Title

ladder — Ladder of powers

Description

`ladder` searches a subset of the ladder of powers (Tukey 1977) for a transform that converts *varname* into a normally distributed variable.

`gladder` and `qladder` each display a graph matrix. `gladder` displays nine histograms of transforms of *varname* according to the ladder of powers. `qladder` displays the quantiles of transforms of *varname* according to the ladder of powers against the quantiles of a normal distribution.

Quick start

Table showing Tukey's ladder of powers transformations for `v`

```
ladder v
```

Same as above, but with separate tables for each level of the categorical variable `catvar`

```
by catvar: ladder v
```

Display transformations graphically using histograms

```
gladder v
```

Same as above, but using quantile plots

```
qladder v
```

Menu

ladder

Statistics > Summaries, tables, and tests > Distributional plots and tests > Ladder of powers

gladder

Statistics > Summaries, tables, and tests > Distributional plots and tests > Ladder-of-powers histograms

qladder

Statistics > Summaries, tables, and tests > Distributional plots and tests > Ladder-of-powers quantile–normal plots

Syntax

Ladder of powers

`ladder` *varname* [*if*] [*in*] [, `generate(`*newvar*`)` `noadjust`]

Ladder-of-powers histograms

`gladder` *varname* [*if*] [*in*] [, *histogram_options combine_options*]

Ladder-of-powers quantile–normal plots

`qladder` *varname* [*if*] [*in*] [, *qnorm_options combine_options*]

`by` and `collect` are allowed with `ladder`; see [U] **11.1.10 Prefix commands**.

Options for ladder

Main

`generate(`*newvar*`)` saves the transformed values corresponding to the minimum χ^2 value from the table. We do not recommend using `generate()` because it is literal in interpreting the minimum, thus ignoring nearly equal but perhaps more interpretable transforms.

`noadjust` is the `noadjust` option to `sktest`; see [R] **sktest**.

Options for gladder

histogram_options affect the rendition of the histograms across all relevant transformations; see [R] **histogram**. Here the `normal` option is assumed, so you must supply the `nonormal` option to suppress the overlaid normal density. Also, `gladder` does not allow the `width(`*#*`)` option of `histogram`.

combine_options are any of the options documented in [G-2] **graph combine**. These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Options for qladder

qnorm_options affect the rendition of the quantile–normal plots across all relevant transformations. See [R] **Diagnostic plots**.

combine_options are any of the options documented in [G-2] **graph combine**. These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Remarks and examples

stata.com

▷ Example 1: ladder

We have data on the mileage rating of 74 automobiles and wish to find a transform that makes the variable normally distributed:

```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)

. ladder mpg
Transformation         Formula               chi2(2)  Prob > chi2
-----------------------------------------------------------------
Cubic                  mpg^3                   43.59        0.000
Square                 mpg^2                   27.03        0.000
Identity               mpg                     10.95        0.004
Square root            sqrt(mpg)                4.94        0.084
Log                    log(mpg)                 0.87        0.647
1/(Square root)        1/sqrt(mpg)              0.20        0.905
Inverse                1/mpg                    2.36        0.307
1/Square               1/(mpg^2)               11.99        0.002
1/Cubic                1/(mpg^3)               24.30        0.000
```

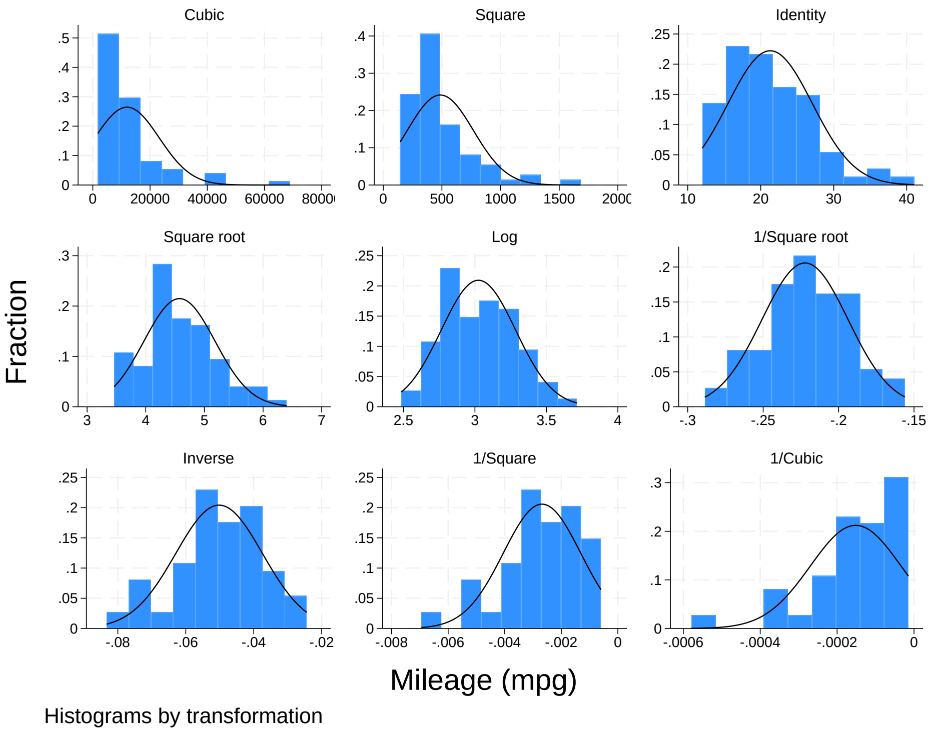
If we had typed `ladder mpg, gen(mpgx)`, the variable `mpgx` containing $1/\sqrt{\texttt{mpg}}$ would have been automatically generated for us. This is the perfect example of why you should not, in general, specify the `generate()` option. We also cannot reject the hypothesis that the inverse of `mpg` is normally distributed and that 1/`mpg`—gallons per mile—has a better interpretation. It is a measure of energy consumption.

◁

▷ Example 2: gladder

`gladder` explores the same transforms as `ladder` but presents results graphically:

```
. gladder mpg, fraction
```



◁

❑ Technical note

`gladder` is useful pedagogically, but be careful when using it for research work, especially with many observations. For instance, consider the following data on the average July temperature in degrees Fahrenheit for 954 U.S. cities:

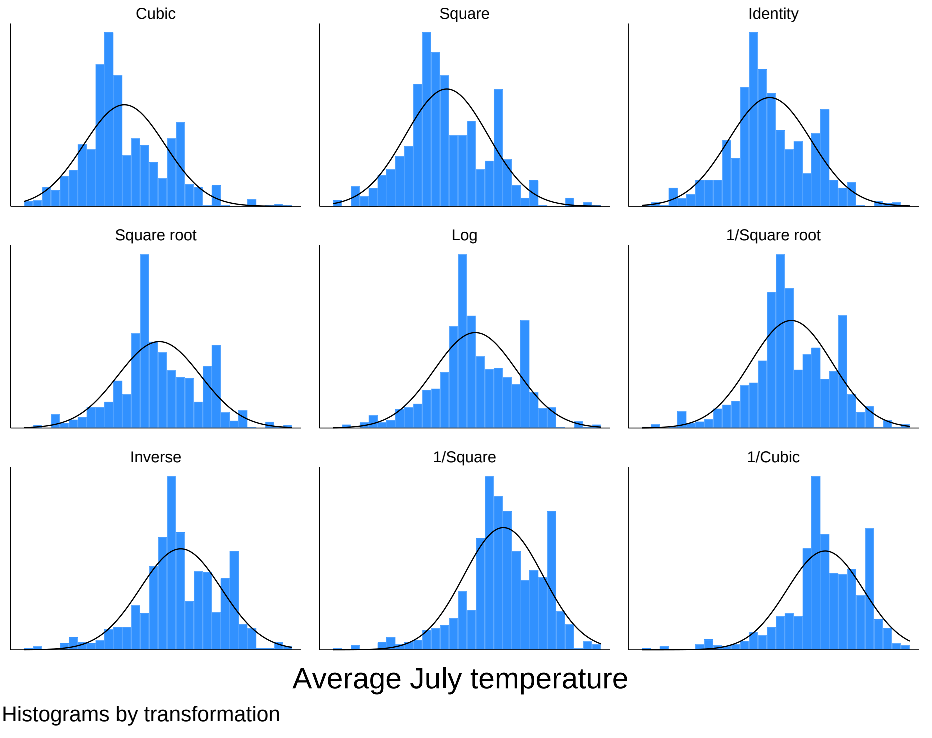
```
. use https://www.stata-press.com/data/r18/citytemp
(City temperature data)

. ladder tempjuly

Transformation         Formula               chi2(2)   Prob > chi2
------------------------------------------------------------------
Cubic                  tempjuly^3              47.49         0.000
Square                 tempjuly^2              19.70         0.000
Identity               tempjuly                 3.83         0.147
Square root            sqrt(tempjuly)           1.83         0.400
Log                    log(tempjuly)            5.40         0.067
1/(Square root)        1/sqrt(tempjuly)        13.72         0.001
Inverse                1/tempjuly              26.36         0.000
1/Square               1/(tempjuly^2)          64.44         0.000
1/Cubic                1/(tempjuly^3)         116.16         0.000
```

From the table, we see that there is certainly a difference in normality between the square and square-root transform. If, however, you can see the difference between the transforms in the diagram below, you have better eyes than we do:

```
. gladder tempjuly, l1title("") ylabel(none) xlabel(none)
```

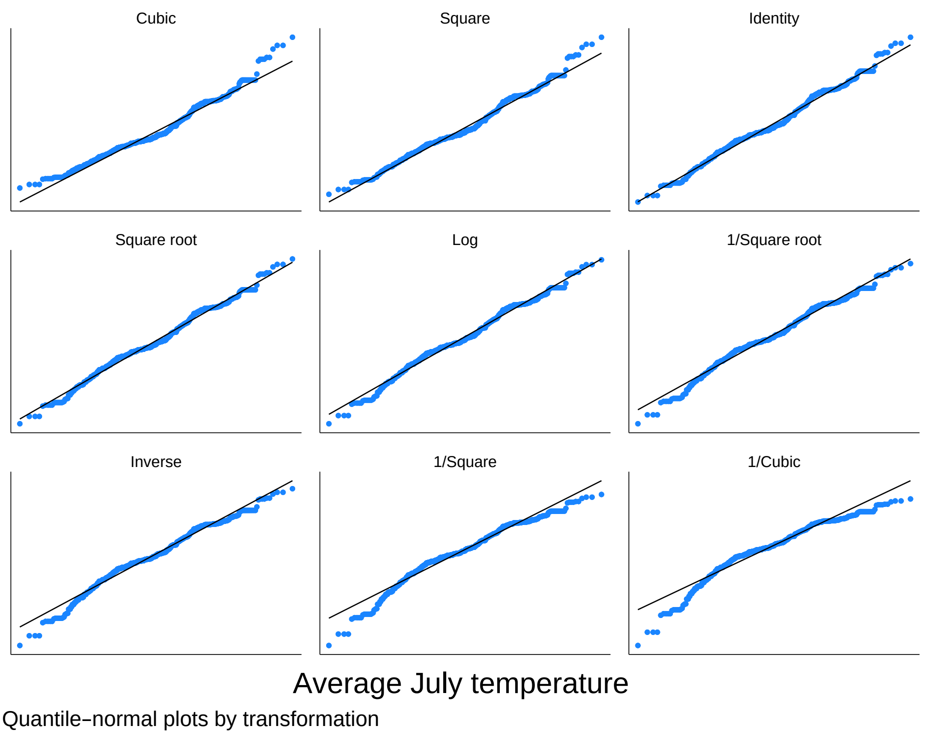


❑

▷ Example 3: qladder

A better graph for seeing normality is the quantile–normal graph, which can be produced by `qladder`.

```
. qladder tempjuly, ylabel(none) xlabel(none)
```



This graph shows that for the square transform, the upper tail—and only the upper tail—diverges from what would be expected. This divergence is detected by `sktest` (see [R] **sktest**) as a problem with skewness, as we would learn from using `sktest` to examine `tempjuly` squared and square rooted.

◁

Stored results

`ladder` stores the following in `r()`:

Scalars

`r(N)`	number of observations
`r(invcube)`	χ^2 for inverse-cubic transformation
`r(P_invcube)`	p-value for normality test after inverse-cubic transformation
`r(invsq)`	χ^2 for inverse-square transformation
`r(P_invsq)`	p-value for normality test after inverse-square transformation
`r(inv)`	χ^2 for inverse transformation
`r(P_inv)`	p-value for normality test after inverse transformation
`r(invsqrt)`	χ^2 for inverse-root transformation
`r(P_invsqrt)`	p-value for normality test after inverse-root transformation
`r(log)`	χ^2 for log transformation
`r(P_log)`	p-value for normality test after log transformation
`r(sqrt)`	χ^2 for square-root transformation
`r(P_sqrt)`	p-value for normality test after square-root transformation
`r(ident)`	χ^2 for untransformed data
`r(P_ident)`	p-value for normality test of untransformed data
`r(square)`	χ^2 for square transformation
`r(P_square)`	p-value for normality test after square transformation
`r(cube)`	χ^2 for cubic transformation
`r(P_cube)`	p-value for normality test after cubic transformation

Methods and formulas

For `ladder`, results are as reported by `sktest`; see [R] **sktest**. If `generate()` is specified, the transform with the minimum χ^2 value is chosen.

`gladder` sets the number of bins to $\min(\sqrt{n}, 10\log_{10}n)$, rounded to the closest integer, where n is the number of unique values of *varname*. See [R] **histogram** for a discussion of the optimal number of bins.

Also see Findley (1990) for a ladder-of-powers variable transformation program that produces one-way graphs with overlaid box plots, in addition to histograms with overlaid normals. Buchner and Findley (1990) discuss ladder-of-powers transformations as one aspect of preliminary data analysis. Also see Hamilton (1992, 18–23) and Hamilton (2013, 129–132).

Acknowledgment

`qladder` was written by Jeroen Weesie of the Department of Sociology at Utrecht University, The Netherlands.

References

Buchner, D. M., and T. W. Findley. 1990. Research in physical medicine and rehabilitation: VIII. Preliminary data analysis. *American Journal of Physical Medicine and Rehabilitation* 69: 154–169. https://doi.org/10.1097/00002060-199006000-00011.

Cox, N. J. 2005. Speaking Stata: Density probability plots. *Stata Journal* 5: 259–273.

Findley, T. W. 1990. sed3: Variable transformation and evaluation. *Stata Technical Bulletin* 2: 15. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 85–86. College Station, TX: Stata Press.

Hamilton, L. C. 1992. *Regression with Graphics: A Second Course in Applied Statistics*. Belmont, CA: Duxbury.

——. 2013. *Statistics with Stata: Updated for Version 12*. 8th ed. Boston: Brooks/Cole.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison–Wesley.

Also see

[R] **boxcox** — Box–Cox regression models

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

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

[R] **lv** — Letter-value displays

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

For suggested citations, see the FAQ on citing Stata documentation.