MSc in Analytics

Course Syllabus

STATISTICAL ANALYSIS
MSCA 3107
Monday 6:00 pm – 9:00 pm.
September 29 – December 1, 2014
Gleacher Center Room x

Yuri Balasanov
ybalasan@uchicago.edu

COURSE DESCRIPTION

This course gives students necessary background in:

- Mathematical foundation of statistical analysis, including introduction to probability theory and probabilistic models underneath statistical experiments;
- Variety of statistical techniques for univariate and multivariate statistical analyses;
- Turning statistical description of the data into analytical tool.

We put special emphasis on covering main steps of building analytics from visualizing data and building intuition about their structure and patterns – to selecting appropriate statistical method – to interpretation of the results and building analytical model. This 10-week course is required for all students of the MSc in Analytics program. Grades will be assigned for this course and the course will appear on your official transcript. All registered students must obtain a passing grade of at least B-.

BOOK

Randall Prium. Foundations and Applications of Statistics. An introduction Using R. 2011. Randall Prium.

SOFTWARE AND HARDWARE

We will be using R and related packages (http://cran.r-project.org)
It is recommended that students have their laptops with R installed during the class. We will use them for data assignments in class.

LEARNING OBJECTIVES

After completing this course, students should be able to:

- Design, conduct and interpret statistical experiments.
- Read and implement results of scientific research requiring background in probability and statistics.

- Learn how to avoid both being fooled by randomness as a result of overfitting and being trapped in black swan events as a result of underestimating the chances.
- Understand and use theoretical distributions to assign probabilities to events.
- Discover patterns hidden in data.
- Understand assumptions behind the common methods of statistical inference.

EVALUATION:

Your course grade will be calculated as follows:

- 40% Weekly Assignments
- 20% Class Participation
- 40% Final Project

GRADING SCALE

A = 93%–100%
A- = 90%–92%
B+ = 87%–89%
B = 83%–86%
B- = 80%–82%
C+ = 77%–79%
C = 73%–76%
C- = 70%–72%
F = 0%–69%

ATTENDANCE

This course will meet once a week on Monday evenings between 6:00 to 9:00 over ten weeks. Your attendance is required and very important not only for your own results in this class, but also for the development of this new course. You are allowed to miss at most two sessions, provided that you make arrangements with the instructor in advance.

FINAL PROJECT

Final course project is submitted in the form of written report summarizing the methods used for the inference, justification of their assumptions and interpretation of the results.

Additional information and details of the assignment will be given by the instructor at least 5 weeks before the due date.

The final project will be graded out of 100 points. You may use R or any other statistical package you feel comfortable with to analyze the data. Yet, it is recommended that you use R since relevant syntax for the analysis required in each session will be provided.

Contact Us 1427 E. 60th Street Chicago, Illinois 60637 773.702.1722 grahamschool.uchicago.edu

WEEKLY ASSIGNMENTS

We plan to end each session with a brief assignment emphasizing the material of the session. We expect that most of the assignments will be completed during the class and finished during the week following the class

LATE WORK

All assignments must be submitted to the Chalk site for the course on the due date before 11:59 pm. If you turn in an assignment late, 10% will be deducted from the total score for each day after the deadline. Assignments turned in more than one week late will not receive credit. In the case of unexpected events, you must contact the instructor before the assignment due date in order to receive a grace period. Students can only receive up to two grace periods in the course.

REQUESTING REASONABLE ACCOMODATIONS

If you are interested in requesting disability accommodations, you may want to begin by reading through the information published on this website https://disabilities.uchicago.edu/. Also, please do communicate your requests as soon as possible to Gregory Moorehead, director of disability services, at 773.702.7776 or gmoorehead@uchicago.edu.

ACADEMIC HONESTY & PLAGIARISM

It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another's statements or ideas of work as one's own. To do so is plagiarism or cheating, offenses punishable under the University's disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously.

Proper acknowledgment of another's ideas, whether by direct quotation or paraphrase, is expected. In particular, if any written or electronic source is consulted and material is used from that source, directly or indirectly, the source should be identified by author, title, and page number, or by website and date accessed. Any doubts about what constitutes "use" should be addressed to the instructor.

At any time during or after the course students are encouraged to help developing this course by providing their feedback to the instructor in any form, as long as it is constructive, respectful and in compliance with the ethical norms of The University of Chicago.

COURSE SCHEDULE

Important Note: Changes may occur to the syllabus at the instructor's discretion. When changes are made, students will be notified via email and in-class announcement.

Contact Us 1427 E. 60th Street Chicago, Illinois 60637 773.702.1722 grahamschool.uchicago.edu

SESSION 1
Problem of statistical analysis; randomness; statistical experiment as check against randomness. Foundations of Probability Theory: random experiment; probability space; random variables; probability distributions, moments; conditional probability, independence and correlation as a measure of linear dependence; perfect correlation; how much of correlation is a lot in the context of linear model?

SESSION 2
Simulation of random variables; uniform distribution; pseudo-random variables: generating and testing; discrete and continuous random variables and their distributions; simulation of linear model

SESSION 3
Method of moments; estimands, estimators, estimates; limit theorems of probability theory; testing statistical hypotheses and defining confidence intervals; estimating mean and variance; method of moments for linear model

SESSION 4
Modeling relationships between variables: linear model framework; learning to think in terms of vectors and matrices: vectors, spans, bases, dot products and projections, orthogonal and orthonormal bases, matrices; linear model in matrix notations

SESSION 5
Least squares method for linear regression model; estimation of simple linear model; inference for simple regression model: distribution of the slope estimator, distribution of the intercept estimator, estimator of variance/standard deviation of residuals and its distribution, confidence intervals and hypotheses testing

SESSION 6
Regression ANOVA; categorical predictors; categorical response; logistic regression model: definition and interpretation

SESSION 7
Robustness of linear model; checking assumption about distribution of residuals; additive linear model with multiple predictors: geometric interpretation, orthogonal vs. non-orthogonal design, comparison of sub-models; more on logistic regression

SESSION 8
One-way ANOVA: ANOVA as linear model, utility test, geometric interpretation, orthogonal vs. non-orthogonal design; contrasts; pairwise comparisons; multiple comparisons

SESSION 9
Linear model with multiple responses and no predictors: reduction of dimensionality; eigenvalues and eigenvectors of a matrix; PCA algorithm and interpretation of the results

SESSION 10
Two-Way ANOVA; more on contrasts: creating a contrast, orthogonality of contrasts, contrast direction vector and statistic for F test, contrast for interaction; examples; ANOVA with factors with more than 2 levels; checking ANOVA model assumptions; stepwise regression

Contact Us 1427 E. 60th Street Chicago, Illinois 60637 773.702.1722 grahamschool.uchicago.edu