BINF 6200/8200 (BINF6200L/8200L) STATISTICS FOR BIOINFORMATICS
SYLLABUS (Fall, 2015)

Instructor: Jun-tao Guo
Contact: Office: Bioinformatics 359, Phone: 704-687-7492, Email: jguo4@uncc.edu
Office Hours: By appointment

Time and Place:
Lecture: Tuesday/Thursday 11-12:15 in Room 301
Lab: Thursday 12:30-13:45 in Room 104

Textbooks:
Required: there is no required textbook for this course.
“Statistical Methods and data Analysis”, R. Lyman Ott, Michael Longnecker

Course Description:
The aim of this 3-credit course is to introduce statistical methods commonly used in bioinformatics. Basic relevant concepts from probability, probability distributions, and statistical inference will be introduced and illustrated by examples from bioinformatics applications. R will be introduced as the statistical analysis software for this course.

Pre/Co-requisites:
Prerequisite: Permission of the department

Learning objectives for BINF6200/8200 Statistics for Bioinformatics

1. Understand the basic concepts of random variables, probability, and Bayes' theorem.
2. Master the key discrete and continuous probability distributions.
3. Master the concept of sampling distributions of a statistic and Central Limit Theorem.
4. Construct confidence intervals and perform hypothesis testing for one-sample, two-sample (paired or independent samples), and multi-sample.
5. Understand the key differences between parametric and non-parametric tests.
6. Have a solid understanding of statistical significance, types of errors, and multiple testing of hypotheses in a bioinformatics context.
7. Perform linear regression and correlation analysis.
8. Perform hypothesis testing and statistical analysis using R.

Instructional Methods:
The course will be presented in a lecture format which will include the following elements as appropriate: presentation of concepts, theories and examples in a standard lecture format, interactive demonstrations of methods, and opportunities for student questions and discussion. Students will get hands-on experience in R programming and statistical analysis in R.
Grading Plan:
Students will be evaluated based on their mastering of the concepts and theories taught in the class, and the ability to use them for solving practical problems. The grade is determined as follows:

Homework assignments: 35%
Mid-term Exam: 25%
Final Exam: 30%
Quizzes: 5%
Classroom participation: 5%. (BINF8200 students will present a research paper or do a research project)

Grades will be assigned on the following scale:
A=90-100 B=78-89 C=65-77 U=0-65

SPECIFY POLICIES THAT APPLY TO THIS COURSE:

1. UNIVERSITY INTEGRITY
All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online at:
http://www.legal.uncc.edu/policies/ps-105.html. A set of links to various resources on plagiarism and how to avoid it is available at the UNCC Library website:
http://library.uncc.edu/display/?dept=instruction&format=open&page=920.

2. ATTENDANCE
Attendance at lecture is required, although exceptions will be made for reasons such as illness or family emergency. Excessive absences will result in a reduced classroom participation score at the instructor’s discretion, and will negatively impact the overall course grade.

TOPICAL OUTLINE OF COURSE CONTENT

• Introduction to R and R programming
• Basic concepts of probability
• Random variables and independence
• Probability distributions (discrete and continuous)
• Sampling distribution and Central Limit Theorem (CLT)
• Classic hypothesis tests: z-test, t-tests, non-parametric tests
• Statistics for microarray data analysis
• Correlation and regression models

Tentative Schedule:

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Overview and introduction to R /Lab: R basics</td>
</tr>
<tr>
<td>Week 2</td>
<td>Basic probability and Statistics, Bayes’ theorem/Lab: R basic structures</td>
</tr>
<tr>
<td>Week 3</td>
<td>Discrete probability distributions /Lab: random simulations and plotting</td>
</tr>
<tr>
<td>Week 4</td>
<td>Continuous probability distributions /Lab: Read in data and sub-setting</td>
</tr>
<tr>
<td>Week 5</td>
<td>Sampling distribution and estimation /Lab: basic R programming</td>
</tr>
<tr>
<td>Week 6</td>
<td>Confidence intervals and hypothesis testing /Lab: sampling distributions in R</td>
</tr>
</tbody>
</table>
Mid-term (October 1, 2015)

Week 7  One-sample hypothesis testing /Lab: hypothesis testing in R
Week 8  Two-sample hypothesis testing /Lab: normality test in R
Week 9  Inference about population variances /Lab: hypothesis testing in R
Week 10 Multi-sample hypothesis testing /Lab: more R programming
Week 11 Non-parametric hypothesis testing
Week 12 Linear regressions and correlation /Lab: linear regression in R

Final exam