BIOST 2011: Principles of Statistical Reasoning
Spring 2018

Course Instructor:	Shyamal Peddada 7126 Parran Hall E-mail: sdp47@pitt.edu
Instructor Office Hours:	Monday 3:00 – 3:50 pm, 7126 Parran Hall or by appointment
Teaching Assistants:	Kristina Boyd KLB223@pitt.edu Avantika Srivastava AVS51@pitt.edu Alex Hurd ALH231@pitt.edu
TA Office Hours:	Alex Hurd: Tuesday, 10:00am to 11:00am, 7121A Parran Hall Avantika Srivastava: Wednesday, 2:00pm to 3:00pm, 7121A Parran Hall Kristina Boyd: Thursday, 9:00am to 10:00am, 7121A Parran Hall
Class Time:	Monday 4:00 – 5:20 pm, G23 Crabtree Hall Wednesday 4:00 – 5:20 pm, G23 Crabtree Hall
Recitation:	**Register for one of the following:** Monday 5:30pm – 6:25pm, G23 Crabtree Hall (Instructor: Kristina Boyd) Tuesday 4:00 – 4:55 pm, A115 Crabtree Hall (Instructor: Avantika Srivastava) Thursday 1:00 – 1:50 pm, A522 Crabtree Hall (Instructor: Alex Hurd)
Prerequisites:	College Algebra with grade of C or better Working knowledge of PC or Mac computer ***Scientific calculator***: A scientific calculator is necessary for in-class exams. ***Computer Software***: The software for BIOST 2011 is **Stata** and is available on all university-supported computers in the recitations and libraries. Personal copies of **Stata** can be downloaded free through the University of Pittsburgh Software Licensing Services.
Required text:	Introductory Applied Biostatistics by Ralph D'Agostino, Sr., Lisa Sullivan and Alexa Beiser; ISBN # 978-0-534-42399-5
Recommended Text:	Using Stata for Quantitative Analysis by Kyle C. (Clayton) Longest; ISBN # 978-1483356631

Catalogue Description

This course acquaints students with the concepts of statistical reasoning as applied to the study of Public Health problems. Students learn the general principles of statistical analysis and acquire the ability to utilize a statistical software package (Stata) as a tool to facilitate the processing, editing, storing, displaying, analysis and interpretation of health research related data.

Course rationale

This is the Biostatistics core course for Graduate School of Public Health for non-Biostatistics majors and provides a basic introduction to the concepts of statistical reasoning as applied to the study of public health problems. This course is designed for public health students that expect to primarily to be able to read and understand statistical procedures in the form of books, journal articles, reports, grants, etc. The course will also give students the ability to perform some basic analyses. Students who intend to be professional research workers in public health areas requiring the daily application of quantitative procedures and statistics should consider taking BIOST 2041 and 2042 (Introduction to Statistical Methods I and II).

Learning Objectives

This course will help students to meet the Biostatistics competencies developed by the Association of Schools of Public Health (ASPH). By the end of this course, each student will be able to:

1. Select quantitative data collection methods appropriate for a given public health context.

- Identify the appropriate statistical procedures to be applied in different public health situations, especially in social science research.
- Identify advantages and challenges of working with different types of data from real public health examples, including ordinal, scale, and quantitative variables.

2. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate.

- Perform exploration of data, including descriptive tables and plots, basic transformations (e.g. categorizing a continuous variable), and inter-rater reliability, and be able to interpret the findings.
- Use statistical software to conduct hypothesis tests, including t-tests, ANOVAs, linear regression, analyses of proportions as well as the non-parametric versions of these tests.
- Perform (using statistical software) and interpret analyses such as computing confidence intervals, sensitivity and specificity analyses, simple linear regression models and basic survival analyses.
-

3. Interpret results of data analysis for public health research, policy or practice.

- Describe results of and draw conclusions from statistical analyses.
- Interpret and evaluate basic statistical analyses as performed in scientific manuscripts.

Ground Rules for Class

- Be respectful of your fellow students, teaching assistants, and instructors. This includes being on time and refraining from using your cell phone during class.

- Students may work together on homework assignments, but each student must submit their own homework. **Copying of another student's assignment for even one homework will result in a score of 0 points for all homework assignments.**

How to Succeed in this Class?

BIOST 2011 can be a challenging course. Here are some important things to remember to keep on track:

- Learning statistics is similar to learning a new language – it is done over time and with lots of practice! Come to class and attend your recitation. Just reviewing the course handouts via Courseweb will not be sufficient.

- It is extremely important to not get behind on readings and assignments. The course content is cumulative so if you get behind, it is very difficult to catch up.

- Although recitation attendance is optional, it is strongly encouraged. This is your opportunity to get feedback and help from the TA and solidify the concepts presented in lecture.

- Utilize the instructor and TA office hours.

- Don't feel shy to ask questions in class or during recitation. No question is a "stupid" question. If you did not understand something, chances are there are many others who did not either.

- Do all homeworks.

Course Website

All readings and course material will be found on the Blackboard site for this class. The website for Blackboard is http://courseweb.pitt.edu. Your login ID and password are the same as for your Pitt account.

Course Requirements

- Lecture attendance.
- Satisfactory completion and submission of all required assignments.
- Satisfactory performance on three exams.

Course grades are based on the following

- Exams (60%)

  You will take three in-class exams, each worth 20% of the overall grade. Exams are closed book and closed note except for one sheet of notes on 8.5"x11" paper, double-sided. This notesheet will be turned in with the exam. You should bring a calculator. The use of computers, cell phones or other internet-attached devices will **NOT** be permitted during

exams. The exams may consist of true/false, multiple choice, and short answer questions. Tentative dates for the exams are:

Exam #1 – Monday, February 12
Exam #2 – Monday, March 26
Exam #3 – Monday, April 23

These exams cannot be taken early or late without a compelling reason and supporting documentation. Students who cannot be present on the day of an exam will be required to take the exam on the first earlier available date. Notification must be given to the instructor in advance.

- Homework (30%)

  There will be ten homework assignments, graded for accuracy and completion. Homework assignments are due via the course website by 3:59pm on the due date, unless otherwise stated. No credit will be given for late assignments. Note: best 9 out 10 homeworks will count towards your grade.

- Discussion Posts (10%)

  There will be six discussion posts on blackboard, graded for completion. Discussion posts will be assigned and due via the course website by 3:59pm on the due date, unless otherwise stated.

Grade scale

98-100%	A+
90-97%	A
89%	A-
85-88%	B+
80-84%	B
79%	B-
70-78%	C
<70%	F

Academic Integrity

All students are expected to adhere to the school's standards of academic honesty. Any work submitted by a student for evaluation must represent his/her own intellectual contribution and efforts, except in the case of home works, which can be collaborative. The GSPH policy on academic integrity, approved by EPCC on 10/14/08, which is based on the University policy, is available online at http://www.publichealth.pitt.edu/Portals/0/Main/Prospective%20Students/Academics/Pitt%20Public%20Health%20Academic%20Handbook-Part%20IV%20G_AcademicIntegrity_AY2014-15.pdf

These guidelines are based on the University policy found here: http://www.provost.pitt.edu/info/acguidelinespdf.pdf

The policy includes obligations for faculty and students, procedures for adjudicating violations, and other critical information. Please take the time to read this policy.

Students committing acts of academic dishonesty, including plagiarism, unauthorized collaboration on assignments, cheating on exams, misrepresentation of data, and facilitating dishonesty by others, will receive sanctions appropriate to the violation(s) committed. Sanctions include, but are not limited to, reduction of a grade for an assignment or a course, failure of a course, and dismissal from GSPH.

All student violations of academic integrity must be documented by the appropriate faculty member; this documentation will be kept in a confidential student file maintained by the GSPH Office of Student Affairs. If a sanction for a violation is agreed upon by the student and instructor, the record of this agreement will be expunged from the student file upon the student's graduation. If the case is referred to the GSPH Academic Integrity Hearing Board, a record will remain in the student's permanent file.

Accommodation for Students with Disabilities

If you have any disability for which you may require accommodation, you are encouraged to notify both your instructor and the Office of Disability Resources and Services, 140 William Pitt Union (412-648-7890) during the first two weeks of the term (http://www.studentaffairs.pitt.edu/drswelcome).

Video / Audio recording of class lectures

Audio recording of the class is permissible provided you first receive approval from the course instructor. You also agree that the recording is for your own personal use and will not be redistributed in any form.

Video recording of the class, in full or in part, is NOT permitted.

Copyright of course material

Unless otherwise stated all course material is 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. As such the material is to be used for academic purposes only. Redistribution of this material to web sites and repositories (e.g., Course Hero) is strictly prohibited.

Course Calendar

Date	Topic	Corresponding Reading	Monday	Wednesday
Week 1 Jan 8 - Jan 12	Intro and Motivation, Summarizing Data	1.1-1.4, 2.1-2.3		
Week 2 Jan 15 - Jan 19	Probability	3.1, 3.4	MLK holiday	**HW #1 due**
Week 3 Jan 22 - Jan 26	Probability, Sampling Distributions	3.5, 4.1-4.3		**HW #2 due**
Week 4 Jan 29 – Feb 2	Intro to Hypothesis Testing, Procedures for μ	5.1-5.3		**HW #3 due**
Week 5 Feb 5 - Feb 9	Procedures for μ	5.1-5.3		**HW #4 due**
Week 6 Feb 12 - Feb 16	Procedures for μ1-μ2	6.1-6.3	**EXAM #1***	
Week 7 Feb 19 – Feb 23	ANOVA	9.1-9.7		**HW #5 due**
Week 8 Feb 26- Mar 2	Categorical Data	7.1-7.3		**HW #6 due**
Week 9 Mar 4 - Mar 11	spring break			
Week 10 Feb 12- Mar 16	Categorical Data	7.1-7.3		**HW #7 due**
Week 11 Mar 19 - Mar 23	Comparing Risks in Two Populations	8.1-8.4		**HW #8 due**
Week 12 Mar 26 - Mar 30	Correlation	10.1-10.2	**EXAM #2****	
Week 13 Apr 2- Apr 6	Regression	10.3-10.5		**HW #9 due**
Week 14 Apr 9- Apr 13	Logistic Regression Analysis	11.1-11.5		**HW #10 due**
Week 15 Apr 16 - Apr 20	Qualitative Analysis, Special topics	12.1-12.6, 13.1-13.3		
Week 16 Apr 23 - Apr 27	Final Exam		**FINAL EXAM**	

Please note: recitations are not included on this schedule but will be necessary to solidify concepts introduced in lecture and to help with homework.

*** Exam #1 covers material from weeks 1-5 (including 5)**

**** Exam #2 covers material from weeks 6-11 (including 11)**

***** Exam #3 covers material from weeks 12-15 (including 15)**

Course Objective Outline

Introduction and Motivation

- Describe the purpose and goals of this course
- Explain why statistics are used and what types of conclusions they provide
- Explain the difference between population parameters and sample estimates
- Identify population, sample, and hypothesis of a research study

Summarizing Data

- Distinguish between continuous and discrete variables
- Describe the shape, center, and spread of data
- Compute (using Stata) and interpret numerical summaries of data
- Generate (using Stata) and interpret graphical displays to summarize data

Probability

- Explain the concept of probability
- Compute simple probabilities from 2x2 tables
- Explain the binomial distribution and its properties
- Calculate binomial probabilities (using Stata)
- Explain the normal distribution and its properties
- Explain and calculate z-scores
- Calculate normal probabilities (using Stata)

Sampling Distributions

- Define sampling distribution
- Explain the central limit theorem

Introduction to Hypothesis Testing

- Identify the five key steps of a hypothesis testing procedure
- Describe the statistical hypotheses (i.e. identify the parameter of interest and the directionality of the hypotheses) based on a research question

Statistical Inference for population mean (μ)

- Estimate a population mean μ
- Compute and interpret a confidence interval for μ
- Explain the concept of precision
- Perform hypothesis tests for μ

Statistical Inference for difference of two population means ($\mu_1 - \mu_2$)

- Compare means from two different populations under different scenarios
- Distinguish between independent and dependent observations

Analysis of Variance

- Compare more than two means using ANOVA
- Distinguish between a fixed and random effect
- Explain the concept of multiple comparisons and when they apply

Categorical Data

- Estimate a population proportion and the difference in two population proportions
- Compute and interpret a confidence interval for p and for p_1-p_2
- Perform hypothesis testing for the comparison of two populations proportions
- Create cross-tabulation tables for discrete variables
- Compute and understand sensitivity and specificity measures
- Explain when to use a chi-square test
- Distinguish between goodness of fit tests and tests of independence

Comparing Risks in Two Populations

- Define and interpret an effect measure
- Compute and interpret a confidence interval for an effect measure
- Compute and interpret Chi-square tests of homogeneity and Fisher's exact tests

Correlation and Regression

- Estimate a population correlation coefficient ρ
- Compute and interpret a sample correlation coefficient r
- Perform hypothesis tests for ρ
- Explain the assumptions that must be met for linear regression
- Explain how to test these assumptions
- Fit a simple linear regression model (using Stata) and interpret the coefficients
- Fit a multiple linear regression model (using Stata) and interpret the coefficients

Logistic Regression Analysis

- Explain when it is appropriate to use logistic regression
- Fit a simple logistic regression model (using Stata) and interpret the coefficients
- Generate (using Stata) and interpret an ROC curve

Qualitative Analysis

- Explain the principles of mixed method evaluations
- Explain three common qualitative methods (observations, interviews, and focus groups) and explain the possible conclusions and limitations
- Compare and contrast the three methods above

Special Topics (depending upon time):

A. Nonparametric Tests
    - Use the nonparametric counterpart tests of the tests we've discussed in earlier lectures (using Stata)
    - Match parametric procedures with the nonparametric counterpart

B. Introduction to Survival Analysis
    - Generate and interpret a Kaplan-Meier curve
    - Compute and interpret median survival time