graph twoway kdensity — Kernel density plots

Description

`graph twoway kdensity` plots a kernel density estimate for `varname` using `graph twoway line`; see [G-2] `graph twoway line`.

Quick start

Kernel density plot of `v1`
```
twoway kdensity v1
```
As above, use the biweight kernel function
```
twoway kdensity v1, kernel(biweight)
```
As above, but specify the half-width of the kernel to be 2
```
twoway kdensity v1, kernel(biweight) bwidth(2)
```
Overlay a kernel density plot on top of a histogram
```
twoway histogram v1 || kdensity v1
```
A separate graph area for each level of `catvar`
```
twoway kdensity v1, by(catvar)
```
A single graph area with curves showing the distribution for `catvar` = 0 and `catvar` = 1
```
twoway kdensity v1 if catvar==0 || kdensity v1 if catvar==1
```

Menu

Graphics > Twoway graph (scatter, line, etc.)
**Graph twoway kdensity — Kernel density plots**

### Syntax

```plaintext
twoway kdensity varname [if] [in] [weight] [ , options]
```

### options

<table>
<thead>
<tr>
<th>option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>bwidth(#)</code></td>
<td>smoothing parameter</td>
</tr>
<tr>
<td><code>kernel(kernel)</code></td>
<td>specify kernel function; default is <code>kernel(epanechnikov)</code></td>
</tr>
<tr>
<td><code>range(# #)</code></td>
<td>range for plot, minimum and maximum</td>
</tr>
<tr>
<td><code>range(varname)</code></td>
<td>range for plot obtained from <code>varname</code></td>
</tr>
<tr>
<td><code>n(#)</code></td>
<td>number of points to evaluate</td>
</tr>
<tr>
<td><code>area(#)</code></td>
<td>rescaling parameter</td>
</tr>
<tr>
<td><code>horizontal</code></td>
<td>graph horizontally</td>
</tr>
<tr>
<td><code>boundary</code></td>
<td>estimate density one <code>bwidth()</code> beyond maximum and minimum; not allowed with <code>range()</code></td>
</tr>
<tr>
<td><code>cline_options</code></td>
<td>change look of the line</td>
</tr>
<tr>
<td><code>axis_choice_options</code></td>
<td>associate plot with alternative axis</td>
</tr>
<tr>
<td><code>twoway_options</code></td>
<td>titles, legends, axes, added lines and text, by, regions, name, aspect ratio, etc.</td>
</tr>
</tbody>
</table>

### `kernel`

| `epanechnikov` | Epanechnikov kernel function; the default                         |
| `epan2`        | alternative Epanechnikov kernel function                          |
| `biweight`     | biweight kernel function                                          |
| `cosine`       | cosine trace kernel function                                      |
| `gaussian`     | Gaussian kernel function                                          |
| `parzen`       | Parzen kernel function                                             |
| `rectangle`    | rectangular kernel function                                       |
| `triangle`     | triangular kernel function                                        |

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

### Options

- `bwidth(#)` and `kernel(kernel)` specify how the kernel density estimate is to be obtained and are in fact the same options as those specified with the command `kdensity`; see [R] `kdensity`.
- `bwidth(#)` specifies the smoothing parameter.
- `kernel(kernel)` specifies the kernel-weight function to be used. The default is `kernel(epanechnikov)`.

See [R] `kdensity` for more information about these options.

All the other `graph twoway kdensity` options modify how the result is displayed, not how it is obtained.
range(#) and range(varname) specify the range of values at which the kernel density estimates are to be plotted. The default is range(m M), where m and M are the minimum and maximum of the varname specified on the graph twoway kdensity command.

range(#) specifies a pair of numbers to be used as the minimum and maximum.

range(varname) specifies another variable for which its minimum and maximum are to be used.

n(#) specifies the number of points at which the estimate is evaluated. The default is n(300).

area(#) specifies a multiplier by which the density estimates are adjusted before being plotted. The default is area(1). area() is useful when overlaying a density estimate on top of a histogram that is itself not scaled as a density. For instance, if you wished to scale the density estimate as a frequency, area() would be specified as the total number of nonmissing observations.

horizontal specifies that the result be plotted horizontally (that is, reflected along the identity line).

boundary specifies that the result be estimated for one bwidth() beyond the maximum and minimum value of varname. boundary cannot be specified with range().

cline_options specify how the density line is rendered and its appearance; [G-3] cline_options.

axis_choice_options associate the plot with a particular y or x axis on the graph; see [G-3] axis_choice_options.

twoway_options are a set of common options supported by all twoway graphs. These options allow you to title graphs, name graphs, control axes and legends, add lines and text, set aspect ratios, create graphs over by() groups, and change some advanced settings. See [G-3] twoway_options.

Remarks and examples

**Typical use**

**Use with by()**
Typical use

The density estimate is often graphed on top of the histogram:

```
. use http://www.stata-press.com/data/r15/lifeexp
   (Life expectancy, 1998)
. twoway histogram lexp, color(*.5) || kdensity lexp
```

Notice the use of `graph twoway histogram`’s `color(*.5)` option to dim the bars and make the line stand out; see [G-4] `colorstyle`.

Notice also the $y$ and $x$ axis titles: “Density/kdensity lexp” and “Life expectancy at birth/x”. The “kdensity lexp” and “x” were contributed by the `twoway kdensity`. When you overlay graphs, you nearly always need to respecify the axis titles using the `axis_title_options ytitle()` and `xtitle()`; see [G-3] `axis_title_options`. 
Use with by()

`graph twoway kdensity` may be used with `by()`:

```
. use http://www.stata-press.com/data/r15/lifeexp, clear
   (Life expectancy, 1998)
. twoway histogram lexp, color(*.5) || kdensity lexp ||, by(region)
```

References


Also see

[R] `kdensity` — Univariate kernel density estimation
[G-2] `graph twoway histogram` — Histogram plots