

Math439: Linear Statistical Models

Fall 2020

Instructor: Dr. Nan LIN
Office: Room 205, Cupples I
Email: nlin@wustl.edu

Time and location: 10:00am-11:15am (Tue and Thur)

Office hours: 9:40-10:40am Friday, or by appointment

Grader information: Jincheng Pang (jinchengp@wustl.edu)

General information

Textbook 1: *Linear Models with R*, by Julian J. Faraway. CRC Press, 2nd Edition, 2014. ISBN: 1439887330 (Available electronically from the Olin library)

Textbook 2: *Introduction to Linear Regression Analysis*, by D.C. Montgomery, E.A. Peck, and G.G. Vining, 5th Edition, Wiley, 2012. ISBN: 0470542810

References: *Applied Linear Statistical Models*, by Michael H. Kutner, Christopher J. Nachtsheim, John Neter and William Li. McGraw-Hill, 2004, 5th edition. ISBN: 0072386886

Class webpage: All homework assignments, handouts, and other information will be available on Canvas (<http://mycanvas.wustl.edu/>). Students should check the class webpage frequently for updates.

Course Description

This is an applied course in linear statistical methods. Topics cover simple/multiple linear regression, analysis of variance (ANOVA) and analysis of covariance (ANCOVA), including testing, estimation, confidence interval procedures, modeling, regression diagnostics and plots, polynomial regression, collinearity, model selection etc.. The computer software R will be used for the majority of the examples and homework problems, but there will also be some significant exposure to SAS.

Prerequisite

Calculus; linear algebra (Math 309 or 429); some fundamentals of computer programming (CSE 131); probability and statistics at the level of Math493. Knowledge of Math494 and R is strongly recommended. More specifically, this class assumes familiarity with the following topics:

- Calculus
- Matrix algebra, such as vectors, matrices, matrix inverse, determinants, linear transformation, quadratic forms.
- Gaussian distributions, t -, χ^2 -, and F -distributions
- Joint, conditional distributions
- Law of large numbers, central limit theorem
- Variance, covariance, correlation
- Sampling distribution
- Estimation, bias, consistency, mean-squared error

- Hypothesis testing, pvalue, significance level
- Confidence intervals.

Computing

Students are **required** to use R or SAS 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. It is strongly recommended you also install RStudio, an integrated development environment (IDE) for R. RStudio can be freely downloaded from <https://rstudio.com/products/rstudio/download/>. An add-on package called `faraway` needs to be installed to run the sample R programs in the textbook.

SAS is a commercial software widely used in many research fields and industries. Its full version is not free but you may download the free University Edition from https://www.sas.com/en_us/software/university-edition.html. SAS runs only on Windows and Unix/Linux. For Mac users, you can use SAS within a virtual machine on a Mac, where the virtual machine runs Windows or Linux. You may set it up following the instruction on the SAS website.

Homework

Homework will be assigned approximately every other week. Homework solutions need to be turned in electronically on Canvas. When needed, include R code and relevant output (be concise by keeping only the necessary output). No late homework is accepted.

Examinations

There will be one midterm exam and one take-home final exam. The midterm exam is closed book and closed notes and will be held in the regular class time. Its coverage includes materials taught in the first 7 weeks. A one-page (one-sided) 8.5 by 11 inch “cheat sheet” is allowed for the midterm exam. Students can bring a calculator to the exams, but sharing calculators is not allowed. The final exam will be a mix of both theoretical questions and data analysis problems that require using a computer.

Make-up exams will **NOT** be given under any circumstances. Be sure to make yourself available on the exam dates. Do NOT make any travel plan overlapping with the exam date. Taking the exam earlier or later is not an option.

If verifiable documentation is given for a legitimate absence, your final exam grade will be reweighed. If a student misses the midterm exam, her/his final exam will count 50% of the final grade, instead of the usual 30%. However, no reweighing will be given if the final exam is missed, and students will receive a course grade of “Incomplete”.

Exam time:

- Midterm: Nov 5
- Final: TBD

Grading

Grades will be based on attendance (10%), homework sets (40%), the midterm (20%), and the final exam (30%). Cr means C or better if you elect “Pass/Fail” or “Credit/No Credit.” 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

Course schedule (tentative and subject to changes)

Week	Topic
1	Introduction to regression analysis and SAS
2	Simple linear regression: ordinary least squares (OLS) estimator
3	Simple linear regression: statistical inference
4	Multiple linear regression: estimation and interpretation
5	Multiple linear regression: basic inference
6	Multiple linear regression: Simultaneous confidence interval and prediction
7	Model adequacy checking and residual analysis
8	Transformation, Midterm on Nov 5
9	Generalized and weighted least squares
10	Model selection: stepwise procedures and criterion-based procedures
11	Collinearity, ridge regression and principal component regression
12	Regularization and Lasso Thanksgiving
13	Indicator variables and ANOVA
14	Multiple comparison and ANCOVA
15	Polynomial regression and linear regression with measurement errors
16	Final exam

Learning Tips

1. Try to show up in all the lectures. Make good notes.
2. Ask questions in class. Your questions may be others' as well. No questions are too elementary, and all deserve to be answered.
3. Discuss with your classmates about your questions.
4. Finish homework in time.

Class Policies

1. Late homework: No credit. Note that students will not be able to submit homework on Canvas after the due time.
2. Exam conflicts: In general, students must attend the midterm unless some verifiable documentation can be provided to justify a legitimate absence, e.g. medical emergency or conflict with attending a research conference. Students are expected to inform the instructor at least one week in advance. If a student misses the midterm due to a legitimate absence, the course grade will be based on the following weights: attendance (10%), homework sets (40%), and the final exam (50%).
3. Collaboration: I encourage discussion of homework in broad, conceptual terms where one student is trying to educate another without giving away the answer, but **all work turned in must be your own**. For example, each student must write his/her own programs in entirety.
4. Academic Integrity: All students are expected to adhere to the university's academic integrity policy (<https://students.wustl.edu/academic-integrity/>). Any student who is found to have cheated on an assignment or exam will receive a zero score for that work, regardless of the extent of the offense.

5. Disclaimer: The instructor reserves the right to make any changes he considers academically advisable. Such changes, if any, will be announced in class. Please note that it is your responsibility to attend the class and keep track of the proceedings.