proportion.plot

Marco Del Giudice (2022). Version 1.2. Contact: marcodg@unm.edu

An R function that draws **proportion plots**, showing the respective percentages of the members of two groups across the distribution of a variable. Group proportions are intuitive, more visually readable than density ratios, and more robust to fluctuations in the data. This function can calculate group proportions from the means and SDs of two normal distributions, from the empirical distributions of two groups, or from idealized normal distributions with the same means and SDs as the empirical distributions. Empirical density functions are estimated with a Gaussian kernel, using the same bandwidth for the two groups. The distributions of the two groups are referenced to the mean and standard deviation of the total population. By default, the function assumes that the two groups are equally represented in the population (e.g., males and females). However, it is also possible to specify the proportions of the two groups in the population, or estimate them from the sample size of the two groups in the data supplied to the function.

Function proportion.plot

```
proportion.plot (x1, x2, bounds = 3, empirical = TRUE, bw.adjust = 1.5,
       pop.composition = NULL, colors = "default", xlab = "")
```

Arguments

x1, x2	either two vectors containing the empirical data for the two groups; or two vectors of length 2, each containing the mean and standard deviation of one group. In the latter case, group proportions are calculated from two normal distributions with the specified means and SDs.
bounds	a value that specifies the bounds for the calculation of the proportion functions. For example, bounds==3 (default) means that group proportions are calculated on the interval from -3 to +3 SDs from the population mean. The plot is adjusted depending on the value of bounds (from a minimum of 2 to a maximum of 4).
empirical	if empirical is TRUE (default) and x1 and x2 are data vectors, group proportions at different levels of the variable are calculated from the empirical distributions of the two groups. If empirical is FALSE, they are calculated from idealized normal distributions with the same means and SDs as the empirical distributions.
bw.adjust	the adjustment coefficient bw.adjust controls the bandwidth of the Gaussian kernel used to estimate group proportions. The common bandwidth for density estimation is determined by averaging the bandwidths for the two groups (selected with Silverman's rule of thumb); when estimating proportions, this bandwidth is multiplied by bw.adjust to stabilize the proportion functions (default is 1.5). If desired, users can specify a different adjustment factor (e.g., bw.adjust=1).
pop.composition	this argument can be used to specify the representation of the two groups in the population. It can be a vector of length 2 containing two numbers proportional to the frequencies of the groups; e.g., pop.composition=c(10, 90) means that

	the first group (`x1`) accounts for 10% of the population, while the second group (`x2`) accounts for the remaining 90%. Alternatively, frequencies can be estimated from the sample sizes of `x1` and `x2`, by setting `pop.composition="data"`. If `pop.composition` is NULL (default), the groups are assumed to be equally represented in the population.
`colors`	colors of the lines in the plot. The "default" color scheme uses blue for `x1` and red for `x2`. The "sexdiff" color scheme uses light blue for `x1` (males) and pink for `x2` (females). Alternatively, users can specify a vector of length 2 with the colors for `x1` and `x2`, respectively.
`xlab`	a label for the X axis of the plot, to name the variable under consideration (empty by default).

Plots

The proportion plot shows the respective percentages of members of the two groups at different levels of the variable under consideration. The bottom X axis shows the actual values of the variable; the top X axis shows the corresponding population standard deviations from the population mean (labeled M). Percentage lines are superimposed on density curves showing the distribution of the variable in the two groups. If `x1` and `x2` are data vectors and `empirical` is TRUE, the plot is based on the empirical distributions of the two groups; otherwise, the plot is based on idealized normal distributions. If based on empirical distributions, the plot includes dotted lines showing the minimum and maximum values of the variable.