

## BIOS 600: Principles of Statistical Inference

Section 001, Fall 2017, 3 credits

Lecture: Tuesday / Thursday: 9:30 – 10:45

Lab (optional): Monday: 11:15 – 12:15

0001 MHRC (Blue Cross & Blue Shield Auditorium)

### Course Introduction

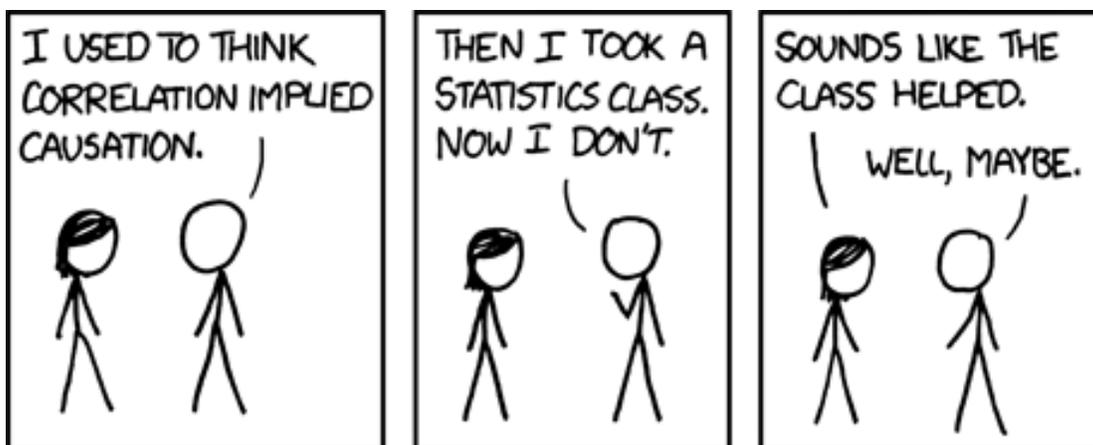
---

Instructor	E-mail	Office Hours
Yue Jiang	yuejiang@live.unc.edu	0001 MHRC Lobby, Thurs, 1:00–3:00
Teaching Assistants	E-mail	Office Hours
Kevin Donovan	kmdono02@live.unc.edu	0001 MHRC Lobby, M/W, 9:00–10:00
Yutong Liu	yutongl@live.unc.edu	0001 MHRC Lobby, Tues, 1:00–3:00
Homework Grader	E-mail	
Kayla Kilpatrick	kkilpat@live.unc.edu	

BIOS 600 is an introductory course in statistics and data science with strong emphasis on applications in public health and medical research. The course is motivated by applications in global public health, which appear frequently throughout the lectures, homeworks, labs, and help sessions.

Topics include descriptive statistics, probability, and statistical inference, including construction of confidence intervals, hypothesis testing, power and sample size calculations, analysis of contingency tables, diagnostic tests and their properties, linear and logistic regression, and an introduction to more advanced topics (e.g., nonparametric methods, and survival analysis).

BIOS 600 meets core competencies recognized by CEPH. Students will use modern statistical software to conduct analysis, with focus on R and Stata. Reading for understanding and translation of statistical results into language accessible to other health science researchers will be stressed.



Source: Randall Munroe, xkcd. <https://xkcd.com/552/>

## **Learning Objectives**

---

- Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect statistical inference.
- Learn to conduct statistical analyses in modern statistical software and interpret analysis results in context.
- Model numerical response variables of different types using one or more explanatory variables in order to learn about relationships between variables.
- Understand how to evaluate which statistical methods are appropriate in answering a research question.
- Learn to evaluate statistical usage in public health and medicine, critiquing data-based claims and evaluating data-based decisions.

## **Council on Education for Public Health (CEPH) Core Competencies**

---

- Describe the roles biostatistics serves in the discipline of public health.
- Distinguish among different measurement scales and the implication for selection of statistical methods to be used.
- Apply descriptive techniques appropriate for summarizing public health data
- Understand basic concepts of probability, random variation, and commonly used probability distributions.
- Apply appropriate statistical methods for inference.
- Understand appropriate methodological strategies according to the type of study design for answering a particular research question.
- Apply descriptive and inferential methods according to the type of study design for answering a particular research question.
- Interpret results of statistical analyses in public health studies.
- Develop written and oral presentations based on statistical analyses for public health professionals and educated lay audiences.
- Apply basic techniques to describe public health characteristics and in public health research and evaluation using vital statistics and public health records.

## Prerequisites

---

- Basic understanding of college algebra
- Willingness to learn R or Stata.

Students may evaluate their math readiness using the *Quantitative Self-Test* on the BIOS 600 FAQ website. This self-test contains a mathematics assessment and resources for math review. BIOS 600 assumes no previous coursework in probability or statistics.

## Suggested Primary Resources

---

- Pagano and Gauvreau, Principles of Biostatistics, 2nd edition (used is fine!)
- Laptop computer or tablet for class
- (Free!) Relevant papers and videos as assigned

## Suggested Supplementary Resources

---

- (Free!) Diez, Barr, and Cetinkaya-Rundel, *OpenIntro Statistics*
- (Free!) Zuur, Ieno, and Meesters, A Beginner's Guide to R

OpenIntro Statistics is available as a free download at the authors' website as hyperlinked above. The guide to R is a free e-book available via UNC Libraries. Note that a student solutions manual for the Pagano and Gauvreau textbook exists. However, you may not necessarily find it useful because only certain problems are solved (and I would not ask you to work them for homework anyway).

## Software

---

R and Stata are the primary software packages for BIOS 600 and were chosen to accommodate student needs in the Gillings School. SAS will not be formally or officially supported, though students are welcome to use it to complete homework assignments or labs. Students will also need to register for a Poll Everywhere account.

- (Free!) Stata option 1: Use Stata via the *UNC ITS Virtual Lab*. Note that to use this resource, students must have an active internet connection. As well, this resource may occasionally be unavailable due to a limit on the number of simultaneous users.
- Stata option 2: Students may order Stata via *UNC Software Acquisition*. Under the Stata Grad-Plan, 'Intercooled Stata' is available for \$125 per year, or \$198 for no expiration date. If you are planning to work in Stata in the absence of an internet connection, consider this option.
- (Free!) R: Available at this *link*.
- (Free!) R Studio: Available at this *link*. R Studio is a free and highly recommended user interface for R. This will make your work in R much easier.

## **Honor Code**

---

Students in BIOS 600 are expected to abide by the UNC Honor Code. All suspected Honor Code violations will be reported to the UNC Dean of Students, who will investigate the case. These investigations typically involve lengthy hearings of the Honor Court, and as outlined in the Instrument of Student Judicial Governance, "The usual sanction for a first academic violation is definite suspension for at least one academic semester and a grade penalty of an 'F' for the course, a portion of the course, or the assignment."

Regarding homework: you are welcome and encouraged to discuss strategies of problem solving with each other and with the TAs, but the work you turn in must be your own. Copying answers will be considered a violation of the UNC Honor Code.

Regarding exams: during the in-class portion of the exams, absolutely no electronic devices, including calculators, are permitted. During the take-home portion of the exams, students may consult any materials with the exception of other students or TAs. All exam questions are only to be directed to the course instructor.

## **Homework**

---

Homework will be assigned approximately weekly and is required. The focus of homework assignments will be on honing skills in using statistical software to analyze data and presenting and interpreting results of data analysis, though other questions will also be used to test understanding.

Homework will be completed online in Sakai. Sakai allows repeated submissions with grading of numeric/multiple choice (but not essay or text box) responses. That is, you may resubmit and correct mistakes on numeric/multiple choice questions. However, the final graded homework will always be the last homework submission on Sakai (careful - no exceptions!), so it is a good idea to save your homework in a separate file until you are ready to submit the final version.

There will be no possibility of making up missed homework assignments, *even if you have a good reason*. Any assignments turned in late will be given a score of 0. However, the lowest homework score will be dropped before semester grades are calculated, and so one homework excuse is automatically granted (note: you will not see this reflected online in Sakai).

## **Lab Reports**

---

Attendance in lab is *optional*, though suggested. The objective of lab is to provide you with hands-on experience conducting data analysis using modern statistical software. As part of lab, you will learn the skills needed to complete these analyses successfully - students have often found that skills practiced/learned during labs have made homeworks easier. During labs, you will work in teams of your choice of up to six students, with the entire team engaging in discussion and answering of questions.

The lab period is Monday 11:15 AM – 12:15 PM in the same room as lecture and is run by TAs. Lab reports are due by Thursday at 9:30 AM, and will be graded and returned the following lab session. Note that the optional lab period is not a time to ask teaching assistants questions regarding homework assignments.

## **Exams**

---

Exams will consist of an in-class and a take-home portion. Absolutely no electronic devices, including calculators, are permitted during the in-class portion, which will be video recorded for academic integrity purposes. The take-home portion of the exams will be due electronically 48 hours after the conclusion of the class period of the in-class portion, with no exceptions a prior written arrangement with the instructor is made. For the take-home portion, students may consult any materials they wish with the exception of other students or TAs. All questions on the take-home portion are to be directed to the course instructor.

I understand that life is not always predictable, and that exams may need to be rescheduled. Please make every effort to contact me by e-mail in advance if this is the case. Make-up exams are not guaranteed if I am not contacted in advance except under extreme extenuating circumstances (e.g., hit by a bus Regina George style). Note that rescheduled exams will be different from in-class examinations and may involve different formats (e.g., essay or oral).

If you need any special accommodation for an exam, you must be registered with *UNC Accessibility Resources and Services*, who will then provide me with an official letter. There will be no exceptions to this rule.

## **Participation: Poll Everywhere**

---

Poll Everywhere will be used to gauge student understanding and participation throughout the course. Students must register to use Poll Everywhere (via phone, laptop, tablet, etc.) at *UNC Poll Everywhere FAQ*. It is your responsibility to be in class on time, with a device registered for Poll Everywhere, and to respond to the polls. You get credit for responding, whether or not you have the correct answer. To get the full credit for Poll Everywhere, you must complete at least 90% of the polls for the semester; otherwise, your credit will be prorated by percentage of polls answered.

## **Class Etiquette**

---

Behavior in and out of the classroom should enhance the learning process. At all times we will use common courtesy and respectful behavior. Attendance is expected - you are responsible for any announcements or content delivered in the classroom, regardless of whether you are in attendance. If you know in advance of a major conflict and would like a class recording, let us know, and we can request it on your behalf. We have done this in the past for religious holidays, medical school, interviews, APHA, and other conflicts.

Occasionally due to class pacing reasons, we may be required to defer a question to office hours after class. Please be understanding in this case - it does not mean that I think your question is a bad one (there are no bad questions!). In fact, it may just be involved, or we may be running behind.

## **Where Do I Go For Help?**

---

The best way is to come to the instructor's and/or teaching assistants' office hours! If you are unable to make it, for course content/homework questions, e-mail is also fine – start with the TAs. For questions regarding course logistics, e-mail the instructor. We will do our best to answer all e-mailed questions within 2 business days. For additional assistance, students may check out the *Academic Enrichment Program*, a free resource at Gillings that provides academic support in a group setting for introductory BIOS courses (see the website linked above for more details).

## Assignment of Grades

---

Each student's numeric grade will consist of 75% exams, 20% homework, and 5% participation. The lab component is optional and is worth 0% (but does provide very helpful practice).

The third exam will be held during the final exam and will primarily focus on material from the last third of the course (though naturally, all material is cumulative). Participation grade is based solely on Poll Everywhere as described above.

Grading Scheme	
Homework	20%
Exam 1	25%
Exam 2	25%
Exam 3	25%
Participation	5%

The Graduate School uses the grades H (clear excellence), P (entirely satisfactory), L (low pass), and F (failure). Graduate students are expected to earn 'P' grades, with remarkable performances rewarded with other grades as appropriate. I reserve the right to curve grades using more generous cut points depending on the overall difficulty of assignments and examinations. Under no circumstances will a student's grade be curved downward.

Undergraduate Scale			
92.5	≤	A	≤ 100
90.0	≤	A-	< 92.5
87.5	≤	B+	< 90.0
82.5	≤	B	< 87.5
80.0	≤	B-	< 82.5
77.5	≤	C+	< 80.0
72.5	≤	C	< 77.5
70.0	≤	C-	< 72.5
65.0	≤	D+	< 70.0
60	≤	D	< 65.0
0	≤	F	< 60.0

Graduate Scale			
92.5	≤	H	≤ 100
70.0	≤	P	< 92.5
60.0	≤	L	< 70.0
0	≤	F	< 60.0

## Modification of Syllabus

---

The instructor reserves the right to make changes to the syllabus, including homework due dates and exam dates. All changes will be announced as early as possible and posted to Sakai.

## Assessment of Instruction

---

Course evaluations will be available online in early December. It is an expectation that all students complete the course evaluations in a timely manner. Please use this forum to give valuable constructive feedback for planning future BIOS 600 courses. However, feel free to raise any issues or concerns with me at any other time during the semester.

## Schedule of Classes

---

### Unit 1: Introduction to Data

---

T - 08/22	Reproducibility and study design	
R - 08/24	Types of variables and graphical displays	
T - 08/29	Data summary measures	

### Unit 2: Probability and Distributions

---

R - 08/31	Basics of Probability	
T - 09/05	Conditional Probability	
R - 09/07	Diagnostic tests, sensitivity, and specificity	HW 1 due
T - 09/12	Binomial distribution	
R - 09/14	Continuous distributions	
T - 09/19	Normal (Gaussian) distribution	

### Unit 3: Foundations for Inference

---

R - 09/21	Sampling distribution of the mean	HW 2 due
T - 09/26	<b>Exam 1</b> in class (Units 1-2; material through 09/19)	
R - 09/28	Confidence intervals	
T - 10/03	Hypothesis testing	HW 3 due
R - 10/05	P-values on trial	

### Unit 4: Inference for Continuous Variables

---

T - 10/10	Comparing two means	
R - 10/12	No class (University Day ceremony)	
T - 10/17	Power and sample size	HW 4 due
R - 10/19	No class (Fall Break)	
T - 10/24	ANOVA	
R - 10/26	Nonparametric methods	HW 5 due

### Unit 5: Inference for Categorical Variables

---

T - 10/31	Analysis of proportions	
R - 11/02	<b>Exam 2</b> in class (Units 1-4; material through 10/26)	
T - 11/07	Comparing categorical variables	HW 6 due
R - 11/09	Comparing categorical variables	

### Unit 6: Modeling

---

T - 11/14	Correlation	HW 7 due
R - 11/16	Linear regression	
T - 11/21	Multiple regression	
R - 11/23	No class (Thanksgiving)	
T - 11/28	Logistic regression	HW 8 due
R - 11/30	Survival analysis	
T - 12/05	Review	
T - 12/12	<b>Exam 3</b> at 8:00 AM (cumulative final exam)	