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

Oscar Torres-Reyna
oscar.torresreyna@rutgers.edu
Department of Economics, Rutgers University, New Brunswick

Spring 2018 19996 CAC AB-4450 Monday 6:10 pm – 9:00 pm

Data analytics is a collection of methods and principles that allow us to extract meaning 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. Through the application of statistical methodology, students will be able to analyze, to critically assess, and to reach valid conclusions when dealing with data.

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 https://dss.princeton.edu/training/ and here http://libguides.princeton.edu/dss. At Rutgers, you may find support at http://libguides.rutgers.edu/data. Please note, this is not a software or a 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. I will be using Sakai's mailtool to communicate with the class, make sure you receive those emails. The syllabus in Sakai will be the most current one and it will supersede any previous versions. Office hours after class as needed, also available for 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. Please note that the grade scale is rounded to two decimals. Bringing a laptop is not mandatory but it is highly recommended.

Course requirements

Each student will be randomly assigned a country. Grading will be as follows:

Type	Topics	Worth	Given	Due
In-class work	Topics covered in previous sessions/case studies/activities	25%	In-class	In-class
Report 1	Topics covered in chapters 1-3, 8, 9 (assigned country)	30%	TBD	TBD
Report 2	Topics covered in chapters 11-14 (assigned country)	45%	TBD	TBD
Final Exam	Conceptual/procedural cumulative test (multiple versions)	25%	TBD	(optional)

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 in groups, however, answers must be submitted *individually* and must show your work. Late submissions will be penalized with five percentage points per day after the due date (i.e. if you get 90% in one report but was submitted 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. Any submission after answers are posted in Sakai will get a zero grade.

In grading the reports presentation plays an important role, in particular, in explaining the steps leading to the answer (including formulas, statistical reasoning, software used, and conclusions). 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.

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

Reports must be typed in a computer (not handwritten). If for any reason a student does not have access to a computer, Rutgers offers several computing labs, see here https://oit-nb.rutgers.edu/computing-lab-locations. Note that there is no mandatory requirement to use Excel or R/RStudio to work on the reports, but you must 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 class.

In-class exercises will be graded as pass(1)/fail(0). A passing grade will be given to those who *show or describe the work done* leading to the answer. A failing grade will be given to those who do not show the work done or show some gibberish to justify an answer. Exercises will not be returned unless clarification is requested. Make-up sessions for exercises might be scheduled later in the semester (grading for these will be stricter). The exercises will be posted in Sakai before class starts. Students will have the first 15-30 minutes of the class to do the exercise on their own, class will continue right after (starting with a discussion of the previous' week exercise). While students must work in groups, class work must be submitted individually at the end of the session. There might be more than one activity during class, all work submitted by the end of the class will be part of the grade for that day. Students must write their names and the table number in which the sat that day.

There are some rules for sitting. A student *must* sit in a different table each week during the following ten weeks. During those ten weeks, every face a student sees in each table *must* also be different. After the tenth week, students are free to sit wherever they want (starting April 9). Given the size of the class it may be difficult to enforce these rules but those caught not following them will face a penalty of one-point deduction per incident in the final grade. A table must have between 2 and 9 students.

The final exam is an optional in-class examination covering major concepts, formulas, and procedures presented in the slides for the entire semester. The final exam can only be used to replace the grade from the in-class work. If the student takes the final test, he/she will forfeit the grade for the in-class work. Once the student notifies his/her decision to take the final exam there is no going back. 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. Be aware, there is going to be *at least* two versions of the test.

Course plan

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

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