STAT 3400 – Applied Regression

COURSE OBJECTIVES: This class will introduce the foundational knowledge, skills, and abilities associated with linear and logistic regression models. Specifically, students will learn to:

- distinguish between methods for developing and evaluating statistical learning models.
- construct predictive regression models using simple and multiple linear regression.
- construct predictive classification models using simple and multiple logistic regression.
- evaluate predictive models using various forms of cross-validation.
- implement model selection techniques such as stepwise, shrinkage, and dimension reduction.
- execute and interpret inferential analyses such as confidence intervals and hypothesis tests.
- construct inferential regression models using simple and multiple linear regression.
- construct inferential classification models using simple and multiple logistic regression.
- diagnose and remedy violations of technical conditions for inferential modeling.
- execute all the above learning objectives using the R coding language.

TEXTBOOK: *An introduction to statistical learning with applications in R* by James Gareth et al. We will cover Chapters 1-6. Access to this electronic textbook is provided to students free of charge.

SCHEDULE AND TOPICS COVERED

Lesson	Section	Topics
1	1	Statistical Learning
2	2.1	Function Estimation
3	2.2.1	Regression Accuracy
4	2.2.2	Bias-Variance Trade-off
5	2.2.3	Classification Accuracy
6	3.1	Simple Linear Regression
7	3.1.2	Accuracy of Coefficients (HT)
8	3.1.2	Accuracy of Coefficients (CI)
9	3.1.3	Accuracy of Model
10	3.2	Multiple Linear Regression
11	3.2.2 (1-2)	Analysis of Variance
12	3.2.2 (3-4)	Prediction Intervals
13	**Exam**	**Chapters 1-3.2**
14	3.3.1	Qualitative Predictors
15	3.3.2	Interaction Terms
16	3.3.3 (1-3)	Technical Conditions
17	3.3.3 (4-5)	Outliers and Leverage
18	3.3.3 (6)	Multicollinearity
19	4.1-2	Classification vs. Regression
20	4.3.1	Simple Logistic Regression
21	4.3.2	Accuracy of Coefficients
22	4.3.3	Accuracy of Model
23	4.3.4	Multiple Logistic Regression
24	4.3.4	Confounding
25	4.3.5	Multinomial Logistic Regression
26	4.4.2	Confusion Matrices
27	4.4.2	ROC Curves
28	**Exam**	**Chapters 3.3-4.4**
29	5.1.1	Validation Sets
30	5.1.2	Leave-One-Out CV
31	5.1.3	k-fold CV (Regression)

32	5.1.4	Bias-Variance Trade-off
33	5.1.5	k-fold CV (Classification)
34	6.1.1	Best Subset Selection
35	6.1.2	Stepwise Selection
36	6.1.3	Model Selection Metrics
37	6.2.1	Ridge Regression
38	6.2.2	The Lasso
39	6.2.3	Tuning Parameters
40	6.3.1	Principal Components Analysis
41	6.3.1	Principal Components Regression
42	6.3.2	Partial Least Squares
43	**Exam**	**Chapters 5.1-6.3**

PREREQUISITES: Requires prerequisite STAT 2600 and STAT 3100 or MATH 4510 (all minimum grade C-). Requires corequisite APPM 3310.

EQUIVALENT COURSES: None

LEARNING OBJECTIVES BY SECTION

Lessons	Topics	Learning Objectives
1-5	Statistical Learning	- Define statistical learning, the associated modeling trade-offs, and common applications. - Distinguish between supervised/unsupervised and parametric/nonparametric models. - Describe the structure and methods for estimating the parameters of a stochastic function. - Distinguish between regression and classification methods for statistical learning models - Explain the primary metrics and methods for assessing the accuracy of a predictive model. - Describe the bias-variance trade-off in the context of modeling flexibility and complexity.
6-12	Regression Models	- Describe the structure and methods for estimating parameters of a linear regression model. - Explain and execute the least squares method to estimate a simple linear regression model. - Construct confidence intervals for simple linear regression parameters using bootstrapping. - Complete hypothesis tests on simple linear regression parameters using randomization. - Explain the results of the Central Limit Theorem and apply them to statistical inference. - Evaluate the prediction accuracy of a linear regression model with common metrics. - Compute and interpret an ANOVA for parameters of a multiple linear regression model. - Construct and interpret prediction intervals for the response of a linear regression model.
13	Exam 1	
14-18	Regression Models	- Implement and interpret qualitative predictors in a multiple linear regression model. - Implement and interpret interaction predictors in a multiple linear regression model. - Diagnose and remedy issues with the technical conditions of a linear regression model. - Diagnose and remedy outliers and high-leverage observations in a training data set - Diagnose and remedy multicollinearity among predictors in a multiple linear regression.
19-27	Classification Models	- Describe the structure and methods for estimating parameters of a logistic regression model. - Explain and execute the maximum likelihood method to estimate a logistic regression model. - Construct confidence intervals for simple logistic regression parameters using bootstrapping. - Complete hypothesis tests on simple logistic regression parameters using randomization. - Evaluate the classification accuracy of a logistic regression model with common metrics. - Diagnose and remedy confounding variables related to a multiple logistic regression model. - Estimate and interpret a multinomial logistic regression model to classify multiple levels. - Construct and interpret a classification confusion matrix from a logistic regression model. - Compute and interpret sensitivity, specificity, and other conditional accuracies for a model. - Construct and interpret an ROC curve to determine the optimal classification threshold.
28	Exam 2	
29-33	Model Validation	- Explain the motivation and methods for splitting sample data into training and testing sets. - Evaluate the accuracy of linear/logistic regression models using the validation set approach. - Evaluate the accuracy of linear/logistic regression models using the LOOCV approach. - Evaluate the accuracy of linear/logistic regression models using the k-fold CV approach. - Describe the bias-variance trade-off inherent in deciding the number of folds in a CV.

34-42	Model Selection	- Distinguish between best subset and stepwise model selection procedures and objectives. - Determine the most accurate regression model using best subset and stepwise procedures. - Explain the pros and cons associated with selecting models based on different metrics. - Distinguish between ridge regression and the lasso shrinkage procedures and objectives. - Determine the most accurate regression model using ridge regression and lasso procedures. - Select the optimal tuning parameter for ridge regression and lasso using cross-validation. - Explain the motivation and methods for reducing the dimensions of a training data set. - Distinguish between PCR and PLS dimension reduction procedures and objectives. - Determine the most accurate regression model using PCR and PLS procedures. - Explain the issues, limitations, and remedies for regression with high-dimensional data.
43	Exam 3	