BIOS 702: Applied Biostatistical Methods I – Fall 2018

Instructor

Clemontina A Davenport, PhD
Biostatistician III, BERD Methods Core
Email: tina.davenport@duke.edu
Office: 11107 Hock Plaza (Suite 1104)
Telephone: 919-694-0332 (Hock)
919-668-1420 (NC Mutual)
Office Hours: Monday 3-4pm or by request

Teaching Assistant

Kara McCormack
Email: kara.mccormack@duke.edu
Office Hours: 11021A TH 4-5pm or by request

Course Website

sakai.duke.edu

Course Schedule

Classes: Monday/Wednesday – 11:45 a.m. – 1:00 p.m., Room 11025 Hock Classroom
First Class: Monday, August 27, 2018
First Exam: Wednesday, September 26, 2018
Second Exam: Wednesday, October 31, 2018
Last Class: Wednesday, November 28, 2018
Graduate Reading Period (No Classes): Saturday, December 2 to Tuesday, December 12
Final Exam: Thursday, December 13, 2018 – 2:00-5:00*
**Subject to change; do not schedule any travel before Tuesday, December 18, until the final exam has been confirmed*. E*arly exams will NOT be permitted.*

Course Description

This course is the first in a two-course sequence (BIOS 702 & BIOS 705) designed to introduce students to the most basic statistical ideas and methods commonly used by professional biostatisticians when addressing *real* questions using *real data* in the biological and medical sciences. While the focus is on applications, the appropriateness and rationale of different methods will link to the statistical theory taught in BIOS 701 and 703. It is a "hands-on" course, where the student will learn to use statistical software to implement the methods being taught.

Prerequisites

Multivariate calculus and linear algebra. Basic math skills are assumed. Proficiency in use of a laptop computer, including installing software, folder navigation, etc. is also assumed.

Co-requisites

BIOS 701, BIOS 703, and BIOS 721

Course Objectives

By the end of the course, the student should be competent to navigate through many of the stages of addressing a research problem by applying statistical methods to relevant data and interpreting the results in context of the problem and in language a collaborator would understand. Focus will be on choosing appropriate methods based on underlying assumptions and intent; understanding the strengths and weaknesses of various approaches relative to each other; skills needed to assess assumptions, implement the methods, and interpret results; and an appreciation of problems for which none of the methods covered are appropriate.[1]

Course Format

The course will include lectures, examples, discussions, and "hands-on" exercises. While you are welcome to bring your laptop to class to take notes and follow along with computational examples, this is not the time to be checking email, using social media, or browsing the web.

Text

There is no required textbook for this course. You may find the course notes to be sufficient, but if not, there are a number of appropriate textbooks available to you. Below are several suggestions.

Content Books:

Daniel WW and Cross CL (2013), *Biostatistics: A foundation for the analysis in the health sciences*, 10th Ed., Wiley.

Rosner B (2011), *Fundamentals of Biostatistics*, 7th Ed. Brooks/Cole, Cengage Learning.

Van Belle G, Fisher LD, Heagerty PJ, and Lumley T (2004), *Biostatistics: A methodology for the health sciences*, 2nd Ed., Wiley.

Computational Aspects:

Kleinman K, and Horton NJ (2010), *Using SAS for data management, statistical analysis, and graphics*, CRC Press, Taylor and Francis Group.

Kleinman K and Horton NJ (2014) *SAS and R: data management, statistical analysis and graphics*, 2nd Ed., CRC Press, Taylor and Francis Group.

Crawley M (2013), *The R Book*, 2nd Ed., Wiley.

[1] This course does not focus on skills needed to work with a collaborator to understand the goals and translate the problem into a question that can be answered using statistical methods. Those skills are addressed in BIOS 703 and BIOS 706.

Grades

The final grade will be a composite of grades for quizzes, homework and exams. The weights for each are given below. However, the instructor reserves the right to alter these a little…not by more than 5 percentage points in each group.

Quizzes (10%)

A short quiz (15 minutes) will be given at the beginning of some Wednesday classes. The quizzes will be graded and returned on Monday. The quizzes will count for about 10% of the final grade.

Homework (25%)

On most weeks, a homework assignment will be given, and will usually be due one week later (on the same day of the week). They will count for about 25% of your grade. These assignments are intended to give you practice implementing what you learn in class and writing clear explanations of the results. They are also intended to provide you with an opportunity to think more deeply about some of the concepts and methods we cover in class. We should have covered the material needed to do the homework before it is assigned, so there is no reason to procrastinate. Your homework should be uploaded to Sakai, and is due by 11:45am (start of class) on the due date. *Late assignments will not be accepted. No exceptions (see below).* You may collaborate with your classmates to get your software working correctly, but beyond that, your work should be your own. Focus will be on your written answers – clarity, accuracy, and completeness of your answers. Failure to be fully responsible for your own work (e.g., copying or paraphrasing someone else's answer) is a violation of the Duke honor code ("Duke Community Standard") given below, and will get you in trouble.

Exams (65%)

There will be two exams and one final exam – all are in class, closed book/notes/technology. The two exams will count for 20% each and the final will count for 25% of the final grade. The dates given above are subject to change, and changes will be announced in class and updated. As course content continues to build on earlier content throughout the course, you may consider each exam to be comprehensive in nature.

Missed Quiz or Homework

A missed quiz or failing to turn in a homework assignment by the due date/time will result in a grade of 0. At the end of the semester, the lowest quiz score and the lowest homework score will be dropped from calculation of the final grade. This is to adjust for unavoidable absences or emergencies since *late homework will not be accepted and missed quizzes will not be made up*. This is not a "free pass" to skip a class or ignore an assignment.

Grading Scale

A+: 97% or higher	B+: 87% - 89.9%	C+: 77% - 79.9%	D+ 67% - 69.9%
A: 93% - 96.9%	B: 83% - 86.9%	C: 73% - 76.9%	D: 63% - 66.9%
A-: 90% - 92.9%	B-: 80% - 82.9%	C-: 70% - 72.9%	D-: 60% - 62.9%

F: Less than 60%

Duke Community Standard

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

Software

The primary software to be illustrated and used in this course is R, but SAS code may be provided. You are free to use whichever you prefer and/or switch between them if you like. You are expected to type your homework (inserting relevant computer output, as needed) using either LaTeX or Microsoft Word.

- R: An environment for statistical computing and graphics. Freely available for Linux, Mac, and Windows from http://cran.r-project.org/
- SAS Statistical analysis software (version 9.4; STAT version 14.1 embedded within). Duke has a site license making SAS free for one year on a Windows or Linux platform, or via the web using the Duke Virtual Computing Lab. Mac OS X users can use the Duke Virtual Computing Lab or install a version of SAS to use within a Virtual Machine in the Mac OS operating system. More information about this is on Sakai.
- LATEX: A document preparation system. Freely available for a number of computer platforms. See "Installation" at http://en.wikibooks.org/wiki/LaTeX for details. The Biostatistical, Statistical, Mathematics, Computer Science and other technical fields often use this.
- Microsoft Word: A common word processing and document preparation system on multiple platforms. Duke provides a license for such. This is more commonly used by those researchers with whom you will collaborate.

Topic List

During this class, we will cover the following general topics – but not in the exact order presented below.

1. Basic concepts and nomenclature of research data (e.g., random variable, population, sample, sampling frame, statistic, parameter, response/outcome/dependent variable, predictor/independent variable, covariate)
2. Overview of types of study designs (observational vs. experimental studies, prospective vs. retrospective)
3. Measurement scales of data (discrete: nominal vs. ordinal, continuous: interval and ratio scales; time-to-event data)

4. Concepts related to probabilistic thinking (e.g., study population, parameter, sample, statistic, underlying population distribution, sampling distribution)
5. Graphical and numerical summaries of data.
6. Characteristics and properties of commonly assumed distributions.
7. Point and interval estimates; summary measures in both discrete and continuous data.
8. Concepts related to making valid inferences (e.g., observational vs. experimental studies, replicates, research hypotheses, null and alternative hypotheses, roles of underlying "truth", evidence, models)
9. Parametric and non-parametric statistical tests concerning a measure of location in a single sample
10. Parametric and non-parametric statistical tests comparing a measure of location in two independent samples.
11. Parametric and non-parametric statistical tests when comparing a measure of location in paired samples.
12. Inference (statistical testing) when assessing the value of a single proportion (hypothesis testing on a parameter representing a probability).
13. Inference (statistical testing) when comparing the values of two proportions (independent and paired)
14. Statistical inference in the comparison of two variances.
15. Contingency table (count data) layout for data arising from different types of sampling (paired, unpaired, stratified, multiple populations)
16. Effect estimates related to count data (e.g., odds ratios, risk ratio, risk difference, sensitivity and specificity of a screening test, positive and negative predictive value of a screening test).
17. Use of the Z, t, Chi Square, and F tables to find critical values (quantiles) and p-values
18. Introduction to linear regression