

Introducción a la estadística

1. Introducción

La Estadística Descriptiva es un método de descripción numérica de conjuntos numerosos. Es por tanto, un método cuantitativo que utiliza el número como soporte objetivo.

Algunos conceptos importantes:

- Población → conjunto de elementos con una o varias características en común.
- Unidad estadística → cada uno de los elementos o individuos de la población.
- Muestra → subconjunto de la población sobre la que se realiza el estudio.

Los elementos de una población vienen descritos por caracteres (color de pelo, peso, estatura...) y se clasifican en:

- Cualitativos → cuando sus modalidades no sean medibles (color de pelo, color de ojos, sexo...)
- Cuantitativos → cuando sus modalidades sean medibles (peso, estatura, número de personas en la cola del supermercado...).

Una variable es un símbolo que notamos X y que toma unos determinados valores. Si estos valores se deben a una observación o medición estadística, diremos que la variable es estadística, y estas pueden ser:

- Discretas → si sus valores posibles son valores aislados (número de hijos de una familia, número de hermanos).
- Continuas → si sus valores están en un intervalo (estatura...)

Si estudiamos un determinado carácter, los datos se suelen recoger en una tabla como la siguiente:

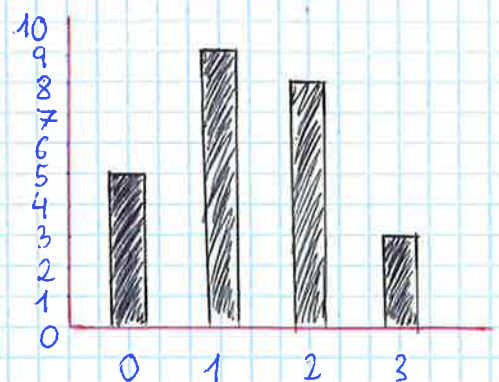
Valores de la variable	Frecuencia absoluta	Frecuencia absoluta acumulada	Porcentaje %
x	f	F	

2. Representación gráfica

2.1. Diagrama de barras

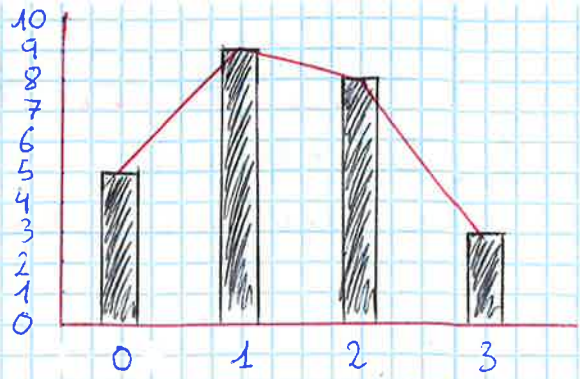
Se trata de representar en un sistema de ejes coordenados los valores de la variable frente a las frecuencias absolutas.

x	f
0	5
1	9
2	8
3	3



2.2. Polígono de frecuencias
Es el polígono que se obtiene al unir los puntos medios, en el diagrama de barras.

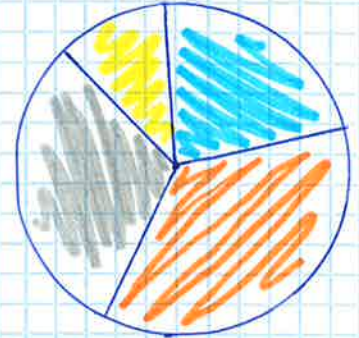
x	f
0	5
1	9
2	8
3	3



2.3. Diagrama de sectores

Consiste en un círculo que descompondremos en tantos sectores como modalidades tenga el carácter.

x	f
0	5
1	9
2	8
3	3



• 0 • 1 • 2 • 3

3. Características estadísticas: estadísticos

Ahora, lo que haremos es dar resúmenes cuantitativos que nos informen de la variable (es lo que se llaman estadísticos o características estadísticas).

Estos resúmenes pueden ser de dos tipos:

- De tendencia central: media, moda y mediana.
- De dispersión: varianza y desviación típica.

Ejemplo 1

Supongamos que estamos estudiando el número de vuelos semanales que realizan 10 pilotos. Los datos obtenidos son los siguientes:

Nº de vuelos	0	1	2	3
Frecuencia absoluta (nº de pilotos que realizan esos vuelos)	2	4	3	1

En primer lugar, construimos la tabla de datos:

Número de vuelos x	Frecuencia absoluta f	Frecuencia absoluta acumulada F	$x \cdot f$
0	2	2	$0 \cdot 2 = 0$
1	4	$2 + 4 = 6$	$1 \cdot 4 = 4$
2	3	$2 + 4 + 3 = 9$	$2 \cdot 3 = 6$
3	1	$2 + 4 + 3 + 1 = 10$	$3 \cdot 1 = 3$
	$N = 10$		$\sum x \cdot f = 13$

• Media: $\bar{x} = \frac{\sum f \cdot x}{N} = \frac{13}{10} = 1.3$

• Moda: $M_0 = 1$ ya que es el valor de la variable (el valor de x) que tiene mayor frecuencia absoluta.

• Mediana: Se puede hacer de 2 formas:

1ª forma \rightarrow como son pocos datos, los ordenamos:

0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 3

y vemos que en el centro hay dos números (1 y 1). En este caso, calculamos la media de dichos números:

$$\frac{1+1}{2} = 1$$

y esa es la mediana: $M_e = 1$

2ª forma \rightarrow de forma aproximada.

Calculamos $\frac{N}{2} = \frac{10}{2} = 5$ y dicho valor lo miramos en la columna de las frecuencias absolutas acumuladas (F). Como no está, se toma el superior, en este caso, 6, y se mira qué valor de la variable le corresponde: en nuestro caso, 1. Pues esa es la mediana, aproximada.

Interpretación: análisis de los datos

Supongamos que estamos estudiando el número de vuelos semanales que realizan 10 pilotos.

Los datos obtenidos son los siguientes:

Nº de vuelos	0	1	2	3
Frecuencia absoluta	2	4	3	1

La media es 1.3 y nos indica, que, por término medio, el número de vuelos es de 1.3, es decir, que por término medio estos pilotos vuelan entre 1 y 2 veces por semana.

La moda es 1, lo que nos indica que lo más frecuente es que vuelen 2 veces por semana.

Y, por último, la mediana es 1, lo que nos dice que hay tantos pilotos que vuelan 1 o más veces, como pilotos que lo hacen 1 vez o menos.

Ejemplo 2

Supongamos que estamos estudiando el número de aciertos de 100 alumnos en una prueba de 30 preguntas. Los resultados obtenidos se recogen en la siguiente tabla:

Acertios	x	f	$x \cdot f$	x^2	$x^2 \cdot f$
[0, 5)	$\frac{5+0}{2} = 2'5$	3	7'50	6'25	18'75
[5, 10)	$\frac{10+5}{2} = 7'5$	10	75'00	56'25	562'50
[10, 15)	$\frac{15+10}{2} = 12'5$	25	312'50	156'25	3906'25
[15, 20)	$\frac{20+15}{2} = 17'5$	38	665'00	306'25	11637'50
[20, 25)	$\frac{25+20}{2} = 22'5$	16	360'00	506'25	8100'00
[25, 30)	$\frac{30+25}{2} = 27'5$	8	220'00	756'25	6050'00
Total		$N=100$	$\sum f \cdot x = 1640$		$\sum f \cdot x^2 = 30275$

Los valores de x son la media de los extremos del intervalo.

$$\text{Media: } \bar{x} = \frac{\sum f \cdot x}{N} = \frac{1640}{100} = 16'40$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f \cdot x^2}{n} - \bar{x}^2 = \frac{30275}{100} - 16'40^2 = 33'79$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = \sqrt{33'79} = 5'81$$

Interpretación: análisis de los datos

Supongamos que estamos estudiando el número de aciertos de 100 alumnos en una prueba de 30 preguntas. Los resultados obtenidos se recogen en la siguiente tabla:

Acertios	x	f
[0, 5)	2'5	3
[5, 10)	7'5	10
[10, 15)	12'5	25
[15, 20)	17'5	38
[20, 25)	22'5	16
[25, 30)	27'5	8
Total		100

En este caso el rango es 30 y, por tanto, no nos proporciona mucha información, ya que es muy grande.

La varianza es $\sigma^2 = 33'79$ y la desviación típica es $\sigma = 5'81$, que son relativamente grandes, lo que nos dice que los datos presentan una agrupación relativamente pequeña respecto de la media.