

# EJERCICIOS RESUELTOS

## 1

La tabla siguiente nos muestra el resultado de una encuesta entre los alumnos de primer curso, analizando el número de suspensos en la primera evaluación :

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

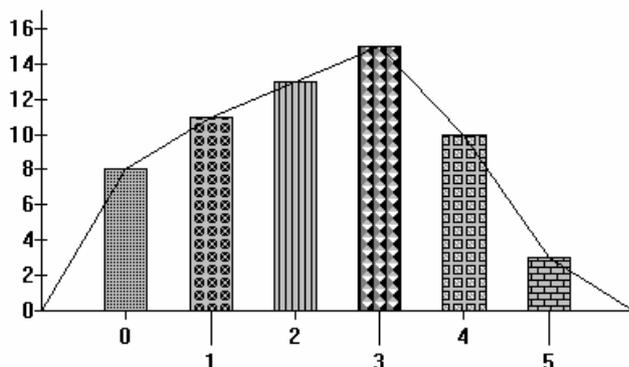
Realicemos un estudio estadístico completo.

Se trata de una variable **cuantitativa discreta**. Esto condicionará algunos procesos del cálculo estadístico.

### RECuento Y TABLA DE FRECUENCIAS

x	recuento	n	r	p	N	R	P
0	////	8	0'1333	13'33	8	0'1333	13'33
1	//// //	11	0'1833	18'33	19	0'3167	31'67
2	//// //	13	0'2167	21'67	32	0'5333	53'33
3	//// //	15	0'2500	25'00	47	0'7833	78'33
4	////	10	0'1667	16'67	57	0'9500	95'00
5	///	3	0'0500	5'00	60	1'0000	100'00
Totales :		<b>N = 60</b>	1'0000	100'00			

### GRÁFICOS ESTADÍSTICOS APROPIADOS PARA ESTE TIPO DE VARIABLE



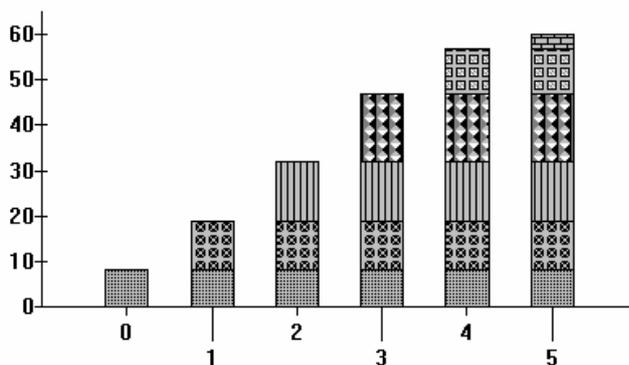
#### DIAGRAMA DE BARRAS :

Sobre el valor de cada variable dibujamos una barra con altura igual a la frecuencia que deseamos representar (en este caso las absolutas n ).

#### POLÍGONO DE FRECUENCIAS :

Obtenidos enlazando los extremos superiores de las barras.

NOTA :Siendo la variable discreta, no tiene sentido dibujar el polígono de frecuencias.



#### DIAGRAMAS ACUMULADOS :

Construidos como los anteriores, son los representativos de las distintas frecuencias acumuladas.

El ejemplo representa las frecuencias absolutas acumuladas (N).

El polígono de frecuencias se construiría enlazando los extremos superiores de las barras.

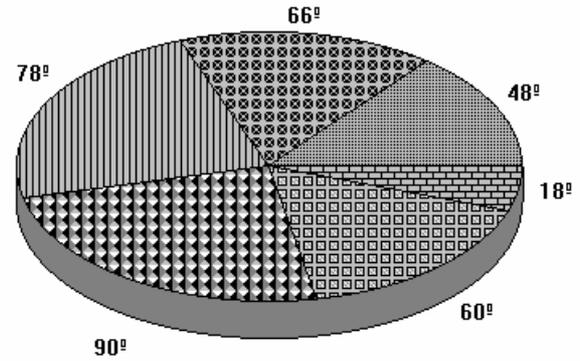
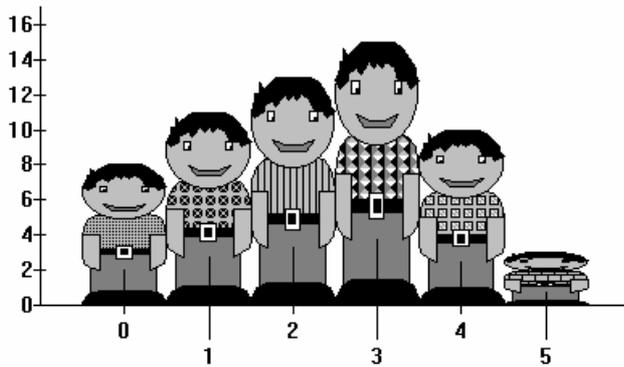
#### PICTOGRAMAS:

Con el mismo principio seguido para la construcción de los diagramas de barras, sustituimos dichas barras por dibujos alusivos a la variable estadística estudiada.

#### DIAGRAMAS DE SECTORES :

Resultan de la división de un círculo en sectores cuya amplitud es proporcional a la frecuencia.

La amplitud de cada sector será :  $\alpha = \frac{n}{N} \cdot 360^\circ = r \cdot 360^\circ$



### MEDIA, VARIANZA Y DESVIACIÓN TÍPICA

x	n	n.x	n.x <sup>2</sup>
0	8	0	0
1	11	11	11
2	13	26	52
3	15	45	135
4	10	40	160
5	3	15	75
<b>N = 60</b>	<b>137</b>	<b>433</b>	

Este tipo de tabla facilita los cálculos.

$$\text{Media} = 137 / 60 = 2,283$$

$$\text{Varianza} = (433 / 60) - \text{media al cuadrado} = 2'005$$

$$\text{Desviación típica} = \text{raíz cuadrada de la varianza} = 1'416$$

$$\bar{x} = \frac{\sum n_i \cdot x_i}{N} = \frac{137}{60} = 2'283 \quad s_x^2 = \frac{\sum n_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{433}{60} - 2'283^2 = 2'005 \quad s_x = \sqrt{s_x^2} = \sqrt{2'005} = 1'416$$

**MODA** = Valor de mayor frecuencia = **3**

### PERCENTILES

Para la determinación de medidas de posición (percentiles), podemos seguir dos procedimientos de cálculo :

1º) Basado en las frecuencias absolutas acumuladas N :

Determinamos el lugar que ocupa :  $L = k \cdot N / 100$

El percentil será el valor cuya frecuencia N primero iguale o supere al lugar L.

2º) Basado en porcentajes acumulados P :

El percentil será el valor cuyo porcentaje P primero iguale o supere al orden k del percentil.

Apliquemos el primer procedimiento para calcular la **mediana** y el **9º decil** :

La mediana (percentil 50) ocupará el lugar :  $L = 50 \cdot 60 / 100 = 30$

El 9º decil (percentil 90) ocupará el lugar :  $L = 90 \cdot 60 / 100 = 54$

x	n
0	8
1	11
2	13
3	15
4	10
5	3
<b>N = 60</b>	

N
8
19
32
47
57
60

← Mediana = 2

← 9º decil = 4

Aplicando el segundo procedimiento descrito, determinemos los **cuartiles 1º y 3º**, así como la **amplitud semi-intercuartílica** :

x	n	r	p	P
0	8	0'1333	13'33	13'33
1	11	0'1833	18'33	31'67
2	13	0'2167	21'67	53'33
3	15	0'2500	25'00	78'33
4	10	0'1667	16'67	95'00
5	3	0'0500	5'00	100'00
<b>N = 60</b>	<b>1'0000</b>	<b>100'00</b>		

← Cuartil 1º (percentil 25) = 1

← Cuartil 3º (percentil 75) = 3

$$\text{Amplitud semi-intercuartílica} = \frac{Q_3 - Q_1}{2} = \frac{3 - 1}{2} = 1$$

## 2

Trabajamos ahora con las edades de 50 jóvenes de nuestro barrio :

1	11	20	15	10	4	12	20	5	23	9	12	13	14	15
24	15	7	8	12	9	9	5	2	20	13	15	7	11	22
20	6	12	4	7	1	18	20	11	10	14	20	11	13	15
21	25	20	22	10										

Como en el ejemplo anterior, realicemos un estudio estadístico completo.

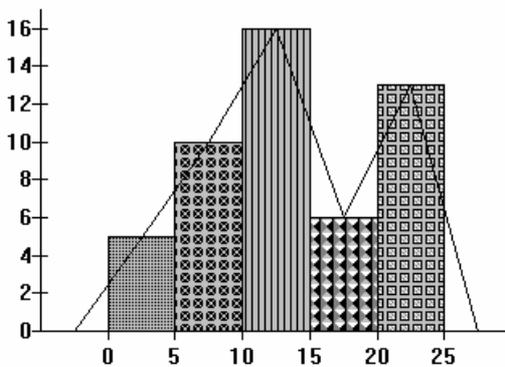
Nos encontramos ante una variable estadística cuantitativa continua. Agruparemos o no las observaciones en intervalos en función de los diferentes valores observados.

### TABLA DE FRECUENCIAS

Observado el valor mínimo (1) y máximo (24), decidimos agrupar los datos en intervalos de 5 años de amplitud, empezando por 0.

Intervalos	recuento	n	r	p	N	R	P
[ 0 , 5 )	////	5	0'10	10	5	0'10	10
[ 5 , 10 )	//// //	10	0'20	20	15	0'30	30
[ 10 , 15 )	//// // // // /	16	0'32	32	31	0'62	62
[ 15 , 20 )	//// /	6	0'12	12	37	0'74	74
[ 20 , 25 ]	//// // // //	13	0'26	26	50	1'00	100
Totales :		<b>N = 50</b>	1'00	100			

### GRÁFICOS ESTADÍSTICOS

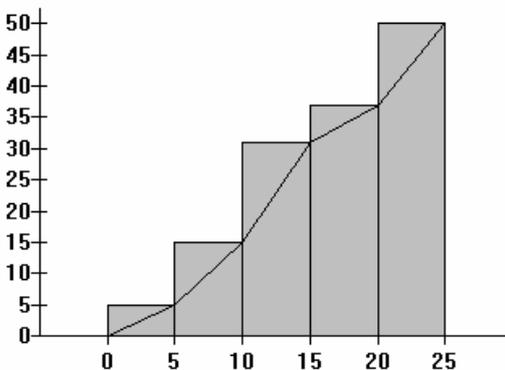


HISTOGRAMA :

Sobre el valor de cada variable dibujamos una franja con altura igual a la frecuencia que deseamos representar (en este caso las absolutas n ).

POLÍGONO DE FRECUENCIAS :

Obtenido enlazando los puntos medios de los extremos superiores de las franjas.



HISTOGRAMAS ACUMULADOS :

Construidos como los anteriores, son los representativos de las distintas frecuencias acumuladas.

El ejemplo representa las frecuencias absolutas acumuladas ( N ).

En este caso, el polígono de frecuencias **NO** se construiría enlazando los puntos medios de los extremos superiores de las franjas, sino como se indica en la figura.

### Cálculo de Moda, Media, Varianza y Desviación típica :

Para el cálculo de la media y la varianza utilizamos la tabla auxiliar siguiente. En ella se incorpora la columna x , que contiene la marca de clase (valor central) de cada intervalo.

La **MODA** (valor de mayor frecuencia) se encuentra en el intervalo [ 10 , 15 ) . Determinemos su valor concreto :

$$Mo = e_i + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} \cdot a_i = 10 + \frac{6}{6 + 10} \cdot 5 = 11'875$$

Intervalos	n	x	n.x	n.x <sup>2</sup>
[ 0 , 5 )	5	2'5	12'5	31'25
[ 5 , 10 )	10	7'5	75'0	562'50
[ 10 , 15 )	16	12'5	200'0	2500'00
[ 15 , 20 )	6	17'5	105'0	1837'50
[ 20 , 25 ]	13	22'5	292'5	6581'25
<b>N = 50</b>			685'0	11512'50

$$\bar{x} = \frac{\sum n_i \cdot x_i}{N} = \frac{685}{50} = 13'7 \quad s_x^2 = \frac{\sum n_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{11512'5}{50} - 13'7^2 = 42 \quad s_x = \sqrt{s_x^2} = \sqrt{42'56} = 6'524$$

Utilizando las frecuencias absolutas acumuladas, calculemos el **decil 2º** y el **percentil 62** :

Lugar que ocupa el decil 2º (percentil 20) =  $20 \cdot 50 / 100 = 10$

Lugar que ocupa el percentil 62 =  $62 \cdot 50 / 100 = 31$

Intervalos	n	N
[ 0 , 5 )	5	5
[ 5 , 10 )	10	15
[ 10 , 15 )	16	31
[ 15 , 20 )	6	37
[ 20 , 25 ]	13	50
<b>N = 50</b>		

⇐ Decil 2º (percentil 20) en [5,10)

Lugar = 10

⇐ Percentil 62 en [10,15)

Lugar = 31

Determinemos sus valores concretos :

$$P_{20} = e_i + \frac{\frac{20 \cdot N}{100} - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{\frac{20 \cdot 50}{100} - 5}{10} \cdot 5 = 7'5$$

$$P_{62} = e_i + \frac{\frac{62 \cdot N}{100} - N_{i-1}}{n_i} \cdot a_i = 10 + \frac{\frac{62 \cdot 50}{100} - 15}{16} \cdot 5 = 15$$

Utilizando los porcentajes acumulados, calculemos el **cuartil 1º** y la **mediana** :

Intervalos	n	r	p	P
[ 0 , 5 )	5	0'10	10	10
[ 5 , 10 )	10	0'20	20	30
[ 10 , 15 )	16	0'32	32	62
[ 15 , 20 )	6	0'12	12	74
[ 20 , 25 ]	13	0'26	26	100
<b>N = 50</b>		1'00	100	

⇐ Cuartil 1º (percentil 25) en [5,10)

⇐ Mediana (percentil 50) en [10,15)

Determinemos sus valores concretos :

$$P_{25} = e_i + \frac{\frac{25 \cdot N}{100} - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{\frac{25 \cdot 50}{100} - 5}{10} \cdot 5 = 8'75$$

$$P_{50} = e_i + \frac{\frac{50 \cdot N}{100} - N_{i-1}}{n_i} \cdot a_i = 10 + \frac{\frac{50 \cdot 50}{100} - 15}{16} \cdot 5 = 13'125$$

### 3

x	n
2	6
3	15
4	10
5	9

De la presente distribución, calculemos :

**Media, varianza y desviación típica.**

**Moda.**

**Mediana, Percentil 82, Cuartiles y amplitud semi-intercuartílica.**

La variable establecida puede ser discreta o continua sin agrupar en intervalos. Realicemos los cálculos en ambos supuestos.

x	n
2	6
3	15
4	10
5	9
	<b>40</b>

N	P
6	15
21	52'5
31	77'5
40	100

n.x	n.x <sup>2</sup>
12	24
45	135
40	160
45	225
<b>142</b>	<b>544</b>

<b>Media</b> $\bar{x} = \frac{\sum n_i \cdot x_i}{N} = \frac{142}{40} = 3'55$	<b>Varianza</b> $\sigma^2 = \frac{\sum n_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{544}{40} - 3'55^2 = 0'9$	<b>Desviación típica</b> $\sigma = \sqrt{0'9975} = 0'99875$
<b>Moda</b> <b>3</b>	<b>Mediana</b> (percentil 50) <b>3</b>	<b>Percentil 82</b> <b>5</b>
<b>Cuartil 1º</b> (percentil 25) <b>3</b>	<b>Cuartil 3º</b> (percentil 75) <b>4</b>	<b>Rango semi-intercuartílico</b> $\frac{Q_3 - Q_1}{2} = \frac{4 - 3}{2} = 0'5$

Los valores anteriores, relativos a percentiles, son válidos si la variable es DISCRETA. En el supuesto de tratarse de una variable CONTINUA (con datos no agrupados), deberíamos entender que el valor identifica el intervalo situado a la izquierda en la siguiente tabla :

Intervalo	x	n	N	P
[1'5,2'5)...	2	6	6	15
[2'5,3'5)...	3	15	21	52'5
[3'5,4'5)...	4	10	31	77'5
[4'5,5'5)...	5	9	40	100
		<b>40</b>		

Los percentiles pedidos se obtendrían del modo siguiente :

<b>Mediana</b>	<b>en [2'5,3'5)</b>	$Me = P_{50} = 2'5 + \frac{\frac{50 \cdot 40}{100} - 6}{15} \cdot 1 = 3'433$
<b>Percentil 82</b>	<b>en [4'5,5'5)</b>	$P_{82} = 4'5 + \frac{\frac{82 \cdot 40}{100} - 31}{9} \cdot 1 = 4'700$
<b>Cuartil 1º</b>	<b>en [2'5,3'5)</b>	$Q_1 = P_{25} = 2'5 + \frac{\frac{25 \cdot 40}{100} - 6}{15} \cdot 1 = 2'767$
<b>Cuartil 3º</b>	<b>en [3'5,4'5)</b>	$Q_3 = P_{75} = 3'5 + \frac{\frac{75 \cdot 40}{100} - 21}{10} \cdot 1 = 3'400$

Interv.	n
[10,12)	5
[12,14)	11
[14,16)	19
[16,18)	21
[18,20]	4

De la distribución de la izquierda, calcular :  
 Media, varianza y desviación típica.  
 Moda  
 Mediana, Percentil 59 y Decil 3º.  
 Desviación media.  
 Coeficientes de asimetría y curtosis.

Interv.	n	a
[10,12)	5	11
[12,14)	11	13
[14,16)	19	15
[16,18)	21	17
[18,20]	4	19
	<b>60</b>	

N	P
5	8'333
16	26'667
35	58'333
56	93'333
60	100'000

n.a	n.a <sup>2</sup>
55	605
143	1859
285	4275
357	6069
76	1444
<b>916</b>	<b>14252</b>

Media $\bar{x} = \frac{\sum n_i \cdot a_i}{N} = \frac{916}{60} = 15'2667$	Varianza $\sigma^2 = \frac{\sum n_i \cdot a_i^2}{N} - \bar{x}^2 = \frac{14252}{60} - 15'2667^2 = 4'$	Desviación típica $\sigma = \sqrt{4'4622} = 2'1124$
--	---	--

Moda	en [16,18)	$Mo = 16 + \frac{4}{4+19} \cdot 2 = 16'3478$
Mediana (percentil 50)	en [14,16)	$Me = P_{50} = 14 + \frac{\frac{50 \cdot 60}{100} - 16}{19} \cdot 2 = 15'4737$
Percentil 59	en [16,18)	$P_{59} = 16 + \frac{\frac{59 \cdot 60}{100} - 35}{21} \cdot 2 = 16'0381$
Decil 3º (percentil 30)	en [14,16)	$D_3 = P_{30} = 14 + \frac{\frac{30 \cdot 60}{100} - 16}{19} \cdot 2 = 14'2105$

Desviación media

$ x - \bar{x} $	$n \cdot  x - \bar{x} $
4'2667	21'3333
2'2667	24'9333
0'2668	5'0668
1'7333	36'4000
3'7333	14'9333
	<b>102'6667</b>

Asimetría y Curtosis

$x - \bar{x}$	$n \cdot (x - \bar{x})^3$	$n \cdot (x - \bar{x})^4$
-4'2667	-388'3615	1657'0090
-2'2667	-128'1019	290'3644
-0'2668	-0'3603	0'0961
1'7333	109'3618	189'5604
3'7333	208'1375	777'0466
	<b>-199'3244</b>	<b>2914'0765</b>

Desviación media	$D = \frac{\sum n_i \cdot  x_i - \bar{x} }{N} = \frac{102'6667}{60} = 1'7111$
Asimetría (-0'3524 < 0) Algo asimétrica hacia la izquierda	$As_1 = \frac{\sum n_i \cdot (x_i - \bar{x})^3}{\sigma^3} = \frac{-199'3244}{2'1124^3} = -0'3524$
Curtosis (-0'5608 < 0) Ligeramente aplanada (Platicúrtica)	$K = \frac{\sum n_i \cdot (x_i - \bar{x})^4}{\sigma^4} - 3 = \frac{2914'0765}{2'1124^4} - 3 = -0'5608$

# 5

La distribución de las estaturas en centímetros de los alumnos de un centro, expresados en porcentajes, es la siguiente:

Estaturas	Porcentajes
Menos de 150	0'3
De 150 a 154	1'6
De 155 a 159	9'4
De 160 a 164	20'5
De 165 a 169	31'5
De 170 a 174	22'5
De 175 a 179	10'7
De 180 y más	3'5

- Siendo abiertos los intervalos primero y el último, ¿ qué valores sería razonable considerar para los límites extremos de esos intervalos ?
- Si suponemos que en el Centro hay 1200 alumnos, ¿ cuáles serían las frecuencias absolutas?
- Calcular la estatura media y la desviación típica.
- ¿ Entre qué estaturas se encuentra la quinta parte de las estaturas centrales ?.

a) Al referirse a intervalos de 5 cm. de amplitud en los restantes casos, debemos considerar que el primer intervalo es de 145 a menos de 150 y, el último, de 180 a 185.

b)

Estaturas	p	n = p . 1200 / 100	n	P	N
[145,150)	0'3	3'6	4	0'3	4
[150,155)	1'6	19'2	19	1'9	23
[155,160)	9'4	112'8	113	11'3	136
[160,165)	20'5	246	246	31'8	382
[165,170)	31'5	378	378	63'3	760
[170,175)	22'5	270	270	85'8	1030
[175,180)	10'7	128'4	128	96'5	1158
[180,185)	3'5	42	42	100'0	1200
			<b>N=1200</b>		

c)

Estaturas	n	x	n.x	n.x <sup>2</sup>
[145,150)	4	147'5	590'0	87025'00
[150,155)	19	152'5	2897'5	441868'75
[155,160)	113	157'5	17797'5	2803106'25
[160,165)	246	162'5	39975'0	6495937'50
[165,170)	378	167'5	63315'0	10605262'50
[170,175)	270	172'5	46575'0	8034187'50
[175,180)	128	177'5	22720'0	4032800'00
[180,185)	42	182'5	7665'0	1398862'50
	<b>1200</b>		201535'0	33899050'00

De aquí resulta :  $\bar{x} = \frac{201535}{1200} = 167'95$

$$s_x^2 = \frac{33899050}{1200} - 167'95^2 = 42'006 \quad s_x = \sqrt{42'006} = 6'481$$

d)

La quinta parte representa el 20%. Con relación al centro (50%), cubrirán desde el 40% al 60%. Se nos pide que calculemos los percentiles 40 y 60 de la distribución de estaturas.

La tabla de porcentajes acumulados del apartado **b)** nos permite deducir que :

Los percentiles 40 y 60 se encuentran en el intervalo [165,170) .

Sus valores concretos son :

$$P_{40} = e_i + \frac{40.N}{100} - N_{i-1} \cdot a_i = 165 + \frac{40.1200}{100} - 382 \cdot .5 = 166'963$$

$$P_{60} = e_i + \frac{60.N}{100} - N_{i-1} \cdot a_i = 165 + \frac{60.1200}{100} - 382 \cdot .5 = 169'471$$

## 6

Partiendo de la siguiente distribución de frecuencias acumuladas, determinar la media, mediana y moda de la siguiente distribución de edades. Analice la relación entre ellas.

Edad	N
[10,12)	4
[12,14)	11
[14,16)	24
[16,18)	34
[18,20]	40

Calculemos los parámetros pedidos, con el fin de observar en qué medida se verifica la relación  $\bar{x} - Mo = 3.(\bar{x} - Me)$

Para obtener las frecuencias absolutas, a partir de las acumuladas, aplicamos el concepto que define a estas últimas. En la práctica, las frecuencias absolutas se obtienen restando la correspondiente acumulada de la anterior.

Edad	N	n	x	n.x	n.x <sup>2</sup>
[10,12)	4	4	11	44	484
[12,14)	11	7	13	91	1183
[14,16)	24	13	15	195	2925
[16,18)	34	10	17	170	2890
[18,20]	40	6	19	114	2166
		40		614	9648

$\bar{x} = \frac{614}{40} = 15'35$	Lugar que ocupa la mediana : $L = 50 \cdot 40 / 100 = 20$ La mediana está en [14,16) : $Me = 14 + \frac{20 - 11}{13} \cdot 2 = 15'3846$	La moda se encuentra en [14 , 16). Su valor concreto es : $Mo = 14 + \frac{10}{10 + 7} \cdot 2 = 15'1765$
------------------------------------	--	--

Comprobemos la relación existente entre ellas :

$$\bar{x} - Mo = 15'35 - 15'1765 = 0'1735$$

$$3.(\bar{x} - Me) = 3.(15'35 - 15'3845) = -0'1035$$

No se verifica la relación esperada, si bien la diferencia no es muy grande.

Esta relación teórica sólo se verifica en situaciones ideales y excepcionales (por ejemplo en distribuciones simétricas, donde  $\bar{x} = Mo = Me$  ).

# 7

Completar la tabla de frecuencias siguiente :

Nº de suspensos	n	N
0	3	
1		10
2	12	
3		30
4		
<b>N =</b>	<b>50</b>	

Nº de suspensos	n	N
0	3	<b>3</b>
1	<b>7</b>	10
2	12	<b>22</b>
3	<b>8</b>	30
4	<b>20</b>	<b>50</b>

coincide con el valor de n

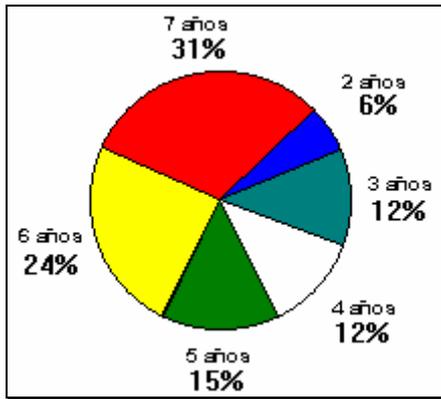
para que al acumular resulte N=10

acumulando 12

para que al acumular resulte N=30

Última acumulada =N=50 y n=20 por diferencia con la anterior

8



Calcular la amplitud semi-intercuartílica de la distribución de las edades de 400 niños, representada a la izquierda.

Conocidos los porcentajes y el total de observaciones (N=400), podemos construir la distribución de frecuencias absolutas :

$$n = p \cdot N / 100$$

x	p	n	P
2	6	24	6
3	12	48	18
4	12	48	30
5	15	60	45
6	24	96	69
7	31	124	100
		400	

← Primer cuartil (percentil 25)  
 ← Tercer cuartil (percentil 75)

La amplitud o rango semi-intercuartílico será pues :  $\frac{Q_3 - Q_1}{2} = \frac{7 - 4}{2} = 1'5$

## 9

Una variable  $X$  tiene por media 12 y desviación típica 3. Si elevamos todos los valores al cuadrado construimos la nueva variable  $Y = X^2$ . ¿Cuál es el valor de su media aritmética ?.

Observemos la expresión de la varianza :  $s_x^2 = \frac{\sum_{i=1}^n n_i \cdot x_i^2}{N} - \bar{x}^2$

La primera parte de la expresión contiene los cuadrados de los valores de la variable  $X$ ; es decir, los valores definidos como la nueva variable  $Y$ .

Con esto :  $s_x^2 = \frac{\sum_{i=1}^n n_i \cdot y_i}{N} - \bar{x}^2 \Rightarrow s_x^2 = \bar{y} - \bar{x}^2 \Rightarrow \bar{y} = s_x^2 + \bar{x}^2 = 3^2 + 12^2 = 153$

## 10

Una variable  $X$  tiene como media 8 y varianza 4. ¿ Qué transformación lineal hemos de realizar con ella, para obtener una nueva variable  $Y$  que tenga por media 42 y desviación típica 10 ?.

Se entiende por transformación lineal a una relación del tipo :  $Y = a + b.X$   
Hemos de calcular los parámetros  $a$  y  $b$  desconocidos.

Haciendo uso de las propiedades de la media y la desviación típica, resulta :

Sobre la media  $\bar{Y} = a + b.\bar{X} \Rightarrow 42 = a + b.8$

En relación con la desviación típica  $s_Y = b.s_X \Rightarrow 10 = b.2 \Rightarrow b = 5 \Rightarrow a = 42 - 5.8 = 2$

La transformación realizada fue :  $Y = 2 + 5.X$

# 11

Las calificaciones de un alumno en dos test de conocimientos fueron 5'4 y 41. El primer test dio como media 5 con varianza 2 y, el segundo, media 38 con varianza 12.

¿ En qué test obtuvo mejor calificación con relación al grupo total de alumnos ?.

Nos encontramos con dos distribuciones de calificaciones medidas en distintas escalas. Para poder comparar tendremos que referir ambas series de valores a otras equivalentes entre sí (igual media y desviación típica).

El proceso de tipificación nos proporciona lo que deseamos (siempre obtendremos una distribución con media 0 y desviación típica 1).

Tipificando ambas calificaciones se obtiene :

$$\text{Nota del test 1}^\circ : 5'4 \rightarrow z_1 = \frac{5'4 - 5}{\sqrt{2}} = 0'283$$

$$\text{Nota del test 2}^\circ : 41 \rightarrow z_1 = \frac{41 - 38}{\sqrt{12}} = 0'866$$

La nota obtenida en el segundo test es superior a la del primero en términos comparativos.

Estatura en cm.	Alumnos
[140,145)	12
[145,150)	35
[150,155)	51
[155,160)	?
[160,165)	7

- a) Determinar la frecuencia desconocida, sabiendo que la estatura media es de 151'5 cm.
- b) Calcule la amplitud semi-intercuartílica.
- c) Moda de la distribución y coeficiente de asimetría que la utiliza.
- d) Percentil correspondiente a una estatura de 153 cm.. Explique su significado.

- e) ¿ Entre qué estaturas se encuentran las 25 centrales ?.
- f) Porcentaje de alumnos que miden más de 157 cm.

a)

	x	n	n.x
[140,145)	142'5	12	1710
[145,150)	147'5	35	5162'5
[150,155)	152'5	51	7777'5
[155,160)	157'5	f	157'5.f
[160,165)	162'5	7	1137'5
		105+f	15787'5+157'5.f

La tabla de cálculos de la media conduce a :

$$151'5 = \frac{15787'5 + 157'5.f}{105 + f}$$

Resolviendo deducimos que : f = 20

b)

	n	N
[140,145)	12	12
[145,150)	35	47
[150,155)	51	98
[155,160)	20	118
[160,165)	7	125
	N=125	

Lugar  $Q_1 = 125 \cdot 25 / 100 = 31'25$   
 $Q_1$  se encuentra en [145,150)

$$Q_1 = 145 + \frac{31'25 - 12}{35} \cdot 5 = 147'75$$

Lugar  $Q_3 = 125 \cdot 75 / 100 = 93'75$   
 $Q_3$  se encuentra en [150,155)

$$Q_3 = 150 + \frac{93'75 - 47}{51} \cdot 5 = 154'5833$$

$$\text{Luego : } Q = \frac{Q_3 - Q_1}{2} = \frac{154'5833 - 147'75}{2} = 3'4167$$

- c) 1º) Moda en [150,155) :  $Mo = 150 + \frac{20}{35 + 20} \cdot 5 = 151'8182$

x	n	n.x	n.x <sup>2</sup>
142'5	12	1710	243675
147'5	35	5162'5	761468'75
152'5	51	7777'5	1186068'75
157'5	20	3150	496125
162'5	7	1137'5	184843'75
	125	18937'5	2872181'25

$$s = \sqrt{\frac{2872181'25}{125} - 151'5^2}$$

$$s = 5'02$$

$$As = \frac{\bar{x} - Mo}{s} = -0'0634$$

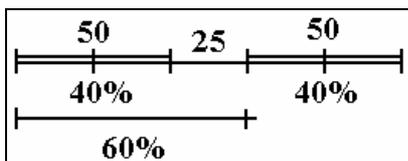
	n	N
[140,145)	12	12
[145,150)	35	47
[150,155)	51	98
[155,160)	20	118
[160,165)	7	125
	N=125	

- d) 153 se encuentra en [150,155)

$$P_k = 150 + \frac{k \cdot \frac{125}{100} - 47}{51} \cdot 5 = 153$$

Resolviendo : k = 62'08 ≈ 62

e)



Lugar =  $125 \cdot 40 / 100 = 50$  ; en [150,155) :

$$P_{40} = 150 + \frac{50 - 47}{51} \cdot 5 = 150'29$$

Lugar =  $125 \cdot 60 / 100 = 75$  ; en [150,155) :

$$P_{60} = 150 + \frac{75 - 47}{51} \cdot 5 = 152'75$$

Entre 150'29 y 152'75

f) 157 se encuentra en [155,160)

$$P_k = 155 + \frac{k \cdot \frac{125}{100} - 98}{20} \cdot 5 = 157$$

Resolviendo :  $k = 84'8\%$  (porcentaje inferiores a 157)

Luego, miden más de 157 cm. :

$$100\% - 84'8\% = 15'2\%$$

Edad	Hombres	Mujeres
22 a 25	7	3
19 a 22	9	5
16 a 19	5	6
13 a 16	11	9
10 a 13	8	2

- a) Determine el número de hombres con edades comprendidas entre los 11 y 15 años.
- b) ¿Cuál de los dos grupos de edades está más disperso ?
- c) Con relación al grupo integrado por los del mismo sexo, ¿quién resulta más joven, un hombre o una mujer de 20 años ?.

	Hombre					Mujer		
	x	n	N	n.x	n.x <sup>2</sup>	n	n.y	n.y <sup>2</sup>
[10,13)	11'5	8	8	92	1058	2	23	264'5
[13,16)	14'5	11	19	159'5	2312'75	9	130'5	1892'25
[16,19)	17'5	5	24	87'5	1531'25	6	105	1837'5
[19,22)	20'5	9	33	184'5	3782'25	5	102'5	2101'25
[22,25)	23'5	7	40	164'5	3865'75	3	70'5	1656'75
		40		688	12550	25	431'5	7752'25

a) 11 pertenece al intervalo [10,13) : 
$$P_k = 10 + \frac{k \cdot \frac{40}{100} - 0}{8} \cdot 3 = 11 \Rightarrow k = 6'67\%$$

15 pertenece al intervalo [13,16) : 
$$P_k = 13 + \frac{k \cdot \frac{40}{100} - 8}{11} \cdot 3 = 15 \Rightarrow k = 38'33\%$$
  
 Entre 11 y 15 el  $38'33 - 6'67 = 31'66\%$ . Luego hay :  $40 \cdot 31'66 / 100 = 12'664 \approx 13$  hombres

b) Calculamos las varianzas de ambos grupos :

$$\bar{x} = \frac{688}{40} = 17'2 \quad ; \quad s_x^2 = \frac{12550}{40} - 17'2^2 = 17'91 \quad ; \quad s_x = \sqrt{17'91} = 4'232$$

$$\bar{y} = \frac{431'5}{25} = 17'26 \quad ; \quad s_y^2 = \frac{7752'25}{25} - 17'26^2 = 12'1824 \quad ; \quad s_y = \sqrt{12'1824} = 3'49$$

Siendo  $17'91 > 12'1824 \Rightarrow$  Grupo hombres más disperso *de forma absoluta*

Pese a ser las medias prácticamente iguales, debemos emplear el coeficiente de variación para estudiar la *variabilidad relativa* de ambos grupos :

$$CV_x = \frac{4'232}{17'2} \cdot 100 = 24'605\% \quad ; \quad CV_y = \frac{3'49}{17'26} \cdot 100 = 20'220\% \Rightarrow \text{hombres más disperso}$$

c) Tipificamos 20 en ambos grupos :

$$Z_{\text{hombre}} = \frac{20 - 17'2}{\sqrt{17'91}} = 0'662 \quad ; \quad Z_{\text{mujer}} = \frac{20 - 17'26}{\sqrt{12'1824}} = 0'785$$

Como  $0'662 < 0'785 \Rightarrow$  Hombre más joven

# 14

La tabla siguiente nos muestra las calificaciones de 10 alumnos, en un test de cálculo matemático, al inicio del curso y al finalizar el mismo.

Alumno	1	2	3	4	5	6	7	8	9	10
Inicio	4	5	1	5	2	3	2	1	1	3
Final	6	8	5	9	3	6	7	6	4	9

- a) Determine la media, desviación típica, mediana y moda de las calificaciones al inicio y al final del curso.  
 b) Calcule la media y desviación típica del incremento o mejora de la calificación obtenida.

a)

Inicio	x	4	5	1	5	2	3	2	1	1	3	27
	x <sup>2</sup>	16	25	1	25	4	9	4	1	1	9	95

$$\bar{x} = \frac{27}{10} = 2'7 \quad ; \quad s_x = \sqrt{\frac{95}{10} - 2'7^2} = 1'487$$

Ordenando valores :

1	1	1	2	2	3	3	4	5	5
---	---	---	---	---	---	---	---	---	---

Mediana = 2'5

Moda = 1

Final	y	6	8	5	9	3	6	7	6	4	9	63
	y <sup>2</sup>	36	64	25	81	9	36	49	36	16	81	433

$$\bar{y} = \frac{63}{10} = 6'3 \quad ; \quad s_y = \sqrt{\frac{433}{10} - 6'3^2} = 1'9$$

Ordenando valores :

3	4	5	6	6	6	7	8	9	9
---	---	---	---	---	---	---	---	---	---

Mediana = 6

Moda = 6

b)

Mejora	d	2	3	4	4	1	3	5	5	3	6	36
	d <sup>2</sup>	4	9	16	16	1	9	25	25	9	36	150

$$\bar{d} = \frac{36}{10} = 3'6 \quad ; \quad s_d = \sqrt{\frac{150}{10} - 3'6^2} = 1'428$$

Media de la diferencia :

$$\bar{d} = \bar{y} - \bar{x} = 6'3 - 2'7 = 3'6 \quad (\text{No es válido para dispersiones})$$

# 15

Nº Suspensos	Alumnos
0	16
1	20
2	14
3	15
4	10
5	5

- a) Determine la media, desviación típica, coeficiente de variación, mediana y moda del número de suspensos.
- b) Coeficiente de asimetría de Fisher.
- c) Puntuación diferencial y tipificada correspondiente a 2 suspensos.

a) De la siguiente tabla de cálculos obtenemos :

$$\bar{x} = \frac{158}{80} = 1'975 \quad s = \sqrt{\frac{496}{80} - 1'975^2} = 1'5164 \quad CV = \frac{1'5164}{1'975} \cdot 100 = 76'78\%$$

Mediana :  $N/2 = 40 \Rightarrow Me = 2$                       Moda = 1

x	n	N	n.x	n.x <sup>2</sup>	$x - \bar{x}$	$n.(x - \bar{x})^3$
0	16	16	0	0	-1'975	-123'2598
1	20	36	20	20	-0'975	-18'5372
2	14	50	28	56	0'025	0'0002
3	15	65	45	135	1'025	16'1534
4	10	75	40	160	2'025	83'0377
5	5	80	25	125	3'025	138'4032
		80	158	496	95'7975	

b)

$$As = \frac{\sum n.(x - \bar{x})^3}{s^3} = \frac{95'7975}{1'5164^3} = 0'3434 \quad \text{Ligeramente asimétrica a la derecha (o positiva)}$$

c)

$$x = 2$$

$$d = x - \bar{x} = 2 - 1'975 = 0'025$$

$$z = \frac{x - \bar{x}}{s} = \frac{0'025}{1'5164} = 0'016$$

# 16

Estatura	Niños
155-159	4
150-154	13
145-149	12
140-144	A
135-139	2
130-134	1

La altura en cm. de los niños de 12 años, examinados durante la última semana en la unidad de crecimiento del centro hospitalario "Crecebien", viene representada en la tabla de la izquierda. Sabiendo que la altura media de los mismos es 147'75 cm., calcular :

- La frecuencia A del tercer intervalo.
- La simetría de la distribución a partir de la comparación de media, mediana y moda.
- El percentil correspondiente a un niño que mide 1'43 m..

x	n	n.x
132	1	132
137	2	274
142	A	142.A
147	12	1764
152	13	1976
157	4	628
<b>TOTAL</b>	<b>32+A</b>	<b>4774+142.A</b>

$$a) \quad \bar{x} = 147'75 = \frac{4774 + 142.A}{32 + A}$$

Resolviendo la ecuación anterior obtenemos el valor de A :  
 $147'75.(32+A)=4774+142.A \rightarrow$   
 $\rightarrow 4728+147'75.A=4774+142.A \rightarrow$   
 $\rightarrow 5'75.A = 46 \rightarrow A = 8$

- Calculemos la mediana y la moda de la distribución :

Moda en [149'5 , 154'5) :

$$Mo = 149'5 + \frac{4}{4 + 12} .5 = 150'75$$

Lugar que ocupa la mediana = 40/2 = 20

$$\text{Mediana en [144'5 , 149'5) : } Me = 144'5 + \frac{20 - 11}{12} .5 = 148'25$$

Intervalos	n	N
[129'5 , 134'5)	1	1
[134'5 , 139'5)	2	3
[139'5 , 144'5)	8	11
[144'5 , 149'5)	12	23
[149'5 , 154'5)	13	36
[154'5 , 159'5)	4	40

Utilizando los coeficientes de asimetría :

$$As_2 = \frac{\bar{x} - Mo}{s} \quad As_3 = \frac{3.(\bar{x} - Me)}{s}$$

y siendo siempre positiva la desviación típica ,concluiremos que la simetría resultará del análisis del signo del numerador.

$$\bar{x} - Mo = 147'75 - 150'75 = -3 < 0$$

$$3.(\bar{x} - Me) = 3.(147'75 - 148'25) = -1'5 < 0$$

Luego es asimétrica izquierda (o negativa).

- La altura 1'43 m. (= 143 cm.) se encuentra en el intervalo [139'5 , 144'5) :

$$P_k = 143 = 139'5 + \frac{\frac{k.40}{100} - 3}{8} .5 \Rightarrow 3'5 = \frac{0'4.k - 3}{8} .5 \Rightarrow \frac{3'5.8}{5} + 3 = 0'4.k \Rightarrow k = \frac{8'6}{0'4} = 21'5$$

Luego corresponde al percentil 21'5.

# 17

X	n
10-12	10
7-9	100
4-6	60
1-3	30

Dada la siguiente distribución de frecuencias., calcular :

- Media y desviación típica.
- Número de observaciones comprendidas entre las puntuaciones directas 3'5 y 9'5.
- Puntuaciones típicas de los percentiles 20 y 80.

Ordenamos los intervalos de menor a mayor, expresándolos mediante sus extremos reales.

Intervalos	n	x	n.x	n.x <sup>2</sup>
[ 0'5 , 3'5 )	30	2	60	120
[ 3'5 , 6'5 )	60	5	300	1500
[ 6'5 , 9'5 )	100	8	800	6400
[ 9'5 , 12'5 ]	10	11	110	1210
<b>Totales</b>	<b>200</b>		<b>1270</b>	<b>9230</b>

N
30
90
190
200

a)  $\bar{x} = \frac{1270}{200} = 6'35$       $s^2 = \frac{9230}{200} - 6'35^2 = 5'8275$       $s = \sqrt{5'8275} = 2'414$

b) De la observación directa de la tabla se concluye que es 160 (60+100).

c) Percentil 20 : Lugar =  $20 \times 200 / 100 = 40$  (Observando N) se encuentra en [ 3'5 , 6'5 )

$$P_{20} = 3'5 + \frac{40 - 30}{60} \cdot 3 = 4 \quad \rightarrow \quad z = \frac{4 - 6'35}{2'414} = -0'9735$$

Percentil 80 : Lugar =  $80 \times 200 / 100 = 160$  (Observando N) se encuentra en [ 6'5 , 9'5 )

$$P_{80} = 6'5 + \frac{160 - 90}{100} \cdot 3 = 8'6 \quad \rightarrow \quad z = \frac{8'6 - 6'35}{2'414} = 0'9321$$

# 18

x	n
0	6
1	12
2	21
3	11

Haciendo uso de coeficientes basados en medidas de posición, estudie la asimetría y el apuntamiento de la distribución.

Tales coeficientes son el de asimetría de Yule y el de curtosis de Kelley.

Obtengamos los percentiles que intervienen en su cálculo a través de la columna de porcentajes acumulados (P) :

x	n	r	p	P
0	6	0'12	12	12
1	12	0'24	24	36
2	21	0'42	42	78
3	11	0'22	22	100
	50			

Cuartil 1º : (25%)	1
Cuartil 3º : (75%)	2
Mediana : (50%)	2
Percentil 10 : (10%)	0
Percentil 90 : (90%)	3

Con ellos :

$$Y = \frac{Q_3 - 2 \cdot Me + Q_1}{Q_3 - Q_1} = \frac{2 - 2 \cdot 2 + 1}{2 - 1} = -1 \quad (\text{asimétrica a la izquierda o negativa})$$

$$K = \frac{Q}{P_{90} - P_{10}} - 0'263 = \frac{Q_3 - Q_1}{P_{90} - P_{10}} - 0'263 = \frac{2 - 1}{3 - 0} - 0'263 = -0'0963 \quad (\text{ligeramente platicúrtica o aplastada})$$

# 19

Determine las medias aritmética, geométrica y armónica de la variable X que toma los valores siguientes :

5, 1, 5, 4, 8.

Media aritmética : 
$$\bar{x} = \frac{\sum x_i}{N} = \frac{5+1+5+4+8}{5} = \frac{23}{5} = 4.6$$

Media geométrica : 
$$\bar{x}_G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N} = \sqrt[5]{5 \cdot 1 \cdot 5 \cdot 4 \cdot 8} = \sqrt[5]{800} = 800^{1/5} = 800^{0.2} = 3.807$$

Media armónica : 
$$\bar{x}_A = \frac{N}{\sum \left( \frac{1}{x_i} \right)} = \frac{5}{\frac{1}{5} + \frac{1}{1} + \frac{1}{5} + \frac{1}{4} + \frac{1}{8}} = \frac{5}{1.775} = 2.817$$

## 20

x	n
1	3
2	10
3	7
	20

Determine las medias aritmética, geométrica y armónica de la distribución.

Generalizamos las expresiones correspondientes al figurar frecuencias :

Media aritmética : 
$$\bar{x} = \frac{\sum n_i \cdot x_i}{N} = \frac{3 \cdot 1 + 10 \cdot 2 + 7 \cdot 3}{20} = \frac{44}{20} = 2'2$$

Media geométrica : 
$$\begin{aligned} \bar{x}_G &= \sqrt[N]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_n^{n_n}} = \sqrt[20]{1^3 \cdot 2^{10} \cdot 3^7} = \\ &= \sqrt[20]{2239488} = 2239488^{1/20} = 2239488^{0'05} = 2'077 \end{aligned}$$

Media armónica : 
$$\bar{x}_A = \frac{N}{\sum \left( \frac{n_i}{x_i} \right)} = \frac{20}{\frac{3}{1} + \frac{10}{2} + \frac{7}{3}} = \frac{20}{10'333} = 1'935$$

## 21

Con el fin de estudiar la edad media y la dispersión de edades en un centro educativo, el director solicita estos datos a los responsables de los distintos niveles, resultando :

- 200 alumnos de Primaria con media 11 años y varianza 2'5.
- 140 alumnos de Secundaria con media 14'6 años y varianza 2.
- 165 alumnos de Bachillerato con media 17'1 años y varianza 0'9.

¿Cuál es la edad media y la varianza del colectivo total de alumnos del centro ?.

$$\text{Media conjunta de los 3 grupos} \quad \bar{X} = \frac{\sum n_i \cdot \bar{x}_i}{\sum n_i} = \frac{200 \cdot 11 + 140 \cdot 14'6 + 165 \cdot 17'1}{200 + 140 + 165} = \frac{7065'5}{505} = 13'99$$

$$\begin{aligned} \text{Varianza conjunta de los 3 grupos} \quad S^2 &= \frac{\sum n_i \cdot S_i^2}{\sum n_i} + \frac{\sum n_i \cdot (\bar{x}_i - \bar{X})^2}{\sum n_i} = \\ &= \frac{200 \cdot 2'5 + 140 \cdot 2 + 165 \cdot 0'9}{505} + \frac{200 \cdot (11 - 13'99)^2 + 140 \cdot (14'6 - 13'99)^2 + 165 \cdot (17'1 - 13'99)^2}{505} = \\ &= \frac{928'5}{505} + \frac{3436'0105}{505} = 1'839 + 6'804 = 8'643 \end{aligned}$$

## 22

De las 10 observaciones de dos variables  $X$ ,  $Y$ , conocemos :

$$\Sigma X = 114 ; \Sigma X^2 = 1410 ; \Sigma Y = 34 ; \Sigma Y^2 = 154 ; \Sigma XY = 398 .$$

Determine la media y varianza de la variable  $V = X - Y$ .

Calculemos la media y varianza de  $X$ , la media y varianza de  $Y$ , así como la covarianza.

$$\bar{X} = \frac{114}{10} = 11'4$$

$$\bar{Y} = \frac{34}{10} = 3'4$$

$$S_X^2 = \frac{1410}{10} - 11'4^2 = 11'04$$

$$S_Y^2 = \frac{154}{10} - 3'4^2 = 3'84$$

$$S_{XY} = \frac{\sum X_i \cdot Y_i}{N} - \bar{X} \cdot \bar{Y} = \frac{398}{10} - 11'4 \cdot 3'4 = 1'04$$

Con ello :

$$\bar{V} = \bar{X} - \bar{Y} = 11'4 - 3'4 = 8$$

$$S_V^2 = S_X^2 + S_Y^2 - 2 \cdot S_{XY} = 11'04 + 3'84 - 2 \cdot 1'04 = 12'8$$

# 23

El estudio de las faltas de asistencia a clase de alumnos de un grupo de 3º de Secundaria produjo los resultados siguientes :

Faltas	1	2	3	4	5	6	7	8
Alumnos	4	3	3	2	3	2	1	2

Determine la mediana y estudie analíticamente y gráficamente el grado de concentración de la distribución.

Los cálculos de la mediana, índice de Gini y curva de Lorenz, se obtienen a partir de la siguiente tabla auxiliar:

$x_i$	$n_i$	$N_i = \sum n_i$	$P_i = (N_i / N) \cdot 100$	$t_i = n_i \cdot x_i$	$T_i = \sum t_i$	$Q_i = (T_i / T) \cdot 100$	$P_i - Q_i$
1	4	4	20	4	4	5'195	14'805
2	3	7	35	6	10	12'987	22'013
3	3	10	50	9	19	24'675	25'325
4	2	12	60	8	27	35'065	24'935
5	3	15	75	15	42	54'545	20'455
6	2	17	85	12	54	70'130	14'870
7	1	18	90	7	61	79'221	10'779
8	2	20	100	16	77	100	0
<b>N = 20</b>			<b>TP = 515</b>	<b>T = 77</b>			<b>TD = 133'182</b>

Uniéndolo el origen del rectángulo (0, 0) con los sucesivos puntos  $(P_i, Q_i)$  obtenemos la **curva de Lorenz** de la derecha.

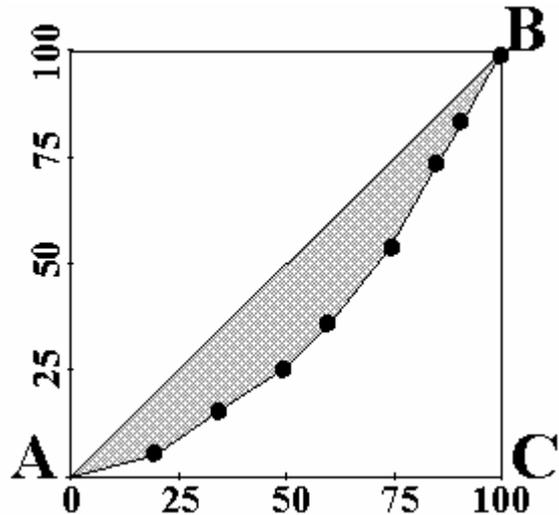
Las sumas TD y TP permiten obtener el **índice de Gini** :

$$G = \frac{TD}{TP - 100} = \frac{133'182}{515 - 100} = 0'3209$$

Concluimos la presencia de una cierta concentración (lo cual también se advierte con la gráfica).

**Mediana = 5**

ya que el primer valor que iguala o supera a 50 en la columna  $Q_i$  es 54'545, el cual corresponde a  $x = 5$ .



## 24

Un análisis del pago de impuesto en el sector de hostelería ofreció los resultados siguientes (importes mensuales por 10.000 pesetas) :

Importe	[0,2)	[2,4)	[4,6)	[6,8)	[8,10)	[10,12]
Empresas	2	6	26	40	21	5

Determine la mediana y estudie analítica y gráficamente el grado de concentración de la distribución.

Los cálculos de la mediana, índice de Gini y curva de Lorenz, se obtienen a partir de la siguiente tabla auxiliar:

	$x_i$	$n_i$	$N_i = \sum n_i$	$P_i = (N_i / N) \cdot 100$	$t_i = n_i \cdot x_i$	$T_i = \sum t_i$	$Q_i = (T_i / T) \cdot 100$	$P_i - Q_i$
[0,2)	1	2	2	2	2	2	0'297	1'703
[2,4)	3	6	8	8	18	20	2'967	5'033
[4,6)	5	26	34	34	130	150	22'255	11'745
[6,8)	7	40	74	74	280	430	63'798	10'202
[8,10)	9	21	95	95	189	619	91'840	3'160
[10,12]	11	5	100	100	55	674	100	0
	<b>N = 100</b>			<b>TP = 313</b>	<b>T = 674</b>			<b>TD = 31'843</b>

Con TD y TP obtenemos el índice de Gini :

$$G = \frac{TD}{TP - 100} = \frac{31'843}{313 - 100} = 0'1495$$

Concluimos que existe una concentración muy baja (lo cuál manifestará también la gráfica de Lorenz).

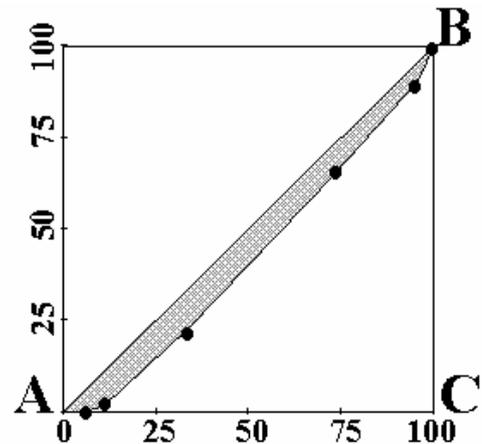
Uniendo el origen del rectángulo (0 , 0) con los sucesivos puntos ( $P_i$  ,  $Q_i$ ) obtenemos la **curva de Lorenz** de la derecha.

### Mediana en el intervalo [6 , 8)

ya que el primer valor que iguala o supera a 50 en la columna  $Q_i$  es 63'798, el cuál corresponde al intervalo indicado.

De aquí :

$$Ml = e_i + \frac{50 - Q_{i-1}}{Q_i - Q_{i-1}} \cdot a_i = 6 + \frac{50 - 22'255}{63'798 - 22'255} \cdot 2 = 7'3357$$



x	f
0	2
1	8
2	10
3	3
4	1

Haciendo uso del cálculo de momentos ordinarios de órdenes 1º al 4º, determine el valor de la media, varianza, asimetría y curtosis de la distribución de la izquierda.

Tabla de cálculo de momentos ordinarios :

		a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>
x	n	n.x	n.x <sup>2</sup>	n.x <sup>3</sup>	n.x <sup>4</sup>
0	2	0	0	0	0
1	8	8	8	8	8
2	10	20	40	80	160
3	3	9	27	81	243
4	1	4	16	64	256
<b>Totales :</b>	<b>24</b>	<b>41</b>	<b>91</b>	<b>233</b>	<b>667</b>

Orden	$a_k = \sum \frac{n}{N} .x^k = \frac{\sum n.x^k}{N}$	m <sub>k</sub>
1	$a_1 = \frac{41}{24} = 1'7083$	$m_1 = 0$
2	$a_2 = \frac{91}{24} = 3'7917$	$m_2 = a_2 - a_1^2 = 3'7917 - 1'7083^2 = 0'8734$
3	$a_3 = \frac{233}{24} = 9'7083$	$m_3 = a_3 - 3.a_2.a_1 + 2.a_1^3 = \dots = 0'2468$
4	$a_4 = \frac{667}{24} = 27'7917$	$m_4 = a_4 - 4.a_3.a_1 + 6.a_2.a_1^2 - 3.a_1^4 = \dots = 2'2954$

Con los momentos calculados :

**Media**

$$\mu = \bar{x} = a_1 = 1'7083$$

**Varianza**

$$\sigma^2 = s_x^2 = m_2 = 0'8734$$

**Coefficiente de asimetría**

$$As = \frac{m_3}{(\sqrt{m_2})^3} = \frac{0'2468}{(\sqrt{0'8734})^3} = 0'3024$$

**Coefficiente de curtosis**

$$K = \frac{m_4}{m_2^2} - 3 = \frac{2'2954}{0'8734^2} - 3 = 0'0091$$

## 26

Haciendo uso del coeficiente de variación, compare la dispersión o variabilidad relativa de las dos variables descritas en cada uno de los apartados siguientes :

- a) El peso medio de los toros de una ganadería es de 410 kg. con desviación típica de 1 kg. y, el peso medio de los perros de una granja es de 8 kg. con igual desviación típica.  
b) Dos fábricas producen tornillos con igual longitud media (50 mm.), siendo la desviación típica de la primera de 2 mm. y de 12 mm. la de la segunda.

a)  $CV_T = \frac{1}{410} \cdot 100 = 0'2439\%$      $CV_P = \frac{1}{8} \cdot 100 = 12'5\%$      $\Rightarrow$  El peso de los perros tiene mayor variabilidad

b)  $CV_A = \frac{2}{50} \cdot 100 = 4\%$      $CV_B = \frac{12}{50} \cdot 100 = 24\%$      $\Rightarrow$  Los de la 2ª tienen mayor variabilidad

X	n <sub>A</sub>	n <sub>B</sub>
0-6	4	4
7-13	6	7
14-20	9	9
21-27	12	8
28-34	9	2

La tabla muestra la comprensión lectora (X) de dos grupos de sujetos educados en niveles socioculturales altos (A) y bajos (B). Si a partir de la puntuación X=19 se considera una comprensión lectora buena, calcular :

- a) El porcentaje de personas en cada grupo con una buena comprensión lectora.  
 b) ¿Cuál de los dos grupos presenta mayor variabilidad ? (Razone adecuadamente su respuesta).

Expresamos los intervalos con extremos reales, obteniendo la tabla de cálculos de percentiles, media y varianza de ambos grupos.

	X	n <sub>A</sub>	N <sub>A</sub>	n <sub>A</sub> .X	n <sub>A</sub> .X <sup>2</sup>	n <sub>B</sub>	N <sub>B</sub>	n <sub>B</sub> .X	n <sub>B</sub> .X <sup>2</sup>
	[-0'5,6'5)	4	4	12	36	4	4	12	36
	[6'5,13'5)	6	10	60	600	7	11	70	700
	[13'5,20'5)	9	19	153	2601	9	20	153	2601
	[20'5,27'5)	12	31	288	6912	8	28	192	4608
	[27'5,34'5)	9	40	279	8649	2	30	62	1922
		40		792	18798	30		489	9867

a)

Calculamos el orden **k** del percentil que es igual a 19. Este nos da el porcentaje de los que tienen menos de 19 puntos, luego, como deseamos saber el porcentaje de los superiores a 19, la respuesta será su diferencia hasta 100.

El valor 19 se encuentra en el intervalo [13'5,20'5) :

En el grupo A :

$$P_k = 19 = 13'5 + \frac{k \cdot 40}{9} - 10 \rightarrow k = 42'68$$

Luego el 57'32% (100 - 42'68) tienen buena comprensión lectora en el grupo A.

En el grupo B :

$$P_k = 19 = 13'5 + \frac{k \cdot 30}{9} - 11 \rightarrow k = 60'24$$

Luego el 39'76% (100 - 60'24) tienen buena comprensión lectora en el grupo B.

b)

Mayor variabilidad la presentará aquel grupo que posea mayor dispersión entre sus valores. Con mayor rigor, si la media es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría), es el coeficiente de variación el más adecuado para medir la variabilidad relativa entre dos series estadísticas (mayor coeficiente indica menor homogeneidad; un menor valor indicará menor dispersión o variabilidad).

Si comparamos mediante las varianzas :

$$\bar{X}_A = \frac{792}{40} = 19'8 \quad ; \quad S_A^2 = \frac{18798}{40} - 19'8^2 = 77'91 \quad ; \quad \bar{X}_B = \frac{489}{30} = 16'3 \quad ; \quad S_B^2 = \frac{9867}{30} - 16'3^2 = 63'21$$

el grupo A presenta una mayor variabilidad.

Si comparamos mediante los coeficientes de variación :

$$CV_A = \frac{S_A}{\bar{X}_A} \cdot 100 = \frac{\sqrt{77'91}}{19'8} \cdot 100 = 44'58\% \quad ; \quad CV_B = \frac{S_B}{\bar{X}_B} \cdot 100 = \frac{\sqrt{63'21}}{16'3} \cdot 100 = 48'78\%$$

luego, concluimos que el grupo B presenta una mayor variabilidad relativa (44'58 < 48'78), en contra de lo obtenido comparando varianzas.

X	n <sub>A</sub>	n <sub>B</sub>
0-6	4	4
7-13	6	7
14-20	9	9
21-27	12	8
28-34	9	2

La tabla muestra la comprensión lectora (X) de dos grupos de sujetos educados en niveles socioculturales altos (A) y bajos (B). Si a partir de la puntuación X=19 se considera una comprensión lectora buena, calcular :

- a) El porcentaje de personas en cada grupo con una buena comprensión lectora.  
 b) ¿Cuál de los dos grupos presenta mayor variabilidad ? (Razone adecuadamente su respuesta).

Expresamos los intervalos con extremos reales, obteniendo la tabla de cálculos de percentiles, media y varianza de ambos grupos.

	X	n <sub>A</sub>	N <sub>A</sub>	n <sub>A</sub> .X	n <sub>A</sub> .X <sup>2</sup>	n <sub>B</sub>	N <sub>B</sub>	n <sub>B</sub> .X	n <sub>B</sub> .X <sup>2</sup>
	[-0'5,6'5)	4	4	12	36	4	4	12	36
	[6'5,13'5)	6	10	60	600	7	11	70	700
	[13'5,20'5)	9	19	153	2601	9	20	153	2601
	[20'5,27'5)	12	31	288	6912	8	28	192	4608
	[27'5,34'5]	9	40	279	8649	2	30	62	1922
		40		792	18798	30		489	9867

a)

Calculamos el orden  $k$  del percentil que es igual a 19. Este nos da el porcentaje de los que tienen menos de 19 puntos, luego, como deseamos saber el porcentaje de los superiores a 19, la respuesta será su diferencia hasta 100.

El valor 19 se encuentra en el intervalo [13'5,20'5) :

En el grupo A :

$$P_k = 19 = 13'5 + \frac{k \cdot 40}{100} - 10 \rightarrow k = 42'68$$

Luego el 57'32% (100 - 42'68) tienen buena comprensión lectora en el grupo A.

En el grupo B :

$$P_k = 19 = 13'5 + \frac{k \cdot 30}{100} - 11 \rightarrow k = 60'24$$

Luego el 39'76% (100 - 60'24) tienen buena comprensión lectora en el grupo B.

b)

Mayor variabilidad la presentará aquel grupo que posea mayor dispersión entre sus valores. Con mayor rigor, si la media es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría), es el coeficiente de variación el más adecuado para medir la variabilidad relativa entre dos series estadísticas (mayor coeficiente indica menor homogeneidad; un menor valor indicará menor dispersión o variabilidad).

Si comparamos mediante las varianzas :

$$\bar{X}_A = \frac{792}{40} = 19'8 ; S_A^2 = \frac{18798}{40} - 19'8^2 = 77'91 ; \bar{X}_B = \frac{489}{30} = 16'3 ; S_B^2 = \frac{9867}{30} - 16'3^2 = 63'21$$

el grupo A presenta una mayor variabilidad.

Si comparamos mediante los coeficientes de variación :

$$CV_A = \frac{S_A}{\bar{X}_A} \cdot 100 = \frac{\sqrt{77'91}}{19'8} \cdot 100 = 44'58\% \quad CV_B = \frac{S_B}{\bar{X}_B} \cdot 100 = \frac{\sqrt{63'21}}{16'3} \cdot 100 = 48'78\%$$

luego, concluimos que el grupo B presenta una mayor variabilidad relativa (44'58 < 48'78), en contra de lo obtenido comparando varianzas.

# EJERCICIOS PROPUESTOS

1

Las edades de los alumnos que asisten a clase de repaso en una academia son las siguientes.

14	16	16	19	17	17	15	17	17	15
19	15	15	16	17	14	15	16	17	16
16	15	16	18	14	15	14	17	13	18
16	16	15	16	17	15	17	14	16	16
18	18	16	18	17	17	17	17	15	16

- Construir la tabla completa de frecuencias.
- Calcular la moda.
- Determinar su media aritmética, varianza y desviación típica.
- Obtener el valor de la mediana, del percentil 29 y de la amplitud semi-intercuartílica.

2

La tabla siguiente contiene los pesos en kg. de los alumnos de un curso.

40	43	58	48	47	41'5	40'5	43	47	52
51'5	57	43	44	56	44	50	50'5	46	42
44	40	45	50	50'5	49'5	41	55	58	51
50	45	43'5	45'5	53	59	39	40	38	39'5

- Agrupar los valores en intervalos de 5 kg. de amplitud, comenzando por 35 kg., realizando un recuento de los mismos y confeccionando la tabla completa de frecuencias
- Calcular la moda de dicha distribución de pesos.
- Determinar su media aritmética, varianza y desviación típica.
- Obtener el valor de la mediana, y del 8º decil.

3

Sea la siguiente distribución de frecuencias:

x	n
1	10
2	15
3	12
4	8

- Calcular la media de esta distribución.
- Si se suma a los valores de  $x_i$  la cantidad A, ¿qué relación guarda la media de la nueva distribución con la de la anterior?. Generalizar este resultado y demostrar que si en una distribución de frecuencias de media m, se sustituyen los valores  $x_i$  por  $x_i + A$ , manteniendo las frecuencias, la media  $m'$  de la nueva distribución verifica:  $m' = A + m$
- Utilizando la igualdad obtenida, ¿cómo podría calcularse más fácilmente la media de la distribución siguiente?

x	n
2752	36
2754	54
2756	24
2758	18

4

Una serie familias se han clasificado por su número de hijos, resultando:

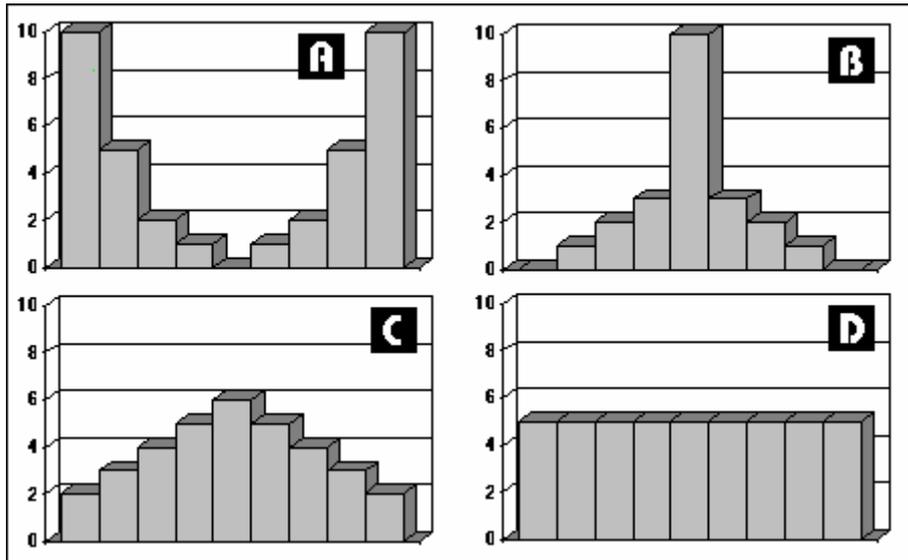
Nº de hijos	0	1	2	3	4	5	6	7	8
Nº de familias	11	13	20	25	14	10	4	2	1

Se pide:

- Calcular la tabla completa de frecuencias.
- Representaciones gráficas.
- Calcular la media, mediana y moda.
- Hallar el recorrido, varianza y desviación típica.

5

Ordenar las cuatro distribuciones siguientes de mayor a menor dispersión.



6

Los precios de una chaqueta en once establecimientos fueron (en pts.):

5000 5200 5300 5600 6000 6400 6500 7200 7300 8400 9000

Calcular la desviación media respecto de la mediana y respecto de la media.

7

Si en una distribución de frecuencias duplicamos las amplitudes de los intervalos, ¿ qué sucederá, aproximadamente, con los valores de las frecuencias ?.

8

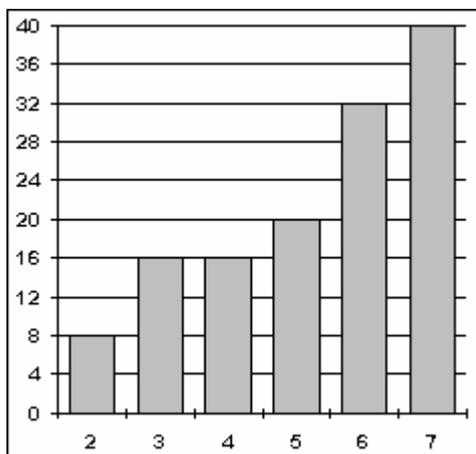
Represente el histograma correspondiente a la siguiente distribución de edades de los trabajadores de una fábrica.

Edades	Nº de trab.
de 20 a menos de 25	15
de 25 a menos de 35	20
de 35 a menos de 45	48
de 45 hasta 65	24

9

Ponga un ejemplo sencillo de una distribución de frecuencias simétrica. Calcule su moda, media y mediana, verificando que los tres parámetros coincidan.

10



A la izquierda se muestra el gráfico representativo de las frecuencias absolutas acumuladas de la distribución de edades de 40 individuos.

- a) Obtenga su media, mediana y moda.
- b) ¿ Cuántos tienen edades inferiores a cinco años y medio ?

11

Una variable X tiene como media 21 y varianza 9. Si se obtiene una nueva variable Y multiplicando los elementos de X por 4 y restándoles 8 unidades, ¿ cuál es el valor del coeficiente de variación de Y ?.

12

Una variable X toma los valores :

2      5      5      6      7

Realizada una transformación lineal con ella, se generó una nueva variable de la que conocemos que su media era 15 y que la puntuación X=2 se transformó en Y=13. Calcule las cuatro puntuaciones Y desconocidas.

13

X	n
0	3
1	9
2	13
3	25
4	16
5	14

Estudie la simetría y el apuntamiento (curtosis) de la distribución de la izquierda.

NOTA :

Obtenga los distintos coeficientes conocidos. Compare los resultados.

14

Nota	Alumnos
9 - 10	2
7 - 8	0
5 - 6	4
3 - 4	14
1 - 2	12

La tabla de la izquierda nos muestra la distribución de calificaciones de los 32 alumnos de un curso.

- a) Determine su media, mediana y moda.
- b) ¿ Qué porcentaje de observaciones tienen nota inferior a 1'62 ?.
- c) ¿ Entre qué valores se encuentra el 70% de las notas centrales ?
- d) Obtenga el coeficiente de variación y la amplitud semi-intercuartílica.

15

Nota	n	N
[0 , 1)		1
[1 , 2)		1
[2 , 3)		5
[3 , 4)	3	
[4 , 5)		11
[5 , 6)	6	
[6 , 7)		19
[7 , 8]		

De la distribución de notas de 20 alumnos, calcular :

- a) Frecuencias absolutas simples (f) y acumuladas (F) que faltan en la tabla.
- b) Coeficiente de variación.
- c) Porcentaje de alumnos con notas inferiores a 2'6.
- d) ¿ Entre qué notas se encuentra el 10% de las calificaciones centrales ?.
- e) Momentos ordinarios y centrales hasta el 4º orden.
- f) Coeficientes de asimetría y curtosis, utilizando los momentos calculados en e).

16

Con el fin de estudiar la distribución de fallos en una pieza de tela, se realizó un recuento de los contenidos en cada metro. Los resultados fueron los siguientes :

Fallos	0	1	2	3	4	5	6	7	8	9
Nº de metros	25	8	4	1	1	1	2	1	3	4

- a) Estudie el grado de concentración de la distribución de fallos a lo largo de la pieza de tela.
- b) Calcule su media y su mediana.

17

La tabla siguiente muestra los fallos cometidos por alumnos en la realización de un test de 120 items.

Errores	[0 , 10)	[10 , 20)	[20 , 30)	[30 , 40)	[40 , 50)	[50 , 60)	[60 , 70)	[70 , 80)
Alumnos	25	20	22	16	29	24	38	26

- a) Estudie el grado de concentración de la distribución de preguntas con respuesta errónea.
- b) Calcule su mediana.

# SOLUCIONES DE LOS EJERCICIOS PROPUESTOS

1

a)

x	n	r	p	N	R	P
13	1	0'02	2	1	0'02	2
14	5	0'10	10	6	0'12	12
15	10	0'20	20	16	0'32	32
16	14	0'28	28	30	0'60	60
17	13	0'26	26	43	0'86	86
18	5	0'10	10	48	0'96	96
19	2	0'04	4	50	1'00	100

b)  $Mo = 16$

c)  $\bar{x} = 16'12$  ;  $s^2 = 1'7856$  ;  $s = 1'3363$

d)  $Me = 16$  ;  $P_{29} = 15$  ;  $Q = 1$

2

a)

Intervalo	n	r	p	N	R	P
[35,40)	3	0'075	7'5	3	0'075	7'5
[40,45)	14	0'350	35'0	17	0'425	42'5
[45,50)	8	0'200	20'0	25	0'625	62'5
[50,55)	9	0'225	22'5	34	0'850	85'0
[55,60]	6	0'150	15'0	40	1'000	100'0

b)  $Mo = 43'636$

c)  $\bar{x} = 47'625$  ;  $s^2 = 36'859$  ;  $s = 6'071$

d)  $Me = 46'875$  ;  $D_8 = 53'889$

3

a)  $\bar{x} = 2'4$

b)  $2'4 + A$

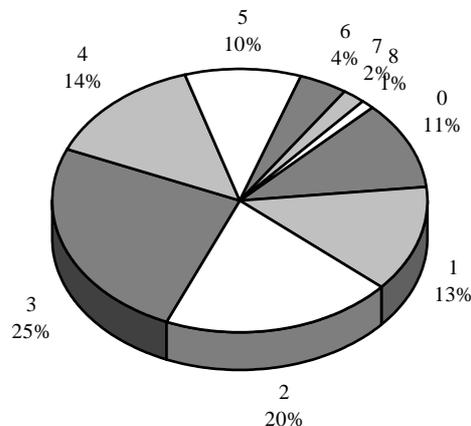
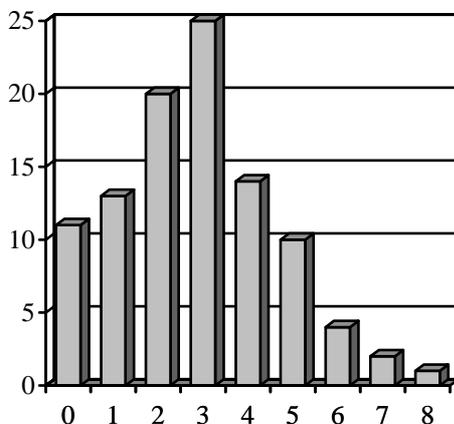
c) Realizando el cambio :  $y = \frac{x - 2754}{2}$

4

a)

x	n	r	p	N	R	P
0	11	0'11	11	11	0'11	11
1	13	0'13	13	24	0'24	24
2	20	0'20	20	44	0'44	44
3	25	0'25	25	69	0'69	69
4	14	0'14	14	83	0'83	83
5	10	0'10	10	93	0'93	93
6	4	0'04	4	97	0'97	97
7	2	0'02	2	99	0'99	99
8	1	0'01	1	100	1'00	100

b)



c)  $\bar{x} = 2'8$  ;  $Me = 3$  ;  $Mo = 3$

d)  $R = 8$  ;  $s^2 = 3'14$  ;  $s = 1'772$

5

A , D , C , B.

6

$$D_{Me} = D_{\bar{x}} = 870$$

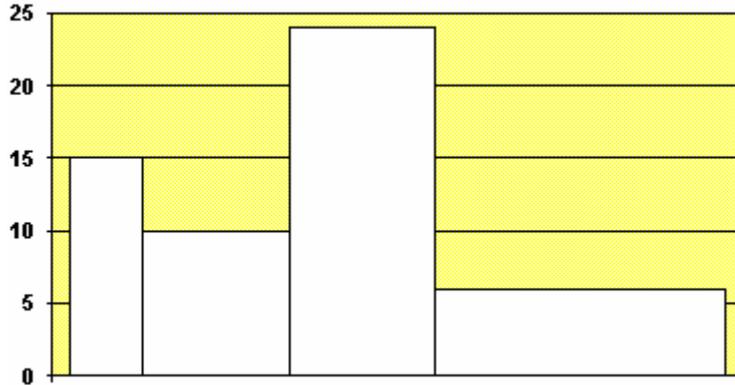
7

Se dividen por dos.

8

Las alturas deben ser proporcionales al área. Dividimos las frecuencias según sea la amplitud del intervalo.

Alturas : 15      10 (20/2)      24 (48/2)      6 (24/4)



9

x	n
0	2
1	8
2	20
3	8
4	2
	40

$$\bar{x} = Me = Mo = 2$$

10

- a)  $\bar{x} = 4'7$  ;  $Me = 5$  ;  $Mo = 6$   
 b) 20

11

$$CV = 15'789$$

12

15 , 15 , 15'667 , 16'333

13

$$As = \frac{\sum n_i \cdot (x_i - \bar{x})^3}{\sigma^3} = -0'299561$$

ligeramente asimétrica a la izquierda

$$As_1 = \frac{\bar{x} - Mo}{\sigma} = 0'036786$$

ligeramente asimétrica a la derecha (prácticamente

simétrica).

$$As_2 = \frac{3 \cdot (\bar{x} - Md)}{\sigma} = -0'110357$$

ligeramente asimétrica a la izquierda

Los coeficientes basados en la moda y la mediana hacen uso de una relación teórica entre los parámetros de centralización. Generalmente no conducen a la misma conclusión, salvo distribuciones claramente asimétricas.

