EJERCICIOS RESUELTOS TEMA 4 Y TEMA 5:
DESCRIPCIÓN NUMÉRICA DE DATOS UNIVARIANTES

Ejercicio 1.

La siguiente tabla muestra el tipo y número de trasplantes realizados en la Unión Europea en 2014:

Tipo de trasplante	*Número*
Trasplante de riñón	19670
Trasplante de hígado	7381
Trasplante de corazón	2146
Trasplante de pulmón	1822
Trasplante de páncreas	818
Otros trasplantes	44

¿Se puede calcular alguna medida que resuma la distribución de frecuencias? En caso afirmativo obtén dicha medida

SOLUCIÓN

En una variable cualitativa o atributo medida en escala nominal, las medidas descriptivas presentadas en el tema carecen de sentido. La única medida que es posible obtener con todo tipo de variables es la moda, es decir la modalidad que presenta una frecuencia más elevada.

De la tabla de frecuencias absolutas del enunciado (e igualmente de las relativas) se observa que la categoría o modalidad más frecuente es la de "*Trasplante de riñon*". Lo expresamos en la forma Moda=”*Trasplante de riñon*”

Ejercicio 2.

Un proveedor de ADSL consulta a 50 de sus clientes sobre el número de dispositivos (excluyendo el teléfono fijo) conectados en el hogar. Las respuestas proporcionadas se recogen en el siguiente cuadro:

0	1	4	0	5	5	4	2	4	5
2	2	1	1	2	3	1	6	4	1
4	3	5	4	3	3	4	0	3	5
6	5	3	3	2	2	0	4	1	3
5	3	5	2	2	1	1	5	2	5

a) Obtén el número medio de dispositivos conectados en el hogar. ¿Cuál es el número de dispositivos conectados más frecuente entre los clientes del proveedor?. ¿Es la distribución unimodal?

b) El proveedor desea conocer además el número mínimo de dispositivos conectados en la mitad de los hogares con mayor número de dispositivos. ¿Qué medida debe calcular? ¿Cuál es su valor?

c) Analiza el grado de homogeneidad en cuanto a número de dispositivos conectados de los clientes del proveedor.

d) Determina el número de dispositivos conectados más frecuente y analiza su representatividad.

e) En base a los valores obtenidos de las medidas de posición, ¿qué puedes concluir sobre la simetría de la distribución?

SOLUCIÓN

a) Se pide la media y la moda de la variable. Para el cálculo de moda, resulta cómodo obtener la distribución de frecuencias. Ésta se recoge en la tabla inferior en la que como vemos el dato con mayor frecuencia es 5 dispositivos. En notación estadística $Mo = 5$. El máximo de la frecuencia es absoluto por lo que la distribución de frecuencias en unimodal.

Para el cálculo de la media aritmética, aplicando la fórmula

$$\underline{x} = \frac{1}{N}\sum_{i=1}^{7} \quad x_i \times n_i = \frac{147}{50} = 2{,}94$$

O, sumando todos los datos y dividiendo por N=50, se obtiene que $\underline{x} = 2{,}94$

Las diferentes distribuciones de frecuencia se recogen en la tabla siguiente:

x_i	n_i	N_i	f_i	F_i
0	4	4	8,00%	8,00%
1	8	12	16,00%	24,00%
2	9	21	18,00%	42,00%
3	9	30	18,00%	60,00%
4	8	38	16,00%	76,00%
5	10	48	20,00%	96,00%
6	2	50	4,00%	100,00%
Total	50		1	

b) El proveedor debe calcular la mediana que es la medida estadística que marca el mínimo valor de la variable número de dispositivos en el hogar de la mitad de los hogares con mayor número de dispositivos. Para calcularla dado que la variable es discreta y disponemos de los datos primarios, nos basamos en las frecuencias relativas acumuladas. En concreto, vemos en la tabla que más de la mitad de los hogares ($F_i = 60\%$) tienen a lo sumo 3 dispositivos, de este modo, el mínimo número de dispositivos de la mitad de los hogares con mayor número de dispositivos podría aproximarse por 3. En notación estadística Me=3.

c) La homogeneidad de una distribución de frecuencias es una característica relacionada con la dispersión de la distribución. De forma más concreta se debe evaluar el coeficiente de variación de Pearson, cuyo valor representa la magnitud de la dispersión entre los datos. Debemos calcular entonces la varianza (y la desviación típica) de la distribución de frecuencias.

Calculando los dos términos de la fórmula abreviada de la varianza, se obtiene que su valor es: $S^2 = \frac{1}{N}\sum_{i=1}^{7} \quad {x_i}^2 \times n_i - \underline{x}^2$ (con los datos tabulados)

$$S^2 = \frac{1}{N}\sum_{i=1}^{50} \quad {x_i}^2 - \underline{x}^2 = \frac{575}{50} - 2{,}94^2 = 2{,}8564$$

Finalmente, obtenemos el valor del coeficiente de variación de Pearson $CV = \frac{S}{\bar{\bar{x}}} = \frac{\sqrt{2{,}86}}{2{,}94} = 57{,}5\%$. El elevado valor (superior al 20% de referencia) de este coeficiente refleja que la distribución del número de dispositivos conectados en los hogares es bastante heterogénea. Este hecho conlleva por ejemplo que el valor de la media no resultará ser representativo de la distribución del número de dispositivos conectados entre los hogares.

d) En el apartado a) de ha obtenido que la moda, o el número más frecuente de dispositivos es de 2. Para analizar su representatividad debemos calcular la medida de dispersión de los datos respecto de la moda, esto es la Desviación Absoluta Media respecto de la Moda. Posteriormente, se deberá obtener en términos relativos, calculando el Índice de Dispersión respecto de la Moda.

De la fórmula $DAMo = \frac{1}{N}\sum_{i=1}^{7} \quad |x_i - Mo| \times n_i$ (con los datos tabulados), o alternativamente

$DAMo = \frac{1}{N}\sum_{i=1}^{50} \quad |x_i - Mo|$ se obtiene que $DAMo = 2{,}14$.

Finalmente, el valor del Índice de Dispersión respecto de la Moda es $IDMo = \frac{DAMo}{Mo} = 42{,}8\%$

Nuevamente, el elevado valor del Índice de Dispersión (distancia global media de los datos a la Moda) conlleva el hecho de que el valor de la Moda (Mo=5 dispositivos) no resulte representativo de la distribución del número de dispositivos conectados en los hogares.

e) Cuando una distribución de frecuencias unimodal es simétrica (perfectamente), los valores de las medidas media aritmética, mediana y moda coinciden. Si esto no ocurre, la posición de estas tres medidas proporciona información acerca del tipo de asimetría. De forma concreta, en distribuciones asimétricas a derecha, la disposición típica de estas tres medidas es: $Me \leq Mo < \underline{x}$. En cambio, para distribuciones asimétricas a izquierda, la disposición de las tres medidas es: $\underline{x} \leq \ Me \leq Mo$. En nuestro caso, por tanto, es posible concluir en base a los valores de las tres medidas de posición que la distribución de frecuencias del número de dispositivos conectados en los hogares será asimétrica a izquierda (cola) o asimétrica negativa.

Ejercicio 3.

Los 50 trabajadores que acudieron una mañana a las oficinas de la Seguridad Social a solicitar algún tipo de pensión debieron responder a la pregunta:
¿Cuántos años de cotización al Sistema Nacional lleva usted hasta el momento actual?

El siguiente cuadro presenta las diferentes respuestas.

3	6	31	4	43	40	31	17	30	39
19	20	5	9	20	26	7	46	32	12
36	26	44	30	24	28	30	1	29	38
47	42	28	28	19	20	3	34	5	26
43	23	42	21	14	9	7	41	19	32

a) ¿Te parece razonable trabajar con esta variable como discreta? ¿Por qué? Para facilitar la comprensión de la información, se decide agrupar los valores de la variable años de cotización en intervalos. ¿Cuántos intervalos recomendarías utilizar? Si los intervalos fueran de amplitud constante, ¿cuál sería su amplitud? Utiliza una amplitud entera de 9 años para obtener los intervalos y su distribución de frecuencias.
b) Ordena los datos, identifica los valores de los años de cotización del 10% de los trabajadores con menor y mayor número de años trabajados. ¿A qué medida de dispersión corresponde la longitud del intervalo entre los dos valores obtenidos? Interpreta dicho recorrido.
c) Con los datos originales, obtén el número medio de años de cotización de los trabajadores de la muestra. Con los datos agrupados, obtén el valor de la media. Compara los valores obtenidos e indica por qué no coinciden.
d) ¿Resulta representativa la media de años de cotización obtenida en el apartado a)?
e) El trabajador con mayor antigüedad lleva 47 años cotizados, mientras que hay un trabajador con solo un año cotizado. ¿Cuál de los dos trabajadores es más atípico respecto de la distribución de los años cotizados del grupo de trabajadores?
f) Se ha obtenido que $C_1 = 14{,}75$ y $C_3 = 33{,}5$. Realiza el estudio de atípicos y elabora el gráfico de cajas de la distribución.

SOLUCIÓN

a) A diferencia del ejercicio anterior, la variable X = número de años de cotización toma muchos valores diferentes, (en concreto 32 valores diferentes) y, por tanto, conviene agruparlos en intervalos. Un criterio habitualmente adoptado para decidir el número de intervalos (en función del número de datos) es tomar $\sqrt{N/2} = \sqrt{25} = 5$

Para ajustar el total de datos, en lugar de comenzar en el mínimo, comenzamos en el 0 y se obtienen los intervalos y el conteo de número de datos que caen en cada intervalo que se refleja en la tabla siguiente:

L_{i-1}	L_i	n_i
0	9	11
9	18	3
18	27	12
27	36	13
36	47	11

b) Ordenando de forma creciente los valores de los años de cotización como se muestra en la tabla inferior (por columnas), identificamos los datos del 10% de los 50 trabajadores (5 trabajadores) con menor y mayor número de años cotizados.

1	5	9	19	21	26	30	32	39	43
3	6	12	19	23	28	30	32	40	43
3	7	14	20	24	28	30	34	41	44
4	7	17	20	26	28	31	36	42	46
5	9	19	20	26	29	31	38	42	47

Con la terminología y notación estadística, éstos corresponden a los deciles $D_1 = P_{0.1}$ y $D_9 = P_{0.9}$ y cuyos valores son: $D_1 = 5$ y $D_9 = 42$. Su interpretación es como sigue: el 80% central de los años cotizados oscila entre 5 y 43 años. La longitud del intervalo correspondiente a esos deciles es el recorrido Decil (R_d) cuyo valor es $R_d = 42 - 5 = 37$ años.

c) Con los datos desagrupados el cálculo del número medio de años cotizados se obtiene como $\underline{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 24{,}58$ años. Para obtener el valor medio con los datos agrupados, la fórmula a utilizar es $\underline{x} = \frac{1}{N}\sum_{i=1}^{5} x_i \times n_i$ donde ahora x_i son las marcas de clase de cada uno de los 5 intervalos. Aplicando la expresión, en los datos agrupados la media que se obtiene es: $\underline{x} = 24{,}52$ años. Aunque los valores de la media obtenidos son muy similares, el valor correcto si disponemos de los datos, es el cálculo desagrupado. La explicación a la diferencia entre los valores obtenidos para la media (a diferencia del ejercicio anterior) reside en que en los datos agrupados se utiliza la marca de clase y no el propio dato como representante del intervalo. Se trata entonces de un valor aproximado (bastante preciso) que utilizamos para aproximar la media cuando, como ocurre en múltiples ocasiones, solo disponemos de los datos agrupados.

d) Para analizar la representatividad de la media, se debe calcular el valor del coeficiente de variación de Pearson que mide la dispersión relativa de los datos respecto de la media aritmética. Comenzamos calculando la varianza de los datos originales; mediante la fórmula abreviada, se obtiene el valor:

$$S^2 = \frac{1}{N}\sum_{i=1}^{50} x_i^2 - \underline{x}^2 = 174{,}12$$

Y el valor del coeficiente de variación es: $CV(\%) = \frac{S}{\bar{\bar{x}}} \times 100 = \frac{\sqrt{174,12}}{24,58} \times 100 = 53,7\%$

En base a su elevado valor, podemos concluir que el número medio de años cotizados es bastante heterogéneo de modo que el número medio de años no resulta representativo de la distribución de los años cotizados de los trabajadores de la muestra.

e) La comparación de datos o valores respecto de la distribución de los datos se realiza mediante la tipificación de éstos. La puntuación tipificada se interpreta como el número de veces que la distancia del dato a la media contiene a la desviación típica. Aunque habitualmente se utiliza para comparar valores tipificados de dos distribuciones (por ejemplo, un trabajador es más atípico en cuanto a sus años de cotización, o en cuanto a su sueldo), también se puede utilizar para comparar el alejamiento de dos o más valores respecto del centro de gravedad (media aritmética).

En definitiva, decidir si es más atípico el valor mínimo del número de años cotizados o si lo es su máximo, consiste simplemente en comparar cuál se encuentra más alejado de la media. (Aunque no sería necesario dividir por la desviación típica porque es la misma, por consistencia con la fórmula, se obtienen las puntuaciones tipificadas).

$$z_{MIN} = \frac{(1 - 24,58)}{\sqrt{174,12}} = -1,79 \qquad z_{MAX} = \frac{(47 - 24,58)}{\sqrt{174,12}} = 1,70$$

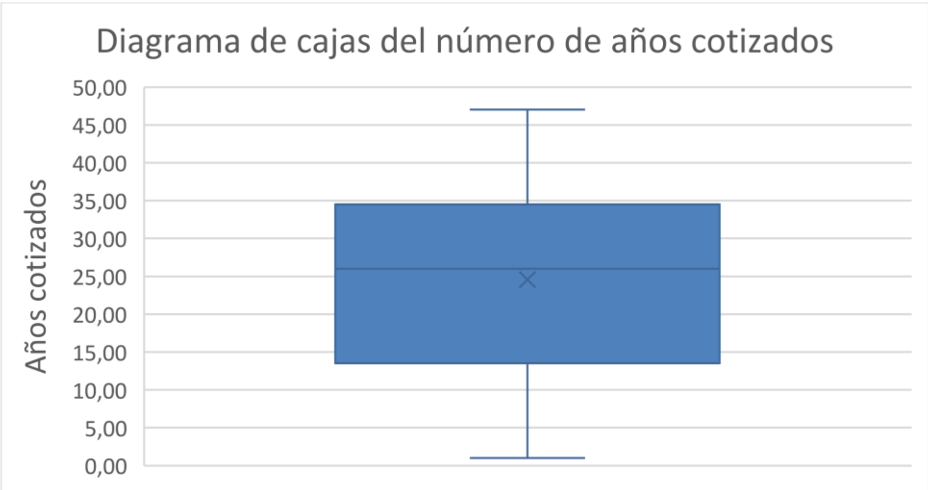
La mayor distancia en valor absoluto corresponde al número mínimo de años cotizados por lo que se concluye que ese valor es más atípico que el del trabajador con el máximo número de años cotizados.

f) Obtenemos el valor de los límites inferior y superior para los atípicos débiles, esto es:

$C_1 - 1,5 \times RI = 14,75 - 1,5 \times (33,5 - 14,75) = -13,75$

$C_3 + 1,5 \times RI = 14,75 + 1,5 \times (33,5 - 14,75) = 61,625$

No hay datos más alejados, por tanto, no hay atípicos (ni débiles, ni fuertes).



Ejercicio 4.

La distribución de la base imponible de las declaraciones del impuesto sobre la renta de las personas físicas (en miles de euros) presentadas en Huesca el primer día de la campaña Renta 2009 ha sido la recogida en la tabla:

Base imponible	Nº de declaraciones
0-5	37
5-15	43
15-50	110
50-60	85
60-100	25

a) ¿Qué base imponible es la que tiene un mayor número de declaraciones?

b) Determina el valor máximo de la base imponible de la mitad de las declaraciones con menor base imponible.

c) ¿Cuál es la base imponible media? Evalúa su grado de representatividad.

d) ¿Entre qué valores de la base imponible se encuentra el 80% central de declaraciones?

e) Sin representar la distribución, argumenta el tipo de asimetría que presenta.

SOLUCIÓN

a) Determinamos el valor de la Moda de la distribución. Tendremos en cuenta la simplificación en las unidades que facilita las operaciones, recordando deshacer el cambio en el resultado final. Más concretamente, la base imponible en miles de € es $Y = \frac{1}{1000}X$ así que calculamos la moda de los datos simplificados: Mo_Y y a partir de ella, el valor de $Mo_X = 1000 \times Mo_Y$ Es cómodo disponer los elementos necesarios para el cálculo en forma de tabla como la que se muestra:

L_{i-1}	L_i	d_i
0	5	7,40
5	15	4,30
15	50	3,14
50	60	8,50
60	100	0,63

Aplicamos la fórmula que aproxima la moda en datos agrupados por intervalo:

$$Mo_Y = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} \times a_i = 50 + \frac{0,63}{3,14 + 0,63} \times 10 = 51,659$$

Y el valor más frecuente de la base imponible e: $Mo_X = 1000 \times Mo_Y = 51.659€$

b) En este apartado se pide el valor de la Mediana ya que es el máximo valor de forma que la mitad de los datos son inferiores a la Mediana. Trabajamos con los datos simplificados porque la Mediana preserva la fórmula del cambio

$$Y = \frac{1}{1000}X \quad \rightarrow Me_X = 1000 \times Me_Y$$

Obteniendo y disponiendo en forma de tabla los elementos necesarios para el cálculo, aplicamos la fórmula que aproxima la mediana en datos agrupados por intervalo:

L_{i-1}	L_i	a_i	f_i	F_i
0	5	5	12,33%	12,33%
5	15	10	14,33%	26,67%
15	50	35	36,67%	63,33%
50	60	10	28,33%	91,67%
60	100	40	8,33%	100,00%

$$Me_Y = L_{i-1} + \frac{0{,}5 - F_{i-1}}{f_i} \times a_i = 15 + \frac{0{,}5 - 0{,}2667}{0{,}3667} \times 35 = 37{,}272$$

Entonces, el valor central de la base imponible de las declaraciones es: $Me_X = 37.272€$

c) Calculamos la media de los datos simplificadas las unidades mediante la fórmula de la media $\underline{y} = \frac{1}{N}\sum_{i=1}^{5} \quad y_i \times n_i$ y la media buscada será $\underline{x} = 1000 \times \underline{y}$. Disponemos los elementos necesarios para el cálculo de la media y también los de la varianza que necesitaremos para evaluar la representatividad de la media:

y_i	$y_i n_i$	$y_i^2 n_i$
2,5	92,5	231,25
10	430	4300
32,5	3575	116187,5
55	4675	257125
80	2000	160000

$\underline{y} = \frac{1}{N}\sum_{i=1}^{5} \quad y_i \times n_i = \frac{1}{300}17072{,}5 = 35{,}91$ y la base imponible media es: $\underline{x} = 35.910€$.

d) Para evaluar la representatividad, debemos obtener el valor del coeficiente de variación de Pearson y para ello la varianza de la distribución. Como este coeficiente es invariante frente a cambios de escala o unidades, podemos continuar trabajando con las unidades simplificadas. Aplicamos la fórmula de la varianza para datos agrupados por intervalos:

$$S_Y^2 = \frac{1}{50}\sum_{i=1}^{5} \quad {y_i}^2 \times n_i - \underline{y}^2 = \frac{1}{50}537844 - 35{,}91^2 = 503{,}4$$

Y el valor del coeficiente de variación es $CV_Y = CV_X = \frac{S_Y}{\underline{y}} = \frac{\sqrt{503,4}}{35,91}$=0.6248. El elevado valor de este coeficiente indica que la media de la base imponible no resulta representativa de la distribución de las declaraciones.

e) Comparando los valores de las medidas de posición $Mo_X = 51.659€$ $Me_X = 37.272€$ y $\underline{x} = 35.910€$ Se puede concluir que la distribución es ligeramente asimétrica a izquierda.

Ejercicio 5.

El ayuntamiento de un municipio A desea estudiar la situación económica de los hogares del municipio y plantea una investigación por muestreo entre las familias. Las familias que han participado en el estudio han indicado el tramo de "renta anual disponible en la unidad familiar"(RFD). Su distribución de frecuencias se recoge en la tabla inferior:

Renta familiar anual disponible (RFD)		*% de familias*
8	10	6,3%
10	14	16,7%
14	20	23,6%
20	35	42,2%
35	50	7,9%
50	65	3,3%

A partir de dicha información responde a las siguientes cuestiones:

a) Determina la renta anual per cápita media para las familias del municipio A. La investigación sobre la renta familiar de un municipio vecino B que ha copiado la idea, concluye que, en éste, la renta media per cápita anual asciende a 11.438,25€ con una desviación típica de 6.358€. ¿Cuál de los dos municipios es más homogéneo en cuanto a renta per cápita familiar disponible?

b) Utilizando los resultados y datos del apartado a) decide cuál de las rentas familiares de dos familias en cada municipio: x_A=22.000€ y x_B=16.400€ está en una mejor situación económica, respecto de las rentas de su municipio.

c) Para paliar la situación de las familias más desfavorecidas, el municipio ha aprobado un paquete de ayudas. Si por las limitaciones de presupuesto, únicamente se puede cubrir al 10% de las familias más desfavorecidas, determina de forma aproximada, la renta familiar máxima para acceder a esas ayudas.

d) Analiza la simetría de la distribución mediante los coeficientes de asimetría de Pearson y Bowley.

SOLUCIÓN:

a) Para el cálculo de la renta familiar disponible media en el municipio[1], aplicamos la expresión de la media, trabajando con las marcas de clase de cada intervalo, y en este caso con los datos que disponemos que son las frecuencias relativas de cada intervalo:

$\underline{y} = \sum_{i=1}^{6} \quad y_i \times f_i = 9 \times 0{,}63 + 12 \times 0{,}167 + 17 \times 0{,}236 + 27{,}5 \times 0{,}422 +$
$42{,}5 \times 0{,}79 + 57{,}5 \times 0{,}33 =$23,443

Por tanto, la renta familiar disponible media del municipio A asciende aproximadamente a los 23.443€

En la segunda parte del apartado, de cara a comparar el grado de homogeneidad de la renta familiar disponible con la de otro municipio, se debe calcular además la varianza para evaluar la dispersión o variabilidad de la renta familiar. Aplicando en este caso la definición de la varianza, trabajando con las marcas de clase y las frecuencias relativas, se obtiene su valor:

$$S_A^2 = \sum_{i=1}^{6} \quad \left(x_i - \underline{x}\right)^2 \times f_i = (9 - 23{,}443)^2 \times 0{,}63 + (12 - 23{,}443)^2 \times 0{,}167 +$$
$$(17 - 23{,}433)^2 \times 0{,}236 + (27{,}5 - 23{,}443)^2 \ \times \ 0{,}422 + (42{,}5 - 23{,}443)^2 +$$
$$\times\, 0{,}79 + (57{,}5 - 23{,}443)^2 \times 0{,}33 = 118{,}718$$

De los datos del enunciado, para el municipio B, $\underline{x}_B = 11.438{,}25€$ y $S_B = 6.358€$ obtenemos el valor del coeficiente de variación como medida de dispersión relativa, cuyo valor en el municipio B es $CV_B = \frac{S_B}{\underline{x}_B} = \frac{6.358}{11.438{,}25}$ =0,5558 mientras que en el municipio A es $CV_A = \frac{S_A}{\underline{x}_A} = \frac{\sqrt{118{,}718}}{23{,}443} = 0{,}4647$

De los valores obtenidos se concluye que, en el municipio A la renta familiar anual disponible, además de ser más alta, resulta ligeramente más representativa que la del municipio B, hay por tanto una mayor homogeneidad en cuanto a la renta en el municipio A.

b) Para comparar la situación de unos valores de una característica en dos o más poblaciones al igual que para comparar valores de dos o más características, se debe proceder tipificando los valores. Los resultados de la tipificación son las denominadas puntuaciones tipificadas; cuanto mayor sean éstas (en valor absoluto) más alejado se encontrará el valor respecto de la media correspondiente a su población. Utilizando la notación estadística, los valores tipificados serán:

$$z_B = \frac{x_B - \underline{x}_B}{S_B} = \frac{16400 - 11438{,}25}{6358} = -0{,}1324 \qquad z_A = \frac{x_A - \underline{x}_A}{S_A} = \frac{22 - 23{,}443}{\sqrt{118{,}718}} = 0{,}601$$

[1] Cabe notar que nuevamente trabajamos con las unidades simplificadas, aunque como ya se ha abordado en algún ejercicio anterior, en este Ejercicio no se formaliza el cambio de escala, simplemente se aplica.

Por lo que la familia del municipio B (aunque su renta es menor que la de la familia del municipio A) se encuentra en mejor situación económica respecto de las rentas del municipio que lo está la familia del municipio A, respecto a las rentas familiares de éste.

c) Al localizar el valor de la renta familiar disponible del 10% de las familias con menor renta, se está pidiendo aproximar el $P_{10\%} = D_1$. De nuevo trabajamos con las unidades simplificadas, facilitando el cálculo, que finalmente llevaremos a las unidades de trabajo ya que los percentiles también preservan la fórmula del cambio de escala:

$$Y = \frac{1}{1000}X \quad \rightarrow \quad P_{10\%\,X} = 1000 \times P_{10\%\,Y}$$

Comenzamos obteniendo las frecuencias relativas acumuladas e identificando al intervalo de clase correspondiente a dicho percentil:

$$F_1 = 6{,}30\% \text{ y } F_2 = 23\%$$

Se trata del segundo intervalo y en él, utilizamos la aproximación para datos agrupados:

$$P_{10\%Y} = L_{i-1} + \frac{0{,}10 - F_{i-1}}{f_i} \times a_i = 10 + \frac{0{,}10 - 0{,}0630}{0{,}167} \times 4 = 10{,}886$$

Así, trasladando a las unidades, se obtiene que $P_{10\%\,X} = 1000 \times 10{,}886 = 10.886€$. Este valor de la renta será aproximadamente el máximo de renta familiar para acceder a las ayudas.

d) Los coeficientes de asimetría de Pearson y de Bowley permiten analizar la simetría de una distribución en base a la relación entre las diferentes medidas de posición.

$$CAB = \frac{C_1 + C_3 - 2Me}{C_3 - C_1} \qquad CAPearson = \frac{\bar{x} - Mo}{S}$$

En definitiva, nos resta por obtener los cuartiles y la Moda de la distribución. Aplicando sus expresiones y trabajando con las unidades simplificadas, se obtienen los valores siguientes:

$$C_{1Y} = L_{i-1} + \frac{0{,}25 - F_{i-1}}{f_i} \times a_i = 14 + \frac{0{,}25 - 0{,}23}{0{,}236} \times 6 = 14{,}51$$

$$C_{3Y} = L_{i-1} + \frac{0{,}75 - F_{i-1}}{f_i} \times a_i = 20 + \frac{0{,}75 - 0{,}4660}{0{,}4220} \times 15 = 30{,}1$$

$$C_{2Y} = L_{i-1} + \frac{0{,}5 - F_{i-1}}{f_i} \times a_i = 20 + \frac{0{,}5 - 0{,}4660}{0{,}4220} \times 15 = 21{,}21$$

$$Mo_Y = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} \times a_i = 10 + \frac{3{,}93}{3{,}93 + 2{,}15} \times 4 = 12{,}22$$

Y sin más que sustituir en los coeficientes de asimetría:

$$CAB = \frac{14{,}508 + 30{,}094 - 2 \times 21{,}208}{830{,}094 - 14{,}50} = 0{,}14$$

$$CAPearson = \frac{23{,}443 - 12{,}221}{\sqrt{118{,}78}} = 1{,}03$$

El signo de ambos coeficientes es positivo, indicando así un cierto grado de asimetría positiva o a la derecha de la distribución respecto de los ejes de simetría (los que referencia cada coeficiente).