

STAT 5010/PSYC 6120 (section 401)

Introduction to Nonparametrics & Log-Linear Models

Instructor: Wei Wang wwa@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 techniques and various regression models. 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), generalized linear models and 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, central limit theorem, point estimation, hypothesis testing, p-value, confidence intervals, and multiple linear regression.

Calculus: continuity, differentiation, and integration.

Linear algebra: vectors, matrices, and matrix multiplication.

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. Sample R code will be provided to help you solve homework problems.

Homework

There will be biweekly 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 instead of a last minute notice.

Late Homework Policy

Timely submission by the due date and time: no penalty

Late >0 to 24h: 25% points off

Late >24 to 48h: 50% points off

Late >48 to 72h: 75% points off

Late >72h: no credit

Grading

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

Notes

- Homework should be submitted via Canvas in one file. Pdf or Word format is preferred. Html format is not recommended.
- If software is used, attach both the code and the output. Otherwise, 50% points will be deducted. Irrelevant output will be penalized.
 - If you use R markdown, you can include the Rmd file.
- Students may work together on homework assignments, but the solution must be in your own words. In cases where responses from different students seem extraordinarily similar, the students involved will be invited to explain the situation. Unsatisfactory explanations will result in zero credit for the homework for all students involved.
- In your work, try to provide some explanation (no need to be in great detail) instead of a simple yes or no. Only writing the result of a calculation is not sufficient.
- You may not use generative AI for your work in this class. The work you produce should be your own, free from any AI assistance. Submitting work that contains AI-generated content will be considered a violation of Penn's Code of Academic Integrity, and suspected use will be referred to the Center for Community Standards & Accountability (CSA).
- Regular classroom attendance and participation is anticipated. If you attended the class and would like to review the lecture, submit the request through Canvas to view the recorded videos. If you are not able to attend a class, submit the request. It will be approved, within reason.

Topics

1. Wilcoxon rank sum test
2. Kolmogorov-Sminov test
3. Sign test
4. Wilcoxon signed rank test
5. Ansari-Bradley test
6. Kruskal-Wallis test
7. Friedman rank test
8. Kendall Correlation Coefficient
9. Spearman Correlation Coefficient
10. Cohen's Kappa
11. Chi-squared test
12. Fisher's test
13. Mantel-Haenszel test
14. McNemar's test
15. Stuart-Maxwell test
16. Linear regression models
17. Logistic regression
18. Poisson regression
19. Generalized linear models
20. Semiparametric regression
21. B-splines
22. Generalized additive models
23. Density estimation
24. Nonparametric regression