ISQS 6350: Multivariate Analysis

Fall 2020

# Instructor

Alireza Sheikh-Zadeh, Ph.D.

# Course Description

The majority of data sets in all disciplines are multivariate, meaning that several measurements, observations, or recordings are taken on each row of the data set, in which there might be some correlation between variables. In a few cases, it may be sensible to isolate each variable and study it separately, but in most instances, all the variables need to be examined simultaneously to grasp the structure and key features of the data fully. For this purpose, the multivariate analysis (MVA) includes methods both for describing and exploring such data and for making formal inferences about them.

# Learning Objectives

By the end of this course, students will be able to:

1. use R to conduct analysis associated with multivariate analysis.
2. generate visualizations for multivariate data that are helpful for exploratory analysis.
3. apply and interpret principal components analysis (PCA), multidimensional scaling, and exploratory factor analysis for dimension reduction.
4. apply and interpret cluster analysis techniques to discover structures and patterns in high-dimensional data.
5. develop confirmatory factor analysis models to discover the latent variables behind a set of variables.

# Teaching Philosophy

My teaching philosophy revolves around promoting critical thinking of my students, as I expect them to be able to explain why they choose a specific method and how that method works for different data sets and assumptions. I usually devote high efforts to design my courses by going beyond traditional textbooks and trying to adapt to industry advancements. I try to avoid providing a "Cookbook." Cookbooks tell you all about the what but nothing about the why. With computers, software, and the Internet readily available, it is easier than ever for students to lose track of the why and focus on the what instead. This class takes exactly the opposite approach.

# How to Succeed in this Course

If you want to be a successful student:

* Be self-motivated and self-disciplined.
* Be willing to "ask questions" if problems arise. The discussion board will be a great platform to ask your questions like you are sitting in the class. I will actively check the discussion board and respond to your concerns.
* Be willing and able to commit to 8 to 24 hours per week for this course.
* Take notes while watching videos.
* Be curious to understand all R coding syntaxes. The best place to ask questions is the discussion board.
* Be able to write codes and apply analysis after watching demonstration videos or reading independently.
* Be able to communicate through writing.
* Accept critical thinking and decision making as part of the learning process.

**In contrast, here are some common behaviors that lead to failing the course.**

* Don't watch videos before starting to work on assignments.
* Wait until the last day to begin assignments.
* Forget about deadlines.
* Ignore blackboard announcements, emails from the instructor and/or your peers regarding course activities.
* Ignore discussion board posts.
* Don't get familiar with the grade book and syllabus.

# Netiquette

Your instructor and fellow students wish to foster a safe online learning environment. All opinions and experiences, no matter how different or controversial they may be perceived, must be respected in the tolerant spirit of academic discourse. You are encouraged to comment, question, or critique an idea, but you are not to attack an individual. Our differences, some of which are outlined in the university's nondiscrimination statement below, will add richness to this learning experience. Please consider that sarcasm and humor can be misconstrued in online interactions and generate unintended disruptions. Working as a community of learners, we can build a pleasant and respectful course ambiance. Please read the Netiquette rules for this course:

* Do not dominate any discussion. Allow other students to join in the discussion.
* Do not use offensive language.
* Present ideas appropriately.
* Be cautious in using the Internet language. For example, do not capitalize all letters since this suggests shouting.
* Avoid using vernacular and/or slang language. This could lead to misinterpretation.
* Keep an "open-mind" and be willing to express even your minority opinion.
* Think and edit before you push the "Send" button.
* Do not hesitate to ask for feedback.

# Textbook and Supplementary Material

## Required*:*

1. Everitt B. & Hothorn, T. (EH) An Introduction to Applied Multivariate Analysis with R, ISBN 978-1-4419-9649-7. errata at <https://cran.r-project.org/web/packages/MVA/vignettes/Ch-Errata.pdf>
2. Manly B.F.J. & Navarro Alberto J.A. (MN) Multivariate Statistical Methods, ISBN 978-1-4987-2896-6
3. R Studio (free) – available from <https://www.rstudio.com/>

## Recommended:

1. Chiu, Yu-Wei. Machine Learning with R Cookbook.
2. Johnson, R. & Wichern, D. Applied Multivariate Statistical Analysis, 6th Edition. Pearson, 2007. ISBN-13: 978-0-13-187715-3.
3. Smola, A. & Vishwanthanan, S.V.N. Introduction to Machine Learning.
4. Marden, J. Multivariate Statistics.
5. Joliffe, I.T. Principal Component Analysis, 2d Edition.

# Course Structure

Module 1- Introduction to Multivariate Data Analysis (MVA)

This module introduces some examples of multivariate data. It also covers the basics of matrix algebra, which helps understand the notation associated with multivariate statistical methods.

Module 2- MVA Visualization

This module proposes various solutions for visualizing multivariate data in more than two dimensions.

Module 3- Principle Component Analysis (PCA)

This module explains the method of PCA, which is a beneficial method for reducing a large number of original variables to a small number of transformed variables.

Module 4- Multidimensional Scaling (MDS)

This module describes the method of MDS, which is a dimension reduction technique, designed to construct a 2-D diagram showing the distances between data units.

Module 5- Exploratory Factor Analysis (EFA)

This module defines the method of EFA as another dimension reduction technique, which is designed to describe a broad set of variables in terms of a smaller number of variables or factors.

Module 6- Clustering

This module describes the idea of cluster analysis to use the value of the variables to devise a scheme for grouping the data units into clusters so that similar objects are in the same group.

Module 7- Confirmatory Factor Analysis (CFA)

This module describes CFA models in which particular manifest variables (measurable variables) are allowed to relate to particular factors. In contrast, other manifest variables are constrained to have zero correlation to some of the factors.

# Assignments and Grading

Your learning will be assessed through a weighted combination of the listed assignments spread throughout the semester. Grades are calculated based on the following. [97%, 100%] A+, [93%, 97%) A, [90%, 93%) A-, [87%, 90%) B+, [83%, 87%) B, [80%, 83%) B-, [70%, 80%) C, [60%, 70%) D, and [00%, 60%) F.

| Assignments | Grading Weight |
| --- | --- |
| Homework | 35% |
| Quiz | 35% |
| Project | 25% |
| Project Peer Review | 5% |
| Total | 100% |

# Homework

Homework exercises are assigned throughout the semester to provide one an opportunity to practice the skills and techniques covered in class and to demonstrate the grasp of related concepts and terminology. Late homework can receive only partial credit, and never a higher score than anyone's on-time submission. Students are always welcome to ask questions from their professor via course message tool or post public questions via discussion boards. *However,* *students are not allowed to share their code or write-up with peers. Any evidence of plagiarism will be reported to the Office of Student Conduct.*

# Quizzes

There will be four quizzes throughout the semester. These quizzes contain multiple-choice or short answer questions and focus mainly on the concepts, definitions, and implications of the most used codes.

# Project

Students will pick the dataset of their interests. Each team will deliver a final project report. The project instruction and associated rubric are posted on Blackboard.

# Project Peer Review

Students review at least two of their peers' drafts (based on the project rubric) and provide brief feedback to them in the blog comments.

# Missed or Late Work

You must communicate your highly legitimate reason for missing an assignment at least 24hrs before the deadline.

Otherwise, any late submissions can receive only partial credit, and never a higher score than anyone's on-time submission.

# Academic Integrity

Operating Procedure 34.12 states: Academic integrity is taking responsibility for one's own class and/or course work, being individually accountable, and demonstrating intellectual honesty and ethical behavior. Academic integrity is a personal choice to abide by the standards of intellectual honesty and responsibility. Because education is a shared effort to achieve learning through the exchange of ideas, students, faculty, and staff have the collective responsibility to build mutual trust and respect. Ethical behavior and independent thought are essential for the highest level of academic achievement, which then must be measured. Academic achievement includes scholarship, teaching, and learning, all of which are shared endeavors. Grades are a device used to quantify the successful accumulation of knowledge through learning. Adhering to the standards of academic integrity ensures grades are earned honestly. Academic integrity is the foundation upon which students, faculty, and staff build their educational and professional careers. Please also see [*Student Handbook and the Code of Student Conduct*](http://www.depts.ttu.edu/dos/handbook/). [Texas Tech University ("University") Quality Enhancement Plan, Academic Integrity Task Force, 2010].

The Master of Science in Data Science Program (MS-DS) at Texas Tech is committed to educating critical thinkers that are free of academic or professional dishonesty. All MS-DS students are upheld to the standard of having integrity in the work they produce. The standard rule is for all MSDS students to be responsible in an ethical and honest manner. Any instance of academic integrity violations as listed in the *MS-DS Program Academic Integrity Policy,* including, but not limited to, cheating on examinations or assignments, plagiarism, collusion, and misrepresenting facts will be taken seriously. All acts of academic misconduct will be reported and adjudicated as prescribed by the *MS-DS Program Academic Integrity Policy*.

# Missed or Late Work

You must communicate your highly legitimate reason for missing an assignment at least 24hrs before the deadline. Otherwise, any late submissions can receive only partial credit, and never a higher score than anyone's on-time submission.

# Students with Disabilities Policy (Operating Procedure 34.22)

Any student who, because of a disability, may require special arrangements in order to meet the course requirements should contact the instructor as soon as possible to make any necessary arrangements. Students should present appropriate verification from Student Disability Services during the instructor's office hours. Please note: instructors are not allowed to provide classroom accommodations to a student until appropriate verification from Student Disability Services has been provided. For additional information, please contact Student Disability Services in West Hall or call 806-742-2405.

# Student Absence for Observance of Religious Holy Day Policy (Operating Procedure 34.19)

"Religious holy day" means a holy day observed by a religion whose places of worship are exempt from property taxation under Texas Tax Code §11.20.2. A student who intends to observe a religious holy day 4 should make that intention known in writing to the instructor prior to the absence. A student who is unable to submit assignments or complete scheduled quizzes due to the observance of a religious holy day shall be allowed to complete those within a reasonable time after the absence.

# Discrimination, Harassment, and Sexual Violence Statement

Texas Tech University is committed to providing and strengthening an educational, working, and living

environment where students, faculty, staff, and visitors are free from gender and/or sex discrimination of

any kind. Sexual assault, discrimination, harassment, and other Title IX violations are not tolerated by the

University. Report any incidents to the Office for Student Rights & Resolution, (806)-742-SAFE (7233) or file

a report online at titleix.ttu.edu/students. Faculty and staff members at TTU are committed to connecting

you to resources on campus. Some of these available resources are: TTU Student Counseling Center, 806-

742-3674, https://www.depts.ttu.edu/scc/(Provides confidential support on campus.) TTU 24-hour Crisis

Helpline, 806-742-5555, (Assists students who are experiencing a mental health or interpersonal violence

crisis. If you call the helpline, you will speak with a mental health counselor.) Voice of Hope Lubbock Rape

Crisis Center, 806-763-7273, voiceofhopelubbock.org (24-hour hotline that provides support for survivors

of sexual violence.) The Risk, Intervention, Safety and Education (RISE) Office, 806-742-2110, https://www.depts.ttu.edu/rise/ (Provides a range of resources and support options focused on prevention education and student wellness.) Texas Tech Police Department, 806-742 3931, http://www.depts.ttu.edu/ttpd/ (To report criminal activity that occurs on or near Texas Tech

campus.)

# Disclaimer

*This syllabus is subject to change.*