

CSC423 Data analysis and Regression
CSC324 Data analysis and statistical software II
Fall 2014/15

Instructor

Dr. John McDonald
Office: CST 831 - CDM Building,
E-mail: jmcdonald@cdm.depaul.edu
Tel. (312) 362-5142.

Office Hours: Monday and Thursday at 3:30-5:00pm

Course Web Page: <http://d2l.depaul.edu>

Course Summary

The course topics include

1. Inference for distributions: inference for a population mean, comparing two population means using paired or independent samples, checking normal assumptions.
2. Multiple regression and correlation, analysis of residuals.
3. Model selection methods
4. Logistic Regression models

Course learning goals, students will be able to:

- Perform data analyses using a statistical software and to interpret the output of their analysis;
- Apply modeling techniques to evaluate the association among variables and predict the values of a variable of interest
- Be informed and critical readers of quantitative arguments,
- Appreciate the role of statistics in empirical research and scientific study, and to gain flexible problem-solving skills applicable to unfamiliar statistical settings.

Prerequisites

IT223 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.

Grading Policy

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 the exam. Course grade is based on the following components:

- **Homework and Participation (35%).** There will be four or five assignments involving both written components and components involving SAS or R programming. Homework submissions that are illegible (e.g. poorly scanned written documents) or poorly organized will be penalized. Assignments submitted 5 days after the due date will not be accepted. Notice however, that a 20% penalty will be applied for late homework.

Participation includes, but is not limited to attendance (watching the course for DL) and accessing materials on D2L weekly. You are also expected to participate in class discussions (in-class or through the D2L discussion forums) and contribute to the class in a positive way.

- **Midterm Exam (35%)** In class - tentatively scheduled on Monday October 20th, 2014. This exam will be written and the only acceptable excuse for missing an exam is and extreme

situation such as illness (which must be documented). If missing an exam will be absolutely unavoidable, see me before the exam takes place or you will receive an automatic zero. Make-up exams are oral.

- **Group project (30%).** Final project is due at the end of the quarter. Details will be provided later in class.

Textbooks and Printed 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 and R will be posted on the course website.

Statistical software:

Students have the option to use either SAS or R. Two or three lab sessions (recorded for online students) will be scheduled during the quarter.

Both SAS 9.3 and R are available in the CDM labs, and SAS is available in all DePaul labs. If you prefer to install the software on your own machine:

- R is freely available at <http://cran.rstudio.com/>.
- The MS Windows-version of SAS 9.3 installer (with installation directions) is now available on Campus Connection at <http://campusconnect.depaul.edu> under Self Service -> Student Resources -> Student Software -> Windows -> SAS -> Installer -> 9.3

The pdf file found there contains installation instructions. Note: SAS 9.3 system requirements can be found at <http://support.sas.com/documentation/installcenter/93/index.html>

Note: The installation file is very big and will take a few hours to download even with a fast connection. Some students have suggested downloading the installation software on a flash drive using the computers on campus as this is significantly faster.

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 in my office (room 831 of the CDM building) during office hours, or by e-mail at other times.

The best way to contact me is through email at jmcdonald@cdm.depaul.edu. You can also call me in my office during office hours at (312) 362-5142.

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 recordings are available through the CDM course websites.
- 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. **Note that it is your responsibility to check the course website often and to watch your e-mail for announcements. Missing an announcement does not excuse you from course requirements**
- To read all the sections in the textbook 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.

Tutors:

The tutors' schedule is at: <http://www.cdm.depaul.edu/advising/Pages/TutoringProgram.aspx>. I'll post names of tutors in statistics on the course website.

Course website and Discussion board

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. Rather than emailing questions, I encourage you to post your questions on the discussion forum at <http://d2l.depaul.edu>. Answers to questions on the discussion board should be in the form of suggestions and discussion about the problem. Posting of homework problem answers will result in the board being suspended.

While I will try to check in to the discussion board from time to time, the discussion board is not the primary method for getting feedback directly from me. If you want to ask me a direct question or get feedback on a problem or project, then contact me directly at jmcdonald@cdm.depaul.edu.

I will also post course announcement on 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>.

Tentative Schedule

The following schedule is tentative. The reading assignments are from the course textbook.

Week	Topic	Reading assignment
1	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 and R. Lab session	Chapter 1: sections 1.1 through 1.9

2	Inference on two population means: two independent samples and paired samples	Chapter 1: Sections 1.10-1.11.
3	Linear regression models, Parameter estimation, Least square estimates	Chapter 2 and 3
4	Multiple linear regression: model diagnostics. Residual Analysis and Categorical variables	Chapter 4: sections 4.1 through 4.8, 4.14.
5	Polynomial models. Computing predictions and prediction errors. Multicollinearity and influential observations.	Chapters 8: sections 8.1 to 8.5. Chapter 5: sections 5.1, 5.2, 5.3, 5.7. Chapter 6.
6	Midterm Exam	
7	Model building and variable selection methods. Building more complex models: non-linear regression, polynomial regression.	Chapter 5 and section 7.7
8	Logistic regression and predictive models for qualitative variables	Chapter 9: sections 9.5,9.6
9	More about logistic regression models. Analysis of variance using regression models.	Chapter 9: sections 9.5,9.6. Chapter 12.
10	Overview of other predictive models (e.g. Poisson models, etc..)	
11	Final project due	