

# DATA ANALYSIS AND BIOSTATISTICS

Docente

**GIOVANNI BELLOMO,  
PHD**

Laboratory of Clinical  
Neurochemistry  
Neurology Clinic,  
University of Perugia  
Perugia, (PG), Italy

**Aula 13  
Edificio C  
Piano -1**

**Ingresso lato edificio D**

Coordinatore

**PROF.SSA ROBERTA SPACCAPELO**

**DOTTORATO DI RICERCA IN PATOLOGIE INFIAMMATORIE E INFETTIVE, STRATEGIE TERAPEUTICHE E BIODIRITTO**

1

**23/11/2023, 10.00-12.00**

## Lesson 1 (2h) - Probability and statistical tests

- Why do you need biostatistics?
- Probability density functions (Normal, Binomial, Student's) and statistical estimates (mean, median, variance, covariance etc.)
- Kolmogorov-Smirnov and Shapiro-Wilk normality tests.
- Parametric statistical tests: t-test
- Non-parametric statistical tests based on rank: U-test)
- Examples on R

2

**30/11/2023, 10.00-12.00**

## Lesson 2 (2h) – Multiple groups comparisons, covariate variables

- Pearson's and Spearman's correlation coefficients.
- One-way analysis of variance (ANOVA)
- Two-way ANOVA
- Analysis of covariance (ANCOVA)
- Linear regression
- Logistic regression
- Examples on R

3

**07/12/2023, 10.00-12.00**

## Lesson 3 (2h) – Power analysis & sample size calculation

- Type I & II statistical errors
- Power and effect size
- Sample size calculation (t-test, correlation, ANOVA)
- Examples on R

4

**14/12/2023, 10.00-12.00**

## Lesson 4 (2 h) – Classification 1

- Unsupervised vs supervised classifiers
- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Clustering: EM-clustering, K-means, Hierarchical clustering
- Examples on R

5

**21/12/2023, 10.00-12.00**

## Lesson 5 (2 h) – Classification 2

- Confusion matrices
- Logistic regression and ROC analysis
- Support vector machines (SVM)
- Regularization techniques (Ridge, LASSO, Elastic-Net)
- Examples on R

6

**10/01/2024, 10.00-12.00**

## Lesson 6 (2 h) – Applied biostatistics

- R fundamentals
- Overview of student's research projects
- How to structure your data analysis pipeline