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:

© 2018, Jeffrey S. Simonoff
(1) Two noncumulative tests (~75%). The first test will be on Tuesday, October 30, from 10:30 AM – 12:00 noon. The second test will be given during the final exam period on Monday, December 17, 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):

© 2018, 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, collinearity, model selection

© 2018, Jeffrey S. Simonoff
E. Hypothesis testing — comparison of groups (independent samples) [SF chapter 17]

The following is a list of the handouts that are available at the class web site, separated by broad coverage. The two parts of the Course Supplement, which will be handed out in class are based on these handouts, but are not identical to them. **The only material you will need to do the homeworks and prepare for the midterm and final exams is in the Course Supplement. YOU SHOULD BRING THE COURSE SUPPLEMENT WITH YOU TO EVERY CLASS SESSION!**

**Background material**
- Data presentation and summary
- Material related to the Foundations of Finance course
- Populations, samples, surveys and statistical inference
  - *Extra data analysis problems*

**Probability**
- Probability — the language of randomness
- Simpson’s paradox
  - *Extra probability problems*

**Random variables**
- Random variables and their properties
- The Binomial distribution
- Continuous distributions
- The normal approximation to the binomial
  - *Extra random variables problems*
- The Central Limit Theorem

**Inference and interval estimation**
- Interval estimation and statistical inference
- Confidence interval for a binomial proportion
  - *Extra inference and interval estimation problems*

**Hypothesis testing**
- Hypothesis testing
- Tests about binomial proportions
  - *Extra hypothesis testing problems*

**Regression**
- Regression analysis
- Youth and the 2008 presidential election
- Getting what you pay for: dinner prices in Manhattan
- Statistical analysis using *Microsoft Excel*
  - *Extra regression problems*

**Two–sample hypothesis testing**
- Two–sample hypothesis tests
- The box office of movie sequels
- Periodontal conditions and coronary artery disease
From *The Words of Mathematics. An Etymological Dictionary of Mathematical Terms Used in English.* (S. Schwartzman)

**statistic** (noun), **statistics** (noun), **statistical** (adjective), **statistician** (noun): the Latin verb *stare* meant the same as its English cognate *stand*; both are descended from the Indo–European root *sta-* “to stand.” From the Latin verb came the noun *status*, literally “a standing, a condition.” The word *state*, which comes from *status*, is therefore a synonym of *condition*. A nation as a whole is also called a state because it is composed of the set of all conditions in a certain geographic area. As a result the term *state* came to be associated with government and politics. The German word *Statistik* was used in the 18th century to refer to political science, which has to do with states and governments. Since the study of political science involves the accumulation of data about the conditions in a country, each individual piece of data came to be known as a statistic, and the collective plural *statistics* came to refer to all the data and their collection and interpretation. Although the etymology of the word statistics is complicated, the meaning of the underlying Indo–European root is still apparent: statistics tell you how things “stand.”

**statistics** (noun): Fiction in its most uninteresting form.
— Evan Esar (*Esar’s Comic Dictionary*)

“When you can measure what you are speaking about, and express it in numbers, then you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind.”
— Lord Kelvin

“Statistics are like a bikini. What they reveal is suggestive, what they conceal is vital.”
— Aaron Levenstein

“To understand God’s thoughts we must study statistics, for these are the measure of His purpose.”
— Florence Nightingale

“You cannot feed the hungry on statistics.”
— David Lloyd George

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”
— H.G. Wells

© 2018, Jeffrey S. Simonoff
“La estadistics, otra mas que nos egaña (Statistics, yet another mistress to deceive us).”
— Spanish proverb

“No matter how much reverence is paid to anything purporting to be ‘statistics,’ the term has no meaning unless the source, relevance, and truth are checked ... The worship of statistics has had the particularly unfortunate result of making the job of the plain, outright liar that much easier.”
— Tom Burnan (The Dictionary of Misinformation)

“Statistics play the same role for a policymaker as a lamppost does for a drunk — for support, rather than illumination.”
— Andrew Lang

“Statistics can be used to support anything — especially statisticians.”
— Franklin P. Jones

“A knowledge of statistics is like a knowledge of foreign languages or of algebra; it may prove of use at any time under any circumstances.”
— Arthur L. Bowley

“It has been said that figures rule the world. Maybe. I am quite sure that it is figures which show us whether it is being ruled well or badly.”
— Johann Wolfgang von Goethe

“A statistician is a person who draws a mathematically precise line from an unwarranted assumption to a foregone conclusion.”
— Lord Schon

“Lord, please find me a one–armed statistician — so I won’t always hear ‘On the other hand . . . !’
— Edmund Muskie, as paraphrased by Kenneth Hammond and Leonard Adelman

“There’s too much abstract willing, purposing,
In this poor world. We talk by aggregates
And think by systems and being used to face
Our evils in statistics, are inclined
To cap them with unreal remedies
Drawn out in haste on the other side.”
— Elizabeth Barrett Browning (Aurora Leigh, Eighth Book)

“Statistics are the poetry of science.”
— F. Emerson Andrews

© 2018, Jeffrey S. Simonoff
“It is indeed easy to lie with statistics, but it is impossible to tell the truth without them.”
— Adrejs Dunkels

“Statistics means never having to say you’re certain.”
— Myles Hollander

“Statistics may be as dull as ditchwater, but they are the essential source from which the River of Knowledge must spring.”
— John Berry

“Before the curse of statistics fell upon mankind we lived a happy, innocent life, full of merriment and go, and informed by fairly good judgment.”
— Hilaire Belloc

“Figures won’t lie, but liars will figure.”
— Charles H. Grosvenor

“The imaginative statistician, far from being a narrow specialist, is often the direct opposite — a person who can walk into almost any scientific or commercial operation, find a numerical way to analyze what is going on, and increase output ... The gains to the economy and productivity of a country — any country, at any level of industrialization — from such mathematical use of intelligence are so great compared to the cost involved that a sound and imaginative statistical generalist ranks high among heavy intellectuals.”
— John A. Wheeler

“Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power in dealing with complicated phenomena is extraordinary.”
— Sir Francis Galton

“Statistics is, or should be, about scientific investigation and how to do it better, but many statisticians believe it is a branch of mathematics... Now I agree that the physicist, the chemist, the engineer, and the statistician can never know too much mathematics, but their objectives should be better physics, better chemistry, better engineering, and in the case of statistics, better scientific investigation. Whether in any given study this implies more or less mathematics is incidental.”
— George E.P. Box

“There are three kinds of lies: lies, damned lies, and statistics.”
— attributed to Benjamin Disraeli by Mark Twain
The Need to Understand Statistical Techniques

Three statisticians and three MBA students were traveling by train to a conference. The statisticians asked the students whether they bought tickets. The students said that they had. “Fools!,” said the statisticians. “We’ve only bought one between us!”

When the conductor appeared, the statisticians hid together in the bathroom. The conductor knocked, and they passed the ticket under the door. He clipped the ticket and slid it back under the door to the statisticians.

The MBA students were very impressed, and resolved to adopt this technique themselves. On the return trip they purchased one ticket between them, and shared the trip with the statisticians, who again asked whether they’d all bought tickets. “No,” they replied, “we’ve bought one to share.”

“Fools!,” said the statisticians, “we’ve not bought any.”

“But what will you do when the conductor comes?”

“You’ll see.”

This time when the conductor appeared, the MBA students hid together in the bathroom. The statisticians walked up to the door and knocked on it. The MBA students slid their ticket under the door, and the statisticians took it and used it as before — leaving the MBA students to be caught by the conductor.

The moral of this story is that you should never use a statistical technique unless you are completely familiar with it.