

ESTADÍSTICA I.

Recuerda:

Población: Es el conjunto de todos los elementos que cumplen una determinada propiedad, que llamamos **carácter estadístico**. Los elementos de la población se llaman individuos.

Muestra: Es cualquier subconjunto de la población. El número de elementos de la muestra se llama tamaño de la muestra.

El **carácter estadístico** es una propiedad que permite clasificar a los individuos de una población. Clasificamos los caracteres estadísticos en **Cualitativos**, los que no podemos medir, y en **Cuantitativos**, los que podemos medir.

Los valores que toman los caracteres estadísticos cuantitativos los llamamos **variables estadísticas**.

Las variables estadísticas pueden ser **discretas** y **continuas**.

Una variable estadística es discreta cuando sólo puede tomar un número finito de valores o infinito numerable.

Representamos sus valores mediante: x_1, x_2, \dots, x_n .

Una variable estadística es continua cuando puede tomar todos los valores posibles dentro de un intervalo de la recta real. Las representamos mediante el intervalo o clase, y tomamos como representante del intervalo la marca de clase del intervalo (punto medio del intervalo). Es necesario que las clases tengan el mismo tamaño.

Llamamos **distribución estadística** al conjunto de todos los datos estadísticos.

Podemos presentar los datos estadísticos mediante tablas con las variables estadísticas y las diferentes frecuencias de estas variables.

Frecuencia absoluta de un valor x_i es el número de veces que aparece este valor x_i . La representaremos por f_i

Frecuencia absoluta acumulada de un valor x_i es la suma de las frecuencias absolutas de los valores menores o iguales que x_i . La representamos por F_i . Sabemos que: $F_i = f_1 + f_2 + \dots + f_i$

Frecuencia relativa de un valor x_i es el cociente entre la frecuencia absoluta de ese valor y el número total de datos que tenemos en la distribución estadística. La representaremos por $h_i = \frac{f_i}{N}$.

Frecuencia relativa acumulada de un valor x_i es el cociente entre la frecuencia absoluta acumulada del valor x_i y el número total de datos de la distribución estadística. La representamos por H_i . Sabemos que: $H_i = \frac{F_i}{N} = \frac{f_1 + f_2 + \dots + f_i}{N} = \frac{f_1}{N} + \frac{f_2}{N} + \dots + \frac{f_i}{N} = h_1 + h_2 + \dots + h_i$

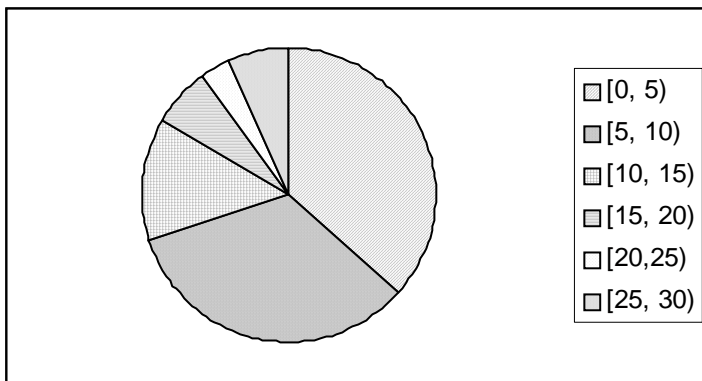
Podemos agrupar los datos haciendo representaciones gráficas. De estas representaciones las más importantes son:

Diagrama de barras: representamos los datos mediante barras de altura proporcional a su magnitud. En el diagrama de barras los datos pueden ser cualitativos o cuantitativos, pero siempre sin agrupar en clases.

Histogramas: Es un diagrama de barras pero los datos son siempre cuantitativos agrupados en clases o intervalos.

El **polígono de frecuencias** es la línea rota que une los puntos medios de los lados superiores de los rectángulos que forman las barras. Tanto el diagrama de barras como el histograma tienen un polígono de frecuencias asociado.

Diagrama de sectores: Dividimos un círculo en sectores de tamaño proporcional a la magnitud de los datos que representan.



Este es un ejemplo de diagrama de sectores, que corresponde a los datos del segundo problema de autoaprendizaje, el del médico de cabecera.

¿Qué hay que hacer para analizar una muestra?

Recoger los datos

Ordenar los datos (en orden creciente o decreciente)

Recuento de frecuencias

Agrupación de los datos: según la variable aleatoria sea discreta o continua.

Construcción de la tabla estadística.

Los **parámetros** son unos números que nos proporcionan una idea, lo más aproximada posible, del comportamiento de todos los elementos de una población en relación al carácter que estudiamos. Los dividimos en dos grupos: los parámetros de centralización y los parámetros de dispersión.

Parámetros de centralización pretenden agrupar o centralizar los datos correspondientes a toda una población en sólo un valor numérico, representando el conjunto total. Los más importantes son: media aritmética, moda y mediana.

La media aritmética: $\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$, donde x_i es cada uno de los valores de la

variable estadística y N es el número total de datos. El cálculo de la media aritmética es muy sencillo e intervienen todos los datos. Presenta un inconveniente, los efectos, que a veces son graves, que producen los valores extremos. Estos valores normalmente, no son significativos.

La moda es el valor de la variable con mayor frecuencia. Si los datos los tenemos agrupados en clases la llamamos clase modal.

La mediana es el valor de la variable que ocupa el lugar central si tenemos un número impar de datos. En caso contrario la mediana coincide con la media aritmética de los valores centrales. La mediana tiene la propiedad de que el 50% de los datos son menores o iguales que ella y el otro 50% son mayores o iguales. Entonces la mediana divide el conjunto de datos en dos subconjuntos iguales.

Parámetros de dispersión tienen como objetivo presentarnos una idea de la proximidad o lejanía de los datos de la población respecto al valor que hemos tomado como valor central. Los más importantes son: el rango, la varianza y la desviación típica. Una desviación típica muy grande indica que hay mucha dispersión de los datos, y una desviación típica muy pequeña indica que los datos están todos alrededor del valor central.

El rango es la diferencia entre el valor máximo y el mínimo de una variable estadística. Si el rango es grande existe la posibilidad de que los datos estén alejados unos de otros, y por tanto, que los valores centrales no sean muy representativos. Por otro lado, si el rango es pequeño, los datos no pueden encontrarse muy distanciados y los valores centrales pueden ser representativos del conjunto.

La varianza es la media aritmética de los cuadrados de las desviaciones de los datos respecto a

$$\text{la media: } \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2.$$

La desviación típica es la raíz cuadrada de la varianza: $\sigma = \sqrt{\sigma^2}$

Cuando en los N datos sólo aparecen k valores diferentes (lógicamente $k < N$), x_1, x_2, \dots, x_k , podemos utilizar la tabla de frecuencias para obtener la media y la varianza:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i f(x_i) = \sum_{i=1}^k x_i h(x_i) \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 f(x_i) - \bar{x}^2 = \sum_{i=1}^k x_i^2 h(x_i) - \bar{x}^2$$

(Recordemos que f es la frecuencia absoluta y h la frecuencia relativa)

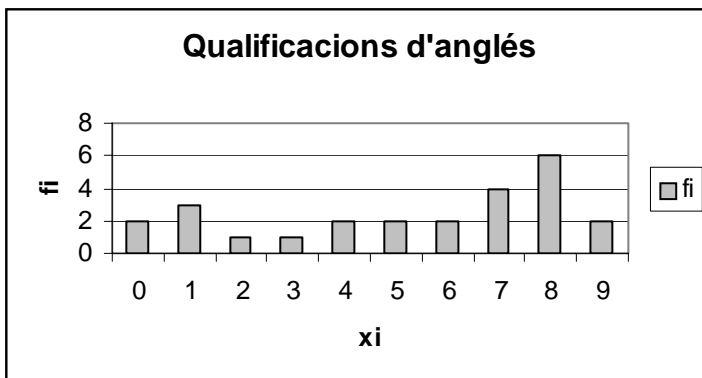
Ejercicios de autoaprendizaje:

1. Un profesor de 3º ESO de inglés, de una clase de 25 alumnos, tiene las notas siguientes en su cuaderno: 5, 3, 4, 1, 2, 8, 9, 7, 6, 8, 6, 7, 9, 8, 7, 7, 1, 0, 1, 5, 8, 0, 8, 8, 4

Se trata de una variable aleatoria discreta.
 Construimos la tabla estadística:

x_i	f_i	F_i	h_i	H_i
0	2	2	$\frac{2}{25} = 0.08$	0.08
1	3	5	$\frac{3}{25} = 0.12$	0.2
2	1	6	$\frac{1}{25} = 0.04$	0.24
3	1	7	$\frac{1}{25} = 0.04$	0.28
4	2	9	$\frac{2}{25} = 0.08$	0.36
5	2	11	$\frac{2}{25} = 0.08$	0.44
6	2	13	$\frac{2}{25} = 0.08$	0.52
7	4	17	$\frac{4}{25} = 0.16$	0.68
8	6	23	$\frac{6}{25} = 0.24$	0.92
9	2	25	$\frac{2}{25} = 0.08$	1
	25		1	

La suma de todas las frecuencias absolutas es el número total de datos, en este caso concreto esta suma se corresponde con el número de alumnos: 25.
 La suma de todas las frecuencias relativas es siempre la unidad.
 Hacemos el correspondiente diagrama de barras:



Hace falta hacer un estudio estadístico, calculando las medidas de centralización y las de dispersión y sacar las conclusiones de estos cálculos.

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
0	2	0	0
1	3	3	3
2	1	2	4
3	1	3	9
4	2	8	32
5	2	10	50
6	2	12	72
7	4	28	196
8	6	48	384
9	2	18	162
	25	132	912

$$\bar{x} = \frac{132}{25} = 5.28 \quad \sigma^2 = \frac{912}{25} - \left(\frac{132}{25}\right)^2 = 8.6016$$

$$\sigma = \sqrt{8.6016} = 2.9328$$

Las medidas de centralización son:

La media aritmética es $\bar{x} = 5.04$, la moda es el 8 y la mediana es 6

Las medidas de dispersión son:

El recorrido o rango es 9, la varianza es 9.1584 y la desviación típica es $\sigma = 3.02628$

Nuestras conclusiones son las siguientes: sabemos que la media es de 5.04, pero hay una dispersión de 3. Podemos decir que las notas de inglés son o muy buenas o malas. Sobre todo podemos decir que, como la moda es 8 y la mediana es 6, hay más gente con buenas notas, pero también hay gente suspendida.

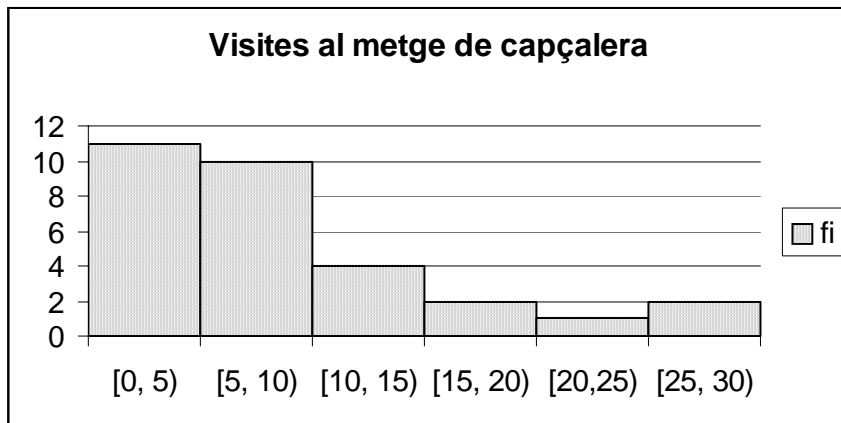
2. Hemos recogido el número de personas que han visitado el médico de cabecera a lo largo del mes de noviembre: 3, 2, 13, 4, 2, 4, 5, 6, 7, 3, 4, 5, 3, 2, 5, 6, 15, 21, 4, 3, 6, 29, 13, 6, 17, 13, 6, 5, 12, 26. Se pide una tabla de frecuencias con su diagrama correspondiente y también un estudio estadístico calculando las medidas de centralización y de dispersión.

Se trata de una variable aleatoria continua. Agrupamos los datos en clases que son intervalos de tamaño 5. La marca de clase es el punto medio de cada intervalo. Construimos la tabla estadística:

Clases	Marca de clase	f_i	F_i	h_i	H_i
[0, 5)	2.5	11	11	$\frac{11}{30} = 0.3\hat{6}$	$0.3\hat{6}$
[5, 10)	7.5	10	21	$\frac{10}{30} = 0.\hat{3}$	0.7
[10, 15)	12.5	4	25	$\frac{4}{30} = 0.1\hat{3}$	$0.8\hat{3}$
[15, 20)	17.5	2	27	$\frac{2}{30} = 0.0\hat{6}$	0.9

[20,25)	22.5	1	28	$\frac{1}{30} = 0.0\hat{3}$	0.9 $\hat{3}$
[25, 30)	27.5	2	30	$\frac{2}{30} = 0.0\hat{6}$	1
			30	1	

Construimos el histograma:



Calcula la media aritmética, la varianza y la desviación típica.

Clases	Marca de clase: x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[0, 5)	2.5	11	27.5	68.75
[5, 10)	7.5	10	75	562.5
[10, 15)	12.5	4	200	2500
[15, 20)	17.5	2	35	612.5
[20,25)	22.5	1	22.5	506.25
[25, 30)	27.5	2	54	1485
		30	412	5735

$$\bar{x} = \frac{412}{30} = 13.7\hat{3} \quad \sigma^2 = \frac{5735}{30} - \left(\frac{412}{30}\right)^2 = 2.56\hat{2}$$

$$\sigma = \sqrt{\sigma^2} = 1.6006$$

La media es $\bar{x} = 13.7\hat{3}$, la varianza es $\sigma^2 = 2.56\hat{2}$ y la desviación típica es: $\sigma = 1.6006$.

Podemos decir que, más o menos, el número de pacientes que ha recibido el médico ha sido, más o menos de 12-13-14.

CON LA CALCULADORA:

Necesitamos poner la calculadora en modo STAT pulsando **2ndF** **STAT ON/F**

Introducir los datos. Por ejemplo: 3, 4, 7, 9, 12 → 3 **DATA M+**; 4 **DATA M+**; 7 **DATA M+**; 9 **DATA M+**; 12 **DATA M+**

Ahora podemos conocer las diferentes medidas:

Para saber:

\bar{x} → necesitamos pulsar $\boxed{x \rightarrow M}$ → en la pantalla aparece $\bar{x} = 7$

σ → necesitamos pulsar $\boxed{2ndF} \boxed{RM}$ → en la pantalla aparece $\sigma = 3.29$

$\sum x$ → necesitamos pulsar $\boxed{2ndF} \boxed{\int}$ → en la pantalla aparece $\sum x = 35$

$\sum n$ → necesitamos pulsar $\boxed{\int}$ → en la pantalla aparece $n=5$

Problemas:

1. Hemos hecho un estudio a 100 familias para estudiar el número de personas que forman la unidad familiar y hemos obtenido la tabla siguiente:

Personas por familia	Número de familias
1	2
2	8
3	19
4	27
5	25
6	14
7	4
9	1

Construir la tabla de frecuencias

Hacer la representación gráfica de la distribución.

2. Los pesos, en kg, de los alumnos de un grupo de 4º ESO son:

54, 73, 52, 58, 66, 52, 53, 67, 53, 54 59, 68, 43, 60, 46, 56, 48, 62, 49, 62, 56, 50, 57, 64, 57, 51.
--

Agrupar los datos en intervalos de tamaño 5 kg, siendo el primero [40, 45), y construir una tabla de frecuencias.

Representar esta distribución de pesos mediante un histograma.

3. Las alturas de todos los alumnos del instituto de 4º ESO están recogidas en la siguiente tabla:

Altura en cm	Alumnos
155 a 160	4
160 a 165	7
165 a 170	18
170 a 175	19
175 a 180	12
180 a 185	11
185 a 190	6
190 a 195	3

Construir la tabla de frecuencias, agrupando los datos en clases y eligiendo la marca de clase adecuada.

Hacer el correspondiente histograma.

4. Las calificaciones finales de Matemáticas de los alumnos de 3º ESO vienen dadas en la tabla siguiente:

Calificaciones	EX	NT	B	SF	IN
Nº Alumnos	2	5	13	2	3

Construir una tabla de frecuencias.

Calcular el tanto por ciento de aprobados y el de suspensos.

Representar gráficamente esta distribución mediante un diagrama de barras y un diagrama de sectores.

5. En una prueba de Biología, de tipo test, de 60 cuestiones, los alumnos han contestado correctamente:

23; 41; 19; 35; 25; 49; 55; 8; 38; 46; 59; 35; 25; 28; 32; 37; 45; 41; 25; 35; 42; 51; 21; 29; 43; 39; 33; 47; 30.

Construir una tabla de frecuencias agrupando los datos en 6 intervalos.

Representar gráficamente esta distribución.

6. Hemos hecho un test de 60 preguntas a 100 alumnos. En la siguiente tabla recogemos el número de preguntas que han contestado correctamente:

Número de preguntas	Número de alumnos
0-9	5
10-19	10
20-29	24
30-39	33
40-49	21
50-60	7

a) Representala gráficamente mediante un histograma. ¿Qué porcentaje de alumnos contestan bien a más de la mitad de las preguntas?

b) Hacer un estudio estadístico, con las medidas de centralización y de dispersión, y sacar las conclusiones oportunas.

7. Llenar la siguiente tabla, con los pesos de todos los alumnos de vuestra clase cuando nacisteis:

Peso en Kg	Nº de Niños/Niñas
2,5 a 2,9	
2,9 a 3,3	
3,3 a 3,7	
3,7 a 4,1	
4,1 a 4,5	

Construir la tabla de frecuencias. Representar gráficamente esta distribución.

Calcular el recorrido o rango.

Calcular el peso medio.

Calcular la desviación típica.

Sacar conclusiones acerca de este estudio.

8. Comparar los resultados académicos de dos clases que han obtenido las siguientes calificaciones finales:

CLASE A	
Insuficientes	7
Suficientes	20
Notables	8
Sobresalientes	2

CLASE B	
Insuficientes	7
Suficientes	15
Notables	4
Sobresalientes	4

a) Para hacer este estudio calcula el tanto por ciento de los alumnos que han conseguido cada una de las calificaciones y representalo gráficamente mediante un diagrama de sectores.

b) Ahora hacer el estudio con las medidas de centralización y de dispersión

9. La cantidad de dinero que llevan los 25 alumnos de una clase del instituto en un momento determinado es:

Euros	0 – 0.5	0.5 -1	1 - 1.5	1.5 - 2	2 – 2.5	2.5 - 3
Nº alumnos	2	4	2	10	4	3

Calcula la media y la desviación típica del dinero que llevan en el bolsillo estos alumnos. Escribir conclusiones a propósito de este pequeño estudio.

10. El peso medio de los alumnos de una clase del instituto es de 65 Kg. Incorporamos una persona al grupo que pesa 65 Kg, ¿cuál será ahora el nuevo peso medio? ¿Cuál es el nuevo peso medio si la persona que se incorpora pesa 70 kg? Conocemos que la persona que se incorpora es un juez, y baja la media a 64 kg. ¿Puede ser verdad?. ¿Y si a la persona que se incorpora le gustan mucho las chucherías?

11. Una familia consta de un matrimonio y cuatro hijos. Todos los miembros de la familia son trabajadores a sueldo. ¿Qué parámetro estadístico, de los seis sueldos, informa mejor de la riqueza familiar: la moda, la media o la varianza? ¿Qué parámetro estadístico informa mejor sobre la diversidad de sueldos: la mediana, la media o la desviación típica?

12. Los aumentos de precios de cinco productos alimentarios han sido, respectivamente, del 21, 31, 34, 48, 3 por ciento. ¿Qué medida de centralización refleja mejor el término medio de los aumentos, la mediana o la media aritmética? Calcular ambos valores.

13. En una clase de 4^{ta} ESO de 25 alumnos, hay 10 alumnos de francés y 15 de inglés. La nota media de los 25 alumnos en Lengua Extranjera es 6.5. Sabemos que la nota media de los alumnos de francés es de 6. ¿Qué media tienen los alumnos de inglés?

14. Calcula todas las medidas de centralización y de dispersión de las calificaciones obtenidas por los alumnos de la clase en Matemáticas:
3, 7, 8, 5, 4, 0, 6, 6, 7, 5, 2, 10, 9, 3, 4, 6, 6, 5, 4, 10, 1, 8, 7, 6, 3, 8, 6, 7, 5, 4.