

T109 Data analyses – Schedule 2024-2025

All courses will be on Thursday mornings from 9 to 12 in the computer room (room 313).

Date	Course	Teacher
19-sept	Estimation – confidence interval	Pascal Campagne
26-sept	Hypothesis testing	Pascal Campagne
03-oct	Experimental design and power analysis	Emeline Perthame
10-oct	Regression 1	Hugo Varet
17-oct	Regression 2	Hugo Varet
24-oct	ANOVA 1	Elise Jacquemet
31-oct	ANOVA 2	Elise Jacquemet
21-nov	Multivariate analysis 1	Vincent Guillemot
05-dec	Multivariate analysis 2	Vincent Guillemot
09-jan	Project	Emeline Perthame

Teaching team:

Emeline Perthame (emeline.perthame@pasteur.fr)

Pascal Campagne (pascal.campagne@pasteur.fr)

Hugo Varet (hugo.varet@pasteur.fr)

Élise Jacquemet (elise.jacquemet@pasteur.fr)

Vincent Guillemot (vincent.guillemot@pasteur.fr)

Course outline

I) **Experimental design and power analysis**

Experimental design

- Definition of an experimental plan from statistical point of view
- Review of frequent experimental designs
- Replication
- Definition of statistical confusion and how to avoid it through randomization & blocking
- A word on pooling

Power analysis

- Reminder of the power of a statistical test
- Application to Student test with R

II) **Estimation, confidence interval and hypothesis testing**

Coming soon

III) **Regression**

This course introduces several concepts of correlation between two quantitative variables before defining the linear regression. Then we go further with the multiple linear regression to take into account several predictors simultaneously. We will focus on the hypotheses that need to be made before performing such an analysis, as well as some situations where data have a structure that can lead to wrong conclusions if not modeled carefully.

IV) **ANOVA**

1. Analysis of variance
 - 1.1 One-way: comparing one factor
 - 2.2 Two-way: comparing two factors
2. What is interaction and how to deal with it?
3. Pairwise comparisons
 - 3.1 Contrasts
 - 3.2 Multiple testing
4. About technical factors: quick introduction to mixed effect models

V) **Multivariate analysis**

This course delves into the exploration of multivariate analysis, with an emphasis on two unsupervised techniques:

1. Principal Component Analysis (PCA) - An integral part of factor methods, PCA enables the unravelling of intricate correlation structures among variables and facilitates the understanding of observation group structures.
2. Hierarchical Clustering - This technique clusters observations or variables, based on a specifically defined distance.

These powerful methods will give you deeper insights into multidimensional data, when traditional methods like scatterplots and boxplots are lacking.

Furthermore, you'll gain hands-on experience by applying these techniques using R, by analyzing biological datasets.