

INST 314 – Statistics for Information Science

Section 0101,
LeFrak Hall (LEF), Room 1220
Tue/Thu. 3:30 PM– 4:45 PM

Instructor: Yla Tausczik

E-mail: ylatau@umd.edu

Phone: (301) 405-2058 (I rarely check voicemail, email instead)

Office: 2118C Hornbake, South Building

Office Hours: Tuesday/Thursday 2:00pm-3:30pm, 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 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 Clean up, standardize, and normalize data to prepare for data 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. 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:

- You must install and use R. R programming language and software is free and available online (<https://www.r-project.org/>). I have posted links to R tutorials on Elms/Canvas.

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).
- R development tools such as Jupyter notebooks or RStudio. Jupyter notebooks are documents that blend together exposition, code, and visualizations. R studio is an integrated development environment for R (<http://web.cs.ucla.edu/~gulzar/rstudio/>).

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.

Required:

- Diez, David M., Barr, Christopher, D., Cetinkaya-Rundel, Mine. *Open Intro Statistics* (https://www.openintro.org/stat/textbook.php?stat_book=os). This book is thorough, easy to understand, and is available for free as a pdf or for a nominal amount as a printed textbook. See course schedule below 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

Homework: There will be a total of 8 homework assignments in weeks when there is no exam or project due. These assignments are meant to assess your mastery of the topics and techniques covered in class. You will be given 2-5 textbook problems that cover material from the last few lectures. 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. 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 25% 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:

Homework	25%
Projects (5)	25%
Exams	50%

- Midterm (25%)
- Cumulative Final (25%)

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.

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 quizzes and exams with an excused absence must be made up within 2 weeks of the original deadline. Missed assignments, quizzes, 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 work must be turned in by the last scheduled day of class Monday December 10th, 2018.

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, quizzes, 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 regarded and your score may go up or down.

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.

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 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/>).

COURSE SCHEDULE

Week	Week	Reading Section(s)	Due (Homework and Projects due Thursdays)
1	Distributions	1.1-1.8	Homework 1
2	Descriptive Statistics	3.1-3.2	Homework 2
3	Two sample t-tests	4.1-4.4	Homework 3
4	Two sample t-tests	5.1-5.4	Homework 4
5	Chi Square Tests	6.3-6.4	Project 1: Hypothesis Testing
6	ANOVA	5.5	Homework 5
7	ANOVA		Homework 6
8	Review and Midterm (Thu. Oct. 18)		
9	Factorial ANOVA	see Elms	Project 2: ANOVA
10	Factorial ANOVA	see Elms	
11	Correlations	7.1	Project 3: Factorial ANOVA
12	Linear Regression	7.2-7.5	Homework 7
13	Multiple Regression	8.1-8.3	Project 4: Linear Regression – due on Tuesday
14	Multiple Regression	8.1-8.3	Homework 8
15	p Hacking & Review	see Elms	Project 5: Multiple Regression
Cumulative Final Exam (Tuesday Dec 18 10:30-12:30pm)			

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