

Relación de ejercicios 1 EDIP

Carlos García, Bora Goker, Javier Gómez,
Ana Graciani, J.Alberto Hoces

2020/2021

Ejercicio 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

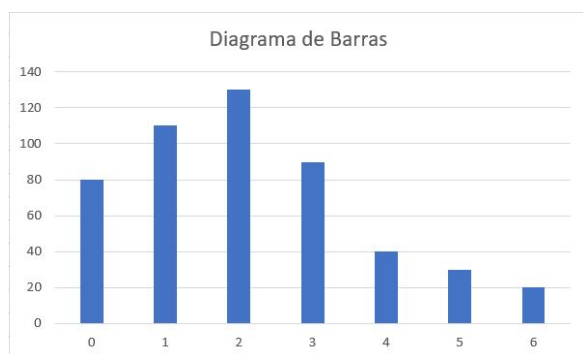
f_i : frecuencias relativas

a) Completar la tabla de frecuencias.

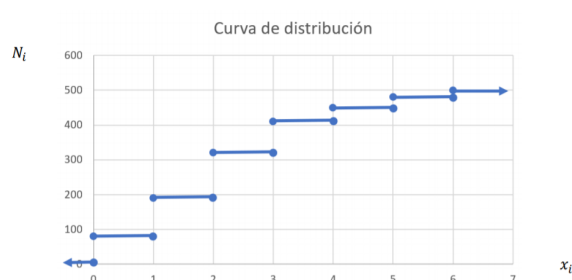
x_i	n_i	N_i	f_i
0	80	80	0,16
1	110	190	0,22
2	130	320	0,26
3	90	410	0,18
4	40	450	0,08
5	30	480	0,06
6	20	500	0,04

$$f_i = \frac{n_i}{n}$$

b) Representar la distribución mediante un diagrama de barras y la curva de distribución.



(a) Diagrama de barras



(b) Curva de distribución

c) Promediar los valores de la variable mediante diferentes medidas. Interpretálas.

i) Media: Puesto que se pueden sumar el número de hijos y la media es única calculamos su media aritmética.

$$\bar{x} = \frac{0 \cdot 80 + 1 \cdot 110 + 2 \cdot 130 + 3 \cdot 90 + 4 \cdot 40 + 5 \cdot 30 + 6 \cdot 20}{500} = 2,14 \approx 2 \text{ hijos}$$

ii) Mediana:

$$\frac{n}{2} = \frac{500}{2} = 250, \quad N_2 < 250 < N_3 \quad \text{Mediana} = 2 \text{ hijos}$$

iii) Moda:

$$n_3 \geq n_j \quad \forall j \neq 3 \quad \text{Moda} = 2 \text{ hijos}$$

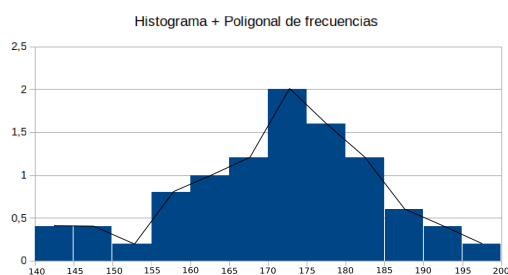
Ejercicio 2. La puntuación obtenida por 50 que presentaron a una prueba de selección, sumadas las puntuación de distintos test, fueron:

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.

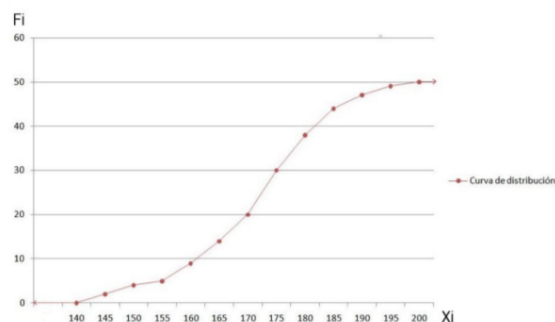
a) Agrupar los datos en intervalos de amplitud 5 desde 140 a 200 y dar la tabla de frecuencias.

I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
[140, 145]	2	2	2/50	2/50	142,5	5	2/5
(145, 150]	2	4	2/50	4/50	147,5	5	2/5
(150, 155]	1	5	1/50	5/50	152,5	5	1/5
(155, 160]	4	9	4/50	9/50	157,5	5	4/5
(160, 165]	5	14	5/50	14/50	162,5	5	5/5
(165, 170]	6	20	6/50	20/50	167,5	5	6/5
(170, 175]	10	30	10/50	30/50	172,5	5	10/5
(175, 180]	8	38	8/50	38/50	177,5	5	8/5
(180, 185]	6	44	6/50	44/50	182,5	5	6/5
(185, 190]	3	47	3/50	47/50	187,5	5	3/5
(190, 195]	2	49	2/50	49/50	192,5	5	2/5
(195, 200]	1	50	1/50	50/50	197,5	5	1/5

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



(a) Histograma



(b) Curva de distribución

Ejercicio 3. La distribución de la renta familiar en el año 2003 por comunidades autónomas se recoge en la siguiente tabla:

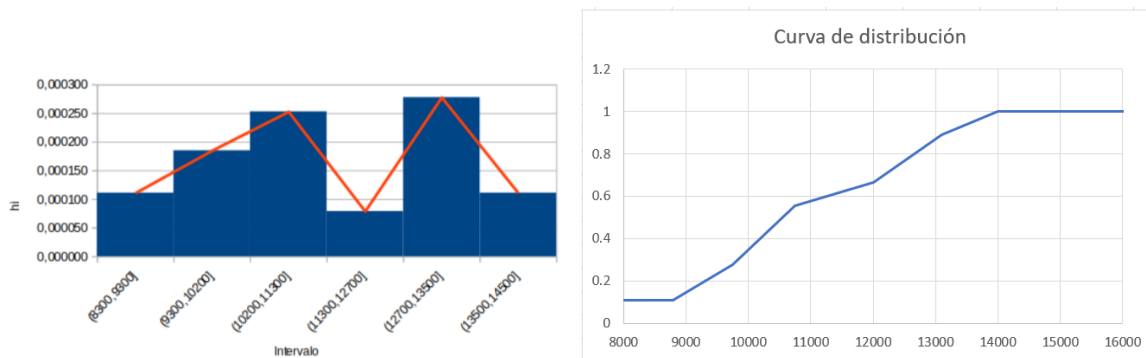
I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
[8300, 9300] , 10200]	2	5		10/18		1100	
			2/18		1200		
	4	18					0,005/18 0,002/18

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

a) Completar la tabla.

I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
[8300, 9300]	2	2	0,11	0,11	8800	1000	0,002/18
(9300 , 10200]	3	5	3/18	5/18	9750	900	0,003/18
(10200, 11300]	5	10	5/18	10/18	10750	1100	0,0045/18
(11300, 12700]	2	12	2/18	12/18	12000	1400	0,0014/18
(12700, 13500]	4	16	4/18	16/18	13100	800	0,005/18
(13500, 14500]	2	18	2/18	18/18	14000	1000	0,002/18

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



(a) Histograma y poligonal de frecuencias

(b) Curva de distribución

- c) ¿Cuántas comunidades presentan una renta menor o igual que 12700 euros? ¿Y cuántas superior a 11300 euros?

12 comunidades presentan una renta menor o igual a 12700€. $N_4 = 12$

$x_4 \rightarrow (11300, 12700]$

8 comunidades presentan una renta superior a 11300 euros. $n_4 + n_5 + n_6 = 8$

Ejercicio 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

- a) Calcular el número medio de piezas defectuosas por caja.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{1}{100} \cdot 436 = 4,36 \text{ piezas defectuosas.}$$

x_i	n_i	$n_i x_i$	N_i	$ x_i - \bar{x} n_i$	$ x_i - Me n_i$	$x_i^2 n_i$
0	6	0	6	26,16	27	0
1	9	9	15	30,24	31,5	9
2	10	20	25	23,6	25	40
3	11	33	36	14,96	16,5	99
4	14	56	50	5,04	7	224
5	16	80	66	10,24	8	400
6	16	96	82	26,24	24	576
7	9	63	91	23,76	22,5	441
8	4	32	95	14,56	14	256
9	3	27	98	13,92	13,5	243
10	2	20	100	11,28	11	200

- b) ¿Cuántas piezas defectuosas se encuentran más frecuentemente en las cajas examinadas?

La moda de una distribución es el valor de mayor frecuencia (el más repetido). En este caso :

$n_6 = n_7 = 16$ $M_o = x_6 = 5$ piezas defectuosas $M_o = x_7 = 6$ piezas defectuosas

Podemos observar que nos hayamos ante una distribución bimodal.

- c) ¿Cuál es el número mediano de piezas defectuosas por caja? Recordemos que la mediana de una distribución es un valor que divide a los individuos de la población en dos efectivos iguales, supuestos ordenados por el valor creciente del carácter.

$$x_i / N_{i-1} < \frac{n}{2} \leq N_i \rightarrow N_i = \frac{n}{2}, \quad Me = \frac{x_i + x_{i+1}}{2} = \frac{4 + 5}{2} = 4,5$$

- d) Calcular los cuartiles de la distribución. Interpretarlos.

$x_i / N_{i-1} < \frac{n}{4} = 25 \leq N_i$. Como en este caso, $N_3 = 25$, hemos de hacer la media aritmética entre x_3 y x_4 :

$$Q_1 = \frac{2+3}{2} = 2,5 \Rightarrow 25 \% \text{ de las cajas presentan menos de 3,5 piezas defectuosas.}$$

$$Q_2 = Me = 4,5 \Rightarrow \text{El } 50 \% \text{ de las cajas presentan menos de 4.5 piezas defectuosas.}$$

$$x_i / N_{i-1} < \frac{3n}{4} \leq N_i \Rightarrow Q_3 = x_7 = 6 \Rightarrow \text{El } 75 \% \text{ de las cajas presentan 6 o menos piezas defectuosas.}$$

- e) Calcular los deciles de orden 3 y 7. Interpretarlos.

$$x_i / N_{i-1} < \frac{3n}{10} \leq N_i \Rightarrow D_3 = 3 \Rightarrow \text{El } 30 \% \text{ de las cajas presentan 3 o menos piezas defectuosas.}$$

$$x_i / N_{i-1} < \frac{7n}{10} \leq N_i \Rightarrow D_7 = 6 \Rightarrow \text{El } 70 \% \text{ de las cajas presentan 6 o menos piezas defectuosas.}$$

- f) Cuantificar la dispersión de la distribución mediante diferentes medidas, interpretando los resultados y señalando las ventajas e inconvenientes de cada una.

- 1) *Recorrido o rango.*

$$R = x_k - x_1 = 10 - 0 = 10$$

Esta medida se interpreta como el rango de valores que pueden tomar los individuos de la población en estudio, en nuestro caso sería desde 0 piezas defectuosas hasta 10. Sin embargo, es una medida poco representativa ya que no nos informa sobre cómo se hallan repartidos los individuos de la población en ese rango.

- 2) *Recorrido intercuartílico.*

$$R_I = Q_3 - Q_1 = 3,5$$

Esta medida representa el valor máximo en el que se diferencia el 50 % central de la distribución. En este caso, las cajas del 50 % central se diferencian en 3.5 piezas defectuosas como máximo. Sin embargo, no informa sobre cómo se hallan distribuidas las cajas dentro de ese 50 % central.

- 3) *Desviación absoluta media respecto a \bar{x}*

$$D_{\bar{x}} = \frac{\sum_{i=1}^k |x_i - \bar{x}| n_i}{n} = \frac{200}{100} = 2$$

- 4) *Desviación absoluta media respecto a Me .*

$$D_{Me} = \frac{\sum_{i=1}^k |x_i - Me|}{n} = \frac{200}{100} = 2$$

Estas dos últimas medidas nos permiten conocer la distancia promedio de la que se separan las cajas con respecto de la media y de la mediana, permitiéndonos hacernos una idea de la representatividad de estas dos medidas. Podemos considerar que son bastantes representativas ya que una diferencia de 2 piezas defectuosas no es alta considerando que estamos ante un rango 10.

5) *Varianza*

$$\sigma^2 = \frac{\sum_{i=1}^k n_i x_i^2}{n} - \bar{x}^2 = \frac{2488}{100} - 4,36^2 = 5,8704$$

Representa la dispersión de los datos de la distribución respecto de la media en piezas defectuosas al cuadrado.

6) *Desviación típica.*

$$\sigma = 2,422890835$$

Representa la dispersión de los datos de la distribución respecto de la media en piezas defectuosas. Esto quiere decir que la mayoría de los datos de la distribución se encuentra entre 1,9371 y 6,78289 piezas defectuosas.

7) *Medidas de dispersión relativas*

Ahora hallaremos las medidas de dispersión relativas, las cuales resultan útiles para comparar nuestra distribución con otras que presentan distinta media, desviación típica, etc:

$$R_R = \frac{R}{\bar{x}} = 2,29358 \text{ (Recorrido relativo)}$$

$$R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0,41176 \text{ (Recorrido semi-intercuartílico)}$$

$$C.V(x) = \frac{\sigma}{\bar{x}} = 0,5557 \text{ (Coeficiente de variación de Pearson)}$$

$$V_{\mu e} = \frac{D_{\mu e}}{\mu_e} = 0,44 \dots \text{ (Índice de dispersión respecto a la mediana)}$$

Ejercicio 5. Dada las siguientes distribuciones:

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

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

Calcular para cada una de ellas:

a) Medias aritmética, armónica y geométrica.

Media aritmética

$$1. \bar{x}_1 = \frac{12 \cdot 0,5 + 13 \cdot 1,5 + 11 \cdot 2,5 + 8 \cdot 3,5 + 6 \cdot 4,5}{12 + 13 + 11 + 8 + 6} = 2,16$$

$$2. \bar{x}_2 = \frac{1 \cdot 0,5 + 6 \cdot 2 + 7 \cdot 4,5 + 12 \cdot 8 + 2 \cdot 11}{1 + 6 + 7 + 12 + 2} = 5,7857$$

Media armónica

$$1. H_1 = \frac{12+13+11+8+6}{\frac{12}{0,5} + \frac{13}{1,5} + \frac{11}{2,5} + \frac{8}{3,5} + \frac{6}{4,5}} = 1,2289325 \approx 1,23$$

$$2. H_2 = \frac{1+6+7+12+2}{\frac{1}{0,5} + \frac{6}{2} + \frac{7}{4,5} + \frac{12}{8} + \frac{2}{11}} = 3,399141631$$

Media geométrica

$$1. G_1 = \sqrt[50]{0,5^{12} \cdot 1,5^{13} \cdot 2,5^{11} \cdot 3,5^8 \cdot 4,5^6} = 1,684688918$$

$$2. G_2 = \sqrt[28]{0,5^1 \cdot 2^6 \cdot 4,5^7 \cdot 8^{12} \cdot 11^2} = 4,769603117$$

b) El valor más frecuente.

Se nos está pidiendo la moda (Mo), la cual se debe encontrar en el intervalo modal (el de mayor densidad de frecuencias. Primero hallemos los h_i :

$h_i^{(1)}$	0,24	0,26	0,22	0,16	0,12
-------------	------	------	------	------	------

Observamos que el intervalo modal es (1, 2].

Ahora recordemos la fórmula (demostrada en clase) para las variables continuas:

$$Mo = e_{Mo-1} + \frac{h_{Mo} - h_{Mo-1}}{(h_{Mo} - h_{Mo-1}) + (h_{Mo} - h_{Mo+1})} a_i \quad (1)$$

$$1 + \frac{0,26 - 0,24}{(0,26 - 0,24) + (0,26 - 0,22)} \cdot 1 = \frac{4}{3} = 1.\widehat{3}$$

$h_i^{(2)}$	0,03571	0,107142	0,083	0,107142	0,03571428
-------------	---------	----------	-------	----------	------------

En este caso tenemos 2 intervalos modales: (1, 3] y (6,10], por lo que habrá dos modas. Volvemos a emplear (1):

(1)

$$1 + \frac{0,107142 - 0,03571}{(0,107142 - 0,03571) + (0,107142 - 0,083)} \cdot 2 = 2,5$$

(2)

$$6 + \frac{0,107142 - 0,083}{(0,107142 - 0,083) + (0,107142 - 0,03571428)} \cdot 4 = 7$$

c) El valor superado por el 50 % de las observaciones.

Dicho valor será la mediana ya que divide a los individuos de la población en 2 efectivos iguales, supuestos ordenados por valor creciente del carácter:

(1)

$$(e_{i-1}, e_i) \quad N_{i-1} < \frac{n}{2} \leq N_i \Rightarrow \frac{n}{2} = 25 \quad y \quad N_2 = 25 \quad Me = e_2 = 2$$

(2)

$$(e_{i-1}, e_i) \quad N_{i-1} < \frac{n}{2} \leq N_i \Rightarrow \frac{n}{2} = 14 \text{ y } N_3 = 14 \quad Me = e_3 = 6$$

d) Recorrido, recorrido intercuartílico y desviación típica. Interpretalos. ¿Qué distribución es más homogénea?

1) *Recorrido*

Se calcula como la diferencia entre el extremo superior del último intervalo de la distribución y el extremo inferior del primer intervalo de la distribución.

(1)

$$R_1 = e_5 - e_0 = 5 - 0 = 5$$

(2)

$$R_2 = e_5 - e_0 = 12$$

2) Recorrido intercuartílico

Primero hemos de hallar los cuartiles Q_1 y Q_3 de cada una de las dos distribuciones para poder calcular el recorrido intercuartílico. Recordemos la fórmula demostrada en clase para cuantiles, la cual también se puede trasladar a cuartiles:

$$C_\alpha = e_{C_\alpha-1} + \frac{n\alpha - N_{C_\alpha-1}}{n_{C_\alpha}} a_{C_\alpha} \quad (2)$$

Busquemos en que intervalo deben encontrarse estos dos cuantiles:

(1)

$$\begin{aligned} \frac{n}{4} &= 12,5 \Rightarrow (1, 2] \\ \frac{3n}{4} &= 37,5 \Rightarrow (3, 4] \end{aligned}$$

Por tanto:

$$\begin{aligned} C_{0,25} &= 1 + \frac{0,25 \cdot 50 - 12}{13} \cdot 1 = 1,03846 \\ C_{0,75} &= 3 + \frac{0,75 \cdot 50 - 36}{8} \cdot 1 = 3,1875 \end{aligned}$$

Así:

$$R_I = Q_3 - Q_1 = 2,14904$$

(2)

$$\frac{n}{4} = 7$$

El intervalo que presenta una frecuencia absoluta acumulada inmediatamente mayor que 7 es el $(1, 3]$. Como $N_2 = 7$ podemos afirmar que $Q_1 = 3$

$$\frac{3n}{4} = 21$$

El intervalo que presenta una frecuencia absoluta acumulada inmediatamente mayor que 21 es $(6, 10]$ por lo que hemos de buscar ahí el Q_3 . Usaremos (2)

$$\begin{aligned} Q_3 &= 6 + \frac{28 \cdot 0,75 - 14}{12} \cdot 4 = 8.\widehat{3} \\ R_I &= Q_3 - Q_1 = 8.\widehat{3} - 3 = 5.\widehat{3} \end{aligned}$$

Recordemos que el R_I representa cuanto se diferencian como máximo los datos que se encuentran en el 50 % de la distribución.

3) *Desviación típica*

(1)

$$\sigma^2 = \frac{\sum_{i=1}^k n_i c_i^2}{n} - \bar{x}^2$$

$$\sigma^2 = \frac{12 \cdot 0,5^2 + 13 \cdot 1,5^2 + 11 \cdot 2,5^2 + 8 \cdot 3,5^2 + 6 \cdot 4,5^2}{50} - 2,16^2 = 1,7444$$

$$\sigma = 1,32075$$

(2)

$$\sigma^2 = \frac{\sum_{i=1}^k n_i c_i^2}{n} - \bar{x}^2$$

$$\sigma^2 = \frac{1 \cdot 0,5^2 + 6 \cdot 2^2 + 7,4 \cdot 4,5^2 + 12 \cdot 8^2 + 2 \cdot 11^2}{28} - 5,7857^2 = 8,52567551$$

$$\sigma = 2,919875941$$

Para poder comparar las desviaciones típicas y determinar que distribución es más homogénea es necesario emplear el coeficiente de variación de Pearson (que es adimensional), ya que las distribuciones presentan distintas medidas y desviaciones típicas:

(1)

$$C.V_1(x) = \frac{\sigma}{\bar{x}} = 0,611458$$

(2)

$$C.V_2(x) = \frac{\sigma}{\bar{x}} = 0,50467116$$

Como la 2ª distribución presenta menor coeficiente de Pearson, sus datos no están tan repartidos como la primera, es decir, la segunda distribución es más HOMOGÉNEA.

Ejercicio 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.

Para ello emplearemos la media armónica, pues se usa para promediar datos que son cocientes de dos magnitudes distintas; esto es, magnitudes relativas (en este caso nuestros datos son velocidades, es decir, $\frac{e}{t}$):

$$v_M = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} = \frac{2}{\frac{1}{60} + \frac{1}{70}} = 64,61538 \text{ km/h}$$

Ejercicio 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.

Como nuestra variable (la rentabilidad) presenta efectos multiplicativos acumulativos, lo más lógico será emplear la media geométrica:

$$G = \sqrt[5]{5 \cdot 6 \cdot 7 \cdot 10 \cdot 12} = 7,5906 \approx 7,6 \%$$

Así, obtenemos que 7.6 % es la rentabilidad neta media en esos 5 años.

Ejercicio 8. Un profesor califica a sus alumnos según el criterio siguiente: 40 % de suspensos, 30 % de aprobados, 15 % notables, 10 % sobresalientes y 5 % 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.

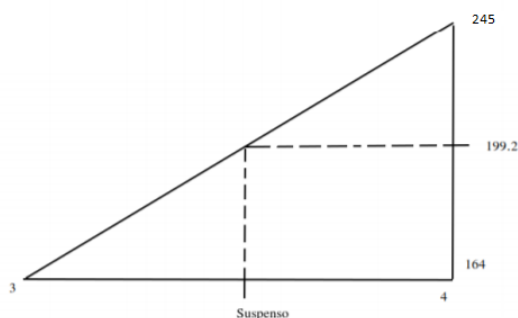
Sabemos que el orden es el siguiente:

Suspensos	Aprobados	Notables	Sobresalientes	Matrículas
40 %	30 %	15 %	10 %	5 %

Para saber la nota máxima de los suspensos hemos de hallar el percentil 40, es decir la nota por la que está debajo el 40 % de la población: $\frac{4n}{10} = 199,2$

	(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
N_i	34	108	164	245	339	409	450	478	494	498

Así, observamos que el percentil buscado está en el intervalo (3,4].



$$\frac{a_i}{n_\alpha} = \frac{e_{p\alpha} - p_\alpha}{N_\alpha - n_\alpha} \iff P_\alpha = e_{p\alpha} + \frac{n_\alpha - N_\alpha}{n_\alpha} a_i$$

$$P_{40} = 4 + \frac{0,4 \cdot 498 - 245}{81} \cdot 1 = 3,434567 \text{ Se suspende por debajo de dicha nota.}$$

Para el cálculo de la nota máxima en aprobados, notables, sobresalientes, y matrículas hemos de emplear la fórmula obtenida:

Aprobados \Rightarrow Percentil 70 $\Rightarrow \frac{7n}{100} = 348,6 \Rightarrow$ La nota máxima de los aprobados está en el intervalo $(5,6]$.

$$P_{70} = 6 + \frac{348,6 - 409}{70} \cdot 1 = 5,137142857$$

Notables \Rightarrow Percentil 85 $\Rightarrow \frac{85n}{100} = 423,3 \Rightarrow$ Buscar en $(6,7]$.

$$P_{85} = 7 + \frac{423,3 - 450}{41} \cdot 1 = 6,34878$$

Sobresalientes \Rightarrow Percentil 95 $\Rightarrow \frac{95n}{100} = 473,1 \Rightarrow$ Buscar en $(7,8]$.

$$P_{95} = 8 + \frac{473,1 - 478}{28} \cdot 1 = 7,825$$

Matrículas \Rightarrow Es obvio, pues el percentil 100 es $P_{100} = 10$

Ejercicio 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

- Si se considera bajos al 3 % de los individuos de menor altura, ¿cuál es la altura máxima que pueden alcanzar?
- Si se condiera altos al 18 % de los individuos de mayor altura, ¿cuál es su altura mínima?
- ¿Qué altura es superada sólo por $1/4$ de los jóvenes?
- Calcular el número de jóvenes cuya altura es superior a 1.75.
- Calcular la altura máxima de los 11 jóvenes más bajos.
- Calcular la altura mínima de los 11 jóvenes más altos.

Ejercicio 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

- a) Calcular la edad más común de los individuos estudiados.
- b) Calcular la edad mínima y máxima del 30 % central de los individuos.
- c) Calcular el recorrido intercuartílico y la desviación típica.
- d) Calcular e interpretar los valores de los coeficientes de asimetría y curtosis.