

screeplot — Scree plot

[Description](#)
[Options](#)
[Also see](#)

[Quick start](#)
[Remarks and examples](#)

[Menu](#)
[Stored results](#)

[Syntax](#)
[References](#)

Description

`screeplot` produces a scree plot of the eigenvalues of a covariance or correlation matrix.

`screeplot` automatically obtains the eigenvalues after `candisc`, `discrim lda`, `factor`, `factormat`, `pca`, and `pcamat`. `screeplot` also works automatically to plot singular values after `ca` and `camat`, canonical correlations after `canon`, and eigenvalues after `manova`, `mca`, `mds`, `mdsmat`, and `mdslong`.

`screeplot` lets you obtain a scree plot in other cases by directly specifying *eigvals*, a vector containing the eigenvalues.

Quick start

Scree plot of the eigenvalues after a command such as `pca` or `factor`
`screeplot`

Add a horizontal line at the mean of the eigenvalues
`screeplot, mean`

Scree plot with confidence intervals after `pca`
`screeplot, ci`

Menu

Statistics > Multivariate analysis > Factor and principal component analysis > Postestimation > Scree plot of eigenvalues

Syntax

```
screepplot [eigvals] [, options]
```

`screep` is a synonym for `screepplot`.

<i>options</i>	Description
Main	
<code>neigen(#)</code>	graph only largest # eigenvalues; default is to plot all eigenvalues
Mean	
<code>mean</code>	graph horizontal line at the mean of the eigenvalues
<code>meanlopts(<i>cline_options</i>)</code>	affect rendition of the mean line
CI	
<code>ci[(<i>ci_options</i>)]</code>	graph confidence intervals (after <code>pca</code> only); <code>ci</code> is a synonym for <code>ci(asymptotic)</code>
Plot	
<code>cline_options</code>	affect rendition of the lines connecting points
<code>marker_options</code>	change look of markers (color, size, etc.)
Add plots	
<code>addplot(<i>plot</i>)</code>	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
<code>twoway_options</code>	any options other than <code>by()</code> documented in [G-3] twoway_options

<i>ci_options</i>	Description
<code>asymptotic</code>	compute asymptotic confidence intervals; the default
<code>heteroskedastic</code>	compute heteroskedastic bootstrap confidence intervals
<code>homoskedastic</code>	compute homoskedastic bootstrap confidence intervals
<code>area_options</code>	affect the rendition of the confidence bands
<code>table</code>	produce a table of confidence intervals
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
<code>reps(#)</code>	number of bootstrap simulations; default is <code>reps(200)</code>
<code>seed(<i>str</i>)</code>	random-number seed used for the bootstrap simulations

Options

Main

`neigen(#)` specifies the number of eigenvalues to plot. The default is to plot all eigenvalues.

Mean

`mean` displays a horizontal line at the mean of the eigenvalues.

`meanlopts(cline_options)` affects the rendition of the mean reference line added using the `mean` option; see [G-3] [cline_options](#).

CI

`ci` [*ci_options*] displays confidence intervals for the eigenvalues. The option `ci` is a synonym for `ci(asymptotic)`. The following methods for estimating confidence intervals are available:

`ci(asymptotic)` specifies the asymptotic distribution of the eigenvalues of a central Wishart distribution, the distribution of the covariance matrix of a sample from a multivariate normal distribution. The asymptotic theory applied to correlation matrices is not fully correct, probably giving confidence intervals that are somewhat too narrow.

`ci(heteroskedastic)` specifies a parametric bootstrap by using the percentile method and assuming that the eigenvalues are from a matrix that is multivariate normal with the same eigenvalues as observed.

`ci(homoskedastic)` specifies a parametric bootstrap by using the percentile method and assuming that the eigenvalues are from a matrix that is multivariate normal with all eigenvalues equal to the mean of the observed eigenvalues. For a PCA of a correlation matrix, this mean is 1.

`ci(area_options)` affects the rendition of the confidence bands; see [G-3] [area_options](#).

`ci(table)` produces a table with the confidence intervals.

`ci(level(#))` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] [20.8 Specifying the width of confidence intervals](#).

`ci(reps(#))` specifies the number of simulations to be performed for estimating the confidence intervals. This option is valid only when `heteroskedastic` or `homoskedastic` is specified. The default is `reps(200)`.

`ci(seed(str))` sets the random-number seed used for the parametric bootstrap. Setting the seed makes sure that results are reproducible. See `set seed` in [R] [set seed](#). This option is valid only when `heteroskedastic` or `homoskedastic` is specified.

The confidence intervals are not adjusted for “simultaneous inference”.

Plot

cline_options affect the rendition of the lines connecting the plotted points; see [G-3] [cline_options](#).

marker_options affect the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] [marker_options](#).

Add plots

`addplot(plot)` provides a way to add other plots to the generated graph; see [G-3] [addplot_option](#).

Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] [twoway_options](#), excluding `by()`. 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](http://www.stata.com)

Cattell (1966) introduced scree plots, which are visual tools used to help determine the number of important components or factors in multivariate settings, such as principal component analysis and factor analysis; see [MV] [pca](#) and [MV] [factor](#). The scree plot is examined for a natural break between the large eigenvalues and the remaining small eigenvalues. The word “scree” is used in reference to the appearance of the large eigenvalues as the hill and the small eigenvalues as the debris

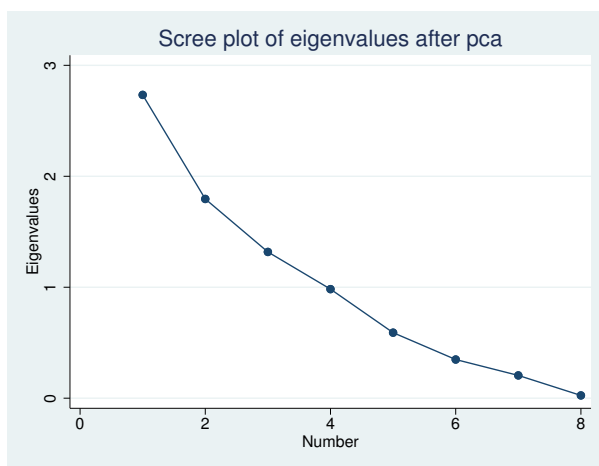
of loose rocks at the bottom of the hill. Examples of scree plots can be found in most books that discuss principal component analysis or factor analysis, including [Rabe-Hesketh and Everitt \(2007\)](#), [Hamilton \(1992, 249–288\)](#), [Rencher and Christensen \(2012\)](#), and [Hamilton \(2013, chap. 11\)](#).

► Example 1: Scree plots after principal component analysis

Multivariate commands, such as `pca` and `factor` (see [\[MV\] pca](#) and [\[MV\] factor](#)), produce eigenvalues and eigenvectors. The `screepplot` command graphs the eigenvalues, so you can decide how many components or factors to retain.

We demonstrate scree plots after a principal component analysis. Say that we have been hired by the restaurant industry to study expenditures on eating and drinking. We have data on 898 U.S. cities:

```
. use http://www.stata-press.com/data/r15/emd
(1980 City Data)
. pca ln_eat - hhsize
(output omitted)
. screepplot
```

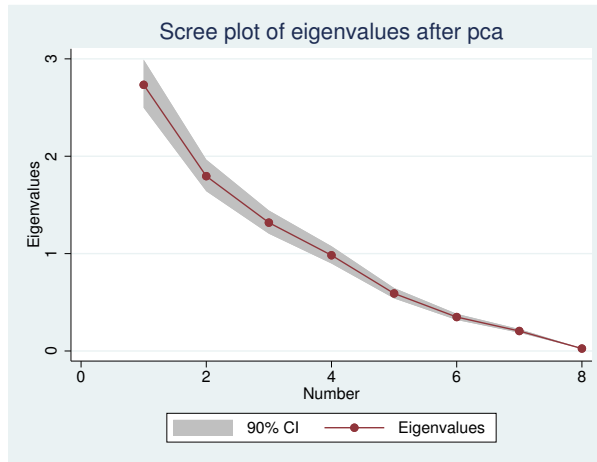


This scree plot does not suggest a natural break between high and low eigenvalues.

We render this same scree plot with the addition of confidence bands by using the `ci()` option. The `asymptotic` suboption selects confidence intervals that are based on the assumption of asymptotic normality. Because the asymptotic theory applied to correlation matrices is not fully correct, we also use the `level()` suboption to adjust the confidence interval. The `table` suboption displays a table of the eigenvalues and lower and upper confidence interval values.

```
. scree, ci(asympt level(90) table)
(caution is advised in interpreting an asymptotic theory-based confidence
interval of eigenvalues of a correlation matrix)
```

	eigval	low	high
Comp1	2.733531	2.497578	2.991777
Comp2	1.795623	1.640628	1.965261
Comp3	1.318192	1.204408	1.442725
Comp4	.9829996	.8981487	1.075867
Comp5	.5907632	.5397695	.6465744
Comp6	.3486276	.3185346	.3815635
Comp7	.2052582	.1875407	.2246496
Comp8	.0250056	.0228472	.027368

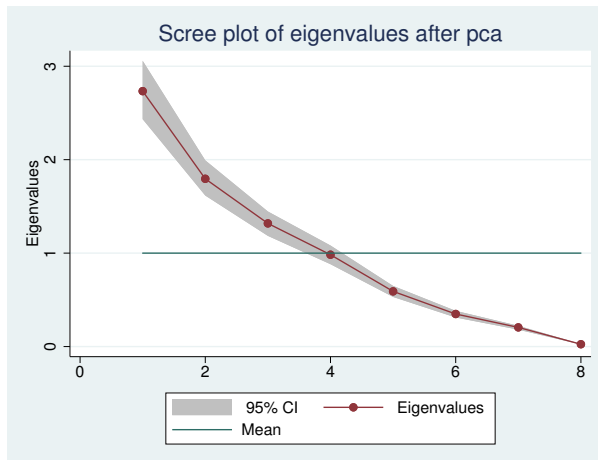


screepLOT warned us about using asymptotic confidence intervals with eigenvalues based on a correlation matrix. screepLOT knew that the eigenvalues were based on a correlation matrix instead of a covariance matrix by examining the information available in the `e()` results from the `pca` that we ran.

Instead of displaying an asymptotic confidence band, we now display a parametric bootstrap confidence interval. We select the `heteroskedastic` suboption of `ci()` that allows for unequal eigenvalues. The `reps()` and `seed()` suboptions control the execution of the bootstrap. The `mean` option provides a horizontal line at the mean of the eigenvalues, here at 1 because we are dealing with a principal component analysis performed on a correlation matrix.

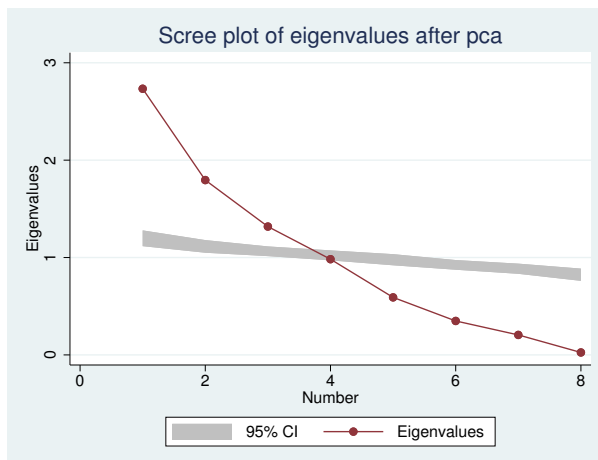
6 screeplot — Scree plot

```
. scree, ci(hetero reps(1000) seed(18228)) mean
```



Our final scree plot switches to computing the bootstrap confidence intervals on the basis of the assumption that the eigenvalues are equal to the mean of the observed eigenvalues (using the homoskedastic suboption of `ci()`). We again set the seed by using the `seed()` suboption.

```
. scree, ci(homo seed(56227))
```



The graph indicates that our data do not support using the homoskedastic assumption.

Stored results

screepLOT stores the following in `r()`:

Scalars

`r(level)` confidence level for confidence intervals

Macros

`r(Ctype)` correlation or covariance

`r(ci)` method for estimating confidence interval

`r(rngstate)` random-number state used for parametric bootstrap

Matrices

`r(ci)` confidence intervals

`r(eigvals)` eigenvalues

Scree is an English word with Old Norse origins. In geomorphology, scree denotes an accumulation of loose, coarse, usually angular debris at the foot of steep rock slopes (Luckman 2004). The accumulation must be sufficiently thick to be independent of the underlying slope; thus thin veneers of rock particles do not qualify. Screens are themselves quite steep; angles between 30° and 40° are common.

Outside geomorphology, and, notably, in some discussions of scree plots, the term is commonly misapplied to imply deposits on much gentler slopes, or those gentler slopes themselves, and by extension parts of curves thought similar to either.

However appropriate in general or in particular, the term *scree plot* is so widely used in multivariate statistics that it is echoed in the name of this Stata command.

References

- Cattell, R. B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research* 1: 245–276.
- 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.
- Luckman, B. 2004. Scree. In *Encyclopedia of Geomorphology*, ed. A. S. Goudie, 915–917. London: Routledge.
- Rabe-Hesketh, S., and B. S. Everitt. 2007. *A Handbook of Statistical Analyses Using Stata*. 4th ed. Boca Raton, FL: Chapman & Hall/CRC.
- Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis*. 3rd ed. Hoboken, NJ: Wiley.

Also see

- [MV] **factor** — Factor analysis
- [MV] **mds** — Multidimensional scaling for two-way data
- [MV] **pca** — Principal component analysis