

CSC423 Data analysis and Regression Spring 2013/14

Instructor

Dr. Raffaella Settimi
Office: CST 716 - CDM Building,
Email: rsettimi@cdm.depaul.edu, Tel. (312) 362-5556.
Skype-ID: raffasw

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

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

Summary of Course

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

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.

Prerequisites

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.

Grading Policy

Course grade is based on the following components:

Homework and Programming assignments (40%). There will be four or five assignments. Assignments will be typically posted on Monday and will be due a week later by Tuesday at midnight.

Assignments submitted 5 days after the due date will not be accepted. Notice that a 20% point penalty will be applied for late homework. Extensions may only be granted for exceptional reasons.

Requests for extensions must be received BY EMAIL before the due date.

Group project (30%). Final project is due at the end of the quarter. Details will be provided later in class. Online students have the option to complete the project by themselves - however online students are encouraged to form teams and work together with inclass students.

Exam (30%) In class - tentatively scheduled on Monday June 9th, 2014. The exam will be open notes. Students should also have a calculator.

Students receiving more than 90% of possible points are guaranteed at least an A-, more than 80% at least a B-, more than 70% at least a C-, and more than 60% at least a D.

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.

ALTERNATIVE BOOK for students interested in a more mathematical/theoretical approach: Introduction to Linear Regression Analysis, 5th ed., by D. Montgomery, E. Peck, G. Vining. Wiley and Sons, 2012. ISBN: 978-0-470-54281-1 (available through DePaul at the Safari Books online website: <http://proquestcombo.safaribooksonline.com/book/math-and-science/9780470542811>)

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 716 of CDM building) during contact hours, or by appointment at other times.

The best way to contact me is through email at rsettimi@cdm.depaul.edu. Most emails will be answered within 24 hours. You can also call me in my office at 312 3625556.

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 at <https://col.cdm.depaul.edu>
- 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 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.

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://col.cdm.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.

Students are encouraged to contact the instructor through email at rsettimi@cdm.depaul.edu, phone (312 3625556) or skype (skype id: raffasw). Most emails will be answered within 24 hours.

Online students must **schedule their final exam** at the course website (<https://d2l.depaul.edu>) during the time frame specified by the instructor. Students living within the Chicago land area are considered *local* and are expected to take their exams at a DePaul University campus. Time slots vary by campus and day. Online students have the option to take the exam with the other in-class students at the official exam time. Online students living outside the Chicago land area (*remote*) will have their exams administered by a qualified proctor. You will need to find an acceptable proctor in your area before you register for your exam. Detailed information about online exams is at <http://blogs.cti.depaul.edu/colwiki/Wiki%20Pages/How%20Do%20I%20Take%20My%20Exams.aspx>

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. Rather than emailing questions, I encourage you to post your questions on the discussion forum at

<http://d2l.depaul.edu>

I will post course announcement 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>.

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. | Chapter 4: sections 4.1 through |

| | | |
|----|--|--|
| | Residual Analysis and Categorical variables | 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 | Model building and variable selection methods. Building more complex models: non-linear regression, polynomial regression. | Chapter 5 and section 7.7 |
| 7 | Logistic regression and predictive models for qualitative variables | Chapter 9: sections 9.5,9.6 |
| 8 | More about logistic regression models. Analysis of variance using regression models. | Chapter 9: sections 9.5,9.6. Chapter 12. |
| 9 | No class | |
| 10 | Overview of other predictive models (e.g. Poisson models, etc..) | |
| 11 | Final exam | |