

DSC 423: Data Analysis and Regression

Course Information

Title: DSC 423: Data Analysis and Regression

Meeting Time: Wednesday 5:45pm to 9:00pm (sections 901)

The class runs from Apr. 3rd to Jun. 5th

Location: CDM 224 for two in-person sessions: the first one on **May. 8th**; the second one on **May 15th**.

Rest of the sessions will be held online via Zoom.

Note: Students who have registered for section 910 are not required to join the classes—they can just watch the recorded lectures that will be made available weekly after each class meeting.

Contact Information

Instructor: Jamshid Sourati

Email: jsourati@depaul.edu

Office Hours: Monday from 12:30pm to 2:00pm.

Location: All offices hours will be held online via Zoom.

Course Overview

The course topics include:

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

Course Objectives

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

- to interpret the output of statistical software 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.

Course Prerequisites:

IT403. Students are expected to understand basic mathematical notation and be familiar with college algebra concepts.

Schedule

Week	Topic	Chapter	Project Milestones
1	Introduction to Class Review of Basics – normal distribution, hypothesis testing	1	
2	Simple Linear Regression – modeling response, model assessment	2, 3	– Forming teams – Selecting project's name
3	Multiple Regression Models – first-order, second-order, interaction models	4	– Selecting a data set
4	Model Building – Two-variable models, qualitative variables	5	– The project's proposal (one to two pages)
5	Variable Screening – Validation, Stepwise Selection	5, 6	
6	Regression Pitfalls – Multicollinearity, variable transformation	7	
7	Residual Analysis – Assumption analysis, detecting lack of fit	8	
8	Advanced Regression Models – Logistic regression, regularization	9, 10	
9	Experimental Design – ANOVA, randomized block design	11, 12	
10	Final project presentation		– Group presentation

Course Management System

We will use DePaul University's Desire2Learn system (d2l.depaul.edu).

Attendance and Participation

Students who have taken the in-person section of the course (901) are expected to attend the two in-person sessions at CDM 224 and other eight synchronous online classes via Zoom. The classes will begin on Apr. 3rd and the first session will be online. Our in-person lectures will be on May 8th and May 15th. For those taking the asynchronous online section of the course (910), there is no expectation to attend the in-person/synchronous online sessions at any specific time. They can just watch the recorded lecture that will be uploaded on D2L weekly after each class meeting. However, everyone is expected to interact with the class on a weekly basis (e.g., through submitting assignments, and participating in the class discussion forums).

Textbooks

Mendenhall and Sincich, Second Course in Regression Analysis, 8th Edition, Prentice Hall/Pearson, 2019.

Software

The statistical packages used in this course will be R. They are available in all DePaul labs. You can also access R remotely by using our CDM terminals (suitable for fast connections).

Please note that this course is **not** a tutorial in R but rather a course in the foundations of regression analysis. You will be provided with some introductory tutorials in R to get you going.

R Software Access—DePaul Labs: Latest version of R is available in the CDM labs and all DePaul labs

Virtual Labs—You can also use DePaul's virtual lab to access the latest version of R on windows and Mac versions. For instructions on how to remotely access the terminal services and how to activate your DePaul account, please visit http://my.cdm.depaul.edu/resources/Terminal_services_guide.pdf

Important Remark 1—Virtual server stalls, gets slow and crashes; if you decide to use the software using virtual lab, start early and keep multiple backups in multiple places/mediums. Late submission or inability to do the assignment due to server issues will not be accepted.

Important Remark 2—For any issues with accessing the virtual lab or the R software, contact Information Services helpdesk at (312) 362-8765, I won't be able to help you with DePaul Software related issues.

Emails

I make every attempt to reply to emails within 24 hours. Include the course number, your name, and your student ID in the title of the email. Emails should be written in a professional manner with clear and concise language.

News Widget

The primary form of communication for this class will be the news widget on the D2L. Subscribe to the widget and ensure that DePaul has your correct email.

Forums

The class forum is the preferred place to ask questions about the class. If you have questions about lecture notes or assignments, please post them there. I read these frequently. All students should subscribe to the forums so that you receive email updates.

Course Evaluation

Assignments: 40%

Quizzes: 30%

Final group project: 30%

Grading Scale

A:	90% - 100%
B:	80% - 90%
C:	70% - 80%
D:	60% - 70%
F:	less than 60%

Plusses and minuses are given for the upper and lower 3% in a letter's range. There is no A+.

Quizzes

Weekly quizzes will be administered through D2L and will test the material in the weekly lectures. These quizzes are open book/notes/friends. Consider the questions yourself because the purpose is to test yourself on the material. This is called '**recall practice**' and will help you remember the material. Quizzes of weeks 6 and 10 are more comprehensive than the rest and only allow for one attempt.

Quiz	%	Attempts
1	1	Unlimited
2	1	
3	1	
4	1	
5	1	
6	10	ONE
7	1	Unlimited
8	1	
9	3	
10	10	ONE
Total	30%	

Homework

Homework assignments will be submitted online on the D2L website. No email submissions will be accepted. There will be 5 homework assignments during the quarter. Work to be submitted for the course is generally due one week after it was assigned; late submissions are allowed with a 5%, 10%, and 15% penalty for a one day, two days, and three days, respectively. No late work will be accepted after three days since the assignment was due. Discussion with peers is allowed with respect to homework assignments, however, each student must turn in original work, i.e., your own write up.

The homework problems will be directly related to each module that they are assigned for, and it is important that you read through the homework **early each week** so that you know what to look for in the lectures and tutorials and have time to think about solution strategies. Each homework will take a significant amount of time and it may not be possible to complete a homework in a single day/weekend; the best way to work on the homework is to work on the materials over time. Note also that it may take up to 24 hours for you to receive a response from an e-mail or posted question, so do not wait until the day the homework is due to ask questions particularly about the meaning or intent of the homework questions. Grading for the homework problems will be based on three components:

HW	Week	%
1	2	8
2	4	8
3	6	12
4	8	12
Total		40

- I. Technical execution of the analysis techniques, which includes the computation of mathematical expressions and the use of scripts to execute the techniques,
- II. Interpretation and conclusions from the analysis,
- III. Explanation of your analysis, including completeness and conciseness. Learning to communicate your results in a concise manner is extremely important, and you will be penalized if you either do

not include enough detail or include far too much detail. In short, if you submit many pages for a problem when a page or two is sufficient, and **if the specific answers to the problem's parts are difficult to find**, you will be penalized. Submitting a stream of R/Python code, with answers included as comments **will result in a deduction**.

The Final Project

The final project will be a group project with three to five students to a project. The project will be to thoroughly analyze and apply course techniques to a large dataset. You will be expected to apply a range of techniques from the course and from your own readings to the data, and to draw conclusions from your analysis. Your project grade will be based on both individual and group performances, including the following components:

- Periodic milestones throughout the quarter (see the schedule table above),
 - Forming team (5%)
 - Data selection (5%)
 - Project proposal (20%)
- A group final summary report for which all members are expected to contribute (20%),
- An individual final report detailing your contributions and individual investigations in the project and reflections on what you learned from the project (20%),
- Final presentation (20%),
- Contribution to team assessed through:
 - Minutes of all team meetings documenting your discussions (5%)
 - Peer evaluations (5%)

(Total = 100%)

The groups will be formed in the first three weeks of the class. Group dynamics play an important role in any project, and you are expected to make every effort to both contribute to the group effort and make the environment safe, comfortable, and respectful for your team members.

Final projects will be presented by each group during the 10th week of the class. Students in the in-person section of the course, 901, are expected to present their work during the online session of the class on the 10th week (Jun. 5th). Students in the asynchronous online section, 910, should record their group presentation and submit their recording by 5:45pm of the same day (Jun. 5th). The final group/individual reports and other materials (minutes of the team meetings and peer evaluations) should be submitted in the week after, **by the end of the day (11:59pm) of Jun. 12th**.

Non-performance as a team-member on a project.

Usually, the peer evaluation and documentation, including the meeting minutes, in addition to an overall desire for excellence, is sufficient motivation for individuals to contribute a fair share to the team project. However, in extreme cases, individuals have been known to completely cease contributing to a team project. If this is the case, a team has the right to notify the instructor that the individual is no longer contributing, and the team no longer wants the individual on the team. **PLEASE INFORM EARLY ON RATHER THAN AT THE LAST MINUTE.**

This is not a decision to be made lightly, as expulsion from a team will result in the **loss of 40% of the final project grade**. Because this is such a serious decision, any team that makes this decision will also experience a point deduction for the remaining members. In this situation, **each remaining team member will lose 10% of the final project grade**.

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.

Plagiarism

Students in this course should be aware of the strong sanctions that can be imposed against someone guilty of plagiarism. If proven, a charge of plagiarism could result in an automatic F in the course and possible expulsion. The strongest of sanctions will be imposed on anyone who submits as his/her own work any assignment which has been prepared by someone else. If you have any questions or doubts about what plagiarism entails or how to properly acknowledge source materials be sure to consult the instructor.

Academic Integrity

Work done for this course must adhere to the University Academic Integrity Policy, which you can review in the Student Handbook or by visiting “Academic Integrity” (academicintegrity.depaul.edu) at DePaul University.

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:

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:

www.cdm.depaul.edu/Current%20Students/Pages/Grading-Policies.aspx

Writing Center

Consider contacting or visiting the Writing Center to discuss your assignments for this course or any others. You may schedule appointments (30 or 50 minutes) on an as-needed or weekly basis, scheduling

up to 3 hours worth of appointments per week. Online services include Feedback-by-Email and IM conferencing (with or without a webcam). All writing center services are free.

Writing Center tutors are specially selected and trained graduate and undergraduate students who can help you at almost any stage of your writing. They will not do your work for you, but they can help you focus and develop your ideas, review your drafts, and polish your writing. They can answer questions about grammar, mechanics, different kinds of writing styles, and documentation formats. They also can answer questions and provide feedback online, through IM/webcam chats and email. Obviously, the tutors won't necessarily be familiar with every class or subject, but they are able to provide valuable help from the perspective of an interested and careful reader as well as a serious and experienced student-writer.

Students with Disabilities

Students seeking disability-related accommodations are required to register with DePaul's Center for Students with Disabilities (CSD) enabling them to access accommodations and support services to assist with their success. There are two office locations:

- Loop Campus – Lewis Center #1420 – (312) 362-8002
- Lincoln Park Campus – Student Center #370 – (773) 325-1677

Students who register with the Center for Students with Disabilities are also invited to contact Dr. Gergory Moorhead, Director of the Center, privately to discuss how he may assist in facilitating the accommodations to be used in a course. This is best done early in the term. The conversation will remain confidential to the extent possible. Please see:

offices.depaul.edu/student-affairs/about/departments/Pages/csd.aspx

Schedule your appointments with enough time to think about and use the feedback you'll receive. Bring your assignment handout and other relevant materials to your appointments.

Student 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.