Statistical Methods in Public Health I
Biostatistics 140.621

September 4 - October 25, 2018

Department of Biostatistics
Johns Hopkins University
Bloomberg School of Public Health

Faculty Instructors:
Marie Diener-West, PhD
Karen Bandeen-Roche, PhD
COURSE SCHEDULE AND READINGS
STATISTICAL METHODS IN PUBLIC HEALTH I (140.621)
FIRST TERM
September 4 - October 25, 2018

Faculty Lecturers:

Marie Diener-West, PhD (Section 140.621.01)
Office W1015 or E3148, 410-502-6894, mdiener@jhu.edu

Karen Bandeen-Roche, PhD (Section 140.621.02)
Office E3527, 410-955-3067, kbandee1@jhu.edu

Department of Biostatistics
Johns Hopkins University
Bloomberg School of Public Health

Lectures: 10:30 am - 12:00 pm – Tuesday and Thursday

Sommer Hall (E2014) - Section 140.621
Sheldon Lecture Hall (W1214) - Section 140.621.02
Overflow Rooms with transmission from Sommer Hall: W3030 and W4030 (starting Sept 6)

Lab 140.921.xx: for review of material through a structured exercise and time for questions:

Lab 01  Monday, 1:30 pm - 3:00 pm - W5030
Lab 02  Tuesday, 1:30 pm - 3:00 pm - W5030
Lab 03  Wednesday, 1:30 pm - 3:00 pm - W5030
Lab 04  Thursday, 1:30 pm - 3:00 pm - W5030
Lab 05  Friday, 1:30 pm - 3:00 pm - W5030
Lab 06  Monday, 3:30 pm - 5:00 pm - W5030
Lab 07  Tuesday, 3:30 pm - 5:00 pm - W5030
Lab 08  Wednesday, 3:30 pm - 5:00 pm - W5030
Lab 09  Thursday, 3:30 pm - 5:00 pm - W5030

Please bring your laptop computers to lab. There is open time for questions with two lab instructors between 3:00 pm -- 3:30 p.m. each day.
COURSE SCHEDULE AND READINGS
STATISTICAL METHODS IN PUBLIC HEALTH I (140.621)
FIRST TERM
September 4 - October 25, 2018

Lab Instructors:
Karen Bandeen-Roche, PhD
Marie Diener-West, PhD
Junrui Di (Lead TA)
Jacob Fiksel (lead TA)
Jason Ji (Lead TA)
Jordan Johns (Lead TA)

Teaching Assistants:
Sophrena Bushey
Lacey Etzkorn
Amy Hong
Albert Kuo
Hojoon Lee
Andrew Leroux
Xintong Li
Tiangle Liu
Yuangen Liu
George Mwinnyaa
Susanna Qiao
Chinenye Ugoji
Prosenjit Kundu
Arkajyoti Saha
Kayode Sosina

Teaching Assistant Office Hours (starting Wednesday, September 5, optional):
Monday through Friday 12:15 pm - 1:15 pm in W2009

Teaching Assistant Office Hours in Computer Lab (starting Wednesday, September 5, optional):
Monday through Friday 2:30 pm - 3:20 pm in W3025
COURSE SCHEDULE AND READINGS
STATISTICAL METHODS IN PUBLIC HEALTH I (140.621)
FIRST TERM
September 4 - October 25, 2018

CoursePlus Site:

Available through CoursePlus: Course schedule, lecture notes, self-evaluation problems, Stata lecture notes, problem sets, data sets and solutions, quiz and exam solutions. Purchase of hardcopy material is included in registration. Lecture capture recordings from each lecture are recorded using Panopto.

Suggested Books:

There are no required books for this course. Most introductory statistics textbooks will provide background information. In addition, there are online resources, such as http://onlinestatbook.com/

Some helpful books are:


Lawrence C. Hamilton, Statistics with Stata 12, 2012, Duxbury, Thomson Brooks/Cole, Belmont,


Handheld Calculator:

A handheld calculator is needed for quizzes and examinations. Basic functions should include (+, -, *, /), logarithms and exponents, simple memory and recall, factorial key.

Statistical Computing Packages:

Stat 15 Intercooled, Stata Press, College Station, Texas

(Buy through http://www.stata.com/order/new/edu/gradplans/student-pricing/ )

This course supports the use of Stata as a statistical analysis package for laboratory exercises and problem sets.

We also support a smaller R Users Group using the R software as the statistical analysis package. The R Users Group is intended for students who have had prior programming experience and have a rationale for learning R. The lab for the R Users Group is on Fridays from 3:30 -4:50 pm. Please complete the “R Application” Quiz in CoursePlus by 11:59 pm on Wed, Sept 5 if you are interested in participating in the R Users Group.
Course Policies:

- Please email your **faculty lecturer** regarding extenuating circumstances or conflicts regarding course deadlines.
- Attendance is required for exams and expected for lectures and labs.
- Laptops and iPads may be used during lecture for class-related purposes. Common courtesy should be followed.
- Availability for course questions: after lecture, during labs, TA office hours, and Stata office hours.

Exam Policy:

Course exam dates will be provided to students on the first day of class, as well as posted in the syllabus on CoursePlus. Students are expected to take examinations at the assigned times and on the assigned dates.

*Previous Conflicts*

Students who have a legitimate conflict (e.g., clinical responsibilities, research presentations including travel, jury duty and other court appearances, weddings and personal travel, that were scheduled before the start of the class) with the scheduled exam dates must inform their faculty lecturer in writing via email at least two weeks prior to the scheduled exam date. If conflicts arise within two weeks of the scheduled exam date, the lecturer should be informed in writing immediately and students must provide documentation (e.g. letter from medical provider, school representative, or conference organizer) of the conflict.

*Sudden Illness*

Students who are not well on or near the exam date must provide medical documentation in the form of a statement on the medical provider’s letterhead and based on a clinic visit within two days of the missed exam.

*Tardiness*

Students who are late for the exam will be given the remaining period of time to complete the exam (that is, they will not be granted additional time). If there is an unforeseen event outside the student’s control (such as a traffic accident) that causes tardiness and the student calls the course coordinator to report the impending tardiness, accommodations may be made at the discretion of the faculty.

*Unexpected Emergency*

Students who experience an unexpected severe personal or family emergency during the course should contact their lecturer as soon as possible. Each case will be handled separately and reasonable effort will be made to allow for completing in a timely fashion missed work and/or exams with the approval of the course faculty lecturers.
Exam Policy (continued):

_Scheduling of Make-up Exams_
Exams will typically be re-scheduled after the posted exam date. Students typically will have no more than one week after the scheduled exam date to take a make-up exam. However, exceptions may be made at the discretion of the faculty for an exam to be rescheduled one or two days prior to the scheduled exam date. Or, for intractable legitimate conflicts, an alternative such as a take-home exam may be offered.

If the lecturer is not notified of conflicts prior to the exam, medical documentation is not provided for an illness or the make-up exam is not scheduled according to the procedures described above, students will receive no points for the exam towards their course grade.

Course Grade:

- 20% completion of 4 problem sets (1 point deducted for each day if turned in late)
- 5% quiz 1 (through CoursePlus)
- 5% quiz 2 (through CoursePlus)
- 35% midterm examination (in class)
- 35% final examination (in class)

Quizzes and examinations are individual work for which a student must work by himself or herself.

Problem sets may be worked on together and discussed. However, _each student must write up the problem set individually using his or her own words_. Copying work is not allowed.

Academic Ethics Code:

- The code, discussed in the Policy and Procedure Memorandum for Students, October 26, 2006, will be adhered to in this class
  [https://my.jhsph.edu/Offices/AcademicIntegrity/Documents/AE-PPM.pdf](https://my.jhsph.edu/Offices/AcademicIntegrity/Documents/AE-PPM.pdf)

- Students enrolled in the Bloomberg School of Public Health of The Johns Hopkins University assume an obligation to conduct themselves in a manner appropriate to the University's mission as an institution of higher education. A student is obligated to refrain from acts which he or she knows, or under the circumstances has reason to know, impair the academic integrity of the University.
COURSE SCHEDULE AND READINGS
STATISTICAL METHODS IN PUBLIC HEALTH I (140.621)
FIRST TERM
September 4 - October 25, 2018

Disability Support Services:

If you are a student with a documented disability who requires an academic accommodation, please contact Ms. Betty Addison in the Office of Career Services and Disability Support: dss@jhsph.edu, 410-955-3034, or Room W1600.

Course Objectives:

Students who successfully master this course will be able to:

1. Use statistical reasoning to formulate public health questions in quantitative terms:
   - Understand the role of statistical reasoning within the scientific model.
   - Understand and apply the counterfactual definition of cause in public health research.
   - Distinguish between continuous, categorical, binary and time-to-event data.
   - Understand that evidence for establishing an association between a risk factor and health outcome is generated by comparing the distribution of the outcome in otherwise similar populations with different levels of the risk factor.
   - Use stratification in design and analysis to minimize confounding and identify risk modification.

2. Design and interpret graphical and tabular displays of statistical information:
   - Create by hand and interpret stem and leaf plots, boxplots, Q-Q plots and frequency tables.
   - Use the statistical analysis package Stata to make basic statistical computations and graphical displays.
   - Graphically compare two groups of observations of otherwise similar units (e.g. people or treatments) and interpret the display.
   - Characterize the distribution of a variable – using the concepts of typical value, variability, and shape.
   - Use a variable transformation, such as the logarithm, to study a right skewed distribution such as hospital costs.
   - Analyze survey information and identify sources of error in variables.
   - Explore study results for associations among multiple variables and interpret the findings.
3. Use probability models to describe trends and random variation in public health data
   - Use the statistical analysis package Stata to make basic statistical computations and graphical displays.
   - Use the concepts of probability to describe the effect of a treatment on a health outcome in a randomized trial.
   - Use the binomial distribution and the Poisson approximation to the binomial to calculate probabilities of events.
   - Use the Gaussian or normal probability model to approximate the distribution of a continuous public health measure and to assess the quality of this approximation.
   - Use a quantile-quantile (Q-Q) plot to compare the shape of an empirical with a theoretical distribution.

4. Use statistical methods for inference, including tests and confidence intervals, to draw public health inferences from data:
   - Generate random numbers and appreciate variation among multiple observations of a random process.
   - Explain the implications of the Central Limit Theorem in determining the sampling distribution of the mean of n observations.
   - Use bootstrapping to determine confidence intervals.
   - Use sampling distribution theory for the mean and for differences between two means to create confidence intervals and hypothesis tests.
   - Use stratification to eliminate the influence of a possible confounding variable in a study of the association of a risk factor and outcome.
   - Use the appropriate two-sample t-test and confidence interval to assess whether average outcome is different between two groups and draw inferences.
   - Use the paired-sample t-test and confidence intervals to assess if the average change is different from zero.
   - Examine the consequence of using an inappropriate unpaired (two-sample) analysis when a paired analysis is appropriate.
   - Define and apply the term “effect modification” or equivalently “interaction” in a randomized trial.

The course is designed to enable students to develop their data analysis skills. Important datasets will be analyzed by the students using the statistical package Stata or R package throughout the 621-623 course sequence.
## COURSE SCHEDULE AND READINGS
### STATISTICAL METHODS IN PUBLIC HEALTH I (140.621)
#### FIRST TERM
##### September 4 - October 25, 2018

<table>
<thead>
<tr>
<th>Class</th>
<th>Date</th>
<th>Topic</th>
<th>Suggested Reading*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sept 4</td>
<td>Perspectives on Data and Statistical Thinking</td>
<td>Chapter 1</td>
</tr>
<tr>
<td>2</td>
<td>Sept 6</td>
<td>Exploring and Organizing Data to Address Public Health Questions: Continuous and Discrete Data</td>
<td>Chapter 2.1 - 2.5, 2.8 - 2.9</td>
</tr>
<tr>
<td>3</td>
<td>Sept 11</td>
<td>Probability Concepts</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>4</td>
<td>Sept 13</td>
<td>Probability Distributions: Binomial and Poisson PROBLEM SET 1 DUE by 11:59 pm</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>5</td>
<td>Sept 18</td>
<td>Probability Distributions: Normal QUIZ 1</td>
<td>Chapter 5</td>
</tr>
<tr>
<td>6</td>
<td>Sept 20</td>
<td>Sampling Strategies</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Sept 25</td>
<td>Summary and Review PROBLEM SET 2 DUE by 11:59 pm</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Sept 27</td>
<td>MIDTERM EXAMINATION</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Oct 2</td>
<td>Introduction to Statistical Inference: Sampling Distributions and the Role of Sample Size</td>
<td>Chapter 6.5</td>
</tr>
<tr>
<td>10</td>
<td>Oct 4</td>
<td>Bootstrap, Estimation and Interval Construction</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Oct 9</td>
<td>Estimation and Hypothesis Testing PROBLEM SET 3 DUE by 11:59 pm</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Oct 11</td>
<td>Inference: Single Sample Mean QUIZ 2</td>
<td>Chapter 7.1-7.4, 7.7</td>
</tr>
<tr>
<td>13</td>
<td>Oct 16</td>
<td>Generalized Linear Models</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Oct 18</td>
<td>Inference: Difference between Two Sample Means Pre-Post Designs and Other Paired Comparisons</td>
<td>Chapter 8.4-8.7, 8.2</td>
</tr>
<tr>
<td>15</td>
<td>Oct 23</td>
<td>Summary and Review PROBLEM SET 4 DUE by 11:59 pm</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Oct 25</td>
<td>FINAL EXAMINATION</td>
<td></td>
</tr>
</tbody>
</table>

*Fundamentals of Biostatistics by Rosner (7th Edition)