

DSC 423 / DSC 323: DATA ANALYSIS AND REGRESSION

Fall 2019 | *In-Class Sessions: M 5:45pm – 9:00 pm @ the assigned classroom per campusconnect*

Instructor: Nandhini Gulasingam
Office Hours: Mondays 4:00 pm– 5:30 pm (Loop – CDM 635) – Times and dates may vary. Check BlueStar via campusconnect
Email: mgulasin@depaul.edu
Phone: (773) 325-4917
Course Website: <http://d2l.depaul.edu>
Drop Dates: <https://academics.depaul.edu/calendar/Pages/default.aspx>
Response Policy: 24-hours within normal business days from course start to end date per campusconnect

I. Course Summary

The course topics include: Inference for distributions, inference for a population mean, comparing two population means using paired or independent samples, checking normal assumptions. Multiple regression and correlation, analysis of residuals. Model selection methods. 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

II. Prerequisite(s)

IT 223 or IT 403 or consent of instructor. Students should be familiar with statistical inference methods, including sampling distributions, confidence intervals and hypothesis testing. A brief review of these topics will be covered in the first lecture of the course.

III. Textbooks and Resources

- **[Required]** A Second Course in Statistics: Regression Analysis, 7th ed., William Mendenhall, Terry L. Sincich, Prentice Hall, 2010 (ISBN: 9780321691699) – Or previous edition.
- **[Optional]** Applied Statistics and the SAS Programming Language, 5th edition, by Cody, R.P. and Smith. J.S. Prentice Hall (2005), ISBN: 0131465325. **Introductory notes on SAS will be posted on the course website.**

VII. Assignments, Activities, and Grading

Homework and Programming Assignments (6)	50%
Group Project	50%

Grading Scale

A	93-100 points	B-	80-82.99 points	D	50-69.99 points
A-	90-92.99 points	C+	77-79.99 points	F	< 50
B+	87-89.99 points	C	73-76.99 points		
B	83-86.99 points	C-	70-72.99 points		

IV. IMPORTANT NOTICES

Homework and Programming Assignments (50%)

Assignments will be typically posted on Monday and will be due a week later by **Monday at 11.59 pm** (there may be exceptions, so refer to the schedule). All assignments should be done **independently and cannot be discussed with anyone** other than your instructor or a CDM assigned tutor. **No assignments or papers will be accepted past the due date** unless a documented medical or personal emergency arises. **These assignments cannot be done in a couple of hours, so start working on them on the day it was given, that way you have 7 days to complete your assignments little by little. Also, applications crash, if you use virtual server to access the software, it will be slow, stall or crash at times, so again, start early, keep multiple backups of your work in multiple places in case you were to lose your files for some reason. NO EXTENSIONS WILL BE GIVEN UNLESS FOR A DOCUMENTED REASON SPECIFIED ABOVE, NO LATE ASSIGNMENTS PAST THE DUE DATE EVEN A COUPLE OF MINUTES LATE WILL BE ACCEPTED. YOU ARE RESPONSIBLE FOR CHECKING THE RIGHT VERSIONS ARE SUBMITTED. THEY WILL NOT BE REGRADED IF YOU SUBMIT THE WRONG FILE(S).**

All assignments should be submitted in a single MS WORD format, not PDF or any other file types will be accepted. If you submit any other file type, it will not be graded.

Group Project (50%)

The purpose of the final project is to demonstrate your ability to apply the knowledge and the techniques learned during this course. The final project for this class is more extensive analysis task, chosen by you from among the topics we discuss. **Detailed project description is posted and is available on D2L under “Group Project” section on the first day of class; please review it on your own. We will discuss the requirements in class on day it is specified to be discussed as per the schedule.**

Final Project:

- **Form Teams (During first 5 Sessions):** Form teams during the first five sessions. Team size will be discussed in class. You can form teams in class or use the D2L Discussion Board topic “Looking for Teammates” to post your information and find your team.
- **Proposal (Nov. 4):** A 1-2 page proposal that includes, project title, teammates, dataset, problem description, the proposed approach and methodology.
- **Presentation (Nov. 18):** In-class presentation, code and data file is due before class.
- **Report (Nov. 25):** A 20-30 pages report is due. See project instructions already available on D2L for detailed information. Team evaluations are also due the same day.

NO EXTENSIONS WILL BE GIVEN FOR ANY PROJECT DELIVERABLE UNLESS FOR A DOCUMENTED REASON SPECIFIED ABOVE UNDER HOMEWORD SECTION. NO LATE DELIVERABLES PAST THE DUE DATE EVEN A COUPLE OF MINUTES LATE WILL BE ACCEPTED. YOU ARE RESPONSIBLE FOR CHECKING THE RIGHT VERSIONS ARE SUBMITTED. THEY WILL NOT BE REGRADED IF YOU SUBMIT THE WRONG FILE(S).

What to Expect

As with any course in statistics and decision science, you are expected to spend a **significant amount of time outside of class** (this will be higher for student with no programming and/or no data backgrounds) doing the required reading and/or additional readings, reviewing lectures and working on assignments and projects. The

best way to learn the coding and techniques learnt in class is to experiment with them on a variety of problems; you can also use the additional code and data provided under examples.zip file for each week. You will, of course, have a range of problems posed on the assignments, but the more you can practice with these techniques, the better you will learn their nuances, and the better prepared you will be to apply them. The topics in this course build on each other, so be sure to monitor your progress carefully in this course and seek assistance immediately if you start to fall behind.

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. **Any questions regarding the course, assignments, project, etc. (other than that of personal nature) should be posted on the discussion forum at <http://d2l.depaul.edu>. If you email me with questions that are not of personal nature, they will not be answered until they are posted under the relevant discussion post topic for that week.**

Course announcements will be posted 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>.

Statistical Software

We will be using SAS software in this course. Three or four lab sessions will be scheduled during the quarter.

SAS Software Access:

DePaul Labs: SAS 9.4 is available in the CDM labs and all DePaul labs.

Virtual Labs

You can also use DePaul's virtual lab to access the SAS 9.4 windows and Mac versions. To access the virtual labs, see instructions posted under SAS Resources section of D2L. If you run into any issues accessing the virtual lab or the software, contact the help-desk at 312-362-8765.

IMPORTANT:

- (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.
- (2) Any issues relating with SAS, contact IS using the above phone number, I won't be able to help you with DePaul Software related issues.

[Optional] SAS Textbook: Applied Statistics and the SAS Programming Language, 5th edition, by Cody, R.P. and Smith. J.S. Prentice Hall (2005), ISBN: 0131465325.

Information for all Students

Any questions regarding the course, assignments, project, etc. (other than that of personal nature) should be posted on the discussion forum at <http://d2l.depaul.edu>. Only questions of personal nature should be emailed.

All students are expected:

- To read this document in full.
- To be familiar with all the course documents and notes posted at the course website.
- To attend all classes. 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 materials and recordings are available at <http://d2l.depaul.edu/> under your course home page.
- To participate actively to 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 read all the sections in the textbook and additional readings 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.
- 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 unobtrusive 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.
- **Talking, Sleeping or being Disruptive: If you talk, sleep or are disruptive while the lecture is starting or in session, you will be asked to leave.**
- **To strictly adhere to the University Academic Integrity Policy**: This course will be subject to the university's academic integrity policy. More information can be found at <https://resources.depaul.edu/teaching-commons/teaching/academic-integrity/Pages/default.aspx>.

Submitting work that is not yours is grounds for an automatic 'F' for the entire course – this includes taking content and ideas from others or consulting others to complete your deliverables other than your instructor.

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.

Accommodations (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 <https://offices.depaul.edu/student-affairs/about/departments/Pages/csd.aspx> for Services and Contact Information.

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>

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>

V. Use of Class Material (to everyone who has access to the D2L Class)

Course material is to be used only by the students in this class for this class purpose only, and cannot be used or redistributed in paper form or be distributed on the internet electronically or via websites. This includes course lecture slides, course videos, example codes, in-class activities and answers, lab exercises and data files, assignment questions and data files, assignment answers, reading materials, project documents and any articles presented in class or for homework.

VI. Schedule

Please note that this schedule is subject to change. The days you have a lab, you will also have a lecture in the assigned classroom at the assigned time per campusconnect

Lecture	Topics	Reading	What's Due
Lecture 1 Sept. 16	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 SAS. Lab session (7:30 PM - 9:00 PM)	Chapter 1: 1.1 - 1.9	
Sept. 23	Class Cancelled due to Medical Emergency		↘ Assignment 1 due Mon. Sept. 23 at 11.59pm
Lecture 2 Sept. 30	Statistical inference and introduction to linear regression models. Multiple linear regression, Parameter estimation, Least square estimates.	Chapters: 2 , 3 Chapter 4: 4.1 - 4.9 Chapter 7: 7.5	
Lecture 3 Oct. 4 (Fri) CDM 224	Multiple linear regression: model diagnostics. Residual Analysis and Categorical variables. Transformations.	Chapter 4: 4.1 - 4.8, 4.12, 4.14 Chapter 5: 5.7 – 5.9 Chapter 7: 7.6 Chapter 8: 8.1 - 8.6	
Lecture 4 Oct. 7	Multiple linear regression: Computing predictions and prediction errors. Multicollinearity and influential observations.	Chapter 3: 3.9 Chapter 4: 4.9, 4.11 Chapter 7: 7.4 Chapter 8: 8.6	↘ Assignment 2 due Mon. Oct. 7 at 11.59pm
Lecture 5 Oct. 14	Multiple linear regression: Model building and variable selection methods. Lab session (7:30 PM - 9:00 PM)	Chapter 6	↘ Assignment 3 due Mon. Oct. 14 at 11.59pm
Lecture 6 Oct. 21	Multiple linear regression: Model validation methods. Discussion of project.	Chapter 4: 4.10-4.14 Chapter 5	↘ Assignment 4 Due Mon. Oct. 21 at 11.59pm
Lecture 7 Oct. 28	Logistic regression: Exploratory analysis, model selection, diagnostics, and predictions. Lab session (7:30 PM - 9:00 PM)	Chapter 9: sections 9.5,9.6	↘ Assignment 5 due Mon. Oct. 28 at 11.59pm
Lecture 8 Nov. 4	Logistic regression: Model validation. Lab session (7:30 PM - 9:00 PM)	Chapter 9: sections 9.5, 9.6. Chapter 12	↘ Assignment 6 due Mon. Nov. 4 at 11.59pm ↘ Project Proposal due Mon. Nov. 4 at 11.59pm
Lecture 9 Nov. 11	Building more complex models: non-linear regression, polynomial regression.	Chapter 4: 4.10-4.14 Chapter 5	
Lecture 10 Nov. 18	Presentation Presentation Slides + SAS Code + Data file		↘ Final Project Presentation, Code, Data due due Mon. Nov. 18 at 5.45 pm
Lecture 11 Nov. 25	Final Project Report Team Evaluations		↘ Final Project Report and Team Evaluations due Mon. Nov. 25 at 11.59 pm