Title

stata.com

spikeplot — Spike plots and rootograms

Description

`spikeplot` produces a frequency plot for a variable in which the frequencies are depicted as vertical lines from zero. The frequency may be a count, a fraction, or the square root of the count (Tukey's rootogram, circa 1965). The vertical lines may also originate from a baseline other than zero at the user's option.

Quick start

Spike plot of `v1`

```
spikeplot v1
```

As above, but apply frequency weights `wvar`

```
spikeplot v1 [fweight = wvar]
```

Plot proportions of the total number of observations instead of frequencies

```
spikeplot v1, fraction
```

Tukey's rootogram of `v2`

```
spikeplot v2, root
```

Menu

Graphics > Distributional graphs > Spike plot and rootogram

Syntax

```
spikeplot varname [if] [in] [weight] [, options]
```

options	Description
Main	
`round(#)`	round *varname* to nearest multiple of # (bin width)
`fraction`	make vertical scale the proportion of total values; default is frequencies
`root`	make vertical scale show square roots of frequencies
Plot	
spike_options	affect rendition of plotted spikes
Add plots	
`addplot(`*plot*`)`	add other plots to generated graph
Y axis, X axis, Titles, Legend, Overall, By	
twoway_options	any options documented in [G-3] ***twoway_options***

`fweights`, `aweights`, and `iweights` are allowed; see [U] **11.1.6 weight**.

Options

Main

`round(#)` rounds the values of *varname* to the nearest multiple of #. This action effectively specifies the bin width.

`fraction` specifies that the vertical scale be the proportion of total values (percentage) rather than the count.

`root` specifies that the vertical scale show square roots. This option may not be specified if `fraction` is specified.

Plot

spike_options affect the rendition of the plotted spikes; see [G-2] **graph twoway spike**.

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, By

twoway_options are any of the options documented in [G-3] ***twoway_options***. These include options for titling the graph (see [G-3] ***title_options***), options for saving the graph to disk (see [G-3] ***saving_option***), and the `by()` option (see [G-3] ***by_option***).

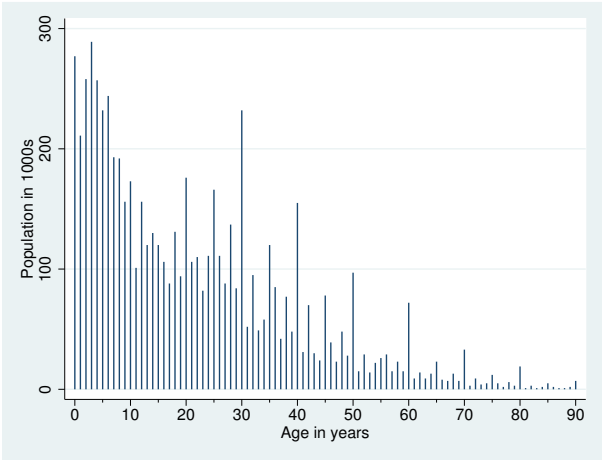
Remarks and examples

stata.com

Example 1

Cox and Brady (1997a) present an illustrative example using the age structure of the population of Ghana from the 1960 census (rounded to the nearest 1,000). The dataset has ages from 0 (less than 1 year) to 90. To view the distribution of ages, we would like to use each integer from 0 to 90 as the bins for the dataset.

```
. use http://www.stata-press.com/data/r14/ghanaage
. spikeplot age [fw=pop], ytitle("Population in 1000s") xlab(0(10)90)
> xmtick(5(10)85)
```



The resulting graph shows a "heaping" of ages at the multiples of 5. Also, ages ending in even numbers are more frequent than ages ending in odd numbers (except for 5). This preference for reporting ages is well known in demography and other social sciences.

Note also that we used the `ytitle()` option to override the default title of "Frequency" and that we used the `xlab()` and `xmtick()` options with *numlist*s to further customize the resulting graph. See [U] **11.1.8 numlist** for details on specifying *numlist*s.

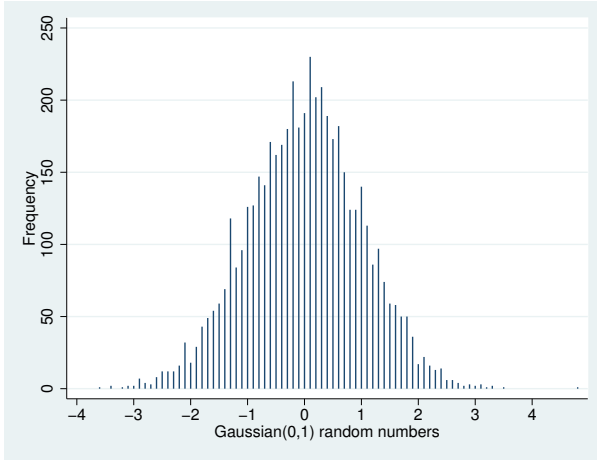
◁

Example 2

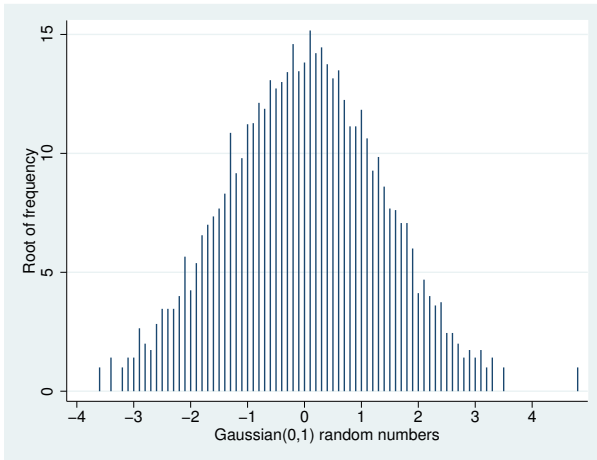
The rootogram is a plot of the square-root transformation of the frequency counts. The square root of a normal distribution is a multiple of another normal distribution.

```
. clear
. set seed 1234567
. set obs 5000
number of observations (_N) was 0, now 5000
. generate normal = rnormal()
. label variable normal "Gaussian(0,1) random numbers"
```

```
. spikeplot normal, round(.10) xlab(-4(1)4)
```




```
. spikeplot normal, round(.10) xlab(-4(1)4) root
```



Interpreting a histogram in terms of normality is thus similar to interpreting the rootogram for normality.

This example also shows how the `round()` option is used to bin the values for a spike plot of a continuous variable. ◁

▷ Example 3

spikeplot can also be used to produce time-series plots. *varname* should be the time variable, and weights should be specified as the values for those times. To get a plot of daily rainfalls, we type

```
. spikeplot day [w=rain] if rain, ytitle("Daily rainfall in mm")
```

The base() option of graph twoway spike may be used to set a different baseline, such as when we want to show variations relative to an average or to some other measure of level. ◁

Acknowledgments

The original version of spikeplot was written by Nicholas J. Cox of the Department of Geography at Durham University, UK, and coeditor of the *Stata Journal* and author of *Speaking Stata Graphics* and by Anthony R. Brady of the Imperial College School of Medicine (1997a, 1997b).

References

Cox, N. J., and A. R. Brady. 1997a. gr25: Spike plots for histograms, rootograms, and time-series plots. *Stata Technical Bulletin* 36: 8–11. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 50–54. College Station, TX: Stata Press.

——. 1997b. gr25.1: Spike plots for histograms, rootograms, and time-series plots: Update. *Stata Technical Bulletin* 40: 12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, p. 58. College Station, TX: Stata Press.

Tukey, J. W. 1965. The future of processes of data analysis. In *The Collected Works of John W. Tukey, Volume IV: Philosophy and Principles of Data Analysis: 1965–1986*, ed. L. V. Jones, 123–126. Monterey, CA: Wadsworth & Brooks/Cole.

Also see

[R] **histogram** — Histograms for continuous and categorical variables