



FACULTAD  
DE CIENCIAS  
ECONÓMICAS

**FAMAF**

Facultad de Matemática, Astronomía,  
Física y Computación



UNC

Universidad  
Nacional  
de Córdoba

**DIPLOMATURA**

**CIENCIA DE DATOS, INTELIGENCIA  
ARTIFICIAL Y SUS APLICACIONES  
EN ECONOMÍA Y NEGOCIOS**



## Docentes:

Dr. Martín Saino - Lic. María Inés Ahumada

## Comunicación:



**DiploDatosEco2021**

[diplodatoseco2021.slack.com](https://diplodatoseco2021.slack.com)

[# estadística-probabilidad](#)

## Materiales y actividades:



Diplomatura en Ciencia de Datos - 2021

[Página Principal](#) / [Mis cursos](#) / [Diplomatura en Ciencia de Datos\\_21](#) / [MÓDULO 2. Unidad 1](#)



## MODULO 2: Métodos cuantitativos en la ciencia de datos

### **FINALIDAD:**

Plantear y abordar paso a paso las distintas herramientas estadísticas y su instrumentación en la resolución de problemas económicos y de negocios.

Sus contenidos se distribuyen en tres unidades:

- **Herramientas exploratorias y probabilidad para el análisis de datos**
- Herramientas multivariadas
- Modelos econométricos

### ***Unidad 1. Herramientas de estadística y probabilidad para análisis***

- Análisis Exploratorio de datos
- Probabilidad y modelización
- Distribución en el muestreo
- Inferencia estadística. Estimación y pruebas de hipótesis
- Modelos de regresión

**Duración:** veinte (20) horas. Encuentros: 30/7 31/7 06/8 07/8 y 13/8



**Encuentros sincrónicos**

**Actividades**

**Evaluación**



# M2-U1: Herramientas exploratorias y probabilidad para el análisis de datos

## Librerías

### [Scipy.stats](#)

- Numpy
- Matplotlib
- Pandas

### [Statmodels](#)

- Seaborn

### [Pandas.profilng](#)

Para visualización interactiva: Bokeh - intertools



## M2-U1: Herramientas exploratorias y probabilidad para el análisis de datos

### Bibliografía

#### BÁSICA

- Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2019). Estadística para negocios y economía. Decimotercera edición. México: Cengage Learning.
- Materiales de Cátedra. [Disponible en el aula virtual del Curso](#).
- Rincón, Luis. Introducción a la Probabilidad (2014) . Departamento de Matemáticas. Facultad de Ciencias UNAM - Circuito Exterior de CU 04510 México CDMX.  
<https://lya.fciencias.unam.mx/lars/indexL.html>
- The Python Data Science Handbook by Jake VanderPlas (O'Reilly). Copyright 2016 Jake VanderPlas, 978-1-491-91205-8.  
<https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

#### ADICIONAL:

- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media [A](#)

# Bibliografía Complementaria

- **Capítulos 2 y 3** de VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data* (1st ed.). O'Reilly Media. Acceso gratuito: <https://jakevdp.github.io/PythonDataScienceHandbook/>
- **Capítulo 2** de Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.







# Repasemos las etapas del EDA

## ¿Qué es el EDA?

**“Exploratory Data Analysis”** es una actitud, un estado de flexibilidad, una voluntad de buscar aquellas cosas que creemos que no están allí y también aquellas que creemos que sí están.

- John Tukey (1977)

- Tipos de datos
- Espacio de la solución (“target”)
- Espacio de las variables (“features”)
- Variables categóricas
- Variables numéricas
- Relación entre solución y variables
- Relación entre variables

- ☐ Valores únicos
- ☐ Valores perdidos (“Missing”)
- ☐ Valores cero
- ☐ Rangos válidos de valores
- ☐ Outliers
- ☐ Distribución de valores (media, mediana, desviación est., percentiles, máx-min, sesgo)

