

# Course Overview

This course develops the foundations of predictive modeling by: introducing the conceptual foundations of regression and multivariate analysis; developing statistical modeling as a process that includes exploratory data analysis, model identification, and model validation; and discussing the difference between the uses of statistical models for statistical inference versus predictive modeling. The high level topics covered in the course include: exploratory data analysis, statistical graphics, linear regression, automated variable selection, principal components analysis, exploratory factor analysis, and cluster analysis. In addition students will be introduced to the R statistical software, and its use in data management and statistical modeling.

---

## Learning Goals

- Develop statistical modeling as a three step process consisting of: (1) exploratory data analysis, (2) model identification, and (3) model validation.
  - Use automated variable selection as a tool for model identification and as a tool for exploratory data analysis in the presence of a large number of predictor variables or a set of unlabeled predictors.
  - Develop a working understanding of the conceptual (theoretical) foundations of linear regression, principal components analysis, factor analysis, and cluster analysis with the objective of applying these techniques appropriately and validating their results.
  - Develop a conceptual and practical understanding of the difference between statistical inference and predictive modeling and how it affects our choices and actions in the statistical modeling process.
  - Apply basic R techniques and procedures for fitting statistical models.
-