

## 1 Problema 1

El número de hijos de las familias de una determinada barriada de una ciudad es una variable estadística de la que se conocen los siguientes datos:

$x_i$	$n_i$	$N_i$	$f_i$
0	80	320	0.16
1	110		0.18
2			
3			
4	40		
5			
6	20		

$n_i$  : frecuencias absolutas  
 $N_i$  : frecuencias absolutas acumuladas  
 $f_i$  : frecuencias relativas

1. Completar la tabla de frecuencias.
2. Representar la distribución mediante un diagrama de barras y la curva de distribución.
3. Promediar los valores de la variable mediante diferentes medidas. Interpretarlas.

## 2 Problema 2

La puntuación obtenida por 50 personas que se presentaron a una prueba de selección, sumadas las puntuaciones de los distintos tests, fueron:

174, 185, 166, 176, 145, 166, 191, 175, 158, 156, 156, 187, 162, 172, 197, 181, 151,  
161, 183, 172, 162, 147, 178, 176, 141, 170, 171, 158, 184, 173, 169, 162, 172, 181,  
187, 177, 164, 171, 193, 183, 173, 179, 188, 179, 167, 178, 180, 168, 148, 173.

1. Agrupar los datos en intervalos de amplitud 5 desde 140 a 200 y dar la tabla de frecuencias.
2. Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.

Ejercicio 2. La puntuación obtenida por 50 personas que se presentaron a una prueba de selección, sumadas las puntuaciones de los distintos tests fueron:

174, 185, 166, 176, 145, 166, 191, 175, 158, 156, 156, 187, 162, 172, 197, 181, 151, 161, 183, 172,  
162, 147, 178, 176, 141, 170, 171, 158, 184, 173, 169, 162, 172, 181, 187, 177, 164, 171, 193, 183,  
173, 179, 188, 179, 188, 179, 167, 178, 180, 168, 148, 173.

POBLACIÓN: Las personas que se presentaron a la prueba de selección

TAMAÑO: 50

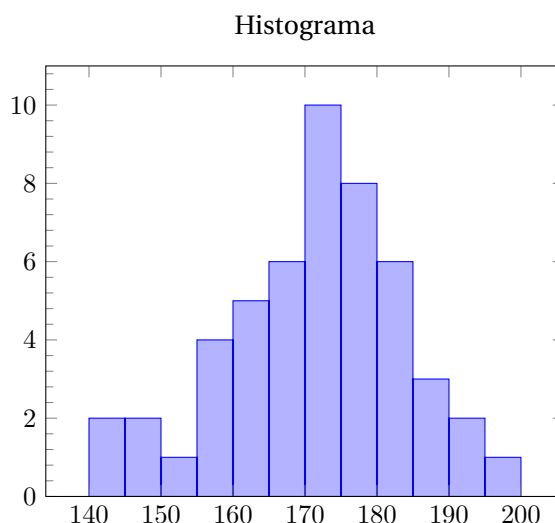
MODALIDADES: Los intervalos que contienen las notas que han obtenido las personas en la prueba

- a) Agrupar los datos en intervalos de amplitud 5 desde 140 a 200 y dar la tabla de frecuencias.

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
(140-145]	2	2	0,04	0,04
(145-150]	2	4	0,04	0,08
(150-155]	1	5	0,02	0,1
(155-160]	4	9	0,08	0,18
(160-165]	5	14	0,1	0,28
(165-170]	6	20	0,12	0,4
(170-175]	10	30	0,2	0,6
(175-180]	8	38	0,16	0,76
(180-185]	6	44	0,12	0,88
(185-190]	3	47	0,06	0,94
(190-195]	2	49	0,04	0,98
(195-200]	1	50	0,02	1

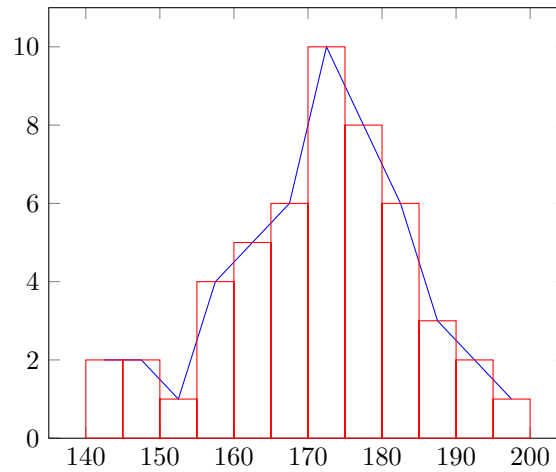
Establecemos los intervalos de amplitud 5, empezando con 140 y terminando con 200. Para cada intervalo, contamos cuantas personas han obtenido una puntuación que esté dentro de dicho intervalo. Estas son las frecuencias absolutas ( $n_i$ ). Para la frecuencia absoluta acumulada ( $N_i$ ), sumamos la frecuencia absoluta de la modalidad que estamos analizando a la frecuencia absoluta acumulada de la modalidad anterior. O lo que es lo mismo, sumamos las frecuencias absolutas de todas las modalidades hasta la que estamos analizando:  $N_i = n_1 + n_2 + \dots + n_i$ . Para frecuencia relativa, dividimos cada  $n_i$  entre n, es decir, entre  $N_{12}$ , la frecuencia absoluta acumulada de la última modalidad. Para la frecuencia relativa acumulada hacemos lo mismo que para la absoluta acumulada pero con la frecuencia relativa: sumamos las frecuencias relativas de todas las modalidades hasta la que estamos analizando:  $F_i = f_1 + f_2 + \dots + f_i$ .

b) Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.



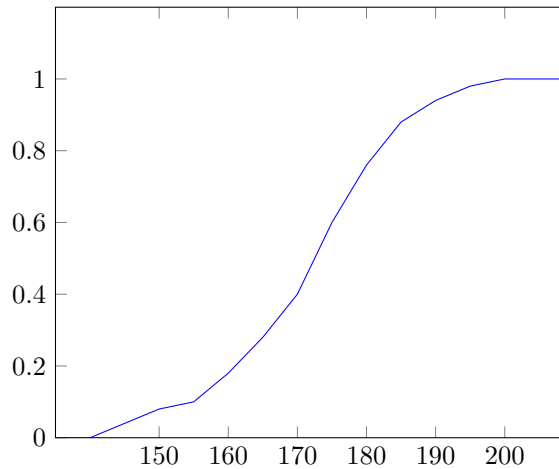
Para el histograma, en el eje x tenemos las modalidades ( $x_i$ ) y en el eje y un múltiplo de  $h_i$  ( $h_i = \frac{f_i}{a}$ ). En este caso, las amplitudes de todos los intervalos son iguales: 5.

Poligonal de frecuencias



Para el poligonal de frecuencias, los ejes son los mismos que para el histograma. En este caso, unimos los puntos que corresponden a las marcas de los intervalos en el histograma. Para hallar las marcas de los intervalos:  $c_i = \frac{e_{i-1} + e_i}{2}$ .

Curva de distribución



En la curva de distribución, en el eje x tenemos los extremos superiores de los intervalos ( $e_i$ ) y en el eje y,  $F(e_i)$ . Pintamos la curva que une los puntos tal que  $F(e_i) = \sum_{j=1}^i f_j = F_i$ . En este caso tenemos que la curva es continua.  $F(e_i)$  es 0 para valores menores que  $e_1$  y 1 para valores mayores que el último intervalo  $e_k$ .

### 3 Problema 3

La distribución de la renta familiar en el año 2003 por comunidades autónomas se recoge en la siguiente tabla:

$I_i$	$n_i$	$N_i$	$f_i$	$F_i$	$c_i$	$a_i$	$h_i$
(8300, 9300] , 10200]	2	5		10/18		1100	
			2/18		12000		
	4	18					0.005/18 0.002/18

$n_i$  : frecuencias absolutas  
 $N_i$  : frec. absolutas acumuladas  
 $f_i$  : frecuencias relativas  
 $F_i$  : frec. relativas acumuladas  
 $c_i$  : marcas de clase  
 $a_i$  : amplitudes  
 $h_i$  : densidades de frecuencia

1. Completar la tabla.
2. Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.
3. ¿Cuántas comunidades presentan una renta menor o igual que 12700 euros? ¿Y cuántas superior a 11300 euros?

## 4 Problema 4

En una determinada empresa se realiza un estudio sobre la calidad de su producción. La distribución siguiente informa sobre el número de piezas defectuosas encontradas en 100 cajas examinadas con 50 unidades cada una de ellas:

Nº piezas defectuosas	0	1	2	3	4	5	6	7	8	9	10
Nº de cajas	6	9	10	11	14	16	16	9	4	3	2

1. Calcular el número medio de piezas defectuosas por caja.
2. ¿Cuántas piezas defectuosas se encuentran más frecuentemente en las cajas examinadas?
3. ¿Cuál es el número mediano de piezas defectuosas por caja?
4. Calcular los cuartiles de la distribución. Interpretarlos.
5. Calcular los deciles de orden 3 y 7. Interpretarlos.
6. Cuantificar la dispersión de la distribución utilizando diferentes medidas, interpretando los resultados y señalando las ventajas e inconvenientes de cada una.

En los lotes de 50 cajas de los productos producidos por una determinada empresa producidos durante un determinado período, con un total de  $n = 100$  cajas, se ha observado el número de piezas defectuosas, presentando ésta un total de  $k = 11$  modalidades, cuya distribución de frecuencias viene representado en la siguiente tabla:

$x_i$	$n_i$	$N_i$	$x_i n_i$	$n_i  x_i - \bar{x} $	$n_i  x_i - Me $	$n_i (x_i - \bar{x})^2$
0	6	6	0	26,16	27,00	114,06
1	9	15	9	30,24	31,50	101,61
2	10	25	20	23,60	25,00	55,70
3	11	36	33	14,96	16,50	20,35
4	14	50	56	5,04	7,00	1,81
5	16	66	80	10,24	8,00	6,55
6	16	82	96	26,24	24,00	43,03
7	9	91	63	23,76	22,50	62,73
8	4	95	32	14,56	14,00	53,00
9	3	98	27	13,92	13,50	64,59
10	2	100	20	11,28	11,00	63,62

a) La variable estadística observada es de carácter discreta, presentando un total de 11 modalidades: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10. En este caso, la media más significativa es la aritmética, cuya expresión viene determinado por:

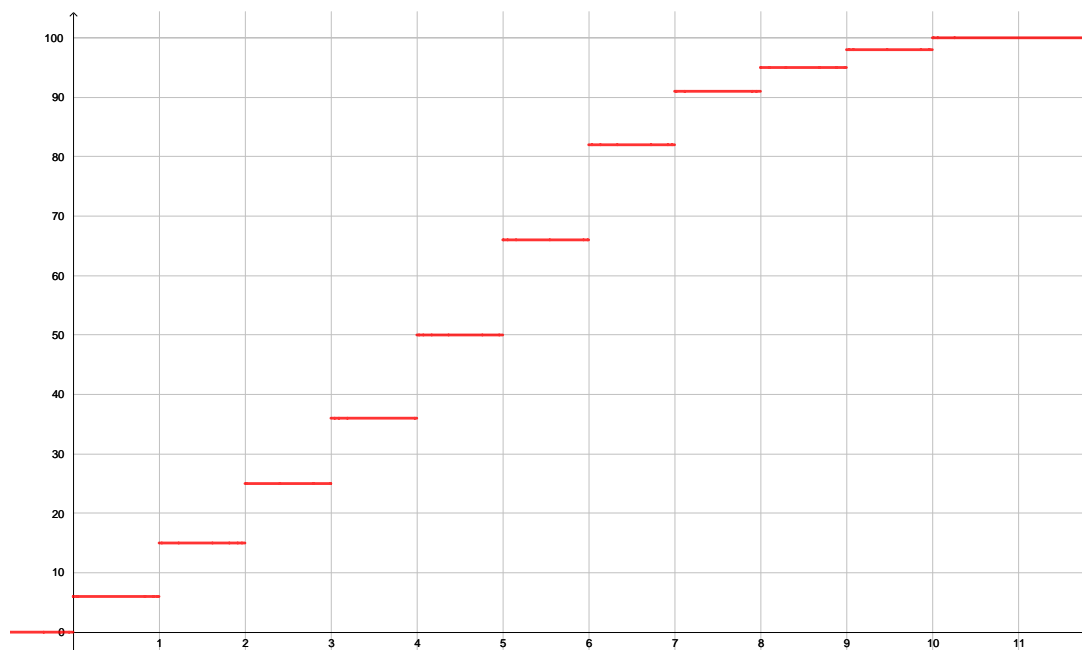
$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i$$

De la tabla podemos obtener, teniendo también en consideración que  $n = 100$  cajas, que  $\bar{x} = 4,36$  piezas.

b) La medida estadística que proporciona información sobre el número de piezas defectuosas que más habitualmente se puede esperar de una caja es la moda. Se define como  $Mo = \max(x_i)$ , con  $i = 1, 2, 3, \dots, k$ . En este caso, se presentan dos modalidades que presentan la misma frecuencia absoluta. Por tanto, podemos afirmar que hay dos valores de moda en esta distribución, siendo más frecuente encontrar un total de 5 y 6 piezas defectuosas por lote examinado.

c) La mediana se define como un valor que divide a una determinada población en dos subgrupos con el mismo número de individuos: una de ellas cuyos individuos pertenecen a una modalidad cuya frecuencia absoluta es inferior a la mediana y otra con frecuencias absolutas superiores.

Para determinar la mediana en esta distribución de frecuencias, representamos la curva de distribución de la variable estadística:



Observamos que  $\frac{n}{2} = 50$ , coincidiendo justamente con el valor  $F(x_5) = n/2$ . Conviene pues tomar la media aritmética entre  $x_5$  y  $x_6$ , de donde  $Me = 4,5$  piezas defectuosas.

d) El cuartil es una medida que permite caracterizar el porcentaje de individuos de la población cuya variable estadística observada es inferior al 25%, 50%, 75% de toda la población, respectivamente. Su cálculo es análogo al de la media (únicamente hay que sustituir  $\frac{n}{2}$  por las fracciones correspondientes). De esta forma, deducimos las siguientes medidas:

- $Q_1 = 2,5$  piezas. Representa que un 25% de los lotes analizados poseen una cantidad de productos defectuosos inferiores o iguales a 2,5 piezas.
- $Q_2 = 4,5$  piezas. Representa que un 50% de los lotes analizados poseen una cantidad de productos defectuosos inferiores o iguales a 4,5 piezas.

- $Q_3 = 6$  piezas. Representa que un 75% de los lotes analizados poseen una cantidad de productos defectuosos inferiores o iguales a 6 piezas.

e) El decil tiene un significado análogo a la de los cuartiles. Por tanto, el decil es una medida que permite caracterizar el porcentaje de individuos de la población cuya variable estadística observada es inferior al 10%, 20%, ..., 90% de toda la población. Su cálculo es, nuevamente, análogo a los casos anteriores:

- $D_3 = 3$  piezas. Representa que un 30% de los lotes analizados poseen una cantidad de productos defectuosos inferiores o iguales a 3 piezas.
- $D_7 = 6$  piezas. Representa que un 70% de los lotes analizados poseen una cantidad de productos defectuosos inferiores o iguales a 6 piezas.

f) Algunas medidas de dispersión que podemos tomar sobre la variable estadística observada son las siguientes:

- Recorrido. Es la magnitud que determina la anchura total de la muestra tomada (esto es, toma la diferencia entre el valor más alto tomado de la variable y el menor valor observado). Su ventaja es que tiene un significado concreto, pero su inconveniente es que varía mucho con fluctuaciones muestrales. En este caso, su valor es  $R = 10$  piezas. Esto quiere decir que hay una variación de 10 productos defectuosos entre los datos de la muestra.
- Recorrido intercuartílico. Es la magnitud que indica la longitud del intervalo que contiene al 50% central de los datos observados. Su virtud es que no varía mucho con fluctuaciones muestrales, pero su inconveniente es que no toma en consideración todos los datos de la población observada. En este caso, su valor es  $R_I = 3,5$  piezas. Ello quiere decir que, del 50% central de la muestra, hay, a lo sumo, una diferencia de 3,5 piezas defectuosas entre ellos.
- Desviación absoluta media respecto a la media aritmética. Indica cómo están los datos distribuidos en función del valor promedio (la media aritmética). Su ventaja es que tiene un significado preciso, pero presenta el inconveniente de que varía mucho en función de fluctuaciones estadísticas (puesto que en su cálculo interviene la media aritmética). Su valor es, en este caso,  $D_{\bar{x}} = 2$  piezas. Representa que los individuos de la población muestran una diferencia de media de 2 piezas defectuosas respecto a la media aritmética.
- Desviación absoluta media respecto a la mediana. Es la medida que representa la media de las distancias de los valores presentados durante la muestra respecto al 50%. Presenta la ventaja de que tiene un significado preciso, pero no es fácil de calcular. Su valor es  $D_{Me} = 2$  piezas, en este caso. Esto significa que los individuos presentan, respecto a la mediana, una diferencia de 2 piezas defectuosas.
- Desviación típica. Es una medida que proporciona información sobre el margen óptimo de diferencia entre los valores medidos de la variable respecto de la media aritmética. Su principal ventaja es que fluctúa poco con variaciones extremas de la muestra, pero tiene el inconveniente de que requiere gran capacidad de cómputo para poder calcularse. Su valor es, en este caso,  $\sigma_x = 2,42$  piezas. Su significado es que, respecto a la media, la diferencia óptima que se debería encontrar en la distribución es de 2,42 piezas.
- Recorrido relativo. Informa sobre el recorrido de la variable respecto de su media aritmética. Su ventaja es que es fácil de calcular, pero tiene la desventaja de que varía mucho con las fluctuaciones estadísticas. En este caso, su valor es  $R_R = 0,8$ . Informa que la amplitud de las modalidades tomadas oscila en torno a 0,8 veces la media aritmética.
- Coeficiente de variación. Representa la desviación típica de la muestra en función de la media aritmética. Su principal virtud es que compara el grado de dispersión entre diferentes variables estadísticas, pero tiene el inconveniente de que no es válida para todas las variables estadísticas (no sería válida para aquellas con  $\bar{x} = 0$  u). Su valor es, en este caso,

$C.V.(X) = 0,55$ . Se trata de una medida adimensional. Sin otra magnitud respecto al cual comparar, esta medida no proporciona información relevante.

## 5 Problema 5

Dadas las siguientes distribuciones:

4pt

$I_i^{(1)}$	(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]
$n_i^{(1)}$	12	13	11	8	6

4pt

$I_i^{(2)}$	(0, 1]	(1, 3]	(3, 6]	(6, 10]	(10, 12]
$n_i^{(2)}$	1	6	7	12	2

Calcular para cada una de ellas:

1. Medias aritmética, armónica y geométrica.
2. El valor más frecuente.
3. El valor superado por el 50 % de las observaciones.
4. Recorrido, recorrido intercuartílico y desviación típica. Interpretarlos. ¿Qué distribución es más homogénea?

Escribamos las tablas con los datos que nos pueden interesar para la resolución del ejercicio:

$I_i^{(1)}$	$c_i^{(1)}$	$n_i^{(1)}$	$N_i^{(1)}$
(0, 1]	0.5	12	12
(1, 2]	1.5	13	25
(2, 3]	2.5	11	36
(3, 4]	3.5	8	44
(4, 5]	4.5	6	50

$I_i^{(2)}$	$c_i^{(2)}$	$n_i^{(2)}$	$N_i^{(2)}$
(0, 1]	0.5	1	1
(1, 3]	2	6	7
(3, 6]	4.5	7	14
(6, 10]	8	12	26
(10, 12]	11	2	28

### Apartado a)

La media aritmética de una variable es la suma de sus valores entre en número total de observaciones. Como los datos están organizados en intervalos de clase, para calcular la media aritmética vamos a suponer que todos los datos de un intervalo son idénticos a la marca de clase de cada intervalo. Por tanto, la media la calculamos de la siguiente manera:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i c_i$$

Para la primera distribución,  $n = 50$ , luego la media aritmética es 2.082.

Para la segunda,  $n = 28$ , luego la media aritmética será 5.786.

La media armónica se usa para promediar datos de magnitudes relativas. La definimos como la inversa de la media aritmética de los valores inversos de la variable (en nuestro caso, usamos las marcas de clase):

$$H = \frac{n}{\frac{n_1}{c_1} + \frac{n_2}{c_2} + \dots + \frac{n_k}{c_k}} = \frac{n}{\sum_{i=1}^k \frac{n_i}{c_i}}$$

Para la primera distribución:  $H = 1.223$ .

Para la segunda distribución:  $H = 3.399$ .

Finalmente la media geométrica se usa cuando se desea promediar datos de una variable que tiene efectos multiplicativos acumulativos en la evolución de una determinada característica con un valor inicial fijo.

Es la raíz  $n$ -ésima del producto de los  $n$  valores (o marcas de clase) de la distribución:

$$G = \sqrt[n]{\prod_{i=1}^k c_i^{n_i}}$$

Para no perder precisión en el resultado, la calcularemos sabiendo que el logaritmo de la media geométrica es la media aritmética de los logaritmos de los valores de la variable:

$$\log G = \log \sqrt[n]{\prod_{i=1}^k c_i^{n_i}} = \frac{1}{n} \sum_{i=1}^k n_i \log c_i$$

Para la primera distribución:  $G = 1.685$ .

Para la segunda distribución:  $G = 1.562$ .

#### **Apartado b)**

Se nos pide calcular la moda de las distribuciones, es decir, el valor de mayor frecuencia:

En el caso primero, este es el intervalo  $(1,2]$ , luego tomamos su marca de clase como la moda: 1.5.

En el segundo caso la moda proviene del intervalo  $(6, 10]$  luego será 8.

#### **Apartado c)**

Si miramos las frecuencias absolutas acumuladas tendremos que tomar la parte inferior del intervalo donde se encuentre  $n/2$ :

Para la primera distribución:  $n/2 = 25$  luego el 50% de la población supera el valor 2.

Para la segunda distribución:  $n/2 = 14$  luego el 50% de la población supera el valor 6.

#### **Apartado d)**

## **6 Problema 6**

Un móvil efectúa un recorrido de 100 km en dos sentidos. En uno va a una velocidad constante de  $V_1=60$  km/h y en el otro va a una velocidad constante de  $V_2=70$  km/h. Calcular la velocidad media del recorrido.

Consideremos un cuerpo que se mueve entre dos puntos  $A$  y  $B$ , alejados una distancia  $d > 0$ , con una velocidad  $v_1$  para ir desde  $A$  hasta  $B$  y una velocidad  $v_2$  para ir desde  $B$  hasta  $A$ . La velocidad de un cuerpo es una magnitud física que se define como la rapidez con la que varía la



posición de un cuerpo. Su expresión matemática viene dado por la siguiente expresión:  $v = \frac{\Delta s}{\Delta t}$ , donde  $v$  es la velocidad del cuerpo,  $s$  es la distancia recorrida y  $t$  es el tiempo transcurrido.

La velocidad media del cuerpo viene determinado por la expresión  $v_{media} = \frac{\Delta s}{\Delta t}$ . Para determinar dicha velocidad, tengamos en cuenta que el cuerpo recorre dos veces la distancia  $d$ , una para ir desde  $A$  hasta  $B$  y otra para ir desde  $B$  hasta  $A$ . Por tanto, sabemos que  $s = 2d$ . Por otra parte, sabemos que el tiempo invertido sería  $t = t_1 + t_2$ , donde  $t_1$  es el tiempo invertido para ir desde  $A$  a  $B$  y  $t_2$  es el tiempo invertido para ir desde  $B$  hasta  $A$ . Por tanto, deducimos que  $v_{media} = \frac{2d}{t_1 + t_2}$ . Obedeciendo a las expresiones  $t_1 = \frac{d}{v_1}$  y  $t_2 = \frac{d}{v_2}$ , además de dividir numerador y denominador por la distancia  $d > 0$ , obtenemos:

$$v_{media} = \frac{2d}{\frac{d}{v_1} + \frac{d}{v_2}} \Rightarrow v_{media} = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}}$$

Por tanto, tenemos que la velocidad media del móvil es la **media armónica** de las dos velocidades. Tomando  $v_1 = 60 \frac{km}{h}$  y  $v_2 = 70 \frac{km}{h}$ , tenemos que la velocidad media del recorrido sería:

$$v_{media} = 64,61 \frac{km}{h} \quad (1)$$

## 7 Problema 7

Las acciones de una empresa han producido los siguientes rendimientos netos anuales:

Año	Rentabilidad
1994	12%
1995	10%
1996	7%
1997	6%
1998	5%

Obtener el rendimiento neto medio en esos cinco años.

## 8 Problema 8

Un profesor califica a sus alumnos según el criterio siguiente: 40% de suspensos, 30% de aprobados, 15% notables, 10% sobresalientes y 5% de matrículas. Las notas obtenidas son las siguientes:

(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]	(5, 6]	(6, 7]	(7, 8]	(8, 9]	(9, 10]
34	74	56	81	94	70	41	28	16	4

Calcular las notas máximas para obtener cada una de las calificaciones.

## 9 Problema 9

Se ha medido la altura de 110 jóvenes, obteniendo:

Altura	(1.55, 1.60]	(1.60, 1.70]	(1.70, 1.80]	(1.80, 1.90]	(1.90, 2.00]
Nº jóvenes	18	31	24	20	17

1. Si se consideran bajos el 3% de los individuos de menor altura, ¿cuál es la altura máxima que pueden alcanzar?
2. Si se consideran altos el 18% de los individuos de mayor altura, ¿cuál es su altura mínima?
3. ¿Qué altura es superada sólo por 1/4 de los jóvenes?

4. Calcular el número de jóvenes cuya altura es superior a 1.75.
5. Calcular la altura máxima de los 11 jóvenes más bajos.
6. Calcular la altura mínima de los 11 jóvenes más altos.

## 10 Problema 10

Realizando una prueba para el estudio del cáncer a 150 personas se obtuvo la siguiente tabla según la edad de los enfermos:

Edad	(10, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 90]
Nº enfermos	15	22	48	40	25

1. Calcular la edad más común de los individuos estudiados.
2. Calcular la edad mínima y máxima del 30% central de los individuos.
3. Calcular el recorrido intercuartílico y la desviación típica.
4. Calcular e interpretar los valores de los coeficientes de asimetría y curtosis.