

PSL Master in Life Sciences – M1

Semester S1, 2026-27

UNBIO1-129 *Data analysis* (6 ECTS)

Course Homepage [URL](#)

1. Short Description of Course

This course develops a model-based approach to data analysis. Starting from likelihood and prior, we will learn to read familiar tools such as least squares, analysis of variance or standard tests as applications of that framework rather than isolated recipes. Topics include linear regression and generalized linear models, confounding and mediation, regularization and model selection, multiple testing, dimensionality reduction methods, and experimental design. The emphasis will be on modeling assumptions, when they might fail, and how to interpret the results. Weekly tutorials will discuss relevant examples and analyze small genomic datasets.

Maximum class size: 40 students

The class is taught in English.

Keywords: statistical modeling, inference, regression, high-dimensional data, genomics

2. Registration

Students external to the ENS Biology Department can attend and validate the class if the primary instructor has been contacted beforehand and agreed (see contact information below). Following approval, students need to register online at least one week before the course starting date.

3. Location and Timeline

Location: Biology Department, 3rd floor, room 306 except Nov 5 in room 324, at 46 rue d'Ulm, Paris
Building entry with access card only (visitors must use the intercom at the entrance). The floor and class rooms are accessible without a badge. The building, floor and class room are accessible to people with disabilities.

Total duration (in person): 45 hours

Homework will consist in simple problems directly relating to preceding lectures/following tutorials.

For exact dates/time of the class this year, refer to the course website.

4. Instructors and Contact Information

Primary instructor: Natanael Spisak (he/him), natanael.spisak@gmail.com

Teaching Assistant: Erell Masson (she/her), massonerell4@gmail.com

Preferred method of contact: email.

5. Course Communications

Course material and homework will be sent by email and made available on Moodle.
Announcements will be sent by email.

6. Course Prerequisites or Co-requisites

Course prerequisites: elementary probability, descriptive statistics
Competency level in Unix: basics (U1)
Level in programming: basic; basics of R syntax

7. Course Format, Teaching Methods, Special Activities

Course format and teaching methods: Lectures will be followed by (simple) homework and in-class tutorials. The tutorials will take place in a classroom with in-house computer technology; however, students are strongly encouraged to bring their own laptop with needed software packages and libraries properly installed prior to class.

8. Expected Learning Outcomes

Upon completion of the course students will be able to:

- Develop models to formulate and test hypotheses about data
- Estimate model parameters using maximum likelihood methods
- Analyze data by linear regression and using generalized linear models
- Interpret model estimates, accounting for uncertainty and confounding
- Estimate statistical power and design experiments
- Follow advanced data analysis and genetics/genomics courses at M2 level

9. Detailed Schedule and Content

Thursdays 9 am. No class on Nov 26, Jan 7, 14. Written exam on January 28.

Topic for the tutorials will be announced weekly.

	Thursdays	9 am	10 am	Content
Principles of statistical inference	Sep 24	Lecture 1		Maximum likelihood estimation. Overview of data types and models
	Oct 1	Tutorial 1		
			Lecture 2	Uncertainty and information. Confidence intervals
	Oct 8	Tutorial 2		
		Lecture 3	Linear regression. Analysis of variance.	
Linear models	Oct 15	Tutorial 3		
			Lecture 4	Generalized linear models: Poisson and logistic regression
	Oct 22	Tutorial 4		
			Lecture 5	Confounding and mediation
	Oct 29	Tutorial 5		
	Nov 5	Tutorial 6		
		Lecture 6	Curse of dimensionality. Large-scale testing	
Inference from high-dimensional data	Nov 12	Tutorial 7		
			Lecture 7	Regularization. Model selection
	Nov 19	Tutorial 8		
			Lecture 8	Principal component analysis. Clustering
	Dec 3	Tutorial 9		
	Dec 10	Tutorial 10		
		Lecture 9	Power and sample size	
Power analysis and experimental design	Dec 17	Tutorial 11		
			Lecture 10	Experimental design and synthesis
	Jan 21	Tutorial 12		

10. Evaluation and Grading

The homework will not be graded but will be a prerequisite to most in-class tutorials.

Written exam will be on January 28.

<u>Assessment Categories</u>	<u>Percentage of final grade</u>
Final exam	100%

Students with a grade below 10/20 may take an oral exam retake; the retake score becomes the final grade, capped at 12/20.

11. Required or Suggested Resources

Readings will be suggested for specific lectures/tutorials. No single textbook is required.

12. Absence, Class Participation, Classroom Behavior Policies

Students are expected to attend and participate in classes; students who stop attending the class may not sit the final exam.

Disclaimer: Information contained in this syllabus may be subject to change with reasonable advance notice, as deemed appropriate by the primary instructor of the course.