

## STATISTICS and DATA ANALYSIS (B01.1305)

Instructor: Professor Cyrus Mohebbi

Telephone: (914) 225-7081

Email: cyrus.mohebbi@morganstanley.com

cmohebbi@stern.nyu.edu

Textbook (Optional): David K. Hildebrand and R. Lyman Ott, and J. Brian Gary  
Basic Statistical Ideas for Managers (Duxbury, 2<sup>nd</sup> edition)

Software: Minitab

### Course Description:

The course COR1-GB.1305 is designed to achieve an understanding of fundamental notions of data presentation and analysis and to use statistical thinking in the context of business problems. The course deals with modern methods of data exploration (designed to reveal unusual or problematic aspects of databases), the uses and abuses of the basic techniques of inference, and the use of regression as a tool for management and for financial analysis. The potential ethical issues related to each topic will be reviewed. Socially and environmentally relevant data will be utilized throughout the course.

### **I. Data Analysis**

#### **A. Data collection [H&O, various locations]**

1. distinction of population and sample [H&O, Section 1:1 and Section 6:1]
2. types of data: nominal, ordinal, interval, and ratio scaled data  
simplicity sometimes reduced to just qualitative and quantitative. As in [H&O, Section 2:1]
3. importance of data quality and use of appropriate sampling mechanisms [H&O, Section 6:1]

#### **B. Summary statistics and graphical methods [H&O, Chapter 2]**

1. histogram, stem-and-leaf plot, box plot
2. measures of location: mean, median, mode
3. measures of scale: standard deviation, range, interquartile range
4. scatter plots

### **II. Applied probability and inference**

#### **A. Basic concept of probability [H&O, Chapter 3]**

1. algebra of events
2. definition of probability: equally likely frequency theory, personal probability
3. conditional probability and independence

**B. Random variable and their properties [H&O, Chapter 4]**

1. definition of random variable
2. probability distribution function
3. mean, variance, and standard deviation of a random variable
4. covariance and correlation of two random variables

**C. Important specific distributions [H&O, Chapter 5]**

1. discrete uniform distribution [not in H&O]
2. hypergeometric distribution [not in H&O]
3. binomial distribution
4. Poisson distribution
5. continuous uniform distribution [not in H&O]
6. Gaussian (normal) distribution

**D. Point and interval estimation [H&O, Chapter 6-7]**

1. Central Limit theorem
2. Confidence intervals for the mean  $\mu$ .
3. Confidence interval for the binomial proportion  $p$

**E. Hypothesis testing: practical and statistical significance [H&O, Chapter 8]**

1. structure of test: null and alternative hypotheses, Type I and II errors, practical significance of rejection of null hypothesis
2. testing hypotheses about  $\mu$  : z-test, t-test, nonparametric alternatives (sign, sign rank tests) and connection to ordinal data
3. testing hypotheses about  $p$  : z-test
4. graphical testing: control charts [H&O, Section 2.3 to 2.5]
5. comparison of groups [H&O, Section 9.1 to 9.4]
6. two-sample t-test
7. nonparametric alternatives (Mann-Whitney, median tests) and connection to original data

**F. Analysis of discrete data [H&O, Section 9.5-9.6]**

1. nominal and ordinal discrete data
2. contingency tables
3. tests of independence

**III. Regression analysis**

**A. Simple linear regression: assumptions, modeling, inference [H&O, Chapter 11]**

1. the linear model: graphical (scatter plot) and theoretical (model building) analysis
2. the principal of least squares
3. assumption of linear least square regression
4. inference in linear regression: t-tests, F-test,  $R^2$

5. prediction
6. checking assumptions in regression: examination of residuals, diagnostics, and residual plots

**B. Multiple regression [H&O, Chapter 12-13]**

1. the multiple regression model
2. t-tests versus the F test
3. residual analysis
4. interpretation of multiple regression coefficient
5. multicollinearity

**C. Other regression models and problems [H&O, Chapters 12-13]**

1. transformations
2. model selections

You are required to attend all classes. Much of the learning in this course occurs in the classroom. If you feel that you cannot attend just about every class (for example, if you have to attend interviews during class time or you have heavy job commitments), please take another class that is more lecture-oriented. Please note that I have a "no excuses" policy; that is, I only note whether you are in class, not why you are not there.

Therefore, you should be in class BEFORE the start of class (I start on time) and you should stay through the end (I also end on time).

You have one free absence that you should use for EMERGENCY purposes, such as jury duty, surgery, funerals, and other such major events. You can simply send me email saying you are using your free absence.

Please arrive promptly for the start of class. Late entrances are disruptive to the discussion. Also, please remember to turn off cell phones and do not use laptop computers in class. If you have a special need for a laptop, please consult me.

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 Moses Center, you must submit a completed Exam Accommodations Form to them at least one week prior to the scheduled exam time to be guaranteed accommodations. Please be assured that issues related to disabilities will be handled as confidentially as possible.