

B6100 – Managerial Statistics – Fall (A) 2016

Professor Costis Maglaras

409 Uris Hall

c.maglaras@gsb.columbia.edu

Class Time: Monday/Wednesday/Friday (first 4 Fridays of semester)

Office Hours: Mondays 6-7 pm, or by appointment (please email (not through canvas))

Course Description

A career in management is likely to involve decision making based on data. Recent advances in computing and technology in general have made an abundance of information readily available to be exploited as part of a decision making process; the so-called Big Data revolution. The main objective of this course is to provide an understanding of basic statistical concepts and tools, and their use in making business decisions. Examples of questions that can be addressed with the aid of statistical analysis include *estimating* product reliability, *testing* investment strategies, testing for *bias* in analysts' recommendations, and *predicting* a product's sales on the basis of its characteristics.

The course will cover the following key concepts:

* *Descriptive statistics* deals with summarizing data, observing patterns in it, and extracting the vital information contained in it.
* *Probability* concepts, such as expected value, variance and covariance, provide a systematic framework for modeling uncertainty and salient features of the data generating process.
* *Sampling* and *Estimation* theory concerns the use of sample data (e.g. surveys/polls) to make inferences, quantify uncertainty, and test hypotheses we are entertaining about the underlying data generating mechanism.
* *Regression* analysis deals with the construction of predictive models based on data. We will focus on linear regression with emphasis on model fitting, significance testing and prediction assessment.

Books and reference material

There is no required textbook for this course.

Prior to the start of the course, students are asked to review a few short videos (available on the course website) that cover some introductory material on data analysis.

Lecture slides will be provided in class and through the course website and will cover all material discussed in class. The course pack, titled *Readings & Cases*, contains more comprehensive notes to supplement the slides; see part A titled “Course Notes.” Part B of the *Readings & Cases* contains a set of problems along with solutions. Working through these problems is the best way to master the material covered in class and prepare for the exams. These problems are meant for self-study and will not be collected or graded. They will also be discussed in weekly review sessions led by teaching assistants. Part C of the course pack contains a few articles of interest and some of the cases covered in class.

For students seeking an additional resource, the text below will be available on reserve in the library.

Levine, Stephan, Krehbiel, & Berenson *Statistics for Managers*, 6th Ed., Prentice-Hall 2010.

It also contains a host of supplementary problems for self-practice, a select subset along with complete solutions will be posted on the course website. These problems will not be collected or graded.

Homework Assignments and Grading

It is highly recommended to work through the problems in part B of the course pack on a weekly basis; students should consider this as informal self-study homework. In addition to these exercises, there will be four hand-in assignments during the course that will be graded. These assignments are of type ‘A’ (see table below), so group work is permitted (and encouraged). Each group submits a single solution write-up, and all members of the group will receive the same grade. Maximal group size is limited to three students; assignments submitted with four or more names will not receive credit.

In order to avoid ambiguity that may lead to unintentional violations of the Honor Code, the description types for assignments have been standardized and are specified in the table below.

We will have a midterm, and a final exam. Exams will be closed book. However, students may bring three sheets of notes to each exam (where you may write on both sides of each sheet).

Grades will be based on the maximum of the following two weighting schemes:

#1: Class contribution 5%; Hand-in assignments 20%; Midterm 30%; Final 45%

#2: Class contribution 5%; Hand-in assignments 20%; Midterm 0%; Final 75%

More specifically, cumulative scores will be computed based on both schemes and the greater of the two will be used to determine the score of each student. This scheme makes it possible to fully recover from a weak score on the midterm.

NOTICE: All requests for reconsideration of assignment or exam scores and requests to take exams at other than the scheduled times, must be made in writing.

Class contribution points amount to actively working toward a constructive learning environment. This means adhering to the core culture (as specified in detail below), and more broadly being engaged and respectful. Class contribution does not grade you on providing correct (or incorrect) answers to questions raised in class, nor should anyone feel obligated to raise points and discussion items merely for the sake of participation credit.

Please note that laptops, tablets and phones are not to be used in this class.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Type*** | ***Designation*** | ***Discussion of concepts***  | ***Preparation of submission***  | ***Grade*** |
| A | group / group  | Permitted with designated group\*  | By the group | Same grade for each member of the group  |
| B | group / individual  | Permitted | Individually(No sharing of any portion of the submission) | Individual |
| C | individual / individual  | None of any kind  | Individually | Individual |
| D | (An optional category to be defined in detail by the individual faculty member)  |

Columbia Core Culture

The purpose of the Columbia Core Culture is to promote a consistent classroom environment of mutual respect, preparation and engagement. Our expectation of you in class is to be:

Present: Your success depends on being on time and present for the entire class every session. Attendance will be part of your grade for class contribution and students are expected to sit in their assigned seats.

Prepared: Bring your nameplate, clicker, and complete any pre-work needed for class discussion. Expect the professor to cold call in class.

Participating: Active participation calls for no electronic devices such as laptops, tablet computers, or smartphones during class, except when the professor **specifically requests this** as part of in-class work.

Course overview

(1) Descriptive statistics: frequency distribution; mean, median and mode; percentiles; sample variance and standard deviation; normal approximation; a prelude to regression. 2 sessions. [CN: Section 1, LSKB: Chapters 2 and 3.]

(2) Modeling uncertainty: definition and basic properties of random variables; probability distributions; expected value and variance; transformations of random variables; normal distribution. 3 sessions. [CN: Section 3 and 4 (optional reading, Section 2), LSKB: Chapters 4, 5 and 6.]

(3) Sampling and Estimation: random sampling; polls; types of samples; distribution of sample estimates; point estimates; accuracy and precision; bias; constructing confidence intervals; statistical significance and p-values. 4 sessions. (CN: Sections 5 and 6, LSKB: Chapters 7, 8 and 9.)

(4) Regression: simple linear regression; estimation of parameters and significance testing; correlation coefficient; interpreting the regression output; multiple linear regression; dummy variables; model selection; examples of non-linear regression; applications. 5 sessions. (CN; Section 7, LSKB: Chapters 13, 14 and 15.)

LSKB refers to the text by Levine, Stephan, Krehbiel, & Berenson, which, as stated earlier, is an optional reference text, and on reserve in the library.

CN refers to the “Course Notes,” part A of the Cases & Readings course pack.

**Course Overview**

Textbook readings are from Levine, Stephan, Krehbiel, and Berenson (LSKB).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Session** | **Date** | **Topic(s)** | **Required Readings** | **Due that day (group/individ.)** |
| **MODULE 1: DESCRIPTIVE STATISTICS****(Course notes Part 1 on Descriptive Statistics)** |
| 1 | **Wed****9/1/16** | Introduction and course overview; recap of descriptive statistics (mean, median, standard deviation, Normal approximation, Value-at-Risk) |  |  |
| 2 | **Fri 9/9/16** | Normal approximation; financial data; regression line; correlation and causation |  |  |
| **MODULE 2: PROBABILITY****(Course notes Parts 3-4 on Random Variables and their Distributions, and Normal Distribution)** |
| 3 | **Mon 9/12/16** | Random variables; expected value; variance |  |  |
| 4 | **Wed 9/14/16** | Normal distribution and normal table; standardizing | LSKB: 6.2 | **HW #1 - group** |
| 5 | **Fri 9/16/16** | Var(aX+bY); linear transformations of normal; applications to hedging and portfolios |  |  |
| *In session 3, we derived the formula for the expected value of the sum of two random variables, such as for example the sum of the returns of two investments in two stocks. In preparation for session 5, please download the file FB-INTC-2015.xlsx that contains daily info about Facebook (FB) and Intel (INTC) from Jan 2, 2013 to Aug 24, 2015.** *Compute daily returns for each of the stocks (e.g. closing price today relative to closing price yesterday).*
* *Construct the daily returns of a portfolio that is half invested in FB and half invested in INTC. You should assume that the weights stay constant throughout the data period, and you can ignore compounding effects throughout.*
* *Compute the expected return and standard deviation of the daily returns of each of the stocks and those of the 50/50 portfolio using the descriptive statistics formulae we discussed in the beginning of the course.*
* *Think how the stock-specific parameters of the 50/50 portfolio relate to those of the individual stocks.*

*Come to class prepared to discuss your findings. You do not need to turn in anything for that calculation.**The case “Gotham Bank Investment Services” in your casebook is good reading after that session. You do not need to prepare answers to the questions at the end of the case.* |
| **MODULE 3: SAMPLING AND ESTIMATION****(Course Notes Parts 5-6 on Sampling Distributions, Estimation and Confidence Intervals, and Hypothesis Testing pg. 85-91)** |
| 6 | **Mon 9/19/16** | Sampling mean  | LSKB: 7.1-7.4 | **HW #2: Variance Associates - group** |
| 7 | **Wed 9/21/16** | Sampling proportions; interpreting opinion polls | LSKB: 7.5 |  |
| 8 | **Fri 9/23/16** | **MIDTERM EXAM** |  |  |
| 9 | **Mon 9/26/16** | Confidence intervals for mean and proportion; Student t distribution | LSKB: 8.1-8.4.  |  |
| 10 | **Wed 9/28/16** | Hypothesis Testing; introduction and structure of a test for the population mean; type-I error; p-value |  |  |
| **MODULE 4: LINEAR REGRESSION****(Course Notes Part 7 on Linear Regression)** |
| 11 | **Fri 9/30/16** | Introduction and simple linear regression  | LSKB: 13.1-13.4 | **HW #3: Analyzing the Analysts - group** |
| 12 | **Mon 10/3/16** | Multiple linear regression + dummy variables | LSKB: 13.7-13.8, 14.1-14.4, 14.6, 15.3 |  |
| 13 | **Wed 10/5/16** | Examples of regression; coming up with a relevant model; estimation of a demand model based on sales data |  |  |
| *In preparation for session 13, please download the file airline-data.xlsx and run linear regressions to estimate the business and youth demand functions. The data file is related but is not identical to the one you studied in your Managerial Economics class in the context of the Airline Pricing on Shuttle Routes case. Please be prepared to discuss your findings. Having a short printout of the two regression outputs would be helpful.* |
| 14 | **Mon 10/10/16** | Applications of regression & examples of non-linear regression (All-around movers) | LSKB: 15.2-15.3 |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 15 | **Wed 10/12/16** | Applications of regression (CAPM); course recap |  | **HW #4: Pricing Diamonds - group** |
| *In your Corporate Finance course, you have been discussing the CAPM in sessions 11-12, where you covered the underpinnings of the model and its main findings and introduced the concept of beta of a stock.* *As part of an application of linear regression in that context, I want you to download WeeklyData-20112013\_Returns.xlsx and run pairwise regressions between each stock and the market return, and between the equal weight portfolio and the market return. The file contains price data and also weekly returns (in the “weekly returns” tab) from Jan 2011 to Sept 2013 for XOM, IBM, JPM, CSCO and the equal weight portfolio. Please come prepared to briefly discuss your regression results.* |