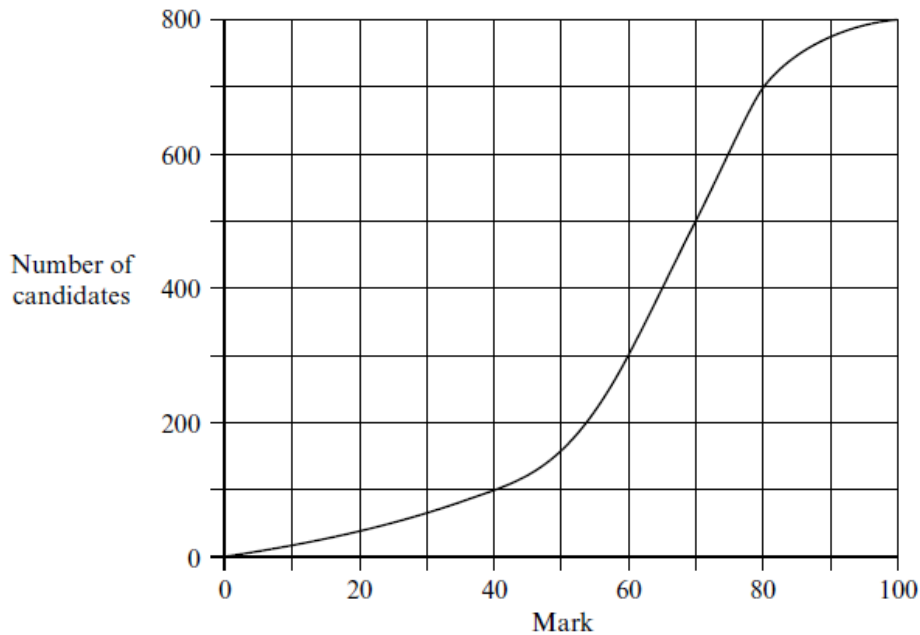


MATEMÁTICAS 1º BI-NM - Serie Estadística Unidimensional y Bidimensional

- 1 Entra en la página web del Instituto Nacional de Estadística y elige una variable numérica de tu interés que disponga de frecuencias absolutas distintas de uno. Con sus datos creas una hoja de cálculo para hallar: la moda, la mediana, los cuartiles, la media aritmética, la varianza y la desviación típica. Redacta un informe con los siguientes puntos: Objeto del estudio. Tabla con los datos, las frecuencias y sus operaciones. Diagrama de frecuencia y de frecuencias acumuladas. Medidas de centralización y dispersión. Diagrama de Caja y Bigotes. Análisis de los resultados.
- 2 The population P is the set of numbers $\{-3, 3, a, b\}$, and has a mean of 0 and a standard deviation of $\sqrt{17}$. Given that $b > a$, determine the values of a and b .
- 3 A test which is marked out of 100 is written by 800 students. The cumulative frequency graph for the results of the test is given below.



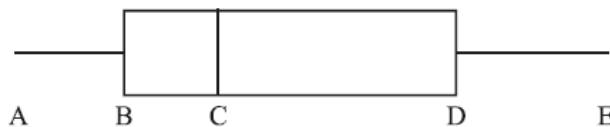
- (a) How many students scored 40 marks or less on the test?
 - (b) The middle 50% of test results lie between the marks a and b , where $a < b$. Write down the values of a and b .
- 4 De enero a septiembre la cantidad media de accidentes automovilísticos por mes fue de 630. De octubre a diciembre la media fue de 810 accidentes por mes.

¿Cuál fue la cantidad media de accidentes automovilísticos por mes para todo el año?

- 5 Se tiene el siguiente conjunto de datos

18 ; 18 ; 19 ; 19 ; 20 ; 22 ; 22 ; 23 ; 27 ; 28 ; 28 ; 31 ; 34 ; 34 ; 36 .

A continuación se muestra el diagrama de caja y bigote para estos datos.



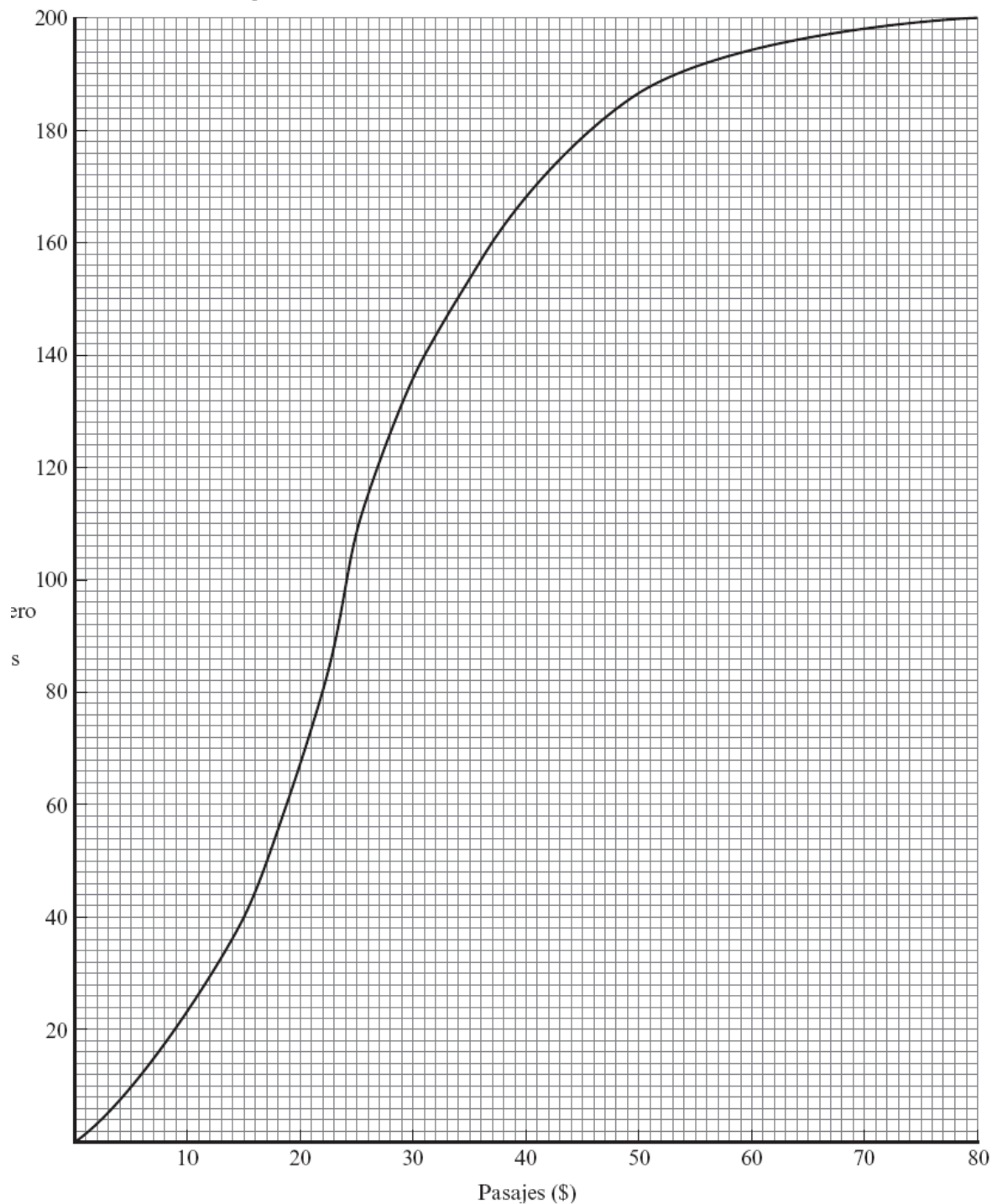
- (a) Escriba los valores de A, B, C, D y E.

A = B = C = D = E =

- (b) Halle el rango intercuartil.

6

- (i) Una empresa tiene 200 taxis. La curva de frecuencias acumuladas a continuación muestra los pasajes, en dólares (\$), cobrados por los taxis en una mañana en particular.



- (a) Use la curva para estimar:

- (i) la mediana de los pasajes;
- (ii) el número de taxis cuyo pasaje es de \$35 o menos.

La empresa cobra 55 centavos por kilómetro de distancia recorrida. No hay ningún otro cargo. Use la curva para contestar las preguntas que siguen.

- (b) La distancia recorrida esa mañana por el 40% de los taxis es de menos de a km. Halle el valor de a .
- (c) ¿Qué porcentaje de los taxis recorre más de 90 km esa mañana?

- 7 En un suburbio de una gran ciudad se vendieron 100 casas en un período de tres meses. La **tabla de frecuencias acumuladas** a continuación muestra la distribución de los precios de venta (en miles de dólares).

Precio de venta P (\$ 1000)	$P \leq 100$	$P \leq 200$	$P \leq 300$	$P \leq 400$	$P \leq 500$
Número total de casas	12	58	87	94	100

- (a) Represente esta información sobre una **curva** de frecuencias acumuladas usando una escala de 1 cm para representar \$ 50 000 en el eje horizontal, y 1 cm para representar 5 casas en el eje vertical.
- (b) Use la curva para hallar el rango intercuartil.

La información anterior se representa en la siguiente distribución de frecuencias.

Precio de venta P (\$ 1000)	$0 < P \leq 100$	$100 < P \leq 200$	$200 < P \leq 300$	$300 < P \leq 400$	$400 < P \leq 500$
Número de casas	12	46	29	a	b

- (c) Halle el valor de a y de b .
- (d) Use valores centrales de intervalo para calcular una estimación del precio de venta medio.
- (e) Las casas que se venden por más de \$ 350 000 se describen como *De Luxe*.
- (i) Use la **gráfica** para estimar el número de casas *De Luxe* vendidas. Exprese su respuesta redondeada al entero más próximo.

- 8 The table below shows the marks gained in a test by a group of students.

Mark	1	2	3	4	5
Number of students	5	10	p	6	2

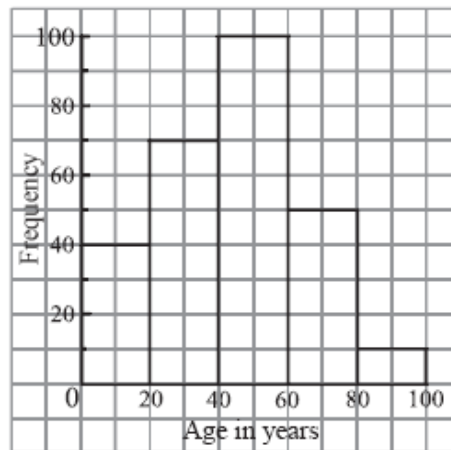
The median is 3 and the mode is 2. Find the **two** possible values of p .

- 9 Una colección de cinco números enteros tiene moda 3, mediana 4 y media 5. Escriba todas las posibles colecciones de cinco números que cumplan lo anterior.

- 10 En una fábrica hay 50 cajas. Sus pesos, w kg, se dividen en 5 clases, tal y como se muestra en la siguiente tabla.

Clase	Peso (kg)	Número de cajas
A	$9,5 \leq w < 18,5$	7
B	$18,5 \leq w < 27,5$	12
C	$27,5 \leq w < 36,5$	13
D	$36,5 \leq w < 45,5$	10
E	$45,5 \leq w < 54,5$	8

- (a) Compruebe que la media estimada del peso de las cajas es igual a 32 kg.
- (b) En la fábrica hay x cajas marcadas como 'Frágil'. Todas ellas pertenecen a la clase E. La media estimada del peso del resto de las cajas de la fábrica es igual a 30 kg. Calcule el valor de x .
- (c) Llegan a la fábrica y cajas nuevas, todas ellas con un peso perteneciente a la clase D. La media estimada del peso de **todas** las cajas que hay ahora en la fábrica es inferior a 33 kg. Halle el valor de y más grande posible.
- 11 The histogram below represents the ages of 270 people in a village.



- (a) Use the histogram to complete the table below.

Age range	Frequency	Mid-interval value
$0 \leq \text{age} < 20$	40	10
$20 \leq \text{age} < 40$		
$40 \leq \text{age} < 60$		
$60 \leq \text{age} < 80$		
$80 \leq \text{age} \leq 100$		

- (b) Hence, calculate an estimate of the mean age.

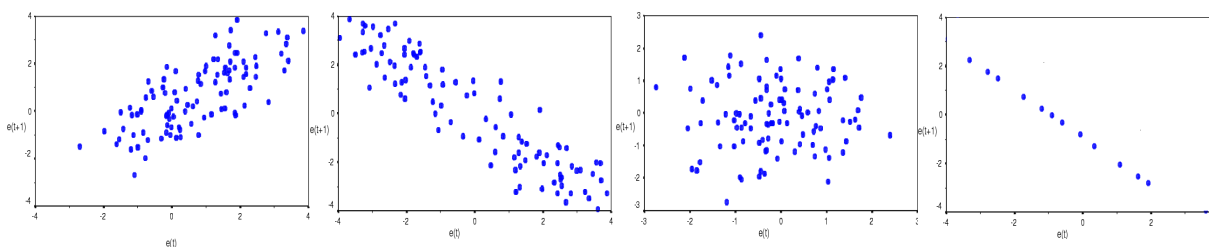
- 12 En la tabla de frecuencias a continuación se muestra el número de horas de sueño de 21 estudiantes.

Halle

- (a) la mediana;
 (b) el cuartil inferior;
 (c) el rango intercuartil.

Horas de sueño	Número de estudiantes
4	2
5	5
6	4
7	3
8	4
10	2
12	1

- 13 Entra en la página web del Instituto Nacional de Estadística y elige dos variables numéricas de tu interés de las que es razonable suponer que tengan alguna relación. Con sus datos creas una hoja de cálculo para hallar: las medias aritméticas, las desviaciones típicas, la covarianza y el coeficiente de correlación lineal. Halla las dos rectas de regresión. Proyecta una estimación y calcúlala. Redacta un informe con los siguientes puntos: Objeto del estudio. Diagrama de Nube de puntos. Tabla con los datos y sus operaciones. Medidas estadísticas calculadas. Valoración de la relación existente entre las dos variables. Ecuaciones de las dos rectas de regresión. Estimación realizada y su valoración.
- 14 Los coeficientes de correlación lineal de cuatro estadísticos, ordenados de mayor a menor, son: 0'7, 0'1, - 0'8 y - 1. Asigna a cada una de las siguientes nubes de puntos el valor adecuado explicando razonadamente la asignación realizada.



- 15 Se ha hecho un test compuesto por 4 preguntas a 10 alumnos de 15 años, 10 alumnos de 16 años y 10 alumnos de 17 años, obteniendo los siguientes aciertos:

Aciertos	0	1	2	3	4
Edad					
15	8	1	1	-	-
16	1	5	-	3	1
17	-	1	-	5	4

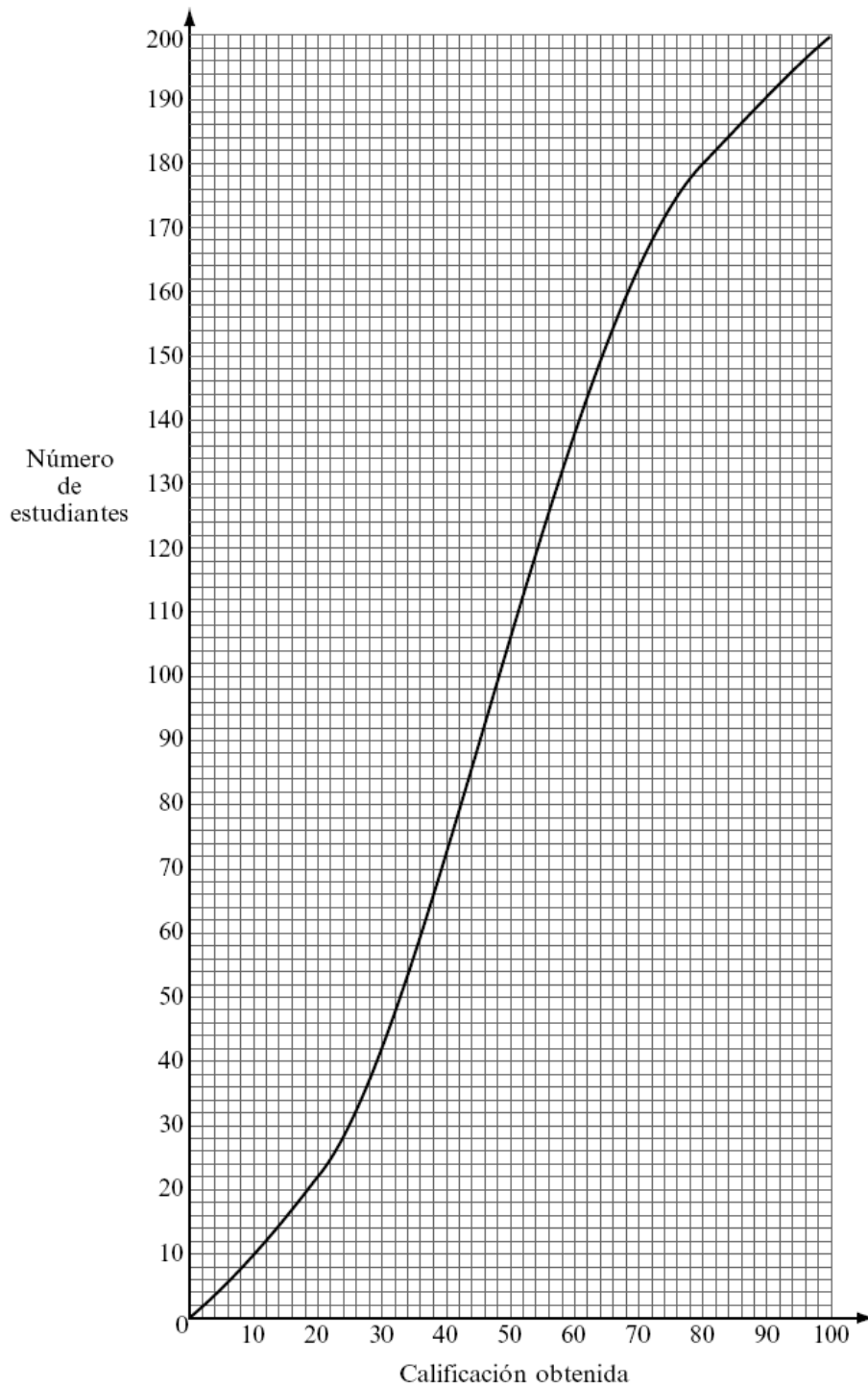
- a) Construye la tabla de frecuencias marginales absolutas y relativas de número de aciertos.
 b) Construye la tabla completa de frecuencias marginales absolutas de edades y con ella calcula su media y varianza.
 c) Representa en un diagrama el estadístico de los aciertos.
- 16 Se ha encuestado a 20 estudiantes sobre el tiempo diario (en minutos) que dedican al ordenador y se ha puesto en relación con la nota media obtenida en el curso anterior. Los resultados son los siguientes:

minutos de ordenador	0 a 20 min.	20 a 40 min.	40 a 60 min.	60 a 80 min.	80 a 100 min.
Calificaciones	8	6	5	4	2
Frecuencia	2	8	6	3	1

- a) Halla la media y la desviación típica de las calificaciones. Interpreta los resultados obtenidos.
 b) Halla la mediana de las calificaciones. Explica brevemente el procedimiento empleado.
 c) Halla el coeficiente de correlación existente entre los tiempos de ordenador y las notas obtenidas e interpreta el resultado.

MATEMÁTICAS 1º BACHILLERATO INTERNACIONAL NIVEL MEDIO - Estadística Unidimensional en clase

- 1 Three positive integers $a, b,$ and $c,$ where $a < b < c,$ are such that their median is 11, their mean is 9 and their range is 10. Find the value of $a.$
- 2 La curva de frecuencias acumuladas que aparece a continuación muestra las calificaciones obtenidas en un examen por un grupo de 200 estudiantes.



- (a) Utilice la curva de frecuencias acumuladas para completar la siguiente tabla de frecuencias.

Calificación (x)	$0 \leq x < 20$	$20 \leq x < 40$	$40 \leq x < 60$	$60 \leq x < 80$	$80 \leq x < 100$
Número de estudiantes	22				20

- (b) El cuarenta por ciento de los estudiantes no supera el examen. Halle la calificación mínima para aprobar.

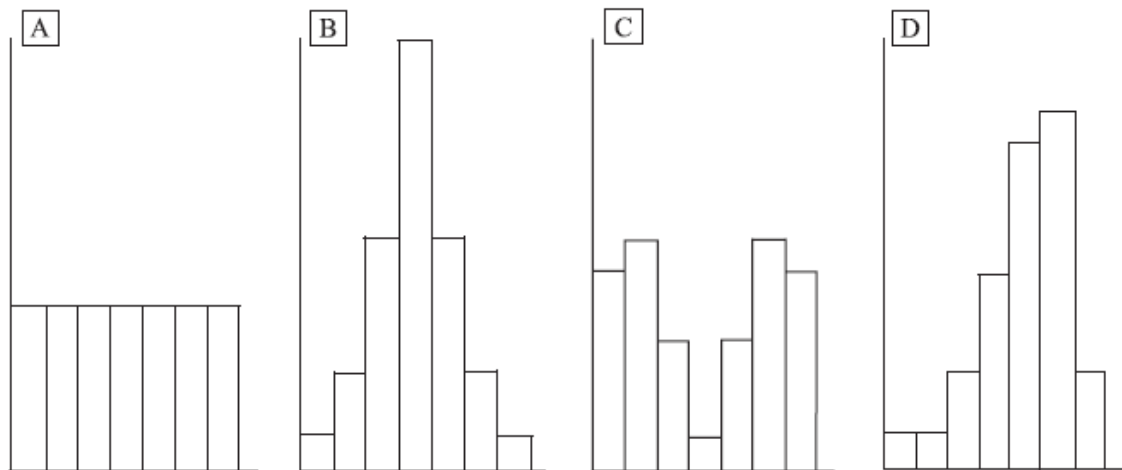
3 Cada uno de los 45 alumnos de una clase ha registrado el número total de minutos, x , que dedicó el lunes a realizar prácticas de laboratorio. Los resultados son $\sum x = 2\,230$.

(a) Halle el número medio de minutos dedicados el lunes por los estudiantes a las prácticas de laboratorio.

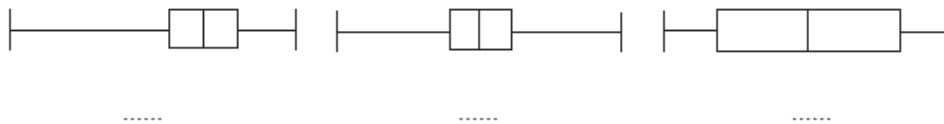
Dos nuevos alumnos se incorporan a la clase e informan que han dedicado respectivamente 37 minutos y 30 minutos.

(b) Calcule la nueva media que resulta al incluir a estos dos alumnos.

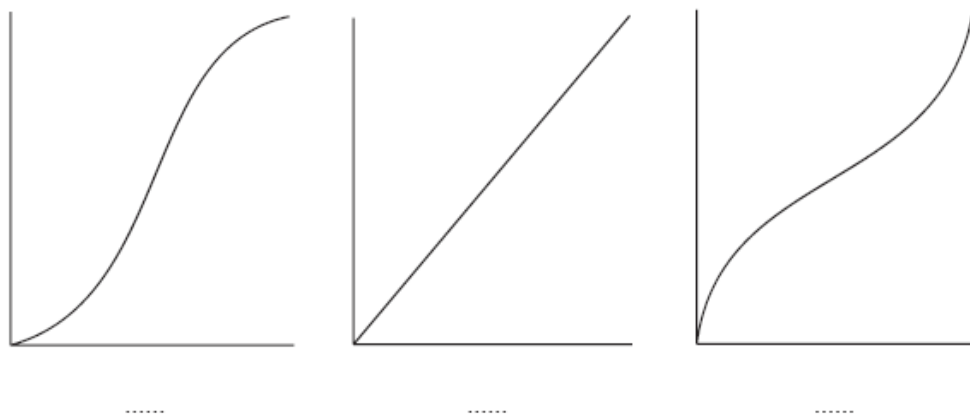
4 Las cuatro poblaciones A, B, C y D tienen el mismo tamaño y el mismo rango. Los siguientes son los histogramas de frecuencias para las cuatro poblaciones.



(a) Cada uno de los tres diagramas de caja y bigotes que aparecen debajo corresponde a una de las cuatro poblaciones. Debajo de cada diagrama, escriba la letra de la población a la que corresponde

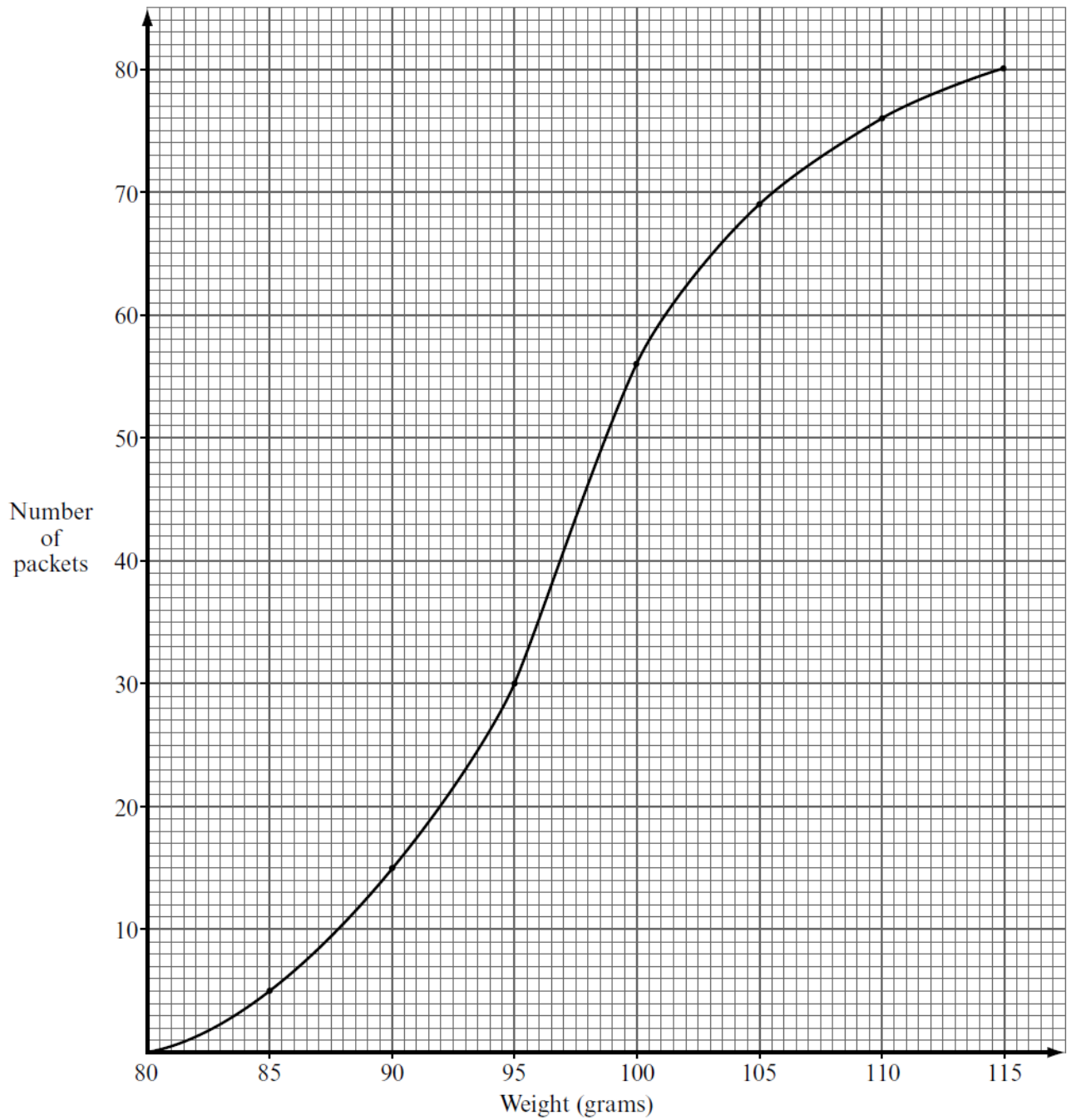


(b) Cada una de las tres gráficas de frecuencia acumulada que aparecen debajo corresponde a una de las cuatro poblaciones. Debajo de cada gráfica, escriba la letra de la población a la que corresponde.



5

- (c) A cumulative frequency graph of the distribution is shown below, with a scale 2 cm for 10 packets on the vertical axis and 2 cm for 5 grams on the horizontal axis.



Use the graph to estimate

- (i) the median;
- (ii) the upper quartile (that is, the third quartile).

Give your answers to the nearest gram.

- (d) Let W_1, W_2, \dots, W_{80} be the individual weights of the packets, and let \bar{W} be their mean. What is the value of the sum

$$(W_1 - \bar{W}) + (W_2 - \bar{W}) + (W_3 - \bar{W}) + \dots + (W_{79} - \bar{W}) + (W_{80} - \bar{W})?$$

- 6 The speeds in km h^{-1} of cars passing a point on a highway are recorded in the following table.

Speed ν	Number of cars
$\nu \leq 60$	0
$60 < \nu \leq 70$	7
$70 < \nu \leq 80$	25
$80 < \nu \leq 90$	63
$90 < \nu \leq 100$	70
$100 < \nu \leq 110$	71
$110 < \nu \leq 120$	39
$120 < \nu \leq 130$	20
$130 < \nu \leq 140$	5
$\nu > 140$	0

- (a) Calculate an estimate of the mean speed of the cars.
- (b) The following table gives some of the cumulative frequencies for the information above.

Speed ν	Cumulative frequency
$\nu \leq 60$	0
$\nu \leq 70$	7
$\nu \leq 80$	32
$\nu \leq 90$	95
$\nu \leq 100$	a
$\nu \leq 110$	236
$\nu \leq 120$	b
$\nu \leq 130$	295
$\nu \leq 140$	300

- (i) Write down the values of a and b .
- (ii) On graph paper, construct a cumulative frequency curve to represent this information. Use a scale of 1 cm for 10 km h^{-1} on the horizontal axis and a scale of 1 cm for 20 cars on the vertical axis.
- (c) Use your graph to determine
- the percentage of cars travelling at a speed in excess of 105 km h^{-1} ;
 - the speed which is exceeded by 15% of the cars.

7 La siguiente población se presenta en orden ascendente.

5; 6; 7; 7; 9; 9; r ; 10; s ; 13; 13; t

La mediana de la población es 9,5. El cuartil superior Q_3 es 13.

(a) Escriba el valor de

(i) r ;

(ii) s .

(b) La media de la población es 10. Halle el valor de t .

8 Sean a , b , c y d enteros tales que $a < b$, $b < c$ y $c = d$.

La moda de estos cuatro números es 11.

El rango de estos cuatro números es 8.

La media de estos cuatro números es 8.

Calcule el valor de cada uno de los enteros a , b , c , d .

9 In a school with 125 girls, each student is tested to see how many sit-up exercises (sit-ups) she can do in one minute. The results are given in the table below.

Number of sit-ups	Number of students	Cumulative number of students
15	11	11
16	21	32
17	33	p
18	q	99
19	18	117
20	8	125

(a) (i) Write down the value of p .

(ii) Find the value of q .

(b) Find the median number of sit-ups.

(c) Find the mean number of sit-ups.

10

A supermarket records the amount of money d spent by customers in their store during a busy period. The results are as follows:

Money in \$ (d)	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120	120 – 140
Number of customers (n)	24	16	22	40	18	10	4

- (a) Find an estimate for the mean amount of money spent by the customers, giving your answer to the nearest dollar (\$). [2 marks]
- (b) Copy and complete the following cumulative frequency table and use it to draw a cumulative frequency graph. Use a scale of 2 cm to represent \$ 20 on the horizontal axis, and 2 cm to represent 20 customers on the vertical axis. [5 marks]

Money in \$ (d)	< 20	< 40	< 60	< 80	< 100	< 120	< 140
Number of customers (n)	24	40					

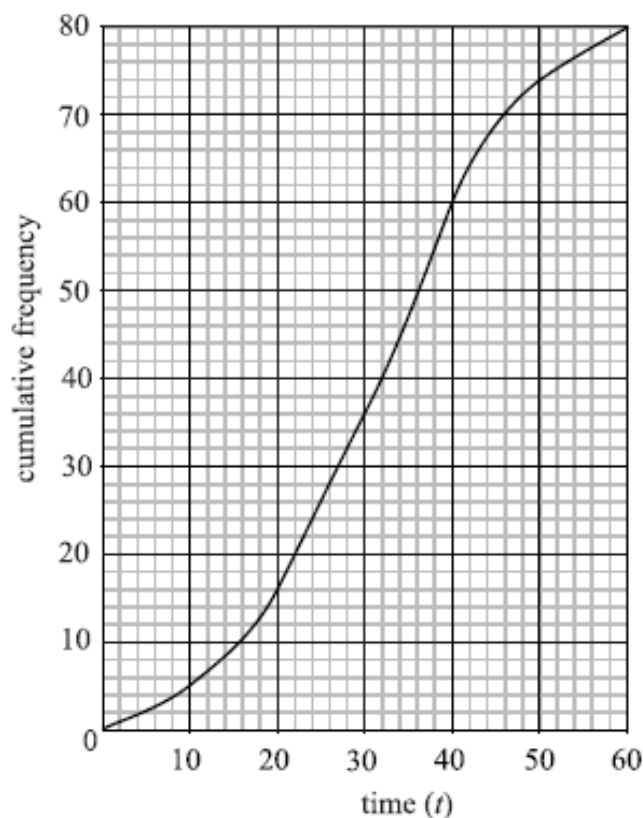
- (c) The time t (minutes), spent by customers in the store may be represented by the equation

$$t = 2d^{\frac{2}{3}} + 3.$$

- (i) Use this equation and your answer to part (a) to estimate the mean time in minutes spent by customers in the store. [3 marks]
- (ii) Use the equation and the cumulative frequency graph to estimate the number of customers who spent more than 37 minutes in the store. [5 marks]

11

The following is a cumulative frequency diagram for the time t , in minutes, taken by 80 students to complete a task.



- (a) Write down the median.
- (b) Find the interquartile range.
- (c) Complete the frequency table below.

Time (minutes)	Number of students
$0 \leq t < 10$	5
$10 \leq t < 20$	
$20 \leq t < 30$	20
$30 \leq t < 40$	24
$40 \leq t < 50$	
$50 \leq t < 60$	6

12 La tabla muestra los resultados de una encuesta sobre el nº de viajes que se han realizado durante el último año (Y) dependiendo de su edad (X).

a) Construye la tabla de frecuencias marginales absolutas y relativas de número de viajes.

b) Construye la tabla completa de frecuencias marginales de las edades

Y \ X		X		
		[20, 36)	[36, 52)	[52, 68)
[0, 3)	7	1	0	
[3, 6)	3	11	1	
[6, 9)	0	5	2	

13 Hemos hecho un estudio de las tasas de paro, en tantos por ciento, existentes en España a lo largo de 15 años (desde 1991 hasta 2005), representando con X la tasa de paro masculina y con Y la tasa de paro femenina. Hemos construido la siguiente tabla en la que la última fila contiene las sumas:

X	Y	X ²	Y ²	X·Y
12,1000	24,0650	146,4100	579,1242	291,1865
14,1800	25,7725	201,0724	664,2218	365,4541
18,7350	29,4050	351,0002	864,6540	550,9027
19,6225	31,6225	385,0425	999,9825	620,5125
18,0450	30,7925	325,6220	948,1781	555,6507
17,4224	29,5540	303,5399	873,4407	514,9020
15,8295	28,1587	250,5732	792,9137	445,7386
13,5658	26,5305	184,0299	703,8648	359,9057
10,9353	22,9319	119,5813	525,8721	250,7678
9,5539	20,4228	91,2779	417,0915	195,1185
7,5426	15,2197	56,8912	231,6399	114,7967
8,2284	16,3657	67,7071	267,8345	134,6636
8,3961	16,0048	70,4943	256,1522	134,3773
8,1543	15,0117	66,4925	225,3512	122,4098
7,0427	12,1571	49,6002	147,7942	85,6190
189,3536	344,0143	2.669,3347	8.498,1153	4.742,0056

a) Halla las medias, las desviaciones típicas y la covarianza redondeando los resultados con cuatro cifras decimales.

b) Halla el coeficiente de correlación e interpreta el resultado.

c) Halla la recta de regresión adecuada para estimar la tasa de paro femenina para cuando la masculina sea del 12 por ciento. Valora la bondad de la estimación.

14 Un fabricante de muebles hace sillas que vende a tiendas.

A lo largo de un período de seis semanas, el costo \$y de producción de x sillas es el que se da en la siguiente tabla.

	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6
Número de sillas x	22	40	32	28	46	44
Costo de producción \$y	3 200	4 600	3 800	3 700	5 100	5 000

(a) Halle la ecuación de la recta de regresión de y sobre x para estos datos.

(b) Las sillas se venden a \$120 cada una. Halle el número mínimo de sillas que debe vender la fábrica por semana para obtener ganancias.

EJEMPLO DE ESTUDIO ESTADÍSTICO DE UNA VARIABLE UNIDIMENSIONAL

Variable objeto de estudio

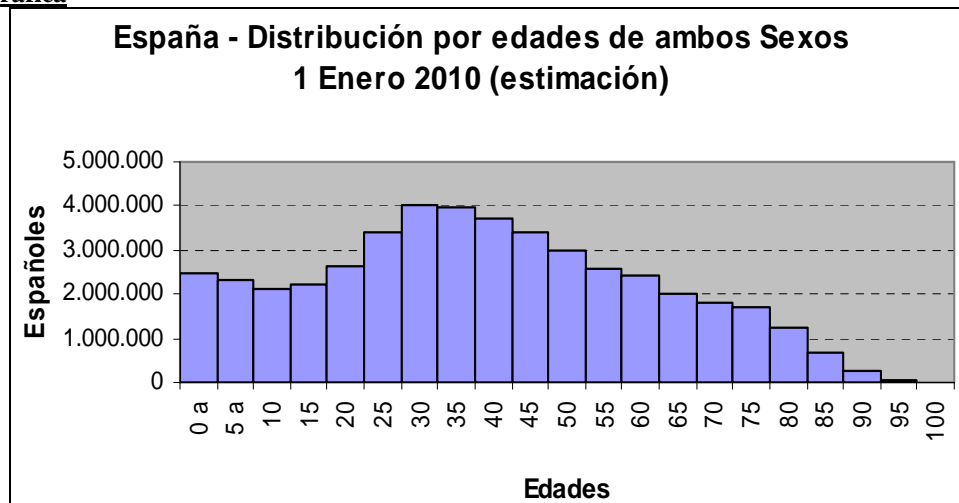
Vamos a estudiar la distribución de población de ambos sexos por segmentos de edad, de cuatro en cuatro años estimada para el 1 de enero de 2010. Están sacados de la web del Instituto Nacional de Estadística español.

(<http://www.ine.es/jaxiBD/tabla.do?per=01&type=db&divi=EPOB&idtab=3>)

Tabla con los datos

Total Nacional	x	f	F	f·x	f·x ²
0 a 4 años	2	2.461.227	2.461.227	4.922.454	9.844.908
5 a 9 años	7	2.297.895	4.759.122	16.085.265	112.596.855
10 a 14 años	12	2.113.106	6.872.228	25.357.272	304.287.264
15 a 19 años	17	2.240.229	9.112.457	38.083.893	647.426.181
20 a 24 años	22	2.619.905	11.732.362	57.637.910	1.268.034.020
25 a 29 años	27	3.379.985	15.112.347	91.259.595	2.464.009.065
30 a 34 años	32	4.030.195	19.142.542	128.966.240	4.126.919.680
35 a 39 años	37	3.945.775	23.088.317	145.993.675	5.401.765.975
40 a 44 años	42	3.711.892	26.800.209	155.899.464	6.547.777.488
45 a 49 años	47	3.423.294	30.223.503	160.894.818	7.562.056.446
50 a 54 años	52	3.013.424	33.236.927	156.698.048	8.148.298.496
55 a 59 años	57	2.595.058	35.831.985	147.918.306	8.431.343.442
60 a 64 años	62	2.411.466	38.243.451	149.510.892	9.269.675.304
65 a 69 años	67	2.015.458	40.258.909	135.035.686	9.047.390.962
70 a 74 años	72	1.786.020	42.044.929	128.593.440	9.258.727.680
75 a 79 años	77	1.706.879	43.751.808	131.429.683	10.120.085.591
80 a 84 años	82	1.226.794	44.978.602	100.597.108	8.248.962.856
85 a 89 años	87	691.867	45.670.469	60.192.429	5.236.741.323
90 a 94 años	92	248.325	45.918.794	22.845.900	2.101.822.800
95 a 99 años	97	63.071	45.981.865	6.117.887	593.435.039
100 y más años	102	7.152	45.989.017	729.504	74.409.408
		45.989.017		1.864.769.469	98.975.610.783

Representación gráfica



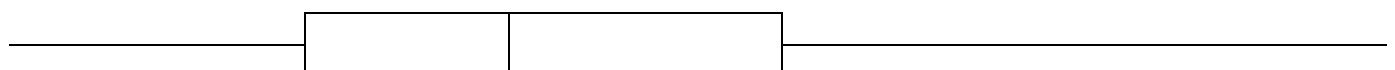
Medidas de centralización y dispersión

Aplicadas las fórmulas, hemos obtenido los siguientes resultados:

Vemos que la población está bastante envejecida, ya que el primer cuarto de la población ocupa aproximadamente 22 años, el segundo cuarto 15, el tercero 20 y el último 47 años. Se observa bien en el diagrama de cajas y bigotes.

Un porcentaje estimable de la población pertenece al intervalo: (18, 63)

Diagrama de Cajas y Bigotes



moda:	32
media:	40,5481
Varianza:	508,0053
desviación típica:	22,5390

Primer Cuartil:	22
Segundo Cuartil:	37
Tercer Cuartil:	57

EJEMPLO DE ESTUDIO ESTADÍSTICO DE UNA VARIABLE BIDIMENSIONAL

Variables objeto de estudio

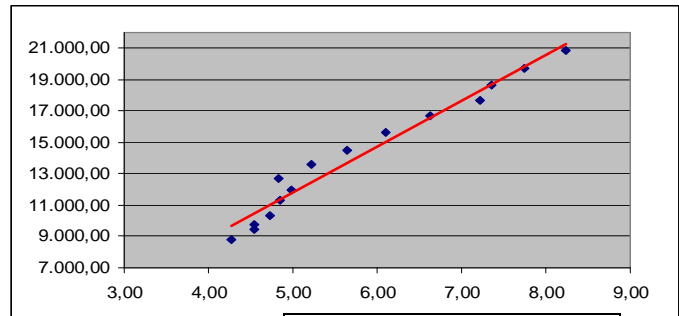
Vamos a estudiar la posible relación entre el crecimiento económico en España, concretamente el PIB medido en euros por habitante, y la natalidad medida con el número de nacimientos por 1.000 mujeres. Disponemos de datos de ambas variables desde el año 1991 hasta el 2005 y los hemos obtenido de la web del Instituto Nacional de Estadística español. (<http://www.ine.es/daco/daco42/sociales06/sociales.htm>)

Tabla con los datos

	nº nacimientos/1000 mujeres	PIB (€/hab)			
	x	y	x ²	y ²	x·y
1991	4,28	8.798,10	18,308	77.406.511,778	37.644,880
1992	4,54	9.444,55	20,571	89.199.582,147	42.836,376
1993	4,54	9.740,83	20,642	94.883.792,185	44.256,399
1994	4,74	10.332,14	22,422	106.753.090,237	48.924,406
1995	4,85	11.353,85	23,570	128.909.811,061	55.121,414
1996	4,99	12.002,66	24,905	144.063.894,161	59.899,300
1997	4,83	12.730,62	23,356	162.068.697,274	61.524,286
1998	5,23	13.581,69	27,337	184.462.365,617	71.012,176
1999	5,64	14.524,98	31,831	210.974.949,376	81.948,044
2000	6,11	15.653,20	37,327	245.022.685,310	95.634,092
2001	6,62	16.715,47	43,858	279.406.834,138	110.699,296
2002	7,22	17.650,33	52,077	311.534.090,172	127.372,078
2003	7,36	18.629,66	54,156	347.064.275,451	137.097,637
2004	7,74	19.678,42	59,908	387.240.023,131	152.310,933
2005	8,24	20.863,89	67,898	435.301.845,973	171.918,442
	86,93	211.700,38	528,165	3.204.292.448,010	1.298.199,759

Representación gráfica

La nube de puntos y la recta de regresión y *sobre x* están representadas en el siguiente diagrama:



Medidas de centralización, dispersión y regresión

Aplicadas las fórmulas, hemos obtenido los resultados de la tabla adjunta

El coeficiente de correlación lineal tiene un valor muy cercano a 1, por lo que se puede calificar de **positiva y muy fuerte** la relación existente entre el PIB y la natalidad. Es decir, que al ir aumentando la riqueza en España también va aumentando la natalidad.

media_x :	5,795
des_tip_x :	1,275
media_y :	14.113,359
des_tip_y :	3.799,027
covar :	4.755,921
coef_corr :	0,982

Rectas de regresión

Si escribimos la ecuación de la recta de regresión y *sobre x* en la forma: $y = m \cdot x + n$ los coeficientes que hemos hallado son:

Recta de regresión y sobre x
m: 2925,1674
n: 10383,4973

Si escribimos la ecuación de la recta de regresión *x sobre y* en la forma: $x = m \cdot y + n$ los coeficientes que hemos hallado son:

Recta de regresión x sobre y
m: 0,0003
n: 1,1446

Estimación de resultados

Vamos a estimar el nº de nacidos por 1.000 mujeres para un PIB de 25.000€ por habitante. Sustituimos $y = 25.000$ en la recta de regresión *x sobre y* obteniendo $x = 9,38$. Es decir, que para dicho PIB habría 9,38 nacidos por 1.000 mujeres en un año.