Math 439
Linear Statistical Models
Lectures: MWF, 11:00-12:00, Mallinckrodt 302
Washington University in St. Louis, Fall 2017

Instructor: José E. Figueroa-López

- Office: Cupples I 203A, Phone: (314)935-7539.
- Office hours: MW 3:10 - 4:00 p.m., F 4:10-5:00, or by appointment.
- e-mail: figueroa@math.wustl.edu (Put "MATH439" as the subject when you email me).

Course description

This course teaches the theory and practice of linear regression models, analysis of variance (ANOVA) and their extensions, including testing, estimation, confidence interval procedures, modeling, regression diagnostics and plots, polynomial regression, colinearity and confounding, model selection, geometry of least squares. The use of the computer (in R) to analyze data will also be emphasized throughout the semester. We will cover most of the material corresponding to the first 10 chapters of the required text, along with some supplementary material where needed.

References

- Introduction to Linear Regression Analysis, by Montgomery, D.C., E.A. Peck and G.G. Vining, 5th Edition, Wiley, 2012 (**Required**)
- Linear models with R, by J. Faraway. (**Required**) *Our library has free online access.*
- Applied Linear Statistical Models, by Kutner, M.H., C.J. Nachtsheim, J. Neter, and W. Li. McGraw-Hill, 2004, 5th edition. (**Recommended**)
- Linear Algebra and Linear Models, by R. Bapat. (**Recommended**) *Our library has free online access.*

Mandatory Prerequisites:

Math 3200 and a course in linear algebra (such as Math 309 or 429), or permission of instructor. Some acquaintance with fundamentals of computer programming (such as CSE 131 or CSE 200). Familiarity with R is useful but not essential.
(See below for more information about the topics that students must be familiar with before taking this class).

Course website:

All homework assignments, handouts, and other information will be available on **Blackboard** (http://bb.wustl.edu/).

Exams:

- **Two midterm exams** and **a comprehensive final** will test your grasp of the material covered in class. The dates for the midterm exams are (tentatively) set as Wednesdays **Oct. 11** and **Nov. 15**. The final is scheduled for **Dec. 19**, 10:30 - 12:30.
- All exams are close-notes and close-book. A one-sided letter-sized page with **only formulas/equations** may be brought to each midterm. Two one-sided pages with formulas can be brought to the final. These notes would be collected (make copies).

- Exam questions can be done by hand with the help of a basic scientific (non-graphing) calculator.
- You should always bring your Washington University Photo ID to exams. Proctors will check student's IDs.
- Make-up exams are strongly discouraged. If you are aware of a conflict, please inform the instructor before the exam.

Homework

- There will be about 7 HWs; Only a few selected problems will be graded and counted towards your HW score.
- The lowest HW score will be dropped;
- Written homework should be submitted at the BEGINNING of class. NO LATE HOMEWORK WILL BE ACCEPTED.
- You will receive no credits for solutions with no work or justifications. The grader and instructor reserves the right to deduct points for messy papers.
- While it is acceptable to briefly discuss individual assignments among students, the student's work that is turn in for grading must reflect his/her understanding of the material ("almost" identical solutions will not be tolerated).

Final Project

In a team setting, students will develop a final project. Each team will present the topic in class and write a team report, both of which will be done during the final week of classes. The topic would be chosen by the students in collaboration with the instructor. This could be based on a research paper, an in depth exploration of a topic related to the course, or a thorough statistical analysis of your own data. The topic should be chosen during the first week of November or even earlier.

Tentative grading procedure (The following is tentative; any changes will be announced in advanced.)

- The final letter grade is given according to the following scale:

$[95, 100]$	A+	$[83, 85)$	B+	$[65, 75)$	C
$[87, 95)$	A	$[77, 83)$	B	$[60, 65)$	D
$[85, 87)$	A-	$[75, 77)$	B-	< 60	Fail

- Weights:

Two Midterm Exams	20% each
One comprehensive final exam	30%
Homeworks	20%
Final Project	10%

- If you register for "Pass/Fail" (or "Credit/No Credit"), you must achieve at least 65 to pass, which is the lowest score for a C.

Attendance:

Class attendance is **strongly** recommended. Experience has shown that students who attend class regularly perform better. Lectures will involve discussion of topics and, more importantly, solving examples that may be similar to those appearing in the exams.

Important Regrading Policy:

Students have only one week to request the regrading of an assignment or an exam after the time that this has been returned to the class.

Further Information About the Prerequisites

Again, Math 3200 and a course in linear algebra (such as Math 309 or 429), or permission of instructor. More specifically, this class assumes familiarity with the following topics:

- Calculus of Several Variables (Math 233 or Math 318), such as partial differentiation, gradients, and the method of Lagrange multipliers.
- Matrix operations, matrix inverse and determinants, linear spaces and transformations, ranks, and quadratic forms.
- Gaussian or Normal Distributions and generalities of joint distributions
- Independence, Variance, and Correlation: concepts and main properties
- Law of Large Numbers and Central Limit Theorem
- Generalities of Estimation (Sampling Distribution of Statistics, Bias, Mean-Square Error, etc).
- Generalities of Hypothesis Testing (HT) and Confidence Intervals (CI), including large-sample HT and CI for the population mean and small-sample HT and CI for normal population distributions. This includes some familiarity of t-, chi-squared, and T-distribution.
- Familiarity with basic programming tools (e.g., loops, etc.)
- Familiarity with R is useful, but not essential.

Further Information About Computing

- R is a free, open-source software package/programming language for statistical computing. Students are required to use R to complete all assignments. R is a free software that can be downloaded from **http://cran.r-project.org/**. It works under major operating systems, including Windows, Linux and Mac OS.
- Many students and other R users prefer to have a prettier interface. One popular user interface is RStudio. This is purely for aesthetics; it is not necessary for anything we will do in R. You can download RStudio for free from http://www.rstudio.com/.
- R **Markdown** (from RStudio) is also another interface to work with R. It would allow you embed your R code and the calculations it produces, in ordinary text. It can also contain figures and equations, etc. Usage R Markdown for your assignments is encouraged. If you use it, you need to submit both your "knitted" file (HTML or PDF, not Word), and the original .Rmd file.
- In case, you choose not to use R Markdown for your HW assignments, you need to submit both a readable write-up, in PDF, and a raw text file containing your R code. Please, commented so that it is clear which pieces of code go with which problem.
- **More R Help Resources are provided in the course's website on Blackboard**

I hope you will enjoy this course. Have a nice semester.