

MS DATA SCIENCE 596: INTERMEDIATE STATISTICAL METHODS 16:960:596:02

FALL 2016, TUESDAY 6:40-9:30 PM, HILL CENTER 552 BUS

1. COURSE INFORMATION

- Instructor: Han Xiao
- Office: Hill Center 451
- Office Hours:
- Email: hxiao@stat.rutgers.edu
- Prerequisite. Undergraduate courses in calculus, linear algebra, probability and mathematical statistics. This course will cover a great deal of materials at a rapid pace and will require some programming skills (R, or other software of your choice, such as SAS). Students who have had difficulty in previous mathematical statistics courses or programming may find that this course requires a considerable amount of time and effort, and should plan accordingly.
- Text 1: *Linear Models with R*, by Julian J. J. Faraway. Chapman & Hall/CRC. 2014, 2ed.
Text 2: *Analysis of Financial Time Series*, by Ruey S. Tsay. Wiley, 2010, 3ed.
- Software: R. Free software available at <http://www.r-project.org/>. If you go to **Manuals** on the left panel of the website, you will find a good introduction *An Introduction to R*. A more advanced reference is *Modern Applied Statistics with S*, by Venables and Ripley. Springer, 2002, 4ed.
- Course website: <http://stat.rutgers.edu/home/hxiao/>
- Course work: eleven homework assignments, midterm and final exams, project.
- Grades: homework (20%), midterm (30%), final (30%), project (20%).

2. SYLLABUS (TENTATIVE)

Week #	Date	Topic	Due
1	TBA	Introduction & Basic concepts in probability and statistics	
2	TBA	Linear regression: estimation	HW1
3	TBA	Linear regression: inference	HW2
4	TBA	Linear regression: diagnostics	HW3
5	TBA	Linear regression: transformation & model selection	HW4
6	TBA	Robust regression	HW5
7	TBA	Nonlinear regression	HW6
8	TBA	Midterm	
9	TBA	Classical decomposition model for time series	HW7 Proposal
10	TBA	State space model	HW8
11	TBA	Linear time series: ARIMA models	HW9
12	TBA	Conditional heteroscedastic models	HW10 Intermediate report
13	TBA	Nonlinear models	HW11
14	TBA	Project presentation	HW12
15	TBA	Final	
16	TBA	Final project report	

3. HOMEWORK

- Homework will be assigned and collected weekly. The lowest grade will be dropped. **So late homework will NOT be accepted. DO NOT COPY from other sources.**
- All homework assignment must be written on standard 8.5 by 11 paper and stapled together. Computer generated output without detailed explanations and remarks will not receive any credit. You may type out your answers, but make sure to use different fonts to distinguish your own words with computer output. Only hard copies are accepted, except under special circumstances. You should also submit the R source code with computing assignments.
- Data analysis is an integral part of the course. The main software package is R. Instructions for using the package will be given and briefly discussed, assuming you have taken the *Regression Analysis* course with R. If you do not have previous exposure to R (or S+), please be aware that you may need to devote considerable time and effort to get started. R is a free software. You may use any other software package of your choice, but no instructions or help will be given from TA or me.

4. PROJECT GUIDELINE

Project is to be carried out by a team of two investigators. Imagine that the upper management is contemplating to disband your team and this is your last chance to show how useful you are, in order to save your job. You can choose to do (but not limited to) one of the following things: (i) finding a strategy that will potentially generate positive returns, (ii) demonstrating your ability to help other team/line of business in generating useful information from data to help their business, (iii) demonstrate your ability to evaluate risk more accurately, including risk calculation and stress testing.

Your **project proposal** needs to include what you plan to do, why it is important to the business, what kind of data you are going to use and a list of possible methodologies you plan to use. Your **intermediate report** needs to include data description, preliminary analysis, the methodologies you are using, and the results you expect to get. The presentation is limited to 10 minutes to your upper management who will decide if you still have a job tomorrow. Your **final report** should contain a report of what you have done, with summaries for a very busy boss who missed your presentation, and with details for an expert consultant who will read all the details of your report and give an expert opinion to the upper management.

Important dates: proposal due on **Mar 03**, intermediate report due on **Apr 07**, final report due on **TBA**. Please submit **hard** copies for the proposal and intermediate report, and the following for the final report **by email**: (i) a **PDF** file of the final report, (ii) data set used for the project, which can be directly read by R, and (iii) R code (please make sure that by running the code, all the results in the report can be reproduced).