BIOM611 Biological Data Analysis
Spring, 2015
Course Syllabus

Introduction

BIOMED 611 is a ½-unit course required for all 1st year BGS students (except GCB and GGB students). It will provide an introduction to important topics in biostatistical concepts and reasoning, and practical experience with elementary data management and 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; and issues in study designs. Lab sessions will help students learn and implement the methods using the R statistical programming language. 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, Austrian Auditorium, Clinical Research Building (CRB).
  Labs 3:30-5:00 Tuesday, Wednesday, and Friday, BRB 252 (need to bring your own laptop and power);
  3:30-5:00 Thursday, Anatomy-Chemistry 202 (limit 15).
  Every student can choose one of the labs during the first two weeks and should stick to the same day of the lab after it is chosen.
- Instructor: Dr. Rui Feng (ruifeng@upenn.edu, Blockley 211)
  TA and lab instructors: Tuesdays: Andrew Smith (andrms@mail.med.upenn.edu)
  Wednesdays: Le Wang (wangle@mail.med.upenn.edu)
  Thursdays and Fridays: Emin Tahirovic (emint@mail.med.upenn.edu)
- Office Hours: Instructor – 4:30-5:30 on Monday; TAs: 5:00-5:30 on Tuesday to Friday

Textbook


Special Instructions

- Lecture and lab notes will be available on https://canvas.upenn.edu/.
- R online resources are available at CRAN: http://www.r-project.org/, and the R manual is available at http://cran.stat.sfu.ca/.
- InVivoStat free software developed specifically for animal researchers, is available through www.invivostat.co.uk . It can generate usable R script.
Course Objectives

- Recognize different types of data arising in biomedical studies
- Understand different study designs
- Understand and know how to do randomizations
- Interpret differences in data distributions via visual displays
- Understand standard normal distribution and resulting probabilities
- Calculate and interpret confidence intervals for population means and proportions
- Interpret and explain type of errors and p-value in hypothesis testing
- Perform a two-sample t-test and interpret the results
- Select an appropriate test for comparing continuous outcomes between two populations
- Perform Analysis of Variance (ANOVA) for comparing continuous outcomes among multiple groups
- Know how to compare proportions between two groups
- Understand why time-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 output from R 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 (Date)</th>
<th>Topic</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (01/12)</td>
<td>Course introduction; introduction to statistics&lt;br&gt;Lab: R installation, data import, manipulation, storage, and export</td>
<td>Ch.1, 3.1, 3.2.1</td>
</tr>
<tr>
<td>2 (01/19)</td>
<td>No lecture on MLK day&lt;br&gt;Lab: Data summary tools, numerical and graphical</td>
<td>Ch 5.2&amp;5.3</td>
</tr>
<tr>
<td>3 (01/26)</td>
<td>Probabilities&lt;br&gt;Lab: Random number generation, Properties of common distributions</td>
<td>Ch 2.2</td>
</tr>
<tr>
<td>4 (02/02)</td>
<td>Population mean and confidence intervals&lt;br&gt;Lab: Means and CIs</td>
<td>Ch 5.2</td>
</tr>
<tr>
<td>5 (02/09)</td>
<td>Hypothesis testing with one sample example&lt;br&gt;Lab: one-sample t-test</td>
<td>Ch 2.3</td>
</tr>
<tr>
<td>6 (02/16)</td>
<td>Two sample t-test&lt;br&gt;Lab: T-test</td>
<td>Ch. 5.4.1 &amp; 5.4.2</td>
</tr>
<tr>
<td>7 (02/23)</td>
<td>Wilcoxon tests&lt;br&gt;Lab: Wilcoxon test</td>
<td>Ch. 5.5.1</td>
</tr>
<tr>
<td>8 (03/02)</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>9 (03/16)</td>
<td>Analysis of Variance (ANOVA)&lt;br&gt;Lab: ANOVA table</td>
<td>Ch. 5.4.3</td>
</tr>
<tr>
<td>10 (03/23)</td>
<td>Multiple comparison&lt;br&gt;Lab: Bonferroni, Tukey, FDR</td>
<td>Ch. 5.4.8</td>
</tr>
<tr>
<td>11 (03/30)</td>
<td>Simple linear regression&lt;br&gt;Lab: Linear regression and diagnosis</td>
<td>Ch. 5.4.7</td>
</tr>
<tr>
<td>12 (04/06)</td>
<td>ANOVA and regression&lt;br&gt;Lab: ANOVA and ANCOVA using regression</td>
<td>Ch. 5.4.6</td>
</tr>
<tr>
<td>13 (04/13)</td>
<td>Power and sample size&lt;br&gt;Lab: Power and sample size calculations</td>
<td>Ch. 3.7.2</td>
</tr>
<tr>
<td>14 (04/20)</td>
<td>Analysis of contingency tables&lt;br&gt;Lab: Chi-square test, Fisher’s exact test</td>
<td>Ch. 5.5.1</td>
</tr>
<tr>
<td>15 (04/27)</td>
<td>Time-to-event data introduction&lt;br&gt;Lab: summary, Kaplan-Meier curve</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>
**Evaluation**

Homework assignments (35%) 4 Canvas Quizzes (10%)

Midterm (25%) Final (30%)

**Grading Scale**

95-100 A+ 90-94.9 A 85-89.9 A-

80-84.9 B+ 75-79.9 B 70-74.9 B-

**Homework**

- **Types of Homework:** Real data problems developed by the instructor and may require some programming 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. Please hand in your homework assignments in class. Late homework will not be accepted.

**Quizzes**

4 quizzes will be held via CANVAS for approximately 20-30 minutes each. The window for each quiz will be open for one week.

**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. No books, no notes, and no electronic devices. No make-up exams.