

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 éstas, 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...)

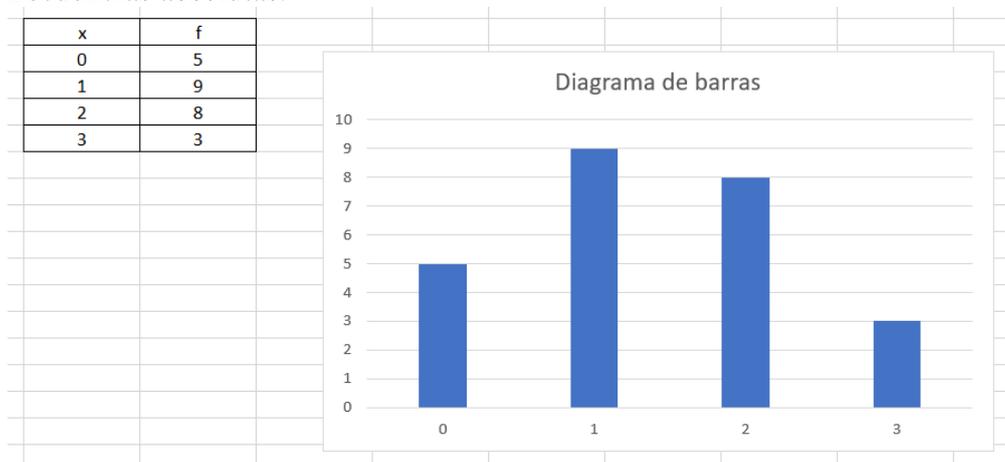
Si estudiamos un determinado carácter, los datos se suelen recoger en una **tabla** como la siguiente:

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

(2) Representación gráfica

(2.1) Diagrama de barras

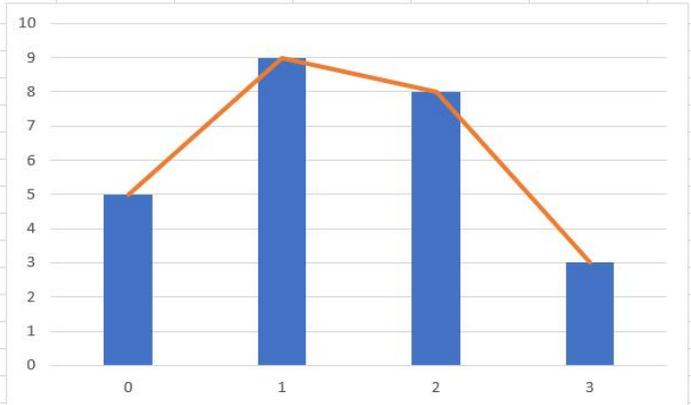
Se trata de representar en un sistema de ejes coordenados los valores de la variable frente a las frecuencias absolutas.



(2.2) Polígono de frecuencias

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

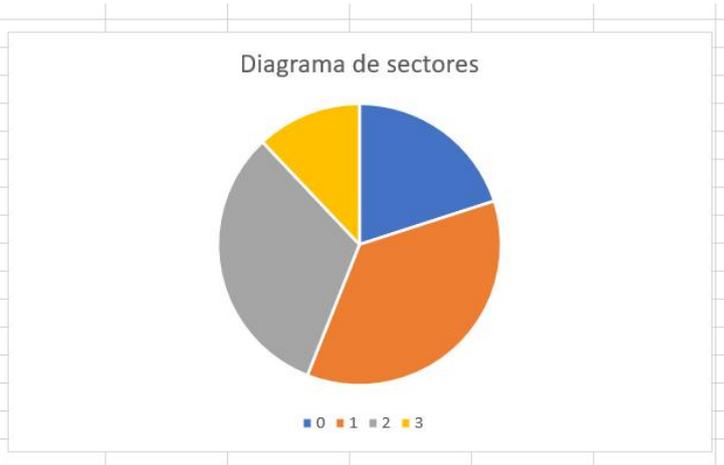
x	f
0	5
1	9
2	8
3	3



(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



(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
- De dispersión

Características estadísticas de tendencia central

Media (aritmética o promedio):

$$\bar{x} = \frac{\sum f \cdot x}{N}$$

Para calcular la media se multiplica cada valor de la variable por su frecuencia absoluta, se suman y se divide entre el número total de datos.

Mediana:

Se ordenan los datos. Si hay un número par de datos la mediana es la media de los dos datos centrales; si el número de datos es impar, la mediana es justamente el dato central.

Si hay muchos datos, en vez de ordenarlos podemos hacer lo siguiente (nos dará un valor aproximado de la mediana): calcular $\frac{N}{2}$ y mirar dicho valor en la columna de las frecuencias absolutas acumuladas y se mira qué valor de la variable le corresponde; si no está (que es lo que pasará la mayoría de las veces), se toma en F_i superior, y se mira el valor de la variable que le corresponde.

Moda:

Valor más frecuente de la variable.

Ejemplo:

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

Mediana:

Se puede hacer de dos formas:

1ª forma: 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: $Me = 1$

2ª forma: 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*.

Características estadísticas de dispersión

Permiten conocer el grado de agrupamiento de los datos en torno a las medidas de centralización y, por tanto, nos dan una medida del error que estamos cometiendo.

Rango

Es la diferencia entre el mayor valor de la variable y el menor valor de la variable.

Varianza:

$$\sigma^2 = \frac{\sum f \cdot (x - \bar{x})^2}{n} \rightarrow \sigma^2 = \frac{\sum f \cdot x^2}{n} - \bar{x}^2$$

Desviación típica:

$$\sigma = +\sqrt{\sigma^2}$$

Ejemplo:

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:

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

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:

Aciertos	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.

Recursos TIC que pueden ser de utilidad:

Construcción de la tabla para la media (Javier Valdés Gómez)

<https://www.youtube.com/watch?v=w2rdWACQig>

Cálculo de la varianza y de la desviación típica (Javier Valdés Gómez)

<https://www.youtube.com/watch?v=gHiLjdX4W30&t=0s>

Estadística con Hojas de Cálculo (Isabel García)

<https://www.youtube.com/watch?v=yg-McatDn7s>

Diagrama de sectores (Isabel García)

<https://www.youtube.com/watch?v=1KRKAYAENoo>

Diagrama de barras (Isabel García)

<https://www.youtube.com/watch?v=2cngaJtAwqw>

Apuntes de Estadística (Elisa Lozano Latorre)

<https://drive.google.com/file/d/1a4uY63HIzHEeHqM7Qklc298XB3GnsLBW/view>

Cómo trabajar con Hojas de Cálculo

Para realizar una misma operación en toda la columna/fila

	A	B	C
1	2	4	8
2	3	5	15
3	4	6	24
4	5	7	35
5		22	82
6			

En el ordenador hay que colocar el cursor/dedo en ese cuadrado/círculo y, arrastrar hacia abajo/derecha.

En la tablet se marca la casilla, se elige la opción **Relleno** y se arrastra hacia abajo/derecha.

Para calcular la media

	A	B	C
1	x	f	x*f
2	2	4	8
3	3	5	15
4	4	6	24
5	5	7	35
6		22	82
7			

	A	B	C	D
1	2	4	8	
2	3	5	15	
3	4	6	24	
4	5	7	35	
5		=SUMA(B1:B4)	82	
6				
7	Media	3,727272727		
8				
9				
10				

	A	B	C
1	2	4	8
2	3	5	15
3	4	6	24
4	5	7	35
5		22	=SUMA(C1:C4)

	A	B	C	D
1	2	4	8	
2	3	5	15	
3	4	6	24	
4	5	7	35	
5		22	82	
6		3,727272727		
7	Media	=C5/B5		

Para calcular los cuadrados

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas

100% € % .0 .00 123 Predetermi.

D2	A	B	C	D
1	x	f	x*f	x^2
2	2	4	8	4
3	3	5	15	9
4	4	6	24	16
5	5	7	35	25
6		22	82	0
8	Media	3,727272727		

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas

100% € % .0 .00 123 Predetermi.

D2:D6	A	B	C	D
1	x	f	x*f	x^2
2	2	4	8	4
3	3	5	15	9
4	4	6	24	16
5	5	7	35	25
6		22	82	0

Para calcular la varianza

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas Extension

100% € % .0 .00 123 11

E2	A	B	C	D	E
1	x	f	x*f	x^2	x^2*f
2	2	4	8	4	16
3	3	5	15	9	45
4	4	6	24	16	96
5	5	7	35	25	175
6		22	82	0	332

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas Extension

100% € % .0 .00 123 10

E6	A	B	C	D	E
1	x	f	x*f	x^2	x^2*f
2	2	4	8	4	16
3	3	5	15	9	45
4	4	6	24	16	96
5	5	7	35	25	175
6		22	82	0	332

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas Extension

100% € % .0 .00 123 11

B10	A	B	C	D	E
1	x	f	x*f	x^2	x^2*f
2	2	4	8	4	16
3	3	5	15	9	45
4	4	6	24	16	96
5	5	7	35	25	175
6		22	82	0	332
8	Media	3,727272727			
9		1,198347107			
10	Varianza	=E6/B6-B8^2			

Para calcular la desviación típica

Hoja de cálculo sin título

Archivo Editar Ver Insertar Formato Datos Herramientas Extension

100% € % .0 .00 123 10

B12	A	B	C	D	E
1	x	f	x*f	x^2	x^2*f
2	2	4	8	4	16
3	3	5	15	9	45
4	4	6	24	16	96
5	5	7	35	25	175
6		22	82	0	332
8	Media	3,727272727			
9		1,198347107			
10	Varianza	=E6/B6-B8^2			
11		1,094690416			
12	Desviación típica	=RAIZ(B10)			

Expresar los resultados con dos decimales

Hay que hacer clic en el Menú **Formato**, seleccionar **Número** y después, otra vez **Número**.

The screenshot shows the Excel interface with the 'Formato' menu open. The 'Número' option is selected, and its sub-menu is displayed. The sub-menu includes the following options:

- Automático (checked)
- Texto sin formato
- Número (1.000,12)
- Porcentaje (10,12%)
- Científico (1,01E+03)
- Contabilidad (- 1.000,12 €)
- Finanzas ((1.000,12))
- Moneda (1.000,12 €)
- Moneda (redondeado) (1.000 €)
- Fecha (26/09/2008)
- Hora (15:59:00)
- Fecha hora (26/09/2008 15:59:00)
- Duración (24:01:00)
- Moneda personalizada
- Fecha y hora personalizadas
- Formato de número personalizado

The spreadsheet data is as follows:

	A	B
1	x	f
2	2	4
3	3	5
4	4	6
5	5	7
6		22
7		
8	Media	3,727272727
9		
10	Varianza	1,198347107
11		
12	Desviación típica	1,094690416
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		