

La descripción de datos

- Tipos de datos
- Frecuencia
- Distribución de frecuencias: diagrama de barras e histogramas
- Medidas de centralización: media aritmética
- Medidas de dispersión: desviación típica y varianza

• Tipos de datos

Variable aleatoria: variable de la que desconocemos el valor que tomará como resultado de una acción.

- Variables aleatorias discretas.- Conjunto finito de valores (ej. lanzamientos de un dado).
- Variables aleatorias continuas.- Conjunto infinito de valores (ej. Tiempos en un cronómetro).

Población o Universo: es el conjunto homogéneo de elementos sobre el que se realiza el estudio.

Muestra: conjunto de N medidas de una cierta cantidad extraídas de la población o universo.

Variable estadística: es cada una de las características observadas sobre los elementos de una población.

Dato estadístico: Cada uno de los valores que toma la variable estadística.

• Frecuencia

Muestra: $x_i, i = 1, \dots, N$

(ejemplo, $N = 1000$ lanzamientos de un dado)

Datos estadísticos: $x_k, k = 1, \dots$ valores diferentes

(ejemplo, 6 resultados posibles $x_k = 1, 2, 3, 4, 5, 6$)

Frecuencia absoluta: Número de veces n_k que el resultado x_k ha tenido lugar:

$$\sum_k n_k = N$$

Frecuencia relativa: Frecuencia absoluta dividida por el número total de medidas N .

$$F_k = \frac{n_k}{N}$$

Condición de normalización:

$$\sum_k F_k = \sum_k \frac{n_k}{N} = \frac{1}{N} \sum_k n_k = \frac{N}{N} = 1$$

Frecuencia relativa acumulada:

$$\sum_{k \leq j} \frac{n_k}{N}$$

- Distribución de frecuencias: diagrama de barras e histogramas

➤ Ejemplo de variable discreta:

26, 24, 26, 28, 23, 24, 25, 24, 26, 25

↓ Ordenación

23, 24, 24, 24, 25, 25, 26, 26, 26, 28

↓ Tabla

Tabla de valores de x hallados y número de apariciones

Valores x_k	23	24	25	26	27	28
Frecuencia, n_k	1	3	2	3	0	1
Frecuencia relativa, $F_k = \frac{n_k}{N}$	0.1	0.3	0.2	0.3	0.0	0.1
Frecuencia relativa acumulada	0.1	0.4	0.6	0.9	0.9	1.

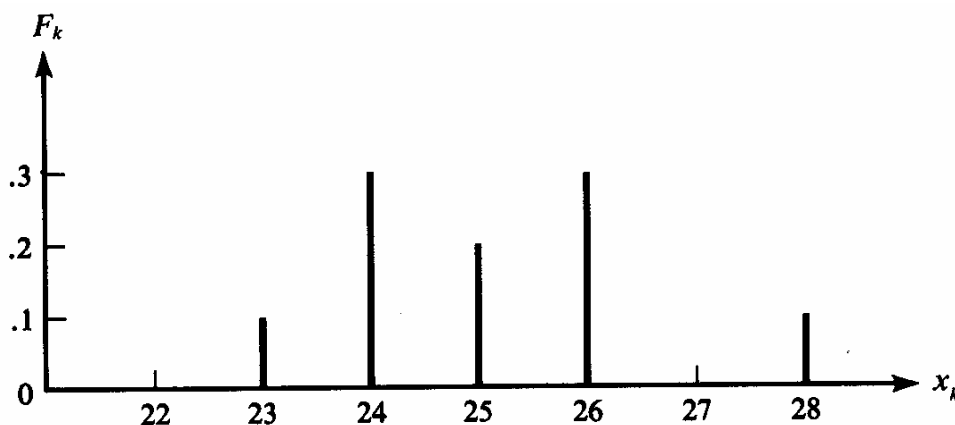


Diagrama de barras de las $N = 10$ medidas de x

➤ Ejemplo de variable continua:

26.4, 23.9, 25.1, 24.6, 22.7, 23.8, 25.1, 23.9, 25.3, 25.4

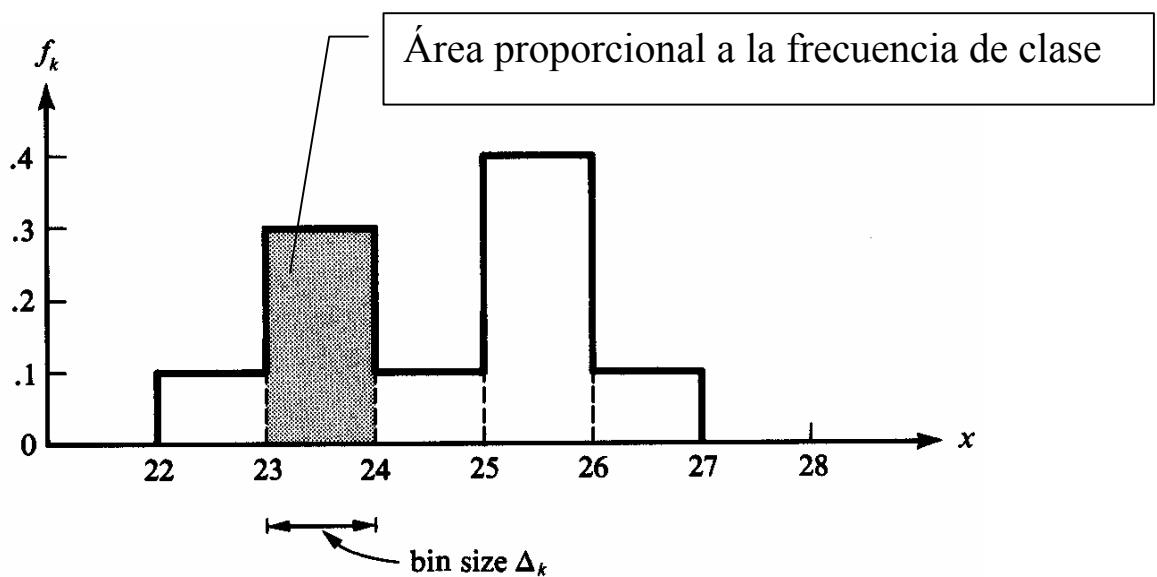
⇓ Ordenación

22.7, 23.8, 23.9, 23.9, 24.6, 25.1, 25.1, 25.3, 25.4, 26.4

⇓ Tabla

Tabla de valores de x agrupados en intervalos de clase (bin)

Valores x_k	22-23	23-24	24-25	25-26	26-27	27-28
Frecuencia n_k	1	3	1	4	1	0
Frecuencia relativa						
$F_k = \frac{n_k}{N}$	0.1	0.3	0.1	0.4	0.1	0.0



Histograma de bins de las 10 medidas de x

- Medidas de centralización: media aritmética

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad ; \quad \bar{x} = \frac{1}{N} \sum_k n_k x_k = \sum_k F_k x_k$$

PROPIEDADES

1. La suma algebraica de las desviaciones de un conjunto de números de su media aritmética es cero:

$$\sum_k n_k (x_k - \bar{x}) = 0$$

2. La suma de los cuadrados de las desviaciones de un conjunto de números de cualquier número a es mínima sólo si $a = \bar{x}$.

$$\sum_k n_k (x_k - a)^2 \quad \text{es mínima solo si } a = \bar{x}$$

3. Si a todos los valores de una variable x se les suma (multiplica por) una cantidad constante A , la media de los nuevos valores también aumentará en (vendrá multiplicada por) esa cantidad.

- Medidas de dispersión: desviación estándar y varianza

➤ Varianza:

$$\sigma_x^2 = \sum_i \frac{(x_i - \bar{x})^2}{N} = \sum_k \frac{n_k}{N} (x_k - \bar{x})^2 = \sum_k F_k (x_k - \bar{x})^2$$

➤ Desviación típica:

$$\sigma_x$$

PROPIEDADES

1. El cálculo de la varianza puede simplificarse de la siguiente manera:

$$\sigma_x^2 = \frac{1}{N} \sum_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_i x_i^2 - \bar{x}^2 = \frac{1}{N} \sum_k n_k x_k^2 - \bar{x}^2$$

2. Si a todos los valores de una variable x se les suma una cantidad constante A , la varianza de los nuevos valores no varía.
3. Si a todos los valores de una variable x se les multiplica por una cantidad constante A , la varianza de los nuevos valores vendrá multiplicada por el cuadrado de esa cantidad