

C4c: Quantitative Methods

Instructors: Shoibal Chakravarty (coordinator shoibalc@atree.org), Som Chattopadhyay, Asmita Sengupta and Jagdish Krishnaswamy
Guest Lecturers: TBA

Lectures: Tuesday and Wednesday (12:00-12:50)

Lab: Thursday (11:00-11:50)

Credits: 2

Course Description

This course will introduce statistical methods and exploratory data analysis using R statistical software (<https://www.r-project.org/>). The goal of this course is two-fold: 1. to equip the student with a good understanding of the basic statistical toolbox that they will use in their research, and 2. to give a working knowledge of programming, data analysis and visualization in the R software environment. This course will emphasize the role of statistics as an integral part of the scientific method. Examples, applications and assignments will be drawn from both social and natural sciences.

Contact time: 3 hours a week including a 1 hour laboratory.

Prerequisites

ATREE Basic Math course or equivalent exposure to mathematics and statistics at the Bachelors or Masters level.

Tentative List of Topics

The list of topics is divided into a **Core** section that we expect the students to learn during the course and an **Electives** section that has advanced topics which we can cover depending on the progress and interests of the class.

Core

[programming]

1. Programming, data analysis and visualization in R.
2. Exploratory data analyses for univariate, bivariate and multi-variate data.
3. Data mining: scraping and tidying data

[Statistical methods]

4. Role of statistics in the scientific method.
5. Probability as an expression of uncertainty. Laws of Probability and Bayes theorem,
6. Probability distributions (discrete and continuous): Bernoulli, Binomial, Poisson, Normal.
7. Sampling distributions. Central Limit Theorem.
8. Fundamentals of frequentist Hypothesis testing. Type 1 and Type 2 errors. P-values, significance testing and the null hypothesis.
9. Expected and actual frequencies: Chi-square analyses
10. Bayesian approach, concept of likelihood and statistical inference.
11. Study design and sampling.
12. Linear models

13. Generalized Linear models
14. Fixed and random effects models

Electives

15. Quantile regression to deal with expression of response at higher levels when multiple unmeasured factors are present.
16. Fundamentals of Time-series analyses. Temporal auto-correlation. Seasonality. Trend analyses. Generalized Least Squares models with auto-correlated errors.
17. Non-linear models. Generalized Additive Models.
18. Concepts of spatial data analyses (geostatistics).
19. Multi-variate data and dimension reduction: factor analyses and principal component analyses.
20. The class is welcome to suggest possible topics!

Course Evaluation

5 assignments of 10 marks each, 5 short quizzes of 5 marks each, and a course project of 25 marks.

Resources:

Introduction to Probability and Statistics Using R, G. Jay Kerns. <http://ipsur.org/>, <http://ipsur.r-forge.r-project.org/book/download/IPSUR.pdf>

Introduction to Statistical Thought, Michael Levine
<http://people.math.umass.edu/~lavine/Book/book.html>

The course will use Rstudio (<https://www.rstudio.com/>) as the user interface for R. See <https://www.rstudio.com/online-learning/#R> for instruction on downloading and installing R and Rstudio, and how to get started with R.

R for Data Science, Garrett Golemund and Hadley Wickham
<http://r4ds.had.co.nz/>

Online courses

Introduction to R on DataCamp <https://www.datacamp.com/courses/free-introduction-to-r>
Try R on Code School <http://tryr.codeschool.com/>

Self-paced short interactive courses in Rstudio
<http://swirlstats.com/students.html>

Schedule of Lectures

Hour No.	Date	Topic	Instructor
1.	9 Jan 2018	Role of statistics in the scientific method	Jagdish Krishnaswamy
2.	10 Jan 2018	Laws of probability and Bayes theorem	Som Chattopadhyay
3.	11 Jan 2018	R Lab: Introduction to R	Shoibal Chakravarty
4.	16 Jan 2018	Probability distributions I	Som Chattopadhyay
5.	17 Jan 2018	Probability distributions II	Som Chattopadhyay
6.	18 Jan 2018	Sampling Distributions and Central limit theorem	Som Chattopadhyay
7.	23 Jan 2018	R Lab: Introduction to R II	Shoibal Chakravarty
8.	24 Jan 2018	Sampling Distributions and Central limit theorem	Som Chattopadhyay
9.	25 Jan 2018	R Lab: R data structures and plotting	Shoibal Chakravarty
10.	30 Jan 2018	Hypothesis and significance Testing I	Som Chattopadhyay
11.	31 Jan 2018	Hypothesis and significance Testing II	Som Chattopadhyay
12.	1 Feb 2018	R Lab: Exploratory data analysis	Shoibal Chakravarty
13.	6 Feb 2018	Chi-square Analysis	Som Chattopadhyay
14.	7 Feb 2018	Exploratory data analyses for univariate, bivariate and multivariate data	Shoibal Chakravarty
15.	8 Feb 2018	R Lab	Shoibal Chakravarty
16.	13 Feb 2018	Exploratory data analyses for univariate, bivariate and multivariate data	Shoibal Chakravarty
17.	14 Feb 2018	Data Mining: Scraping and tidying data	Shoibal Chakravarty
18.	15 Feb 2018	R Lab: Data Mining: Scraping and tidying data	Shoibal Chakravarty
19.	20 Feb 2018	Linear Models	Shoibal Chakravarty/Asmita Sengupta
20.	21 Feb 2018	Generalized Linear Models	Shoibal Chakravarty/Asmita Sengupta
21.	22 Feb 2018	R Lab	Shoibal Chakravarty/Asmita Sengupta
22.	27 Feb 2018	Generalized Linear Models	Shoibal Chakravarty/Asmita Sengupta
23.	28 Feb 2018	Fixed and random effects models	Shoibal Chakravarty/Asmita Sengupta
24.	1 Mar 2018	R Lab	Shoibal Chakravarty/Asmita Sengupta
25.	6 Mar 2018	Fixed and random effects models	Shoibal Chakravarty/Asmita Sengupta

26.	7 Mar 2018	Elective/extra lecture: TBA	Shoibal Chakravarty/Asmita Sengupta
27.	8 Mar 2018	R Lab	Shoibal Chakravarty/Asmita Sengupta
28.	13 Mar 2018	Elective/extra lecture: TBA	TBA
29.	14 Mar 2018	Elective/extra lecture: TBA	TBA
30.	15 Mar 2018	R Lab	Shoibal Chakravarty
31.	20 Mar 2018	Bayesian approach and concept of likelihood	Jagdish Krishnaswamy
32.	21 Mar 2018	Bayesian approach and concept of likelihood	Jagdish Krishnaswamy
33.	22 Mar 2018	R Lab	Jagdish Krishnaswamy
34.	27 Mar 2018	Bayesian approach and statistical inference	Jagdish Krishnaswamy
35.	28 Mar 2018	Bayesian approach and statistical inference	Jagdish Krishnaswamy
36.	29 Mar 2018	R Lab	Jagdish Krishnaswamy
37.	3 Apr 2018	Concepts of spatial data analysis	Jagdish Krishnaswamy
38.	4 Apr 2018	Multi-variate data and dimension reduction: factor analyses and principal component analyses	Jagdish Krishnaswamy
39.	5 Apr 2018	R Lab	Jagdish Krishnaswamy
40.	10 Apr 2018	Elective/extra lecture: TBA	TBA
41.	11 Apr 2018	Elective/extra lecture: TBA	TBA
42.	12 Apr 2018	R Lab	TBA
43.	19 Apr 2018	Optional R Lab	TBA