**CSCI E106 Data Modeling**

Fall 2018

*Syllabus*

Instructor: Dr. Hakan Gogtas <[hakangogtas@yahoo.com](mailto:hakangogtas@yahoo.com)>

Lectures: Mondays, 5:50-7:50 pm, live web conference

Office Hours: By appointment or before/after the class

Required Sections: Tuesdays, 5:50-7:50 pm, live web conference

Course Outline:

The course explores data modeling methodologies with the goal of understanding how to choose, apply, and interpret appropriate statistical designs and analyses for practical data problems. Topics covered include understanding the relationships in the data, theory and application of linear and non-linear regression models, model building steps, diagnostic of models, and remedial measures.

Course Objectives:

* Select and apply an appropriate data modeling technique on a real-world data set
* Apply a data modeling technique using a statistical computing package and assess the performance of the model
* Design an experiment and interpret the results of the analysis

Prerequisites: Knowledge of programming equivalent to CSCI 7 or CSCI 10, linear algebra, probability, statistics equivalent to STAT 104, and proficiency in R.

Textbook References:

* [Required] Applied Linear Statistical Models, 5th Edition, by Kutner, Nachtsheim, Neter & Li
* [Optional] Mastering Scientific Computing with R, by Paul Gerrard, Radia M. Johnson
* [Optional] Linear Models with R, Julian J. Faraway

Grading: Attendance at every session is required. Weekly assignments, including problem set zero (PSET0), have equal weights. Late submissions are not accepted. Students should use PSET0 to determine if the level of this course is appropriate for their academic background. Grades for PSET0 will be provided before Add/Drop Deadline.

Participation 5%

Assignments 25%

Midterm Exam 30%

Final Exam 40%

Computing: Students need to be familiar with the statistical software package R, available to download for free at <http://cran.us.r-project.org/>, and <http://www.rstudio.com/> for Mac and Windows. Students need to install R and RStudio on their laptops, or create a free account at <http://www.rstudio.cloud> before the semester starts.

Tentative Topics:

* Simple Linear Regression Models
* Multiple Linear Regression Models
* Variable/Model Selection
* Regression Diagnostics and Transformations
* Polynomials and Factors
* Autoregressive Models
* Non-Linear Regression Models
* Logistics Models
* Poisson Regression Models
* Generalized Linear Models

Accommodations for students with disabilities

The Extension School is committed to providing an accessible academic community. The Accessibility Office offers a variety of accommodations and services to students with documented disabilities. Please visit <https://www.extension.harvard.edu/resources-policies/accessibility-student-services> for more information.

Academic Integrity

You are responsible for understanding University’s Academic Integrity policies available at <https://www.extension.harvard.edu/resources-policies/student-conduct/academic-integrity> There are no excuses for failure to uphold academic integrity. Not knowing the rules, misunderstanding the rules, running out of time, submitting the wrong draft, or being overwhelmed with multiple demands are not acceptable excuses. To support your learning about academic citation rules, please visit the Resources to Support Academic Integrity https://www.extension.harvard.edu/resources-policies/resources/tips-avoid-plagiarism where you will find links to the Harvard Guide to Using Sources and two free online 15-minute tutorials to test your knowledge of academic citation policy. The tutorials are anonymous open-learning tools.