4.- Tablas estadísticas

Al considerar sólo un carácter de tipo cuantitativo en una población, estamos subordinando dicho estudio, al de una única variable, y diremos que estamos realizando un **estudio unidimensional** de dicha población.

Una vez que se poseen los datos numéricos relativos a dicha variable, deberemos presentarlos de forma **ordenada, coherente y sistemática**. Este es el objetivo de la utilización de tablas y gráficos estadísticos.

Según el tipo de variable y el número de datos que se manejan, aparecen los distintos tipos de tablas, siendo las más usuales las siguientes:

- **Tablas de tipo I**

Si la variable tiene pocas modalidades y son pocos los datos que manejamos, nos basta con presentar estos ordenadamente.

Ejemplo:

Si estamos estudiando el número de hijos de los cinco vecinos de un bloque de viviendas, presentaremos los datos escribiendo:

0, 1, 2, 2, 4

- **Tablas de tipo II**

Se utilizan cuando manejamos una variable discreta con pocas modalidades, pero contamos con un número considerable de datos.

Ejemplo:

Si tenemos el número de días que han faltado al trabajo en un trimestre los 50 empleados de una empresa

1	0	0	3	1	2	2	1	3	0
0	1	2	1	1	3	1	0	0	1
0	0	0	2	0	0	5	1	0	0
4	1	1	3	1	0	0	0	2	1
1	1	1	0	0	0	4	3	0	2

presentamos los datos en la siguiente tabla:

x_i	n_i
0	20
1	16
2	6
3	5
4	2
5	1
	N=50

$\mathbf{x_i}$ representa los valores correspondientes a cada modalidad de la variable, $\mathbf{n_i}$ el número de individuos de la población que presentan la modalidad $\mathbf{x_i}$, y **N** el número total de individuos.

También se utilizan para presentar datos correspondientes a atributos.
Ejemplo:
Si estudiamos la nacionalidad de los turistas que pasan sus vacaciones en una determinada región de la costa española, podemos hacer la siguiente presentación:

Nacionalidad de turistas	
Nacionalidad	n_i
Alemana	252
Francesa	144
Inglesa	175
Italiana	96
Otras	19
	N=690

- **Tablas de tipo III**

Se usan cuando manejamos variables discretas con un número elevado de modalidades, o bien variables continuas, cuando también es grande el número de datos de que disponemos
Ejemplo:
Si tenemos las estaturas de un grupo de 50 alumnos

1,68 1,65 1,82 1,60 1,72 1,58 1,57 1,60 1,79 1,68
1,92 1,65 1,72 1,55 1,55 1,65 1,70 1,73 1,60 1,65
1,88 1,64 1,81 1,78 1,60 1,80 1,82 1,80 1,83 1,68
1,60 1,72 1,60 1,66 1,65 1,68 1,66 1,60 1,64 1,75
1,65 1,70 1,75 1,65 1,70 1,63 1,68 1,70 1,55 1,58

podemos presentar los datos en la forma:

l_i	x_i	n_i
[1,55-1,60)	1,575	5
[1,60-1,65)	1,625	11
[1,65-1.70)	1,675	13
[1,70-1,75)	1,725	8
[1,75-1,80)	1,775	5
[1,80-1,85)	1,825	6
[1,85-1,90)	1,875	1
[1,90-1,95)	1,925	1
		N=50

Para elaborar una tabla de tipo III seguiremos el siguiente esquema:

- Señalaremos los valores **máximo M** y **mínimo m** de la variable. La diferencia entre ambos **R=M-m** se denomina **rango o recorrido** de la misma. En el ejemplo M=1,92, m=1,55 y R=1,92-1,55=0,47
- Dividiremos el rango de la variable en intervalos de la misma longitud, siempre que sea posible, salvo que la propia naturaleza de los datos, o los fines del estudio que estamos haciendo requieran lo contrario. Los intervalos se denominan **intervalos de clase.** En el ejemplo hemos dividido el rango en los 8 intervalos que figuran en la columna izquierda de la tabla.
- La clasificación debe ser **exhaustiva** y **mutuamente excluyente** de forma que cada uno de los datos iniciales pertenezca a uno y sólo uno de estos intervalos. Estos no pueden solaparse, ni pude haber parte del rango que no esté contenida en uno de ellos. En el ejemplo el valor 1,85 se contabiliza dentro del intervalo [1,85-1,90) y no en el [1,80-1,85). Con el corchete indicamos que incluimos el extremo inferior, y con el paréntesis, que no incluimos el extremo superior.
- Si los valores extremos M y m, conducen a una clasificación poco manejable en la práctica, pueden sustituirse por alguno próximo, que la haga viable. En el ejemplo hemos sustituido el valor 1,92 por 1,95.
- Tomaremos como representante de todos los valores contenidos en un intervalo, el punto medio del mismo, al que representaremos por x_i y llamaremos **marca de clase.** Con esto perdemos precisión, pero ganamos en operatividad, puesto que hacemos manejables datos que inicialmente no lo eran. Cuánto mayor sea el número de intervalos considerados, mayor es la precisión en los resultados obtenidos, pero es más penoso el proceso para llegar a su obtención. En el ejemplo podemos contemplar también esta otra clasificación:

l_i	x_i	n_i
[1,55-1,65)	1,60	16
[1,65-1,75)	1,70	21
[1,75-1,85)	1,80	11
[1,85-1,95)	1,90	2
		N=50

Si estamos clasificando una población en cuatro grupos sociales, clase baja, media-baja, media-alta y alta, esta imposición puede llevarnos al estudiar el salario mensual, a la siguiente clasificación:

l_i	x_i	n_i
[0-500)	250	2500
[500-1200)	850	38000
[1200-3000)	2100	8000
[3000-15000)	9000	1500
		N=50000

El estudio del número de días de funcionamiento de una cierta componente electrónica, puede llevarnos a la siguiente tabla:

l_i	x_i	n_i
[0-20]	10	2
(20-40]	30	12
(40-50]	45	25
(50-60]	55	45
(60-90]	75	13
>90		3
		N=100

Salvo que no nos sea posible, debemos huir de estos dos tipos de tablas, puesto que condicionan mucho y hacen muy penosos los estudios posteriores.

5.- Distribuciones de frecuencia

Sea x_i una variable que presenta k modalidades ordenadas $x_1,x_2,\ldots,x_k$, y sean $n_1,n_2,\ldots,n_k$ el número de individuos de la población, que presenta cada una de las modalidades anteriores respectivamente. Cada uno de estos números diremos que es la **frecuencia absoluta** de la correspondiente modalidad. Evidentemente $n_1+n_2+\ldots+n_k = N$ siendo N el número total de individuos de la población, y lo expresaremos simbólicamente escribiendo

$$\sum_{i=1}^{k} n_i = N .$$

6.- Representaciones gráficas estadísticas

Se utilizan para expresar de forma más clara y evidente la información recogida en las tablas estadísticas. Los más importantes son:

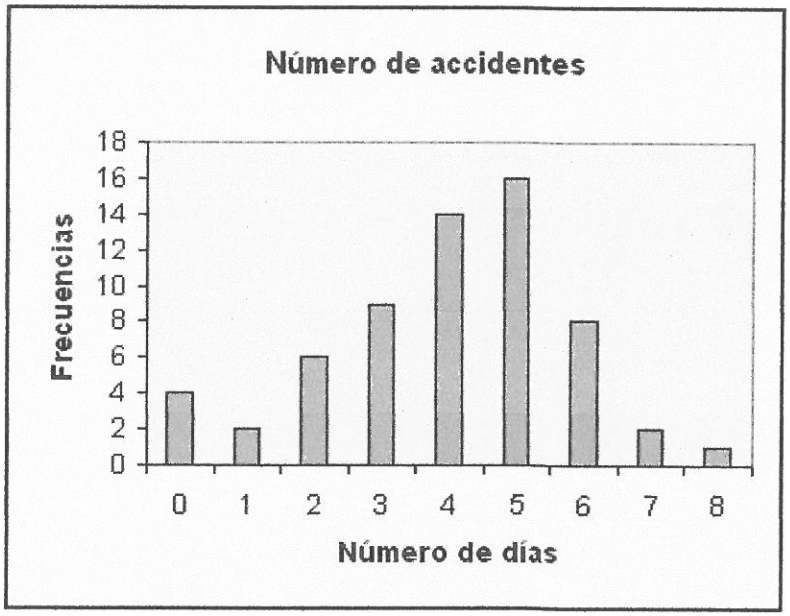
- **Diagramas de barras**

Sirven para presentar datos recogidos en tablas de tipo II. Se construyen trazando dos ejes cartesianos perpendiculares. Sobre el eje horizontal se señalan puntos, que representan a las distintas modalidades recogidas en la tabla, y sobre cada uno de ellos se dibuja un segmento perpendicular al eje, cuya **longitud** sea **directamente proporcional** a la frecuencia de cada modalidad. Se logra fácilmente construyendo previamente una escala de medida sobre el eje vertical.

Ejemplo:

Presentamos la tabla y el diagrama de barras de un estudio sobre el número de accidentes de tráfico durante dos meses, en una determinada comarca.

Número de accidentes	
x_i	n_i
0	4
1	2
2	6
3	9
4	14
5	16
6	8
7	2
8	1
N	62

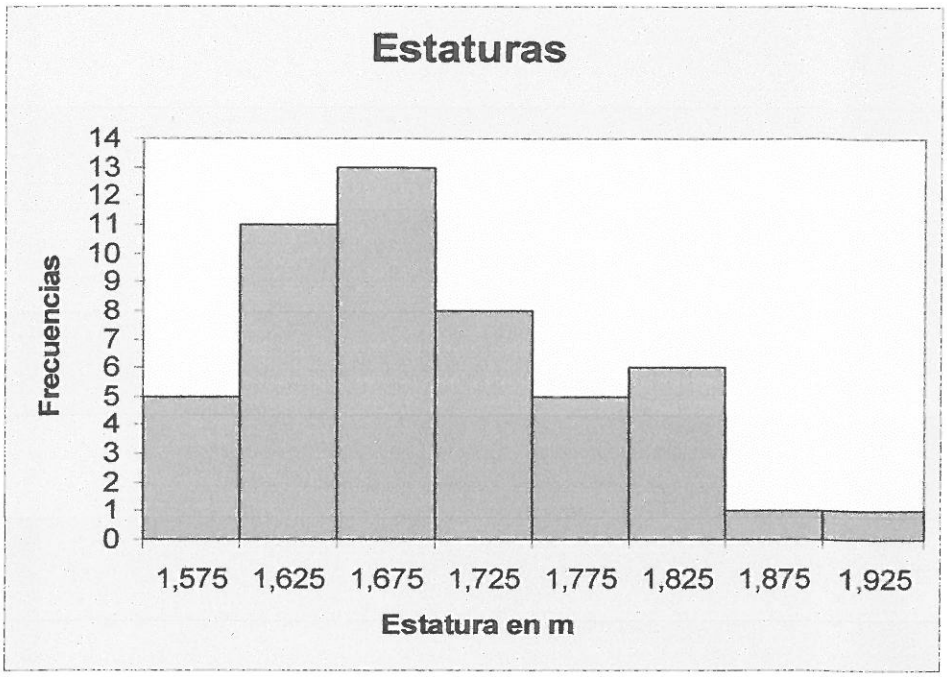


- **Histogramas**

Sirven para presentar datos recogidos en tablas de tipo III. Para construirlos se señalan sobre el eje horizontal los intervalos de clase, y tomándolos como base, se construyen sobre ellos rectángulos cuyas **áreas** sean **directamente proporcionales** a las frecuencias. Si los intervalos son de igual longitud nos basta con que las alturas de los rectángulos sean directamente proporcionales a las frecuencias, y para ello nos ayudaremos de una escala construida sobre el eje vertical.

Ejemplo:

Presentamos los datos del ejemplo de las estaturas con el histograma:

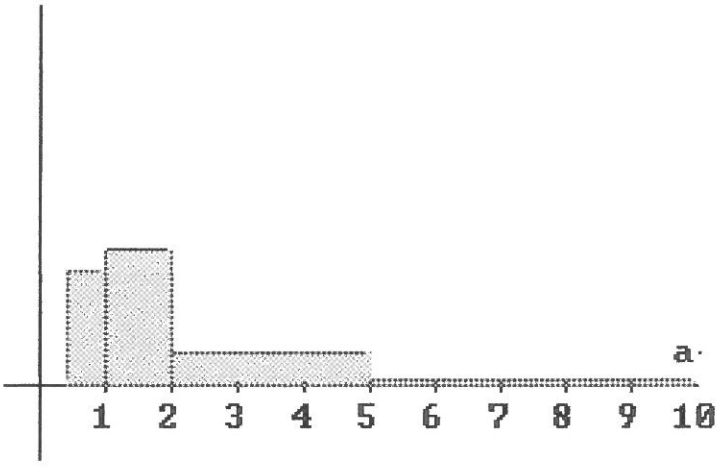


Cuando los intervalos de clase son de distinta longitud, para construir el histograma, añadimos una columna a la tabla, en la que figuren las longitudes de los intervalos, que representamos por a_i. Añadimos otra columna en la que hallamos los valores $h_i = \frac{n_i}{a_i}$, que se denominan densidades de frecuencia.

Cada rectángulo tiene por base a_i y por altura h_i. Logramos así, que las áreas de los rectángulos sean directamente proporcionales a las correspondientes frecuencias, que es la propiedad que define al histograma.

Ejemplo:

l_i	n_i	a_i	h_i
[40,100)	10	60	0,17
[100,200)	20	100	0,20
[200,500)	15	300	0,05
[500,1000]	5	500	0,01



Llamamos **frecuencia relativa** al cociente de dividir la frecuencia absoluta por el número total de individuos. La frecuencia relativa de la modalidad $\mathbf{x_i}$ es pues $f_i = \frac{n_i}{N}$. Evidentemente:

$$\sum_{i=1}^{k} f_i = f_1 + f_2 + ... + f_k = \frac{n_1}{N} + \frac{n_2}{N} + ... + \frac{n_k}{N} = \frac{n_1 + n_2 + ... + n_k}{N} = \frac{N}{N} = 1$$

Llamamos **frecuencia absoluta acumulada ascendente** a:

$$N_i^{\uparrow} = n_1 + n_2 + ... + n_i = \sum_{i=1}^{i} n_i \, .$$

Llamamos **frecuencia relativa acumulada ascendente** a:

$$F_i^{\uparrow} = f_1 + f_2 + ... + f_i = \sum_{i=1}^{i} f_i \, .$$

Estos valores se presentan añadiendo las correspondientes columnas a la tabla inicial.

Ejemplo:

l_i	x_i	n_i	N_i	f_i	F_i
[1,55-1,60)	1,575	5	5	0,10	0,10
[1,60-1,65)	1,625	11	16	0,22	0,32
[1,65-1.70)	1,675	13	29	0,26	0,58
[1,70-1,75)	1,725	8	37	0,16	0,74
[1,75-1,80)	1,775	5	42	0,10	0,84
[1,80-1,85)	1,825	6	48	0,12	0,96
[1,85-1,90)	1,875	1	49	0,02	0,98
[1,90-1,95)	1,925	1	50	0,02	1,00
		50			

- **Diagrama de sectores**

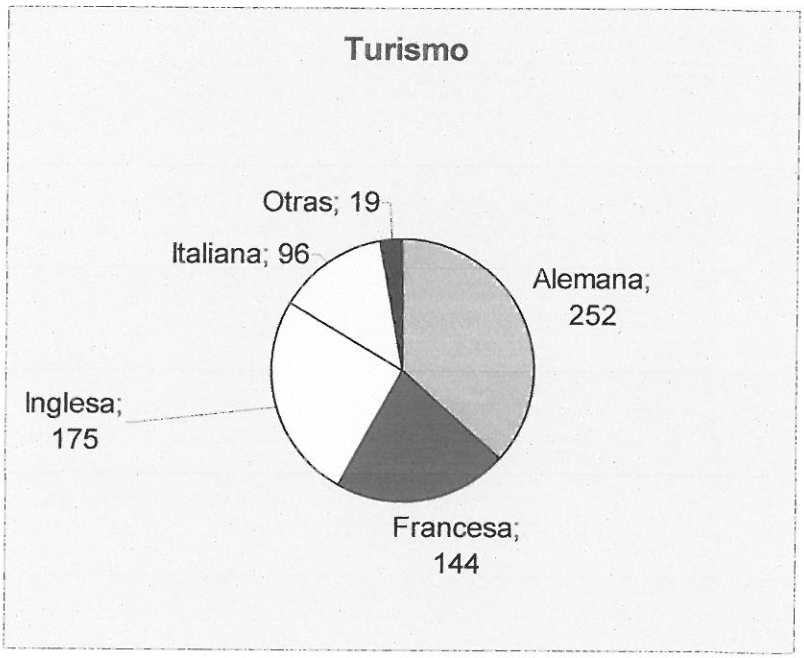
Se utilizan para representar datos recogidos en tablas de tipo II, cuando se quiere resaltar la importancia de cada modalidad con respecto al total de la población.

Para construirlos se dibuja un círculo, que representa a toda la población, y se divide en sectores circulares, cuya **amplitud** sea directamente proporcional a la frecuencia de la modalidad que representa.

Ejemplo:

Construimos el diagrama de sectores correspondiente al estudio de las nacionalidades de los turistas.

Nacionalidad de turistas		
Nacionalidad	n_i	Amplitud
Alemana	252	131,5
Francesa	144	75,1
Inglesa	175	91,3
Italiana	96	50,1
Otras	19	9,9
	N=690	



Añadimos una columna a la tabla para calcular la amplitud de cada sector aplicando la fórmula $a_i = \frac{n_i.360}{N}$.

Si queremos resaltar la importancia de cada modalidad con respecto a las demás usaremos el diagrama de barras.