

ESTADÍSTICA



Informática



Organización de datos



Definiciones

La **organización de datos** es el proceso mediante el cual **ordenamos, clasificamos y estructuramos la información recolectada** en una observación, encuesta o experimento, con el fin de facilitar su análisis estadístico.

Antes de aplicar cualquier técnica estadística (como medidas de tendencia o de dispersión), es fundamental **organizar los datos correctamente**.

Datos

Tipos de datos:

- **Datos cualitativos:** describen cualidades o categorías (ej. color, género, estado civil).
- **Datos cuantitativos:** representan cantidades o valores numéricos.
 - **Discretos:** valores contables (número de hijos, llamadas por día).
 - **Continuos:** valores medibles con infinitas posibilidades dentro de un rango (peso, altura, tiempo).

¿Por qué organizarlos?

- Para visualizar fácilmente cómo se distribuyen.
- Para identificar repeticiones o patrones.
- Para facilitar el cálculo de medidas estadísticas (media, mediana, cuartiles, etc.).

Tabla

Modelos matemáticos

◆ Distribución de frecuencias

La forma más común de organizar datos es a través de **tablas de frecuencia**, que resumen los datos indicando cuántas veces aparece cada valor o intervalo.

A. Para datos no agrupados:

Valor (x)	Frecuencia (f)
3	2
4	5
5	3

- f : frecuencia absoluta
- $N = \sum f$: número total de datos

Frecuencias

B. Para datos agrupados (intervalos):

Intervalo (clase)	f	F	f%
0–10	4	4	20%
10–20	6	10	30%
20–30	10	20	50%

- F : frecuencia acumulada
- $f\%$: frecuencia relativa en porcentaje



Ejercicio

Para un conjunto de datos construir la tabla de distribución



Datos

22, 45, 39, 34, 31, 29, 54, 61, 47, 58,
33, 26, 41, 49, 35, 36, 28, 53, 42, 44,
38, 46, 43, 30, 40, 27, 59, 55, 60, 52,
34, 48, 32, 30, 37, 33, 62, 31, 36, 56,
44, 29, 40, 35, 45, 43, 34, 38, 39, 42,
41, 50, 46, 44, 48, 47, 30, 33, 29, 27,
26, 45, 55, 59, 58, 52, 49, 50, 51, 54,
31, 32, 35, 36, 37, 38, 39, 41, 40, 43,
44, 42, 46, 47, 48, 49, 50, 51, 52, 53,
54, 55, 56, 57, 58, 59, 60, 61, 62, 63

Ordenar



Paso 2: Ordenar los datos (mínimo y máximo)

- Mínimo: 22
- Máximo: 63

Esto nos da el rango total:

$$\text{Rango} = 63 - 22 = 41$$

Intervalos



Paso 3: Calcular el número de intervalos

Para esto usamos la **Regla de Sturges**:

$$k = 1 + \log_2(n) \approx 1 + 3.322 \log_{10}(n)$$

$$k \approx 1 + 3.322 \log_{10}(100) = 1 + 3.322 \cdot 2 = 1 + 6.644 = 7.644 \approx 8 \text{ intervalos}$$



Amplitud

► **Paso 4: Calcular la amplitud del intervalo**

$$\text{Amplitud} = \frac{\text{Rango}}{k} = \frac{41}{8} = 5.125 \approx 6$$

(Tomamos 6 para redondear hacia arriba y cubrir todo el rango.)

□ Paso 5: Construir los intervalos

Comenzamos desde 22 (mínimo) y sumamos la amplitud (6):

Intervalo	Rango numérico
1	22 - 27
2	28 - 33
3	34 - 39
4	40 - 45
5	46 - 51
6	52 - 57
7	58 - 63

Tabla



Paso 6: Contar frecuencias

Ahora contamos cuántos datos caen en cada intervalo.

Conteo

Intervalo (minutos)	f (frecuencia)
22 – 27	6
28 – 33	14
34 – 39	20
40 – 45	21
46 – 51	19
52 – 57	10
58 – 63	10

Cálculos

+ Paso 7: Calcular frecuencias acumuladas y relativas

Intervalo (minutos)	f	F	f%
22 – 27	6	6	6.0%
28 – 33	14	20	14.0%
34 – 39	20	40	20.0%
40 – 45	21	61	21.0%
46 – 51	19	80	19.0%
52 – 57	10	90	10.0%
58 – 63	10	100	10.0%

Ejercicio

18, 25, 30, 35, 42, 47, 53, 56, 58, 60,
21, 23, 27, 33, 39, 45, 51, 54, 59, 61,
19, 20, 29, 37, 41, 46, 49, 52, 57, 62,
17, 26, 31, 34, 40, 44, 48, 50, 55, 63,
22, 24, 28, 36, 38, 43, 60, 61, 58, 59,
18, 25, 32, 35, 42, 47, 53, 56, 57, 60,
21, 23, 27, 33, 39, 45, 51, 54, 58, 61,
19, 20, 29, 37, 41, 46, 49, 52, 57, 62,
17, 26, 31, 34, 40, 44, 48, 50, 55, 63,
22, 24, 28, 36, 38, 43, 60, 61, 58, 59

Tabla respuesta

Tabla de distribución de frecuencias agrupadas

Intervalo (minutos)	f	F	f%	F%
17 – 23	14	14	14%	14%
24 – 30	16	30	16%	30%
31 – 37	18	48	18%	48%
38 – 44	18	66	18%	66%
45 – 51	18	84	18%	84%
52 – 58	12	96	12%	96%
59 – 65	4	100	4%	100%

Medidas de Posición



Definiciones

Las **medidas de posición** son herramientas estadísticas que permiten **dividir un conjunto de datos ordenados en partes iguales**, con el fin de analizar la distribución de los valores. Se utilizan principalmente para entender **cómo se distribuyen los datos respecto a su ubicación relativa** dentro del conjunto.

- **Cuartiles (Q)**: Dividen los datos en **4 partes iguales**. Se identifican como:
 - Q_1 : primer cuartil (25% de los datos están por debajo)
 - Q_2 : segundo cuartil (equivale a la mediana, 50%)
 - Q_3 : tercer cuartil (75% de los datos están por debajo)



Más medidas

- **Quintiles (K)**: Dividen los datos en **5 partes iguales**. Por ejemplo:
 - K_1 : 20%, K_2 : 40%, ..., K_4 : 80%
- **Deciles (D)**: Dividen los datos en **10 partes iguales**. Por ejemplo:
 - D_1 : 10%, D_5 : 50%, D_9 : 90%
- **Percentiles (P)**: Dividen los datos en **100 partes iguales**. Son los más detallados:
 - P_{25} : 25%, P_{50} : 50% (la mediana), P_{90} : 90%, etc.

Estas medidas permiten identificar el lugar que ocupa un dato dentro del conjunto.

FIN