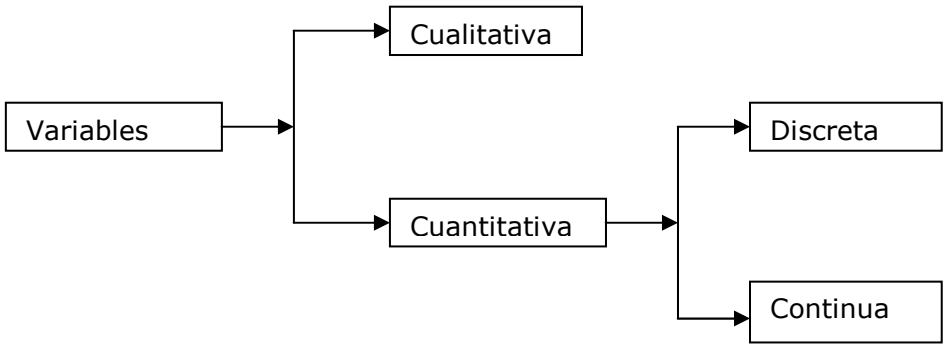
Elaboración de Tablas ó Cuadros

La elaboración de tablas o cuadros, facilita el análisis y la presentación de la información.

Para elaborar los cuadros, se debe, antes que todo, identificar las variables, características ó datos que se investigaron, lo cual permite una mejor clasificación de lo observado. Las características ó variables de un fenómeno pueden ser: cualitativas y cuantitativas.



1. Variable Cualitativa:
Denominadas también atributos, son todas aquellas características ó cualidades que se pueden describir mediante palabras. Podemos distinguir dos tipos:

a. Variable Cualitativa Nominal:
Una variable cualitativa nominal presenta modalidades que no admiten un criterio de orden. Por ejemplo: El estado civil, con las siguientes modalidades: soltero, casado, separado, divorciado y viudo. Los empleados de una empresa clasificados por cargos, las marcas de aceite de uso domestico, genero: hombre o mujer, los colores preferidos de un grupo de estudio.

Tabla 1.
Frecuencia de colores preferidos del grupo estudiado

Color	Frecuencia
Rojo	2
Azul	3
Verde	3
Morado	1
Amarillo	1

Figura 1.
Frecuencia de colores preferidos del grupo estudiado.

Se puede observar que los colores preferidos de mayor frecuencia son el Azul y el Verde, cada uno con una frecuencia de 3.

b. Variable Cualitativa Ordinal:
Una variable cualitativa ordinal presenta modalidades no numéricas, en las que existe un orden. Por ejemplo: La nota en un examen: Pésimo, malo, regular, bueno y excelente. Puesto conseguido en una prueba deportiva: 1º, 2º, 3º, Medallas de una prueba deportiva: oro, plata, bronce. Las ventas clasificadas por sucursal, por días, meses, etc.

2. Variable Cuantitativa:
Es la que se expresa mediante un número, por tanto se pueden realizar operaciones aritméticas con ella. Podemos distinguir dos tipos:

a. Variable Discreta:
Es aquella que toma valores numéricos enteros, es decir no admite valores intermedios entre dos valores específicos. Por ejemplo: El número de hermanos de 5 amigos: 2, 1, 0, 1, 3. La inasistencia a clase de un grupo de alumnos.

Tabla 2.
Frecuencia de inasistencia a clase del grupo estudiado

Inasistencia por días	Frecuencia
0 días	2
1 día	2
2 días	1
3 días	3
4 días	1
5 días	0
6 días	0
7 días	1

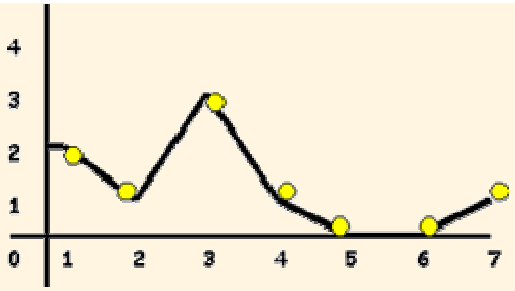


Figura 2.
Frecuencia de inasistencia a clase del grupo estudiado

Se puede observar de la Figura 2, que en la muestra de sujetos estudiados, tres días es la mayor frecuencia de inasistencia.

b. Variable Continúa:
Es aquella que puede tomar valores comprendidos entre dos números. Son consideradas como tales las expresadas en forma de medidas de superficies, peso, longitud, volumen, tiempo, temperatura, Por ejemplo: La altura de los 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.
Además las expresadas en medidas de relación tales como porcentajes, tasas, puntuaciones.

Ejercicio:

Clasifica las siguientes variables estadísticas:
a) Número de músculos de los animales vertebrados.
b) Intención de voto.
c) Velocidad que, en un instante dado, llevan las motocicletas que circulan por las carreteras y calles de una gran ciudad española.
d) Talla de pantalones de los alumnos de tu centro.
e) Tipos de zumos que prefieren los adolescentes.
f) Temperatura mínima en tu ciudad cada día del año.
g) Las marcas de los coches que circulan en España.
h) Deporte practicado por los chicos y chicas de tu centro.
i) La duración de cada pila eléctrica producida por una empresa durante un semestre.

Distribución de Frecuencias
La tabla de frecuencias tiene como finalidad presentar en forma ordenada los valores que toman las diferentes características ó variables, en tal forma que permitan al lector tener una visión del conjunto, aclarando el texto del informe o completándolo. Bajo este principio los datos se clasifican y ordenan de acuerdo a ciertas características cualitativas y/o cuantitativas, indicándose el número de veces que se repiten.

Símbolos Utilizados en las Tablas de Frecuencia

N = Tamaño poblacional.
n = Tamaño de la muestra.
x_i = Identificación para valores observados sin ordenar.
y_i = representa los valores que toma la variable ya ordenada si ésta es discreta ó la marca de clase en la variable continúa.
n_i = Frecuencia absoluta, número de veces que se repite cada valor de la variable.

h_i = Frecuencia relativa, ($n_1 i/n$) es un valor porcentual obtenido al dividir la frecuencia absoluta entre el tamaño de la muestra (n).
Ni = Frecuencia absoluta acumulada.
Hi = Frecuencia relativa acumulada.
m = Número de intervalos ó número de marcas de clase.
C = Amplitud o tamaño de cada intervalo
$Y'_{i-1} - Y'_i$ = Los intervalos en que se divide la variable continua. Siendo Y'_{i-1} el límite inferior y Y'_i el límite superior del intervalo.

Ejemplos Distribución de Frecuencia Variable Discreta:
1. Supongamos que se tienen 300 cajas en la bodega de un hospital y cada uno de ellos contiene medicamentos. Se desea examinar las cajas con el fin de saber el número de medicamentos que han sufrido imperfectos en el transporte del laboratorio hasta la bodega. Por motivos de tiempo, espacio físico y personal disponible se toma la decisión de revisar un 10% de las cajas.

N = 300 cajas (tamaño de la población)
n = 30 cajas (tamaño de la muestra)

El primer paso de cualquier estudio estadístico es recoger los datos. Cada caja seleccionada se simboliza por x_i = donde el subíndice i toma los valores desde 1 hasta n. Cada x_i tendrá como valor el correspondiente ala característica observada.

$$x_1 = 2 \quad x_6 = 3 \quad x_{11} = 0 \quad x_{16} = 2 \quad x_{21} = 2 \quad x_{26} = 2$$
$$x_2 = 1 \quad x_7 = 2 \quad x_{12} = 2 \quad x_{17} = 3 \quad x_{22} = 2 \quad x_{27} = 3$$
$$x_3 = 1 \quad x_8 = 1 \quad x_{13} = 3 \quad x_{18} = 0 \quad x_{23} = 1 \quad x_{28} = 2$$
$$x_4 = 0 \quad x_9 = 2 \quad x_{14} = 1 \quad x_{19} = 3 \quad x_{24} = 4 \quad x_{29} = 1$$
$$x_5 = 3 \quad x_{10} = 4 \quad x_{15} = 2 \quad x_{20} = 2 \quad x_{25} = 3 \quad x_{30} = 2$$

La anterior información la denominaremos datos sin agrupar ó datos originales.

Una vez que tenemos los datos recogidos, pasamos a hacer el recuento ó tabulación: determinando los valores que toma la variable y contando el número de veces que aparece cada una. Por lo tanto los valores que toma la variable son, 0, 1, 2, 3, y 4

Tabla 3. Distribución de Frecuencias

y_i	n_i	h_i (%)	N_i	H_i (%)
0	3	10	3	10
1	6	20	9	30
2	12	40	21	70
3	7	23	28	93
4	2	7	30	100
Σ	30	100		

y_i = Valores que toma la variable
n_i = Número de caja con imperfectos. La sumatoria de los datos de esta columna debe coincidir con el tamaño de la muestra.

h_i **=** Frecuencia relativa, la sumatoria de los datos debe dar el 100%

En el cálculo de las *frecuencias relativas* se efectúa de la siguiente forma: $h_i = \frac{n_i}{n}$

Entonces:

$$h_1 = \frac{n_1}{n} = \frac{3}{30} = 0{,}10 \times 100 = 10\%$$

$$h_2 = \frac{n_2}{n} = \frac{6}{30} = 0{,}20 \times 100 = 20\%$$

$$h_3 = \frac{n_3}{n} = \frac{12}{30} = 0{,}40 \times 100 = 40\%$$

$$h_4 = \frac{n_4}{n} = \frac{7}{30} = 0{,}23 \times 100 = 23\%$$

$$h_5 = \frac{n_5}{n} = \frac{2}{30} = 0{,}07 \times 100 = 7\%$$

N_i **=** La obtención de las *frecuencias absolutas acumuladas* se hace por medio de sumas sucesivas. Los datos de esta columna no se suman y la ultima frecuencia absoluta acumulada será igual al tamaño de la muestra (*n.)*

$$N_1 = n_1 = 3$$
$$N_2 = N_1 + n_2 = 3 + 6 = 9$$
$$N_3 = N_2 + n_3 = 9 + 12 = 21$$
$$N_4 = N_3 + n_4 = 21 + 7 = 28$$
$$N_5 = N_4 + n_5 = 28 + 2 = 30$$

H_i **=** Al igual que en las frecuencias absolutas acumuladas, se procede para obtención de las frecuencias relativas acumuladas. Los datos de esta columna no se suman y la última frecuencia relativa acumulada será igual al 100%

$$H_1 = h_1 = 10$$
$$H_2 = H_1 + h_2 = 10 + 20 = 30$$
$$H_3 = H_2 + h_3 = 30 + 40 = 70$$
$$H_4 = H_3 + h_4 = 70 + 23 = 93$$
$$H_5 = H_4 + h_5 = 93 + 7 = 100$$

Se podrá observar en la tabla 4, que la frecuencia absoluta nos indica, que de las 30 cajas examinadas, 3 de ellas no presentaron medicamentos defectuosos, 6 cajas presentaron 1 medicamento defectuoso, y así sucesivamente. La frecuencia relativa indica lo mismo pero términos de porcentaje así: un 10% de las cajas examinadas no presentaron medicamentos defectuosos, el 20% de las cajas tienen 1 medicamento defectuoso, etc. La frecuencia absoluta acumulada nos indica que de

las 30 cajas hay 21 de ellas que contienen hasta 2 medicamentos defectuosos que corresponden al 70% (frecuencia relativa acumulada), etc.

2. Preguntamos a 20 alumnos el número de miembros de su familia, y sus respuestas fueron:

3, 5, 4, 3, 5, 6, 8, 3, 3, 5, 7, 5, 6, 5, 4, 4, 7, 4, 5, 3

Elabora una tabla estadística completa con todos los tipos de frecuencias existentes.

Solución

Miembros por familia Y_i	**Frecuencia absoluta** n_i	**Frecuencia absoluta acumulada** N_i	**Frecuencia relativa** h_i	**Frecuencia relativa acumulada** H_i
3	5	5	0,25	0,25
4	4	9	0,2	0,45
5	6	15	0,3	0,75
6	2	17	0,1	0,85
7	2	19	0,1	0,95
8	1	**20**	0,05	**1**
Σ	20			

3. Una fábrica de gaseosas proyecta lanzar al mercado un nuevo sabor. Se realiza un test de aceptación de dicho sabor en una muestra de 30 niños, utilizando una escala de 10 puntos, para medir el grado de aceptación. Los puntos obtenidos de los 30 niños fueron:
2,6,8,7,4,5,10,6,6,7,6,7,3,8,7,6,8,6,5,4,7,8,5,7,7,6,7,7,2,7.
Se pregunta:

A. ¿Cuál es la población?
B. ¿Cuál es la muestra?
C. ¿Cuál es la variable?
D. ¿La variable es cualitativa o cuantitativa?
E. ¿Cuál es la escala de medida?
F. Construya una tabla de distribución de frecuencias
G. ¿Cuantas clases tiene la variable?
H. Interprete N_2, N_5, h_5, H_3

Solución
¿Cuál es la población? Niños
¿Cuál es la muestra? 30 niños
¿Cuál es la variable? Grado de aceptación del sabor de la gaseosa
¿La variable es cualitativa o cuantitativa? Cuantitativa discreta
¿Cuál es la escala de medida? Intervalo
Construya una tabla de distribución de frecuencias

Yi	**ni**	**hi**	**Ni**	**Hi**
2	2	7	2	7
3	1	3	3	10
4	2	7	5	17
5	3	10	8	27
6	7	23	15	50
7	10	33	25	83
8	4	13	29	97
10	1	3	30	100
Total	**30**	**100**		

¿Cuantas clases tiene la variable? Tiene 8 clases
Interprete
N_2, 3 niños dieron un puntaje entre 2 y 3
N_5, 15 niños dieron un puntaje entre 2 y 6
h_5, El 23% de los niños dieron un puntaje de 6
H_3, El 17% de los niños dieron un puntaje entre 2 y 4

Distribución de Frecuencia para Variable Continua[1]:
La distribución de frecuencias agrupadas o tabla con datos agrupados se emplea si las variables toman un número grande de valores o la variable es continua.

1. Se determina el Rango ó Recorrido: Valor máximo – valor mínimo que toma Xi

$$R = X\max - X\min$$

2. Se determina el número de intervalos ó marcas de clase *m* el cual no debe ser menor de 5 ni mayor de 16; una de las formas de obtener m es aplicando la regla de Sturges, con la cual se obtiene una aproximación aceptable, ***m*=1+3.3 Log n**, ó mediante $m = \sqrt{n}$

3. Una vez determinado el número de intervalos, se determina su amplitud mediante la formula:

$$C = \frac{X\max - X\min}{m} = \frac{R}{m}$$

Ejemplo:
Supongamos que el jefe de ventas investiga los precios (en miles $) de cierto artículo en 40 almacenes diferentes y encuentra los siguientes datos:

76	85	80	88	74	65	91	89
76	83	71	70	86	67	68	73
77	71	75	75	68	74	72	75
84	75	75	73	87	68	79	70
72	63	77	89	60	72	83	88

Solución
1. Se determina el Rango ó Recorrido:

$$R = X\max - X\min$$
$$R = 91 - 60$$
$$R = 31$$

2. Se determina el número de intervalos:

$$m = 1 + 3.3\,Log\,n$$
$$m = 1 + 3.3\,Log\,40$$
$$m = 6{,}28$$
$$m = 6$$

[1] Ciro Martínez Bencardino

3. Se determina el tamaño de cada intervalo o su amplitud:

$$C = \frac{X\max - X\min}{m} = \frac{R}{m}$$

$$C = \frac{31}{6} = 5{,}1666$$

$$C = 6$$

Para facilitar los cálculos se aproxima *C* a 6, siempre al número inmediatamente superior por pequeña que sea la fracción, por lo tanto se altera el valor del rango.

anteriormene	*ahora*
$C = \frac{31}{6} = 5{,}16$	$C = \frac{36}{6} = 6$

El rango se incrementa en 5 unidades, de 31 pasó a 36. El incremento debe ser distribuido ojalá proporcionalmente, sumando unas unidades al límite superior (91+3=94) y restándole otras al límite inferior (60-2=58).

Para la elaboración de los intervalos, se inicia con la determinación del valor de $X\min$ en el nuevo rango, siendo en nuestro caso 58, el cual se toma como límite inferior del primer intervalo, luego se procede a agregarle el valor de la amplitud para así obtener el limite superior, que será a su vez el limite inferior del segundo intervalo, al cual se le agrega nuevamente el valor de C para obtener el límite superior del segundo intervalo, y así sucesivamente hasta conformar la columna de la variable continua.

$Y'_{i-1} - Y'_i$	y_i	n_i	h_i (%)	N_i	H_i (%)
58,1 – 64	61	2	5	2	5
64,1 – 70	67	7	17,5	9	22,5
70,1 – 76	73	16	40	25	62,5
76,1 – 82	79	4	10	29	72,5
82,1 – 88	85	8	20	37	92,5
88,1 - 94	91	3	7,5	40	100
	Total	**40**	**100**		

Se observa también que a cada uno de los limites inferiores de los intervalos se les agregó 0,1 con el fin de facilitar la clasificación de cada observación, así por ejemplo $X_1 = 76$ estaría considerada en el intervalo 64,1 – 70 y no en el intervalo 70,1 – 76.

Debe quedar claro que la amplitud del intervalo sigue siendo 6 y que el 0,1 es usado únicamente como ayuda para la clasificación.

La columna simbolizada por y_i se denomina marca de clase y se puede obtener de las siguientes formas:

1. Como promedio de los límites de cada intervalo:

$$y_1 = \frac{y'_0 + y'_1}{2} = \frac{58+64}{2} = 61$$

$$y_2 = \frac{y'_1 + y'_2}{2} = \frac{64+70}{2} = 67$$

$$y_3 = \frac{y_2' + y_3'}{2} = \frac{70+76}{2} = 73$$

$$\ldots \; y_6 = \frac{y_5' + y_6'}{2} = \frac{88+94}{2} = 91$$

2. Si la amplitud del intervalo es constante, basta con determinar la primera marca de clase, de acuerdo con el método anterior, luego se le va sumando el valor de la amplitud:

$$y_1 = \frac{y_0' + y_1'}{2} = \frac{58+64}{2} = 61$$
$$y_2 = y_1 + C = 61 + 6 = 67$$
$$y_3 = y_2 + C = 67 + 6 = 73$$
$$y_4 = y_3 + C = 73 + 6 = 79$$

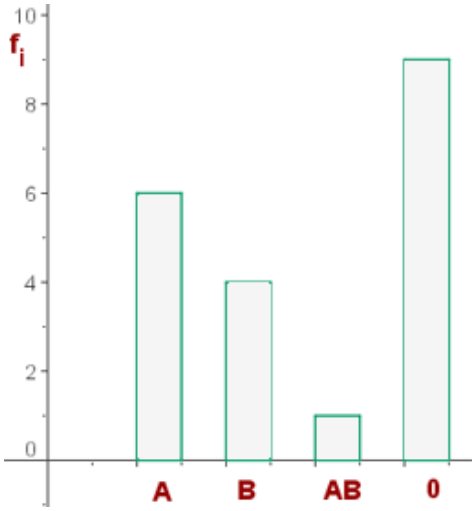
Y así sucesivamente.

Elaboración de Gráficas
Las tablas estadísticas muestran la información de forma esquemática y están preparadas para cálculos posteriores. La misma información estadística puede mostrarse de forma global y más expresiva, utilizando los **gráficos estadísticos.** Los gráficos poseen un fuerte poder de comunicación de los resultados de un estudio estadístico.

Diagrama de Barras ó Histograma
Un diagrama de barras se utiliza para representar datos cualitativos o datos cuantitativos de tipo discreto y continuo.
Se representan sobre unos ejes de coordenadas, en el eje de abscisas (X) se colocan los valores de la variable ó la amplitud del intervalo, y sobre el eje de ordenadas (Y) 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:

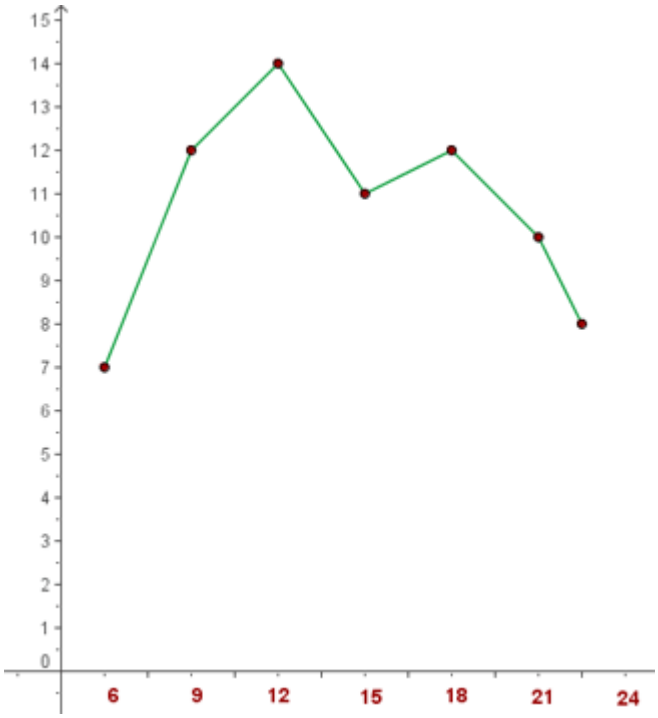
Grupo sanguíneo	fi
A	6
B	4
AB	1
0	9
	20



Polígono de Frecuencia
Un polígono de frecuencia 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.
Para construir el polígono de frecuencia se toma la marca de clase que coincide con el punto medio de cada rectángulo.

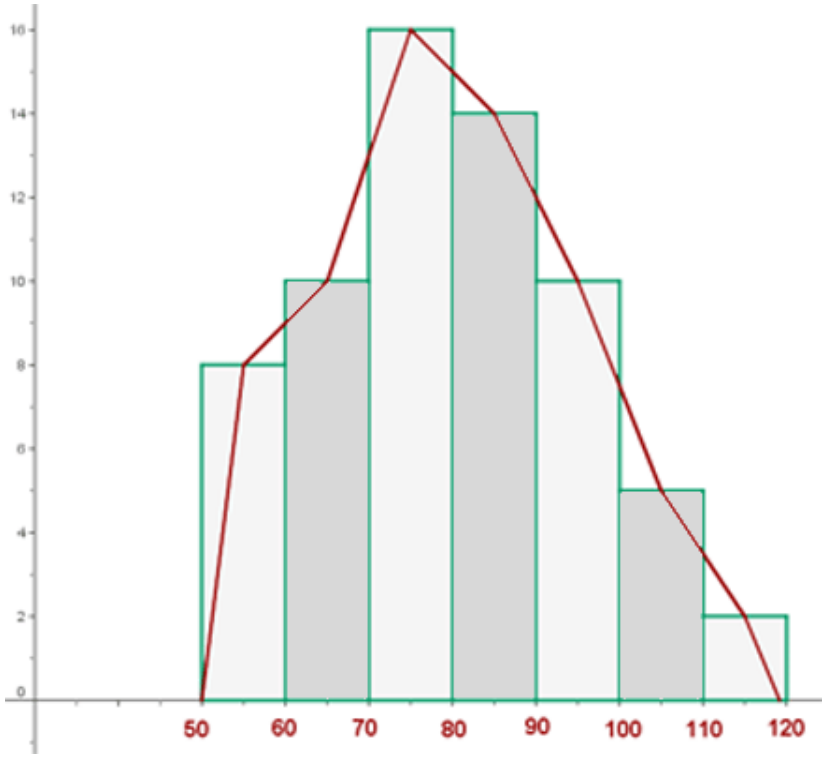
Ejemplo
1. Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

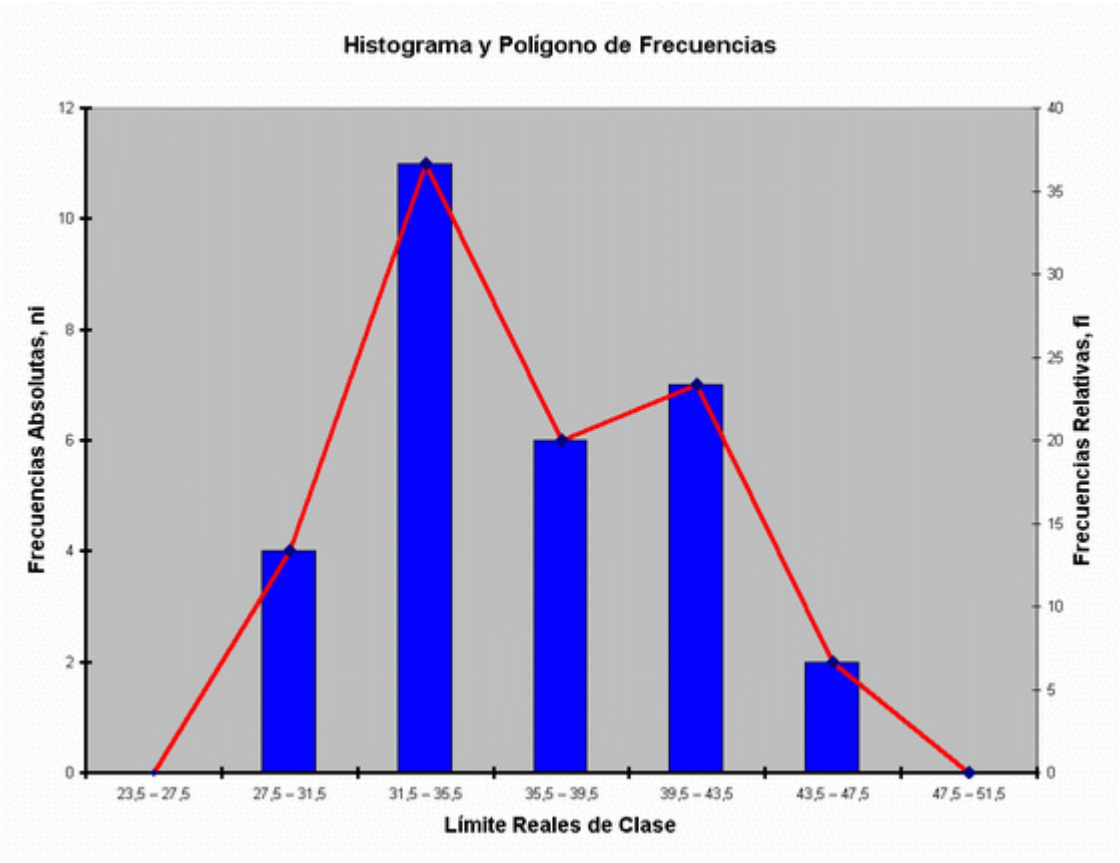
Hora	Temperatura
6	7º
9	12°
12	14°
15	11°
18	12°
21	10°
24	8°



2. El peso de 65 personas adultas viene dado por la siguiente tabla:

Intervalos	Y_i	f_i	F_i
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34
[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	105	5	63
[110, 120)	115	2	65
Total		65	


16
14
12
10
8
6
4
2
0
50
60
70
80
90
100
110
120


Histograma y Polígono de Frecuencias
Frecuencias Absolutas, ni
Frecuencias Relativas, fi
Límite Reales de Clase
23,5 – 27,5
27,5 – 31,5
31,5 – 35,5
35,5 – 39,5
39,5 – 43,5
43,5 – 47,5
47,5 – 51,5

Diagramas Acumulativos (ojivas)
En ocasiones se desea mostrar la distribución de datos, en forma acumulada. Las frecuencias acumuladas se pueden formar sobre una base "menor que" ó "mayor que" y se obtienen sumando en orden ascendente o descendente las frecuencias.

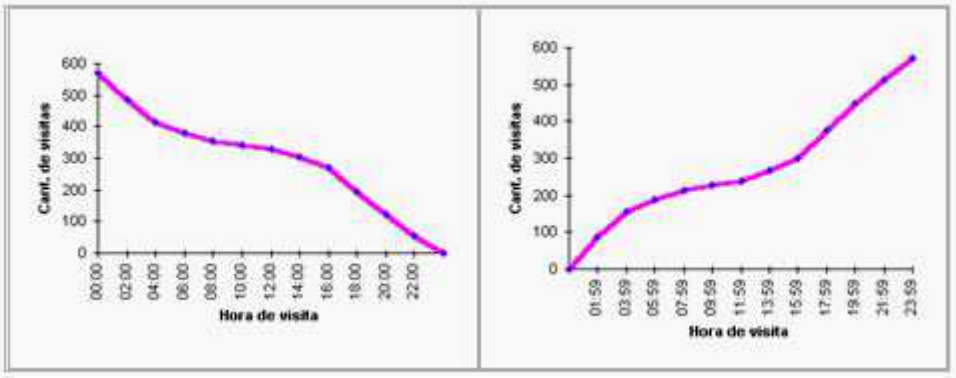


Diagrama de Sectores
Se puede utilizar para todo tipo de *variables*, pero se usa frecuentemente para las variables cualitativas.
Los datos se representan en un círculo, de modo que el ángulo de cada sector es proporcional a la frecuencia absoluta correspondiente.

$$\alpha = \frac{360°}{N} \cdot f_i$$

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

Ejemplo
En una clase de 30 alumnos, 12 juegan baloncesto, 3 practican la natación, 9 juegan al fútbol y el resto no practica ningún deporte.

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

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

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