



**Professor: Yong Zheng**

*Address:* Perlstein Hall, Room 221; 10 W 33rd St, Chicago IL 60616

*Telephone:* 312 567 3575

*Email:* [yzheng66@iit.edu](mailto:yzheng66@iit.edu)

*Office(s):* Main Campus - Perlstein Hall, Room 221D

*Office Hours:* Tuesdays 1:00 PM – 4:00 PM

**Note:** This syllabus is applicable to the following sections

- IT-D 827-01 and ITMD 527-01, Live Section
- IT-D 827-02, IT-D 827-04 and ITMD 527-02, Online Section
- ITMD 527-03, Remote Students from India

**Course Catalog Description:** Data analytics is involved with multiple useful techniques which can be adopted to visualize and analyze data, make predictions, or extract knowledge/patterns to build intelligent applications or serve for business intelligence. This course focuses more on statistical data analytics. The course will introduce the basic concepts in data analytics, cover the topics in linear regressions and classification. The course will introduce different statistical and analytical software, particularly R and SAS, in order to perform statistical data analytics. The course will provide rich practice experience based on real-world data, such as financial and social media data sets.

**Prerequisites:** There are no prerequisites. It is better for students to have background in database and programming languages. **Credit:** 3-0-3

**Course Outcome:** At the completion of the course, each student will have the capability to perform statistical data analytics by using popular tools, such as R or SAS. Particularly, students should be able to understand the statistic basics, including descriptive statistics, data variability, statistical distributions, hypothesis test. In addition, students should be able to build predictive models based on multiple linear regressions, logistic regressions and other classification models.

**Lecture Days, Time & Place:**

Time: Tuesdays & Thursdays, 08:35 AM to 09:50 AM

Place: Stuart Building (SB) 212, IIT Main Campus

**Important Dates:**

The following information are collected from official IIT Academic Calendar

Jan 14 Spring Classes Begin

Jan 21 Martin Luther King, Jr. Day - No Classes

Feb 6 Spring Final Exam Schedule Published Online

Mar 18-23 Spring Break Week, No Classes

Mar 25 Fall Class Schedule Published Online

Apr 1 Last Day to Withdraw from Full Semester Classes

Apr 8 Fall Registration Begins

May 6-11 Final Exam Week

May 15 Final Grades Due



**Schedule of Topics:** Updates to the schedule, readings and assignments will be posted to Blackboard on a weekly basis and announced in class. The schedule below is tentative and subject to change.

Session	Date	Topic
1	Jan 15	Intro: Data Analytics, Data and Data Types
	Jan 17	Statistical Basis: Descriptive Statistics
2	Jan 22, 24	Statistical Basis: Estimating Population from Samples
3	Jan 29	Statistical Basis: One-Sample Hypothesis Testing
	Jan 31	Statistical Basis: Two-Sample Hypothesis Testing and Using R
4	Feb 5, 7	Linear Regression: Supervised Learning and Multiple Linear Regression
5	Feb 12, 14	Linear Regression: Feature Selections and Improve Your Models
6	Feb 19, 21	Linear Regression: Improve Linear Regression Models
7	<b>Feb 26</b>	<b>In-Class Practice 1: Linear Regression Models</b>
	Feb 28	Analysis of variance (ANOVA)
8	Mar 5	Exam1 Reviews
	Mar 7	Data Preprocessing Techniques
9	Mar 12	Classification: Logistic Regression Models
	Mar 14	<b>Exam 1: Exam of Linear Regression Models</b>
10	<b>Feb 19, 21</b>	<b>Spring Break Week, No Classes</b>
11	Mar 26	Classification: KNN and Naïve Bayes
	<b>Mar 28</b>	<b>In-Class Practice 2: Basic Classification Models</b>
12	Apr 2, 4	Classification: Decision Trees and Ensemble Classifications
13	Apr 9	Advanced Topics in Classification: Multi-Label Classifications
	Apr 11	Exam2 Reviews
14	Apr 16	<b>In-Class Practice 3: Advanced Classification Models</b>
	Apr 18	<b>Exam 2: Exam of Classification Models</b>
15	Apr 23, 25	Advanced Topics: Other Regression Models
16	Apr 30, May 2	Advanced Topics: Data Mining and Machine Learning
17	<b>May 6-11</b>	<b>Final Project Presentations (You should present in the presentation day and cannot book a flight which departures in this period)</b>

**Textbook:** The textbook for this course is NOT mandatory, but highly suggested and recommended.

- William Mendenhall, Terry Sincich. "A Second Course In Statistics: Regression Analysis (7<sup>th</sup> Edition)". Pearson Education, ISBN: 9780321691699; Note: You can also use the 6<sup>th</sup> Edition
- Peter Flach. "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press; 1 edition; ISBN-10: 1107422221, ISBN-13: 978-1107422223

**Other Books:** other learning materials which may be helpful

- Alvin C. Rencher and G. Bruce Schaalje. "Linear Models in Statistics (2nd Edition)". John Wiley & Sons, Inc., 2008; ISBN 978-0-471-75498-5; You may find an eBook online.



- Geoff Der and Brian S. Everitt. "A Handbook of Statistical Analyses using SAS (2nd Edition)". Chapman & Hall/CRC, 2002; You may find an eBook online.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. "An Introduction to Statistical Learning with Applications in R"; Springer; You may find an eBook online.

**Manuals:** These are eBooks related to statistical programming in this course

- W. N. Venables, D. M. Smith and the R Core Team. "An Introduction to R";
- SAS Team. "Step-by-Step Programming with Base SAS 9.4";

**Online Resources:** A list of helpful online resources can be found below.

- KDnuggets, Data Mining Community's Top Resource for Data Mining and Analytics info. <http://www.kdnuggets.com/>
- Statistics Concepts, <http://www.statsoft.com/Textbook/Elementary-Statistics-Concepts>
- R tutorial, <https://www.tutorialspoint.com/r/>
- SAS tutorial, <http://support.sas.com/training/tutorial/>

**Readings/Videos:** Readings for the class will be assigned from the textbook as well as in the form of online reading. Online resources and videos will be linked from or embedded in a Blackboard page. It is essential that you do all readings and/or view the videos before coming to class on the assigned date. These materials are a necessary and integral part of the class and will form the basis for any class discussions on the topic. Specific readings are assigned by topic above.

**Course Objectives:** At the conclusion of this course, each successful student will able to:

- Understand and be clear about statistic basics, including descriptive statistics, inference, variance and deviation, data variability, statistical distribution, confidence interval, statistical significance test, hypothesis test, significance levels, etc
- Be able to draw different plots in order to visualize data
- Be able to build different predictive models, including linear regression & classification model
- Be able to perform statistical analytics and interpret the data and analytical results
- Be familiar with SAS and R for statistical data analytics

**Course Notes:** Copies of the course lecture notes in the form of a PDF of the PowerPoint presentation accompanying each lecture will be provided for each student on Blackboard. This should be useful if you must miss a class. You should be aware that note taking is encouraged and should help your understanding of the material.

**Course Web Site:** <http://blackboard.iit.edu/>

**Blackboard:** The course will make intensive use of Blackboard (<http://blackboard.iit.edu/>) for communications, assignment submissions, group project coordination, providing online resources and administering examinations. All remote students will view the course lectures online via Blackboard, and online readings will be found on Blackboard.

**Guest Lectures:** Guest lecturers may be featured as part of course topics. When a guest speaker is expected you should make an extra effort to be seated and ready prior to class time. Guest lectures may be in the evening in which case class will not be held during a scheduled morning period. A question & answer/ discussion period will be held at the end of each lecturer's presentation.

**Attendance:** If you are in a live section of the class and will not be able to attend class, please notify me via email prior to class time. Post-notifications will not be considered. Online students are welcome to attend the live sections if there are available seats. Roster sheets will be randomly distributed to the class, and students in the live sections are responsible to sign. For online sections, the online roster sheets will be randomly distributed too, and online students should sign by Friday, 11:59 PM on that week.

**Assignments:** There will be three types of assignments: 1). *Written assignments* which examine students' knowledge and skills in related topics; 2). *Final individual or team projects*. **No late submissions are allowed for final projects. Note that, all assignments except the final project are individual assignments, you cannot work together and submit the same or similar answers.**

**Late submissions:** **1).** For late submissions, a 15% penalty based on the grades they earned will be applied. For example, the total score is 100 for an assignment, and one student finally got a score of 90. But his submission was late and passed the due date. Just apply a 15% penalty. His or her final score will be  $90 * 0.85 = 76.5$ . **2).** The submission will be discarded if it was submitted later than a week. For example, due date is Sep 1st, 11:59 PM. Students can submit late work, but their submission must be prior to Sep 8th, 11:59 PM. Otherwise, he or she will get a ZERO score.

**Blog and Quizzes:** Each student will have access to the discussion board on the Blackboard system. Students are encouraged to post questions and discussions in the forums. Students can help each other in the following topics: understanding assignments, discussing challenges in assignments, questions about lectures or learnings, sharing learning materials and job positions, helping debugging, and so forth. Students are NOT allowed to share answers in assignments or exams.

I may give quizzes at my discretion and may use them for verification that you have completed assigned course readings or have read the blog entries. Quizzes may be online via Blackboard. As they are discretionary, the weight of quizzes in grading is also left to my discretion and will be included in your class participation grade. If I see a regular pattern of comments on other student's blog entries, I will not need to give any quizzes in this area.

**Examinations:** The exams will be an in class, closed books, closed notes, closed laptop and written test. The Final will be a team project.

#### Academic Honesty:

**Plagiarism:** All work you submit in this course **must be your own**. You must fully attribute **all** material directly quoted in papers and you must document all sources used in the preparation of the paper using complete, APA-style bibliographic entries. Including directly quoted material in an assignment without attribution is always plagiarism and will always be treated as such by me. No more than thirty-three percent of material included in any paper may be direct quotes.

For any concept questions in the assignments, such as "what is classification?". You can learn by searching answers from Internet, but you cannot simply copy the original texts online in your assignments. You should use your own language/texts as answers based on your understandings.

I will process all written, coding and project submissions using automated utilities to check for potential violations in assignments, including programming projects. Students who have copied codes, texts, or answers from other students or the internet without attribution in their assignments, exams or final projects are considered as plagiarism.

For the first time of plagiarism, the assignment or exam will be graded as a zero score. The student will receive a warning. If the plagiarism happens more than one time by a same student, an Academic Honesty Violation Report (AHVR) will be filed, and the student will automatically receive a final grade of E for this course as per the IIT and ITM academic honesty policies. **There is no excuse for not understanding this policy** and if you do not understand it please let me know and I will be happy to discuss it with you until you do.

**Grading:** Grading criteria for this course will be as follows:

<b>A</b>	<i>Outstanding work reflecting substantial effort</i> .....	90-100%
<b>B</b>	<i>Adequate work fully meeting that expected of a graduate student</i> .....	80-89.99%
<b>C</b>	<i>Satisfactory work meeting minimum expectations</i> .....	60-79.99%
<b>F</b>	<i>Unsatisfactory work</i> .....	0-59.99%

The final grade for the class will be calculated as follows:

Regular Assignments .....	<b>27%</b>
Exam 1 .....	<b>25%</b>
Exam 2 .....	<b>20%</b>
Final Project and Presentations .....	<b>25%</b>
Class Participation .....	<b>3%</b>

**Other Class Resources:** Online readings and other class resources may be found at on Blackboard.

**Our Contract:** This syllabus is my contract with you as to what I will deliver and what I expect from you. If I change the syllabus, I will issue a revised version of the syllabus; the latest version will always be available on Blackboard. Revisions to readings and assignments will be communicated via Blackboard.

**Disabilities:** Reasonable accommodations will be made for students with documented disabilities. In order to receive accommodations, students must obtain a letter of accommodation from the Center for Disability Resources and make an appointment to speak with me as soon as possible. My office hours are listed on the first page of the syllabus. The Center for Disability Resources (CDR) is located in 3424 S. State St., room 1C3-2 (on the first floor), telephone 312.567.5744 or [disabilities@iit.edu](mailto:disabilities@iit.edu).