Stern School of Business
Statistics and Data Analysis
COR1-GB-1305.06

Professor: William Greene, Departments of Economics and IOMS

BS Ohio State University, 1972 (Operations Research); MA Wisconsin, 1974 (Economics); PhD Wisconsin, 1976 (Econometrics); History: Cornell 1976-1982; Real world, 1982-1983; Return to ivory tower at Stern (then GBA) NYU, 1983-; Robert Stansky Professor, 2010-. Permanent affiliations: University of Lugano, American University, University of Sydney, Monash University, University of Queensland, Curtin University, Queensland University of Technology, University College London. Publications: Articles - see vita on home page; Books: *Modeling Ordered Choices*, 2010, *Econometric Analysis*, *7th Ed* (2012); *Applied Choice Analysis* (2006,2015); Software, *NLOGIT* (www.nlogit.com), Editor in Chief, *Foundations and Trends in Econometrics*. Editor in Chief, *Journal of Productivity Analysis*, Associate Editor, *Journal of Economic Education*, *Journal of Choice Modeling*, *Economics Letters*. Research interests: econometric methodology, discrete choice modeling, efficiency and productivity analysis, health economics, transportation, nonlinear estimation, entertainment and media.
Office: MEC 7-90, Ph. 998-0876, Fax. 995-4218
Email: wgreene@stern.nyu.edu
Home Page: http://people.stern.nyu.edu/wgreene

Abstract

This course has two broad objectives: (1) This course will provide students with an understanding of fundamental notions of data presentation and analysis. We will develop tools to enable students to use statistical thinking in the context of business problems. The course deals with modern methods of data exploration (partly to reveal unusual or problematic aspects of data sets), the uses and abuses of the basic techniques of statistical inference, and the use of linear regression as a tool for management and financial analysis. (2) There is randomness everywhere in life and in the environment. We will develop models of probability and random variables that help to understand the randomness of everyday life and the business environment.

Prerequisites

I will assume that students are familiar with routine algebra, exponents and logarithms as well as graphical tools such as the slope and intercept of a straight line. Algebra will be used freely throughout the course. We may have rare occasions to use calculus, but these will be sparing at most.

Course Requirements and Course Grades

Final grades for the course will be determined on the basis of the following components and weights:

* Mid-term exam: 30%.
  (Sample midterm questions) (**2007 Midterm Exam with Solutions**) (**2008 Midterm Exam with Solutions**) (**2009 Midterm Exam with Solutions**)
* Final exam: 30%**.**
  (**Sample problems for study for the final**) (**Notes for sample problems**) (**2007 Final Exam with Solutions**) (**2008 Final Exam with Solutions**) (**2009 Final Exam with Solutions**)
* In class short (10 minutes) quizzes 15% (5 @ 3%) 15% (in aggregate)
* Homework assignments (details below): (5 @ 3%) 15% (in aggregate). Students may work in groups of up to 4 and submit a single report for the group.
* Model development project (details below) 10%. Students may work in groups of up to 4 and submit a single report for the group.

Official policy at Stern mandates that grades in core classes follow a distribution in which no more than 35% of students receive A or A-. Homework assignments are mandatory. Late submissions will be accepted only with a persuasive justification, but not after the solutions have been posted. All examinations are open book, open notes, closed telephone, closed PDA, closed iPhone, closed iPad, closed Droid, closed laptop, closed tablet, open mind. Do bring a conventional hand calculator (not a cell phone that includes a calculator) to both exams. Links to copies of past exams appear below.

Honor Code: Of course.

Course Materials

* **Text** for this course is *Statistics for Business: Decision Making and Analysis (Pearson, 2nd Edition)* by Robert Stine and Dean

Foster: (http://www.amazon.com/Statistics-Business-Decision-Analysis-Edition/dp/0321836510).

* **Software** for the course will be Minitab, Release 17. Version 17 is available from the Professional Bookstore for about $100, but you can rent a copy for about $30 for the semester from the website, http://www.onthehub.com/minitab - at the site, click on e-Store. (**Introduction to Minitab**) (**How to use Minitab on the Stern Citrix Server**) (**Useful notes on Minitab**)

Other Stuff

* Please remember to turn off your cell phone before you come to class.
* Please try to arrive early. Late entrances are disruptive.
* As a general rule, laptops are an annoyance during class, particularly when you are checking your email, playing with Facebook, tweeting or watching YouTube videos while others are studying statistics. If you absolutely must use your laptop to take notes, please be respectful of the interests of your colleagues.
* There will be no makeups for the quizzes.

Course Outline and Schedule

Materials: (**Introductory notes for the course - Notes 0: Introduction - Right click to download**

Session 1: Introduction to statistics; data description and presentation; types of data; Minitab.
Reading: Text:, Chapters 1 and 2, Sections 3.1-3.3, 4.1-4.3, Pages 472-477.
Materials: **(Slides for this session - Notes 1: Data Presentation.)**
Minitab Project Files: (**Basic Statistics**)

Session 2: Sampling, Descriptive statistics: mean, median, mode, standard deviation, correlation.
Reading: Text, Page 313, Sections 3.4, 4.4, Chapter 6, Pages 144-146.
Materials: **(Slides for this session - Notes 2: Descriptive Statistics.)** (**An article about nonrandom sampling and biased statistical analysis**)
Minitab Project Files: (**US Gasoline Market**) (**Basic Statistics**)

Session 3: Probability, conditional and unconditional probability, independence, joint probability, Bayes Theorem.
Reading: Text, Chapters 7, 8. **Probabilities and the Gulf oil spill.**
Materials **(Slides for this session - Notes 3: Probability.)** (**Sample Probability Exercises**) (**Some famous and fun problems in probability**)

Session 4: Expected value, applications of expected value.
Reading: Text, Section 3.4, 4.3, Chapter 9
Materials: **(Slides for this session - Notes 4: Expected Value.) (Notes about credit default swaps)**

Session 5: Random variables.
Reading: Text, Chapter 9.
Materials: **(Slides for this session - Notes 5: Random Variables.)**

Session 6: Covariance and correlation.
Reading: Text Sections 6.1-6.6, 10.1-10.4
Materials: **(Slides for this session - Notes 6: Bivariate Random Variables.)**

Session 7: Discrete distributions, Bernoulli, binomial.
Reading: Text Sections 11.1 - 11.3.
Materials: **(Slides for this session - Notes 7: Bernoulli and Binomial Distributions)**

Session 8: Discrete distributions, Poisson Model.
Reading: Text. Section 11.4, 18.4.
Materials: **(Slides for this session - Notes 8: Poisson Distributions)** (**A science experiment to produce Poisson outcomes**)

Session 9: The normal distribution.
Readings: Text Chapter 12.
Materials: **(Slides for this session - Notes 9: The Normal Distribution.)** (**Sample Problems**) (**What is the margin of error?**)

Session 10: Samples and sampling distributions, normal distribution, large samples, law of large numbers, central limit theorem.
Reading: Text Pages 155-157, Section 10.5, Chapter 13, Sections 14.1-14.3, (**DataStor case**)
Materials: **(Slides for this session - Notes 10: The Central Limit Theorem and the Law of Large Numbers.)** (**Random Walk Models for Stock Prices**)
Minitab Project Files: (**Cleared Calls**) (**WHO Data**) (**Basic Statistics**)

Session 11: Central Limit Theorem, normal approximations, lognormality, random walk.
Reading: Text Chapter 13
Materials: **(Slides for this session - Notes 11: Normal Approximation and Random Walks.)** (**Seminar: That U.S. 37th Ranking by WHO**) (**Notes for seminar**) (**Lognormal Random Walks for Stock Prices**)
Minitab Project Files: (**WHO Data**)

Session 12: Statistical inference, point estimates and confidence intervals
Reading: Text Chapter 15.

Materials: (Slides for this session - Notes 12: Statistical Inference.)
Minitab Project Files: (Credit Application data)

Session 13: Statistical tests - 1
Reading: Text Chapter 16.
Materials: (Slides for this session - Notes 13: Testing Hypotheses Part 1)
Minitab Project Files: (Credit Application data)

Session 14: Statistical tests - 2
Reading: Text Chapters 17.
Materials: (Slides for this session - Notes 14; Testing Hypotheses Part 2.)
Minitab Project Files: (Credit Application data) (Econometrics Midterm Grades data)

Session 15: Hypothesis testing,
Reading: Text Chapters 5, 18
Materials: (Slides for this session - Notes 15: Hypothesis Testing) (Notes on the Power of a Test)
Minitab Project Files: (Credit Application data) (US Gasoline Market) (Sale Prices for Monet Paintings) (German Health Survey Data) (Sydney/Melbourne Travel Choice)

Session 16: Linear regression.
Reading: Text Chapter 19, 20, 21.
Materials: (Slides for this session - Notes 16: Linear Regression.) (A controversial regression study) (Slides for an application of modeling) (Handout for application) (Regression Analysis by WHO)
Minitab Project Files: (WHO Data) (Drug Couriers) (Trends in Frequent Flyers) (US Gasoline Market)

Session 17: Linear regression model, sample and population.
Reading: Text Chapter 19, 21.
Materials: (Slides for this session - Notes 17: Regression Modeling)
Minitab Project Files: (Movie Success) (US Gasoline Market)

Session 18: Least squares linear regression, residual analysis, analysis of variance.
Reading: Text Chapter 19, 20
Materials: (Slides for this session - Notes 18: Regression Analysis.)
Minitab Project Files: (Movie Success) (WHO Data)

Session 19: Correlation and covariation.
Reading: Text Chapter 20, Section 22.2.
Materials: (Slides for this session - Notes 19: Regression and Correlation.)
Minitab Project Files: (Sale Prices for Monet Paintings) (Trends in Frequent Flyers) (Cost Data for U.S. Utilities) (Salaries) (US Gasoline Market)

Session 20: Aspects of regression to the mean, measurement error, truncation, selection.
Reading: Section 21.4.
Materials: (Slides for this session - Notes 20: Specifying the Regression Model.)
Minita Project Files: (US Gasoline Market) (Sale Prices for Monet Paintings)

Session 21: Multiple regression - 1,
Reading: Text Sections 23.1, 23.2.
Materials: (Slides for this session - Notes 21: Multiple Regression Part 1) (Multiple Regression with Minitab)
Minitab Project Files: (US Gasoline Market) (Sale Prices for Monet Paintings)

Session 22: Multiple regression - 2
Reading: Text Chapter 24.
Materials: (Slides for this session - Notes 22: Multiple Regression Part 2)
Minitab Project Files: (US Gasoline Market) (Sale Prices for Monet Paintings) (WHO Data) (Computer Prices) (UK Electronics Stores) (House Prices)

Session 23: Multiple regression - 3
Reading: Text Section 22.2, 24.2, 24.3.
Materials: (Slides for this session - Notes 23: Multiple Regression Part 3)
Minitab Project Files: (US Gasoline Market) (WHO Data) (Profits and R&D Data) (Movie Madness)

Session 24: Multiple regression - 4
Reading: Text Section 23.4, Chapter 25.
Materials: (Slides for this session - Notes 24 Multiple Regression Part 4)
Minitab Project Files: (WHO Data) (Sale Prices for Monet Paintings) (Longley Data)

Session 25: Modeling Qualitative Data
Reading: None
Materials: (Slides for this session - Notes 25: Analyzing Qualitative Data.) (A Study about Obesity) (The Netflix Prize)
Minitab Project Files: (Credit Application data) (German Health Survey Data) (Sydney/Melbourne Travel Choice)

Individual Problem Sets and Assignments

Students may work with one of their colleagues on these homework assignments and submit your assignment as a group. (Groups may have no more than four students.). All data sets for the assignments are linked below. You can left click to open them in Minitab, or right click to download them to your own computer.

Assignment 1. Data Description and basic probability. Problem set 1 Problem set 1 solutions
Data sets: (HOG-Ex0201.mpj) (HOG-Ex0202.mpj) (HOG-Ex0218.mpj) (HOG-Ex0222.mpj) (97employ.mpj) (WHO-HealthStudy.mpj)

Assignment 2. Probability and Random Variables, Expected Value, Poisson and Hypergeometric Distributons, Normal Distribution. Problem set 2 Problem set 2 solutions
Data sets: (WHO-HealthStudy.mpj) (Easton.mpj) (salary.mpj) (Movies9OCT2003.mpj)

Assignment 3. Statistical Inference. Problem set 3 Problem set 6 solutions
Data sets: (German Health Survey Data) (Sale Prices for Monet Paintings)

Assignment 4. Basic Regression. Problem set 4 Problem set 4 solutions
Data sets: (WHO-HealthStudy.mpj) (EconGrades.mpj) (heating.mpj) (KansasCtyPopn.mpj) (WSJ-Height-Income.mpj)

Assignment 5. Multiple Regression. Problem set 5 Problem set 5 solutions
Data sets: (UKElectronics.mpj) (GermanHealth.mpj) (MoreMoviemadness data.mpj) (Credit Application data)

Model Development Project

Notes on the model development project **Data for Model Development** (Minitab) (Excel)