

# Relación de problemas Tema 1

## Introducción a la Estadística Descriptiva unidimensional

*Estadística Descriptiva e Introducción a la Probabilidad*  
1º DGIIM, 2017/2018

Alumnos (**Grupo 1**):  
Miguel Ángel Fernández Gutiérrez  
Eladia Gómez Morales  
Marta Gómez Sánchez  
Daniel González Alfert

Fecha de entrega:  
24 de marzo de 2018

- 
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.  
b) Representar la distribución mediante un diagrama de barras y la curva de distribución.  
c) Promediar los valores de la variable mediante diferentes medidas. Interpretarlas.

---

*Solución:*

Antes de nada, caracterizaremos nuestro *puñetero asterisco*:

- ⊗ **Población:** las familias de una determinada barriada ( $n = 500$ ).  
**Variable estadística:** el número de hijos de cada una de las familias.  
**Número de modalidades:**  $k = 7$

**Nota:** caracterizaremos nuestro *puñetero asterisco* al principio de cada uno de los problemas.

Dicho esto, procederemos a hacer cada uno de los apartados del problema:

a) La tabla completa es la siguiente:

$x_i$	$n_i$	$N_i$	$f_i$	$x_i n_i$
0	80	80	0.16	0
1	110	190	0.22	110
2	130	320	0.26	260
3	90	410	0.18	270
4	40	450	0.08	160
5	30	480	0.06	150
6	20	500	0.04	120
	500			1070

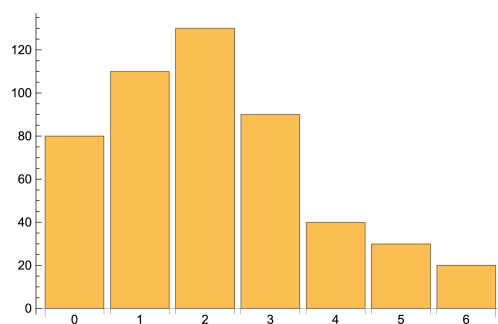
$n_i$  : frecuencias absolutas

$N_i$  : frecuencias absolutas acumuladas

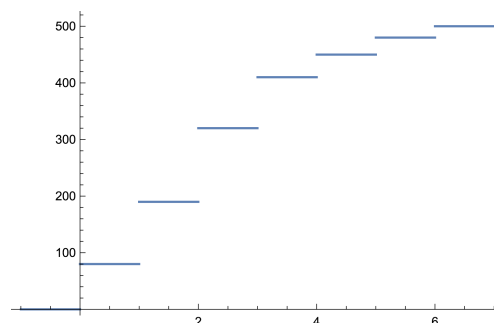
$f_i$  : frecuencias relativas

**Nota:** de ahora en adelante, insertaremos en las tablas las columnas necesarias para realizar los cálculos pertinentes al problema.

b) Las gráficas son las siguientes:



Gráfica de barras



Curva de distribución

c) Lo promediaremos con las siguientes medidas estadísticas:

- **Media:** usamos la **media aritmética**.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i = 2,14$$

El valor es representativo de los datos, pues se trata de una variable unidimensional, que no es acumulativa ni depende del cociente de otras dos magnitudes.

*Por lo expuesto anteriormente, el resto de medidas de posición (media armónica y media geométrica) no tienen sentido en este caso.*

- **Mediana:**  $Me = 2$ , puesto que  $n/2 = 250 < N_3$ , y  $Me = x_3 = 2$ .

La mitad de las familias tienen dos o menos hijos.

- **Moda:**  $Mo = 2$ , que es la variable con mayor  $n_i$ .

La mayoría de las familias tienen dos hijos.

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.

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

*Solución:*

⊗ **Población:** las familias de una determinada barriada.

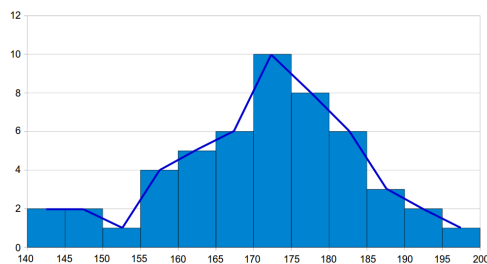
**Variable estadística:** el número de hijos de cada una de las familias ( $n = 50$ ).

**Número de modalidades:**  $k = 12$ .

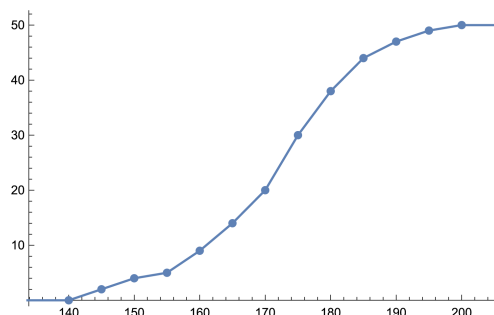
- a) La tabla completa es la siguiente:

$I_i$	$n_i$	$N_i$	$f_i$	$F_i$	$c_i$	$a_i$	$h_i$
[140, 145]	2	2	0.04	0.04	142.5	5	0.008
(145, 150]	2	4	0.04	0.08	147.5	5	0.008
(150, 155]	1	5	0.02	0.1	152.5	5	0.004
(155, 160]	4	9	0.08	0.18	157.5	5	0.016
(160, 165]	5	14	0.1	0.28	162.5	5	0.02
(165, 170]	6	20	0.12	0.4	167.5	5	0.024
(170, 175]	10	30	0.2	0.6	172.5	5	0.04
(175, 180]	8	38	0.16	0.76	177.5	5	0.032
(180, 185]	6	44	0.12	0.88	182.5	5	0.024
(185, 190]	3	47	0.06	0.94	187.5	5	0.012
(190, 195]	2	49	0.04	0.98	192.5	5	0.008
(195, 200]	1	50	0.02	1	197.5	5	0.004
	50						

- b) Las gráficas son las siguientes:



Histograma y poligonal de frecuencias



Curva de distribución

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]	2							$n_i$ : frecuencias absolutas
( , 10200]		5		10/18		1100		$N_i$ : frec. abs. acumuladas
			2/18		12000			$f_i$ : frecuencias relativas
	4						0.005/18	$F_i$ : frec. rel. acumuladas
		18					0.002/18	$c_i$ : marcas de clase
								$a_i$ : amplitudes
								$h_i$ : densidades de frecuencia

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

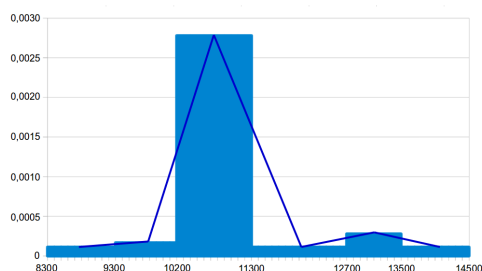
*Solución:*

- ⊗ **Población:** las comunidades autónomas de España, contando a Ceuta y Melilla como una ( $n = 18$ ).  
**Variable estadística:** la renta familiar.

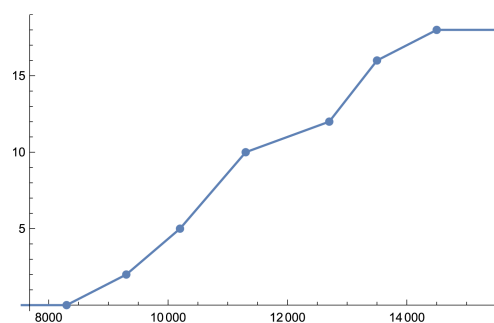
- a) La tabla completa es:

$I_i$	$n_i$	$N_i$	$f_i$	$F_i$	$c_i$	$a_i$	$h_i$
(8300, 9300]	2	2	2/18	2/18	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	0.05/18
(11300, 12700]	2	12	2/18	12/18	12000	1400	0.002/18
(12700, 13500]	4	16	4/18	16/18	13100	800	0.005/18
(13500, 14500]	2	18	2/18	1	14000	1000	0.002/18

- b) Las gráficas son las siguientes:



Histograma y poligonal de frecuencias



Curva de distribución

- c) 12 comunidades presentan una renta menor o igual que 12700 euros (simplemente  $N_4 = 12$ ).  
 8 comunidades presentan una renta superior a 11300 euros (basta calcular  $n - N_3 = 18 - 10 = 8$ , puesto que al ser las comunidades que presentan una renta superior a 11300, la cota inferior de  $I_4$ , debemos de contar también el  $n_4$ , y  $n_4 + n_5 + n_6 = n - N_3$ ).

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

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

*Solución:*

- ⊗ **Población:** las cajas de una determinada empresa ( $n = 100$ ).  
**Variable estadística:** el número de piezas defectuosas.

La tabla completa es:

$x_i$	$n_i$	$N_i$	$x_i n_i$	$x_i^2 n_i$
0	6	6	0	0
1	9	15	9	9
2	10	25	20	40
3	11	36	33	99
4	14	50	56	224
5	16	66	80	400
6	16	82	96	576
7	9	91	63	441
8	4	95	32	256
9	3	98	27	243
10	2	100	20	200
	100		436	2488

- a) Hay una media de 4,36 piezas defectuosas:

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

- b) Se encuentran más frecuentemente 5 y 6 piezas defectuosas.  
c)  $Me = 4,5$ , puesto que  $n/2 = 50 = N_5$ , por tanto,  $Me = \frac{x_5 + x_6}{2} = \frac{4+5}{2} = 4,5$ .

d) El procedimiento es análogo a la mediana:

- **Primer cuartil:**  $Q_1 = 2,5$ , puesto que  $n/4 = 25 = N_3$ , por tanto,  $Q_1 = \frac{x_3+x_4}{2} = \frac{2+3}{2} = 2,5$ .  
*Interpretación: un cuarto de las cajas tiene como máximo 2 piezas defectuosas.*
- **Segundo cuartil:**  $Q_2 = Me = 4,5$ .  
*Interpretación: la mitad de las cajas tiene como máximo 4 piezas defectuosas.*
- **Tercer cuartil:**  $Q_3 = 6$ , puesto que  $3n/4 = 75 < N_7$ , por tanto,  $Q_3 = x_7 = 6$ .  
*Interpretación: tres cuartos de las cajas tienen como máximo 6 piezas defectuosas.*

e) Del mismo modo:

- **Decil de orden 3:**  $D_3 = 3$ , puesto que  $3n/10 = 30 < N_4$ , por tanto,  $D_3 = x_4 = 3$ .  
*Interpretación: tres décimos de las cajas tienen como máximo 3 piezas defectuosas.*
- **Decil de orden 7:**  $D_7 = 6$ , puesto que  $6n/10 = 60 < N_6$ , por tanto,  $D_7 = x_6 = 6$ .  
*Interpretación: siete décimos de las cajas tienen como máximo 6 piezas defectuosas.*

f) Realizaremos las siguientes medidas de la dispersión:

- **Recorrido o rango:**  $R = x_{11} - x_1 = 10 - 0 = 10$ .  
*La distancia entre el menor y el mayor valor de la variable estadística es de 10.*
- **Recorrido intercuartílico:**  $R_I = Q_3 - Q_1 = 6 - 2,5 = 3,5$ .  
*La longitud del intervalo en el que está incluido el 50 % central de los datos es de 3,5.*
- **Varianza:** podemos calcularla, de acuerdo al *teorema de König*, del siguiente modo:

$$Var(X) = \sigma^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \left( \frac{1}{n} \sum_{i=1}^k x_i n_i \right)^2 = \frac{2488}{100} - \frac{436^2}{100^2} = 5,87 \text{ piezas}^2$$

- **Desviación típica:**

$$\sigma = +\sqrt{\sigma^2} = \sqrt{5,87} = 2,42$$

- **Coefficiente de variación:**

$$C.V.(X) = \frac{\sigma_x}{|\bar{x}|} = \frac{2,42}{4,36} = 0,56$$

*Al estar el coeficiente de variación acotado entre 0 y 1, y estar éste a la mitad, los datos están un tanto dispersos.*

5. Dadas las siguientes distribuciones:

$I_i^{(1)}$	(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]
$n_i^{(1)}$	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:

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

*Solución:*

- ⊗ Aquí no conocemos el *puñetero asterisco*, así que lo mencionaremos igualmente: *Estamos ante una población de tamaño  $n$ , en la que se ha observado cierta variable estadística  $X$ , que ha presentado  $k$  modalidades distintas  $(x_1, x_2, \dots, x_i, \dots, x_k)$ , con una distribución de frecuencias  $\{x_i, n_i\}_{i=1, \dots, n}$ .*

Las tablas tipo 3 para este problema son:

$I_i^{(1)}$	$n_i^{(1)}$	$N_i^{(1)}$	$c_i^{(1)}$	$a_i^{(1)}$	$c_i n_i^{(1)}$	$c_i^2 n_i^{(1)}$	$I_i^{(2)}$	$n_i^{(2)}$	$N_i^{(2)}$	$c_i^{(2)}$	$a_i^{(2)}$	$c_i n_i^{(2)}$	$c_i^2 n_i^{(2)}$
(0, 1]	12	12	0.5	1	6	3	(0, 1]	1	1	0.5	1	0.5	0.25
(1, 2]	13	25	1.5	1	19.5	29.25	(1, 3]	6	7	2	2	12	24
(2, 3]	11	36	2.5	1	27.5	68.75	(3, 6]	7	14	4.5	3	31.5	141.75
(3, 4]	8	44	3.5	1	28	98	(6, 10]	12	26	8	4	96	768
(4, 5]	6	50	4.5	1	27	121.5	(10, 12]	2	28	11	2	22	242
	50				108	320.5		28				162	1176

a) Realizaremos cada una de las medias:

• **Media aritmética:**

$$\bar{x}^{(1)} = \frac{1}{n^{(1)}} \sum_{i=1}^k c_i n_i^{(1)} = \frac{108}{50} = 2,16 \quad \bar{x}^{(2)} = \frac{1}{n^{(2)}} \sum_{i=1}^k c_i n_i^{(2)} = \frac{162}{28} = 5,79$$

• **Media armónica:**

$$H^{(1)} = \frac{n^{(1)}}{\sum_{i=1}^k \frac{n_i^{(1)}}{c_i^{(1)}}} = \frac{50}{40,67} = 1,23 \quad H^{(2)} = \frac{n^{(2)}}{\sum_{i=1}^k \frac{n_i^{(2)}}{c_i^{(2)}}} = \frac{28}{8,24} = 3,40$$

• **Media geométrica:**

$$G^{(1)} = \sqrt[n^{(1)}]{\prod_{i=1}^k c_i^{n_i^{(1)}}} = 1,68 \quad G^{(2)} = \sqrt[n^{(2)}]{\prod_{i=1}^k c_i^{n_i^{(2)}}} = 4,77$$

*Conclusión: deberíamos especificar qué tipo de variable es, puesto que una de ellas tendrá sentido dependiendo del caso.*

b) Estamos ante una distribución bimodal, con  $Mo_1 = 2,498$ ,  $Mo_2 = 7,003$ .

c) Sería, simplemente, calcular la mediana para cada caso. Tenemos que:

(1)  $Me^{(1)} = 2$ , puesto que  $n^{(1)}/2 = 25 = N_2^{(1)}$ , así que basta tomar la cota superior del intervalo  $I_2$ , pues a partir de ella estará la mitad de la población.

(2)  $Me^{(2)} = 6$ , puesto que  $n^{(2)}/2 = 14 = N_3^{(2)}$ , así que basta tomar la cota superior del intervalo  $I_3$ , pues a partir de ella estará la mitad de la población.

d) Calcularemos e interpretaremos cada una de las medidas estadísticas:

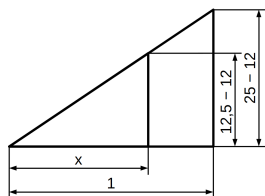
• **Recorrido:**

(1)  $R^{(1)} = e_5^{(1)} - e_0^{(1)} = 5 - 0 = 5$

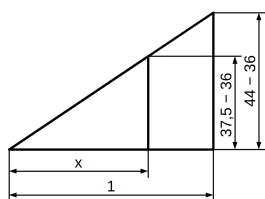
(2)  $R^{(2)} = e_5^{(2)} - e_0^{(2)} = 12 - 0 = 12$

*Interpretación: la distancia entre el mayor y menor valor posible de la variable es de 5 y 12, respectivamente.*

• **Recorrido intercuartílico:** calcularemos los cuartiles a continuación:

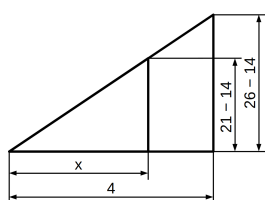


$$\frac{x}{1} = \frac{0,5}{13} \rightarrow x = 0,0384 \Rightarrow Q_1^{(1)} = 1,0384$$



$$\frac{x}{1} = \frac{1,5}{8} \rightarrow x = 0,1875 \Rightarrow Q_3^{(1)} = 3,1875$$

Ya que  $n^{(2)}/4 = 7$ , que es la cota superior de  $I_2$ ,  $Q_1^{(2)} = 3$ .



$$\frac{x}{4} = \frac{7}{12} \rightarrow x = 2,3 \Rightarrow Q_3^{(2)} = 8,3$$

Por tanto, los recorridos son:

(1)  $R_I^{(1)} = Q_3^{(1)} - Q_1^{(1)} = 3,1875 - 1,0384 = 2,1491$

(2)  $R_I^{(2)} = Q_3^{(2)} - Q_1^{(2)} = 8,3 - 3 = 5,3$

*Interpretación: la longitud del intervalo en el que está incluido el 50 % central de los datos es de 2,1491 y 5,3, respectivamente.*

**Nota:** hemos realizado los cálculos suponiendo que la distribución es uniforme.



- Desviación típica:

(1)

$$\sigma^{(1)} = \sqrt{\frac{1}{n^{(1)}} \sum_{i=1}^k c_i^2 n_i^{(1)} - \left( \frac{1}{n^{(1)}} \sum_{i=1}^k c_i n_i^{(1)} \right)^2} = \sqrt{\frac{320,5}{50} - \frac{108^2}{50^2}} = 1,32$$

(2)

$$\sigma^{(2)} = \sqrt{\frac{1}{n^{(2)}} \sum_{i=1}^k c_i^2 n_i^{(2)} - \left( \frac{1}{n^{(2)}} \sum_{i=1}^k c_i n_i^{(2)} \right)^2} = \sqrt{\frac{1176}{28} - \frac{162^2}{28^2}} = 2,92$$

*Interpretación: para ver qué distribución es más homogénea, calculamos el **coeficiente de variación de Pearson**:*

$$C.V.(X^{(1)}) = \frac{\sigma_x^{(1)}}{|\bar{x}^{(1)}|} = \frac{1,32}{2,16} = 0,61 \quad C.V.(X^{(2)}) = \frac{\sigma_x^{(2)}}{|\bar{x}^{(2)}|} = \frac{2,92}{5,79} = 0,50$$

*Vemos que la primera distribución está más dispersa.*

6. Un móvil efectúa un recorrido de 100km en dos sentidos. En uno va a una velocidad constante de  $v_1 = 60\text{km/h}$  y el otro va a una velocidad constante de  $v_2 = 70\text{km/h}$ . Calcular la velocidad media del recorrido.

---

*Solución:*

⊗ **Población:** los móviles que efectúan el recorrido ( $n = 2$ ).

**Variable estadística:** la velocidad de cada móvil.

Para resolver este problema, simplemente hemos de tener en cuenta la fórmula de la velocidad:

$$v \text{ (velocidad)} = \frac{d \text{ (distancia)}}{t \text{ (tiempo)}}$$

Por tanto, para calcular la velocidad media del recorrido,  $v_m$ , tendremos que aplicar la fórmula:

$$v_m \text{ (velocidad media)} = \frac{d_{\text{total}} \text{ (distancia total recorrida)}}{t_{\text{total}} \text{ (tiempo total empleado)}}$$

1. Distancia total recorrida:

$$d_{\text{total}} = d_1 + d_2 = 100\text{km} + 100\text{km} = 200\text{km}$$

2. Tiempo total empleado:

$$t_{\text{total}} = t_1 + t_2 = \frac{d_1}{v_1} + \frac{d_2}{v_2} = \frac{100\text{km}}{60\text{km/h}} + \frac{100\text{km}}{70\text{km/h}} = 1,67\text{h} + 1,43\text{h} = 3,10\text{h}$$

3. Velocidad media:

$$v_m = \frac{d_{\text{total}}}{t_{\text{total}}} = \frac{d_1 + d_2}{t_1 + t_2} = \frac{200\text{km}}{3,10\text{h}} = 64,62\text{km/h}$$

Por tanto, la velocidad media del recorrido es de 64,62km/h.

Por otra parte, podríamos haber realizado la media armónica, dando el mismo resultado:

$$v_m = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} = \frac{2}{\frac{1}{60\text{km/h}} + \frac{1}{70\text{km/h}}} = 64,62\text{km/h}$$

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.

---

*Solución:*

⊗ **Población:** los años en los que una empresa ha obtenido beneficios.

**Variable estadística:** la rentabilidad de la empresa.

**Nota:** para poder considerar el *puñetero asterisco*, las modalidades de la variable estadística deberían estar ordenadas de menor a mayor.

El rendimiento de la empresa es un cálculo que se acumula a lo largo de los años. De este modo:

$$c_1 = 1,12c_0 \quad c_2 = 1,1c_1 \quad c_3 = 1,07c_2 \quad c_4 = 1,06c_3 \quad c_5 = 1,05c_4$$

Queremos obtener el rendimiento medio, en otras palabras, el  $i$  que cumple que:

$$\begin{aligned} c_0(1+i)^5 &= c_5 = c_0 * 1,12 * 1,1 * 1,07 * 1,06 * 1,05 \\ \rightarrow i - 1 &= G = \sqrt[5]{1,12 * 1,1 * 1,07 * 1,06 * 1,05} - 1 = 1,08 - 1 = 0,08 \end{aligned}$$

Por tanto, el rendimiento neto medio es del 8 %. Nótese que hemos hecho una *media geométrica*.

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.

*Solución:*

⊗ **Población:** los estudiantes de un profesor.

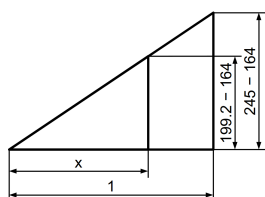
**Variable estadística:** la nota que ha sacado cada estudiante.

La tabla tipo 3 para este problema es:

$I_i$	$n_i$	$N_i$	$c_i$	$a_i$
(0, 1]	34	34	0.5	1
(1, 2]	74	108	1.5	1
(2, 3]	56	164	2.5	1
(3, 4]	81	245	3.5	1
(4, 5]	94	339	4.5	1
(5, 6]	70	409	5.5	1
(6, 7]	41	450	6.5	1
(7, 8]	28	478	7.5	1
(8, 9]	16	494	8.5	1
(9, 10]	4	498	9.5	1
	498			

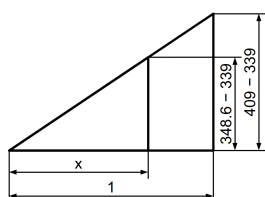
Para poder calcular las notas máximas para obtener cada una de las calificaciones, utilizaremos el concepto de *percentil*:

- **Suspenseo:** bastará calcular el percentil 40,  $P_{40}$ . Esa será la nota máxima de los que han suspendido.



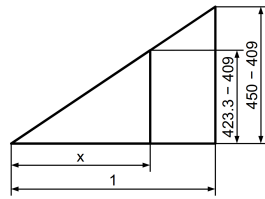
$$\frac{x}{1} = \frac{35,2}{81} \rightarrow x = 0,4345 \Rightarrow P_{40} = 3,4345$$

- **Aprobado:** bastará calcular el percentil 70,  $P_{70}$ . Esa será la nota máxima de los que tienen un aprobado.



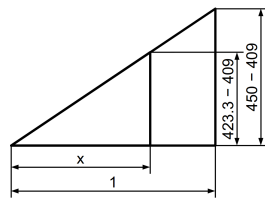
$$\frac{x}{1} = \frac{9,6}{70} \rightarrow x = 0,1371 \Rightarrow P_{70} = 5,1371$$

- **Notable:** bastará calcular el percentil 85,  $P_{85}$ . Esa será la nota máxima de los que tienen un notable.



$$\frac{x}{1} = \frac{14,3}{41} \rightarrow x = 0,3487 \Rightarrow P_{85} = 6,3487$$

- **Sobresaliente:** bastará calcular el percentil 95,  $P_{95}$ . Esa será la nota máxima de los que tienen un notable.



$$\frac{x}{1} = \frac{23,1}{28} \rightarrow x = 0,825 \Rightarrow P_{95} = 7,825$$

- **Matrícula:** la nota máxima será un 10, puesto que es la cota superior del último intervalo.

**Nota:** hemos realizado los cálculos suponiendo que la distribución es uniforme.

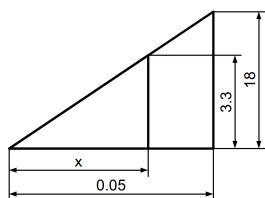
9. Se ha medido la altura de 110 jóvenes, obteniendo:

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 consideran bajos el 3 % de los individuos de menor altura, ¿cuál es la altura máxima que pueden alcanzar?
- Si se consideran altos el 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.

*Solución:*

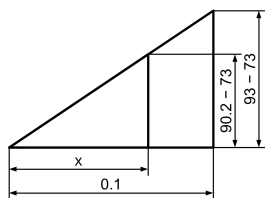
- a) Bastaría calcular el percentil  $P_3$ , pues el 3 % estará por debajo de éste.



$$\frac{x}{3,3} = \frac{0,05}{18} \rightarrow x = 0,009 \Rightarrow P_3 = 1,559$$

*Interpretación:* el 3 % de los jóvenes miden como máximo 1,559m.

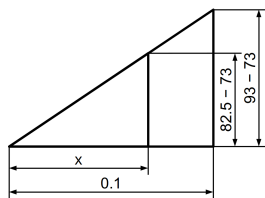
- b) Bastaría calcular el percentil  $P_{82}$ , pues el 18 % estará por encima de éste.



$$\frac{x}{17,2} = \frac{0,1}{20} \rightarrow x = 0,086 \Rightarrow P_{82} = 1,886$$

*Interpretación:* el 18 % de los jóvenes más altos miden más de 1,886m.

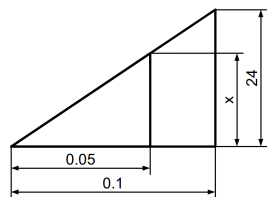
- c) Bastaría calcular el percentil  $P_{75}$ , pues el 25 % estará por encima de éste.



$$\frac{x}{9,5} = \frac{0,1}{20} \rightarrow x = 0,0475 \Rightarrow P_{75} = 1,8475$$

*Interpretación:* un cuarto de los jóvenes mide más de 1,8475m.

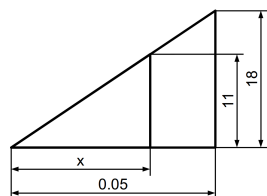
- d) Tomamos el intervalo donde se encuentra el 1,75,  $I_3 = (1,70, 1,80]$ , y suponiendo que la distribución es homogénea, calculamos el valor correspondiente dentro de la función de distribución.



$$\frac{x}{24} = \frac{0,05}{0,1} \rightarrow x = 12$$

A continuación sumamos a esto  $n_4$  y  $n_5$ , obteniendo que el número de jóvenes cuya altura es superior a 1,75m es de:  $12 + 20 + 17 = 49$  jóvenes.

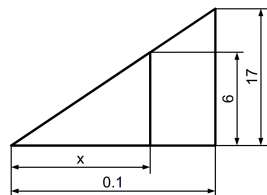
- e) Tomamos el intervalo donde se encuentra el joven número 11 si los ordenamos de menor a mayor altura, es decir, el intervalo  $I_1 = (1,55, 1,60]$ , y suponiendo que la distribución es homogénea, calculamos el valor correspondiente dentro de la función de distribución.



$$\frac{x}{0,05} = \frac{11}{18} \rightarrow x = 0,0305 \Rightarrow \text{Altura}_{11} = 1,5805$$

*Interpretación: el joven número 11 mide como máximo 1,5805m .*

- f) Tomamos el intervalo donde se encuentra el joven número 99 si los ordenamos de menor a mayor altura, es decir, el joven número 6 en el intervalo  $I_5 = (1,90, 2,00]$ , y suponiendo que la distribución es homogénea, calculamos el valor correspondiente dentro de la función de distribución.



$$\frac{x}{0,1} = \frac{6}{17} \rightarrow x = 0,03529 \Rightarrow \text{Altura}_{99} = 1,9353$$

*Interpretación: el joven número 99 mide más de 1,9353m .*

**Nota:** hemos realizado los cálculos suponiendo que la distribución es uniforme.

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

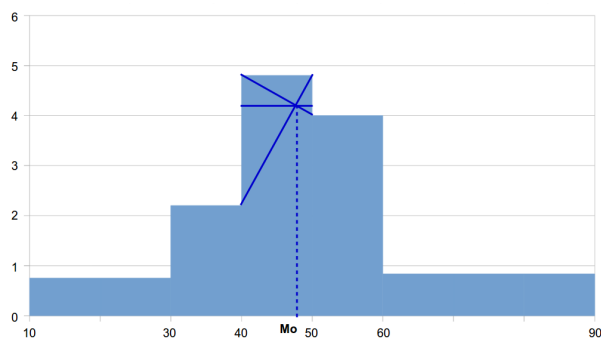
- Calcular la edad más común de los individuos estudiados.
- Calcular la edad mínima y máxima del 30 % central de los individuos.
- Calcular el recorrido intercuartílico y la desviación típica.
- Calcular e interpretar los valores de los coeficientes de asimetría y curtosis.

*Solución:*

La tabla completa de este problema es:

$I_i$	$n_i$	$N_i$	$f_i$	$c_i$	$c_i n_i$	$c_i^2 n_i$	$a_i$	$h_i$
[10, 30]	15	15	0.1	20	300	6000	20	0.75
(30, 40]	22	37	0.147	35	770	26950	10	2.2
(40, 50]	48	85	0.32	45	2160	97200	10	4.8
(50, 60]	40	125	0.267	55	2200	121000	10	4
(60, 90]	25	150	0.167	75	1875	140625	30	0.833
	150		$\approx 1$		7305	391775		

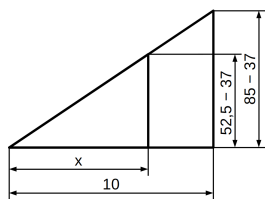
- a) Calcularemos la moda. Para ello, aplicamos el *teorema de Tales* sobre el histograma y obtenemos la siguiente proporción:



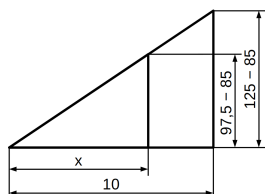
$$\frac{26}{Mo-40} = \frac{8}{50-Mo} \Rightarrow Mo = 47,6471$$



- b) Calcularemos el percentil 35,  $P_{35}$ , y el percentil 65,  $P_{65}$ .



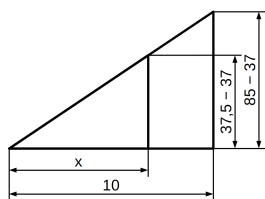
$$\frac{x}{15,5} = \frac{10}{48} \rightarrow x = 3,229 \Rightarrow P_{35} = 43,229$$



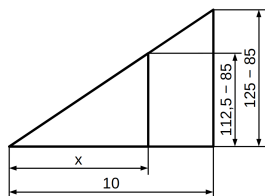
$$\frac{x}{12,5} = \frac{10}{40} \rightarrow x = 3,125 \Rightarrow P_{65} = 53,125$$

La edad mínima del 30 % central de los individuos es de 43,229 años, mientras que la máxima es de 53,125 años.

- c) Calcularemos los cuartiles, de forma análoga a problemas anteriores:



$$\frac{x}{10} = \frac{0,5}{48} \rightarrow x = 0,1042 \Rightarrow Q_1 = 40,1042$$



$$\frac{x}{10} = \frac{27,5}{40} \rightarrow x = 6,875 \Rightarrow Q_3 = 56,875$$

- **Recorrido intercuartílico:**  $R_I = Q_3 - Q_1 = 56,875 - 40,1042 = 16,7708$ .  
La longitud del intervalo en el que está incluido el 50 % central de los datos es de 16,7708.
- **Desviación típica:** la calcularemos mediante el teorema de König:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^k c_i^2 n_i - \left( \frac{1}{n} \sum_{i=1}^k c_i n_i \right)^2} = 15,497$$

**Nota:** hemos realizado los cálculos suponiendo que la distribución es uniforme.

d) Calcularemos los coeficientes de asimetría y curtosis:

- **Coeficiente de asimetría:** lo calcularemos mediante el *coeficiente de asimetría de Fisher*:

$$\gamma_1(X) = \frac{\mu^3}{\sigma_x^3} = \frac{1}{n} \sum_{i=1}^k \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^3 = -0,66$$

Como  $\gamma_1(X) < 0$ , estamos ante una distribución *asimétrica por la izquierda* o *negativa*.

- **Coeficiente de curtosis:** lo calcularemos mediante el *coeficiente de curtosis de Fisher*:

$$\gamma_2(X) = \frac{\mu^4}{\sigma_x^4} - 3 = \frac{1}{n} \sum_{i=1}^k \left( \frac{c_i - \bar{x}}{\sigma_x} \right)^4 - 3 = -1,72$$

Como  $\gamma_2(X) < 0$ , estamos ante una distribución *platicúrtica*.