INTRODUCTION TO DATA ANALYSIS

Fall session, October 22nd – November 20th, 2018
Winter session, January 14th – February 8th, 2019

Aims and course schedule

The C3BI offers a data analysis and an optional introduction to image analysis courses, primarily intended to the first year PhD students of the Institut Pasteur. New students starting at the Institute are automatically registered to the course, but applications from 2nd and 3rd year PhD students and postdocs are encouraged. These applications will be accepted within the seat limit of the course (35 per session). The course is mandatory for 1st year PhD students, unless they have a background in mathematics or physics. This document aims at answering legitimate questions you may have about the course.

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 covers two domains in two parts: data analysis and image analysis.

The data analysis module 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 optional image analysis module will introduce the basic principles of image analysis, or how to extract quantitative information from microscopy images. The course is aimed at people who have no or very little experience in the field. It will be oriented towards
practical use, and short lectures will be followed by hands-on sessions and tutorials. It should help experienced microscopists and beginners who have never had any formal training in image quantification.

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 22\textsuperscript{nd}, 2018 and the second one starting January 14\textsuperscript{th}, 2019. Four lectures will be fully dedicated to practice.

**Course program.**
The exact schedule will be specified on the course web page.

**Computer Science module**
This one-lecture module is intended for students who are not familiar with basic computer science notions such as computer architecture, file system organization, file format and programming language. It is highly recommended to all students. At the end of this lecture, there will be time left for questions regarding the needed configuration of students’ personal laptop for the Data and Image Analysis modules.

**Data analysis module**
Each lecture will be dedicated to a particular point of R programming and / or statistics:
1. Introduction to R and RStudio 1: objects and basic functions, data import/export
2. Introduction to R and RStudio 2: objects and basic functions, data import/export
4. Estimation: proportion, mean, variance, standard deviation, standard error to the mean, min, max, rank, median, quartiles, quantiles, outliers and robust estimation.
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, RMarkdown
15. Practice #3
16. Practice #4

**Image analysis module**
1. Introduction. Typical challenges in image analysis. (Lecture)
2. Fundamentals. (Lecture & practicals)
   a. Navigating in images.
   b. Bit-depth and image representations.
   c. Histograms
   d. Manufactured images.
e. Display adjustments. View of the data vs data.

f. Lookup tables and multi-channel images.

g. Interlude: Digital images ethics.

h. Dealing with 3D images: Z-stacks.

i. Spatial calibration. Sizes and distances.

j. Conclusion.

3. Quantification of image data. (Lecture & practicals)

   a. Introduction.

   b. Extracting objects.

   c. Local maxima and noise.

   d. Image processing in brief.

   e. Other detection techniques.

   f. Automated segmentation of objects.

   g. Thresholding.

   h. Morphological operations in brief.

   i. Edge images and the watershed algorithm.

   j. Deformable contours.

   k. Segmentation by machine learning.

4. Example applications: extracting data from time-lapse images. (Practicals)

   a. Analysis of bleb growth.

   b. Calcium response of bronchial cells.

   c. FRAP on a moving target.

Course web page.

You can access the course web page through this link: https://c3bi.pasteur.fr/training-introduction-to-data-analysis-2018-19/

Course material, program, schedule and practical information are made available on this webpage.

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.

- **PC – Windows based:**
  - Intel i3
  - Windows 7
  - 4GB RAM
  - 256 GB HD

- **Apple Mac:**
  - mid-2010 mac book
  - OSX 10.10
  - 4GB RAM
  - 256 GB HD

We will send instructions as how to install the required software shortly before the course. Also, at the end of the Computer Science module, students will be invited to have their laptop checked by the C3BI teaching team if necessary.
**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 you have 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 an 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.