**Major References**

This is an advanced topics class. The topics to be covered are taken from different texts. I will not assign a required textbook for this class. Instead, I provide a list of eBooks that are freely available in WCU’s library in the following

1. **Introduction to Computer-Intensive Methods of Data Analysis in Biology** https://ebookcentral.proquest.com/lib/wcupa/reader.action?docID=261114&ppg=7
2. **Applied Regression and Modeling: A Computer Integrated Approach**

https://ebookcentral.proquest.com/lib/wcupa/reader.action?docID=4560113&ppg=46

1. **Modern Regression Techniques Using R: A Practical Guide**

https://ebookcentral.proquest.com/lib/wcupa/reader.action?docID=743577&ppg=105

1. **Regression Analysis with R**: Design and Develop Statistical Nodes to Identify Unique Relationships Within Data at Scale

https://ebookcentral.proquest.com/lib/wcupa/reader.action?docID=5259460&ppg=1

1. **Regression Models for Categorical, Count, and Related Variables: Applied Approach**, byJohn Hoffmann, University of California Press, 2016.

https://search.ebscohost.com/login.aspx?direct=true&AuthType=shib&db=e000xna&AN=1293234&site=ehost-live&scope=site&custid=s3916018&ebv=EB&ppid=pp\_vi

1. **Practical Time Series Analysis**

https://ebookcentral.proquest.com/lib/wcupa/reader.action?docID=5064689&ppg=11

**STA321 TENTATIVE TOPICS**

Following is the list of tentative topics to be covered in the semester. I may modify the list as we move forward during the semester. The course web page will provide the up-to-date list of topics every week.

**Week 1:** Setting up computing tools - getting started with R, RStudio, and R Markdown

1. Introduction: class structure, topics, assessments, and logistics.
2. Install R, RStudio, and possibly MikTex
3. Create R Markdown document- Knit HTML, PDF, and WORD file
4. Getting start with R: basic operations, vectors, data frames (R data sets).
5. Install and load R libraries

**Week 2:** Nonparametric Bootstrap Inferences

Review of simple random sampling (SRS) plan

Sampling from empirical distribution – Bootstrap sampling

Bootstrap confidence intervals

Bootstrap hypothesis tests

**R Applications**: Case study – CI and testing about population means

**Week 3: Review:** Correlation and Simple Linear Regression

1. Relationship between two numerical variables
2. Linear relationship: strength of linear correlation -coefficient of correlation
3. Least square regression model – structure, assumptions, and interpretation
4. Diagnostics, R square and interpretation
5. **R Applications**: Case study - simple linear regression (SLR) with R.

**Week 4:** Multiple Regression

1. Turn categorical predictor variables to dummy variables
2. Assumptions, Goodness-of-fit measures, and diagnostics
3. Variable selection methods
4. Summarizing output and interpreting coefficients and R square
5. **R Applications**: Case study - MLR with R

**Week 5**. Nonparametric Bootstrapping Regression Modeling

1. Bootstrapping records
2. Bootstrapping residuals
3. Bootstrap confidence intervals of regression coefficients
4. Bootstrapping test for regression coefficients
5. **R Applications**: Case study – Bootstrapping regression modeling
6. **Mini** **Project #1**: Project Report – multiple regression model: Normal and Bootstrap

**Week 6:** Categorical Regression – Simple Logistic Regression

1. Practical question and model formulation
2. Structure of the model
3. Interpretation of the regression coefficients
4. **R Applications**: Case study – fitting logistic regression with R

**Week 7:** Categorical Regression – Multiple Logistic Regression

1. Models with only categorical predictor variables – Dummy variable
2. Interpretation of coefficients
3. Variable selection methods and criteria
4. **R Applications**: Case study - logistic regression with R
5. **Mini Project #2**: Data analysis – multiple logistics regression model

**Week 8:** Frequency Regression - Simple Poisson Regression

1. Practical question and model formulation
2. Model structure and interpretation
3. Regression on rates
4. **R Applications**: Case-study - fitting counts and rates in biology

**Week 9:** Frequency Regression - Multiple Poisson Regression

1. Poison regression with more than one predictor variable
2. Categorical predictor variable – dummy variables must be defined
3. Issues of overdispersion and underdispersion
4. **R Applications**: Case-study -multiple Poisson regression model with medical data

**Week 10**: Frequency Regression - Negative Binomial Regression

1. Negative binomial is not a universal solution but an option
2. Obsolete the assumption of equality of mean and variance in Poisson
3. Negative binomial regression assumes a different relationship on the mean and variance
4. **R Applications**: Case study – multiple negative binomial regression for counts and rates
5. **Mini Project #3**: Data analysis – Regression with count and rate response

**Week 11**: Concepts of Time Series

1. The nature of time series data – Dependency between observations
2. The purpose for time series – forecasting
3. Autocorrelation and moving average
4. Measures of goodness-of-forecasting
5. Types and approaches to time series modeling
6. **R Applications**: Illustrative examples of time series using R

**Week 12**: Moving Average and LOESS Smoothing

1. Decompose a time series: trend, seasonality, and random error.
2. Additive and multiplicative models
3. Moving average (MA) and double MA
4. LOESS smoothing model: fitting, forecasting, and evaluating
5. **R Applications**: Case study – LOESS smoothing and forecasting

**Week 13**: Exponential Smoothing Models

1. Single exponential smoothing models
2. Holt’s trend models
3. Holt-Winters’ seasonal models
4. Model selection – measures of accuracy.
5. **R Applications**: Case-study – examples of exponential smoothing.
6. **Mini Project #4**: Data analysis – Modeling data with both trend and seasonality

**Week 14:** Wrap up and Prepare Presentation

1. The types of models: Linear models, GLM, and time series model (smoothing)
2. Choose one of the three projects to present to class
3. Submit a PPT presentation by the end of Friday via D2L.
4. Each presentation allows 15 minutes: 10 minutes for presenting and 5 minutes for QA.