

20-March-2017

Below are step-wise descriptions of an example data set and Program R code for generating confidence intervals around means.

- Email questions, comments, and ideas on how to improve either the analysis itself or the methodological description to Michael Venarsky ([mvenarsky@gmail.com](mailto:mvenarsky@gmail.com)). Meaningful contributions will be credited to the person.

### Data files

Spider\_v1.csv description:

- SITE = sampling site
- DATE = sampling date
- HABITAT = type of habitat samples. Pool is depositional habitat dominated by fine substrates. Riff is erosional habitat dominated by large cobbles and boulders. Wood is sticks and logs submerged in the stream.
- Each line is a replicate. Note that replicates are NOT uniquely identified (e.g., Pool-1, Pool-2, etc.). Doing this will cause the code to return a “resampled data table” that is identical to the original data set.
- M\_HABITAT = riparian spider biomass (mg dry mass m<sup>-1</sup> stream bank)
- M\_VALLEY = riparian spider biomass (mg dry mass m<sup>-1</sup> stream valley)

### Bootstrapping data

Bootstrap\_CI\_v1.R:

Pre-analysis setup:

- Check code 20-March-2017
- Constructed using R version 3.3.1 (2016-06-21) -- "Bug in Your Hair". Copyright (C) 2016 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit).
- Michael Venarsky wrote code using dplyr, tidyr, lubridate, and zoo packages on 28-March-2015. Several stackoverflow.com users provided assistance to Michael Venarsky. See questions by stackoverflow user Vesuccio.
- For those not familiar with dplyr or tidyr, consider loading the swirl package which is an interactive tutorial for dplyr and tidyr implemented in R studio.
- Links to dplyr and tidyr cheat sheets:
  - <http://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>
  - <https://www.rstudio.com/resources/cheatsheets/>

- Include three folders your working directory: data, results, check. CSV files will be written to each of these folders. Note that CSV files are written to the “check” folder only when you delete the “#” at the beginning of the line of code.
- Clear all data from previous runs, load required packages, set working directory and import all data files.

#### Step 1

- `Resample_function_biomass` resamples data set with replacement.

#### Step 2

- `N_resample` defines the number of resampling events. We conducted 1000.
- Loop 1 begins by running the `Resample_function_biomass` function.

#### Step 3

- Calculate average riparian spider biomass per site per habitat. Units are mg dry mass. Note that DATE was not included in this analysis.

#### Step 4

- `N_resample` defines the number of times Loop 1 is run. We conducted 1000. If the `N_resample` is 1, the output from the analyses are exported. If the `N_resample` is >1, then the results from each resampling are sequentially assembled into a table and the exported as a CSV file.

#### Step 5

- Calculate 95% quantiles.