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

**Description**

`swilk` performs the Shapiro–Wilk $W$ test for normality for each variable in the specified varlist. Likewise, `sfrancia` performs the Shapiro–Francia $W'$ test for normality. See [MV] `mvtest normality` for multivariate tests of normality.

**Quick start**

**Shapiro–Wilk test of normality**

Shapiro–Wilk test for `v1`

```
swilk v1
```

Separate tests of normality for `v1` and `v2`

```
swilk v1 v2
```

Generate new variable `w` containing $W$ test coefficients

```
swilk v1, generate(w)
```

Specify that average ranks should not be used for tied values

```
swilk v1 v2, noties
```

Test that `v3` is distributed lognormally

```
generate lnv3 = ln(v3)
swilk lnv3
```

**Shapiro–Francia test of normality**

Shapiro–Francia test for `v1`

```
sfrancia v1
```

Separate tests of normality for `v1` and `v2`

```
sfrancia v1 v2
```

As above, but use the Box–Cox transformation

```
sfrancia v1 v2, boxcox
```

Specify that average ranks should not be used for tied values

```
sfrancia v1 v2, noties
```

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

Syntax

Shapiro–Wilk normality test

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

Shapiro–Francia normality test

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

**swilk_options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>generate(newvar)</code></td>
</tr>
<tr>
<td><code>lnnormal</code></td>
</tr>
<tr>
<td><code>noties</code></td>
</tr>
</tbody>
</table>

**sfrancia_options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>boxcox</code></td>
</tr>
<tr>
<td><code>noties</code></td>
</tr>
</tbody>
</table>

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

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

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

swilk can be used with $4 \leq n \leq 2000$ observations. sfrancia can be used with $10 \leq n \leq 5000$ observations; however, if the boxcox option is specified, it can be used with $5 \leq n \leq 1000$ observations.

Also see [R] sktest for the skewness and kurtosis test described by D’Agostino, Belanger, and D’Agostino (1990) with the empirical correction developed by Royston (1991b). While the Shapiro–Wilk and Shapiro–Francia tests for normality are, in general, preferred for nonaggregated data (Gould and Rogers 1991; Gould 1992b; Royston 1991b), the skewness and kurtosis test will permit more observations. Moreover, a normal quantile plot should be used with any test for normality; see [R] diagnostic plots for more information.

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/r14/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 (1991d). 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/r14/cancer
(Patient Survival in Drug Trial)
generate lnstudytime = ln(studytime)
swilk lnstudytime
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>W</th>
<th>V</th>
<th>z</th>
<th>Prob&gt;z</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnstudytime</td>
<td>48</td>
<td>0.92731</td>
<td>3.311</td>
<td>2.547</td>
<td>0.00543</td>
</tr>
</tbody>
</table>

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(studytime) is not distributed normally, we now test that ln(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)
```

<table>
<thead>
<tr>
<th>Transform</th>
<th>k</th>
<th>[95% Conf. Interval]</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(studytim-k)</td>
<td>-11.01181</td>
<td>-infinity</td>
<td>-.9477328</td>
</tr>
</tbody>
</table>

```
swilk lnstudytimek, lnnormal
```

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>W</th>
<th>V</th>
<th>z</th>
<th>Prob&gt;z</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnstudytimek</td>
<td>48</td>
<td>0.97064</td>
<td>1.337</td>
<td>1.261</td>
<td>0.10363</td>
</tr>
</tbody>
</table>
```

We cannot reject the hypothesis that ln(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(p): significance
- r(z): z statistic
- r(W): W or W’
- r(V): V or V’

Methods and formulas

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.

The relative merits of the Shapiro–Wilk and Shapiro–Francia tests versus skewness and kurtosis test have been a subject of debate. The interested reader is directed to the articles in the Stata Technical Bulletin. Our recommendation is to use the Shapiro–Francia test whenever possible, that is, whenever dealing with nonaggregated or ungrouped data (Gould and Rogers 1991; Gould 1992b); see [R] swilk. If normality is rejected, use sktest to determine the source of the problem. As both D’Agostino, Belanger, and D’Agostino (1990) and Royston (1991c) mention, researchers should also examine the normal quantile plot to determine normality rather than blindly relying on a few test statistics. See the qnorm command documented in [R] diagnostic plots for more information on normal quantile plots.

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


Also see

[R] **Inskew0** — Find zero-skewness log or Box–Cox transform

[R] **lv** — Letter-value displays

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

[R] **diagnostic plots** — Distributional diagnostic plots

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