

# Análisis de datos para administradores de negocios

NICOLÁS GARCÍA PEÑALOZA

Curso con enfoque empresarial  
**Administración de negocios**

2025

## 1 Primer 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.

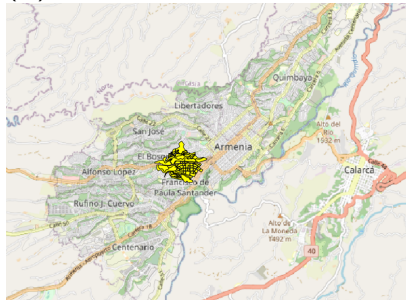
*Depende de cómo se conforma la muestra.*

Validez

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

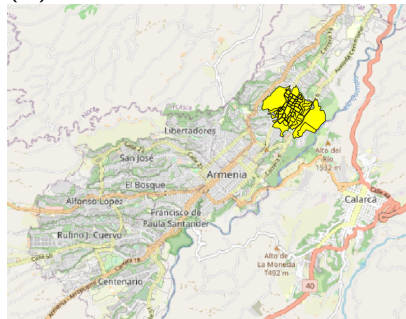
Quién nos dice la verdad

(A)



Sector del terminal de transportes de Armenia

(B)



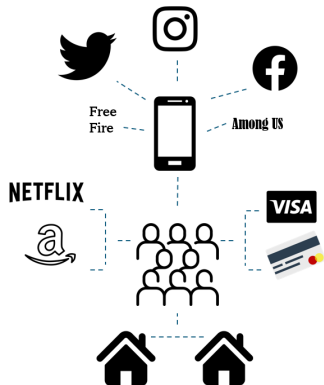
Sector de Providencia, Catellana y Avenida Centenario.

*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.**

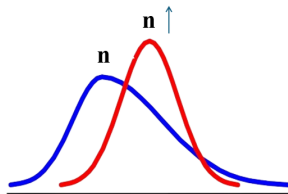
*Durante el piloto (siempre hagan piloto cuando van a recoger datos a través de encuestas) encontré algo inesperado en los datos. **Los ingresos mensuales de los recicladores eran mayores a los gastos.** Esto no tenía sentido dado lo que conocíamos.*

*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.*

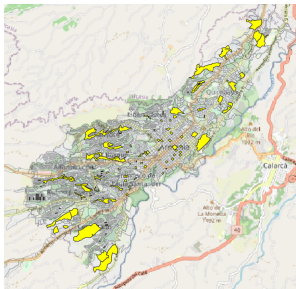
# Teorema Central del Límite

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.

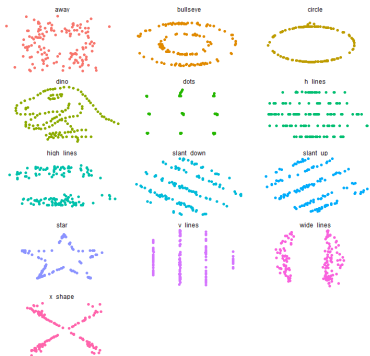




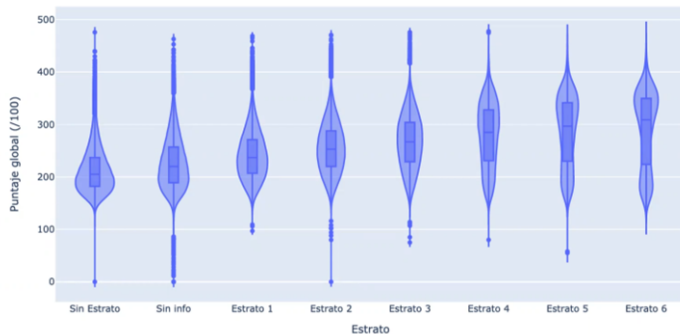
# Por qué es importante visualizar los datos:

## ¿Qué es el mapa de Snow?

Un estudio de la Universidad de Delaware publicado en **Quarterly Journal of Economics** presentó que «una adecuada representación gráfica tiene un impacto considerable en la forma en la que los resultados de una investigación son percibidos, validados y comunicados, tanto por los expertos como por el público general.»



*Ser pilo paga:* «El diseño del programa surgió de estudios previos que mostraban que cerca del **60% de los 27,000 puntajes más altos de la prueba Saber 11 de 2012 eran obtenidos por estudiantes de estrato 1, 2 y 3** y que, dos años después de la prueba de Estado, alrededor del 23% de los jóvenes no habían logrado acceder a una institución de educación superior (IES) por falta de recursos económicos.»»



# Errores en las presentaciones

## Factores de confusión

**Los problemas estadísticos no terminan con los resultados;** aunque puedan tener relación estadística, ¿representan realmente causalidad?

- Datos ocultos
- Escala
- Distribución
- Planteamiento erróneo del problema
- Representación errónea
- Contaminado por población

 Últimas Noticias Opinión Política Judicial Economía Mundo Bogotá Deportes Colombia + 20 Empleos

Suscribirse Iniciar Sesión

Home > Economía

## Comercio electrónico creció más del 100%

La cifra de compras a través de internet ascendió a US\$8.283 millones en 2013.

Redacción Negocios  
28 de enero de 2015 - 08:37 a. m.





La información **nos UNE** **UNETE**

La cifra de compras hechas en internet ascendió a los US\$8.283 millones en 2013. Aumentó compras hechas a través de móviles. Según el primer estudio de comercio electrónico en Colombia durante 2013 la cifra de compras hechas a través de plataformas digitales ascendió a US\$8.283 millones.

Los dispositivos por medio de los cuales se realizaron estas compras también tuvo un comportamiento diferente en comparación con 2012 pues mientras disminuyeron las transacciones hechas en desktop, pasando de 81% a 64%, las que se realizaron con dispositivos móviles aumentaron de 16% a 36%.



Salvo más  
Veloz a propósito



si se contenido exclusivo

Argentina Colombia España México Perú Mundo Juegos Olímpicos Últimas Noticias Entretenimiento Deportes Violencia Cultura

infobae

Colombia Últimas Noticias Violencia Economía Entretenimiento Newsletters

COLOMBIA >

**Desempleo en Colombia bajó de nuevo en diciembre: terminó el año en el 10%, según el Dane**

De enero a diciembre de 2023, la tasa global de desocupación bajó 1,1 puntos porcentuales respecto al mismo periodo de 2022

Por Luis Benito

21 Dic 2024 12:04 a.m. | Actualizado 21 Dic 2024 12:13 p.m. GMT



El desempleo en Colombia bajó en diciembre en comparación con el mismo mes de 2022 - crédito Colprensa.

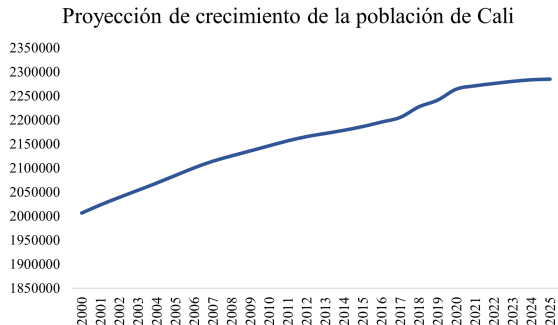
El Departamento Administrativo Nacional de Estadística, Dane, reveló que **el desempleo**

PUBLICADO  
Google ha cerrado el anuncio

# Israel Palestina - Datos ocultos



# Proyecciones - Escala, Datos ocultos



ARCHIVO

## LA CABO HERIDA EN ACCIÓN

Cuando Rosa María se acercó a su compañero para tratar de ayudarlo con un detenido que traía del brazo no se percató del peligro que venía detrás de él. Lo único que oyó fue un disparo. Ella cayó al piso y nunca más, hasta el día de hoy, volvió a caminar.

### Comparison



REDACCIÓN EL TIEMPO

28 de agosto 2004, 12:00 A.M.

Actualizado: 28.08.2004 00:00

Unirse a whatsapp

100%

Composers



**S**on pocas las mujeres policía que han resultado heridas en misión del servicio en Colombia. Y se cuentan con los dedos de una mano las que han quedado con incapacidades graves, como Rosa María Sánchez Bermúdez.

Conforme a los criterios de



 The Trust Project

[MÁS INFORMACIÓN >](#)

La cabo primero, a pesar de que se jubiló ya hace 10 años de la Policía, por su historia fue elegida para encabezar el último desfile de Independencia, que partió de la Plaza de Bolívar de Bogotá y culminó en el Parque Nacional.

Según los registros de la Policía, mientras desde el 2003 hasta la fecha, el conflicto ha dejado 142 policías hombres con discapacidad permanente (de un total de cerca de 950 heridos), esta suboficial y otras tres de sus colegas son las únicas mujeres en Colombia que han quedado discapacitadas en acciones de servicio.



# Mapa de calor - Población

Mapa de distribución poblacional de indígenas en el Quindío.



- A medida que los tonos tiendan a ser oscuros, hay una población superior a cien habitantes.

- A medida que los tonos tiendan a grises, su población está por debajo de cien habitantes.

Fuente: Elaboración propia, DANE.

# Análisis Exploratorio de Datos:

## Colguemos a los datos del hilo de la verdad

**Todo estudio debe superar la fase de Análisis Exploratorio de Datos**, es esencial para cualquier investigación. Para esto debemos hacernos esta pregunta: ¿Cómo se puedo resumir mi información?”

- ¿Hay presencia irregular de los datos?
- ¿Existen patrones de comportamiento regulares?
- ¿Se presentan datos atípico?
  - ▶ ¿Qué hacemos con ellos?
- ¿Hay presencia considérame de NA?
  - ▶ ¿Podemos hacer imputación de datos?
- ¿Existe heterogeneidad de  $x_1$  entre grupos?

# Storytelling con datos:

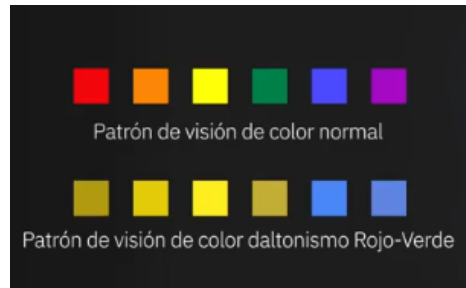
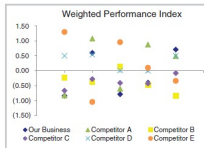
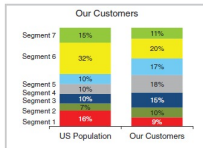
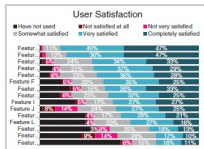
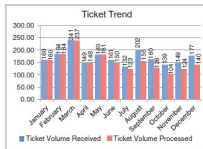
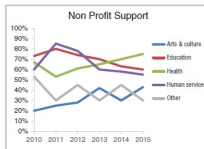
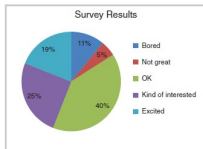
## El público importa

«Los gráficos no son un mero adorno, sino que, diseñados de manera adecuada, son una poderosa herramienta para comunicar resultados» *Walter Sosa Escudero*

- 1 Elejir el mejor descriptivo que me exprese mis datos, tomar en cuenta la naturaleza de los datos.
  - ▶ Conozca el contexto.
  - ▶ Piense cuáles son las variables que dominan el análisis.
- 2 Kiss (Keep it simple).
- 3 Siempre, **siempre**, conserve la misma paleta de colores.
  - ▶ Busque colores suaves a la vista

# Cómo contamos la historia

Nadie se propone hacer un mal gráfico



# ¿Qué debemos tener en cuenta?

## Vendamos una buena historia

Mostrar datos no es suficiente. Hay que contarlos como una historia visual que guíe a la audiencia a comprender y actuar.

- 1 Entender el contexto
- 2 Escoge la visual adecuada
- 3 Elimina el desorden (clutter)
- 4 Enfoca la atención
- 5 Piensa como diseñador
- 6 Eche una buena carreta

# Tipos de metodologías en Data Science:

Recordemos que nosotros no hacemos ciencia de datos, pero la ciencia de datos en esencia es investigación, por ende; las metodologías que allí se encuentra son insumo claro para articular nuestras investigaciones.

**Veamos algunas de las metodologías que pueden ser extrapolables y tomar partido a la hora de investigar:**

- KDD (Knowledge Discovery in Databases)
- SEMMA (Sample, Explore, Modify, Model and Access)
- CRISP-DM (Cross-Industry Standard Process for Data Mining)
- ASUM DM
- TDPS

*El sistema T<sub>E</sub>X fue diseñado y desarrollado por Donald Knuth en la década del 70. Es un sofisticado programa para la construcción de tipografía de textos científicos y es la mejor opción disponible para edición de textos con contenido matemático tales como artículos, reportes, libros entre otros.*

L<sup>A</sup>T<sub>E</sub>X es un completo conjunto de comandos de marcado que se utiliza con el potente programa de composición T<sub>E</sub>X para la preparación de una amplia variedad de documentos, desde artículos científicos e informes hasta libros complejos.

Paper

Tesis

Distribuidores de L<sup>A</sup>T<sub>E</sub>X: Overleaf