

BIOS 600, Section 2, Fall 2013
Principles of Statistical Inference
Course Information

Robert M. Hamer, Ph.D., Professor of Biostatistics and Psychiatry

Class Location and Time: 8:00 AM-9:15 AM, Tuesdays, Thursdays, Rosenau 0133 (Auditorium)

Recitation Section Location and Time: 3:30 – 4:50 Tuesday, McGavran-Greenberg 2306

(The recitation section required, and will be devoted to computer applications, and help sessions)

Professor

Robert M. Hamer, Ph.D.

Office:	Medical School Wing B Room 336
Email:	hamer@unc.edu
Phone:	919 843 5508
Office Hours:	By appointment

Teaching Assistants: To be arranged. The TAs will have regular office hours, and you can always see me by appointment. Also, the class meets starting at 8 AM on Tuesdays and Thursdays. I am usually sitting in the Hooker Wing Atrium after about 7 AM on Tuesdays and Thursdays and you can see me then. The TAs should be your first line of inquiry about the class material.

Introduction and Welcome

Welcome on Bios 600, Principles of Statistical Inference. While some of you, or perhaps nearly all of you, are here because this is a required course, it is my hope that during this semester I will convince you of at least the following three things: (1) The world in which we live throws quantitative and statistical information at us, day after day, minute after minute, and the ability to interpret and understand this information is essential to functioning well in our *personal* and *professional* lives, (2) although this information may seem simple at times, there are often hidden complexities which we need to automatically be able to recognize and think through in order to use this information easily and correctly, and (3) an attitude of skepticism towards nearly everything we read and hear will serve us well in trying to sort out truth from fiction, and this applies specially to science. I am not accusing scientists of deliberately misleading us, but contend that often the articles we read, both in our scientific journals and in the lay press, often have not probed deeply enough into the information they present to present an accurate picture.

Richard Feynman (if you don't know who Richard Feynman was, use your favorite search engine or Wikipedia) said at the 1974 Caltech commencement address, "*The first principle is that you must not fool yourself - and you are the easiest person to fool.*" You should take this to heart, because when we want to believe something, we often if not usually manage to perceive reality in such a way that it confirms what we want to believe. He also said, "*For a successful technology, reality must take precedence over public relations, for Nature cannot be fooled.*" We can substitute "science" for "technology" here, or whatever it is we do, and the concept holds. We cannot fool Mother Nature, and the harder we try, the longer and more difficult it is to find out we were trying.

Finally, we are going to have a discussion of history and philosophy of science. Not a long one, but I believe a useful one. Although Professor Feynman also said, "Philosophers say a great deal about what is absolutely necessary for science, and it is always, so far as one can see, rather naive, and probably wrong," I don't believe that he meant we can get away with "doing" science, or "using" science, or "reading" science without some understanding of what science is, why we do things the way we do, and how to interpret the results.

I leave you with a quote from *Daniel Patrick Moynihan* (find out who he was, too), who said, "Everyone is entitled to his own opinion, but not his own facts." We have to be very careful in distinguishing our facts from

our opinion. I'd add a corollary (or is it a lemma?): that picking and choosing which facts to which to attend and which to ignore is the same as making up one's own facts.

Oh, and one more thing. In October, 2005, Stephen Colbert introduced a word in his debut broadcast. The word was "truthiness." The Merriam-Webster dictionary defines truthiness as:

1 : "truth that comes from the gut, not books" (Stephen Colbert, Comedy Central's "The Colbert Report," October 2005)

2 : "the quality of preferring concepts or facts one wishes to be true, rather than concepts or facts known to be true" (American Dialect Society, January 2006)

I am afraid that much of what we read in our scientific and other professional journals falls under the heading of truthiness, and not, unfortunately truth. Maybe even most of it. Probably even most of it. I do not want to participate in turning you into cynics. I do want to help you be skeptics, to be able to think critically about what you are reading, and to attempt to separate the wheat from the nice-sounding chaff. I would like to help you understand not to contribute to the noise in our literature.

EMAIL: My email address is hamer@unc.edu. However, unfortunately, I get lots of email. Today, August 12, as of 2:41 PM, I have received 32 emails, not counting whatever I've deleted or sent to my spam folder. If I don't recognize the name of the sender, and the subject line is not informative enough to make me certain that the email is not spam, I sometimes delete the email without reading. Even if the subject is informative, if I'm particularly busy, I might just put off reading it. Over the course of the day, it might then scroll off the screen and out of my mind. Therefore, to be sure I don't miss email from you, please make the first three words of the subject line of email you send to me "BIOS 600 EMAIL" so that, for example, if you send me an email asking whether section 5.3 is covered, the subject line might be "BIOS 600 EMAIL Section 5.3". That will help me try not to neglect email from this class.

[UNC's Honor Code](#)

"It shall be the responsibility of every student at the University of North Carolina at Chapel Hill to obey and to support the enforcement of the Honor Code, which prohibits lying, cheating or stealing..." Because I was an undergraduate student here, and a graduate student here, the honor code means a great deal to me. (I would hope it would mean a great deal to me even had I not been a student here.) I will not be present in the room during midterms and the final exam; I will be in the Hooker Research Wing Atrium, available for questions, but not obsessively watching you to be sure you're not cheating. You are on your honor not to cheat.

Grading (subject to slight change):

Homework: 0%

(There is homework and you should do it, because that's how you learn, but we don't grade it. If you don't do the homework you won't be able to do the quizzes and exams and won't do well in the course. If there are several repetitive problems assigned and you feel comfortable, skip the others. Try doing problems not assigned.)

Weekly Quizzes 10% (drop the 3 lowest scores including zeros if some are not taken **no makeup quizzes**),
Midterm 40%(takes two class periods),
Final: 50%

(Quizzes, Midterm and Final are closed book, but you will be allowed one 8.5 by 11 sheet (both sides) of formulae or other text per quiz and all of your quiz formula sheets can be used cumulatively for the midterm and final examinations.)

NOTE THAT THIS MEANS THAT DURING QUIZZES, MIDTERMS, AND FINAL EXAMINATIONS YOUR COURSE NOTES SHOULD BE CLOSED AND YOUR BOOK SHOULD BE CLOSED. TABLES ARE ON-LINE, AND YOU SHOULD PRINT THEM AND BRING THEM TO CLASS; THUS THERE IS NO NEED FOR THE BOOK DURING EXAMS. ADDITIONALLY, CELL PHONES, TABLETS, AND COMPUTERS ARE NOT ALLOWED DURING EXAMS. THUS, COMPUTERS, CELL PHONES, AND TABLETS SHOULD BE LEFT UNOPENED DURING QUIZZES, MIDTERMS, AND EXAMS.

The final average (10% quizzes, 40% midterm, 50%final) will be used to assign a letter grade for the course as follows:

Graduate Grade	Final Average		Undergraduate Grade	Final Average
H	90 to 100		A	95 to 100
P	70 to <90		A-	90 to < 95
L	65 to < 70		B+	85 to < 90
F	Below 65		B	80 to < 85
			B-	75 to < 80
			C+	72 to < 75
			C	68 to < 72
			C-	65 to < 68
			D+	63 to < 65
			D	60 to < 63
			F	Below 60

Grades will not be curved. I have gotten pretty good over the years at writing test items which are of the same difficulty, so if there are differences between classes in the grade distribution from year to year, it is most likely due to actual differences in the class, rather than the tests. If you are prepared, and if you study and do the homework, it ought to be very difficult to fail this course. (It might be difficult to get an H, too.) You will notice that the “P” range covers a reasonably wide range, from 70 to 90%. That reflects my feeling that you have done acceptably well if you get 70% of the material correct, but haven’t done outstandingly well unless your average is over 90%. I’m a little more lenient with undergraduates in the lowest C, a C-, goes down to 65%. **You will be able to drop the course after the midterm, and I would urge students who have not done well on the midterm to consider dropping and retaking the course. I do not enjoy giving out F, L, or D grades.** I have had semesters where I’ve failed no one because those who would have failed dropped the course. I have had semesters where I failed students who did not drop the course and should have dropped it. Other than that, the grades fall where they fall.

Quizzes: Quizzes occur approximately weekly when we have covered an entire chapter or major unit, usually after 2 lectures, and on the day when we start the next chapter or major unit. Dates are listed in the syllabus. They comprise the first 10 minutes of classes on the days they occur. This is to ensure that (a) you are not late for class, and (b) you come to most classes and thus have a better chance of learning the material. We drop the 3 lowest quizzes to allow for religious holidays, illness, or a just plain off day. That means that if you miss some of the quizzes, you will get zero for those quizzes, and those zeros are the lowest grades. That means you can miss up to three quizzes and have those resulting zero scores dropped. *There are no make-up quizzes.* I’ll repeat that, because every semester I get at least one student asking to make up a quiz: **There are no make-up quizzes.** Repeat after me: there are no make-up quizzes. Quizzes are graded on a 10-point scale. If you have been keeping up with the material, you should have no trouble doing well on the quizzes. Due to Fall Break, while quizzes will occur at the beginning of the semester on Tuesdays, they will switch to Thursdays after the break. (The academic calendar may be viewed at <http://registrar.unc.edu/academic-calendar/>)

Homework: There are weekly homework assignments, which we will not collect. Although homework does not count toward your grade, if you can’t do the homework, it is unlikely that you’ll be able to do the quizzes, and unlikely that you will be able to do the midterm and final, and likely that you will fail the course. Most of the homework assignments are exercises from the book, but a few of them involve material which will be posted on the web.

Midterm and Final: The midterm and final examination are multiple choice. I have had as many as 120-140 students in this class, and I'm simply not going to grade that many problems. The midterm will take two class periods and will consist of 50 questions total, and the final exam will consist of 50 questions total. These questions are all multiple choice questions, and usually have 4 or 5 choices. I understand that some of you would rather have problem sets, but I just really can't do it. I may decide to grade some computer assignments and/or essays.

Computer Usage: There will be no computer usage required in this course at the moment. If you wish to use computers for the homework assignments, we encourage you to do so. Many of you are quite good at doing calculations with Excel. Just remember that you won't have computers in class for the midterm and final; all you'll have is notes, paper, and your calculator. Further, if you get a fancy calculator (one that does many statistics automatically) you may carry the arithmetic to more decimal places than I do when I construct the items, and you may get a different answer than I. If you do, that is your responsibility. I am of mixed mind about the use of computers in a class like this. Unpleasant as it may be, actually doing the arithmetic (with a calculator) does make sure you understand the steps to follow. I hate doing the arithmetic with a calculator, both because it is tedious and because if I make an error entering a number, I may never know it. That's why when I do arithmetic with a calculator; I usually do it at least twice, just to make sure I get the same answer each time. If we had software that you could use to take the drudgery out of it, but still have to follow all the steps in the computations, that might be useful, but of course, that would still leave the problem of how you would do the QUIZZES, midterm, and final.

Calculators: Get a calculator. You will need it for the homework assignments, quizzes, midterm and final. It does not have to be a complicated calculator; the minimal requirements are that it have the usual four functions, a square root key, and an exp(X) key (sometimes written as " e^x "), and at least one memory. A small number of assignments and exam questions will require that you calculate exp(X), which is the base of the natural log system raised to a power. Some calculators have a natural log key and an inverse key, and exp(X) is the inverse of the natural log (and visa versa). I think calculators like this cost \$10 or less these days. If you get a fancy calculator that can do t-tests, standard deviations, regression, etc., and rely on that instead of learning to do the calculations, you will probably simply not know how to do many of the test questions and may get the test questions wrong. You can't use the calculator built into your cell phone (if you have one) or your PDA, because these things must remain closed during quizzes, midterms and the final. You should carry calculations out to the precision of your calculators although the answers may be rounded in the multiple choices.

Text: Daniel, Wayne W., Biostatistics: A Foundation for Analysis in the Health Sciences. Tenth Edition. New York: Wiley, 2013. **Datasets from the book can be downloaded from the publisher's website.** This book, like all statistics books, and indeed like most textbooks, is unreasonably expensive. I have been told that it is possible to get this book over the web, from international sources, more cheaply, but that'll depend on such things as the exchange rate and shipping charges. If any of you find a really cheap place to get the book, I'd like to know. You really do need the Tenth Edition, so page numbers and exercises will match. I know and you know that the idea of a new edition every 3-4 years is really a rip-off designed by publishers to make all the used books worthless, but unfortunately, that's the world in which we live. (You might be able to live with an earlier edition if you can get access to the Tenth Edition for homework assignments.)

Other Interesting and Perhaps Useful Books:

The Cartoon Guide to Statistics, Gonick and Smith ISBN 0-06-273102-5

(No kidding, this book contains very clear discussions of statistical concepts.)

Statistics as Principled Argument, Robert P. Abelson ISBN 0-8058-0528-1 (Paperback)

Interesting Material on the Web: The Khan Academy (<http://khanacademy.com>)

The Khan Academy is an extraordinarily interesting site which contains several thousand approximately 10 minute videos of lectures, including about 100 on probability and statistics. They cover most (but not all) of the material I cover in this course, although somewhat more informally than I would. A person named Salman

Khan does all the lecturing. You can read about him both in Wikipedia and on the web site itself. You cannot simply substitute these lectures for coming to class, if for no other reason that whenever one teaches an elementary statistics course one simplifies some ambiguous definitions (like what a “large” sample is), and different statisticians or different books may make different decisions.

There is a great deal of supplementary material on the web page, and I may add more during the semester. You may find it educational and we may discuss it. Or you may find it completely uninteresting. I would strongly urge you to read the material on philosophy of science from the book by Maxwell and Delaney that I have posted; it is unfortunate that we “do science” without any real idea of the logical and philosophical underpinnings of what we do.

Important Remarks (at least I think they are):

(1) Taking a first biostatistics course has aspects of learning a new language. Although I allow you to make a sheet of paper corresponding to each quiz, and to use them cumulatively on the midterm and final examination, there is a great deal of memorization which is still required. If you don't know which formula you need for a particular problem, or don't know the symbols for a term, or don't know precisely what a term means, then all the formulas listed on a piece of paper won't help. You won't know what formula to look for; you won't understand the differences between similar formulas and symbols, or when to use which. You have to understand the material, not just use your formula sheets, and not just memorize terms. Memorization, understanding, and precision are all essential to learning and using biostatistics correctly. I try very hard (even if I don't always succeed) to be very precise in how I put things in this class.

(2) It is traditional to say that there is no such thing as a stupid question. Of course there is. However, it is far more likely that any question you wish to ask will be reasonable, and not stupid. I am very open to questions during class, because I need to understand if I am getting the material across, and if you need to ask a question, then possibly someone else needs to ask the same question. Please feel free to ask questions. The worst I'll do is to tell you we should cover that after class, or to ask one of the TAs, etc.

(3) One of my goals is to help you digest and use quantitative and statistical information you encounter in the popular press as well as your scientific literatures critically and skeptically. It is astonishing how much of the quantitative information we read in newspapers, magazines, and our scientific journals is just plain nonsense. I hope to make you very critical consumers of such information.

Final Remark:

Many of you are taking this course only because you are required to. I urge you to remember that throughout your careers, you will be reading articles that contain statistical information. It will do you a great deal of good to be able to understand it and digest it critically, rather than simply allow your eyes to skim over it and skip to the discussion section. Since so much of what is done with data in the literature is illogical, wrong, inappropriate, and misinterpreted, you cannot rely on the fact that something was published to be able to infer that it is true or useful. In terms of enjoying the material, although for some of you it might be difficult, it will (I really mean this) become easier if you open your minds to the aesthetic beauty inherent in the material, which is about using mathematics (reasonably simple mathematics) to make system out of chaos and make educated guesses about important things we care about when we can't measure them. As a final remark, I will tell you that I am not much of a mathematician. When I got past concrete material such as calculus and linear algebra, differential equations, and finite mathematics, to courses where one really dealt with abstract objects, manipulating them in abstract ways, I found it difficult and didn't do very well. I found it interesting and beautiful, but I didn't do very well. I would like you to convince yourselves that this material we are about to cover is interesting and beautiful; if you can do that, it will be easier for you. As a bonus, it'll make you a better backgammon, craps, or card player. Although I do not have set office hours, I will be very happy to talk with you.

I hope you have an enjoyable voyage in this course.