2017:01:220:212:02 Econ Data Analytics:
Introduction to Data Management, Statistics and Regression

Fall 2017

Oscar Torres-Reyna

oscar.torresreyna@rutgers.edu

Department of Economics, Rutgers University, New Brunswick

21364   CAC AB-4225   Tuesday/Thursday 8:10 am – 9:30 am

Data analytics is a collection of methods and principles that allows us to learn from data. This course will focus on analyzing quantifiable information. The goal is to introduce basic elements of data collection, cleaning and preparation, as well as statistical principles, methods, and programing for data analysis. This will provide the foundation on which students can reason and reach valid conclusions when conducting scientific analysis. Through the application of statistical methodology, students will be able to analyze and critically assess the flow of information in their field.

One important aspect of data analytics is the use of software. This course will introduce the use of *Excel* and *R,* which are widely available. *Excel* is part of Microsoft Office, and *R* is a programing language used for statistical analysis. *R* is open-source (i.e. free), and available at http://cran.r-project.org/. The learning curve for R is quite steep, students may find *RStudio* a more user-friendly interface (you are still required to install *R*). Free for academic purposes, *RStudio* is an integrated development environment (IDE) designed for *R*, it is available in the following link: http://www.rstudio.com/.

Among the many online resources, introductory tutorials on using Excel, R (and other software) for data analysis can be found here http://dss.princeton.edu/training/ and here http://libguides.princeton.edu/dss. At Rutgers, you will find support at http://libguides.rutgers.edu/data. Please note, this is not a software or programming class.

Textbook

Key statistical ideas and formulas presented in the lectures come from the following book:

- Johnson, Richard A., Gouri K. Bhattacharyya, *Statistics. Principles and Methods*, 7th edition, John Wiley & Sons, 2014.
- [Complement] Salkind, Neil J., *Statistics for People Who (Think They) Hate Statistics*, 4th ed, 2017.

Logistics and grade scale

Presentations and other class materials will be made available through SAKAI or email as needed. The syllabus on SAKAI will be the most current one and will supersede any previous versions. Office hours after class as needed, also available to questions over email at any time. The grade scale used for the class is 90-100 = A / 85-89.99 = B+ / 80-84.99 = B / 75-79.99 = C+ / 60-74.99 = C / 50-59.99 = D / 0-49.99 = F.

Course requirements

During the first couple of weeks, each student will be assigned a country. Grading will be as follows:

Type	Topics	Worth	Given	Due
Assignments	Topics covered during class	10%	At random	On document
Report 1	Topics covered in chapters 1-3 (assigned country)	30%	Oct 10	Oct 17
Report 2	Topics covered in chapters 11, 12, and 14 (assigned country)	40%	Nov 7	Nov 14
Final-Test	Conceptual/procedural cumulative test (multiple versions)	20%	December 19	

Reports 1 and 2 are take-home exams on specific topics using data from the assigned country. The goal is to provide hands-on experience in data analysis using real world data. For the reports, you may work with other students, however, answers must be submitted *individually* and must show your work. All reports must be submitted in hard copy on the due date. Late submissions will be penalized with five percentage points per day after the due date. For example, if you get

90% in one report but was handed in one day late, the grade will go down to 85%. Late submissions without penalty allowed only in cases of documented health or medical emergency. Missing reports will get a zero grade.

The criteria for grading the reports will be based on how well the student explained the steps to get to the answer (as brief as possible), including formulas, statistical reasoning, etc. A key component in grading the reports is good *presentation*, which include but not limited to a clear identification of answers, and whether the reasoning in answering the question is logical and well specified. Answers *must* be posted at the front of your report along with your *name* and the *country* assigned (see template in SAKAI). There will be a penalty of five points if it is hard to identify your answers.

Note that there is no mandatory requirement to use Excel or R/RStudio to work on the reports, but you have to use some software to work on them (reports cannot be done with a calculator). Students are free to use whatever software they want, however only Excel or R/RStudio will be supported during the class.

The final-test is an examination covering major concepts, formulas, and procedures presented in the slides. There are no make-ups for this exam, no "early birds" allowed, please plan to be available on the week the test is given. Students should review the notes provided and the corresponding contents from the textbook. Be aware, there is going to be *at least* two versions of the test.

Random assignments will cover topics from previous sessions, and will be graded as pass (1 point) or fail (0 points). Due date will be specified on the document. Late assignments will not be accepted. A passing grade will be one that not only has the right answer but also shows the work. A passing grade will be one that even though the answer was not correct it clearly describes the work and shows the effort. A failing grade will be one that has the right answer but does not show the work. A failing grade will be one that has the wrong answer and the work was not clearly described.

Course plan

Topic	Chapter
Introduction	slides
Data in economics, examples	slides
Data collection, preparation, visual and descriptive statistics	1, 2
Bivariate regression and correlation	3
Simple linear regression	11
Multiple linear regression	12
Analysis of variance	14
Categorical data	13
Inferences from large samples	8
Inferences from small samples	9

Academic integrity - Get familiar with the university's policy on academic integrity, it will be enforced in this class: http://academicintegrity.rutgers.edu/academic-integrity-policy/

Absence reporting - It is expected that students attend all sessions. However, if are going to miss class(es), please use the University absence reporting website https://sims.rutgers.edu/ssra/ to indicate the date and reason for your absence. Reporting your absence does not excuse from your course responsibilities.

Accommodations - If you need special accommodation due to disability, check the procedures and guidelines set by the Office of Disability Services: https://ods.rutgers.edu.

Useful links related to student wellness:

- Office for Violence Prevention & Victim Assistance (VPVA): http://vpva.rutgers.edu/
- Counseling, ADAP & Psychiatric Services (CAPS): http://rhscaps.rutgers.edu/
- *Just In Case* web app, suicide prevention hotline for students: https://goo.gl/BN9Vb9