

CSC 334: Advanced Data Analysis

Session: Spring

Tuesdays 5:45-9:00, 1/9 - 3/20

Class: Lewis 1111

Office: CDM 820

Instructor: Jonathan F. Gemmell

Email: jonathan.gemmell@gmail.com

Office Phone: (312) 362-8261

Office Hours: Tuesdays 1:00-4:00pm

Please email ahead of time

Course Description: The 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: CSC 324 or consent of instructor. Students are expected to understand basic mathematical notation and be familiar with college algebra concepts.

Course Management System: DePaul University's Desire2Learn system (d2l.depaul.edu).

Calendar: The calendar on the D2L contains all the quizzes, labs, exams and lecture schedules. Make sure to review this schedule.

News Widget: The primary form of communication for this class will be the news widget on the D2L. Please make sure you subscribe to the widget and 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 a quiz question, the exams or lecture notes, please post them there. I read these frequently. All students should subscribe to the forums so that you receive email updates.

Suggested Textbooks: 1) Hair, Black, Babin, & Anderson, "Multivariate Data Analysis", Published by Prentice Hall, ISBN-13: 9780138132637, 2010 (7th edition). 2) Johnson & Wichern, "Applied Multivariate Statistical Analysis", Published by Prentice Hall, ISBN-13: 9780131877153, 2008 (6th edition).

Software: The statistical package used in this course is R. It is available in all DePaul labs. If you are familiar with SAS or SPSS, you may complete your assignments with this software.

Grades:	Class contract:	2.5%	Final Grades:	A:	90% - 100%
	Quizzes:	(7 of 9) x 2.5%		B:	80% - 90%
	Labs:	4 x 5%		C:	70% - 80%
	Project:	30%		D:	60% - 70%
	Midterm:	10%		F:	less than 60%
	Final Exam:	20%			

- Pluses and minuses are given for the upper and lower 3% in a letter's range. There is no A+.
- You may not pass the class if you do not receive a passing grade on the final exam and the project.

Class Contract: All students must complete the class contract. It is meant to ensure that you have read the syllabus and understand the requirements for the course.

Quizzes: There are 9 quizzes. The lowest is dropped. A quiz may be taken as many as 10 times and only the highest grade is recorded. Quizzes are taken online through the D2L. Quizzes are due at 5:45 on the day of the class in which they will be reviewed. There will never be an extension for the quizzes.

The purpose of the online quizzes is not to test your knowledge. You may (and are encouraged) to work through the problems with your fellow classmates, seek out the help of tutors, ask questions on the forums or even offer potential solutions on the forums.

The purpose of the online quizzes is to prepare you for the exams (which will test your knowledge). The questions are very indicative to what will be on the exam. Thus, when taking a quiz, you should endeavor to learn the underlying material and not simply get the right answers.

Labs: Labs are meant to test your knowledge about R and the statistical approaches. These labs are meant to give you experience working with statistical tools. You are expected to work alone. Late labs receive a penalty of 2% per hour.

Project: The class project is a significant part of the class and thus requires a significant amount of effort. Students will work in a cohorts of 3-5. Weekly milestones are meant to keep you on pace. The project concludes with a 20-page report and a presentation.

Exams: The midterm is given during week 5. The final exam is given during week 11; please see the calendar. The exam may not be rescheduled except under extreme extenuating circumstances. Exams are closed-book and closed-notes (one page). Stand alone calculators are OK, but not cell phones.

Proctored Exams: DL students may take the exams 1) in class, 2) on campus in the DL office or 3) with a proctor. DL students must take the exam in the allotted window. Exams may not be rescheduled except under extreme extenuating circumstances. Registration begins on the D2L by using the "Proctored Exam" tab. Please see: <http://www.cdm.depaul.edu/onlinelearning/Pages/FAQ.aspx#exams>

Attendance: Attendance is not required. However, students are responsible for all material and announcements presented in class. DL students are expected to watch the lectures online.

Online Teaching Evaluation: 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. Students complete the evaluation online in CampusConnect.

Academic Integrity Policy: This course will be subject to the academic integrity policy passed by faculty. More information can be found at <http://academicintegrity.depaul.edu/>

Plagiarism: The university and school policy on plagiarism can be summarized as follows: 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.

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 College of Computing and Digital Media. Any consequences resulting from a poor grade for the course will not be considered as valid reasons for such a request.

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), and make sure that you have contacted the Center for Students with Disabilities (CSD) at:

Student Center, LPC, Suite #370

Phone number: (773)325.1677

Fax: (773)325.3720

TTY: (773)325.7296