Goodness of Fit Tests

Documentation on `libcdhc.a`

and

A GRASS Tutorial on `s.normal`

James Darrell McCauley*

September 12, 1994

Abstract

The methods used by the GRASS program `s.normal` are presented. These are various goodness of fit statistics for testing the null hypothesis of normality. Other additional tests found in `cdhc` a C programming library, are also documented (this document serves two puposes: a tutorial for the GRASS geographic information system and documentation for the library).

1 Introduction

This document is a programmer's manual for `cdhc`, a C programming library useful for testing whether a sample is normally, lognormally, or exponentially distributed. Prototypes for library functions[1] are given in the margins near corresponding mathematical explanations. Hence, it is also a user's guide for programs using `cdhc`. Readers should be equipped with at least one graduate course in probability and statistics. Much of the background and derivation/justification of each test has been omitted. A good text for more background information is *Goodness-of-Fit Techniques* by D'Agostino and Stephens [13] (see also references in text).

1.1 Hypothesis Testing

Before beginning the description of the tests, a few definitions should be given. The general framework for mosts tests is that the *null* hypothesis H_0 is that a random variable x follows a particular distribution $F(x)$. Generally, the *alternative* hypothesis is that x does not follow $F(x)$ (with no additional usuable

*USDA National Needs Fellow, Department of Agricultural Enginering, Purdue University. Email: `mccauley@ecn.purdue.edu`

[1] Each function in the library returns a pointer to static double. The `cdhc` library was inspired by Johnson's STATLIB collection of FORTRAN routines for testing distribution assumptions [22]. Some functions in `cdhc` are loosely based on Johnson's work (they have been completely rewritten, reducing memory requirements and number of computations and fixing a few bugs). Others are based on algorithms found in *Applied Statistics*, *Technometrics*, and other related journals.

information; the Kotz Separate Families test in §8 is one exception). This may differ from the way that some have learned hypothesis testing in that some tests are set up to reject the null hypothesis in favor of the alternative.

A *simple* hypothesis implies that $F(x)$ is completely specified, e.g., $x \sim N(0;1)$. A *composite* hypothesis means that one (or more) of the parameters of $F(x)$ is not completely specified, e.g., $x \sim N(\ ;\ )$. That is, the composite hypothesis may be:

$$H_0 : F(x) = F_0(x;\ )$$

where $\ = [\ _1;:::;\ _p]'$ is a p vector of *nuisance* parameters whose values are unknown and must be estimated from data.

1.2 Probability Plots

In addition to these analytical techniques, graphical methods are valuable supplements. The most important graphical technique is probability plotting. A *probability plot* is a plot of the cumulative distribution function $F(x)$ on the vertical axis versus x on the horizontal axis. The vertical axis is scaled such that, if the data fit the assumed distribution, the resulting plot will lie on a straight line. Special plotting paper may be purchased to do these plots; however, most modern scientific plotting programs have this capability (e.g., `gnuplot`). Each test presented below should be used in conjunction with a probability plot.

1.3 Shape of Distributions

Through much of the literature are references to Johnson curves: S_U or S_B (see §2, page 3). These refer to a system of distributions introduced by Johnson [21] where a standard normal random variable Z is translated to $(Z -\ )=\ $ and transformed using T:

$$Y = T\left(\frac{Z -\ }{\ }\right): \qquad (1)$$

Three families in Johnson's [21] system are:

1. a family of bounded distributions, denoted by S_B, where:

$$Y = T\left(\frac{e^x}{1 + e^x}\right); \qquad (2)$$

2. a family lognormal distributions where:

$$Y = T\left(e^x\right); \qquad (3)$$

3. and a family of unbounded distributions, denoted by S_U, where:

$$Y = \sinh(x) = T\ ($$

1.4 Miscellaneous

Many tests are presented here without mention of their relative merits. Users are advised to consult the cited literature to determine which test is appropriate for their situation. Sometimes a certain test will have more *power* than another; that is, a test may have a better ability to reject a model when the model is incorrect.

2 Moments: b_2 and $\sqrt{b_1}$

```
double*
omnibus_moments(x,n)
double *x;
int n;
```
Returns $\left[\sqrt{b_1}; b_2\right]'$.

Let $x_1; x_2; :::; x_n$ be the n observations with mean:

$$m_1 = \frac{1}{n} \sum_{j=1} n x_j: \tag{5}$$

The central moments are defined as:

$$m_i = \frac{1}{n} \sum_{j=1} n \left(x_j - m_i\right)^i; \; i = 2; 3; 4: \tag{6}$$

The sample skewness $\left(\sqrt{b_1}\right)$ and kurtosis (b_2) are defined as:

$$\sqrt{b_1} = m_3 {=} m_2^{3/2} = \sqrt{n} \left(\sum_{j=1}^{n} (x_i - \bar{x})^3 \right) = \left(\sum_{j=1}^{n} (x_i - \bar{x})^2 \right)^{3/2} \tag{7}$$

and

$$b_2 = m_4 {=} m_2^2: \tag{8}$$

These are invariant under both origin and scale changes [4]. When a distribution is specified, these are denoted as $\sqrt{\ _1}$ and $\ _2$.

For a standard normal, $\sqrt{\ _1} = 0$ and $\ _2 = 3$. To use either or both of these statistics to test for departure from normality, these are sometimes transformed to their standardized to their normal equivalent deviates, $X\left(\sqrt{b_1}\right)$ and $X\left(b_1\right)$.

For $X\left(\sqrt{b_1}\right)$, D'Agostino and Pearson [11] gave coefficients and ($n = 8$ to 1000) for:

$$X\left(\sqrt{b_1}\right) = \quad \sinh^{-1}\left(\sqrt{b_1} {=} \ \right) \tag{9}$$

that transforms $\sqrt{b_1}$ to a standard normal using a Johnson S_U approximation (Table 5). An equivalent approximation [10] that avoids the use of tables is given by:

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

2. Compute:

$$Y = \sqrt{b_1}\left[\frac{(n+1)(n+3)}{6(n-2)}\right]^{\frac{1}{2}}; \tag{10}$$

2

4. Compute the third standardized moment of b_2:

$$\sqrt{\ _1(b_2)} = \frac{6\left(n^2 - 5n + 2\right)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}: \tag{20}$$

5. Compute:

$$A = 6 + \frac{8}{\sqrt{\ _1(b_2)}}\left[\frac{2}{\sqrt{\ _1(b_2)}} + \sqrt{1 + \frac{4}{\sqrt{\ _1(b_2)}}}\right]: \tag{21}$$

6. Compute:

$$Z = \left(\left(1 - \frac{2}{9A}\right) - \left[\frac{1 - 2{=}A}{1 + y\sqrt{2{=}(A-4)}}\right]^{\frac{1}{3}}\right){=}\sqrt{2{=}(9A)} \tag{22}$$

where Z is a standard normal variable with zero mean and variance of one.

Example: For the sample data given in Table 4 ($n = 584$), $b_2 = 1{:}9148$. Suppose that we wish to test the hypothesis of normality:
H_0: $\ _2 = 3$ (normality)
versus the one-sided alternative
H_1: $\ _2 > 3$ (non-normality)
at a level of significance of 0.05. We would reject H_0 if Z (eqn. 22) is larger than 1.645 (Table 1). Following the procedure given above, $E(b_2) = 2{:}9897$, $Var(b_2) = 0{:}0401$, $y = -26{:}8366$, $\sqrt{\ _1(b_2)} = 0{:}0989$, $A = 2163$, and $Z = -131{:}7$. Therefore, we cannot reject H_0.

2.1 Omnibus Tests for Normality

3 Geary's Test of Normality

```
double*
geary_test(x,n)
double *x;
int n;
Returns [√a; y]'.
```

Let $x_1; x_2; \ldots; x_n$ be the n observations. The ratio of the mean deviation to the standard deviation is given as:

$$a = \frac{1}{n\sqrt{m_2}}\sum_{j=1}^{n}|x_i - \bar{x}| \tag{23}$$

where $\bar{x} = \sum_{i=1}^{n} x_i$ and m_2 is defined by eqn. 6. This ratio can be transformed a standard normal [10] via

$$y = \frac{\sqrt{n}\,(a - 0.7979)}{0.2123}. \tag{24}$$

This test is valid for $n \geq 41$.

More generally, Geary [17] considered tests of the form

$$a\,(c) = \frac{1}{nm_2^{c/2}} \sum_{j=1}^{n} |x_i - \bar{x}|^c \text{ for } c \geq 1 \tag{25}$$

where $a\,(1) = a$ of eqn. 23, and $a\,(4) = b_2$ of eqn. 8.

D'Agostino and Rosman [12] conclude that Geary's a test has good power for symmetric alternatives and skewed alternatives with ${}_2 < 3$ when compared to other tests, though for symmetric alternatives, b_2 (eqn. 8) can sometimes be more powerful and for skewed alternatives, W (eqn 70) or W' (eqn 73) usually dominate a. The Geary test (eqns. 23-24) is seldom used today—D'Agostino [10] include it in his summary work because it is of "historical interest."

Example: For the sample data given in Table 4 ($n = 584$), $a = 0.8823$. Suppose that we wish to test the hypothesis of normality:

H_0: normality

versus the two-sided alternative

H_1: non-normality

at a level of significance of 0.05. From eqn. 24, $y = 9.9607$.

---

4 Extreme Normal Deviates

```
double*
extremes(x,n)
double *x;
int n;
```
Returns $[\mathrm{x_n} - \bar{\mathrm{x}}, \mathrm{x_1} - \bar{\mathrm{x}}]'$.

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be the n observations. Given a known normal deviation , the largest and smallest deviation from a normal population may be computed:

$$u_n = \frac{x_n - \bar{x}}{} \tag{26}$$

and

$$u_1 = -\frac{x_1 - \bar{x}}{}, \tag{27}$$

respectively. These statistics are potentially useful for detecting outliers for populations with a known but an unknown mean. Table 25 in Pearson and Hartley [25] gives percentage points for this statistic. Pearson and Hartley [25] also give examples of the use of extreme deviates when an estimator of (independent of the sample) is known and when a combined "internal" and "external" estimate is used.

5 EDF Statistics for Testing Normality

[Note: This section follows closely the presentation of Stephens [35].]

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be the n observations. Suppose that the continuous distribution of x is $F(x)$. The empirical distribution function (EDF) is $F_n(x)$ defined by:

$$F_n(x) = \frac{1}{n} \text{(number of observations} \leq x\text{)} ; \ -\infty < x < \infty \tag{28}$$

or

$$\begin{array}{rcll} F_n(x) & = & 0; & x < x_1 \\ F_n(x) & = & \frac{1}{n}; & x_i \leq x < x_{i+1}; \quad i = 1; : : : ; n-1 \\ F_n(x) & = & 1; & x_n \leq x: \end{array}$$

Thus $F_n(x)$ is a step function calculated from the data. As $n \to \infty$, $|F_n(x) - F(x)|$ decreases to zero with probability one [35].

EDF statistics that measure the difference between $F_n(x)$ and $F(x)$ are divided into two classes: supremum and quadratic. On a graph of $F_n(x)$ and $F(x)$ versus x_i, denote the largest vertical distance when $F_n(x) > F(x)$ as D^+. Also, let D^- denote the largest vertical distance when $F_n(x) < F(x)$. These two measures are supremum statistics. Quadratic statistics are given by the Cramér–von Mises family

$$Q = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \quad (x)\, dF(x) \tag{29}$$

where (x) is a weighting function [35].

To compute these statistics, the Probability Integral Transformation is used: $z = F(x)$ where $F(x)$ is the Gaussian distribution. The new variable, z, is uniformly distributed between 0 and 1. Then z has distribution function $F^*(z) = z$, $0 \leq z \leq 1$. A sample $x_1; x_2; : : : ; x_n$ gives values $z_i = F(x_i)$, $i = 1; : : : ; n$, and $F_n^*(z)$ is the EDF of values z_i. For testing normality,

$$z_{(i)} = \Phi\left(\left(x_{(i)} - \hat{\ }\right) = \hat{\ }\right) \tag{30}$$

where $\hat{\ }$ and $\hat{\ }$ are estimated from the data and $\Phi(\cdot)$ denotes the cumulative probability of a standard normal. For testing if the data follows an exponential distribution Exp$(\ ;\ )$, where is known to be zero, $\hat{\ }$ is estimated by $\bar{x}$ (the sample mean) and

$$z_{(i)} = 1 - \exp\left(-x_{(i)} = \bar{x}\right) : \tag{31}$$

Now, EDF statistics can be computed by comparing $F_n^*(z)$ and a uniform distribution for z. These take the same values as comparisons between $F_n(x)$ and $F(x)$:

$$F_n(x) - F(x) = F_n^*(z) - F^*(z) = F_n^*(z) - z: \tag{32}$$

After ordering z-values, $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(n)}$ and computing $\bar{z} = \sum_{i=1}^{n} z_i = n$, the supremum statistics are

$$D^+ = \max_{i=1,\ldots,n} \left(i = n - z_{(i)}\right) \tag{33}$$

and

$$D^- = \max_{i=1,\ldots,n} \left(z_{(i)} - (i-1) = n\right) : \tag{34}$$

5.1 Kolmogorov D

```
double*
kolmogorov_smirnov(x,n)
double *x;
int n;
Returns [Dn; D]'.
```

```
double*
kolmogorov_smirnov_exp(x,n)
double *x;
int n;
Returns [De; D]'.
```

The most well-known EDF statistic is Kolmogorov's D, computed from supremum statistics:

$$D = \sup_x |F_n(x) - F(x)| = \max\left(D^+; D^-\right): \tag{35}$$

The modified form for testing a completely specified distribution [35]:

$$D^* = D\left(\sqrt{n} + 0{:}12 + 0{:}11 = \sqrt{n}\right): \tag{36}$$

For testing a normal distribution with and unknown, the modified equation is [35]:

$$D^n = D\left(\sqrt{n} - 0{:}01 + 0{:}85 = \sqrt{n}\right): \tag{37}$$

For testing an exponential distribution with and unknown, D does not need modified [35].

Example: For the sample data given in Table 4 ($n = 584$), $D^n = 4{:}0314$ and $y =$. Suppose that we wish to test the hypothesis of normality:

H_0: normality

versus the two-sided alternative

H_1: non-normality

at a level of significance of 0.05.

5.2 Kuiper's V

```
double*
kuipers_v(x,n)
double *x;
int n;
Returns [Vn; V]'.
```

```
double*
kuipers_v_exp(x,n)
double *x;
int n;
Returns [Ve; V]'.
```

Kuiper's [24] V is another statistic computed from supremum statistics:

$$V = D^{+} + D^{-}: \tag{38}$$

The modified form for testing a completely specified distribution [35]:

$$V^{*} = V\left(\sqrt{n} + 0{:}155 + 0{:}24\sqrt{n}\right): \tag{39}$$

For testing a normal distribution with and unknown, the modified equation is [35]:

$$V^{n} = V\left(\sqrt{n} + 0{:}05 + 0{:}82{=}\sqrt{n}\right): \tag{40}$$

For testing an exponential distribution with and unknown, V the modified equation is [35]:

$$V^{e} = \left(V - 0{:}2{=}\sqrt{n}\right)\left(\sqrt{n} + 0{:}24 + 0{:}35{=}\sqrt{n}\right): \tag{41}$$

5.3 Pyke's Statistics

For some purposes, eqns. 33 and 34 may be modified to [26]:

$$C^{+} = \max_{0 \le i \le n}\left(\frac{i}{n+1} - z_{(i)}\right);\ z_{(0)} = 0; \tag{42}$$

and

$$C^{-} = \max_{0 \le i \le n}\left(z_{(i)} - \frac{i}{n+1}\right) \tag{43}$$

(following the modification of notation by Durbin [16]). Then,

$$C = \max\left(C^{+}; C^{-}\right): \tag{44}$$

Durbin [16] notes that these modifications to eqns. 33 and 34 are related to the fact that $E\left(z_{(i)}\right) = i{=}(n+1)$. Percentage points were given by Durbin [15].

5.4 Brunk's B

As an alternative to Kuiper's V (eqn. 38), Brunk [6] suggests:

$$B = C^{+} + C^{-} \tag{45}$$

where C^{+} and C^{-} are given by eqns. 42 and 43.

5.5 Cramér–von Mises W^2

```
double*
cramer_von_mises(x,n)
double *x;
int n;
```
Returns $[W^{2,n}; W^2]'$.

```
double*
cramer_von_mises_exp(x,n)
double *x;
int n;
```
Returns $[W^{2,e}; W^2]'$.

Quadratic statistics are computed from the Cramér–von Mises family given in eqn 29. When $(x) = 1$ in eqn 29, the statistic is the Cramér–von Mises statistic W^2:

$$W^2 = \sum_{j=1}^{n} (Z_i - (2j - 1) = (2n))^2 + \frac{1}{12n} \qquad (46)$$

(When $(x) = (F(x)(1 - F(x)))^{-1}$, this yields the Anderson–Darling statistic given below in §5.7 [35].) The modified form for testing a completely specified distribution [35]:

$$W^{2,*} = \left(W^2 - 0{:}4{=}n + 0{:}6{=}n^2\right) = (1 + 1{=}n)\,: \qquad (47)$$

For testing a normal distribution with and unknown, the modified equation is [35]:

$$W^{2,n} = W^2\,(1{:}0 + 0{:}5{=}n)\,: \qquad (48)$$

For testing an exponential distribution with and unknown, the modified equation is [35]:

$$W^{2,e} = W^2\left(1{:}0 + 2{:}8{=}n - 3{=}n^2\right): \qquad (49)$$

5.6 Watson U^2

```
double*
watson_u2(x,n)
double *x;
int n;
```
Returns $[U^{2,n}; U^2]'$.

```
double*
watson_u2_exp(x,n)
double *x;
int n;
```
Returns $[U^{2,e}; U^2]'$.

$$U^2 = W^2 - n\left(\bar{Z} - 0{:}5\right)^2 \qquad (50)$$

where W^2 is the Cramér–von Mises statistic (§5.5). The modified form for testing a completely specified distribution [35]:

$$U^{2,*} = \left(U^2 - 0{:}1{=}n + 0{:}1{=}n^2\right) = (1 + 0{:}8{=}n)\,{:} \tag{51}$$

For testing a normal distribution with and unknown, the modified equation is [35]:

$$U^{2,n} = U^2\,(1{:}0 + 0{:}5{=}n)\,{:} \tag{52}$$

For testing an exponential distribution with and unknown, the modified equation is [35]:

$$U^{2,e} = U^2\left(1{:}0 + 2{:}3{=}n - 3{=}n^2\right){:} \tag{53}$$

5.7 Anderson–Darling A^2

```
double*
anderson_darling(x,n)
double *x;
int n;
Returns [A^{2,n}; A^2]'.
```

```
double*
anderson_darling_exp(x,n)
double *x;
int n;
Returns [A^{2,e}; A^2]'.
```

Anderson and Darling [1] present another EDF test statistic which is sensitive at the tails of the distribution (rather than near the median). When $(x) = (F(x)(1 - F(x)))^{-1}$ in eqn.(29), this yields the Anderson–Darling statistic [1, 35]:

$$A^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln z_j + \ln(1 - z_{n-j+1})\right]{:} \tag{54}$$

Equivalently [35],

$$A^2 = -n - \frac{1}{n}\sum_{j=1}^{n}\left[(2j-1)\ln z_j + (2n+1-2j)\ln(1 - z_j)\right]{:} \tag{55}$$

Anderson and Darling [1] give the following asymptotic significance values of A^2:

Significance Level	Significance Point
0.10	1.933
0.05	2.492
0.01	3.857

Anderson and Darling [1] state that sample size should be at least 40; however, Stephens [35] give the same asymptotic values (for more significance levels) for a sample size ≥ 5.

For testing a completely specified distribution, A^2 is used unmodified. For testing a normal distribution with and unknown, the modified equation is [35]:

$$A^{2,n} = A^2 \left(1.0 + 0.75/n + 2.25/n^2\right). \tag{56}$$

For testing an exponential distribution with and unknown, the modified equation is [35]:

$$A^{2,e} = A^2 \left(1.0 + 5.4/n - 11/n^2\right). \tag{57}$$

5.8 Durbin's Exact Test

```
double*
durbins_exact(x,n)
double *x;
int n;
```
Returns $\left[\mathsf{K_m}, \sqrt{\mathsf{n}}\mathsf{K_m}\right]'$.

Durbin [14] presented a modified Kolmogorov test. The discussion that follows has been adapted from Durbin's work [14].

Let $x_1, x_2, \ldots, x_n$ be the n i.i.d. observations and suppose that it is desired to test the hypothesis that they come from the continuous distribution $F(x)$. If the null hypothesis is true, then $u_j = F(x_j)$ $(j = 1, \ldots, n)$ are independent $U(0, 1)$ variables and are randomly scattered on the (0,1) interval. Clustering may indicated a departure from the null hypothesis. Denoting the ordered u's by $0 \le u_{(1)} \le \cdots \le u_{(n)} \le 1$, let $c_1 = u_{(1)}$, $c_2 = u_{(j)} - u_{(j-1)}$ $(j = 2, \ldots, n)$, and $c_{n+1} = 1 - u_{(n)}$.

Since the interest is in relative magnitudes of c's, these are ordered: $c_{(1)} \le c_{(2)} \cdots \le c_{(n)}$. Then, the following transformation is applied:

$$g_j = (n + 2 - j)\left(c_{(j)} - c_{(j-1)}\right) \left(c_{(0)} = 0;\ j = 1, \ldots, n + 1\right). \tag{58}$$

Durbin [14] shows that the g's, which depend on the *ordered* intervals, have the same distribution as the *unordered* c's.

Letting

$$w_r = \sum_{j=1}^{r} g_j \tag{59}$$

it follows that $w_1, \ldots, w_n$ have the same distribution as the ordered $U(0, 1)$ variables $u_{(1)}, \ldots, u_{(n)}$.

From eqns. 58 and 59, w_j can be expressed as:

$$w_j = c_{(1)} + \cdots + c_{(j-1)} + (n + 2 - j)\, c_{(j)}, \ (j = 1, \ldots, n), \tag{60}$$

where $c_{(1)} \le \cdots \le c_{(n)}$ is the ordered set of intervals.

In addition to two other test, Durbin [14] introduces the *modified Kolmogorov test*. The test statistic is:

$$K_m = \max_{r=1,\ldots,n} \left(\frac{r}{n} - w_r\right). \tag{61}$$

The test procedure is to reject when K_m is greater than the value tabulated for a one-sided Kolmogorov test.

Example: For the sample data given in Table 4 ($n = 584$), $K_m = 0{:}4127$. To test the hypothesis of normality:

H_0: normality

versus the one-sided alternative

H_1: non-normality

at a level of significance of 0.05, we would reject H_0 if K_m is larger than 0.895 (critical value of D for $= 0{:}05$. Therefore, we cannot reject H_0.

---

6 Chi-Square Test

double*
chi_square(x,n)
double *x;
int n;
Returns $\left[\mathrm{x}^2; \mathrm{k}-3\right]'$.

double*
chi_square_exp(x,n)
double *x;
int n;
Returns $\left[\mathrm{x}^2; \mathrm{k}-2\right]'$.

According to Shapiro [34], the chi-square goodness of fit test is the oldest procedure for testing distributional assumptions. It is useful for testing normality and exponentiality when the number of observations is large (because its power is poor for small samples when compared to other tests). It is also useful when data are discrete [34].

The basic idea is to divide the n data into k cells and compare the observed number in each cell with the expected number in each cell. The resulting statistic is distributed as a chi-square random variable with $k - 1 - t$ degrees of freedom, where t is the number of parameters estimated. The number of cells is taken as

$$k = (\text{int})4\left[0{:}75\,(n-1)^2\right]^{1/5}: \tag{62}$$

The ratio $n{=}k$ should be at least 5; otherwise another test should be used [34]. In this implementation, k is decremented by one until $n{=}k \geq 5$.

what should the notation be for rounding? For ceil, we use $\lceil x \rceil$. For floor, we use $\lfloor x \rfloor$.

Let $x_{(1)}; x_{(2)}; :::; x_{(k)}$ be the upper boundaries of cells. Choose $x_{(i)}$ so that the probability of being in any cell is the same:

$$P\left(x \leq x_{(i)}\right) = \frac{i}{k}; \; i = 1; 2; :::; k \tag{63}$$

In thse implmentation, only the case of raw data, as opposed to pre-tabulated data, is considered (i.e., equal probability cells).

For testing the normality hypothesis, let $x_{(0)} = -\infty$ and $x_{(k)} = \infty$. The values of $x_{(i)}$ are:

$$x_{(i)} = \bar{x} + s\,Z_{i/k} \tag{64}$$

where $\bar{x}$ and s are estimated mean and variance parameters and $Z_{i/k}$ are percentiles of the standard normal distribution. The test statistic is

$$x^2 = \frac{k}{n}\sum_{i=1}^{k} f_i^2 - n \tag{65}$$

where f_i is the number of observations in cell i. The hypothesis of normality is rejected at an level if x^2 is greater x^2_α, a ^{2} random variable with $k-3$ degrees of freedom.

Example: For the sample data given in Table 4 ($n = 584$), $x^2 = 952.7$ with $= 45$ degrees of freedom. Since $^2_{45,0.05} \approx 30.33$ (Table 2), we reject H_0 at an $= 0.05$ level.

For testing the exponentiality hypothesis, let $x_{(0)} = 0$ and $x_{(k)} = \infty$. The values of $x_{(i)}$ are:

$$x_{(i)} = -\frac{1}{\ } \ln\left(1 - \frac{i}{k}\right); i = 1; 2; \ldots; k-1. \tag{66}$$

The parameter is estimated from

$$\hat{\ } = n\left(\sum_{i=1}^{n} x_i\right)^{-1} \tag{67}$$

where x_i is the ith observation in the sample. Equation (65) is the statistic used for testing exponentiality. The hypothesis of exponentiality is rejected at an level if x^2, a ^{2} random variable with $k-2$ degrees of freedom.

Example: For the sample data given in Table 4 ($n = 584$), $x^2 = 308.11$ with $= 46$ degrees of freedom. Since $^2_{46,0.05} \approx 31.16$ (Table 2), we reject H_0 : exponentiality, at an $= 0.05$ level.

---

7 Analysis of Variance Tests

7.1 Shapiro-Wilk W

```
double*
shapiro_wilk(x,n)
double *x;
int n;
Returns [W; S^2]'.
```

```
double*
shapiro_wilk_exp(x,n)
double *x;
int n;
Returns [W; S^2]'.
```

Recall the description of a probability plot given on page 2. Ordered observations are plotted against expected values of order statistics from the distribution being tested. The plot tends to be linear if the distributional assumption is correct. If a genearlized least squares is performed, an F-type ratio could be used to test the fit of a linear model. This was the basis of test introduced by Shapiro and Wilk [32]. Foregoing many of the details in the derivation, the test procedures for normality and exponentiality are given below.

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be the n ordered observations and let

$$S^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2 : \tag{68}$$

Calculate

$$b = \sum_{i=1}^{k} a_{n-i+1} \left(x_{n-i+1} - x_i \right) \tag{69}$$

where $k = n{=}2$ if n is even, $k = (n-1){=}2$ if n is odd, and a_{n-i+1} are found in Table 6. Then a test of normality for small samples ($3 \leq n \leq 50$) is defined as

$$W = \frac{b^2}{S^2} \tag{70}$$

Small values of W indicate non-normality ("lower-tail"). Hence if the computed value of W is less than the W_α shown in Table 7, the hypothesis of normality is rejected.

Example: Using the first 40 observations from the sample data given in Table 4, $W = 0{:}0000245$. Using $\;= 0{:}05$ and Table 7, $W_{0.05} = 0{:}940$. Since $W < W_{0.05}$, we reject H_0.

For testing exponentiality, no tabulated constants are needed for calculation of b:

$$b = \sqrt{\frac{n}{n-1}} \left(\bar{x} - x_1 \right) \tag{71}$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i : \tag{72}$$

This assumes that the origin parameter is unknown. It also differs from the test of normality in that it is a two-tailed procedure. That is, too small or too large a value of the test statistic indicates non-exponentiality [34].

Example: Using the first 40 observations from the sample data given in Table 4, $W = 0.0909$. Using $= 0.05$ and Table 8, $W_{0.025} = 0.0148$ and $W_{0.975} = 0.0447$. Since W is not contained in the interval $[W_{0.025}, W_{0.975}]$, we reject H_0: exponentiality.

7.2 Modified Shapiro–Francia W'

```
double*
shapiro_francia(x,n)
double *x;
int n;
Returns [W'; S^2]'.
```

The W test of normality in the previous section for sample sizes of 50 or less. Shapiro and Francia [31] extended the W test for n up to 99 by replacing the values a_{n-i+1} in Table 6 b_{n-i+1} in Table 9. The test procedure follows.

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be the n ordered observations. Then a test of normality for large samples is defined as:

$$W' = \frac{b'}{S^2} \tag{73}$$

The numerator b' is defined as:

$$b' = \sum_{i=1}^{k} b_{n-i+1} \left(x_{n-i+1} - x_i \right) \tag{74}$$

where $k = n/2$ if n is even and $k = (n-1)/2$ is n is odd. Significant values, determined empirically by Shapiro and Francia [31] are given in Table 10. D'Agostino [10] notes that the values given by Shapiro and Francia [31] in the lower tail were "higher than what they should be" since too few samples were used in determining these significance levels.

Example: Using the first 99 observations from the sample data given in Table 4, $W' = 1.0139$. Using $= 0.05$ and Table 10, $W'_{0.05} = 0.976$. Since $W' > W'_{0.05}$, we cannot reject H_0.

7.3 Weisberg-Bingham $\tilde{W}'$

```
double*
weisberg_bingham(x,n)
double *x;
int n;
Returns [W̃'; S^2]'.
```

An alternative way of computing b' is to note that the vector $[b_1, b_2, \ldots, b_n]'$ is equivalent to $m'/(m'm)^{1/2}$ where $m' = (m_1, m_2, \ldots, m_n)$ denotes a vector of expected normal order statistics. One approximation for normal order statistics attributed to Blom [3] is:

$$E\,(r, n) = -\Phi^{-1}\left(\frac{r - \quad}{n - 2\quad + 1}\right) \tag{75}$$

with a recommended "compromise value $\quad = 0.375$ [28]." Define this new statistic as $\tilde{W}'$. So, instead of hardcoding constants (as done in §7.1-7.2), this approximation is used. Since $\tilde{W}'$ is essentially the same as W', the table of critical values for W' (Table 10) may be used.

7.4 D'Agostino's D Test of Normality

```
double*
dagostino_d(x,n)
double *x;
int n;
Returns [D, y]'.
```

D'Agostino [10] presents a modified Shapiro-Wilk W test that eliminates the need for a table of weights. The test statistic is given as

$$\begin{aligned} D &= T/\left(n^2\sqrt{m_2}\right) && (76)\\ &= T/\left(n^{3/2}\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2}\right) \end{aligned}$$

where

$$T = \sum_{i=1}^{n}\left(i - \frac{1}{2}(n+1)\right)x_i. \tag{77}$$

An approximate standard variable is

$$y = \frac{\sqrt{n}\,(D - 0.28209479)}{0.02998598}. \tag{78}$$

Significant values are given in Table 3.

Example: For the sample data given in Table 4 ($n = 584$), $D = 0.2859$ and $y = 3.0667$. Suppose that we wish to test the hypothesis of normality:

H_0: normality

versus the two-sided alternative

H_1: non-normality

at a level of significance of 0.005. From Table 3 (linearly interpolating), we reject H_0 if $y < -3.006$ or $y > 2.148$. Therefore, we cannot reject H_0.

7.5 Royston's Modification

```
double*
royston(x,n)
double *x;
int n;
Returns [W; P]'.
```

Royston [30] also presented a modified W statistic for n up to 2000 that did not require extensive use of tabulated constants. If $m' = (m_1; m_2; : : : ; m_n)$ denotes a vector of expected values of standard normal order statistics and $V = (v_{ij})$ denote the corresponding $n \times n$ covariance matrix, then W may be written as:

$$W = \left[\sum_{i=1}^{n} a_i x_{(i)}\right]^2 = \sum_{i=1}^{n} \left(x_{(i)} - \bar{x}\right)^2 \tag{79}$$

where

$$a' = m'V^{-1}\left[\left(m'V^{-1}\right)\left(V^{-1}m'\right)\right]^{1/2}: \tag{80}$$

Let $a^* = m'V^{-1}$; The following approximation for a^* is used:

$$\hat{a}^* = \begin{cases} 2m_i; & \text{i=2,3,...,n-1} \\ \left(\frac{\hat{a}_1^2}{1-2\hat{a}_1^2}\sum_{i=2}^{n-1}\hat{a}_i^{*2}\right)^{1/2}; & \text{i=1, i=n} \end{cases} \tag{81}$$

where

$$\hat{a}_1^2 = \hat{a}_n^2 = \begin{cases} g(n-1); n \le 20 \\ g(n); n > 20 \end{cases} \tag{82}$$

and

$$g(n) = \frac{\Gamma\left(\frac{1}{2}[n+1]\right)}{\sqrt{2\Gamma\left(\frac{1}{2}n+1\right)}}: \tag{83}$$

The function $g(n)$ is approximated using:

$$g(n) = \left[\frac{6n+7}{6n+13}\right]\left(\frac{\exp(1)}{n+2}\left[\frac{n+1}{n_2}\right]^{n-2}\right)^{1/2} \tag{84}$$

Royston [30] used eqns. 81–84 for the range $7 \le n \le 2000$, but exact values of a_i for $n < 7$.

Royston [30] used the following normalizing transformation:

$$y = (1-W)^\lambda \tag{85}$$

so that

$$z = \left[(1-W)^\lambda - {}_y\right] = {}_y \tag{86}$$

can be compared with the upper tail of a standard normal. Large values of z indicate non-normality of the original sample.

This implementation in `cdhc` closely follows Royston's published FORTRAN code [29, 28]. It returns W and a corresponding P value (smallest level at which we could have preset   and still have been able to reject H_0). It also utilizes algorithms by Hill [19] and Wichura [36].

8 Kotz Separate Families T_f'

```
double*
kotz_families(x,n)
double *x;
int n;
Returns [T'_f; T_f]'.
```

Kotz [23] developed a test where the null hypothesis H_0 is that the sample $x_1; x_2; :::; x_n$ came from a lognormal distribution, and the alternate hypothesis is that the parent population was normal. The test statistic, given as:

$$T_f' = \frac{\log \frac{\hat{\beta}_2}{\beta_{2,\hat{\alpha}}}}{2\sqrt{n}\{\frac{1}{4}\,(e^{4\hat{\alpha}_2} + 2e^{3\hat{\alpha}_2} - 4) - \hat{\ }_2 - \frac{\hat{\alpha}_2\left(2e^{\hat{\alpha}_2}-1\right)^2}{2\left(2e^{\hat{\alpha}_2}-1\right)^2} + \frac{3}{4}e^{\hat{\alpha}_2}\}^{1/2}} \tag{87}$$

is asymptotically normal [8].

Example: For the sample data given in Table 4 ($n = 584$), $T_f' = -0{:}6021$. Suppose that we wish to test the hypothesis

H_0 : lognormal

versus

H_1 : normal

at a level of significance of 0.05. We would reject H_0 if T_f' is larger than 1.645. Therefore, we reject H_0.

The discussion that follows explains in more detail how this statistic is calculated and how it was derived. The remainder of this section was taken directly from the work of Kotz [23] (pages 123,124–126).

... A test for this special situation was considered by Roy [27], where he bases his decision on the statistic

$$R = \frac{L_l}{L_n} \tag{88}$$

where L_l denotes the likelihood of the sample under the lognormal hypothesis and L_n that under the normal hypothesis. If $R > 1$ one accepts lognormality, and if $R < 1$ normality is accepted. More recently Cox [7, 8] has elaborated on Roy's heuristic approach, and has derived a general class of tests to discriminate between hypotheses that are *separate* (in the sense that an arbitrary simple hypothesis in H_0 cannot be obtained as a limit—in the parameter space—of a simple hypothesis in H_1. We will now apply Cox's general theory to testing lognormality against normality...

Suppose $x_1; x_2; :::; x_n$ is a random sample from a certain population. The null hypothesis, H_f, is that the p.d.f. of the x's is log-normal and the alternate hypothesis, H_g, is that the p.d.f. is normal, that is, for H_f

$$f\,(y;\ \ ) = \frac{1}{\sqrt{2}} \exp - \left(\frac{(\log y - \ \ )^2}{2} \right); -\infty < y < \infty: \tag{89}$$

and for H_g:

$$g(y;\beta) = \frac{1}{y\sqrt{2\pi\beta_2}} \exp - \left(\frac{(y - \beta_1)^2}{2\beta_2} \right);\ y > 0. \tag{90}$$

From the maximum likelihood equations we find that

$$\hat{\alpha}_1 = \frac{1}{n} \sum \log x_i;\ \hat{\alpha}_2 = \frac{1}{n} \sum (\log x_i - \hat{\alpha}_1);\tag{91}$$

and analogous equations for $\hat{\beta}_1$ and $\hat{\beta}_2$.

Under H_f, the log-normal null hypothesis, as the sample size n increases to infinity, $\hat{\alpha}_1 \to \alpha_1$, $\hat{\alpha}_2 \to \alpha_2$, $\hat{\beta}_{1,\alpha} \to \beta_{1,\alpha}$, and $\hat{\beta}_{2,\alpha} \to \beta_{2,\alpha}$ where

$$\hat{\beta}_{1,\alpha} = \exp\left(\alpha_1 + \frac{\alpha_2}{2}\right) \tag{92}$$

and

$$\hat{\beta}_{2,\alpha} = \exp(2\alpha_1 + \alpha_2)\left[\exp(\alpha_2) - 1\right]. \tag{93}$$

Cox's test is based on the log likelihood ratio

$$L_{fg} = \sum_{i=1}^{n} \log \frac{f(x_i;\hat{\alpha})}{g\left(x_i;\hat{\beta}\right)} \tag{94}$$

and his test statistic is given by

$$T_f = L_{fg} - E_{\hat{\alpha}}(L_{fg}) \tag{95}$$

where $E_{\hat{\alpha}}(L_{fg})$ is the expected value under H_f when α takes the value $\hat{\alpha}$. Writing

$$F = \log f(x;\alpha);\ F_{\alpha_i} = \frac{\partial \log f(x;\alpha)}{\partial \alpha_i};\ i = 1, 2 \tag{96}$$

$$F_{\alpha_i\alpha_j} = \frac{\partial^2 \log f(x;\alpha)}{\partial \alpha_i \partial \alpha_j};\ G = \log g(x;\beta) \tag{97}$$

$$G_{\beta_i} = \frac{\partial \log g(x;\beta)}{\partial \beta_i};\ \text{etc.,} \tag{98}$$

Cox shows that T_f is asymptotically normal with zero mean and variance

$$V_\alpha(T_f) = nV_\alpha(F - G) - \sum \frac{C_\alpha^2(F - G; F_{\alpha_i})}{V_\alpha(F_{\alpha_i})} \tag{99}$$

where $V_\alpha(\cdot)$, $C_\alpha(\cdot)$, denote variance and covariance under H_f.

In our case it can be shown that

$$T_f = \frac{n}{2} \log \frac{\hat{\beta}_2}{\hat{\beta}_{2,\hat{\alpha}}} \tag{100}$$

Results of the following type are used in the derivation of $V_\alpha(T_f)$:

$$E_\alpha\left[x^2 \log x\right] = (\alpha_1 + 2\alpha_2)\exp(2\alpha_1 + 2\alpha_2) \tag{101}$$

$$E_\alpha\left[x^2\log^2 x\right] = \left({}_2 + {}^2_1 + 4 {}_1 {}_2 + 4 {}^2_2\right)\exp\left(2 {}_1 + 2 {}_2\right) \tag{102}$$

$$E_\alpha\left[(\log x)(\log x -)\right] = {}_2 \tag{103}$$

$$E_\alpha\left[(\log x)(\log x - {}_1)^2\right] = {}_2 ({}_1 + 2 {}_2) \tag{104}$$

$$E_\alpha\left[(\log x - {}_1)(\log x - {}_1)^2\right] = 2 {}_2 {}_2. \tag{105}$$

Using these results, after a considerable amount of simplification, we get

$$V_\alpha(T_f) = n\left[\frac{1}{4}\left(e^{4\alpha_2} + 2e^{3\alpha_2} + 3e^{\alpha_2} - 4\right) {}_2 - \frac{{}_2\left(2e^{\alpha_2} - 1\right)^2}{2\left(2e^{\alpha_2} - 1\right)^2}\right] \tag{106}$$

Cox [8] has shown that

$$T_f' = \frac{T_f}{\sqrt{V_\alpha(T_f)}} \tag{107}$$

is asymptotically standardized normal. In our case we get, after substituting the estimators for the parameters,

$$T_f' = \frac{\log\frac{\hat\beta_2}{\beta_{2,\hat\alpha}}}{2\sqrt{n}\{\frac{1}{4}\left(e^{4\hat\alpha_2} + 2e^{3\hat\alpha_2} - 4\right) - \hat{}_2 - \frac{\hat\alpha_2\left(2e^{\hat\alpha_2}-1\right)^2}{2\left(2e^{\hat\alpha_2}-1\right)^2} + \frac{3}{4}e^{\hat\alpha_2}\}^{1/2}} \tag{108}$$

9 Utility Functions

This section describes some useful functions included in `cdhc` but not necessarily described in the previous sections, e.g., normal order statistics, normal probabilities, inverse normals.

References

[1] T. W. Anderson and D. A. Darling. A test of goodness of fit. *J. Am. Statist. Ass.*, 49(268):765–769, 1954.

[2] F. J. Anscombe and J. W. Tukey. The examination and analysis of residuals. *Technometrics*, 5:141–160, 1963.

[3] G. Blom. *Statistical Estimates and Transformed Beta-variates.* John Wiley & Sons, New York, 1958.

[4] K. O. Bowman and L. R. Shenton. *Moment* ($\sqrt{b_1}, b_2$) *Techniques*, chapter 7, pages 279–330. Volume 68 of D'Agostino and Stephens [13], 1986.

[5] Barry W. Brown and James Lovato. *CDFLIB: Library of FORTRAN Routines for Cumulative Distribution Functions, Inverses, and Other Parameters.* Department of Biomathematics, The University of Texas, M. D. Anderson Cancer Center, Box 237, 1515 Holcombe Boulevard, Houston, Texas 77030, June 1993. Public domain software from STATLIB.

[6] H. D. Brunk. On the range of the difference between hypothetical distribution function and Pyke's modified empirical distribution function. *Ann. Math. Statist.*, 33:525–532, 1962.

[7] D. R. Cox. Tests of separate families of hypotheses. In *Proc. 4th Berkeley Symp.*, volume 1, pages 105–123, 1961.

[8] D. R. Cox. Further results on tests of separate families of hypotheses. *J. R. Statist. Soc. B.*, 24:406–424, 1962.

[9] Noel A. C. Cressie. *Statistics for Spatial Data.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, NY, 1991.

[10] Ralph B. D'Agostino. *Tests for the Normal Distribution*, chapter 9, pages 367–420. Volume 68 of D'Agostino and Stephens [13], 1986.

[11] Ralph B. D'Agostino and Egon Sharpe Pearson. Tests for departure from normality. empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60(3):613–622, 1973.

[12] Ralph B. D'Agostino and Bernard Rosman. The power of Geary's test of normality. *Biometrika*, 61(1):181–184, 1974.

[13] Ralph B. D'Agostino and Michael A. Stephens. *Goodness-of-Fit Techniques*, volume 68 of *Statistics: textbooks and monographs.* Marcel Dekker, Inc, New York, 1986.

[14] J. Durbin. Some methods of constructing exact tests. *Biometrika*, 48(1&2):41–55, 1961.

[15] J. Durbin. Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika*, 59:1–15, 1969.

[16] J. Durbin. *Distribution Theory for Tests Based on the Sample Distribution Function.* Number 9 in CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, PA, 1973.

[17] R. C. Geary. The frequency distribution of $\sqrt{b_1}$ for samples of all sizes drawn at random from a normal population. *Biometrika*, 34(1):68–97, 1947.

[18] J. F. Hart. *Computer Approximations.* John Wiley & Sons, New York, 1968.

[19] I. D. Hill. Algorithm as 66: The normal integral. *Appl. Statist.*, 22:424–427, 1973.

[20] David Caster Hoaglin. *Summarizing Shape Numerically: The g-and-h Distributions*, chapter 11, pages 461–512. John Wiley & Sons, New York, 1985.

[21] N. L. Johnson. Systems of frequency curves generate by methods of translation. *Biometrika*, 36(?):149–176, 1949.

[22] Paul Johnson. *CDH: Tests of Composite Distributional Hypotheses for the Analysis of Biological & Environmental Data.* University of California at Davis, 1420 Lake Blvd #29, Davis, California 95616, January 1994. Public domain software from STATLIB.

[23] Samuel Kotz. Normality vs. lognormality with applications. *Communications in Statistics*, 1(2):113–132, 1973.

[24] N. H. Kuiper. Tests concerning random points on a circle. In *Proc. Koninkl. Neder. Akad. van. Wetenschappen*, volume 63, pages 38–47, 1960.

[25] Egon Sharpe Pearson and H. O. Hartley, editors. *Biometrika Tables for Statisticians*. Biometrika Trust, London, 3rd (with additions & corrections) edition, 1976.

[26] R. Pyke. The supremum and infimum of the poisson process. *Ann. Math. Statist.*, 30:568–576, 1959.

[27] A. D. Roy. A further statistical note on the distribution of individual output. *Economic J.*, 60:831–836, 1950.

[28] J. P. Royston. Algorithm as 177: Expected normal order statistics (exact and approximate). *Appl. Statist.*, 31(2):161–165, 1982.

[29] J. P. Royston. Algorithm as 181: The W test for normality. *Appl. Statist.*, 31(2):176–180, 1982.

[30] J. P. Royston. An extension of shapiro and wilk's W test for normality to large samples. *Appl. Statist.*, 31(2):115–124, 1982.

[31] S. S. Shapiro and R. S. Francia. An approximate analysis of variance test for normality. *J. Am. Statist. Ass.*, 67(337):215–216, 1972.

[32] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3&4):591–611, 1965.

[33] S. S. Shapiro and M. B. Wilk. An analysis of variance test for the exponential distribution (complete samples). *Technometrics*, 14(2):355–370, 1972.

[34] Samuel S. Shapiro. *How to Test Normality and Other Distributional Assumptions*, volume 3 of *The ASQC Basic References in Quality Control: Statistical Techniques*. Am. Soc. Qual. Control: Statistics Div., 310 W. Wisconsin Ave., Milwaukee, Wisconsin 53203, revised edition, 1990.

[35] Michael A. Stephens. *Tests Based on EDF Statistics*, chapter 4, pages 97–191. Volume 68 of D'Agostino and Stephens [13], 1986.

[36] M. J. Wichura. Algorithm as 241: The percentage points of the normal distribution. *Appl. Statist.*, 37(3):477–484, 1988.

Table 1: Cumulative Standard Normal Distribution.
Area Under the Normal Curve from

$$-\infty \text{ to } z = \frac{X_i -}{}:$$

Computed by the author using algorithm 5666 for the error function, from Hart *et al.* [18].

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

Table 2: Cumulative Chi-Square Distribution.
Computed by the author using CDFLIB [5], with the exception of items marked with a dagger (†), which were found in *Biometrika Tables for Statisticians* (1966), 3rd. Ed., University College, London, as cited by Shapiro [34].

	α									
ν	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.0000393†	0.000157†	0.000982†	0.0158†	0.102†	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.8
6	0.676	0.872	1.24	1.64	2.20	10.6	12.6	14.5	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.6	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.9	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.9	32.0	35.1	38.0	41.6	44.2
24	9.89	10.9	12.4	13.9	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7

According to Shapiro [34], for situations with larger than 30 degrees of freedom, $^{2}_{\nu,\alpha} = 0.5\left(z_\alpha + \sqrt{2\quad - 1}\right)^2$, where z_α is the 100 % point of the standard normal distribution, e.g., $z_{0.05} = -1.645$ from Table 1.

Table 3: Signficant Values of D'Agostino's D Test (y statistic of eqn. 78). Reproduced from D'Agostino [10].

	Percentiles									
n	0.5	1.0	2.5	5	10	90	95	97.5	99	99.5
10	-4.66	-4.06	-3.25	-2.62	-1.99	0.149	0.235	0.299	0.356	0.385
12	-4.63	-4.02	-3.20	-2.58	-1.94	0.237	0.329	0.381	0.440	0.479
14	-4.57	-3.97	-3.16	-2.53	-1.90	0.308	0.399	0.460	0.515	0.555
16	-4.52	-3.92	-3.12	-2.50	-1.87	0.367	0.459	0.526	0.587	0.613
18	-4.47	-3.87	-3.08	-2.47	-1.85	0.417	0.515	0.574	0.636	0.667
20	-4.41	-3.83	-3.04	-2.44	-1.82	0.460	0.565	0.628	0.690	0.720
22	-4.36	-3.78	-3.01	-2.41	-1.81	0.497	0.609	0.677	0.744	0.775
24	-4.32	-3.75	-2.98	-2.39	-1.79	0.530	0.648	0.720	0.783	0.822
26	-4.27	-3.71	-2.96	-2.37	-1.77	0.559	0.682	0.760	0.827	0.867
28	-4.23	-3.68	-2.93	-2.35	-1.76	0.586	0.714	0.797	0.868	0.910
30	-4.19	-3.64	-.291	-2.33	-1.75	0.610	0.743	0.830	0.906	0.941
32	-4.16	-3.61	-2.88	-2.32	-1.73	0.631	0.770	0.862	0.942	0.983
34	-4.12	-3.59	-2.86	-2.30	-1.72	0.651	0.794	0.891	0.975	1.02
36	-4.09	-3.56	-2.85	-2.29	-1.71	0.669	0.816	0.917	1.00	1.05
38	-4.06	-3.54	-2.83	-2.28	-1.70	0.686	0.837	0.941	1.03	1.08
40	-4.03	-3.51	-2.81	-2.26	-1.70	0.702	0.857	0.964	1.06	1.11
42	-4.00	-3.49	-2.80	-2.25	-1.69	0.716	0.875	0.986	1.09	1.14
44	-3.98	-3.47	-2.78	-2.24	-1.68	0.730	0.892	1.01	1.11	1.17
46	-3.95	-3.45	-2.77	-2.23	-1.67	0.742	0.908	1.02	1.13	1.19
48	-3.93	-3.43	-2.75	-2.22	-1.67	0.754	0.923	1.04	1.15	1.22
50	-3.91	-3.41	-2.74	-2.21	-1.66	0.765	0.937	1.06	1.18	1.24
60	-3.81	-3.34	-2.68	-2.17	-1.64	0.812	0.997	1.13	1.26	1.34
70	-3.73	-3.27	-2.64	-2.14	-1.61	0.849	1.05	1.19	1.33	1.42
80	-3.67	-3.22	-2.60	-2.11	-1.59	0.878	1.08	1.24	1.39	1.48
90	-3.61	-3.17	-2.57	-2.09	-1.58	0.902	1.12	1.28	1.44	1.54
100	-3.57	-3.14	-2.54	-2.07	-1.57	0.923	1.14	1.31	1.48	1.59
150	-3.409	-3.009	-2.452	-2.004	-1.520	0.990	1.233	1.423	1.623	1.746
200	-3.302	-2.922	-2.391	-1.960	-1.491	1.032	1.290	1.496	1.715	1.853
250	-3.227	-2.861	-2.348	-1.926	-1.471	1.060	1.328	1.545	1.779	1.927
300	-3.172	-2.816	-2.316	01.906	-1.456	1.080	1.357	1.528	1.826	1.983
350	-3.129	-2.781	-2.291	-1.888	-1.444	1.096	1.379	1.610	1.863	2.026
400	-3.094	-2.753	-2.270	-1.873	-1.434	1.108	1.396	1.633	1.893	2.061
450	-3.064	-2.729	-2.253	-1.861	-1.426	1.119	1.411	1.652	1.918	2.090
500	-3.040	-2.709	-2.239	-1.850	-1.419	1.127	1.423	1.668	1.938	2.114
550	-3.019	-2.691	-2.226	-1.841	-1.413	1.135	1.434	1.682	1.957	2.136
600	-3.000	-2.676	-2.215	-1.833	-1.408	1.141	1.443	1.694	1.972	2.154
650	-2.984	-2.663	-2.206	-1.826	-1.403	1.147	1.451	1.704	1.986	2.171
700	-2.969	-2.651	-2.197	-1.820	-1.399	1.152	1.458	1.714	1.999	2.185
750	-2.956	-2.640	-2.189	-1.814	-1.395	1.157	1.465	1.722	2.010	2.199
800	-2.944	-2.630	-2.182	-1.809	-1.392	1.161	1.471	1.730	2.020	2.221
850	-2.933	-2.621	-2.176	-1.804	-1.389	1.165	1.476	1.737	2.029	2.221
900	-2.923	-2.613	-2.710	-1.800	-1.386	1.168	1.481	1.743	2.037	2.231
950	-2.914	-2.605	-2.164	-1.796	-1.383	1.171	1.485	1.749	2.045	2.241
1000	-2.906	-2.599	-2.159	-1.792	-1.381	1.174	1.489	1.754	2.052	2.249
1500	-2.845	-2.549	-2.123	-1.765	-1.363	1.194	1.519	1.793	2.103	2.309
2000	-2.807	-2.515	-2.101	-1.750	-1.353	1.207	1.536	1.815	2.132	2.342

Table 4: Sample Data. Diameters at Breast Height (cm) of 584 Longleaf Pine Trees.
Locations and Diameters at Breast Height (dbh, in centimeters) of all 584 Longleaf Pine Trees in the 4 hectare Study Region. The x coordinates are distances (in meters) from the tree to the southern boundary. The y coordinates are distances (in meters) from the tree to the eastern boundary. Reproduced from Table 8.1 of Cressie [9].

x	y	dbh	x	y	dbh	x	y	dbh	x	y	dbh
200.0	8.8	32.9	199.3	10.0	53.5	193.6	22.4	68.0	167.7	35.6	17.7
183.9	45.4	36.9	182.5	47.2	51.6	166.1	48.8	66.4	160.7	42.4	17.7
162.9	29.0	21.9	166.4	33.6	25.7	163.0	35.8	25.5	156.1	38.7	28.3
157.6	42.8	11.2	154.4	36.2	33.8	150.8	45.8	2.5	144.6	25.4	4.2
142.7	25.4	2.5	144.0	28.3	31.2	143.5	36.9	16.4	123.1	14.3	53.2
113.9	13.1	67.3	114.9	8.1	37.8	101.4	9.3	49.9	105.7	9.1	46.3
106.9	14.7	40.5	127.0	29.7	57.7	129.8	45.8	58.0	136.3	44.2	54.9
106.7	49.4	25.3	103.4	49.6	18.4	89.7	4.9	72.0	10.8	0.0	31.4
26.4	5.4	55.1	11.0	5.5	36.0	5.1	3.9	28.4	10.1	8.5	24.8
18.9	11.3	44.1	28.4	11.0	50.9	41.1	9.2	47.5	41.2	12.6	58.0
33.9	21.4	36.9	40.8	39.8	65.6	49.7	18.2	52.9	6.7	46.9	39.5
11.6	46.9	42.7	17.2	47.9	44.4	19.4	50.0	40.3	26.9	47.2	53.5
39.6	47.9	44.2	38.0	50.7	53.8	19.1	45.2	38.0	32.1	35.0	48.3
28.4	35.5	42.9	3.8	44.8	40.6	8.5	43.4	34.5	11.2	40.2	45.7
22.4	34.3	51.8	23.8	33.3	52.0	24.9	29.8	44.5	9.0	38.9	35.6
10.4	61.2	19.2	30.9	52.2	43.5	48.9	67.8	33.7	49.5	73.8	43.3
46.3	80.9	36.6	44.1	78.0	46.3	48.5	94.8	48.3	45.9	90.4	20.4
44.2	84.0	40.5	37.0	64.3	44.0	36.3	67.7	40.9	36.7	71.5	51.0
35.3	78.3	36.5	33.5	81.6	42.1	29.3	83.8	15.6	22.4	84.1	18.5
17.1	84.7	43.0	27.3	89.4	28.9	27.9	90.6	21.3	48.4	99.5	30.9
43.6	98.4	42.7	39.0	97.3	37.6	14.9	91.2	47.1	6.1	96.2	44.6
10.7	98.6	44.3	22.2	100.0	26.1						
						32.7	99.1	25.9	0.9	100.0	41.4
93.5	96.2	59.5	85.1	90.6	26.1	92.8	61.5	11.4	91.3	69.5	33.4
95.9	59.7	35.8	93.4	71.5	54.4	89.6	86.3	33.6	99.5	78.9	35.5
100.6	53.1	7.4	103.5	72.1	36.6	104.7	74.0	19.1	104.0	67.1	34.9
104.2	64.7	37.3	105.0	59.8	16.3	111.8	73.2	39.1	112.4	69.8	36.5
110.0	65.9	25.0	120.4	79.2	46.8	109.4	62.5	18.7	109.7	62.9	23.2
113.3	60.4	20.4	118.0	69.3	42.3	126.5	69.2	38.1	125.1	68.2	17.9
114.2	54.6	39.7	110.6	51.5	14.5	147.3	73.8	33.5	146.7	73.0	56.0
148.1	86.2	66.1	138.2	73.4	26.3	135.7	70.7	44.8	134.9	72.7	24.2
98.0	27.7	39.0	93.5	28.7	15.1	82.3	16.8	35.6	79.2	25.3	21.6
84.2	29.0	17.2	88.8	35.1	22.3	82.5	36.3	18.2	75.6	28.1	55.6
72.9	36.2	23.2	79.1	43.6	27.0	50.0	48.8	50.1	59.9	34.4	45.5
60.5	13.0	47.2	60.2	11.4	37.8	66.5	15.9	31.9	70.4	6.6	38.5
70.7	2.2	23.8	71.7	1.9	46.3	179.5	92.6	2.8	186.1	91.0	3.2
178.3	92.4	5.8	178.6	91.8	3.5	186.2	90.3	2.3	185.2	89.9	3.8
185.5	89.8	3.2	185.8	89.1	4.4	186.5	88.8	3.9	176.7	92.3	7.8
177.7	91.5	4.7	184.0	89.0	4.8	11.0	34.4	44.1	17.5	21.9	51.5
4.3	31.3	51.6	5.9	8.1	33.3	1.9	68.5	13.3	1.8	71.0	5.7
1.1	82.5	3.3	2.4	95.3	45.9	4.6	94.0	32.6	3.1	79.5	11.4
3.9	72.1	9.1	4.1	70.9	5.2	7.9	68.7	4.9	14.8	81.8	42.0
9.4	67.7	32.0	15.9	78.7	32.8	16.6	78.8	22.0	18.2	80.3	20.8
174.1	135.6	7.3	173.0	127.4	3.0	174.0	125.7	2.2	177.3	121.0	2.2
177.6	120.3	2.2	195.7	144.1	59.4	197.0	142.5	48.1	178.2	112.6	51.5
173.8	112.7	50.3	172.8	124.4	2.9	162.7	114.6	19.1	164.6	120.9	15.1
80.4	90.7	21.7	71.0	88.8	42.4	73.0	85.6	40.2	56.7	95.3	37.4
66.5	86.2	40.1	67.0	84.7	39.5	62.9	87.9	32.5	61.8	89.0	39.5
51.9	94.5	35.6	60.9	71.6	44.1	61.0	69.8	42.2	61.7	66.2	39.4
57.3	68.4	35.5	54.2	76.4	39.1	76.1	52.9	9.5	67.2	57.6	48.4
81.9	58.5	31.9	90.1	59.6	30.7	135.3	126.6	15.0	135.0	124.0	24.5

Table 4(continued).

x	y	dbh	x	y	dbh	x	y	dbh	x	y	dbh
136.2	122.1	15.0	129.7	127.0	22.2	134.8	120.2	27.5	136.9	116.8	10.8
137.0	116.0	26.2	128.9	124.2	10.2	127.5	125.0	18.9	127.6	121.7	44.2
129.7	119.0	13.8	126.6	121.1	16.7	133.4	77.1	35.7	129.9	76.1	12.1
126.5	77.3	35.4	129.1	83.1	32.7	134.4	87.0	30.1	130.7	90.1	28.4
130.9	90.7	16.5	132.0	94.5	12.7	136.8	96.7	5.5	137.7	98.0	2.5
157.8	99.9	3.0	187.1	98.1	3.2	190.6	92.1	3.2	185.4	93.1	4.0
186.6	92.2	3.6	185.9	91.7	3.8	184.3	92.1	4.3	188.2	91.2	3.3
104.4	145.1	6.3	104.9	145.0	18.4	101.5	148.4	5.4	102.4	148.7	5.4
123.4	128.9	26.0	123.8	135.1	22.3	127.0	133.8	35.2	109.6	145.9	24.1
112.4	145.0	6.9	133.1	144.8	61.0	139.4	143.1	20.6	140.4	143.6	6.5
184.1	88.2	2.8	183.5	88.5	4.8	183.0	88.0	5.4	176.1	91.0	4.3
175.6	90.2	4.0	173.8	89.9	3.2	164.9	93.7	2.8	163.0	95.3	4.9
163.2	94.1	3.5	162.4	94.5	2.9	161.5	94.9	2.4	162.2	94.3	3.3
161.0	94.7	2.1	157.7	95.7	2.0	154.9	96.2	3.9	154.6	92.7	5.0
152.9	93.7	2.3	153.2	93.2	2.2	168.2	73.0	67.7	151.6	93.0	2.9
151.4	93.4	2.4	157.6	67.2	56.3	149.4	63.0	39.4	149.4	64.3	59.5
167.3	54.6	42.4	157.4	51.5	63.7	181.5	66.1	66.6	196.5	55.2	69.3
189.9	85.2	56.9	155.1	149.2	23.5	154.5	148.4	9.1	162.9	119.9	29.9
158.4	113.4	14.9	153.9	108.3	38.7	156.1	116.0	31.5	156.5	118.9	27.8
156.8	122.3	28.5	159.0	126.1	21.6	161.0	131.9	2.0	161.3	132.8	2.6
160.6	132.6	2.3	161.3	134.9	3.5	159.7	129.8	3.6	161.7	136.1	2.6
161.1	136.4	2.0	160.1	133.0	2.0	159.0	133.6	2.7	160.0	134.8	2.6
160.2	135.5	2.2	159.1	136.5	2.7	154.7	126.8	30.1	151.9	127.5	16.6
151.3	124.7	10.4	151.0	127.3	11.8	150.4	123.0	32.3	149.6	124.6	33.5
146.2	127.1	30.5	146.1	127.4	10.5	144.4	131.8	13.8	143.3	131.5	22.8
140.6	137.7	31.7	143.2	125.4	10.1	127.1	119.9	14.5	120.7	115.6	12.0
115.3	112.6	2.2	134.1	105.2	2.3	134.6	104.1	3.2	135.6	103.3	3.0
128.9	102.6	50.6	116.3	106.5	2.6	104.3	104.0	50.0	111.5	100.0	52.2
100.5	149.7	5.2	100.0	145.5	5.2	100.8	145.0	6.7	100.9	143.5	14.0
100.3	140.8	12.7	101.5	120.8	59.5	99.3	110.6	52.0	99.2	106.0	45.9
102.0	137.1	18.0	105.4	115.7	43.5	103.6	134.2	3.3	103.9	139.4	4.3
102.6	141.6	7.4	102.0	143.3	10.1	102.1	144.4	23.1	103.5	141.3	8.1
102.9	143.8	5.7	105.7	138.2	13.3	106.6	135.1	12.8	108.5	133.2	11.6
105.2	142.3	6.3	139.7	145.8	20.0	145.5	148.4	8.9	146.4	148.4	27.6
105.8	149.8	4.5	96.7	149.1	9.2	66.5	150.0	2.3	55.7	148.5	5.0
54.7	146.8	4.0	57.1	144.0	21.8	61.7	145.3	10.9	60.1	143.7	14.9
77.7	144.8	45.0	67.2	139.3	16.4	80.7	133.2	43.3	85.1	133.5	55.6
94.7	143.7	10.6	81.2	125.0	45.9	81.9	123.2	45.2	83.8	123.1	35.5
84.8	121.4	43.6	82.9	119.2	44.6	82.1	116.4	38.8	84.3	114.8	34.9
96.7	142.6	17.0	92.0	109.0	50.4	96.1	146.6	2.0	78.5	102.5	33.8
78.7	103.0	51.1	59.5	107.4	21.8	56.5	105.5	46.5	64.3	132.1	5.6
152.7	146.7	19.6	155.8	145.4	32.3	161.2	138.1	3.7	161.0	138.1	2.7
162.1	136.9	2.5	166.2	132.0	2.5	168.7	133.4	2.4	169.3	133.7	7.2
57.9	140.7	7.0	57.5	142.3	11.8	57.3	141.7	8.5	56.0	137.7	9.5
53.4	139.3	7.0	53.1	136.0	10.5	54.0	137.7	6.6	54.5	136.7	6.6
53.3	137.8	8.8	52.1	139.3	11.6	48.0	114.4	48.2	44.2	129.6	36.2
39.4	136.8	44.9	42.7	124.0	43.0	38.1	134.4	37.5	37.1	131.9	31.5
37.6	125.4	39.9	31.2	127.9	35.5	40.1	112.2	51.7	29.3	118.6	36.5
23.8	114.5	40.2	141.0	127.8	7.8	140.1	127.3	17.0	140.9	121.4	36.4
135.0	132.3	19.6	139.3	122.9	15.0	142.0	117.2	28.8	140.4	117.2	20.1

Table 4(continued).

x	y	dbh	x	y	dbh	x	y	dbh	x	y	dbh
138.5	121.5	39.3	28.7	158.8	37.9	33.7	162.3	40.6	23.1	160.8	33.0
11.3	158.9	35.7	18.2	168.2	20.6	21.5	172.3	22.0	15.9	168.3	16.3
15.4	172.8	5.6	14.0	174.2	7.4	6.8	179.6	42.3	6.0	184.1	43.8
1.6	194.9	53.0	43.6	197.3	48.1	39.4	195.5	41.9	37.1	196.1	48.0
23.7	193.9	75.9	21.5	187.9	40.4	27.7	188.7	40.9	32.3	178.9	39.4
32.6	168.6	40.9	37.7	176.9	17.6	107.5	138.5	17.8	107.9	139.5	3.7
116.5	122.6	19.0	114.5	127.7	11.2	115.3	127.4	27.6	115.3	128.1	14.5
119.0	127.4	34.4	119.4	127.7	20.0	94.7	179.8	2.9	89.3	185.0	7.3
90.8	174.0	52.7	95.3	158.4	8.7	90.9	162.1	3.6	90.2	162.1	4.6
90.2	161.7	11.4	90.6	160.8	11.0	93.0	158.0	18.7	78.4	172.4	5.6
76.2	171.4	2.1	75.8	171.0	3.3	75.7	169.7	11.5	82.7	163.5	2.6
76.7	166.3	4.4	74.7	167.1	18.3	119.4	170.8	7.5	74.2	164.3	17.2
73.9	162.7	4.6	81.7	156.7	32.0	79.5	156.3	56.7	56.8	116.0	46.0
62.2	137.7	7.8	58.2	125.1	54.9	54.1	115.5	45.5	59.5	138.1	9.2
58.6	140.3	13.2	58.8	141.5	15.3	57.9	137.3	8.5	153.5	159.9	2.2
155.9	183.7	58.8	160.4	176.6	47.5	171.3	185.1	52.2	182.8	187.4	56.3
182.5	196.0	39.8	176.3	197.7	38.1	161.9	199.4	38.9	199.5	179.4	9.7
197.6	176.9	7.4	196.3	192.4	22.1	195.7	180.5	16.9	196.2	177.1	5.9
196.3	176.0	10.5	193.7	185.8	9.5	191.7	189.2	45.9	194.5	173.8	11.4
192.7	177.3	7.8	188.9	182.1	14.4	190.1	174.4	8.3	186.9	179.4	30.6
26.9	111.3	44.4	17.9	111.0	38.7	34.4	104.2	41.5	31.9	103.2	34.5
20.6	101.5	31.8	14.1	103.1	39.7	2.9	122.8	23.3	6.4	125.9	37.7
2.2	142.2	43.0	11.7	116.2	39.2	14.2	116.5	40.4	15.6	118.1	36.7
13.6	127.4	48.4	11.1	134.8	27.9	7.2	141.7	46.4	12.2	140.1	38.5
23.0	132.7	39.4	30.2	133.9	50.0	27.7	136.5	51.6	3.4	148.8	38.7
15.4	145.6	39.6	16.7	146.4	29.1	24.3	145.7	44.0	0.4	175.2	50.9
0.0	177.5	50.8	7.9	151.0	43.0	33.2	151.2	44.5	36.6	150.6	29.8
42.2	153.7	44.3	24.5	153.4	51.2	40.4	179.3	37.7	41.0	176.6	36.8
43.9	182.2	33.6	44.7	184.6	47.9	45.6	175.2	32.0	47.5	175.9	40.3
51.2	177.9	42.5	55.0	159.3	59.7	58.0	180.3	44.2	54.6	188.7	30.9
58.9	180.0	39.5	63.9	178.6	48.7	64.3	178.9	32.8	65.6	179.3	47.2
61.0	184.9	42.1	63.1	183.3	43.8	86.1	186.9	30.5	65.8	194.9	28.3
90.0	195.1	10.4	94.3	196.1	15.0	91.9	197.1	7.4	86.5	197.4	15.3
87.5	199.3	17.5	93.9	199.2	5.0	92.4	199.3	12.2	81.8	198.9	9.0
99.0	158.1	2.4	94.1	187.2	13.7	95.4	182.9	13.1	97.1	168.4	12.8
79.2	155.6	27.0	61.6	158.2	2.6	70.3	153.1	4.9	79.8	151.8	35.0
110.1	150.4	23.7	116.1	156.8	42.9	114.0	165.1	14.2	103.2	154.4	3.3
112.3	167.0	28.4	110.4	167.3	10.0	110.6	166.4	6.4	107.0	165.0	22.0
105.6	160.6	4.3	104.0	162.4	10.0	104.0	166.1	9.2	103.7	167.2	3.7
108.6	182.1	66.7	105.7	182.6	68.0	102.8	169.7	23.1	101.5	171.8	5.7
100.4	170.5	11.7	144.1	199.0	40.4	138.3	197.9	43.3	142.7	197.2	60.2
118.8	188.0	55.5	142.3	173.3	54.1	143.8	156.0	22.3	145.3	155.6	21.4
151.2	192.2	55.7	153.7	176.5	51.4	186.9	174.7	23.9	181.2	176.9	5.2
181.1	176.1	7.6	177.2	174.5	27.8	182.8	162.9	49.6	180.0	160.2	51.0
189.1	156.3	50.7	196.9	151.4	43.4	171.4	161.6	55.6	169.1	160.0	4.3
162.5	157.3	2.5	156.7	155.3	23.5	154.1	150.8	8.0	87.7	200.0	11.7

Table 5: Coefficients for transforming $\sqrt{b_1}$ to a standard normal using a Johnson S_U approximation.
Reproduced from Table 4 of D'Agostino and Pearson [11].

n	δ	$1/\lambda$	n	δ	$1/\lambda$	n	δ	$1/\lambda$
8	5.563	0.3030	62	3.389	1.0400	260	5.757	1.1744
9	4.260	0.4080	64	3.420	1.0449	270	5.835	1.1761
10	3.734	0.4794	66	3.450	1.0495	280	5.946	1.1779
			68	3.480	1.0540	290	6.039	1.1793
11	3.447	0.5339	70	3.510	1.0581	300	6.130	1.1808
12	3.270	0.5781						
13	3.151	0.6153	72	3.540	1.0621	310	6.220	1.1821
14	3.069	0.6473	74	3.569	1.0659	320	6.308	1.1834
15	3.010	0.6753	76	3.599	1.0695	330	6.396	1.1846
			78	3.628	1.0730	340	6.482	1.1858
16	2.968	0.7001	80	3.657	1.0763	350	6.567	1.1868
17	2.937	0.7224						
18	2.915	0.7426	82	3.686	1.0795	360	6.651	1.1879
19	2.900	0.7610	84	3.715	1.0825	370	6.733	1.1888
20	2.890	0.7779	86	3.744	1.0854	380	6.815	1.1897
			88	3.772	1.0882	390	6.896	1.1906
21	2.884	0.7934	90	3.801	1.0909	400	6.976	1.1914
22	2.882	0.8078						
23	2.882	0.8211	92	3.829	1.0934	410	7.056	1.1922
24	2.884	0.8336	94	3.857	1.0959	420	7.134	1.1929
25	2.889	0.8452	96	3.885	1.0983	430	7.211	1.1937
			98	3.913	1.1006	440	7.288	1.1943
26	2.895	0.8561	100	3.940	1.1028	450	7.363	1.1950
27	2.902	0.8664						
28	2.910	0.8760	105	4.009	1.1080	460	7.438	1.1956
29	2.920	0.8851	110	4.076	1.1128	470	7.513	1.1962
30	2.930	0.8938	115	4.142	1.1172	480	7.586	1.1968
			120	4.207	1.1212	490	7.659	1.1974
31	2.941	0.9020	125	4.272	1.1250	500	7.731	1.1959
32	2.952	0.9097						
33	2.964	0.9171	130	4.336	1.1285	520	7.873	1.1989
34	2.977	0.9241	135	4.398	1.1318	540	8.013	1.1998
35	2.990	0.9308	140	4.460	1.1348	560	8.151	1.2007
			145	4.521	1.1377	580	8.286	1.2015
36	3.003	0.9372	150	4.582	1.1403	600	8.419	1.2023
37	3.016	0.9433						
38	3.030	0.9492	155	4.641	1.1428	620	8.550	1.2030
39	3.044	0.9548	160	4.700	1.1452	640	8.679	1.2036
40	3.058	0.9601	165	4.758	1.1474	660	8.806	1.2043
			170	4.816	1.1496	680	8.931	1.2049
41	3.073	0.9653	175	4.873	1.1516	700	9.054	1.2054
42	3.087	0.9702						
43	3.102	0.9750	180	4.929	1.1535	720	9.176	1.2060
44	3.117	0.9795	185	4.985	1.1553	740	9.297	1.2065
45	3.131	0.9840	190	5.040	1.1570	760	9.415	1.2069
			195	5.094	1.1586	780	9.533	1.2073
46	3.146	0.9882	200	5.148	1.1602	800	9.649	1.2078
47	3.161	0.9923						
48	3.176	0.9963	205	5.202	1.1616	820	9.763	1.2082
49	3.192	1.0001	210	5.255	1.1631	840	9.876	1.2086
50	3.207	1.0038	215	5.307	1.1644	860	9.988	1.2089
			220	5.359	1.1657	880	10.098	1.2093
52	3.237	1.0108	225	5.410	1.1669	900	10.208	1.2096
54	3.268	1.0174						
56	3.298	1.0235	230	5.461	1.1681	920	10.316	1.2100
58	3.329	1.0293	235	5.511	1.1693	940	10.423	1.2103
60	3.359	1.0348	240	5.561	1.1704	960	10.529	1.2106
			245	5.611	1.1714	980	10.634	1.2109
			250	5.660	1.1724	1000	10.738	1.2111

Table 6: Coefficients $\{a_{n-i+1}\}$ for the Shapiro-Wilk W Test for Normality. Reproduced from Table 5 of Shapiro and Wilk [32].

i	n									
	2	3	4	5	6	7	8	9	10	
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	
2	–	0.0000	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	
3	–	–	–	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141	
4	–	–	–	–	–	0.0000	0.0561	0.0947	0.1224	
5	–	–	–	–	–	–	–	0.0000	0.0399	
	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	–	–	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	–	–	–	–	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	–	–	–	–	–	–	0.0000	0.0163	0.0303	0.0422
10	–	–	–	–	–	–	–	–	0.0000	0.0140
	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	–	–	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	–	–	–	–	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	–	–	–	–	–	–	0.0000	0.0084	0.0159	0.0227
15	–	–	–	–	–	–	–	–	0.0000	0.0076
	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2463	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	–	–	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	–	–	–	–	0.0000	0.0057	0.0110	0.0158	0.0203	0.0244
19	–	–	–	–	–	–	0.0000	0.0053	0.0101	0.0146
20	–	–	–	–	–	–	–	–	0.0000	0.0049
	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3964
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2737
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2368
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2098
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1878
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	–	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	–	–	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	–	–	–	–	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	–	–	–	–	–	–	0.0000	0.0037	0.0071	0.0104
25	–	–	–	–	–	–	–	–	0.0000	0.0035

Table 7: Critical Values of the Shapiro-Wilk W for Testing Normality. Reproduced from Table 6 of Shapiro and Wilk [32].

n					
	0.01	0.02	0.05	0.10	0.50
3	0.753	0.756	0.767	0.789	0.959

Table 8: Critcal Values of the Shapiro-Wilk W for Testing Exponentiality. Reproduced from Table 1 of Shapiro and Wilk [33].

n	α										
	0.005	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99	0.995
3	.2519	.2538	.2596	.2697	.2915	.5714	.9709	.9926	.9981	.9997	.99993
4	.1241	.1302	.1434	.1604	.1891	.3768	.7514	.8581	.9236	.9680	.9837

Table 9: Coefficients $\{b_{n-i+1}\}$ for the Shapiro-Francia W' Test for Normality.
Reproduced from Table 1 of Shapiro and Wilk [31].

Table 10: Percentage Points for W' Test Statistic
Reproduced from Table 1 of Shapirio and Francia [31].

n	P										
	0.01	0.05	0.10	0.15	0.20	0.50	0.80	0.85	0.90	0.95	0.99
35	0.919	0.943	0.952	0.956	0.964	0.976	0.982	0.985	0.987	0.989	0.992
50	.935	.953	.963	.968	.971	.981	.987	.988	.990	.991	.994
51	0.935	0.954	0.964	0.968	0.971	0.981	0.988	0.989	0.990	0.992	0.994
53	.938	.957	.964	.969	.972	.982	.988	.989	.990	.992	.994
55	.940	.958	.965	.971	.973	.983	.988	.990	.991	.992	.994
57	.944	.961	.966	.971	.974	.983	.989	.990	.991	.992	.994
59	.945	.962	.967	.972	.975	.983	.989	.990	.991	.992	.994
61	0.947	0.963	0.968	0.973	0.975	0.984	0.990	0.990	0.991	0.992	0.994
63	.947	.964	.970	.973	.976	.984	.990	.991	.992	.993	.994
65	.948	.965	.971	.974	.976	.985	.990	.991	.992	.993	.995
67	.950	.966	.971	.974	.977	.985	.990	.991	.992	.993	.995
69	.951	.966	.972	.976	.978	.986	.990	.991	.992	.993	.995
71	0.953	0.967	0.972	0.976	0.978	0.986	0.990	0.991	0.992	0.994	0.995
73	.956	.968	.973	.976	.979	.986	.991	.992	.993	.994	.995
75	.956	.969	.973	.976	.979	.986	.991	.992	.993	.994	.995
77	.957	.969	.974	.977	.980	.987	.991	.992	.993	.994	.996
79	.957	.970	.975	.978	.980	.987	.991	.992	.993	.994	.996
81	0.958	0.970	0.975	0.979	0.981	0.987	0.992	0.992	0.993	0.994	0.996
83	.960	.971	.976	.979	.981	.988	.992	.992	.993	.994	.996
85	.961	.972	.977	.980	.981	.988	.992	.992	.993	.994	.996
87	.961	.972	.977	.980	.982	.988	.992	.993	.994	.994	.996
89	.961	.972	.977	.981	.982	.988	.992	.993	.994	.995	.996
91	0.962	0.973	0.978	0.981	0.983	0.989	0.992	0.993	0.994	0.995	0.996
93	.963	.973	.979	.981	.983	.989	.992	.993	.994	.995	.996
95	.965	.974	.979	.981	.983	.989	.993	.993	.994	.995	.996
97	.965	.975	.979	.982	.984	.989	.993	.993	.994	.995	.996
99	.967	.976	.980	.982	.984	.989	.993	.994	.994	.995	.996