Estadística Descriptiva e Introducción a la Probabilidad

Relación 1

Daniel Alconchel Vázquez
Mario García Márquez
Nars El Farissi
Alberto Diaz Cencillo
Javier Garrues Apecechea

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		0.16
1	110		
2		320	
3			0.18
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.

(*) En una población de tamaño $n = 500$ familias de una barriada, se ha observado una variable estadística X = "nº hijos de cada familia" que ha presentado 7 modalidades distintas $\{0, 1, 2, 3, 4, 5, 6\}$ con una distribución de frecuencia $\{x_i, n_i\}$ $i = 1, \dots, 7$

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
	500		1

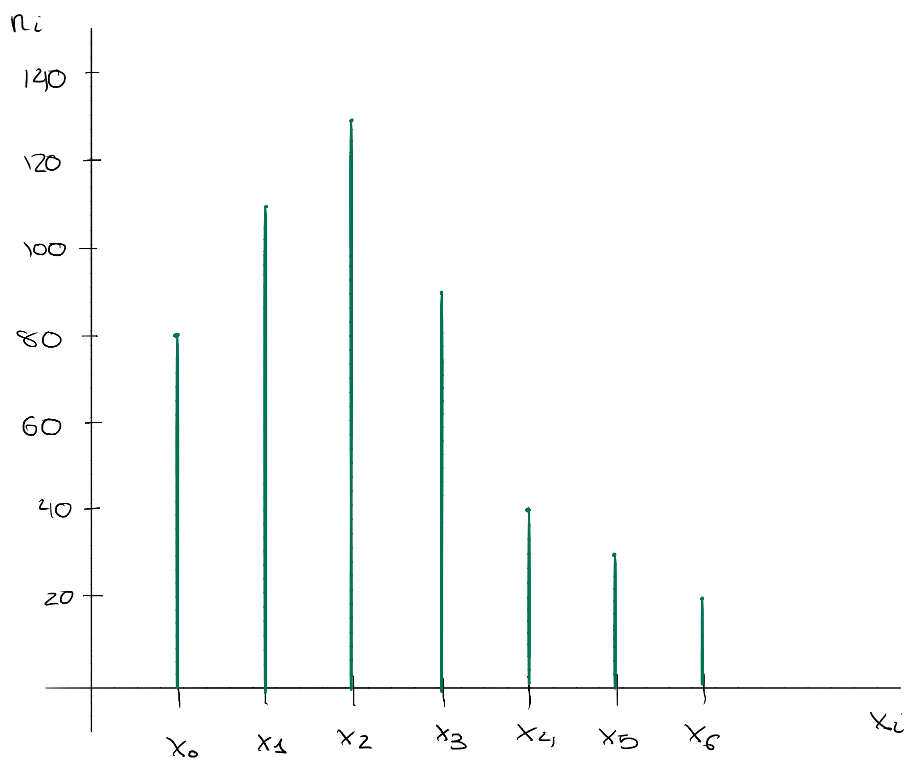
Para calcular el tamaño total de la población (n) hemos usado:

$$f_i = \frac{n_i}{n} \Rightarrow n = \frac{n_i}{f_i} = \frac{n_1}{f_1} = \frac{80}{0,16} = 500 \text{ familias}$$

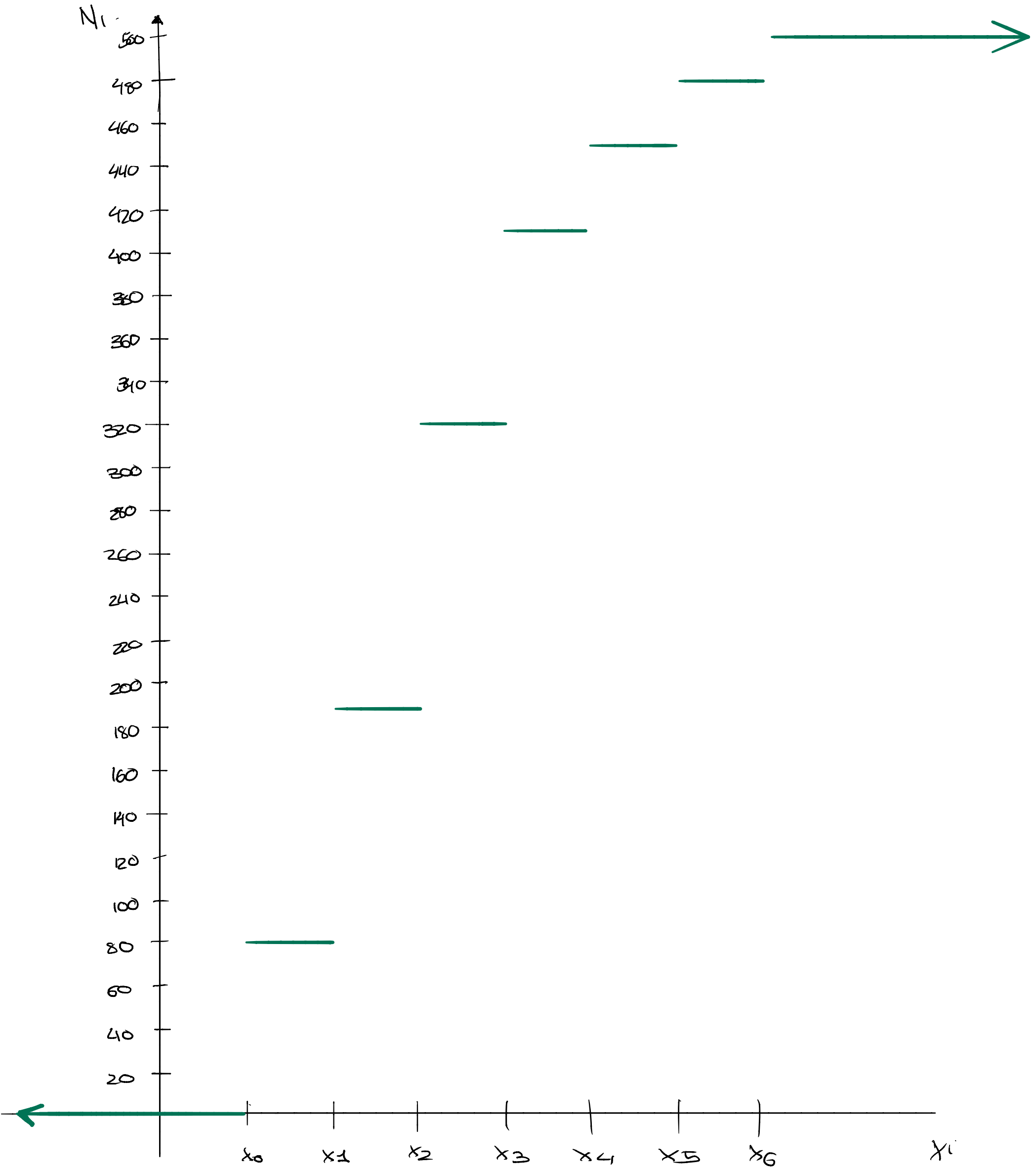
Para rellenar el resto de la tabla.

- $N_1 = n_1 = 80$
- $N_2 = N_1 + n_2 = 190$
- $f_2 = \frac{n_2}{n} = 0,22$
- $n_3 = N_3 - N_2 = 130$
- $f_3 = \frac{n_3}{n} = 0,26$
- $n_4 = f_4 \cdot n = 90$
- $N_4 = n_4 + n_3 = 410$
- $N_5 = n_5 + N_4 = 450$
- $f_5 = \frac{n_5}{n} = 0,08$
- $N_7 = n = 500$
- $N_6 = N_7 - n_7 = 480$
- $n_6 = N_6 - N_5 = 30$
- $f_6 = \frac{n_6}{n} = 0,06$
- $f_7 = \frac{n_7}{n} = 0,04$

Diagrama de barras



Curva de distribución



c)

Media aritmética:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i = 2,12 \text{ hijos / familia}$$

Mediana:

$Me = x_i : N_i > \frac{n}{2}$ $\quad \frac{500}{2} = 250$ $\quad$ Como $N_3 > 250 \Rightarrow Me = 2$ hijos

Moda:

$Mo = 2$ hijos

Vemos que la moda y la mediana coinciden. Además, la media aritmética tiene un valor próximo a estas, por lo que es un buen representante del centro de gravedad de la distribución

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

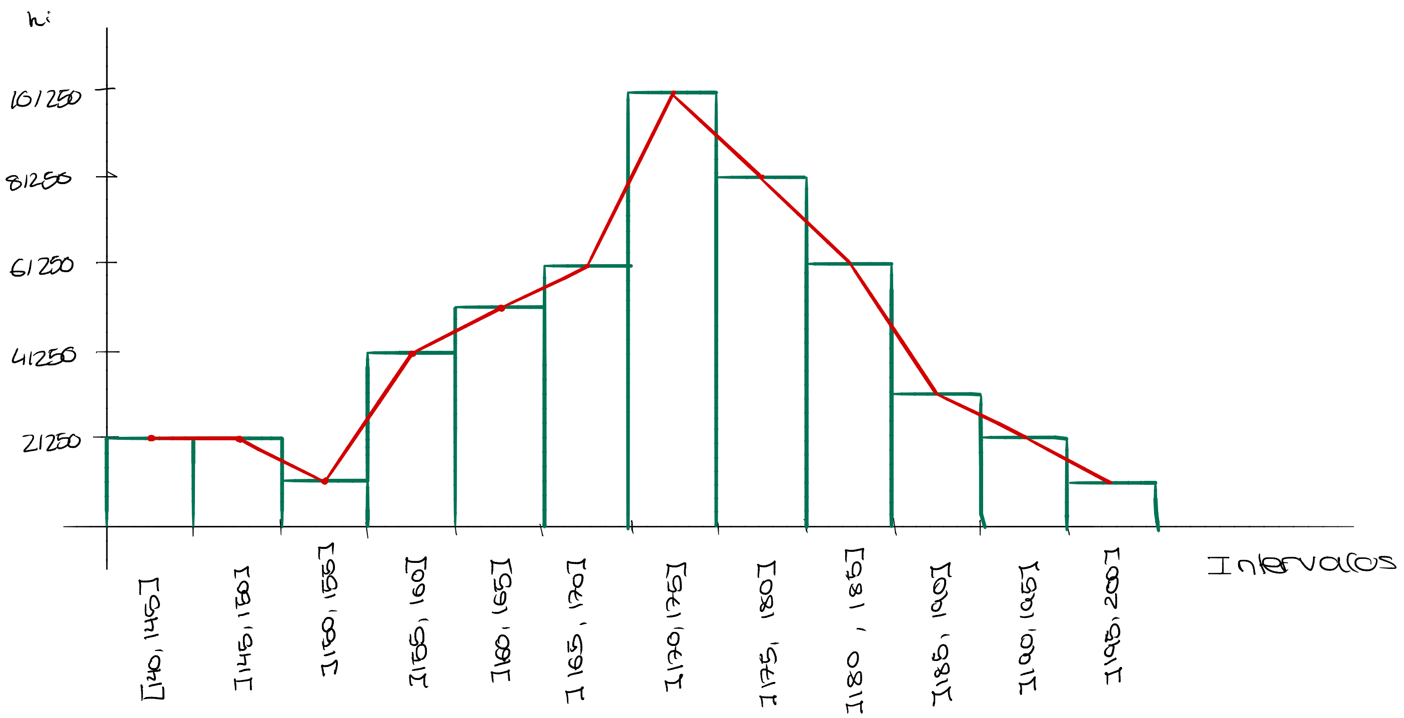
$$174, 185, 166, 176, 145, 166, 191, 175, 158, 156, 156, 187, 162, 172, 197, 181, 151,$$
$$161, 183, 172, 162, 147, 178, 176, 141, 170, 171, 158, 184, 173, 169, 162, 172, 181,$$
$$187, 177, 164, 171, 193, 183, 173, 179, 188, 179, 167, 178, 180, 168, 148, 173.$$

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

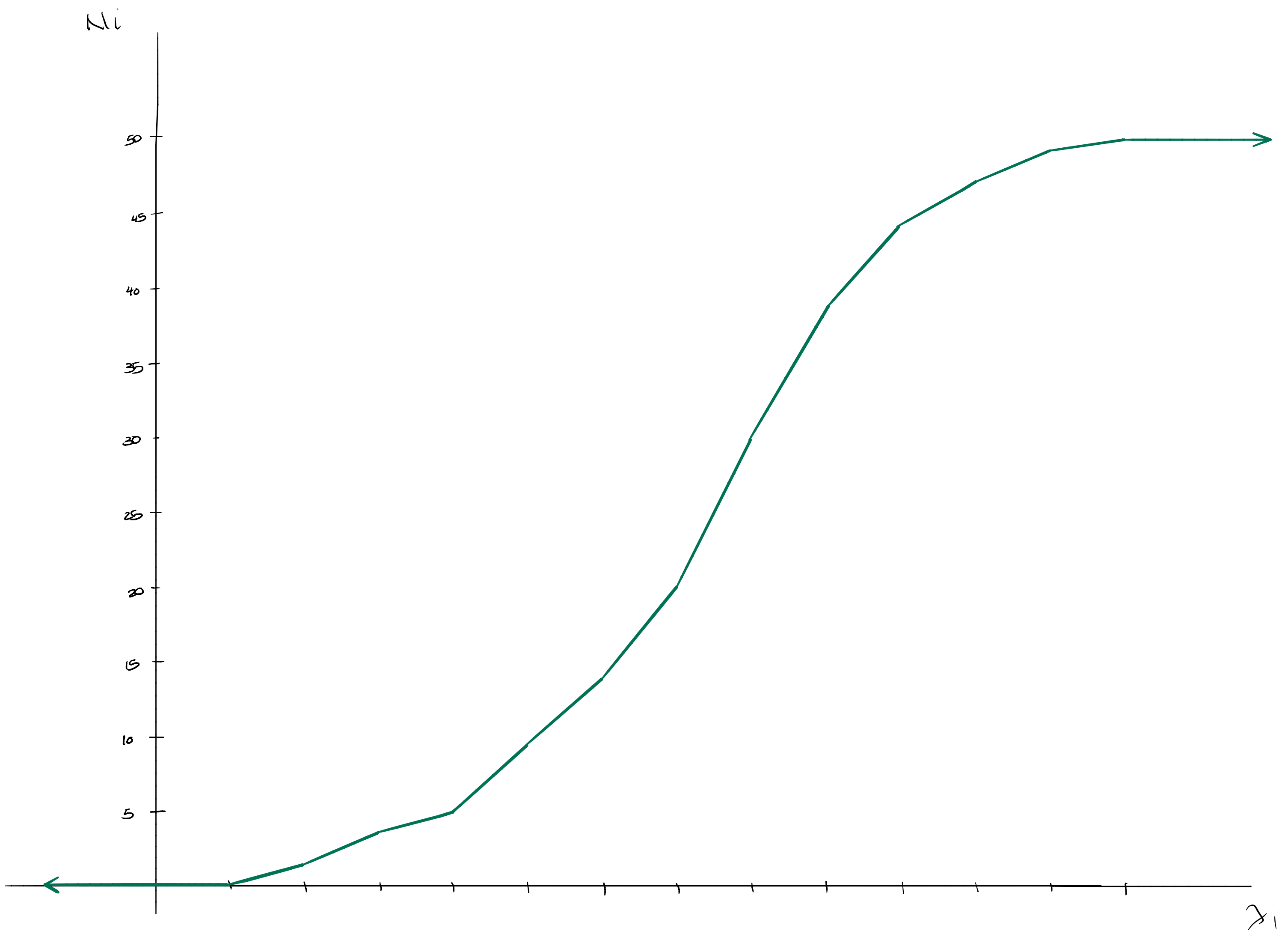
b) Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.

Intervalo	n_i	N_i	f_i	c_i	a_i	$h_i = \frac{n_i}{a_i n}$
[140,145]	2	2	0,04	142,5	5	2/250
]145, 150]	2	4	0,04	147,5	5	2/250
]150, 155]	1	5	0,02	152,5	5	1/250
]155, 160]	4	9	0,08	157,5	5	4/250
]160, 165]	5	14	0,1	162,5	5	5/250
]165, 170]	6	20	0,12	167,5	5	6/250
]170, 175]	10	30	0,2	172,5	5	10/250
]175, 180]	8	38	0,16	177,5	5	8/250
]180, 185]	6	44	0,12	182,5	5	6/250
]185, 190]	3	47	0,06	187,5	5	3/250
]190, 195]	2	49	0,04	192,5	5	2/250
]195, 200]	1	50	0,02	197,5	5	1/250
	n=50		1			

Histograma y poligonal



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						
$, 10200]$		5					
				10/18		1100	
			2/18		12000		
	4						0.005/18
		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.

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?

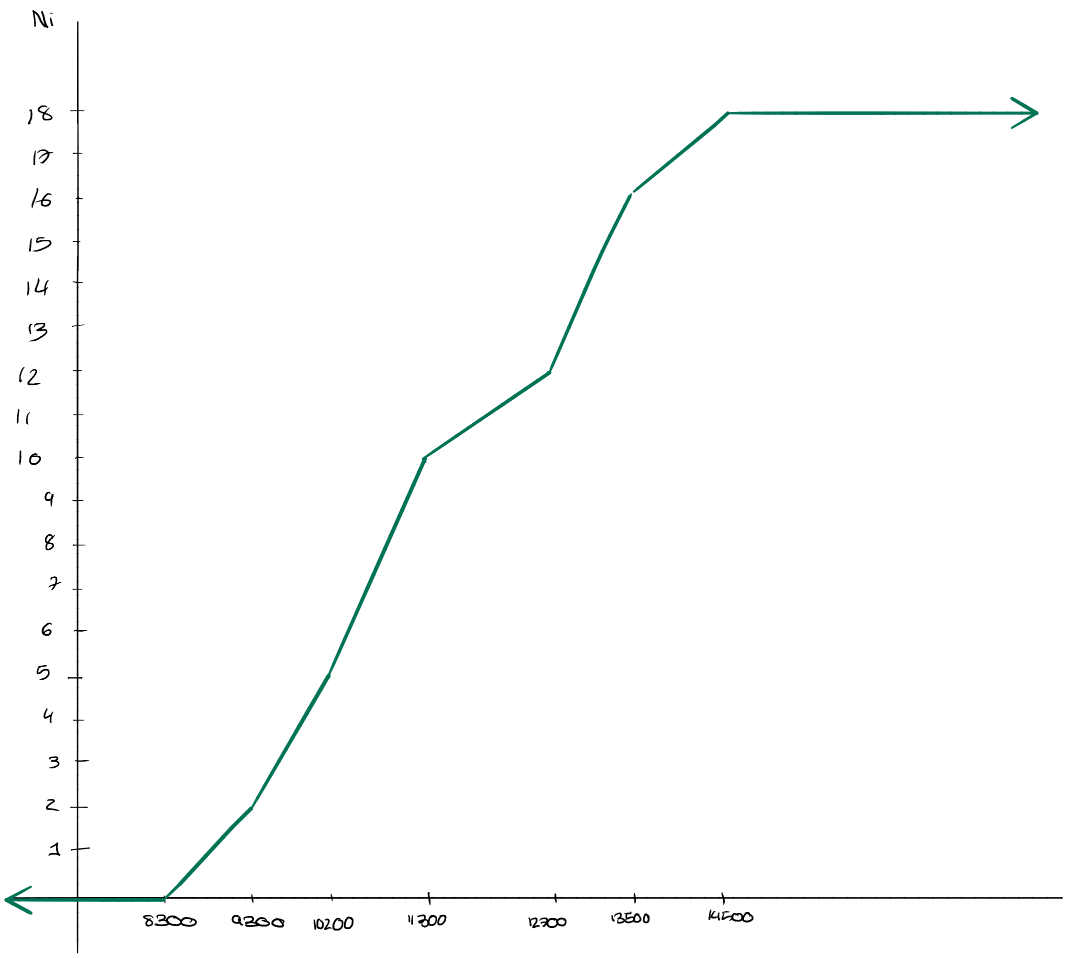
(*) En una población de $n = 18$ comunidades autónomas, se ha observado una variable estadística X = Renta familiar media en el año 2003, que ha presentado 6 modalidades distintas con distribución de frecuencia $\{x_i, n_i\}$ $i = 1, \ldots 6$

Es importante aclarar que este problema no tiene interés estadístico, ya que no existe ningún país con 18 comunidades autónomas. Si el problema quisiera referirse a España sería más apropiado hablar de núcleos urbanos o utilizar otra terminología más apropiada.

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	$0{,}00\bar{3}/18$
$(10200, 11300]$	5	10	5/18	10/18	10750	1100	$0{,}00\bar{45}/18$
$(11300, 12700]$	2	12	2/18	12/18	12000	1400	0,0015/18
$(12700, 13500]$	4	16	4/18	16/18	13100	800	0.005/18
$(13500, 14500]$	2	18	2/18	18/18	14000	1000	0.002/18
	$n=18$		1				

Histograma y poligonal de frecuencias

Curva de distribución



c) 12 comunidades presentan una renta menor o igual a 12700 euros y 8 comunidades presentan una renta superior a 11300 euros.

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.

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

c) ¿Cuál es el número mediano de piezas defectuosas por caja?

d) Calcular los cuartiles de la distribución. Interpretarlos.

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

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

(*) En una población de tamaño $n = 100$ cajas se ha observado una variable estadística X = número de piezas defectuosas por caja, que ha presentado 11 modalidades distintas, con distribución de frecuencia $\{x_i, n_i\}$ $i = 1, \ldots, 11$

a) Este apartado equivale a calcular la media aritmética:

$$\bar{x} \; \frac{1}{100} \sum_{i=1}^{11} x_i n_i = 4{,}36 \text{ piezas defectuosas por caja}$$

b) Nos piden calcular la moda (La modalidad con mayor frecuencia)
En este caso los números de piezas defectuosas que se encuentran más frecuentemente son 5 y 6.

c) Nos pide calcular la mediana:

$\frac{n}{2} = 50 \Rightarrow Me = 4{,}5$ (Tomamos la media aritmética de 4 y 5 porque en $x_i = 4$ se alcanza $N_i = 50$, pero no sería del todo representativo, ya que no supera dicha frecuencia absoluta acumulada)

d) Los cuartiles son:

$Q_1 = P_{25} = 2{,}5$. Por la misma explicación que con la mediana tomamos la media aritmética de 2 y 3

$\frac{nr}{100} = \frac{\cancel{100} \cdot 25}{\cancel{100}}$

Esto quiere decir que el 25% no tienen más de 2 piezas defectuosas

$Q_2 = P_{50} = 4,5$. Esto quiere decir que el 50 % de las cajas no tienen más de 4 piezas defectuosas

$\frac{nr}{100} = 50$

→ Por la misma explicación de los casos anteriores tomamos la media ($Q_2 = P_{50} = Me$)

$Q_3 = P_{75} = 6$. Más del 75 % de las cajas tienen 6 o menos piezas defectuosas

$\frac{nr}{100} = 75$

e) Los deciles de orden 3 y 7 son:

$P_{30} = 3$ ⇒ Más del 30 % de las cajas tienen 3 o menos piezas defectuosas

$\frac{nr}{100} = 30$

$P_{70} = 6$ ⇒ Más del 70 % de las cajas tienen 6 o menos piezas defectuosas.

$\frac{nr}{100} = 70$

f)

- Recorrido $= x_k - x_1 = 10 - 0 = 10$
  Nos indica que ninguna caja va a tener más de 10 piezas defectuosas en comparación a otra.
  - Ventaja: Fácil de calcular
  - Inconveniente: No aporta gran información

- Recorrido intercuartílico $= Q_3 - Q_1 = 3,5$
  Indica que el 50 % de las cajas están a una distancia menor o igual a 3,5 , es decir, el 50 % de las cajas tienen menos de 4 piezas defectuosas de diferencia.
  - Ventaja: Fácil de calcular
  - Desventaja: No da información de como es la distribución entre el primer y el tercer cuartil, ni fuera de estos.

- Desviación absoluta media respecto a la media aritmética ($\bar{x}$):

$$D\bar{x} = \frac{\sum_{i=1}^{11} |x_i - \bar{x}| n_i}{100} = \text{aproximadamente 2 piezas defectuosas}$$

- Ventaja: Da información sobre la dispersión de la distribución.

- Desviación absoluta media respecto a la mediana (Me):

$$DMe = \frac{\sum_{i=1}^{11} |x_i - Me| n_i}{100} = \text{aprox. 2 piezas defectuosas}$$

- Ventajas: Da información sobre la dispersión de la distribución respecto a la mediana.

- Varianza:

$$\sigma^2 = \frac{1}{100} \sum_{i=1}^{11} n_i (x_i - \bar{x})^2 = 5,88 \text{ (piezas defectuosas)}^2$$

- Inconveniente: Da unidades cuadráticas

- Desviación típica:

$$\sigma = \sqrt{\sigma^2} = 2,42 \text{ piezas defectuosas}$$

- Recorrido relativo:

$$RR = \frac{R}{\bar{x}} = \frac{10 - 0}{\bar{x}} = 2,294$$

- Ventaja: Fácil de calcular y al ser adimensional, se puede comparar con otras distribuciones.

- Recorrido semi-intercuartílico:

$$RSI = \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0,412$$

- Ventaja: Fácil de calcular, es adimensional por lo que se puede comparar con otras distribuciones e indica hacia que lado está más desviado la distribución.

- Coeficiente de variación de Pearson:

$$C.V.(x) = \frac{\sigma_x}{|\bar{x}|} = 0{,}556$$

-Ventaja: Fácil de calcular y al ser adimensional, se puede comparar con otras distribuciones.

- Índice de dispersión respecto a la mediana:

$$V_{Me} = \frac{D_{Me}}{Me} = 0{,}44$$

-Ventaja: Fácil de calcular y al ser adimensional, se puede comparar con otras distribuciones.

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:

a) Medias aritmética, armónica y geométrica.

b) El valor más frecuente.

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

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

I_i	n_i	N_i	c_i	$n_i c_i$	h_i	$n_i c_i^2$
(0,1]	12	12	0,5	6	12	3
(1,2]	13	25	1,5	19,5	13	29,5
(2,3]	11	36	2,5	27,5	11	68,75
(3,4]	8	44	3,5	28	8	98
(4,5]	6	50	4,5	27	6	121,5
	$n=50$			2,16		6,41

a) Calcular las distintas medias

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{5} c_i n_i = 2{,}16$$

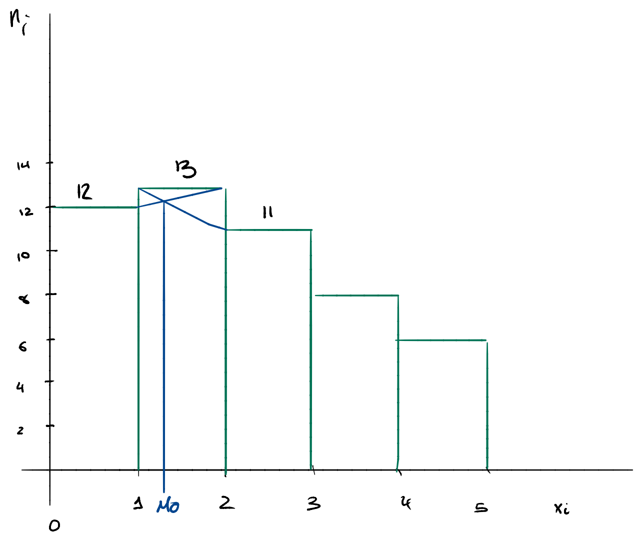
$$G_1 = \sqrt[50]{\prod_{i=1}^{5} c_i^{n_i}} = 1{,}684$$

$$H_1 = \frac{n}{\sum_{i=1}^{5} \frac{n_i}{c_i}} = 1{,}228$$

b) Nos piden calcular la moda

$$\frac{13-12}{13-12+13-11} = \frac{Mo-1}{2-1}$$

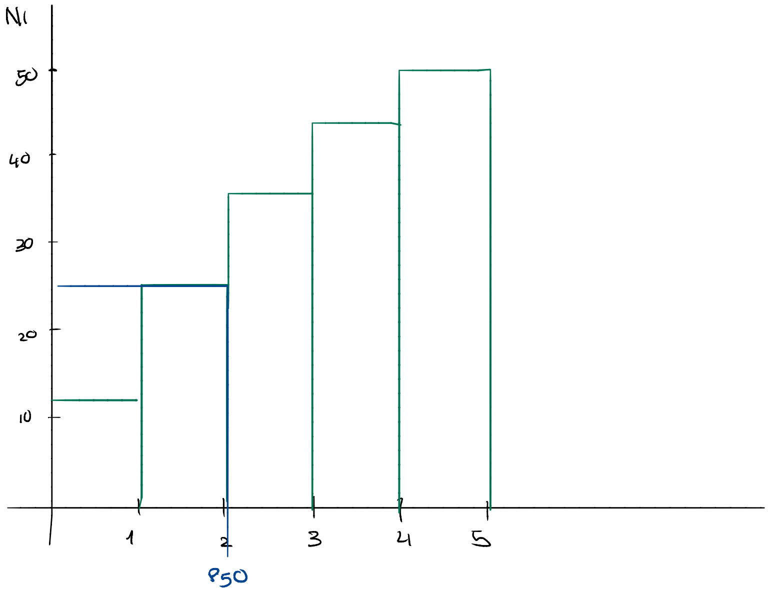
$$Mo = \frac{4}{3}$$



Nos pide calcular el percentil 50

$$\frac{nr}{100} = \frac{50 \cdot 50}{100} = 25$$

$P_{50} = 2$



d) Para calcular el recorrido intercuartílico necesitamos los cuartiles 3 y 1 :

$Q_1 = P_{25}$

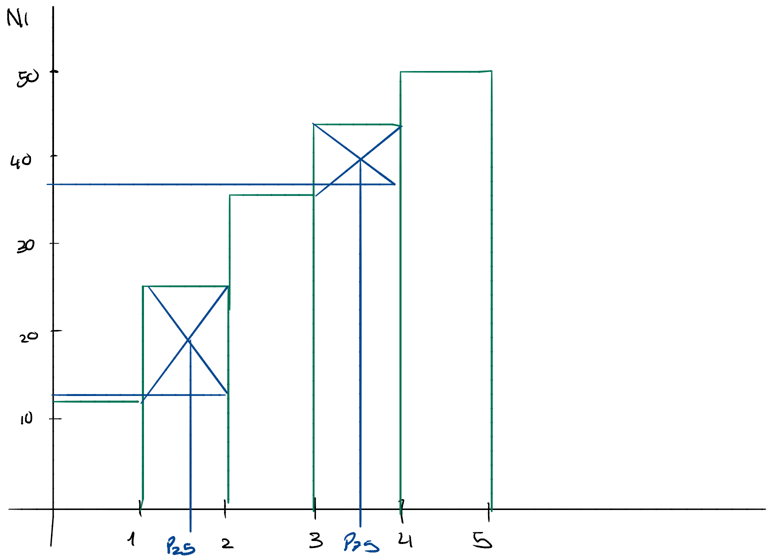
$$\frac{nr}{100} = 12{,}5$$

$$\frac{P_{25} - 1}{2 - 1} = \frac{12{,}5 - 12}{26 - 12} \Rightarrow P_{25} = \frac{29}{28}$$

$Q_3 = P_{75}$

$$\frac{nr}{100} = 37{,}5$$

$$\frac{P_{75} - 3}{4 - 3} = \frac{37{,}5 - 36}{44 - 36} \Rightarrow P_{75} = \frac{51}{16}$$



Luego $R_I = Q_3 - Q_1 = \frac{51}{6} - \frac{29}{28} = \frac{241}{112}$

$R = x_k - x_1 = 5 - 0 = 5$

$\sigma_x = +\sqrt{\sigma_x^2} = \sqrt{m_2 - m_1^2} = \sqrt{6{,}41 - 2{,}16^2} = 1{,}32$

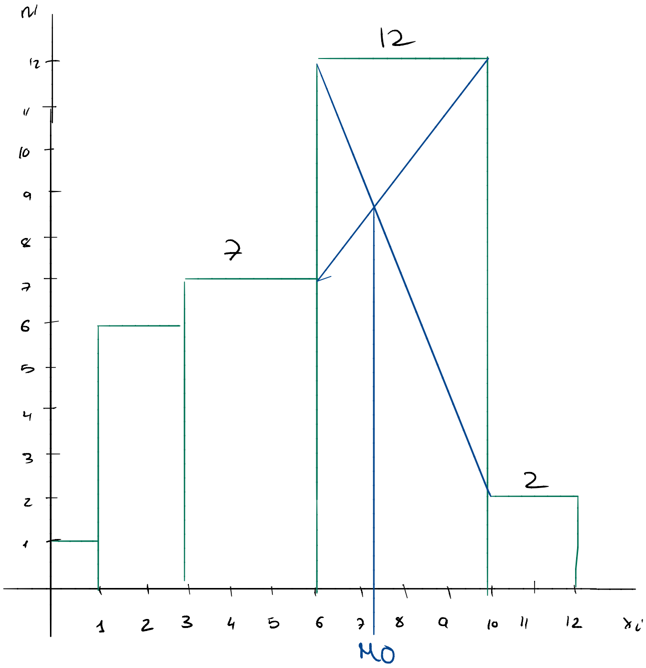
I_i	n_i	N_i	c_i	$n_i c_i$	$n_i c_i^2$	h_i
(0,1]	1	1	0,5	0,5	0,25	1
(1,3]	6	7	2	12	24	3
(3,6]	7	14	4,5	31,5	141,75	2,33
(6,10]	12	26	8	96	768	3
(10,12]	2	28	11	22	242	1
	28			162	1176	

a) Calcular las distintas medias

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{5} c_i n_i = \; 5,78$$

$$G_1 = \sqrt[28]{\prod_{i=1}^{5} c_i^{n_i}} = 4,769$$

$$H_1 = \frac{n}{\sum_{i=1}^{5} \frac{n_i}{c_i}} = 5,7857$$



b) Calcular la moda:

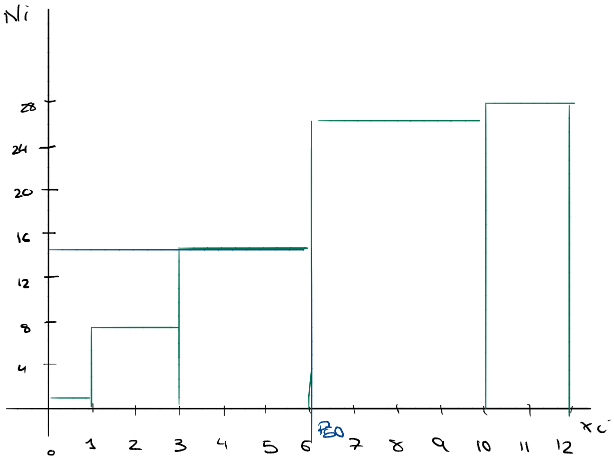
$$\frac{12-7}{12-7+12-2} = \frac{Mo-6}{10-6}$$

$$Mo = \frac{22}{3}$$

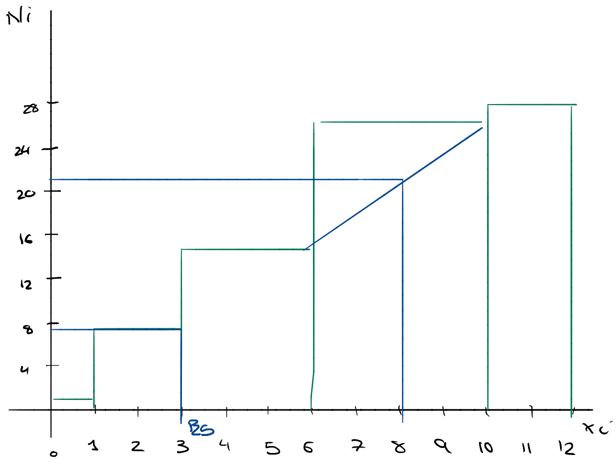
c) Hay que calcular el percentil 50:

$$C_{50} = \frac{nr}{100} = \frac{28 \cdot 50}{100} = 14$$

$$P_{50} = 6$$



d) Para el recorrido intercuartílico necesitamos Q_1, Q_3:



$Q_1 = P_{25} = 3$

$$C_{25} = \frac{nr}{100} = \frac{28 \cdot 25}{100} = 7$$

$Q_3 = P_{75}$

$$C_{75} = \frac{nr}{100} = \frac{75 \cdot 28}{100} = 21$$

$$\frac{P_{75} - 6}{10 - 6} = \frac{21 - 14}{26 - 14}$$

$$P_{75} = \frac{25}{3}$$

Luego $R_I = Q_3 - Q_1 = 16/3$

$R = x_k - x_1 = 12 - 0 = 12$

$$\sigma_x = +\sqrt{\sigma_x^2} = \sqrt{m_2 - m_1^2} = \sqrt{42 - \left(\frac{81}{14}\right)^2} = 2{,}92$$

Para ver que distribución es más homogénea calculamos el coeficiente de variación de Pearson:

$$CV(x_1) = \frac{\sigma_{x_1}}{|\bar{x}_1|} = \frac{1{,}32}{2{,}16} = \frac{11}{18} \simeq 0{,}61$$

$$CV(x_2) = \frac{\sigma_{x_2}}{|\bar{x}_2|} = \frac{2{,}92}{81/14} = 0{,}5$$

Como $CV(x_2) < CV(x_1) \Rightarrow$ La segunda distribución es más homogénea

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.

(*) En una población de tamaño $n = 2$ se ha observado una variable estadística X = velocidad en $\frac{Km}{h}$, que ha presentado 2 modalidades distintas, $60 \frac{Km}{h}$, $70 \frac{Km}{h}$, con distribución de frecuencia. $\{60, 1\}, \{70, 1\}$

Primer tramo $60 \frac{Km}{h}$:

$$x = \cancel{x_0}^{\,0} + vt \Rightarrow t_1 = \frac{100 \cancel{Km}}{60 \frac{\cancel{Km}}{h}} = \frac{10}{6} \text{ horas}$$

Segundo tramo $70 \frac{Km}{h}$

$$t_1 = \frac{100 \cancel{Km}}{70 \frac{\cancel{Km}}{h}} = \frac{10}{7} \text{ horas}$$

$$\text{Velocidad media} = \frac{\Delta x}{\Delta t} = \frac{200 \text{ Km}}{t_1 + t_2} = \boxed{64{,}6 \frac{Km}{h}}$$

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.

En una población de tamaño $n=5$ se ha observado una variable estadística X = Rentabilidad de las empresas, que ha presentado $k=5$ modalidades distintas

x_i	n_i	N_i	c_i
5	1	1	1,05
6	1	2	1,06
7	1	3	1,07
10	1	4	1,1
12	1	5	1,12

Como se trata de una variable con rendimientos acumulativos, calcularemos la media geométrica:

$$G = \sqrt[5]{1{,}05 \cdot 1{,}06 \cdots 1{,}12} - 1 = 7{,}97\ \%$$

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

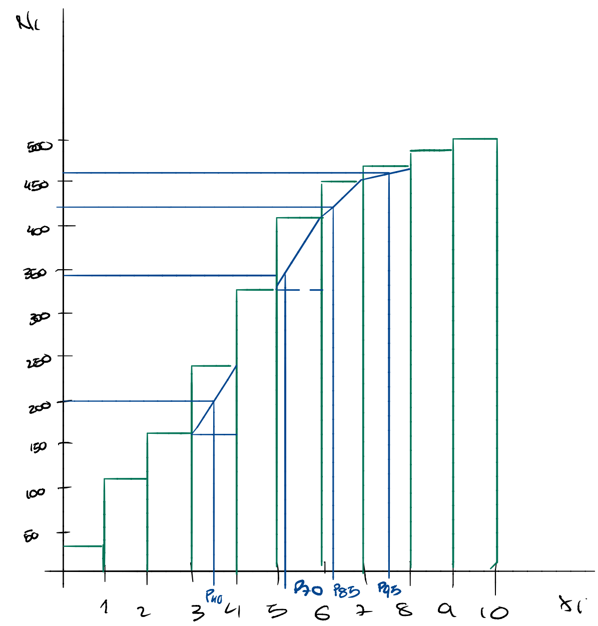
Calcular las notas máximas para obtener cada una de las calificaciones.

El ejercicio nos pide calcular los percentiles

x_i	(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]	(5, 6]	(6, 7]	(7, 8]	(8, 9]	(9, 10]
n_i	34	74	56	81	94	70	41	28	16	4
N_i	34	108	164	245	339	409	450	478	494	498

En una población $n = 498$ se ha observado una variable estadística $X =$
$=$ calificaciones de los alumnos que ha presentado 10 modalidades distintas

Hay que calcular el percentil 40, 70, 85, 95:



$$\frac{nr}{100} = 199,2$$

$$\frac{200 - 164}{245 - 164} = \frac{P_{40} - 3}{4 - 3} \Rightarrow P_{40} = 3,44 \text{ de nota}$$
(Máximo del suspenso)

$$\frac{nr}{100} = 348,6$$

$$\frac{348,6 - 339}{409 - 339} = \frac{P_{70} - 5}{6 - 5} \Rightarrow P_{70} = 5,137 \text{ de nota}$$
(máximo del aprobado)

$$\frac{nr}{100} = 423,3$$

$$\frac{P_{85} - 6}{7 - 6} = \frac{423,3 - 409}{450 - 409} \Rightarrow P_{85} = 6,348 \text{ de nota}$$
(Máximo del notable)

$$\frac{nr}{100} = 473,1$$

$$\frac{P_{95} - 7}{8 - 7} = \frac{473,1 - 450}{478 - 450} \Rightarrow P_{95} = 7,825 \text{ de nota}$$
(Nota máxima sobresaliente)

El máximo de la matrícula es 10, ya que buscamos la frecuencia absoluta acumulada asociada al 100 % de la población.

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

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

b) Si se consideran altos el 18 % de los individuos de mayor altura, ¿cuál es su altura mínima?

c) ¿Qué altura es superada sólo por 1/4 de los jóvenes?

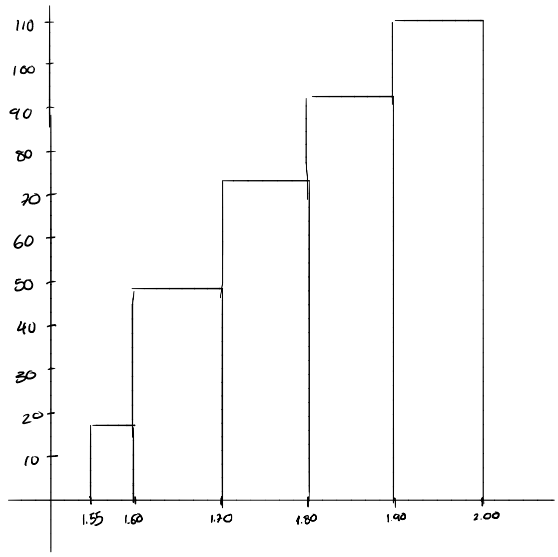
d) Calcular el número de jóvenes cuya altura es superior a 1.75.

e) Calcular la altura máxima de los 11 jóvenes más bajos.

f) Calcular la altura mínima de los 11 jóvenes más altos.

En una población de tamaño $n = 110$ jóvenes, se ha observado una variable estadística X = Altura, que ha presentado $K = 5$ modalidades distintos con distribución de frecuencia $\{x_i, n_i\}$ $i = 1, \dots 5$

X_i	n_i	N_i
(1.55, 1.60]	18	18
(1.60, 1.70]	31	49
(1.70, 1.80]	24	73
(1.80, 1.90]	20	93
(1.90, 2.00]	17	110
	110	

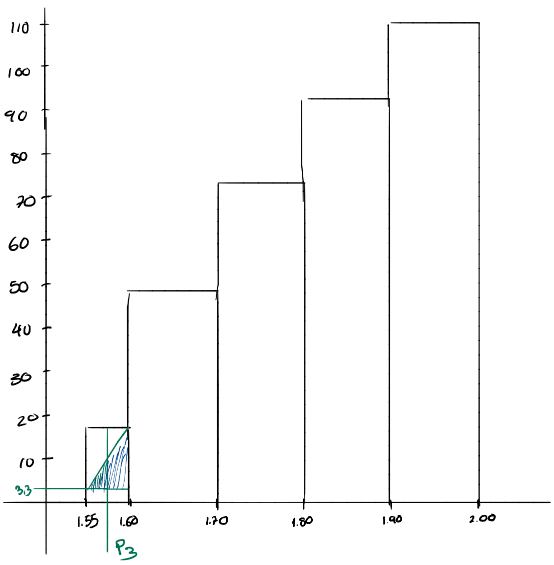


a) Hay que calcular el percentil 3

$$C_3 = \frac{n \cdot r}{100} = \frac{110 \cdot 3}{100} = 3{,}3$$

$$\frac{P_3 - 1.55}{1.60 - 1.55} = \frac{3.3 - 0}{18 - 0}$$

$P_3 = 1.5591$ metros

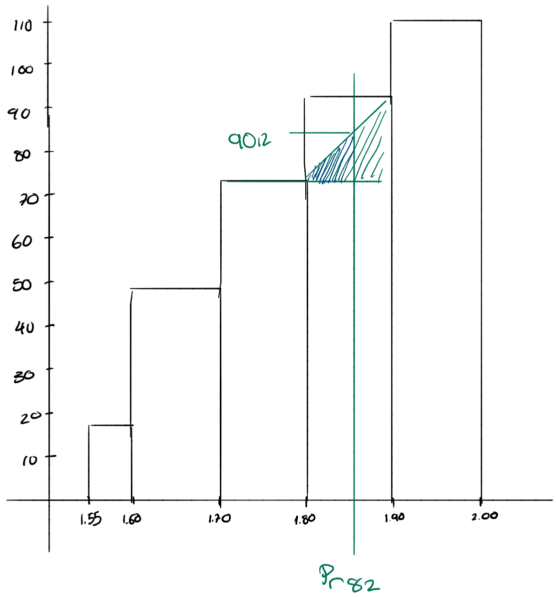


b) Hay que calcular el percentil 82:

$$C_{82} = \frac{110 \cdot 82}{100} = 90,2$$

$$\frac{Pr_{82} - 1.80}{1.90 - 1.80} = \frac{90.2 - 73}{93 - 73}$$

$Pr_{82} = 1,886$ metros

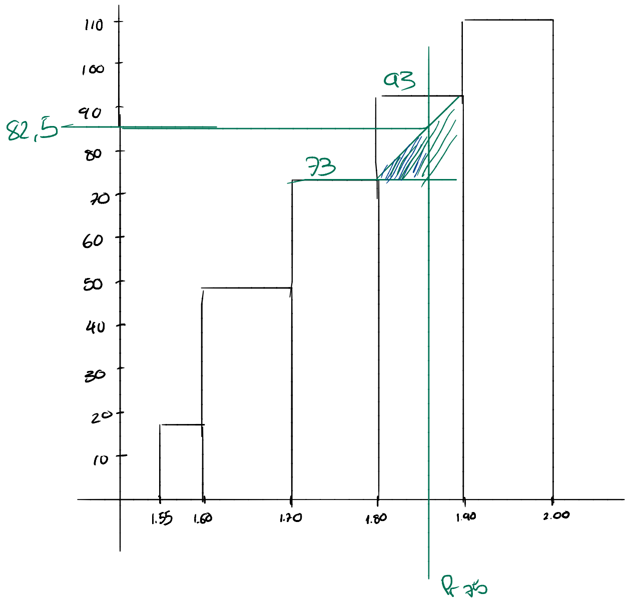


c) Nos piden calcular el percentil 75:

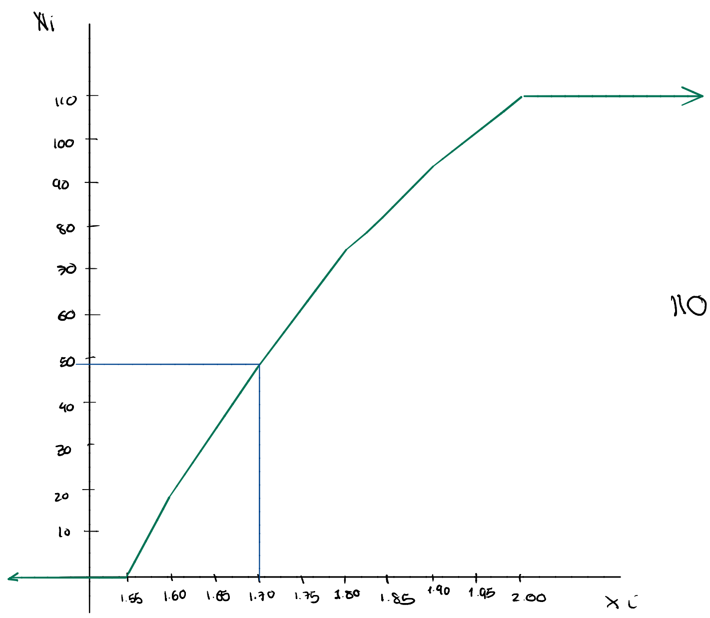
$$C_{75} = \frac{110 \cdot 75}{100} = 82,5$$

$$\frac{Pr_{75} - 1.80}{1.90 - 1.80} = \frac{82,5 - 73}{93 - 73}$$

Pr_{75}: 1.8475 metros



d)



110 − 49 = 61 jóvenes

e) Hay que calcular el siguiente percentil:

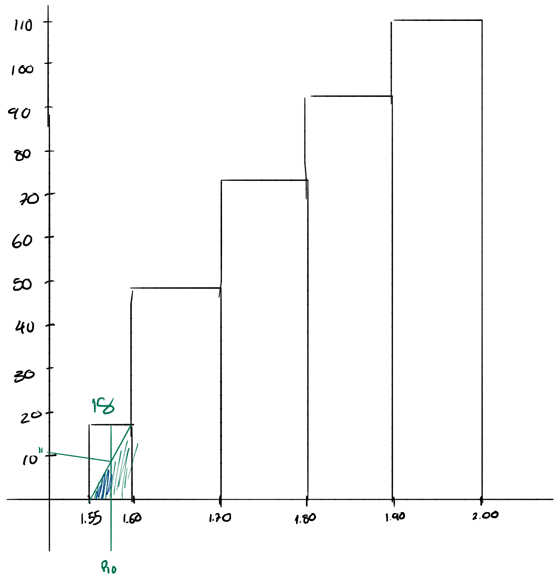
11 personas de 110 = $\frac{11}{110} = 0,1 = 10\%$

Luego, hay que calcular el P_{10}

$$C_{10} = \frac{10 \cdot 110}{100} = 11$$

$$\frac{P_{10} - 1.55}{1.60 - 1.55} = \frac{11 - 0}{18 - 0}$$

$P_{10} = 1.58$ metros

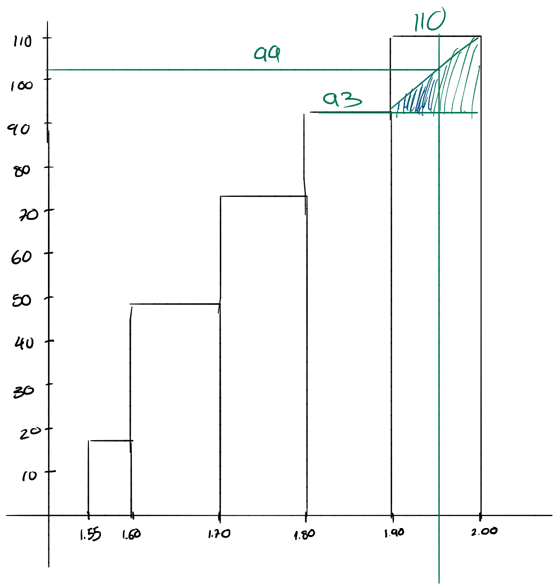


f) Hay que calcular el P_{90}:

$$C_{90} = \frac{90 \cdot 110}{100} = 99$$

$$\frac{P_{90} - 1.90}{2.00 - 1.90} = \frac{99 - 93}{110 - 93}$$

$P_{90} = 1.935$ metros



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.

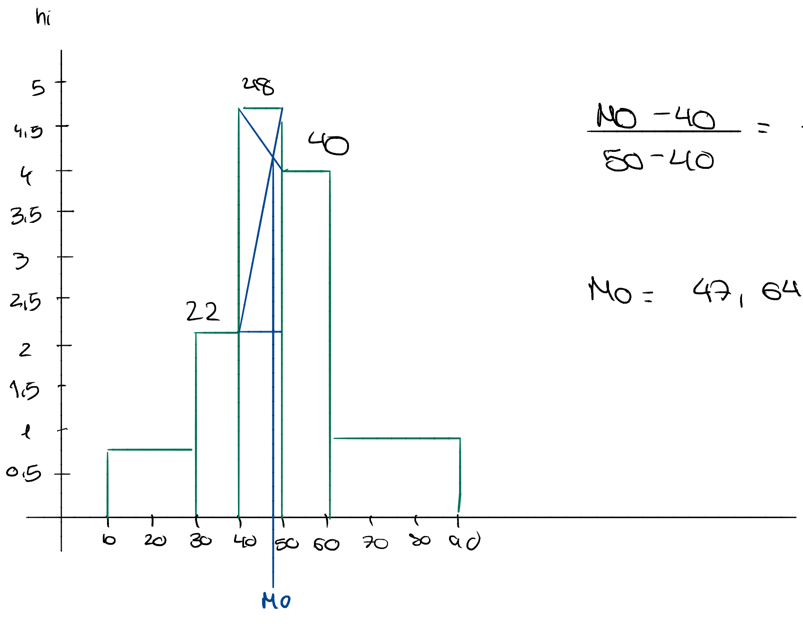
c) Calcular el recorrido intercuartílico y la desviación típica.

d) Calcular e interpretar los valores de los coeficientes de asimetría y curtosis.

En una población de tamaño $n = 150$ una variable estadística X = edad de los enfermos, que ha presentado $K = 5$ modalidades distintas

x_i	n_i	N_i	c_i	$n_i c_i$	$n_i c_i^2$	h_i
(10, 30]	15	15	20	300	6000	3/4
(30, 40]	22	37	35	770	26950	2,2
(40, 50]	48	85	45	2160	97200	4,8
(50, 60]	40	125	55	2200	121000	4
(60, 90]	25	150	75	1875	140625	5/6
	150			7305	391775	

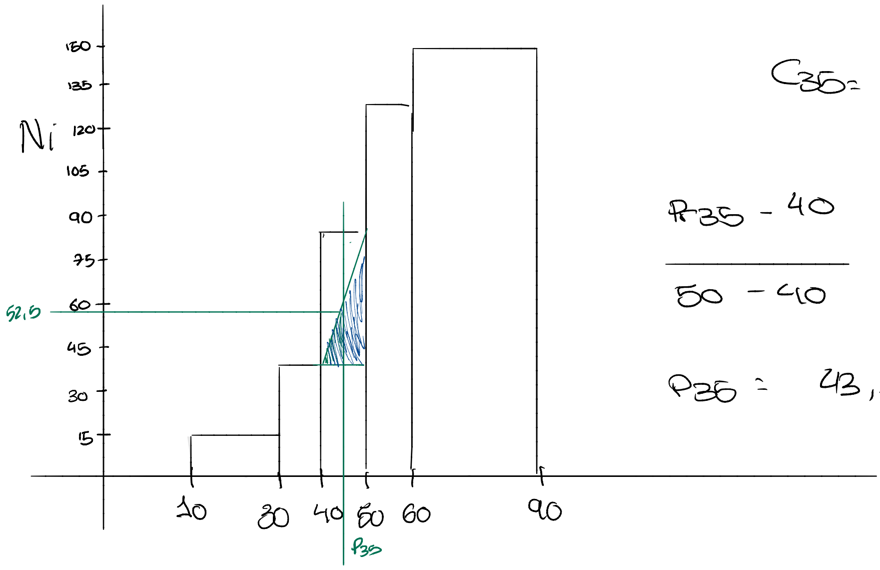
a) Nos pide calcular la moda, que en este caso es un intervalo modal



$$\frac{Mo - 40}{50 - 40} = \frac{48 - 22}{48 - 22 + 48 - 40}$$

$Mo = 47,647$ años

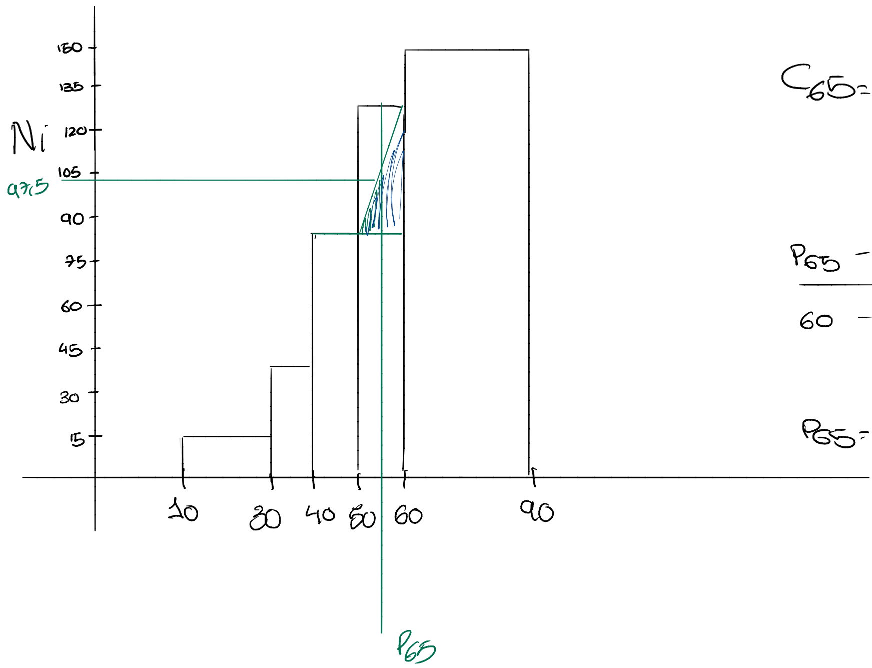
b) Equivale a calcular P_{35} y P_{65}



$$C_{35} = \frac{35 \cdot 150}{100} = 52{,}5$$

$$\frac{P_{35} - 40}{50 - 40} = \frac{52{,}5 - 37}{85 - 37}$$

$P_{35} = 43{,}22$ años (el mínimo 30%)



$$C_{65} = \frac{65 \cdot 150}{100} = 97{,}5$$

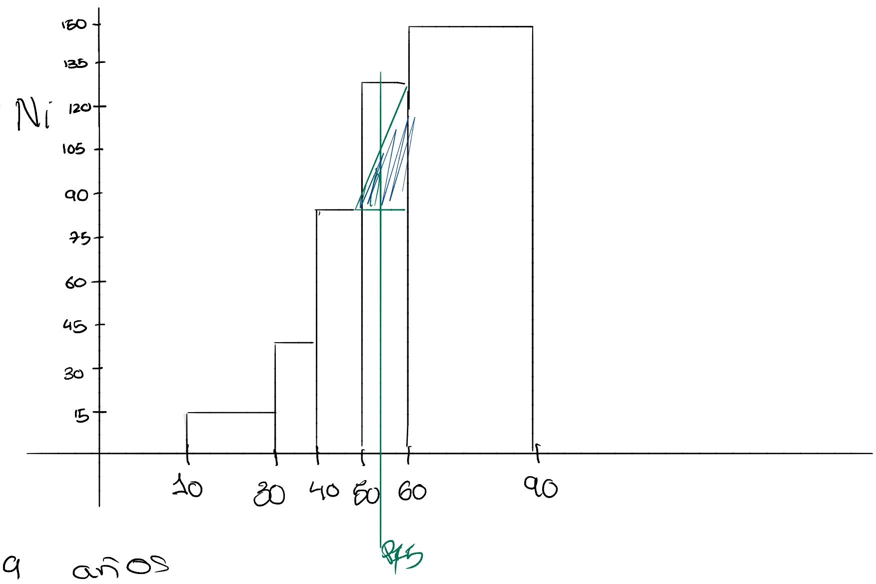
$$\frac{P_{65} - 50}{60 - 50} = \frac{97{,}5 - 85}{125 - 85}$$

$P_{65} = 53{,}125$ años (Edad máxima)

c) $RI = Q_3 - Q_1$

$Q_3 = P_{75}$

$$C_{75} = \frac{150 \cdot 75}{100} = 112{,}5$$



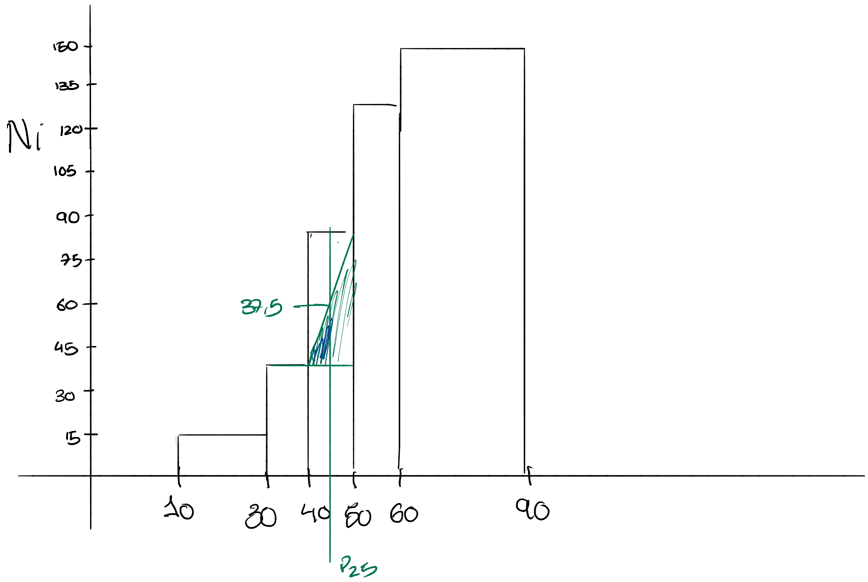
$$\frac{P_{75} - 50}{60 - 50} = \frac{112{,}5 - 85}{125 - 85} = 56{,}9 \text{ años}$$

$Q_1 = P_{25}$

$C_{25} = \frac{25 \cdot 150}{100} = 37,5$

$\frac{P_{25}-40}{50-40} = \frac{37,5-37}{85-37}$

$P_{25} = 40,1$ años



$R_I = Q_3 - Q_1 = 16,8$ años

c) calculamos:

$\sigma_x = +\sqrt{\sigma_x^2} = \sqrt{m_2 - m_1^2} = \sqrt{2611,83 - 48,7^2} = 15,49$ años

d) Coeficiente de asimetría

$$\gamma_1 = \frac{\mu_3}{\sigma_x^3} = \frac{1}{x}\sum_{i=1}^{n} n_i \left(\frac{x_i-\bar{x}}{\sigma_x}\right)^3 \Rightarrow \gamma_1(x) = 0,0917 > 0 \Rightarrow$$

$\Rightarrow$ La distribución es asimétrica por la derecha

coeficiente de curtosis de Fisher:

$\gamma_2(X) = \frac{\mu_4}{\sigma^4} - 3 = -0,277196 \Rightarrow$ La distribución es mesocúrtica

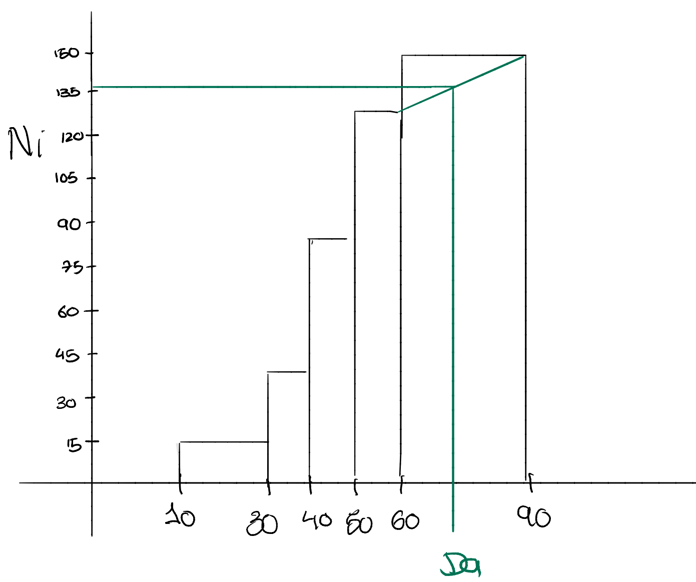
Coeficiente de curtosis de Kelley:

$K = \frac{1}{2}\frac{Q_3-Q_1}{D_9-D_1} - 0,263$

$C_{90} = \frac{nr}{100} = \frac{150 \cdot 90}{100} = 135$

$\frac{135-125}{150-125} = \frac{D_9-60}{90-60}$

$D_9 = 72$ años



$$C_{10} = \frac{nr}{100} = \frac{150 \cdot 10}{100} = 15$$

$D_9 = 30$ años

$$K = \frac{1}{2} \cdot \frac{56{,}875 - 40{,}104167}{72 - 30} - 0{,}263 =$$

$$= -0{,}0632976 \Rightarrow$$

⇒ La distribución es mesocúrtica