University of New Hampshire

COURSE SYLLABUS – Fall 2021

MATH 739/839 – Applied Regression Analysis - 4 Credits

Lecture: MWF: 8:10 am – 9:30 pm, Kingsbury S320 or using Zoom

---

INSTRUCTOR: Michelle Capozzoli, Ph.D.

Kingsbury N315D

Email: michelle.capozzoli@unh.edu

Phone number: 603-862-4407

Office Hours: Depending on circumstances, office hours can be face to face or virtual using Zoom. Please sign up in myCourses for a time slot that works for you: Monday and Wednesday 1:00 to 2:00 pm and Friday 12:00 to 1:00 pm. If the posted hours conflict with your schedule, please email me to set up an alternative time.

Description of Course Activities and Requirements

This is a writing intensive course and will focus on scientific writing. It is expected that answers to assignments will be thoughtfully written, using correct spelling and grammar. Graphs and statistical output should be well labeled and appropriately support your discussion. Below are some resources to support your writing:

- https://owl.purdue.edu/owl/research_and_citation/using_research/writing_with_statistics/index.html
- https://magazine.amstat.org/blog/2009/09/01/writingscipapersep09/
- http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWstats.html
- https://owl.purdue.edu/owl/purdue_owl.html

The final grade will be based on participation, homework assignments and a final project. There will be no exams.

Participation – Presentations, activities, discussions and/or short assignments on reading/lectures. These activities are geared towards enhancing the learning experience and assist students in mastering the material.

Homework– Regular homework assignments will be posted to myCourses. Problems will be graded based on completeness and correctness of the response. It is expected that all answers will be submitted showing all necessary work with software output supporting your discussions.

Final Project – A final project report is required, which applies the methods learned in the course. The project is an important part of the final grade and a thorough effort, in addition to a well-written project report is expected. Details for the project will be provided during the course.

Grading Percentages

Participation	20%
Homework	40%
Final Project	40%
Total	**100%**

Grades - Course grade will be based on earned scores as follows:

Grade	Percentage
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	0-59

Course Topics – Regression Analysis is a multivariable technique that statistically quantifies how several "predictor" variables are related to a "response" variable of interest. Applied Regression Analysis introduces various statistical methods to (1) identify such a relationship, (2) estimate parameters that characterize the relationship, and (3) apply the estimated model for prediction of future outcomes of the response variable. Furthermore, statistical tests are discussed that determine whether or not certain variables affect the response. Such tests are also utilized for model building in the case where one needs to decide which among a large number of predictor variables are important for the response variable. A major portion of the course deals with continuous (measurements) variables, but discrete variables (counts) will also be considered in topics such as Analysis of Covariance, and Logistic Regression.

The methods are introduced by way of applications from many fields of study, from the physical to the social sciences. Real world case studies are used throughout the course. Software use is an integral part of the course and allows students to gain considerable hands-on experience.

The course will cover the following topics as time allows:

Simple Linear Regression (SLR)
- Least Squares Estimation of model parameters
- Inference on Model parameters
- Inference on Model prediction
- Checking model adequacy with residual plots
- Correlation and Regression
- Introduction to Maximum Likelihood Estimation

Multiple Linear Regression (MLR)
- Brief of review of basic matrix manipulations for regression analysis
- Least Squares Estimation of MLR model parameters
- Properties of Least Squares Estimators
- Inference on Model parameters
- Inference on model predictions
- Brief introduction to multicollinearity

Checking Model Adequacy
- Various types of Residual Analysis
- Partial Regression and Leverage Plots
- PRESS statistic
- Detection and mitigation of Outliers
- Lack of Fit testing

Transformation and Weighting to Cope with Model Inadequacies
- Variance stabilizing transformations
- Linearizing transformations
- Generalized and Weighted Least Squares

Diagnosing Influence and Leverage
- Leverage definition
- Cook's D
- DFFITS and DFBETAS

Indicator Variable Regression

Multicollinearity
- What is multicollinearity?
- Understanding the causes of multicollinearity
- Basics of eigenvalues and principle components
- Negative effects of multicollinearity on fitted regression models
- Diagnosing multicollinearity with variance inflation factors and variance proportions
- Dealing with multicollinearity with Principle Components and Ridge Regression.

Model Building and Variable Selection
- Selection criteria such as PRESS and Mallows Cp
- Stepwise model building

All Subsets Regression
Model averaging.

Validation of Regression Models

Logistic Regression

Generalized Linear Model (If time allows)

Note: The above may change at the instructor's discretion.
