Title

stata.com

swilk — Shapiro–Wilk and Shapiro–Francia tests for normality

Syntax

Shapiro–Wilk normality test

`swilk` *varlist* [*if*] [*in*] [, *swilk_options*]

Shapiro–Francia normality test

`sfrancia` *varlist* [*if*] [*in*] [, *sfrancia_options*]

swilk_options	Description
Main	
`generate(`*newvar*`)`	create *newvar* containing W test coefficients
`lnnormal`	test for three-parameter lognormality
`noties`	do not use average ranks for tied values

sfrancia_options	Description
Main	
`boxcox`	use the Box–Cox transformation for W'; the default is to use the log transformation
`noties`	do not use average ranks for tied values

`by` is allowed with `swilk` and `sfrancia`; see [D] **by**.

Menu

swilk

Statistics > Summaries, tables, and tests > Distributional plots and tests > Shapiro-Wilk normality test

sfrancia

Statistics > Summaries, tables, and tests > Distributional plots and tests > Shapiro-Francia normality test

Description

`swilk` performs the Shapiro–Wilk W test for normality, and `sfrancia` performs the Shapiro–Francia W' test for normality. `swilk` can be used with $4 \leq n \leq 2000$ observations, and `sfrancia` can be used with $5 \leq n \leq 5000$ observations; see [R] **sktest** for a test allowing more observations. See [MV] **mvtest normality** for multivariate tests of normality.

Options for swilk

Main

`generate(`*newvar*`)` creates new variable *newvar* containing the W test coefficients.

`lnnormal` specifies that the test be for three-parameter lognormality, meaning that $\ln(X-k)$ is tested for normality, where k is calculated from the data as the value that makes the skewness coefficient zero. When simply testing $\ln(X)$ for normality, do not specify this option. See [R] **lnskew0** for estimation of k.

`noties` suppresses use of averaged ranks for tied values when calculating the W test coefficients.

Options for sfrancia

Main

`boxcox` specifies that the Box–Cox transformation of Royston (1983) for calculating W' test coefficients be used instead of the default log transformation (Royston 1993a). Under the Box–Cox transformation, the normal approximation to the sampling distribution of W', used by `sfrancia`, is valid for $5 \leq n \leq 1000$. Under the log transformation, it is valid for $10 \leq n \leq 5000$.

`noties` suppresses use of averaged ranks for tied values when calculating the W' test coefficients.

Remarks and examples

stata.com

▷ Example 1

Using our automobile dataset, we will test whether the variables `mpg` and `trunk` are normally distributed:

```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)
. swilk mpg trunk
                    Shapiro-Wilk W test for normal data
    Variable |   Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
         mpg |    74    0.94821      3.335     2.627    0.00430
       trunk |    74    0.97921      1.339     0.637    0.26215
. sfrancia mpg trunk
                   Shapiro-Francia W' test for normal data
    Variable |   Obs       W'          V'        z       Prob>z
-------------+--------------------------------------------------
         mpg |    74    0.94872      3.650     2.510    0.00604
       trunk |    74    0.98446      1.106     0.195    0.42271
```

We can reject the hypothesis that `mpg` is normally distributed, but we cannot reject that `trunk` is normally distributed.

The values reported under W and W' are the Shapiro–Wilk and Shapiro–Francia test statistics. The tests also report V and V', which are more appealing indexes for departure from normality. The median values of V and V' are 1 for samples from normal populations. Large values indicate nonnormality. The 95% critical values of V (V'), which depend on the sample size, are between 1.2 and 2.4 (2.0 and 2.8); see Royston (1991b). There is no more information in V (V') than in W (W')—one is just the transform of the other.

◁

▷ Example 2

We have data on a variable called `studytime`, which we suspect is distributed lognormally:

```
. use http://www.stata-press.com/data/r13/cancer
(Patient Survival in Drug Trial)
. generate lnstudytime = ln(studytime)
. swilk lnstudytime
                    Shapiro-Wilk W test for normal data
    Variable |   Obs        W           V         z       Prob>z
-------------+--------------------------------------------------
 lnstudytime |    48    0.92731       3.311     2.547     0.00543
```

We can reject the lognormal assumption. We do *not* specify the `lnnormal` option when testing for lognormality. The `lnnormal` option is for three-parameter lognormality.

◁

▷ Example 3

Having discovered that $\ln(\texttt{studytime})$ is not distributed normally, we now test that $\ln(\texttt{studytime} - k)$ is normally distributed, where k is chosen so that the resulting skewness is zero. We obtain the estimate for k from `lnskew0`; see [R] **lnskew0**:

```
. lnskew0 lnstudytimek = studytime, level(95)
       Transform |          k     [95% Conf. Interval]        Skewness
-----------------+-----------------------------------------------------
  ln(studytim-k) |  -11.01181    -infinity  -.9477328       -.0000173
. swilk lnstudytimek, lnnormal
        Shapiro-Wilk W test for 3-parameter lognormal data
    Variable |   Obs        W           V         z       Prob>z
-------------+--------------------------------------------------
lnstudytimek |    48    0.97064       1.337     1.261     0.10363
```

We cannot reject the hypothesis that $\ln(\texttt{studytime} + 11.01181)$ is distributed normally. We do specify the `lnnormal` option when using an estimated value of k.

◁

Stored results

`swilk` and `sfrancia` store the following in `r()`:

Scalars

`r(N)`	number of observations	`r(W)`	W or W'
`r(p)`	significance	`r(V)`	V or V'
`r(z)`	z statistic		

Methods and formulas

The Shapiro–Wilk test is based on Shapiro and Wilk (1965) with a new approximation accurate for $4 \leq n \leq 2000$ (Royston 1992). The calculations made by `swilk` are based on Royston (1982, 1992, 1993b).

The Shapiro–Francia test (Shapiro and Francia 1972; Royston 1983; Royston 1993a) is an approximate test that is similar to the Shapiro–Wilk test for very large samples.

Samuel Sanford Shapiro (1930–) earned degrees in statistics and engineering from City College of New York, Columbia, and Rutgers. After employment in the U.S. Army and industry, he joined the faculty at Florida International University in 1972. Shapiro has coauthored various texts in statistics and published several papers on distributional testing and other statistical topics.

Acknowledgment

`swilk` and `sfrancia` were written by Patrick Royston of the MRC Clinical Trials Unit, London and coauthor of the Stata Press book *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*.

References

Brzezinski, M. 2012. The Chen–Shapiro test for normality. *Stata Journal* 12: 368–374.

Genest, C., and G. J. Brackstone. 2010. A conversation with Martin Bradbury Wilk. *Statistical Science* 25: 258–273.

Gould, W. W. 1992. sg3.7: Final summary of tests of normality. *Stata Technical Bulletin* 5: 10–11. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 114–115. College Station, TX: Stata Press.

Royston, P. 1982. An extension of Shapiro and Wilks's *W* test for normality to large samples. *Applied Statistics* 31: 115–124.

——. 1983. A simple method for evaluating the Shapiro–Francia *W*' test of non-normality. *Statistician* 32: 297–300.

——. 1991a. sg3.2: Shapiro–Wilk and Shapiro–Francia tests. *Stata Technical Bulletin* 3: 19. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, p. 105. College Station, TX: Stata Press.

——. 1991b. Estimating departure from normality. *Statistics in Medicine* 10: 1283–1293.

——. 1992. Approximating the Shapiro–Wilk *W*-test for non-normality. *Statistics and Computing* 2: 117–119.

——. 1993a. A pocket-calculator algorithm for the Shapiro–Francia test for non-normality: An application to medicine. *Statistics in Medicine* 12: 181–184.

——. 1993b. A toolkit for testing for non-normality in complete and censored samples. *Statistician* 42: 37–43.

Shapiro, S. S., and R. S. Francia. 1972. An approximate analysis of variance test for normality. *Journal of the American Statistical Association* 67: 215–216.

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

Also see

[R] **lnskew0** — Find zero-skewness log or Box–Cox transform

[R] **lv** — Letter-value displays

[R] **sktest** — Skewness and kurtosis test for normality

[MV] **mvtest normality** — Multivariate normality tests