

Statistics 598c

Statistical Methods For Bioinformatics and Computational Biology

Fall 2011

Location: TTh 10:30 -11:45 pm, REC 307

Instructor: Olga Vitek, office HAAS 120, mail HAAS 164, phone 496-9544, email ovitek@stat.purdue.edu

Goals of the course: The course introduces statistical methods and algorithms for modern high-throughput quantitative experiments in molecular biology. Emphasis will be made on design and analysis of transcriptomic, proteomic and metabolomic experiments, and on their interpretation in the broader context of functional investigations.

The course will overview the biological and technological background underlying these investigations, the theoretical and computational aspects of the methods, and will provide hands-on experience with using free and open-source software R. At the end of the course the students will have a broad view of computational problems in this field, and will be able to independently interpret such datasets.

Pre-requisite: The course is designed for a broad audience, and students from both statistics and life sciences background are welcome to attend. A prior exposure to statistical methodology, e.g. from STAT503, STAT 512, STAT 514 or STAT 524, is required. A prior exposure to R is desirable, but not required. Prior knowledge of biology is not required.

Software: Homeworks and projects will be carried out using R. Access to R is required. Please install R from <http://lib.stat.cmu.edu/R/CRAN/> prior to the course. Instructions for accessing Bioconductor will be provided during the course.

Course web page: <http://www.stat.purdue.edu/~ovitek/STAT598C-Fall11.html>

Daily updates on the schedule, handouts and homework assignments will be posted on the course page.

Course mailing list: The course mailing list will allow us to communicate outside of the lecture hours. You are encouraged to ask and answer questions on the list. All important announcements will be made through the list.

Textbook: There textbook for the course is

Sorin Draghici, *Statistics and Data Analysis for Microarrays using R and Bioconductor*, 2nd edition, CRC press <http://www.crcpress.com/product/isbn/9781439809754> The 1st edition is not a substitute. The 2nd edition will be in print in September 2011. The textbook is particularly recommended to life scientists who need extra help with statistical methods and R.

Office hours: Tue 11:45am-12:45pm and Wed 9:30-10:30am, or by appointment

Homework: There will be approximately 8 homework assignments due in class. Permission for late submissions should be obtained from the instructor in advance. Depending on the topic, different problem sets will be given to students in statistics and students in life sciences.

Projects: The course will have three projects.

- **Literature review:** The students will select a topic of interest, and review the related literature. The project can be done in groups of **2-3 students**. The project should be prepared in the format of a scientific paper, contain a relatively broad discussion of the topic, and cover at least one paper that was not presented in detail in the class.

- **Data analysis:** The project involves analysis of experimental data, preparation of a report written in the format of a scientific paper, and presentation of the results during the final week of class. The project will be done in groups of **2-3 students** that will consist of a statistician and a life scientist whenever possible. If you already work on a research project in this area you are welcome to use a dataset from your research, provided that you make an extra effort for the class.
- **Review of a data analysis project** During the last week of the class, each student will prepare a short review of a data analysis project by another project group.

Examinations: There will be no midterm and no final exam.

Final Grade: The final grade will be computed as follows:

Homework: 30%

Literature review project: 30%

Data analysis project: 30% Review of data analysis: 10%

The final letter grades will follow the usual scale:

90-100 = A, 80-89 = B, 70-79 = C, 60-69 = D, 0-59 = F.

The grades will be made available through Blackboard.

Registration: All students are asked to formally register for the course. No auditors will be allowed.