

EJERCICIOS RESUELTOS TEMA 3: TABULACIÓN Y REPRESENTACIÓN GRÁFICA DE DATOS UNIVARIANTES

Ejercicio 1.

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

<i>Tipo de trasplante</i>	<i>Número</i>
Trasplante de riñón	19670
Trasplante de hígado	7381
Trasplante de corazón	2146
Trasplante de pulmón	1822
Trasplante de páncreas	818
Otros trasplantes	44

- a) Obtén la distribución de frecuencias de los trasplantes de órganos en la Unión Europea en 2014.
- b) ¿Es posible calcular las frecuencias relativas acumuladas? ¿Por qué?
- c) Representa gráficamente de forma adecuada la información recogida en la tabla.

SOLUCIÓN

- a) Las frecuencias relativas se calculan mediante la expresión:

$$f_i = n_i / N$$

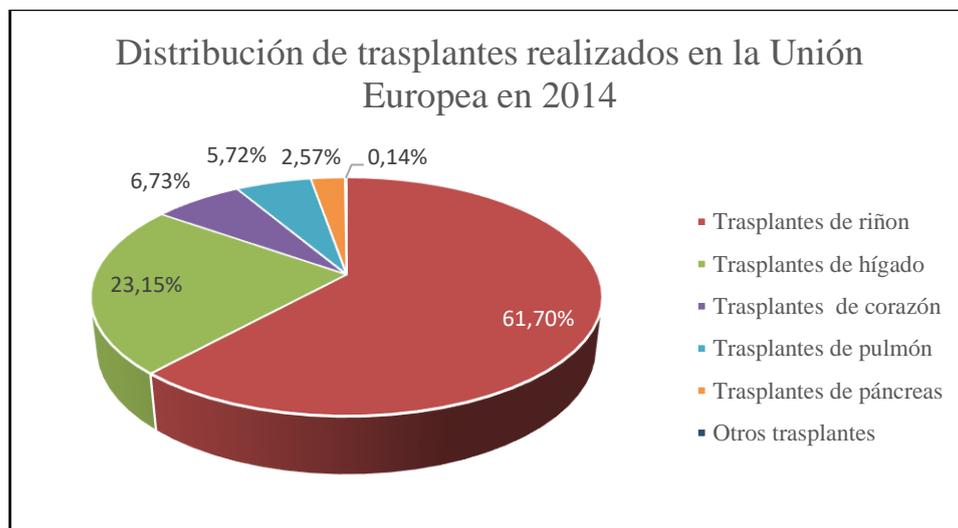
Los resultados expresados en % se recogen en la tabla siguiente:

<i>Tipo de trasplante</i>	<i>f_i</i>
Trasplante de riñón	61,70%
Trasplantesde hígado	23,15%
Trasplante de corazón	6,73%
Trasplante de pulmón	5,72%
Trasplante de páncreas	2,57%
Otros trasplantes	0,14%

- b) No es posible calcular las frecuencias relativas acumuladas, puesto que para su cálculo se necesita que la variable objeto de estudio esté ordenada; que tenga una escala, al menos de orden. En el caso de las variables nominales no es posible acumular, puesto

que los datos podrían estar expuestos en la tabla en cualquier orden y eso no variaría la interpretación de dicha tabla.

c) Como se ha visto anteriormente en teoría el gráfico que mejor se adapta a unos datos de tipo nominal es el diagrama de sectores.



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

- Elabora la tabla de frecuencias absolutas y relativas tanto ordinarias como acumuladas.
- ¿En cuántos hogares hay más de 4 dispositivos conectados?
- ¿Qué porcentaje de hogares tiene como mucho dos dispositivos conectados?
- Representa gráficamente los datos observados

SOLUCIÓN

a) Para construir la tabla de frecuencias correspondiente a la variable $X =$ número de dispositivos, en primer lugar, debemos saber cuántos valores diferentes o modalidades existen entre las 50 respuestas dadas. Estos valores son $x_i = 0, 1, 2, 3, 4, 5$ y 6 .

A continuación, hacemos el recuento de las veces que se observa cada uno de ellos, obteniendo así las frecuencias absolutas n_i . A partir de dichas frecuencias, calculamos las frecuencias absolutas acumuladas:

$$N_i = \sum_{j \leq i} n_j$$

y las frecuencias relativas tanto ordinarias como acumuladas:

$$F_i = N_i / N$$

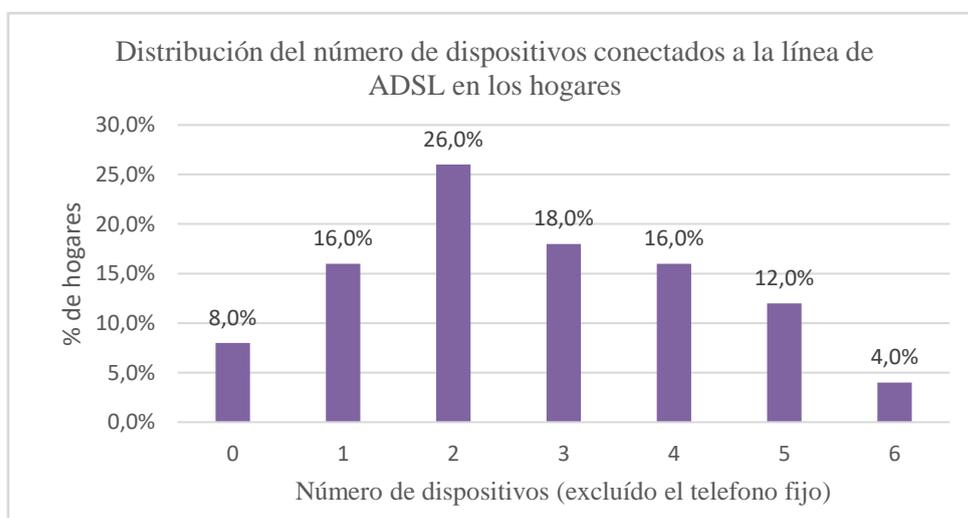
Las diferentes distribuciones de frecuencia se recogen en la tabla siguiente:

x_i	n_i	N_i	f_i	F_i
0	4	4	8,0%	8,0%
1	8	12	16,0%	24,0%
2	13	25	26,0%	50,0%
3	9	34	18,0%	68,0%
4	8	42	16,0%	84,0%
5	6	48	12,0%	96,0%
6	2	50	4,0%	100,0%
Total	50		100,0%	

b) Para calcular el número de hogares con un número de dispositivos superior a 4, utilizando únicamente los datos de la tabla, bastará con sumar las 10 familias con 5 dispositivos más las 2 familias con 6 dispositivos, haciendo un total de 12.

c) $F_3 = 0,42$ es la frecuencia relativa acumulada correspondiente al valor $x_i = 2$. Por tanto, un 42% de las familias tienen, como mucho, 2 dispositivos.

d) El gráfico que mejor se adapta a los datos (variable cuantitativa discreta) es el diagrama de barras.



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

- Construye la tabla de frecuencias agrupadas con 5 intervalos de amplitud 10 y determinar las marcas de clase, frecuencias relativas y absolutas tanto ordinarias como acumuladas.
- Representa gráficamente los datos.

SOLUCIÓN

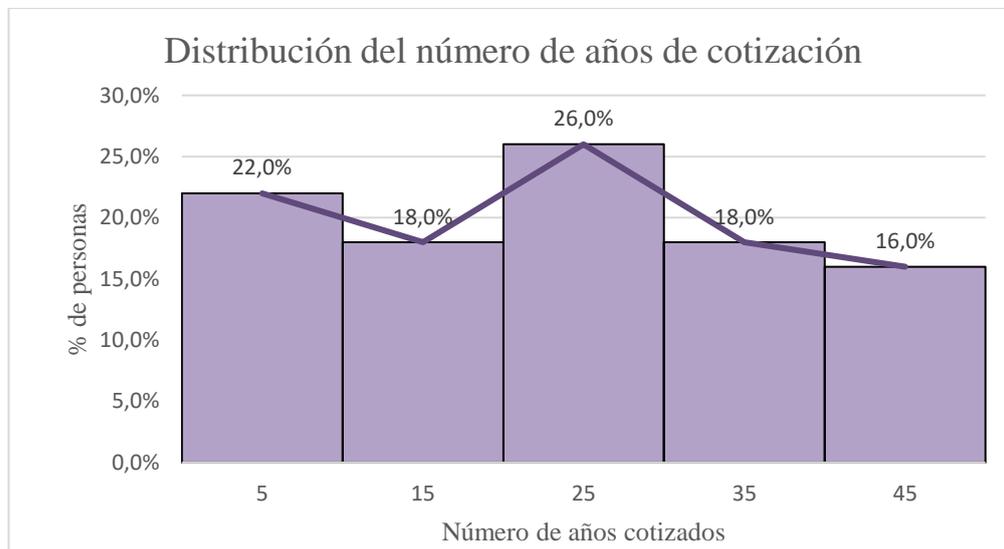
a) A diferencia del ejercicio anterior, la variable $X =$ número de años de cotización; toma muchos valores diferentes, (en concreto 32 valores diferentes) y, por tanto, conviene agruparlos en intervalos. Como se pide organizar los datos en una tabla de 5 intervalos de amplitud 10, es conveniente obtener los valores mínimo y máximo del conjunto de datos objeto de estudio. En este caso, el valor mínimo es 1 y el valor máximo es 47.

A partir de aquí, el siguiente paso consiste en concretar qué intervalos tomar, de forma que tanto los límites inferior y superior, como la marca de clase, sean entendidos fácilmente por un usuario no necesariamente experto en técnicas estadísticas. Por tanto, parece lógico tomar como primer intervalo el $[0, 10]$ y como último el $(40, 50]$.

Una vez determinados los intervalos se procede al conteo de los valores que caen dentro de cada uno, obteniendo así las frecuencias absolutas n_i . Las marcas de clase corresponden al punto medio de cada intervalo. Finalmente, a partir de n_i se calculamos, igual que en el ejercicio anterior, las frecuencias absolutas acumuladas N_i y las frecuencias relativas tanto ordinarias f_i como acumuladas F_i . Los resultados se muestran en la tabla que aparece a continuación:

$[L_{i-1}, L_i)$	n_i	x_i	N_i	f_i	F_i
[0,10]	11	5	11	0,22	0,22
(10,20]	9	15	20	0,18	0,40
(20,30]	13	25	33	0,26	0,66
(30,40]	9	35	42	0,18	0,84
(40,50]	8	45	50	0,16	1,00
Total	50			1,00	

b) El gráfico que mejor se adapta a este tipo de datos es el histograma o alternativamente, el polígono de frecuencias. En el eje de abscisas (horizontal) se han señalado las marcas de clase, mientras que en el eje de ordenadas (vertical) hace referencia a las frecuencias relativas.



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 siguiente:

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) Calcula el porcentaje de declaraciones con base imponible menor o igual que 50000 euros.
- c) Porcentaje de declaraciones con una base imponible mayor que 60.000 euros.
- d) Calcular el porcentaje de declaraciones con una base imponible entre 50.000 y 60.000 euros.
- e) Porcentaje de declaraciones con una base imponible menor o igual que 55.000 euros.
- f) Representa gráficamente esa distribución de frecuencias.

SOLUCIÓN

Para responder a los 4 apartados es conveniente construir primero la tabla de frecuencias completa, lo cual suele ser de utilidad en la práctica totalidad de los Ejercicios. A partir de ella se pueden resolver las cuestiones planteadas.

$[L_{i-1}, L_i)$	a_i	n_i	$d_i=n_i/a_i$	N_i	f_i	F_i
[0,5]	5	37	7,40	37	0,1233	0,1233
(5,15]	10	43	4,30	80	0,1433	0,2667
(15,50]	35	110	3,14	190	0,3667	0,6333
(50,60]	10	85	8,50	275	0,2833	0,9167
(60,100]	40	25	0,63	300	0,0833	1,0000
Total		300			1,0000	

Si solo nos fijamos en la frecuencia absoluta, está claro que el valor mayor corresponde al tercer intervalo, (15,50]; pero si nos damos cuenta de que se trata de intervalos de diferente amplitud, llegamos a la conclusión de que lo importante es saber cuál es el número de declaraciones por unidad (de base imponible). En ese caso, la característica que hay que mirar dentro de la tabla de frecuencias es su densidad. La mayor densidad se da en el cuarto intervalo (50,60], con un valor de 8,5. La frecuencia absoluta (85) es menor que la del intervalo (15,50], pero en términos relativos hay más declaraciones en el tramo (50,60] que en el tramo (15,50].

a) El porcentaje de declaraciones con una base imponible menor de 50.000 euros corresponde a la frecuencia relativa acumulada (F_3) hasta el intervalo (15,50], es decir, un $12,33 + 14,33 + 36,67 = 63,33\%$.

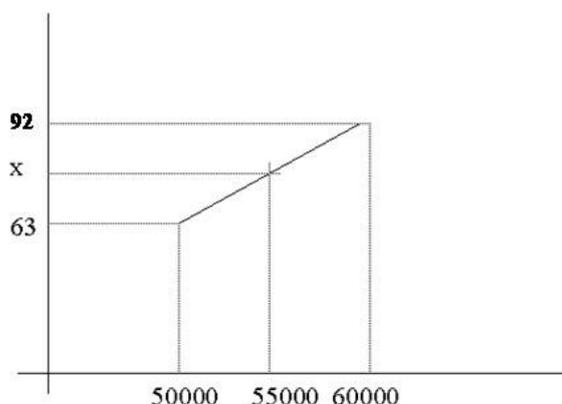
b) De la tabla de frecuencias vemos que 25 es el número de declaraciones con una base imponible mayor que 60.000 euros, por tanto, expresado en porcentaje, es un $8,33\%$.

c) De nuevo de la tabla de frecuencias vemos que 85 es el número de declaraciones,

con una base imponible entre 50.000 y 60.000 euros, por tanto, expresado en porcentaje, es un 28,33%.

d) En este caso, de la tabla podemos obtener que el porcentaje de declaraciones con una base imponible menor o igual que 50.000 euros es 63,33% y que el porcentaje de declaraciones con una base imponible menor o igual que 60.000 euros es 91,67%. Por tanto, para calcular el porcentaje de declaraciones con una base imponible menor o igual que 55.000 euros lo que hacemos es interpolar:

NOTA: Esto ocurrirá siempre que haya que realizar un cálculo en el que nos pregunten por algún valor del interior de un intervalo, no por alguno de los extremos. En ese caso será necesario interpolar:



Resolviendo tenemos que:

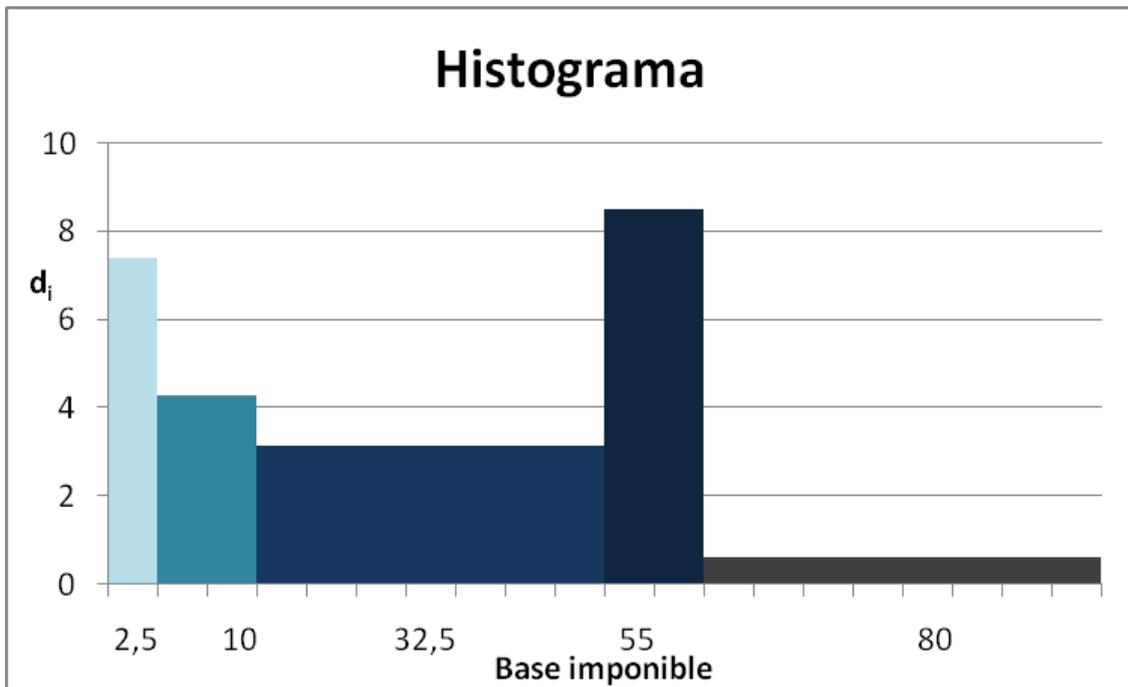
$$x = \frac{(55000 - 50000)(91,67 - 63,33)}{60000 - 50000} + 63,33 = \mathbf{77,5\%}$$

Por tanto, un 77,5% de declaraciones tenían una base imponible menor o igual a 55000 euros.

e) En este caso hay que representar los datos mediante un histograma, pero teniendo en cuenta que los intervalos son de diferente amplitud.

f) El gráfico que mejor se adapta a este tipo de datos es el histograma o alternativamente, el polígono de frecuencias. En el eje de abscisas (horizontal) se han señalado las marcas de clase, mientras que en el eje de ordenadas (vertical) hace referencia a las densidades.

NOTA: Este tipo de gráficos no se pueden realizar directamente con la hoja de cálculo Excel, puesto que dibuja los intervalos siempre con la misma amplitud. Si se quiere hacer el gráfico con EXCEL hay que asignar a cada intervalo tantas filas de la hoja como veces contiene al intervalo con la mínima amplitud y poniendo como altura la densidad de frecuencia. Si esto se hace así aparece el histograma que se muestra a continuación.



Ejercicio 5.

La siguiente tabla muestra la distribución de edades en Aragón según el Censo de 2001(en cientos):

Edad (años)	Personas
[0 –20]	65
(20 –30]	199
(30 –40]	475
(40 –50]	865
(50 –60]	685
(60 –70]	768
(70 –100]	714

- Determina la amplitud de los intervalos, marcas de clase, frecuencias absolutas acumuladas y frecuencias relativas ordinarias y acumuladas.
- Calcula el porcentaje de personas con una edad superior a 55 años.
- Dibuja el histograma con la hoja de cálculo EXCEL

SOLUCIÓN

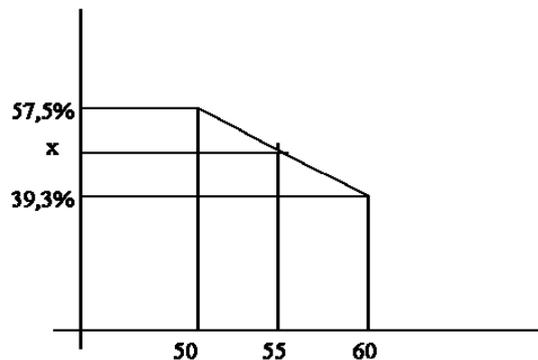
- La tabla de frecuencias completa aparece a continuación:

$[L_{i-1}, L_i)$	n_i	x_i	a_i	N_i	f_i	F_i
[0 –20]	65	10	20	65	0,017	0,017
(20 –30]	199	25	10	264	0,053	0,070
(30 –40]	475	35	10	739	0,126	0,196
(40 –50]	865	45	10	1604	0,229	0,425
(50 –60]	685	55	10	2289	0,182	0,607
(60 –70]	768	65	10	3057	0,204	0,811
(70 –100]	714	85	30	3771	0,189	1,000
Total	3771					

b) Para calcular el porcentaje de personas con edad superior a 55 años, se interpola dentro del intervalo [50,60) a partir de los valores:

% de personas con edad superior a 50 años = $100 - 42,5 = 57,5\%$.

% de personas con edad superior a 60 años = $100 - 60,7 = 39,3\%$.



Resolviendo tenemos que:

$$x = \frac{(55 - 50)(57,5 - 39,3)}{(60 - 50)} + 39,3 = 48,4$$

obteniéndose que un 48,4% de personas con una edad igual o superior a 55 años.

c) Calculamos las densidades de frecuencia al ser los intervalos de distinta amplitud

$[L_{i-1}, L_i)$	n_i	a_i	d_i
[0 –20]	65	20	3,25
(20 –30]	199	10	19,9
(30 –40]	475	10	47,5
(40 –50]	865	10	86,5
(50 –60]	685	10	68,5
(60 –70]	768	10	76,8
(70 –100]	714	30	23,8

Para dibujarlo se asignan dos filas al primer intervalo, tres al último y una al resto y se crea la siguiente tabla antes de representar el gráfico:

Edad	d_i
	3,25
[0 –20]	3,25
(20 –30]	19,9
(30 –40]	47,5
(40 –50]	86,5
(50 –60]	68,5
(60 –70]	76,8
(70 –100]	23,8

Y ya se representa normalmente, pero cuidando de eliminar los tramos interiores de las barras

