COMMENTARIES

A Suggestion for Using Powerful and Informative Tests of Normality

RALPH B. D'AGOSTINO, ALBERT BELANGER, AND RALPH B. D'AGOSTINO, JR.*

For testing that an underlying population is normally distributed the skewness and kurtosis statistics, $\sqrt{b_1}$ and b_2, and the D'Agostino–Pearson K^2 statistic that combines these two statistics have been shown to be powerful and informative tests. Their use, however, has not been as prevalent as their usefulness. We review these tests and show how readily available and popular statistical software can be used to implement them. Their relationship to deviations from linearity in normal probability plotting is also presented.

KEY WORDS: $\sqrt{b_1}$, b_2; D'Agostino–Pearson K^2; Kurtosis; Normal probability plot; Skewness.

1. INTRODUCTION

Tests of normality are statistical inference procedures designed to test that the underlying distribution of a random variable is normally distributed. There is a long history of these tests, and there are a plethora of them available for use (D'Agostino 1971; D'Agostino and Stephens 1986, chap. 9). Major studies investigating the statistical power of these over a wide range of alternative distributions have been undertaken, and a reasonably consistent picture has emerged as to which of these should be recommended for use. See D'Agostino and Stephens (1986, chap. 9) for a review of these power studies. The Shapiro–Wilk W test (Shapiro and Wilk 1965), the third sample moment ($\sqrt{b_1}$) and fourth sample moment (b_2) tests, and the D'Agostino–Pearson K^2 test combining these (D'Agostino and Pearson 1973) emerge as excellent tests. The W and K^2 tests share the fine property of being omnibus tests, in that they have good power properties over a broad range of nonnormal distributions. The $\sqrt{b_1}$ and b_2 tests have excellent properties for detecting nonnormality associated with skewness and nonnormal kurtosis, respectively. The extensive power studies just mentioned have also demonstrated convincingly that the old warhorses, the chi-squared test and the Kolmogorov test (1933), have poor power properties and should not be used when testing for normality.

Unfortunately, the preceding results have not been disseminated very well. The chi-squared and Kolmogorov tests are still suggested in textbooks for testing for normality. Major statistical packages such as SAS and SPSSX perform the excellent Shapiro–Wilk W test for sample sizes up to 50. For larger samples, however, they supply the poor-power Kolmogorov test. These statistical packages do give skewness and kurtosis measures. They are not, however, the $\sqrt{b_1}$ and b_2 statistics. Rather they are functions of these, the so-called Fisher g statistics (Fisher 1973). The documentation on this latter point is very incomplete. In our experience, many users are unaware of it, and descriptive evaluation of normality or nonnormality is confused because of it. Hypothesis testing using the powerful $\sqrt{b_1}$ and b_2 is not presented or even suggested.

In this article, we discuss the skewness, $\sqrt{b_1}$, and kurtosis, b_2, statistics and indicate how they are excellent descriptive and inferential measures for evaluating normality. Further, we relate the Fisher g skewness and kurtosis measures produced by the SAS and SPSSX software packages to $\sqrt{b_1}$ and b_2 and show how a simple program (SAS macro) can be used to produce an excellent, informative analysis for investigating normality. This analysis contains separate tests based on $\sqrt{b_1}$ and b_2 and the K^2 test, which combines $\sqrt{b_1}$ and b_2 for an omnibus test. Finally, we indicate how the preceding can be used in conjunction with normal probability plotting. The latter gives an informative graphical component to an analysis for normality.

2. POPULATION–MOMENTS DESCRIPTION OF NORMALITY AND NONNORMALITY

A population, or its random variable X, is said to be normally distributed if its density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0. \end{array} \quad (1)$$

Here μ and σ are the mean and standard deviation, respectively, of it. Of interest here are the third and fourth standardized moments given by

$$\sqrt{\beta_1} = \frac{E(X - \mu)^3}{[E(X - \mu)^2]^{3/2}} = \frac{E(X - \mu)^3}{\sigma^3} \qquad (2)$$

and

$$\beta_2 = \frac{E(X - \mu)^4}{[E(X - \mu)^2]^2} = \frac{E(X - \mu)^4}{\sigma^4}, \qquad (3)$$

where E is the expected value operator. These moments measure skewness and kurtosis, respectively, and for the normal distribution they are equal to 0 and 3, respectively. The nonnormality of a population can be described by values of its central moments differing from the normal values. The normal distribution is symmetric, so $\sqrt{\beta_1} = 0$. A nonnormal distribution that is asymmetrical has a value of $\sqrt{\beta_1} \neq 0$ (see Fig. 1); $\sqrt{\beta_1} > 0$ corresponds to skewness to the right and $\sqrt{\beta_1} < 0$ corresponds to skewness to the left.

The word *kurtosis* means "curvature," and it has traditionally been measured by the fourth standardized moment β_2. For the normal distribution, its value is 3. Figure 1 displays two nonnormal distributions in which $\beta_2 \neq 3$. Unimodal distributions whose tails are heavier or thicker than the normal have $\beta_2 > 3$. These distributions also tend to have higher peaks in the center of the distribution, and in the past these distributions were often described in terms of the high peaks (leptokurtic). Unimodal distributions whose
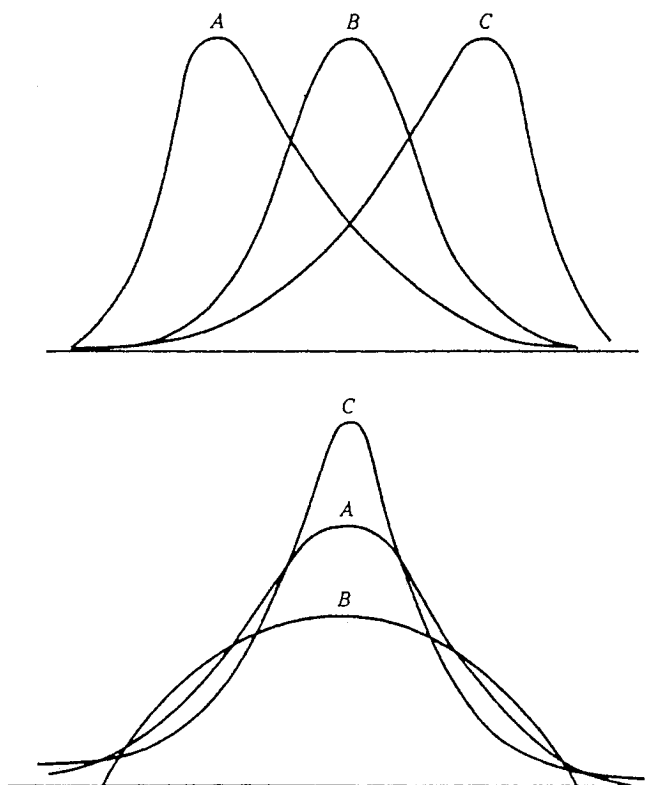

Figure 1. Illustration of Distributions With $\sqrt{\beta_1} \neq 0$ and $\beta_2 \neq 3$; Distributions Differing in Skewness and Differing in Kurtosis from the Normal Distribution; Top panel: A, $\sqrt{\beta_1} > 0$; B, $\sqrt{\beta_1} = 0$; C, $\sqrt{\beta_1} < 0$; Bottom panel: A, $\beta_2 = 3$; B, $\beta_2 < 3$; C, $\beta_2 > 3$.

tails are lighter than the normal tend to have $\beta_2 < 3$. In terms of their peak, it tends to be broader than the normal (platykurtic). Readers are referred to D'Agostino and Stephens (1986) for further discussion of these and to Balanda and MacGillivray (1988) for a detailed discussion of kurtosis. D'Agostino and Stephens (1986) also gave examples of well-known nonnormal distributions indexed by $\sqrt{\beta_1}$ and β_2.

3. SAMPLE MOMENTS AS INDICATORS OF NONNORMALITY

Karl Pearson (1895) was the first to suggest that the sample estimates of $\sqrt{\beta_1}$ and β_2 could be used to describe nonnormal distributions and used as the bases for tests of normality. For a sample of size n, $X_1, \ldots, X_n$ the sample estimates of $\sqrt{\beta_1}$ and β_2 are, respectively,

$$\sqrt{b_1} = m_3/m_2^{3/2} \qquad (4)$$

and

$$b_2 = m_4/m_2^2, \qquad (5)$$

where

$$m_k = \Sigma(X_i - \bar{X})^k/n \qquad (6)$$

and $\bar{X}$ is the sample mean

$$\bar{X} = \Sigma X_i/n. \qquad (7)$$

As descriptive statistics, values of $\sqrt{b_1}$ and b_2 close to 0 and 3, respectively, indicate normality. To be more precise the expected values of these are 0 and $3(n - 1)/(n + 1)$ under normality. Values differing from these are indicators of nonnormality. The signs and magnitudes of these give information about the type of nonnormality [e.g., $\sqrt{b_1} > 0$ corresponds to positive skewness and $b_2 > 3(n - 1)/(n + 1)$ relates to heavy tails in the population distribution].

4. TESTS OF NORMALITY BASED ON SAMPLE MOMENTS

The $\sqrt{b_1}$ and b_2 statistics are the bases for powerful tests of normality (D'Agostino and Stephens 1986, chap. 9).

4.1 Tests of Skewness ($\sqrt{b_1}$)

Here the null hypothesis is H_0: normality versus the alternative; H_1: nonnormality due to skewness. For alternatives ($\sqrt{\beta_1} \neq 0$), a two-sided test based on $\sqrt{b_1}$ is performed. For directional alternatives ($\sqrt{\beta_1} > 0$ or $\sqrt{\beta_1} < 0$), one-sided tests are performed. Tables of critical values are available (D'Agostino and Stephens 1986, chap. 9). For sample sizes $n > 8$, a normal approximation that is easily computerized is available. It is obtained as follows (D'Agostino 1970):

1. Compute $\sqrt{b_1}$ from the sample data.
2. Compute

$$Y = \sqrt{b_1}\left\{\frac{(n + 1)(n + 3)}{6(n - 2)}\right\}^{1/2}, \qquad (8)$$

$$\beta_2(\sqrt{b_1}) = \frac{3(n^2 + 27n - 70)(n + 1)(n + 3)}{(n - 2)(n + 5)(n + 7)(n + 9)}, \qquad (9)$$

$$W^2 = -1 + \{2(\beta_2(\sqrt{b_1}) - 1)\}^{1/2}, \quad (10)$$

$$\delta = 1/\sqrt{\ln W}, \quad (11)$$

and

$$\alpha = \{2/(W^2 - 1)\}^{1/2}. \quad (12)$$

3. Compute

$$Z(\sqrt{b_1}) = \delta \ln(Y/\alpha + \{(Y/\alpha)^2 + 1\}^{1/2}). \quad (13)$$

$Z(\sqrt{b_1})$ is approximately normally distributed under the null hypothesis of population normality.

4.2 Tests of Kurtosis (b_2)

Here the null hypothesis is H_0: normality versus the alternative; H_1: nonnormality due to nonnormal kurtosis. Again a two-sided test (for $\beta_2 \neq 3$) or one-sided tests (for $\beta_2 > 3$ or $\beta_2 < 3$) can be performed. Again elaborate tables are available (D'Agostino and Stephens 1986, chap. 9). Moreover, a normal approximation due to Anscombe and Glynn (1983) is available. It is valid for $n \geq 20$ and is as follows:

1. Compute b_2 from the sample data.
2. Compute the mean and variance of b_2,

$$E(b_2) = \frac{3(n - 1)}{n + 1} \quad (14)$$

and

$$\text{var}(b_2) = \frac{24n(n - 2)(n - 3)}{(n + 1)^2(n + 3)(n + 5)}. \quad (15)$$

3. Compute the standardized version of b_2,

$$x = (b_2 - E(b_2))/\sqrt{\text{var}(b_2)}. \quad (16)$$

4. Compute the third standardized moment of b_2,

$$\sqrt{\beta_1(b_2)} = \frac{6(n^2 - 5n + 2)}{(n + 7)(n + 9)} \sqrt{\frac{6(n + 3)(n + 5)}{n(n - 2)(n - 3)}}. \quad (17)$$

5. Compute

$$A = 6 + \frac{8}{\sqrt{\beta_1(b_2)}} \left[\frac{2}{\sqrt{\beta_1(b_2)}} \right.$$
$$\left. + \sqrt{\left(1 + \frac{4}{\beta_1(b_2)}\right)} \right]. \quad (18)$$

6. Compute

$$Z(b_2) = \left(\left(1 - \frac{2}{9A}\right) - \left[\frac{1 - 2/A}{1 + x\sqrt{2/(A - 4)}}\right]^{1/3} \right) / \sqrt{2/(9A)}. \quad (19)$$

$Z(b_2)$ is approximately normally distributed under the null hypothesis of population normality.

Both $Z(\sqrt{b_1})$ and $Z(b_2)$ can be used to test one-sided and two-sided alternative hypotheses.

4.3 Omnibus Test

D'Agostino and Pearson (1973) presented a statistic that combines $\sqrt{b_1}$ and b_2 to produce an omnibus test of normality. By *omnibus*, we mean it is able to detect deviations from normality due to either skewness or kurtosis. The test statistic is

$$K^2 = Z^2(\sqrt{b_1}) + Z^2(b_2), \quad (20)$$

where $Z(\sqrt{b_1})$ and $Z(b_2)$ are the normal approximations to $\sqrt{b_1}$ and b_2 discussed in Sections 4.1 and 4.2. The K^2 statistic has approximately a chi-squared distribution, with 2 df when the population is normally distributed.

5. NUMERICAL EXAMPLE

Table 1 contains a sample of cholesterol values from a sample of 62 subjects from the Framingham Heart Study. The data are presented as a stem-and-leaf plot. From these data we obtain

$$\sqrt{b_1} = 1.02, \quad Z(\sqrt{b_1}) = 3.14, \quad p = .0017,$$

$$b_2 = 4.58, \quad Z(b_2) = 2.21, \quad p = .0269,$$

and

$$K^2 = 14.75, \quad p = .0006.$$

The preceding p values are the levels of significance for the corresponding two-sided tests. For the Kolmogorov test, $p = .087$. The data are clearly nonnormal. The $\sqrt{b_1}$ and b_2 statistics quantify the nature of the nonnormality. The data distribution is skewed to the right and heavy in the tails. The Kolmogorov test gives no information about this nonnormality and only indicates marginally nonnormality.

6. THE FISHER g STATISTICS

Both SAS and SPSSX routinely give skewness and kurtosis statistics in their descriptive statistics output. Unfor-

Table 1. Cholesterol Data From the Framingham Heart Study

	Stem-and-leaf plot	Number
39	3	1
38		
37		
36		
35	3	1
34		
33	46	2
32	7	1
31		
30	008	3
29		
28	35	2
27	00288	5
26	347778	6
25	444668	6
24	03678	5
23	0000122244668	13
22	0556	4
21	0125678	7
20	02	2
19	28	2
18	4	1
17		
16	7	1

NOTE: The descriptive statistics are sample size, n = 62; mean, $\bar{x}$ = 250.0; standard deviation, S = 41.4; skewness, $\sqrt{b_1}$ = 1.02; kurtosis, b_2 = 4.58.

tunately, neither give $\sqrt{b_1}$ and b_2. Rather, they give the Fisher g statistics defined as follows:

$$\text{skewness} \quad g_1 = \frac{n\Sigma(X - \overline{X})^3}{(n - 1)(n - 2)S^3} \qquad (21)$$

and

$$\text{kurtosis} \quad g_2 = \frac{n(n + 1)\Sigma(X - \overline{X})^4}{(n - 1)(n - 2)(n - 3)S^4}$$
$$- \frac{3(n - 1)^2}{(n - 2)(n - 3)}, \qquad (22)$$

where

$$S^2 = \frac{n}{n - 1} m_2 = \frac{\Sigma(X - \overline{X})^2}{n - 1} \qquad (23)$$

is the sample variance.

These are related to $\sqrt{b_1}$ and b_2 via the following:

$$\sqrt{b_1} = \frac{(n - 2)}{\sqrt{n(n - 1)}} g_1 \qquad (24)$$

and

$$b_2 = \frac{(n - 2)(n - 3)}{(n + 1)(n - 1)} g_2 + \frac{3(n - 1)}{(n + 1)}. \qquad (25)$$

The BMDP statistics software package does compute $\sqrt{b_1}$ and b_2. All of the preceding software do not perform tests of normality based on skewness and kurtosis.

7. RECOMMENDATIONS

The tests just described based on $\sqrt{b_1}$ and b_2 are excellent and powerful tests. We recommend that for all sample sizes $\sqrt{b_1}$ and b_2 should be computed and examined as descriptive statistics. For all sample sizes $n \geq 9$, tests of hypotheses can be based on them. In particular, for $n > 50$, where the Shapiro–Wilk test is no longer available, we recommend these tests and the D'Agostino–Pearson K^2 test as the tests of choice. The justification for this is not only because of their fine power but also because of the information they supply on nonnormality. In conjunction with the use of standard statistical software, such as SAS, SPSSX, and BMDP, the skewness and kurtosis measures they produce can be used as inputs to simple programs (macros) to perform these tests. In the appendix, we supply one such simple macro that can be used with SAS and that will provide two-tailed tests.

8. NORMAL PROBABILITY PLOT

Another component in a good data analysis for investigating normality of data and an item again often not well handled routinely in computer packages is the normal probability plot. This plot is a graphical presentation of the data that will be approximately a straight line if the underlying distribution is normal. Deviations from linearity correspond to various types of nonnormality. Some of these deviations reflect skewness and/or kurtosis. Others reflect features such as the presence of outliers, mixtures in the data, or truncation (censoring) in the data. Readers are referred to D'Agostino

and Stephens (1986, chap. 2) for a detailed discussion of probability plotting.

A normal probability plot is simply a plot of the inverse of the standard normal cumulative on the horizontal axis and the ordered observations on the vertical axis. The inverse of the normal cumulative is usually defined in such a way to enhance the linearity of the plot, and one common procedure is to let the normal probability plot employ Blom's (1958) approximation. In this case, the normal probability plot is a plot of

$$Z = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right) \quad \text{on} \quad X_{(i)}, \qquad (26)$$

where $X_{(i)}$ is the ith ordered observation from the ordered sample $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ and

$$Z = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right) \qquad (27)$$

is the Z value such that

$$\frac{i - 3/8}{n + 1/4} = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \qquad (28)$$

for $i = 1, \ldots, n$.

Figure 2 is a normal probability plot of the data of Section 5. The expected straight line can be obtained by connecting the +'s on the graph. Figure 3 contains a number of forms that the normal probability plot will produce in the presence of nonnormality. For the present data set, its skewness to the right is very evident in the plot.

A program for the normal probability plot applicable to SAS is part of the macro given in the appendix.

9. CONCLUSION

We have discussed the uses of $\sqrt{b_1}$ and b_2 as descriptive and inferential statistics for evaluating the normality of data. We have made specific recommendations for their uses. Further we have reviewed briefly the normal probability plot, which can be used in conjunction with $\sqrt{b_1}$ and b_2 for a graphical analysis. A good complete normality analysis would consist of the use of the plot plus the statistics. The use of these is superior to what is routinely given in standard computer software. Serious investigators should consider using the materials of this article in their data analysis.

APPENDIX: A MACRO FOR USE WITH SAS STATISTICAL SOFTWARE

The following macro takes as input a variable name and a data set name. It produces as output the results of a univariate descriptive analysis (PROC UNIVARIATE), skewness and kurtosis measures and test statistics, the D'Agostino–Pearson omnibus normality test statistic, p levels, and a normal probability plot.

```
%MACRO NORMTEST(VAR,DATA);
PROC UNIVARIATE NORMAL PLOT DATA=&DATA;
    VAR &VAR; OUTPUT OUT=XXSTAT N=N
        MEAN=XBAR STD=S SKEWNESS=G1
        KURTOSIS=G2;
```
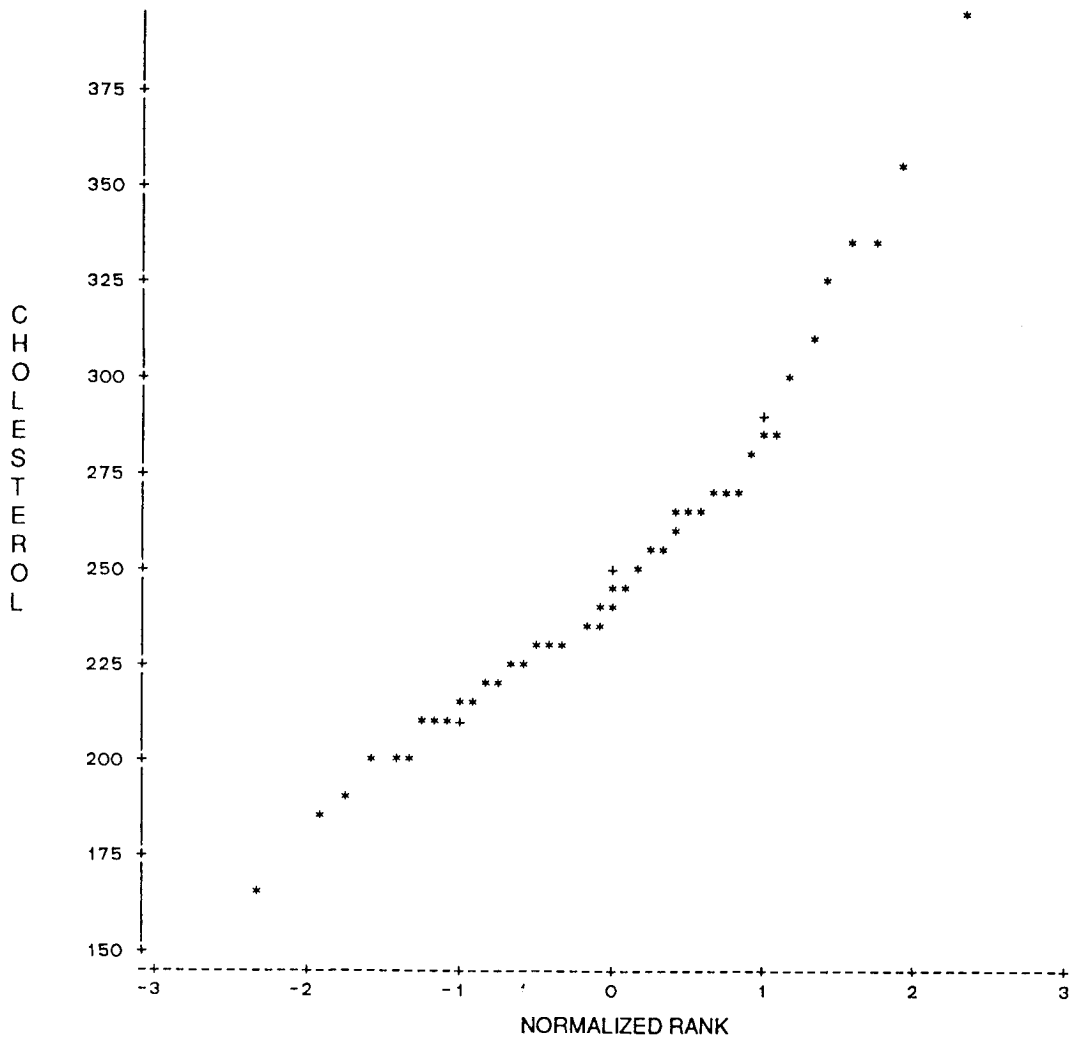
Figure 2. Normal Probability Plot of Cholesterol Data.

```
DATA;
   SET XXSTAT;
   DO _Z_ = −1,0,1; _X_=XBAR+_Z_*S; OUTPUT;
      END;
   KEEP _X_ _Z_;
DATA; SET &DATA _LAST_;
PROC RANK TIES = MEAN NORMAL = BLOM; VAR
      &VAR; RANKS BLOMRANK;
OPTIONS LS = 80;
PROC PLOT NOLEGEND;
   PLOT &VAR*BLOMRANK = '*' _X_*_Z_ = ' + '/
      OVERLAY HAXIS = −3 TO 3 BY .5;
   LABEL BLOMRANK = "NORMALIZED RANK"
      &VAR = "NORMAL PROBABILITY PLOT FOR
      &VAR";
DATA;
   SET XXSTAT;
   SQRTB1 = (N−2)/SQRT(N*(N−1))*G1;
   Y = SQRTB1*SQRT((N+1)*(N+3)/(6*(N−2)));
   BETA2 = 3*(N*N+27*N−70)*(N+1)*(N+3)/
      ((N−2)*(N+5)*(N+7)*(N+9));
   W = SQRT(−1+SQRT(2*(BETA2−1)));
   DELTA = 1/SQRT(LOG(W));
   ALPHA = SQRT(2/(W*W−1));
```

```
Z_B1 = DELTA*LOG(Y/ALPHA+SQRT((Y/
      ALPHA)**2+1));
B2 = 3*(N−1)/(N+1)+(N−2)*(N−3)/
      ((N+1)*(N−1))*G2;
MEANB2 = 3*(N−1)/(N+1);
VARB2 = 24*N*(N−2)*(N−3)/
      ((N+1)*(N+1)*(N+3)*(N+5));
X = (B2−MEANB2)/SQRT(VARB2);
MOMENT = 6*(N*N−5*N+2)/
      ((N+7)*(N+9))*SQRT(6*(N+3)*(N+5)/
      (N*(N−2)*(N−3)));
A = 6+8/MOMENT*(2/MOMENT+SQRT(1+4/
      (MOMENT**2)));
Z_B2 = (1−2/(9*A))−((1−2/A)/(1+X*SQRT(2/
      (A−4))))**(1/3))/SQRT(2/(9*A));
PRZB1 = 2*(1−PROBNORM(ABS(Z_B1)));
      PRZB2 = 2*(1−PROBNORM(ABS(Z_B2)));
CHITEST = Z_B1*Z_B1+Z_B2*Z_B2;
      PRCHI = 1−PROBCHI(CHITEST,2);
FILE PRINT;
PUT "NORMALITY TEST FOR VARIABLE &VAR "
      N= /
   @20 G1 = 8.5 @33 SQRTB1 = 8.5 @50 "Z =" Z_
      B1 8.5 " P =" PRZB1 @6.4/
```

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.