**STA 4102**

**RI: Computational Statistics**

Instructor: Dr. Lianfen Qian

Office: SE 244

Office Hours: TBA

Email: Lqian@fau.edu

**Catalog description**: The course covers computer algorithms for visualization, evaluation, simulation, random number generation, sampling from prescribed distributions. Simulations, graphics for data display, computation of probabilities and percentiles, hypothesis testing, simple linear regression and multiple regression are covered.  This is a research intensive course.

**Textbook:**  *Statistical Computing with R* by Maria L. Rizzom Chapman & Hall/CRC, 2007.

**Course description:** With the increasing power of computers and the modern developments in statistical computing, statisticians are able to handle far more complex problems today than a few years ago. Statistical computing enhances one’s understanding of statistical theories and ability to handle complex problems and data mining. This course is an exposition of statistical methodology that focuses on ideas and concepts, and makes extensive use of graphical presentation. It covers the following topics: Computer algorithms for evaluation, simulation and visualization, random number generation, sampling from prescribed distributions. Introduction to formal inference, Simple and multiple linear regression, non-linear regression, Generalized linear models, Multi-level models, Time series and repeated measures, Classification and regression tree (CART), Multivariate data exploration and discrimination.

This course address undergraduate research and inquiry components on (1) knowledge, (2) formulating questions, (3) plan of action, (4) critical thinking, (5) ethical conduct, and (6) communication.

**Course objectives**: Students who successfully completes the course will grasp the coverage of statistical methodology and graphical presentation. Students will be able to write codes in R and modify computing algorithms. Students will be able to understand and use basic vocabulary, data curation and management, statistical thinking, statistical modeling including model selection skills, and computational skills. Students will have hands-on experience with random number generation, data structure, writing algorithm for simulation to discover knowledge, and reproducible research for real data analysis.

**Research Intensive Designation:** This course contains an assignment or multiple assignments designed to help students conduct research and inquiry at an intensive level.  If this class is selected to participate in the university-wide assessment program, students will be asked to complete a consent form and submit electronically some of their research assignments for review.  Visit the Office of Undergraduate Research and Inquiry (OURI) for additional opportunities and information at <http://www.fau.edu/ouri>.

The URI portion of the course will address all six Student Learning Objectives:

1. **Knowledge**: Common base of knowledge required for effective data preprocessing, data visualization, data learning, model selection and reproducible research. Students will grasp a set of key skills on statistical learning and computing in data analytics. Students will also show knowledge of tools and practical skills needed to preprocessing and managing data from various sources for both structured and unstructured massive data.
2. **Formulation of Questions**: Students are required to develop research statement in which they specifically address their research questions for real data sets or simulation study. Students are expected to formulate their research questions into subject related hypotheses which are clear and concise to the research problem, ready to be tested or answered through statistical modeling and simulation. When appropriate, the students should be able to break down principal problems into smaller solvable sub-problems.
3. **Plan of Action**: Students will create a plan of action for individual term projects of this research intensive course that will encompass the following elements: (i) scope of the study; (ii) literature review; (iii) planning context; (iv) problem statement and research methodology; (v) analysis and report findings; (vi) presenting the results to general audience. The students will develop hypothesis if needed, identify methods of analysis and select appropriate statistical techniques. Using the course timeline as a template, each student is expected to develop her/his own project management plan with specific tasks related to the topic in consideration.
4. **Critical Thinking:** Students will demonstrate critical thinking skills by formulating research questions, applying appropriate selection criteria for real data analysis or simulation study, taking into consideration multiple perspectives, and examining implications and consequences of an action or planning alternative.
5. **Ethical conduct:** All students are required to familiarize themselves with the rules of academic integrity. Student projects involving primary data collection through website will be credited in their term written research projects. Students are required to be loyal to the original data with confidential information removed. A class module will be provided with a discussion of when statistics were used inappropriately to arrive at a faulty conclusion for unethical purposes.
6. **Communication:** Students will be required to write and present their individual project reports professionally. They are required to submit research report (e.g., analysis, findings and recommendations), and develop a webpage uploaded to rpubs.com for instant to communicate research results as outlined in SLO-3. Students are expected to demonstrate knowledge of writing technical report and orally presenting their findings. Advanced visualization techniques are also required for students to incorporate research findings in planning documents and present them through a real data project.

**Reference book:** *Data Analysis and Graphics Using R: An Example-based Approach*, John Maindonald and John Braun. 3rd edition, 2010, Publisher: Cambridge University Press, ISBN: 9780521762939

**Prerequisites:** MAC 2312 and STA 2023 or equivalent with minimum grade C

**Grading policy**:

Biweekly homework/lab assignment: 20%

Midterm Exam 30%

Midterm project: 20%

Final project: 30%

Grading scale: A/A-: 90-100%, B+/B/B-: 80-89%, C+/C: 70-79%, D: 60-69%, F: <59%.

**Course outline:**

Week 1: Introduction to R/Rstudio and Review Probability & Statistics

Week 2: Methods for Generating Random Variables

Week 3: Visualization of Multivariate Data

Week 4: Monte Carlo Integration

Week 5: Monte Carlo Variance Reduction

Week 6: Monte Carlo Methods in Inference

Week 7: Discussion of Midterm Project

Week 8: Bootstrap and Jackknife

Week 9: Permutation Tests

Week 10: Markov Chain Monte Carlo Methods

Week 11: Non-parametric Density Estimation

Week 12: Numerical Methods in R Language

Week 13-14: Discussion and Prepare for Final Project

Week 15: Final Project Presentation

**Description of the RI assignments:**

**Assignment 1 (Knowledge Discovery via Simulation):** In week 4 of the course, students will produce a tech report on a small size simulation study via random number generation to evaluate some statistical theory. Students will publish the computing algorithm to a repository website. Students will submit the first tech report on their findings. This submission will be limited to five page preliminary tech report which will include a description of the scientific problem, statistical thinking and formulation of the problem, a description of the methods and a flow chart of the algorithm used to knowledge discovery via simulation, a summary of the findings and a conclusion.

**Assignment 2 (Knowledge Discovery in Databases)**: In week 8 of the course, students will prepare a tech report and publish the computing algorithm to a repository website. Students submit the second tech report on the data visualization and knowledge extraction, simulation mimic the data structure to verify the reliability of the data analysis for reproducible research. This will be at an intermediate level with room for improvement but strictly follow the guideline of tech reporting. This submission will be limited to eight-page tech report which will include a description of the real data, statistical thinking and formulation of the problem, data visualization, simulation formulation to mimic the setting of the real data and verify the reliability of the data analysis, a summary of the simulation results and reproducible data analysis, and a conclusion.

**Assignment 3 (Reproducible Research Final Project):** In weeks 9-14 of the course, students will write a tech report according to the tech report guideline and publish the computing algorithm to a repository website. Students have opportunity to propose improvements of the methodology or quality of data analysis at the end of the report. Students will present their simulation results with setting motivated by real data and mimic real data. Their data analysis will be reproduced by simulation study. Reproducibility measure will be reported. Students will present their results to classmates, even public audience, and submit the final tech report by the end of the semester. This submission will be limited to 10-15 page tech report/scientific paper which will include a description of the real data, statistical thinking and formulation of the problem for simulation, data visualization and parameter settings, a description of the methods considered and model selection criteria, a summary of the findings with reproducible performance measure, comments on complication and limitations and conclusions. Students are encouraged to present their findings either through a poster presentation or an oral presentation at the FAU undergraduate research symposium held on the Boca Raton Campus each Spring Semester (<http://www.fau.edu/ouri/undergrad_symposium.php>) and submit the final report for possible publication in a journal such as Florida Atlantic University Undergraduate Research Journal (FAURJ).

**Incomplete grades**

A grade of I (incomplete) will only be given under certain conditions and in accordance with the academic policies and regulations put forward in FAU's University Catalog. The student has to show exceptional circumstances why requirements cannot be met. A request for an incomplete grade has to be made in writing with supporting documentation, where appropriate.

**Classroom etiquette policy**

University policy on the use of electronic devices states: “In order to enhance and maintain a productive atmosphere for education, personal communication devices, such as cellular telephones and pagers, are to be disabled in class sessions.”

**Disability policy statement**

In compliance with the Americans with Disabilities Act (ADA), students who require reasonable accommodations due to a disability to properly execute coursework must register with Student Accessibility Service (SAS) ---in Boca Raton, SU 133 (561-297-3880) ; in Davie, LA 203 (954-236-1222); or in Jupiter, SR 110 (561-799-8585)---and follow all SAS procedures.

**Academic integrity**

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty, including cheating and plagiarism, is considered a serious breach of these ethical standards, because it interferes with the University mission to provide a high quality education in which no student enjoys an unfair advantage over any other. Academic dishonesty is also destructive of the University community, which is grounded in a system of mutual trust and places high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. For more information, see <https://www.fau.edu/ctl/4.001_Code_of_Academic_Integrity.pdf>