BIOM611 Biological Data Analysis
Spring, 2015
Tentative Syllabus

Introduction

BIOMED611 is a ½-unit course required for all 1st year BGS students (except GCB students). It will provide an introduction to important topics in biostatistical concepts and reasoning and instructions on practices in data analysis. Specific topics include tools for describing and summarizing data; inference methods on population means and proportions; statistical hypothesis testing; group comparisons; simple linear regression; categorical data analysis; time to event data analysis; power and sample size in study designs. Labs will help students learn and implement the methods using R. Though there are some formulae and computational elements to the course, the emphasis is on interpretation, concepts, and applications.

Description

• Time and Place:
  Lecture 3:30-4:30 Monday, Class of ‘62 Auditorium, John Morgan Building.
  Labs 3:30-5:00 Tuesday to Friday, BRB 259
  Each student is required to attend the lecture and one of the labs during the week;
  Everyone should stick to the same day of the lab after it is chosen.

• Instructor: Dr. Rui Feng (ruifeng@mail.med.upenn.edu, Blockley 211)
  TA and lab instructors: Emin Tahirovic (head)
    Others (3-4) TBA

• Office Hours: instructor – 4:30-5:30 on Monday; TAs: 5:00-5:30 on Tuesday to Friday

Textbook

• The Design and Statistical Analysis of Animal Experiments (2014, Simon T. Bate and Robin A. Clark, Cambridge)

Special Instructions

• Lecture notes, R programs and datasets for examples will be available on https://canvas.library.upenn.edu/.
• R online resources are available at CRAN: http://www.r-project.org/ and R manual is available at http://cran.stat.sfu.ca/.
Course Objectives

- Recognize different types of data arising in biomedical studies
- Interpret differences in data distributions via visual displays
- Calculate standard normal scores and resulting probabilities
- Calculate and interpret confidence intervals for population means and proportions
- Interpret and explain a p-value
- Perform a two-sample t-test and interpret the results
- Select an appropriate test for comparing continuous outcomes in two or multiple populations
- Choose an appropriate method for comparing proportions between two groups
- Understand and interpret relative risks and odds ratios when comparing two populations
- Understand why timed to event data requires its own type of analysis techniques
- Construct a Kaplan-Meier estimate of the survival function that describes the "survival experience" of a cohort of subjects
- Interpret the result of a log-rank test in the context of comparing the "survival experience" of multiple cohorts
- Describe different kinds of studies
- Understand confounding and interaction in studies
- Interpret output from the statistical software package related to the various estimation and hypothesis testing procedures covered in the course
- Use R to manipulate data and perform various analyses
## Discussion Topics

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (01/14)</td>
<td>Course introduction; introduction to statistics</td>
<td>Ch.1</td>
</tr>
<tr>
<td></td>
<td>Lab: R introduction, installation, data import</td>
<td></td>
</tr>
<tr>
<td>2 (01/19)</td>
<td>Random variables, samples, populations, probabilities</td>
<td>Ch 2.1 &amp; 2.2 &amp; 3.2.1</td>
</tr>
<tr>
<td></td>
<td>Lab: Data manipulation, random number generation</td>
<td></td>
</tr>
<tr>
<td>3 (01/26)</td>
<td>Data summaries: statistics and graphic tools</td>
<td>Ch 5.2&amp;5.3</td>
</tr>
<tr>
<td>4 (02/02)</td>
<td>Population mean and confidence intervals</td>
<td>Ch 5.2</td>
</tr>
<tr>
<td>5 (02/09)</td>
<td>Hypothesis testing with one sample example</td>
<td>Ch 2.3</td>
</tr>
<tr>
<td>6 (02/16)</td>
<td>Two sample t-test and Wilcoxon tests</td>
<td>Ch. 5.4.1 &amp; 5.4.2 &amp; 5.5.1</td>
</tr>
<tr>
<td>7 (02/23)</td>
<td>ANOVA</td>
<td>Ch. 5.4.3</td>
</tr>
<tr>
<td>8 (03/02)</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>9 (03/16)</td>
<td>Simple linear regression</td>
<td>Ch. 5.4.7</td>
</tr>
<tr>
<td>10 (03/23)</td>
<td>Regression diagnosis</td>
<td>Ch. 5.4.7</td>
</tr>
<tr>
<td>11 (03/30)</td>
<td>Multiple comparison</td>
<td>Ch. 5.4.8</td>
</tr>
<tr>
<td>12 (04/06)</td>
<td>Power and sample size</td>
<td>Ch. 3.7.2</td>
</tr>
<tr>
<td>13 (04/13)</td>
<td>Analysis of contingency tables</td>
<td>Ch. 5.5.1</td>
</tr>
<tr>
<td>14 (04/20)</td>
<td>Time to event data introduction</td>
<td>Ch. 5.5.3</td>
</tr>
<tr>
<td>15 (04/27)</td>
<td>Cox PH model</td>
<td>Ch. 5.5.3</td>
</tr>
<tr>
<td>16</td>
<td>Reading week</td>
<td></td>
</tr>
<tr>
<td>17 (05/04-05/12)</td>
<td>Final</td>
<td></td>
</tr>
</tbody>
</table>


Attendance and Lecture Policies

Use of computers during lecture is distracting for both the students using them and those sitting nearby. Laptops, tablets, smartphones, or other electronic devices need to be silenced and you cannot use any during lectures.

All students are required to bring their own laptops and power to the lab, with R appropriately installed after the 1st lab.

Evaluation

Homework assignments (35%) 4 Canvas Quizzes (15%)  
Midterm (25%) Final (25%)

Grading Scale

95-100 A+ 90-94.9 A 89-89.9 A-  
85-88.9 B+ 80-84.9 B 79-79.9 B-  
75-78.9 C+ 70-74.9 C 69-69.9 C-  
65-68.9 D+ 60-64.9 D

Homework

• Types of Homework: Two types of homework are required of this course.  
  (i) Short Problems: Taken directly from the text and meant to reinforce and review concepts presented in class.  
  (ii) Longer Problems: Developed by the instructors and may require an extension or application of the methodological concepts developed in class or in the text. They may be based on scenarios encountered by collaborative statisticians and require creative thinking and integration of concepts to solve.

• Policy: Each homework has equal value unless indicated otherwise by the instructor. A homework assignment is due on most Tuesdays. They will be available on the class web page a couple of weeks before they are due. I will not hand out printed copies of the assignments in class. Late homework will not be accepted. Do not e-mail me/TA your homework assignments; we will not save them. You must hand in your homework assignments on paper.

Quizzes

4 quizzes will be held via CANVAS for approximately 15-20 minutes. No make-ups.

Exams

The exams will be cumulative. The main emphasis of the exams will be testing your knowledge of what the appropriate statistical test is to use in a particular situation and how to interpret
the results. You will not be allowed to use reference books or notes or cell phones, tablets, or advanced calculators (models advanced than TI-30X or TI-36X or equivalent) during the exams, but you will need a simple calculator. There will no make-up exam. Under extreme circumstances, taking exams in next year will be allowed with permission.