THE PROCEDURE DURBIN

M. E. McBride, Clarkson College of Technology
R. P. Parks, Washington University

Testing the hypothesis that a sample of observations \( x_1, \ldots, x_n \) is a random sample from a normal distribution \( \mathcal{N}(\mu, \sigma^2) \) is possible through a variety of statistical techniques. One of the earlier techniques was Pearson's \( \chi^2 \) test. More recent research has centered on the Shapiro-Wilk test and the Kolmogorov test. This paper describes four such tests of normality and presents a SAS procedure which performs the necessary computations. Three of the tests are due to a 1961 Biometrika article by J. Durbin. The fourth test is due to a 1965 Biometrika article by Shapiro and Wilk.

The first section of the paper describes the four alternative tests. Each statistical test is presented with the necessary information to understand the use of the test. The second section provides a description of the procedure DURBIN, its features, how to call it, and an example run. The final section presents the results of a small Monte Carlo study to illustrate the power of the alternative tests.

THE TEST STATISTICS

The procedure DURBIN calculates four alternative tests of normality on a random sample of \( n \) deviates \( x_1, \ldots, x_n \). Three of the tests, due to J. Durbin [Biometrika 1961], are exact tests in that the tests yield exact probabilities of rejection when the null hypothesis is true. Furthermore, one does not need to know the mean and variance of the hypothesized normal distribution in order to achieve the exact tests. The fourth test, due to Shapiro and Wilk [Biometrika 1965], is based upon an empirical distribution developed by Shapiro and Wilk. The Shapiro-Wilk test is also scale and origin invariant.

The problem is to test whether a set of \( n \) deviates \( x_1, \ldots, x_n \) is a random sample from a distribution of the form \( F(x) \). Letting \( u_i = F(x_i) \) \((i=1, \ldots, n)\), then the hypothesis that the \( u_i \)'s have distribution function \( F(x) \) is equivalent to the hypothesis that the \( u_i \)'s are distributed uniformly over the interval zero to one, i.e., \( U(0,1) \). Departure from the null hypothesis will be indicated by shorter (or longer) intervals between the ordered \( u_i \)'s than would be found in a random scatter. The heart of the Durbin procedure is to transform the ordered \( u_i \)'s so that any uneven spacing is highlighted, i.e., smaller intervals are made smaller and larger intervals are made larger. Thus, letting

\[
\begin{align*}
(3) & \quad c_0 = u(1) \\
(4) & \quad c_i = u(1) - u(i-1) \quad i=2, \ldots, n \\
(5) & \quad c_{n+1} = 1 - u(n),
\end{align*}
\]

The \( c_i \)'s are now ordered since it is their relative magnitudes which are of interest

\[
(4) \quad c(1) \leq c(2) \leq \ldots \leq c(n+1).
\]

Another transformation is made to set the ordered \( c_i \)'s into a more manageable form:

\[
(5) \quad g_i = (n+2-i) \left( c(i) - c(i-1) \right) \quad i=1, \ldots, n+1
\]

Note that each \( g_i \geq 0 \) and that

\[
\sum_{i=1}^{n+1} g_i = \sum_{i=1}^{n} c(i) = 1.
\]

Durbin has shown the remarkable result that the \( g_1, \ldots, g_{n+1} \) which depend on the ordered intervals have the same distribution as the unordered intervals \( c_1, \ldots, c_{n+1} \). This is the key result to Durbin's transformation.

Letting

\[
(6) \quad w_r = \sum_{i=1}^{r} g_i
\]

it follows that the \( w_1, \ldots, w_n \) have the same distribution as the ordered \( n \) \( U(0,1) \) variables \( u(1), \ldots, u(n) \). Thus, any test statistic which depends on the ordered \( u_i \)'s (e.g., Kolmogorov-Smirnov) have the same properties under the null hypothesis when the \( w_i \)'s are used. The transformation is made in the hope of gaining additional power for the test, particularly in alternatives where the differences between the two distribution functions are nowhere large but the frequency function differs.

Durbin proposed three different goodness of fit tests. All are hypotheses that the \( w_i \)'s are as ordered \( U(0,1) \) variables. In order to make computation easier, note that the \( w_i \)'s can be rewritten from (4) and (6) to

\[
(7) \quad w_i = c(1) + \ldots + c(i-1) + (n+2-i) c(i) \quad i=1, \ldots, n.
\]

The first test is the modified median test. The test is based on the median of the \( w_i \)'s. The probability element required is that in a sample of \( n \) uniform variables \( w_r \) are greater than \( w_r \), one is in the range \( w_r + dw_r \), and \( r \) are less than \( w_r \). The test statistic is

\[
(8) \quad m_r = \frac{(r+1-r)}{w_r} - w_r
\]

which has an \( F \) distribution with \( 2(n+1-r), 2r \) degrees of freedom. Since under the alternative,
w, is expected to decrease with departures from the null hypothesis, a one-tail F-test is appropriate.

The second test statistic is the modified Kolmogorov test. This test considers the difference between the sample and population distribution functions corresponding to \( w_1, \ldots, w_n \). The test statistic is

\[
K = \max_{i=1}^{n} \left( \frac{1}{n} - w_i \right).
\]

The test statistic \( K \) is expected to increase with departures from the null hypothesis, and thus a one-tail test is appropriate. Appropriate tables for this exact test can be found in Miller (1956).

The third test statistic proposed by Durbin is the modified probability product. The test statistic derives from the observation that

\[
P = \prod_{i=1}^{n} u_i
\]

and that \( \prod_{i=1}^{n} u_i \) is distributed as \( \exp(-1/2 \chi^2) \).

The test statistic is

\[
P = 2 \log \prod_{i=1}^{n} w_i.
\]

The test statistic \( P \) is tested as an \( \chi^2 \) with \( 2n \) degrees of freedom.

A serious drawback to the above tests is the presence of nuisance parameters. In the Kolmogorov test, an unknown mean and variance prohibits exact tests. Durbin suggested a randomization procedure to get exact tests in a situation where the mean and variance are unknown.

The desire is to test the hypothesis that \( x_1, \ldots, x_n \) are independent observations from a normal distribution with unknown mean \( \mu \) and variance \( \sigma^2 \). Define \( \bar{x} = \sum x_i \) and \( s^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2 \). Let \( \bar{x} \) and \( s \) be observations of random variables drawn independently of \( x_1, \ldots, x_n \) and distributed as the sample mean and variance (i.e., \( \bar{x} \) and \( s^2 \)) of an independent sample of size \( n \) from a \( N(0,1) \) distribution. Now transform the \( x_1, \ldots, x_n \) to \( x_1^*, \ldots, x_n^* \) through the following relationship

\[
x_i^* = \frac{x_i - \bar{x}}{s}, \quad i = 1, \ldots, n.
\]

Durbin has shown that the \( x_1^*, \ldots, x_n^* \) are distributed \( N(0,1) \). Thus, any of the exact tests discussed above are applicable to the \( x_1^*, \ldots, x_n^* \).

The price that is paid for being able to conduct exact tests is entering a degree of randomization to the test. Therefore, any two researchers could arrive at different results for the same set of data due to the random draws of \( \bar{x} \) and \( s^2 \).

An alternative test of normality, which has been empirically shown to have greater power than almost all other tests, is the Shapiro-Wilk test. The test resulted from work analyzing analysis of variance tests on probability plots of the ordered sample \( x_{(1)}, \ldots, x_{(n)} \) against the expected value of the order statistics of the hypothesized null distribution. If the hypothesized distribution is true, then the plot should be linear. The test statistic is the ratio of the squared slope of the probability plot regression line and the usual symmetric sample sum of squares about the mean.

Formally, given a random sample \( x_1, \ldots, x_n \), the test statistic is defined as

\[
w = b^2 / s^2
\]

where \( s^2 = \sum (x_i - \bar{x})^2 \) and \( b \) is, up to a multiplicative constant, the best linear unbiased estimate of the slope of a linear regression of the ordered observations, \( x_{(i)} \), on the expected values, \( m_i \), of the standard normal order statistics. The slope coefficient \( b \) may be calculated as

\[
b = \sum_{i=1}^{n-1} a_{n-i+1} (x_{(n-i+1)} - x_{(i)})
\]

where

\[
R = n/2 \text{ for even } n \text{ and } R = (n+1)/2 \text{ for odd } n \text{ (the median does not affect } b \text{ when } n \text{ is odd). The } a_i \text{'s are the normalized 'best linear unbiased' coefficients for order statistics tabulated in Sarhan and Greenberg (1966). Small values of } w \text{ indicate nonnormality.}

Although the \( w \) statistic has been empirically shown to have greater power than other tests of normality (e.g., Shapiro, Wilk, and Chen (1968)), it has some serious limitations. First, the \( a_i \)'s are known exactly only for \( n \leq 20 \). Approximations are used for \( n > 20 \), which are

\[
a_i = \frac{\Gamma(1/2h)}{\sqrt{2} \Gamma(1/2(h+1))}
\]

where \( h = n/2 \text{ for even } n \text{ and } h = (n+1)/2 \text{ for odd } n \). The rest of the \( a_i \)'s are

\[
a_i = \frac{2m_i}{c_i}
\]

where \( c_i = -2.722 + 4.083n \) and the \( m_i \)'s are available from Harter (1961).

For \( n \leq 20 \), \( c_i \) is known exactly and Shapiro-Wilk calculated the above linear interpolation for \( n = 2(1)20 \). Although the line is quite good for \( n = 2(1)20 \), one has no idea of how good the approximation is for \( n > 20 \), nor for the approximations of the \( a_i \) by the expected order statistics \( m_i \). Second, the null distribution of the \( w \) statistic is unknown. Shapiro and Wilk tabulated an empirical null distribution for \( n \leq 50 \), and provided normal approximations of it through Johnson's \( s_i \) transformation (see Shapiro and Wilk (1968)). Hence the test is an approximate one and has unknown properties for \( n > 50 \) (a very unusual case in residual analysis in econometrics).

THE PROCEDURE

The procedure DURBIN is PL-1 based and was constructed by modifying the code in PROC RANK. This insures efficient reading of the data and ranking of the observations. The procedure can handle any number of variables and observations and BY groups.
The data is stored internally in single precision, but calculations use double precision. Missing values are ignored. The SAS-based routines for $X$ and $F$ probabilities are used. The probabilities reported for the Shapiro-Wilk statistic use Johnson's $s_i$ transformation to a normal variate (Shapiro-Wilk (1965)) and the three parameters of that transformation are calculated from regression $N, \log(N), 1/N^2$ and $N^2$, using the tabled numbers in Shapiro-Wilk (1968) for $n = 2(1)50$. These probabilities appear to have at least a 10% error compared to the tabled distribution reported in Shapiro-Wilk (1968).

All of the Durbin statistics are exact calculations. The nulls transformation given in equation (1) is obtained by either summing 12 uniform deviates or a fast normal generator. Both are based on a simple uniform deviate generator (see Knuth (1971)). The fast normal generator operates as $x_i = z_{i(2n-1)} - 1, 2, and s = v_1^2 + v_2^2$. If $s < 1$, then a standard normal deviate is $v_1 \sqrt{-2 \ln s}$. Optionally, the fast normal or the sum of 12 uniform deviates can be used to calculate $x_i$ and $s_i$ in equation (1).

The Shapiro-Wilk statistic is calculated using the approximations in equations (13) and (14) for all values of $n$. The $m_i, n$ are calculated by

$$m_i, n = N \left[ (i-\alpha)/(n-2i+1) \right]$$

where $\alpha = 0.325711 + 0.058212 \cdot \log_{10}(n)$. According to Harter, the $m_i, n$ are very good approximations to the expected values of the standard normal rank order statistics for $n \leq 400$. As mentioned before, for $n > 20$ the approximations given in (13) and (14) are of unknown accuracy, although the calculated $a_i$'s differ from the tabled $a_i$'s in Shapiro-Wilk by at most 3% for $n \leq 50$. There is much greater error in the calculation of the null distribution of the Shapiro-Wilk than in its calculation.

The two options of the procedure are: which normal generator is used to calculate $x_i$ and $s_i$ in (11) and to allow $x_i = 0$ and $s_i = 1$. These options are in effect for all variables and BY groups.

### Output

The Durbin procedure provides for each variable the following output: (a) the variable name and the number of observations, (b) the mean and standard deviation, (c) the Durbin corrections, (d) the modified median test (with associated significance level), (e) the modified probability product test (with associated significance level), (f) the modified Kolmogorov test (significance level is not printed, see Miller (1956) for appropriate tables), (g) the Shapiro-Wilk test (with one minus the significance level), and (h) the variable labels if the line size parameter is greater than 80. The Durbin procedure does not produce an output data set.

### PROC Durbin Statement

```
PROC DURBIN options and parameters;
DATA = data_set_name
```

The DATA parameter tells DURBIN the SAS data set to be used. If it is left off, the last created data set will be used.

The options provided in PROC DURBIN are:

1 - the Durbin nuisance corrections are taken to be $X = 0$ and $s^2 = 1$, i.e., there are no nuisance corrections. The tests are valid assuming the null distribution is standard normal, i.e., $N(0,1)$. Without this option, the null distribution is only assumed normal and the nuisance corrections are calculated.

2 - if this option is specified, the fast normal generator is used to calculate the Durbin nuisance corrections. Otherwise, 12 uniform deviates are summed to form a normal variate.

### VARIABLES list_of_variables;

The test statistic will be computed for the variables listed in the VARIABLES statement. If the VARIABLES statement is omitted, then the program will compute the statistics on all variables. WARNING, only numeric variables can be used in the procedure.

```
BY list_of_variables;
```

If a BY statement is included, the data set must already be sorted by the variables in the BY statement. You can use the SORT procedure to sort the data set. The statistics for one group of observations will be printed directly after the statistics for the preceding group of observations. Only three lines will separate the statistics for the two groups.

### Treatment of Missing Values

If a variable included in the variables statement has missing values, these observations are completely excluded from the analysis.

### Example

Figure 1 provides a sample run of the procedure.
DATA TEST;
   TR: TRIAL+1; N=0; OBS: N+1;
10   UNIFORM=UNIFORM(11111);
11   NORMAL=NORMAL(11111);
12   LABEL NORMAL=NORMAL GENERATED BY SAS;
13   LABEL UNIFORM=UNIFORM GENERATED BY SAS;
14   OUTPUT: IF N<19 THEN GOTO OBS;
15   IF TRIAL=2 THEN GOTO TR;

NOTE: DATA SET WORK.TEST HAS 60 OBSERVATIONS AND 4 VARIABLES. 361 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.44 SECONDS AND 102K.

PROC DURBIN: BY TRIAL: VAR UNIFORM NORMAL;
NOTE: THE PROCEDURE DURBIN USED 1.54 SECONDS AND 170K AND PRINTED PAGE 2.
NOTE: SAS USED 170K MEMORY.

NOTE: BARR, GOODNIGHT, SALL AND HELWIG
SAS INSTITUTE INC.
P.O. BOX 10066
RALEIGH, N.C. 27605

<table>
<thead>
<tr>
<th>NAME/OBS</th>
<th>MEAN</th>
<th>ST.DEV.</th>
<th>DURBIN CORRECTIONS</th>
<th>MEAN</th>
<th>ST.DEV.</th>
<th>MEDIAN</th>
<th>PRODUCT</th>
<th>MOD KS</th>
<th>S-W</th>
<th>LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIFORM</td>
<td>0.450</td>
<td>0.233</td>
<td>-0.107</td>
<td>1.305</td>
<td>1.515</td>
<td>44.126</td>
<td>0.207</td>
<td>0.950</td>
<td>0.633 UNIFORM GENERATED BY SAS</td>
<td></td>
</tr>
<tr>
<td>NORMAL</td>
<td>-0.060</td>
<td>1.048</td>
<td>0.197</td>
<td>0.990</td>
<td>1.561</td>
<td>44.956</td>
<td>0.196</td>
<td>0.963</td>
<td>0.410 NORMAL GENERATED BY SAS</td>
<td></td>
</tr>
<tr>
<td>TRIAL=2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNIFORM</td>
<td>0.511</td>
<td>0.290</td>
<td>0.293</td>
<td>0.845</td>
<td>0.850</td>
<td>37.427</td>
<td>0.130</td>
<td>0.939</td>
<td>0.767 UNIFORM GENERATED BY SAS</td>
<td></td>
</tr>
<tr>
<td>NORMAL</td>
<td>0.410</td>
<td>0.926</td>
<td>0.003</td>
<td>1.159</td>
<td>0.870</td>
<td>52.124</td>
<td>0.205</td>
<td>0.915</td>
<td>0.915 NORMAL GENERATED BY SAS</td>
<td></td>
</tr>
<tr>
<td>TRIAL=3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNIFORM</td>
<td>0.402</td>
<td>0.319</td>
<td>0.118</td>
<td>1.114</td>
<td>0.865</td>
<td>35.041</td>
<td>0.176</td>
<td>0.882</td>
<td>0.880 UNIFORM GENERATED BY SAS</td>
<td></td>
</tr>
<tr>
<td>NORMAL</td>
<td>-0.082</td>
<td>1.060</td>
<td>-0.132</td>
<td>1.075</td>
<td>0.695</td>
<td>46.218</td>
<td>0.127</td>
<td>0.904</td>
<td>0.948 NORMAL GENERATED BY SAS</td>
<td></td>
</tr>
</tbody>
</table>
THE POWER OF THE TESTS

A small Monte Carlo study was run to illustrate the power of the four tests against various alternatives. This set of trials was not intended to be conclusive evidence as to the power of the tests, but is provided only for expository purposes. Random samples of size 10, 25, and 50 were generated for seven alternative distributions: Chi-square, two exponentials, two normals, log normal, and uniform. These samples were then processed through the procedure Durbin. The trials were repeated 200 times. The results are tabulated in Table I for the modified median, modified probability product, and Shapiro-Wilk tests. 1

There are three major items to notice about the results. First, as is expected, the power of all three tests goes up as sample size increases. The power of the tests seems quite good against the five non-normal distributions, except in distinguishing between uniform and normal. Second, a ranking of the tests, in terms of power would be Shapiro-Wilk, modified probability product, and modified median. This is consistent with the findings of Shapiro and Wilk that even though the S-w test is based upon an empirical distribution, it still has good power against most alternative distributions and relative to other tests of normality. Finally, the ability of the Shapiro-Wilk test to accept normality when the underlying distribution actually is normal (i.e., type error) is quite poor. This could be due to the nature of the approximations being made in the calculation of the w statistic and in the calculation of the null distribution.

Footnotes

1 The notation of Durbin (1961) is followed for ease of following this article and Durbin's article. All proofs can be found in his article.

2 Again, Shapiro and Wilk's notation is being used in describing their test statistic.

3 The results of the modified Kolmogorov test were not tabulated due to a time factor involving determining rejection regions.

References


<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-sq.</th>
<th>Exp. 1</th>
<th>Exp. 2</th>
<th>Log(Norm)</th>
<th>Normal 1</th>
<th>Normal 2</th>
<th>Uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>38.6</td>
<td>19.4</td>
<td>16.9</td>
<td>30.8</td>
<td>5.5</td>
<td>6.0</td>
<td>8.5</td>
</tr>
<tr>
<td>Prob. Prod.</td>
<td>66.7</td>
<td>41.3</td>
<td>37.8</td>
<td>57.7</td>
<td>16.4</td>
<td>19.4</td>
<td>19.2</td>
</tr>
<tr>
<td>S-W</td>
<td>94.0</td>
<td>75.6</td>
<td>75.6</td>
<td>82.1</td>
<td>28.4</td>
<td>28.4</td>
<td>33.8</td>
</tr>
<tr>
<td>N = 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>93.0</td>
<td>55.7</td>
<td>58.7</td>
<td>78.6</td>
<td>8.46</td>
<td>10.0</td>
<td>18.4</td>
</tr>
<tr>
<td>Prob. Prod.</td>
<td>98.0</td>
<td>74.6</td>
<td>76.1</td>
<td>90.0</td>
<td>16.9</td>
<td>14.4</td>
<td>30.4</td>
</tr>
<tr>
<td>S-W</td>
<td>100.0</td>
<td>99.5</td>
<td>99.5</td>
<td>100.0</td>
<td>16.4</td>
<td>16.4</td>
<td>62.7</td>
</tr>
<tr>
<td>N = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>100.0</td>
<td>82.6</td>
<td>85.1</td>
<td>97.5</td>
<td>11.0</td>
<td>8.5</td>
<td>28.9</td>
</tr>
<tr>
<td>Prob. Prod.</td>
<td>100.0</td>
<td>93.5</td>
<td>91.5</td>
<td>99.5</td>
<td>15.4</td>
<td>17.4</td>
<td>40.8</td>
</tr>
<tr>
<td>S-W</td>
<td>100.0</td>
<td>99.5</td>
<td>99.5</td>
<td>100.0</td>
<td>34.8</td>
<td>34.8</td>
<td>61.2</td>
</tr>
</tbody>
</table>