Math 542: Analysis of Variance and Regression
Spring 2023

Instructor: Dmitrii M. Ostrovskii – dostrovs@usc.edu – KAP 406H
Schedule and Classroom: M/W/F 12-12.50 pm @ CPA 252
Office hours: Monday 1-2.30 pm, Friday 2-3.30 pm, or by appointment

Outline of the course. This course gives an introduction to two of the most widely used statistical primitives: linear regression and analysis of variance. Our primary focus will be on theory, but we shall also see some applications in data analysis. If time permits, we shall proceed with a more modern outlook on linear and nonlinear regression in the end of the semester.

Contact. The best way to contact me is by email (please put M542 into the subject field), or in person during the office hours. We can also arrange an appointment, in person or via zoom.

Prerequisites. Working knowledge – at the advanced undergraduate level – of Linear Algebra (Math 471), Probability (Math 407 or 505a), and Mathematical Statistics (as per Math 541a).

Covered topics. Some topics are optional due to limited time; they are marked with "?."

- random vectors and their characteristics;
- the multivariate normal distribution;
- quadratic forms of random vectors;
- moment generating functions;
- the linear regression model;
- estimation of regression coefficients & maximum likelihood estimation;
- estimation with linear constraints;
- orthogonal design matrices;
- generalized least squares;
- hypothesis testing in regression: the F-test and goodness-of-fit testing;
- confidence intervals in regression: Student's t, Bonferroni, maximum modulus t, Scheffé's S-Method;
- prediction intervals;
- polynomial regression;
- ANOVA: one-way layout, two-way layout, random effects models, and invariance; analysis of covariance;
- regularization methods including ridge regression and the LASSO;

? multiple testing and the False Discovery Rate.
? exponential families and generalized linear models (GLMs).

As a general remark, please note that *this syllabus is not a legally binding document.*

Homework assignments. There will be 4-5 homework assignments: "theoretical" problems (mostly assigned from Seber & Lee) and data analysis assignments. For the latter ones, you can use your favorite language, such as Matlab, Python, or R, but I do not recommend using C, C++, or other low-level languages. Some advice:

- Always start to work on homework assignments as early as possible. (Well, at least *look* in there, simply in order to estimate the difficulty of what you'll have to deal with!)
- It is OK to work in a group. However, always try to solve the problem yourself first – especially those of you who do research or planning to pursue it. If you are blocked, use the office hours to ask a question. Taking notes during a discussion is a useful practice.
- Always write the solution on your own; suspected plagiatrism will receive 0 credit, and further sanctions in case of a relapse. I recommend typesetting solutions in LaTeX; writing by hand is possible, but please make sure that it is readable.

Deadlines. Homework assignments should be turned in by 11:59 pm on the due date. If you cannot make it, please contact me in advance and explain the reason.

Exams. Two midterm exams will take place during the lectures on Wednesday, March 1 and Wednesday, April 12 (possibly Friday, April 14). The (comprehensive, two-hour) **final exam** will start **at 11 am on Friday, May 5** in our regular classroom.

Grading. 25% for homework assignments, each of the two midterms, and the final exam. The final letter grades will be decided based on the distribution of the raw percentage scores.

Attendance. I shall not keep track of attendance – I shall rely on your willingness to learn.

Literature. Our main textbook, and the main source of homework problems, is *Linear Regression Analysis*, 2nd edition, by Seber & Lee. Other useful references are *Introduction to Linear Regression Analysis* by Montgomery et al., and *Applied Regression Analysis* by Draper & Smith; *Handbook of Statistical Analyses in R* by Everitt & Hothorn is a source of R examples. Finally, as a "linear algebra manual" you can use Appendix A in Seber & Lee, as well as the appendix in *Convex Optimization* by Boyd and Vandenberghe; the latter book is available online: `https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf`).

Special Accommodations. Any student requesting academic accommodations based on special needs is required to register with DSP each semester. A letter of verification for approved accommodations can be obtained from DSP. DSP is located in GFS 120 and is open 8:30 a.m. until 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Academic Integrity. USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. The Student Guidebook contains the Student Conduct Code in Section 11.00, while the recommended sanctions are located in Appendix A: `http://www.usc.edu/dept/publications/SCAMPUS/gov/`. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: `http://www.usc.edu/student-affairs/SJACS/`.