

Analítica de Datos

Syllabus 2025 - I

Pontificia Universidad Javeriana

Clase

- 🕒 Lunes 4:10pm–6:45pm
- 👤 Carlos Cardona Andrade
- ✉ carlos.cardonaa@javeriana.edu.co
- https://github.com/ccardonaandrade/analitica_2025
- 📖 OpenIntro Statistics, 4th ed.
- 📖 Introduction to Modern Statistics, 2th ed.
- 📖 R for Data Science, 2th ed.
- 📖 Data Visualization: A practical introduction

Resumen

Descripción: La gestión empresarial moderna exige una toma de decisiones basada en datos, superando las limitaciones de métodos puramente intuitivos o experienciales. Este curso introduce la estadística como herramienta fundamental para transformar datos en decisiones empresariales efectivas. Los estudiantes desarrollarán competencias en:

- Recopilación y procesamiento sistemático de datos
- Análisis exploratorio y técnicas de inferencia estadística
- Modelización e interpretación de resultados
- Evaluación crítica de argumentos basados en datos

El curso enfatiza la aplicación práctica de la estadística en contextos empresariales, investigación científica y políticas públicas. Como señaló S. Wilks (1950), 'El pensamiento estadístico será una cualificación tan esencial para la ciudadanía eficiente como la capacidad de leer y escribir'

El aprendizaje de la programación estadística es inherente a la práctica del análisis de datos. Por ello, a lo largo de este curso también enseñaremos el lenguaje de programación estadística R.

Prerrequisitos: Este curso no asume ningún conocimiento previo en estadística o programación. Además, sólo se requiere aritmética básica.

Recomendaciones

1. **Sean amable.**
2. **Asuman la responsabilidad** de su propia educación e intenten **aprender** tanto como puedan.
3. **Hagan tu propio trabajo.**
4. Desarrollen su **intuición**—*por ejemplo*, ¿por qué la regresión funciona en una situación y falla en otra?
5. **Aprendan R.** Luchen mientras lo intentas—y utilicen **Google** para resolver sus dudas.
6. Asiste a las **horas de oficina**.¹
7. **Pidan ayuda temprano**—no esperen hasta el final del semestre.

Software

R: Usaremos el lenguaje de programación estadística **R**, y utilizaremos **RStudio** para interactuar con R.

Aprender R: requerirá tiempo y esfuerzo, pero es una herramienta poderosa y versátil valorada por muchos empleadores. Dediquen el esfuerzo y tiempo necesarios, y serán recompensado. Los computadores en la universidad ya tienen instalado R, pero recomiendo enfáticamente que instalen R y RStudio en su propio computador.

Si están preocupados por aprender R—o quieren aprender más rápido—les sugiero que consulten los siguientes recursos en línea, gratuitos.

- [Introducción a R de DataCamp](#)
- [R Bootcamp de TeamLeada](#)
- [Guía para principiantes en R de Computerworld](#)

El equipo de RStudio ha preparado un excelente [material de aprendizaje](#).

Quices, talleres y parciales

Lab: This course includes a lab, which is integral to learning the material in (and passing) this course. For now, we are requesting that you attend the lab for which you registered. The lab includes both general econometrics instruction and computing tips necessary to complete the homework assignments—linking the lecture material to R—as well as topics which the lecture may not be cover. **The lab is the best way you can get quick feedback and help in this course.** The GEs will also post a video for you to watch before the remote lab meeting/call.

See above for lab times.

Problem Sets

- You will **turn in assignments online via Canvas**.
- Assignments will be due approximately every 1–2 weeks.
- Assignments **must be in your own words. Do not copy.**

¹Dos artículos relacionados de NPR sobre horas de oficina: [Estudiantes universitarios: Cómo hacer que las horas de oficina sean menos intimidantes](#) y [Descubriendo un gran misterio de la universidad: Las horas de oficina](#).

- See below for **late policy**.

Feel free to work together on the assignments. Unless explicitly stated, **each student is required to write and submit independent answers**. This means that word-for-word copies will not be accepted and will be viewed as academic dishonesty. In other words: You must place answers **in your own words**. **Copying from other people (even if you worked with them) or from previous assignments is considered cheating**.

Late policy

- We accept assignments **up to 48 hours late**, but we **subtract 2 percentage points for each hour it is late**.
- For example, you turn in an assignment 12 hours late and would have received 85%. We subtract $12 \times 2 = 24$ percentage points, meaning you will receive $85\% - 24\% = 61\%$.
- No exceptions.

Exams

- We will give the **“in-class” midterm online on February 11, 2021 from 10:15am–11:45am**.
- We will give the **final exam on Tuesday, March 16, 2021 from 8am–10am**.

We will not offer early exams. Each exam will be accompanied by a more open-ended project.

Evaluación

Las notas para esta clase se asignarán en función de las siguientes tareas: tareas aproximadamente quince-nales, dos exámenes parciales, y un proyecto final. La nota final se determinará según los siguientes pesos para cada una de las actividades:

Quices y Talleres	30%
Parcial I	15%
Parcial II	15%
Proyecto Final	40%

Libros y otras lecturas

Este curso busca destacar la amplia gama de recursos educativos gratuitos disponibles en línea y fomentar una mentalidad proactiva y autodidacta, esencial para el aprendizaje efectivo de la programación.

Libros de Estadística: Hay dos libros de estadística recomendados para este curso:

1. **OpenIntro Statistics** de Diez, Cetinkaya-Rundel y Barr (**OIS**)
2. **Introduction to Modern Statistics, 2th ed.** de Cetinkaya-Rundel y Hardin (**IME**)

Ambos libros los encuentran gratis en línea. Estos textos ofrecen otra perspectiva complementaria sobre el material que cubrimos en las clases. El cronograma del curso (más abajo) incluye lecturas sugeridas para cada tema. Recomiendo que lean las lecturas asignadas previo a cada clase.

Libros de R: Para aprender R, seguiremos los siguientes textos que también son gratuitos:

1. **R for Data Science, 2th ed.** de Grolemund y Wickham (**R4DS**)
2. **Data Visualization: A practical introduction** de Healy (**DV**)

¿Quieren profundizar más? Consulten **Advanced R**.

Cronograma tentativo del curso

La tabla a continuación presenta el plan actual para el esquema del curso y las lecturas asignadas de los libros de texto. Ocasionalmente, asignaremos artículos para que los leas para la clase, el laboratorio o como parte de tus tareas. Publicaré estos artículos en Canvas. Como sugiere el título de esta sección, el cronograma y los temas pueden estar sujetos a cambios.

Cronograma tentativo

Clase	Fecha	Tema	Lectura sugerida
01	27/01	Intro & Preliminares	ItE 1-6
02	03/02	Tidyverse & ggplot	ItE 1-6; MM 2
03	10/02	Explorando datos categóricos	ItE 1-6; MM 2
04	17/02	Explorando datos numéricos	ItE 1-7
05	24/02	Intro a la Estadística inferencial	ItE 7
06	03/03	Inferencia para datos categóricos	ItE 7
07	10/03	Inferencia para datos numéricos	
08	17/03	Parcial I	ItE pp. 68-75
	24/03	Festivo	ItE 11
09	31/03	Muestreo y Encuestas	ItE 11
10	07/04	Experimentos	ItE 12
	14/04	Semana Santa	
11	21/04	Intro a Regresión Lineal	ItE 12 & 13
12	28/04	Parcial II	ItE pp. 68-75
13	05/05	Más de Regresión Lineal	ItE 9; MM 3
14	12/05	Regresión Logística	ItE 9; MM 3
15	19/05	Análisis de Texto	ItE 9; MM 3
16	26/05	Presentación Trabajo Final	ItE pp. 68-75