gamma$ led to the choice of $\gamma = .3175$ as used above in (3).

## 8. EXAMPLE

To illustrate the procedure for performing the $r$ test for normality, consider the $n = 7$ example of Shapiro and Wilk (1965, page 606): $y_1 = 6$, $y_2 = 1$, $y_3 = -4$, $y_4 = 8$, $y_5 = -2$, $y_6 = 5$, $y_7 = 0$.

1) The ordered sample is $X_1 = -4$, $X_2 = -2$, $X_3 = 0$, $X_4 = 1$, $X_5 = 5$, $X_6 = 6$, $X_7 = 8$.

2) The uniform [0, 1] order statistic medians $m_i$ (as computed from (3)) are $m_1 = .0943$, $m_2 = .2284$, $m_3 = .3642$, $m_4 = .5000$, $m_5 = .6358$, $m_6 = .7716$, $m_7 = .9057$.

3) The normal $N(0, 1)$ order statistic medians $M_i$ (as computed from $M_i = \Phi^{-1}(m_i)$ where the normal percent point function ($p$) is given, for example, by Abramowitz and Stegun, 1964, p. 933, formula 26.2.23) are $M_1 = -1.31493$, $M_2 = -.74388$, $M_3 = -.34681$, $M_4 = 0$, $M_5 = .34681$, $M_6 = .74388$, and $M_7 = 1.31493$.

4) From lemma 1 (or the computationally shorter lemma 2): $\Sigma M_i X_i = 23.46425$, $\Sigma M_i^2 = 4.80535$, $\Sigma(X_i - \bar{X})^2 = 118$, and therefore $r = 23.46425/\sqrt{(4.80535)(118)} = .98538$.

5) From Table 1 with $n = 7$, it is seen that .98538 is well above the 5% critical value; in fact, the observed $r$ falls between the 75% and 90% points of the null distribution. On the basis of the $r$ test, there is no evidence to contradict the hypothesis of normality.

We note in passing that if a 7 decimal place algorithm (rather than algorithm (3)) were used to compute the uniform [0, 1] order statistic medians $m_i$ in step 2 above, then the resulting $r$ value would be identical to the above computed $r$ value for all 5 decimal places; this is an empirical verification of the adequacy of algorithm (3).

## 9. EXTENSIONS

Although there exist other obvious extensions of the normal probability plot correlation coefficient $r$ (for example, to incomplete samples), by far the most important extension is that $r$ may be readily employed *per se* as a test statistic for non-normal distributional hypotheses. This point was initially alluded to and discussed in several previous sections. Continuing further along these lines and probing deeper into what the generalized probability plot correlation coefficient might possibly do in a non-normal distributional context, one may conceive of a "Maximum Probability Plot Correlation Coefficient" *criterion* whereby one selects the "best" distribution out of a finite family of admissible distributions on the basis of the maximum value of $r$ where $r$ has been computed for each member of the family. This "best" distribution under the above criterion would of course correspond to that distribution in the family which would have yielded the most linear probability plot. In working with distributional families defined by a continuous tail length or shape parameter (for example, the Tukey $\lambda$ family, the extreme value type II family, and the Weibull family), the above selection criterion becomes in effect an *estimation criterion* for the tail length/shape parameter of that family. In a practical note, the "Maximum Probability Plot Correlation Coefficient" criterion has been routinely used by the author for several years and has been found to be a valuable tool for distributional model selection. The application of this criterion to several "real-world" data sets is presented and demonstrated in Filliben (1972).

## 10. CONCLUDING REMARKS

As a test for normality in complete samples, the normal probability plot correlation coefficient has the following features:

1) *Composite Hypotheses:* $r$ can be used as a test of the composite hypothesis of normality, that is, with the values of the location and scale parameters unspecified.

2) *Conceptual Simplicity:* Two fundamentally simple concepts have been combined to form $r$—the probability plot and the correlation coefficient. Good distributional fits result in near-linear probability plots, and a simple measure of linearity is the product moment correlation coefficient. It is to be noted that the use of the probability plot correlation coefficient

should complement (rather than replace) the use of probability plots themselves.

3) *Computer Implementability:* The calculation of the $r$ statistic is completely computer implementable—no coefficients need be stored.

4) *Unlimited Sample Size:* The mechanics for computing the $r$ statistic is not limited to any sample size. Percent points for $r$ and $n > 100$ will be given in another paper.

5) *Power:* $r$ compares favorably with the 7 other normal test statistics for $n = 20$ and $n = 50$ for longer-tailed and skewed alternatives. For $n = 100$, $r$ has good absolute power for longer-tailed, skewed, and most shorter-tailed alternatives.

6) *Extendibility:* An important and distinguishing property of the $r$ test is that it is readily extendible to non-normal distributional hypotheses. Further, the generalized $r$ statistic is applicable *per se* (in conjunction with the "Maximum Probability Plot Correlation Coefficient" criterion) to the problem of selecting the "best-fit" distribution from a set or family of admissible distributions.

Internally documented ANSI Fortran subroutines for computing the normal probability plot correlation coefficient and the generalized probability plot correlation coefficient (for all of the distributions and distributional families considered in Table 2) are available from the author. Fortran subroutines which apply the maximum probability plot correlation coefficient criterion to a 44-member symmetric distributional family, to the extreme value family, and to the Weibull family are also available from the author. Finally, the author would like to express his thanks to the referees for several extremely helpful suggestions in regard to the presentation of this paper.

## REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. (eds.) (1964). *National Bureau of Standards Applied Mathematics Series, 55: Handbook of Mathematical Functions*, Washington, D. C: U. S. Government Printing Office.

[2] BLOM, G. (1958). *Statistical Estimates and Transformed Beta-Variables.* New York: Wiley.

[3] D'AGOSTINO, R. B. (1971). An omnibus test of normality for moderate and large size samples, *Biometrika 58*, 341–348.

[4] D'AGOSTINO, R. B. (1973). Monte Carlo power comparison of the $W'$ and $D$ tests of normality *Comm. Statist. 1*, 545–551.

[5] DAVID, H. A., HARTLEY, H. O., and PEARSON, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation, *Biometrika 41*, 482–493.

[6] DOWNTON, F. (1966). Linear estimates with polynomial coefficients. *Biometrika 53*, 129–141.

[7] FILLIBEN, J. J. (1969). Simple and robust linear estimation of the location parameter of a symmetric distribution, Unpublished Ph.D. Dissertation, Princeton University.

[8] FILLIBEN, J. J. (1972). Techniques for tail length analysis, *Proceedings of the Eighteenth Conference on the Design of Experiments in Army Research and Testing*, Durham: U. S. Army Research Office.

[9] FISHER, R. A. (1929). Moments and product-moments of sampling distributions, *Proc. Lond. Math. Soc., (2), 30*, 199.

[10] FISHER, R. A. (1930). The moments of the distribution for normal samples of measures of departure from normality, *Proc. Roy. Soc. A, 130*, 16.

[11] GEARY, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality, *Biometrika, 27*, 310–332.

[12] GEARY, R. C. (1936). Moments of the ratio of the mean deviation to the standard deviation for normal samples, *Biometrika 28*, 295–307.

[13] GEARY, R. C. (1947). Testing for normality, *Biometrika 34*, 209–242.

[13a] HARTER, H. L. (1961). Expected values of normal order statistics, *Biometrika 48*, 151–165.

[14] JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika 36*, 149–176.

[15] KENDALL, M. G. and STUART, A. (1961), *The Advanced Theory of Statistics, 2.* London: Griffin.

[16] PEARSON, E. S. (1930). A further development of tests for normality, *Biometrika 22*, 239–249.

[17] PEARSON, E. S. (1931). Note on tests for normality, *Biometrika 22*, 423–424.

[18] PEARSON, E. S. (1963). Some problems arising in approximating to probability distributions, using moments, *Biometrika 50*, 95–112.

[19] PEARSON, E. S. (1965). Tables of percentage points of $\sqrt{b_1}$ and $b_2$ in normal samples; a rounding off, *Biometrika 52*, 282–285.

[20] PEARSON, E. S. and STEPHENS, M. A. (1964). The ratio of range to standard deviation in the same normal sample, *Biometrika 51*, 484–487.

[21] Rand Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates*, Glencoe, Ill.: The Free Press Publishers.

[22] SHAPIRO, S. S. and FRANCIA, R. S. (1972). Approximate analysis of variance test for normality, *J. Am. Statist. Ass. 67*, 215–216.

[23] SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika 52*, 591–611.

[24] SHAPIRO, S. S. and WILK, M. B. (1968). Approximation for the null distribution of the $W$ statistic, *Technometrics 10*, 861–866.

[25] SHAPIRO, S. S., WILK, M. B., and CHEN, H. J. (1968). A comparative study of various tests for normality, *J. Am. Statist. Ass. 63*, 1343–1372.

[26] WILLIAMS, P. (1935). Note on the sampling distribution of $\sqrt{\beta_1}$, where the population is normal, *Biometrika 27*, 269–271.