DATA 229
Statistical Programming

Course Description
Students in this course will use a statistical programming language to perform effective data analysis. Students will acquire statistical programming skills including reading data, accessing statistical packages, writing functions, debugging, profiling code, organizing code and commenting code.

Illinois Articulation Initiative (IAI) number: N/A

Credit and Contact Hours:
Lecture   **3**
Lab   **0**
Credit Hours   **3**

Prerequisites: Grade of "C" in Math 228 or equivalent.

Books, Supplies, and Supplementary Materials

A.   **Textbooks**
      Wickham, H. and Grolemund, G., *R for Data Science*. Riley. https://r4ds.had.co.nz/

B.   **Other Required Materials**
      R and *R-Studio* free downloads.

Methods of Instruction:
        Hybrid, Online

General Education Student Learning Outcome
Quantitative Literacy: Students possess the ability to reason and solve quantitative problems from an array of contexts.

Course Learning Outcomes (CLOs)
1. Students describe the statistical programming language in the context of big data.
2. Students apply basic skills of a statistical programming language.
3. Students use the statistical programming language to access data.
4. Students construct statistical programming language functions.
5. Students use the statistical programming language advanced data structures.

6. Students analyze real-life problems using the statistical programming language.
7. Students explain the solutions for their analysis.

Lesson Learning Outcomes (LLOs)
1. Students describe the statistical programming language.
2. Students describe the benefits of a statistical programming language.
3. Students apply the statistical programming language to big data.
4. Students explain the statistical programming language packages.
5. Students use variables.
6. Students demonstrate vectors.
7. Students construct control statements.
8. Students write loops.
9. Students create functions.
10. Students interpret missing data.
11. Students use comma separated value (CSV) files.
12. Students employ spreadsheet data.
13. Students select databases.
14. Students manipulate data included with the statistical programming language.
15. Students use data from websites.
16. Students memorize syntax.
17. Students use function arguments.
18. Students use return values.
19. Students use data frames.
20. Students employ lists.
21. Students construct matrices.
22. Students create arrays.
23. Students formulate the problem.
24. Students collect relevant data.
25. Students analyze the data.
26. Students apply the statistical programming language.
27. Students prepare solution documents.
28. Students justify solutions orally.
29. Students present visualizations.

TOPICAL OUTLINE/TIMELINE

Week	LLOs Learned	CLOs Learned	PLOs Learned
Week One	1,2	1	1
Week Two	3,4	1	4
Week Three	5,6,7	2	4
Week Four	8,9,10	2	4
Week Five	5,6,7,8,9,10	2	4
Week Six	11,12,13	3	2,4
Week Seven	14,15	3	2,4
Week Eight	Midterm		
Week Nine	16,17,18	4	2,4
Week Ten	19,20	5	2,4
Week Eleven	21,22	5	2,4
Week Twelve	23,24,25,26	6	2,3,4
Week Thirteen	23,24,25,26	6	2,3,4
Week Fourteen	23,24,25,26	6	2,3,4

| Week Fifteen | 27,28,29 | 7 | 4,5 |
| Week Sixteen | 27,28,29 | 7 | 4,5 |

Graded Assignments

The individual instructor will determine which items he or she considers essential for the student to memorize without error and test accordingly. The individual instructor will determine the types of projects that the student will complete during the class

Final Course Grading Scale

Grade   Percentage
A   90-100%
B   80-89%
C   70-79%
D   60-69%
F   lower than 60%

Faculty Commitment

Faculty members are committed to providing a quality learning experience through thoughtful planning, implementation, and assessment of course activities. They are also committed to being readily available to students throughout the semester by returning e-mails and phone calls within 48 hours and to returning graded course work within a week. Furthermore, they are committed to selecting appropriate course materials and making them available in an organized and timely manner.

Student Commitment

For every credit hour a student is enrolled in, they should expect to spend at least 2 hours outside of class studying, working on assignments, and preparing for class each week of the fifteen-week semester. For example, for this four credit-hour class, students can expect to spend four hours per week in class actively engaged in learning the material by participating in face-to-face classes or viewing lectures and instructional material online. In addition, students should expect to spend another eight hours per week outside of class completing homework and assignments, posting to discussion boards online, or studying for quizzes and tests. This means students should spend a minimum of 12 hours per week engaged in achieving the learning outcomes for this course. If you are not achieving your desired results in this class, you should consider increasing your prep time outside of class, in addition to using available resources such as instructor office hours and tutoring services.

By registering for this course, you commit yourself to active participation in course activities as well as the submission of all assignments and exams on time. Furthermore, you commit to accessing the course site and checking your JJC e-mail several times a week.

TOPICAL OUTLINE

Unit	Topic	Assignments
1	Statistical Programming Intro I • Define statistical programming • Describe how statistical programming is prevalent in different languages • Identify the benefits of statistical programming in different situations	Orientation quiz Homework Project: Short research paper describing data science and statistical programming.
2	Statistical Programming Intro II • Apply statistical programming • Explain when different packages would be used and for what purposes	Homework Project: Short research outline and presentation describing common statistical packages and benefits to each
3	Employ Statistical Programming I • Define and use variables • Define and demonstrate vectors • Define and construct control statements	Quiz: Topics to date Homework: Swirl exercises
4	Employ Statistical Programming II • Define and implement repetition. • Define the purpose and benefits of functions. • Implement functions • Interpret missing data	Homework: Swirl exercises
5	Employ Statistical Programming III • Locate more complex data sets • Perform more complex analysis	Homework: Swirl exercises
6	Spreadsheets and databases • Define and use CSV files • Analyze spreadsheet data using a common spreadsheet application. • Implement selection from databases	Quiz: Topics to date
7	Data extraction • Collect and use data from websites • Manipulate data	Homework Project: Discover, extract, and manipulate data sets for case study situations
8	Midterm exam	
9	Employ Statistical Programming IV • Memorize syntax • Describe the need for function arguments. • Implement functions with arguments. • Describe the need for function return values • Implement functions with return values.	Homework: Swirl exercises
10	Framing data I • Define and use data frames	Homework Project: Outline research topic

	• Implement lists programmatically.	
11	Framing data II • Describe the need for and construct matrices • Describe the need for and implement arrays	Quiz: Topics to date Homework: Swirl exercises
12-14	Analyze real-life problems using the statistical programming language I • Formulate research problem • Collect relevant data • Analyze research problem • Apply the appropriate statistical programming language to address the problem.	Homework Project: model data using a common data modelling technique for research project
15-16	Presentations of data analysis • Describe the importance of presenting data via an oral presentation • Present a data analysis project as an oral presentation.	Homework Project: Present an in-class data analysis project, using both oral and visual techniques.
17	• Final exam	