

# Análisis de datos para administradores de negocios

NICOLÁS GARCÍA PEÑALOZA

Curso con enfoque empresarial  
**Administración de negocios**

2025

# Unidad

- 1 Primer unidad
- 2 Segunda unidad
- 3 Tercera unidad

## ► Introducción.

# Introducción:

- 1 ¿Cuáles y cómo utilizar los paquetes estadísticos?
- 2 ¿Por qué es tan importante la estadística?
- 3 Nociones elementales en estadística para el manejo de los datos.
- 4 ¿Por qué es importante la visualización de los datos?
- 5 EDA.
- 6 Storytelling con datos.
- 7 Metodologías.
- 8 Presentación de informes tipo científicos.

# Conozcamos los programas

Programa	Creadores/Año	Descripción
<u>R</u>	Ross Ihaka y Robert Gentleman/1996	Es un lenguaje y un entorno para la computación estadística y los gráficos.
<u>Python</u>	Guido van Rossum/1989	Es un lenguaje de programación interpretado, orientado a objetos, de alto nivel y semántica dinámica.
<u>Qgis</u>	Gary Sherman/2002	Proporciona un visor de datos geográficos rápido con creciente gama de capacidades a través de sus funciones básicas y complementos.

**Fuente:** Elaboración propia

# La importancia de la estadística

## La estadística es la gramática de la ciencia

La Estadística nos permite tomar decisiones informadas y realizar juicios inteligentes a pesar de la presencia de incertidumbre y variación. Esta disciplina ofrece potentes métodos para obtener información en una amplia gama de campos, como los negocios, la medicina, la agricultura, las ciencias sociales y la ingeniería.

**Las empresas dependen de decisiones basadas en datos para optimizar procesos, reducir riesgos y mejorar resultados.**

- Identificación de tendencias
- Optimización de recursos
- Soporte a la investigación
- Evaluación de riesgos

El análisis de datos lo podemos dividir de la siguiente manera:

- Análisis estadístico **Descriptivo**.
- Análisis estadístico **Predictivo**.

Hay otra ventana que podemos encontrar que es el **Generativo**.

Analítica I

Analítica II

# Donde todo empieza

## Regresión lineal estimación mediante mínimos cuadrados ordinarios -MCO-

En 1886 Francis Galton planteó que, a pesar de la tendencia de los padres de estatura alta a procrear hijos altos y los padres de estatura baja, hijos bajos, la estatura promedio de los niños de padres de una estatura determinada tendía a desplazarse, o “regresar”, a la estatura promedio de la población total. Es decir, la estatura de los hijos de padres altos o bajos tiende a dirigirse a la estatura promedio de la población.

- $y = \beta x + \varepsilon$  Ecuación Poblacional

- $\hat{y} = \hat{\beta}x + \varepsilon$  Ecuación Estimada

Hoy el análisis de regresión se basa en el estudio de la dependencia de una variable (**variable dependiente**) respecto de una o más variables (**variables explicativas**) con el objetivo de estimar o predecir la media o valor promedio poblacional



# Diferencia de los modelos con estadística Paramétrica y no-paramétrica

## Estadística clásica vs el teorema de Bayes

Modelos Paramétricos (Enfoque Frecuentista)	Modelos no-paramétricos (Enfoque Bayesiano)
Asumen una distribución	No asumen distribución específica
Fija y conocida	No fija, crece con los datos
Menos flexible, pero más interpretable	Más flexible, pero menos interpretable
Menos datos	Más datos
Máxima verosimilitud (MLE)	Teorema de Bayes

**Fuente:** Elaboración propia

# La elaboración de experimentos

## La validez estadística en la medición

### La Metodología Estadística se emplea bajo:

- Definición del problema
- Definición de los objetivos
- Definición de las variables de interés
- Diseño del experimento\*
- Recolección de la información\*
- Procesamiento de los datos
- Análisis descriptivo o exploratorio de datos
- Inferencia estadística
- Conclusiones y recomendaciones

⇒ **Construir una medición debe tomar en cuenta dos principios validez y confiabilidad.**

# Valor de la información

## » Valor Interno.

- Al repetir la medición bajo las mismas condiciones/escenario ¿El resultado es el mismo o similar?

*Refleja la situación que se pretende medir.*

## » Valor Externo.

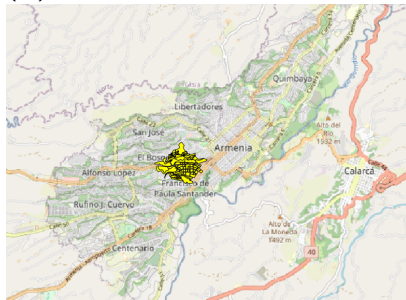
- Grado en el que la situación se puede generalizar a otras situaciones no medidas.

*Depende de cómo se conforma la muestra.*

Validez

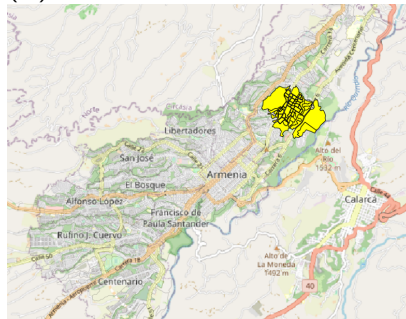
# ¿Cómo espero que sea la distribución del ingreso en este sector?

(A)



Sector del terminal de transportes de Armenia

(B)

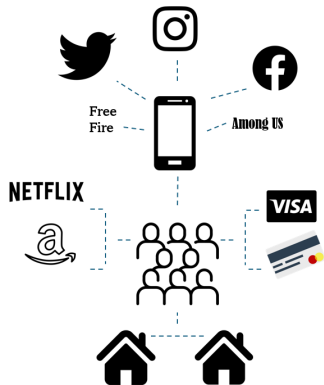


Sector de Providencia, Catellana y Avenida Centenario.

*Nota: Si un estudio no tiene validez interna, **no puede tener validez externa.***

*¿Puede tener valor mis conclusiones si tengo validez interna pero no externa?*

# El dato y las conclusiones. ¿Y la estadística?



## Estadística



- Hay correlación entre las crisis financieras y las actividades macroeconómicas influido en los responsables políticos. Y los bancos centrales han logrado jugar un papel de corrector de estas crisis.
  - Bernanke, Diamond and Dybvig Nobel 2022
- Las mujeres se están convirtiendo en más liberales y los hombres continúan su tendencia estable.
  - Financial Times

# Caso real...

## **Digitación de un censo de recicladores que se hizo en la ciudad de Cali.**

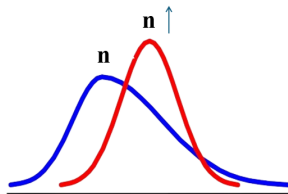
*Durante el piloto (siempre hagan piloto cuando van a recoger datos a través de encuestas) encontré algo inesperado en los datos. **Los ingresos mensuales de los recicladores eran mayores a los gastos.** Esto no tenía sentido dado lo que conocíamos.*

*Así que antes de empezar a sacar conclusiones, me fui como encuestadora a pilotear un par de encuestas adicionales. **Descubrí que cuando le preguntábamos por los gastos mensuales, los recicladores no pensaban en los gastos diarios, o tal vez sí pero no sabía como multiplicar ese valor diario por 30, o un valor semanal por 4 y por ende no lo incluían en su respuesta.***

*¿Qué hicimos? Recapacitamos a los encuestadores en este tema para que esta pregunta preguntaran por rubros más importantes (alimentación, vivienda, servicios y otros) haciendo énfasis en gastos diarios y documentando todos estos elementos.*

# Teorema Central del Límite

Si tenemos una variable aleatoria  $x$  independiente, lo que representa que  $x_1$  **no depende de la ocurrencia de  $x_n$** , un  $n$  **lo suficientemente grande** y una **varianza finita**, yo esperaría que  $\sum_1^n x_1 \sim N(n\mu, \sigma\sqrt{n})$



# Ingreso promedio de Armenia

Cuando se decide aleatorizar la muestra, la media estimada resultantes se espera sea igual a la media del total de la población. Entonces, se dice que la estimación es **insesgada**. Cada muestra aleatoria puede proporcionar una estimación diferente, lo que da lugar a cierta incertidumbre ( $\sigma^2$ ). En este contexto, **a medida que el tamaño de la muestra  $n$  aumenta, la estimación tiende a parecerse más al valor real de la población**, reduciendo así la incertidumbre asociada con la estimación.

