M19-512 Intermediate Biostatistics for Clinical Research

Course Overview

1. Instructors:
   - Jingqin Rosy Luo, Ph.D, Professor of Surgery, Biostatistics and Medicine, jingqinluo@wustl.edu
   - Feng Gao, Ph.D, Professor of Surgery, Biostatistics, Medicine, feng@wustl.edu

2. Course TA
   - Yifei Xu, MS: xyifei@wustl.edu

3. Course time arrangements:
   - Fall 2: 10/23~12/15/2023 Mondays and Wednesdays 9am~12pm
     (WashU calendar: Thanksgiving Break – no classes Nov 22-26 (Wednesday-Sunday)
     - 9~10:30 Lecture
     - 10:30~12:00 Lab

4. PREREQUISITES: M19-511 or consent by instructor
5. TARGET AUDIENCE: medical students, clinicians, clinical and population health researchers

6. Office hours
   - Before/after class
   - Email scheduling for zoom or in person meeting

7. Course organization
   - (1) Lecture notes
   - (2) Lab assignment
   - (3) HW: usually due 1 week after assignment (submitted by 11:59pm of the due date, see table below)
   - (4) Final Project: data analysis project using your own data or public data
     - Project background and objective
     - hypothesis
     - modeling
     - result presentation
     - result interpretation

8. Class schedule, Syllabus and homework schedule:

<table>
<thead>
<tr>
<th>#</th>
<th>Date</th>
<th>Handout # / HW #</th>
<th>Topics</th>
<th>HW Due Date</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th></th>
<th>Date</th>
<th>Topic</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10/23</td>
<td>Overview, Note 1, Lab 1 / HW 1</td>
<td>Course overview Recap on R Intro to R markdown Intro to some course datasets EDA and data visualization</td>
<td>10/30</td>
</tr>
<tr>
<td>2</td>
<td>10/25</td>
<td>Note 2, Lab 2 / HW 2</td>
<td>Simple linear regression: data visualization, model estimation by least square, interpretation, prediction</td>
<td>11/1</td>
</tr>
<tr>
<td>3</td>
<td>10/30</td>
<td>Note 3, Lab 3 / HW 3</td>
<td>Multiple linear regression</td>
<td>11/06</td>
</tr>
<tr>
<td>4</td>
<td>11/1</td>
<td>Note 4, Lab 4 / HW 4</td>
<td>Linear Regression model assumption &amp; diagnostic</td>
<td>11/08</td>
</tr>
<tr>
<td>5</td>
<td>11/06</td>
<td>Note 5, Lab 5 / HW 5</td>
<td>Binomial Distribution, MLE, and Simple Logistic Regression</td>
<td>11/13</td>
</tr>
<tr>
<td>6</td>
<td>11/08</td>
<td>Note 6, Lab 6 / HW 6</td>
<td>Multiple Logistic Regression</td>
<td>11/15</td>
</tr>
<tr>
<td>7</td>
<td>11/13</td>
<td>Note 7, Lab 7 / HW 7</td>
<td>Logistic regression: Goodness of fit assessment</td>
<td>11/20</td>
</tr>
<tr>
<td>8</td>
<td>11/15</td>
<td>Note 8, Lab 8 / HW 8</td>
<td>Logistic Regression: Case-Control and Matched Designs</td>
<td>11/21</td>
</tr>
<tr>
<td>9</td>
<td>11/20</td>
<td>Note 9, Lab 9 / HW 9</td>
<td>Multinomial Logistic Regression</td>
<td>11/30</td>
</tr>
<tr>
<td></td>
<td>11/22</td>
<td>Thanksgiving break, no class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11/27</td>
<td>Note 10, Lab 10 / HW 10</td>
<td>Logistic Regression for Ordinal Outcome</td>
<td>12/4</td>
</tr>
<tr>
<td>11</td>
<td>11/29</td>
<td>Note 11, Lab 11 / HW 11</td>
<td>Poisson Regression and Zero-Inflated Poisson Model</td>
<td>12/6</td>
</tr>
<tr>
<td>12</td>
<td>12/4</td>
<td>Note 12, Lab 12 / HW 12</td>
<td>Intro to Survival Analysis, KM method</td>
<td>12/11</td>
</tr>
<tr>
<td>13</td>
<td>12/6</td>
<td>Note 13, Lab 13 / Final project</td>
<td>Survival analysis: Cox Proportional Hazards Model</td>
<td>None</td>
</tr>
<tr>
<td>14</td>
<td>12/11</td>
<td>Note 14, Lab 14 / Final project</td>
<td>Cox model assessment and handling</td>
<td>None</td>
</tr>
<tr>
<td>15</td>
<td>12/13</td>
<td>Note 15 / Final project</td>
<td>Time varying covariates</td>
<td>12/20</td>
</tr>
</tbody>
</table>

9. Course description & objectives: This course is a continuation of the introductory Biostatistics course (M19-511). The topics include basic statistical concepts and methods for various types of data (continuous, categorical, count, and time-to-event outcome data). Through lectures, R labs, homework assignments, and in-class exam, students will learn the concepts and methods commonly used for analyzing various data types, and will implement those methods using R software.

10. Competency: After completing the course, students should
(1) understand the basic statistical concepts and methods for various types of data
(2) be able to frame and address research questions using these concepts and methods
(3) be able to perform data analyses on these types of data using R software, and
(4) be able to interpret the results in the context of clinical research.
11. Grading: Your grade will be based on:
   - Class participation (10%)
   - HW: (60% total)
   - In class exam or Final project (30%)
12. Grading Scale
   A+: 97-100; A: 93-96; A-: 90-92; B+: 87-89; B: 83-86; B-: 80-82; C+: 77-79; C: 73-76; C-: 70-72
13. ATTENDANCE AND PARTICIPATION
   Class attendance is required. More than two unexcused absence from class result in 0 point from class participation.
14. POLICY ON LATE ASSIGNMENTS
   Late assignments will result in a deduction of 20% of the assignment for each day late (including weekends) unless prior approval is obtained from the instructor or a compelling situation prevents prior approval (i.e. documented health issues or family emergencies).
15. Course textbooks
    (1) Categorical data analysis by Argesti
    (2) Applied survival analysis using R (free online book)
        https://link.springer.com/book/10.1007/978-3-319-31245-3
16. Software: The course uses R/Rstudio. Please download from the link @https://posit.co/download/rstudio-desktop/ to your laptop and bring your laptop to classes for the computing lab assignments.