

CSC423: DATA ANALYSIS AND REGRESSION / CSC 324: DATA ANALYSIS & STATISTICAL SOFTWARE II

Spring 2018 | CDM, DePaul University

Instructor: Nandhini Gulasingam
Office Hours: Mon. 4:45 pm– 5:30 pm (Loop – CDM 703) and
Office: Thu. 4:45 pm– 5:30 pm (Loop – CDM 703)
Email: mgulasin@depaul.edu
Phone: (773) 325-4917
Website: <http://d2l.depaul.edu>

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)

IT223 or IT403 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

- A Second Course in Statistics: Regression Analysis, 7th ed., William Mendenhall, Terry L. Sincich, Prentice Hall, 2010 (ISBN: 9780321691699) – Or previous edition.
- Introductory notes on SAS will be posted on the course website.

VII. Assignments, Activities, and Grading

CSC 324 is taught in parallel to CSC 423. The structure of the class is identical. However, the undergrad section will have fewer requirements on homework assignments, project, and exam.

| | |
|------------------------------------------|-----|
| Homework and Programming Assignments (6) | 50% |
| Group Project and Reports | 50% |

Grading Scale

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

IV. IMPORTANT NOTICES

Homework and Programming Assignments (50%)

There will be five or six assignments. Assignments will be typically posted on Monday and will be due a week later by **Tuesday 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. **No assignments or papers will be accepted past the due date** unless a documented medical or personal emergency arises.

Group Project and Report (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. Final projects will include a presentation to the rest of the class at the end of the quarter, in place of a final exam.

Whenever it is possible, it is recommended that the online students attend the final presentations to participate in the live discussions of the final projects and to complete critiques of the other projects. Alternate arrangements will be provided for online students to do their presentations or submit it as a video recording.

Deliverables for the Final Project:

- **Proposal (Week 8)**: One page proposal that includes, project title, team mates, dataset, problem description, the proposed approach and methodology, and at least three references other than text book or class notes.
- **Presentation (Week 11)**: Each project is to be presented using PowerPoint, and the PPT file will have to be submitted to be published on course web site.
- **Reports (Week 11)**: Report will include a 1-page non-technical report and a 5-8 pages technical report. The electronic copy should be in a zip file format consisting all program source code, data and the reports.

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>.**

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>.

Information 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 <http://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.

Statistical Software

We will be using SAS software in this course. Three or four lab sessions (recorded for online students) 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.

Optional text for SAS: 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

Students are encouraged to contact the instructor for any question related to the course. You can see me during office hours. The best way to contact me is through email at mgulasin@depaul.edu. Most emails will be answered within 24 hours. You can also call me in my office at 773-325-4917.

All students are expected:

- To read this document in full.
- To attend all classes (online students are expected to watch each lecture). 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 be familiar with all the course documents and notes posted at the course website.
- 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.
- **To strictly adhere to the University Academic Integrity Policy**, that is published in the Student Handbook or at the Academic Integrity site at DePaul University (<http://academicintegrity.depaul.edu>). Violations of the University Academic Integrity Policy include (but are not limited to): (a) using or providing unauthorized assistance or materials on course assignments; (b) possessing unauthorized materials during an examination; (c) submitting as one's own any material that is copied from published or unpublished sources such as the Internet, print, computer files without proper acknowledgement that it is someone else's; (d) submitting as one's own work a report, examination, paper, computer file, lab report or other assignment which has been prepared by someone else. If you are unsure about what constitutes unauthorized help on an exam or assignment, or what information requires citation and/or attribution, please ask your instructor. If proven, violations may result in the failure of the assignment, failure of the course, and/or additional disciplinary actions. **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. The only exception to this rule is discussing the project with your teammates.**

Accommodations (Students with Special Needs)

Any student who requires assistance is asked to contact the University's Office of Students with Disabilities (Phone 773/325-1677, TTY 773/325-7296, Fax 773/325-7396, website

<http://studentaffairs.depaul.edu/studentswithdisabilities/>). They will be able to assist both student and faculty. If

you have a condition that requires accommodation from the Productive Learning Strategies program (PLuS

Program) please contact them at the Student Center room 370 (Phone 773/3251677 or online:

<http://studentaffairs.depaul.edu/plus/>

V. Schedule

Please note that this schedule is subject to change

| Week | Topics | Reading | What's Due |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Week 1 Mar. 26 | 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) @ Daley 505 | Chapter 1: 1.1 - 1.9 | |
| Week 2 Apr. 2 | Statistical inference and introduction to linear regression models. Multiple linear regression, Parameter estimation, Least square estimates. | Chapter 2 Chapter 3 Chapter 4: 4.1 - 4.9 Chapter 7: 7.5 | ➤ Assignment 1 due Tu. Apr. 3 at 11.59pm |
| Week 3 Apr. 9 | 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 | ➤ Assignment 2 due Tu. Apr. 10 at 11.59pm |
| Week 4 Apr. 16 | 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 Chapters 8: 8.6 | ➤ Assignment 3 due Tu. Apr. 17 at 11.59pm |
| Week 5 Apr. 23 | Multiple linear regression: Model building and variable selection methods. Lab session (7:30 PM - 9:00 PM) @ CDM 527 | Chapter 6 | ➤ Assignment 4 due Tu. Apr. 24 at 11.59pm |
| Week 6 Apr. 30 | Multiple linear regression: Model validation methods. Discussion of project. | Chapter 4: 4.10-4.14 Chapter 5 | ➤ Assignment 5 due Tu. May. 1 at 11.59pm |
| Week 7 May. 7 | Logistic regression: Exploratory analysis, model selection, diagnostics, and predictions. Lab session (7:30 PM - 9:00 PM) @ Daley 505 | Chapter 9: sections 9.5,9.6 | |
| Week 8 May. 14 | Logistic regression: Model validation. Lab session (7:30 PM - 9:00 PM) @ Daley 505 | Chapter 9: sections 9.5, 9.6. Chapter 12 | ➤ Assignment 6 due Tu. May. 15 at 11.59pm ➤ Project Proposal due Tu. May. 15 at 11.59pm |
| Week 9 May. 21 | Building more complex models: non-linear regression, polynomial regression. | Chapter 4: 4.10-4.14 Chapter 5 | |

| | | | |
|---------------------------|----------------------------------------------------------------------------------------------------------------|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Week 10 May. 28 | Memorial Day – University Closed | | |
| | Work on Final Project Deliverables and Presentations | | |
| Week 11 Jun. 4 | Group Presentations (in-class) Final Project Reports + Code + Data Team Evaluations | | <ul style="list-style-type: none"> ➤ Presentations (in-class) ➤ Presentations slides due on D2L before class starts for in-class and online students – due Jun. 4 at 5:45 pm ➤ Final Project Reports and Team Evaluations due Mon. Jun. 4 at 11.59 pm |