

# APPLIED STATISTICS

(Applied Biostatistics, Statistics and R)

Course Outline

Current version September 28, 2024

Erlis Ruli

Department of Statistical Sciences

University of Padova

Second Cycle Degree in

Molecular Biology

Quantitative and Computational Biosciences.

## General concepts

- statistical test
- p-value
- type I and type II error
- power of a test
- confidence interval
- some criteria for sample size calculation
- parametric and non-parametric test statistics
- use of central limit theorem
- multiple testing
- correlation and regression
- principal component analysis
- cluster analysis

## Statistical toolbox

- analysing a proportion (*Mendel, Spock, diphtheria*).
- comparing two proportions
  1. using the normal approximation (*genes and alcohol*);
  2. Fisher's exact test (refer to computer lab session).
- analysing the mean/median of a homogeneous population
  1. mean of a normal distribution with known variance (*INR*);
  2. mean of a normal distribution with unknown variance – one-sample *t* test (*monarch butterflies*);
  3. large samples (*BMD*);
  4. median/sign test (*xenope*);
  5. median of a symmetric distribution – one-sample Wilcoxon test (*xenope*).

- analysing two groups
  1. two normal populations/comparing means – two-sample  $t$  test (*cuckoos and Darwin*);
  2. large samples/comparing means (*pressure reduction and calcium intake*);
  3. paired data (*xenope*);
  4. two-sample Wilcoxon test (*medicinal herbs*).
- comparing several groups/one-way analysis of variance/Kruskal-Wallis test (*schizophrenia*).
- Bonferroni and Holm correction; FDR at glance (*schizophrenia*).
- verifying normality and Shapiro-Wilk test (*cuckoos and Darwin, monarch butterflies, ...*).
- two-way analysis of variance; the idea of interaction (*poisons and antidotes*).
- correlation: interpretation of (linear) correlation coefficient; limits of application; correlation and causality (*several examples*).
- linear regression: least-squares; interpretation of regression coefficients; significance testing for regression coefficients (*several examples*).
- logistic regression: definition; interpretation and significance testing for regression coefficients (*shuttle, diabetes*).
- a glimpse of principal components: definition; use; interpretation of importance measures and loadings; biplot (*crabs*).
- a brief look into cluster analysis: purpose; distance measures among statistical units; hierarchical algorithms; dendrogram; kmeans (*red vines of the Lange region*).