Ejercicios resueltos

5.2-1 Dada la siguiente tabla de ingresos mensuales, calcular la media, la mediana y el intervalo modal.

Ingresos	Frecuencia
Menos de 400€	5
[400, 700)	20
[700, 1.000)	35
[1.000, 1.500)	70
[1.500, 2.500)	50
Más de 2.000€	15

Solución

La media no se puede calcular ya que el último intervalo está abierto hasta el infinito.

Para calcular la mediana nos fijamos en que hay 195 datos, por lo que el lugar intermedio es el 98, luego la mediana está en el intervalo [1.000, 1.500). Aplicamos la fórmula:

$$\mathrm{Me} = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i = 1.000 + \frac{\frac{195}{2} - 60}{70} \cdot 500 = 1.000 + \frac{37{,}5}{70} \cdot 500 =$$
$$= 1.000 + 0{,}5357 \cdot 500 = 1.000 + 267{,}85 = 1.267{,}85$$

El intervalo modal es el [1.000, 1.500)

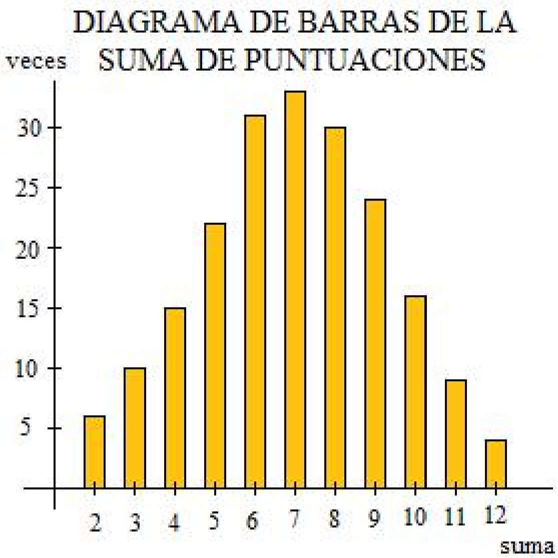
5.2-2 Se lanzan dos dados 200 veces y cada vez se recoge la suma de las puntuaciones, obteniendo los siguientes resultados:

Suma	2	3	4	5	6	7	8	9	10	11	12
Veces	6	10	15	22	31	33	30	24	16	9	4

Realiza el diagrama de barras, calcula la media y las frecuencias relativas.

Solución

El diagrama de barras es:



Para calcular la media y las frecuencias relativas completamos la tabla:

X_i	n_i	$x_i \cdot n_i$	N_i	f_i	F_i
2	6	12	6	0,03	0,03
3	10	30	16	0,05	0,08
4	15	60	31	0,075	0,155
5	22	110	53	0,11	0,265
6	31	186	84	0,155	0,42
7	33	231	117	0,165	0,585
8	30	240	147	0,15	0,735
9	24	216	171	0,12	0,855
10	16	160	187	0,08	0,935
11	9	99	196	0,045	0,98
12	4	48	200	0,02	1
TOTAL	200	1392		1	

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{N} = \frac{1.392}{200} = 6,96$$

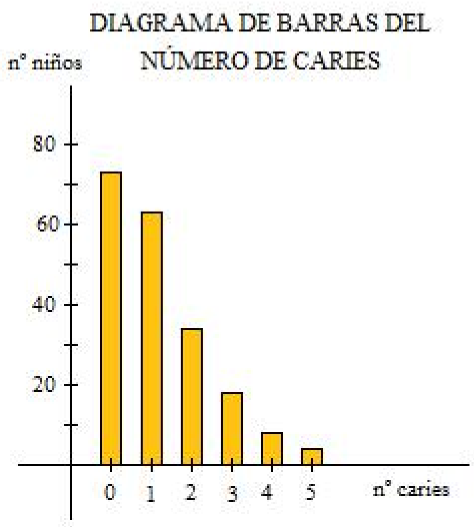
5.2-3 Un dentista observa el número de caries en 200 niños, obteniendo el siguiente resultado:

Nº de caries	Nº de niños
0	73
1	63
2	34
3	18
4	8
5	4

Dibuja un diagrama de barras y calcula la media, moda, mediana y desviación típica.

Solución

El diagrama de barras queda:



Para calcular las medidas calculamos la tabla:

X_i	n_i	$x_i \cdot n_i$	N_i	x_i^2	$x_i^2 \cdot n_i$
0	73	0	73	0	0
1	63	63	136	1	63
2	34	68	170	4	136
3	18	54	188	9	162
4	8	32	196	16	128
5	4	20	200	25	100
TOTAL	200	237			589

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i \cdot n_i}{N} = \frac{237}{200} = 1,185$$

La mediana será la media de los valores que ocupen los lugares 100 y 101, si observamos la tabla de frecuencias acumuladas, vemos que ambos valores valen 1, por lo que **Me = 1**

El valor que más se repite es 0, luego **Mo = 0**

Para calcular la desviación típica calculamos primero la varianza:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 \cdot n_i}{N} - \overline{x}^2 = \frac{589}{200} - 1,185^2 = 2,945 - 1,4042 = 1,5408$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,5408} = 1,2412$$

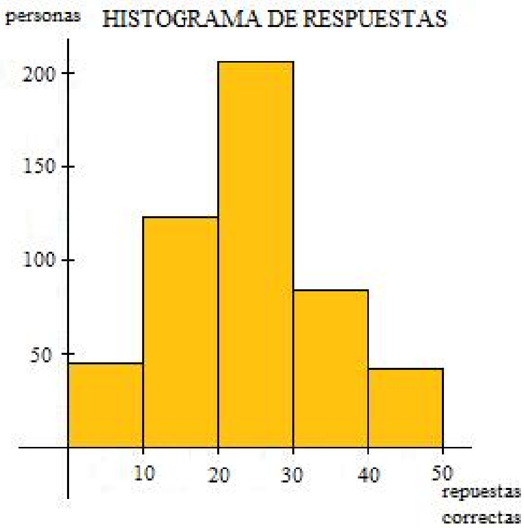
5.2-4 Los resultados de un test de 49 preguntas realizado a 500 personas han dado los siguientes resultados:

Respuestas correctas	[0,10)	[10,20)	[20,30)	[30,40)	[40,50)
Número de personas	45	123	206	84	42

Dibujar el histograma y calcular la media y la desviación típica.

Solución

Para representar el histograma nos fijamos en que los intervalos tienen todos amplitud 10.



Para calcular la media y la desviación típica calculamos la tabla:

Intervalo	X_i Marca de clase	n_i	$x_i \cdot n_i$	x_i^2	$x_i^2 \cdot n_i$
[0,10)	5	45	225	25	1.125
[10,20)	15	123	1.845	225	27.675
[20,30)	25	206	5.150	625	128.750
[30,40)	35	84	2.940	1.225	102.900
[40,50)	45	42	1.890	2.025	85.050
	TOTAL	500	12.050		345.500

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i \cdot n_i}{N} = \frac{12.050}{500} = 24,1$$

$$\sigma^2=\frac{\sum_{i=1}^{n} x_i^2 \cdot n_i}{N} - \overline{x}^2 = \frac{345.500}{500} - 24,1^2 = 691 - 580,81 = 110,19$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{110,19} = 10,49$$

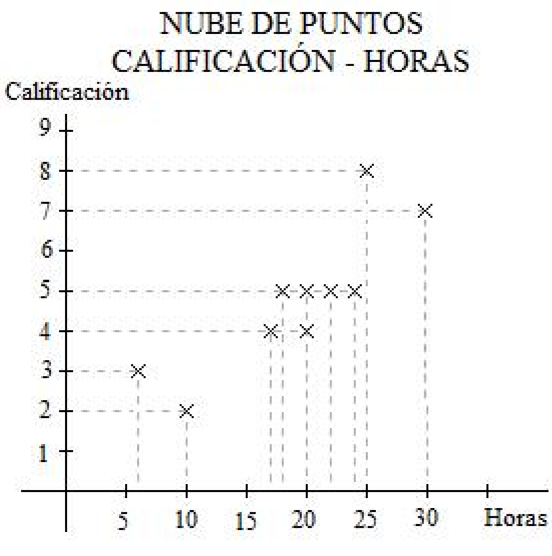
5.2-5 El número de horas dedicado al estudio de una asignatura por 10 estudiantes y la calificación obtenida en dicha asignatura por cada uno de ellos viene dado en la siguiente tabla:

X: horas de estudio	25	10	20	30	17	22	24	6	18	20
Y: calificación	8	2	5	7	4	5	5	3	5	4

Dibuja la nube de puntos y decide a partir de ella si puede existir algún tipo de correlación entre las variables.

Solución

El diagrama de dispersión queda:



Observando la nube de puntos podemos decir que existe una correlación lineal relativamente fuerte y directa, es decir, cuántas más horas de estudio más nota.

5.2-6 En una determinada región se realiza la medición de temperatura mínima en puntos de distinta altitud, obteniendo los siguientes resultados:

X: Altitud (m)	120	300	210	250	160	350	70	380
Y: Temperatura (ºC)	8	5	7	3	7	3	8	2

Calcula e interpreta el coeficiente de correlación lineal entre las variables.

Solución

Para calcular el coeficiente de correlación lineal rellenamos la tabla:

	X_i	Y_i	$X_i \cdot Y_i$	$X_i^{\ 2}$	$Y_i^{\ 2}$
	120	8	960	14.400	64
	300	5	1.500	90.000	25
	210	7	1.470	44.100	49
	250	3	750	62.500	9
	160	7	1.120	25.600	49
	350	3	1.050	122.500	9
	70	8	560	4.900	64
	380	2	760	144.400	4
Totales	1.840	43	8.170	508.400	273

$$\overline{X} = \frac{1.840}{8} = 230$$

$$\overline{Y} = \frac{43}{8} = 5,375$$

$$\sigma_X^2 = \frac{508.400}{8} - 230^2 = 63.550 - 52.900 = 10.650$$

$$\sigma_X = \sqrt{10.650} = 103,2$$

$$\sigma_Y^2 = \frac{273}{8} - 5,375^2 = 34,125 - 28,89 = 5,235$$

$$\sigma_Y = \sqrt{5,235} = 2,288$$

$$\sigma_{XY} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y} = \frac{8.170}{8} - 230 \cdot 5,375 =$$

$$= 1.021,25 - 1.236,25 = -215$$

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{-215}{103,2 \cdot 2,288} = \frac{-215}{236,121} = -0,91$$

Como el coeficiente de correlación es próximo a -1 la correlación lineal es fuerte y como es negativo la correlación es inversa, es decir al aumentar una variable disminuye la otra.

5.2-7 Las mediciones de peso y talla de una muestra de ocho estudiantes se recogen en la siguiente tabla:

X: peso (Kg.)	75	60	72	80	65	68	73	55
Y: talla (cm.)	180	158	170	169	170	176	175	160

a) Calcula el coeficiente de correlación lineal y explica el tipo de correlación existente.

b) Calcula la recta de regresión de y sobre x.

c) ¿Qué altura cabe esperar para un estudiante que pese 70 Kg?

Solución

a) Para calcular el coeficiente de correlación lineal rellenamos la tabla:

	X_i	Y_i	$X_i \cdot Y_i$	X_i^2	Y_i^2
	75	180	13.500	5.625	32.400
	60	158	9.480	3.600	24.964
	72	170	12.240	5.184	28.900
	80	169	13.520	6.400	28.561
	65	170	11.050	4.225	28.900
	68	176	11.968	4.624	30.976
	73	175	12.775	5.329	30.625
	55	160	8.800	3.025	25.600
Totales	548	1.358	93.333	38.012	230.926

$$\bar{X} = \frac{548}{8} = 68,5$$

$$\overline{Y} = \frac{1.358}{8} = 169,75$$

$$\sigma_X^2 = \frac{38.012}{8} - 68,5^2 = 4.751,5 - 4.692,25 = 59,25$$

$$\sigma_X = \sqrt{59,25} = 7,697$$

$$\sigma_Y^2 = \frac{23.0926}{8} - 169,75^2 = 28.865,75 - 28.815,062 = 50,688$$

$$\sigma_Y = \sqrt{50,688} = 7,119$$

$$\sigma_{XY} = \frac{93.333}{8} - 68,5 \cdot 169,75 = 11.666,625 - 11.627,875 = 38,75$$

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{38,75}{7,697 \cdot 7,119} = \frac{38,75}{54,794} = 0,71$$

Como el coeficiente de correlación no es demasiado próximo a 1 existe correlación lineal moderada y como es positivo la correlación es directa, es decir al aumentar una variable aumenta la otra.

b) Para calcular la recta de regresión de y sobre x aplicamos la fórmula:

$$y - \overline{y} = \frac{\sigma_{XY}}{\sigma_x^2}(x - \overline{x}) \Rightarrow y - 169,75 = \frac{38,75}{59,25}(x - 68,5)$$

$$y - 169,75 = 0,654 \cdot (x - 68,5)$$

$$y - 169,75 = 0,654x - 44,8$$

$$y = 0,654x + 124,95$$

c) Sustituimos x por 70 en la recta de regresión

$$y = 0,654 \cdot 70 + 124,95$$

$$y = 45,78 + 124,95$$

$$y = 170,73$$

Cabe esperar un peso de 170,73 Kg.