

<http://aestudiar.wordpress.com/>

<http://www.ebookpp.com/ej/ejercicios-resueltos-de-estadistica-pdf.html>

http://www.vitutor.net/2/11/distribucion_frecuencias.html

La **Estadística** trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

Un **estudio estadístico** consta de las siguientes fases:

Recogida de datos.

Organización y representación de datos.

Análisis de datos.

Obtención de conclusiones.

Conceptos de Estadística

Población

Una **población** es el conjunto de todos los elementos a los que se somete a un estudio estadístico.

Individuo

Un **individuo** o **unidad estadística** es cada uno de los elementos que componen la población.

Muestra

Una **muestra** es un conjunto representativo de la población de referencia, el número de individuos de una muestra es menor que el de la población.

Muestreo

El **muestreo** es la reunión de datos que se desea estudiar, obtenidos de una proporción reducida y representativa de la población.

Valor

Un **valor** es cada uno de los distintos resultados que se pueden obtener en un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos dos valores: cara y cruz.

Dato

Un **dato** es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos 5 datos: cara, cara, cruz, cara, cruz.

VARIABLES ESTADÍSTICAS

Una **variable estadística** es cada una de las **características o cualidades** que poseen los **individuos de una población**.

TIPOS DE VARIABLE ESTADÍSTICAS

Variable cualitativa

Las **variables cualitativas** se refieren a **características o cualidades** que **no** pueden ser medidas con **números**. Podemos distinguir dos tipos:

Variable cualitativa nominal

Una **variable cualitativa nominal** presenta **modalidades no numéricas** que **no admiten un criterio de orden**. Por ejemplo:

El estado civil, con las siguientes modalidades: soltero, casado, separado, divorciado y viudo.

Variable cualitativa ordinal o variable cuasicuantitativa

Una **variable cualitativa ordinal** presenta **modalidades no numéricas**, en las que existe un **orden**. Por ejemplo:

La nota en un examen: suspenso, aprobado, notable, sobresaliente.

Puesto conseguido en una prueba deportiva: 1º, 2º, 3º, ...

Medallas de una prueba deportiva: oro, plata, bronce.

Variable cuantitativa

Una **variable cuantitativa** es la que se expresa mediante un **número**, por tanto se pueden realizar **operaciones aritméticas** con ella. Podemos distinguir dos tipos:

Variable discreta

Una **variable discreta** es aquella que toma **valores aislados**, es decir **no admite valores intermedios** entre dos valores específicos. Por ejemplo:

El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

Variable continua

Una **variable continua** es aquella que puede tomar **valores comprendidos entre dos números**. Por ejemplo:

La altura de los 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.

En la práctica medimos la altura con dos decimales, pero también se podría dar con tres decimales.

Una **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

Tipos de frecuencia

Frecuencia absoluta

La **frecuencia absoluta** es el **número de veces** que aparece un determinado **valor** en un estudio estadístico.

Se representa por f_i .

La **suma de las frecuencias absolutas** es igual al número total de datos, que se representa por N .

$$f_1 + f_2 + f_3 + \dots + f_n = N$$

Para indicar resumidamente estas sumas se utiliza la letra griega Σ (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^{i=n} f_i = N$$

Frecuencia relativa

La **frecuencia relativa** es el **cociente** entre la **frecuencia absoluta** de un determinado valor y el **número total de datos**.

Se puede expresar en tantos por ciento y se representa por n_i .

$$n_i = \frac{f_i}{N}$$

La suma de las frecuencias relativas es igual a 1.

Frecuencia acumulada

La **frecuencia acumulada** es la **suma de las frecuencias absolutas** de todos los **valores inferiores o iguales** al **valor** considerado.

Se representa por **F_i** .

Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**. Se puede expresar en tantos por ciento.

Ejemplo

Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

x_i	Recuento	f_i	F_i	n_i	N_i
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097

29	I	6	9	0.194	0.290
30	II	7	16	0.226	0.516
31	III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Este tipo de **tablas de frecuencias** se utiliza con **variables discretas**.

Distribución de frecuencias agrupadas

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**.

Se **agrupan** los **valores** en **intervalos** que tengan la **misma amplitud** denominados **clases**. A cada **clase** se le asigna su **frecuencia correspondiente**.

Límites de la clase

Cada **clase** está **delimitada** por el **límite inferior de la clase** y el **límite superior de la clase**.

Amplitud de la clase

La **amplitud de la clase** es la **diferencia** entre el **límite superior e inferior de la clase**.

Marca de clase

La **marca de clase** es el **punto medio** de cada **intervalo** y es el **valor** que representa a todo el **intervalo** para el **cálculo** de algunos **parámetros**.

Construcción de una tabla de datos agrupados

3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

1º se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

2º Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos de queramos poner.

Es conveniente que el número de intervalos oscile entre 6 y 15.

En este caso, $48 - 3 = 45$, incrementamos el número hasta $50 : 5 = 10$ intervalos.

Se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece intervalo, se cuenta en el siguiente intervalo.

	c_i	f_i	F_i	n_i	N_i
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10, 15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.2775

[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	42.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1
		40		1	

Diagrama de sectores

Un **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.

$$\alpha = \frac{360^\circ}{N} \cdot f_i$$

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

Ejemplo

En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 4 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360^\circ}{30} \cdot 12 = 144^\circ \quad \alpha_2 = \frac{360^\circ}{30} \cdot 3 = 36^\circ$$

$$\alpha_3 = \frac{360^\circ}{30} \cdot 4 = 48^\circ \quad \alpha_4 = \frac{360^\circ}{30} \cdot 6 = 72^\circ$$

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



Diagrama de barras

Un **diagrama de barras** se utiliza para presentar **datos cualitativos** o **datos cuantitativos de tipo discreto**.

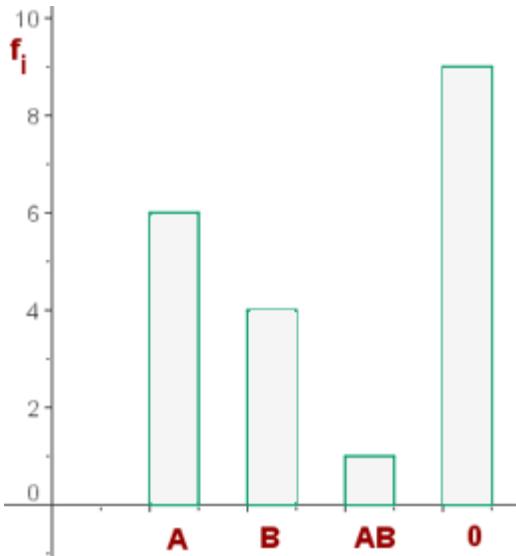
Se representan sobre unos ejes de coordenadas, en el **eje de abscisas** se colocan los **valores de la variable**, y sobre el **eje de ordenadas** las **frecuencias absolutas o relativas o acumuladas**.

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	f_i
A	6
B	4
AB	1
0	9
	20



Histograma

Un **histograma** es una **representación gráfica** de una **variable** en forma de **barras**.

Se utilizan para **variables continuas** o para **variables discretas**, con un gran número de datos, y que se han agrupado en **clases**.

En el **eje abscisas** se construyen unos **rectángulos** que tienen por **base la amplitud del intervalo**, y por **altura**, la **frecuencia absoluta** de cada **intervalo**.

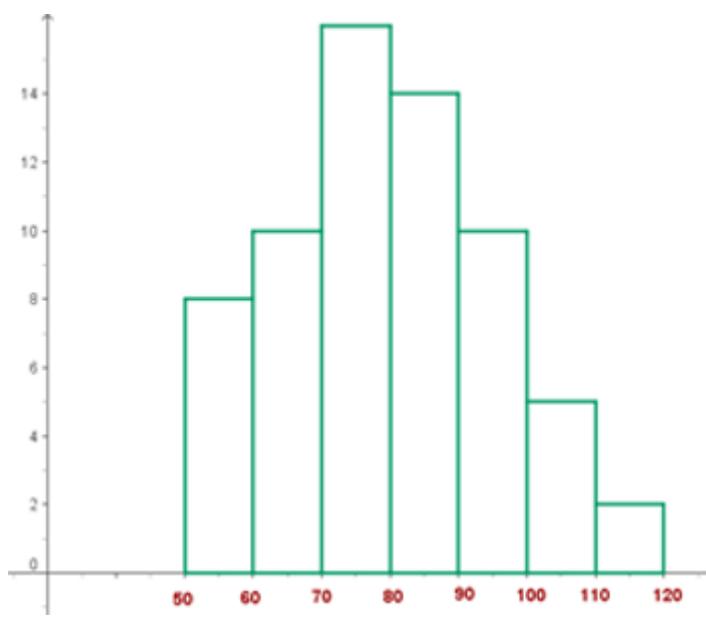
La **superficie** de cada **barra** es **proporcional** a la **frecuencia** de los **valores** representados.

Ejemplo

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

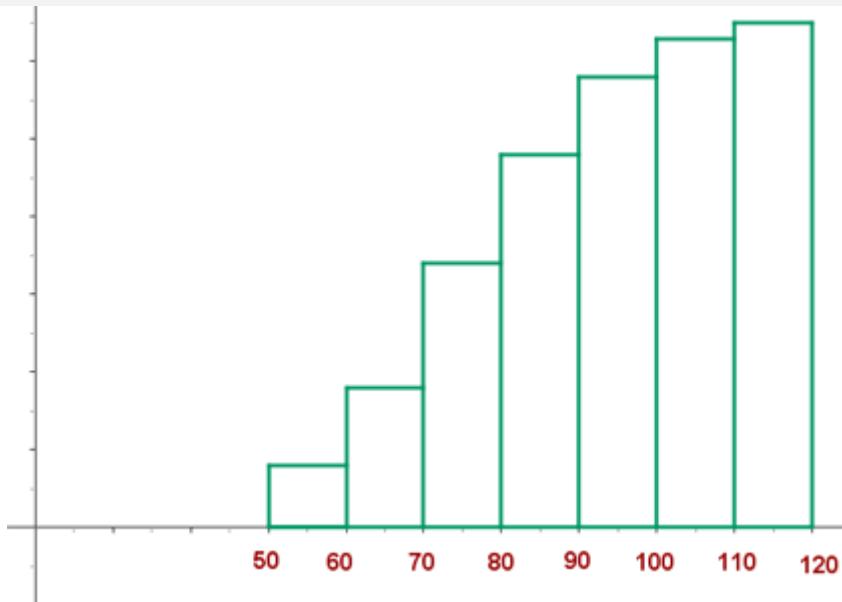
	c _i	f _i	F _i

[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)	110	5	63
[110, 120)	115	2	65
		65	



Histograma de frecuencias acumuladas

Si se representan las **frecuencias acumuladas** de una **tabla de datos agrupados** se obtiene el **histograma de frecuencias acumuladas**.



Histogramas con intervalos de amplitud diferente

Para **costruir** un **histogramas** con **intervalo de amplitud diferente** tenemos que **calcular** las **alturas** de los **rectángulos** del **histograma**.

$$h_i = \frac{f_i}{a_i}$$

h_i es la altura del intervalo

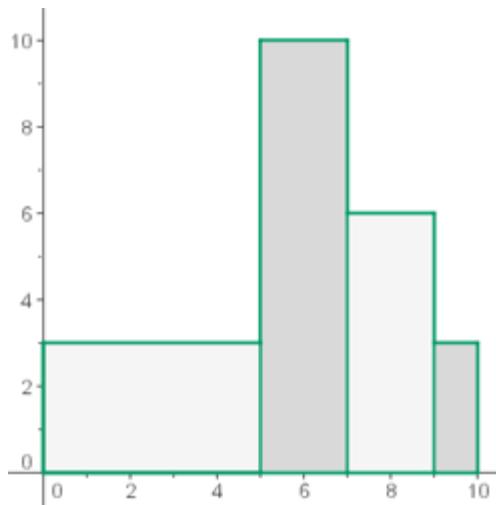
f_i es la frecuencia del intervalo

a_i es la amplitud del intervalo

Ejemplo

En la siguiente tabla se muestra las calificaciones (suspenso, aprobado, notable y sobresaliente) obtenidas por un grupo de 50 alumnos.

	f_i	h_i
[0, 5)	15	3
[5, 7)	20	10
[7, 9)	12	6
[9, 10)	3	3
	50	



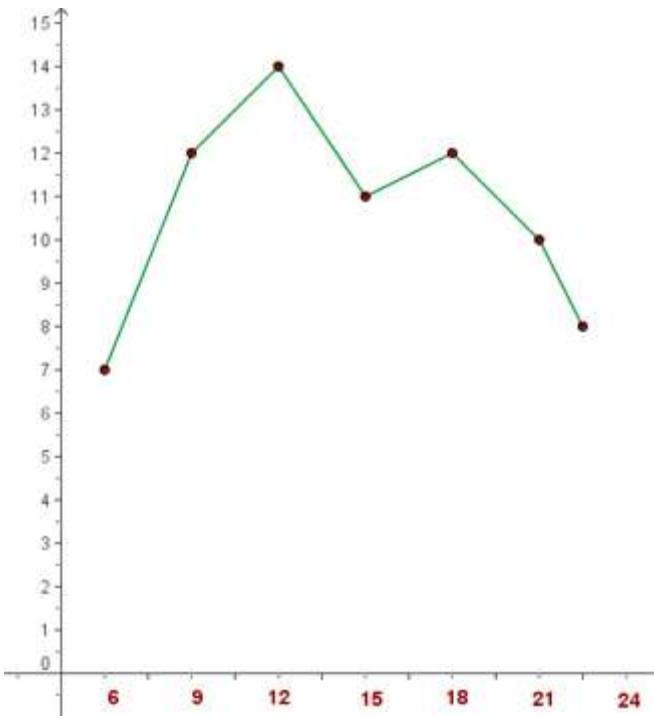
Un **polígono de frecuencias** se forma uniendo los **extremos** de las **barras** de un [diagrama de barras](#) mediante **segmentos**.

También se puede realizar trazando los **puntos** que representan las **frecuencias** y uniéndolos mediante **segmentos**.

Ejemplo

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

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



Polígonos de frecuencia para datos agrupados

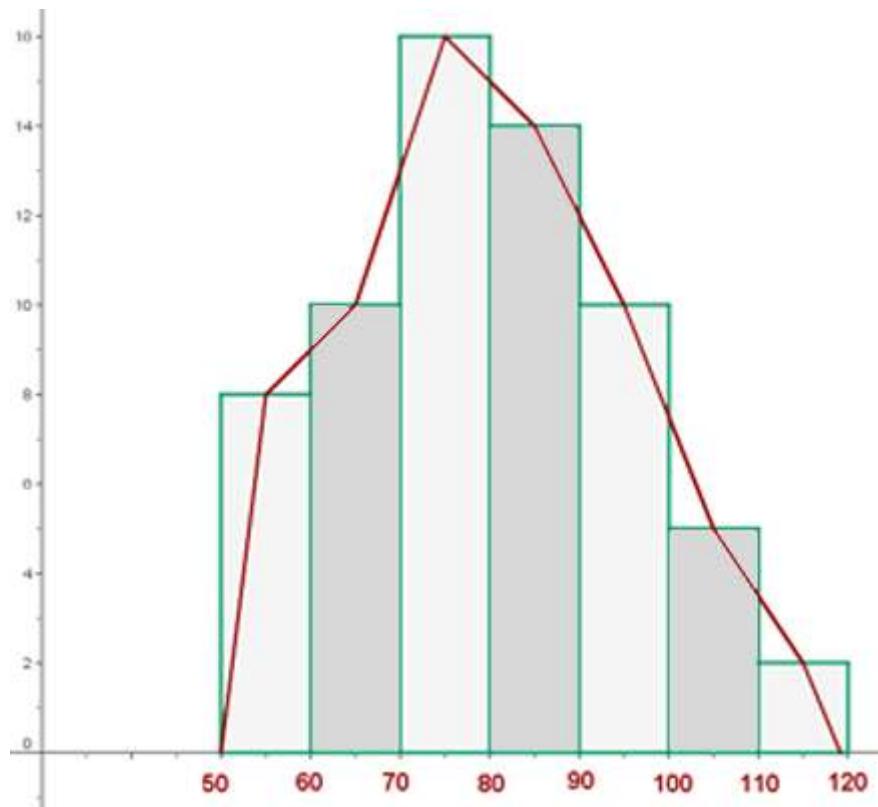
Para construir el **polígono de frecuencia** se toma la **marca de clase** que coincide con el **punto medio** de cada **rectángulo** de un [histograma](#).

Ejemplo

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

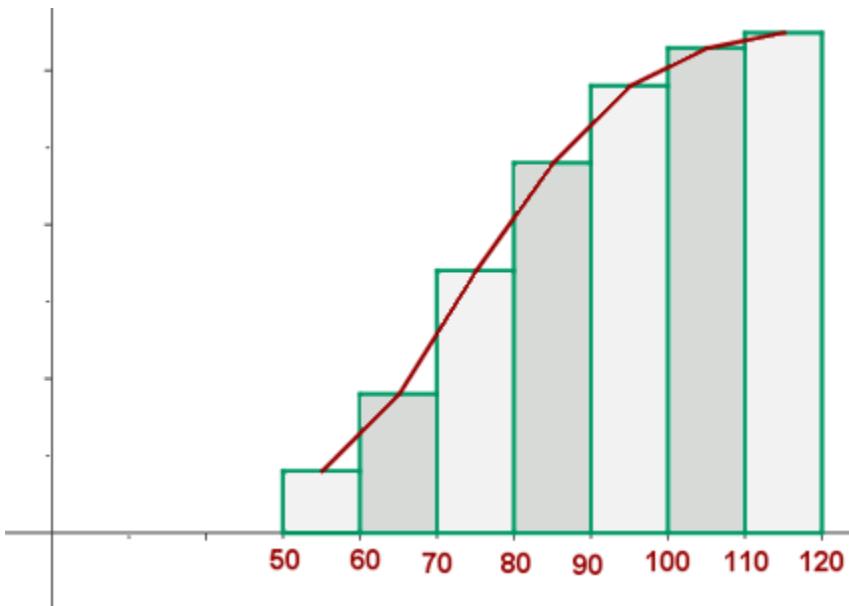
	c_i	f_i	F_i
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34

[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	110	5	63
[110, 120)	115	2	65
		65	



Polígono de frecuencias acumuladas

Si se representan las **frecuencias acumuladas** de una **tabla de datos agrupados** se obtiene el **histograma de frecuencias acumuladas** o su correspondiente **polígono**.



Moda

La **moda** es el **valor** que tiene **mayor frecuencia absoluta**.

Se representa por **M_o** .

Se puede hallar la **moda** para **variables cualitativas y cuantitativas**.

Hallar la moda de la distribución:

2, 3, 3, 4, 4, 4, 5, 5 **$M_o = 4$**

Si en un grupo hay **dos o varias puntuaciones** con la **misma frecuencia** y esa frecuencia es la máxima, la **distribución** es **bimodal o multimodal**, es decir, tiene **varias modas**.

1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9 **$M_o = 1, 5, 9$**

Cuando todas las **puntuaciones** de un grupo tienen la **misma frecuencia**, **no** hay **moda**.

2, 2, 3, 3, 6, 6, 9, 9

Si **dos puntuaciones adyacentes** tienen la **frecuencia máxima**, la **moda** es el **promedio** de las dos puntuaciones adyacentes.

0, 1, 3, 3, 5, 5, 7, 8 **Mo = 4**

Cálculo de la moda para datos agrupados

1º Todos los intervalos tienen la misma amplitud.

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

L_{i-1} es el límite inferior de la clase modal.

f_i es la frecuencia absoluta de la clase modal.

f_{i-1} es la frecuencia absoluta inmediatamente inferior a la en clase modal.

f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal.

a_i es la amplitud de la clase.

También se utiliza otra **fórmula** de la **moda** que da un **valor aproximado** de ésta:

$$Mo = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot a_i$$

Ejemplo

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

	f_i
[60, 63)	5
[63, 66)	18
[66, 69)	42
[69, 72)	27
[72, 75)	8
	100

$$Mo = 66 + \frac{(42 - 18)}{(42 - 18) + (42 - 27)} \cdot 3 = 67.846$$

$$Mo = 66 + \frac{27}{18 + 27} \cdot 3 = 67.8$$

2º Los intervalos tienen amplitudes distintas.

En primer lugar tenemos que hallar las alturas.

$$h_i = \frac{f_i}{a_i}$$

La clase modal es la que tiene mayor altura.

$$Mo = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

La **fórmula** de la **moda aproximada** cuando existen distintas amplitudes es:

$$Mo = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

$$Mo = 5 + \frac{10 - 3}{(10 - 3) + (10 - 6)} \cdot 2 = 6.27$$

$$Mo = 5 + \frac{6}{3 + 6} \cdot 2 = 6.33$$

Mediana

Es el **valor** que ocupa el **lugar central** de todos los **datos** cuando éstos están **ordenados de menor a mayor**.

La **mediana** se representa por **Me**.

La **mediana** se puede **hallar** sólo para **variables cuantitativas**.

Cálculo de la mediana

1 Ordenamos los **datos** de **menor a mayor**.

2 Si la serie tiene un **número impar** de **medidas** la **mediana** es la **puntuación central** de la misma.

2, 3, 4, 4, 5, 5, 5, 6, 6 **Me = 5**

3 Si la serie tiene un **número par** de puntuaciones la **mediana** es la **media** entre las dos **puntuaciones centrales**.

7, 8, 9, 10, 11, 12 **Me = 9.5**

Cálculo de la mediana para datos agrupados

La **mediana** se encuentra en el **intervalo** donde la **frecuencia acumulada** llega hasta la **mitad de la suma de las frecuencias absolutas**.

$$\frac{N}{2}$$

Es decir tenemos que buscar el intervalo en el que se encuentre $\frac{N}{2}$.

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

L_{i-1} es el límite inferior de la clase donde se encuentra la mediana.

$$\frac{N}{2}$$

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas.

F_{i-1} es la **frecuencia acumulada** anterior a la clase mediana.

a_i es la amplitud de la clase.

La **mediana** es **independiente** de las **amplitudes** de los **intervalos**.

Ejemplo

Calcular la **mediana** de una distribución estadística que viene dada por la siguiente tabla:

	f_i	F_i
[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65

[69, 72)	27	92
[72, 75)	8	100
	100	

$$100 / 2 = 50$$

Clase modal: [66, 69)

$$Me = 66 + \frac{50 - 23}{42} \cdot 3 = 67.93$$

Media aritmética

La **media aritmética** es el **valor** obtenido al **sumar** todos los **datos** y **dividir** el resultado entre el **número** total de **datos**.

\bar{x} es el símbolo de la **media aritmética**.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{N}$$

Ejemplo

Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio.

$$\bar{x} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = 80 \text{ Kg}$$

Media aritmética para datos agrupados

Si los **datos** vienen **agrupados** en una tabla de frecuencias, la expresión de la **media** es:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{N}$$

Ejercicio de media aritmética

En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. **Calcula la puntuación media.**

	x_i	f_i	$x_i \cdot f_i$
[10, 20)	15	1	15
[20, 30)	25	8	200
[30, 40)	35	10	350
[40, 50)	45	9	405
[50, 60)	55	8	440
[60, 70)	65	4	260

[70, 80)	75	2	150
	42	1 820	

$$\bar{X} = \frac{1820}{42} = 43.33$$

Propiedades de la media aritmética

1 La **suma** de las **desviaciones** de todas las puntuaciones de una distribución respecto a la **media** de la misma igual a **cero**.

$$\sum(X_i - \bar{X}) = 0$$

Las suma de las desviaciones de los números 8, 3, 5, 12, 10 de su media aritmética 7.6 es igual a 0:

$$8 - 7.6 + 3 - 7.6 + 5 - 7.6 + 12 - 7.6 + 10 - 7.6 =$$

$$= 0.4 - 4.6 - 2.6 + 4.4 + 2.4 = 0$$

2 La **media aritmética** de los **cuadrados** de las **desviaciones** de los valores de la variable con respecto a un **número** cualquiera se hace **mínima** cuando dicho **número** coincide con la **media aritmética**.

$$\sum(X_i - \bar{X})^2 \text{ Mínimo}$$

3 Si a todos los valores de la variable se les **suma** un mismo **número**, la **media aritmética** queda **aumentada** en dicho **número**.

4 Si todos los valores de la variable se **multiplican** por un mismo **número** la **media aritmética** queda **multiplicada** por dicho **número**.

Observaciones sobre la media aritmética

1 La **media** se puede **hallar** sólo para **variables cuantitativas**.

2 La **media** es **independiente** de las **amplitudes** de los **intervalos**.

3 La **media** es muy sensible a las **puntuaciones extremas**. Si tenemos una distribución con los siguientes pesos:

65 kg, 69kg , 65 kg, 72 kg, 66 kg, 75 kg, 70 kg, 110 kg.

La **media** es igual a 74 kg, que es una **medida de centralización** poco representativa de la distribución.

4 La **media** no se puede calcular si hay un intervalo con una **amplitud indeterminada**.

Cuartiles

Los **cuartiles** son los **tres valores** de la variable que **dividen** a un **conjunto** de **datos ordenados** en **cuatro partes iguales**.

Q₁, Q₂ y Q₃ determinan los valores correspondientes al **25%, al 50% y al 75%** de los **datos**.

Q₂ coincide con la **mediana**.

Cálculo de los cuartiles

1 Ordenamos los **datos** de **menor a mayor**.

2 Buscamos el lugar que ocupa cada **cuartil** mediante la expresión $\frac{k \cdot N}{4}$, $k = 1, 2, 3$.

Número impar de datos

2, 5, 3, 6, 7, 4, 9

2, 3, 4, 5, 6, 7, 9

↓ ↓ ↓

Q_1 Q_2 Q_3

Número par de datos

2, 5, 3, 4, 6, 7, 1, 9

1, 2, 3, 4, 5, 6, 7, 9

2.5 4.5 6.5

↓ ↓ ↓

Q_1 Q_2 Q_3

Cálculo de los cuartiles para datos agrupados

$$\frac{k \cdot N}{4}, k = 1, 2, 3$$

En primer lugar buscamos la **clase** donde se encuentra $\frac{k \cdot N}{4}$, en la **tabla de las frecuencias acumuladas**.

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \quad k = 1, 2, 3$$

Ejercicio de cuartiles

Calcular los cuartiles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Cálculo del primer cuartil

$$\frac{65 \cdot 1}{4} = 16.25$$

$$Q_1 = 60 + \frac{16.25 - 8}{10} \cdot 10 = 68.25$$

Cálculo del segundo cuartil

$$\frac{65 \cdot 2}{4} = 32.5$$

$$Q_2 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.0625$$

Cálculo del tercer cuartil

$$\frac{65 \cdot 3}{4} = 48.75$$

$$Q_3 = 90 + \frac{48.75 - 48}{10} \cdot 10 = 90.75$$

Deciles

Los **deciles** son los **nueve valores** que **dividen** la serie de **datos** en **diez partes iguales**.

Los **deciles** dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D_5 coincide con la **mediana**.

Cálculo de los deciles

$$\frac{k \cdot N}{10}, \quad k = 1, 2, \dots, 9$$

En primer lugar buscamos la clase donde se encuentra $\frac{k \cdot N}{10}$, en la tabla de las frecuencias acumuladas.

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i \quad k = 1, 2, \dots, 9$$

Ejercicio de deciles

Calcular los deciles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Cálculo del primer decil

$$\frac{65 \cdot 1}{10} = 6.5$$

$$D_1 = 50 + \frac{6.5 - 0}{8} \cdot 10 = 58.12$$

Cálculo del segundo decil

$$\frac{65 \cdot 2}{10} = 13$$

$$D_2 = 60 + \frac{13 - 8}{10} \cdot 10 = 65$$

Cálculo del tercer decil

$$\frac{65 \cdot 3}{10} = 19.5$$

$$D_3 = 70 + \frac{19.5 - 18}{16} \cdot 10 = 70.94$$

Cálculo del cuarto decil

$$\frac{65 \cdot 4}{10} = 26$$

$$D_4 = 70 + \frac{26 - 18}{16} \cdot 10 = 75$$

Cálculo del quinto decil

$$\frac{65 \cdot 5}{10} = 32.5$$

$$D_5 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.06$$

Cálculo del sexto decil

$$\frac{65 \cdot 6}{10} = 39$$

$$D_6 = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

Cálculo del séptimo decil

$$\frac{65 \cdot 7}{10} = 45.5$$

$$D_7 = 80 + \frac{45.5 - 34}{14} \cdot 10 = 88.21$$

Cálculo del octavo decil

$$\frac{65 \cdot 8}{10} = 52$$

$$D_8 = 90 + \frac{52 - 48}{10} \cdot 10 = 94$$

Cálculo del noveno decil

$$\frac{65 \cdot 9}{10} = 58.5$$

$$D_9 = 100 + \frac{58.5 - 58}{5} \cdot 10 = 101$$

Percentiles

Los **percentiles** son los **99 valores** que **dividen** la serie de **datos** en **100 partes iguales**.

Los **percentiles** dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P₅₀ coincide con la **mediana**.

Cálculo de los percentiles

$$\frac{k \cdot N}{100}, \quad k = 1, 2, \dots, 99$$

En primer lugar buscamos la clase donde se encuentra $\frac{k \cdot N}{100}$, en la tabla de las frecuencias acumuladas.

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \quad k = 1, 2, \dots, 99$$

Ejercicio de percentiles

Calcular el percentil 35 y 60 de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
	65	

Percentil 35

$$\frac{65 \cdot 35}{100} = 22.75$$

$$P_{35} = 70 + \frac{22.75 - 18}{16} \cdot 10 = 72.97$$

Percentil 60

$$\frac{65 \cdot 60}{100} = 39$$

$$P_{60} = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

Las **medidas de dispersión** nos informan sobre cuánto se alejan del centro los valores de la distribución.

Las **medidas de dispersión** son:

Rango o recorrido

El **rango** es la **diferencia** entre el **mayor** y el **menor** de los **datos** de una distribución estadística.

Desviación media

La **desviación respecto a la media** es la **diferencia** entre cada **valor** de la variable estadística y la **media aritmética**.

$$D_i = x - \bar{x}$$

La **desviación media** es la **media aritmética** de los **valores absolutos de las desviaciones respecto a la media**.

La **desviación media** se representa por $D_{\bar{x}}$

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{N}$$

$$D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N}$$

Ejemplo

Calcular la **desviación media** de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$D_{\bar{x}} = \frac{|9-9|+|3-9|+|8-9|+|8-9|+|9-9|+|8-9|+|9-9|+|18-9|}{8} = 2.25$$

Desviación media para datos agrupados

Si los datos vienen agrupados en una **tabla de frecuencias**, la expresión de la **desviación media** es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + \dots + |x_n - \bar{x}|f_n}{N}$$

$$D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}|f_i}{N}$$

Ejemplo

Calcular la **desviación media** de la distribución:

	x_i	f_i	$x_i \cdot f_i$	$ x - \bar{x} $	$ x - \bar{x} \cdot f_i$
[10, 15)	12.5	3	37.5	9.286	27.858
[15, 20)	17.5	5	87.5	4.286	21.43
[20, 25)	22.5	7	157.5	0.714	4.998
[25, 30)	27.5	4	110	5.714	22.856
[30, 35)	32.5	2	65	10.174	21.428
		21	457.5		98.57

$$\bar{x} = \frac{457.5}{21} = 21.786$$

$$D_{\bar{x}} = \frac{98.57}{21} = 4.69$$

Varianza

La **varianza** es la **media aritmética del cuadrado de las desviaciones respecto a la media** de una distribución estadística.

La varianza se representa por σ^2 .

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$$

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \dots + (x_n - \bar{x})^2 f_n}{N} \quad \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{N}$$

Para simplificar el **cálculo de la varianza** vamos a utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{N} - \bar{x}^2 \quad \sigma^2 = \sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2$$

Varianza para datos agrupados

$$\sigma^2 = \frac{x_1^2 f_1 + x_2^2 f_2 + \dots + x_n^2 f_n}{N} - \bar{x}^2 \quad \sigma^2 = \sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2$$

Ejercicios de varianza

Calcular la varianza de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8} = 15$$

Calcular la varianza de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225

[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Propiedades de la varianza

- 1 La **varianza** será siempre un **valor positivo o cero**, en el caso de que las puntuaciones sean iguales.
- 2 Si a todos los **valores** de la variable se les **suma** un **número** la **varianza no varía**.
- 3 Si todos los **valores** de la variable se **multiplican** por un **número** la **varianza queda multiplicada** por el **cuadrado** de dicho **número**.

4 Si tenemos varias distribuciones con la misma **media** y conocemos sus respectivas **varianzas** se puede calcular la **varianza total**.

Si todas las muestras tienen el mismo tamaño:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}$$

Si las muestras tienen distinto tamaño:

$$\sigma^2 = \frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \dots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \dots + k_n}$$

Observaciones sobre la varianza

1 La **varianza**, al igual que la media, es un índice muy sensible a las puntuaciones extremas.

2 En los casos que no se pueda hallar la media tampoco será posible hallar la **varianza**.

3 La **varianza** no viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

Desviación típica

La **desviación típica** es la **raíz cuadrada de la varianza**.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

La **desviación típica** se representa por σ .

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N}} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \dots + (x_n - \bar{x})^2 f_n}{N}} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos a utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{N} - \bar{x}^2} \quad \sigma = \sqrt{\sum_{i=1}^n \frac{x_i^2}{N} - \bar{x}^2}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \dots + x_n^2 f_n}{N} - \bar{x}^2} \quad \sigma = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

Ejercicios de desviación típica

Calcular la **desviación típica** de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \sqrt{\frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8}} = 3.87$$

Calcular la **desviación típica** de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$

[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30, 40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60, 70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
			42	1 820
				88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42} - 43.33^2} = 14.797$$

Propiedades de la desviación típica

- 1 La **desviación típica** será siempre un **valor positivo o cero**, en el caso de que las puntuaciones sean iguales.
- 2 Si a todos los **valores** de la variable se les **suma** un **número** la **desviación típica no varía**.

3 Si todos los **valores** de la variable se **multiplican** por un **número** la **desviación típica** queda **multiplicada** por dicho **número**.

4 Si tenemos varias distribuciones con la misma **media** y conocemos sus respectivas **desviaciones típicas** se puede calcular la **desviación típica total**.

Si todas las muestras tienen el mismo tamaño:

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}}$$

Si las muestras tienen distinto tamaño:

$$\sigma = \sqrt{\frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \dots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \dots + k_n}}$$

Observaciones sobre la desviación típica

1 La **desviación típica**, al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.

2 En los casos que no se pueda hallar la media tampoco será posible hallar la **desviación típica**.

3 Cuanta más pequeña sea la **desviación típica** mayor será la **concentración de datos** alrededor de la **media**.

Coeficiente de variación

El **coeficiente de variación** es la relación entre la **desviación típica** de una muestra y su **media**.

$$C.V = \frac{\sigma}{\bar{X}}$$

El **coeficiente de variación** se suele expresar en **porcentajes**:

$$C.V = \frac{\sigma}{\bar{X}} \cdot 100$$

El **coeficiente de variación** permite comparar las **dispersiones** de dos distribuciones distintas, siempre que sus **medias** sean **positivas**.

Se calcula para cada una de las distribuciones y los valores que se obtienen se comparan entre sí.

La **mayor dispersión** corresponderá al valor del **coeficiente de variación mayor**.

Ejercicio

Una distribución tiene $x = 140$ y $\sigma = 28.28$ y otra $x = 150$ y $\sigma = 25$. ¿Cuál de las dos presenta mayor dispersión?

$$C.V_1 = \frac{28.28}{140} \cdot 100 = 20.2\%$$

$$C.V_2 = \frac{25}{150} \cdot 100 = 16\%$$

La primera distribución presenta mayor dispersión.

Puntuaciones típicas

Puntuaciones diferenciales

Las **puntuaciones diferenciales** resultan de **restarles** a las **puntuaciones directas** la **media aritmética**.

$$x_i = X_i - \bar{X}$$

Puntuaciones típicas

Las **puntuaciones típicas** son el resultado de **dividir** las **puntuaciones diferenciales** entre la **desviación típica**. Este proceso se llama **tipificación**.

Las **puntuaciones típicas** se representan por **z**.

$$z = \frac{X_i - \bar{X}}{\sigma}$$

Observaciones sobre puntuaciones típicas

La **media aritmética** de las **puntuaciones típicas** es **0**.

La **desviación típica** de las **puntuaciones típicas** es **1**.

Las **puntuaciones típicas** son **adimensionales**, es decir, son independientes de las unidades utilizadas.

Las **puntuaciones típicas** se utilizan para **comparar** las **puntuaciones** obtenidas en distintas distribuciones.

Ejemplo

En una clase hay 15 alumnos y 20 alumnas. El peso medio de los alumnos es 58.2 kg y el de las alumnas y 54.4 kg. Las desviaciones típicas de los dos grupos son, respectivamente, 3.1 kg y 5.1 kg. El peso de José es de 70 kg y el de Ana es 65 kg. ¿Cuál de ellos puede, dentro del grupo de su sexo, considerarse más grueso?

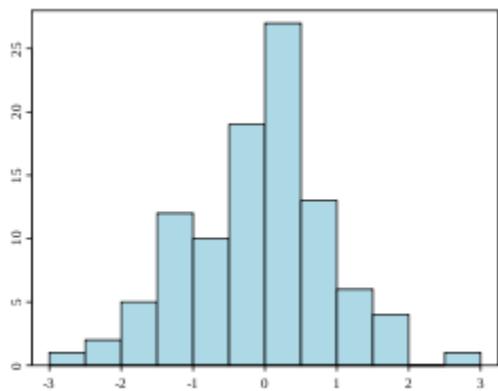
$$z_1 = \frac{70 - 58.2}{3.1} = 3.81$$

$$z_2 = \frac{65 - 54.4}{5.1} = 2.47$$

Histograma

De Wikipedia, la enciclopedia libre

Saltar a: [navegación](#), [búsqueda](#)



Histograma.

En [estadística](#), un **histograma** es una representación [gráfica](#) de una [variable](#) en forma de barras, donde la superficie de cada barra es proporcional a la [frecuencia](#) de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

En términos matemáticos, puede ser definida como una función inyectiva (o mapeo) que acumula (cuenta) las observaciones que pertenecen a cada subintervalo de una partición. El *histograma*, como es tradicionalmente entendido, no es más que la representación gráfica de dicha función.

Se utiliza cuando se estudia una variable continua, como franjas de edades o altura de la muestra, y, por comodidad, sus valores se agrupan en clases, es decir, valores continuos. En los casos en los que los datos son cualitativos (no-numéricos), como sexto grado de acuerdo o nivel de estudios, es preferible un [diagrama de sectores](#).

Los histogramas son más frecuentes en [ciencias sociales](#), humanas y económicas que en ciencias naturales y exactas. Y permite la comparación de los resultados de un proceso.

Tipos de histograma

- Diagramas de barras simples

Representa la frecuencia simple (absoluta o relativa) mediante la altura de la barra la cual es proporcional a la frecuencia simple de la categoría que representa.

- Diagramas de barras compuesta

Se usa para representar la información de una [tabla de doble entrada](#) o sea a partir de dos variables, las cuales se representan así; la altura de la barra representa la frecuencia simple de las modalidades o categorías de la variable y esta altura es proporcional a la frecuencia simple de cada modalidad.

- Diagramas de barras agrupadas

Se usa para representar la información de una tabla de doble entrada o sea a partir de dos variables, el cual es representado mediante un conjunto de barras como se clasifican respecto a las diferentes modalidades.

- Polígono de frecuencias

Es un gráfico de líneas que de las frecuencias absolutas de los valores de una distribución en el cual la altura del punto asociado a un valor de las variables es proporcional a la frecuencia de dicho valor.

- Ojiva porcentual

Es un gráfico acumulativo, el cual es muy útil cuando se quiere representar el rango porcentual de cada valor en una distribución de frecuencias.

En los gráficos las barras se encuentran juntas y en la tabla los números poseen en el primer miembro un corchete y en el segundo un parentesis, por ejemplo: [10-20)

[editar] Construcción de un histograma

- **Paso 1**

Determinar el rango de los datos. Rango es igual al dato mayor menos el dato menor.

- **Paso 2**

Obtener los números de clases, existen varios criterios para determinar el número de clases (o barras) -por ejemplo la [regla de Sturgess](#)- . Sin embargo ninguno de ellos es exacto. Algunos autores recomiendan de cinco a quince clases, dependiendo de cómo estén los datos y cuántos sean. Un criterio usado frecuentemente es que el número de clases debe ser aproximadamente a la raíz cuadrada del número de datos. Por ejemplo, la raíz cuadrada de 30 (número de artículos) es mayor que cinco, por lo que se seleccionan seis clases.

- **Paso 3**

Establecer la longitud de clase: es igual al rango dividido por el número de clases.

- **Paso 4**

Construir los intervalos de clases: Los intervalos resultan de dividir el rango de los datos en relación al resultado del PASO 2 en intervalos iguales.

- **Paso 5**

Graficar el histograma: En caso de que las clases sean todas de la misma amplitud, se hace un gráfico de barras, las bases de las barras son los intervalos de clases y altura son la frecuencia de las clases. Si se unen los puntos medios de la base superior de los rectángulos se obtiene el polígono de frecuencias.