Statistics 6358: Biostatistics and Bioinformatics Fall 2021
Tuesday and Thursday
11:00 – 12:20
Heroy 129

Instructor: Associate Professor Monnie McGee, PhD
Office: Heroy 142
Email: mmcgee@smu.edu
Office Hours: Tuesday and Thursday 3:30 – 4:30 and Friday by appointment.

Mask Policy: Masks are required per current university and Dallas County policy. This masking policy is subject to change during the semester, and any changes will be posted clearly in Canvas announcements.

Course Objectives: By the end of this course, the student should be able to

- Explain basic genomic & proteomic structure of living organisms
- Analyze data from various types of high-throughput technology and their applications in basic biological research and medicine.
- State the basic statistical issues relevant to many types of high-throughput technology
- Apply statistical methods that account for these issues, mostly with the aid of R and Bioconductor.
- Interpret the results of a high-throughput data analysis using appropriate databases and other tools

Prerequisites: Statistics 6327, or permission of instructor.

Course Texts:

You do not need to purchase any text for this course. Readings will be available on Canvas as necessary, and you will be given ample time to read important papers before they are discussed in class.

Software (all open source – therefore – FREE!):

Please download the latest version of R to your computer before the semester begins.

R (free download from http://cran.r-project.org/). You will need the base installation to start. There are many R add-on packages that we will be using during the semester.

Bioconductor (free download from http://bioconductor.org). Bioconductor is an R-add suite of packages that deal with the analysis of high-throughput biological data. You must install R first before installing Bioconductor.

Outline of Topics:

The plan is to spend 2-4 class sessions (1 – 2 weeks) on each of the topics below. The technology will be introduced on the first day, a statistical method on the second, and then a hands-on problem set or R lab on the third day. Subject to change according to pace and interest of the students and instructor.

- **Overview: Molecular biology and Bioconductor**
- **EDA, Preprocessing and Gene MicroChips**
- **Overdispersion & RNA-seq gene-level analysis**
- **Differential expression and Multiple testing adjustments**
- **Classification and Clustering via Gene Expression**
- **Over/Under Representation and the Chi-Square test**
- **Biological Networks & Graphical Models**
- **Markov Chains and Sequencing**
- **Flow Cytometry & Cross Sample Matching**
- **Cell types & compositional data**
- **Mass spectrometry**
- **Lessons from Ecology**

Grading: Weblems (10%), Problem Sets (30%), Final Project (30%), Oral Presentation (20%), Interim Final Project Deadlines (10%).

Weblems (10%): Weblems are small web-based problems designed to give you experience with finding data or using various genomics websites. There will be roughly one per week.

Problem Sets (30%): Problem sets will be assigned roughly every 10 days – 2 weeks. Assignments will be a mix of theoretical and computational problems. We will often start these in class.

Final Project (30%): due during finals period. See the documents on the course website for intermittent due dates, details, and project ideas. This is your final exam.

Oral Presentation (20%): During the last week of class (November 30 and December 2), each student will be assigned a time to present his/her final project.

Interim Project Deadlines (10%): If you make a reasonable attempt and turn it in on time, you will receive full marks for this category. Don't miss this chance for an easy A! Here are the deadlines (subject to change with the pace of the course, blah, blah):

> *Choosing a project topic*: **October 29 by 5 p.m**. Submit a document containing *a brief summary* of the project. I will mention potential project ideas as they come up in class. For now, one idea is to select a paper referenced in class and reproduce or tweak its findings. You can also choose a technology not explored in the course or texts (e.g. CyTOF, gel electrophoresis, fMRI, protein binding, etc.) and explain how data from this technology are analyzed and utilized. These are not the only acceptable project ideas. Please talk to me about your idea before the deadline if you are unsure. **The main requirement for a project is that it is must be more than a simple data analysis.**
>
> *Example Code Submission*: **November 4 by 5 p.m.** Upload R (or Python or SAS) code that implements part of the project to Canvas. I must be able to run it on my computer; therefore, any data files must be included.
>
> *Rough Draft of Project*: **November 18 (Tuesday) by 5 p.m**. Submit a written draft of the project. *To get credit, the draft must have at least one figure and at least one table.* The figure and the table should be fully explained in the text, and an outline of other pieces of

the project should also be included. There must also be a bibliography. In general, the more you give me, the better I will be able to review your project. Comments on drafts will be returned before Thanksgiving break.

From The Provost (i.e. University Policies over which I have little control)

Incompletes will only be given in the case of extraordinary circumstances that prevent you from finishing the semester. You must have completed at least 50% of the course with a passing grade to be eligible for an incomplete.

Disability Accommodations: Students needing academic accommodations for a disability must first register with Disability Accommodations & Success Strategies (DASS). Students can call 214-768-1470 or visit http://www.smu.edu/Provost/ALEC/DASS to begin the process. **Once registered, students should then schedule an appointment with the professor as early in the semester as possible, present a DASS Accommodation Letter, and make appropriate arrangements**. Please note that accommodations are not retroactive and require advance notice to implement.

Religious Observance: Religiously observant students wishing to be absent on holidays that require missing class should notify me in writing by February 4 (e-mail is OK), and should discuss with me, in advance, acceptable ways of making up any work missed because of the absence. (See University Policy No. 1.9.)

Excused Absences for University Extracurricular Activities: Students participating in an officially sanctioned, scheduled University extracurricular activity will be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work. (University Undergraduate Catalogue)