Ejercicios Resueltos de Estadística:
Tema 1: Descripciones univariantes

1. Los datos que se dan a continuación corresponden a los pesos en Kg. de ochenta personas:

(a) **Obténgase una distribución de datos en intervalos de amplitud 5, siendo el primer intervalo [50; 55].**

(b) **Calcúlese el porcentaje de personas de peso menor que 65 Kg.**

(c) **¿Cuántas personas tienen peso mayor o igual que 70 Kg. pero menor que 85?**

60; 66; 77; 70; 66; 68; 57; 70; 66; 52; 75; 65; 69; 71; 58; 66; 67; 74; 61; 63; 69; 80; 59; 66; 70; 67; 78; 75; 64; 71; 81; 62; 64; 69; 68; 72; 83; 56; 65; 74; 67; 54; 65; 65; 69; 61; 67; 73; 57; 62; 67; 68; 63; 67; 71; 68; 76; 61; 62; 63; 76; 61; 67; 67; 64; 72; 64; 73; 79; 58; 67; 71; 68; 59; 69; 70; 66; 62; 63; 66;

<u>SOLUCIÓN:</u>

(a) Como se trata de efectuar una distribución de datos agrupados, debemos obtener primero los intervalos correspondientes, situando los datos en sus lugares respectivos:

L_{i-1} - L_i	n_i	N_i
[50;55)	2	2
[55; 60)	7	9
[60; 65)	17	26
[65;70)	30	56
[70; 75)	14	70
[75; 80)	7	77
[80; 85]	3	80
	80	

(b) Observando la columna de frecuencias acumuladas se deduce que existen $N_3 = 26$ individuos cuyo peso es menor que 65 Kg., que en términos de porcentaje corresponden a:

$$\frac{26}{80} \cdot 100 = 32{,}5\%$$

(c) El número de individuos con peso comprendido entre 70 y 85 Kg. es:

$n_5 + n_6 + n_7 = 14 + 7 + 3 = 24$

lo que es equivalente a: $N_7 - N_4 = 80 - 56 = 24$

2. Dada la distribución siguiente, constrúyase una tabla estadística en la que aparezcan las frecuencias absolutas, las frecuencias relativas y las frecuencias acumuladas relativas crecientes:

x_i	1	2	3	4	5	6
n_i	5	7	9	6	7	6

SOLUCIÓN:

La tabla que se obtiene es la siguiente:

x_i	n_i	f_i	$F_i\downarrow$
1	5	0,125	0,125
2	7	0,175	0,300
3	9	0,225	0,525
4	6	0,15	0,675
5	7	0,175	0,85
6	6	0,15	1
	40	1	

3. Las edades de los empleados de una determinada empresa son las que aparecen en la siguiente tabla:

Edad	Nºempleados
Menos de 25	22
Menos de 35	70
Menos de 45	121
Menos de 55	157
Menos de 65	184

Sabiendo que el empleado más joven tiene 18 años, escríbase la distribución de frecuencias acumuladas decrecientes (o «más de»).

SOLUCIÓN:

Es preciso obtener, en principio, la distribución de frecuencias absolutas:

L_{i-1} - L_i	n_i
[18; 25) [25; 35) [35; 45) [45; 55) [55; 65]	22 48 51 36 27
	184

A la vista de la tabla anterior, la distribución pedida es:

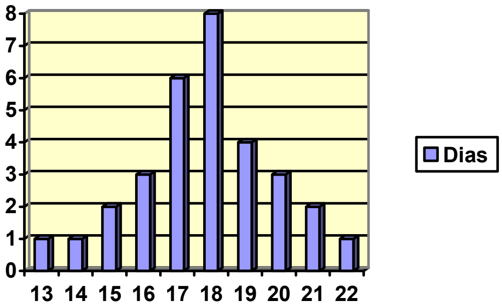
Edad	*N.° de empleados*
Más de 18	184
Más de 25	162
Más de 35	114
Más de 45	63
Más de 55	27

4. Las temperaturas medias registradas durante el mes de mayo en Madrid, en grados centígrados, están dadas por la siguiente tabla:

Temperatura	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
N.° de días	**1**	**1**	**2**	**3**	**6**	**8**	**4**	**3**	**2**	**1**

Construýase la representación gráfica correspondiente.

SOLUCIÓN:



---

5. Dada la distribución de frecuencias:

x_i	n_i
1	9
2	22
3	13
4	23
5	8
6	25

(a) **Constrúyase una tabla en la que aparezcan frecuencias absolutas, frecuencias relativas, frecuencias acumuladas absolutas crecientes (o «menos de») y decrecientes (o «más de»).**

(b) **Represéntese mediante un diagrama de barras la distribución dada y su correspondiente polígono de frecuencias.**

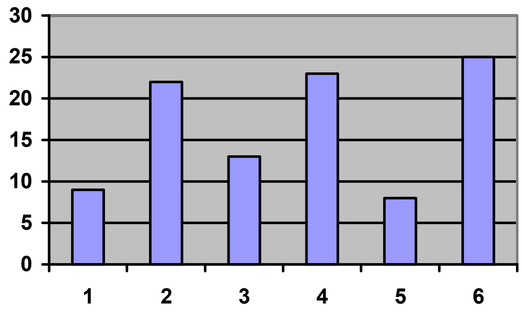
(c) **Obténgase el polígono de frecuencias absolutas acumuladas crecientes y decrecientes.**

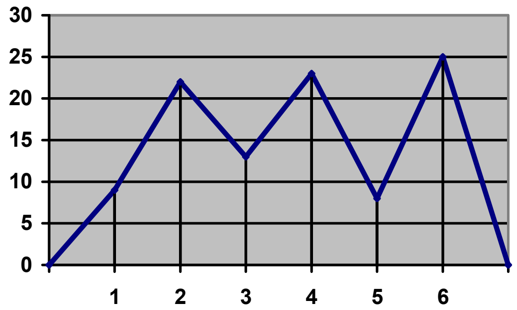
<u>SOLUCIÓN:</u>

(a) La tabla pedida es la siguiente:

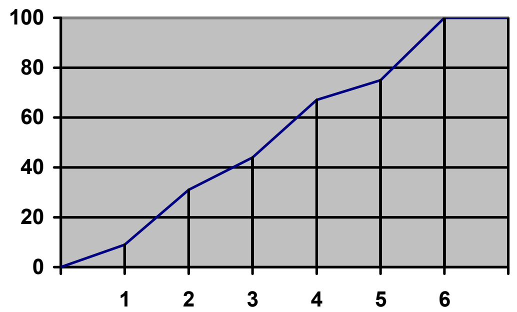
x_i	n_i	f_i	$N_i\downarrow$	$N_i\uparrow$
1	9	0,09	9	100
2	22	0,22	31	91
3	13	0,13	44	69
4	23	0,23	67	56
5	8	0,08	75	33
6	25	0,25	100	25
	100	1		

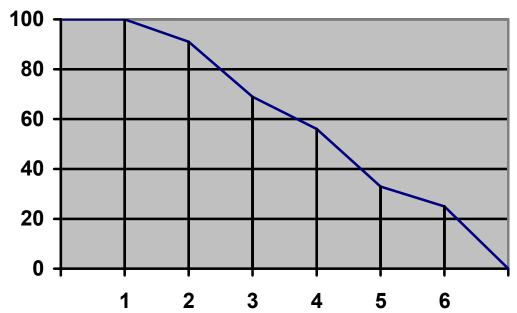
(b)





(c)





---

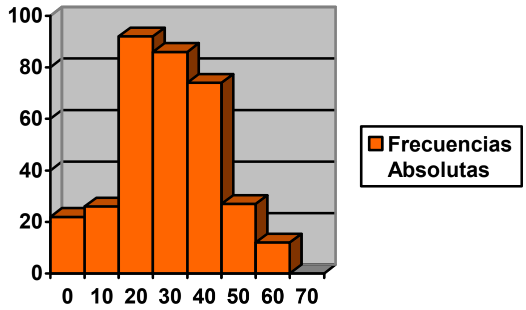
6. Represéntese gráficamente la siguiente distribución de frecuencias:

L_{i-1}-L_i	n_i
0-10	**22**

10-20	26
20-30	92
30-40	86
40-50	74
50-60	27
60-70	12

SOLUCIÓN:

Como es una distribución de datos agrupados, o de tipo III, cuyos intervalos tienen amplitudes iguales *(a* = 10), su representación gráfica es el histograma siguiente, en el que se han colocado como alturas las frecuencias absolutas:



7. Dada la siguiente distribución de frecuencias:

L_{i-1}-L_i	n_i
1-3	3
3-7	29
7-8	35
8-10	26
10-13	6
13-20	1

(a) **Constrúyase una tabla en la que aparezcan las marcas de clase, las frecuencias absolutas y relativas y las frecuencias absolutas acumuladas crecientes (o «menos de») y decrecientes (o «más de»).**

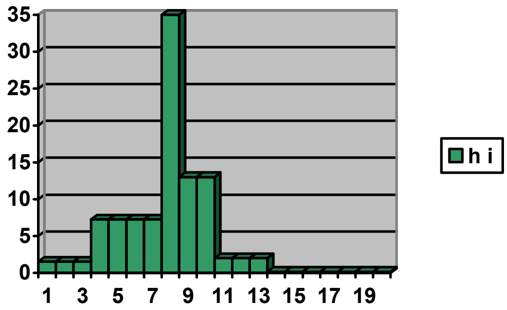
(b) **Represéntese la distribución mediante un histograma y su correspondiente polígono de frecuencias.**

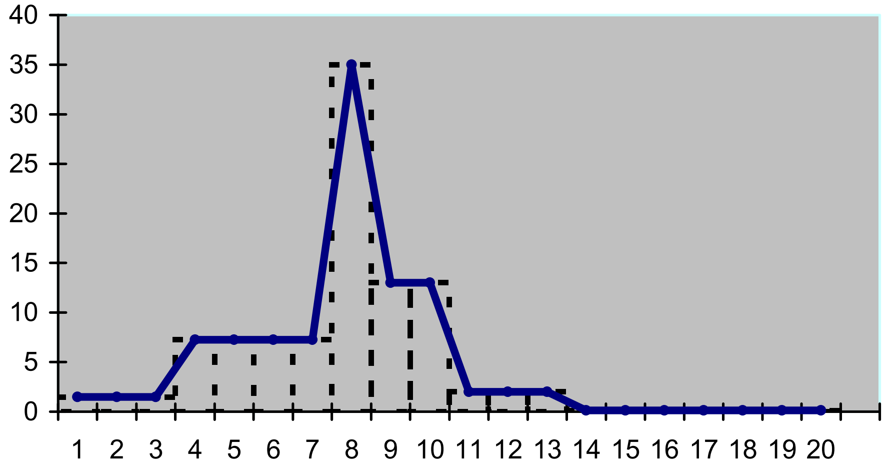
SOLUCIÓN:

(a) La tabla pedida es la siguiente, en la que se han añadido, además, la columna de las amplitudes de los intervalos y la columna de las alturas correspondientes para construir el histograma.

L_{i-1}-L_i	n_i	x_i	f_i	$N_i\downarrow$	$N_i\uparrow$	a_i	h_i
[1;3)	3	2	0,03	3	100	2	1,5
[3;7)	29	5	0,29	32	97	4	7,25
[7; 8)	35	7,5	0,35	67	68	1	35
[8; 1)	26	9	0,26	93	33	2	13
[10;13)	6	11,5	0,06	99	7	3	2
[13;20]	1	16,5	0,01	100	1	7	0,143
	100		1				

(b) Con la primera y última columna de la tabla anterior se obtienen el siguiente histograma y su polígono de frecuencias:





---

8. Encuestados cincuenta matrimonios respecto a su número de hijos, se obtuvieron los siguientes datos:

2 ; 4 ; 2 ; 3 ; 1 ; 2 ; 4 ; 2 ; 3 ; 0 ; 2 ; 2 ; 2 ; 3 ; 2 ; 6 ; 2 ; 3; 2; 2; 3; 2; 3; 3; 4;1 ; 3 ; 3 ; 4 ; 5 ; 2 ; 0 ; 3 ; 2 ; 1; 2; 3; 2; 2; 3; 1 ; 4 ; 2 ; 3 ; 2 ; 4 ; 3 ; 3 ; 2

Constrúyase una tabla estadística que represente dichos datos:

SOLUCIÓN:

Efectuando el recuento de los datos se obtiene:

x_i	n_i
0	2
1	4
2	21
3	15
4	6
5	1
6	1
	50

9. Calcula la media, la varianza, la desviación típica y el coeficiente de variación de Pearson tras Tras encuestar a 25 familias sobre el número de hijos que tenían, se obtuvieron los siguientes datos,

Nº de hijos(X_i)	**0**	**1**	**2**	**3**	**4**	
Nº de familias(n_i)	**5**	**6**	**8**	**4**	**2**	**25**

SOLUCIÓN:

Las cuatro distribuciones de frecuencia serán:

X_i	n_i	f_i	N_i	F_i
0	5	0'20	5	0'20
1	6	0'24	11	0'44
2	8	0'32	19	0'76
3	4	0'16	23	0'92
4	2	0'08	25	1
	25	**1**		

La *Media Aritmética* de las veinticinco familias encuestadas será:

$$a = \frac{\sum_{i-1}^{5} x_i \cdot n_i}{n} = \frac{0 \cdot 5 + 1 \cdot 6 + 2 \cdot 8 + 3 \cdot 4 + 4 \cdot 2}{25} = \frac{42}{25} = 1{,}68$$

es decir, las familias encuestadas tienen un número medio de hijos de 1'68.

El *Recorrido* será *R* = 4 - 0 = 4.

La *Varianza* es:

$$s^2 = 4'24 - (1'68)^2 = 1'4176.$$

Y la *Desviación Típica s* = 1'85.

Para este ejemplo el *Coeficiente de Variación de Pearson*, V_p, toma el valor:

$$v_p = \frac{1{,}19062}{1{,}68} \cdot 100 = 70{,}869$$

En cuanto a la simetría, el *Coeficiente de Variación de Pearson*, A_p,es igual a:

$$A_p = \frac{1{,}68 - 2}{1{,}1906} = -0{,}2688$$

Con lo que la distribución es ligeramente asimétrica a la izquierda.

---

10. Calculo de la media aritmética, la mediana y la moda. Se analizó el IVA que se aplica, en diversos países europeos, a la compra de obras de arte. Los resultados obtenidos fueron los siguientes:

<u>PAIS</u>	
España	**0,16**
Italia	**0,20**
Bélgica	**0,06**
Holanda	**0,06**
Alemania	**0,07**
Portugal	**0,17**
Luxemburgo	**0,06**
Finlandia	**0,22**

<u>SOLUCIÓN:</u>

Ahora realizamos las cuatro distribuciones de frecuencias:

Xi	ni	fi	Ni	Fi
0,06	3	0,375	3	0,375
0,07	1	0,125	4	0,500
0,16	1	0,125	5	0,625
0,17	1	0,125	6	0,750
0,20	1	0,125	7	0,875
0,22	1	0,125	8	1
Total	**8**	**1**		

Calculamos la media aritmética:

$$a = \frac{\sum x_i \cdot n_i}{n} = \frac{1}{8} = 0{,}125.$$

Ahora calculamos la mediana:

$$Me = \frac{x_{j-1} + x_j}{2} = \frac{0{,}07 + 0{,}16}{2} = 0{,}115.$$

Por último, el valor mas frecuente, correspondiente a la *moda*, es el valor:

x_j = 0,06. Por tanto:

M_d = **0,06.**

---

11. Con los mismos datos del ejercicio anterior vamos a calcular los cuartiles:

SOLUCIÓN:

Como sabemos el segundo cuartil es igual a la mediana:

$P_{2/4} = M_e =$ **0,115.**

Para determinar los otros dos cuartiles $p_{1/4}$ Y $p_{3/4}$, debemos establecer primero las desigualdades:

$$N_{j-1} < \frac{r}{k} \cdot n < N_j$$

Para los casos r/k = 1/4 y r/k = 3/4.

Para el primer cuartil:

$$\frac{1}{4} \cdot 8 = 2 < 3 = N_1$$

Es decir menor que la primera frecuencia absoluta acumulada, por tanto:

$P_{1/4} = \mathbf{0{,}06.}$

Ahora calculamos el tercer cuartil:

$$N_4 = 6 = \frac{3}{4} \cdot 8 = 6 < 7 = N_5$$

$$p_{3/4} = \frac{0{,}17 + 0{,}2}{2} = \mathbf{0{,}185.}$$

---

12. Del siguiente ejercicio calcular la varianza y la desviación típica.

X	Intervalo	f.absoluta	f.acumulada	f.relativa	f.r.acumulada	f.x	x^2	f. x^2
52	50-54	7	7	0,078	0,078	364	2704	18928
56	54-58	10	17	0,111	0,189	560	3136	31360
60	58-62	16	33	0,178	0,367	960	3600	57600
64	62-66	20	53	0,222	0,589	1280	4096	81920
68	66-70	18	71	0,2	0,789	1224	4624	83232
72	70-74	11	82	0,122	0,911	792	5184	57024
76	74-78	8	90	0,089	1	608	5776	46208
448		90		1		5788		376272

SOLUCIÓN:

Varianza:

$$S^2 = [\,\Sigma f \cdot x^2 - [\,(\Sigma f \cdot x)^2 / N\,]\,] / (N - 1)$$

$$S^2 = [\,376272 - [\,(5788)^2 / 90\,]\,] / (90 - 1)$$

$\mathbf{S^2 = 45{,}402.}$

Desviación típica:

(Raiz cuadrada de la varianza.)

S = 6,74

---

13. Para los siguientes datos, calcular:

A) El intervalo de intercuartil.
B) La desviación del cuartil.

97 72 87 57 39 81 70 84 93 79
84 81 65 97 75 72 84 96 94 77

SOLUCIÓN:

A)

$$RQ = \frac{Q_3 - Q_1}{2} = IQ$$

$$RQ = \frac{15}{2}$$

$$= RQ = 7{,}5$$

B)

$$IQ = Q_3 - Q_1$$

$$= 87 - 72$$

$$= IQ = \mathbf{15}$$

14. Unos grandes almacenes disponen de un aparcamiento para sus clientes. Los siguientes datos que se refieren al número de horas que permanecen en el aparcamiento una serie de coches:

4	5	5	1	7	4	4	3	6	5
3	2	4	4	3	6	6	4	5	5
6	4	3	3	4	5	4	3	2	4
5	2	4	7	3	6	2	2	4	1
2	1	3	7	3	1	5	1	7	2
4	4	2	4	5	3	6	3	5	3

Se pide:

A- Obtener la tabla de frecuencias para ese conjunto de datos. Interpretar la tabla.
B- Obtener la tabla de frecuencias ascendente y descendente.
C- Determinar e interpretar la tercera cuartilla y el centil del 42%.
D- Calcular el tiempo medio de permanencia de los coches en el aparcamiento. Interpretar el resultado y los elementos que intervienen.

SOLUCIÓN:

A- El primer paso para construir la tabla de frecuencias es determinar el número de valores diferentes en observación, k, que en este caso es 7. A continuación podemos ver que esos 7 valores van desde el 1, x $_1$, al 7 $_7$, y podemos determinar la frecuencia absoluta y relativa de cada uno de esos valores. Una vez calculadas las frecuencias resulta la siguiente tabla de frecuencias.

$x_1 (n^o\ horas)$	1	2	3	4	5	6	7
$n_i (n^o\ coches)$	5	8	12	15	10	6	4
$f_i (\%coches)$	8.33	13.33	20	25	16.67	10	6.67

En esta tabla aparecen por filas el número de horas que permanecen los coches en el aparcamiento, el número de coches que han aparcado durante cada número de horas y la proporción de coches en % que han estado aparcados durante cada número de horas. Una de las columnas, por ejemplo la cuarta, nos dice que 15 coches, que representa el 25% de los coches analizados, han estado aparcados durante 4 horas en el aparcamiento.

B- La tabla de frecuencias ascendente es

$x_i(n^o\ horas)$	1	2	3	4	5	6	7
$\sum_{j=1}^{i} n_j(n^o\_coches\_acumulados)$	5	13	25	40	50	56	60
$\sum_{j=1}^{i} f_j(proporción\_acumulada)$	8.33	21.67	41.67	66.67	83.33	93.33	100

La tabla de frecuencias descendente es:

$x_i(n^o\ horas)$	1	2	3	4	5	6	7
$\sum_{j=i}^{7} n_j(n^o\ coches\_acumulados)$	60	55	47	35	20	10	4
$\sum_{j=i}^{7} f_j(proprción\_acumulada)$	100	91.67	78.34	58.34	33.34	16.67	6.67

C- La tercera cuartilla es el centil 75%, luego el ser N = 60 calculamos 0.75*60=45 que al ser entereo, la fórmula aplicada será

$$c_{0.75} = \frac{x_{(45)} + x_{(46)}}{2} = \frac{5+5}{2} = 5 \text{ horas}$$

Su significado es que el 75% de los coches analizados estacionan en el aparcamiento a lo sumo, o como máximo, 5 horas.

Para calcular el centil 42% hallamos 0.42*60=25.2, que al no ser entero, deberemos utilizar la otra fórmula.

$$c_{0.42} = x_{([25.2]+1)} = x_{(26)} = 4 \text{ horas}$$

Su significado es que el 42% de los coches analizados estacionan en el aparcamiento a lo sumo, o como máximo, 4 horas.

D- Según la primera fórmula, el tiempo medio de permanencia de los coches en el aparcamiento es

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i * x_i}{N} = \frac{231}{60} = 3.85 \text{ horas}$$

Se calcula dividiendo el tiempo total de permanencia de todos los coches en el aparcamiento, 231 horas, entre los coches analizados, 60.
En la segunda fórmula se calcula el tiempo medio como resultado de las aportaciones que hacen a dicho valor los productos de los diferentes valores del número de horas que han estado los coches aparcados, x_i, por la proporción de lcoches, f_i, que han estado aparcados durante cada número de horas. Por tanto,

$$\overline{X} = \sum_{i=1}^{k} f_i * x_i = 3.85 \text{ horas}$$

En promedio, cada coche ha estado estacionado 3 horas y 51 minutos, y el tiempo total de permanencia en el aparcamiento de los 60 coches ha sido 231 horas.

15. Un fabricante de neumáticos ha recabado, de los diferentes concesionarios, información sobre la cantidad de miles de kilómetros recorridos por un modelo concreto de esos neumáticos hasta que se ha producido un pinchazo o un reventón del neumático. Los concesionarios la han proporcionado los siguientes datos:

52.452	50.432	37.748	51.831	73.808	61.065	35.807	57.277
48.698	65.854	75.850	36.949	75.548	69.010	61.477	65.585
44.411	41.886	34.754	59.888	59.449	67.632	89.116	69.483
63.692	70.003	65.996	55.989	49.677	46.502	67.467	64.398
84.588	40.709	50.238	61.390	85.720	45.313	46.724	61.752
55.643	55.912	46.681	66.519	59.168	66.313	35.884	28.625
47.012	71.360	78.635	41.715	72.635	41.463	48.996	48.172
79.426	67.662	53.324	49.011	29.480	41.128	30.252	33.412
48.240	57.884	55.257	84.656	48.662	10.504	60.951	38.420
74.239	60.727	56.155	86.070	90.565	53.751	76.580	68.629
51.179	74.582	58.708	48.035	67.124	41.830	61.030	58.267
61.979	4.3068	41.539	62.215	51.269	82.919	34.182	37.654
80.502	35.342	44.719	37.402				

Se pide:

- **a- Construir una taba de frecuencias para esos datos tomando como número de intervalos el que proporciona la fórmula de Sturgess. Interpretas la tabla.**
- **b- Construir las tablas de frecuencias acumuladas ascendente y descendente.**
- **c- Dibujar el histograma de frecuencias relativas sin acumular y acumulado.**
- **d- Calcular las principales medidas de tendencia central e interpretarlas.**
- **e- Obtener las medidas de dispersión más importantes e interpretarlas.**

f- **Analizar la asimetría y el apuntamiento de la distribución de frecuencias resultante.**
g- **Si el fabricante quiere proponer un kilometraje para realizar el cambio de neumáticos, ¿qué valor propondría para que solo 3 de cada 10 coches hayan tenido un pinchazo o reventón antes de ese kilometraje?**

<u>SOLUCIÓN:</u>

a- La fórmula de Sturgess propone como número k de intervalos, para agrupar un conjunto de N observaciones en intervalos.

k=1+ [3.3*log N]

En este caso N=100, luego k=7. ahora debemos propones el límite inferior del primer intervalo y el límite superior del último intervalo. Al ser el valor mínimo 4.3068 se propone 4 como límite inferior del primer intervalo, y al ser 7 intervalos se propone como anchura 13 para cada uno de ellos, para que sea un valor entero, con lo cual el límite superior del último intervalo es 95.

La tabla de frecuencias será:

$Intervalo\_I_i$	$4 < x \le 17$	$17 < x \le 30$	$30 < x \le 43$
$Frecuencia$ $absoluta\_n_i$	2	2	19
$Frecuencia$ $relativa\_f_i$	.02	.02	.19

I_i	$43 < x \le 56$	$56 < x \le 69$	$69 < x \le 82$	$82 < x \le 95$
n_i	27	29	14	7
f_i	.27	.29	.14	.07

En esta tabla aparecen por filas los intervalos, junto con la frecuncia absoluta y la frecuencia relativa. Por ejemplo la cuarta columna se puede interpretar diciendo que el 27% de estos neumáticos han recorrido entre 43000 y 5600 Km hasta que se ha producido un pinchazo o reventón.

b- La tabla de frecuencias acumuladas ascendente sería:

$Intervalos\_I_i$	(4,17]	(17,30]	(30,43]	(43,56]	(56,69]	(]69,82	(82,95]
$\sum_{j=1}^{i} n_j$	2	4	23	50	79	93	100

y la tabla de frecuencias acumuladas descendente quedaría

$Intervalos\_I_i$	(4,17]	(17,30]	(30,43]	(43,56]	(56,69]	(69,82]	(82,985]
$\sum_{j=1}^{k} n_j$	100	98	96	77	50	21	7

c- El histograma de frecuencias relativas se represena es la figura 1 y el de frecuencias acumuladas en la figura 2.

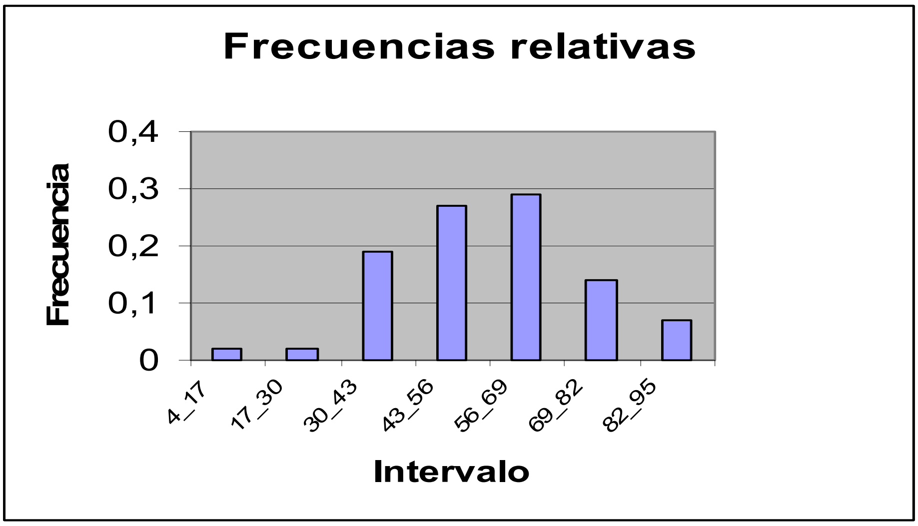


Figura 1

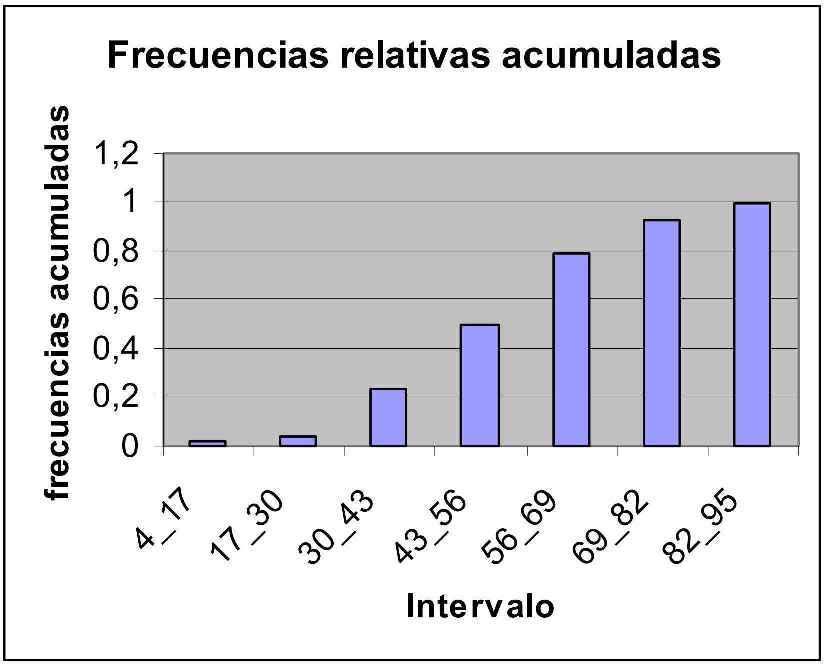


Figura 2

d- Para calcular las medidas de tendencia central trabajamos con la tabla de frecuencias del apartado a. resulta que la media aritmética es

$$\overline{X} = 55870 \text{ Km}$$

Se interpreta diciendo que son los 100 neumáticos analizados se han recorrido 5587000 de Km antes de un pinchazo o reventón.

La mediana será

$$Me = 56000 \text{ Km}$$

Significa que la mitad de los neumáticos han recorrido a lo sumo 56000 Km antes de un pinchazo o reventón.
La moda será

$$Mo = 56 + 13 * \frac{2}{2+15} = 57529 \text{ Km}$$

Significa que la cantidad más frecuente, de kilómetros recorridos antes de un pinchazo, a sido 57529 Km.

e- La desviación típica es

$$s = 16899 \text{ Km}$$

y nos informa sobre lo que se dispersan los kilómetros recorridos por los diferentes neumáticos respecto del kilometraje medio.
El coeficiente de variación de Pearson será

$$g = \frac{s}{x} * 100\% = 30.24$$

Al tomar un valor inferior al 100% resulta que la mediana es representativa, y al ser dicho valor del 30% nos informa que el valor de la desviación típica es el 30% del valor de la media.

f- Los coeficientes de asimetría de Pearson son en este caso

$$V_1 = \frac{55.87 - 57.529}{16.899} = -0.09817$$

$$V_2 = \frac{55.87 - 56.00}{16.899} = -0.02308$$

Para calcular el coeficiente g_1 calculamos

$$m_3 = \sum_{i=i}^{7} f_i (x_i - \overline{x})^3 = -421.39015$$

Luego g_1=-0.08732, resultado de dividir m_3 entre s^3. a la vista de este

coeficiente de asimetría la distribución resulta ser ligeramente asimétrica a la izquierda, lo que significa que algo menos de la mitad de los neumáticos pinchan o revientan antes de los 5600 Km, valor mediano de la distribución.
Para el cálculo del coeficiente de curtosis g_2 necesitamos

$$m_4 = \sum_{i=1}^{7} f_i (x - \overline{x})^4 = 234594.7408$$

Luego

$$g_2 = \frac{m_4}{s^4} - 3 = -0.12343$$

Esto significa que la distribución es de tupo platicúrtica, algo menos apuntada que la distribución normal de media 55870 km y desviación típica 16899 km. Por tanto, en los intervalos $\overline{X} \pm$ ks con $k \in N$ habrá menos proporción de observaciones que en dicha distribución normal.

g- Propondría un kilometraje tal que el 70% de los neumáticos no hayan pinchado o reventado antes de este kilometraje. Por tanto, buscamos el centil del 30%, que vendrá dado por

$$c_{0.3} = 43 + 13 * \frac{7}{27} = 46.37$$

Luego el fabricante propondría cambiar los neumáticos a los 46370 km.

---

16. La tabla siguiente nos proporciona los valores de la media y la desviación típica de dos variables así como su coeficiente de correlación lineal para dos muestras diferentes:

$Muestra$	$n^o\ de\_observaciones$	$\overline{x}$	$\overline{y}$	s_x	s_y	r_{xy}
1	600	5	12	2	3	0.6
2	400	7	10	3	4	0.7

Se pide:

a- Recta de regresión de Y sobre X en cada muestra.

b- Si consideramos la muestra que resulta de agrupar las dos muestras en una sola de tamaño 1000, obtener el nuevo coeficiente de correlación lineal de Pearson y explicar el hecho de que sea inferior a los de cada una de las muestras tomadas por separado.

SOLUCIÓN:

a- La recta de regresión de Y sobre X en cada muestra es

$$y = \overline{y} + \frac{m_{11}}{s_x^2}(x - \overline{X})$$

Como la información dada es la del coeficiente de correlación lineal,

$$r_{xy} = \frac{m_{11}}{s_x s_y}$$

se tiene que la recta de regresión es

$$y = \overline{Y} + r_{xy}\frac{s_y}{s_x}(x - \overline{X})$$

Luego, sustituyendo, las rectas de regresión de Y sobre X en cada una de las dos muestras son:

Muestra 1: y=12+0.9*(x-4)
Muestra 2: y=10+0.93*(x-7)

b- Se trata de calcular el coeficiente de correlación lineal de Pearson en la nueva muestra de tamaño 1000, que notaremos por $r_{xy,t}$ y que será

$$r_{xy,t} = \frac{m_{11,T}}{s_{x,T}s_{y,T}}$$

donde $m_{11,T}$ es la covarianza en la muestra total y $s_{x,T}$, $s_{x,T}$ las desviaciones típicas de X e Y en la muestra total. Para obtener estas cantidades necesitamos $\overline{X}_T$ e $\overline{Y}_T$, medias de X e Y en la muestra total, que se calculas como un promedio entre las medias de X e Y en las muestras 1 y 2, notadas por $\overline{X}_1$, $\overline{Y}_1$, $\overline{X}_2$, $\overline{Y}_2$, según las relaciones siguientes

$$\overline{X}_T = \frac{\overline{X}_1 * 600 + \overline{X}_2 * 400}{1000}$$

$$\overline{Y}_T = \frac{\overline{Y}_1 * 600 + \overline{Y}_2 * 400}{1000}$$

Sustituyendo se obtiene que $\overline{X}_T = 5.8$ e $\overline{Y}_T = 11.2$.
Por otra parte si $m_{11,h}$ denota la covarianza en la muestra h, se tiene que

$m_{11,1}$=2*3*0.6=3.6
$m_{11,2}$=3*4*0.7=8.4

Como

$$m_{1,h} = \frac{\sum n_{ij,h}x_{i,h}y_{j,h}}{N_h} - \overline{X}_h\,\overline{Y}_h$$

resulta que:

$$\frac{\sum n_{ij,1}x_{i,1}y_{j,1}}{N_1} = 3.6 + 5.12 = 63.6$$

$$\frac{\sum n_{ij,2}x_{1,2}y_{l,2}}{N_2} = 8.4 + 7*10 = 78.4$$

Luego en la muestra total con N=1000 se tiene que

$$\frac{\sum n_{ij}x_i y_j}{N} = \frac{\sum n_{ij,1}x_{i,1}y_{j,1} + \sum n_{ij,2}x_{ij,2}x_{i,2}y_{j,2}}{N} =$$
$$= \frac{63.6*600 + 78.4*400}{1000} = 69.52$$

Por tanto

$$\mathrm{m}_{11,T} = \frac{\sum n_{ij}x_i y_j}{N} - \overline{X}_T\ \overline{Y}_T = 69.52 - 5*8*11.2 = 4.56$$

P ara obtener $\mathrm{s}_{x,t}$ y $\mathrm{s}_{y,t}$ utilizamos que

$$\mathrm{s}_{x,T} = \sqrt{a_{2,xT} - \overline{X}_T^{\,2}} \quad \text{y} \quad \mathrm{s}_{y,T} = \sqrt{a_{2,yT} - \overline{Y}_T^{\,2}}$$

donde

$$a_{2,xT} = \frac{a_{2,x1}*600 + a_{2,x2}*400}{1000}$$

$$a_{2,yT} = \frac{a_{2,y1}*600 + a_{2,y2}*400}{1000}$$

siendo

$$a_{2,Xh} = s_{X,h}^2 + \overline{X}_h^{\,2}$$

$$a_{2,Yh} = s_{Y,h}^2 + \overline{Y}_h^{\,2} \qquad \text{para h=1,2.}$$

Operando se obtiene que $\mathrm{a}_{2,X1}$=29, $\mathrm{a}_{2,X2}$=58, $\mathrm{a}_{2,Y1}$=153 y $\mathrm{a}_{2,Y2}$=116

Luego

$$a_{2,XT} = 0.6*29+0.4*58=40.6$$

$$a_{2,YT} = 0.6*153+0.4*116=138.2$$

de donde

$$s_{X,T} = \sqrt{40.6-5.8^2} = 2.6382$$

$$s_{Y,T} = \sqrt{138.2-11.2^2} = 3.5721$$

Luego resulta que

$$r_{xy,T} = \frac{4.56}{2.6382*3.5721} = 0.4838$$

Con lo cual el coeficiente de correlación lineal entre X e Y en la muestra total de 1000 observaciones es inferior al que hay en cada una de las dos muestras por separado. La explicación de este hecho es la siguiente: en cada muestra parcial se puede dar un mayor grado de relación lineal que en la muestra total porque las observaciones se encuentran mas agrupadas en torno a una recta que cuando las juntamos, ya que al formar la muestra total la nube de puntos resultante estará formada por las nubes de puntos de las muestras parciales y presentará un menor ajuste a una recta.

---

17. En una compañía aérea se sabe que, por término medio, el 65% de los vuelos tiene retraso. La distribución de los vuelos retrasados es la siguiente:

Duración del retraso (centésimas de hora)	**Numero de vuelos**
0-10	**2000**
10-20	**3000**
20-30	**2500**
30-50	**2000**
50-100	**500**

Se pide:

a- Determinas el retraso medio y la desviación típica del tiempo de retraso para los vuelos retrasados.

b- Determinar el centil del 60% e interpretarlo.

c- La compañía ha determinado que por cada vuelo con retraso se producen unas pérdidas fijas de 17000 pts y unas pérdidas variables de 10000 pts por cada minuto de retraso. ¿Entre qué cantidades se encuentran al menos las tres cuartas partes de las pérdidas generadas por cada vuelo retrasado?

d- Resolver el apartado a- para el total de los vuelos. ¿Es representativa la nueva media? En caso negativo propones razonadamente otra medida de centralización.

SOLUCIÓN:

a- Sea la variable estadística X: tiempo de retraso de un vuelo retrasado, y consideremos la tabla de frecuencias siguiente obtenida a partir de la dada con las marcas de clase

x_i	5	15	25	40	75
f_i	0.2	0.3	0.25	0.2	0.05

En esta tabla se verifica que

$\overline{X} = 23.5$ centésimas de hora

s_x =16.6658 centésimas de hora

b- De la tabla de frecuencias acumuladas siguiente

$[a_{i-1}, a_i)$	[0,10)	[10,20)	[20,30)	[30,50)	[50,100)
$\sum_{j=1}^{i} f_j$	0.2	0.5	0.75	0.95	1

se observa que el centil 60% se encuentra en el intervalo [20,30), luego

$c_{0.6}$ =20+z

por una regla de tres

$10 \rightarrow 0.25$

$z \rightarrow 0.1$

$$z = \frac{0.1*10}{0.25} = 4$$

Así $c_{0.6}$ =24 centésimas de hora y significa que el 60% de los vuelos retrasados (con menos tiempo de retraso) han tenido un retraso de a lo sumo 24 centésimas de hora y significa que el 60% de los vuelos retrasados (con menos tiempo de retraso) han tenido un retraso de a lo sumo 24 centésimas de horas.

c- Sea la variable estadística Y: pérdidas que se producen por un vuelo con retraso, se verifica que

$$Y = 17000 + 10000 * \frac{6}{10} X$$

ya que X*: tiempo de retraso de un vuelo retrasado en minutos se relaciona con X por la igualdad X*=$\frac{6}{10}X$.

Por aplicación de la desigualdad de Chebyshev se sabe que al menos las tres cuartas partes de las pérdidas generadas por cada vuelo retrasado se encuentran entre $\overline{Y}-2s_y$ e $\overline{Y}+2s_y$. Como

$$\overline{Y} = 17000 + 6000 * \overline{X} = 17000 + 6000 * 23.5 = 158000 pts$$
$$s_y = 6000 * s_x = 6000 * 16.6658 = 999948 pts$$

Resulta que: "Entre 0 pts y 3579896 pts se encuentran al menos las tres cuartas partes de las pérdidas generadas por cada vuelo retrasado".

Como

$$g_x = \frac{s_x}{|\overline{X}|} = \frac{16.6658}{23.5} = 0.7092$$ y

$$g_y = \frac{s_y}{|\overline{Y}|} = \frac{99994.8}{158000} = 0.6329$$

se deduce que hay más variabilidad en los tiempos de retraso.

d- Al considerar el total de los vuelos hay que modificar la tabla del anunciado por la tabla siguiente

Con la nueva variable estadística X*: tiempo de retraso de un vuelo cualquiera en centésimas de 1 hora.

Se verifica que

$\overline{X}^*$=15.275 centésimas de hora y s $_{X^*}$ =17.5 centésimas de hora.

Como g $_{X^*} = \frac{17.5}{15.275} = 1.1457 > 1$ la nueva media no es representativa al existir observaciones extremas. Una medida de centralización que evita este problema es la mediana. Para esta distribución se verifica que Me = 11.026 centésimas de hora.

---

18. En una clínica se han registrado durante un mes las longitudes en metros que los niños andan el primer día que comienzan a caminar, obteniéndose los siguientes resultados:

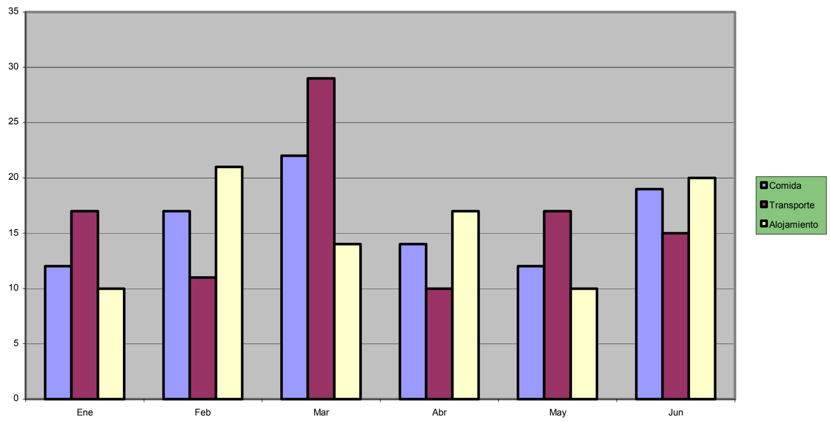
Número de metros	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Número de niños	**2**	**6**	**10**	**5**	**10**	**3**	**2**	**2**

Construir la distribución de frecuencias adecuada para la variable longitud y realizar los gráficos pertinentes que la representen.

SOLUCIÓN:

La tabla de frecuencias relativa a la variable se presenta a continuación:

X_i	n_i	N_i	f_i	F_i
1	2	2	0.05	0.05
2	6	8	0.15	0.2
3	10	18	0.25	0.45
4	5	23	0.125	0.575
5	10	33	0.25	0.825
6	3	36	0.075	0.9
7	2	38	0.05	0.95
8	2	40	0.05	1



19.- La distribución de los costes salariales de los 100000 empleados de una multinacional se presenta en la tabla siguiente:

Salarios	**Nº de empleados**
0 – 15000	**2145**
15000 – 20000	**1520**
20000 – 25000	**840**
25000 – 30000	**955**
30000 – 35000	**1110**
35000- 40000	**2342**
4000 – 50000	**610**
50000 – 100000	**328**

100000 - 300000	**150**

Calcular el salario medio por trabajador, el salario más frecuente y el salario tal que la mitad de los restantes sea inferior a él. Calcular también el primer cuartel salarial y el percentil 75.

SOLUCIÓN:

La tablas siguiente contiene los elementos relativos a la distribución d frecuencias de la variable salario (X) necesarios para realizar los cálculos pedidos en el problema.

$L_{(i-1)}$	L_i	n_i	Marcas = X_i	X_i*n_i	N_i	c_i	$D_i = n_i/c_i$
0	15000	2145	7500	16087500	2145	15000	0.143
15000	20000	1520	17500	26600000	3665	5000	0.304
20000	25000	840	22500	18900000	4505	5000	0.168
25000	30000	955	27500	26262500	5460	5000	0.191
30000	35000	1110	32500	36075000	6570	5000	0.222
35000	40000	2342	37500	87825000	8912	5000	0.4684
40000	50000	610	45000	27450000	9522	10000	0.061
50000	100000	328	75000	24600000	9850	50000	0.00656
100000	300000	150	200000	30000000	10000	200000	0.00075
		10000		293800000			

Para hallar el salario medio por trabajador calculamos la medida de la variable X.

$$\frac{293800000}{1000} = 29380$$

Para hallar el salario más frecuente se calcula la moda de la variable X. Para ello hemos de tener presente que los intervalos de la distribución de frecuencias son desiguales, por lo que l intervalo modal será el correspondiente al mayor valor de d_i, es decir será el intervalo (35000-40000).por lo tanto la moda se calcula como sigue:

$$M_0 = L_{i-1} + \frac{d_{+1}}{d_{i-1} + d_{i+1}} c_i = 35000 + \frac{0,061}{0,222+0,061} 5000 = 36077,74$$

Para hallar el salario tal que la mitad de los restantes sea inferior a él se calcula la mediana. Para llo, como N/2 = 5000, el intervalo mediano será (25000-3000) ya que $N_{i-1}<N/2>N_i$ es equivalente en este problema a 4505 < 50000< 5460.la mediana se calculará como sigue:

$$M_e = L_{i-1} + \frac{N/2 - N_{i-1}}{n_i} c_i = 25000 + \frac{1000/2 - 4505}{955} 5000 = 27591,62$$

Para calcular el primer cuartil (primer cuartil de orden 4) observamos que como N/4 = 2500, el intervalo relativo al primer cuartel será (15000-20000) ya que $N_{i-1}<2500<N_i$ es equivalente en este problema a 2145<2500<3655.El primer cuartel se calculará como sigue:

$$Q_{1,4} = L_{i-1} + \frac{N/4 - N_{i-1}}{n_i} c_i = 15000 + \frac{10000/4 - 2145}{1520} 5000 = 16167{,}76$$

El primer cuartil se interpreta como el valor de la variable para el que la cuarta parte de los valores menores que él y las tres cuartas partes resultantes son superiores.

Para calcular el percentil 75 (cuantil 75 de orden 100), observamos que como 75N7100 = 7500, el intervalo al percentil 75 será (3500 – 40000) ya que $N_{i-1}<7500<n_i$ es equivalente en este problema a 6570<7500<8912.El percentil 75 se calculará como sigue:

$$Q_{75,100} = L_{i-1} + \frac{75N/100 - N_{i-1}}{ni} c_i = 35000 + \frac{75*10000/100 - 6570}{2342} 5000 = 36985{,}48$$

El percentil 75 se interpreta como el valor de la variable para que el 75% de los valores son inferiores a él y el 25% restante son superiores.

20. Los rendimientos de cinco inversiones distintas realizadas por un individuo y las cantidades iniciales invertidas n unidades monetarias son los siguientes:

Cantidades iniciales	**Rendimientos**
200.000	**1000**
360.000	**900**
250.000	**500**
240.000	**800**
180.000	**1200**

Calcular el rendimiento medio por unidad monetaria invertida para el total de inversiones del individuo.

<u>SOLUCIÓN:</u>

Como se trata d promediar rendimientos por unidad, estamos ante un caso de aplicación del concepto de media armónica. Calcularemos por tanto el rendimiento medio por unidad monetaria invertida para el total de inversión del individuo como la media armónica de los rendimientos de cada inversión ponderada de las cantidades iniciales desembolsadas en cada inversión.

$$H = \frac{N}{\sum(1/x_i) * n_i} = \frac{200000 + 360000 + 250000 + 240000 + 180000}{200000 + 360000 + 250000 + 240000 + 180000} = 793{,}5$$

21. En el cuadro siguiente se presentan los consumos de electricidad en España en miles de millones de Kw/hora desde diciembre n 1985 hasta diciembre de 1986.

Meses	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov
Consumo	10,1	10,7	9,96	9,46	9,54	8,92	8,95	9,58	7,86	8,96	9,57	9,57

A partir de los incrementos unitarios de consumo de cada mes calcular el incremento

unitario anual medio acumulativo.

SOLUCIÓN:

Al tratarse del cálculo de una media acumulativa, el promedio más adecuado es la media geométrica. Se trata por tanto de calcular la media geométrica de los incrementos unitarios mensuales. Estos incrementos se calculan a continuación.

$\frac{10,7}{10,1} = 1,06$ $\frac{9,96}{10,7} = 0,93$ $\frac{9,46}{9,96} = 0,95$ $\frac{9,54}{9,46} = 1,008$ $\frac{8,92}{9,54} = 0,93$ $\frac{8,95}{8,92} = 1,003$

$\frac{9,58}{8,95} = 1,07$ $\frac{7,86}{9,58} = 0,82$ $\frac{8,96}{7,86} = 1,14$ $\frac{9,17}{8,96} = 1,02$ $\frac{9,57}{9,17} = 1,04$ $\frac{10,2}{9,57} = 1,06$

La media geométrica de estos incrementos unitarios mensuales se calcula como sigue:

$$G=\sqrt[12]{1,06*0,93*0,95*1,008*0,93*1,003*1,07*0,82*1,14*1,02*1,04*1,06} = 1,01$$

22. En la siguiente tabla se muestran las diferentes cantidades de IVA que se imponen en la compra de una obra de arte.

País	**IVA**
España	**0,16**
Italia	**0,20**
Bélgica	**0,06**
Holanda	**0,06**
Alemania	**0,07**
Portugal	**0,17**
Luxemburgo	**0,06**
Finlandia	**0,22**

Determine el recorrido, la varianza, la desviación típica, la cuasivarianza, la cuasidesviación típica, el coeficiente de variación de Pearson, el coeficiente de asimetría de Pearson y el coeficiente de asimetría de Fisher.

SOLUCIÓN:

- El recorrido:

$$R = (x)\max - (x)\min = 0,22 - 0,06 = 0,16$$

- La varianza:

$$s^2 = \frac{1}{n}\sum_{i=1}^{6} x_i^2 n_i - a^2 = \frac{0,1586}{8} = 0,125^2 = 0,0042$$

- De la misma forma la desviación típica se obtiene haciendo la raíz cuadrada de la varianza:

$$s = \sqrt{s^2} = \sqrt{0,048} = 0,06481$$

- La cuasivarianza:

$$S^2 = \frac{ns^2}{n-1} = \frac{8*0,0042}{7} = 0,0048$$

- De la misma forma la cuasidesviación típica:

$$S = \sqrt{S^2} = \sqrt{0,0048} = 0,06928$$

- El coeficiente de variación de Pearson:

$$V_p = \frac{s}{a}*100 = \frac{0,06481}{0,125}*100 = 51,848$$

- El coeficiente de asimetría de pearson:

$$A_p = \frac{a - M_d}{s} = \frac{0,125 - 0,06}{0,06481} = 1,00293$$

- Por ultimo el coeficiente de asimetría de Fisher:

$$A_f = \frac{\sum_{i=1}^{6}(x_i - a)^3 * n_i}{n * S^3} = \frac{0,000423}{8*0,0003325} = 0,159$$

---

23. Dados los siguientes datos:

Tamaño tabla	**n**
3-6	**37**
6-11	**198**
11-16	**191**
16-21	**149**
21-26	**79**
26-31	**46**
31-41	**55**
41-51	**51**
51-76	**26**
76-101	**25**
101-201	**25**
201-501	**11**
501-1000	**2**

Calcula la media aritmética, la mediana.

SOLUCIÓN:

- La media aritmética:

$$a = \frac{1}{n}\sum_{i=1}^{k} x_i * n_i = \frac{27678,5}{895} = 30,93$$

- La mediana:

Para ello nos valemos del cálculo de frecuencias absolutas acumuladas:
Este tipo de datos nos dice el número de datos que hay igual o inferiores a uno determinado.

Se calcula con la siguiente formula: $N_i = \sum_{j=1}^{i} n_j = N_{i=1} + n_j$

Así de esta forma:

Tamaño tabla	Frecuencia acumulada
3-6	37
6-11	235

11-16	426
16-21	575
21-26	654
26-31	700
31-41	755
41-51	806
51-76	832
76-101	857
101-201	882
201-501	893
501-1000	895

Respecto a esta tabla calculamos la mediana:

$$N_3 = 426 < \frac{n}{2} = \frac{895}{2} = 447,5 < 575 = N_4$$

Con lo que podemos decir que la mediana esta en el intervalo [16,21)siendo la mediana el valor:

$$M_e = x_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} * c_j = 16 + \frac{447,5 - 426}{149} * 5 = 16,72$$

24. Con los datos del ejercicio anterior calcular el primer cuartel y el sexto decil.

SOLUCIÓN:

a) El primer cuartil:

$$37 < \frac{1}{4} * n = 223,75 < 235$$

Será $p_{1/4} \in [6,11)$ y, en concreto:

$$p_{1/4} = x_{j-1} + \frac{\frac{1}{4} * n - N_{j-1}}{n_j} * c_j = 6 + \frac{\frac{1}{4} * 895 - 37}{198} * 5 = 10,716$$

Que será igual al centil 25

b) Calcular el sexto decil

Para ello acotamos el valor:

$$\frac{6}{10} * n = \frac{60}{100} * n = \frac{6}{10} * 895 = 537$$

Por las frecuencias absolutas acumuladas:

$$N_3 = 426 < 537 < 575 = N_4$$

$$p_{6/10} = x_{j-1} + \frac{\frac{6}{10} * n - N_{j-1}}{n_j} * c_j = 16 + \frac{537 - 426}{149} * 5 = 19,72$$

25. Mediante los datos del ejercicio numero 23 calcular las medidas de dispersión

SOLUCIÓN:

- El recorrido:

$$R = x_{\max} - x_{\min} = 750{,}5 - 4{,}5 = 746$$

- La varianza es:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - a)^2 * n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - a^2 = \frac{3713428{,}25}{895} - 30{,}93^2 = 3192{,}47$$

- La cuasivarianza

$$S^2 = \frac{n * s^2}{n-1} = \frac{895 * 3192}{894} = 3195{,}988$$

- La desviación típica:

$$s = \sqrt{s^2} = \sqrt{3192{,}417} = 56{,}50$$

- La cuasidesviación típica:

$$S = \sqrt{S^2} = \sqrt{3195{,}988} = 56{,}53$$

- El coeficiente de variación de Pearson:

$$V_p = \frac{s}{a} * 100 = \frac{56{,}50}{30{,}93} * 100 = 182{,}67$$

- El coeficiente de asimetría de Pearson:

$$A_p = \frac{a - Md}{s} = \frac{30{,}93 - 9{,}78}{56{,}50} = 0{,}374$$

26. En un colegio de un pequeño pueblo de la comunidad valenciana se han recogido los siguientes datos de información sobre cuantos niños se matriculan de cada sexo cada año, según se muestra en la siguiente tabla:

Año	**Niños**	**Niñas**
1995	**32**	**43**
1996	**27**	**24**
1997	**29**	**32**
1998	**29**	**31**
1999	**31**	**31**

Se debe calcular la frecuencia absoluta y la frecuencia relativa para los datos correspondientes a los niños y a las niñas y disponer los datos mediante un diagrama de sectores o de pastel en cada caso.

SOLUCIÓN:

1- Para empezar calcularemos las frecuencias correspondientes a los niños para esto utilizaremos la siguiente formula:

- En el caso de la frecuencia absoluta contaremos el número de veces que se repite un determinado valor, por lo que para x_1=0 n_1=32; x_2= 0 n_2=27; x_3=1 n_3=29; x_4=0 n_4=31;

- En el caso de las frecuencias relativas para hallar este valor dividiremos la frecuencia absoluta por el número total de datos. Así de esta forma la formula utilizada es: f_i=n_i/n

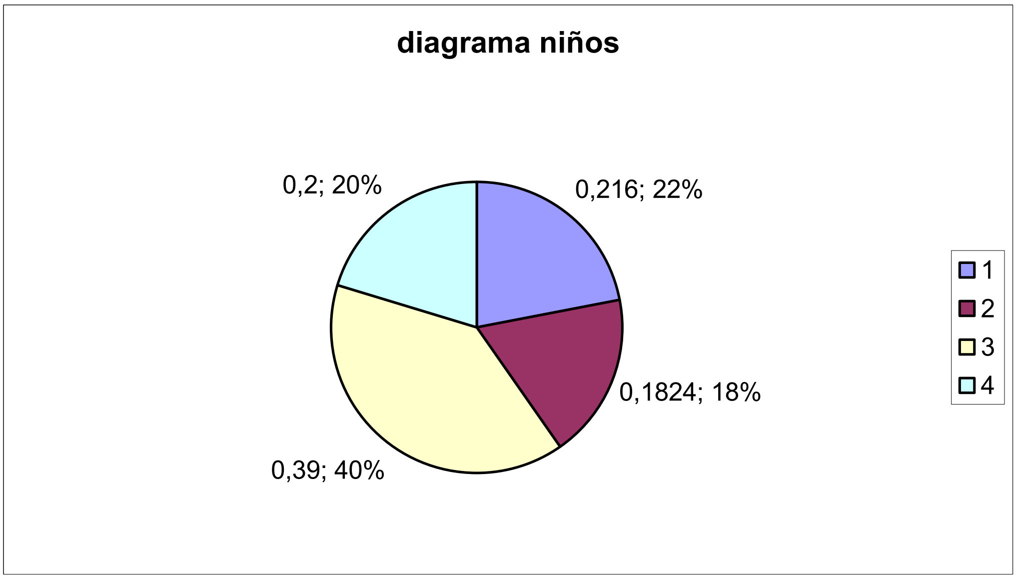
Por lo que obtenemos al calcularlo: $f_{1=}$ 32/148= 0,216; f_2= 27/148= 0,1824; f_3= 58/148= 0,39; f_4= 31/148= 0,20;

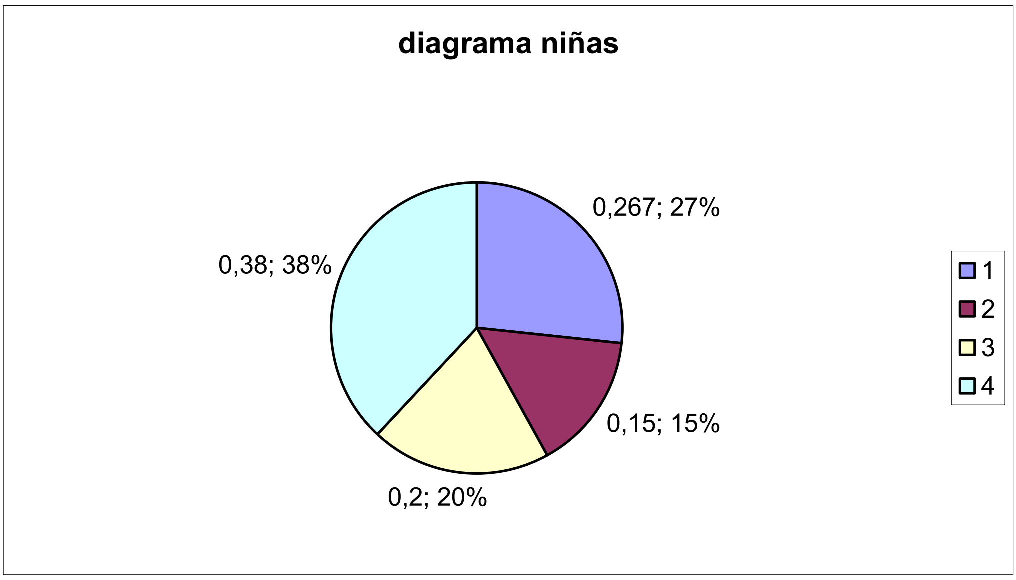
2- de la misma forma calcularemos las frecuencias correspondientes para las niñas:

- Para las frecuencias absolutas: x_1=0 n_1=43; x_2= 0 n_2=24; x_3=0 n_3=32; x_4=1 n_4=31;
- Para las frecuencias relativas: $f_{1=}$ 43/161= 0,267; f_2= 24/161= 0,15; f_3= 32/161= 0,20; f_4= 62/161= 0,38;

3- Pasemos a la representación grafica, para ello debemos partir de las frecuencias relativas y calcular cada porción del diagrama mediante esta formula:

$\frac{1}{f_i} = \frac{360^\circ}{\alpha}$ así de esta forma obtenemos:





27. Establecido un balance de explotación sobre las ocho sucursales de una cadena de almacenes, resultó la siguiente estimación:

Sucursal	Beneficios sobre ventas	Ventas totales
	28	500
	15	126
	24	432
	32	870
	17	180
	23	450
	18	912
	25	713

Obténgase el porcentaje medio de beneficios sobre las ventas totales de la cadena.

SOLUCIÓN:

El porcentaje medio se obtiene como media aritmética ponderada de los beneficios, siendo el peso respectivo la venta total de cada sucursal, es decir,

d_i	p_i	j_i-p_i
28	500	**14000**
15	126	**1890**
24	432	**10368**
32	870	**27840**
17	180	**3060**
23	450	**10350**
18	912	**16416**
	713	**17825**
	4183	**101479**

De donde:

$$x = \frac{101479}{4183} = 24{,}32\%$$

---

28. Una prestigiosa frutería tiene como norma clasificar los mangos según su tamaño, de cara a la venta, en superiores y normales. Los superiores son aquellos cuyo peso es superior a 450 g. De una partida, representativa de los mangos que recibe normalmente, se ha obtenido la distribución de frecuencias siguientes:

Peso	Núm. De mangos
250-300	3
300-350	10
350-400	15
400-450	25

450-500	32
500-550	20
550-600	19
600-650	4
650-700	2

a- Un exquisito aristócrata ha acordado con el frutero quedarse con los mangos cuyo peso sea superior a 625 gramos. ¿Qué porcentaje de mangos se destinarán a este aristócrata?

b- El frutero compra la partida de mangos a 300 pts el kg. Los normales se venden a 600 pts/kg, os superiores a 800 pts /kg, mientras que el aristócrata se os deja a 700 pts/kg. ¿Cuánto espera ganar este frutero es esta partida?

SOLUCIÓN:

a- Primero habrá que calcular el número de mangos cuyo peso es supereior a 625 gamos. Bajo la hipótesis de distribución uniforme de la frecuencia en los intervalos, resulta que en el intervalo (625-800] hay 5 mangos de la partida. Por tanto se apartarán para el aristócrata el 3.85% de los mangos recibidos.

b- Ya que el frutero decide retirar de la venta aquellos cuyo peso sea a lo sumo de 317.5 g, la distribución del peso de los mangos normales, vaqriables notada por X_N, será

Peso	(317.5-350]	(350-400]	(400-450]
Nº de mangos	6	15	25

La distribución del peso de los mangos superiores, excluídos los destinados al aristócrata, será la de la variable notada por X_S según esta tabla

Peso	(450-500]	(500-550]	(550-600]	(600-625]
Nº	32	20	19	1

La del aristócrata es X_A

Peso	(625-700]	(700-800]
Nº	3	2

El peso medio de los mangos normales seá la media de la distribución de frecuencias de la variable X_N, tomando como valores de las variable las marcas de clase de los intervalos. RTesulta se la variable las marcas de clase de los intervalos. Resulta ser la variable las marcar de clase de los intervalos. Resulta ser

$$\overline{X}_N = \frac{333.75 * 6 + 375 * 15 + 4 * 25}{46} = 396.79 \text{ g.}$$

De manera análoga se obtiene que

$$\overline{X}_S == 517.19 \text{ g. y } \overline{X}_A = 697.5 \text{ g.}$$

Por lo tanto en esta partida el frutero espera tener 46*396.79 = 18.25 kg de mangos normales, 72*517.19 =37.24 kg de mangos superiores y 5*697.5=3.49kg de mangos destinados al aristócrata. Con lo cual espera ganar por esta partida la cantidad de

300*18*253+500*37.24+400*3.49=25490 pts.

---

29. En una ciudad, analizamos el nivel de vida a través de la renta anual familiar. Se recoge información sobre 50 familias. Los datos en millones de pesetas, son los siguientes:

3’2	**1’1**	**3’3**	**0’2**	**2**
1’3	**0’8**	**0’4**	**3’8**	**2’6**
2’3	**3’4**	**2’8**	**1’7**	**1’2**
3’2	**3’2**	**2’6**	**1’1**	**2’4**
2’6	**1’6**	**0’9**	**2**	**1’8**
3’6	**1’3**	**2’7**	**2’3**	**2’3**
1’7	**2’9**	**1’2**	**2’2**	**2**
1’3	**1’8**	**0’8**	**2’3**	**1’4**
0’9	**1’1**	**2’1**	**1’7**	**1’2**
2’3	**1’6**	**2’2**	**1’7**	**2’1**

Obtener medidas que indiquen la localización, la dispersión, la asimetría y la curtosis. Repetir el problema agrupando los datos en intervalos de amplitud 0’5 y posteriormente en intervalos de amplitud 1. Comprobar si existen grandes diferencias.

SOLUCIÓN:

En este problema deseamos comprobar si, al agrupar los datos en intervalos, la información original aportada por los datos en cierta forma se conserva, o por el contrario, hay diferencias relevantes.

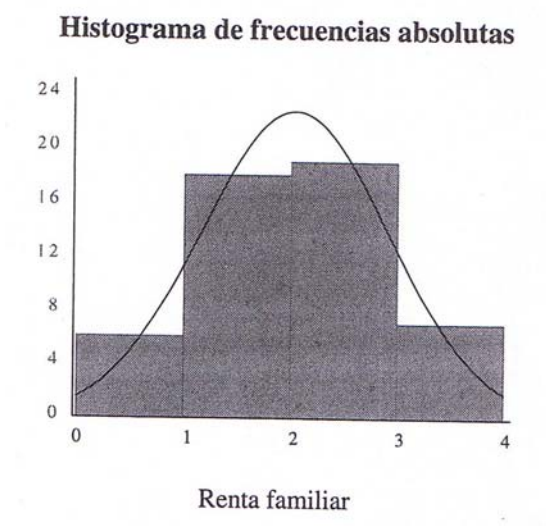
Teniendo en cuenta los resultados concretos procedentes de diferentes familias, recogemos esta información:

Nº de datos: 50
Mínimo: 0’2 millones
Máximo: 3’8 millones
Media: 1’964
Moda: 2’3
Varianza: 0’7095
Desv. Típica: 0’8423
Primer Cuartel: 1’3
Mediana: 2
Tercer Cuartel: 2,6
Asimetría: 0’1697
Curtosis: -0’5984
Coef. de Pearson: 0’4289

En este primer análisis, las rentas son valores que oscilan entre 200.000 ptas. y 3’8 millones; la renta media familiar es de 1.964.000 ptas.; es una distribución que tiende a ser simétrica (el coeficiente de asimetría es igual a 0’1697) y el coeficiente de curtosis es negativo, que indica que la distribución está por debajo de la distribución normal tipificada, es decir, es platicúrtica.

Agrupemos los datos en intervalos de amplitud 0’5; como la renta toma valores positivos y no superan el valor 4, podemos considerar rango 0-4.

L_{i-1}-L_i	x_i	n_i	N_i
0’0-0’5	0’25	2	2
0’5-1’0	0’75	4	6
1’0-1’5	1’25	10	16
1’5-2’0	1’75	8	24
2’0-2’5	2’25	13	37
2’5-3’0	2’75	6	43
3’0-3’5	3’25	5	48
3’5-4’0	3’75	2	50



Como esta información se puede calcular las siguientes medidas:

Media: 1’99
Varianza: 0’7324
Desv. Típica: 0’8558
Moda: 2’2143
Mediana: 2’0385
Asimetría: 0’046
Curtosis: -0’5888
Índice de Pearson: 0’4301

No hay mucha variación respecto del caso anterior; lo más significativo es que la renta media ahora es 1990.000; si nos guiamos por este procedimiento, comparando este valor medio con el primero, estaremos sustrayendo a cada familia unas 26.000 ptas.

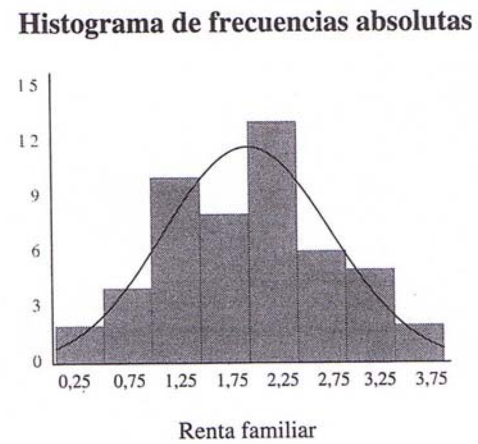
Los intervalos con mayor frecuencia están situados en el centro; agrupar los datos en un sentido u otro hace que el coeficiente de asimetría cambie (el nuevo valor es 0’046), aunque en todos los casos toma valores cercanos a cero.

A pesar de que el intervalo con marca 2’25 es el de mayor frecuencia (por encima de la gráfica de la distribución normal), los intervalos adyacentes reflejan lo contrario. Debido a esta situación, el coeficiente de curtosis es negativo.

Si consideramos intervalos con mayor amplitud:

L_{i-1}-L_i	x_i	n_i	N_i
0-1	0’5	6	6
1-2	1’5	18	24
2-3	2’5	19	43
3-4	3’5	7	50

Media: 2’04
Varianza: 0’7684
Desviación Típica: 0’8766
Moda: 2’28
Mediana: 2’0526
Asimetría: -0’0331
Curtosis: - 0’6989
Coef. Pearson: 0’4297



De acuerdo con estos resultados, la renta media aumenta en 41000 ptas. de los datos originales, siendo el resto de valores muy similares. Observar que ahora el coeficiente de asimetría es negativo, pero en todos los casos muy próximo a cero, con lo cual la distribución se puede considerar prácticamente simétrica.

30. ¿Se producen alteraciones en las *medidas de posición* al realizar un cambio de origen?

SOLUCIÓN:

El cambio de origen supone una traslación del tipo y = x + a. Las medidas de posición son afectadas de la siguiente forma:

Media

$$\overline{Y} = \frac{\sum_{i=1}^{n} y_i \cdot n_i}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} (x_i + a) \cdot n_i}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{\sum_{i=1}^{n} n_i} + \frac{a \cdot \sum_{i=1}^{n} n_i}{\sum_{i=1}^{n} n_i} = \overline{x} + a$$

Moda

$N_k = \text{máx.}(n_1, n_2, \_\_\_, n_n), \Rightarrow M_o(x) = x_k$;por consiguiente, la moda de la variación del valor correspondiente a la frecuencia n_k:

$M_o(Y) = y_k = x_k + a = M_o(x) + a$

**Si el valor i-ésimo intervalo es el de mayor altura, tanto la moda de X como la de Y estarán situadas en él:

$$M_o(Y) = L_{i-1} + a + c_i \cdot \frac{h_{i+1}}{h_{i-1} + h_{i+1}} = M_o(x) + a$$

Mediana

*Si los datos están sin agrupar en intervalos, por ser frecuencias de X y de Y las mismas, la mediana de Y será la mediana de X trasladada en “a” unidades.

**Si la distribución esta agrupada en intervalos de clase $N / 2 < N_i \Rightarrow$ la mediana se encuentra en el i-ésimo intervalo tanto para X como para Y.

$$Me(Y) = L_{i-1} + a + c_i \cdot \frac{\frac{N}{2} - N_{i-1}}{n_i} = Me(x) + a$$

Cuartiles, deciles y percentiles

Podemos observar que en estas medidas se produce el mismo cambio que en la variable. En el caso de los percentiles:

$P_j(Y) = P_j(x) + a$ con j=1,___,99.

---

31. Para lanzar un nuevo producto al mercado, una empresa estudia el tiempo de publicidad, en segundos, empleando en los medios audiovisuales por otra empresa que produce un producto similar.

Duración	Nº de Anuncios
0-20	3
20-25	17
25-30	13
30-40	9
40-60	8

¿Cuál es la duración media aproximada de los anuncios?¿Es representativa?
¿Cuál es la duración más frecuente?
¿A partir de que valor un anuncio es de los veinte más largos?

Estudiad la forma de la distribución.
Si cada segundo cuesta mil cuatrocientas pesetas, ¿cuál es el gasto aproximado que realiza la otra empresa en la publicidad de ese producto?

SOLUCIÓN:

a) $\overline{X}$ = 29'7 segundos. V_X = 0'358170667 moderadamente representativa.
b) Mo = 24'7273 segundos.
c) P_{60} = 28'8461538 segundos.
d) A_P = 0'4675 > 0. A_P = 0'7831 > 0. C = 0'3234 > 0.
g_1 = 0'06454 > 0.
La distribución presenta asimetría positiva o por la derecha.

g2 = -0'08595 > 0. La distribución es moderada platicúrtica.

El gasto aproximado será de 2079000 ptas.

32. La distribución del importe de las facturas por reparación de carrocería (en miles de ptas.) de una muestra de 80 vehículos en un taller, viene dad por la siguiente tabla:

Importe	Nº de vehículos
0-60	10
60-80	20
80-120	40
120-180	10

a) Calcular el importe medio. Estudiar la representatividad en esta medida.
b) Calcular la mediana y estudiar su representatividad.
c) ¿Cuál es el importe más habitual?
d) ¿Qué interpretación tiene en este caso los deciles? Calcular el tercer decil.
e) ¿Cuál es el importe mínimo pagado por las 75 reparaciones más baratas.
f) Estudiar la concentración del importe de las facturas.

SOLUCIÓN:

a) $\overline{X}$ = 90000 ptas. V_X = 0'36. Es moderadamente representativa.
b) Me = 90000 ptas. D_{Me} = 25. $V_{Me} = 0.2\hat{7}$. Es representativa.
c) Hay dos modas: Mo_1 = 77143 ptas. y Mo_2 = 85714 ptas.
d) D_3 = 74000 ptas.
e) P_{75} = 110000 ptas.
f) $I_G = 0'\hat{2}$.

33. Dos compañías aseguradoras tienen formas diferentes de pagar a sus empleados. La compañía A lo hace mediante un sueldo fijo mensual y la compañía B a través de un porcentaje sobre los seguros realizados. La distribución de los salarios por categorías es:

Compañía A		Compañía B	
Sueldo (miles ptas.)	Nº empleados	Sueldo (miles ptas.)	Nº empleados
50-80	35	50-80	21
80-100	21	80-100	25
100-150	14	100-140	34
		140-200	15

a) Por término medio,¿gana más un empleado de la compañía A o de la B?
b) Calcular y comentar la representatividad de los sueldos medios.
c) ¿Cuál es el sueldo más frecuente en la compañía A?
d) Aunque en la compañía B el sueldo se gana por méritos, ¿crees que el reparto de salarios por categorías es equitativo?
e) Si en la compañía B el salario fuese el anterior más un fijo de 10000 pesetas, ¿cuál sería el salario medio y la desviación típica?

SOLUCIÓN:

a) Sean:

X = «sueldo (en miles de pesetas) de los empleados de la compañía A».
Y = «sueldo (en miles de pesetas) de los empleados de la compañía B»

$\overline{X} = 84.5$

$\overline{Y} = 107.842105$

b) $V_X = 0.27273876$, $V_Y = 0.31479111$, los sueldos están menos dispersos en la empresa A.

c) Mo = 80000 pesetas.

d) $I_G(Y)$ = ‘.200456171

e) Z = Y + 10

$\overline{Z} = \overline{Y} + 10 = 107.842105 + 10 = 117.842105$

$S_Z = S_Y$

---

34. Las notas finales de 100 estudiantes de una Escuela Superior son las siguientes:

11	**46**	**58**	**25**	**48**	**18**	**41**	**35**	**59**	**28**
35	**2**	**37**	**68**	**70**	**31**	**44**	**84**	**64**	**82**
26	**42**	**51**	**29**	**59**	**92**	**56**	**5**	**52**	**8**
1	**12**	**21**	**6**	**32**	**15**	**67**	**47**	**61**	**47**
43	**33**	**48**	**47**	**43**	**69**	**49**	**21**	**9**	**15**
11	**22**	**29**	**14**	**31**	**46**	**19**	**49**	**51**	**71**
52	**32**	**51**	**44**	**57**	**60**	**43**	**65**	**73**	**62.**
3	**17**	**39**	**22**	**40**	**65**	**30**	**31**	**16**	**80**
41	**59**	**60**	**41**	**51**	**10**	**63**	**41**	**74**	**81**
20	**36**	**59**	**38**	**40**	**43**	**18**	**60**	**71**	**44**

Determinar:

1- El número de estudiantes con nota superior a 80

2- La nota del estudiante nº 38 en orden a la peor puntuación de la distribución del tipo III:

SOLUCIÓN:

Li-1 ─ li						ni	Ni
0-10	~~||||~~	|||				8	8
10-20	~~||||~~	~~||||~~	||			12	20
20-30	~~||||~~	~~||||~~				10	30
30-40	~~||||~~	~~||||~~	||||			14	44
40-50	~~||||~~	~~||||~~	~~||||~~	~~||||~~	|	21	65

50-60	卌 卌 卌 |	16	81
60-70	卌 卌	10	91
70-80	卌	5	96
80-90	|||	3	99
90-100	|	1	100
		100	

Podemos decir que:

1º Número de estudiantes con nota superior a 50 e inferior a 80:

$$N_6 + N_7 + N_8 = 16 + 10 + 5 = 31$$

2º Nota del estudiante número 38: De 30 a 40 puntos

---

35. Dada la siguiente distribución de frecuencias:

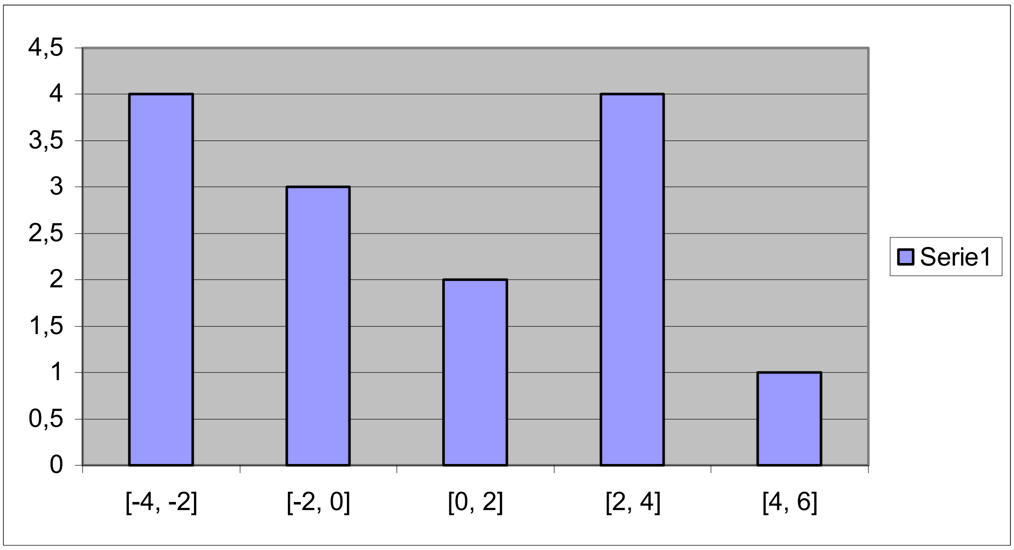
Li-1 — Li	**ni**
-4 — -2	**4**
-2 — 0	**3**
0 — 2	**2**
2 — 4	**4**
4 — 6	**1**

1º Representarla gráficamente
2º Obtener la serie de frecuencias acumuladas
3º Representar la distribución de frecuencias acumuladas

SOLUCIÓN:

1º Representar gráficamente el histograma:

Por tratarse de una distribución con intervalos de igual amplitud, podemos tomar la ni, como altura, obteniéndose:

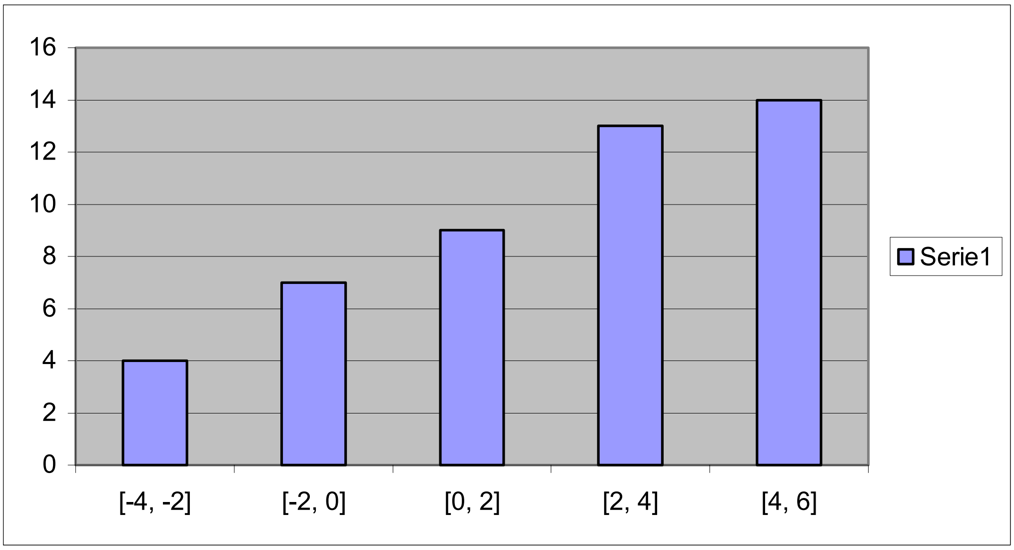


2º Serie de frecuencias acumuladas Ni:

Li-1 — Li	ni	Ni
-4 — -2	4	4

-2 — 0	3	7
0 — 2	2	9
2 — 4	4	13
4 — 6	1	14

3º Representación gráfica de la distribución de frecuencias acumuladas:



Donde se han tomado hi=Ni por tratarse de una distribución de intervalos de igual amplitud.

---

36. Hallar la mediana de la siguiente distribución de frecuencias:

Li-1— Li	**ni**
0 — 1	**12**
1 — 2	**13**
2 — 3	**11**
3 — 4	**8**
4 — 5	**6**

SOLUCIÓN:

Li-1— Li	ni	Ni
0 — 1	12	12
1 — 2	13	25
2 — 3	11	36
3 — 4	8	44
4 — 5	6	50
	50	

$$\frac{N}{2} = \frac{50}{2} = 25$$

---

37. Se ha observado la vida de 280 bombillas obteniéndose la siguiente distribución:

Vida en horas	**Nº de bombillas**
0 ─ 500	**4**
500 ─ 1000	**21**
1000 ─ 1500	**107**
1500 ─ 2000	**78**
2000 ─ 2500	**44**
2500 ─3000	**24**
	280

Hallar la moda.

SOLUCIÓN:

Se trata de una distribución del tipo III con intervalos constantes.

$$Mo = Li-1+\frac{ni+1}{ni-1+ni+1}a1 = 1000+\frac{78}{21+78}\times 500 = 1000+394 = 1394$$

38. En una clínica se han registrado durante un mes las longitudes en metros que los niños andan el primer dia que comienzan a caminar, obteniéndose los siguientes resultados.

Número de metros	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Número de niños	**2**	**6**	**10**	**5**	**10**	**3**	**2**	**2**

Construir la distribución de frecuencias adecuada para la variable longitud y realizar los gráficos pertinentes que la representen.

SOLUCIÓN:

Dado que se trata de una variable cuantitativa con valores sin agrupar, podemos comenzar realizando su representación mediante un diagrama de barras situado sobre el eje de abscisas los valores de la variable X, y sobre el eje de ordenadas los valores de sus frecuencias absolutas n_i. Asimismo, si sobre el eje de ordenadas situamos las frecuencias absolutas acumuladas N_i, obtenemos el diagrama de barras acumuladas. La tabla de frecuencias relativas a la variable se presenta a continuación.

La figura 1 muestra el diagrama de barras asociado a la variable y a la figura 2 muestra el diagrama de barras acumulado.

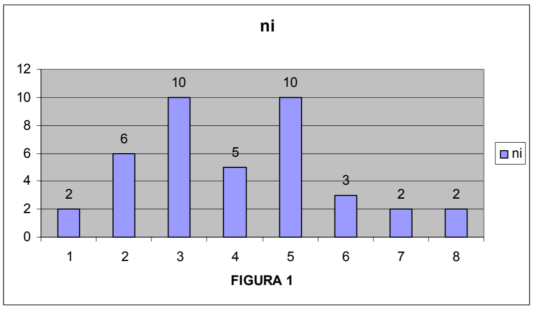
Distribución de frecuencias.

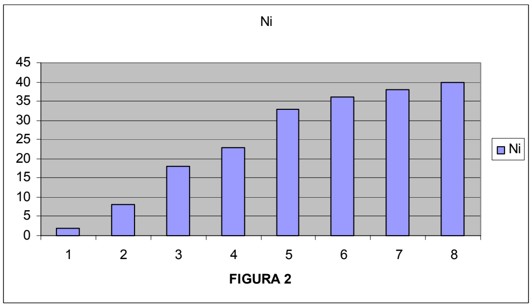
X_i	n_i	N_i	f_i	F_i
1	2	2	0,05	0,05
2	6	8	0,15	0,2
3	10	18	0,25	0,45
4	5	23	0,125	0,575
5	10	33	0,25	0,825
6	3	36	0,075	0,9
7	2	38	0,05	0,95
8	2	40	0,05	1

$f_k=n_k/N \quad \rightarrow \quad N=n_1+\ldots+n_k=N_k$

$F_k=N_k/N$

Graficos.





---

39. La distribución de los costes salariales de los 100 000 empleados de una multinacional se presenta en la tabla siguiente:

Salarios	**Nº de empleados**
0-15000	**2145**
15000-20000	**1520**
20000-25000	**840**
25000-30000	**955**
30000-35000	**1110**
35000-40000	**2342**
40000-50000	**610**
50000-100000	**328**
100000-300000	**150**

Calcular el salario medio por trabajador, el salario más frecuente y el salario tal que la mitad de los restantes sea inferior a él. Calcular también el primer cuartel salarial y el percentil 75.

SOLUCIÓN:

La tabla siguiente contiene los elementos relativos a la distribución de frecuencia de la variable salario (X) necesarios para realizar los cálculos pedidos en el problema.

Salarios	Nº de empleados= n_i	Marcas = x_i	$x_{i*}\, n_i$	N_i	c_i	$d_i = n_{i/}\, c_i$
0-15000	2145	7500	16087500	2145	15000	0,143
15000-20000	1520	17500	26600000	3665	5000	0,304
20000-25000	40	22500	18900000	4505	5000	0,168
25000-30000	55	27500	26262500	5460	5000	0,191
30000-35000	110	32500	36075000	6570	5000	0,222
35000-40000	342	37500	87825000	8912	5000	0,4684
40000-50000	10	45000	27450000	9522	10000	0,061
50000-100000	28	75000	2460000	9850	50000	0,0056
100000-300000	50	200000	30000000	10000	200000	0,00075
	0000		293800000			

Para hallar el salario medio por trabajador calculamos la media de la variable X.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = \frac{293800000}{10000} = 29380$$

Para hallar el salario más frecuente se calcula la moda de la variable X. Para ello hemos de tener presente que los intervalos de la distribución de frecuencias son desiguales, por lo que el intervalo modal será el correspondiente al mayor valor de d_i, es decir será el intervalo (35000 – 40000). Por lo tanto la moda se calcula como sigue:

$$M_0 = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i = 35000 + \frac{0,061}{0,222 + 0,061} 5000 = 36077,74$$

Para hallar el salario tal que la mitad de los restantes sea inferior a él se calcula la mediana. Para ello, como N/2 = 5000, el intervalo mediano será (25000 – 3000) ya que $N_{i-1} < N/2 < N_i$ es equivalente en este problema a 4505 < 5000 <5460. La mediana se calculará como sigue:

$$M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i = 25000 + \frac{\frac{10000}{2} - 2145}{955} 5000 = 27591,62$$

Para calcular el primer cuartel (primer cuantil de orden 4) observamos que como N/4 = 2500, el intervalo relativo al primer cuartel será (15000-20000) ya que $N_{i-1}<2500<N_i$ es equivalente en este problema a 2145<2500<3665.El primer cuartel se calculará como sigue:

$$Q_{1,4} = L_{i-1} + \frac{\frac{N}{4} - N_{i-1}}{n_i} c_i = 15000 + \frac{\frac{10000}{4} - 2145}{1520} 5000 = 16167,76$$

El primer cuartel se interpreta como el valor de la variable para el que la cuarta parte de los valores son menores que él y las tres cuartas partes restantes son superiores.

Para calcular el percentil 75 (cuantil 75 de orden 100), observamos que como 75N/100 = 7500, el intervalo relativo al percentil 75 será (35000-40000) ya que $N_{i-1}<7500<N_i$ es equivalente en este problema a 6570<7500<8190. El percentil 75 se calculará como sigue:

$$_{5,100} = L_{i-1} + \frac{\frac{75N}{100} - N_{i-1}}{n_i} c_i = 35000 + \frac{75*10000/100 - 6570}{2342} 5000 = 36985,48$$

El percentil 75 se interpreta como el valor de la variable para el que el 75% de los valores son inferiores a él y el 25% restante son superiores.

El percentil 75 también podrá haberse calculado como el tercer cuartil (cuantil 3de orden 4). Como 75N/100=7500=3N/4, el tercer cuartil se calcularía como sigue:

$$Q_{3,4} = L_{i-1} + \frac{\frac{3N}{4} - N_{i-1}}{n_i} c_i = 35000 + \frac{\frac{3*10000}{4} - 6570}{2342} 5000 = 36985,48$$

El tercer cuartil se interpreta como el valor de la variable para el que las tres cuartas partes de los valores son inferiores a él y la cuarta parte restante es superior. Como las tres cuartas partes son el 75%, el percentil 75 coincide con el tercer cuartil.

---

40. Los rendimientos de cinco inversiones distintas realizadas por un individuo y las cantidades iniciales invertidas en unidades monetarias son los siguientes:

Cantidades iniciales	**Rendimientos**
200 000	**1000**
360 000	**900**
250 000	**500**
240 000	**800**
180 000	**1200**

Calcular el rendimiento medio por unidad monetaria invertida para el total de inversiones del individuo.

SOLUCIÓN:

Como se trata de promediar rendimientos por unidad, estamos ante un caso de aplicación del concepto de media armónica. Calcularemos por tanto el rendimiento medio pos unidad monetaria para el total de inversiones del individuo como la media armónica de los rendimientos de cada inversión ponderada por las cantidades iniciales desembolsadas en cada inversión.

$$H = \frac{N}{\sum_{i=1}^{k} \frac{1}{x_i}} = \frac{200000 + 360000 + 250000 + 240000 + 180000}{\frac{200000}{1000} + \frac{360000}{900} + \frac{250000}{500} + \frac{240000}{800} + \frac{180000}{1200}} = 793,5$$

Aunque en este problema es menos adecuada, podríamos haber utilizado también la media aritmética ponderada, que se calcula como sigue:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i = \frac{200000*100 + 360000*900 + 250000*500 + 240000*800 + 180000*120}{200000 + 360000 + 250000 + 240000 + 180000}$$

$$\overline{X} = 859,35$$

También podría utilizarse la media geométrica, ya que las cantidades a promediar son no nulas y positivas. Para hallar esta media es conveniente aplicar logaritmos (en este caso neperianos) y calcular el valor final como se indica a continuación:

$$G = \sqrt[N]{x_1^{n1} x_2^{n2} \cdots x_k^{nk}} \Rightarrow \ln(G) = \frac{1}{N}\sum_{i=1}^{k} n_i \ln(x_i) =$$

$$= \frac{200000 * \ln(100) + 360000 * \ln(900) + 250000 * \ln(500) + 240000 * \ln(800) + 180000 * \ln(120)}{200000 + 360000 + 250000 + 240000 + 180000}$$

$=6´72 \Rightarrow G=e^{6,72}=828,82$

---

41. En el cuadro siguiente se presentan los consumos de electricidad en España en miles de millones de de kw/hora desde diciembre en 1985 hasta diciembre de 1986.

Meses	Consumo
Dic	10.1
Ene	10.7
Feb	9.96
Mar	9.46
Abr	9.54
May	8.92
Jun	8.95
Jul	8.58
Ago	7.86
Sep	8.96
Oct	9.17
Nov	9.57
Dic	10.2

A partit de los incrementos unitarios de consumo de cada mes calcular el incremento unitario anual medio acumulativo.

<u>SOLUCIÓN:</u>

Al tratarse de cálculo de una media unitaria acumulativa, el promedio más adecuado es la media geométrica. Se trata por tanto de calcular la media geométrica de los incrementos unitarios mensuales. Estos incrementos se calculan a continuación.

$$\frac{10,7}{10,1} = 1,06 \quad \frac{9,96}{10,7} = 0,93 \quad \frac{9,46}{9,96} = 0.95 \quad \frac{9,54}{9,45} = 1,008 \quad \frac{8,92}{9,54} = 0,93 \quad \frac{8,95}{8,92} = 1,003$$

$$\frac{9,58}{8,95} = 1,07 \quad \frac{7,86}{9,58} = 0,82 \quad \frac{8,96}{7,86} = 1,14 \quad \frac{9,17}{8,96} = 1,02 \quad \frac{9,57}{9,17} = 1,04 \quad \frac{10,2}{9,57} = 1,06$$

La media geométrica de estos incrementos unitarios mensuales se calcula como sigue:

$$G = \sqrt[12]{1,06 * 0,93 * 0,95 * 1,008 *,0,93 * 1,003 * 1,07 * 0,82 * 1,14 * 1,02 * 1,04 * 1.06} = 1,01$$

---

42. Supongamos que un automóvil recorre 60 km a una velocidad de 50 km/h y 40km/h a una velocidad de 70 km/h ¿Cuál será la velocidad media del automóvil en todo el recorrido?

SOLUCIÓN:

Al tratarse de cálculo de una velocidad media utilizaremos la media armónica, que se cálcula como se indica a continuación:

$$H = \frac{N}{\sum_{I=1}^{k} \frac{1}{x_1} n_1} = \frac{60+40}{\frac{60}{50}+\frac{40}{70}} = 56{,}45 \text{ km/h}$$

En este caso, cualquier otro promedio que se utilice no produce la velocidad media.

---

43. En una distribución discreta de de 6 valores, a saber: -10, 3, a, 10, 1, 0, sabemos que su desviación típica es igual al coeficiente de variación de Pearson. Se pide:

a) hallar la media de la distribución
b) hallar el valor desconocido de a

SOLUCIÓN:

a) como CV=s entonces se tiene que la media aritmética vale 1, puesto que el coeficiente de variación de Pearson es el coeficiente entre la desviación típica y la media aritmética.
b) Aplicando el resultado obtenido en el apartado anterior se tiene que:

1= (-10+3+a+10+1+0) / 6

Despejando a de la expresión se tiene que a=2

---

44. Un examen consta de 5 preguntas en las que dos alumnos A y B obtienen las siguientes calificaciones según el orden de las preguntas:
A: 5,8,6,5,4 **B: 3,7,8,6,3**

a) **¿Cuál de los dos alumnos tuvo mejor nota sabiendo que los ejercicios 1, 3 y 4 puntúan la mitad que los ejercicios 2 y 5?**
b) **si consideramos que todas las preguntas valen igual, ¿Qué alumno obtendría mejor calificación si utilizamos la media geométrica?**

SOLUCIÓN:

a) se calcula la media ponderada con los pesos que se indican para cada uno de los alumnos siendo el alumno con mayor media el que tuvo mejor nota.

Los pesos para los problemas 1, 3 y 4 será 1 y para los problemas 2 y 5 será 2. Así obtendremos los siguientes resultados:

A= (5*1 + 8*2 + 6*1 + 5*1 + 4*2) / (1+2+1+1+2)=5.714

B= (3*1 + 7*2 + 8*1 + 6*1 + 3*2) / (1+2+1+1+2)=5.286

Por tanto fue el alumno A el que obtuvo mejor calificación.

b) Primero recordemos las expresiones de la media geométrica

$G = \sqrt[n]{ (\prod x_i^{fi}) }$

Si consideramos la media geométrica, el alumno A obtiene una calificación de 4.448 y el alumno B una de 4.967

---

45. Una empresa ha realizado un test físico entre sus empleados para comprobar la capacidad de esfuerzo que posee cada uno de ellos. Una de las medidas que componen el mismo es el número de pulsaciones después de una determinada actividad física, que esta altamente relacionada con las que se realizan a lo largo de una jornada laboral. Los datos conseguidos han sido distribuidos en una tabla de frecuencias. La tabla resultante es la que se presenta:

Numero de pulsaciones	**Numero de empleados**
70 – 75	3
75 – 80	3
80 – 85	7
85 – 90	10
90 – 95	12
95 – 100	8

Se pide:

a) **media aritmética, mediana, cuartil inferior, percentil 60 y desviación típica.**
b) **¿Qué tanto por cien de empleados tuvieron menos de 83 pulsaciones?**

SOLUCIÓN:

a) media= 88.198
Me= 89.25
Q_1= 83.393
P_{60}= 91.167
b) 23.721%

---

46. En el marco de un estudio sobre la posible incidencia que tiene la religión profesada por los distintos matrimonios en la presencia de una mayor ó menor frecuencia de divorcios, se ha tomado una muestra aleatoria a nivel mundial de tamaño 32000

Religión \ Divorcio	Divorcio	No divorcio
Católicos	1435	7565
Ateos	845	2155
Musulmanes	160	7840
Protestantes	610	4390
Otros	1250	5750

SOLUCIÓN:

a) **Basando tus razonamientos y afirmaciones en las frecuencias relativas que resulten mas informativas para este estudio señala cual es la religión donde los matrimonios presentan una mayor probabilidad de terminar en divorcio y cual es en la que se dan menos.**

En este caso las frecuencias que proporcionan más información son las frecuencias relativas

condicionales de Divorcio/Religión.
Frec. Relativa (Divorcio/ Católicos) = 1435/(1435+7565)=15.94% de los matrimonios católicos acaban en divorcio
Frec. Relativa (Divorcio/Ateos) = 845/(845+2155)=28.17% de los matrimonios ateos acaban en divorcio
Frec. Relativa (Divorcio/Musulmanes) =160/(160+7840)=2% de los matrimonios musulmanes acaban en divorcio
Frec. Relativa (Divorcio/Protestantes) = 610/(610+4390)= 12.2% de los matrimonios musulmanes acaban en divorcio.
Frec. Relativa (Divorcio/ Otros) = 1250/(1250+5750)=17.9% de los otros matrimonios acaban en divorcio

A partir de los datos se observa que en el caso de los ateos hay mas probabilidad de que los matrimonios acaben en divorcio 28.17%. En la religión musulmana ocurrirá justo lo contrario con solo un 2% de divorcios.

b) Obtener las frecuencias marginales absolutas y relativas de la variable divorcio.

Frecuencias marginales de la variable divorcio:
Frecuencias absolutas marginales: divorcio si: 4300 divorcio no: 27700
Frecuencias relativas marginales: divorcio si: 4300/32000=13.44%
divorcio no:27700/32000=86.56%

---

47. Para estudiar la eficacia de un tratamiento sobre las resistencias de un determinado hormigón se ha realizado un ensayo sobre 15 probetas. Se han medido los días transcurridos hasta que el hormigón alcance la resistencia de 40MPa y los datos han sido los siguientes:

15	**13**	**10**	**28**	**12**	**17**	**18**	**14**
15	**9**	**16**	**13**	**10**	**19**	**11**	

<u>SOLUCIÓN:</u>

a) Indicar la población, la variable aleatoria implicada y de que tipo es esta ultima.

La población es todo el hormigón de ese tipo. La variable aleatoria son el número de días transcurridos hasta alcanzar los 40MPa.Cuantitativa y discreta

b) Dibujar el diagrama Box-Whisker y comentar sus características principales

9 10 10 $\boxed{11}$ 12 13 13 $\boxed{14}$ 15 15 16 $\boxed{17}$ 18 19 28

Mediana=14
Cuartil 1=11
Cuartil 3=17
IIC=Intervalo intercuartilico=17-11=6
C3+1.5*IIC=17+1.5*6=26
C1-1.5*IIC=11-1.5*6=2
El bigote de la izquierda llegará, como mucho a 2.Como el valor observado posterior es 9, el bigote de la izquierda llegará a 9.
El bigote de la derecha llegará, como mucho a 26.Como el valor observado inmediatamente

anterior es 19,el bigote de la derecha llegara a 19
No existen puntos aislados por la izquierda.
Existe un valor observado superior a 26 (el valor 28), que se representará como un punto aislado en el diagrama Box-whisker

c) ¿Qué signo cabe esperar que tenga el coeficiente de curtosis de los datos?(Justifica la respuesta)

Se trata de una muestra con una distribución simétrica y 1 dato aislado, probablemente dato anómalo, pues dista mucho del resto de los datos, por lo tanto el coeficiente de curtosis será positivo.

---

48. La siguiente tabla muestra la cantidad de tierra de las regiones de un cierto país, junto con el porcentaje de tierra cultivada en cada region.

REGION	CANTIDAD DE TIERRA	% CULTIVADO
Norte	**421**	**46.7**
Sur	**350**	**21.0**
Este	**259**	**8.7**
Oeste	**80**	**18.8**

Calcula el porcentaje de tierra cultivada en la totalidad del pais

---

SOLUCIÓN:

$$\sum xi = 421 + 350 + 259 + 80 = 1110$$

La tierra cultivada será:

$$\frac{46.7 * 421 + 21 * 350 + 8.7 * 259 + 18.8 * 80}{100} = 307.68$$

Por lo tanto, el porcentaje de tierra cultivada será:

$$\% = \frac{307.68 * 100}{1100} = 27.97$$

---

49. Los alumnos de último curso de Bachillerato de un Instituto eligen carrera según los datos de la tabla siguiente:

CARRERA	*MEDICINA*	*DERECHO*	*CIENCIAS*	*LETRAS*	*INEF.*	*OTRAS*
ALUMNOS	250	176	127	314	103	30

Construir la distribución de frecuencias adecuada para la variable carrera elegida por los alumnos y realizar los gráficos pertinentes que la representen.

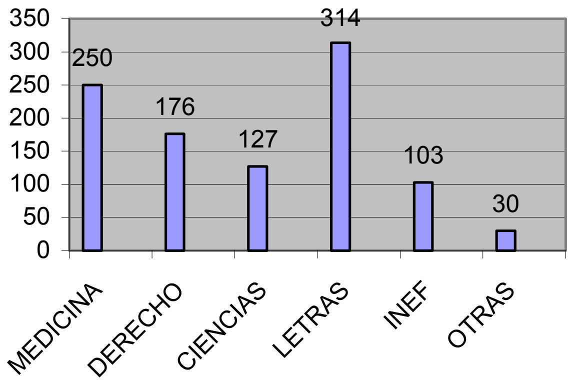
SOLUCIÓN:

Dado que se trata de una variable cualitativa, podemos comenzar realizando su representación mediante un diagrama de rectángulos, que se construye asignando a cada modalidad de la variable cualitativa un rectángulo con altura igual (o proporcional) a su frecuencia absoluta n_i y con base constante. La tabla de frecuencias relativa a la variable se presenta a continuación.

CARRERA	ALUMNOS (n_i)	$f_i = n_i / N$	$\square_i = 360f_i$
Medicina	250	0,25	90
Derecho	176	0,176	63,36
Ciencias	127	0,127	45,72
Letras	314	0,314	113,04
Inef	103	0,103	37,08
Otras	30	0,03	10,8
	N=1000	1	360

El diagrama de barras se presenta en la siguiente Figura.

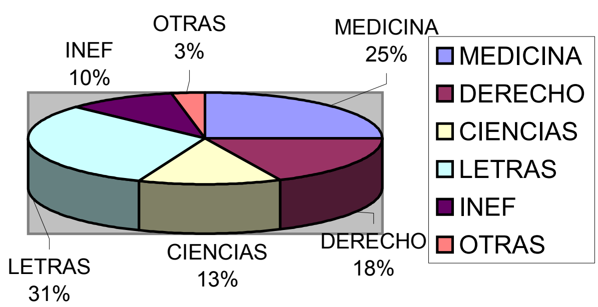
ESTUDIANTES POR CARRERAS



Figura

También podría realizarse la representación de la distribución de frecuencias de muestra variable cualitativa mediante el *diagrama de sectores con porcentajes* de la Figura 1-14. Los porcentajes relativos a cada carrera se calculan mediante $100f_i$ y los ángulos centrales de cada sector se calculan mediante $360f_i$.

ESTUDIANTES POR CARRERAS



50. Las puntuaciones obtenidas por 100 opositores en el último ejercicio se presentan en el cuadro siguiente:

7	**3**	**2**	**4**	**5**	**1**	**8**	**6**	**1**	**5**
3	**2**	**4**	**9**	**8**	**1**	**0**	**2**	**4**	**1**
2	**5**	**6**	**5**	**4**	**7**	**1**	**3**	**0**	**5**
8	**6**	**3**	**4**	**0**	**10**	**2**	**5**	**7**	**4**
0	**2**	**1**	**5**	**6**	**4**	**3**	**5**	**2**	**3**
9	**7**	**3**	**4**	**3**	**5**	**7**	**4**	**6**	**5**
6	**1**	**0**	**5**	**7**	**8**	**5**	**2**	**3**	**10**
4	**6**	**2**	**1**	**1**	**2**	**6**	**7**	**4**	**5**
4	**7**	**6**	**3**	**5**	**0**	**2**	**8**	**2**	**7**
8	**5**	**2**	**7**	**1**	**4**	**6**	**3**	**5**	**6**

1. **Construir la distribución de frecuencias adecuada para las puntuaciones.**
2. **Hallar el porcentaje de alumnos que aprobó la oposición.**
3. **Hallar el porcentaje de alumnos que sacaron notas superiores a 6.**
4. **Si sólo hay 20 plazas ¿En qué nota hay que situar el aprobado?**
5. **Realizar las representaciones gráficas de la distribución adecuadas para este problema.**

SOLUCIÓN:

Para construir la distribución de frecuencias de la variable aleatoria X que representa las distintas calificaciones, tabulamos los datos haciendo un recuento de los opositores que obtienen cada calificación (frecuencias absolutas de cada calificación) y derivando el resto de las columnas de la tabla de frecuencias tal y como se indica a continuación:

X_i	n_i	$f_i = n_i / N$	N_i	$F_i = N_i /N$
0	6	0,06	6	0,06
1	10	0,1	16	0,16
2	13	0,13	29	0,29
3	11	0,11	40	0,4
4	13	0,13	53	0,53
5	16	0,16	69	0,69
6	11	0,11	80	0,8
7	10	0,1	90	0,9
8	6	0,06	96	0,96
9	2	0,02	98	0,98
10	2	0,02	100	1
	N = 100	□ $f_i = 1$		

Puesto que las frecuencias relativas pueden interpretarse como el peso relativo de cada valor en la distribución, el porcentaje de alumnos que aprobó la oposición (o sea, que obtuvieron un 5) será la frecuencia relativa correspondiente al valor 5 de la variable, es decir, el 16 por ciento (0,16).

Puesto que las frecuencias absolutas acumuladas correspondientes a un valor dado de la variable pueden interpretarse como el número de valores iguales o inferiores a ese valor dado, resulta que para el valor 6 de la variable hay 80 opositores que obtuvieron una calificación inferior o igual a 6. Por lo tanto habrá 20 opositores (100-80=20) que han obtenido una calificación superior a 6. Este resultado quiere decir que en caso de haber sólo 20 plazas, la nota mínima para superar la oposición hay que situarla por encima del 6. Es decir, superarán la oposición los alumnos que obtengan más de un 6.

El diagrama de barras y el polígono de frecuencias suelen ofrecer información sobre la simetría y la normalidad de la distribución. En este caso vemos que estas representaciones no se desvían demasiado de una campana de Gauss, lo que indica que puede admitirse la normalidad de los datos. En cuanto a la simetría se observa que la parte izquierda de la distribución aglomera más frecuencia, por lo que podría haber una asimetría débil en esa dirección. No obstante, podría admitirse también la simetría, al igual que la normalidad con un margen de error no muy elevado.

---

51. Los valores relativos al número de empresas y trabajadores en una determinada región son los siguientes:

1. **Construir la distribución de frecuencias adecuada a los datos.**
2. **Hallar el número de empresas con más de 300 trabajadores.**
3. **Hallar el porcentaje de empresas con más de 100 trabajadores y menos de 400**

trabajadores	Nº de empresas
0-100	25
100-200	37
200-300	12
400-500	22
500-600	21
600-700	13
700-800	5
800-900	3
900-1000	2

<u>SOLUCIÓN:</u>

[Li-1;Li)	ci	ni	fi	Ni	Fi
[0;100)	50	25	0,178571	25	0,178571
[100;200)	150	37	0,264286	62	0,442857
[200;300)	250	12	0,085714	74	0,528571
[400;500)	450	22	0,157143	96	0,685714
[500;600)	550	21	0,15	117	0,835714
[600;700)	650	13	0,092857	130	0,928571
[700;800)	750	5	0,035714	135	0,964286
[800;900)	850	3	0,021429	138	0,985714
[900;1000)	950	2	0,014286	140	1

Si observamos la columna de frecuencias absolutas acumuladas *Ni* de la tabla constatamos que la frecuencia absoluta acumulada hasta empresas con 300 trabajadores es de 74 lo que quiere decir que con más de 300 trabajadores existen $140 - 74 = 66$

Si observamos la columna de frecuencias relativas acumuladas fi de la tabla, tenemos que el potcentaje de empresas con 400 trabajadores o menos es de 0,5286(52.865%), es decir,

el mismo que el correspondiente a 300 trabajadoreso menos (no se registraron empresas con un número de trabajadores comprendido ente 300 y 400). Por otra parte, el porcentaje de empresas con 100 trabajadores o menos es de 0,1786 (17,86%), lo que indica que con más de 100 y menos de 400 tenemos $0{,}5286 - 0{,}1786 = 0{,}35(35\%)$.

---

52. Un examen consta de 5 preguntas en las que dos alumnos A y B obtienen las siguientes calificaciones segun el orden de las preguntas:

A: 5, 8, 6, 5, 4.
B: 3, 7, 8, 6, 3.

a) **¿cual de los dos alumnos tuvo mejor nota sabiendo que los ejercicios 1,3 y 4 puntuan la mitad que los ejercicios 2 y 5?**

b) **si consideramos que todas las preguntas valen igual, ¿que alumno obtendra mejor calificacion si utilizamos la media geometrica? ¿y si usamos la media cuadratica?**

SOLUCIÓN:

a) Se calcula la media ponderada con los pesos que se indican para cada uno de los alumnos siendo el alumno con mayor media el que obtuvo mejor nota.
Los pesos para los problemas 1, 3 y 4 sera 1 y para los problemas 2 y 5 sera 2. asi, obtendremos los siguientes resultados:

$$A = \frac{(5x1+8x2+6x1+5x1+4x2)}{(1+2+1+1+2)} = 5{,}714$$

$$B = \frac{(3x1+7x2+8x1+6x1+3x2)}{(1+2+1+1+2)} = 5{,}286$$

Por tanto fue el alumno A el que obtuvo mejor calificacion.

b) Primero recordemos las expresiones de la media geometrica y la media cuadratica:

$$G= \sqrt{\prod_{i-1}^{k} Xi^{fi}} \qquad\qquad C= \sqrt{\frac{1}{n}\sum_{i-1}^{k} fiXi^2}$$

Si consideramos la media geometrica el alumno A obtiene una calificacion de 5.448 y el alumno B una de 4.967.
Si consideramos la media cuadratica el alumno A obtiene un resultado de 5.762 y el alumno B de 5.779.

---

53. En una distribucion discreta de 6 valores, a saber: -10,3,a,10,1,0, sabemos que su desviacion tipica es igual al coeficiente de variacion de Pearson. Se pide:

a) **Hallar la media de la distribucion**
b) **Hallar el valor desconocido de a.**

SOLUCIÓN:

a)
Como CV=s entonces se tiene que la media aritmetica vale 1, puesto que el coeficiente de

variacion de Pearson es el cociente entre la desviacion tipica y la media aritmetica.

b)
Aplicando el resultado obtenido en el apartado anterior se tiene que:

$$1=\frac{(-10+3+a+10+1+0)}{6}$$

Despejando a de la expresion anterior se tiene que a=2.

---

54. En un aparcamiento cobran por cada minuto que esta estacionado el vehiculo un euro y veinte centimos. La ocupación del parking en un dia fué:

TIEMPO DE ESTACIONAMIENTO	NUMERO DE VEHICULOS
0-60	1240
60-120	3575
120-180	746
180-240	327
240-360	218
360-1440	44

a) Obtener el tiempo medio de estacionamiento.
b) A partir de que cantidad de tiempo un vehiculo esta estacionado mas que el 85% de los vehiculos?

SOLUCIÓN:

LI-1- LI	XI	NI	XI.NI	CI	NI
0-60	30	1240	37200	60	1240
60-120	90	3575	321750	60	4815
120-180	150	746	111900	60	5561
180-240	210	327	68670	60	5888
240-360	300	218	65400	120	6106
360-1440	900	44	39600	1080	6150
		6150	644520		

Por lo tanto, el tiempo medio de aparcamiento es:

$$\overline{X}=\frac{644\,520}{6150}=104.8$$

La medida de posicion que indica a partir de que cantidad de tiempo un vehiculo esta estacioNado mas que el 85% de los vehiculos es el percentil 85.

$$\frac{85N}{100}=5227.5$$

Y la primera frecuéncia acumulada que lo supera es N3=5561, con lo que el P85 esta en el intervalo (120,180):

$$P_{85}=L_{i-1}+C_i\frac{85N/100-N_{i-1}}{n_i}=120+60\frac{5227.5-4815}{746}=153.176944$$

Es decir, a partir de 153.176944 minutos un vehiculo esta estacionado mas que el 85% de los vehiculos.

55. Los porcentajes de participación de los alumnos en las actividades extraescolares durante los trimestres lectivos de los dos últimos cursos sufrió el siguiente aumento: el primer trimestre 8%, el segundo 12%, el tercero 18%, el primer trimestre del ultimo curso 27%, el segundo 40,5% , el tercero 60,75%. Calcular la media geométrica del porcentaje de participación de los alumnos en esas actividades.

SOLUCIÓN:

$$Mg = \sqrt[N]{\prod_{i=1}^{n} x_i^{ni}}$$

Siendo en nuestro caso n = 6, $x_1 = 8$, $x_2 = 12$, $x_3 = 18$, $x_4 = 27$, $x_5 = 40,5$, $x_6 = 60,75$
y $n_i = 1$ para todo i = 1,2,3,4,5,6, con lo que N = 6. Aplicando la formula anterior tenemos:

$$Mg = \sqrt[6]{8*12*18*27*40,5*60,75} = \sqrt[6]{114791256} = \underline{22,04540769}$$

Tambien se puede utilizar para el calculo de la media geometrica la formula anterior tomando cualquier tipo de logaritmos. Usando logaritmos en base 10:

$$Mg = \text{antilog } \frac{\sum_{i=1}^{n} n_i \log(x_i)}{N} =$$

$$= \text{antilog } \left(\frac{\log(8)+\log(12)+\log(18)+\log(27)+\log(40,5)+\log(60,75)}{6} \right)$$

$$= \text{antilog}(1,343318135) = \underline{22,04540769.}$$

Entonces la media geometrica de la participación de los alumnos es <u>22,0454%.</u>

56. La cajera de una tienda va anotando los precios y las cantidades de los productos que ha adquirido un cliente. En el ticket de compra aparece esta relación:

Producto	**nº unidades**	**Precio/unidad**
Azúcar	**5**	**156**
Aceite girasol	**10**	**115**
Leche semidesnatada	**15**	**64**
Zumo	**6**	**75**
lata de refrescos	**12**	**50**
botella de vino	**2**	**139**

¿Cual será el precio superado por la mitad de los productos?

SOLUCIÓN:

La pregunta se puede formular de otro modo: ¿Cuál es el valor que divide la distribución en dos partes?; es decir, ¿Cuál es el valor de la mediana?

Recordemos que para su cálculo, los valores deben estar ordenados. La cajera, posiblemente por comodidad o por falta de tiempo, no sigue esa estrategia. Va registrando según llegan los artículos. Colocando los precios de menor a mayor con el correspondiente numero de unidades y hallamos las frecuencias absolutas acumuladas.

Producto	Precio	Nº unidades	Ni
lata de refrescos	50	12	12
Leche semidesnatada	64	15	27
Zumo	75	6	33
aceite de girasol	115	10	43
botella de vino	169	2	45
Azúcar	156	5	50

La mitad de los valores es N/2 = 50/2 =25. La primera frecuencia acumulada que lo supera es $N_2 = 27$. Esto significa que el precio correspondiente a la mediana es el de un litro de leche semidesnatada: <u>64 pesetas</u>

---

57. Completar la siguiente tabla para el estudio de la concentración de una distribución de frecuencias y calcular el indice de gini. Comentar el resultado.

xi	**ni**	**Ni**	**pi**	**si**	**Ai**	**qi**
10	**90**	**90**		**900**	**900**	
20	**50**	**140**			**1900**	
40	**30**	**170**				
60	**20**	**190**		**1200**		**0,86**
70	**10**	**200**		**700**	**5000**	**1**

<u>SOLUCIÓN:</u>

Los valores de p_i se obtienen dividiendo las frecuencias acumuladas absolutas, N_i, entre el total de datos, N:

Ni	pi = Ni/N
90	90/200 = 0,45
140	140/200 = 0,7
170	170/200 = 0,85
190	190/200 = 0,95
200	200/200 = 1

Los valores de s_i se calculan multiplicando los valores de la variable por las frecuencias respectivas, $s_i = x_i n_i$, y los de A_i son los acumulados de la columna anterior, s_i:

xi	ni	si = xi ni	$\boxed{Ai = \sum_{k=1}^{i} s_k}$
10	90	900	900
20	50	20*50 = 1000	1900
40	30	40*30 = 1200	1900+1200 = 3100
60	20	1200	3100+1200 = 4300
70	10	700	5000

Dividiendo A_i entre A_n obtenemos q_i:

Ai	qi = Ai/An
900	900/5000 = 0,18
1900	1900/5000 = 0,32
3100	3100/5000 = 0,62
4300	0,86
5000	1

Las diferencias $p_i - q_i$:

pi	qi	pi -qi
0,45	0,18	0,45 - 0,18 = 0,27
0,7	0,38	0,70 - 0,38 = 0,32
0,85	0,62	0,85 - 0,62 = 0,23
0,95	0,86	0,95 - 0,86 = 0,09
2,95		0,91

Los valores de la última fila se han omitido, pues en la fórmula del índice de gini sólo se suma hasta el penúltimo, n-1.

Así, el índice de Gini será:

$$I_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = \frac{0,91}{2,95} = \underline{0,30847458.}$$

Podemos afirmar que la concentración tiende a ser baja.

---

58. Cierta empresa se dedica a la elaboración y a la venta directa de 4 productos. Estos tienen diferentes precios y, cada dia, se venden determinadas cantidades.
Dicha información se recoge en la siguiente tabla:

Artículo	**Precio unitario**	**cantidad**
A	**20**	**300**
B	**35**	**225**
C	**50**	**150**
D	**70**	**50**
		725

Las condiciones del mercado provocan un aumento de la producción en una unidad en todos los productos, lo que hace bajar los precios según aparece en la tabla:

Artículo	**Precio unitario**	**cantidad**
A	**18**	**301**
B	**33**	**226**
C	**46**	**151**
D	**63**	**51**

		729

¿Cómo varia el ingreso total, teniendo en cuenta estos cambios? ¿Y el ingreso medio? Desde el punto de vista de la dispersión, ¿hay cambios significativos?

SOLUCIÓN:

Realizamos los cálculos para los primeros datos en la tabla:

Artículo	Precio unitario	Cantidad	pi qi	pi qi
A	20	300	6000	120000
B	35	225	7875	275625
C	50	150	7500	375000
D	70	50	3500	245000
		725	24875	1015625

El volumen total de producción es N = 725. El ingreso total es

$$\sum_{i=1}^{n} p_i q_i = \underline{24875,}$$

Y el ingreso medio

$$\overline{X} = \frac{\sum_{i=1}^{n} p_i q_i}{N} = \frac{24875}{725} = \underline{34{,}3103448 \text{ u.m.}}$$

Varianza:

$$S^2_x = \frac{\sum_{i=1}^{n} p_i^2}{N} - \overline{X}^2 = \frac{1015625}{725} - 34{,}3103448^2 = \underline{223{,}662307,}$$

Y la desviación típica

$$S_x = \sqrt{223{,}662307} = \underline{14{,}9553438.}$$

Con las nuevas condiciones del mercado, construimos la siguiente tabla:

Artículo	Precio unitario	Cantidad	pi qi	pi qi
A	18	301	5418	97524
B	33	226	7458	246114
C	46	151	6946	319516
D	63	51	3213	202419
		729	23035	865573

En este segundo caso, el ingreso total

$$\sum_{i=1}^{n} p_i q_i = \underline{23035,}$$

Y el ingreso medio

$$\overline{Y} = \frac{\sum_{i=1}^{n} p_i q_i}{N} = \frac{23035}{729} = \underline{31,59808.}$$

La varianza

$$S^2_y = \frac{\sum_{i=1}^{n} p_i^2 q_i}{N} - \overline{Y}^2 = \frac{865573}{729} - 31,59808^2 = \underline{188,904304}$$

y la desviación típica

$$S_y = \sqrt{188,904304} = \underline{13,744246.}$$

De acuerdo con esto, el ingreso total y el medio es superior en el primer caso.
Desde el punto de vista de la dispersión, podemos comparar ambos resultados a traves del coeficiente de variación de pearson:

$$V_x = \frac{S_x}{\overline{X}} = \frac{14,9553438}{34,3103448} = 0,435884392$$

$$V_y = \frac{S_y}{\overline{Y}} = \frac{13,744246}{31,59808} = 0,434971.$$

Que indica que no han cambiado mucho los resultados en relacion con la media.

---

59. Una fábrica de coches desea estudiar el consumo de un nuevo modelo de coche que quiere lanzar al mercado. Para ello realiza cien pruebas echando diez litros de gasolina y viendo que distancia en kilómetros recorre el coche. Los resultados de las pruebas fueron los siguientes:

85	**90**	**91**	**88**	**91**	**91**	**86**	**92**	**90**	**89**
91	**87**	**88**	**88**	**90**	**90**	**89**	**90**	**90**	**89**
91	**87**	**90**	**84**	**91**	**88**	**90**	**88**	**88**	**88**
92	**90**	**90**	**90**	**93**	**90**	**89**	**92**	**91**	**92**
89	**88**	**91**	**89**	**90**	**90**	**88**	**90**	**89**	**86**
90	**88**	**88**	**94**	**91**	**90**	**92**	**87**	**90**	**91**
92	**88**	**92**	**92**	**88**	**89**	**88**	**91**	**89**	**91**
91	**88**	**88**	**92**	**89**	**87**	**88**	**88**	**91**	**88**
89	**90**	**93**	**89**	**91**	**92**	**89**	**85**	**86**	**91**
89	**87**	**88**	**88**	**93**	**90**	**95**	**89**	**92**	**89**

a) obtener la distribución de frecuencias y su representación grafica.

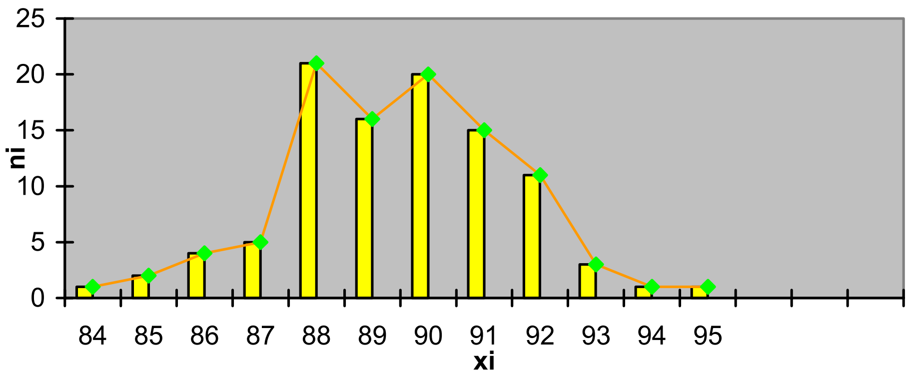
b) Agrupar estos valores en los intervalos [83.5-86.5), [86.5-89.5), [89.5-92.5) y [92.5-95.5). Obtener la correspondiente distribución de frecuencias con las marcas de clase, las amplitudes de los intervalos y las alturas. Representar gráficamente la distribución.

SOLUCIÓN:

a) la distribución de frecuencias es:

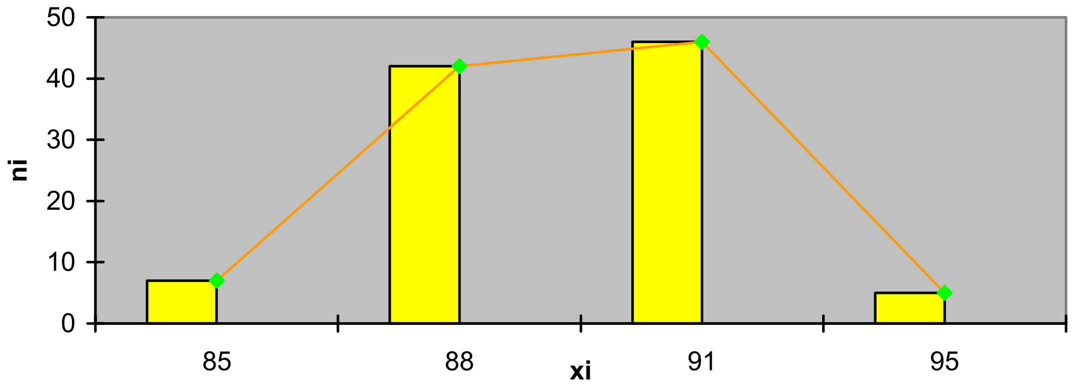
x_i	n_i	f_i	N_i	F_i
84	1	0.01	1	0.01
85	2	0.02	3	0.03
86	4	0.04	7	0.07
87	5	0.05	12	0.12
88	21	0.21	33	0.33
89	16	0.16	49	0.49
90	20	0.20	69	0.69
91	15	0.15	84	0.84
92	11	0.11	95	0.95
93	3	0.03	98	0.98
94	1	0.01	99	0.99
95	1	0.01	100	1
	100	1		

Para representar gráficamente esta distribución de frecuencias, usaremos un diagrama lineal de barras con su correspondiente polígono de frecuencias.



b) agrupando los intervalos tenemos:

L_{i-1}-L_i	x_i	n_i	f_i	N_i	F_i	c_i	h_i
83.5-86.5	85	7	0.07	7	0.07	3	2.333
86.5-89.5	88	42	0.42	49	0.49	3	14
89.5-92.5	91	46	0.46	95	0.95	3	15.333
92.5-95.5	94	5	0.05	100	1	3	1.667
		100	1				



60. Los resultados en el análisis del valor calórico (Kcal/ración) de 20m marcas de galletas normales y 12 integrales, considerando como ración 5 o 6 galletas (30 gramos) son los siguientes:

Normales	Kcal	integrales	kcal
Canente	125	cura	135
Cura	125	fomesa	135
Curra	150	dan	130
Dan	135	desa	135
Desa	150	erus	125
Erus	130	gafin	140
Fomesa	130	les	150
Foleda	145	mali	135
Fura	135	naria	135
Gafin	145	sanli	145
Gelo	130	suno	150
Hela	150	veras	130
Hipu	140		
Les	150		
Mali	140		
Neria	145		
Pros	130		
Riz	130		
Suno	130		
Veras	140		

a) Escribir y representar la distribución de frecuencias
b) Calcular la media aritmética, mediana y moda

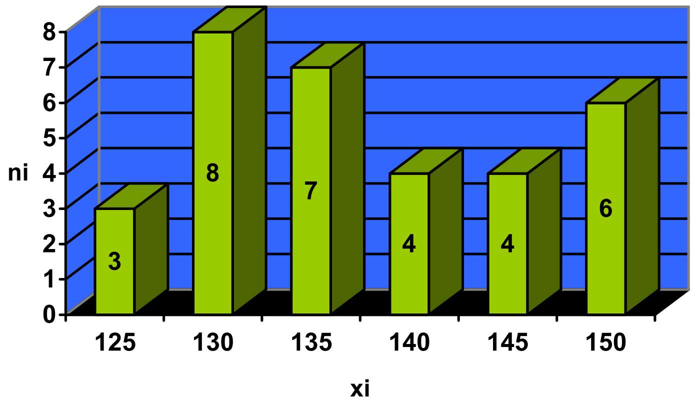
SOLUCIÓN:

a) sea X la variable "calorías por ración de galletas" y n_i las frecuencias absolutas, o sea número de veces que se repite cada dato x_i de la variable la distribución de frecuencias es:

x_i	n_i	N_i	f_i	F_i
125	3	3	0.09375	0.09375
130	8	11	0.25	0.34375
135	7	18	0.21875	0.5625
140	4	22	0.125	0.6875
145	4	26	0.125	0.8125

150	6	32	0.1875	1
	32		1	

Donde N_i son las frecuencias absolutas acumuladas, f_i son las frecuencias relativas y F_i son las frecuencias relativas acumuladas.



b) completamos la tabla con $x_i n_i$ para calcular la media X

x_i	n_i	x_i n_i
125	3	375
130	8	1040
135	7	945
140	4	560
145	4	580
150	6	900
	32	4400

Media: 137.5

Para la mediana veamos entre que valores de la columna de frecuencias acumuladas N_i esta N/2= 16:

x_i	n_i	N_i
125	3	3
130	8	11
135	7	18
1140	4	22
145	4	26
150	6	32
	32	

N/2 =16 esta entre N_2 y N_3 como la distribución es de datos sin agrupar, Me = x_3 = 135.

La moda, por ser datos sin agrupar es el valor de la variable que más se repite, es decir, el dato que mas frecuencia absoluta tenga. Como la n_i máxima es $n_2 = 8$, la moda Mo = x_2 = 130.