

**EJERCICIOS RESUELTOS TEMA 2: Estadísticos de resumen****Ejercicio 1**

A continuación tiene dos distribuciones por sexo y salario declarado en el primer empleo de un grupo de titulados.

Salario en €	Hombres en %	Mujeres en %
< 600	10,1	17,8
De 600 a < 900	16,7	23,8
De 900 a < 1200	9,3	26
De 1200 a < 3000	57,2	31,2
≥ 3000	6,7	1,2
Total	(N=96)	(N=96)

- ¿Qué población (la de hombres o la de mujeres) presenta un salario medio mayor?
- ¿Cómo ha resuelto el problema de los intervalos abiertos (1º y último) para el cálculo de la media? Explíquelo.
- ¿Qué porcentaje de varones gana más de 900 €?
- Realice la representación gráfica de las frecuencias absolutas de las mujeres.

**Solución:**

a) Para conocer las medias de ambas distribuciones a partir de las tablas, necesitamos cerrar los intervalos abiertos. Un criterio razonable, entre los posibles, para cerrar los intervalos sería considerar el límite inferior en 450 € al mes y establecer el límite del intervalo superior en 4.500 €, en el supuesto de que en muy pocos casos la retribución será superior a esa cantidad y elevar la cuantía distorsionaría el valor de la media que pretendemos calcular.

Una vez fijados los límites inferior y superior construimos los intervalos de la distribución. Como tenemos los datos agrupados, establecemos **las marcas de clase** para poder calcular los valores medios de las distribuciones de hombres y mujeres:

Salario en €	Marca de clase $x_i$
[450-600)	525
[600-900)	750
[900-1200)	1050
[1200-3000)	2100
[3000-4500]	3750

Salario en €	Marcas de clase $X_i$
$(450+600)/2$	525
$(600+900)/2$	750
$(900+1200)/2$	1050
$(1200+3000)/2$	2100
$(3000+4500)/2$	3750

**Convirtiendo los porcentajes en frecuencias absolutas** podemos calcular la media para los hombres. Como sabemos el total (N) para ambas distribuciones, basta con hacer una simple regla de tres. Fijémonos en la primera categoría:

Si 96 casos suponen el 100%, entonces 10,1% supondrán:  $\frac{96 \cdot 10,1}{100} = 9,696$

Como las frecuencias absolutas hacen el recuento en números enteros de los individuos de cada categoría (no hay 9,696 hombres), aproximamos al valor entero más próximo (10 hombres):

Realizamos la misma operación con todas las categorías:

$x_i$	$n_i$	$x_i \cdot n_i$
525	10	5250
750	16	12000
1050	9	9450
2100	55	115500
3750	6	22500
Total	96	164700

$$\bar{X}_h = \frac{\sum_{i=1}^{i-1} x_i n_i}{n} = \frac{164700}{96} = 1715,63$$

Es importante ofrecer el resultado interpretado de manera correcta, sin olvidar en qué unidades se expresa la variable: el salario medio de los hombres de este grupo de titulados es de **1715,63 euros**

De manera análoga calculamos la media para las mujeres:

$x_i$	$n_i$	$x_i \cdot n_i$
525	17	8925
750	23	17250
1050	25	26250
2100	30	63000
3750	1	3750
Total	96	119175

$$\bar{X}_m = \frac{\sum_{i=1}^{i-1} x_i n_i}{n} = \frac{119175}{96} = 1241,406$$

El salario medio de las mujeres tituladas de este grupo es de **1241,406 euros**

Puede verse la gran diferencia entre los salarios medios de ambas poblaciones: el salario medio de los hombres es casi 1,4 veces el de las mujeres. O, análogamente, el salario medio de las mujeres no llega a las tres cuartas partes (0,72) del salario de los hombres. Para hallar estas diferencias en términos proporcionales, tal como las hemos visto aquí, hay que calcular la proporción de un salario sobre otro:

$$\frac{1715,63}{1241,406} = 1,38 \text{ es lo que supone el salario de los hombres respecto al de las mujeres}$$

O análogamente:

$$\frac{1241,406}{1715,63} = 0,72 \text{ es lo que supone el salario de las mujeres respecto de los hombres}$$

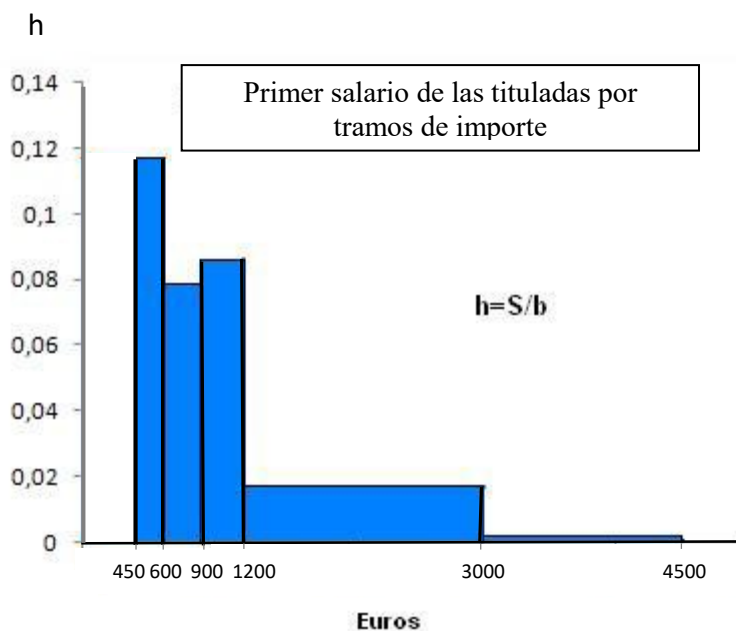
b) Ha quedado explicado en el apartado a)

c) El porcentaje de varones que gana más de 900 € se puede obtener directamente de la tabla sumando los porcentajes de las tres categorías superiores:

% de hombres que ganan más de 900€	% de mujeres que ganan más de 900€
<b>73,2%</b>	<b>58,4%</b>

d) La representación gráfica adecuada para esta distribución es un histograma, ya que tenemos una variable de intervalo. Al tratarse de intervalos desiguales, la mejor forma de ofrecer una representación gráfica es haciendo que las áreas correspondientes a cada polígono (cada una de las barras del histograma) sean proporcionales al número de tituladas en cada intervalo, dividiendo los porcentajes (S) por la amplitud de cada intervalo (b), para obtener así la altura de los polígonos. En realidad, aplicamos la fórmula del área del rectángulo para conocer la altura:

(Superficie = base x altura)



Como la tabla nos proporciona S (el porcentaje de cada categoría) y b (la amplitud del intervalo), solo tenemos que hallar la altura. Por ejemplo, para el primer intervalo de salario, la altura sería:

$$h = \frac{17,8}{150} = 0,1186$$

### Ejercicio 2

En la siguiente tabla, aparecen los datos que se registraban sobre de la edad de los profesores de enseñanza pública no universitaria en la Comunidad Murciana en el curso 2002/2003.

Grupo de edad	Nº Profesores
21 a 25	475
26 a 30	1.652
31 a 35	2.046
36 a 40	2.454
41 a 45	2.506
46 a 50	2.390
51 a 55	1.945
56 a 60	1.149
61 a 65	212
Más de 65	38
<b>TOTAL</b>	<b>14.867</b>

- a) Calcule las frecuencias relativas para cada grupo de edad, así como las frecuencias acumuladas relativas.
- b) Obtenga la edad media del profesorado
- c) Calcule la mediana de la distribución.
- d) Represente gráficamente las frecuencias absolutas.

### Solución

a) Las frecuencias relativas son la proporción en tanto por uno de las frecuencias absolutas de cada categoría sobre el total de casos. En notación matemática:

$$fr_i = \frac{n_i}{N}$$

En el primer caso, grupo de 21 a 25, el resultado será:

$$fr_1 = \frac{n_1}{N} = \frac{475}{14867} = 0,03194886$$

Los resultados se detallan a continuación:

Grupo de edad	Frecuencia Absoluta (ni)	Frecuencia Relativa (fr)	Frecuencia relativa acumulada (Fra)
21 a 25	475	0,0319	0,0319
26 a 30	1652	0,1111	0,1430
31 a 35	2046	0,1376	0,2806

36 a 40	2454	0,1650	0,4456
41 a 45	2506	0,1686	0,6142
46 a 50	2390	0,1608	0,7750
51 a 55	1945	0,1308	0,9058
56 a 60	1149	0,0773	0,9831
61 a 65	212	0,0143	0,9974
Más de 65	38	0,0026	1
TOTAL	14867	1	

b) Para el cálculo de la media y de otras medidas de tendencia central o de dispersión es necesario, en el caso de intervalos, obtener las marcas de clase. Para ello deben tenerse en cuenta dos hechos acerca de la naturaleza y codificación de los datos, el primero sobre los límites reales, y el segundo sobre el intervalo de cierre en las categorías abiertas.

Como la variable es la edad, debe tenerse en cuenta que la codificación de los datos se realiza en función de la edad cumplida. Es decir, en este caso el valor inferior de la categoría es el límite real inferior. Sin embargo esto no sucede con el superior. Por ejemplo una persona que tenga de edad exacta 25 años y 11 meses y 25 días, se habrá codificado en edad cumplida como una persona de 25 años y por tanto estará en el primer intervalo. Es decir en este caso el primer intervalo tiene como límite inferior 21, mientras que el límite superior es 25,99999999  $\approx$  26

El caso del último intervalo, su límite real inferior será 66 años. Obsérvese que quienes tienen entre 65 y 66 años, estarán en el intervalo 61 y 65. El límite superior del último intervalo será 70 años (si bien en este colectivo profesional la edad de jubilación es de 65 años, cuando estos profesionales ocupan ciertos cargos de gestión pueden retrasar la misma hasta el momento de cumplir 70 años).

A partir de los límites reales, la marca de clase será el punto medio de ambos, es decir:

$$X_i = \frac{L_{\text{inf}} + L_{\text{sup}}}{2}$$

Para el primer intervalo de 21 a 25 años, obtendremos:

$$X_i = \frac{L_{\text{inf}} + L_{\text{sup}}}{2} = \frac{21 + 26}{2} = \frac{47}{2} = 23,5$$

El resto de los límites reales y marcas de clase se detallan a continuación:

Grupo de edad	Límite Inferior $L_{\text{inf}}$	Límite Superior $L_{\text{sup}}$	Marca de Clase $x_i$	$n_i$
[21-26)	21	26	23,5	475
[26-31)	26	31	28,5	1652
[31-36)	31	36	33,5	2046
[36-41)	36	41	38,5	2454
[41-46)	41	46	43,5	2506

[46-51)	46	51	48,5	2390
[51-56)	51	56	53,5	1945
[56-61)	56	61	58,5	1149
[61-66)	61	66	63,5	212
[66-69]	66	70	68	38
TOTAL				14867

Para saber la edad media calculamos la media aritmética de la variable edad utilizando la marca de clase:

$$\bar{x} = \frac{\sum x_i n_i}{N} = \frac{23,5 \cdot 475 + 28,5 \cdot 1652 + (\dots) + 68 \cdot 38}{14867} = 42,61 \text{ años}$$

c) Dado que tenemos la variable agrupada en intervalos, podemos observar fácilmente en cuál de ellos se encontrará la Mediana obteniendo las frecuencias acumuladas absolutas o acumuladas relativas. Si nos fijamos en las frecuencias absolutas relativas obtenidas en el apartado a), al ser la Mediana el valor de la variable que deja por encima de él al 50% de los casos y por debajo a la otra mitad de casos, esta se encontrará en el intervalo cuya frecuencia relativa acumulada sea mayor o igual a 0,50 (o 50% si lo expresamos en porcentaje).

La Mediana se encuentra, por tanto, entre los 41 y 45 años.

También podemos comprobarlo con las frecuencias absolutas acumuladas:

Edad	$n_i$	$N_a$
21 a 25	475	475
26 a 30	1652	2127
31 a 35	2046	4173
36 a 40	2454	6627
<b>41 a 45</b>	<b>2506</b>	<b>9133</b>
46 a 50	2390	11523
51 a 55	1945	13468
56 a 60	1149	14617
61 a 65	212	14829
Más de 65	38	14867
TOTAL	14867	

La mitad de los casos es  $\frac{N}{2} = \frac{14867}{2} = 7433,5$

Como puede observarse, la mediana estará entre 41 y 45 años que son los valores en los que se acumulan los primeros 7433,5 casos. (Hasta 40 años hay 6627 casos y hasta 45 hay 9133, por lo tanto los 7433,5 estarán entre ambos límites)

NOTA: Sobre el cálculo de la Mediana

Es posible calcular con mayor precisión el valor de la mediana (así como el de los cuantiles) cuando los valores de la variable se presentan, como en este caso, en forma de intervalo. Los programas estadísticos como el SPSS o el PSpP trabajan con microdatos y ofrecen un valor concreto para estas medidas. Sin embargo, la forma de cálculo que utilizan no es única, por lo que aquí únicamente vamos a mostrar una de las posibles fórmulas para acercarnos de forma más precisa a la mediana cuando las variables se presentan en valores agrupados.

Para este ejercicio, el cálculo preciso lo ofrece la siguiente fórmula:

$$Me = L_i + \frac{\frac{N}{2} - N_{a-1}}{n_i} C = 41 + \frac{7433,5 - 6627}{2506} \times 5 = 41 + \frac{806,5}{2506} \times 5 = 41 + 1,609 = 42,609$$

Donde:

C es la amplitud del intervalo que contiene la mediana:  $C = L_{Sup} - L_{inf} = 46 - 41 = 5$

$L_i$ : Límite inferior del intervalo que contiene la mediana

$N_{a-1}$ : Frecuencia absoluta acumulada del intervalo anterior

N: total de casos

$n_i$ : frecuencia absoluta del intervalo que contiene la mediana

Por lo tanto la edad mediana será de un poco más de 42 años y medio.

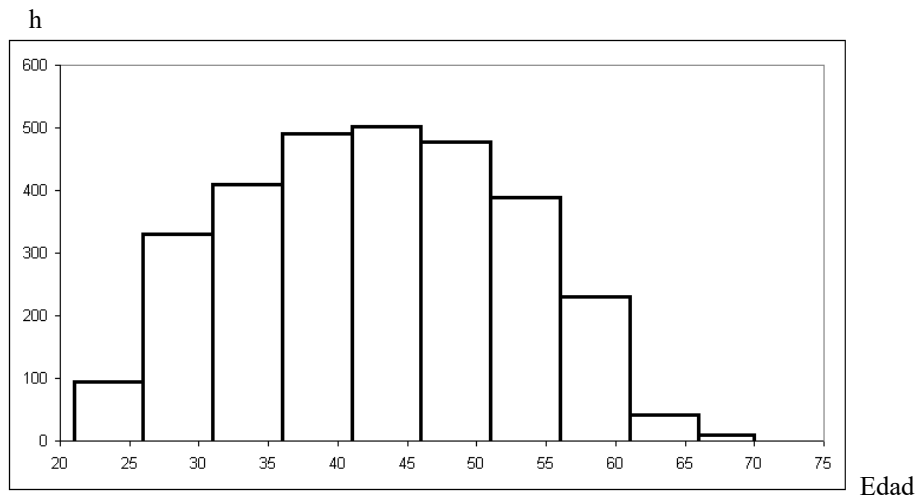
d) La representación gráfica correcta es un histograma. Todos los intervalos tienen la misma amplitud 5, a excepción del último que sólo tiene 4 ( $70-66=4$ ).

El histograma es una representación de áreas proporcional a las frecuencias de cada intervalo. Por tanto, consideraremos que los rectángulos del histograma tendrán un área o superficie (S) igual a la frecuencia absoluta  $n_i$ . Como el área de un rectángulo es base por altura ( $b \cdot h$ ), tendremos que  $h = n_i/b$ .

En nuestro caso las alturas de los rectángulos serán  $h = n_i/5$ , la quinta parte de las frecuencias, excepto en el último caso en que serán la cuarta parte, dado que la amplitud del intervalo es 4. La siguiente tabla nos proporciona las alturas para la representación gráfica.

Edad	Frecuencia ( $n_i$ )=área	Amplitud (Base=b)	Altura ( $h=n_i/b$ )
[21-26)	475	5	95
[26-31)	1652	5	330,4
[31-36)	2046	5	409,2
[36-41)	2454	5	490,8
[41-46)	2506	5	501,2
[46-51)	2390	5	478
[51-56)	1945	5	389
[56-61)	1149	5	229,8
[61-66)	212	5	42,4
[66-69]	38	4	9,5

Distribución por edad del profesorado de enseñanza pública no universitaria.  
Comunidad Murciana. Curso 2002/03



### Ejercicio 3

Según datos del INE, estos son los datos, según sexo, de los menores condenados en 2016:

Edad	Menores condenados según edad	
	Hombres	Mujeres
14	1.712	550
15	2.444	675
16	2.935	705
17	3.255	652
<b>Total</b>	<b>10.346</b>	<b>2.582</b>

- ¿Qué podemos decir comparando la proporción total de menores condenados según sexo?
- ¿Cuál de las dos poblaciones presenta una mayor edad media?
- ¿Cuál de estas dos poblaciones está mejor representada por su media? Justifique su respuesta.
- Indique la moda de la distribución.
- Represente gráficamente los datos.

### Solución

a) Para comparar ambas poblaciones, hallamos el peso que suponen respecto al total de menores condenados. Si el total de menores condenados es la suma de hombres y mujeres, entonces:

Total menores condenados:  $10346 + 2582 = 12928$

Proporción de hombres:  $\frac{10346}{12928} = 0,8003$  o bien 80,03% del total

Proporción de mujeres:  $\frac{2582}{12928} = 0,1997$  o bien 19,97% del total



Concluimos que la población de menores condenados está compuesta mayoritariamente por hombres (80,03%), en comparación con las mujeres, que solo representan la quinta parte del total (19,97%).

b) Debemos hallar la media aritmética para cada población. Dado que la variable está expresada en forma discreta (años cumplidos), no es necesario calcular las marcas de clase:

$$\bar{x} = \frac{\sum x_i n_i}{N}$$

$$\bar{x}_h = \frac{14 \cdot 1712 + 15 \cdot 2444 + 16 \cdot 2935 + 17 \cdot 3255}{10346} = \frac{23968 + 36660 + 46960 + 55335}{10346} = \frac{162923}{10346} = 15,747 \text{ años}$$

$$\bar{x}_m = \frac{14 \cdot 550 + 15 \cdot 675 + 16 \cdot 705 + 17 \cdot 652}{2582} = \frac{7700 + 10125 + 11280 + 11084}{2582} = \frac{40189}{2582} = 15,565 \text{ años}$$

La edad media de los menores condenados es solo ligeramente superior a la de las menores condenadas.

c) Aunque sabemos que la media aritmética es una medida de tendencia central muy útil para resumir la información de una distribución, debemos saber hasta qué punto es representativa estudiando la dispersión, es decir, la medida en que los valores de la variable se alejan o no de la media según la distribución de los casos. Para ello calcularemos la desviación típica de ambas distribuciones:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{N}}$$

Para su cálculo, podemos ayudarnos confeccionando una tabla:

Hombres:

$x_i$	$n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
14	1.712	-1,747	3,052	5225,024
15	2.444	-0,747	0,558	1363,752
16	2.935	0,253	0,064	187,84
17	3.255	1,253	1,570	5110,35
<b>Total</b>	10.346	-----	-----	11886,966

$$S_H = \sqrt{\frac{11886,966}{10346}} = 1,0719 \text{ años}$$

Mujeres:

$x_i$	$n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
14	550	-1,565	2,4492	1347,06
15	675	-0,565	0,3192	215,46
16	705	0,435	0,1892	133,386
17	652	1,435	2,0592	1342,598
<b>Total</b>	2.582	-----	-----	3038,504

$$S_M = \sqrt{\frac{3038,504}{2582}} = 1,0848 \text{ años}$$

Comparando la desviación típica, vemos que, aunque la diferencia es muy pequeña, la dispersión es ligeramente mayor en la distribución de mujeres. También podemos asegurar nuestra conclusión calculando la proporción de dispersión por unidad de media. Esto lo conseguimos estandarizando la medida de dispersión con el Coeficiente de variación (Cv), que es el cociente entre la desviación típica y la media, y permite calcular las unidades de desviación típica por unidad de media en cada distribución. A mayor valor del Cv, mayor será  $S_x$  respecto a  $\bar{x}$ , por lo que la dispersión será mayor.

$$Cv = \frac{S_x}{\bar{x}}$$

$$\text{Para los hombres: } Cv = \frac{1,0719}{15,747} = 0,0681$$

$$\text{Para las mujeres: } Cv = \frac{1,0848}{15,565} = 0,0697$$

El Cv corrobora que la dispersión es muy similar en ambas distribuciones, aunque algo mayor en el caso de las mujeres menores condenadas.

d) La moda es el valor de la variable que más se repite, es decir, que mayor frecuencia tiene.

Hombres:  $Mo = 17$  años

Mujeres:  $Mo = 16$  años

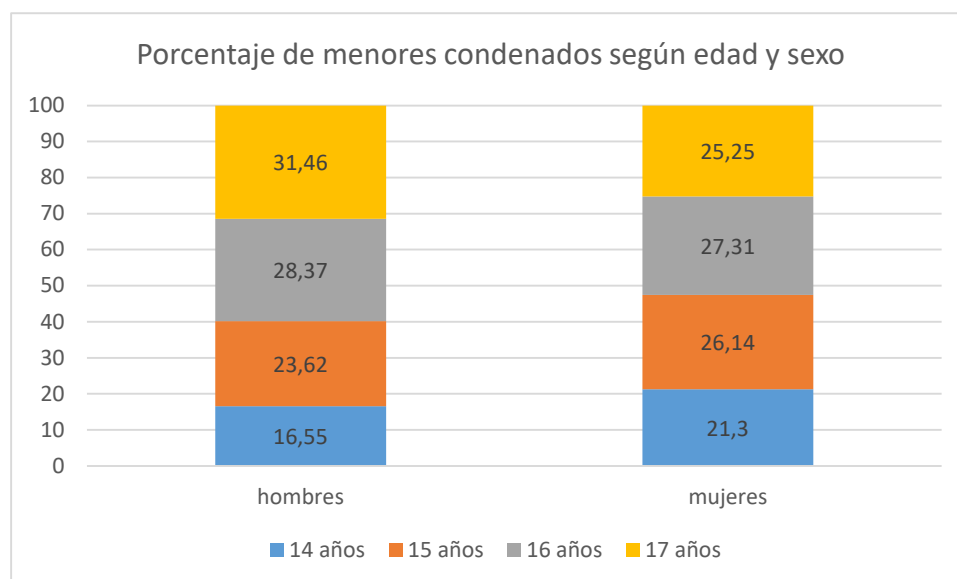
e) La tabla del INE contiene datos de interés que pueden ser representados gráficamente de distinta forma.

Dado que contamos con los datos de dos poblaciones (hombres y mujeres) podemos considerarlas como dos categorías de una misma variable “sexo” y representar, respecto al total, la proporción de hombres y mujeres entre los menores condenados, prescindiendo de momento de la variable edad. Como la variable sexo es cualitativa, con un nivel de medición nominal, y solo posee dos categorías, puede ser muy útil su representación a través de un diagrama de sectores o “de tarta”:



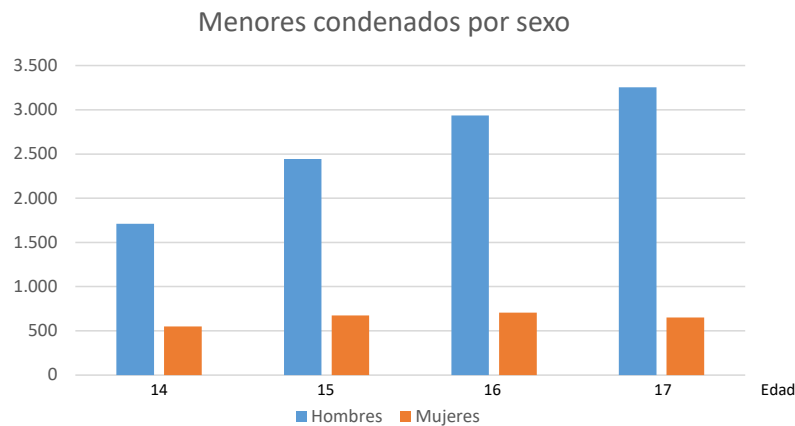
El tamaño de los sectores permite apreciar claramente la diferencia proporcional entre hombres y mujeres.

Para mostrar las diferencias proporcionales de las distintas categorías de edad podemos utilizar un gráfico de barras apiladas (o compuestas). Ello permitirá comparar el peso de cada edad para cada una de las poblaciones (hombres y mujeres). Para realizar el gráfico hemos hallado previamente las frecuencias relativas de la edad en cada población:



El gráfico permite comparar, por ejemplo, el peso de las categorías inferior y superior para ambos sexos. Vemos que la barra correspondiente a los 14 años es más pequeña para los hombres que para las mujeres, situación que se invierte para los 17 años. Es decir, las condenas entre los menores varones van obteniendo un mayor peso conforme aumenta la edad, mientras que en las mujeres menores, las condenas se encuentran más repartidas entre los distintos tramos de edad.

Si nuestro interés se centra en comparar las frecuencias absolutas de ambas poblaciones poniendo de manifiesto la gran diferencia cuantitativa entre menores condenados y condenadas, teniendo además en cuenta la edad, podemos representar un gráfico de barras agrupadas donde unimos, para cada categoría de edad, las frecuencias de hombres y mujeres:



#### Ejercicio 4

En el Barómetro del Centro de Investigaciones Sociológicas de diciembre de 2003, se preguntó a los entrevistados: “Actualmente, entre todos los miembros del hogar (incluido el entrevistado) y por todos los conceptos, ¿de cuántos ingresos netos disponen por término medio en su hogar al mes?”

Las respuestas de los entrevistados que contestaron a la pregunta se presentan en la siguiente tabla:

Ingresos	%
Menos o igual a 300 euros	1,9
De 301 a 600 euros	13,9
De 601 a 900 euros	19,8
De 901 a 1200 euros	23,3
De 1201 a 1800 euros	20,5
De 1801 a 2400 euros	10,3
De 2401 a 3000 euros	5,3
De 3001 a 4500 euros	2,8
De 4501 a 6000 euros	1,9
Más de 6000 euros	0,3
(N=1690)	

- Calcule el primer cuartil de la distribución de los ingresos declarados por los hogares de la muestra.
- Calcule la mediana de la distribución.
- Calcule el noveno decil.
- Comente los resultados

#### Solución

El cuartil y la mediana se pueden obtener a partir de las tablas de frecuencias relativas expresadas en porcentajes. Los datos agrupados en intervalos de ingresos están ordenados y podemos calcular las frecuencias relativas acumuladas  $Fr_a$ .

Ingresos €	%	% acumulado
1-300	1,9	1,9
301-600	13,9	15,8
601-900	19,8	35,6
901-1200	23,3	58,9
1201-1800	20,5	79,4
1801-2400	10,3	89,7
2401-3000	5,3	95
3001-4500	2,8	97,8
4501-6000	1,9	99,7
6001 o más	0,3	100
	100	

En la tabla observamos que en los tres primeros intervalos están agrupadas el 35,6% de las familias de la muestra que tienen menores ingresos. Por tanto el primer cuartil estará contenido en el tercer intervalo de 601 a 900 €. Tal y como se presentan los datos iniciales, hemos de considerar que los límites reales del intervalo serán 600,5 y 900,5.

Una vez identificado el intervalo que contiene al  $Q_1$ , podemos utilizar una fórmula para ofrecer una medida más precisa del cuartil (o de cualquier otro cuantil) al igual que hicimos en un ejercicio anterior con la Me. Esta medida ofrecerá el valor en euros por debajo del cual se encuentra exactamente el 25% de las familias de la muestra. Para ello tomaremos como valor de  $N = 100$  y la frecuencia acumulada en porcentajes:

$$Q_1 = L_i + \frac{\frac{N}{4} - N_{a-1}}{n_i} C_i = 600,5 + \frac{25 - 15,8}{19,8} 300 = 739,8939$$

Se obtendría el mismo resultado si calculamos la frecuencia acumulada absoluta (en número de casos) para  $N = 1690$ :

$$Q_1 = L_i + \frac{\frac{N}{4} - N_{a-1}}{n_i} C_i = 600,5 + \frac{\frac{1690}{4} - 267,02}{334,62} 300 = 739,8939$$

Un 25% de las familias entrevistadas dicen no superar los 739,89 € de ingresos al mes, mientras que un 75% restante superarían esa cantidad. Procediendo de manera análoga calculamos el segundo cuartil o Mediana:

$$M_e = L_i + \frac{\frac{N}{2} - N_{a-1}}{n_i} C_i = 900,5 + \frac{50 - 35,6}{23,3} 300 = 1085,9077$$

El 50% de las familias entrevistadas dicen no superar los 1.085,91 €.

El noveno decil es el valor que deja por encima al 10% de la población con mayores ingresos. Coincide, obviamente, con el percentil 90. En la tabla vemos que corresponde con el intervalo 2401-3000. Para calcularlo de forma precisa utilizamos esta fórmula tomando  $N=100$ :

$$D_9 = L_i + \frac{\frac{9N}{10} - Na - 1}{n_i} C_i = 2400,5 + \frac{90 - 89,7}{5,3} 600 = 2434,46$$

Solo el 10% de las familias supera los 2.434,46 euros.

### Ejercicio 5

Queremos estudiar la estructura de la distribución de adultos condenados en España en 2016 a partir de los siguientes datos obtenidos en el INE:

Edad	Frecuencia absoluta ( $n_i$ )
De 18 a 20	23.945
De 21 a 25	39.862
De 26 a 30	39.491
De 31 a 35	39.552
De 36 a 40	37.472
De 41 a 50	55.459
De 51 a 60	24.948
De 61 a 70	8.178
Más de 70	2.619

Hallar e interpretar las siguientes medidas:

- Intervalo de valores donde se encuentran, respectivamente, la Moda y la Mediana
- $Q_3$ ,  $D_4$  y  $P_{27}$

### Solución

a) Si los intervalos tuvieran la misma amplitud, la **Moda** se podría obtener directamente de la tabla, pues es el valor que más se repite de la distribución, es decir, que mayor frecuencia absoluta presenta. Sin embargo, como **los intervalos tienen diferente amplitud, es necesario corregir la frecuencia de cada uno de ellos según la amplitud del intervalo**, es decir, según el número de unidades de la variable (años) que describa cada intervalo. Por eso calculamos la amplitud de cada intervalo y seguidamente hallamos  $h$ , exactamente igual a como hacemos para representar un histograma con intervalos de distinta amplitud:  $h = n_i / b$

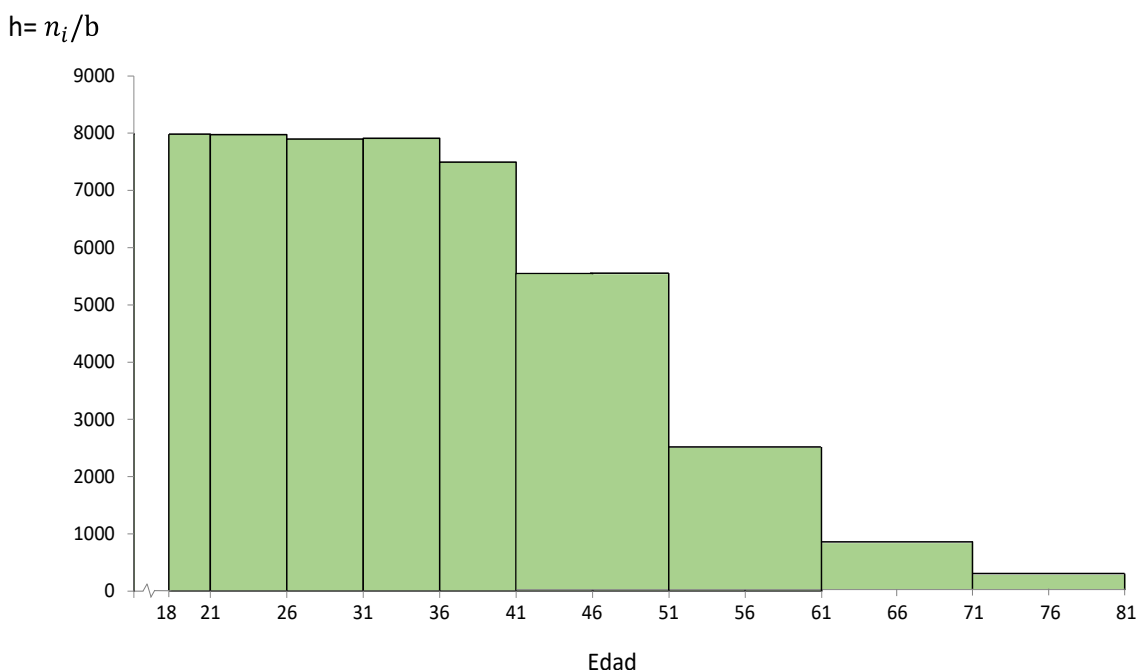
Edad	Frecuencia ( $n_i$ )=área	Amplitud (Base= $b$ )	Altura ( $h=n_i/b$ )
[18-21)	23.945	3	7981,7
[21-26)	39.862	5	7972,4
[26-31)	39.491	5	7898,2
[31-36)	39.552	5	7910,4
[36-41)	37.472	5	7494,4
[41-51)	55.459	10	5545,9
[51-61)	24.948	10	2494,8
[61-71)	8.178	10	817,8
[71-80]	2.619	10	261,9

Se ha cerrado el último intervalo en la edad de 80 años, entendiendo que es poco probable que haya una cantidad importante de individuos que superen tal edad en el colectivo de condenados. De esta forma evitamos distorsionar por exceso el rango de la variable.

Al dividir la frecuencia por la amplitud del intervalo, lo que obtenemos es la frecuencia absoluta por unidad de la variable. Vemos entonces que la **Moda** se encuentra en [18-21), es decir, entre 18 y 20 años, pues la frecuencia absoluta de ese intervalo se reparte entre menos unidades de la variable (3 años) que la frecuencia absoluta del intervalo [41-51), que ha de repartirse entre 10 unidades de la variable (una amplitud de 10 años).

Si representamos gráficamente la distribución a través de un histograma, se visualiza la moda en el intervalo de mayor altura ( $h$ ), es decir, en [18-21), aunque a muy poca distancia del intervalo [21-26). Se observa a partir del gráfico que la población condenada se concentra significativamente a edades jóvenes, descendiendo paulatinamente a partir de los 40 años.

**Distribución según edad de los adultos condenados en España (2016)**



Para conocer la **Mediana** necesitamos la distribución de las Frecuencias acumuladas. Al calcular el total N, podemos comprobar que N coincide con la frecuencia acumulada absoluta del último intervalo:

Edad	Frecuencia absoluta ( $n_i$ )	Frecuencia absoluta acumulada ( $N_a$ )
De 18 a 20	23.945	23.945
De 21 a 25	39.862	63.807
De 26 a 30	39.491	103.298
De 31 a 35	39.552	142.850
De 36 a 40	37.472	180.322
De 41 a 50	55.459	235.781
De 51 a 60	24.948	260.729
De 61 a 70	8.178	268.907
Más de 70	2.619	<b>271.526</b>
<b>TOTAL (N)</b>	<b>271.526</b>	

Me:  $\frac{N}{2} = \frac{271526}{2} = 135.763$  Esta cifra NO ES la mediana. Es la posición que ocupa el valor mediano de la distribución teniendo en cuenta que los casos se encuentran ordenados de menor a mayor valor de la variable (de menor a mayor edad). Así, solo sabemos que la mediana se encuentra en la posición 135.763 de los 271.526 casos.

Entonces, ¿cuál es la Mediana? Nos piden solo el intervalo, así que localizamos la posición 135.763 en la distribución de frecuencias acumuladas absolutas. El intervalo donde se encuentra la Mediana, que es el valor que divide a la distribución en dos partes iguales, incluirá esa posición. Por tanto, Me está en el intervalo: **“De 31 a 35 años”, o bien [31-36)**

b) Siguiendo la misma lógica, calculamos:

- $Q_3$ : Cuartil tercero

$$\frac{3N}{4} = \frac{814578}{4} = 203.644,5$$

El tercer cuartil es el valor de la distribución que deja por debajo de sí las  $\frac{3}{4}$  partes de los casos con valores de la variable más bajos (con edades más bajas) y, por encima  $\frac{1}{4}$  de casos con las edades más altas. Según la posición obtenida,  $Q_3$  se encuentra en el intervalo **“De 41 a 50”, o bien [41-51)**

- $D_4$ : Decil cuarto

$$\frac{4N}{10} = \frac{1086104}{10} = 108.610,4$$

El cuarto decil deja por debajo las  $\frac{4}{10}$  partes de los casos con menores edades (en realidad, el 40%), mientras que por encima deja las  $\frac{6}{10}$  partes de los casos con mayor edad (el 60%). Según la posición obtenida,  $D_4$  se encuentra, al igual que la mediana, en el intervalo **“De 31 a 35”, o bien [31-36)**



- P<sub>27</sub>: Percentil 27

$$\frac{27N}{100} = \frac{7331202}{100} = 73.312,02$$

Siguiendo el mismo procedimiento que en los casos anteriores, según su posición el percentil 27 se encuentra en el intervalo **“De 26 a 30”, o bien [26-31)**. El P<sub>27</sub> deja por debajo al 27% de los casos de la distribución con menor valor de la variable, y por encima, lógicamente, al 73% restante. Nótese que si dispusiéramos de la distribución de **frecuencias acumuladas relativas** y sus correspondientes porcentajes, podemos localizar directamente el intervalo en el que se encuentra cada **percentil**:

Edad	Frecuencia absoluta (n <sub>i</sub> )	Frecuencia relativa (fr <sub>i</sub> )	Frecuencia relativa acumulada (Fra <sub>i</sub> )
18-20	23.945	0,0882 (8,82%)	0,0882 (8,82%)
21-25	39.862	0,1468 (14,68%)	0,2350 (23,50%)
26-30	39.491	0,1455 (14,55%)	0,3805 ( <b>38,05%</b> )
31-35	39.552	0,1457 (14,57%)	0,5262 (52,62%)
36-40	37.472	0,1380 (13,80%)	0,6642 (66,42%)
41-45	28.839	0,1062 (10,62%)	0,7704 (77,04%)
46-50	26.620	0,0980 (9,80%)	0,8684 (86,84%)
51-60	24.948	0,0919 (9,19%)	0,9603 (96,03%)
61-70	8.178	0,0301 (3,01%)	0,9904 (99,04%)
71 o más	2.619	0,0096 (0,96%)	1 (100%)
<b>TOTAL (N)</b>	<b>271.526</b>	<b>1 (100%)</b>	

En este intervalo se encuentra P<sub>27</sub>

### Ejercicio 6

En la siguiente tabla se muestran datos sobre los suicidios consumados en España durante determinado período según el sexo y la edad.

Edad	Ambos sexos		Varón	Mujer
De 13 a 19	33	22	11	
De 20 a 29	180	147	33	
De 30 a 39	306	233	73	
De 40 a 49	295	228	67	
De 50 a 59	273	202	71	
60 y más	719	556	163	
Total	1.806	1.388	418	

Calcule las edades medias y las desviaciones típicas de varones y de mujeres. ¿Qué población, hombres o mujeres, está mejor representada por su media? Justifique su respuesta.

**Solución**

Para calcular las medias necesitamos en primer lugar cerrar el intervalo abierto “60 años o más”. Desgraciadamente la tabla no detalla los valores a partir de los 60 años. Como podemos ver, el número de suicidios aumenta con la edad, y la tasa de suicidios en personas con edades avanzadas será superior respecto de las edades inferiores. Por tanto tiene sentido llevar el límite superior de ese intervalo hasta un valor alto. Hemos cerrado el intervalo en 94 años siendo 95 el valor del límite superior.

Construimos una tabla con las marcas de clase de cada intervalo. También incluimos otras columnas para facilitar el cálculo de la edad media:

$$\bar{x} = \frac{\sum x_i n_i}{N}$$

Edad	$x_i$	$n_i$	$n_i$	$x_i n_i$	$x_i n_i$
		Hombres	Mujeres	Hombres	Mujeres
13 a 19	16,5	22	11	363	181,5
20 a 29	25	147	33	3675	825
30 a 39	35	233	73	8155	2555
40 a 49	45	228	67	10260	3015
50 a 59	55	202	71	11110	3905
60 a 94	77,5	556	163	43090	12632,5
Total		1388	418	76653	23114

$$\bar{x}_{\text{Hombres}} = \frac{76653}{1388} = 55,23 \text{ años}$$

$$\bar{x}_{\text{Mujeres}} = \frac{23114}{418} = 55,30 \text{ años}$$

Como podemos ver, a pesar de las mayores frecuencias de suicidios en los hombres, las edades medias casi coinciden.

Para calcular las desviaciones típicas de ambos colectivos emplearemos la fórmula:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{N}}$$

Construimos las tablas que nos faciliten los cálculos:

Edad	$x_i$	$n_i$ Hombres	$n_i$ Mujeres	$(x_i - \bar{x})^2 n_i$ Hombres	$(x_i - \bar{x})^2 n_i$ Mujeres
13 a 19	16,5	22	11	33000,28	16559,84
20 a 29	25	147	33	134336,38	30296,97
30 a 39	35	233	73	95355,93	30082,57
40 a 49	45	228	67	23860,86	7108,03
50 a 59	55	202	71	10,69	6,39
60 a 94	77,5	556	163	275749,81	80332,92
Total		1388	418	562313,95	164386,72

$$S_{\text{Hombres}} = \sqrt{\frac{562313,95}{1388}} = 20,13 \text{ años}$$

$$S_{\text{mujeres}} = \sqrt{\frac{164386,72}{418}} = 19,83 \text{ años}$$

La dispersión de las edades también es muy semejante en ambos casos, con un valor ligeramente superior en el caso de los hombres. Para confirmarlo, podemos calcular el coeficiente de variación (Cv):

$$Cv = \frac{S_x}{\bar{x}}$$

$$\text{Para los hombres: } Cv = \frac{20,13}{55,23} = 0,364$$

$$\text{Para las mujeres: } Cv = \frac{19,83}{55,30} = 0,359$$

El Cv confirma que la dispersión de las edades es muy semejante en ambas poblaciones. La media representaría solo ligeramente mejor a las mujeres que a los hombres, aunque se trata de una diferencia mínima.

### Ejercicio 7

Vamos a estudiar la distribución de renta en dos poblaciones, medida en los ingresos mensuales. Para ello contamos con las frecuencias absolutas ( $n_i$ ) y las frecuencias acumuladas ( $F_a$ ):

Población A			Población B		
€ mensual	$n_i$	$F_a$	€ mensual	$n_i$	$F_a$
[0-500)	75	75	[0-500)	10	10
[500-1000)	152	227	[500-1000)	47	57
[1000-1500)	170	397	[1000-1500)	152	209
[1500-2000)	201	598	[1500-2000)	221	430
[2000-2500)	81	679	[2000-2500)	186	616
[2500-3000)	27	706	[2500-3000)	101	717
[3000-4000)	12	718	[3000-4000)	65	782
$\geq 4000$	2	720	$\geq 4000$	28	810
TOTAL	720		TOTAL	810	

Se pide lo siguiente:

- ¿En qué tramo de ingresos se encuentra la Mediana de la distribución? Interpretéla
- ¿Qué tramo de ingresos deja por debajo al 25% de los individuos que menos ganan?
- ¿A partir de qué tramo se encuentra el 10% de los individuos más ricos?
- Según estos indicadores, ¿qué población parece mostrar mayor desigualdad en la distribución de la renta?

### Solución

a) Vamos a comparar la mediana de ambas distribuciones.

Para la población A:

El tramo de ingresos donde se encuentra la mediana está en la posición:  $\frac{N}{2} = \frac{720}{2} = 360$

La mediana de A se encuentra en [1000-1500), es decir, entre los 1000 y los 1499,99 euros.

Para la población B:  $\frac{N}{2} = \frac{810}{2} = 405$

La mediana de B se encuentra en [1500-2000), es decir, entre los 1500-1999,99 euros.

Comparando ambas medianas, vemos que en la población A el nivel de renta que deja por debajo (y por encima) al 50% de la población se sitúa en un tramo inferior al de la población B. Esto quiere decir que la mitad de la población que menos ingresos tiene, no llega a los 1500 euros en la población A, mientras sí los sobrepasa en la población B.

b) Para identificar el tramo de ingresos que deja por debajo al 25% de los individuos que menos ganan podemos calcular dos cuantiles equivalentes: el primer cuartil o bien el percentil 25. Ambos dan, lógicamente, un resultado idéntico. Calculemos el  $Q_1$ :

$$\text{Para la población A: } \frac{N}{4} = \frac{720}{4} = 180 \quad Q_1 : [500-1000)$$

$$\text{Para la población B: } \frac{N}{4} = \frac{810}{4} = 202,5 \quad Q_1 : [1000-1500)$$

El tramo de renta que divide al 25% que menos gana respecto del 75% que más ingresos tiene es de [500-1000) euros en la población A y de [1000-1500) en la población B. Por tanto, podemos ver que ese 25% de individuos con menos ingresos en cada población, llega a rentas inferiores en la población A que en la población B.

c) Se pide el percentil 90, o bien el decil 9, que es el nivel de ingresos que deja por debajo al 90% de la población y por encima al 10%. Calculemos el  $P_{90}$  (si calculamos la distribución de frecuencias relativas acumuladas, en porcentaje, lo podríamos obtener directamente de la tabla):

$$\text{Para la población A: } \frac{90N}{100} = \frac{64800}{100} = 648 \quad P_{90} : [2000-2500)$$

$$\text{Para la población B: } \frac{90N}{100} = \frac{72900}{100} = 729 \quad P_{90} : [3000-4000)$$

Vemos que el 10% de los individuos que más ingresos tienen en cada población, lo hacen a partir de una renta superior en la población B que en la población A. Esto es equivalente a decir que el 90% de la población A no llega en ningún caso a los 2500 euros. En cambio, en la población B, en ese 90% restante habría incluso individuos que podrían sobrepasar los 3000 € (sin llegar a los 4000).

d) Teniendo en cuenta el cálculo de cuantiles que hemos efectuado, parece que el reparto de la renta es más desigual en la población A, pues los cuantiles de posición baja (como el  $Q_1$ ), es decir, los valores menores de la variable, engloban a una proporción mayor de individuos que en la población B. Alternativamente, los cuantiles de posición alta (como el  $P_{90}$ ), es decir, los valores altos de ingresos, incluyen una proporción mucho menor de individuos en la población A que en la población B. En definitiva, hay más pobres y menos ricos en la población A.

### Ejercicio 8

Se ha obtenido del INE la evolución de la ocupación y el paro de la población española. Las medidas corresponden al tercer trimestre de los años comprendidos en el periodo 2011-2018:

Población activa en España (en miles)

	PARADOS	OCUPADOS	ACTIVOS
<b>2011T3</b>	4.998,0	18.484,5	23.482,5
<b>2012T3</b>	5.824,2	17.667,7	23.491,9
<b>2013T3</b>	5.943,4	17.230,0	23.173,4
<b>2014T3</b>	5.427,7	17.504,0	22.931,7
<b>2015T3</b>	4.850,8	18.048,7	22.899,5
<b>2016T3</b>	4.320,8	18.527,5	22.848,3
<b>2017T3</b>	3.731,7	19.049,2	22.780,9
<b>2018T3</b>	3.326,0	19.528,0	22.854,0

Construya un gráfico de áreas para visibilizar la estructura de la población activa durante el período considerado.

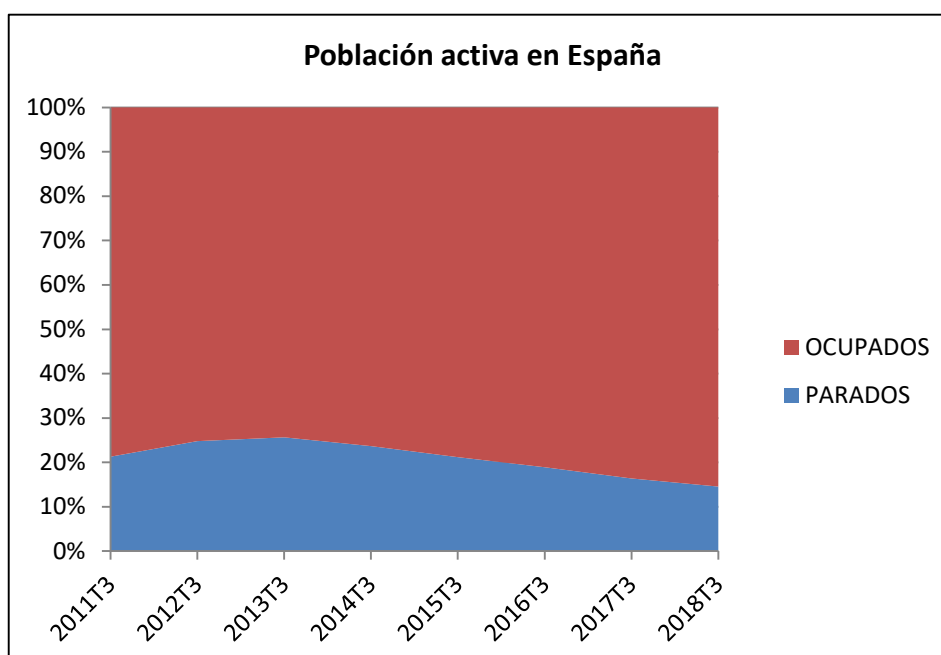
### Solución

La tabla ofrece la población activa (en miles) desagregada en población ocupada y población en paro. Se trata de obtener un gráfico que haga visible la diferencia entre “parados” y “ocupados” viendo su evolución en el tiempo. Para ello utilizaremos un gráfico de áreas.

Dado que en este gráfico la suma de las dos superficies que corresponden a “parados” y “ocupados” debe representar la población activa total, tenemos que calcular previamente el porcentaje de “ocupados” y de “parados” respecto del total de activos para cada año del período:

	PARADOS (%)	OCUPADOS (%)	ACTIVOS (%)
<b>2011T3</b>	21,3	78,7	100,0
<b>2012T3</b>	24,8	75,2	100,0
<b>2013T3</b>	25,6	74,4	100,0
<b>2014T3</b>	23,7	76,3	100,0
<b>2015T3</b>	21,2	78,8	100,0
<b>2016T3</b>	18,9	81,1	100,0
<b>2017T3</b>	16,4	83,6	100,0
<b>2018T3</b>	14,6	85,4	100,0

Representamos el gráfico situando el período en el eje de abscisas y la frecuencia relativa (%) en el eje de ordenadas, quedando como un gráfico de áreas apiladas:



Podemos observar que el paro asciende hasta el año 2013, donde superaba el 25% de la población activa, para ir descendiendo paulatinamente, respecto al total de activos, a partir de ese año y hasta el final del período.

### Ejercicio 9

Según datos del INE, el flujo de inmigración procedente del extranjero durante el segundo semestre de 2017 fue el siguiente, que se ha desagregado por tramos de edad:

<b>Edad</b>	<b>Inmigrantes</b>
0-4 años	16.962
5-9 años	15.576
10-14 años	14.987
15-19 años	23.129
20-24 años	39.656
25-29 años	45.663
30-34 años	34.667
35-39 años	26.280
40-44 años	20.026
45-49 años	15.629
50-54 años	12.258
55-59 años	9.613
60-64 años	8.682
65-69 años	7.130
70-74 años	4.046
75-79 años	2.323
80-84 años	1.346
85-89 años	637

Se pide representar el gráfico de frecuencias acumuladas (ojiva) y localizar, en el propio gráfico, la mediana de la distribución

### Solución

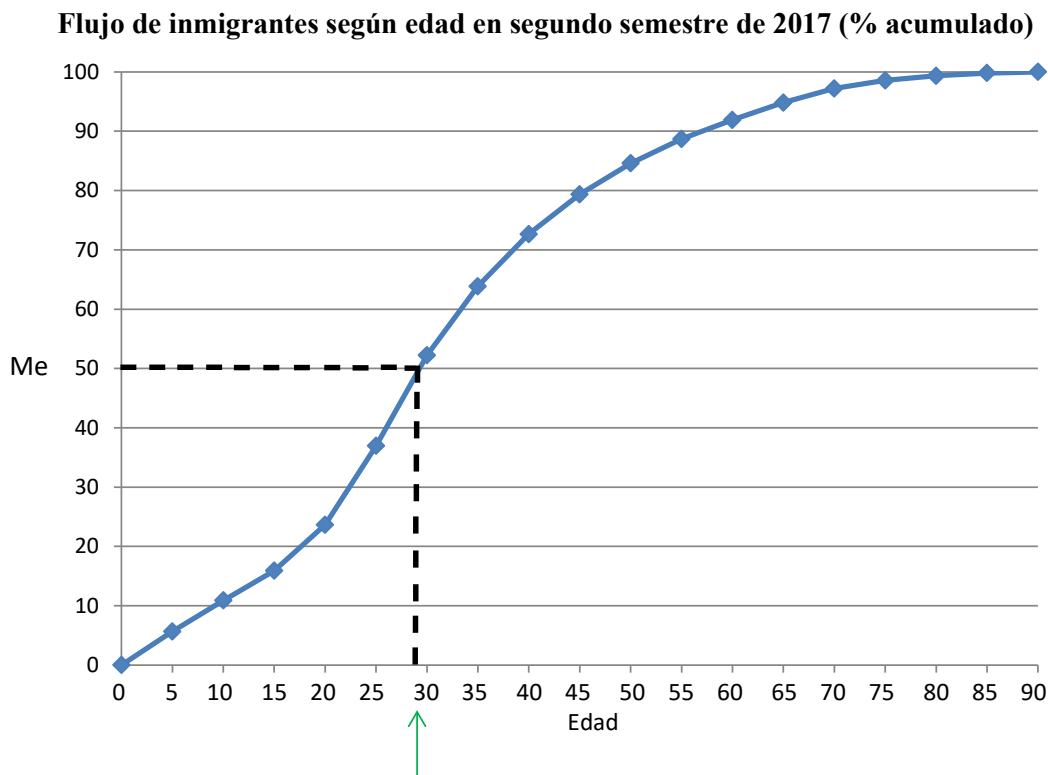
El gráfico de frecuencias acumuladas (ojiva) permite observar la proporción de casos que se encuentran por encima o por debajo de un determinado punto. Por eso, es interesante construirlo con las frecuencias relativas, pues desde el propio gráfico se podrán analizar los diferentes cuantiles. Para ello, sumamos las frecuencias absolutas para obtener el N total, construimos la tabla de frecuencias relativas (en porcentaje) y seguidamente el porcentaje acumulado para cada tramo de edades:

$x_i$	$n_i$	%	% Acumulado
0-4 años	16.962	5,68	5,7
5-9 años	15.576	5,22	10,9
10-14 años	14.987	5,02	15,9
15-19 años	23.129	7,75	23,7
20-24 años	39.656	13,28	36,9
25-29 años	45.663	15,29	52,2
30-34 años	34.667	11,61	63,8
35-39 años	26.280	8,80	72,6
40-44 años	20.026	6,71	79,3
45-49 años	15.629	5,23	84,6
50-54 años	12.258	4,11	88,7
55-59 años	9.613	3,22	91,9
60-64 años	8.682	2,91	94,8
65-69 años	7.130	2,39	97,2
70-74 años	4.046	1,35	98,6
75-79 años	2.323	0,78	99,3
80-84 años	1.346	0,45	99,8
85-89 años	637	0,21	100,0
TOTAL	298.610	100,0	

Dado que la variable edad está agrupada en intervalos, para construir el gráfico situaremos los puntos que unen los distintos segmentos de la ojiva en el límite superior de cada intervalo, pues estamos representando frecuencias acumuladas.

Podemos apreciar que la ojiva describe, siempre que los valores de la variable estén dispuestos de menor a mayor, una línea creciente cuya pendiente es más o menos pronunciada según la proporción de casos que acumule cada valor de la variable. En este caso, vemos que la curva se hace más vertical en las edades que van desde los 18 hasta aproximadamente los 40 años, para ir suavizándose paulatinamente a partir de esa edad. Ello indica que una proporción importante del flujo de inmigrantes se produce a edades adultas jóvenes, en etapas de la vida preparadas para formar parte de la población activa. Las entradas a edades infantiles y ancianas, de menor magnitud, podrían corresponder a reagrupamientos familiares.





Desde el gráfico podemos localizar directamente el intervalo al que corresponde la mediana: intervalo entre los 25 y 29 años, lo cual indica que ya el 50% del flujo migratorio hacia nuestro país se sitúa en edades bastante jóvenes, por debajo de los 30 años. Invitamos al alumno a localizar otros cuantiles de interés para el estudio de la estructura de edades de la población inmigrante.

### Ejercicio 10

La encuesta sobre Desigualdad y Movilidad Social del CIS (noviembre de 2017) muestra los siguientes resultados sobre el grado de acuerdo, medido en una escala del 1 al 6, sobre la afirmación: *No tengo miedo de expresar mis opiniones, incluso cuando son opuestas a las de la mayoría de la gente*:

No tengo miedo de expresar mis opiniones...		Porcentaje (%)
Totalmente en desacuerdo	1	3,10
	2	5,60
	3	12,29
	4	18,09
	5	25,46
Totalmente de acuerdo	6	34,49
	NS/NC	0,97
Total (N) = 2.482		100

Se pide calcular el Coeficiente de Variación de la distribución obteniendo previamente las frecuencias absolutas de la distribución.

### Solución

Para calcular el Coeficiente de Variación (Cv) necesitamos calcular previamente la media aritmética ( $\bar{x}$ ) y la desviación típica ( $S_x$ ) de la distribución:

$$Cv = \frac{S_x}{\bar{x}}$$

Calculamos primero la **media**  $\bar{x}$ , y para ver la lógica de su cálculo, hallaremos las frecuencias absolutas ( $n_i$ ) de cada valor de la variable.

$$\bar{x} = \frac{\sum x_i n_i}{N}$$

En la tabla tenemos las frecuencias relativas (fr) en forma de porcentaje, y obtenemos las frecuencias absolutas a partir de ellas:

$$n_i = fr \cdot N \quad \text{sabiendo que } fr = \frac{\%}{100}$$

$x_i$	%	fr	$n_i$
1	3,10	0,031	77
2	5,60	0,056	139
3	12,29	0,1229	305
4	18,09	0,1809	449
5	25,46	0,2546	632
6	34,39	0,3439	856
NS/NC	0,97	0,0097	24
Total	100	1	2.482

Como las **frecuencias absolutas  $n_i$**  siempre serán números enteros (ya que representan al número de individuos que han optado por cada una de las posiciones sobre el grado de acuerdo con la afirmación del cuestionario) si en su cálculo aparecen decimales debemos redondear al valor entero más próximo. Así, por ejemplo, para hallar la frecuencia absoluta del primer valor de la variable  $x_i = 1$  hemos procedido de la siguiente forma:

$$n_i = fr \cdot N \quad n_i = 0,031 \cdot 2.482 = 76,942 \approx 77$$

Y así sucesivamente para cada valor de la variable.

Es muy importante tener en cuenta que solo podemos calcular medidas como la media y la desviación típica en variables de intervalo o de razón cuyos valores vienen expresados numéricamente (cuantitativamente). Si nos fijamos, los valores de la variable del enunciado son números del 1 al 6 que representan posiciones en una escala de acuerdo o desacuerdo. Pero la categoría NS/NC, (quienes “no saben” o “no contestan”) no se expresa numéricamente: es una categoría cualitativa que indica que esos individuos no han señalado ninguna posición numérica. Por eso **no podemos incluir los NS/NC dentro del cálculo de la media porque no representan ninguna posición dentro de la escala numérica de la variable**. Por tanto, su frecuencia absoluta ( $n_i$ ) se ignora, tanto para el cálculo de la media como de la desviación típica, y es necesario también restarla del total (N) a partir del cual calculamos esas medidas.

Para el cálculo de la media y de la desviación típica  $N = 2.482 - 24 = 2.458$

$$\bar{x} = \frac{\sum x_i n_i}{N} = \frac{11.362}{2.458} = 4,6224573 \text{ es la posición media}$$

Para el cálculo de la **desviación típica** podemos optar por cualquiera de estas dos fórmulas:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{N}} \quad \text{o bien} \quad S_x = \sqrt{\frac{\sum x_i^2 \cdot n_i}{N} - \bar{x}^2}$$

Optaremos en este ejercicio por la segunda formulación. Construimos esta tabla:

$x_i$	$n_i$	$x_i^2$	$x_i^2 \cdot n_i$
1	77	1	77
2	139	4	556
3	305	9	2.745
4	449	16	7.184
5	632	25	15.800
6	856	36	30.816
Total (N)	2.458		$\Sigma = 57.178$

$$S_x = \sqrt{\frac{\sum x_i^2 \cdot n_i}{N} - \bar{x}^2} = \sqrt{\frac{57.178}{2.458} - 4,6224573^2} =$$
$$= \sqrt{23,262002 - 21,36711} = 1,37655 \text{ puntos}$$

La desviación típica  $S_x$  es una medida de dispersión. En este caso nos indica que, por término medio, las puntuaciones señaladas por los individuos se separan alrededor de 1,377 puntos de la posición media (que hemos visto que era de 4,62).

Sin embargo, el ejercicio nos pide calcular otra medida de dispersión, el Coeficiente de Variación, que nos permitiría comparar la dispersión de esta distribución con la de cualquier otra población (siempre que los valores de las variables sean positivos). A mayor CV, peor representatividad de la media, y viceversa. En este caso:

$$Cv = \frac{S_x}{\bar{x}} = \frac{1,37655}{4,6224573} = 0,298$$

Si lo expresamos en porcentaje, el Cv indica una dispersión relativa respecto a la media de alrededor del 30%.



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

La autoría de este trabajo corresponde al equipo docente de la asignatura Estadística Social del Grado en Criminología de la UNED: Beatriz Mañas Ramírez, Alejandro Almazán Llorente y José María Arribas Macho.

[http://www2.uned.es/socioestadistica/Crim/Ejercicios\\_resueltos\\_Tema2\\_Estadisticos\\_de\\_resumen.pdf](http://www2.uned.es/socioestadistica/Crim/Ejercicios_resueltos_Tema2_Estadisticos_de_resumen.pdf)