



Introducción a la Estadística

Panorama general

Instructor: Juan Luis Palacios Soto

palacios.s.j.l@gmail.com



scidata
Matemáticas para
la ciencia de datos

- 1 Definición
- 2 Ramas de la Estadística
- 3 ¿Dónde y cuándo se aplica?
- 4 Tipos de datos
- 5 Configuración de Excel

scidata
matemáticas para
la ciencia de datos

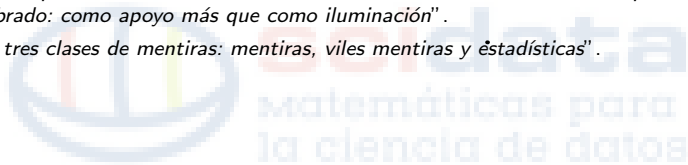
El alcohol es causante del 35 % de los accidentes automovilísticos. Luego, el 65 % de los accidentes automovilísticos son causados por personas sobrias. En conclusión: es más peligroso manejar sobrio!



scidata
matemáticas para
la ciencia de datos

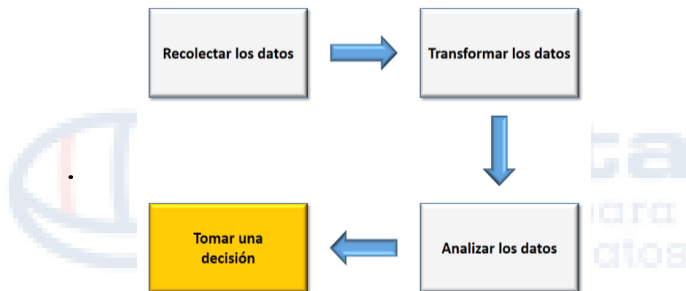
El alcohol es causante del 35 % de los accidentes automovilísticos. Luego, el 65 % de los accidentes automovilísticos son causados por personas sobrias. En conclusión: es más peligroso manejar sobrio!

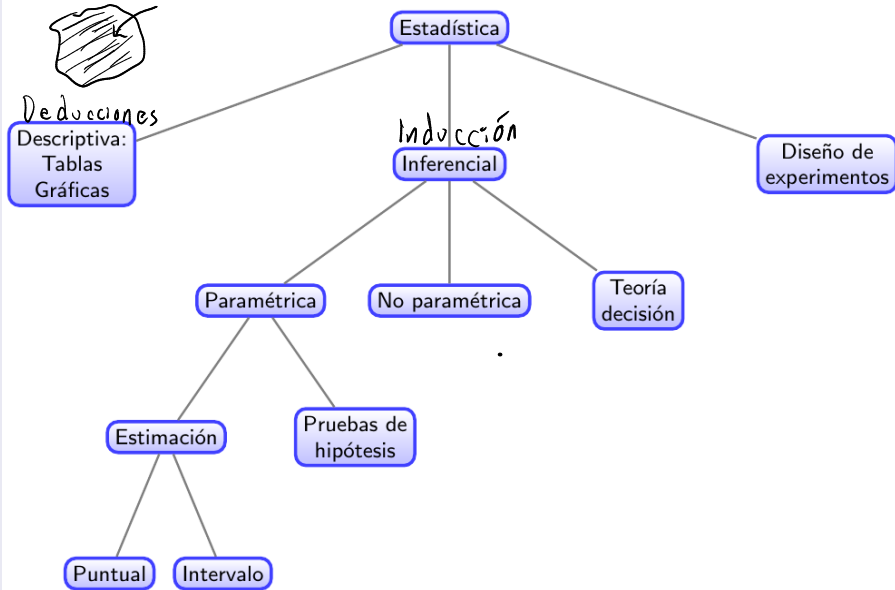
- ❶ “ *La estadística puede usarse para sustentar cualquier cosa, en especial a los estadísticos.*”
- ❷ “*Las cifras no mienten; los mentirosos calculan las cifras*”.
- ❸ “*Algunas personas utilizan la estadística como los borrachos utilizan los postes de alumbrado: como apoyo más que como iluminación*”.
- ❹ “*Hay tres clases de mentiras: mentiras, viles mentiras y estadísticas*”.



Definición (Definición de Estadística)

La **Estadística** es una rama de las matemáticas que se encarga de la recolección, estudio, transformación y análisis de datos para la **toma de decisión**.





¿Para qué y dónde se aplica la Estadística?

Describir los datos

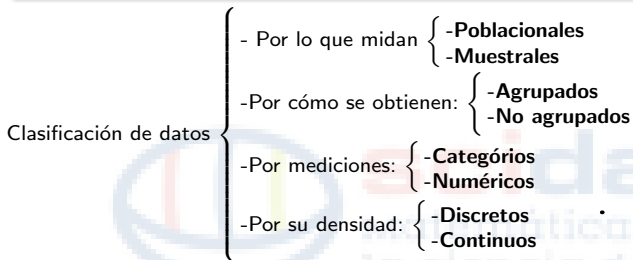
**Analizar y presentar
los datos**

Realizar predicciones

Mejorar procesos

Definición (Datos)

Un dato es el resultado de la observación de un “experimento”. Tenemos tres tipos de clasificación principalmente.

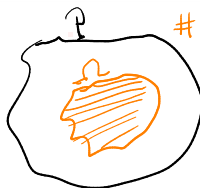


Definición (Por lo que midan)

Los datos se clasifican por lo que miden en **muestrales** o **poblacionales**.



Pop. \rightarrow .



$\#P = N$
 $\#\Omega = n$ $0 < n < N$

Definición (Por cómo se obtienen)

*Los datos se clasifican por cómo se obtiene en **agrupados y no agrupados**.*

Definición (Datos agrupados)

*Los **datos agrupados** son aquellos que están clasificados en función a un criterio, mostrando una frecuencia para cada clase o grupo formado. Es decir, los datos agrupados están separados por categorías, y cada dato u observación solo puede pertenecer a una categoría (no a dos o más)*

Definición (Datos no agrupados)

*Los **datos no agrupados** constituyen el conjunto de datos que no han sido clasificados y que son presentados en una tabla de datos en forma individual, es decir que no forman parte de un conjunto.*

•

Datos no agrupados

Tabla de datos				
158	165	174	153	192
183	145	189	159	157
172	189	175	167	205
149	173	180	182	188
177	175	159	175	170
162	181	168	193	163
151	163	156	156	159
159	162	171	155	151
162	150	184	178	168
188	201	196	174	153

Datos agrupados

Edad (años)	Marca de clase x_i	Frecuencia f_i	$x_i \cdot f_i$
[0-2)	1	7	7
[2-4)	3	8	24
[4-6)	5	8	40
[6-8]	7	7	49
	Σ	30	120

F_i :
7
15
23
30

$$[0-2) = [0, 2)$$

↑
Toma

↑
No toma

$$[3, 8) \rightarrow 5.5$$

$$C_1 [0, 2]$$

$$C_2 [2, 4]$$

Definición (Por mediciones)

*Los datos se clasifican por mediciones en **categoricos (cualitativos)** y **numéricos (cuantitativos)**.*

Datos categóricos



Datos numéricos

5	7
2	3

Definición (Datos categóricos o cualitativos)

*Un dato **cualitativo o categórico** representa una cualidad de una observación que se le asigna un caracter que no es numérico y si lo es no tienen sentidos las operaciones algebraicas. Los datos cualitativos se dividen en nominal y ordinal.*



scidata
matemáticas para
la ciencia de datos

Definición (Datos nominal)

Los datos **nominales** son etiquetas que no representan números operables, por lo que no tienen orden ni jerarquía entre ellos.

Ejemplo (Datos nominales)

Ejemplos de datos nominales.

- ① Lugar de nacimiento
- ② Número de seguro social
- ③ Número de empleado. 741, 747 NS47...
- ④ Equipo de fútbol preferido de la gente
- ⑤ Marca de automóvil compacto preferido en México
- ⑥ Sexo: Femenino, masculino
- ⑦ Estado civil de una persona
- ⑧ Tipo de sangre
- ⑨ Resultado de lanzar una moneda, un dado

Definición (Datos ordinales)

Los datos **ordinales** a diferencia de los datos nominales, éstos sí presentan un orden o jerarquía entre ellos.

Ejemplo (Datos ordinales)

A continuación veamos algunos ejemplos de datos ordinales.

- ① Calidad del restaurante: pésimo, malo, regular, bueno, excelente.
- ② Calificación de una materia.
- ③ Número de lista de clase.
- ④ Nivel de contaminación del aire en una ciudad: nulo, bajo, medio, alto, excesivo.
- ⑤ Nivel de estudios: prepa, licenciatura, maestría, etc.
- ⑥ Nivel de satisfacción de un cliente.
- ⑦ Calidad de un producto
- ⑧ Escala de Likert: totalmente de acuerdo, de acuerdo,...

Definición (Datos numéricos o cuantitativos)

Un dato **cuantitativo** se representa por medio de un número, sin embargo no siempre tiene sentido las operaciones algebraicas entre ellos. Los datos cuantitativos se dividen en dos: de intervalo y de razón.



scidata
matemáticas para
la ciencia de datos

Definición (Datos de intervalo)

Los **datos de intervalo** se caracterizan porque entre ellos puedes realizar operaciones de restas y no tienen un cero “verdadero” (consecuencia de esto último es que $2x$ no representa lo doble de x).

Ejemplo (Datos de intervalo)

- 1 *Temperatura de una placa metálica.*
- 2 *Edad de una persona*
- 3 *Talla de una persona*
- 4 *Cintura*
- 5 *Rendimiento académico de un alumno.*

Richter

Definición (Datos de razón)

Los datos **de razón** a diferencia de los datos de intervalo, es que con estos podemos verlos como los números reales. Se pueden realizar operaciones algebraicas entre ellos y tiene la propiedad del orden.

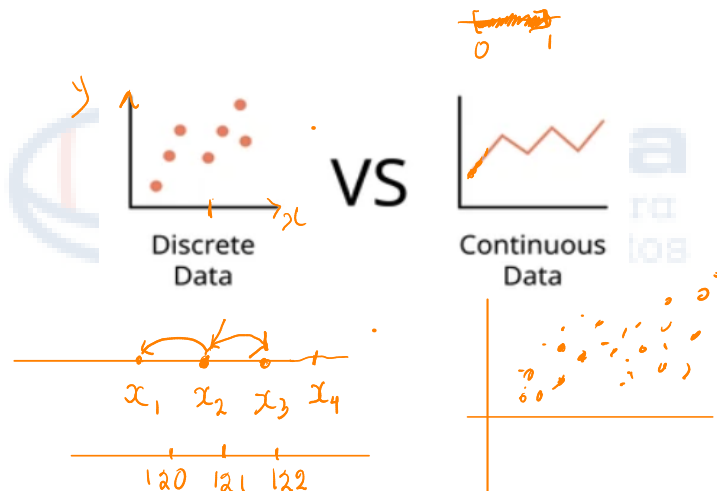
Ejemplo (Datos de razón)

A continuación veamos algunos ejemplos de datos de razón.

- 1 *Peso de una persona.*
- 2 *Altura de una persona.*
- 3 *Longitud, área, volumen.*
- 4 *Masa, densidad.*
- 5 *Ventas, gastos, ingresos.*
- 6 *Números de fallas en un proceso* ← *discreto*
- 7 *Número de artículos defectuosos en una línea de producción.*
- 8 *Tiempo de vida de un artículo* ← *Continuo.*

Definición (Datos discretos y continuos)

Un conjunto de datos A es llamado **discreto** si podemos contar cuántos puntos hay, de lo contrario diremos que es **continuo**.

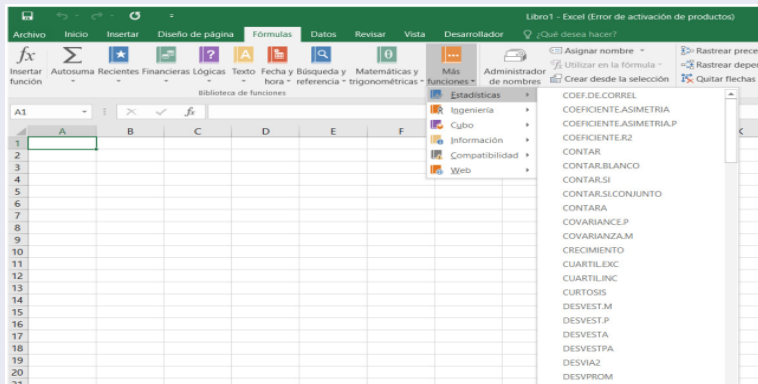


Definición (Proceso del Análisis Estadístico)

El proceso del análisis estadístico de datos es un proceso que implica la definición, recopilación, clasificación, organización, interpretación, presentación y modelado de datos. Este proceso se utiliza para entender e interpretar los datos para tomar decisiones basadas en información real y tangible.



Configuración de Excel



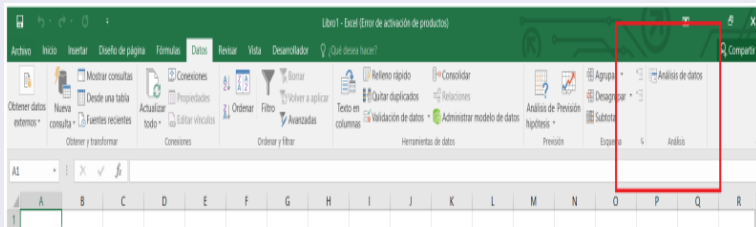
Configuración de Análisis de Datos

Sigue esta ruta:

Archivo → Opciones → Complementos → Herramientas para el análisis

Al seleccionar esta opción dar click en **Aceptar**

La pestaña de **Datos** tendrá activada una subzona de **Análisis de datos** como se muestra en la siguiente imagen:



Si el paso anterior no te funcionó, una alternativa es:

Archivo → Opciones → Complementos → Herramientas para el análisis

Selecciona ahora la opción **Complementos de Excel** y da click en Ir. En el cuadro de diálogo que se abrirá, selecciona la casilla de **Herramientas para análisis**. Da click en **Aceptar**.

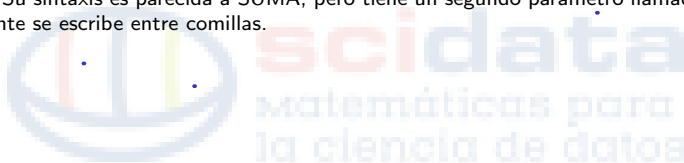


Funciones SUMA y SUMA.SI

La función SUMA calcula la suma de un rango seleccionado. Admite varias sintaxis:

- 1 Explícitamente: $\text{SUMA}(\text{num}_1, \text{num}_2, \dots, \text{num}_n)$.
- 2 Por fila o columna: $\text{SUMA}(\text{primer elemento}:\text{último elemento})$.
- 3 Por rango: $\text{SUMA}(\text{rango})$.

Por su parte, SUMAR.SI calcula la suma de un rango siempre que los datos cumplan con alguna condición. Su sintaxis es parecida a SUMA, pero tiene un segundo parámetro llamado criterio, que generalmente se escribe entre comillas.

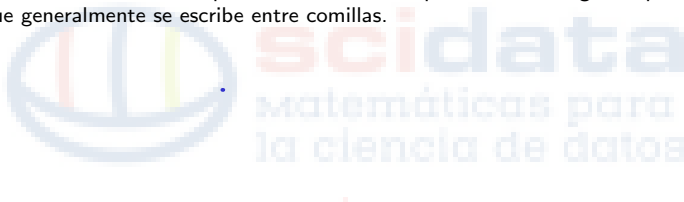


Funciones CONTAR y CONTAR.SI

La función CONTAR calcula el total de celdas que contienen datos numéricos de un rango seleccionado. Admite las sintaxis

- 1 Por fila o columna: CONTAR(primer elemento:último elemento).
- 2 Por rango: CONTAR(rango).

Por su parte, CONTAR.SI retorna el total de celdas de un rango siempre que los datos cumplan con alguna condición. Su sintaxis es parecida a CONTAR, pero tiene un segundo parámetro llamado criterio, que generalmente se escribe entre comillas.

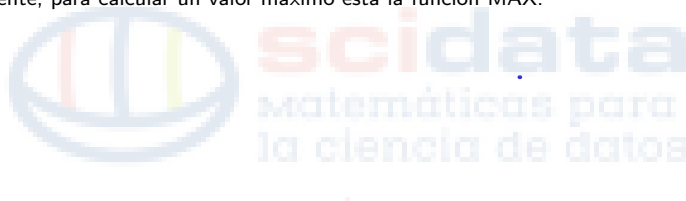


Funciones MIN y MAX

La función MIN retorna el valor mínimo de un rango seleccionado. Admite las sintaxis:

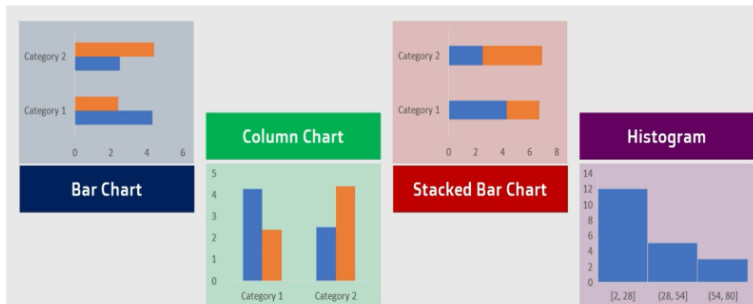
- 1 Explícitamente: $\text{MIN}(num_1, num_2, \dots, num_n)$.
- 2 Por fila o columna: $\text{MIN}(\text{primer elemento}:\text{último elemento})$.
- 3 Por rango: $\text{MIN}(\text{rango})$.

Análogamente, para calcular un valor máximo está la función MAX.



Gráficos de barras

Pueden usarse para hacer comparaciones absolutas.



Gráficos de líneas

Pueden usarse para visualizar cambios de los datos en el tiempo. Pueden identificar tendencias para realizar predicciones, control de calidad, etc.



Nubes de puntos

Pueden usarse para establecer relaciones entre dos medidas diferentes. Pueden identificar valores atípicos (outliers).

