

Master of Science in Analytics

Course Syllabus

Introduction to Statistical concepts

MSCA 31000

Autumn 2020

Week of August 24 – Week of September 21

Location: Remote

Dr. Anil Chaturvedi
anilchaturvedi@uchicago.edu

COURSE DESCRIPTION

This course levels the playing field by providing statistical knowledge that all Data Scientists must possess as a pre-requisite. It provides general exposure to basic statistical concepts that are necessary for students to understand the content presented in more advanced courses in the program. The course covers theoretical distributions and the way these distributions are used to assign probabilities to events in some depth. The course also introduces students to descriptive statistical methods to explore and summarize data, methodologies for sampling units for measurement or analysis, drawing inferences on the basis of knowledge gained from samples to populations, assessing relationships between variables, and making predictions based upon relationships between variables. This 50-credit course is either required or strongly recommended for all students entering the MSc in Analytics program. Letter grades will be assigned for this course and the course will appear on your official transcript. Students must obtain a passing grade of at least B-.

REQUIRED TEXTBOOK

David Lane (Editor). *Introduction to Statistics*. (Version 2.0). Free online book at <http://onlinestatbook.com>, developed by Rice University, University of Houston Clear Lake, and Tufts University. You must use the web version, do not use the PDF version for the homework or readings. This book's simulations work best on Apple devices. If you are using PC, Internet Explorer might work the best. Simulations do not work with Google Chrome and Microsoft Edge. Use the mobile version of the book if you are using an iPad.

OPTIONAL TEXTBOOK:

Students can download the free textbook “Introduction to Statistical Learning using R” from <https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf>. It comes with a lot of freely available resources – Videos and other resources.

SOFTWARE

R and related packages (<https://cran.r-project.org/>)

LEARNING OBJECTIVES

After completing this course, students should be able to:

- Produce and interpret graphic and numerical summaries of data.
- Understand and use theoretical distributions to assign and compute probabilities to random events.
- Graphically and numerically describe the association between different variables.
- Perform Hypothesis tests for statistical inferences of populations means and proportions
- Perform hypothesis test for comparing means and proportions of two populations
- Understand and apply the results of the Central Limit theorem to construct Confidence Intervals
- Understand and apply simple linear regression to build a linear relationship between a numeric outcome variable and a predictor variable
- Perform analysis of variance

EVALUATION:

Your course grade will be calculated as follows:

- 80% Weekly Assignments
- 20% Quizzes.

GRADING SCALE

A = 94%–100%
A- = 90%–93%
B+ = 87%–89%
B = 84%–86%
B- = 80%–83%
C+ = 77%–79%
C = 74%–76%
C- = 70%–73%
F = 0%–69%

ATTENDANCE

This course will meet five times. All course goals, session learning objectives, and assessments are supported through in class lecture, activities, and discussions. Your attendance is thus required and paramount to your success in this course. You are allowed to miss at most one session, provided that you make arrangements with the instructor in advance.

FINAL PROJECT (Due last class)

Final project is a three to five pages (excluding appendices) written report summarizing the results of an analysis of a data set presented in the case study “Angry Moods.” The case study is about controlling anger and can be found at the end of the book.

You will analyze the Angry Moods data set throughout the course in several steps. The final project will put together all of your findings from earlier sessions with the findings from additional analyses you will complete at the end of the course. Therefore, it is important that you start writing the final project as you obtain results in earlier sessions. The goal of the project is to develop a coherent and concise story or knowledge on how different people deal with anger. It is important that your story is substantiated on the basis of results of your data analyses. Put your detailed data analyses results in an appendix and submit together with the written report.

In addition to a brief executive summary of the entire project at the beginning, the written report must have the following sections.

- Background: Brief description of the study, the data set, and the general research question (10 points).
- Hypotheses: Specific questions that you would like to answer by analyzing the data set (20 points).
- Findings: Summary of your results (30 points).
- Discussion: Your interpretation of the findings, e.g., overall story of how different people deal with anger (30 points).
- Conclusion: The (possibly actionable) insight(s) you would like your readers to take away (5 points).
- References: List of sources you have consulted.

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. Make sure that your writing is clear and free of grammatical and spelling mistakes and that you use **APA format for reporting statistical results and citations (5 points)**. An example project on a different topic is posted on the Canvas site for the course.

WEEKLY ASSIGNMENTS

There are five set of weekly assignments, due the Sunday of the week following the class. Please name your assignment as name of the assignment+ your name. For instance, the name for Assignment 1 might be: **Assignment1_yourname**. Each assignment is about answering the questions listed in the course schedule for the respective session. Whenever necessary, R syntax will be provided. Most assignments have questions relevant for the final project. Your answers to these questions will be graded as part of the assignment, not the final project.

Note that you do not have to type your answers; yet, you are responsible for making yourself understood.

LATE WORK

All assignments must be submitted to the Canvas site for the course on the due date before 11:59 pm. If you turn in an assignment late, 10% credit 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.