Course Information
Term: Summer 2017  
Credits: 4

Class Meetings
Parkvale 305 A/B  
Group A  
Group B

Recitations
Parkvale 305 A/B  
Group A  
Group B

Labs
Parkvale 222

TA Office Hours
TBD

Instructors
TBD

TAs
TBD

Course Description
This course focuses on basic concepts and methods of statistics and their application to problems in the health and biomedical sciences. We will cover graphical and numerical descriptions and summarization of data, basic probability theories, probability distributions, point and confidence interval estimation, and hypothesis testing with emphasis on one-, two-, and multi-sample comparisons involving continuous and categorical data. Correlation, simple linear regression, and nonparametric tests will also be introduced. Scholars will work with clinical datasets and learn to analyze the data and explain their findings. This rigorous first course in biostatistics will serve as a prerequisite for other biostatistics courses. At the completion of the course, you will be able to:

- Appreciate the concepts of random variation and bias
- Know the basic statistical procedures used to analyze data
- Apply these techniques utilizing a standard statistical package
- Appreciate the wide range of applications of biostatistical methods used to solve problems in medicine and public health
- Know some specific applications in a specialized area of interest
• Recognize pitfalls in interpreting biomedical and public health data

Course Format
This course will be in a hybrid – or flipped – format. Rather than attending lectures in person, you will complete designated readings, video lectures, and assignments on your own time, and then come to class to discuss and apply what you’ve learned. The hybrid format allows you to:

• Engage with the content at your own pace and on your own schedule
• Rewind, rewatch, or jump ahead as necessary
• Use class time for interaction and clarification
• Spend less total time in class

Required Materials

Software: Stata 15, a data analysis and statistical software (www.stata.com). Stata 14 can be accessed from our ICRE computing lab, the University computer labs at Posvar Hall, or in Falk Library at Scaife Hall. A yearly license can be purchased through University of Pittsburgh Computing Services and Systems Development (CSSD) for $5. CSSD is located in Bellefield Hall. Stata is available for PC, Mac, and Linux machines, so make sure you specify the platform you are using when you order the program. If you have decided to purchase Stata, it is best to have it in your computer by July 3. Note: In some cases you may be able to get your department/grant to pay for software. Contact Ms. Juliana Tambellini at tambellinijm2@upmc.edu if you need assistance.

We highly recommend the book, Statistics with Stata by Hamilton LC. (Updated for Version 12, 8th ed.) Duxbury Press: Pacific Grove, CA. We will put one copy of this book on reserve for your convenience.

Optional Materials


Weekly schedule
This class is split into two groups for class sessions and recitations. Check CourseWeb to find your group.

Weekly Responsibilities
Each week, you will complete the designated online materials (lecture videos, readings, and assignments) on your own time before coming to class. Be sure to take notes as you work, write down any questions you have, and bring your questions with you to class. Note: These materials are required,
not optional. We expect you to come to class fully prepared to discuss and apply what you’ve learned. We also expect you to ask for clarification and help when you need it.

Suggested Weekly Workflow
Managing time in a hybrid course can be challenging. While it’s up to you how you distribute your work during the week, we suggest the following workflow:

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do readings; watch videos; prepare for Tuesday recitation</td>
<td>Prepare for Wednesday class</td>
<td>Do readings; watch videos</td>
<td>Prepare for Friday class</td>
<td>Do readings; watch videos</td>
<td>Review; study</td>
<td></td>
</tr>
<tr>
<td>Attend recitation</td>
<td>Attend class; submit HW assignments</td>
<td>Attend lab; complete lab assignment</td>
<td>Attend class; submit HW assignments</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Course Requirements
In addition to completing weekly online modules and attending class sessions, you will complete 16 assignments, 7 labs, a midterm exam, and a final exam. You are encouraged to work together on homework and lab assignments, but you should write up your results individually and independently. A hard copy of homework and lab assignments should be submitted by 5:00 p.m. on the due date. No late submissions will be accepted. You will find the task descriptions and deadlines for assignments in the “Assignments” section of CourseWeb.

Grading
Your course grade will be based on your performance on homework assignments, lab assignments, midterm exam, and final exam.

- Homework: 30% (best 13 of 16 homework scores)
- Lab assignments: 5%
- Midterm exam: 30%
- Final exam: 35%

Note: In calculating final grades, we will take into account the degree of difficulty of exams and the average performance of students.

Help
Please do not hesitate to contact your TA via email if you have any questions about course content. You can generally expect an answer within 24 hours, if not sooner.

If you have any questions about scheduling or technology (CourseWeb, GoToMeeting, the sign-in system for attendance, etc.), please contact the course administrator, Juliana Tambellini at: tambellinijm2@upmc.edu.
Course Policies
You are responsible for knowing and following these course policies. Please read them carefully.

Attendance Policy
You are required to complete the assigned online self-paced modules before coming to class. We encourage you to actively engage with these modules, taking notes as you watch and read, and writing down any questions or comments you have. Bring your notes with you to class.

Attendance at weekly class meetings is required. You must also physically show up to class to take the midterm and final exams. Attendance in recitations is also required, though you are free to join remotely via GoToTraining if you cannot attend in person.

Please sign in to each class on the computer provided in suite lobby. Also sign in at labs and recitations. If you encounter a problem with the sign-in system, please let us know immediately.

Recording Policy
ICRE Produced Recordings: ICRE faculty and/or staff may video and/or audio record this course (hereby referred to as "Recordings"). By enrolling in this course, you hereby give the University of Pittsburgh and the Institute for Clinical Research Education, through its faculty, employees, agents, licenses or assigns, the irrevocable and worldwide right to use your name, voice, likeness and/or image in all forms and media (to include internet websites and online course website). You waive your right to inspect or approve the finished version(s) of the Recordings, including any copy that may be created in connection therewith. You understand that you will not be paid for your participation in the Recording and that you are not entitled to your own copy of the Recording. You understand that the University of Pittsburgh is not responsible for any unauthorized use of the Recording. You have read this syllabus and have no questions about the contents and are an adult over the age of 18.

Student Produced Recordings: To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student’s own private use.

Academic Integrity
You are expected to comply with the University of Pittsburgh’s Policy on Academic Integrity detailed here: http://www.provost.pitt.edu/info/ai1.html. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

Disabilities
If you have a disability for which you are requesting an accommodation, you are encouraged to contact both your instructor and the Office of Disability Resources and Services, 216 William Pitt Union, 412-648-7890 / 412-3837355 (TTY), as early as possible in the term. Disability Resources and Services will verify your disability and determine reasonable accommodations for this course.
Copyright

Course materials may be protected by copyright. United States copyright law, 17 USC section 101, et seq., in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials. See Library of Congress Copyright Office and the University Copyright Policy.

Competencies addressed in this course:
This course focuses on the Data Analysis competency domain. Every session covers elements of the Applied Analytic Techniques competency, with the following learning objective: “Determine and apply a range of appropriate statistical techniques to answer research questions and explain the implications of missing data on conclusions drawn from statistical results.”
Module 1: Introduction to Biostatistics

At the conclusion of this module, you should be able to:

- Understand the course objectives and organization
- Identify common mistakes or mistaken uses in statistics
- Apply descriptive statistics to the measure of location and variability
- Summarize data by using numerical and graphical methods

Before coming to class, complete the following online materials:

**Unit 1: Intro to Biostatistics**

<table>
<thead>
<tr>
<th>Watch</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lecture 1: Course overview</td>
</tr>
<tr>
<td>• Lecture 2: Statistics and common mistakes of using statistics</td>
</tr>
<tr>
<td>• Lecture 3: Descriptive statistics – measure of location and variability</td>
</tr>
<tr>
<td>• Lecture 4: Descriptive statistics – measure of variability</td>
</tr>
<tr>
<td>• Lecture 5: Data description – numerical summaries</td>
</tr>
<tr>
<td>• Lecture 6: Data description – graphical summaries</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Rosner, Chapters 1-2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Do</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Assignment #1</td>
</tr>
</tbody>
</table>

Module 2: Probability

At the conclusion of this module, you should be able to:

- Explain basic concepts of probability
- Use probability rules/ laws to solve problems
- Compute conditional probabilities
- Define sensitivity, specificity, PV+, and PV-
- Apply probability rules in screening tests

Before coming to class, complete the following online material:

**Unit 1: Probability I**

<table>
<thead>
<tr>
<th>Watch</th>
</tr>
</thead>
</table>

**Unit 2: Probability II**

<table>
<thead>
<tr>
<th>Watch</th>
</tr>
</thead>
</table>
Module 3: Discrete Probability Distributions

At the conclusion of this module, you should be able to:
- Understand concepts of a random variable and its distribution
- Calculate the mean and variance of binomial and Poisson distribution
- Calculate probability mass function (PMF) and cumulative distribution function (CDF) of binomial and Poisson distribution
- Tell the difference between binomial and Poisson distribution

Before coming to class, complete the following online material:

Unit 1: Discrete Probability Distributions

Watch
- Lecture 1: Discrete random variables and probability distributions preliminaries
- Lecture 2: The probability mass function
- Lecture 3: Cumulative distribution function
- Lecture 4: Binomial distribution
- Lecture 5: The Poisson distribution

Read
- Rosner, Chapter 4

Do
- Assignment #4
At the conclusion of this module, you should be able to:

- Calculate the expected value and variance of continuous random variable
- Explain the PDF and CDF of normal distribution
- Be familiar with the properties of normal distribution
- Use normal tables to solve problems
- Standardize any normal distribution
- Use normal distribution to approximate Binomial and Poisson distributions

Before coming to class, complete the following online material:

**Unit 1: Continuous Probability Distributions I**
- Lecture 1: Continuous random variables and probability distributions
- Lecture 2: Normal distributions
- Lecture 3: Using normal tables

**Unit 2: Continuous Probability Distributions II**
- Lecture 1: Properties of normal distribution
- Lecture 2: Standardization of a normal random variable
- Lecture 3: Normal approximation to binomial and Poisson distributions

**Read**
- Rosner, Chapter 5

**Do**
- Assignment #5

---

Module 5: Estimation

At the conclusion of this module, you should be able to:

- Explain concepts of sampling distribution
- Apply law of large numbers and central limit theorem
- Estimate the population mean and variance from a sample

Before coming to class, complete the following online material:

**Unit 1: Estimation**
- Lecture 1: Sampling distributions I
- Lecture 2: Sampling distributions II
- Lecture 3: Laws of large numbers and central limit theorem
- Lecture 4: Point and interval estimations of population mean
- Lecture 5: Point and interval estimations of population variance

**Read**
- Rosner, Chapter 6

**Do**
Module 6: Inference

At the conclusion of this module, you should be able to:
- Formulate null and alternative hypotheses for hypothesis testing
- Be familiar with test statistics and interpret $P$ value
- Make inference based on one-sample $t$ test
- Identify and represent type I error, type II error, and power on a plot
- Make inference based on two-sample hypothesis tests
- Explain the relationship between hypothesis testing and confidence interval

Before coming to class, complete the following two units:

<table>
<thead>
<tr>
<th>Unit 1: Inference I</th>
<th>Unit 2: Inference II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch</td>
<td>Watch</td>
</tr>
<tr>
<td>• Lecture 1: General concepts of hypothesis testing</td>
<td>• Lecture 1: Two-sample $F$ tests</td>
</tr>
<tr>
<td>• Lecture 2: Test statistics and $P$ value</td>
<td>• Lecture 2: Two-sample $t$ tests</td>
</tr>
<tr>
<td>• Lecture 3: One-sample $t$ test</td>
<td>• Lecture 3: Paired-sample $t$ tests</td>
</tr>
<tr>
<td>• Lecture 4: Graphical representations of type I error, type II error, and power</td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>Read</td>
</tr>
<tr>
<td>• Rosner, Chapter 7</td>
<td>• Rosner, Chapter 8</td>
</tr>
<tr>
<td>Do</td>
<td>Do</td>
</tr>
<tr>
<td>• Assignment #8</td>
<td>• Assignment #9</td>
</tr>
</tbody>
</table>

Midterm review

Midterm
Module 7: Sample Size and Power

At the conclusion of this module, you should be able to:

- Estimate power and sample size for one- and two-sample Z test
- Know the factors that affect power

Before coming to class, complete the following online material:

Unit 1: Sample Size and Power

Watch
- Lecture 1: Review type I and type II error and power
- Lecture 2: Power estimation for one-sample Z test
- Lecture 3: Factors affecting power
- Lecture 4: Sample size estimation

Read
- Rosner, Chapters 7.6, 7.7, and 8.9

Do
- Assignment #10

Module 8: Categorical Data

At the conclusion of this module, you should be able to:

- Conduct one- and two-sample tests of proportions and confidence intervals
- Estimate power and sample size for proportion tests
- Run chi-square and Fisher exact test for contingency tables
- Test equal proportions of one or two groups of binary data
- Analyze data from contingency tables
- Analyze matched binary data
- Measure agreement between two raters

Before coming to class, complete the following online material:

Unit 1: Categorical Data I

Watch
- Lecture 1: One-sample inference for proportions
- Lecture 2: One-sample interval estimation for proportions
- Lecture 3: Two-sample inference for

Unit 2: Categorical Data II

Watch
- Lecture 1: Power and sample size estimation for proportions
- Lecture 2: Chi-square test for 2*2 contingency tables
- Lecture 3: Fisher exact test for matched
<table>
<thead>
<tr>
<th>Lectures</th>
<th>Inference for count data</th>
<th>McNemar test for matched binary data</th>
<th>Kappa statistic for agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Read</strong></td>
<td><strong>Read</strong></td>
<td><strong>Read</strong></td>
<td></td>
</tr>
<tr>
<td>• Rosner, Chapter 10</td>
<td>• Rosner, Chapter 10</td>
<td>• Rosner, Chapter 10</td>
<td></td>
</tr>
<tr>
<td><strong>Do</strong></td>
<td><strong>Do</strong></td>
<td><strong>Do</strong></td>
<td></td>
</tr>
<tr>
<td>• Assignment #11</td>
<td>• Assignment #12</td>
<td>• Assignment #12</td>
<td></td>
</tr>
</tbody>
</table>

**Module 9: One-Way Analysis of Variance (ANOVA), and Correlation and Simple Linear Regression I**

**Before coming to class, complete the following online material:**
- Test equal means among more than two groups using one-way analysis of variance (ANOVA)
- Assess assumptions for ANOVA
- Know post-hoc procedures for the adjustment of multiple comparisons
- Describe when nonparametric tests are necessary
- Carry out proper nonparametric tests, when necessary

**Unit 1: One-Way Analysis of Variance (ANOVA)**
- **Watch**
  - Lecture 1: General concepts
  - Lecture 2: Hypothesis and ANOVA table
  - Lecture 3: Assumptions
  - Lecture 4: Post-hoc procedures
- **Read**
  - Rosner, Chapter 12
- **Do**
  - Assignment #13

**Unit 2: Correl. and Simple Linear Regression I**
- **Watch**
  - Lecture 1: Correlation coefficient
  - Lecture 2: Simple linear regression – model and estimations
  - Lecture 3: Simple linear regression – hypothesis testing
  - Lecture 4: Simple linear regression – ANOVA table
- **Read**
  - Rosner, Chapter 11
- **Do**
  - Assignment #14

**Module 10: Simple Linear Regression II, and Nonparametric Statistics**

**At the conclusion of this module, you should be able to:**
- Measure association between two continuous variables
- Fit and interpret simple linear regression models
- Understand residuals
- Use appropriate methods to assess the assumptions of a simple linear regression model
- Interpret residual plots

**Before coming to class, complete the following online material:**

**Unit 1: Correl. and Simple Linear Regression II**

<table>
<thead>
<tr>
<th>Watch</th>
<th>Unit 2: Nonparametric Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lecture 1: Confidence interval</td>
<td>• Lecture 1: Motivation of nonparametric tests</td>
</tr>
<tr>
<td>• Lecture 2: Prediction interval</td>
<td>• Lecture 2: Sign test and Wilcoxon signed rank test</td>
</tr>
<tr>
<td>• Lecture 3: Model diagnostics</td>
<td>• Lecture 3: Wilcoxon rank sum test and Kruskal-Wallis test</td>
</tr>
<tr>
<td></td>
<td>• Lecture 4: Spearman rank correlation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Read</th>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Rosner, Chapter 11</td>
<td>• Rosner, Chapters 9, 11.12, and 12.7</td>
</tr>
</tbody>
</table>

| Do                                                                   | Do                                                                                           |
|                                                                     |                                                                                               |
| • Assignment #15                                                    | • Assignment #16                                                                              |

**Module 11: Logistic Regression, Survival Data Analysis, and Correlated Data Analysis**

**Before coming to class, complete the following online material:**

- Describe the basic ideas of logistic regression
- Describe the basic ideas of survival analysis and correlated data analysis

**Unit 1: Logistic Regression**

<table>
<thead>
<tr>
<th>Watch</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lecture 1: Logistic regression</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No reading</td>
</tr>
</tbody>
</table>

| Do                                                                   |
|                                                                     |
| • No assignment                                                      |

**Unit 2: Survival/Correlated Data Analyses**

<table>
<thead>
<tr>
<th>Watch</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lecture 2: Survival analysis</td>
</tr>
<tr>
<td>• Lecture 3: Correlated data analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No reading</td>
</tr>
</tbody>
</table>

| Do                                                                   |
|                                                                     |
| • No assignment                                                      |
Final Exam review

Final Exam