

STAT 5010/PSYC 6120
Introduction to Nonparametrics & Log-Linear Models
Section 401 Spring 2023

Instructor: Wei Wang wei.wang@pennmedicine.upenn.edu

Class hours: Tuesday and Thursday, 12:00 pm - 1:30 pm (EST).

Office hours: After class, email or by appointment.

Course description:

The course covers commonly used nonparametric (semiparametric) statistical methods. Topics include the Wilcoxon rank sum test, signed rank test, the Kruskal-Wallis test, two sample tests on proportions, smoothing methods (kernel smoothing and spline smoothing), and regression models such as generalized additive models.

References:

Nonparametric Statistical Methods, M. Hollander, D. A. Wolfe, and E. Chicken.

An Introduction to Categorical Data Analysis, A. Agresti.

Nonparametric Statistical Methods Using R, J. Kloeke and J. W. McKean.

Course Prerequisite: STAT 5000.

Required Background:

Undergraduate level probability and statistics: conditional probability, random variables, distributions, sample mean and sample variance, convergence in distribution, central limit theorem, point estimation, hypothesis testing, confidence intervals, and multiple linear regression.

Calculus: limit, supremum, infimum, continuity, differentiation, integration, Taylor expansion, etc.

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

R programming experience: scatter plots, histograms, and data management skills.

Software: We will use the free statistical computing software R (<http://www.r-project.org/>) frequently in class. You will apply what you learn in class to solve your homework and final exam problems.

Homework: There will be biweekly homework assignments. Without a convincing reason, late homework will not be given full credit (25 points off every 24 hours). If you are not sure about your situation, ask the instructor in advance. No last minute notice.

Exam: There will be a take home final exam.

Format:

You are strongly advised to type your homework or exam solutions. Try your best to copy and paste everything including the relevant R code and output into a single file.

Grading: The final grade will be based 60% on homework, and 40% on the final exam. The solution has to be in your own words. No plagiarism or cheating. Identical solutions will be marked zero. Points up to 50% will be deducted if no R code or R output. Independent work is expected for the final exam. No discussion or communication with other people. Otherwise, it will be considered cheating, and the exam paper will be marked 0. Regular classroom attendance and participation is required.