DSC 423 Data Analysis and Regression

Fall 2023

Instructor

Prof. Alvin Chin
Office: Online
Email: alvin.chin@depaul.edu, Tel: (312) 480-8616

Office Hours: Thursday at 4:00-5:30 PM on Zoom (link to be provided when appointment made with BlueStar)

Class: Wednesdays at 5:45-9:00 PM, Lewis Center Room 1216 (Section 702) and online Zoom (Section 711)

Course Web page: http://d2l.depaul.edu

Summary of Course

The course topics include

1. Inference for distributions: inference for a population mean, comparing two population means using paired or independent samples, checking normal assumptions.
2. Multiple regression and correlation, analysis of residuals.
3. Model selection methods
4. Logistic Regression models

At the end of this course, students will be able:

- to perform data analyses using a statistical software and to interpret the output of their analysis;
- to apply modeling techniques to evaluate the association among variables and predict the values of a variable of interest
- to be informed and critical readers of quantitative arguments,
- to appreciate the role of statistics in empirical research and scientific study, and to gain flexible problem-solving skills applicable to unfamiliar statistical settings.

Prerequisites

IT403 or consent of instructor. Students should be familiar with univariate data analysis methods, descriptive statistics and data visualization methods, sampling techniques for data collection, and statistical inference methods including simple linear regression, confidence intervals and hypothesis testing. A brief review of these topics will be covered in the first lecture of the course.

Grading Policy

Grading is on assignments, participation, a mid-term exam and a final project in place of the final exam.

Homework and Programming assignments (40%). There will be four assignments. Assignments will be typically posted on Tuesday and will be due a week later by the following Tuesday at 11:59 PM. Assignments submitted 5 days after the due date will not be accepted. Notice that a 20% point penalty will be applied for late homework. Extensions may only be granted for exceptional reasons. **Requests for extensions must be received BY EMAIL before the due date**.

Group project (30%). Final project is due at the end of the quarter. Details will be provided later in class. The project presentation will be on Wednesday, November 8, 2023 and project report is due on Tuesday, November 14, 2023.

Mid-term Exam (20%) Online using Examity - scheduled on Wednesday October 5th, 2023. The exam will be closed book with only a calculator and one double-sided page of notes allowed.

Participation (10%) In class – students will be evaluated based on participation in the class and/or on the class online discussion board through asking and posting questions and discussion

Students receiving more than 90% of possible points are guaranteed at least an A-, more than 80% at least a B-, more than 70% at least a C-, and more than 60% at least a D.

Textbooks and Printed Resources
A Second Course in Statistics: Regression Analysis, 8th ed., William Mendenhall, Terry L. Sincich, Prentice Hall, 2020, ISBN: 13:978013516379-5 or previous edition (7th edition is fine). You can get the digital copy 8th edition from any of these sources:

A Second Course in Statistics: Regression Analysis 8th Edition [8 ed.] 2018040176, 9780135163795, 013516379X - EBIN.PUB (free PDF download)

https://www.vitalsource.com/educators/products/pearson-etext-for-a-second-course-in-statistics-william-mendenhall-terry-v9780137515264 ($43.96 for 120-day access)

http://www.mypearsonstore.com/bookstore/second-course-in-statistics-regression-analysis-013516379X ($10.99 per month)

Introductory notes on R will be posted on the course website.

Statistical software:

Students will use the R programming language. There will be an introductory lab session on R during the first week.

R is available in the CDM labs. If you prefer to install the software on your own machine, go and download it from http://cran.rstudio.com/ . You can then install RStudio (IDE for R) to run R scripts at https://www.rstudio.com/products/rstudio/download/ .

Information for all students
Students are encouraged to contact the instructor for any question related to the course. You can see me in my office during contact hours only. I am not available outside of these hours.

The best way to contact me is through email at alvin.chin@depaul.edu. Most emails will be answered within 24 hours, but do not expect a response right away during the day, as I will be working at my company office during that time.

All students are expected:
- To read this document in full!
- To attend all classes (in-person or online). If you miss a class, it is your responsibility to watch the lecture recording and to get copies of the notes or documents handed out in class. All lecture recordings are available at the D2L course site.
- To participate actively in class discussions and activities and to work on the in-class problems and exercises that are designed to improve students' understanding of the class topics.
- To be familiar with all the course documents and notes posted at the course website.
- To read all the sections in the textbook relevant to the lecture before coming to class. The reading assignments are listed in the schedule included in this syllabus. Lecture notes are meant to complement the course textbook NOT TO REPLACE IT.
- To strictly adhere to the University Academic Integrity Policy, that is published in the Student Handbook or at the Academic Integrity site at DePaul University (http://academicintegrity.depaul.edu).

Violations of the University Academic Integrity Policy include (but are not limited to): (a) using or providing unauthorized assistance or materials on course assignments; (b) possessing unauthorized materials during an examination; (c) submitting as one's own any material that is copied from published or unpublished sources such as the Internet, print, computer files without proper acknowledgement that it is someone else's; (d) submitting as one's own work a report, examination, paper, computer file, lab report or other assignment which has been prepared by someone else. If you are unsure about what constitutes unauthorized help on an exam or assignment, or what information requires citation and/or attribution, please ask your instructor. If proven, violations may result in the failure of the assignment, failure of the course, and/or additional disciplinary actions.

Tutors: You can use the tutoring program and get access to a tutor at http://www.cdm.depaul.edu/advising/Pages/TutoringProgram.aspx.

Group Project and Report

The purpose of the project is to demonstrate your ability to apply the knowledge and the techniques during this course. You can choose a particular topic and dataset for this project, or choose one from which I will select. There will be a final project presentation to the class at the end of the quarter. You can form teams of 3-5 students.

It is recommended that the online students attend the final presentations to participate in the live discussions and listen to the other projects. Alternate arrangements will be provided for online students to do their presentations or submit it as a video recording.

Deliverables for the Final Project

- Team List (Sept. 20): Team list to be posted on D2L under the discussion topic "Project Team List"
- Project Proposal (Sept. 27): One page proposal that includes project title, team members, dataset to be used (including the source), problem description, proposed approach and methodology, to be submitted on D2L
- Project Presentation (Nov. 8): Present the project in class and show results
- Project Report (Nov. 14): Report will include Abstract, Introduction, Problem Description, Dataset, Approach and Methodology, Observations and Results, Conclusion and Future Work. Details of the format will be communicated in class and on D2L. Please also submit an electronic copy of your program source code, a readme.txt file on how to run your code, data and the report

News and course information

You are expected to log in to the course website regularly, at least every other day, so you can keep up with announcements and course updates, and read the new posts on the discussion board. Participation to the discussion boards is important as it helps students share ideas and learn from each other. Rather than emailing questions, I encourage you to post your questions on the discussion forum at http://d2l.depaul.edu

I will post course announcement on the News page of the course D2L site. Some information will be sent by email (check your spam folder since your email application may filter the D2L emails as spam) and make sure that DePaul has your correct email. You also have the option (highly recommended!) to subscribe to the news page, and the discussion forum at http://d2l.depaul.edu.

Tentative Schedule

The following schedule is tentative. The reading assignments are from the course textbook.

Week	Class Date	Topic	Reading assignment
1	Sept. 6	A review of basic concepts relevant to our course, students should have already covered much of this material elsewhere. Inference for the mean of a population. Introduction to R. Lab session.	Chapter 1: sections 1.1 through 1.9
2	Sept. 13	Inference on two population means: two independent samples. Linear regression models: simple linear regression, parameter estimation, least square estimates	Chapter 1: Section 1.10, Chapter 2 and 3
3	Sept. 20	Multiple linear regression, categorical variables, computing predictions and prediction errors. Assignment #1 handed out. Project team list due.	Chapter 4: sections 4.1 through 4.8, 4.14
4	Sept. 27	Residual analysis, model building, predictions and prediction errors. Assignment #1 due. Project proposal due.	Chapters 8: sections 8.1 to 8.5, Chapter 5: sections 5.1, 5.2, 5.3, 5.7, Chapter 3.9
5	Oct. 4	Multicollinearity and influential observations. Building more complex models: non-linear regression, polynomial regression. Midterm review. Assignment #2 handed out.	Chapter 8.6, Chapter 7, Chapter 5.6
6	Oct. 11 (Zoom)	Midterm exam (1st half of class). Variable selection methods and model validation. (2nd half of class).	Chapter 5, Section 5.11 and Chapter 6
7	Oct. 18	Interaction model, logistic regression and predictive models for qualitative variables. Assignment #2 due. Assignment #3 handed out.	Case Study 2, Chapter 9: sections 9.5,9.6
8	Oct. 25	More about logistic regression models. Analysis of variance using regression models.	Chapter 9: sections 9.5,9.6. Chapter 12.
9	Nov. 1	Industrial applications of logistic regression and machine learning. Assignment #3 due. Assignment #4 handed out.	
10	Nov. 8	Final project presentations. Assignment #4 due.	
11	Nov. 14	Final project report due.	

Course Policies

Changes to Syllabus

This syllabus is subject to change as necessary during the quarter. If a change occurs, it will be thoroughly addressed during class, posted under Announcements in D2L and sent via email.

Online Course Evaluations

Evaluations are a way for students to provide valuable feedback regarding their instructor and the course. Detailed feedback will enable the instructor to continuously tailor teaching methods and course content to meet the learning goals of the course and the academic needs of the students. They

are a requirement of the course and are key to continue to provide you with the highest quality of teaching. The evaluations are anonymous; the instructor and administration do not track who entered what responses. A program is used to check if the student completed the evaluations, but the evaluation is completely separate from the student's identity. Since 100% participation is our goal, students are sent periodic reminders over three weeks. Students do not receive reminders once they complete the evaluation. Please see https://resources.depaul.edu/teaching-commons/teaching/Pages/online- teaching-evaluations.aspx for additional information.

Academic Integrity and Plagiarism

This course will be subject to the university's academic integrity policy. More information can be found at https://offices.depaul.edu/oaa/faculty-resources/teaching/academic-integrity/Pages/default.aspx.

Academic Policies

All students are required to manage their class schedules each term in accordance with the deadlines for enrolling and withdrawing as indicated in the University Academic Calendar. Information on enrollment, withdrawal, grading and incompletes can be found at: http://www.cdm.depaul.edu/Current%20Students/Pages/PoliciesandProcedures.aspx

Incomplete Grades

An incomplete grade is a special, temporary grade that may be assigned by an instructor when unforeseeable circumstances prevent a student from completing course requirements by the end of the term and when otherwise the student had a record of satisfactory progress in the course. All incomplete requests must be approved by the instructor of the course and a CDM Associate Dean. Only exceptions cases will receive such approval. Information about the Incomplete Grades policy can be found at http://www.cdm.depaul.edu/Current%20Students/Pages/Grading-Policies.aspx

Students with Disabilities

DePaul University is committed to ensuring equal access to its educational and extracurricular opportunities for students with disabilities. The Center for Students with Disabilities (CSD) offers reasonable academic accommodations and services to support our students. We also serve as a resource to the many university departments that have a responsibility to accommodate students.

Please see https://offices.depaul.edu/student-affairs/about/departments/Pages/csd.aspx for Services and Contact Information.

Online office hours for OL courses

Faculty should be accessible to online students via phone and email.ss

Course Policies as Suggested by the Dean of Students Office

Attendance: Students are expected to attend each class and to remain for the duration. Coming 15 minutes late or leaving 15 minutes early constitutes an absence for the student. The overall grade for participation drops one-third after any absence. Three absences for any reason, whether excused or not, may constitute failure for the course.

Class Discussion: Student participation in class discussions will be measured in two ways. First, students are highly encouraged to ask questions and offer comments relevant to the day's topic. Participation allows the instructor to "hear" the student's voice when grading papers. Secondly, students will be called upon by the instructor to offer comments related to the reading assignments. Students must keep up with the reading to participate in class discussion.

Attitude: A professional and academic attitude is expected throughout this course. Measurable examples of non-academic or unprofessional attitude include but are not limited to: talking to others when the instructor is speaking, mocking another's opinion, cell phones ringing, emailing, texting or using the internet whether on a phone or computer. If any issues arise a student may be asked to leave the classroom. The professor will work with the Dean of Students Office to navigate such student issues.

Civil Discourse: DePaul University is a community that thrives on open discourse that challenges students, both intellectually and personally, to be Socially Responsible Leaders. It is the expectation that all dialogue in this course is civil and respectful of the dignity of each student. Any instances of disrespect or hostility can jeopardize a student's ability to be successful in the course. The professor will partner with the Dean of Students Office to assist in managing such issues.

Cell Phones/On Call: If you bring a cell phone to class, it must be off or set to a silent mode. Should you need to answer a call during class, students must leave the room in an undisruptive manner. Out of respect to fellow students and the professor, texting is never allowable in class. If you are required to be on call as part of your job, please advise me at the start of the course.