### Introduction to Data Science STA 2XXX

**Instructor:**

Dr. Daniel Maxwell

Marston Science Library L301B, danielmaxwell@ufl.edu  
(352) 294-1034

**Office Hours:**

Wednesday (2:00 to 5:00 pm)

Required Texts

Carlson, K., & Winquist, J. (2018). *An introduction to statistics: An active learning approach* (2nd ed.) Thousand Oaks, CA: Sage Publications. ISBN 9781483378732.

This is a dramatically revised second edition so you cannot use the first edition of the book.

Stowell, S (2014). *Using R for statistics* (1st ed.). Apress. ISBN 978-1484201404

An eBook version of this text is available for free on the library’s website.

Required Software

You will be required to use the free Rstudio integrated development environment (IDE). Instructions on how to install this software package are provided in Module 1: R and RStudio.

Office Hours, Tutoring, and Academic Support

I have scheduled office hours each week, but I am available at other times as well. If you would like to meet with me, just send an email with times you are available and we can set up a time to meet.

We will also have student tutors hold regular office hours each week. These hours will be posted on Canvas as soon as they are available.

Course Objectives

The three main objectives of this course are to teach you: (1) basic statistics; (2) the fundamentals of writing code in Python or R; and (3) basic data management. To be a Data Scientist, you will need to be skilled in these three areas, and this class introduces the concepts and tools you need to master before taking advanced data science classes or conducting exploratory data analysis in the workplace.

Student Learning Objectives

The Student Learning Objectives (SLOs) for this course are:

1. Students will be able to navigate within the RStudio (IDE)
2. Students will be able to manage datasets using R. Specifically, students will be able to:
3. Utilize R’s basic datatypes and type conversion functions
4. Import data from files in .csv and .txt formats
5. Create a dataframe and the add columns and/or rows to it
6. Demonstrate proficiency with R operators
7. Explain how R handles missing or null values
8. Demonstrate proficiency creating subsets from an existing dataframe
9. Students will be able to create basic graphs in R. Specifically, students will be able to:
10. Implement sound design principles in the creation of graphs
11. Demonstrate proficiency with R’s graphical parameters
12. Combine multiple plots into a single output
13. Demonstrate proficiency creating bar plots, histograms, boxplots, and other graphs
14. Students will be able to compute and interpret basic descriptive statistics (e.g., mean, median, mode, standard deviation, z scores).
15. Students will be able to use null hypothesis significance testing procedures appropriately. Specifically, students will be able to:
    1. State the null and research hypotheses
    2. Compute degrees of freedom and define a critical region
    3. Compute statistics using the R programming language
    4. Compute an effect size
    5. Summarize the results of a study
16. Students will demonstrate an understanding of the logic of hypothesis testing including p-values, Type I errors, Type II errors, statistical power, and effect sizes.
17. Students will be able to compute and interpret confidence intervals.
18. Students will be able to determine which statistic is appropriate in different situations.
19. Students will be able to create and interpret graphs.
20. Students will be able to generate and interpret R output for all of the statistics described above.
21. Students will appreciate the value of statistics.

Course Communication

All course-related communication and questions should be posted to the Canvas learning environment. Private questions should be sent to the instructor via Canvas -- the course management system.

Course Structure

The course has been designed to minimize lectures and to maximize your active learning of the material. Although I provide brief video lectures in Canvas, the vast majority of your time will be spent working on activities that were designed to help you develop an understanding of the material.

In each module, you will read the chapter and complete reading questions. Afterwards, you will complete activity worksheets on the same topic as the reading. After completing the reading, reading questions, and activities, you will then take a practice test on each chapter. Finally, before taking the exams you will also complete a unit practice exam.

Reading Questions

Reading questions are embedded in each reading assignment. You must submit your answers to these reading questions on or before the due-date by entering them into the Canvas quiz. After you enter your answers, Canvas will indicate which items you answered incorrectly. If any of your responses were incorrect you can reread the relevant portion of the chapter, correct your answer, and then resubmit your answers. Your grade on the reading questions will be the ***AVERAGE*** of your first and second attempt. *You may only take the reading questions twice*. These reading questions are not intended to be difficult. If you actually read the assigned chapters you will find that most of the reading questions are straight forward.

At the end of the course, reading questions will be worth 150 points.

Activities and Practice Exams

The activities in this course will prepare you for exams. You will be able to finish many activities during class but you will be expected to complete activities outside of class. All of the answers to these activities will be provided on Canvas. Your goal when completing the activities is to understand the material rather than simply get the right answers. To earn points for completing the activities, *you will need to turn in the activities no later than midnight the day indicated on the syllabus*. The final answers for all activities are provided on Canvas, therefore, you must show your work. If you turn in your activity on time, complete, and with your work shown, you will earn full credit for that assignment. If it is only partially complete, your work is not shown, or it is turned in late, you will receive partial credit. Under extenuating circumstances, extensions may be given, but only if I approve the extension in advance. *Under no circumstances will activities be accepted after the exam covering that chapter was given*.

I have structured this course so that you will take a practice exam before the actual exam. This will give you time to determine what you understand and what you still need to work on. If the practice test does not go well, I encourage you to review the material that you missed and then do the chapter practice tests on Canvas. I have also scheduled the exams so that you have five days between taking the practice exam and taking the graded exam. This gives you ample opportunity to review, meet with me, and/or seek help from a tutor. Unit sample exams must be taken on the day listed on the syllabus. Under extenuating circumstances, extensions may be given, but only if I approve the extension in advance.  *Under no circumstances will practice exams be accepted after the exam covering that chapter was given*.

At the end of the course, activities and practice exams will be worth 250 points.

Coding Activities

For these assignments, you will first need to create a script file. In the script file, add comments to identify the code for each question. Here’s a template of how this should be done. Please note: the pound sign (#) identifies user-supplied comments in the code.

# Your Name

# Date

# Coding Assignment Number

# Question 1

Your code goes here

For each line or block of code executed, include console output here

# Question 2

Your code goes here

For each line or block of code executed, include console output here

# Etc…

Once you’ve tested your code by executing each line in the console, save the file and then convert it to a .pdf file for upload to Canvas. A quick way to do that is to copy-and-paste everything in the script console into a MS Word document and then save the Word document as a .pdf file, using the File -> Save As option.

If you’re adventurous, explore the possibility of creating an RMarkdown document and then “knitting” it to a .pdf. Be sure to use the template provided above to construct your document. Also, you’ll need to install either MikTex or MacTex on your computer in order to create a .pdf document using RStudio’s “knit” function.

At the end of the course, the R coding assignments will be worth 80 points, 20 points per assignment.

Student Evaluation and Assessment

Your grade for this course will be based on 4 exams (100 points each), a comprehensive final exam (120 points), reading question assignments (150 points as a set), R coding assignments (80 points), discussion forum participation (50 points), and activities and practice exams (200 points as a set).

Your final grade will be the percent of points you receive out of the 1000 points that are possible. The following grade cuts will be used:

93 to 100% = A 87 to 89.9% = B+ 77 to 79.9% = C+ 67 to 69.9% = D+

90 to 92.9% = A- 83 to 86.9% = B 73 to 76.9% = C 63 to 66.9% = D

80 to 82.9% = B- 70 to 72.9% = C- 60 to 62.9% = D-

Below 60% = F

At the end of the semester I will use the following formula to compute your final course grade:

Final Grade = ((Test1% + Test2% +Test3% + Test4% + (proportion correct on all reading questions \* 150) + Activity and practice test points + R coding assignments + discussion forum points + (proportion correct on Final \* 120))/1000. The table below lists total points for each deliverable as a % of the total.

|  |  |  |
| --- | --- | --- |
| Deliverable | Points | % Value |
| Unit Exams | 400 | 40 |
| Final Exam | 120 | 12 |
| Activities + Practice Exams | 200 | 20 |
| Reading Questions | 150 | 15 |
| R Coding Assignments | 80 | 8 |
| Discussion Forum Participation | 50 | 5 |

Academic Honesty and “Working Together” on Homework

The policy for *exams* is INDIVIDUAL WORK. Group work during exams will not be permitted. The time for learning from each other is during the activities, additional problems, and practice tests. However, it is important to note that it is inappropriate to copy any portion of another student’s work even if that student took the course during a previous semester. “Working together” is not equivalent to “copying.” Working together means that each person is actively engaged in trying to understand the material not just “getting the right answer to turn in.” If you have any questions about what constitutes appropriate “working together” you should ask me. Of course, I expect you to uphold the honor code of the University in its entirety: “On my honor, I have neither given nor received unauthorized aid in doing this assignment.” Using another’s completed course work is unauthorized aid. I expect you to do more than simply follow rules. I expect you to act honorably in all of your exchanges with me, and I will act honorably when I interact with you.

Makeup Policy

Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at:

<https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx>.

Course Schedule & Assignment Due Dates

|  |  |  |
| --- | --- | --- |
| Date | Reading Questions | Activities |
| Week 1 | Getting Started in R | * Getting Started in R (Worksheet) |
| Week 2 | Basic Data Management in R | * Data Management in R (Worksheet) |
|  | Basic Graphing in R | * Graphs in R (Worksheet) |
| Week 3 | Ch 1 RQ: Frequency Distributions | * Activity 1-1 (R) |
|  | Ch 2 RQ: Central Tendency | * Activity 2-1 |
| Week 4 | Ch 3 RQ: Variability | * Activity 3-1 |
|  | Ch 4 RQ: z Scores | * Activity 4-1 * Practice Exam and Review |
| Week 5 |  | **Exam 1 (100 points)** |
|  | Ch 5 RQ:  Distribution of Sample Means and z for a Sample Mean | * Activity 5-1 |
| Week 6 |  | * Activity 5-2 * Applied Activity on sampling error (not in book); confidence interval; error bars; boxplot |
|  | Ch 6 RQ: Hypothesis Testing | * Activity 6-1 |
| Week 7 |  | * Activity 6-2 * Activity 6-3 |
|  |  | * Activity 6-4 * Practice Exam and Review |
| Week 8 |  | **Exam 2 (100 points)** |
|  | Ch 7 RQ: Single Sample t | * Activity 7-1 (R) |
| Week 9 | Ch 8 RQ: Confidence Intervals | * Activity 8-1 (R) |
|  | Ch 9 RQ: Related Samples t | * Activity 9-1 (R) * Activity 9-2 (R—I’ll give you the output file) |
| Week 10 | Ch 10 RQ: Independent Samples t | * Activity 10-1 |
|  |  | * Activity 10-3 (choose stat) * Activity 10-2 (R) |
| Week 11 |  | * Activity 10-5 * Activity 10-4 (collect data in R) |
|  |  | * Practice Exam and Review |
| Week 12 |  | **Exam 3 (100 points)** |
|  | Ch 11 RQ: One-way ANOVA | * Activity 11-1 * Activity 11-2 (R—I’ll give you the output file) |
| Week 13 |  | * Activity 11-4 * Activity 11-3 (R) * Activity 11-6 (choose stat) * Practice Exam and Review 11-5 is not required, but gives more practice writing up one-way ANOVAs |
|  |  | * Practice Exam and Review 11-5 is not required but gives more practice writing up one-way ANOVAs |
| Week 14 | **Break** |  |
|  |  | **Exam 4 (100 points)** |
|  | Ch 13 RQ: Correlation and Regression | * Activity 13-1 (R) * Activity 13-3 (R) |
| Week 15 |  | * Activity 13-5 (choose stat) * Activity 13-4 * Assessment Surveys |
|  | Ch 14 RQ: Chi Square | * Activity 14-1 * Activity 14-2 * Activities 13-1, 13-4, and 14-1 |
| Week 16 |  | * Complete Practice Final |
|  | **Comprehensive Final Exam (120 points)** |  |

Students with Disabilities

Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc/) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

Online Course Evaluations

“Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at <https://evaluations.ufl.edu/results/>.