

Nonparametric Statistics

University of Kentucky

STA 621, Fall 2012

Credit: 3.0

Lecture: 12:30 p.m - 1:45 p.m., Tuesdays and Thursdays
Room 335, Multidisciplinary Sciences Building (MDS 335)

Instructor: Patrick Breheny, Ph.D.

Office: Room 205D, Multidisciplinary Sciences Building

Phone: 218-2077

e-mail: patrick.breheny@uky.edu

Office Hours: Whenever I'm in my office, or by appointment

Course description: This course provides an overview of nonparametric statistics. “Nonparametric” can be defined quite broadly, and this course aims to acquaint its students with as many of those areas as possible. “Nonparametric” means something rather different now than it has in the past; this course will take a modern view of nonparametric statistics, although we will spend some time on traditional rank-based nonparametrics.

My goal is to introduce students to the wide range of interesting nonparametric ideas in statistics. Some of those ideas are theoretical, others are computational and methodological. Theory will be introduced when it is interesting and relevant, although some proofs may be skipped in the interest of time. My objectives are:

- To broaden your mind in terms of thinking about data and inference in a nonparametric way
- To introduce useful nonparametric methods that you can apply while analyzing data
- To familiarize you with important topics in nonparametric statistics that you may wish to research
- To bring you up to speed concerning terminology and concepts in nonparametric statistics so that you can more easily read research articles in the field

Text:

- Wasserman, L. (2006): *All of Nonparametric Statistics*. Springer.

Prerequisite: STA 606.

Course website: The course notes, assignments, data sets, and other relevant materials will be made available on the course web site: <http://web.as.uky.edu/statistics/users/pbreheny/621/F12/index.html>

Grading: Your grade will be based on six projects, each worth $\approx 1/6$ of your grade. Each project will consist of three sections: (1) Mathematical concepts and derivations (2) Simulation study (3) Analysis of real data.

The grading scale will be as follows:

85-100	A
70-84	B
55-69	C
Below 54	E

These projects comprise your entire grade; consequently, they are longer and more involved than a typical homework assignment and you should expect to spend more time on them. I will distribute the projects well in advance of their due date and let you know when we have covered sufficient material to start on certain problems, as I would not advise waiting until the last minute to work on the entire project.

Assignments may be turned in either electronically or as a physical copy. However, if they are turned in electronically, they must be turned in as a `.pdf` file (*i.e.*, no Word documents). Each assignment will also involve writing code (to analyze data, run simulations, make plots, etc.). Please turn in this code separately and electronically.

All electronic submissions should be made via DROPitTOme at <http://www.dropitto.me/pbreheny> and should follow the following naming convention: last name, assignment number, proper extension. So, for example, if Adam Smith is turning in assignment 1, he would name the file `Smith1.pdf`. The associated code would be `Smith1.R`. If you wish to break up your code into separate files, you may submit them as `Smith1a.R`, `Smith1b.R`, and so on.

All assignments may be resubmitted for partial credit for the points lost on the original submission. The final grade will then consist of a weighted average of the original submission and the resubmission, with the resubmission receiving $1/3$ of the weight. So for example, if Adam Smith got a 51/60 on the original submission, but a 60/60 on the resubmission, his final grade for the assignment would be 54/60. There is no deadline for the resubmissions, nor is there a limit to the number of times you may resubmit an assignment.

Computing: The simulation study and real data analysis portion of each project will involve a computer. I assume that you are familiar with the basics of R programming; I will introduce and demonstrate helpful additional functions and code during class.

Proofreading: Despite my best efforts, my notes occasionally have typographical errors. If you see them, please tell me about them! Doing so will benefit not only you, but also your classmates and any future students of this course. I will announce corrections on the course

webpage, or if the mistake is critical, send out an e-mail.

Electronic communication: I will occasionally send e-mails to the class through e-mail (to the account listed for you in the campus directory), so please check that account regularly.

Academic honesty: Academic honesty is highly valued at the University of Kentucky. You must always submit work that represents your original words or ideas. You may discuss the assignments and your solutions with other students, but your writeup must be your own. Specifically, any copying of mathematical solutions, or copy/paste operations on typed documents or code are strictly forbidden.

Complaints: Students with suggestions or complaints should see me first, and if we cannot come to an agreement, I will direct you to the head of the department.

Disabilities: If anyone has a disability requiring special accommodations, please let me know as soon as possible, so that these arrangements can be made.

Religious observances: If a religious observance prevents you from taking an exam or finishing an assignment or project, please let me know in advance so that we can make arrangements for you to make up the work.

Inclement weather: The University of Kentucky has a detailed policy for decisions to close in inclement weather. The snow policy is described in detail at <http://www.uky.edu/MicroLabs/documents/temp/policies-weather.htm> or you can call (859) 257-5684.

I look forward to getting to know you, and I hope that we have a great semester together!

Course outline:

1. The empirical distribution function
 - (a) Empirical distribution functions
 - (b) Statistical functionals
 - (c) Influence functions
 - (d) The jackknife
 - (e) The bootstrap
 - (f) Bootstrap confidence intervals
 - (g) Empirical likelihood
2. Hypothesis testing
 - (a) Permutation tests
 - (b) Rank tests
 - i. Invariance and optimality
 - ii. Linear rank statistics
 - iii. Power and relative efficiency
 - (c) Bootstrap tests
3. Density estimation
 - (a) The bias-variance tradeoff
 - (b) Cross-validation
 - (c) Kernels
 - (d) Kernel density estimation
 - (e) The curse of dimensionality
 - (f) Kernel density classification
4. Nonparametric regression
 - (a) Local regression
 - (b) Basis expansions, splines, and penalized regression
 - (c) Quantile regression
 - (d) Nonparametric approaches to multiple regression
 - i. Additive models
 - ii. Thin-plate splines
 - iii. Regression trees