

STAT 5160 Advanced Statistical Inference II

Section 001

Instructor: Wei Wang wwa@upenn.edu

Class hours: Tuesday and Thursday, 10:15 am - 11:45 am (EST).

Office hours: After class, email or by appointment.

Course description

The course covers both the theory and the application of frequently used regression models. Topics include multivariate normal distribution, multiple linear regression, generalized linear models, EM algorithm, finite mixture models, and linear and generalized linear mixed effects models. If time permits, non/semi-parametric techniques will also be studied.

References

Introduction to Linear Regression Analysis, DC. Montgomery, EA Peck, GG Vining.

Generalized Linear Models, P. McCullagh, J. Nelder.

The EM Algorithm and Extensions, G. J. McLachlan, T. Krishnan.

Finite Mixture Models, G. McLachlan, D. Peel.

Mixed-Effects Models in S and S-PLUS, J. C. Pinheiro and D. M. Bates.

Generalized, Linear and Mixed Models, C. E. McCulloch, S. R. Searle.

Prerequisite: STAT 5150

Undergraduate level probability and statistics: conditional expectation/probability/variance, exponential family, maximum likelihood estimation, Fisher information, point estimation, hypothesis testing, and confidence intervals.

Calculus: continuity, differentiation, integration, Taylor expansion, etc.

Linear algebra: vectors, matrices, matrix multiplication, matrix determinant and inverse, Jacobian matrix and Hessian matrix.

R programming experience: scatter plots, histograms, boxplots, simple simulation, and data summaries.

Software

We will use the free statistical computing software R (<http://www.r-project.org/>) frequently in class. Sample R code will be provided to help you solve homework problems.

Homework

The homework will be assigned biweekly. Without a convincing reason, late homework will not be given full credit (25 points off every 24 hours). If you are not certain about your situation, ask the instructor in advance instead of a last minute request.

Grading

The final grade will be based ~60% on homework, and 40% on the take home final exam.

Notes

- Pdf format is preferred, and html format is not recommended.
- If R is used, attach both the code and the output. Otherwise, 50% points will be deducted. Irrelevant R output will be penalized.
- The solution must be in your own words. Identical solutions will be marked zero.
- Try to provide some explanation of your answer instead of a simple yes or no.
- You are not allowed to use generative AI (e.g, tools like ChatGPT) for your work. Using such tools in this course will be considered a violation of Penn's Code of Academic Integrity and will be reported to the Center for Community Standards and Accountability.
- Regular classroom attendance and participation is anticipated. If you are not able to attend some lectures, you can request to view the recorded videos. The request will be approved, within reason.
- Independent work is expected for the final exam. Discussion or communication with other people is not allowed. Otherwise, it will be considered cheating, and the exam paper will be marked 0.