Tema 5 La información que recibimos

1.- Conceptos

La **Estadística** trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

Un **estudio estadístico** consta de las siguientes fases:

- Recogida de datos.
- Organización y representación de datos.
- Análisis de datos.
- Obtención de conclusiones.

En la mayoría de los estudios estadísticos es necesario la creación de un cuestionario

A.- Cuestionario

Documento básico para obtener la información en la gran mayoría de las investigaciones y estudios.

Los cuestionarios pueden ser:

• **abierto:** cada uno puede contestar lo que quiera.
• **limitado:** con un número prefijado de posibles respuestas.
• **cerrado:** más cómodo para el entrevistado pero que puede "deformar" el estudio, ya que el listado de posibles respuestas va a depender del encuestador y su buen criterio.

B.- Población

Una **población** es el conjunto de todos los elementos a los que se somete a un estudio estadístico.

C.- Individuo

Un **individuo** o **unidad estadística** es cada uno de los elementos que componen la población.

D.- Muestra

Una **muestra** es un conjunto representativo de la población de referencia, el número de individuos de una muestra es menor que el de la población.

E.- Muestreo

El **muestreo** es la reunión de datos que se desea estudiar, obtenidos de una proporción reducida y representativa de la población.

F.- Valor

Un **valor** es cada uno de los distintos resultados que se pueden obtener en un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos dos valores: cara y cruz.

G.- Dato

Un **dato** es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos 5 datos: cara, cara, cruz, cara, cruz.

H.- Variable

Una **variable estadística** es cada una de las **características o cualidades** que poseen los i**ndividuos de una población**.

Tipos de variable estadísticas

Variable cualitativa

Las **variables cualitativas** se refieren a **características o cualidades** que **no** pueden ser medidas con **números**

Variable cuantitativa

Una **variable cuantitativa** es la que se expresa mediante un **número**, por tanto se pueden realizar o**peraciones aritméticas** con ella.

Actividades 1,2,3

2.-Distribución de frecuencias

La **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

2.1.-Frecuencia

La **frecuencia absoluta** es el **número de veces** que aparece un determinado **valor** en un estudio estadístico.

Se representa por $\mathbf{f_i}$.

La **suma de las frecuencias absolutas** es igual al número total de datos, que se representa por **N**.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

Ejemplo:

1.- Dato: altura, en centímetros, de las primeras diez personas que pasan por la calle:

Datos 1: 167, 169, 165, 178, 177, 169, 181, 176, 168 y 175.

Datos 2: 174, 199, 197, 187, 206, 189, 188, 203, 188 y 178

Altura de personas

Datos 1

Altura	Frecuencia
165 – 168	3
169 – 172	2
173 – 176	2
177 – 180	2
181 – 184	1

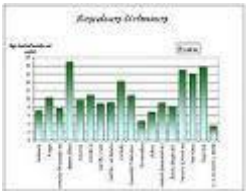
Datos 2

Altura	Frecuencia
173 – 177	1
178 – 182	1
183 – 187	1
188 – 192	3
193 – 197	1
198 – 202	1
203 – 207	2

2.2.- Representamos los datos en una gráfica

Una **gráfica estadística** es la mejor forma de disponer de toda la información que se ha recogido con una simple “ojeada”

Existen múltiples modelos de gráficas estadísticas, aunque los más difundidos son:



Gráfica de barras

Diagrama de sectores

2.2.1.- Diagrama de barras

Un **diagrama de barras** se utiliza para de presentar **datos cualitativos**

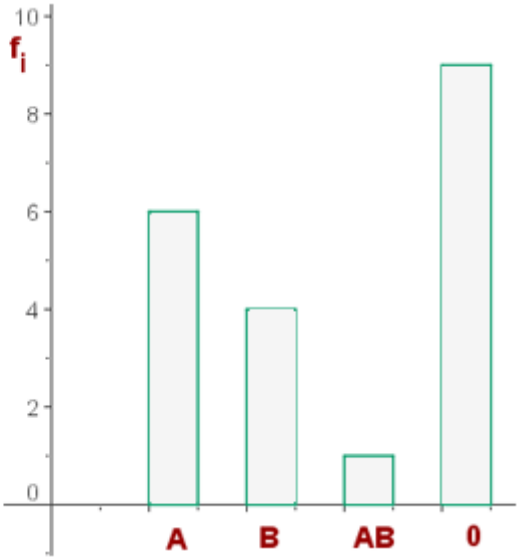
Se representan sobre unos ejes de coordenadas, en el **eje de abscisas** se colocan los **valores de la variable**, y sobre el **eje de ordenadas** las **frecuencias absolutas o relativas o acumuladas**.

Los **datos** se representan mediante **barras** de una **altura proporcional** a la **frecuencia**.

Ejemplo

Un estudio hecho al conjunto de los 20 alumnos de una clase para determinar su grupo sanguíneo ha dado el siguiente resultado:

Grupo sanguíneo	f_i
A	6
B	4
AB	1
0	9
	20



Polígono de frecuencias

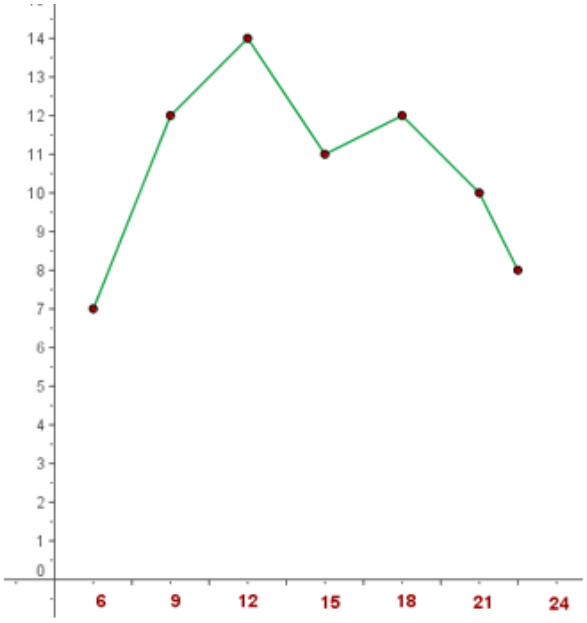
Un **polígono de frecuencias** se forma uniendo los **extremos** de las **barras** mediante **segmentos**

También se puede realizar trazando los **puntos** que representan las **frecuencias** y uniéndolos mediante **segmentos**.

Ejemplo

Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

Hora	Temperatura
6	7º
9	12º
12	14º
15	11º
18	12º
21	10º
24	8º



Actividades 4,5,6,7

2.2.2.- Diagrama de sectores

Un **diagrama de sectores** se puede utilizar para todo tipo de *variables*, pero se usa frecuentemente para las **variables cualitativas**.

Los **datos** se representan en un **círculo**, de modo que el **ángulo** de cada **sector** es **proporcional** a la **frecuencia absoluta** correspondiente.

$$\alpha = \frac{360°}{N} \cdot f_i$$

El diagrama circular se construye con la ayuda de un transportador de ángulos.

Ejemplo

En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 4 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360°}{30} \cdot 12 = 144° \qquad \alpha_2 = \frac{360°}{30} \cdot 3 = 36°$$

$$\alpha_3 = \frac{360°}{30} \cdot 9 = 108° \qquad \alpha_4 = \frac{360°}{30} \cdot 6 = 72°$$

	Alumnos	Ángulo
Baloncesto	12	144°
Natación	3	36°
Fútbol	9	108°
Sin deporte	6	72°
Total	30	360°

Actividades 8, 9 y 10

3.- Calculamos parámetros y los interpretamos

Los parámetros estadísticos se dividen en dos categorías: centrales y de dispersión

Centrales: Nos indican en torno a qué valor (centro) se distribuyen los datos.

Las **medidas de centralización** son: Media aritmética y moda

Dispersión: Las **medidas de dispersión** nos informan sobre cuanto se alejan del centro los valores de la distribución.

Las **medidas de dispersión** son: Rango o recorrido, , varianza y desviación típica

3.1.- Centrales

a.-La media

La **media aritmética** es el **valor** obtenido al **sumar** todos los **datos** y **dividir** el resultado entre el **número** total de **datos**.

$\bar{X}$ es el símbolo de la **media aritmética**.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

Ejemplo

Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio.

$$\bar{x} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = 80\ Kg$$

b.-La moda

La **moda** es el **valor** que tiene **mayor frecuencia absoluta**.

Se representa por $\mathbf{M_o}$.

Hallar la **moda** de la distribución:

2, 3, 3, 4, 4, 4, 5, 5 $\mathbf{M_o = 4}$

3.2.- Dispersión

a.- Rango

El **rango** es la **diferencia** entre el **mayor** y el **menor** de los **datos** de una distribución estadística.

Ejemplo:

La altura de 10 personas son: 167, 169, 165, 178, 177, 169, 181, 176, 168, 175

El rango: diferencia entre el valor de las observaciones mayor y el menor

Rango=181− 165= 16

b.- Varianza

La **varianza** es la **media aritmética del cuadrado de las desviaciones respecto a la media** de una distribución estadística.

La varianza se representa por σ^2. (sigma)

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N} \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}$$

Sirve para identificar si los datos están cercanos a la media o no,

Ejemplo

La altura de 10 personas son: 167, 169, 165, 178, 177, 169, 181, 176, 168, 175

	Datos 1	(Dato – Media)	(Dato – Media)²
	167	-5,5	30,25
	169	-3,5	12,25
	165	-7,5	56,25
	178	5,5	30,25
	177	4,5	20,25
	169	-3,5	12,25
	181	8,5	72,25
	176	3,5	12,25
	168	-4,5	20,25
	175	2,5	6,25
Total:	1725		272,5
Media:	172,5		
Varianza:	27,25		

Ejemplo 2

La altura de 10 personas son: 174, 199, 197, 187, 206, 189, 188, 203, 188,178

	Datos 2	(Dato – Media)	(Dato – Media)²
	174	-16,9	285,61
	199	8,1	65,61
	197	6,1	37,21
	187	-3,9	15,21
	206	15,1	228,01
	189	-1,9	3,61
	188	-2,9	8,41
	203	12,1	146,41
	188	-2,9	8,41
	178	-12,9	166,41
Total:	1909		964,9
Media:	190,9		
Varianza:	96,49		

A la vista de estos resultados podemos decir que los datos correspondientes a la segunda captura están más dispersos con respecto al valor de la media

c.- Desviación típica

La **desviación típica** es la **raíz cuadrada de la varianza**.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

La **desviación típica** se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}}$$

Ejemplo

La altura de 10 personas son: 167, 169, 165, 178, 177, 169, 181, 176, 168, 175

$s = 27'25 = 5'22$

Coeficiente de variación

Se calcula según la fórmula $C.V. = \frac{desviación\ típica}{media}$ en tanto por ciento

Ejemplo

La altura de 10 personas son: 167, 169, 165, 178, 177, 169, 181, 176, 168, 175

$CV = \frac{5'22}{172'5} = 0{,}01386$ porcentualmente será 1,38 %

Ejemplo 2

La altura de 10 personas son: 174, 199, 197, 187, 206, 189, 188, 203, 188,178

$$CV = \frac{9'82}{190'9} = 0{,}0514$$

Porcentualmente será 5,14 %

Actividades 11, 12, 13, 14, 15, 16, 17,18 y 19