

This reproducibility package contains data and R code for the figures in our EJM technical spotlight:

Beyond differences in means: robust graphical methods to compare two groups in neuroscience

Guillaume A. Rousselet, Cyril R. Pernet, Rand R. Wilcox
European Journal of Neuroscience (in press)

The accepted version of the article is also available on [biorxiv](#).

If you have any question or suggestion, don't hesitate to get in touch:

Guillaume.Rousselet@glasgow.ac.uk

Also, please do get in touch if you use the tools described in the article. We would love to be able to showcase examples in our teaching, in workshops and in other publications. More importantly, we'd like to know the article is helping to improve standards in neuroscience and other disciplines.

There is an R script for each figure. To run the scripts, set your working directory to the `rcode` folder, or place the content of that folder in your working directory. That folder also contains example datasets. Matlab code implementing the main functions is also available.

The datasets are licensed CC0, but please cite the original authors if you re-use them.

The R scripts are also licensed CC0. If you use large sections of the code, consider citing the EJM article.

The main analyses are done using the `rogme` R package (MIT license), which can be installed using these commands:

```
install.packages("devtools")
devtools::install_github("GRousselet/rogme", ref = "v0.1.0-alpha.ejm")
```

This particular version of the package was used to create the examples in the article.

You can get the latest version of the package here:

```
devtools::install_github("GRousselet/rogme")
```

`rogme` is an adaptation of code originally developed by Rand Wilcox, and available in the `WRS` package or as a text file - see details and updates of his code here:

<http://dornsife.usc.edu/labs/rwilcox/software/>

Dependencies

To run the R scripts you will need these packages:

`ggplot2`

cowplot

readr

retimes

tidyr

ggbeeswarm

royme will prompt you if you need to install other packages.

Abbreviations

pb = percentile bootstrap

ci = confidence interval

se = standard error

R functions

All the functions rely on the [Harrell-Davis quantile estimator](#), computed by the `hd()` function.

Shift function

The original shift function approach (Doksum & Sievers 1976 - see paper) is implemented in the functions `sband()` and `wband()` in the `WRS` R package.

In the same package, the shift function for deciles using *pb* *se* estimation is computed using:

- `shifthd()` for independent groups
- `shiftdhd()` for dependent groups

You can access these functions by using the file `Rallfun-v33.txt` in the `rcode` folder. For that, type in the R console:

```
source("\rcode\Rallfun-v33.txt")
```

Shift function for deciles using *pb* *ci* without *se* estimation:

- `qcomhd()` for independent groups
- `Dqcomhd()` for dependent groups

In `rogme`, the shift function can be calculated using:

- `shifthd()` or `shifthd_pbc()` for independent groups
- `shiftdhd()` or `shiftdhd_pbc()` for dependent groups

You can see the shift function in action [here](#) and [here](#).

Limitations To make the shift function even more useful, it would be worth quantifying how its slope and non-linear patterns relate to differences in variance and skewness between groups. Another important development would be to consider a hierarchical approach, in which shift functions and difference asymmetry functions are built for each participant or condition, and are then averaged across participants to make group inferences. Such inferences could also be made on features of the functions, for instance on the size of the shift, the slope of the function, or the location of a non-linearity in the function. The shift function can also be extended to a more general ANOVA framework, to consider various linear contrasts, including interaction terms. More complex designs will be challenging to handle, but powerful tools are being developed to visualise large multivariate datasets, and explore them systematically, notably with the goal of combining exploratory data analysis with rigorous inferential statistics (Cook et al. 2016).

Cook, D., Lee, E.K. & Majumder, M. (2016)
Data Visualization and Statistical Graphics in Big Data Analysis.
Annu Rev Stat Appl, 3, 133–159.
<https://github.com/dicook/Annual.Reviews.paper>

Difference asymmetry function

Rand Wilcoxon's version:

- `qwmwhd()` for independent groups
- `difqpci()` for dependent groups

In `rogme`, these functions have been renamed:

- `asymhd()` for independent groups
- `asymdhd()` for dependent groups
- `plot_diff_asym()` to plot the results

You can see the difference asymmetry function in action [here](#). The difference asymmetry function is a new approach to quantify asymmetries in difference distributions. To understand this approach, we first need to consider how difference scores are usually characterised. It helps to remember that for continuous distributions, the Mann–Whitney–Wilcoxon (MWW) U statistics is the sum of the number of times observations in group X are larger than observations in group Y. Concretely, this calculation requires to compute all pairwise differences between X and Y, and then count the number of positive differences. And the MWW test assesses $P(X > Y) = 0.5$. Essentially, the MWW test is a non-parametric test of the hypothesis that the distributions are identical. The MWW test does not compare the medians of the marginal distributions as often stated; it is based on an estimate of the median of the difference scores, but it uses the wrong standard error when distributions differ (Cliff, 1996). A more powerful test is Cliff's delta, which uses $P(X > Y) - P(X < Y)$ as a

measure of effect size. Wilcox (2012)'s approach is an extension of the MWW test: the idea is to get a sense of the asymmetry of the difference distribution by computing a sum of quantiles = $q + (1-q)$, for various quantiles estimated using the Harrell-Davis estimator.

Cliff, N. (1996) Ordinal methods for behavioral data analysis. Erlbaum, Mahwah, N.J.
Wilcox, R.R. (2012) Comparing Two Independent Groups Via a Quantile Generalization of the Wilcoxon-Mann-Whitney Test. Journal of Modern Applied Statistical Methods, 11, 296–302.

One sample estimation

- `quantiles_pbci()` to compute pb ci of several quantiles of one distribution
- `plot_dec_ci()` to plot the results

Matlab functions

The Matlab code is available in the folder `matlabcode`. The same functions are also available on [github](#), and some of them have been described [here](#). Updated and new Matlab functions will be posted to github.

Shift function

Shift function for deciles using *pb se* estimation:

- `shiftd` for independent groups
- `shiftdhd` for dependent groups

Shift function for deciles using *pb ci* without se estimation:

- `shiftd_pbci` for independent groups
- `shiftdhd_pbci` for dependent groups

In addition to the plot option available for each function, the four types of Matlab shift functions can be plotted using `shift_fig` (available in the `matlabcode` folder and [here](#)).

Difference asymmetry function

- `diffall_asym` for independent groups
- `diff_asym` for dependent groups

One sample estimation

- `decilesci` to compute *ci* of the deciles of one distribution, with *pb* estimation of the se of the deciles.
- `decilespbci` to compute *pb ci* of the deciles of one distribution, without se estimation.

Other R resources

There are many R functions and packages implementing very useful data visualisation techniques. Here are some of our favourites:

R package	Functionality	Link
ggplot2 extensions	extends ggplot2 capabilities	https://www.ggplot2-exts.org
ggbeeswarm	beeswarm scatterplots	https://github.com/eclarke/ggbeeswarm
yarr	superimposed scatter & density plots + inferences for base R (aka pirate plots)	https://github.com/ndphillips/yarr
beeswarm	beeswarm plots for base R	http://www.cbs.dtu.dk/~eklund/beeswarm/
Tufte in R	collection of examples in base R, lattice & ggplot2 to replicate visualisations introduced by Edward Tufte	http://motioninsocial.com/tufte/

For more complex datasets, this article lists a lot of resources and R packages:

Cook, D., Lee, E.K. & Majumder, M. (2016)
Data Visualization and Statistical Graphics in Big Data Analysis.
Annu Rev Stat Appl, 3, 133–159.
<https://github.com/dicook/Annual.Reviews.paper>

Other Matlab toolboxes

Again, there are many Matlab options, so we only flag a few here:

- <https://github.com/piermorel/gramm>
- <http://neurostatscyrilpernet.blogspot.co.uk/2016/08/show-me-data.html?m=1>