Statistical Methods of Bioinformatics & Proteomics/Genomics
The University of Toledo
BRIM Program in Bioinformatics & Proteomics/Genomics
BIPG5200/7200 Statistical Methods in BPG, Section 001, CRN #45720

Instructor: Sadik Khuder, PhD
Email: Sadik.Khuder@utoledo.edu
Office Hours: Tuesdays 1-4 pm
Office Location: 0012 RHC
Office Phone: 419-383-4089
Term: Fall

Class Location: Online Lectures
Class Day/Time: Wednesdays/8 AM
Lab Location: 127 HEB
Lab Day/Time: By arrangement
Credit Hours: 3cr HR

COURSE/CATALOG DESCRIPTION
This course introduces the application of probability and statistics to bioinformatics analysis. This is an introductory course, though it is assumed students have a basic background in statistics. Topics include stochastic processes, Markov chains, statistical basis for DNA sequence analysis, evolutionary models, and statistical analysis of functional genomic data. Includes computer laboratory sessions. May be taken concurrently with BIPG510/710 (Fundamentals of BPG).

COURSE OVERVIEW
This course introduces students to statistical methods commonly used in bioinformatics. These methods are no more complex mathematically than those such as ANOVA or Student’s t-test, but are not typically covered in standard statistics courses. Students will learn to use statistical programs (especially R) and related bioinformatics resources locally and on the internet. Lectures and lab discussion will emphasize the statistical models and methods underlying the computational tools. The course will focus on the application of the newer statistical methods and the reasoning behind these applications. More emphasis will be placed on the analysis of genomic and proteomic experiments and students will learn statistical techniques to handle RNAseq and microarray data.

STUDENT LEARNING OUTCOMES
After completion of the course, successful students should be able to:

1. State the fundamental concepts of statistics in bioinformatics.
2. Describe the formulation of stochastic models for genomic and proteomic data.
3. Apply statistical techniques to analyze microarray data and interpret the results generated.
4. Apply statistical techniques to analyze RNA-seq data and interpret the results generated.
5. Use statistical tests commonly employed in bioinformatics.
6. Describe modern statistical methods and software to solve complex problems in bioinformatics.
7. Interpret statistical results as reported in the bioinformatics literature.

PREREQUISITES AND COREQUISITES
Any general introductory statistics course (or permission of course director).
REQUIRED TEXTS AND ANCILLARY MATERIALS
There is no required text for this course. All the required materials will be available on the course web site. Readings will consist of original literature, review articles, and R based books (available free online).

UNIVERSITY POLICIES
Policy Statement on Non-Discrimination on the basis of Disability (ADA).
The University is an equal opportunity educational institution. Please read The University’s Policy Statement on Nondiscrimination on the Basis of Disability Americans with Disability Act Compliance.

ACADEMIC ACCOMMODATIONS
The University of Toledo is committed to providing equal access to education for all students. If you have a documented disability or you believe you have a disability and would like information regarding academic accommodations/adjustments in this course please contact the Student Disability Services Office.

GRADING
Assignments 30%

There are 11 homework assignments in this course. These assignments are intended to improve skills in bioinformatics and statistical analysis of genomic and proteomic data. Each is worth 3%, with the lowest-scoring one automatically dropped.

Project 20%

A project will explore a topic of the course in greater depth. A written report, along with a PowerPoint presentation, is required. A project that addresses a question through data analysis, with a written report and summary of conclusions is sufficient. More detailed descriptions of the project along with suggested topics will be posted to the course webpage at the appropriate times during the term. The tentative due date for project submission is November 25th.

Final examination 50%

The final exam will be a take-home exam. The exam will cover all the materials presented in the course.
COURSE SCHEDULE
All lectures and labs for this course will be posted online on Blackboard. The following illustrates the types and range of topics covered, but individual lectures may change as needed to keep the course up to date.

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Assignments/Quizzes</th>
</tr>
</thead>
</table>
| 8/26 | Introduction  
Review of probability and statistics  
Modeling DNA | 1 |
| 9/02 | Overview of microarray technologies  
Preprocessing Affymetrix microarray data | 2 |
| 9/09 | Analysis of multiple DNA Sequences  
Protein sequences and substitution matrices | 3 |
| 9/16 | Other Microarray Platforms | 4 |
| 9/23 | Analyzing differential gene expression I | Project update |
| 9/30 | Analyzing differential gene expression II | 5 |
| 10/07 | RNA-seq  
Analyzing RNA-seq data | 6 |
| 10/14 | Biological annotation & visualization | 7 |
| 10/21 | Proteomics  
Analyzing proteomic data | 8 |
| 10/28 | Evolutionary models  
Phylogenetic tree estimation | 9 |
| 11/04 | Phylogenetics  
Infectious disease | Project update |
| 11/11 | Markov Chains  
Hidden Markov model | 10 |
| 11/18 | Cluster Analysis | 11 |
| 11/25 | Discriminant analysis  
ANOVA Models | Project submission |
| 12/02 | Applied General Bioinformatics  
Applied Genomic & proteomic | Final Exam |

The University of Toledo is committed to providing equal access to education for all students. If you have a documented disability or you believe you have a disability and would like information regarding academic accommodations/adjustments in this course please contact the Student Disability Services Office.