INTRODUCTION TO DATA ANALYSIS

Fall session, October 16th – November 17th, 2017
Winter session, January 8th – February 2nd, 2018

Aims and course schedule

Under the initiative of our direction, the C3BI has set up a data analysis course. This course is (primarily) dedicated to first year PhD students of the Institut Pasteur (each of them entering our institute will be automatically registered to the course), but applications from 2nd and 3rd year PhD students and postdocs working at the Institute are encouraged. Applications will be accepted within the limit of the number of places available. First year PhD students with a background in mathematics or physics will be allowed to ask for an exemption. In order to answer legitimate questions you may have about the course, we encourage you to spend a few minutes reading the following.

Course description

The course will cover a broad range of notions that are mandatory for students seeking to design an experiment, explore and analyze the data, interpret the results and generate relevant scientific plots for publication purposes. It will provide fundamental knowledge in statistics, including uni- and multi-variate descriptive analyses, usual probability distributions and their application in biology, estimation, sampling and hypothesis testing. Practical work will be carried out in R through the RStudio interface. The course will start with several lectures dedicated to R programming, during which basic notions of statistics will be revised through exercises.
The course will consist in a series of three-hour lectures per day, four days a week over four weeks mixing closely theory and practical work. There will be two sessions, the first one starting October 16th, 2017 and the second one starting January 8th, 2018. Four lectures will be fully dedicated to practice.

Course program

Each lecture will be dedicated to a particular point of R programming and / or statistics:

1. Introduction to R and RStudio: objects and basic functions, data import, file system and path
2. Sampling, random variables, laws, central limit theorem, law convergence. R: import/export and manipulate data.
3. Estimation 1: proportion, mean, variance, standard deviation, standard error to the mean, min, max. Graph 1.
5. Confidence intervals (mean, proportion, variance)
6. Hypothesis testing (comparison of means, t-test, comparison of proportions)
7. Practice #1: basic data description and analysis
8. Correlation between 2 continuous variables + hypothesis testing, introduction to simple linear models
9. Linear models and variable selection
10. Logistic regression, ROC curves, ANOVA
11. Practice #2: linear models
12. Principal component analysis
13. Clustering
14. Experimental design and power analysis
15. Practice #3, RMarkdown
16. Practice #4

Course web page

You can access the course web page through this link: https://c3bi.pasteur.fr/training-introduction-to-data-analysis-2017-18/
Course material, program, schedule and practical information will be available on this webpage soon.

**Course requirement**

In order to follow the course all students need to bring a laptop. Please check that your computer meets the minimum requirements listed below.


The week before the course starts, students will be invited to have their laptop checked by the C3BI teaching team if necessary. The schedule will be announced on the course web page.

**Contact**

For any question you may ask, please contact us at: ida@pasteur.fr

**FAQ**

Based on our first experience of this course, we suggest you to read the following FAQ before contacting the teaching team for questions.

**Q.** What happens if I miss lectures because of missions or meetings?
**A.** Just send an email to marie-agnes.dillies@pasteur.fr to be excused. Try to catch up on missing information with the course material available on the course website. Course assistants will answer your question when you are back, if any.

**Q.** May I be (partially) exempted for the course?
**A.** Yes you can ask for an exemption if you have a background in bioinformatics, computer science, mathematics, physics or any related application field. You can ask for a part-time exemption if you followed some classes covering some of the course topics.

**Q.** How can I ask for exemption?
**A.** Please follow the “exemption” link on the course website.
Q. How can I make sure that my laptop suits the course needs?
A. Requirements about the laptop configuration are posted on the course web page. If you have any doubt you can come to the Hub to have your machine checked. Dedicated times slots are announced on the course web page.

Q. What if I have no laptop?
A. Please contact us: c3bi-trainings@pasteur.fr

Q. Can I validate the course as a Doctoral School module?
A. Yes. Just send a copy of your attendance certificate to the Doctoral School after the end of the course.