Applied Statistics and Data Analysis II (EN.553.414/614)

Spring 2020

Logistics

- **Instructor:** Jesús Arroyo (jesus.arroyo@jhu.edu)
- **Lecture:** Tue & Thu 3:00pm - 4:15pm, ~~Shaffer 100~~ https://wse.zoom.us/s/209372597
- **Office hours:** Wednesdays 3:00pm - 5:00pm, ~~Clark Hall 319 A~~ https://wse.zoom.us/s/209372597, and by appointment
- **Teaching assistant:** Cong Mu (cmu2@jhu.edu)
- **TA office hours:** Thursdays 11:00am - 1:00pm, ~~Whitehead 212~~ https://jhubluejays.zoom.us/j/488434702?pwd=cGlNN2RZV3lBaVkzczFBNmpENnFOZz09

Course information
An introduction to more advanced concepts and techniques in linear and generalized linear models for applied statistics and data analysis. Real-life datasets are used in lectures and computer assignments. Topics covered include:

1. Generalized linear models.
2. Categorical data analysis.
3. Linear and generalized linear mixed models.
4. High dimensional linear regression.
5. Semiparametric and nonparametric regression.

Prerequisites
EN 553.413/613 (Linear regression models). The statistical programming language R will be used for the class, including homeworks and exams.

Suggested references
No textbook is required for the course. Lecture notes will be posted on Blackboard. The following are suggested references.

- A. Agresti, *Foundations of Linear and Generalized Linear Models**, Wiley (2013).
- A. Agresti, *Categorical Data Analysis**, 3rd edition, Wiley (2012).
- S. Wood, *Generalized additive models using R: An introduction with R**, 2nd edition, CRC Press (2017).
- E. Demidenko, *Mixed Models: Theory and Applications with R**, 2nd edition, Wiley (2013).
- A. Galecki and T. Burzykowski, *Linear mixed-effects models using R**, Springer (2013).
- G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R**, Springer (2013).

- T. Hastie, R. Tibshirani and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press (2015).
- D. Ruppert, M. P. Wand and R. J. Carroll, *Semiparametric Regression*, Cambridge University Press (2003).

* Available online through the JHU library.

Grading Policy

- Homework assignments: 30% of the total grade.
- Midterm exams: 40% (20% + 20%) of the total grade
- Term project: 30% (5% proposal + 25% final report) of the total grade.

Important Dates

Midterm 1 ..... Thursday, March 5, 2020 (in class)
Midterm 2 ..... ~~Thursday, April 23, 2020 (in class)~~ TBD
Term project proposal ..... ~~Thursday, March 26, 2020~~ Tuesday, April 7, 2020 (by email)
Term project report ..... Thursday, May 7, 2020 (by email)
Drop deadline ..... Sunday, March 8, 2020
Withdrawal deadline ..... Friday, April 27, 2020

Homework policy
There will be homework assignments every 2-3 weeks. Homework assignments are due by the end of the class. Late homework submissions will not be graded. Should there be exceptional circumstances for which you will not be able to complete or return one homework assignment in time, at the instructor's discretion, that homework assignment will be dropped and your grade computed based on the remaining homework submissions.

Term paper
For the term paper, the students will complete a data analysis and write a short report (10 pages or less) summarizing the work. The project will be done in small groups of 2-3 students. The goal of the project is to select a dataset from which interesting and challenging questions can be asked, and answer them with at least some of the methods from the class. The report should provide a clear description of the dataset, the scientific questions of interest, the technical approach used to address or answer these questions, and state what conclusions, if any, can be derived from the analysis. More details about the proposal and term paper requirements will be announced during the semester.

Grading disputes
Your assignment scores and exams score will be entered and stored on the Blackboard page for the course. You are responsible for keeping track of your scores and to notify the course instructor should there be any missing grades or discrepancies. Your assignments and exams will be returned to you after they had been graded. Please keep all returned assignments and exams. Grading dispute for an assignment will be considered only if there is a copy of the graded assignment. Grading dispute for an exam should be made at the end of the class session in which the exam was returned. A grading dispute might entails a regrading of the whole submission.

Grading scale Grading is rounded to the nearest percentage point. Potential (upward) curving of the grades might also be done at the instructor discretion. Guaranteed letter grades, regardless of curving, are as follows,

- 90% and above: A
- 85% — 89%: A-
- 80% — 84%: B+
- 75% — 79%: B
- 70% — 74%: B-
- 65% — 69%: C+
- 60% — 64%: C
- 55% — 59%: C-
- 45% — 54%: D
- 44% or below: F

Class outcomes
Specific outcomes for this course are that students will learn more sophisticated models and techniques for analyzing data (in comparison to that discussed in EN.550.413 or EN. 553.413/613), along with the various difficulties and subtleties that might appear. This course will address the following Criterion 3 Student Outcomes

- An ability to apply knowledge of mathematics, statistics, and computational methods to analyze data (3(a) and 3(b))
- An ability to identify, formulate and solve engineering problems (3(e))
- An understanding of professional and ethical responsibility (3(f))
- An ability to communicate effectively (3(g))
- A recognition of the need for and an ability to engage in life-long learning (3(i))

Ethics
The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful. Ethical violations include cheating on exams, plagiarism, reuse of assignments, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Students are permitted and indeed encouraged to discuss lecture materials and homework problems with one another, but it is expected that the writing up of answers will be done privately. Copying by one student of another student's homework solutions is considered an ethics violation in this course. Specifically, please do not share code or output; please do not collaborate on interpretation; please do not access or use solutions from any source (including past years' homework solutions) before your homework assignment is submitted. Report any violations you witness to the instructor.

You can find more information about university misconduct policies on the web at these sites:

- For undergraduates: http://e-catalog.jhu.edu/undergrad-students/student-life-policies/
- For graduate students: http://e-catalog.jhu.edu/grad-students/graduate-specific-policies/.

Special accomodaions
Any student who may need a special accommodations in this class must obtain an accommodation letter from Student Disability Services, 385 Garland, (410) 516-4720, studentdisabilityservices@jhu.edu.