

# INST 314 – Statistics for Information Science

Section 0105

**Instructor:** Abdirisak Mohamed

**E-mail:** [amoham70@umd.edu](mailto:amoham70@umd.edu) (DO NOT email me via ELMS)

**Office:** HBK 4120K

**Office Hours:** Online and by appointment

**Teaching Assistant:** Courtney Douglass

**E-mail:** [cdoug88@umd.edu](mailto:cdoug88@umd.edu)

**Office:** HBK 4117M

**Office hours:** Online and 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.

**Prerequisite:** Must have completed or be concurrently enrolled in INST201; or must have completed or be concurrently enrolled in INST301. And minimum grade of C- in INST201 and INST301; and MATH115; and STAT100; and minimum grade of C- in MATH115 and STAT100.

## 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 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 course activities.

Required:

- You must install and use R. R programming language and software is free and available online (<https://www.r-project.org/>). You may want to use R Studio, which will be used predominately through the class; R studio is an integrated development environment for R (<http://web.cs.ucla.edu/~gulzar/rstudio/>). I have posted links to R tutorials on Elms/Canvas.

Optional

- Microsoft Excel or Open Office Calc. 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>).
- GitHub/Git. Links will be posted on Elms/Canvas

### 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, D., Barr, C., and Çetinkaya-Rundel, M. (2015), *OpenIntro Statistics*, 3<sup>rd</sup> ed., OpenIntro. <https://www.openintro.org/stat/textbook.php> Available as a free pdf download, color hardcover of black & white paperback for purchase, and free interactive website.
- Verzani, J. (2004), *Using R for Introductory Statistics*, Chapman & Hall/CRC. ISBN/ASIN: 1584884509 ISBN-13: 9781584884507 <https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>
  - Available as a free pdf download.

### Recommended Extra Reading:

- Rice, J.A. (2006). *Mathematical Statistics and Data Analysis*, 3<sup>rd</sup> ed. Cengage Learning. ISBN-13: 978-8131519547. This book provides more mathematical detail about the techniques we will cover in the class.
- Golemund, G. & Wickham, H. (2017), *R for Data Science*, O'Reilly Publishing, Inc. ISBN-13: <http://r4ds.had.co.nz/index.html> . This is a free book available in HTML and a physical copy is available for purchase. This book focuses on working with data in R, rather than statistics, but many of the techniques may be useful to prepare your data for statistical analysis.

## **Course Activities**

### **Homework**

There will be a total of 8 homework assignments. These assignments are meant to assess your mastery of the topics and techniques covered in class. These will be textbook-style problems that cover material from the last few lectures. Homework assignments will be submitted via ELMS.

### **Quizzes**

There will be a total of 7 quizzes. Each is designed to test your knowledge from recent lectures (since the last quiz) and provide you rapid feedback to improve your understanding of the material. Quizzes will be taken on Elms/Canvas on the scheduled Day.

### **Projects**

There will be a total of 3 projects. These are your opportunity to apply concepts learned in class to real problems and data sets from ideas and hypotheses you develop and test. These assignments will be memo-style narrative reports with statistical analysis. You may work with your colleagues to figure out the underlying concepts and problem-solving processes but are expected to work *individually* to write your own project paper. Projects will be submitted via ELMS

### **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

20% Homework including DataCamp

10% Quizzes

30% Projects

40% Exams

Midterm 20%

Cumulative Final 20%

Grades will be assigned based on the total percent 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 think that there might be a problem.

A-	90.0-92.9%	A	93.0-100%		
B-	80.0-82.9%	B	83.0-86.9%	B+	87.0-
	89.9%				
C-	70.0-72.9%	C	73.0-76.9%	C+	77.0-
	79.9%				
D-	60.0-62.9%	D	63.0-66.9%	D+	67.0-
	69.9%				
		F	0-59.9%		

## Excused Absences

If an assignment due date or scheduled exam is a religious holiday for you or other university approved activity, please let me know at least one week in advance, so an alternate due date can be set. Missed exams with an excused absence must be made up within 2 weeks of the original deadline. Missed 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. Late assignments will be penalized.

Late assignments will be penalized by 50% if they are turned in within one week of the due date and not accepted if they are more than one week late.

## Re-grading

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. Re-grading 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 miss-graded, and an explanation for why you think it should be given a different score. For any re-grade requests, the entire assignment will be regarded and your score may go up or down.

**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 (online) 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/>

## **Class and Campus Cancellations**

This is an online course and is not limited to campus closures and weather events.

### **INST 314, Section 105 Course Calendar**

#### **Week One: 8/26 – 9/1, *Statistical Foundations***

**Readings:** Week 1 Materials ELMS/Canvas

**DUE: Homework 1, September 1, 11:59 pm**

#### **Week Two: 9/2 – 9/8, *Basic Data Management***

**Readings:** Week 2 Materials ELMS/Canvas

**DUE: Quiz 1, September 8, 11:59 pm**

#### **Week Three: 9/9 – 9/15, *Descriptive Statistics and Graphs***

**Readings:** Week 3 Materials ELMS/Canvas

**DUE: Homework 2, September 15, 11:59 pm**

#### **Week Four: 9/16 – 9/22, *Hypothesis Testing***

**Readings:** Week 4 Materials ELMS/Canvas

**DUE: Quiz 2, September 22, 11:59 pm, Homework 3, September 22, 11:59 pm**

#### **Week Five: 9/23 – 9/29, *Chi-Square Tests***

**Readings:** Week 5 Materials ELMS/Canvas

**DUE: Quiz 3, September 29, 11:59 pm, Homework 4, September 29, 11:59 pm**

**Week Six: 9/30 – 10/6, *Central Limit Theorem***

**Readings:** Week 6 Materials ELMS/Canvas

**DUE: Quiz 4, October 6, 11:59 pm, Homework 5, October 6, 11:59 pm**

**Week Seven: 10/7 – 10/13, *One Sample Z & T Tests***

**Readings:** Week 7 Materials ELMS/Canvas

**DUE: Project 1, October 13, 11:59 pm**

**Week Eight: 10/14 – 10/20, *Midterm Review***

**Readings:** Week 8 Materials ELMS/Canvas

**DUE: Midterm Exam, October 20, 11:59 pm**

**Week Nine: 10/21 – 10/27, *One-Way ANOVA***

**Readings:** Week 9 Materials ELMS/Canvas

**DUE: Quiz 5, October 27, 11:59 pm**

**Week Ten: 10/28 – 11/3, *Two Sample Z & T Tests***

**Readings:** Week 10 Materials ELMS/Canvas

**DUE: Homework 6, November 3, 11:59 pm**

**Week Eleven: 11/4 – 11/10, *Correlations***

**Readings:** Week 11 Materials ELMS/Canvas

**DUE: Quiz 6, November, 11:59 pm,**

**Week Twelve: 11/11 – 11/17, *Linear Regression***

**Readings:** Week 12 Materials ELMS/Canvas

**DUE: Project 2, November 17, 11:59 pm**

**Week Thirteen: 11/18 – 11/24, *Multiple Regression***

**Readings:** Week 13 Materials ELMS/Canvas

**DUE:** Quiz 7, November 24, 11:59 pm, Homework 7, November 24, 11:59 pm

**Week Fourteen:** 11/25 – 12/1, *Bayesian and Causal Inference*

**Readings:** Week 14 Materials ELMS/Canvas

**DUE:** Quiz 8, December 1, 11:59 pm, Homework 8, December 1, 11:59 pm

**Week Fifteen:** 12/2 – 12/8, *Final Exam Review*

**Readings:** Week 15 Materials ELMS/Canvas

**DUE:** Project 3, December 8, 11:59 pm

**Final Exam Week**