# ESTADÍSTICA DESCRIPTIVA E INTRODUCCIÓN A LA PROBABILIDAD: RELACIÓN 1

## *GRUPO: B4*


*AUTORES:*

- Alejandro Moreno Guerrero
- Pablo Nieto López
- Daniel Pérez Ruiz
- David Suárez González
- Daniel Zufrí Quesada

EDSP: RELACIÓN 1

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

a) COMPLETAR LA TABLA

| $x_i$ | $n_i$ | $N_i$ | $f_i$ |
|---|---|---|---|
| 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 |

(*) En una población de familias de una barriada, de tamaño 500, se ha observado una variable estadística X = "nº de hijos de cada familia", que ha presentado 7 modalidades diferentes {0, 1, ..., 6} con una distribución de frecuencias $\{x_i, n_i\}_{i=1,...,7}$

→) Para calcular el tamaño total de la población (n). Hemos utilizado lo siguiente:

→) Como $f_i = \frac{n_i}{n} \Rightarrow \boxed{n = \frac{n_i}{f_i}}$ ; $n = \frac{n_1}{f_1} = \frac{80}{0'16} = 500$ familias

→) Cálculo de datos:

→) $N_1 = n_1 = 80$

→) $N_2 = N_1 + n_2 = 190$

→) $f_2 = \frac{n_2}{n} = 0'22$

→) $n_3 = N_3 - N_2 = 130$

→) $f_3 = \frac{n_3}{n} = 0'26$

→) $n_4 = f_4 \cdot n = 90$

→) $N_4 = n_4 + N_3 = 410$

→) $N_5 = n_5 + N_4 = 450$

→) $f_5 = \frac{n_5}{n} = 0'08$

→) $N_7 = 500$

→) $N_6 = N_7 - n_7 = 480$

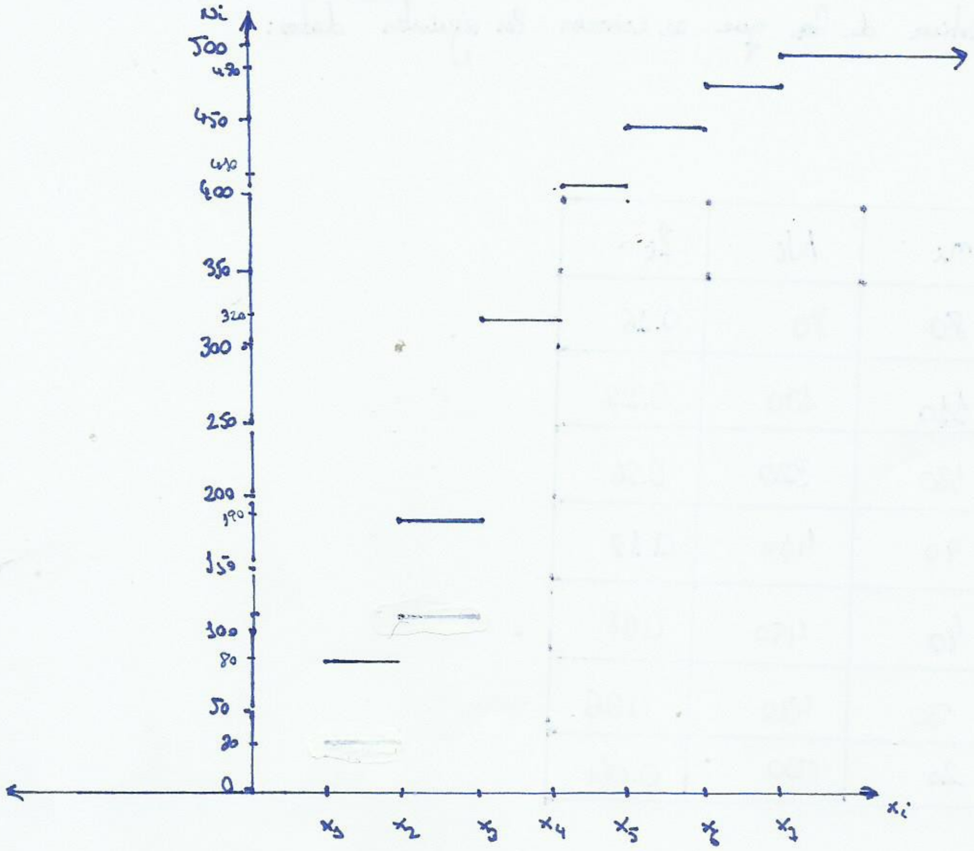
→) $n_6 = N_6 - N_5 = 30$

→) $f_6 = \frac{n_6}{n} = 0'06$

→) $f_7 = \frac{n_7}{n} = 0'04$

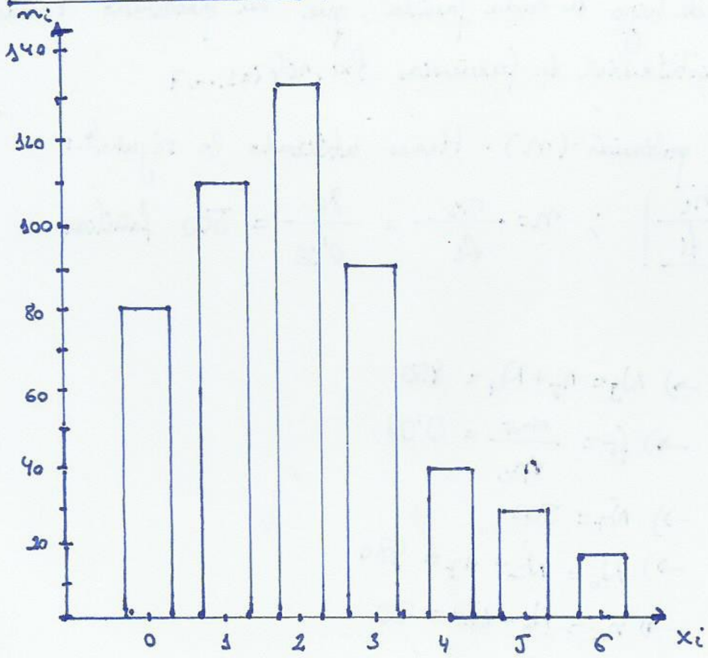
b) Representar la distribución mediante un diagrama de barras y la curva de distribución.

→ CURVA DE DISTRIBUCIÓN




Recordemos que esta función está definida en todo $\mathbb{R}$.

→ DIAGRAMA DE BARRAS



EDIP : RELACIÓN 1

**1.** c) Promediar los valores de la variable mediante diferentes medidas. Interpretarlas:

→) MEDIA: Calcularemos la única que tenga realmente sentido. En este caso, la media aritmética:

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

$$\bar{x} = \frac{1}{500} \cdot (80 \cdot 0 + 110 \cdot 1 + 130 \cdot 2 + 90 \cdot 3 + 40 \cdot 4 + 30 \cdot 5 + 20 \cdot 6) = 2'12 \text{ hijos/familia}$$

→) MEDIANA:

$$\boxed{Me = x_i : \; N_i > n/2} \qquad \frac{500}{2} = 250 \; ; \quad N_3 > 250 \Rightarrow \boxed{Me = 2 \text{ hijos}}$$

→) MODA:

$$\boxed{Mo = x_i : \; x_i \geq n_j \;\; \forall j = 1, \ldots, k} \Rightarrow Mo = 2 \text{ hijos}$$

→) La mediana y la moda coinciden en este caso y la media de hijos por familia es de 2'12 hijos por familia.

$\boxed{2.}$ La puntuación obtenidas por 50 personas que se presentaron a una prueba de selección, sumadas las puntuaciones de los distintos test fueron las mostradas.

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

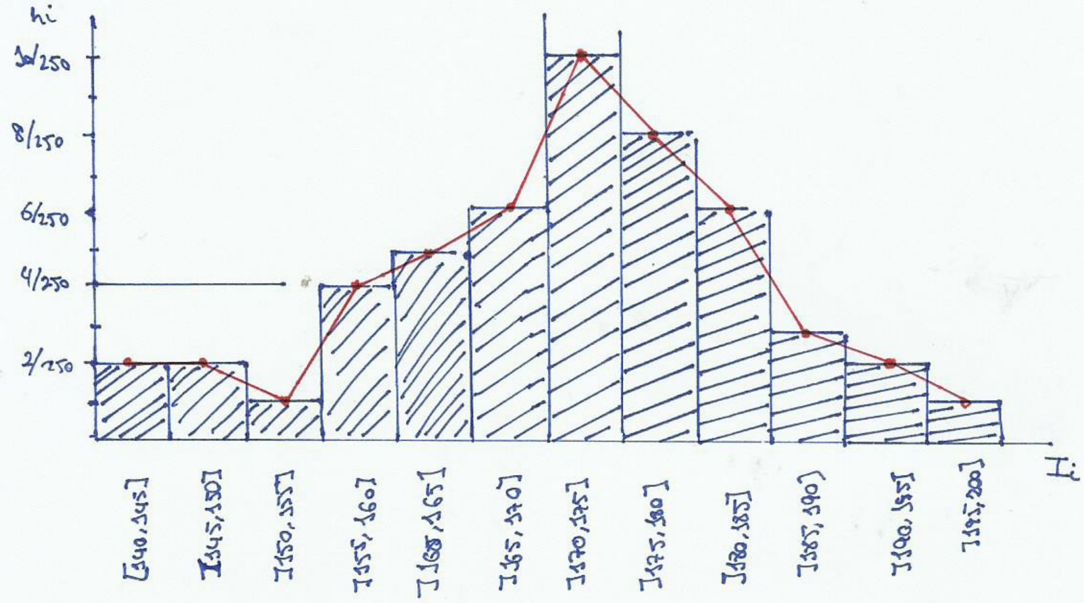
→) ⊛ En una población de tamaño 50 (población de personas), se ha observado una variable estadística $X$ = "Resultados de un test", que ha presentado 12 modalidades distintas, con una distribución de frecuencias $\{x_i, n_i\}_{i=1,\dots,12}$

| $I_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ | $c_i$ | $a_i$ | $h_i$ |
|---|---|---|---|---|---|---|---|
| ⊛ [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 |
| ]175, 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 |
| $\sum_{i=1}^{12} n_i$ → | n=50 | | 1 | ← $\sum_{i=1}^{12} f_i$ | | | |

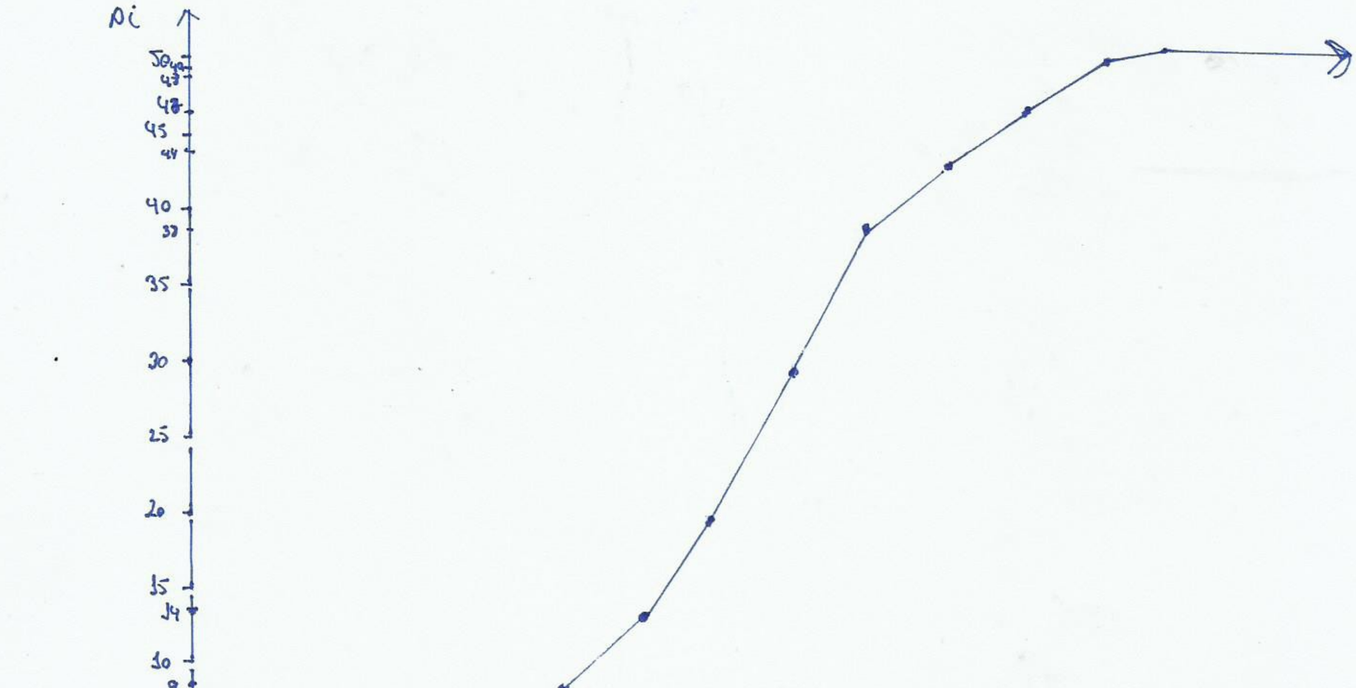
EDIP: RELACIÓN 1.

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

→) HISTOGRAMA




→) CURVA DE DISTRIBUCIÓN



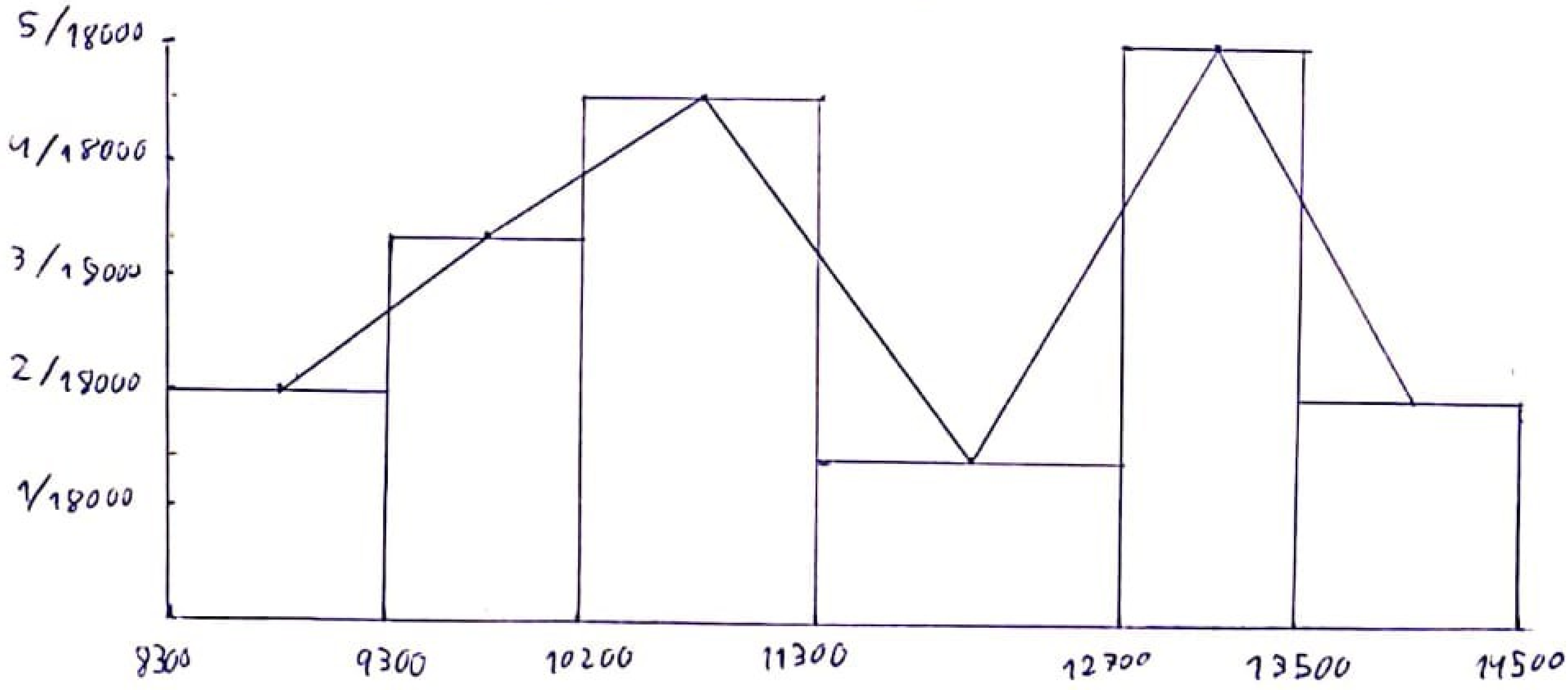
③

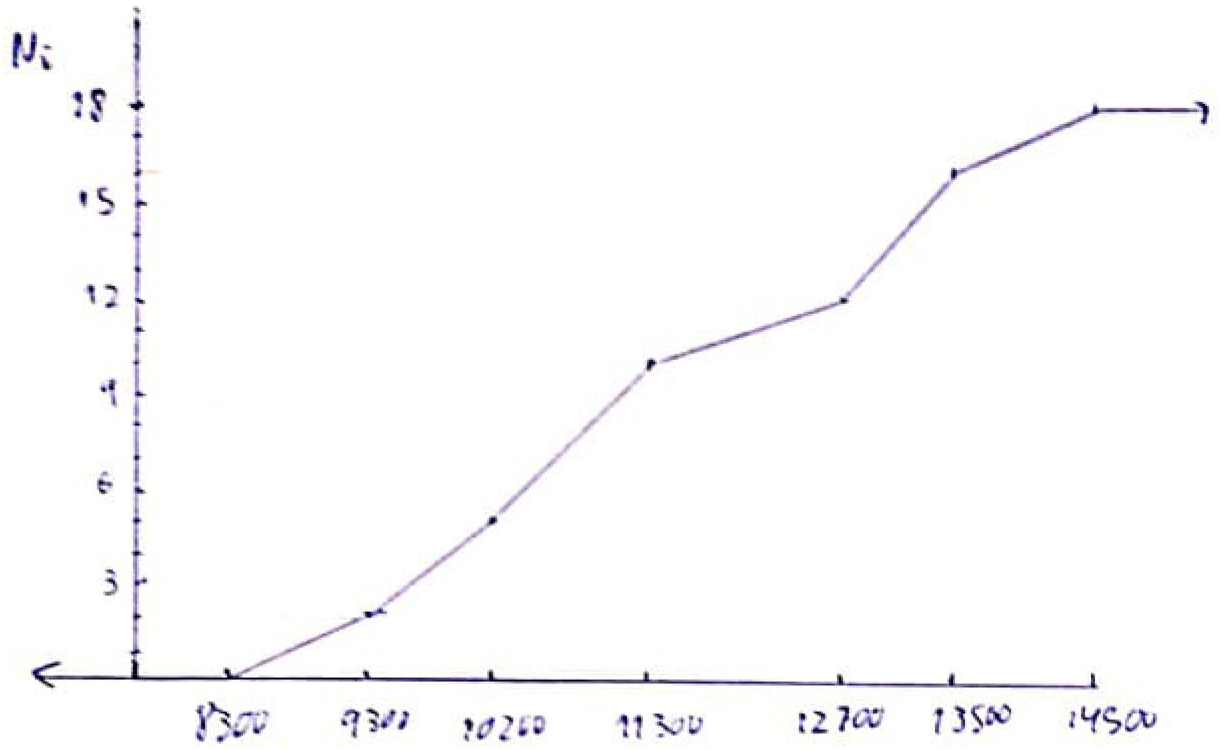
En la población de las comunidades autónomas, de tamaño 18, se ha observado la variable estadística de la cuantía de la renta familiar media en el año 2003, que ha presentado 6 modalidades.

a)

| $I_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ | $c_i$ | $a_i$ | $h_i$ |
|---|---|---|---|---|---|---|---|
| (8300, 9300] | 2 | 2 | 1/9 | 1/9 | 8800 | 1000 | 0,002/18 |
| (9300, 10200] | 3 | 5 | 1/6 | 5/18 | 9750 | 900 | 0,01/54 |
| (10200, 11300] | 5 | 10 | 5/18 | 5/9 | 10750 | 1100 | 0,05/198 |
| (11300, 12700] | 2 | 12 | 1/9 | 2/3 | 12000 | 1400 | 0,01/126 |
| (12700, 13500] | 4 | 16 | 2/9 | 8/9 | 13100 | 800 | 0,005/18 |
| (13500, 14500] | 2 | 18 | 1/9 | 1 | 14000 | 1000 | 0,002/18 |

b) Histograma y poligonal de frecuencias:






Curva de distribución

c) 12 comunidades presentan una renta menor o igual a 12700 euros.

8 comunidades presentan una renta superior a 11300 euros.

(4)

En una población de 100 cajas se ha observado la variable estadística del número de piezas defectuosas por caja, la cual ha presentado 11 modalidades: 0, 1, 2 ... 10.

a) El número de piezas defectuosas por caja es:

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

b) Los números de piezas defectuosas que se encuentran más frecuentemente son 5 y 6.

c) El número mediano de piezas defectuosas por caja es 4,5

d) Los cuartiles son:

- $Q_1 = 2,5$. El 25% de las cajas no tiene más de 2 piezas defectuosas.
- $Q_2 = 4,5$. El 50% de las cajas no tiene más de 4 piezas defectuosas.
- $Q_3 = 6$. Más del 75% de las cajas tiene menos de 7 piezas defectuosas.

e) Los deciles de orden 3 y 7 son:

- De orden 3: 3. Más del 30% de las cajas tiene menos de 4 piezas defectuosas
- De orden 7: 6. Más del 70% de las cajas tiene menos de 7 piezas defectuosas.

f)

- Recorrido $\Rightarrow x_K - X_1 = 10 - 0 = 10$. Indica que ninguna caja contiene más de 10 piezas defectuosas más que otra.
  - Ventaja: es sencilla de calcular
  - Inconveniente: la poca 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
  - Ventaja: es fácil de calcular
  - Inconveniente: no da información sobre cómo es la distribución entre el primer y el tercer cuartil ni fuera de estos.

· Desviación absoluta media respecto a $\bar{x}$:

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

- Ventaja: da información sobre la dispersión de la distribución

· Desviación absoluta media respecto a la mediana:

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

- Ventaja: da información sobre la dispersión de la distribución respecto a la mediana

· Varianza:

$$\sigma^2 = \frac{1}{100} \sum_{i=1}^{10} n_i (x_i - \bar{x})^2 = 5{,}87 \text{ (piezas defectuosas)}^2$$

- Ventaja: Permite comparar la dispersión de dos distribuciones con la misma media.
- Inconveniente: da unidades cuadráticas

· Desviación típica:

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

- Ventaja: da las misma unidad que la de ~~las mo~~ las ~~población~~ modalidades

· Recorrido relativo:

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

- Ventaja: es independiente de la magnitud de las modalidades

· Recorrido semi-intercualtílico:

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

· Ventaja: Indica hacia qué lado está más desviada la distribución

· Coeficiente de variación de Pearson:

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

Ventaja: permite comparar distribuciones con diferentes medias

· Índice de dispersión respecto a la mediana:

$$V_{Me} = \frac{D_{Me}}{M_e} = 0{,}444$$

Ventaja: es independiente de la magnitud de las modalidades

| $I_i^{(1)}$ | $n_i$ | $N_i$ | $c_i$ | $n_ic_i$ | $n_ic_i^2$ | $(c_i-\bar{x})^2$ |
|---|---|---|---|---|---|---|
| (0,1] | 12 | 12 | 0,5 | 6 | 3 | 2,7556 |
| (1,2] | 13 | 25 | 1,5 | 19,5 | 29,25 | 0,4356 |
| (2,3] | 11 | 36 | 2,5 | 27,5 | 68,75 | 0,1156 |
| (3,4] | 8 | 44 | 3,5 | 28 | 98 | 1,7956 |
| (4,5] | 6 | 50 | 4,5 | 27 | 121,5 | 5,4756 |
| | 50 | | | 108 | 641 | |

| $I_i^{(2)}$ | $n_i$ | $N_i$ | $c_i$ | $n_ic_i$ | $n_ic_i^2$ | $(c_i-\bar{x})^2$ | $h_i$ |
|---|---|---|---|---|---|---|---|
| (0,1] | 1 | 1 | 0,5 | 0,5 | 0,25 | 1 | 1 |
| (1,3] | 6 | 7 | 2 | 12 | 24 | 3 | 3 |
| (3,6] | 7 | 14 | 4,5 | 31,5 | 141,75 | 1,6530 | 2,333 |
| (6,10] | 12 | 26 | 8 | 96 | 768 | 4,9031 | 3 |
| (10,12] | 2 | 28 | 11 | 22 | 242 | 27,1889 | 1 |
| | 28 | | | 162 | 1176 | | |

$\boxed{5}$

| $I_i^{(1)}$ | (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 |

a) calcula medias aritmética, armónica y geométrica

$$\bar{x}_1 = \frac{0{,}5 \cdot 12 + 1{,}5 \cdot 13 + 2{,}5 \cdot 11 + 3{,}5 \cdot 8 + 4{,}5 \cdot 6}{50} = 2{,}16$$

$$\bar{x}_2 = \frac{0{,}5 \cdot 1 + 2 \cdot 6 + 4{,}5 \cdot 7 + 8 \cdot 12 + 11 \cdot 2}{28} = 5{,}7857$$

$$H_1 = \frac{50}{\frac{12}{0,5} + \frac{13}{1,5} + \frac{11}{2,5} + \frac{8}{3,5} + \frac{6}{4,5}} = 1,228$$

$$H_2 = \frac{28}{\frac{1}{0,5} + \frac{7}{2} + \frac{14}{4,5} + \frac{26}{8} + \frac{28}{11}} = 5,7857$$

$$G_1 = \sqrt[50]{0,5^{12} \cdot 1,5^{13} \cdot 2,5^{11} \cdot 3,5^{8} \cdot 4,5^{6}} = 1,684$$

$$G_2 = \sqrt[28]{0,5^{1} \cdot 2^{6} \cdot 4,5^{7} \cdot 8^{12} \cdot 11^{2}} = 4,769$$

b) Calcular valor más frecuente:

$$M_{o1} = 1 + \left[\frac{13-12}{(13-12)+(13-11)}\right](2-1) = 1 + \frac{1}{3} \cdot 1 = \frac{4}{3}$$

$$M_{o2} = 10 + \left[\frac{2-12}{(2-12)+(2-0)}\right](12-10) = 10 + \left(\frac{-10}{-8}\right) \cdot 2 =$$

$$= 10 + \frac{10}{4} = \frac{50}{4} = 12,5$$

c) Calcular el valor superado por el 50% de las observaciones.
(Equivale a calcular la mediana)

$$C_{50}^{(1)} = \frac{50 \cdot 50}{100} = 25 = N_2 \rightarrow P_{50}^{(1)} = 2$$

$$C_{50}^{(2)} = \frac{28 \cdot 50}{100} = 14 = N_3 \rightarrow P_{50}^{(2)} = 6$$

d) Recorrido, recorrido intercuartílico y desviación típica.

Recorrido

$R_1 = 5$

$R_2 = 12$

En el segundo caso hay mayor dispersión en los datos estudiados.

Recorrido intercuartílico

$$C_{75}^{(1)} = \frac{50 \cdot 75}{100} = 37,5 \qquad P_{75}^{(1)} = 3 + \frac{37,5-36}{8}(4-3) = \frac{51}{16}$$

$$C_{75}^{(2)} = \frac{28 \cdot 75}{100} = 21 \qquad P_{75}^{(2)} = 6 + \frac{21-14}{12}(10-6) = \frac{25}{3}$$

$$C_{25}^{(1)} = \frac{50 \cdot 25}{100} = 12,5 \qquad P_{25}^{(1)} = 1 + \frac{12,5-36}{13}(2-1) = \frac{27}{26}$$

$$C_{25}^{(2)} = \frac{28 \cdot 25}{100} = 7 \qquad P_{25}^{(2)} = C_{25}^{(2)} = 3$$

Por tanto:

$$R_I^{(1)} = P_{75}^{(1)} - P_{25}^{(1)} = 2,149$$

$$R_I^{(2)} = P_{75}^{(2)} - P_{25}^{(2)} = 5,333$$

En el primer caso observamos que el 50% de la distribución se encuentra en el intervalo $[0,851, 3]$, mientras que en el segundo se encuentra en el intervalo $[7/6, 6.5]$. La dispersión es por tanto menor en el caso 1.

Desviación típica

$$\sigma_1 = +\sqrt{\frac{641}{100} - 2,16^2} = 1,32 \qquad \sigma_2 = +\sqrt{42 - \left(\frac{102}{27}\right)^2} = 2,92$$

De nuevo vemos que en el caso 1 la dispersión es menor ya que la desviación típica es una medida de la dispersión óptima.

Cotas

$$\underline{Cota}^{(1)} = \sqrt{\min (x_i - \bar{x})^2} = 0,34 \qquad \overline{Cota}^{(1)} = \sqrt{\max (x_i - \bar{x})^2} = 2,34$$

$$\underline{Cota}^{(2)} = \sqrt{\min (x_i - \bar{x})^2} = 1,285 \qquad \overline{Cota}^{(2)} = \sqrt{\max (x_i - \bar{x})^2} = 5,28$$

Nos dice cuán dispersa está una distribución según el valor de la misma. En el primer caso la dispersión es menor.

Varianza

$$\sigma_1^2 = 1,744 \qquad \sigma_2^2 = \frac{1671}{196}$$

Cotas

$$\underline{Cota}^{(1)} = \min(x_i - \bar{x})^2 = 0,1156 \qquad \underline{Cota}^{(2)} = \min(x_i - \bar{x})^2 = 1,653$$

$$\overline{Cota}^{(1)} = \max(x_i - \bar{x})^2 = 5,4756 \qquad \overline{Cota}^{(2)} = \max(x_i - \bar{x})^2 = 27,94$$

Coeficiente variación Pearson

$$CV_{x_1} = \frac{\sigma_1}{|\bar{x}_1|} = \frac{+\sqrt{1744}}{2,16} = 0,6115$$

$$CV_{x_2} = \frac{\sigma_2}{|\bar{x}_2|} = \frac{+\sqrt{\frac{1671}{196}}}{\frac{81}{14}} = 0,505$$

A pesar de los resultados anteriores, se tiene que la segunda distribución es más homogénea.

⑥ En una población de coches de tamaño 2 se ha observado la velocidad en $\frac{km}{h}$, que ha presentado 2 modalidades distintas: $60 \frac{km}{h}$, $70 \frac{km}{h}$, con distribución de frecuencias ~~{60, 70; 1,1}~~ {60,1}, {70,1}.

Calcula la velocidad media de recorrido.

Por definición: $V_{media} = \frac{d_{total}}{t_{total}}$. Despejando el tiempo para $V_1, V_2$:

$$t_1 = \frac{100 km}{60 km/h} = 5/3\, h \qquad t_2 = \frac{100 km}{70 km/h} = \frac{10}{7} h$$

Aplicando la fórmula de Vmedia:

$$V_{media} = \frac{d_{total}}{t_{total}} = \frac{2 \cdot 100 km}{\left(\frac{5}{3} + \frac{10}{7}\right) h} = \frac{840}{13}\, km/h = 64,62 \frac{km}{h} \,//$$

⑦

En una población de tamaño $n=5$ se ha observado una variable estadística $\underline{X}$,* que ha presentado 5 modalidades distintas con la siguiente distribución de frecuencias:

| $x_i$ | $n_i$ | $N_i$ |
|---|---|---|
| 5 | 1 | 1 |
| 6 | 1 | 2 |
| 7 | 1 | 3 |
| 10 | 1 | 4 |
| 12 | 1 | 5 |

* $\underline{X}$ = La rentabilidad de la empresa durante los años 1994-1998.

Por tratarse de una variable que estudia rendimientos, la media que calcularemos será la geométrica:

$G = \sqrt[n]{\prod_{i=1}^{5} c_i} - 1 \Rightarrow c_1 = 1'05 \ ; c_2 = 1'06; \ c_3 = 1'07 \ ; \ c_4 = 1'1 \ ; \ c_5 = 1'12$

Entonces $G = \sqrt[n]{1'05 \cdot 1'06 \cdot 1'07 \cdot 1'1 \cdot 1'12} - 1 \simeq 0'0797 = 7'97\%$

⑧

En una población de tamaño $n=498$ se ha observado una variable estadística $\underline{X}$ = Notas de los alumnos, que ha presentado 10 modalidades distintas.

- Suspenso $(C_{40})(P_{40})$

$C_{40} = \frac{n \cdot r}{100} = \frac{40 \cdot 498}{100} = 199,2 \Rightarrow (3,4]$

$P_{40} = 3 + \frac{199'2 - 164}{81}(4-3) = 3'43$ (Nota máxima de los suspensos)

- Aprobados $(C_{70})(P_{70})$:

$C_{70} = \frac{70 \cdot 498}{100} = 348'6 \Rightarrow (5,6]$

$P_{70} = 50 + \frac{348'6 - 339}{70}(6-5) = 5'137$ (Nota máxima de los aprobados).

- Notables $(C_{85})(P_{85})$:

$C_{85} = \frac{85 \cdot 498}{100} = 423,3 \Rightarrow (6,7]$

$P_{75} = 6 + \frac{423,3 - 409}{41}(7-6) = 6'348$ (Nota máxima de los notables)

• Sobresalientes $(C_{95})(P_{95})$:

$C_{95} = \frac{95 \cdot 498}{100} = 473,1 \Rightarrow (7,8]$

$P_{95} = 7 + \frac{473'1 - 450}{28}(8-7) = 7'825$ (Nota máxima de los sobresalientes)

• Matrículas: Buscamos la frec. absoluta acumulada asociada al 100%, que trata del máximo de la distribución. Por tanto, la nota máxima de las matrículas es 10.

(9) En una población de 110 jóvenes se ha observado la variable estadística de ~~sus alturas~~ la altura de estos que ha presentado 5 modalidades con distribución de frecuencias $\{x_i, n_i\}_{i=1,\dots,5}$.

| $x_i$ | $n_i$ | $N_i$ |
|---|---|---|
| (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) Buscamos la frecuencia absoluta acumulada asociada al 3%.

$$C_3 = \frac{nr}{100} = \frac{3 \cdot 110}{100} = 3{,}3 < N_1$$

Luego tenemos que:

$$P_3 = 1{,}55 + \frac{3{,}3}{18}(1{,}60 - 1{,}55) = \underline{\underline{1{,}559 \text{ metros}}} \simeq$$

$$\simeq \underline{1{,}56 \text{ m}}$$

b) En este caso hemos de ~~calcular~~ ~~calculando~~ el percentil 82:

$$P_r = e_{i-1} + \frac{\frac{nr}{100} - N_{i-1}}{n_i}(e_i - e_{i-1})$$

(fórmula que también empleamos en el apartado a).

$$C_{82} = \frac{nr}{100} = \frac{82 \cdot 110}{100} = 90,2 < N_4$$

$$P_{82} = 1,80 + \frac{90,2 - 73}{20}(1,90 - 1,80) = \underline{1,886 \text{ metros}} \simeq \underline{1,89 \text{ m}}$$

c) El enunciado nos pide el percentil 75:

$$C_{75} = \frac{nr}{100} = \frac{75 \cdot 110}{100} = 82,5 < N_4$$

$$P_{75} = 1,80 + \frac{82,5 - 73}{20}(1,90 - 1,80) = \underline{1,8475 \simeq 1,85 \text{ metros}}$$

d) Para ello calcularemos el número de jóvenes que miden 1,75 metros o más:

$$P_r = 1,75 \text{ m}$$

Despejamos $C_r$ de la formula de $P_r$:

$$C_r = \frac{P_r - e_{i-1}}{e_i - e_{i-1}} n_i + N_{i-1} = \frac{1,75 - 1,70}{1,80 - 1,70} \cdot 24 + 49 = 61 \text{ jóvenes.}$$

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

Luego los jóvenes que miden más de 1,75 metros son: $110 - 61 = \underline{49 \text{ jóvenes}}$

e) $C_{10} = \frac{nr}{100} = \frac{110 \cdot 1\cancel{0}}{1\cancel{00}} = 11 < N_1$

$$P_{10} = 1,55 + \frac{11}{18}(1,60 - 1,55) = \underline{1,58 \text{ metros}}$$

f) $C_{90} = \frac{nr}{100} = \frac{90 \cdot 110}{100} = 99 < N_5$

$$P_{90} = 1,90 + \frac{99 - 93}{17}(2,00 - 1,90) = 1935 \simeq \underline{1,94 \text{ metros}}$$

(10)

En una población de 150 personas con cáncer se ha observado la variable estadística de ~~sus edades que h~~ la edad de los enfermos que ha presentado 5 modalidades diferentes

| $x_i$ | $n_i$ | $N_i$ | $c_i$ | $n_i c_i$ | $n_i c_i^2$ | $\sqrt{(c_i-\bar{x})^2}$ | $n_i(c_i-\bar{x})^4$ |
|---|---|---|---|---|---|---|---|
| (10, 30] | 15 | 15 | 20 | 300 | 6000 | 28,7 | 10176978,24 |
| (30, 40] | 37 | 35 | 35 | 770 | 26950 | 13,7 | 1303418,836 |
| (40, 50] | 48 | 85 | 45 | 2160 | 97200 | 3,7 | 8995,9728 |
| (50, 60] | 40 | 125 | 55 | 2200 | 121000 | 6,3 | 63011,844 |
| (60, 90] | 25 | 150 | 75 | 1975 | 140625 | 26,3 | 11960876,4 |
| | 150 | | | 7305 | 391775 | | 23513281,3 |

a) Moda:

$$M_o = e_{i-1} + \frac{h_1 - h_{i-1}}{2h_i - h_{i-1} - h_{i+1}} (e_i - e_{i-1})$$

$$M_o = \frac{810}{17} = \underline{\underline{47,6471 \text{ años}}}$$

b) Para resolver el apartado necesitamos calcular los percentiles 35 y 65:

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

$$P_r = e_{i-1} + \frac{\frac{nr}{100} - N_{i-1}}{n_i}(e_i - e_{i-1})$$

$$P_{35} = 40 + \frac{52,5 - 37}{48}(50 - 40) = 43,22 \text{ años.}$$

La edad mínima del 30% central ~~son~~ es de 43,22 años.

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

$$P_{65} = 50 + \frac{97,5 - 85}{40}(50 - 40) = \underline{53,125 \text{ años}}$$

La edad máxima del 30% central es de 53,125 años.

c)

Recorrido intercuartílico: $R_I = Q_3 - Q_1$

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

$$Q_3 = 50 + \frac{112{,}5 - 85}{40}(50-40) = 56{,}875 \text{ años.}$$

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

$$Q_1 = 50 + \frac{37{,}5 - 37}{48}(50-40) = 40{,}1 \text{ años}$$

$$R_I = 56{,}875 - 40{,}1 = 16{,}775 \text{ años}$$

Desviación típica:

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

$$\bar{x} = \frac{20 \cdot 15 + 35 \cdot 22 + 45 \cdot 48 + 55 \cdot 40 + 75 \cdot 25}{20} = 48{,}7$$

$$\sigma = + \sqrt{\frac{391775}{150} - 48{,}7^2} \approx 15{,}49 \text{ años}$$

d) Coeficiente de asimetría:

$$\gamma_1 = \frac{\mu_3}{\sigma_X^3} = \frac{1}{X} \sum_{i=1}^{n} n_i \left( \frac{x_i - \bar{X}}{\sigma_X} \right)^3$$

$\gamma_1(x) = 0{,}0917 > 0 \Rightarrow$ 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:

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

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

$$D_9 = P_{90} = 60 + \frac{135 - 125}{25}(90 - 60) = 72 \text{ años}$$

$$C_{10} = \frac{nr}{100} = \frac{150 \cdot 10}{100} = 15 \text{ años} \rightarrow D_1 = 30 \text{ años.}$$

$$K = \frac{1}{2} \frac{56{,}875 - 40{,}104167}{72 - 30} - 0{,}263 = -0{,}0632976$$

La distribución es mesocúrtica.