INST 314 – Statistics for Information Science
Section 0101,
Key Hall (KEY), Room 0126
Tue/Thu. 3:30 PM– 4:45 PM

Instructor: Yla Tausczik
E-mail: ylatau@umd.edu
Phone: (301) 405-2058
Office: 2118C Hornbake, South Building
Office Hours: Tuesdays and Thursdays 2:00pm-3:15pm or by appointment.

TA: Mona Swami
Email: mswami@umd.edu
Office: 0215A (CLIS) Hornbake Library
Office Hours: Mondays 3:00-5:00pm, Wednesdays 12:00-1:00pm

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 in class labs, homework assignments, projects, and exams, you will practice working with these techniques and develop statistical reasoning skills.

LEARNING OBJECTIVES

After successfully 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 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 in class labs, homework, projects and exams for this course. Every student will need access

to this software at home. In addition you must bring a charged laptop to class with this software and/or find a "lab partner" to work with in class who brings a charged laptop.

Required:

- R software. R is a programming language and software environment for data analysis. R is free and available online (https://www.r-project.org/). I have posted links to R tutorials on Elms/Canvas.
- Jupyter Notebooks (http://jupyter.org/install). Jupyter notebooks are documents that combine together exposition, code, output and visualizations. We will use Jupyter Notebooks for completing labs and creating data science reports.

Optional:

- Microsoft Excel, Open Office Calc, or Google Spreadsheets. Microsoft Excel is available for Macintosh through the university's TERPware website (https://terpware.umd.edu). Open Office Calc is a free software spreadsheet application available online (https://www.openoffice.org/product/calc.html). Google Spreadsheets can be found on Google Drive (https://www.google.com/drive/). You may find it helpful to inspect some of your data as spreadsheets (if it is small enough).

Readings: Completing the required reading for the class is essential to understanding the core statistical concepts. In order to learn, you must review the material multiple times. The required reading listed below in the course schedule is given for the main textbook used in the class. There are a few advanced topics that are beyond the course textbook; readings for these sections will be posted on Elms/Canvas. The readings sometimes cover topics that go beyond what is covered in class, you are only responsible for the topics covered in class.

Required:

Online Statistics by David Lane (http://onlinestatbook.com/). This book is thorough, easy to understand, and is available for free on the web, as an e-book, or as a pdf. We will not follow the book in sequence and we will not cover all sections. See course schedule below and Elms/Canvas for required reading.

Optional Extra Reading:

- Radziwill, Nicole M. *Statistics (The Easier Way) with R: an informal text on applied statistics*. This book provides even more accessible explanations, procedures, and examples.
- Rice, John A. *Mathematical Statistics and Data Analysis*. This book provides more mathematical detail about the techniques we will cover in the class.

COURSE ACTIVITIES

In Class Labs: This course will use a flipped course design. Each lecture will consist of a very short lecture (~20 minutes) followed by in class exercises. Thus, class time will focus on developing statistical skills and reasoning through actively engaging with data.

In order to do the labs successfully it will be important to do the readings before class as well as to attend class. You can work on these as pairs in class, but all work and code must be created independently. If you work with someone else in class you must include the name of your partner in your turned in lab. Completed labs will be composed using Jupyter notebooks and submitted via Elms/Canvas.

Homework: There will be a total of 8 homework assignments. You will be given 2-5 textbook problems that cover material from the last few lectures. These assignments are meant to assess your mastery of the topics and techniques covered in class. Completed solutions will be submitted via Elms/Canvas.

Projects: There will be a total of 5 projects. These are your opportunity to apply concepts learned in class to real problems and data sets. These assignments will be 2 to 3 page reports composed using Jupyter notebooks. 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 Elms/Canvas.

Exams: There will be one midterm and one cumulative final each worth 20% of your final grade. 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.

Grading:

In Class Labs	20%
Homework	20%
Projects	20%
Exams	40%

- Midterm (20%)
- Cumulative Final (20%)

Grades will be assigned based on the total percent of points earned, using the following rubric. Grades will be rounded to the nearest 10th of a percent. Please come and talk to me early if you are think that there might be a problem. If you are consistently getting grades that are lower than you would like you will need to change the way you are studying and practicing. Please come get help during office hours.

A	90.0-100% (A- 90.0-92.9%)
B	80.0-89.9% (B+ 87.0-89.9%); B- 80.0-82.9%)
C	70.0-79.9% (C+ 77.0-79.9%; C- 70.0-72.9%)
D	60.0-69.9% (D+ 67.0-69.9%; D- 60.0-62.9%)
F	0-59.9%

COURSE POLICIES

Excused Absences: If an assignment due date or exam is a religious holiday for you, please let me know at least one week in advance, so an alternate due date can be set. Missed homework and exams with an excused absence must be made up within 2 weeks of the original deadline. Missed homework or exams without a documented, excused absence cannot be made up and will receive a score of 0.

Late Work: Timely submission of the completed assignments is essential. The due date of each assignment will be stated clearly in the assignment description 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. All other work must be turned in by the last scheduled day of class Monday December 9th, 2019.

Regrading: Fairness in giving grades is very important to me, at the same time both our time is best spent on helping you learn the material. Regrading of assignments and exams must be turned in within one week of receiving the graded work. They must be submitted as a written document in which you include the graded work, an explanation of what you believe was missgraded, and an explanation for why you think it should be given a different score. For any regrade requests, the entire assignment will be regraded and your score may go up or down.

Extra Credit: I very rarely offer extra credit opportunities. I believe that the labs, homework, projects, and exams are the best way to practice the course objectives and to show mastery of the material. If you are having difficulty scoring well on these assignments I'm happy to work with you during office hours to help you study more effectively and to improve your grades. In addition, if you can demonstrate on the comprehensive final exam that you have learned more than your grade reflects I will raise your final grade to be within one letter grade of your grade on the final exam (before any curve is applied). For example, if you scored 90.0% on the final before any curve you would receive at least a B in the class regardless of your other grades.

Other Policies: Other policies relevant to undergraduate courses are found here: http://ugst.umd.edu/courserelatedpolicies.html. Topics that are addressed in these various policies include academic integrity, student and instructor conduct, accessibility and accommodations, attendance and excused absences, grades and appeals, copyright and intellectual property.

GETTING HELP

Feel free to email the TA or me about the course material. I will respond to your emails within 48 hours. However, I usually do not respond to emails in the evenings or on the weekends. If you know you need help please do so **at least** 48 hours before a deadline.

Please visit the TA or me during office hours if you want extra help. We won't give you the answers to the assignments, but we will go over the material with you and help answer your questions. 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 hours so 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.

ACCOMMODATIONS

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/).

Week	Day	Unit	Topic	Reading Section(s)	Due
1	Tu 8/27	Getting Started	Descriptive Stats	III: 4, 13	
1	Th 8/29		Distributions	I: 11	
2	Tu 9/3		Scales of Measurement	I: 9	
2	Th 9/5		Sampling Procedures	VI: 5	Lab 1
3	Tu 9/10	One sample t-tests	Hypothesis Testing	I: 5, XI: 3	Homework 1
3	Th 9/12		Central Limit Theorem	IX: 6	
4	Tu 9/17		One sample t-test	XI: 4, XII: 2	Project 1
4	Th 9/19		Effect Size	XIX: 1	
5	Tu 9/24		Power	XIII: 2, 6	Homework 2
5	Th 9/26		Confidence Intervals	X: 8	Lab 2
6	Tu 10/1	Two sample t-tests	Two sample t-test	XII: 4	Project 2
6	Th 10/3		Effect Size, Graphing	XI: 6, 7	Lab 3
7	Tu 10/8	Chi Square Tests	Goodness of Fit	XVII: 2, 3	Homework 3
7	Th 10/10		Test of Independence	XVII: 5	Lab 4
8	**Tu 10/15**	**Midterm**			
8	Th 10/17	ANOVA	One Way ANOVA	XV: 4	
9	Tu 10/22		Assumptions		Homework 4
9	Th 10/24		Post hoc tests	XII: 6	Project 3
10	Tu 10/29		Two way ANOVA		Lab 5
10	Th 10/31	Linear Regression	Correlations	IV: 3, 6	Homework 5
11	Tu 11/5		Linear Model	XIV: 2	Project 4
11	Th 11/7		Assumptions	XIV: 6	Lab 6
12	Tu 11/12	Multiple Regression	Linear Model	XIV: 9	Homework 6
12	Th 11/14		Coefficients		
13	Tu 11/19		Dummy Coding	See Elms	Lab 7
13	Th 11/21	Logistic Regression	Generalized Linear Model	See Elms	Homework 7
14	Tu 11/26		Assumptions	See Elms	Project 5
14	Th 11/28		Thanksgiving		
15	Tu 12/3		Log Odds	See Elms	Lab 8
15	Th 12/5	Review			Homework 8

Cumulative Final Exam (Tuesday Dec 17 10:30-12:30pm)

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