

**Econ Data Analytics: Introduction to Data Management, Statistics and Regression**  
**Oscar Torres-Reyna (ot50@scarletmail.rutgers.edu)**  
**Department of Economics, Rutgers University, New Brunswick**  
**Spring 2023      01/17-05/10      04862 01:220:212:01**  
**Classroom AB-4400    Tuesdays 5:40 pm – 8:40 pm    Office hours: after class as needed**

Data analytics is a collection of methods and principles to extract meaning from data. The goal of this course is to introduce basic elements of data collection, cleaning and preparation, data visualization, statistical principles, methods, and programming for data analysis. It is expected that students will be able to analyze, to critically assess, and to reach valid conclusions when dealing with real-world data.

This course will have a strong focus on software application. It will introduce the use of *Excel* and *R* for data analysis. *Excel* is part of Microsoft Office (install latest version, see <https://oit.rutgers.edu/microsoft-office-students>). *R* is open-source and available at <http://cran.r-project.org/>. Free for academic purposes, *RStudio* is a user-friendly interface designed for *R*, it is available at the following link: <http://www.rstudio.com/>

**Textbook:** Key statistical ideas and formulas presented in the lectures come from the following books (optional):

- Moore, David, George McCabe, and Bruce Craig, *Introduction to the Practice of Statistics*, 10th edition, W. H. Freeman. Any other edition works as well. [MMC] This will be the main book.
- Johnson, Richard A., Gouri K. Bhattacharyya, *Principles and Methods*, 7<sup>th</sup> edition, John Wiley & Sons, 2014. ISBN-13: 978-0470904114, ISBN-10: 8126557745. This book is a good reference on statistical analysis. Any other edition works as well. [JB]
- Salkind, Neil J., *Statistics for People Who (Think They) Hate Statistics*, 4th ed, 2017. ISBN-13: 978-1412979597, ISBN-10: 1412979595. This book is a good reference on how to use Excel for statistical analysis. [S]

**Logistics**

- It is a requirement for this class to have a functioning computer with the required software properly installed (see links in the intro above), a good WiFi connection, and full access to the course in Canvas. If you are running into tech issues, contact OIT at [help@oit.rutgers.edu](mailto:help@oit.rutgers.edu).
- Attendance is mandatory.
- For class activities you will need to access websites to extract data. Make sure to have Firefox, Chrome, or Microsoft Edge installed in your laptop. Safari is not recommended for data extraction.
- Grading will be through weekly activities directly connected to the lectures. Each activity is meant to simulate research projects where students will apply quantitative techniques to analyze real-world data. Activities are worth a few points that will accumulate through the semester up to a final grade (see the grade scale and course plan below).
- All class lectures and other materials will be made available through Canvas.
- Most of class materials will be in the form of PDF, ASCII, and Excel documents. A collection of supplementary videos will be available on key data analysis procedures. If you have questions, please stop after class, or send me an email.
- I will be using Canvas's email tool to communicate with the class, make sure you receive those emails.
- The syllabus in Canvas will be the most current and it will supersede any previous versions.

**Grade scale in points**

Grade	Minimum # points needed
A	266
B+	240
B	214
C+	188
C	162
D	149
F	0

Rounding criteria: half point goes up to the next integer, for example 3.50 goes to 4, 3.49 goes to 3. This applies only to the calculation of the final grade.

**Course plan:** The first two weeks we will go over lessons 1-4, no graded activities will be given during this time. Weekly in-class activities will start the third week.

The grading in this course will be based on points which will accumulate each week up to a final total of 300 which will determine your final grade according to the grade scale shown above. The content of the activities is cumulative, questions could be related to the same week's lecture and/or to material from previous weeks. Activities are modeled after content from the lecture slides. Activities should be considered an extension to the lectures.

Each in-class activity is broken into two parts: 1) reproduce the main components of data analysis: collecting, reading, cleaning, and preparing data; and 2) a Canvas quiz with multiple-choice or calculation questions requiring the application of statistical methods and/or about concepts shown in class (checkpoint questions). During class, students will have full support from the instructor and will clarify any questions related to the activity except those leading to a specific answer. Students not able to finish the activity in class, will have until Friday, of the same week, no later than 11:59 pm to submit their answers. Further questions about the activity can be sent over email no later than Thursday 10 pm, activity related questions sent after this time will not get a reply. Notice that activities are not timed but they are due by the deadline. There are no midterm or final exams.

Students have one week to clarify and/or verify their grade for an activity on a specific week, after then there will be no further grade revisions for any previous activity.

**Missing work:** Because of the way the class is structured, there are no extensions under any circumstances. There is, however, the possibility of recovering some missing points over a  $t$  number of weeks according to the following formula:

$$Points\_Week_{t+1} = Points\_Week_{t+1} + \alpha \left( \beta \sum_1^t Missed\_Points \right)$$

Where:

$\alpha$  = the proportion of points obtained in the activity on week  $t+1$ .

$\beta = 1 - (n * 0.10)$ ;  $n$  = number of missing activities

Let's say that you missed three activities,  $n = 3$ , making  $\beta = 0.7$ . Let's also assume that the total points for those three missed activities is 50 points. On week 4 ( $t+1$ ), you did the activity and got 16 out of 20 points for that week, making  $\alpha = 16/20 = 0.8$ . The total number of points for week  $t+1$  would be  $16 + 0.8(0.7*50) = 16 + 0.8*35 = 16 + 28 = 44$ .

Please note that you have the option to apply this formula towards the end of the semester in the last activity to account for any number of missing activities (this is the default option). You also have the choice of using this procedure at any point during the semester up to two times (you will need to request the recovery option via email *before* taking the  $t+1$  activity, otherwise it will not be applied for that week). The use of this procedure a third time will require a letter from your dean or the proper academic authority speaking on your behalf as per university policies regarding excused work

**Student code of conduct:** Students can work, discuss, brainstorm with other students in the class but must submit answers individually through Canvas. By taking each activity, you accept the following honor pledge:

*On my honor, I have neither received nor given any unauthorized assistance on this assignment.*

**Course completion requirements:** Students are encouraged to review and to reproduce the examples presented in the lecture slides. This will help to have a better understanding of the procedures shown and to generate questions that may clarify further doubts. Reviewing lecture material will help with the activities which are intended to reinforce the content of the lectures. The goals of the activities are to help students to develop their analytic and data management skills, to focus, to prioritize and to develop collaboration skills. If you are struggling with the material, find the work too challenging, or have questions about the course material, please send me an email. I would strongly suggest doing this earlier rather than later.

**Attendance bonus:** All students start out with an attendance bonus worth five points. In each lecture, we will check the presence of ten randomly selected students. You can miss two sessions without losing the bonus, if you notified me by email *before* the beginning of class (= email time stamp before 5:40 pm). No need to explain the reason for your absence. You cannot send more than two such emails during the entire semester.

**Course Outline:** For practical reasons, the lecture notes will follow the book by Moore, McCabe, and Craig. Students wanting more in-depth information about each statistical topic are welcome to read the chapter/sections indicated below:

Lesson 1: "?"

Lesson 2: The Tools, The Example [Introducing Excel/R]

Lesson 3: It is about randomness [MMC 4.1, 4.2, 4.5]

Lesson 4: Where the true value lies [MMC 4.3, 4.4]

Lesson 5: Describing data [MMC 1.1, 1.2, 1.3, 4.4]

Lesson 6: Visualizing data [MMC 2.1, 2.2]

Lesson 7: Correlation: beware [MMC 2.3, 2.5, 2.7]

Lesson 8: Intro to linear regression (part 1) [MMC 2.4]

Lesson 9: Inference in large samples [MMC 6.1, 6.2, 6.3, 6.4]

Lesson 10: Inference in small samples [MMC 7.1, 7.2]

Lesson 11: Inference for Proportions [MMC 5.3, 8.1, 8.2]

Lesson 12: Analysis of Variance [MMC 12.1]

Lesson 13: Intro to linear regression (bivariate, part 2) [MMC 10.1]

Lesson 14: Intro to linear regression (multiple, part 3) [MMC 11.1]

Lesson 15: Intro to linear regression (categorical/interactions/LPM, part 4, lecture notes)

Lesson 16: Categorical Data [MMC 9.1, 9.2]

Lesson 17: Collecting and preparing data [MMC 3.1, 3.2, 3.3]

Lesson 18: "?+"

---

## Campus resources

**Help with data sources and statistical software:** <https://libguides.rutgers.edu/data>

**Need a space to study on campus** - See here <https://scheduling.rutgers.edu/ecosystem-for-learning/study-spaces>

**Peer group tutoring** - See here <https://rlc.rutgers.edu/student-services/peer-tutoring>

**IT help and support** - See here <https://it.rutgers.edu/help-support>

**Academic integrity** - Get familiar with the university's policy on academic integrity, it will be enforced in this class:

<https://nbprovost.rutgers.edu/academic-integrity-students>

**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>
- Crisis Intervention: <http://health.rutgers.edu/medical-counseling-services/counseling/crisis-intervention>