

UNIDAD 8. ESTADÍSTICA

Clasificación de las variables estadísticas

Se define una **variable estadística** como la característica que se estudia de cada uno de los individuos de la muestra en la que se aplica el estudio que se realiza.

Si la variable estadística se expresa en números entonces se llama **variable cuantitativa** si no **variable cualitativa**.

Las variables numéricas se clasifican en **discretas y continuas**.

Una variable es **discreta** si los diferentes valores que toma están aislados. Por ejemplo, la edad de un grupo de personas.

Una variable es **continua** si puede tomar todos los valores comprendidos en un intervalo determinado. Por ejemplo, las alturas de los alumnos de una clase.

Las variables cualitativas se clasifican en **nominales y ordinales**.

Una variable cualitativa es **nominal** si los diferentes caracteres de la variable no son ordenables. Por ejemplo, el estado civil de un grupo de personas.

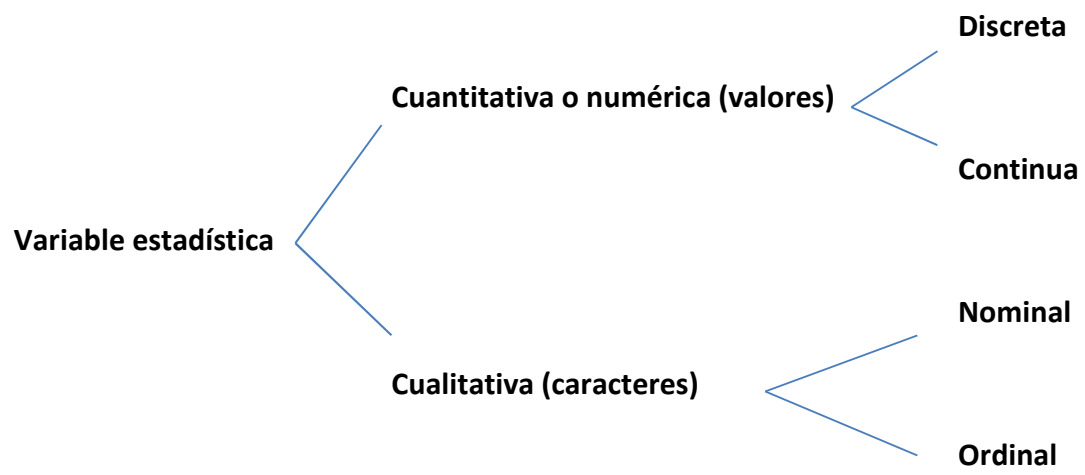
Una variable cualitativa es **ordinal** si los diferentes caracteres de la variable son ordenables. Por ejemplo, las notas de evaluación de matemáticas de una clase.

Edad	Frecuencia absoluta
15	4
16	6
17	8
18	2
	20

Altura	Frecuencia absoluta
(165,170]	3
(170,175]	7
(175,180]	8
(180,185]	2
	20

Estado civil	Frecuencia absoluta
Casado	30
Soltero	25
Separado	15
Viudo	10
	80

Calificaciones	Frecuencia absoluta
I	10
S	7
B	4
N	2
E	2
	25



Frecuencias acumuladas

N : número de individuos de la muestra

F : frecuencia absoluta

f : frecuencia relativa $f = \frac{F}{N}$

Edad	Frecuencia absoluta	Frecuencia relativa	Tanto por ciento
15	4	0,2	20
16	6	0,3	30
17	8	0,4	40
18	2	0,1	10
	20	1	100

La **frecuencia absoluta acumulada** de un determinado valor de la variable es el resultado de sumar las frecuencias absolutas de los resultados anteriores.

La **frecuencia relativa acumulada** de un determinado valor de la variable es el resultado de sumar las frecuencias relativas de los resultados anteriores.

EL **tanto por ciento acumulado** de un determinado valor de la variable es el resultado de sumar los tantos por cientos de los resultados anteriores.

Edad	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada	Tanto por ciento	Tanto por ciento acumulado
15	4	4	0,2	0,2	20	20
16	6	10	0,3	0,5	30	50
17	8	18	0,4	0,9	40	90
18	2	20	0,1	1	10	100
	20		1		100	

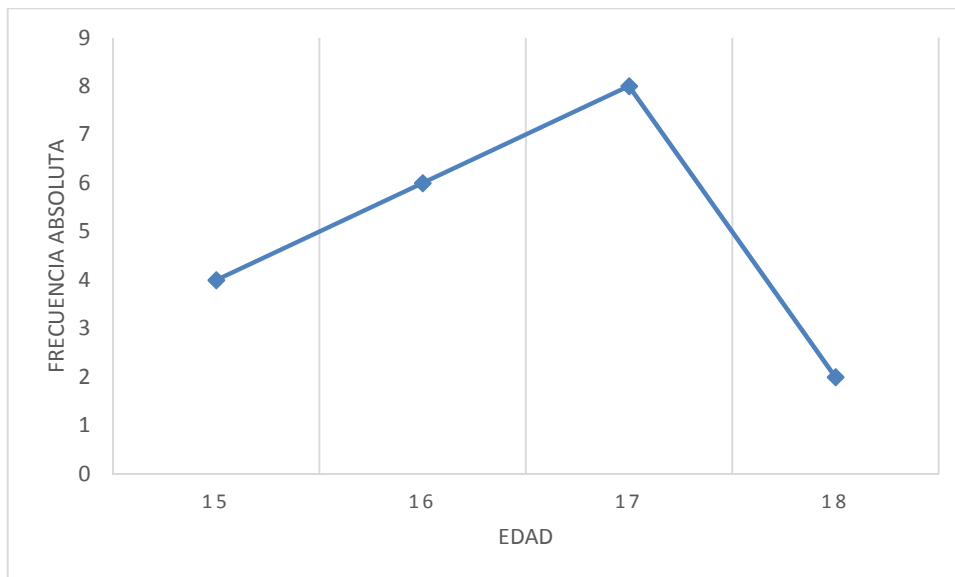
Altura	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia Relativa acumulada	Tanto por ciento	Tanto por ciento acumulado
(165,170]	3	3	0,15	0,15	15	15
(170,175]	7	10	0,35	0,5	35	50
(175,180]	8	18	0,4	0,9	40	90
(180,185]	2	20	0,1	1	10	100
	20		1		100	

Polígonos de frecuencias y histogramas

Considero una variable cuantitativa discreta

Edad	Frecuencia absoluta
15	4
16	6
17	8
18	2
	20

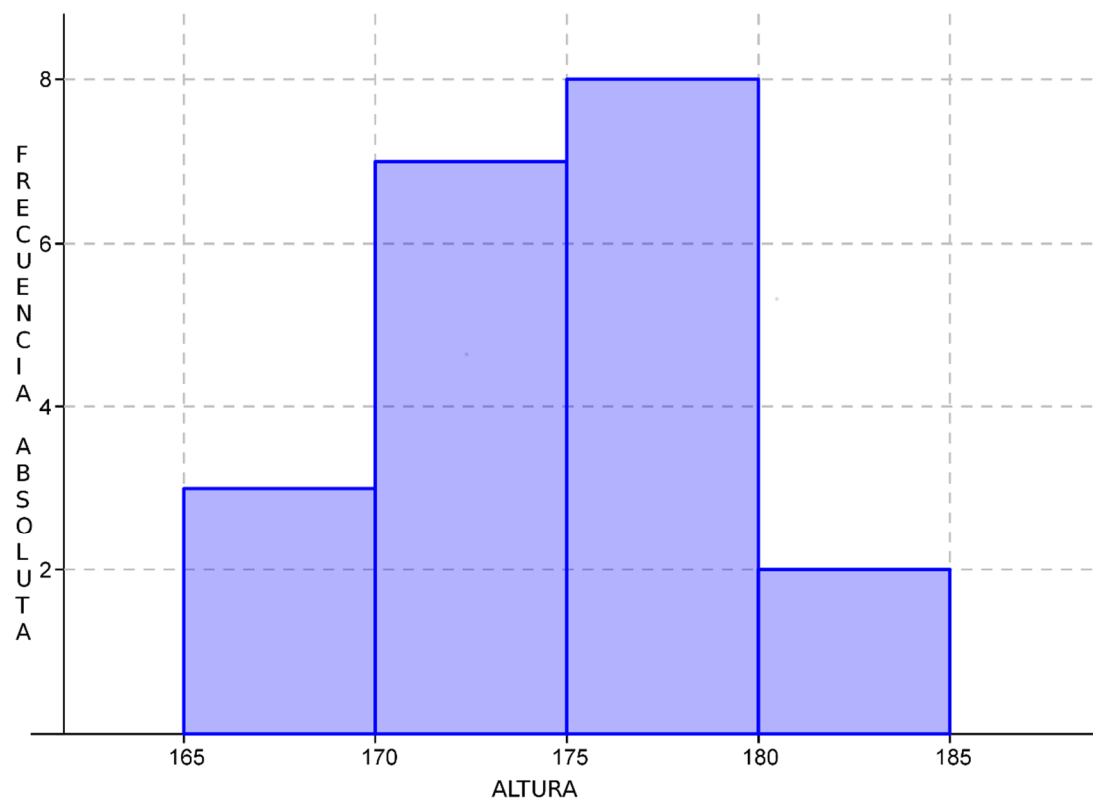
El **polígono de frecuencias absolutas** se obtiene uniendo con una poligonal los puntos definidos por las frecuencias absolutas



Considero una variable cuantitativa continua

Altura	Frecuencia absoluta
(165,170]	3
(170,175]	7
(175,180]	8
(180,185]	2
	20

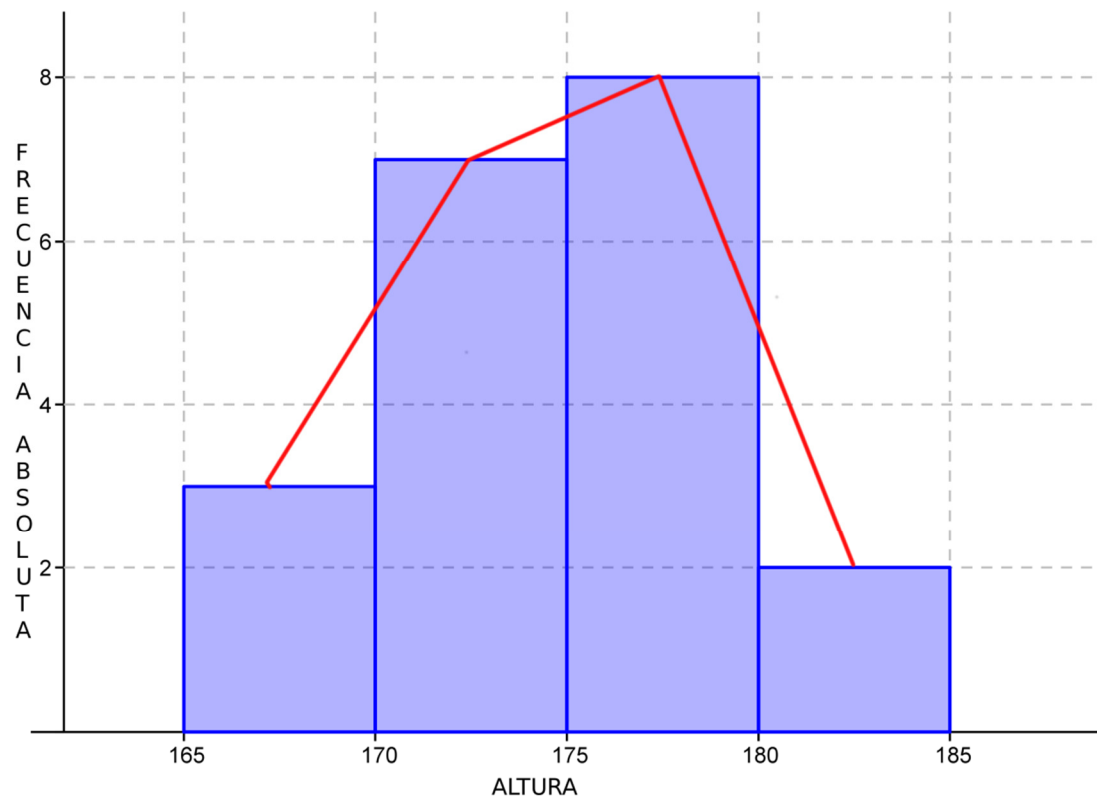
El **histograma de frecuencias absolutas** se obtiene dibujando rectángulos de altura la frecuencia absoluta y base en los extremos de cada intervalo.



Las **marcas de clase** de los intervalos son los puntos medios de los extremos.

Intervalo	Marca de clase
(165,170]	167,5
(170,175]	172,5
(175,180]	177,5
(180,185]	182,5

Podemos superponer en el histograma el polígono de frecuencias definido con las marcas de clase.



Parámetros de centralización de una variable discreta

La **moda** es el valor más frecuente de la variable.

La **mediana** es el valor que ocupa la posición central de todos los valores de la variable una vez ordenados estos en orden creciente.

La **media** es el resultado que se obtiene al sumar todos los valores de la variable dividido entre el número de individuos de la muestra.

Edad	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada	Tanto por ciento	Tanto por ciento acumulado
15	4	4	0,2	0,2	20	20
16	5	9	0,25	0,45	25	45
17	9	18	0,45	0,9	45	90
18	2	20	0,1	1	10	100
	20		1		100	

La **moda** es **17**

La **mediana** es **17**

Cálculo de la **media**

$$\frac{15 \cdot 4 + 16 \cdot 5 + 17 \cdot 9 + 18 \cdot 2}{20} = 15 \cdot 0,2 + 16 \cdot 0,25 + 17 \cdot 0,45 + 18 \cdot 0,1 = \mathbf{16,45}$$

Parámetros de centralización de una variable continua

La **moda o clase modal** es el intervalo más frecuente de la variable.

La **mediana** es el valor que ocupa la posición central de la variable.

La **media** es el resultado que se obtiene al sumar todos los valores de la variable dividido entre el número de individuos de la muestra.

Altura	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia Relativa acumulada	Tanto por ciento	Tanto por ciento acumulado
(165,170]	3	3	0,15	0,15	15	15
(170,175]	5	8	0,25	0,4	25	40
(175,180]	10	18	0,5	0,9	50	90
(180,185]	2	20	0,1	1	10	100
	20		1		100	

La **clase modal** es el intervalo (175,180]

Cálculo de la **media**:

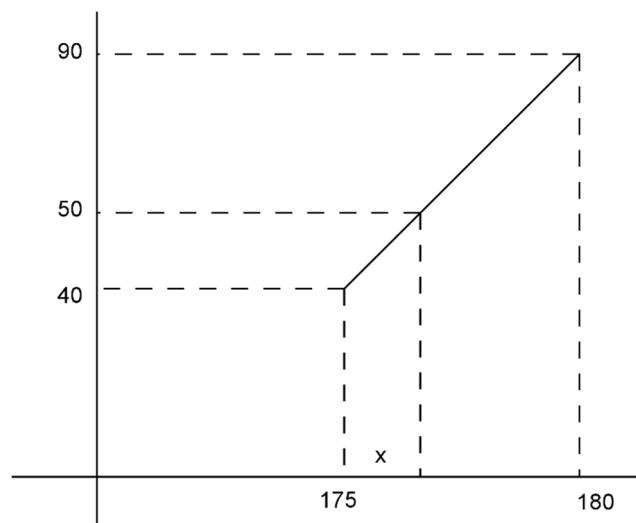
Calculamos como en el caso discreto tomando las marcas de clase

$$\frac{167,5 \cdot 3 + 172,5 \cdot 5 + 177,5 \cdot 10 + 182,5 \cdot 2}{20} =$$

$$= 167,5 \cdot 0,15 + 172,5 \cdot 0,25 + 177,5 \cdot 0,5 + 182,5 \cdot 0,1 = \mathbf{175,25}$$

Cálculo de la **mediana**:

La mediana se encuentra en el intervalo (175,180] . Por lo tanto la mediana es $175 + x$



Para calcular x planteamos la siguiente proporción: $\frac{180-175}{90-40} = \frac{x}{50-40} \Rightarrow x = \frac{5 \cdot 10}{50} = 1$

La **mediana** es $175 + 1 = \mathbf{176}$