Recuerda, lo que está en azul no es obligatorio copiar. Para tener unos apuntes que te ayuden a estudiar es necesario copiar la teoría y los ejercicios en orden. Haz los ejercicios a lápiz por si te equivocas poder rectificar de forma más cómoda.

TEMA 8: ESTADÍSTICA UNIDIMENSIONAL

8.1 Elementos de la Estadística unidimensional.

Es la parte de las Matemáticas que estudia los métodos de recogida, presentación, análisis e interpretación de datos. También se ocupa de hacer predicciones a la vista de éstos datos. En estos días hemos vivido claramente la necesidad de recoger datos, número de infectados por covid 19, para analizar la información y tomar decisiones. Supongo que habéis visto gráficos y de esa forma poder "predecir" el futuro.

Los elementos de la Estadística que debemos conocer son:

Población es el conjunto de todos los elementos del que se quiere estudiar una o varias características a las que llamaremos **variables.**
Podemos estudiar en la población canina de Mijas la variable, raza canina.
Podemos estudiar en la población de los habitantes de Mijas la variable, estatura.
Podemos estudiar en la población de la flota automovilística de Mijas la variable, marca.
Podemos estudiar en la población países de la OTAN la variable, renta per cápita.

Cada elemento de la población se llamará **individuo,** un individuo puede ser una persona, un perro, un país, etc.

Si la población es muy grande, resulta imposible estudiar una variable sobre cada uno de los individuos, por ello necesitaremos tomar una parte de la población sobre la que estudiaremos la variable. Esa parte de la población se llama **muestra**, de la cual intentaremos extraer conclusiones para toda la población.

Para que una muestra sea representativa es necesario:
a) la elección sea aleatoria de forma que todos los individuos tengan la misma probabilidad de ser elegidos.
b) Conservar la proporción, por ejemplo, mantener el % de hombres y mujeres.

El número de elementos de la muestra se llama **tamaño** y se denota por N.

Atendiendo al contenido de las variables se pueden clasificar en:

1) **Variables cuantitativas**: cuando las variables son números y medidas. Estas a su vez se pueden clasificar en:
   a. **Variable Cuantitativa Discreta**: cuando toma valores aislados (número de hijos en las familias, número de coches en una familia, edad de los alumnos de un instituto).
   b. **Variable Cuantitativa Continua**: cuando los valores varían en un intervalo (estatura, peso, sueldo de los empleados de una empresa).

2) **Variables Cualitativas**: cuando las variables son cualidades, por ejemplo: color de ojos, profesión, sistema de gobierno, deporte favorito.

La **estadística descriptiva** analiza algunos caracteres de los individuos de un grupo dado sin extraer conclusiones para un grupo mayor. Para ello:
1) Selecciona los caracteres dignos de estudio.
2) Análisis de cada carácter para conocer los valores que toman.
3) Completar la tabla estadística y, si fuera necesario, su representación gráfica.
4) Cálculo de los parámetros estadísticos.

En cambio, la **estadística inferencial** trabaja con muestras y a partir de ella quiere predecir (inferir) características de toda la población. Tendremos que tener mucho cuidado para elegir la muestra, el grado de confianza del resultado obtenido.

Una vez recogida la información sobre una población o muestra, necesitamos organizarla para ello utilizamos las **tablas de frecuencias** (es el paso 3 que mencionamos anteriormente). En ella mostramos en las filas las distintas **variables** (x_i) o intervalos. Y en las columnas distintos valores que nos ayudaran a estudiar las características de la variable y que vamos a definir a continuación:
a) **Frecuencia absoluta** número de veces que una variable aparece en la muestra, se denota por f_i.
b) Si la frecuencia absoluta se divide entre el número total de elementos de la muestra se llama **frecuencia relativa** y se denota $h_i = f_i/N$.
c) La **frecuencia acumulada** de x_i es el número de veces que aparece en la muestra dicha variable o variables de inferior valor. Se denota por F_i y se calcula de la siguiente forma $F_i = f_1 + f_2 + \ldots + f_i$.
d) Si la frecuencia acumulada se divide entre el número total de elementos de la muestra se llama **frecuencia relativa acumulada** y se denota $H_i = F_i/N$. La usaremos el año que viene.

Ejemplo resuelto 1: Tabla de frecuencias en una variable cuantitativa discreta. Para ello vamos a utilizar los siguientes datos: en la clase de 1º bachillerato C hay 4 alumnos no tienen hermanos, 10 alumnos que tienen 1 hermano, 12 que tienen 2 hermanos y 6 que tienen 3 hermanos.

a) ¿Quién es la población? Los alumnos de 1º bachillerato C.
b) ¿Quién es un individuo? Un alumno
c) ¿Qué variable, x_i, vamos a estudiar? El número de hermanos.
d) Completa la tabla de frecuencias:

Variable	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada	
	Moda	D sectores o %	Mediana		D sectores
x_i	f_i	h_i	F_i	H_i	$h_i \cdot 360°$
$x_1 = 0$	$f_1 = 4$	$h_1 = \frac{4}{32} = 0{,}125$	$F_1 = f_1 = 4$	$H_1 = \frac{4}{32} = 0{,}125$	45°
$x_2 = 1$	$f_2 = 10$	$h_2 = \frac{10}{32} = 0{,}3125$	$F_2 = f_1 + f_2 = 4 + 10 = 14$	$H_2 = \frac{14}{32} = 0{,}4375$	112,5°
$x_3 = 2$	(*2) $f_3 = 12$	$h_3 = \frac{12}{32} = 0{,}375$	$F_3 = f_1 + f_2 + f_3 = 14 + 12 = 26$	$H_3 = \frac{26}{32} = 0{,}8125$	135°
(*1)$x_4=3$	$f_4 = 6$	(*4) $h_4 = \frac{6}{32} = 0{,}1875$	$F_4 = f_1 + f_2 + f_3 = 26 + 6 = 32 = N$	$H_4 = \frac{32}{32} = 1$	67,5°
	(*3) N = 4 + 10 + 12 + 6 = 32	La suma siempre da 1			La suma siempre da 360°

$x_4 = 3$ significa que la cuarta variable es 3, es decir, tener 3 hermanos. f_3 significa que hay 12 personas que tienen 2 hermanos (x_3) La suma de todos los f_i siempre tiene que ser igual al tamaño de la población o la muestra. (*4) La frecuencia relativa viene a ser como el porcentaje $h_4 = 0{,}1875$, significa que 18,75% de la población tiene 3 hermanos (x_4).
Puedes ver este vídeo de youtbe https://www.youtube.com/watch?v=xq6tBKbg3HQ

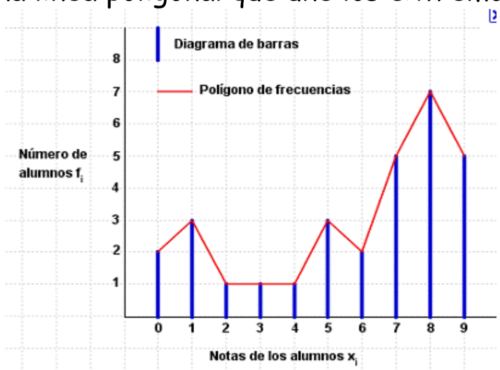
Ejemplo resuelto 2: En la clase de 1° bachillerato C se ha estudiado la paga semanal y se han obtenido los siguientes datos hay 6 alumnos que reciben menos de 10 €, 12 alumnos que reciben más de 10 pero menos de 20 y 14 que reciben más de 20 pero menos de 30.

a) ¿Quién es la población? Los alumnos de 1° de bachillerato C.
b) ¿Qué variable vamos a estudiar? Tipo de variable. La paga semanal de un alumno. Variable cuantitativa continua.
c) ¿Cómo se calcula la marca de clase? Sumamos los extremos y dividimos por 2.

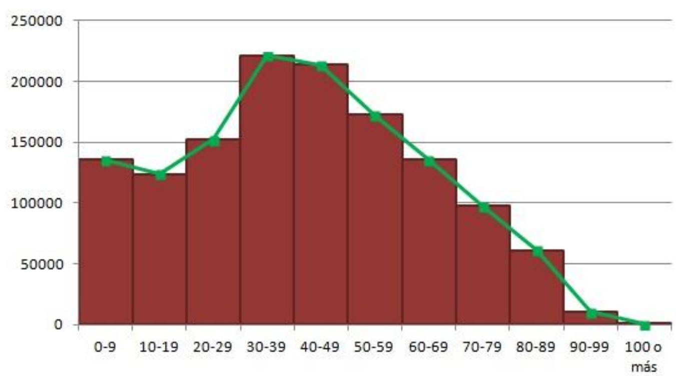
Intervalo	Marca de clase	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada	Diagrama de sectores
	$\frac{x_i + x_{i+1}}{2}$	f_i	h_i	F_i	$h_i \cdot 360°$
[0, 10)	$x_1 = 5$	$f_1 = 6$	$h_1 = \frac{6}{32} = 0{,}1875$	$F_1 = f_1 = 6$	67,5°
[10, 20)	$x_2 = 15$	$f_2 = 12$	$h_2 = \frac{12}{32} = 0{,}375$	$F_2 = f_1 + f_2 = 6 + 12 = 18$	135°
[20, 30)	$x_3 = 25$	$f_3 = 14$	$h_3 = \frac{14}{32} = 0{,}4375$	$F_3 = F_2 + f_3 = 18 + 14 = 32$	157,5°
		N = 6 + 12 + 14 = 32	La suma siempre da 1		360°

¿Qué significa f_2 = 12? Hay 12 alumnos de 1º bach C que tienen una paga semanal entre 10 y 20 €.
¿Qué significa F_2 = 18? Hay 18 alumnos de 1º bach C que tienen una paga semanal entre 0 y 20 €.
¿Qué significa h_3 = 0,4375? Hay un 43,75% de ellos que tienen una paga semanal entre 20 y 30 €.

Para mostrar los resultados obtenidos utilizamos las **representaciones gráficas**:
a) **Diagramas de barras:** Se utilizan para comparar datos cualitativos o cuantitativos discretos. La altura de la barra indica la f. absoluta. Del diagrama de barras podemos obtener el **Polígono de frecuencias** que es una línea poligonal que une los extremos de las barras del diagrama de barras.

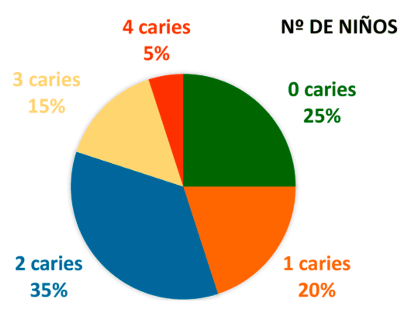


b) **Histograma** formado por rectángulos cuya base es la variable en forma de intervalo (variable cuantitativa continua) y la altura será la frecuencia absoluta, también aquí podemos trazar el polígono de frecuencias marcando el punto medio de la parte alta del rectángulo y unimos dichos puntos.



Edad de los habitantes de una comunidad.

c) **Diagrama de sectores:** se utilizan para comparar distintas modalidades de un carácter. El área del sector circular es proporcional a la frecuencia absoluta, se calcula teniendo en cuenta h_i.360. Se utiliza en variables cualitativas, cuantitativas tanto discretas como continuas, pero cuando hay pocas variables. Si son muchas variables, los "quesitos" son muy pequeños y no aportan información.



Este tipo de diagramas es muy utilizado en prensa y comunicaciones. Lo vemos para dar la información del coronavirus, elecciones electorales,...

Para hacer un diagrama de sectores podéis ver el siguiente vídeo, el problema es que dibuja con transportador de ordenador, pero explica muy bien la tabla https://www.youtube.com/watch?v=aaV-ROo2Z8M

Para ver cómo utilizar el transportador de ángulos, mira el siguiente vídeo poco antes del minuto 8 https://www.youtube.com/watch?v=J6rf4MxCiqQ

Si nos dan muchos datos es necesario agruparlos en intervalos, es decir, se trata de una variable cuantitativa continua. Para organizarlos en intervalos seguiremos los siguientes pasos:

1) La diferencia entre el valor más pequeño y el mayor, calculamos b – a, llamado recorrido.
2) Se decide el número de intervalos (no debe ser menor de 6 ni mayor de 15, a veces viene indicado en el enunciado del ejercicio).
3) Aumentamos b – a hasta un número que sea múltiplo del número de intervalos.
4) El extremo inferior del primer intervalo será menor o igual que a y el extremo superior del último intervalo es mayor o igual que b. Los iguales dependerán de donde tenemos el corchete (todos los corchetes a la izquierda o todos a la derecha).

Ejercicio 1: En un bloque de 25 viviendas se estudia el número de coches por familia obteniendo los siguientes datos: 0 2 1 2 3 1 2 1 3 0 2 1 1 1 0 4 3 1 1 0 2 2 4 1 1. Completa su tabla estadística:

x_i	f_i	h_i	F_i	$h_i \cdot 360°$		
$x_1 = 0$	$f_1 =$					
$x_2 = 1$	$f_2 =$					
$x_3 = 2$	$f_3 =$					
$x_4 = 3$	$f_4 =$					
$x_5 = 4$	$f_5 =$					
	$N =$					

a) ¿Cuál es la población?
b) ¿Cuál es la variable que se estudia?
c) ¿Qué tipo de variable es?
d) ¿Cuántas variables tiene el estudio?
e) ¿Qué significa $x_3 = 2$?
f) ¿Qué significa $f_3 = 6$?
g) ¿Qué significa $F_3 = 20$?
h) ¿Qué porcentaje de las familias tienen 3 coches?
i) Dibuja su diagrama de barras y su polígono de frecuencias.

j) Dibuja su diagrama de sectores, no olvides poner al lado del diagrama la relación entre los colores y la variable.

Ejercicio 2: En el mismo bloque de 25 viviendas se estudia los sueldos de cada una de las familias (si hay varias personas que trabajan en la misma familia, se suman los sueldos) obteniendo:

	x_i	f_i	h_i	F_i	$h_i \cdot 360°$		
[500, 1000)	x_1 =	$f_1 = 3$					
[1000, 1500)	x_2 =	$f_2 = 8$					
[1500, 2000)	x_3 =	$f_3 = 12$					
[2000, 2500)	x_4 =	$f_4 = 2$					
		N = 25					

a) ¿Cuál es la población?
b) ¿Cuál es la variable que se estudia?
c) ¿Qué tipo de variable es?
d) Completa su tabla estadística (recuerda hasta cuatro decimales).
e) ¿Qué significa $f_3 = 12$?
f) ¿Qué significa $h_3 = 23$?
g) ¿Qué porcentaje de las familias tienen un sueldo menor a 1000 €?
h) Dibuja su histograma y su polígono de frecuencias.
i) Dibuja su diagrama de sectores, no olvides poner al lado del diagrama la relación entre los colores y la variable.

8.2 PARÁMETROS ESTADÍSTICOS.

Cuando quiero estudiar una característica, hemos aprendido a "ordenar" los datos, ¿Qué significan? ¿Qué conclusiones podemos extraer? Para obtener información de una distribución o tabla se buscan una serie de valores que se llaman **parámetros estadísticos**. Hay dos tipos:

<u>Parámetros de centralización</u> nos indica en torno a qué valor (centro) se distribuyen los datos. Hay varios tipos:

1) **Media aritmética,** se denota por $\overline{x}$, y se calcula mediante la siguiente fórmula $\overline{x} = \frac{\sum f_i x_i}{N}$.

2) **Moda** es el valor de la variable que presenta una mayor frecuencia. Se denota por Mo.
3) **Mediana** es el valor de la variable que cumple que la mitad de los valores están por debajo de él y la otra mitad por encima. Si el número de datos es par, a la mediana se le asigna el valor medio de los dos términos centrales. Se denota por Me. La mediana se encuentra en la variable x_k si $F_k > \frac{N}{2}$. Si $F_k = \frac{N}{2}$ entonces Me = $\frac{x_k + x_{k+1}}{2}$.

Parámetros de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución, son necesarias para aquellas distribuciones en que la media no es el valor más representativo por ejemplo si hay sueldos bajos y uno muy alto. Para ello tenemos:

1) **Recorrido** es la diferencia entre el mayor y el menor de los valores de la variable.

2) **Varianza** $\sigma^2 = \frac{\sum f_i x_i^2}{N} - \bar{x}^2$

3) **Desviación típica** es la raíz cuadrada positiva de la varianza. Cuánto mayor sea, mayor es la dispersión. Entre el 50% y el 70% de los datos se encuentran en el intervalo ($\bar{x}$ - σ $\bar{x} + \sigma$) y el 90% en el intervalo ($\bar{x}$ - 2σ $\bar{x} + 2\sigma$)

4) **Coeficiente de variación** se denota por C.V. = $\frac{\sigma}{\bar{x}}$. Nos sirve para comparar dos o más distribuciones. Cuánto más cerca está del 0 menos dispersa es la distribución.

Ejemplo 3: Este ejemplo complementa al ejemplo que nos hablaba del número de hermanos del alumnado de 1º bach C. Vamos a completar la tabla estadística y calcularemos sus parámetros de centralización y dispersión.

x_i	f_i	F_i	$h_i \cdot 360º$	$x_i\, f_i$	$x_i^2\, f_i$		
$x_1 = 0$	$f_1 = 4$	$F_1 = 4$	45º	0	0		
$x_2 = 1$	$f_2 = 10$	$F_2 = 14$	112,5º	10	10		
$x_3 = 2$	$f_3 = 12$	$F_3 = 26$	135º	24	48		
$x_4 = 3$	$f_4 = 6$	$F_4 = 32 = N$	67,5º	18	54		
	N = 32		360º	52	112		

Calculemos sus medidas de centralización

Media $\bar{x} = \frac{52}{32} = 1{,}4857$ Moda Mo = 2, es decir, la moda es "tener 2 hermanos"

Mediana Me = 2 porque $F_3 = 26 > N/2$.

Calculemos sus parámetros de dispersión

Recorrido es 3 – 0 = 3 Varianza $\sigma^2 = \frac{112}{32} - (1{,}4857)^2 = 1{,}2927$ **no lo borres de la calculadora, lo necesitas para calcular el siguiente parámetro** $\sigma = \sqrt{\text{varianza}} = 1{,}1370$

($\bar{x}$ - σ $\bar{x} + \sigma$) = (0,193, 2,7784) significa que entre el 50 y 70 % tienen 1 ó 2 hermanos.

Coeficiente de variación C.V. = $\frac{\sigma}{\bar{x}} = \frac{1{,}137}{1{,}4857}$ = 0,7653

Puedes repasar los parámetros de centralización en el siguiente vídeo https://www.youtube.com/watch?v=CrItHF8aJ3M y los parámetros de dispersión en https://www.youtube.com/watch?v=Vg5PD7FROgo

Ejemplo 4: Este ejemplo complementa al ejemplo 2 que nos hablaba de la paga semanal del alumnado de 1º bach C. Completa la tabla estadística y calcula sus parámetros de centralización y dispersión.

	x_i	f_i	h_i	F_i	$h_i \cdot 360°$	$x_i\, f_i$	$x_i^2\, f_i$
[0, 10)	$x_1 = 5$	$f_1 = 6$	$h_1 = 0{,}1875$	$F_1 = 6$	67,5º	30	150
[10, 20)	$x_2 = 15$	$f_2 = 12$	$h_2 = 0{,}375$	$F_2 = 18$	135º	180	2700
[20, 30)	$x_3 = 25$	$f_3 = 14$	$h_3 = 0{,}4375$	$F_3 = 32$	157,5º	350	8750
		N = 32	1		360º	560	11600

Calculemos sus medidas de centralización: Media $\overline{x} = \frac{560}{32} = 17{,}5$ Intervalo modal es [20, 30)

Intervalo Mediano es [10, 20) porque $F_2 = 18 > N/2$.

Calculemos sus parámetros de dispersión : Recorrido es 30 – 0 = 30 Varianza $\sigma^2 = \frac{11600}{32} - (17{,}5)^2 = 56{,}25 \rightarrow \sigma = \sqrt{\text{varianza}} = 7{,}5$ $(\overline{x} - \sigma \quad \overline{x} + \sigma)$ = (48,75, 63,75) significa que entre el 50 y 70 % tienen una paga entre (48,75, 50). Coef de variación C.V. = $\frac{\sigma}{\overline{x}} = \frac{7{,}5}{17{,}5}$ = 0,4286, por tanto la paga semanal está menos dispersa que el número de hermanos

Para repasar los parámetros de una variable continua tenéis el siguiente vídeo https://www.youtube.com/watch?v=ISbnLcFFrNY , es suficiente con ver los cinco primeros minutos, los demás son de percentiles que no vamos a dar.

Ejercicio 3: En el ejercicio 3 se estudió en un bloque de 25 viviendas el número de coches por familia, utilizando la misma tabla añade dos columnas $x_i\, f_i$ y $x_i^2\, f_i$.

a) Completa las columnas $x_i\, f_i$ y $x_i^2\, f_i$.
b) Calcula los parámetros de centralización: media, mediana y moda.
c) Calcula los parámetros de dispersión: varianza, desviación típica y coeficiente de variación.
d) Calcula el intervalo $(\overline{x} - \sigma \quad \overline{x} + \sigma)$

Ejercicio 4: En el ejercicio 4 estudiamos los sueldos de 25 viviendas, utilizando la misma tabla añade dos columnas $x_i\, f_i$ y $x_i^2\, f_i$.

a) Completa las columnas $x_i\, f_i$ y $x_i^2\, f_i$.
b) Calcula los parámetros de centralización: media, intervalo mediano e intervalo modal.
c) Calcula los parámetros de dispersión: varianza, desviación típica y coeficiente de variación.
d) Compara la dispersión de los ejercicios 5 y 6.

**En este documento tenemos la teoría y 4 ejercicios. Para completarla necesitamos más ejercicios que están en un documento distinto ya que trabajaremos con el folio apaisado.