Estadística Descriptiva e Introducción a la Probabilidad.
Relaciones de prácticas

ELÍAS MONGE SÁNCHEZ
DANIEL MORÁN SÁNCHEZ
JESÚS MUÑOZ VELASCO

Marzo 2023

Índice

1. Relación 1

1.1. 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		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

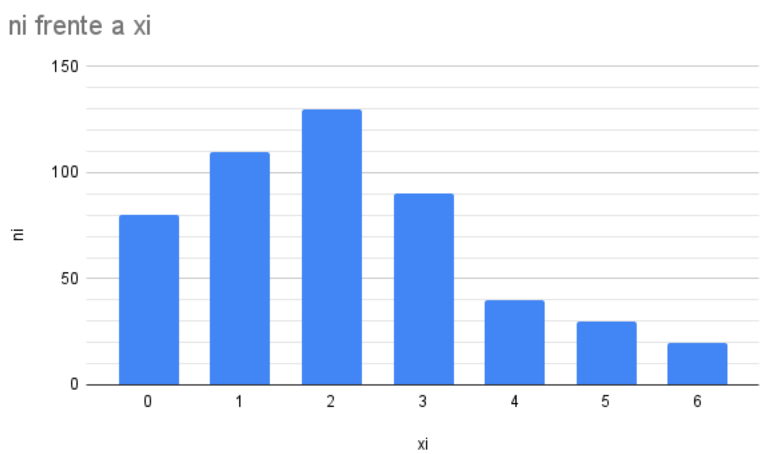
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, \ldots, x_k$

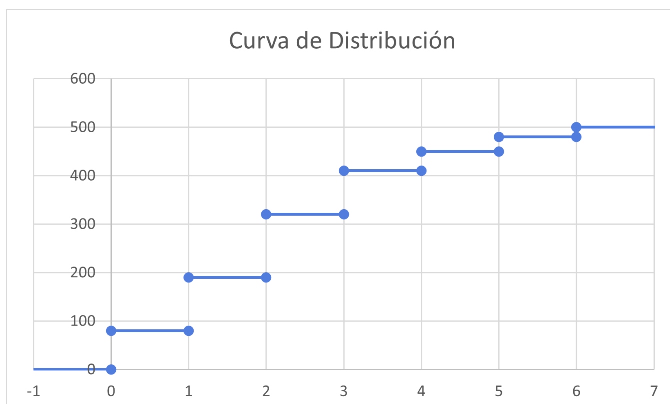
a). Completar la tabla de frecuencias

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**

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

Diagrama de barras





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

Las medias se obtienen mediante las siguientes fórmulas:

Media aritmética

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

La media aritmética nos da una idea bastante acertada de en torno a qué valor medio se agrupan los datos. Mirando la tabla y el diagrama de barras, vemos que en el entorno cercano al valor 2,14 las frecuencias son más altas, mientras que cuanto más nos alejamos más disminuyen las frecuencias.

Media armónica

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

Esta media no es posible de calcular, pues existe una modalidad nula que no tiene inverso. Lo único que se podría argumentar aquí es que tomando límite el valor de H es 0.

Media geométrica

$G = \sqrt[n]{\prod_{i=1}^{k} f_i x_i} = 0$

La media geométrica tiene el inconveniente de que, como se ha dado el caso, si alguna modalidad vale 0, su valor es cero, por lo que no aporta ninguna información añadida.

Media cuadrática

$Q = \sqrt{\sum_{i=1}^{k} f_i x_i^2} = 7,1$

La media cuadrática al menos ha dado un valor no nulo, pero no parece guardar mucha relación con la distribución ya que, de hecho, se sale del rango. Esto es porque no es una media adecuada para esta distribución al tener un cuadrado, y es que se usa sobre todo para promediar superficies.

1.2. Ejercicio 2:

La puntuación obtenida por 50 personas que se presentaros a una prueba de selección, sumadas las puntuaciones de los 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.

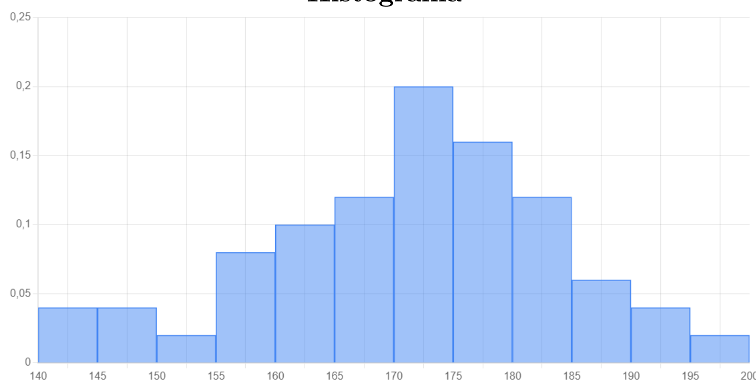
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

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

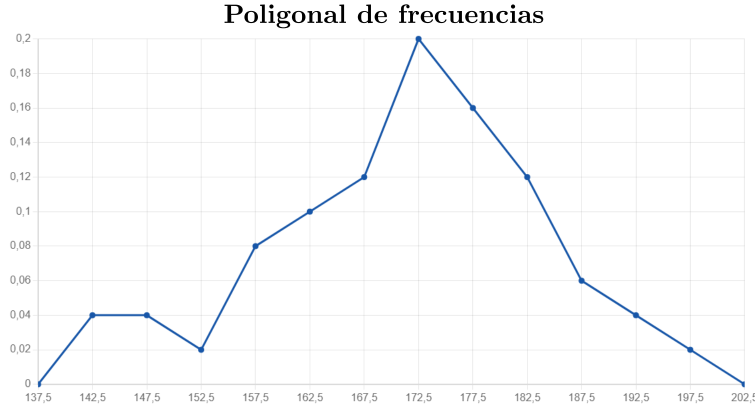
x_i	n_i	N_i	f_i	F_i	h_i
(140,145]	2	2	2/50	2/50	4/500
(145,150]	2	4	2/50	4/50	4/500
(150,155]	1	5	1/50	5/50	2/500
(155,160]	4	9	4/50	9/50	8/500
(160,165]	5	14	5/50	14/50	10/500
(165,170]	6	20	6/50	20/50	12/500
(170,175]	10	30	10/50	30/50	20/500
(175,180]	8	38	8/50	38/50	16/500
(180,185]	6	44	6/50	44/50	12/500
(185,190]	3	47	3/50	47/50	6/500
(190,195]	2	49	2/50	49/50	4/500
(195,200]	1	50	1/50	50/50	2/500

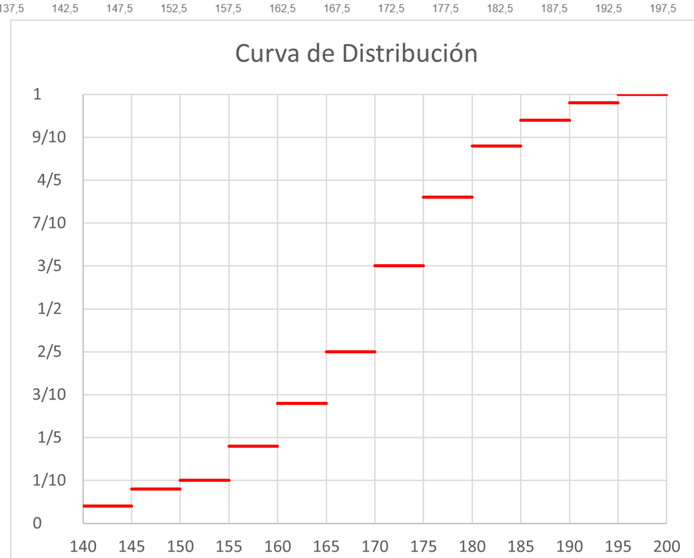
b). Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.

Histograma



Poligonal de frecuencias





1.3. 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]	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

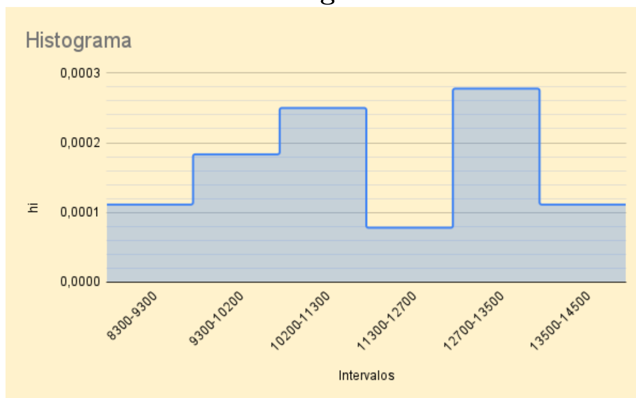
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

a). Completar la tabla

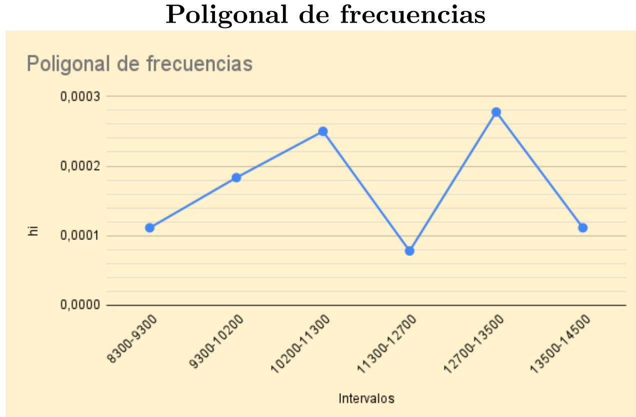
I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
(8300,9300]	2	**2**	**2/18**	**2/18**	**8800**	**1000**	**0.002/18**
(**9300**,10200]	**3**	5	**3/18**	**5/18**	**9750**	**900**	**0.0033/18**
(10200,11300]	**5**	**10**	**5/18**	10/18	**10750**	1100	**0.0045/18**
(11300,12700]	**2**	**12**	2/18	**12/18**	12000	**1400**	**0.0014/18**
(12700,13500]	4	**16**	**4/18**	**16/18**	**13100**	**800**	0.005/18
(13500,14500]	**2**	18	**2/18**	**1**	**14000**	**1000**	0.002/18

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

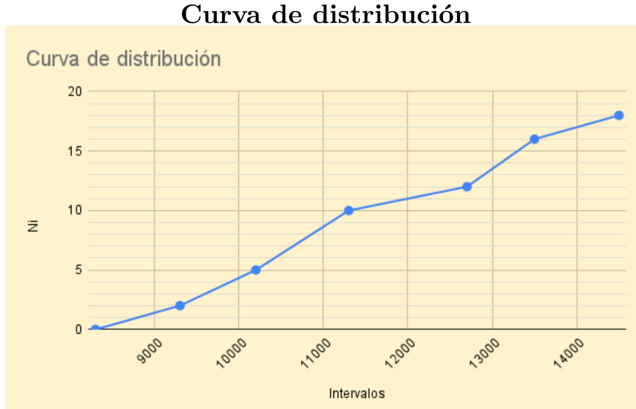
Histograma



Poligonal de frecuencias



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 euros (N_4) y

8 comunidades presentan una renta superior a 11300 euros $(n - N_3)$.

1.4. 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 de 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

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, \ldots, x_k$

x_i	n_i	N_i
0	6	6
1	9	15
2	10	25
3	11	36
4	14	50
5	16	66
6	16	82
7	9	91
8	4	95
9	3	98
10	2	100

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

$$\bar{x} = \frac{6+9+10+11+14+16+16+9+4+3+2}{11} = 9{,}0909\, piezas$$

b). ¿Cuántas piezas defectuosas se encuentra más frecuentemente en las cajas examinadas?

Este apartado hace referencia a la moda:

$$Mo_1 = 5$$

$$Mo_2 = 6$$

En este caso la variable es de tipo cuantitativa discreta con $n_6, n_7 > n_i \, \forall i \in \{1, ..., 11\} \,/\, i \neq 6, 7$
Por lo que se encuentra 5 o 6 piezas más frecuentemente.

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

$$Me = x_i \,/\, \frac{n}{2} = N_i \Rightarrow Me = 4\, piezas$$

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

$$Q_1 = x_i \,/\, \frac{n}{4} = N_i \Rightarrow Q_1 = 2\,piezas$$

$$Q_2 = x_i \,/\, \frac{n}{2} = N_i \Rightarrow Q_2 = Me = 4\,piezas$$

$$Q_3 = x_i \,/\, N_{i-1} < \frac{3n}{4} \leq N_i \Rightarrow Q_2 = 6\,piezas$$

El cuartil 1 nos indica que el 25 % de las cajas tiene menos o hasta 2 piezas defectuosas.
El cuartil 2 nos indica que el 50 % de las cajas tiene menos o hasta 4 piezas defectuosas.
El cuartil 3 nos indica que el 75 % de las cajas tiene menos o hasta 6 piezas defectuosas.

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

$$D_3 = x_i \,/\, N_{i-1} < \frac{3n}{10} \leq N_i \Rightarrow D_3 = 3\,piezas$$

$$D_7 = x_i \,/\, N_{i-1} < \frac{7n}{10} \leq N_i \Rightarrow D_7 = 6\,piezas$$

El decil 3 nos indica que el 30 % de las cajas tiene menos o hasta 3 piezas defectuosas.
El decil 7 nos indica que el 70 % de las cajas tiene menos o hasta 6 piezas defectuosas.

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

Medidas de dispersión absolutas:

$$Recorrido\ o\ rango : R = x_k - x_1 = 10 - 0 = 10\,piezas$$

Nos indica como de esparcidos están los datos.
Su principal desventaja es que es difícil de interpretar y es poco preciso ya que puede haber datos aislados que modifiquen ampliamente este resultado.

$$Recorrido\ intercuartilico : R_I = Q_3 - Q_1 = 6 - 2 = 4\,piezas$$

Indica como de esparcido está el 50 % central de los datos.

Al igual que el dato anterior es poco interpretable.

$$Desviación\ absoluta\ media\ respecto\ a\ \bar{x} : D_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{n} = 4{,}7673$$

Indica cómo de cercanos están los datos a la media (de media).
Cuanto más pequeño sea este resultado mayor representatividad tendrá la media.
Su desventaja es que al ser una medida absoluta es difícil de interpretar si es “grande” o “pequeña”.
Además es costosa de calcular.

$$Desviación\ absoluta\ media\ respecto\ a\ Me : D_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{n} = 2$$

Indica cómo de cercanos están los datos a la mediana (de media).
Cuanto más pequeño sea este resultado mayor representatividad tendrá la mediana.
Al igual que el resultado anterior, al ser una medida absoluta es difícil de interpretar si es “grande” o “pequeña”.
También es difícil de calcular.

$$Varianza : Var(X) = \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i}{n} = 28{,}25$$

Representa la variabilidad de los datos con respecto a la media (al igual que la desviación absoluta media con respecto a la media).
De igual forma es difícil de interpretar y no nos da una medida representativa de la bondad de la media.

$$Dexviación\ tipica : \sigma = +\sqrt{\sigma^2} = +\sqrt{28{,}25} = 5{,}315$$

Es la raíz cuadrada de la Varianza y por ello hereda las ventajas e inconvenientes de la misma.
Al igual que la varianza, este resultado trata de indicar la medida de representatividad de la media.

Medidas de dispersión relativas:

$$Coeficiente\ de\ apertura : C_A = \frac{x_k}{x_1} = \frac{10}{0} \Rightarrow imposible\ de\ calcular\ para\ esta\ distribución$$

La principal desventaja es la que se puede apreciar en este caso, que al aparecer un valor de x mínimo igual a 0, esta medida es imposible de calcular.

$$Recorrido\ relativo : R_R = \frac{R}{\bar{x}} = \frac{10}{9{,}09} = 1{,}1$$

Indica el número de veces que la media está contenida en el recorrido.
Al igual que el recorrido su principal desventaja es que un valor extremo puede afectar en gran medida a este resultado.

$$Recorrido\ semi-intercuartilico : R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{6-2}{6+2} = \frac{4}{8} = 0{,}5$$

La ventaja de este dato es que no se ve afectado por los datos extremos.
Trabaja con el 50 % central de los datos.

$$Coeficiente\ de\ variación\ de\ Pearson : C.V.(X) = \frac{\sigma_x}{|\bar{x}|} = \frac{5{,}315}{9{,}09} = 0{,}585$$

Nos indica la dispersión relativa de los datos.
Al ser una medida relativa, es más fácil de interpretar y nos va a permitir comparar diferentes distribuciones.

$$Indice\ de\ dispersión\ respecto\ a\ la\ mediana : V_{Me} = \frac{D_{Me}}{Me} = \frac{2}{4} = 0{,}5$$

Al igual que el resultado anterior da una medida de dispersión de los datos pero en este caso con respecto a la mediana.
Igual que el anterior resultado, nos va a permitir comparar distintas distribuciones.

1.5. Ejercicio 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.

Las medias se obtienen mediante las siguientes fórmulas, tomando las marcas de clase:

$$\bar{x} = \sum_{i=1}^{k} f_i c_i$$

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

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

Medias	Distribución 1	Distribución 2
Aritmética	2,16	5,7857
Armónica	1,2289	3,3991
Geométrica	1,3419	1,5850

b). El valor más frecuente

Las modas en variable continua se obtienen mediante la siguiente fórmula, aplicando la semejanza de triángulos en el intervalo modal (aquel con mayor frecuencia):

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

$Mo(1) = 1,314285714$
$Mo(2) = 6,5$

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

Las medianas se calculan haciendo una interpolación en el intervalo mediano, pero si ocurre, como es el caso, que para algún i, $N_i = \frac{n}{2}$, entonces $Me = e_i$. Así, las medianas son:

$Me(1) = e_2 = 2$
$Me(2) = e_3 = 6$

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

El recorrido no es más que el rango total en el que se encuentran los datos, es decir, el máximo menos el mínimo.

$R^{(1)} = 5$

$R^{(2)} = 12$

El recorrido intercuartílico requiere un cierto cálculo, ya que se obtiene al restar los cuartiles Q_3 y Q_1 de ambas distribuciones:

$Q_1 = P_{25} = e_{i-1} + \frac{0{,}25n - N_{i-1}}{n_i}(e_i - e_{i-1})$

$Q_3 = P_{75} = e_{j-1} + \frac{0{,}75n - N_{j-1}}{n_j}(e_j - e_{j-1})$

Siendo e_i y e_{i-1} los extremos del intervalo I_i y e_j y e_{j-1} los extremos del intervalo I_j de tal manera que $N_{i-1} < \frac{n}{4} \leq N_i$ y $N_{j-1} < \frac{3n}{4} \leq N_j$. Es decir que I_i es el intervalo bajo el que se encuentra el 25 por ciento de los datos y I_j el intervalo bajo el que se encuentra el 75 por ciento de los datos.

Para la primera distribución, $i = 2$ (segundo intervalo) y $j = 4$ (cuarto intervalo).
Para la segunda distribución, $i = 2$ y $j = 4$ también, aunque en este caso da la casualidad de que $N_2 = \frac{n}{4}$, luego $Q_1^{(1)} = e_2 = 3$.

$R_I^{(1)} = Q_3^{(1)} - Q_1^{(1)} = 3,1875 - 1,0385 = 2,1490$

$R_I^{(2)} = Q_3^{(2)} - Q_1^{(2)} = 8,3333 - 3 = 5,3333$

Comparando sendos valores con el recorrido de cada distribución, se interpreta que ambas distribuciones tienen una centralización del 50 por ciento de los datos similar:

$\frac{R_I^{(1)}}{R^{(1)}} = \frac{2{,}1490}{5} = 0,4280$

$\frac{R_I^{(2)}}{R^{(2)}} = \frac{5{,}3333}{12} = 0,4444$

De aquí se deduce que el recorrido intercuartílico de la segunda distribución representa una mayor fracción del recorrido total, y que por tanto dicha distribución es ligeramente más homogénea en cuanto a la centralización del 50 por ciento de los datos.
No obstante, cabe destacar que la primera distribución está desplazada a

la izquierda y la segunda desplazada hacia la derecha, información que nos proporcionaría el coeficiente de asimetría de Fisher, por ejemplo.

La desviación típica se define como la raíz cuadrada de la varianza, y se calcula de la siguiente manera:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}$$

Por lo tanto:

$\sigma^{(1)} = \sqrt{\sigma^{2(1)}} = \sqrt{1,7444} = 1,3208$

$\sigma^{(2)} = \sqrt{\sigma^{2(2)}} = \sqrt{8,5255} = 2,9198$

Este coeficiente nos proporciona información sobre cuánto se desvían los datos de la media aritmética. Si lo comparamos con el recorrido total obtenemos:

$\frac{\sigma^{(1)}}{\frac{1}{2}R^{(1)}} = \frac{1{,}3208}{2{,}5} = 0,5280$

$\frac{\sigma^{(2)}}{\frac{1}{2}R^{(2)}} = \frac{2{,}9198}{6} = 0,4866$

Aquí observamos que en la primera distribución los datos se desvían de la media un 52,8 por ciento de la mitad del recorrido, mientras que en la segunda se desvían un 48,66 por ciento de la mitad del recorrido, lo que quiere decir que en la segunda se desvían menos en relación con el recorrido, concluyendo con esto y con la interpretación del recorrido intercuartílico que la segunda distribución es ligeramente más homogénea.

1.6. 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.
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, \ldots, x_k$

La media más adecuada para esta situación es la media armónica ya que se trata de una magnitud compuesta (km/h). El cálculo sería:

$$\bar{x} = \frac{200}{\frac{100}{70} + \frac{100}{60}} = 64{,}615\, km/h$$

1.7. 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.

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, \ldots, x_k$
La media más adecuada para esta situación es la media geométrica ya que la variable presenta un comportamiento multiplicativo acumulativo. El cálculo sería:

$$\bar{x} = G = \sqrt[5]{1{,}12 \times 1{,}1 \times 1{,}07 \times 1{,}06 \times 1{,}05} = 1{,}08$$

Traducido en términos de rendimiento, el rendimiento neto medio ha sido del 8 %. $((1{,}08 - 1) \times 100 = 8)$

1.8. 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.

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

l_i	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]
n_i	34	74	56	81	94
N_i	34	108	164	245	339
f_i	0,0682731	0,1485944	0,1124498	0,1626506	0,1887550
F_i	0,0682731	0,2168675	0,3293173	0,4919679	0,6807229
h_i	0,0682731	0,1485944	0,1124498	0,1626506	0,1887550

l_i	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]
n_i	70	41	28	16	4
N_i	409	450	478	494	498
f_i	0,1405622	0,0823293	0,0562249	0,0321285	0,0080321
F_i	0,8212851	0,9036145	0,9598394	0,9919679	1,0000000
h_i	0,1405622	0,0823293	0,0562249	0,0321285	0,0080321

Nota máxima para estar suspenso: percentil 40:

$$3 + \frac{498 \times \frac{40}{100} - 164}{81} = 3{,}435$$

Nota máxima para estar aprobado: percentil 70:

$$5 + \frac{498 \times \frac{70}{100} - 339}{70} = 5{,}137$$

Nota máxima para estar notable: percentil 85:

$$6 + \frac{498 \times \frac{85}{100} - 409}{41} = 6{,}348$$

Nota máxima para estar sobresaliente: percentil 95:

$$7 + \frac{498 \times \frac{95}{100} - 450}{28} = 7{,}825$$

Nota máxima para ser matricula: percentil 100: 10

$$9 + \frac{498 \times \frac{100}{100} - 494}{4} = 10$$

1.9. 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

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

La altura máxima para ser bajo es el percentil 3, porque los que queden por debajo del 3 % se consideran bajos: (hacer interpolación en (1.60, 1.70]):

$$1,60 + \frac{110 \times \frac{3}{100} - 0}{18} \times 0,05 = 1,5592$$

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

La altura máxima para no ser alto es la misma que la altura mínima para ser alto, que es el percentil 82: (hacer interpolación en (1.80, 1.90]):

$$1,80 + \frac{110 \times \frac{82}{100} - 73}{20} \times 0,1 = 1,886$$

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

Cuartil 3: (hacer interpolación en (1.80, 1.90]):

$$1,80 + \frac{110 \times \frac{3}{4} - 73}{20} \times 0,1 = 1,8475$$

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

Nº de sujetos mayores de 1.75: n -(hacer interpolación inversa en (1.70, 1.80]):

$$110 - \frac{100}{110} \times (24 \times \frac{1,75 - 1,7}{0,1} + 49) = 54,5454$$

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

Altura superada por 11 de la muestra: percentil 10: (hacer interpolación en (1.55, 1.60])

$$1,55 + \frac{110 \times \frac{10}{100} - 0}{18} \times 0,05 = 1,5805$$

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

Altura superada por 99 de la muestra: percentil 90:(hacer interpolación en (1.90, 2.00]):

$$1,90 + \frac{110 \times \frac{90}{100} - 93}{17} \times 0,1 = 1,9352$$

1.10. 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

Edad	(10, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 90]
Nº enfermos (n_i)	15	22	48	40	25
N_i	15	37	85	125	150
f_i	0,1	0,1467	0,32	0,2667	0,1667
F_i	0,1	0,2467	0,5667	0,8333	1
a_i	20	10	10	10	30
h_i	0,005	0,01467	0,032	0,02667	0,00556

a). Calcular la edad más común de los individuos estudiados. Para ello calcularemos la moda:

$$Mo = 40 + \frac{0{,}032 - 0{,}01467}{(2 \times 0{,}032 - 0{,}01467 - 0{,}02667) \times 10} = 47{,}65\, años$$

b). Calcular la edad mínima y máxima del 30 % central de los individuos. Percentil 35:

$$P_{35} = 40 + \frac{\frac{150 \times 35}{100} - 37}{48} \times 10 = 43{,}23\, años$$

Percentil 65:

$$P_{65} = 50 + \frac{\frac{150 \times 65}{100} - 85}{40} \times 10 = 53{,}125\, años$$

c). Calcular el recorrido intercuartílico y la desviación típica.

Recorrido intercuartílico:

$R_I = Q_3 - Q_1 = P_{75} - P_{25} =$
$(50 + \frac{\frac{3}{4}150 - 85}{40} \times 10) - (40 + \frac{\frac{1}{4}150 - 37}{48} \times 10) = 56,875 - 40,104 = 16,771$

Desviación típica:

$\bar{x} = \sum_{i=1}^{k} f_i c_i =$
$0,1 \times 15 + 0,1467 \times 35 + 0,32 \times 45 + 0,2667 \times 55 + 0,1667 \times 75 = 48,2055$

$\sigma^2 = 0,1 \times (15 - 48,2055)^2 + 0,1467 \times (35 - 48,2055)^2 + 0,32 \times (45 - 48,2055)^2 + 0,2667 \times (55 - 48,2055)^2 + 0,1667 \times (75 - 48,2055)^2 = 271,125$

$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} f_i(c_i - \bar{x})^2} = \sqrt{271,125} = 16,466$

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

Coeficiente de asimetría de Fisher:

$\gamma_1 = \sum_{i=1}^{k} f_i(\frac{c_i - \bar{x}}{\sigma})^3 = -0,1611$

Coeficiente de asimetría de Pearson:

$A_p = \frac{\bar{x} - Mo}{\sigma} = \frac{48,2055 - 47,65}{16,466} = 0,0337$

De aquí se puede deducir que la distribución está bastante centrada, pues ambos coeficientes dan valores de diferente signo aunque muy próximos a cero. El coeficiente de Fisher con signo negativo está más lejano a cero, así que se puede decir que la distribución es ligeramente asimétrica hacia la izquierda.

Coeficiente de curtosis de Fisher:

$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{212561,932}{73511,019} - 3 = -0,1084$

Coeficiente de curtosis de Kelley:

$K = \frac{R_I}{2(D_9 - D_1)} = \frac{16,771}{2(72-30)} = 0,1997$

Nuevamente ambos coeficientes tienen valores próximos a cero aunque con signos distintos. Esto da a entender que la distribución es más bien mesocúrtica, aunque al ser el coeficiente positivo el que más se aleja del cero, se puede decir que es ligeramente leptocúrtica.