

**DSC 424: Advanced Data Analysis**  
**Tuesday, 5:45PM-9:00PM**  
**Online**  
**Fall 2020**

***Instructor Information***

Instructor: Stephanie Besser, M.S.P.A., M.S.A.S., M.S.A., M.A.C.J.C.  
Office: Online  
Office Hours: Tuesday, 5:15pm-5:45pm  
Tuesday, 9:00pm-10:00pm  
Email: [sbesser1@depaul.edu](mailto:sbesser1@depaul.edu)  
Course Page: <http://d21.depaul.edu>

***Course Description***

The course will teach advanced statistical techniques to discover information from large sets of data in a lecture discussion format, including R program instruction and examples. 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.

***Course Objectives***

Students should be able to do the following after completing this course:

1. Selects appropriate statistical methods for their final project research questions by understanding the statistical theory behind the multivariate statistical methods.
2. Demonstrates these statistical methods in R, and applies them to the final group project datasets.
3. Interprets the R output both statistically and within the application of the field the methods are being applied.
4. Deconstructs critically journal articles on the statistical methods covered in class and discusses them in class.
5. Reports statistical results in written form through homework assignments and the final project.
6. Analyzes and shows statistical findings to audiences with various understanding of statistics through the final project executive summary and journal article.
7. Designs a presentation of findings and presents them to an audience through final project course presentations.

***Course Topics***

At the end of this course, the student should have a basic understanding of the following topics and be able to identify which approach is appropriate for a given dataset and data analysis task to be performed:

- Multivariate linear regression (least-square estimation & normal equations, model building & variable selection)
- Principal component analysis & Factor analysis (Eigen-values and eigenvectors, scree plots, dimension reduction, factor rotation)
- Canonical Correlation (to assess the relationship between two sets of variables)
- Discriminant analysis (Fisher's discriminant function)
- Cluster analysis Part I (similarity measures, hierarchical clustering & non-hierarchical clustering).
- Cluster analysis Part II (advanced clustering techniques)
- Multidimensional Scaling (if time permits)

***Recommended Book***

- Everitt and Dunn, "Applied Multivariate Data Analysis", Published by Wiley, ISBN-13: 978-0470711170, 2010 (2<sup>nd</sup> edition).

### ***Prerequisites***

DSC 423 (Formerly CSC 423): Data Analysis and Regression. We also assume that you have had some experience in linear algebra and in particular, matrix arithmetic.

### ***Grading***

Grading in this course will be based on a combination of homework, programming and participation assignments, periodic quizzes, the midterm exam, which will be held during the 5<sup>th</sup> week of class, and the final project. The final grade will be computed based on the following weights:

- Homework/programming assignments: 30%,
- **Take-Home Midterm** exam on **October 13<sup>th</sup>**: 35%
- **Final project** due on **November 27<sup>th</sup>**: 35%

The midterm exam is mandatory and you must take it to pass the course. Makeup exams/quizzes are only given in extreme circumstances (severe illness, etc.) which must be documented. If there is a chance that work or other commitments will cause you to be absent for one of these, you should enroll in the online section of the course, as it gives more flexibility in terms of date and time for the exam.

### ***Graduate Students***

- School of Computing Students in graduate classes in the school will use the A/B/C/D/F grading basis.

### **GRADING SCALE:**

93 - 100: A  
90 - 92: A -  
87 - 89: B +  
83 - 86: B  
80 - 82: B -  
77 - 79 C +  
73 - 76 C  
70 - 72 C -  
67 - 69 D +  
60 - 66 D  
60 - 0 F

### ***Course Drop Dates***

**Last day to drop AQ2020 classes with no penalty – Tuesday September 22, 2020**  
(100% refund of tuition if applicable and no grade on transcript)

**Last day to withdraw from AQ2020 classes – Tuesday October 27, 2020**

See Academic Calendar: <https://academics.depaul.edu/calendar/Pages/default.aspx>

### ***Homework/Programming Assignments, Papers' Reviews, and Exam Policies***

#### **Homework/programming assignments**

There will be homework/programming assignments, which are due at the beginning of class one or two weeks after they are assigned. Late assignments will be accepted up to one lecture later than the assigned due date with a 25% penalty. No assignments will be accepted beyond a week after the due date. The assignments must be submitted online at <https://d21.depaul.edu>. No assignments will be accepted via e-mail.

### **Tentative Homework Assignment Due Dates**

- Assignment 1 Due Sunday, September 27<sup>th</sup>, 2020
- Assignment 2 Due Sunday, October 11<sup>th</sup>, 2020
- Assignment 3 Due Sunday, November 1<sup>st</sup>, 2020
- Assignment 4 Due Sunday, November 15<sup>th</sup>, 2020

### ***Midterm:***

There will be a midterm exam given on **October 13<sup>th</sup>, 2020**. The midterm is a closed book and notes exam, but students are allowed to bring a calculator (no phones or internet connected devices are allowed).

### ***The Final Project:***

Depending upon the class size, the final project will be a group project **with four 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 grade in the project will be based on both individual (40%) and group (60%) performance, including the following components:

- Periodic milestones throughout the quarter which will entail both group and individual work
- Minutes of all team meetings documenting your discussions
- A group final summary report in the form of a journal article for which all members are expected to contribute
- Peer evaluations

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 on the 10<sup>th</sup> and 11<sup>th</sup> weeks of class, November 17<sup>th</sup>, 2020 and November 24<sup>th</sup>, 2020.** Each group must present their projects. Online students are encouraged to participate either through helping to build the presentation or by recording a part of the presentation to be played during the presentation.

### ***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.

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. In this situation, **each remaining team member will lose 10% of the final project grade**.

### ***What to Expect***

As with any course in mathematics and computer science, you are expected to spend a significant amount of time outside of class reviewing lectures and working on homeworks/projects. The best way to learn mathematical or statistical techniques is to experiment with them on a variety of problems. You will, of course, have a range of problems posed on the homeworks, but the more you can experiment with these techniques on real datasets, the better you will learn their nuances, and the better prepared you will be to apply them in novel situations.

The topics in this course build on each other, much in the same way as in any programming or math course. Be sure to monitor your progress carefully in this course and see me immediately if you miss a class or start to fall behind so that we can discuss how you can resume and complete the course.

### *Software*

The use of the RStudio statistical system will be taught in class and will be required for some homework problems. It will be the only officially supported platform for the course and the homework assignments. However, the final project will allow you to use any software you wish to complete the analysis (e.g. SPSS, SAS, R, Python, etc.) as long as the software has sufficient features to complete the assignment/project. Instructor permission will be required in order to use external programs for the final project, which will be asked as part of the introductory project assignments. You may use more than one software package to work on your final project. Different packages have different strengths and you should learn to leverage each package for its strengths.

### *E-Mail questions*

I receive many e-mails each week, as well as teach multiple classes, so in order to help insure that you receive a prompt response from me, you are required to preface each e-mail's subject line with **DSC 424:<subject>** (note: capital letters on DSC and no space). **If you do not do this, you may not receive a response to your e-mail.** Outside of office hours, the best way to reach me is by email, whereby I will return emails within a timely manner, usually within 24-48 hours. Online students can email me to schedule alternative office hours during the quarter.

### *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, and if your DePaul email is being forward to a secondary email address that the forwarding process is working properly. **DO NOT WAIT UNTIL THE LAST MINUTE TO SEND ME EMAILS REGARDING ASSIGNMENTS OR THE FINAL PROJECT!**

### *Attendance*

It is expected that you will attend every class; it is the single most important action you can take in mastering the course objectives. You are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss. For distance learning students, this means viewing the classes in a timely manner, participating in the discussion forums, and being sure to email me with any course or concept questions as soon as possible.

### Use of Class Material

**COURSE MATERIALS ARE TO BE USED ONLY BY THE STUDENTS IN THIS CLASS AND CANNOT BE REDISTRIBUTED IN PAPER FORM OR ON THE INTERNET VIA ANY WEBSITES.** This includes course lectures, homework questions, homework answers, the midterm review, or any articles presented in class or for homework.

### *For online students:*

Recordings of each lecture will be available a few hours after the "live" class, and can be found at the course website <https://d2l.depaul.edu>. Online students are expected to watch the lectures every week and to keep up with the course information posted on the course website.

### **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.

## **University Policies**

**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.

### **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=332>.

**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 undistruptive 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.

### **Academic Integrity and Plagiarism**

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>.

### **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>

### ***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 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**

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. 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.

### **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.