STA4XX Modern Applied Regression Analysis

Catalog Description

This is an applied regression course focusing on the strategies and applications of linear, nonlinear, and nonparametric regression models. Topics include concepts of bootstrap regression modeling, smoothing methods, missing values and imputation, Outliers, and multi-collinearity detection and remedies in multiple linear regression models. A formal programming language will be used extensively for data analysis.

Pre-requisites

STA319 or permission of instructor.

Syllabus Components

- **Course Objectives**: Upon successful completion of this course, students will be equipped with skills and strategies in building both parametric and nonparametric regressions models with software programs such as R and SAS.
- **Learning Objectives**: By the end of this course, students will be able to:
  - check the model assumptions and perform diagnostics
  - build nonlinear regressions
  - build both parametric and non-parametric models
  - identify appropriate models for a particular real-world problem
  - identify the best models from a set of candidate models
  - use computer program languages to perform data analysis
  - write statistical analysis report
- **Learning Outcomes**: By the end of the class, students should be able to
  - have a clear understanding of the regression modeling process
  - have a conceptual understanding of nonlinear and nonparametric regression modeling.
  - evaluate the performance of different types of regression models
  - apply appropriate models to solve specific real-world problems
  - summarize the analysis results and draft formal reports
- **Textbooks (optional)**
  - Applied Linear Statistical Models, (any edition), by Kutner et al., McGraw-Hill
  - Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, by Harrell, Springer.

  - Applied Regression Analysis and Generalized Linear Models, by Fox, Sage.

- **Class Notes:** Class notes will be provided
- **Coverage:** See the list of tentative topics
- **Assessment:** There will be weekly assignments and projects
  - 10 Weekly data analysis assignments (small data analysis, project proposal, etc) (40%)
  - 2 projects (SAP and final report) (40%, 20% each project)
  - Class attendance and participation (10%)
  - Project presentation (10%)

Suggested Weekly Topics

Week 1: Introduction and Probability Review

- Introduction
  - Types of regression modeling
  - Regularity assumptions for all regressions
- Software
  - Install R, Rstudio, and LaTex.
  - Create a SAS account to use SAS OnDemand (SAS Studio)
- Review of probability distributions
  - Continuous distributions: normal related distributions
  - Discrete distributions: binomial and Poisson distributions

Week 2: Bootstrap Methods

- Non-parametric Bootstrap inference
  - assumptions and algorithm
  - confidence intervals
  - hypothesis-testing
- Parametric Bootstrap method

Week 3: Data Generation Process and Regression Models

- Types of regression models
  - Linear and nonlinear regression
  - Parametric and nonparametric regression
- Design-specific regression Models
  - Cross-sectional regression
  - Longitudinal regression (not to cover)
- Overfitting and underfitting issues
- Multiple linear regression

  - Structure and assumptions
  - Diagnostics and remedies
- Case study

Week 4: One-predictor Regression: SLR and Piecewise SLR

- Single predictor regression - SLR
- Non-linearity: piece-wise linear regression model
  - Concepts of spline
  - "knot value" determination
- Isotonic regression
- Case study

Week 5: One-predictor Regression: Smooth Spline Regression

- Polynomial regression
  - Collinearity
  - Centralizing predictor variable
- Concepts of basis spline (B-spline) regression
  - piecewise polynomial regression
  - order of spline
  - smoothing parameter

Week 6: One-predictor Regression: Nonparametric smoothing Regression

- Motivation and assumptions
- Local averaging methods
- Local regression method (LOESS)
- Kernel regression (optional)
- Case study

Week 7: One-predictor Regression: Bootstrap Regression

- Non-parametric Bootstrap linear regression
  - Sampling cases
  - sampling residuals
- Parametric Bootstrap regression
- Case study

Week 8: Multi-predictor Regression: MLR Tasks

- Model structure and potential violations
  - Residual distribution
  - Inappropriate regression function
  - Collinearity
- Variable selection and creation (feature extraction)

- Model selection methods
  - Error-based measures: R^2, R^2_{adj}, cross-validation, etc
  - likelihood-based measures: AIC and BIC

Week 9: MLR: Outliers and Influential Points

- Detection of influential points
  - Understand residual plots: Cooks distance and studentized residuals
  - Leave-one-out algorithm
  - Measures of influential points: DEFITS and DEBETAS
- Handling influential points
  - Deletion
  - Change the model
  - Robust regression such as iteratively reweighted least square (IRLS)
- case study

Week 10: MLR: Detecting and Handling Collinearity

- Consequence of collinearity
- Detection of collinearity: focusing on VIF
- Handling collinearity
  - Dropping variables
  - biased regression - ridge regression
- Case study

Week 11: Multi-predictor Regression: Variable Selection

- Analysis of Prediction
  - Methods of quantifying variable importance
  - Algorithmic-based cross-validation approaches
- Analysis of Association
  - Statistical significance vs practical importance
  - Methods of quantifying practical importance of predictors
- Best practices of variable selection
- Case study

Week 12: Handling Missing Values: Imputation by Regression

- Concepts of imputation
- Single and multiple imputations
- Imputing via prediction
  - Numerical variables: linear regression
  - Categorical variables: categorical regression
  - Count variables: Poisson and relevant regression

Week 13: Additive Models

- Equivalence of one-dimension of nonparametric smoothing
- Extending multiple linear regression to additive models
- Assumptions on the response
- Potential issues of overfitting and underfitting
- Prediction with additive models

Week 14: Project Report

- Finalizing the term project report.