11.29.2022

Data Analytics for Information Professionals
INST 627
Fall 2022

Tuesday 6:00 – 8:45 PM
Hornbake 2119

Instructor: Christopher Antoun (antoun@umd.edu)
Grader: Maitri Patel (mdp29@umd.edu)

Data Analytics for Information Professionals provides an overview of the basic statistical methods that transform raw data into usable information. The course exposes students to key statistical concepts through lectures and readings. The course provides an introduction to the applied practice of data analysis through exercises conducted outside of class and a final project, structured around R as the main software package. Students will gain an understanding of a wide range of analysis methods as well as the process of selecting an appropriate method for a given research question and type of data.

Learning Objectives: After completing this course you will be able to:

-Select and apply appropriate statistical methods;
-Use R for basic data manipulation and analysis; and
-Communicate the results of an analysis effectively.

Course Requirements:

Reading assignments are mostly from the *Online Stat Book* (hereafter OSB). It is free and available online: http://onlinestatbook.com/2/index.html.

R tutorials are from Mike Marin and hosted on YouTube: https://www.youtube.com/playlist?list=PLqzoL9-eJTNDw71zWePXyHx3_cm_fMP8S

Please read and watch the assigned material for each class session beforehand.

R software is free and available online: https://www.r-project.org/. You are encouraged to use R Studio (the free version), which is an integrated development environment for R (https://www.rstudio.com/).

Grades for the course will be based on:

- Participation in class demonstrating understanding of the readings and in-class material (10% of grade)

- Practical exercises (40% of grade). There will be four exercises. They are due by 11:59 PM on the day before class.
- Midterm exam (25% of grade).
- Group project (25% of grade). Over the course of the project you will identify a relevant dataset, develop a research question, form an analysis plan, carry out the analysis, and report on the results. There will be a few assignments specific to the group project, including a project proposal, a progress report (i.e., update), a presentation, and a final paper.

Attending Class:

This course uses a hybrid format with most students meeting in-person and a few students joining online. The recurring Zoom "meeting" link for INST 627 is: https://umd.zoom.us/j/96137972542 (passcode: 6271).

To facilitate discussions and foster an engaging and safe learning environment, students are encouraged to follow these suggestions when using Zoom in this course:
- Please sign in with your full first and last name. This makes it possible to know who attends and who is speaking.
- In general, please use a computer/laptop to access the meetings. Smartphones are okay but not ideal, given their small screens.
- Turn on your video when possible. It is helpful to be able to see each other, just as in an in-person class. If you have limited Internet bandwidth or no webcam, it is ok to not use video. You can use the Virtual Background function in Zoom if you do not want to share your actual background.
- Please mute your microphone when you are not talking. This helps eliminate background noise.
- Do not take any screenshots without explicit permission by all participants.

Questions may arise outside of class. If you email me a course-related question, I may either 1) post your question (removing your name for anonymity) and my answer as a Canvas Announcement so others can see it; or 2) post your question in a Canvas Discussion Forum (removing your name for anonymity) and encourage others to answer it. The goal of this approach is to facilitate peer-to-peer teaching (I usually don't understand something until I try to explain it to someone else) and to have everyone benefit from the information contained in the answer.

Accommodations for Students with Disabilities

In order to receive services you must contact the Disability Support Services (DSS) office to register for services. Please call the office to set up an appointment to register with a DSS counselor. Contact the DSS office at 301-314-7682; http://www.counseling.umd.edu/DSS/

Academic conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at the web site for the Office of the President (https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III-100A.pdf). Knowledge of these rules is your responsibility and ignorance of them does not excuse misconduct.

On the weekly exercises in INST 627, you may work with your colleagues to figure out the underlying concepts and problem-solving processes. However, are expected to work individually to answer the specific problems that are assigned.

Schedule and Reading Assignments

Topics	Required Readings/Tutorials	Due
W1: Aug 30 **Introduction, Measurement, & Design**		
W2: Sep 6 **Descriptive Statistics Overview**	-OSB Chapter 1, Sections 2-5, 7, 9, 11 Chapter 2, Section 5 Chapter 3, Sections 2-4, 12-13, 16 Chapter 6, Section 7 -R tutorials 1-7	• R installed • Find an R tutorial online. Post its name on Canvas with a brief review (~100 words) of its strengths and weaknesses.
W3: Sep 13 **R Practice**	-"Presenting and summarising data" (PDF on Canvas) -R tutorials 8-10	
W4: Sep 20 **Probability & Sampling** **(Guest Lecture – Chuck Huber)**	-OSB Chapter 7, Section 3 Chapter 9, Sections 2, 6 Chapter 11 Sections 2, 3, 6 -"Samples and populations" (PDF) -R tutorials 11-12	• Exercise 1 due by Sep 19
W5: Sep 27 **Hypothesis Testing** (one-sample t-tests)	-OSB Chapter 11, Sections 4-8 Chapter 12, Sections 2 -"Hypothesis testing and P values" (PDF) -R tutorial 13	

W6: Oct 4 **Hypothesis Testing** (two-sample t-tests)	-OSB Chapter 10, Sections 7-9, 11 Chapter 12, Sections 4 -"Comparison of means" (PDF) -R tutorials 14-15	
W7: Oct 11 **Chi-Square**	-OSB Chapter 17, Sections 2, 3, 5 -R tutorial 16	• Exercise 2 due by Oct 10
W8: Oct 18 **MIDTERM EXAM** **(Open Book)**		
W9: Oct 25 **Analysis of Variance** (ANOVA), One Way	-OSB Chapter 15, Sections 2-4 -"One-way analysis of variance" (PDF) -R tutorial 17	
W10: Nov 1 **Analysis of Variance** (ANOVA), Two Way	-OSB Chapter 15, Sections 6, 8	• Project Part 1 due by Oct 31
W11: Nov 8 **Correlations &** **Linear Regression**	-OSB Chapter 14, Sections 2-6 -"Correlation and Regression" (PDF) -R tutorials 18-20	• Exercise 3 due by Nov 7
W12: Nov 15 **Multiple Linear** **Regression**	-OSB Chapter 14, Section 9 -R tutorials 21-22	• Project Part 2 due by Nov 14
W13: Nov 22 **Review; Group Work**	(Meet Via Zoom)	• Exercise 4 due by Nov 21
W14: Nov 29 **Logistic Regression**	-"Logistic regression example in R" (PDF) -"Generalized Linear Models" (http://data.princeton.edu/R/glms.html)	• Project part 3 due by Dec 5
W15: Dec 6 **Presentations of project** **Dec 13: Final paper due**		

If there are updates to the schedule, they will be posted to Canvas.

Required R Tutorials

1. Downloading and Installing R
   (https://www.youtube.com/watch?v=cX532N_XLls/)
2. Import Data (https://www.youtube.com/watch?v=qPk0YEKhqB8)

3. Introduction to R (https://www.youtube.com/watch?v=UYclmg1_KLk – data set for this tutorial is published on Canvas)
4. Introduction to R II (https://www.youtube.com/watch?v=1BcGnHwUT6k)
5. Vectors in R (https://www.youtube.com/watch?v=2TcPAZOyV0U)
6. Subsetting Data (https://www.youtube.com/watch?v=jGf7WNh-LX8)
7. Basic Plots (http://www.cyclismo.org/tutorial/R/plotting.html)
8. Summary Statistics (https://www.youtube.com/watch?v=ACWuV16tdhY&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=20)
9. Basic Probability Distributions (http://www.cyclismo.org/tutorial/R/probability.html)
10. Z scores (https://www.youtube.com/watch?v=peEsXbdMY_4&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=26)
11. Calculating p-values (http://www.cyclismo.org/tutorial/R/pValues.html)
12. Calculating Confidence Intervals (http://www.cyclismo.org/tutorial/R/confidence.html)
13. One sample t-test (https://www.youtube.com/watch?v=kvmSAXhX9Hs)
14. Installing packages (https://www.youtube.com/watch?v=3RWb5U3X-T8&index=11&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU)
15. Two sample t-test (https://www.youtube.com/watch?v=RlhnNbPZC0A&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=29)
16. Chi Square test of independence (https://www.youtube.com/watch?v=POiHEJqmiC0&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=34)
17. Analysis of Variance (ANOVA) https://www.youtube.com/watch?v=lpdFr5SZR0Q
18. Scatterplots (https://www.youtube.com/watch?v=FEAS3akVxD8&index=19&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU)
19. Correlations (https://www.youtube.com/watch?v=XaNKst8ODEQ&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=36)
20. Linear Regression (https://www.youtube.com/watch?v=66z_MRwtFJM&list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU&index=37)
21. Multiple Linear Regression https://www.youtube.com/watch?v=q1RD5ECsSB0
22. Checking Linear Regression Assumptions https://www.youtube.com/watch?v=eTZ4VUZHzxw