

Caffeine Multispecies : Data Analysis

philippe.stocker42@gmail.com
philippe.stocker@univ-rouen.fr

September 19th 2023

1 Statistical Approach

1.1 Data Collection and Curation

The study employed a comprehensive multi-biomarker strategy to evaluate the ecotoxicological effects of caffeine on four marine organisms: *Pachygrapsus marmoratus*, *Magallana gigas*, *Littorina littorea*, and *Actinia equina*. A dataset was meticulously compiled to focus on these specific species and was categorized by variables such as batch, species, tissue types, and temporal intervals.

1.2 Statistical Software

Data analysis was conducted using R software (version 4.0.5), taking advantage of packages including tidyverse for data manipulation, ggplot2 for data visualization, broom for tidying up statistical outputs, knitr for dynamic report generation, and ggpvr for enhancing ggplot2 plots. Spatial data regarding the sample locations were mapped using QGIS software version 3.0.2.

1.3 Data Treatment

Prior to any statistical inference, tests for normality and homogeneity of variance were performed using the Shapiro-Wilk test and Bartlett's test, respectively. Data were presented as mean \pm standard deviation within each stratified group (i.e., batch, species, tissue types, and temporal intervals).

1.4 Normalization

Mean and standard deviation values of each biomarker were normalized between 0 and 1 within species groups to allow for inter-species comparisons.

1.5 Statistical Tests

We used Welch's t-test to assess the significance of differences between concentrations and control groups for each biomarker. This method was chosen for its robustness against unequal variances and sample sizes, which are commonly encountered in ecotoxicological studies.

1.6 Significance Levels

Differences were considered statistically significant when the p-value was less than 0.05. To facilitate interpretation, p-values were converted into asterisks for significance representation on plots. These annotations were positioned above the standard deviation bars.

1.7 Data Presentation

Visual plots were constructed using ggplot2, with asterisks (or 'NS' for non-significant results) placed just above the top of the standard deviation bars to indicate the level of significance.