Nonparametric Statistics

Chunlin Wang ©

Department of Statistics, School of Economics and
Wang Yanan Institute for Studies in Economics
Xiamen University

Fall 2021

General Information

- **Instructor information:**
  - **Instructor:** Chunlin Wang
  - **E-mail**: wangc@xmu.edu.cn
  - **Office**: Economics Building A121
  - **Office hours**: Tuesday 15:00-17:00, or by e-mail appointment.
- **Teaching assistant (TA) information**:
  - **TA**: Xinyu Wang
- **Class schedule**: Wednesday 10:10-11:50 & Friday 14:30-16:10 in Jia Geng 4 - 206

Course Description

- Nonparametric statistics, as the opposite to parametric statistics, is an important branch of Statistics.
- In parametric statistics, the data are assumed to come from a specific family of distributions/models characterized by a finite # of parameters, e.g. normal distribution, linear regression model.
- Nonparametric methods differ in that the form of data distribution/model structure is not specified ("distribution-free"), and hence they are more flexible and robust in a wider range of applications.
- There is a close connection between nonparametric regression and modern statistical machine learning.
- This course serves the first introduction to nonparametric statistics, which would be very useful for your thesis research, future studies and applications.

Course Description

- The course materials include: my lecture notes, sample R codes and data sets. The midterm and final exams will not exceed the coverage of the lecture notes.
- In general, there is no textbook required in this course.

  Reference textbooks:
  – *"Introduction to Modern Nonparametric Statistics"*, by James J. Higgins, 2004.
  – *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"*, by T. Hastie, R. Tibshirani, J. H. Friedman, 2009.
- We will use R as the programming language in this course. R is free and can be downloaded from:
  `http://www.r-project.org/`

Course Materials

- I will use SPOC (`https://l.xmu.edu.cn/`) to upload course materials (class notes, assignments, solutions to assignments and supplements) and make some announcement.
- Please join in the course QQ group. TA will also upload course materials in the QQ group. You may use it for course-related discussions.
- Please check this course webpage, QQ group and your email account regularly to keep yourself up-to-date.

PREREQUISITES

- Students are responsible to have good prior knowledge in probability theory, mathematical statistics and regression analysis.
  - We will review some basic knowledge in class, but the students MUST have taken a whole semester of the above courses before enrolling in this course.
  - Statistical programming language `R` will be used in lectures. Although I will spend some time on how to start using `R`, you are encouraged to get familiar with `R` after classes.

GRADING SCHEME

- **Grading policy**

Pop quiz and attendance	Homework assignments	Midterm exam	Final exam	Project
5%	15%	25%	25%	30%

GRADING SCHEME (CONTINUED)

- **Pop quiz and attendance (5%)**: Quizzes may be given randomly in class to check the understanding of the lectures and the attendance.
- Students should attend every lecture. If you miss more than 3 classes in random check, or 6 classes in successive check, without approval for a valid reason*, you will directly fail this course.
- If you miss over 1/3 of the total classes, even if you have been approved for valid reasons*, you will directly fail this course.

GRADING SCHEME (CONTINUED)

- **Homework assignments (15%)**: There will be 4-5 homework assignments. The due dates will be posted later. Please submit in class before the lecture.
  - Late homework is NOT acceptable, unless a valid reason* is reported to the instructor by email at least 2 DAYS before the due day.
  - If you are approved to submit a late homework assignment, the length of extension will be determined case by case.
  - If you are not satisfied with the way a homework/test being marked, you can submit it to me for re-marking but only within 7 DAYS from the time that a particular homework was first made available for pick-up.

GRADING SCHEME (CONTINUED)

- **Midterm and final exams (25% + 25%)**: We will have two exams and both are closed-book written exams.
  - The midterm exam will be scheduled in around the 7th or 8th teaching week.
  - The specific times and locations of both exams are to be announced later.
  - If you have to miss an exam, a valid reason* should be reported to the instructor at least 5 DAYS before the exam date.
  - For anyone absent from an exam without approval, you will receive a grade of zero on that exam.

GRADING SCHEME (CONTINUED)

- **Course project (30%)**: You need to finish an individual course project.
  - Topics of the course project are not pre-assigned, as long as they are related to what you have learned from this course.
  - Oral presentation (10 mins) may be scheduled in the end of semester.
  - The topic, content, writing and oral presentation of a project are equally important.
- The course project is crucial and I expect it to be the initial version of your thesis.
- The thesis writing class in the next term will further help you polish and finish your thesis.
- More details on project will be discussed later.

TENTATIVE TEACHING SCHEDULE

Week	Chapter	Contents
1	1	Introduction to Nonparametric Statistics
2,3	2	One-sample methods
4,5,6	3	Two-sample methods
7,8	4	K-sample methods
9	5	Paired comparisons and blocked design
10	6	Tests for trends and association
11	7	Nonparametric bootstrap method
12,13,14	8	Introduction to nonparametric regression
-	-	Presentations and final review($\star$)

($\star$: depending on time availability)

GOALS OF THIS COURSE

- In general, this course can be split into two parts: (i) classical nonparametric inferences based on ranks together with resampling methods, and (ii) modern nonparametric regression techniques.
- You are expected to gain solid knowledge in following topics:
- One-sample methods:
  - Inference for the median
  - Inference on the population cdf and percentiles
  - A comparison of statistical tests

GOALS OF THIS COURSE (CONTINUED)

- Two-sample methods:
  - The permutation test
  - Wilcoxon rank-sum test
  - Mann-Whitney test
  - A confidence interval for a shift parameter
  - Scoring systems
  - Tests for equality of scale parameters
  - Selecting among two-sample tests
  - Large-sample approximations
- *K*-sample methods:
  - *K*-sample permutation tests
  - Kruskal-Wallis test
  - Multiple comparisons

GOALS OF THIS COURSE (CONTINUED)

- Two-sample methods:
  - The permutation test
  - Wilcoxon rank-sum test
  - Mann-Whitney test
  - A confidence interval for a shift parameter
  - Scoring systems
  - Tests for equality of scale parameters
  - Selecting among two-sample tests
  - Large-sample approximations
- K-sample methods:
  - K-sample permutation tests
  - Kruskal-Wallis test
  - Multiple comparisons

GOALS OF THIS COURSE (CONTINUED)

- Paired comparisons and blocked design:
  - Permutation tests for paired comparison
  - Wilcoxon signed-rank test
- Tests for trends and association:
  - Permutation test for correlation and slope
  - Spearman rank correlation
  - Kendall's tau
- Nonparametric bootstrap method:
  - Bootstrap variance estimation
  - Bootstrap estimation of sampling distribution
  - Bootstrap confidence intervals
- Introduction to nonparametric regression:
  - Kernel method
  - Spline method
  - Generalized additive models

GOALS OF THIS COURSE (CONTINUED)

- Paired comparisons and blocked design:
  - Permutation tests for paired comparison
  - Wilcoxon signed-rank test
- Tests for trends and association:
  - Permutation test for correlation and slope
  - Spearman rank correlation
  - Kendall's tau
- Nonparametric bootstrap method:
  - Bootstrap variance estimation
  - Bootstrap estimation of sampling distribution
  - Bootstrap confidence intervals
- Introduction to nonparametric regression:
  - Kernel method
  - Spline method
  - Generalized additive models

GOALS OF THIS COURSE (CONTINUED)

- Paired comparisons and blocked design:
  - Permutation tests for paired comparison
  - Wilcoxon signed-rank test
- Tests for trends and association:
  - Permutation test for correlation and slope
  - Spearman rank correlation
  - Kendall's tau
- Nonparametric bootstrap method:
  - Bootstrap variance estimation
  - Bootstrap estimation of sampling distribution
  - Bootstrap confidence intervals
- Introduction to nonparametric regression:
  - Kernel method
  - Spline method
  - Generalized additive models

GOALS OF THIS COURSE (CONTINUED)

- Paired comparisons and blocked design:
  - Permutation tests for paired comparison
  - Wilcoxon signed-rank test
- Tests for trends and association:
  - Permutation test for correlation and slope
  - Spearman rank correlation
  - Kendall's tau
- Nonparametric bootstrap method:
  - Bootstrap variance estimation
  - Bootstrap estimation of sampling distribution
  - Bootstrap confidence intervals
- Introduction to nonparametric regression:
  - Kernel method
  - Spline method
  - Generalized additive models

POLICIES

- This course strictly follows the teaching policies and regulations of Xiamen University, School of Economics and WISE.
- Any type of academic dishonesty and cheating activity is ZERO tolerance in this course, including but not limited to:
  - copy answers for homework assignments, project & exams etc;
  - use others ID information to attend and write a test or exam;
  - talking or using electronic devices in a test or exam.
  - Once a cheating activity is identified, those involved will receive a grade of zero on that assessment, minimum.
  - Anyone suspected of academic dishonesty will be reported to the Undergraduate Affairs Office immediately.

POLICIES (CONTINUED)

- Only legitimate reason, such as, illness, emergency, or taking a language examination is acceptable for the application for leave. Valid approval note and relevant supporting documentary proof(s) should be provided to the instructor.
- Taking pictures and recording videos are strictly prohibited during the lectures, unless with my permission.
- All the course materials are for class use only. Please do not circulate outside this class. All copy rights are reserved.
- Please turn all electronic devices into silence before the start of lectures; do not disturb others.
- For more information, please refer to
  - `http://jwc.xmu.edu.cn/`
  - `http://soe.xmu.edu.cn/programs/ba/`
  - `http://wise.xmu.edu.cn/iuec/`

Some Suggestions

- Attend every lectures (with your mind active).
- Take some notes in class, although my course materials will be made available for you.
- Start working on your course project as early as possible.
- Put your effort on the assignments and exams (practice, practice and practice).
- Write down your answers clearly and easy to follow and read.
- Do not expect to finish the assignments in one day.
- Take advantage of my office hours.

Additional Remarks

- The best way to reach me outside of class and office hours is by e-mail.
- Your feedback on any aspects about this course is always welcome and very appreciated!

QUESTIONS? COMMENTS?

READY?

LET'S GET STARTED!