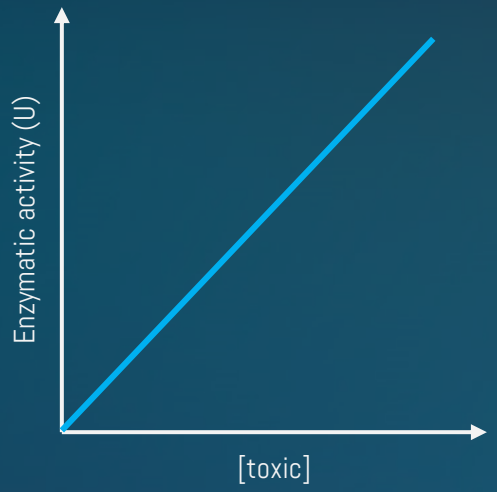
04
STATISTICS IN ECOTOXICOLOGY

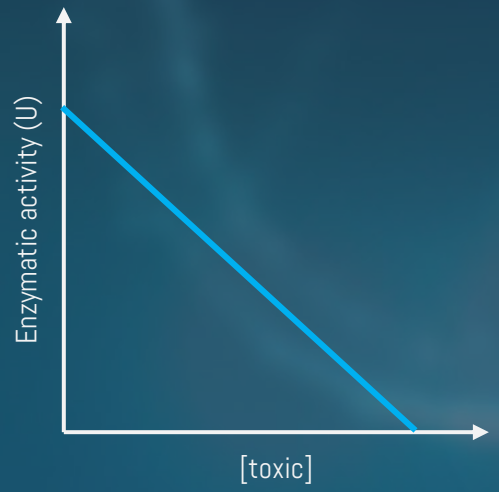
Uni- and multi- variate statistics in ecotoxicology
Meta-analysis
Indexes

04 STATISTICS IN ECOTOXICOLOGY : LINEAR AND NON-LINEAR REGRESSIONS



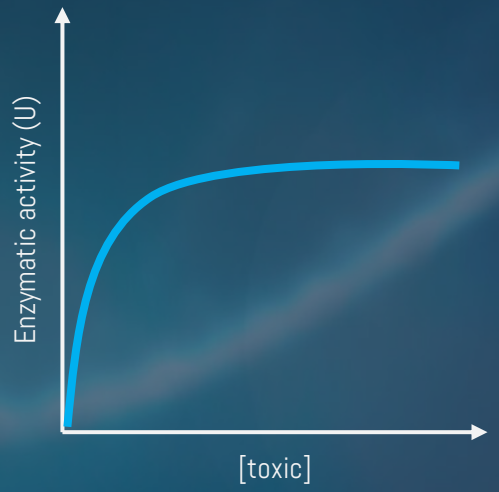
DIRECT LINEAR COORELATION

This would be the perfect biomarker. Its activity is proportional to the concentration of the toxic substance.



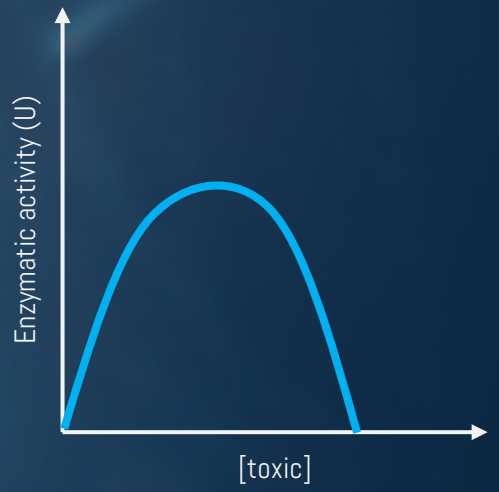
INVERSE LINEAR COORELATION

The activity of some enzymes can be impaired by certain toxic substances.



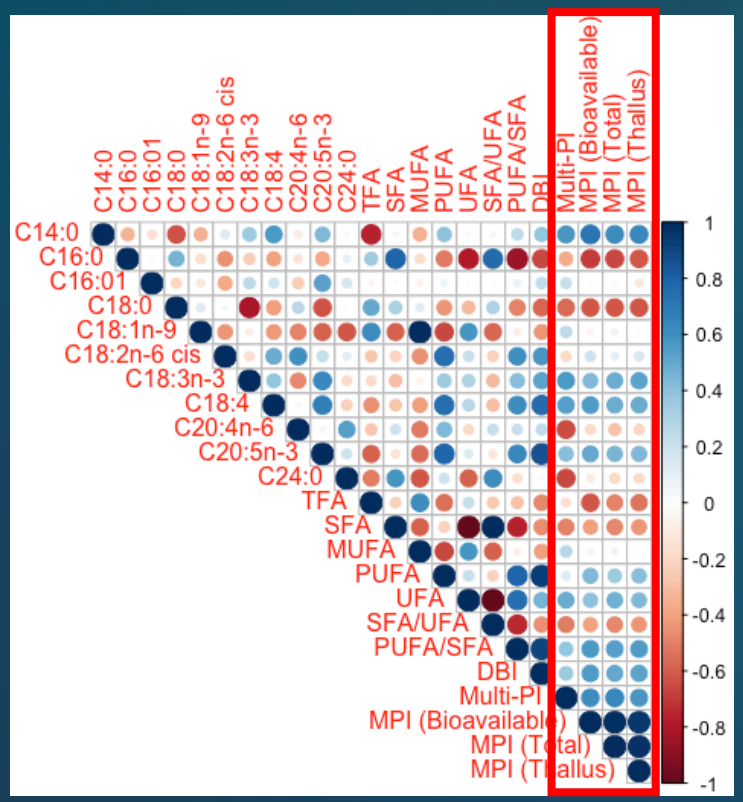
EXPONENTIAL

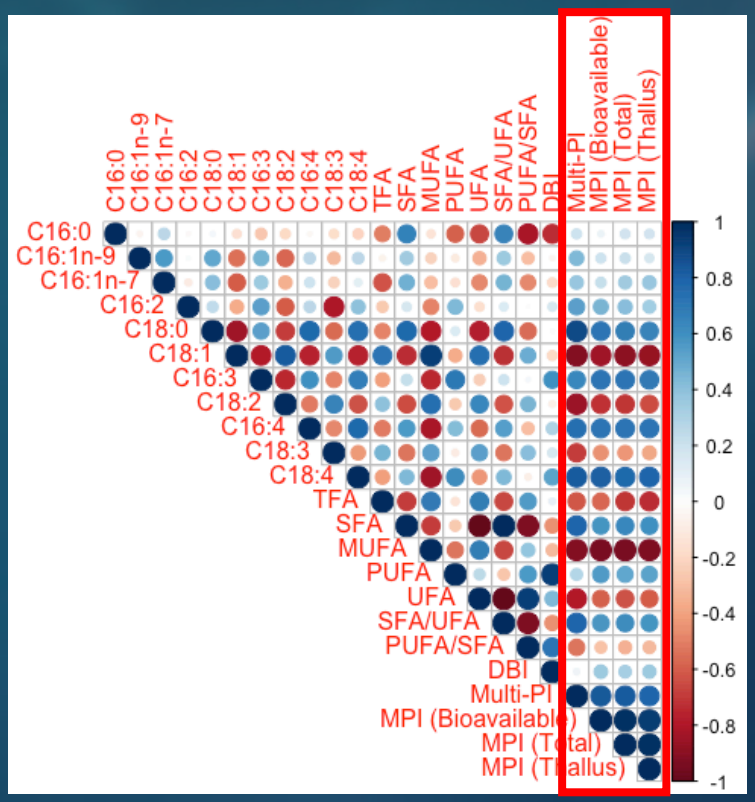
The enzyme increases its activity in response to stress until a maximum velocity after which increasing toxic concentrations do not produce any effect.



PARABOLIC FUNCTION

The enzyme exhibits an increase in its activity until a certain concentration after which it is inhibited. The activity should be only used as biomarker in the positive slope range of concentrations.

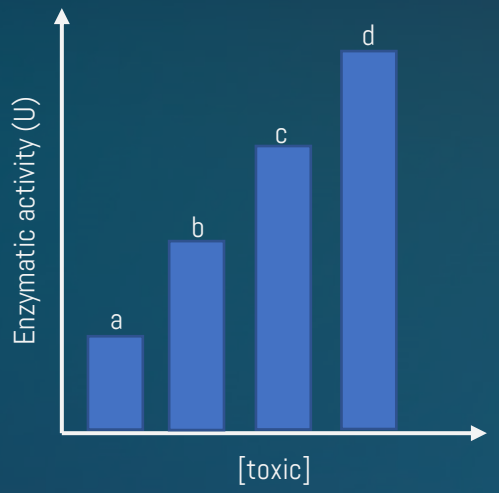




LINEAR CORRELATIONS

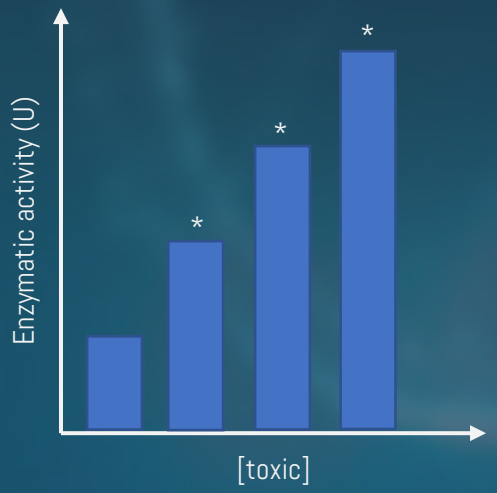
- Ideally the exogenous concentration and the internal organism concentrations has a direct correlation.
- Different species have different physiological behaviors, due to their intrinsicic resistance and tolerance traits;
- The same biochemical trait can respond to differently (inhibited or enhanced) in different species.
- Not all species have the exact same biomarkers.

04 STATISTICS IN ECOTOXICOLOGY : UNIVARIATE COMPARISIONS

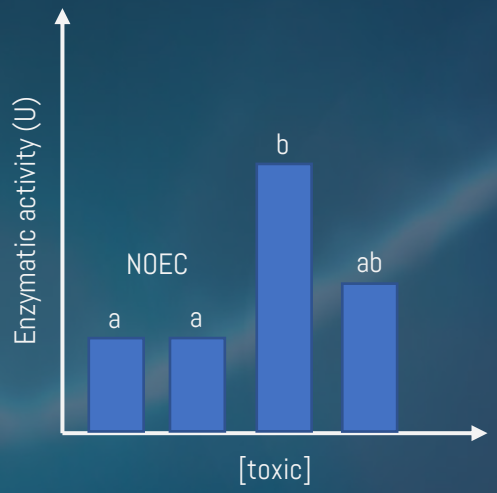


BETWEEN EFFECTS AND COMPARISON TOWARDS THE CONTROL

To evaluate incremental responses of the applied toxic towards all the tested concentrations (pair-wise comparisions).

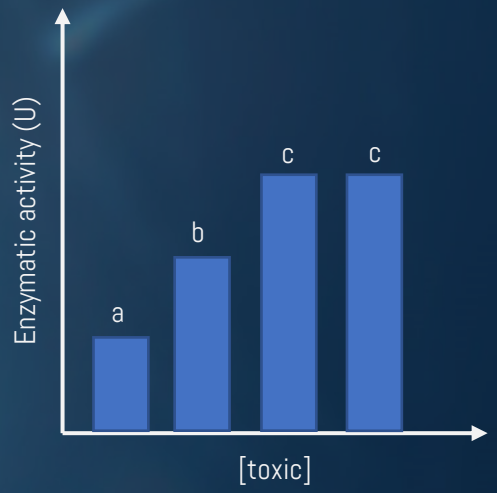


Classical ecotoxicology, comparison towards the control only.



NON-MONOTONIC

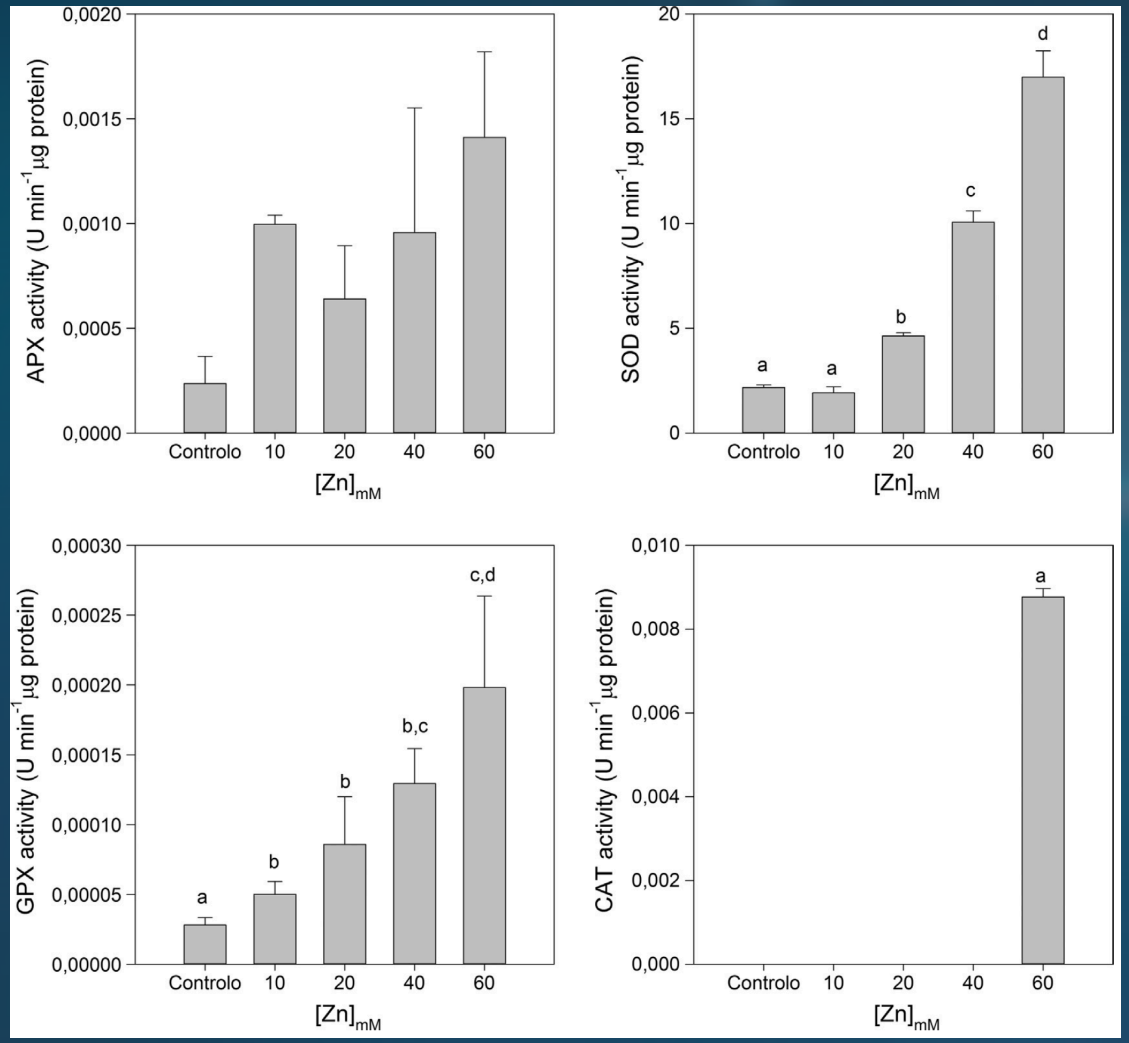
Linear increase until a toxicity threshold after which there is inhibition of activity or decrease of the biomarker concentration.



EXPONENTIAL

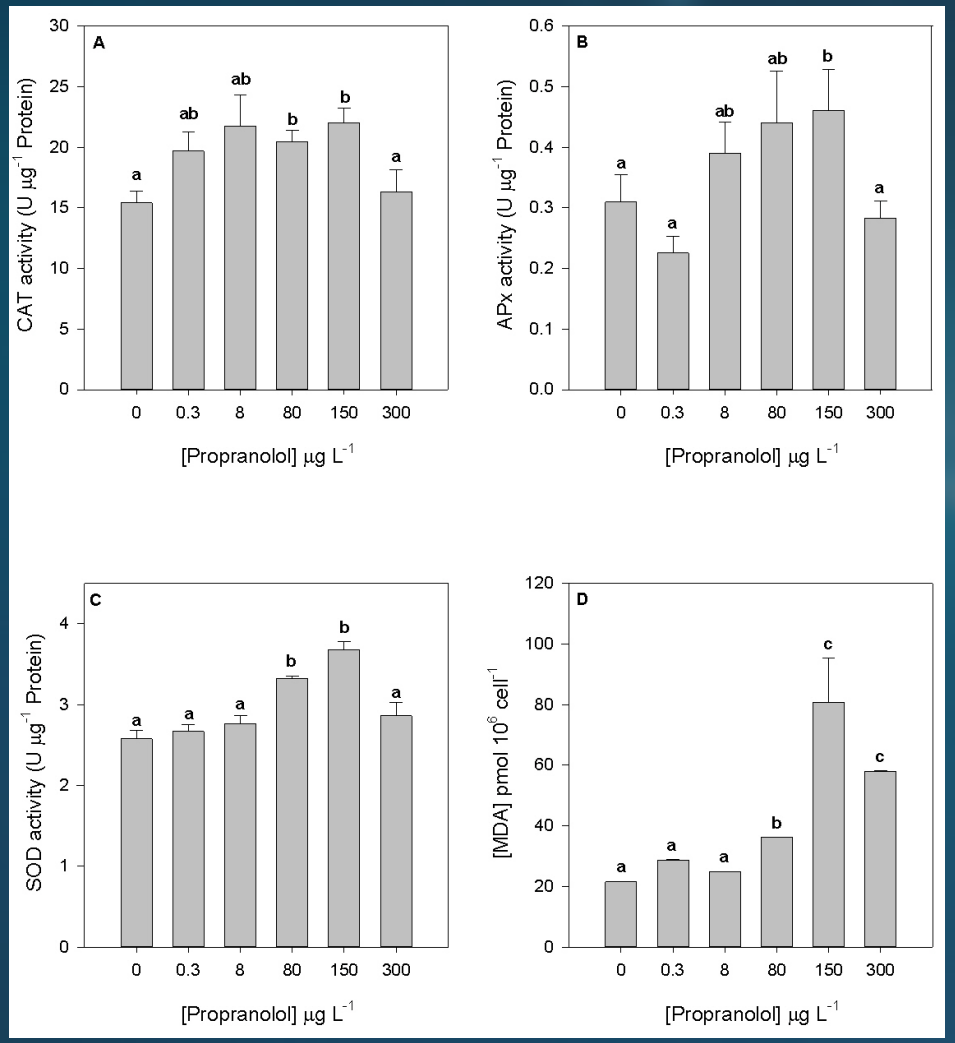
Linear increase until a toxicity threshold after there is a stabilization (saturation) and the biomarker stops responding.

04 STATISTICS IN ECOTOXICOLOGY : UNIVARIATE COMPARISIONS



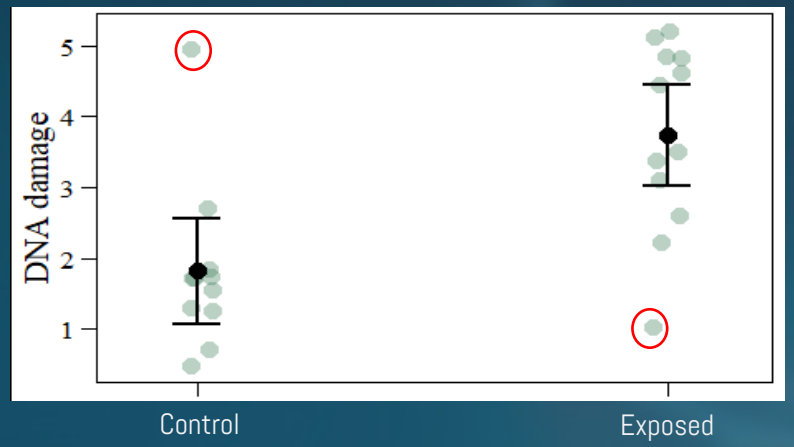
IDEALLY...

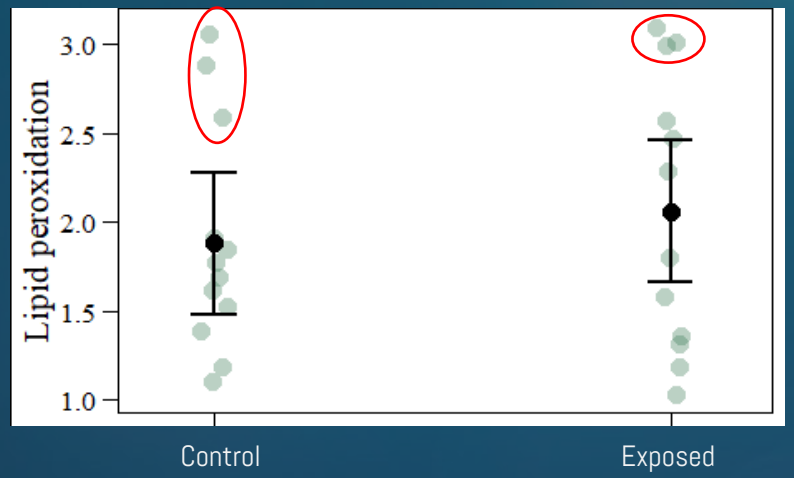
- Several biomarkers present a clear tendency.
- Due to the intrinsic variation some may not exhibit statistical differences.
- Some biomarkers might even only be activated at almost lethal doses.



REALITY MOST OF THE TIMES

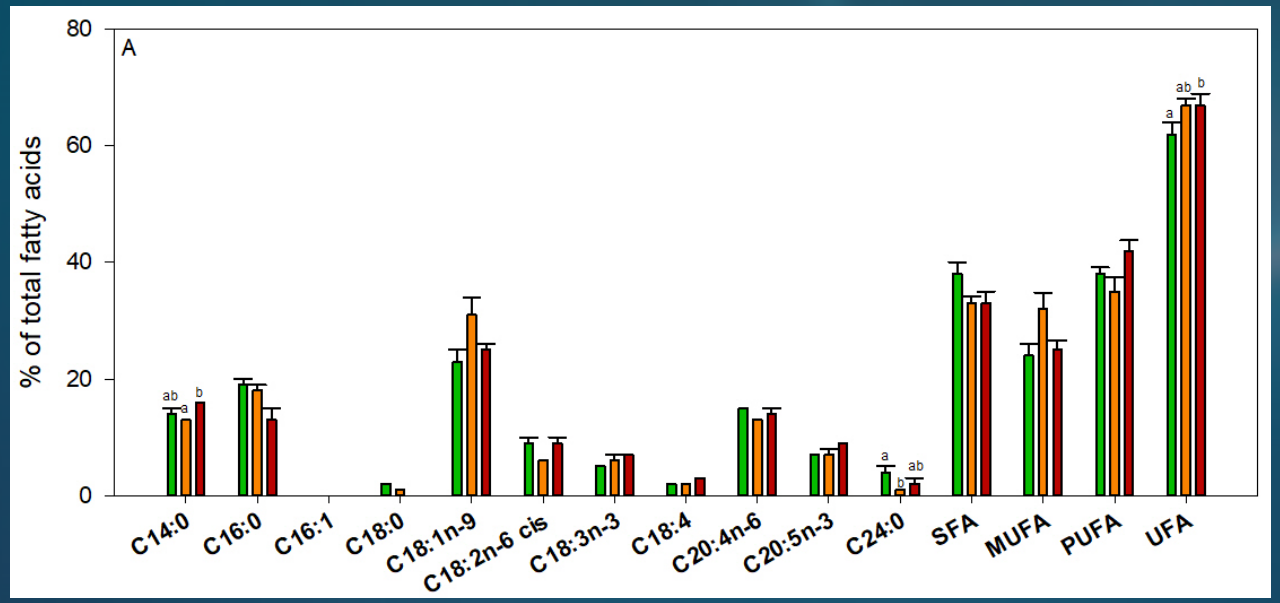
- Not all biomarkers have a monotonic evident pattern.
- Different biomarkers have completely different behaviours towards a same toxic.
- While some have biphasic responses other can have linear tendencies.





OUTLIER ASSESSMENT

- Non-clonal organisms have a high degree of variability.
- Operator performance and complex protocols can also introduce a degree of variance (besides biological replicates, technical analytical replicates are also recommended).
- Its important to evaluate the existence of severe or moderate outliers.

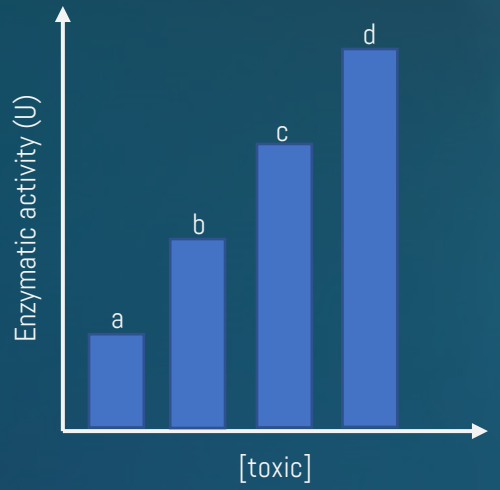


APPARENT NO EFFECTS

- Some biochemical traits that result from metabolic transformations with a significant co-variation can fail to reveal univariate statistical differences.
- These result from the insignificant and discrete differences between metabolites, that individually do not reflect significant differences.
- In this cases multivariate statistics can be a useful tool.

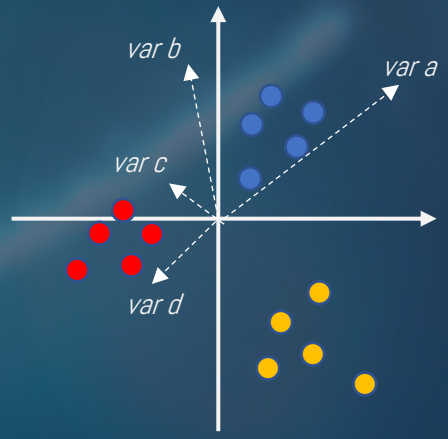
04 STATISTICS IN ECOTOXICOLOGY : UNIVARIATE VERSUS MULTIVARIATE STATISTICS

UNIVARIATE



Univariate is a term commonly used in statistics to describe a type of data which consists of observations on only a single characteristic or attribute. A simple example of univariate data would be the salaries of workers in industry. Like all the other data, univariate data can be visualized using graphs, images or other analysis tools after the data is measured, collected, reported, and analysed.

MULTIVARIATE

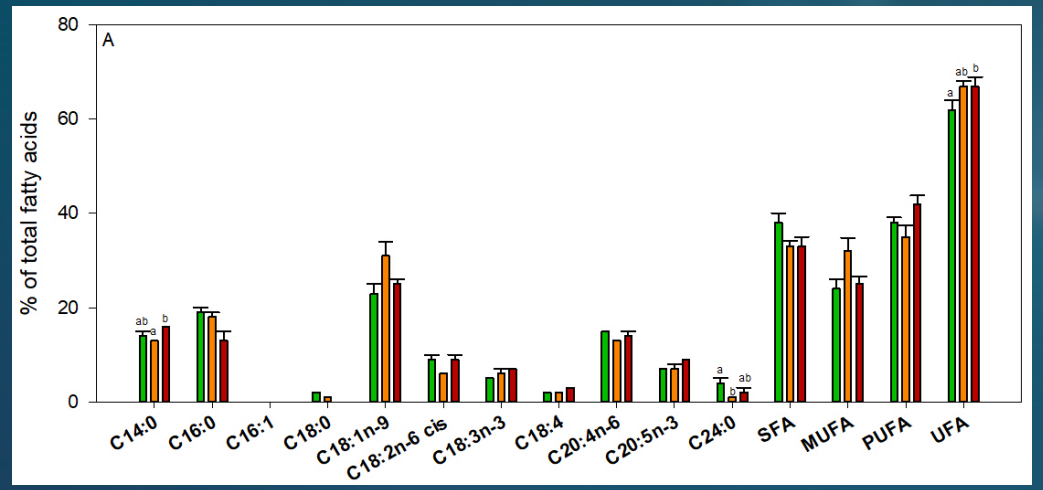


Multivariate statistics is a subdivision of statistics encompassing the simultaneous observation and analysis of more than one outcome variable. Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical application of multivariate statistics to a particular problem may involve several types of univariate and multivariate analyses in order to understand the relationships between variables and their relevance to the problem being studied.

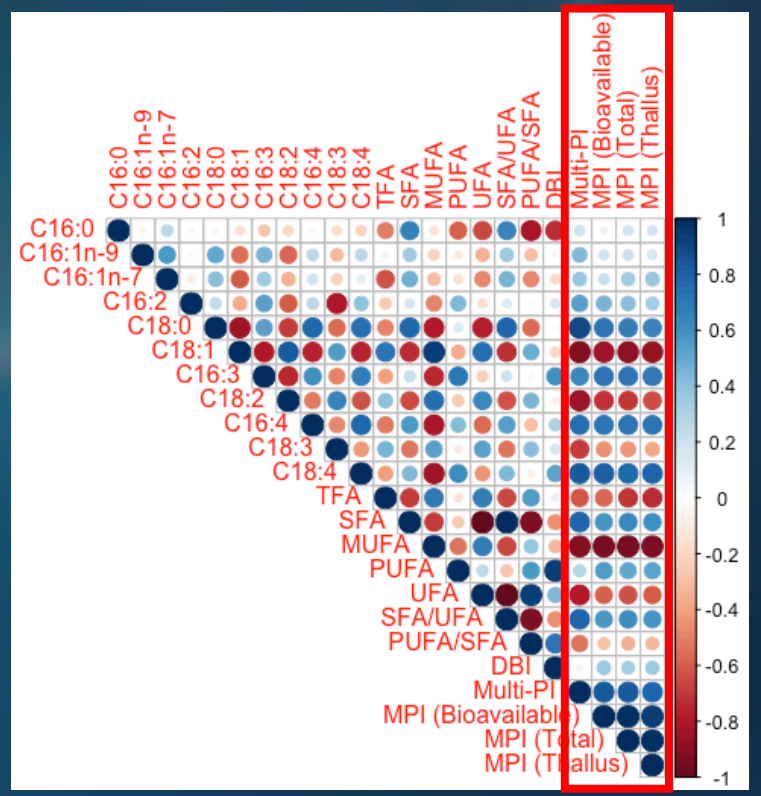
04 STATISTICS IN ECOTOXICOLOGY : MULTIVARIATE STATISTICS

- **Multivariate analysis of variance (MANOVA)** extends the analysis of variance to cover cases where there is more than one dependent variable to be analysed simultaneously.
- **Multivariate regression** attempts to determine a formula that can describe how elements in a vector of variables respond simultaneously to changes in others.
- **Principal components analysis (PCA)** creates a new set of orthogonal variables that contain the same information as the original set. It rotates the axes of variation to give a new set of orthogonal axes, ordered so that they summarize decreasing proportions of the variation.
- **Factor analysis** is similar to PCA but allows the user to extract a specified number of synthetic variables, fewer than the original set, leaving the remaining unexplained variation as error.
- **Canonical correlation analysis** finds linear relationships among two sets of variables;
- **Redundancy analysis (RDA)** is similar to canonical correlation analysis but allows the user to derive a specified number of synthetic variables from one set of (independent) variables that explain as much variance as possible in another (independent) set.
- **Correspondence analysis (CA)**, finds a set of synthetic variables that summarise the original set, assuming chi-squared dissimilarities among records (cases).
- **Multidimensional scaling** comprises various algorithms to determine a set of synthetic variables that best represent the pairwise distances between records.
- **Discriminant analysis**, attempts to establish whether a set of variables can be used to distinguish between two or more groups of cases.
- **Linear discriminant analysis (LDA)** computes a linear predictor from two sets of normally distributed data to allow for classification of new observations.
- **Clustering systems** assign objects into groups (called clusters) so that objects (cases) from the same cluster are similar to each other.
- Recursive partitioning creates a **decision tree** that attempts to correctly classify members of the population based on a dichotomous dependent variable.
- **Artificial neural networks** extend regression and clustering methods to non-linear multivariate models.

NO APPARENT DIFFERENCES

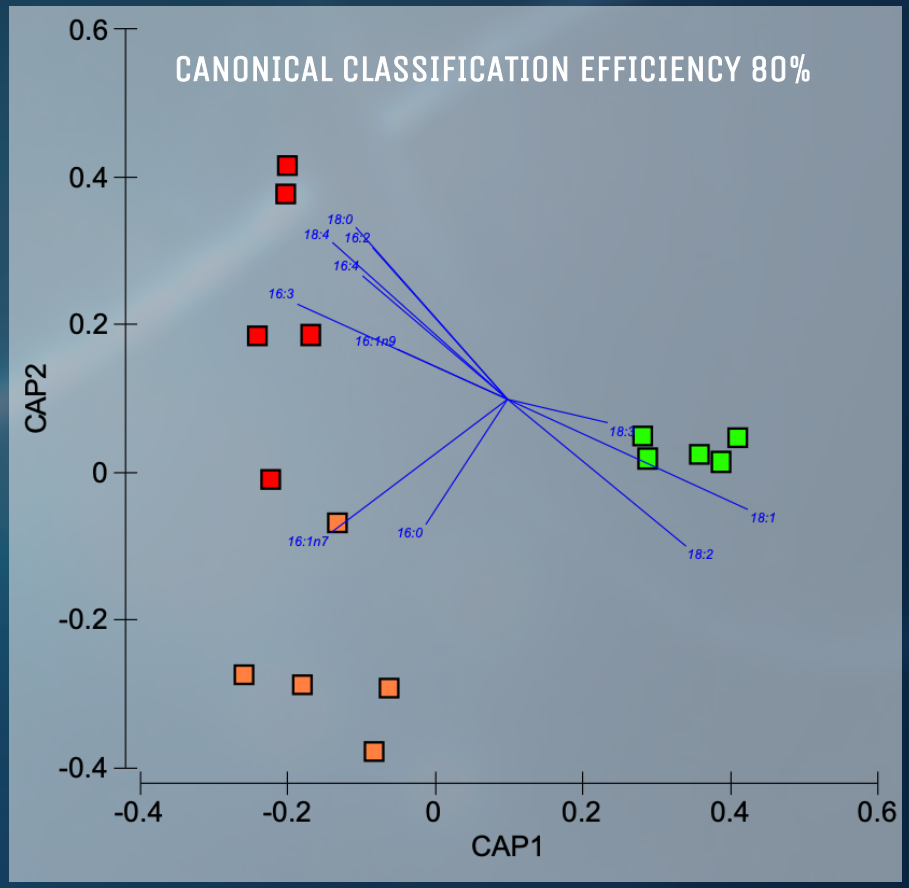


CORRELATION ANALYSIS REVEAL SOME TENDENCIES



DISCLOSING NO APPARENT NO EFFECTS

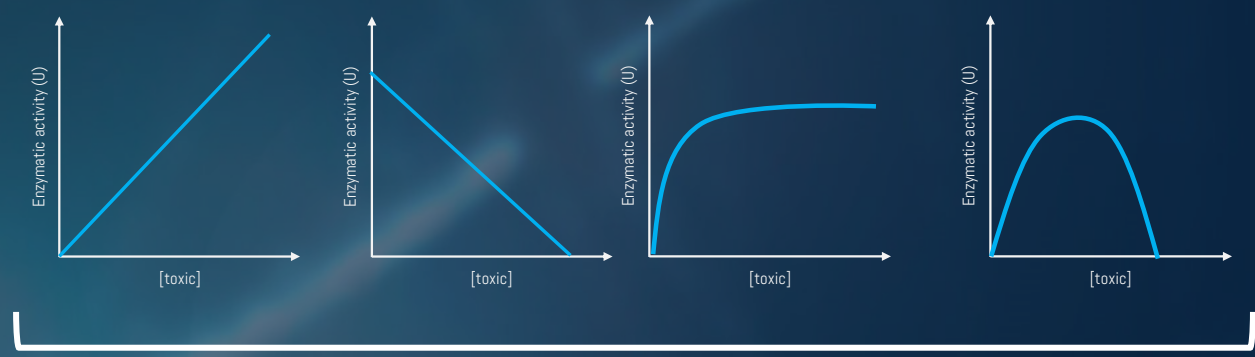
- Called "CAP" for "Canonical Analysis of Principal coordinates," this method will allow a constrained ordination to be done on the basis of any distance or dissimilarity measure.
- Canonical tests using permutations are also given, and we show how the method can be used
    - place a new observation into the canonical space using only interpoint dissimilarities,
    - to classify observations and obtain misclassification or residual errors, and
    - to correlate the original variables with patterns on canonical plots.
- Misclassification error or residual error is used to obtain a non-arbitrary decision concerning the appropriate dimensionality of the response data cloud (number of PCO axes) for the ensuing canonical analysis.



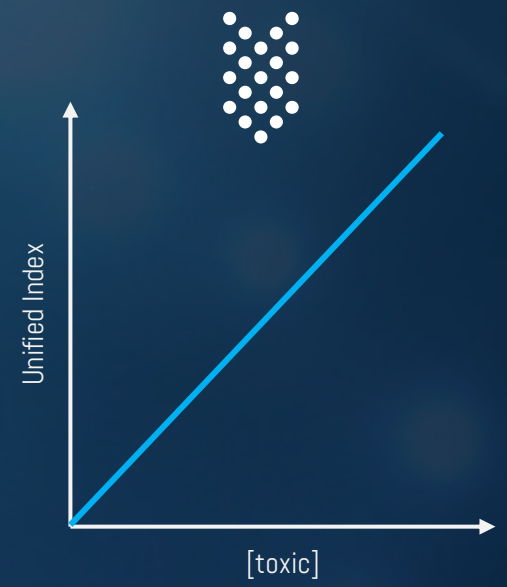
Orig. group	Class 1	Class 2	Class 3	Total	%correct
Class 1	4	1	0	5	80
Class 2	1	4	0	5	80
Class 3	0	1	4	5	80

04 STATISTICS IN ECOTOXICOLOGY : MULTIVARIATE INDEXES

- Indexes aim to integrate a wide number of variables into a single numeric value.
- Advantages: easier communication to non-specialist audience of the ecotoxicological results attained, summarized into scaled values (for e.g. 1-0).
- Disadvantages: loss of physiological meaning.
- Variables must be normalized and weighted so that the mathematical value of a variable does not interfere just by having a different order of magnitude.
- Requires normally final index weighting using repetitive mathematical function (sin, cos for e.g.), power to an inverse number or logarithmic scales.
- Variable selection and weighting can be preformed having as basis a statistical approach or empirical knowledge.



Set of individual biomarkers



04 STATISTICS IN ECOTOXICOLOGY : MULTIVARIATE INDEXES

NON-SELECTIVE METHOD — BIOEFFECT ASSESSMENT INDEX (BAI)

ALL BIOMARKERS ARE INCLUDED AND SCORED ACCORDING TO THEIR QUARTIL POSITION

WEIGHTED INDEX — INTEGRATED BIOMARKER RESPONSE (IBR) INDEX

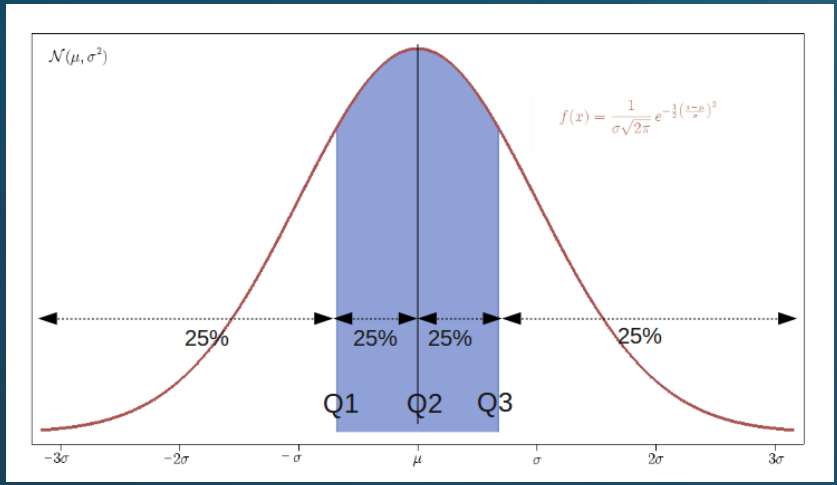
VARIABLES ARE INCLUDED AND WEIGHTED AND ITS INFLUENCE DECIDED BY THE OPERATOR EMPIRICAL KNOWLDGEMENT

STATISTICAL SELECTION

VARIABLES ARE SELECTED AND WEIGHTED DUE TO ITS STATISTICAL POWER AND WEIGHT AND OVERALL EXPLANANTION FOR THE WHOLE DATASET DIFFERENCES AMONG SAMPLE GROUPS

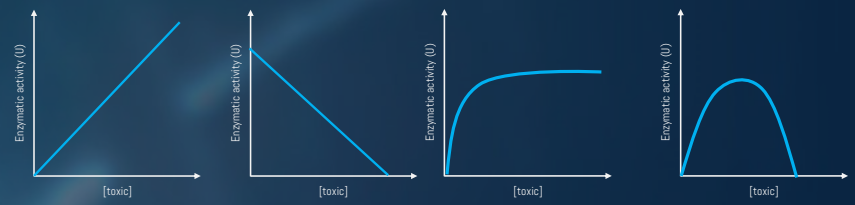
NON-SELECTIVE METHOD – BIOEFFECT ASSESSMENT INDEX (BAI)

ALL BIOMARKERS ARE INCLUDED AND SCORED ACCORDING TO THEIR QUARTIL POSITION



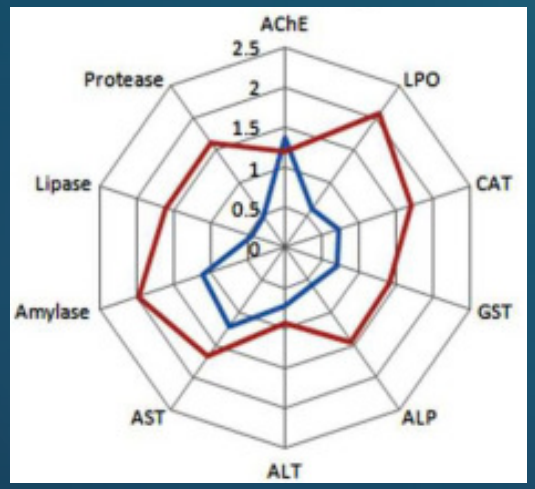
- Analyse the normal distribution of the biomarker data population.
- Each biomarker value is then scored according to its position in the quartile interspace
  - Values > 3rd quartil: 40
  - Values between 3rd and 2nd quartils: 30
  - Values between 2nd and 1st quartils: 20
  - Values < 1st quartil: 10

$$BAI = \sum Biomarker_{index}^{1/n}$$



WEIGHTED INDEX – INTEGRATED BIOMARKER RESPONSE (IBR) INDEX

VARIABLES ARE INCLUDED AND WEIGHTED AND ITS INFLUENCE DECIDED BY THE OPERATOR EMPIRICAL KNOWLDGEMENT



- Values standardized using the biomarker average and standard deviation and in a second step trough its minimum value observed;
- Values are considered negative or positive in the case of a biological effect corresponding respectively to an inhibition or a stimulation;

$$IBR = \sum_{t=15}^{n} A_i$$

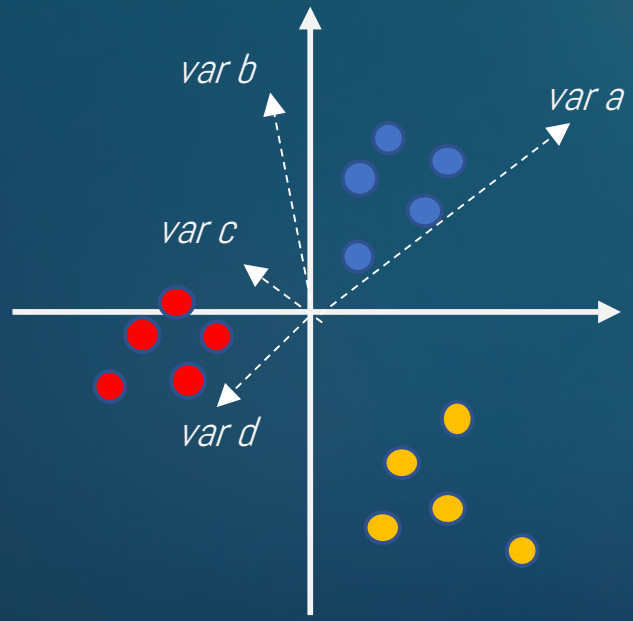
being Ai the area between two consecutive clockwise scores in a given star plot:

$$A_i = \frac{S_i}{2} \sin\beta (S_i \cos\beta + S_{i+1} \sin\beta)$$

$$\beta = \tan^{-1} \frac{S_{i+1} \sin\alpha}{S_i - S_{i+1} \cos\alpha}$$

STATISTICAL SELECTION

VARIABLES ARE SELECTED AND WEIGHTED DUE TO ITS STATISTICAL POWER AND WEIGHT AND OVERALL EXPLANANTION FOR THE WHOLE DATASET DIFFERENCES AMONG SAMPLE GROUPS

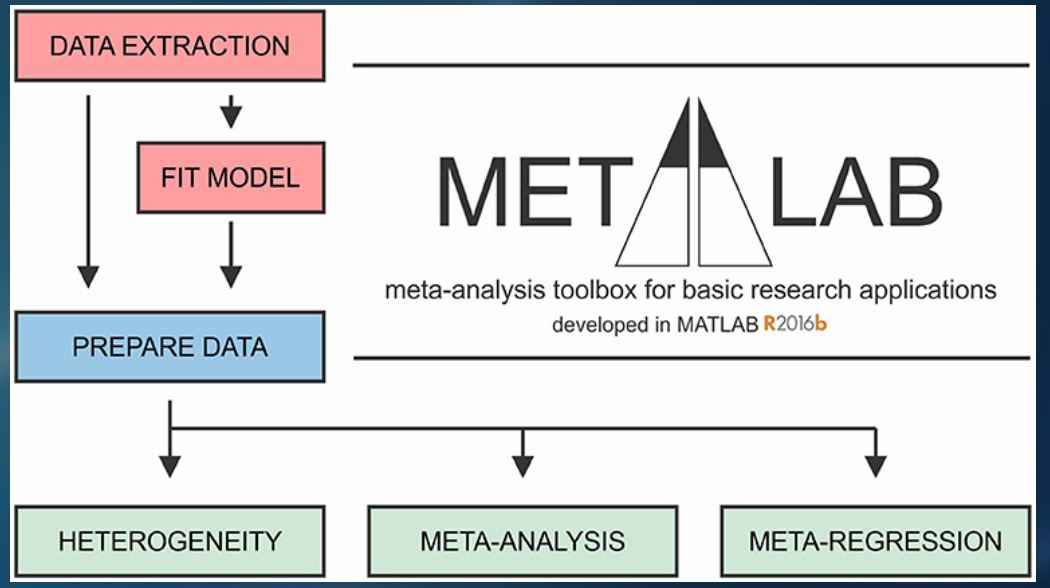


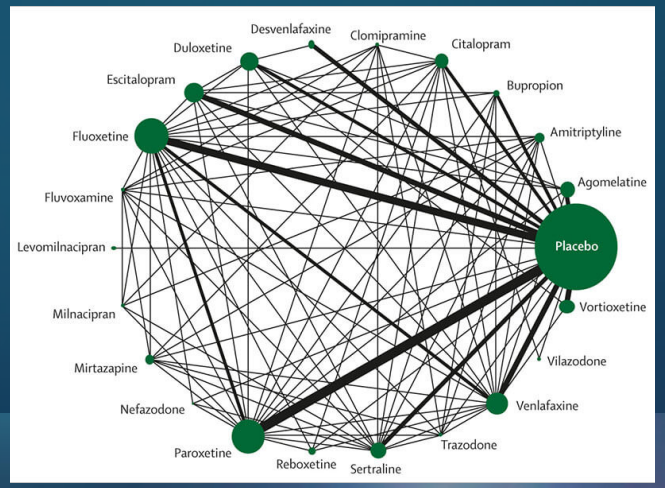
- Biomarkers to be included are selected using a statistical explanation threshold (top 5 most significant, higher dissimilarity towards the control, higher explanatory power);
- Selected values are normalized fore its statistical weight (PCA eigenvalues, similarity percentage, statistical relative contribution).
- Values are normalized typically using a sin or 1/x function to scale each biomarker score within a limit range so that all the weighted biomarkers vary between a pre-defined range

$$E = \frac{a}{1 + {}^{x}/_{x_0}{}^{b}} \qquad PCA - based\ index = \sum W_i\, E_i$$

Quantitative approach for systematically combining results of previous research to arrive at conclusions about the body of research.

- Quantitative : numbers
- Systematic : methodical
- combining: putting together
- previous research: what's already done
- conclusions: new knowledge





04 STATISTICS IN ECOTOXICOLOGY : META-ANALYSIS

PUBLICATIONDATABASE ANALYSIS

Using scientific search engines (SCOPUS for e.g.) and appropriate keywords collected the highest number of references with data on the target subject.

PUBLICATION ANALYSIS

Analyse the scientific content of the studies and decide which ones to keep and to throw out according to a methodological decision tree.

METHODOLOGICAL EVALUATION

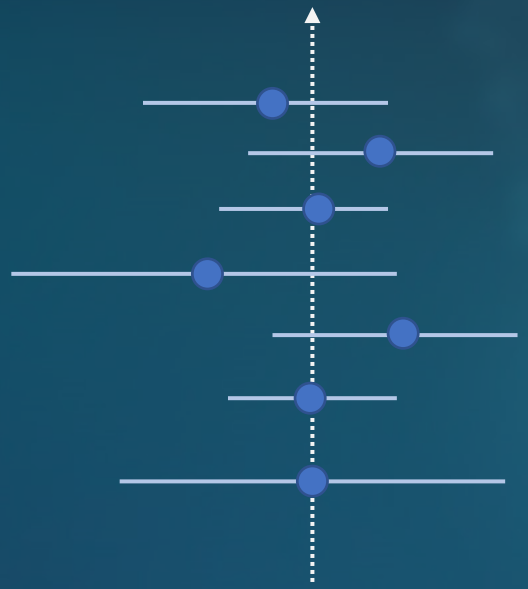
Evaluate the methods used and if they are preformed according to the predefined standards and adequate for being included in your comparative database.

DATA EXTRACTION

Extract all the data of the variables of interest: average, standard deviation, number of replicates.

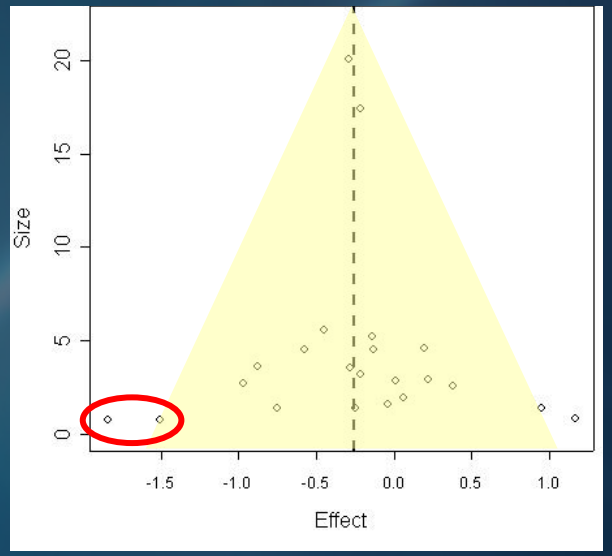
04 STATISTICS IN ECOTOXICOLOGY : META-ANALYSIS

FOREST PLOT



- The dotted line passes across null, or 1.0
- The Risk Estimate of each study is lined up on each side of the dotted line, with 95% CI spread as the line

FUNNEL PLOT



- Plots the effect size against the sample size of the study
- To study a funnel plot, look at its LOWER LEFT corner, that's where negative or null studies are located
- If EMPTY, this indicates "PUBLICATION BIAS"
- Note that here, the plot fits in a funnel, and that the left corner is not all that empty, but we cannot rule out publication bias