Thematic Course
PhD in "Scienze Economiche e Statistiche"
SEAS Department
University of Palermo

Academic Year	2021-2022
Subject	Large-scale Inference and Bioinformatics
Instructor	Giorgio Bertolazzi
Course description	This course will provide the primary technical skills for analyzing a huge amount of data. The students will be introduced to the usage of the shall, the usage of computer clusters, and parallel computing. The statistical methodology about large-scale inference and resampling approaches will be contextualized in the computational biology field. However, those statistical and technical methods can be used to analyze any type of big data. Students are expected to be confident in computer cluster usage and parallel computing. Moreover, the students will create their own R programs that run from the shall.
Learning Objectives	Students completing this course should be able to: • Run parallel programs • Connect and use computer clusters from the shall • Create R programs that run from the shall • Understand the bootstrapping approach • Adjust p-values for multiple comparisons
Suggested readings	• An introduction to bootstrap – Efron, Tibshirani • Papers in a dedicated dropbox folder
Course Activity (hrs)	8h
Credits	2
Assessment Method	The student will create a parallelized program that runs from the shall
Teaching Methods	Each section utilizes a combination of lecturing, computer lab, and class discussion.
Calendar	January/February 2022
Contacts	giorgio.bertolaz@gmail.com

Calendar of Classes

Lecture	Date	Topic	Duration
1	Jan.-Feb.	**Introduction to computational biology**; analysis of gene expression **Large scale inference**; multiple comparison procedures	2h
2	Jan.-Feb.	**Large scale inference**; differential expression analysis, clustering, gene set enrichment, application of gene signatures to survival analysis	2h
3	Jan.-Feb.	**Non-parametric methods based on resampling**; bootstrapping, jackknife, bootstrap testing, cross-validation	2h
4	Jan.-Feb.	**Informatics for the analysis of big data**; usage of computer clusters, parallel computing, usage of the shall from R console, creation of R programs	2h