In addition an extra fault code has been introduced:

IFAULT = 9 if $\theta_{\min}$ has been reached and the maximum likelihood estimate of μ found but moving θ away from $\theta_{\min}$ slightly increases the log-likelihood; the estimate of μ returned is the estimate conditional on $\theta = \theta_{\min}$; the estimates are not the overall maximum likelihood estimates.

References

Prentice, R. L. (1986) Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement error. *J. Am. Statist. Ass.*, **81**, 321–327.

Qu, Y., Beck, G. J. and Williams, G. W. (1990) Polya–Eggenberger distribution: parameter estimation and hypothesis tests. *Biometr. J.*, **32**, 229–242.

Smith, D. M. (1983) Algorithm AS 189: Maximum likelihood estimation of the parameters of the beta binomial distribution. *Appl. Statist.*, **32**, 196–204.

Remark AS R94

A Remark on Algorithm AS 181: The *W*-test for Normality

By Patrick Royston†

Royal Postgraduate Medical School, London, UK

[Received December 1992. Final revision March 1995]

Keywords: Censored data; Goodness of fit; Regression tests of non-normality

Language

Fortran 77

Description and Purpose

Shapiro and Wilk's (1965) *W*-statistic is a well-established and powerful test of departure from normality. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of n observations. The numerator of W is proportional to the square of the 'best' (minimum variance, unbiased) linear estimator of the standard deviation, and the denominator is the sum of squares of the observations about the sample mean. W may also be written as the square of the Pearson correlation coefficient between the ordered observations and the weights $\mathbf{a} = (a_1, \ldots, a_n)$ which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics, W has a simple interpretation as an approximate measure of the

†*Address for correspondence*: Medical Statistics Unit, Royal Postgraduate Medical School, Ducane Road, London, W12 0NN, UK.
E-mail: proyston@rpms.ac.uk

straightness of the normal quantile–quantile (Q–Q) probability plot. An extension to censored samples is described by Verrill and Johnson (1988).

Royston (1982a) presented an algorithm for the W-test, including a method for obtaining the P-value of the test. However, Royston (1992) observed that Shapiro and Wilk's (1965) approximation to the weights $\mathbf{a}$ used in that algorithm, together with an extension for large samples provided by Royston (1982b), was inadequate for $n > 50$. He gave an improved approximation for the weights and described a simple normalizing transformation of the distribution of W, enabling its P-value to be determined for any n in the range $3 \leqslant n \leqslant 5000$. Full details are given by Royston (1992).

Royston (1993a) used Verrill and Johnson's (1988) percentage points for $\sqrt{W}$ in singly censored samples to extend the P-value algorithm to the censored case. Royston's (1993a) method gives approximate P-values for $20 \leqslant n \leqslant 5000$ and for $0 \leqslant \delta \leqslant 0.8$, where δ is the proportion censored ($\delta = 0$ corresponds to the complete sample). The approximation is acceptable for practical data analysis for $P < 0.1$, the critical region of most interest.

The present algorithm embodies the methods described by Royston (1992, 1993a). For censored samples, it expects right (upper) censoring, which is achieved without loss of generality for left-censored samples by changing the sign of each observation.

Structure

SUBROUTINE SWILK(INIT, X, N, N1, N2, A, W, PW, IFAULT)

Formal parameters

INIT	Logical	input:	.FALSE. if weights A(N2) are to be calculated, .TRUE. otherwise
		output:	.TRUE. unless IFAULT output as 1 or 3
X	Real array (N1)	input:	sample values in ascending order
N	Integer	input:	the total sample size
N1	Integer	input:	the number of uncensored observations (N1 $\leqslant N$)
N2	Integer	input:	$[N/2]$, i.e. $\frac{1}{2}N$ if N is even, $\frac{1}{2}(N-1)$ if N is odd
A	Real array (N2)	output:	weights (see INIT)
W	Real	output:	the W-statistic
PW	Real	output:	the P-value for W
IFAULT	Integer	output:	a fault indicator: = 0 for no fault; = 1 if N1 < 3; = 2 if $N > 5000$ (a non-fatal error); = 3 if N2 < $N/2$, so insufficient storage for A; = 4 if N1 > N or (N1 < N and $N < 20$); = 5 if the proportion censored $(N - N1)/N > 0.8$; = 6 if the data have zero range (if sorted on input);

= 7 if the data are not in ascending order

The uncensored observations must be stored in ascending order in X[1], . . ., X[N1] before SWILK is invoked. If there is no censoring then N1 = N, the total sample size. For a given value of N, SWILK must be called at least once with INIT set to .FALSE. to ensure calculation of the correct weights.

Failure Indications

All calculations are carried out for samples larger than 5000, but IFAULT is returned as 2. Although W will be calculated correctly, the accuracy of its P-value cannot be guaranteed.

Auxiliary Algorithms

The following auxiliary routines are required.

(a) FUNCTION POLY(C, NORD, X) evaluates a polynomial on X, namely C(1) + C(2)*X + C(3)*X**2 + . . . + C(NORD)*X**(NORD − 1), e.g. as supplied with algorithm AS 181 (Royston, 1982a);
(b) FUNCTION ALNORM(X, UPPER)—if UPPER is .TRUE., the upper normal tail area at X, e.g. algorithm AS 66 (Hill, 1973);
(c) FUNCTION PPND(P)—the normal deviate corresponding to P, e.g. algorithm AS 111 (Beasley and Springer, 1977).

Constants

The value of SMALL, used for example for 'fuzzy' comparisons of successive values of X in checking the sort order, is set to 10^{-19}. SMALL may be made much smaller in a double-precision implementation of SWILK.

Precision

Fortran 32-bit single precision should prove adequate. However, if a double-precision version is required, all REAL declarations should be changed to DOUBLE PRECISION and the exponents of all constants should be changed from E to D (e.g. 1.0D0 instead of 1.0E0). Double-precision versions of functions POLY, ALNORM and PPND will also be needed.

Validation

Table 1 shows the estimated probability of SWILK rejecting the null hypothesis of normality when it is true. The nominal type I error rate is 0.05. The values were obtained by using Monte Carlo simulation with 10000 replicates for each combination of δ and n.

Each estimate has a standard error of about 0.002. Although there is some evidence of systematic error in the rejection rate when $\delta > 0$, nevertheless the type I error rate is sufficiently close to 0.05 for practical purposes.

TABLE 1
Empirical rejection rates of the null hypothesis of normality†

n	*Rejection rates for the following values of* δ:				
	0.0	*0.2*	*0.4*	*0.6*	*0.8*
25	0.052	0.051	0.052	0.045	0.055
50	0.049	0.054	0.053	0.047	0.058
100	0.048	0.053	0.051	0.045	0.057
200	0.048	0.053	0.056	0.046	0.061

†The nominal rate is 0.05.

Related Algorithms

Algorithm AS 248 (Davis and Stephens, 1989) includes several tests of departure from normality by using empirical distribution function statistics. It also tests for departure from uniformity and exponentiality but does not cater for censored samples.

Test Data

The following is an ordered pseudorandom sample of size 25 from a log-normal distribution:

0.139, 0.157, 0.175, 0.256, 0.344, 0.413, 0.503, 0.577, 0.614, 0.655, 0.954, 1.392, 1.557, 1.648, 1.690, 1.994, 2.174, 2.206, 3.245, 3.510, 3.571, 4.354, 4.980, 6.084, 8.351.

For $\delta = \{0, 0.2, 0.4, 0.6, 0.8\}$, the results are

$$W = \{0.83467, 0.77613, 0.73455, 0.83230, 0.66402\}$$

and

$$P = \{0.000914, 0.000338, 0.000878, 0.089028, 0.030112\}.$$

Additional Comments

Values of W and P from the original (Royston, 1982a) and present algorithms agree fairly closely for $N \leqslant 50$ but diverge for higher sample sizes. Users should trust the results from the present algorithm. For example, they will now observe the expected convergence of the W- and Shapiro–Francia W'-test statistics towards their common asymptotic distribution (Verrill and Johnson, 1988). A convenient algorithm for computing the W'-test is described by Royston (1993b).

Note that the power of the W-test to detect non-normality is inversely related to δ, the amount of censoring, and may in some circumstances be very poor. The trend to lower power is seen in the above example (see the 'Test data' section).

References

Beasley, J. D. and Springer, S. G. (1977) Algorithm AS 111: The percentage points of the normal distribution. *Appl. Statist.*, **26**, 118–121.
Davis, C. S. and Stephens, M. A. (1989) Algorithm AS 248: Empirical distribution function goodness-of-fit tests. *Appl. Statist.*, **38**, 535–543.
Hill, I. D. (1973) Algorithm AS 66: The normal integral. *Appl. Statist.*, **22**, 424–427.
Royston, J. P. (1982a) Algorithm AS 181: The *W* test for normality. *Appl. Statist.*, **31**, 176–180.
——— (1982b) An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl. Statist.*, **31**, 115–124.
——— (1992) Approximating the Shapiro–Wilk *W*-test for non-normality. *Statist. Comput.*, **2**, 117–119.
——— (1993a) A toolkit for testing for non-normality in complete and censored samples. *Statistician*, **42**, 37–43.
——— (1993b) A pocket-calculator algorithm for the Shapiro–Francia test for non-normality: an application to medicine. *Statist. Med.*, **12**, 181–184.
Shapiro, S. S. and Wilk, M. B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
Verrill, S. and Johnson, A. (1988) Tables and large-sample distribution theory for censored-data correlation statistics for testing normality. *J. Am. Statist. Ass.*, **83**, 1192–1197.

Remark AS R95

A Remark on Algorithm AS 226: Computing Non-central Beta Probabilities

By Ming Long Lam†

SPSS Inc., Chicago, USA

[Received May 1993. Revised May 1995]

Frick (1990) improved the algorithm BETANC of Lenth (1987) for computing cumulative probabilities for the non-central beta distribution. The following changes will correct algorithm AS R84 and further enhance algorithm AS 226.

Frick (1990) proposed to start the partial sum from an index X0 to avoid floating point underflow. However, the starting index is still zero in the code. Therefore, the code should be corrected by changing the statement

XJ = ZERO

to

XJ = X0.

In remark AS R84, the error bound is 1 – Φ(UALPHA) + ERRMAX where Φ is the standard normal cumulative distribution function. Frick (1990) used

†*Address for correspondence*: SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611-3962, USA.
E-mail: ming@spss.com