CLRE 253 – BIOSTATISTICS I, 2 UNITS

Summer 2020

Course Instructor
Florin Vaida, PhD, fvaida@ucsd.edu; Office hours: by request

Course Teaching Assistant
Rishi Deka, rdeka@ucsd.edu; Office hours: TBA

Course Description
This course offers a rigorous, intensive introduction to Biostatistics. The first part of the course introduces the fundamentals of statistical inference: summarizing and visualizing data, probability, normal and binomial distribution, sampling, central limit theorem, confidence intervals, hypothesis testing. The second part of the course gives you the skills to understand and conduct the most commonly used analyses for continuous (scale) and binary response variables in the case of one-, two-, and three or more groups. The statistical methods will include independent samples and paired samples t-tests, tests based on ranks, Pearson’s chi-square and Fisher’s exact tests for binary variables, paired tests for continuous and binary variables, one-way ANOVA, correlation, and simple linear regression. Basic issues of study design will be discussed (clinical trials versus observational studies; sample size calculation; limitations of direct group comparisons in non-randomized studies). Data analyses will be conducted in R/Rstudio using datasets based on real-world biomedical studies and applications. This course is a pre-requisite for taking elective statistical courses.

Course Organization
Two 50-minute sessions per week. Case studies will be introduced and discussed for each key statistical method. We will also conduct interactive data analysis sessions using R/Rstudio. Students will work in small groups and individually.

Course Goals
Provide the knowledge, tools, and practical experience for understanding the basis of statistical inference, and for performing most commonly used basic statistical analyses for cross-sectional observational studies and randomized clinical trials.

Course Objectives
1. Develop a solid understanding of basic statistical concepts, and of the common statistical methods appropriate for biomedical data.
2. Develop the ability to read critically and understand the aspects of study design and statistical analysis from studies reported in the scientific biomedical literature.
3. Develop facility with examining and analyzing biomedical datasets using the appropriate statistical methods, in R/Rstudio.

Prerequisites and Preparation
Install R and Rstudio on your laptop prior to first class, from http://www.r-project.org/index.html and https://www.rstudio.com/products/rstudio/. Bring your laptop at all lectures and perform the analyses in real time, with the instructor.
**Course Materials/Resources** – Required and Recommended
- **Other Materials:** Detailed course notes are provided
- **Computer software:** R and Rstudio will be used in class, for homework, and for the final exam.
- **Online resources:** Canvas, at canvas.ucsd.edu; extensive online resources on R are available.

**Course Schedule**

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Date</th>
<th>Topic</th>
<th>Reading</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Jul 1/2</td>
<td>Study design; data summaries &amp; visualization</td>
<td>BCB Ch2,3</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>Jul 8/9</td>
<td>Probability; normal distribution</td>
<td>BCB Ch4</td>
<td>Hw1 due</td>
</tr>
<tr>
<td>L3</td>
<td>Jul 15/16</td>
<td>Central limit theorem, confidence intervals</td>
<td>BCB Ch4,5</td>
<td>Hw2 due</td>
</tr>
<tr>
<td>L4</td>
<td>Jul 22/23</td>
<td>Hypothesis testing, single group</td>
<td>BCB Ch5</td>
<td>Hw3 due</td>
</tr>
<tr>
<td>L5</td>
<td>Jul 29/30</td>
<td>Inference for means of two groups</td>
<td>BCB Ch5,6</td>
<td>Hw4 due</td>
</tr>
<tr>
<td>L6</td>
<td>Aug 5/6</td>
<td>Inference for proportions</td>
<td>BCB Ch4,5</td>
<td>Hw5 due</td>
</tr>
<tr>
<td>L7</td>
<td>Aug 12/13</td>
<td>Comparing proportions, two groups</td>
<td>BCB Ch5,6</td>
<td>Hw6 due</td>
</tr>
<tr>
<td>L8</td>
<td>Aug 19/20</td>
<td>Comparing 3 or more groups: ANOVA</td>
<td>BCB Ch6,7</td>
<td>Hw7 due</td>
</tr>
<tr>
<td>L9</td>
<td>Aug 26/27</td>
<td>Correlation and simple linear regression</td>
<td>BCB Ch8</td>
<td>Hw8 due</td>
</tr>
<tr>
<td>L10</td>
<td>Sep 2/3</td>
<td>Review</td>
<td></td>
<td>Hw9 due</td>
</tr>
<tr>
<td>Final</td>
<td>Sep 9/10</td>
<td>Final Exam</td>
<td>BCB Ch1-8</td>
<td></td>
</tr>
</tbody>
</table>

**Course Components**
- Weekly assignments, to be submitted through Canvas.
- Homework is due at 11:59pm, on the following week. Late homework is not acceptable.
- The only acceptable format is .pdf. Do not submit raw R output.
- One homework submission for a **group of two students**. Groups are assigned by instructor via Canvas.

**Final Exam**
- In-class, open-book, written 1h:50min exam, last week of the course. The exam will involve data analysis in R, and/or hand calculations.
- A make-up exam date will be set up for students with extenuating circumstances. These should make arrangements with the instructor at least one month in advance.
- The best tool for preparing the final exam is practice on the prior final exams, posted on course website. **Note however that we have switched from using SPSS to R, so past exam solutions use SPSS.**

**Grading Policy**
- The course grade will have two components: Homework (35%) and Final exam (65%).
- All students receive a final grade bonus of 0-2 percentage points, based on completed evaluation by the class as a whole (0% = 0 points, 100% = 2 points, with linear interpolation in-between).
Grade thresholds use the UCSD grading scheme:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total Score</th>
<th>Grade</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>94-100</td>
<td>C+</td>
<td>77-79.9</td>
</tr>
<tr>
<td>A-</td>
<td>90-93.9</td>
<td>C</td>
<td>74-76.9</td>
</tr>
<tr>
<td>B+</td>
<td>87-89.9</td>
<td>C-</td>
<td>70-73.9</td>
</tr>
<tr>
<td>B</td>
<td>84-86.9</td>
<td>D-/D/D+</td>
<td>≥61/64/67</td>
</tr>
<tr>
<td>B-</td>
<td>80-83.9</td>
<td>F</td>
<td>0-60.9</td>
</tr>
</tbody>
</table>

- Plagiarism will be dealt with according to the UCSD Policy of Integrity of Scholarship
- See your grades on Canvas canvas.ucsd.edu (if you are matriculated- in MAS Program), or on MyExtension https://myextension.ucsd.edu/ (if you are non-matriculated/CREST, concurrent enrollment)

Course Policy and Expectations (classroom rules of conduct)
- Time Commitment: Expect to spend 4-6 hours a week outside of class.
- Attendance Policy: The general CREST attendance policy applies, no more than 3 missed sessions are allowed. You have to be present in both halves of the class.
  - Academic Integrity (Plagiarism): http://academicintegrity.ucsd.edu/
  - Late homework is not acceptable.

Communication with lecturers
The best way to reach us is via email. We will respond within 24 hours under usual circumstances.

Student Evaluation of Course and Faculty
Course and faculty evaluations provide important feedback to instructors to improve course content and teaching methodology. The evaluations are also an important factor in faculty advancement. A link to one-time course evaluations is available on the course website. All students will receive a final grade bonus of 0-2 percentage points, based on evaluation completion by the class as a whole (0% = 0 points, 100% = 2 points, with linear interpolation in-between).

Technical Requirements
The students should bring their laptop with R and Rstudio installed to every class, and conduct analyses at the same time with the instructor.

Accommodations: If you have a disability that may impact your academic performance, you may request accommodations by submitting documentation to: https://students.ucsd.edu/well-being/disability-services/