

Fall 2020, Stat 579: Statistical Analysis With Missing Data

Instructor: Dr. Yan Lu, yanlu@unm.edu

Class website: <https://learn.unm.edu>

Class Time/Place: TTr 12:30am-1:45pm by zoom

Instructor Office hours: TTr 1:45pm-2:30pm by zoom, or by appointment

Prerequisites: To be enrolled in this class, you need to know basic calculus (integrals, derivatives and Taylor expansion); Basic probability theory (density functions, conditional probabilities, conditional independence, expectations, basic statistical distributions, multivariate normal etc); Matrix notations and manipulations; Basic statistics theory (random sample, point estimator such as maximum likelihood estimator, interval estimation etc); Linear models; and reasonable R programming skills.

Ideally, you want to have the first year statistics courses Stat 440/540 Regression, Stat 445/545 Anova, Stat 461/561 Probability and Stat 453/553 Statistical Inference ready.

Textbook: No textbook required

Main resources: lectures notes developed by Marie Davidian and Anastasios A. Tsiatis at the North Carolina State University. The lecture notes are available at: <https://www4.stat.ncsu.edu/~davidian/st790/notes.html>

Dr. Davidian has graciously granted the instructor permission to use her course materials. Please feel free to let me know if you find any typo, mistake or something to improve. I will let Dr. Davidian know.

Reference:

- I incorporate slides from Dr. Mauricio Sadinle (University of Washington) for teaching, as we both use lecture notes from Dr. Davidian.
- Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. Hoboken, New Jersey: Wiley.
- Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. London: Chapman & Hall.

- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis With Missing Data*. New York: Wiley.
- Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, Florida: Chapman & Hall.

Topics:

Missing data are in almost every area, particularly in research involving human subjects such as in survey and health sciences. This course introduces methodologies for dealing with missing data in statistical analyses. For missing at random (MAR) cases, we introduce naive imputation methods, likelihood-based approaches, Bayesian and multiple imputation approaches, and inverse-probability weighting methods. For nonignorable missingness (MNAR), we introduce pattern-mixture models, and sensitivity analysis. Computational tools such as the Expectation-Maximization algorithm and the Gibbs sampler will be introduced. This course is intended for students who are interested in methodological research and applications.

Computing: R will be used in this class.

R:

<http://cran.r-project.org>

Rstudio:

<https://www.rstudio.com/products/rstudio/download>

To see how to install R and Rstudio in windows, visit

<https://www.youtube.com/watch?v=eD07NznguA4>

for Mac

<https://www.youtube.com/watch?v=GFImMj11MRI>

Grading: Homework, 100% (bi-weekly homework)

Stat 579

A 90%-100%

B 80%-89%

C 65%-79%

D under 64%