

INST 627 – Data Analytics for Information Professionals

Section 0101,
Hornbake, Room 105
Wed 2:00 PM– 4:45 PM

Instructor: Yla Tausczik

E-mail: ylatau@umd.edu

Phone: (301) 405-2058

Office: 2118C Hornbake, South Building

Office Hours: Tuesdays 1:00pm-3:00pm or by appointment

Advances in hardware and software technologies have led to a rapid increase in the amount of data collected, with no end in sight. Decision making in the coming decades will depend, to an ever greater extent, on extracting meaning and knowledge from all that data. In this class we focus on one branch of statistics, inferential statistics, to help us reason about data. By gathering datasets, formulating proper statistical analyses and executing these analyses, information professionals play a significant role in bridging the gap between raw data and decision making.

This course will introduce basic concepts in data analytics including study design, measure construction, data exploration, hypothesis testing, and statistical analysis. The course also provides an overview of commonly used data manipulation and analytic tools. Through homework assignments, projects, and in-class activities, you will practice working with these techniques and develop statistical reasoning skills.

LEARNING OBJECTIVES

After completing this course you will be able to:

- Select and evaluate various types of data to use in decision making;
- Use prescriptive and descriptive analyses to reach defensible, data-driven conclusions;
- Select and apply appropriate statistical methods;
- Use MS Excel and SPSS or R for basic data manipulation and analysis;
- Critically evaluate data analyses and develop strategies for making better decisions.

COURSE MATERIALS

Software

The following software is necessary for you to successfully complete the homework, exams, and project for this course.

- **(Required)** Microsoft Excel 2007, 2010 or 2013 (or Excel 2011 for Macintosh)
- If you do not have access to a computer that has Excel 2007, 2010 or 2013 (or Excel 2011 for Macintosh) installed, consider downloading Office 2013

Professional Plus (for Windows) or Office 2011 for Macintosh through the university's TERPware website (<https://terpware.umd.edu>).

- **(Required)** You must have either SPSS or R. SPSS Statistical Analysis software (also called PAWS) - The student version available at TERPware website (<https://terpware.umd.edu>) is fine. R programming language and software is free and available online (<https://www.r-project.org/>). You may find the R commander graphical user interface easier to use (<http://www.rcommander.com/>). If you are specializing in data analytics or plan to take Digging into Data course you must use R. Otherwise you can choose between R and SPSS.
- (Recommended) Microsoft Access, MySQL, or some other relational database may be helpful for working with large datasets.

Readings and Online Resources:

There are many good texts and online sources for information on decision-making, statistical techniques and data tools. The main textbook used for the class will be The Online Stats Book (<http://onlinestatbook.com/2/index.html>) developed primarily by Rice University. This book is thorough, easy to understand, and is available for free in several formats. In addition, we will be collectively keeping an updated list of tutorials and references on Elms that prove to be useful during the class.

COURSE ACTIVITIES

Homework

Most weeks you will have an assignment that is designed to assess your mastery of the topics and techniques covered the previous week and provide you feedback to improve your understanding of the material. There will be a total of 11 homework assignments. Your 10 best scores on the homework assignments will be used to calculate your final grade, while the lowest score will be dropped.

You may work with your colleagues to figure out the underlying concepts and problem-solving processes, but are expected to work *individually* to answer the specific problems that are assigned. Completed assignments will be submitted via Canvas/ELMS. Timely submission of the completed assignments is essential. The due date of each assignment will be stated clearly in the assignment description. If an assignment due date is a religious holiday for you, please let me know at least one week in advance, so an alternate due date can be set. Late assignments will be penalized by 10% if they are turned in within one week of the due date and 50% if they are more than one week late.

Group Project

In groups of 2-3 you will prepare a data-related analytic project. Over the course of the project you will identify an interesting dataset, develop a research question, formulate an appropriate statistical analysis, 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 (update), a poster, and a final paper. Additional details about the group project will be provided in ELMS/Canvas and discussed in class.

Exams

There will be a midterm worth 20% of the course grade and a final worth 20%. These exams provide an opportunity for you to test your understanding of the concepts, techniques, and problems associated with statistical reasoning. In order to learn and understand the material fully it is important to review and revisit it multiple times.

Quizzes

Throughout the semester there will be periodic quizzes. These quizzes will not count toward your grade in the class. They are a way for you to test your knowledge and get rapid feedback.

Grading

30% Homework	150 pts.
30% Group Project	
Proposal (Sept. 23)	20 pts.
Update (Nov. 4)	10 pts.
Poster (Dec. 9)	20 pts.
Paper (Dec. 9)	100 pts.
40% Exams	
Midterm (October 21)	100 pts.
Final (Dec. 16)	100 pts.

Grades will be assigned based on the total number of points earned, using the following rubric. Please come and talk to me early if you think that there might be a problem. No extra credit will be given.

A	500-450pts (A- 465-450pts)
B	449-400pts (B+ 449-435pts ; B- 415-400pts)
C	399-350pts (C+ 399-385pts ; C- 365-350pts)
D	349-300pts (D+ 349-335pts ; D- 315-300pts)
F	299-0pts

OFFICE HOURS

Please visit me during office hours. This is an opportunity to ask questions about the material covered in the reading materials or in lecture. If you are having trouble in the course please talk to me as soon as possible. If you do poorly or lower than you expected on the first exam, it is imperative that you come to office ho that we can figure out the problem early.

ACADEMIC DISHONESTY

Cheating in any form (copying, falsifying signatures, plagiarism, etc.) will not be tolerated. It will result in a referral to the Office of Student Conduct irrespective of scope and circumstances, as required by university rules and regulations. There are severe consequences of academic misconduct, some of which are permanent and reflected on the student's transcript. If you have any questions regarding the University's policies on scholastic dishonesty, please see <http://osc.umd.edu/OSC/Default.aspx>.

It is very important that you complete your own assignments, and do not share files (excluding raw data), partial work or final work.

University of Maryland Code of Academic Integrity

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit <http://shc.umd.edu/SHC/Default.aspx>.

ACCOMODATIONS

Please come and see me as soon as possible if you think you might need any special accommodations for disabilities. In addition, please contact the Disability Support Services (301-314-7682 or <http://www.counseling.umd.edu/DSS/>). Disability Support Services will work with us to help create appropriate academic accommodations for any qualified students with disabilities. If you experience psychological distress during the course of the semester you can get professional help at the Counseling Center (301-314-7651 or <http://www.counseling.umd.edu/>).

COURSE SCHEDULE

Date	Topics	Required Readings	Due
9/2	Introduction	Rice's Online Stats Book Chapter I, Sections 4 , 5	Software Installed
9/9	Measurement & Design	Rice's Online Stats Book Chapter I, Sections 7 , 9 Chapter VI, Section 7	Co-developing questions and data
9/16	Descriptive Statistics	Rice's Online Stats Book Chapter I, Sections 11 Chapter II, Section 5 Chapter III, Section 3 , 13 , 16	Designing a Data Analytics Project
9/23	Sample, Population, Sampling Distributions	Rice's Online Stats Book Chapter VII, Section 3 Chapter IX, Sections 2 , 6 Chapter XI 2 , 3	Project Proposal
9/30	Hypothesis Testing (One-sample t-test)	Rice's Online Stats Book Chapter XI, Sections 4 , 6 , 7 , 8 Chapter XII, Section 2	Probability
10/7	Hypothesis Testing (Two-sample t-test)	Rice's Online Stats Book Chapter X, Sections 7 , 8 , 9 , 11 Chapter XII, Sections 4	Hypothesis Testing and Power
10/14	Chi-Squared	Rice's Online Stats Book Chapter XVII, Sections 2 , 3	T-Tests

10/21	Midterm Exam		Qualitative Data
10/28	Analysis of Variance (ANOVA)	Rice's Online Stats Book Chapter XV, Sections 2,3,4	
11/4	Correlations & Linear Regression	Rice's Online Stats Book Chapter XIV, Sections 2,4,5,6	<ul style="list-style-type: none"> • One-way ANOVA • Project Update
11/11	Multiple Regression	Rice's Online Stats Book Chapter XIV, Section 9	<ul style="list-style-type: none"> • Regression • Project Update Peer Feedback
11/18	Factorial Designs – Main Effects	Rice's Online Stats Book Chapter XV, Sections 6,8	<ul style="list-style-type: none"> • Multiple Regression
11/25	Factorial Designs - Interactions	Rice's Online Stats Book Chapter XIV, Section 9	Factorial Designs
12/2	Logistic Regression	UCLA tutorials (SPSS , R)	Work on Group Project
12/9	Group Presentations, Review, and Course Evaluations		<ul style="list-style-type: none"> • Final Project Poster • Final Project Paper
12/16 1:30pm-3:30pm	Final Exam		

This schedule is for planning purposes and may change. See ELMS/Canvas for current information and deadlines.