STA 305-01 Regression Analysis

Professor: David Holleran, PhD

Email: holleran@tcnj.edu

Semester/Year: Fall 2024

Class M/R 9:30AM-10:50AM

Class Location: Physics 228

Credit Hours: 4

Phone: 609-771-3366

Office: SSB 308

Office Hours: W 9:00AM-10:30AM (Office)

By appointment (Zoom)

Course Synopsis

Regression analysis is one of the most important topics in the discipline of statistics. Nearly all statistical procedures that involve relationships among variables have their foundation in regression modeling. Moreover, statistical learning procedures such as random forest have their basis in regression analysis. The mathematical foundations for regression analysis will be thoroughly explicated in this course. Topics such as geometry, trigonometry, calculus, and linear algebra will be covered to provide the student with the baseline mathematical knowledge of regression procedures. This course is primarily, if not entirely, application based, however. The statistical package R will be used in class meetings, course assignments, and examinations as a supplementary analytical tool. It is critical, therefore, that each student develop a fluency in R scripting and usage throughout the course.

Course Objectives

This course will pursue, at a minimum, to develop within the dedicated student the ability to:

- Understand the mathematical basis of every regression procedure covered in the course;

- Formulate, estimate and evaluate regression models as the basis of supervised and unsupervised learning algorithmic methods for prediction and inference;

- Understand the role of link functions in modeling the relationship between outcome variable and the systematic component of the equation;

- Apply graphical and numerical post estimation diagnostics to examine model performance in training and validation data sets through the use of various methods of cross-validation;

- Apply data visualization in meaningful ways to convey effect sizes, understand functional form, and summarize effects of covariate patterns from regression models;

- Understand and apply model calibration procedures with the goal of improving predictions, fit and classification accuracy in validation data.

Learning Outcomes

Upon completion of this course, students will be able to demonstrate mastery in the following areas:

1. Formulate and estimate regression models for various types of outcome variables and functional forms with the purpose of analyzing and solving problems.

2. Evaluate and validate regression models for various types of outcome variables and functional forms with the purpose of analyzing and solving problems.

3. Visualize statistical data in substantively meaningful ways to examine functional form, evaluate statistical models and construct alternative models with the goal of analyzing and solving problems.

4. Formulate and deploy an analytical protocol using an ensemble of statistical learning methods with the goal of analyzing and solving problems.

Required Materials

- *Regression Analysis By Example Using R.* 6th edition. Ali S. Hadi & Samprit Chatterjee. Wiley. ISBN: 9781119830870.

- Notebook (or laptop or other device) for notetaking.

Supplemental Learning Opportunities

Periodically, I will post recorded material on Canvas to supplement what is covered in class meetings. The recordings will also allow for not only additional learning opportunities but asynchronous learning opportunities as well.

Course Expectations

Attendance will be taken at the beginning of every class meeting. The final grade for the course will not be affected by attendance, however. If you do not attend a class meeting, then it is your responsibility to obtain class lecture notes from a classmate. I will not under any circumstances provide lecture notes or tutoring for a class session missed by a student. Moreover, I will not be responsible for helping a student get "caught up" with the rest of the class if he/she is late for a class meeting. Please review the TCNJ policies concerning attendance (http://policies.tcnj.edu/policies/digest.php?docId=9134).

Student Behavior

It is unlikely that unruly student behavior will be an issue in this course. Nevertheless, I ask you to refrain from engaging in disruptive behavior during class meetings. Disruptive behavior includes (but is not limited to) engaging in disruptive discussion with another classmate when you have not been asked to do so, text messaging on your cell phone, internet use and instant messaging on the personal computer, sending/checking/reading e-mail, sending/receiving cell phone calls, internet "surfing", and continually arriving late to class meetings.

Assessment

Grades will not be given in this class; they will be *earned*. A final grade for this course will be the result of:

Regression Analysis Diary	10 points
R Assignments	40 points
Exam 1	15 points
Mid Term Exam	15 points
Final Assignment (Exam)	20 points
Total Points	100 points

A:	100-93 Total Points	C:	76-73 Total Points
A-:	92-90 Total Points	C-:	72-70 Total Points
B+:	89-87 Total Points	D+:	69-67 Total Points

B:	86-83 Total Points	D:	66-60 Total Points
B-:	82-80 Total Points	F:	< 60 Total Points
C+:	79-77 Total Points		

Regression Analysis Diary

Each student will maintain brief diary entries on Canvas on topics covered each week in the course. The diary entries are meant to be educational but also offer a degree of fun as students learn to apply their knowledge acquired in the course. Additional details will be provided on Canvas as diary entry assignments are posted.

R Assignments

Students will be periodically assigned tasks in R where they will use their regression analysis skills on a data set provided for each assignment. All work is to be completed solely by each individual student unless it is noted that it is a group or collaborative effort. Students will submit R script as well as typewritten results that will be properly formatted to include statistical tables and visualizations and summarized in report form. Additional details will be provided on Canvas for each assignment.

Exams

There will be two in-class examinations this semester. The examinations will be administered during the regularly scheduled class meeting. The final examination will be a final assignment. The final assignment is to be completed by each individual student with no additional help from others. Additional details on the final assignment will be provided at a later date.

Academic Integrity

Cheating, plagiarism, or any other form of academic misconduct will be handled and punished in accordance with the regulations governing such behavior as outlined in The College of New Jersey student handbook. You are urged to read The College of New Jersey 2024-2025 Undergraduate Catalog pertaining to policies on academic honesty.

Withdrawal from the Course

If you wish to withdraw from the course you are cautioned to follow formal procedures outlined by TCNJ. Consult The College of New Jersey 2024-2025 Undergraduate Catalog for official policies on withdrawal from classes.

Incomplete for the Course

Except for the gravest of emergencies, a grade of "incomplete" for the course will not be granted to a student. Any missing grades as a result of student absence during a quiz or examination will be assumed to be zeros and will be averaged as such.

Students with Differing Abilities

The College of New Jersey complies with Section 503/504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990 (ADA). Students with differing abilities who seek academic accommodations must first take appropriate documentation to the Office of Differing Abilities Services Eickhoff Hall Room 159. http://differingabilities.pages.tcnj.edu/.

Office Hour Meetings

I will be available each week to meet with students during regularly scheduled office hours or by appointment made in advance. A meeting during office hours will not consist of me providing a private tutorial of material already covered in class to the student. If you were absent you must obtain the notes from a classmate. I will not repeatedly provide examples of subject matter when I already covered the

subject matter in class with multiple examples. An office hour meeting should consist of the student asking questions related to the material covered in class so that the student can reach a deeper understanding of the material. **You needn't make an appointment for my regularly scheduled office hours reported at the top of this syllabus and next to my office door** <u>with the exception of hours held via Zoom videoconference.</u> The Zoom videoconference appointments allow for a greater diversity and availability of office hour meetings with me; moreover, in the case of applications in R, I am able to screen share what I am doing in R with the student so that they may work in parallel as needed.

Email

Per the TCNJ policy manual, the TCNJ email account is your official and primary email account. Thus, if you wish to communicate with me via email you must use your TCNJ account. You are permitted to have email forwarded from your TCNJ account to an external account. You are responsible, however, for checking your TCNJ account. I will only send messages to your TCNJ account. I will do my best to respond within 24 hours to your requests sent via email during the week. Email messages sent to me after 5pm on Friday of each week will hopefully be answered as early as 9am on the following Monday.

Canvas

The online course system Canvas will be used for this course and will constitute the primary interface (in addition to TCNJ email) when we are not meeting as part of regularly scheduled class meetings. The assignments and schedule for the course this semester will be managed on Canvas. The schedule is featured on the calendar. Be sure to check Canvas each day for announcements concerning the course. Canvas also features a discussion board where each student can post questions concerning course material and lectures 24 hours a day 7 days a week; thus, between office hours, email, discussion board, and chat sessions each student has *ample* opportunities to seek clarification on course material. If you post a comment to the discussion board on Canvas please notify me via email.

STA 305-01
Fall 2024*

Date	Topic(s)/Exam(s)	Readings/Assignment(s)
Thursday, August 29	Simple linear regression I	pp.71-79
Tuesday, September 3	Simple linear regression II	pp.80-90
Thursday, September 5	Simple linear regression III	pp.90-101
Monday, September 9	Multiple linear regression I	pp.109-117
Thursday, September 12	Multiple linear regression II	pp.117-121
Monday, September 16	Multiple linear regression III	pp.121-123
Thursday, September 19	**Exam 1**	
Monday, September 23	Multiple linear regression IV	pp.123-134
Thursday, September 26	Multiple linear regression V	pp.134-146
Monday, September 30	Regression Diagnostics I	pp.151-164
Thursday, October 3	Regression Diagnostics II	pp.165-191
Monday, October 7	*Fall Break – No Class Meeting*	
Thursday, October 10	Qualitative Variables I	pp.193-203
Monday, October 14	Different Slopes and Different Intercepts	pp.203-210
Thursday, October 17	**Mid Term Exam**	
Monday, October 21	Same Slopes and Different Intercepts	pp.210-217
Thursday, October 24	Transformation of Variables I	pp.225-234
Monday, October 28	Transformation of Variables II	pp.234-243
Thursday, October 31	Weighted Least Squares	pp.243-249
Monday, November 4	Transformation of Variables II	pp.234-243
Thursday, November 7	Weighted Least Squares	pp.257-275
Monday, November 11	Correlated Errors	pp.277-297
Thursday, November 14	Collinear Data	pp.301-321
Monday, November 18	Principal Components Analysis	pp.327-351
Thursday, November 21	Ridge Regression	pp.351-360
Monday, November 25	Logistic Regression I	pp.409-415
Thursday, November 28	*No Class Meeting*	
Monday, December 2	Logistic Regression II	pp.415-432
Thursday, December 5	Logistic Regression III	

*Schedule is subject to change.