DSC 424: Advanced Data Analysis
Tuesdays, 5:45-9:00pm CST
CDM 220 (Weeks 3-10), Weeks 1 and 2 will be held over Zoom at the instructor's Zoom link

Instructor Information

Name: Ronan Johnson
Office: CDM, Room 822 (subject to change)
Office Hours:
In person: Tuesdays, 10am-Noon CST
Zoom: By appointment
Email: sjohn165@depaul.edu
Zoom: https://depaul.zoom.us/my/ronanej
Course page: http://d2l.depaul.edu

Course Description

This course will teach advanced statistical techniques to discover information from large sets of data. The course topics include visualization techniques to summarize and display high dimensional data, dimensional reduction techniques such as principal component analysis and factor analysis, clustering techniques for discovering patterns from large datasets, and classification techniques for decision making. The methods will be implemented using standard computer packages.

Prerequisites

DSC 423: Data Analysis and Regression. You are expected to have mastered simple linear regression and multiple linear regression with stepwise variable selection techniques. Previous experience with R and linear algebra is helpful but not required.

Course Goals

At the end of this course, students should be able to conduct and interpret the following analysis techniques and be able to identify which approach is appropriate for a given data set and task to be performed:

- Multivariate linear regression
    - Model building, variable selection, and regularization techniques
- Principal component analysis & Factor analysis
    - Eigenvectors and eigenvalues, dimension reduction, factor rotation, scree plots
- Correspondence analysis to find relationships between categorical variables
- Linear discriminant analysis including Fisher's discriminant function
- Cluster analysis
    - Similarity measures; hierarchical, density-based, and spectral clustering techniques
- Multidimensional scaling
- Canonical Correlation analysis to assess the relationship between two sets of variables

Recommended Books:

- Brian S. Everitt, Graham Dunn, "Applied Multivariate Data Analysis" (2nd Edition). ISBN 978-0470711170.
- Hair, Black, Babin, & Anderson, “Multivariate Data Analysis” (8th edition). ISBN 978-9353501358

Lectures

Lectures will be traditional in-class presentations, discussions, and exercises. As per University policy during Winter quarter 2022, the first two lectures will be held online at the instructor's Zoom link (https://depaul.zoom.us/my/ronanej). Lectures 3-10 are expected to be held on-campus in **CDM room 220**. Please be aware that this may change due to extenuating circumstances due to the COVID-19 pandemic. Students are expected to check their emails and course announcements on d2l regularly to stay updated on University policies. All lectures will take place between **5:45-9:00pm CST** unless otherwise announced in rare cases. All lectures will be recorded and posted promptly to d2l.

Attendance

Formal attendance will not be taken, with the exception of Week 1 for the purposes of financial aid reporting. If you do not attend in-person during Week 1, you will be asked to submit a text file to the appropriate submission folder on d2l stating your intent to take the course.

Students are expected to either attend or watch all lectures in a timely manner.

Grading

Final grades will be determined by weekly homework assignments and quizzes, a midterm exam, and a final group project. Each category will be weighted as follows:

- Homework: 30%
- Quizzes: 10%
- Midterm: 25%
- Final project milestones: 20%
- Final project report: 15%

Coursework:

- There will be **six (6) homework assignments**, assigned during modules 1, 2, 3, 4, 7, and 8. Each assignment will take a significant amount of time; They are usually not possible to complete in a single day/weekend. Students are highly encouraged to start early and ask questions during class time. A portion of the lecture each week will be dedicated to answering homework questions.

- There will be a total of **ten (10) quizzes** to accompany the lectures. They will cover the material from each week's lecture with the exception of the Module 6 quiz which will act as a review for the midterm. Three (3) attempts will be given per quiz. Only the highest grade will be recorded in the grade book. The Module 0 quiz covers the syllabus and has 10 attempts. You must get a 100% on this quiz before any assignment submissions will be graded.

- The **midterm** will be given during class time on **Tuesday, February 8th**. The exam will start promptly at **5:45pm CST**. Students will have two (2) hours to complete the exam and will be allowed a single 3x5 inch index card (both sides) for notes. The exam will cover all material from Modules 0 through 5. Students enrolled in the on-campus section are expected to take the exam in person unless extenuating circumstances are communicated to the instructor *before the exam*. Online asynchronous students should see the Online Students section of this syllabus for more information about exam options.

- The final project is to be completed in groups of three (3). A very small number of two or four person groups will be allowed if even groups of three are not possible. Groups will be formed after the drop date. There will be **four (4) milestones** assigned throughout the quarter and a **single final written report**. All milestones and the final report will be submitted as a group to the appropriate group submission folder by the posted due date. As with the homework, groups are encouraged to submit partial work before the due date. Only the most recent submissions will be graded unless clearly indicated otherwise. Peer evaluations will be submitted individually to a separate submission folder. More detailed information will be available on d2l and discussed during lecture.

All due dates are listed in this syllabus under the Course Schedule section. They will also be clearly indicated in the assignment descriptions and on the submission folders and quiz sections on d2l.

Assignment submissions

All assignments, examples, lectures, quizzes, and course notes will be available only on the online learning platform d2l at http://d2l.depaul.edu. All assignments are to be submitted to the appropriate submission folder on d2l by the posted due date. No submissions will be accepted by email.

Students are encouraged to submit partially finished assignments early in order to avoid missing the due date and therefore eligible for partial credit. Only the most recent submission will be graded unless clearly marked otherwise. Submissions will not be graded before the due date.

Submissions are expected to be in the stated required format, typically a single pdf document. Students are expected to put their names somewhere visible in the assignment document and to use intuitive, readable file names. Submissions not in the proper format may be marked down.

Students are responsible for ensuring all work submitted to the d2l submission folder is complete and accurate. It is highly recommended that you download your file(s) after submission to ensure they have been uploaded properly and that the uploaded file is the one the student intended. Corrupted files, or files that cannot be opened, may be given a grade of zero (0).

Quizzes are to be completed on d2l by the posted due date.

Late Policy

Due to the scope of this course, deadlines must be strictly adhered to. This benefits students by keeping the schedule on track, allowing the instructor adequate time to give detailed feedback, and keeping the workload evenly spaced. All work submitted after the posted due date will be given a grade of **zero (0)** with only two exceptions:

1. Each student will be granted a single 48-hour extension on one (1) homework assignment, no questions asked. The first late homework submission will be considered use of this extension.

This extension cannot be used on the quizzes, project milestones, midterm, or final report. All further homework submitted after the due date will be given a zero (0).

2. Extenuating circumstances may occasionally require special accommodations. If a student is experiencing such a circumstance, they are expected to contact the instructor as early as possible to make any necessary arrangements. If a student must delay or miss several assignments or the midterm, they will be advised to either withdraw or drop the course.

If a student or group makes a submission both before and after the due date, only the submission uploaded *before* the due date will be graded. If a student would like the later version to be graded, they must clearly indicate so in the notes section of the submission. This will constitute the use of the single 48-hour extension. If the student has already used their extension, or if the assignment is part of the final project, then only the on-time submission will be graded.

Software

This class is taught using the R programming language in RStudio, but you are free to use any appropriately capable software system or programming language to complete assignments. However, be aware that I may not be able to provide technical assistance for any given package. If you aren't used to programming, R may seem intimidating at first, but it's a very useful tool for any data analyst or statistician; You are all highly encouraged to take this opportunity to learn this versatile software!

That being said, this is not a course about R. Rather, we will be using R to perform the calculations necessary to form the basis of our analysis. You will be given sample code each week, and we will cover examples together in class, but you will be expected to modify and debug code on your own. This will likely take some Googling of error messages.

The vast majority of the code you'll need to write will be heavily based on the provided examples. You are free to directly copy and modify any example code I give you for your assignments and final project without fear of being accused of plagiarism. There will be **no** R code on the midterm.

What to Expect

This class is dense and highly technical. We have a lot to cover every week, so we will be keeping up a fairly quick pace. You will be expected to spend a significant amount of time outside of class reviewing notes, working on assignments, and doing some independent research. If you fall behind, it will likely be difficult for you to catch up. For many, data analysis can be a tough subject to wrap your head around. You'll be thinking about the world in ways you may have never had to before. I'll be asking you to think critically and express your conclusions in clear, precise writing. I will get picky about your wording so your conclusions can't be misinterpreted. You'll likely get frustrated.

Probably the most unique thing about data analysis versus other areas of mathematics is that there are no "correct" answers. You will have to justify your reasoning with the concepts we cover in class, but as long as you make strong arguments, you'll get a good score even if I (the instructor) don't necessarily agree with you. It's not uncommon for two students to come to opposite conclusions on an assignment and get similar scores.

On that note, I welcome debate! If you disagree with a conclusion I make, please speak up (especially during lecture), and we'll have a discussion.

Course Schedule

Week/date	In class	Work assigned	Work due
Week 1 January 3rd-9th	**Tuesday, January 4th** **5:45PM at** **https://depaul.zoom.us/my/ronanej** Lecture: Modules 0 and 1 Course and syllabus overview Linear algebra Multiple regression review	Quiz Module 0 Quiz Module 1 Homework Module 1 Intent to attend	**Sunday, January 9th** at **11:59PM CST** Quiz Module 0 Quiz Module 1
Week 2 January 10th-16th Last day to drop: January 16th	**Tuesday, January 11th** **5:45PM at** **https://depaul.zoom.us/my/ronanej** Lecture: Module 2 Regularized Regression	Quiz Module 2 Homework Module 2	**Sunday, January 16th** at **11:59PM CST** Quiz Module 2 Homework Module 1
Week 3 January 17th-23rd	**Tuesday, January 18th** **5:45PM at CDM 220** Lecture: Module 3 Eigenvectors/Eigenvalues Principal component analysis	Quiz Module 3 Homework Module 3 Project Milestone 1	**Sunday, January 23rd** at **11:59PM CST** Quiz Module 3 Homework Module 2 Project Milestone 1
Week 4 January 24th-30th	**Tuesday, January 25th** **5:45PM at CDM 220** Lecture: Module 4 Factor analysis Rotation	Quiz Module 4 Homework Module 4 Project Milestone 2	**Sunday, January 30th** at **11:59PM CST** Quiz Module 4 Homework Module 3
Week 5 January 31st - February 6th	**Tuesday, February 1st** **5:45PM at CDM 220** Lecture: Module 5 Correspondence analysis Multiple correspondence analysis	Quiz Module 5 Quiz Module 6	**Friday, February 4th** at **11:59PM CST** Homework Module 4 Project Milestone 2 **Sunday, February 6th** at **11:59PM CST** Quiz Module 5 Quiz Module 6
Week 6 February 7th-13th	**Tuesday, February 8th** **5:45PM at CDM 220** Midterm Exam (2 hours)	Project Milestone 3	None (work on Milestone 3!)
Week 7 February 14th-20th	**Tuesday, February 15th** **5:45PM at CDM 220** Lecture: Module 7 Linear Discriminant Analysis Fisher's Linear Discriminant	Quiz Module 7 Homework Module 7	**Sunday, February 20th** at **11:59PM CST** Quiz Module 7 Project Milestone 3

Week 8 February 21st-27th	**Tuesday, February 22nd** **5:45PM at CDM 220** Lecture: Module 8 Cluster analysis techniques	Quiz Module 8 Homework Module 8 Project Milestone 4	**Sunday, February 27th** at **11:59PM CST** Quiz Module 8 Homework Module 7
Week 9 February 28th - March 6th	**Tuesday, March 1st** **5:45PM at CDM 220** Lecture: Module 9 Canonical Correlation Analysis	Quiz Module 9	**Sunday, March 6th** at **11:59PM CST** Quiz Module 9 Homework Module 8 Project Milestone 4
Week 10 March 7th-13th	**Tuesday, March 8th** **5:45PM at CDM 220** Lecture: Module 10 Overflow week; Misc topics Final report questions/discussion	Final project report Peer evaluations	None
Finals Week March 14th-20th	None	None	**Tuesday, March 15th** at **11:59PM CST** Final project report Peer evaluations

Changes to the Syllabus

This syllabus is intended to be unchanged for the duration of the quarter, but modifications and adjustments may be necessary. Significant changes are unlikely, but should they be needed, any modifications will be discussed in class and announced through email and announcements on d2l.

If you see an error on the syllabus, d2l, assignment descriptions, etc, *please* let me know ASAP so I can correct it. You will never be held responsible for anything caused by a mistake *I* (the instructor) made.

Online Students

This class has both synchronous in-person and online asynchronous sections. If you are an online asynchronous student, you are welcome to join any online lectures during the scheduled lecture time. You may also attend lectures in person, provided you have the necessary vaccination documentation or exemption on file with the University.

There will be two options for online students to take the midterm exam:

1. Proctoring. If have the necessary medical documentation, you may schedule an on-campus proctor at the Loop campus. Alternatively, you may schedule a session with a remote proctor if you cannot come to campus. Proctored exams will be pencil and paper.
2. Online exam. I ask you to coordinate accommodations with me in advance if you choose to take the exam online.

Online students will need to email me indicating their choice of examination method no later than **Tuesday, February 1st** (one week before the midterm exam). The exam *must* be taken during **Week 6**.

School Policies:

Attitude and Civil Discourse:
A professional and academic attitude is expected throughout this course. Measurable examples of nonacademic 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.

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 the instructor at the start of the course.

COVID-19 Health and Safety Precautions
Keeping our DePaul community safe is of utmost importance in the pandemic. Students, faculty and staff are expected to (1) wear a mask as required at all times while indoors on campus; (2) refrain from eating and drinking in classrooms; (3) keep current with their COVID-19 vaccinations or exemptions; (4) stay home if sick; (5) participate in any required COVID-19 testing; (6) complete the online Health and Safety Guidelines for Returning to Campus training; and (7) abide by the City of Chicago Emergency Travel Advisory. By doing these things, we are Taking Care of DePaul, Together. The recommendations may change as local, state, and federal guidelines evolve. Students who do not abide by the mask requirement may be subject to the student conduct process and will be referred to the Dean of Students Office. Students who have a medical reason for not complying with any requirements should register with DePaul's Center for Student with Disabilities (CSD).

Respect for Diversity and Inclusion at DePaul University as aligned with our Vincentian Values
At DePaul, our mission calls us to explore "what must be done" in order to respect the inherent dignity and identity of each human person. We value diversity because it is part of our history, our traditions and our future. We see diversity as an asset and a strength that adds to the richness of classroom learning. In my course, I strive to include diverse authors, perspectives and teaching pedagogies. I also encourage open dialogue and spaces for students to express their unique identities and perspectives. I am open to having difficult conversations and I will strive to create an inclusive classroom that values all perspectives. If at any time, the classroom experience does not live up to this expectation, please feel free to contact me via email or during office hours.

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/teachingcommons/teaching/Pages/onlineteachingevaluations.aspx for additional information.

Email

Email is the primary means of communication between faculty and students enrolled in this course outside of class time. Students should be sure their email listed under "demographic information" at http://campusconnect.depaul.edu is correct.

Academic Integrity Policy

I expect that you have read and understood DePaul's policy on Academic Integrity:
http://academicintegrity.depaul.edu/
It is part of this syllabus; follow it.

Plagiarism

The university and school policy on plagiarism can be summarized as follows: Students in this course, as well as all other courses in which independent research or writing play a vital part in the course requirements, 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 a report, examination paper, computer file, lab report, or other 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.

Incomplete

An incomplete grade is given only for an exceptional reason such as a death in the family, a serious illness, etc. Any such reason must be documented. Any incomplete request must be made at least two weeks before the final, and approved by the Dean of the School of Computer Science, Telecommunications and Information Systems. Any consequences resulting from a poor grade for the course will not be considered as valid reasons for such a request. Students must formally request an incomplete by filling out a Request for Incomplete Grade form, available at the CDM main office, and submitting it to me.

Preferred Name & Gender Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the quarter so that I may make appropriate changes to my records. Please also note that students may choose to identify within the University community with a preferred first name that differs from their legal name and may also update their gender. The preferred first name will appear in University related systems and documents except where the use of the legal name is necessitated or required by University business or legal need. For more

information and instructions on how to do so, please see the Student Preferred Name and Gender Policy at http://policies.depaul.edu/policy/policy.aspx?pid=3327

Resources for Students with Disabilities

Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential. To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class).

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 (312) 362-8002
- Lincoln Park Campus (773) 325-1677
- Email: csd@depaul.edu

Students who register with the Center for Students with Disabilities are also invited to contact Dr. Gregory 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.