

## BIOS 600: Principles of Statistical Inference

Fall 2012, 3 credits

### Location and Time:

Class: 9:30-10:45 Tuesday and Thursday, 133 Rosenau Hall

Labs: Thursday or Friday in HSL 307; see below for times

**Instructor:** Amy Herring, Professor of Biostatistics

Office: 3104-D McGavran-Greenberg

e-mail: [aherring@bios.unc.edu](mailto:aherring@bios.unc.edu)

office hours: Tuesdays and Thursdays 10:45-11:15 and by appointment

| Teaching Assistant      | E-mail <sup>†</sup>  | Lab Session   | Office Hour & Location |
|-------------------------|--|---------------|------------------------|
| Andy Ni                 | <a href="mailto:andyni@live.unc.edu">andyni@live.unc.edu</a>       | 3:00-3:50 F   | 3:00-4:00 M McG 1303   |
| Kyung Su Kim            | <a href="mailto:kskim@email.unc.edu">kskim@email.unc.edu</a>       | 10:00-10:50 F | 2:00-3:00 Tu McG 1302  |
| Sebastian Teran Hidalgo | <a href="mailto:shidalgo@email.unc.edu">shidalgo@email.unc.edu</a> | 1:00-1:50 F   | 3:00-4:00 Tu McG 1302  |
| Lia Weiner              | <a href="mailto:liaj@live.unc.edu">liaj@live.unc.edu</a>           | 1:00-1:50 Th  | 2:00-3:00 W HC 0015    |
| Ho Jin Yang             | <a href="mailto:hojiny@live.unc.edu">hojiny@live.unc.edu</a>       | 2:00-2:50 Th  | 8:00-9:00 Th McG 1303  |
| Elizabeth Koehler       | <a href="mailto:elizk@live.unc.edu">elizk@live.unc.edu</a>         | 2:00-2:50 F   | 11:30-12:30 F McG 1303 |

<sup>†</sup>You are welcome to e-mail us at any time. However, please use the Sakai Discussion Forum to post questions related to homework or course material more generally so that others can benefit from any clarifications that are provided.

BIOS600 is an introductory course in probability theory and statistical inference 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, and laboratory sessions. Topics include descriptive statistics, construction of confidence intervals, hypothesis testing, power and sample size calculations, analysis of contingency tables, diagnostic tests and their properties, and an introduction to numerous more advanced topics (linear and logistic regression, nonparametric methods, survival analysis, and sampling theory). Students will use statistical software to conduct analysis. Reading for understanding and translation of statistical results into language accessible to other health science researchers will be stressed.

### Class Participation:

Class participation is strongly encouraged! **Poll Everywhere** will be used to gauge understanding and facilitate participation. To use the software, students will need to have access either to a cell phone with texting capabilities or a laptop with wireless internet. Directions on setting up your account are on the [help.unc.edu](http://help.unc.edu) website; please register before Thursday.

**Learning Objectives:**

- Gain a basic working knowledge of important statistical topics including (but not limited to) elementary probability theory and probability distributions, random variation, descriptive statistics, graphical displays of data, basic hypothesis testing, regression, and nonparametric methods.
- Understand how to evaluate which methods are appropriate in answering a research question for a given study design.
- Learn to conduct basic statistical analysis in modern statistical software and interpret analysis results.
- Learn to evaluate straightforward statistical usage in public health and medicine, with a focus on relevant research publications.
- Gain the understanding of basic concepts necessary to interact knowledgeably with biostatisticians in planning, conducting, analyzing, and reporting public health and medical research (and know how to determine when a biostatistician should be consulted).

**Prerequisites:**

- Basic understanding of algebra and arithmetic (at the level of college algebra). Evaluate your math readiness using the [Quantitative Self-Test](#) on the Sakai website. This self-test contains a mathematics assessment (that will not affect your course grade) and resources for math review. BIOS 600 assumes no previous coursework in probability or statistics.
- Access to Stata or another statistical software package (SAS, R, etc.) for data analysis. Stata tutorials will be led by the TA's early in the course, and other software will not necessarily be supported by TA's or the instructor.

**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 at lectures and labs is expected. You are responsible for any announcements or content delivered in the classroom, whether you are in attendance or not.
- Please make every effort to arrive on time, and I promise to do so as well. If you must arrive late or leave early, please do so quietly.

- Use of electronic devices should be limited to viewing course notes and using Poll Everywhere. I am happy to take important phone calls and texts on your behalf should this be necessary. Online shopping, games, and other distracting electronic behaviors may result in confiscation of your device, which can be picked up after class from me on your first infraction and from your department chair in the case of repeated infractions.
- Eating in class is permitted only if you bring enough to share with all 150 of us. Properly contained beverages should not be a problem.
- Occasionally due to class pacing concerns, I 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 was a bad one (there are no bad questions!), in fact it may just be involved, or we may be running behind.

#### **Texts:**

- Primary Resources
  - Pagano and Gauvreau, Principles of Biostatistics, 2nd edition
  - Relevant papers as assigned
  - Access to cell phone with text messaging or laptop with wireless internet during class time
- Supplemental Resources
  - Pagano and Gauvreau, Student Solutions Manual for Principles of Biostatistics (I feel obliged to inform you of the existence of this book, but I do not think you will necessarily find it useful because (a) only certain problems are solved, and (b) I would not ask you to work those for homework anyway!)
  - Kelley and Donnelly, The Humongous Book of Statistics Problems: Translated for People Who Don't Speak Math (Not required, but some students from the Fall 2011 BIOS 600 course recommend it as a supplement.)
  - Gonick and Smith, The Cartoon Guide to Statistics, if you find yourself wanting more!

**Software:** Stata is the primary software for BIOS 600. Most TA's will be able to support SAS and R for many tasks. Excel should not be used.

- Stata Option A (free): Use Stata via the [UNC ITS Virtual Lab](http://virtuallab.unc.edu) at <http://virtuallab.unc.edu>. Drawbacks include the fact that sometimes this resource may not be available due to a limit on the number of simultaneous users. We will use Stata by connecting to the Virtual Lab during learning lab sessions.
- Stata Option B (fee): Students can order Stata at a greatly reduced price by linking to UNC Software Acquisition. Options range from 'Small Stata' starting at \$32 for 6 months to 'Intercooled Stata' at \$98 for one year. Small Stata will be sufficient for BIOS 600, but if you plan to use Stata for any other purpose, you should strongly consider the standard (intercooled) version. You can see a comparison of the two versions on the Stata website.

### **Grading and Assessments:**

#### **Honor Code:**

Students in BIOS600 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."

#### **Homework:**

Homework will be assigned approximately weekly and is required. Much of the homework will be completed online in Sakai. You are welcome to discuss strategies of problem solving with each other and with the TA's or myself, but the work you turn in must be your own. Copying answers will be considered a violation of the UNC honor code.

Assignments will be based on numerous important health studies, including the China Health and Nutrition Survey (PI: Dr. Barry Popkin, UNC). We thank the China Health and Nutrition Survey, funded by NIH (R01-HD30880, DK056350, and R01-HD38700), the Carolina Population Center, the Chinese CDC, Dr. Barry Popkin, and Mr. Jim Terry for providing these data.

There will be no possibility of making up missed homework assignments. However, the lowest two homework scores will automatically be dropped before grades are averaged. (You will not see this reflected online in Sakai.)

**Exams:**

All exams are in-class. Absolutely no electronic devices are permitted during exams. Exams will be video recorded for academic integrity.

I understand that life is not always predictable and that examinations may need to be rescheduled. Please make every effort to contact me by e-mail in advance if this is the case. Make-up examinations are not guaranteed if I am not contacted in advance (of course, if you are in the emergency room, I will understand!). Note that rescheduled exams will be different from in-class examinations and may involve different formats (e.g., essay or oral).

*Important note:* 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.

**Assignment of Grades:**

Each student's numeric grade will be 25% exam 1, 25% exam 2, 25% exam 3, and 25% homework. The grading scale is given below. I reserve the right to curve grades using more generous cutpoints depending on the overall difficulty of the assessments. The brackets indicate the letter grade that is provided to a borderline score (for example, a grade must be at least 90.0 to get a grade of A-; any grade  $\geq 87.5$  but less than 90.0 is a B+).

| Numerical<br>Grade | Undergraduate<br>Letter<br>Grade | Graduate<br>Letter<br>Grade |
|--------------------|----------------------------------|-----------------------------|
| $\geq 92.5$        | A                                | H                           |
| [90.0, 92.5)       | A-                               | P                           |
| [87.5, 90.0)       | B+                               | P                           |
| [82.5, 87.5)       | B                                | P                           |
| [80.0, 82.5)       | B-                               | P                           |
| [77.5, 80.0)       | C+                               | P                           |
| [72.5, 77.5)       | C                                | P                           |
| [70.0, 72.5)       | C-                               | P                           |
| [65.0, 70.0)       | D+                               | L                           |
| [60.0, 65.0)       | D                                | L                           |
| $< 60.0$           | F                                | F                           |

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 from the scale as appropriate. Class participation (including poll participation rates) may modify the association between the numeric average and assigned letter grade in exceptional cases.

## Schedule

| Date                   | Topic  |
|------------------------|--|
| Tuesday, August 21     | Introduction, reproducibility                  |
| Thursday, August 23    | Data presentation                              |
| Week 1 Lab             | Stata computing basics                         |
| Tuesday, August 28     | Data summary measures                          |
| Thursday, August 30    | Elementary probability theory                  |
| Week 2 Lab             | Measures of location and spread                |
| Tuesday, September 4   | Conditional probability                        |
| Thursday, September 6  | Diagnostic tests, sensitivity, and specificity |
| Week 3 Lab             | Diagnostic tests and ROC curves                |
| Tuesday, September 11  | Binomial and Poisson distributions             |
| Thursday, September 13 | Case studies: probability                      |
| Week 4 Lab             | Case studies: calculating probabilities        |
| Tuesday, September 18  | Normal (Gaussian) distribution                 |
| Thursday, September 20 | Exam 1 (in class)                              |
| Week 5 Lab             | Case study: normal (Gaussian) distribution     |
| Tuesday, September 25  | Sampling distribution of the mean              |
| Thursday, September 27 | Confidence intervals                           |
| Week 6 Lab             | Case study: confidence intervals               |
| Tuesday, October 2     | Hypothesis testing                             |
| Thursday, October 4    | Hypothesis testing                             |
| Week 7 Lab             | Hypothesis tests about the mean                |
| Tuesday, October 9     | Power calculations                             |
| Thursday, October 11   | Comparing two means                            |
| Week 8 Lab             | Comparing means                                |
| Tuesday, October 16    | Analysis of variance (ANOVA)                   |
| Thursday, October 18   | Fall break                                     |
| Tuesday, October 23    | Nonparametric methods                          |
| Thursday, October 25   | Case study: comparison of means                |
| Week 10 Lab            | Case study: ANOVA                              |
| Tuesday, October 30    | Exam 2 (in class)                              |
| Thursday, November 1   | Analysis of proportions                        |
| Week 11 Lab            | Case study: proportions                        |
| Tuesday, November 6    | Analysis of contingency tables                 |
| Thursday, November 8   | Analysis of contingency tables                 |
| Week 12 Lab            | Case study: Contingency Tables                 |
| Tuesday, November 13   | Correlation                                    |
| Thursday, November 15  | Linear regression                              |
| Week 13 Lab            | Case study: correlation and regression         |
| Tuesday, November 20   | Multiple regression                            |
| Thursday, November 22  | Thanksgiving holiday                           |
| Tuesday, November 27   | Logistic regression                            |
| Thursday, November 29  | Survival Analysis                              |
| Week 15 Lab            | Bringing it all together                       |
| Tuesday, December 4    | Sampling                                       |
| Tuesday, December 11   | Final exam, 8:00am                             |