Course Description

A career in management is likely to involve decision making based on data. In today's revolution of computer and information technology there is an abundance of readily available information to be exploited as part of a decision making process. It is important to be able to organize and use data intelligently and correctly. The main focus of this course is to provide an understanding of basic statistical (inference) tools that are useful—or necessary—in managerial decision making. Statistical inference will allow one to make generalizations based on sample data, and answer questions such as 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 of statistical analysis and inference:

- **Descriptive statistics** deals with summarizing data, observing patterns in it, and extracting the vital information contained in it.

- **Probability** concepts provide a systematic framework for dealing with uncertainty that form the basis for statistical inference.

- **Sampling and Estimation** theory concerns the use of sample data for estimation and as a guide for statistical inference. How do we construct point estimates of unknown parameters and interpret confidence intervals? How reliable are election time exit polls? How large of a sample should we use? How can we use a geographically small and targeted marketing survey in predicting nationwide sales?

- **Hypothesis testing** involves the analysis of statistical evidence in verifying or rejecting a statement about the world (the hypothesis).

- **Regression** analysis deals with the construction of predictive models based on statistical data.
**Books and reference material**

Lecture slides will be provided in class and through the course website. These will cover the material discussed in class. In addition, a set of notes will be handed out as part of the course’s *Casebook*, under the title “Course Notes”. This should serve as the secondary source for the material covered. These two sets of handouts will provide the most important source of information for the course. The *Casebook* also contains a few articles of interest and some of the cases covered in class. Finally, there are a few short videos that cover some introductory material that should be reviewed prior to the start of the course, and some additional supporting videos that would support some of the material covered in class.

There is one supplementary textbook that could serve as an additional source for the material, but it is not required.


The “Course Notes” also contain practice problems organized by topic that should serve as weekly homework assignments. These assignments form an integral part of the course. Students should take these assignments seriously, as is it essential to learning the material. Weekly homework will not be collected or graded; answers are provided in the “Course Notes.”

I have also selected a few sample problems from the LSKB book that I list at the end of this handout. These tend to be simpler, are also optional, and their solutions are provided at the end of the *Casebook*. Homework will be discussed in the review sessions run by the course’s teaching assistants.

**Homework Assignments and Grading**

In addition to the weekly homework, there will be four hand-in assignments during the course; some will be short case write-ups, while some others will be more involved assignments that will primarily serve in helping students digest the material of the course and familiarizing students with the basic capabilities of spreadsheet software. Students can work in groups of up to six on the hand-in assignments. This means that students can work on the hand-in assignments with their study groups (but cannot discuss the answers with anybody else), and submit a single solution write-up. All members of the group will receive the same grade.

In order to avoid ambiguity that may lead to unintentional violations of the Honor Code, the description types for assignments have been standardized. They are specified in the table below. All of our assignments are of type A.
<table>
<thead>
<tr>
<th>Type</th>
<th>Designation</th>
<th>Discussion of concepts</th>
<th>Preparation of submission</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>group / group</td>
<td>Permitted with designated group*</td>
<td>By the group</td>
<td>Same grade for each member of the group</td>
</tr>
<tr>
<td>B</td>
<td>group / individual</td>
<td>Permitted</td>
<td>Individually</td>
<td>Individual</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(No sharing of any portion of the submission)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>individual / individual</td>
<td>None of any kind</td>
<td>Individually</td>
<td>Individual</td>
</tr>
<tr>
<td>D</td>
<td>(An optional category to be defined in detail by the individual faculty member)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

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

- #1: Class participation 5%; Hand-in assignments 20%; Midterm 30%; Final 45%
- #2: Class participation 5%; Hand-in assignments 20%; Midterm 20%; Final 55%

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 makes it possible to recover from a weak score on the midterm.

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

**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 participation 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 tells you as part of in-class work.
Course overview

(1) Descriptive statistics: frequency distribution; mean, median and mode; percentiles; variance and standard deviation; introduction to regression. 2 sessions. (LSKB: Chapters 2 and 3.)

(2) Probability: definition and basic properties of random variables; probability distributions; expected value and variance; transformations of random variables; normal distribution. 3 sessions. (LSKB: Chapters 4, 5 and 6.)

(3) Sampling and Estimation: random sampling; polls; types of samples; distribution of sample estimates; central limit theorem; point estimates; accuracy and precision; bias; constructing confidence intervals; hypothesis tests, p-values. 4 sessions. (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. (LSKB: Chapters 13, 14 and 15.)
Detailed syllabus

(1): Descriptive Statistics
(Course notes Part 1 on Descriptive Statistics)

Session 1: Introduction and course overview; recap of descriptive statistics material (mean, median, standard deviation, Normal approximation, Value-at-Risk)

Session 2: Normal approximation; financial data; regression line; correlation and causation

(2): Probability
(Course notes Parts 3-4 on Random Variables and their Distributions, and Normal Distribution)

Session 3: Random variables; expected value; variance

## HW1 due ##

Session 4: Normal distribution and normal table; standardizing. (LSKB: 6.2)

Session 5: 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 IBM-XOM-2013.xlsx that contains daily info about IBM and ExxonMobil (XOM) from Jan 2, 2013 to Aug 19, 2013.

- 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 XOM and half invested in IBM. You should assume that the weights stay constant throughout the 8 month 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 relate to those of the 50/50 portfolio relate to the 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.

(3): Sampling and Estimation
(Course Notes Parts 5-6 on Sampling Distributions, Estimation and Confidence Intervals, and Hypothesis Testing pg. 85-91)

Session 6: Sampling mean  (LSKB: 7.1-7.4)

## HW2 due: Variance Associates ##
Session 7: Sampling proportions; interpreting opinion polls (LSKB: 7.5)

Session 8: Midterm

Session 9: Confidence intervals for mean and proportion; Student t distribution (LSKB: 8.1-8.4)

   Skim through the note “Dart Investment Funds” in your casebook

Session 10: Hypothesis Testing: introduction and structure of a test for the population mean; type-I error; p-value

(4): Linear Regression
(Course Notes Part 7 on Linear regression)


   ## HW3 due — Analyzing the analysts ##

Session 12: Multiple linear regression + dummy variables (LSKB: 13.7-13.8, 14.1-14.4, 14.6, 15.3)

Session 13: 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.

Session 14: Applications: GM case; Examples of non-linear regression; All-around movers (LSKB: 15.2-15.3)

   Skim through the GM case.

   ## HW4 due — Pricing Diamonds ##

Session 15: Applications of regression (CAPM); course recap

   In your Corporate Finance course you have been discussing the CAPM in sessions 11 and sessions 12, where you covered the underpinnings of the model, 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.