Syllabus

Feature	Considerations
Instructor Information	• Course Co-Directors: Peng Wei, PhD; pwei2@mdanderson.org; Yisheng Li, PhD; ysli@mdanderson.org; Jian Wang, PhD; jianwang@mdanderson.org; • TA: Ms. Ziqiao Wang; ZWang21@mdanderson.org
	• Office Hours: by appointment • TA: Wednesday 4pm-5pm
	• Course number and title: GS01 1273 Modern Nonparametrics
	• Semester: Spring 2021
	• Credit Hours: 3
	• Format: Face-to-face
	• When: Mondays 1:00pm-4:00pm • Where: Zoom
	This course seeks to introduce students to the many developments in modern nonparametrics, including resampling methods, nonparametric and semiparametric regression models that have occurred over the last several decades. Topics include the bootstrap, jackknife, cross-validation, permutation tests, classification tree, random forests, nonparametric smoothing and regression, spline regression, and functional data analysis. While the course will focus on applications, time will be devoted to derivations and theoretical justifications of methods. The statistical software R will be used for the homework exercises.
	This course is intended primarily for second-year M.S. and Ph.D. biostatistics/quantitative sciences students, but is open to other students with prerequisites.
Textbook and Materials	• Required Textbooks: 1. Efron, B. and Tibshirani, R. (1994) An introduction to the bootstrap. Chapman & Hall: New York. 2. Ruppert, D., Wand, M.P., Carroll, C.J. (2003) Semiparametric regression. Cambridge University Press. 3. Hastie, T., Tibshirani, R., and and Friedman, J. (2009) The Elements of Statistical Learning. 2nd Edition. Springer-Verlag. • Lecture notes

	•
Course Expectations	Students are expected to participate in-class activities, complete homework assignments on time, apply the methods covered in the course to solve a quantitative biomedical problem, and summarize and present the results as a final course project.
Course Learning Objectives	• At the end of the course, students should be able to: 1. Apply the bootstrap properly to standard error estimation, confidence interval and hypothesis testing 2. Apply the permutation tests properly to independent and dependent data 3. Choose and apply appropriate nonparametric regression techniques to biomedical problems 4. Report and present results from the application of resampling and nonparametric regression techniques
List of Topics	• Topics include the bootstrap, jackknife, cross-validation, permutation tests, CART, random forest, nonparametric smoothing and regression, and splines.
Learning Activities	• Lectures, homework assignments including computer simulations and data analyses, a midterm exam and a final project
	• The statistical software R will be used extensively for in-class examples, homework assignments and the final project.
Student Assessment And Grading Criteria	• The course will be letter graded. Evaluation will be based on class participation, homework assignments, a midterm and a final project (a written report and a 20-min presentation).
	• The weights of all components are: homework (25%), midterm (30%), final project (45%).
	• Grading Criteria: A: >= 80%, B: <80% and >=70%, C: <70% and >=60%, F: <60%
	• There will be bi-weekly homework assignments. You can work with other students on the homework assignments; however, all turned in work should be your own. There will be a 20% penalty per day for late homework unless you provide a written medical excuse for not turning in homework on time.
Prerequisites and/or Technical Requirements	• PH1910 and PH1911 (UT-SPH) Theory of Biostatistics I & II (or equivalent) and Linear Regression or Consent of Instructor.
	• Majority of the homework assignments will require R programming. At least working knowledge of R is required. http://www.r-project.org/
	•
	•

	•
Policies and Procedures	• Homework assignments allow limited discussions; however, all turned in work should be your own
	• Withdrawal Information: will follow the GSBS general guidelines
	• Incomplete Grades - Students must have extenuating circumstances to request an incomplete. - Incompletes must be made-up by the end of the following semester or the grade will become a fail.
	• Classroom attendance is expected in person. Part of the grade is based on classroom participation.
	• Academic Dishonesty policies: -All as in HOOP are followed - Plagiarism is unacceptable and will result in an automatic fail for the class and further action if the GSBS determines this is necessary.
Course Calendar SEPARATE DOCUMENT	• Tentative course calendar: • Week 1: Review of probability and statistics; Empirical distribution and plug-in estimators (Wei) • Week 2: Bootstrap estimate of standard error (Wei) • Week 3: Bootstrap for regression models; confidence intervals (Wei) • Week 4: Confidence intervals (Wei) • Week 5: Permutation and bootstrap-based hypothesis testing (Wei) • Week 6: Miscellaneous topics in bootstrap; take-home midterm exam (Wei) • Week 7: Tree based methods (Wang) • Week 8: Tree based methods (Wang) • Week 9: Review of linear and linear mixed models (Li) • Week 10: Scatterplot smoothing; penalized splines (Li) • Week 11: Penalized splines (Li) • Week 12: Semiparametric models (Li) • Week 13: General and generalized additive models (Li) • Week 14: Final project presentation (Wei, Wang, Li)