Syllabus and Basic Information for Regression Analysis–STA4210, Spring 2019

Instructor: Rohit Patra

1 Basic Information

Objective The course is primarily on the linear regression model, for which the main techniques are rooted in the method of least squares. Procedures are motivated by applications. Mathematical results are stated and explained, and occasionally derived. A course in mathematical statistics is helpful but not strictly required. The focus is on carrying out and explaining the methods. After some review of basic statistics, the course proceeds systematically through the simple regression model, the matrix formulation of this model, the multiple regression model, and a number of related tools such as model diagnostic measures, collinearity statistics, and variable selection procedures. Computations will be carried out in the R statistical programming language.

Prerequisite STA 3024 or STA 3032 or STA 4321 or MAS 3114 or MAS 4105.

Class Schedule :Room– **AND 0134**

- Tuesday—**10:40am–12:35pm** (Period 4 and 5)
- Thursday—**10:40am–11:30am** (Period 4)

Website We will use Canvas. You will find the class notes, Homeworks, and Grades on there.

Required Materials

- **Textbook:** Kutner, Nachtsheim, Neter, and Li, Applied Linear Statistical Models, 5th ed., Volume I (Chapters 114, Appendix A) The Data Sets are available online.

- **Scientific calculator** You need one which will compute the mean and standard deviation automatically. You will use it for tests. A graphing calculator is allowed.
- **Statistical Software** We will use the free statistical computing language R; download it in the first week of the semester from http://www.r-project.org. Also download Rstudio from http: www.rstudio.com (Desktop free license).

Instructor Rohit Patra, 221 Griffin-Floyd Hall
Email: rohitpatra@ufl.edu

Office Hours: Monday and Wednesday at 4:30-6

TA **Ma, Mingliang**, Office-TBA
Email: mamingliang@ufl.edu

TA Office Hours:
- Monday 10:30-11:30
- Wednesday 10:30-12:30
- Friday 11:30-12:30

Feel free to email me or the TA with questions if you cannot make the regular office hours.

Grades: The final grade is based on the following composite score:

Homeworks	— 20%
Exam 1: Tuesday, Feb 11, in class	— 25%
Exam 2: Tuesday, March 17, in class	— 25%
Exam 3: Tuesday, April 21, in class	—30%

Homework: There will approximately 6-8 homeworks with approximately 10 questions each. The homeworks will generally be due at 11:59pm a week from the day of assignment. Late submissions will not be accepted under any circumstances. Some but not all of the problems on each homework will be graded. Some homeworks will require you to use R and to produce a written report of a data analysis. No makeup homeworks will be offered. However, homeworks with the lowest **two** grades will be **dropped**.

Please read the file titled "HomeworkInstructions.pdf" in your canvas for more instructions on how to submit homeworks.

Exams: On each exam, there will be some short-answer questions, both multiple-choice and fill-in-the-blank types. Other questions will require a written response; for these questions, it is not sufficient to simply give a numerical or one-word answer. There will be short calculations required; you need a calculator for the exams. On each exam, there will be at least one question which describes a regression scenario, gives R output from the analysis, and asks you to interpret the output.

Attendance Classroom lecture attendance and participation is fully expected, even if not strictly enforced. You are responsible for learning all material presented during lecture, and any topic covered in lecture is a potential exam topic (unless otherwise stated). You should not use cell phones in class.

Integrity Please familiarize yourself with the Student Honor Code and Academic Honesty Guidelines outlined in your University of Florida Student Guide and at http://www.dso.ufl.edu/studentguide/studentrights.php.

Accommodations To request classroom accommodation, please be certain that you have made all necessary arrangements with the Dean of Students Office, and obtain from them documentation to submit to the instructor at the time of your request. A request must be made to the instructor at least 10 days in advance of the date for which the accommodation is requested.

Exam Aid:

- You will need a graphing or scientific calculator for the exam.
- **1st Exam**: You will be allowed a handwritten formula sheet on **1** side of an $A4$ size paper (this is the most common size of paper).
- **2nd Exam**: handwritten formula sheet on **2** sides of $A4$ size paper.
- **3rd Exam**: handwritten formula sheet on **3** sides of $A4$ size paper.

Tentative Grades: A =90 or above, , $A-$ = 86-89, $B+$ = 81-85, B = 71-80, $B-$ = 66-70, $C+$ = 61-65, C = 56-60, $C-$ =51-55, $D+$ = 46-50, D = 41-45, $D-$ = 36-40, E = 35 or below. These numbers are tentative and are subject

to change over the course of the semester. The tests will be hard the final cutoffs are very likely to be lower than these numbers.

Grades: After each exam I will send you an email with your current grade. The grade will take into account all the exams, quizzes, and homeworks up to that point in time. It should be only taken as a rough guide on your current standing in the class. The calculation of your final average will be done outside of Canvas; the formula used by Canvas will not necessarily produce the final average according to the course grading scheme.

Feedback: I have created an assignment on Canvas. There you can submit anonymous feedback. You should be able to submit **as many feedbacks** as you want throughout the semester.

Course objectives:

1. Be able to use `R` software.
2. With two quantitative variables, be able to carry out correlation and simple regression analyses, and to correctly interpret such analyses.
3. Be able to carry out and interpret inference procedures for simple linear regression.
4. Know the simple and multiple linear regression models, and be able to state and explain the standard methods of estimation for these models.
5. Know the multiple linear regression model in its matrix form, including all the common variations on this model (e.g., continuous predictors, categorical predictors, square and interaction terms).
6. Know what is the general F-test; given a particular multiple regression scenario, be able to apply the general F test.
7. Be able to carry out and interpret inference procedures for the various types of multiple regression model, including the quadratic regression model.
8. Know what are the purposes of diagnostic methods in simple and multiple regression; be able to carry out several common diagnostic procedures and interpret them.

9. Know several measures of model performance (R^2, adjusted R^2, and Akaikes Information Criterion), how to compute and interpret them for a multiple regression model.
10. Know what multicollinearity is, why it is an issue in multiple regression, and how to analyze and deal with its presence.