Text


[The text is recommended, but not required; you might very well find that you are glad to have it now or in the future, but you will not directly need it during the semester.]

Supported software

Minitab version 18/19, Minitab, Inc.

[The package is available available on the Stern network and for rental through the website http://estore.onthehub.com. I highly recommend that you either purchase or rent the package. You must have ready access to it during the semester.]

Anonymous quote of the day

“If I had only one day left to live, I would live it in my statistics class — it would seem so much longer.”

The goals of this course

The following conversation is one that I have had many times when meeting someone for the first time:

Them: So, what do you do for a living?
    Me: I’m a faculty member in the business school at NYU.

Them: Oh, really? What do you teach?
    Me: Statistics.

Them: I had a statistics course in college — I hated it!

The most important purpose of this course is to try to make sure you never say something like that!

For many people, the word “statistics” elicits one (or more) of three impressions: sports statistics (interesting for the fan, but ultimately not of much real–world value), lots of dry numbers that are probably useful to the specialist (but thank goodness I’ll
never have to worry about them!), and carefully chosen and manipulated figures used by politicians to fool the electorate into voting for them (or against the other side). As is true of most generalizations, there is an element of truth to this point of view, but it unfortunately masks a much more important truth. Quantitative reasoning (sometimes called numeracy, in analogy with the word literacy) has become crucial for everyday living, and essential for the practice of business. Statistical reasoning and methodology provide the tools to become numerate.

The underlying principle of this course is that the world is full of randomness, and the only way to understand that randomness is to examine it systematically. We will talk about various statistical methodologies, and of course I hope that you know how to use these methodologies when the course is over. Even more importantly, however, I hope that you will know how to think about randomness, and about data. This is a very applied course — we will talk about applications of statistics in many different fields, both business–related and non–business–related. You will see many analyses of real data, and you will spend lots of time doing your own statistical analyses of real data using the computer and learning to interpret the results of those analyses. You will spend relatively little time learning (or memorizing) formulas. We will talk about practical issues in how to think about randomness, and will discuss some basic probability (the language of randomness). We will talk about practical issues in how to think about data, and will discuss graphical and methodological ways to highlight what is going on in data. Finally, we will discuss ways to summarize relationships in data using statistical models, and demonstrate the ability to highlight structure in data by doing so.

The idea that is ultimately most important to grasp in a class like this is not any specific methodology, but rather the principle of statistical inference. What is statistical inference? One definition is that it is probabilistic generalization from data. This simple phrase summarizes all of the reasons behind the structure of the course. First, probabilistic refers to recognizing and expressing the uncertainty that is part of any inference. This can only be done by knowing something about probability, which is the language of randomness, so we will spend some time talking about probability. Second, generalization refers to the point that we are usually interested in claims that go beyond the data that we have; we want to know about what we don’t have yet, not what we already have. This depends on knowing how to apply statistical methods that are tuned to the questions at hand, but calculation is less important than understanding, since often statistical software can do the calculation work for us. Prediction, as opposed to description, is usually a big part of useful generalizations. Third, from data refers to the fact that we must be explicit about the evidence that is actually available to us. In statistics, context always matters; this is in direct contrast to mathematics, which is the study of objects independent of context (the rules and strategies of geometric or logical proofs are the same, no matter the context). Mathematics is about moving from the general to the specific (deduction), while statistics is about moving from the specific to the general (induction), and that is impossible without understanding the natures of both.

Administrative structure of class
The grade in this class will be based on a total of 260 points that can be earned, in the following ways:

© 2019, Jeffrey S. Simonoff
(1) Two noncumulative tests (∼75%). The first test will be on Thursday, October 31, from 10:30 AM – 12:00 noon. The second test will be given during the final exam period on Monday, December 16, from 9:00 AM – 11:00 AM. No makeups will be scheduled for the tests, so make sure that you do not miss them. In particular, be sure that you do not arrange for a flight that leaves New York before the scheduled final exam date and time. I will be grading the tests, and I will be giving out practice problems before it, as well as the last two years’ tests, to help you in your preparations. If you have a qualified disability and will require academic accommodation during this course, please contact the Moses Center for Students with Disabilities (CSD, 212-998-4980) and provide me with a letter from them verifying your registration and outlining the accommodations they recommend. If you will need to take an exam at the CSD, you must submit a completed Exam Accommodations Form to them at least one week prior to the scheduled exam time to be guaranteed accommodation.

(2) Homeworks (∼25%). Note: you MUST do the homeworks! Failure to do the homeworks can result in a penalty to your grade greater than 25% of the grade! Even more importantly, you will discover that doing the homeworks is by far the best way to learn the material and prepare for the examinations. Late homeworks will be subject to progressively larger penalties based on the number of days late the homework is handed in. Assignments will not be accepted after the answer sheet has been given out in class. You should also show your work, or your thought processes, when doing the homeworks, since you might otherwise lose some or all credit. The homeworks will be graded by the teaching assistant, and if you have a question about grading, you should go to her first to discuss it.

In an April 1998 memo the Dean’s Office mandated that grades in core course classes follow a distribution where no more than 35% of the class receives A or A-.

The use of laptop/notebook computers will not be permitted during class. They are very distracting to other students, and the fact is that they are usually not being used for purposes related to class. If you wish to use a laptop computer to take notes during class, you must inform me in writing of your intention to do this (e-mail is fine), and then you must send me (via e-mail) a copy of the notes that you took in class that day within one hour of the end of each class.

The Stern Code of Conduct states that you commit to “Exercise integrity in all aspects of our academic work including, but not limited to, the preparation and completion of exams, papers and all other course requirements by not engaging in any method or means that provides an unfair advantage.” Further, you commit to “Refrain from behaving in ways that knowingly support, assist, or in any way attempt to enable another person to engage in any violation of the Code of Conduct. Our support also includes reporting any observed violations of this Code of Conduct or other School and University policies that are deemed to have an adverse effect on the NYU Stern community.” This applies to this class in the following specific ways (in addition to general prohibitions on cheating, plagiarism, and so on):

© 2019, Jeffrey S. Simonoff
(1) I encourage you to ask me any questions you wish, on any subject related to the course, in class, in my office, or by e–mail. If there is some reason that I can’t answer the question, I’ll let you know.

(2) Not only are you allowed to work with classmates on homework, I encourage you to do so. I do ask that each person turn in their own copy of the homework, however. Obviously, you should actually have worked on any assignment you turn in under your name, whether by yourself or with other people.

(3) The examinations will be open book and open notes. You should bring a calculator to the exam (you will not be permitted to share a calculator with someone else), and are welcome to bring any books, notes, and tables you wish (I will not be providing any tables to you, for example, so bring whatever you think you might need). You will not be permitted to bring a laptop computer, and you will not be permitted to use any wireless device during the examination, such as cellular phones or tablets. I will give you copies of exams from recent years to help in your studying.

I strongly urge you to bring in to me examples of statistics, probabilistic reasoning, and so on, that you see in newspapers or magazines, whether they are directly relevant to material being discussed in class or not. Such material might very well then be incorporated into class discussion.

Syllabus

Here is a version of the syllabus separated by topic, with more details of each topic:

I. Applied Probability
   A. Basic concepts of probability — definitions of probability, conditional probability, independence [SF chapters 7, 8]
   B. Random variables and their properties — definition, probability distribution, mean, variance, covariance [SF chapters 9, 10]
   C. Specific distributions — binomial [SF chapter 5 section 2]
   D. The normal (Gaussian) distribution [SF chapter 11]

II. Statistical Inference
   A. Sampling distributions and the Central Limit Theorem [SF chapter 14]
   B. Point and interval estimation — confidence interval for the mean, prediction interval for a future observation, confidence interval for a proportion [SF chapter 15]
   C. Hypothesis testing — structure of tests, tests for the mean, tests for a proportion [SF chapter 8, chapter 16]

III. Regression Analysis [SF chapters 21–24]
   A. Assumptions of regression — the linear model, the principle of least squares, assumptions
   B. Inference — determination of estimates, \( t \)-tests, \( F \)-test, \( R^2 \), prediction
   C. Checking assumptions — residual plots, diagnostics
   D. Multiple regression — the model, inference, interpretation of coefficients, colinearity, model selection

© 2019, Jeffrey S. Simonoff
E. Hypothesis testing — comparison of groups (independent samples) [SF chapter 17]

The only material you will need to do the homeworks and prepare for the midterm and final exams is in (the two parts of) the Course Supplement. YOU SHOULD BRING THE COURSE SUPPLEMENT WITH YOU TO EVERY CLASS SESSION!