



# Socioestadística I

## Análisis estadístico en Sociología

### Capítulo 2

#### ESTADÍSTICA DESCRIPTIVA UNIVARIANTE. ANÁLISIS COMPARATIVO

#### I. ANÁLISIS COMPARATIVO EN SOCIOESTADÍSTICA

En Sociología se hace continuamente uso de las comparaciones para avanzar el pensamiento sociológico, pero siempre que hacemos comparaciones necesitamos un referente, algo que nos permita discernir diferencias apreciables entre dos fenómenos.

La sociología comparada no es una rama de la sociología, es la sociología misma, en tanto que cesa de ser meramente descriptiva y aspira a explicar los hechos.

Por todo ello, cuando hacemos sociología comparada, debemos plantearnos dos preguntas: ¿Que debe compararse? - ¿Cómo debe hacerse la comparación?

Ejemplo: El número medio de hijos que han tenido las mujeres casadas o que han estado casadas en dos periodos de tiempo distintos. En el censo de 1950, el número medio de hijos, era de 3,1 hijos por mujer; después de 20 años el número medio de hijos por mujer ha sido de 2,8 hijos. Que es lo que se puede explicar de la comparación entre estas dos estadísticas. Pues que se ha producido un descenso de la fecundidad en el periodo que va de 1950 hasta 1970.

Por tanto, en socioestadística, en muchas ocasiones, las comparaciones que se realizan, simplemente, son una comparación del mismo fenómeno en periodos de tiempo distintos, pero a la vez necesarias, esas comparaciones, para explicar el fenómeno.

#### I.2. Tipos de comparación. ¿Qué debe compararse?.

Para establecer los tipos de comparación en socioestadística, debemos tener claro una serie de aspectos, con ello podremos determinar que es lo que se debe comparar:

1. Primero, depende de que es lo que se está estudiando, considerar si es o se trata de un tema estrictamente teórico.
2. Plantear o formular el problema. Será más fácil saber que datos son necesarios reunir para hacer la comparación o para tratar de explicar el problemas.
3. Y por último, establecer el marco teórico y el marco de referencia, ósea, identificar todas las variables y los objetos o fenómenos que se van a medir para que se pueda contrastar consistentemente grupos comparables.



Con estas consideraciones, podemos establecer los siguiente tipos de comparaciones:

**a. Comparaciones entre grupos** y dentro de estas, ya sean las realizadas en el mismo estudio o las realizadas entre diferentes estudios.

- Es el modelo de comparación ideal desde el punto de vista científico.
- Es el realizado entre un grupo experimental (a los que se les somete a algún tipo de tratamiento conocido) y un grupo de control (a los que no se les somete a ningún tipo de tratamiento).
- Consiste por tanto, en la comparación realizada entre los dos grupos en dos momentos de tiempo distintos, esto es, antes y después de someter al primer grupo al tratamiento especial.
- Los grupos que se comparan, pueden ser constituidos por individuos, cosas u objetos no personales (ejemplo: instituciones sociales, organismos...).
- Comparamos un serie de puntuaciones sobre medidas de dimensiones o aspectos definidores del grupo. Por tanto, hay que resumir dichas puntuaciones por medio del estadístico que se considera más oportuno, comparándose de este modo los estadísticos resúmenes de cada grupo

**b. Comparaciones entre un grupo y un individuo.**

- En este tipo, hay que delimitar y definir las características del grupo que se compara con las correspondientes al individuo, o el de un caso individual. (Ejemplo: comparar los resultados escolares de un individuo con los correspondientes a la media de la clase de la que es miembro el individuo).

**c. Comparación entre el resultado de un estudio y un resultado estándar**

- Cuando nos referimos a un resultado estándar, no referimos a los resultados de un estudio anterior que nos sirve de referencia.
- También, tenemos que destacar, que hay muchas teorías que nos sirven para realizar comparaciones y a veces, además de para contrastar, podemos llegar a modificar, rechazar la teoría que sirve de referencia o estándar.



### 1.3. Operaciones básicas de comparación. ¿Como debe hacerse la comparación?

Dentro de la estadística descriptiva se pueden realizar dos operaciones fundamentales. La primera se refiere a la organización y ordenación de los datos o medidas obtenidas en algún tipo de distribución y la segunda se refiere al tratamiento aritmético de dichos datos.

#### 1.3.1. Organización de los datos.

Los datos siempre hay que organizarlos y ordenarlos siguiendo algún tipo de criterio que nos permita observar con mayor facilidad la distribución de los datos. Todo ello hay que establecerlo o tenerlo en cuenta en nuestro marco teórico, en nuestro planteamiento de hipótesis, etc.

La mejor forma de verlo, es plantear un ejemplo donde se observa la organización de nuestros datos según el objetivo que perseguimos en la investigación, según el marco teórico de la investigación.

Tabla 1. Residentes Europeos en España. 1979. Ordenador por orden alfabético.

País	Nº residentes
Alemania	18.444
Austria	1.145
Bélgica	3.764
Dinamarca	2.009
Finlandia	892
Francia	14.891

Tabla 2. Residentes Europeos en España. 1979. Ordenador por nº de residentes.

País	Nº residentes
Alemania	18.444
Francia	14.891
Bélgica	3.764
Dinamarca	2.009
Austria	1.145
Finlandia	892



Tabla 3. Residentes Europeos en España. 1979. Ordenador por áreas regionales.

Europa del Este	
País	Nº residentes
Polonia	88
Hungría	36

Mediterráneo	
País	Nº residentes
Grecia	425
Italia	9.192
<b>TOTAL</b>	<b>9.617</b>

Península Ibérica	
País	Nº residentes
Portugal	21.801
<b>TOTAL</b>	<b>21.801</b>

Resto Europa Occidental	
País	Nº residentes
Alemania	18.444
Gran Bretaña	17.330
<b>TOTAL</b>	<b>35.774</b>

La agrupación realizada, pues, ha permitido realizar una comparación con la que se puede analizar de forma más relevante los datos originales. El marco teórico en el que se inscriba el análisis cuantitativo debe de ser, en toda investigación empírica, el criterio básico que se ha de seguir para agrupar los datos y poder realizar una comparación significativa. En el caso de la última tabla (el número de residentes según sus países de procedencia y organizados por zonas regionales), nos da más posibilidades a la hora de hacer comparaciones porque tenemos más argumento de comparación. No solo el número de residentes, sino también el país y el área de procedencia.



### 1.3.2. Distribuciones.

#### 1.3.2.1. Distribución de frecuencias.

Cuando se maneja un número amplio de datos, conviene distribuirlos en **CLASES** o **CATEGORÍAS** y determinar el número de casos que pertenecen a cada clase, este número se denomina **FRECUENCIA DE CLASE** y se simboliza por la letra (*f* o *f<sub>i</sub>*). El número total de casos es la suma de la columna de las frecuencias y se representa por **N** o  $\sum f_i$ .

El número de clase o categoría que seleccionemos vendrá determinado por las necesidades y sabiendo que siempre perdemos información porque los datos vienen agrupados y no hacemos distinciones, pero por otra parte, se gana en comodidad.

Edad (años)	f <sub>i</sub>
De 18 a 20	10
21 a 25	14
26 a 35	23
36 a 45	20
46 a 60	29
61 a 74	24
<b>TOTAL</b>	<b>N=120</b>

→ INTERVALO DE CLASE O CATEGORÍA

FRECUENCIA DE CLASE ←

Se simboliza al intervalo de clase o categoría de la siguiente forma: de 18 a 20.

Los números extremos de cada clase o categoría, se denomina los **LÍMITES DE CLASE**. El mayor es el **LÍMITE SUPERIOR** y el menor es el **LÍMITE INFERIOR**.

La clase o categoría que no tienen límite inferior o límite superior, se conocen con el nombre de **INTERVALO DE CLASE O CATEGORÍA ABIERTA**. Ejemplo: "menos de 21 años", "más de 60 años".

Si las edades se registran con aproximación de meses, el intervalo de la categoría "de 21 a 25 años" incluye todos los individuos con edades comprendidas entre los 20,5 años y los 25,5 años, a esto, le denominamos los **límites REALES O VERDADEROS DE CLASE O CATEGORÍA**. El menor es el límite real inferior y el mayor es el límite real superior.

Estos límites se obtienen de la siguiente ecuación:



$$\frac{\text{Límite Superior de un intervalo} + \text{Límite Inferior del intervalo siguiente}}{2}$$

Ejemplo: El intervalo real superior del intervalo de 21 a 25 años sabiendo que el siguiente es el intervalo de 26 a 35 años.

$$\frac{25+26}{2} = \frac{52}{2} = 25,5$$

No es conveniente utilizar estos límites como intervalos porque no son mutuamente excluyentes y podríamos incurrir en ambigüedad, por esta razón no se suele utilizar estas categorías.

AMPLITUD / TAMAÑO / LONGITUD de la clase o categoría es la diferencia entre los límites reales que forman cada clase o categoría. Puede ser idéntico o no.

Intervalos	Límites Reales	Diferencia
Menos de 21	20,5 - 17,5	3
21 a 25	25,5 - 20,5	5
26 a 35	35,5 - 25,5	10
36 a 45	45,5 - 35,5	10
46 a 60	60,5 - 45,5	15
Más de 60	75,5 - 60,5	15

\* La muestra son individuos mayores de 18 años y menores de 65 años

El punto intermedio del intervalo de clase o categoría se obtiene sumando el intervalo inferior y el superior de la clase o categoría y dividiéndolo por dos. PUNTO MEDIO DE LA CLASE O CATEGORÍA y se simboliza por  $x_i$

Ejemplo: punto intermedio en el intervalo de 21 a 25 años

$$\frac{21+25}{2} = \frac{46}{2} = 23$$



Intervalos	fi	Punto Medio Xi	Límites Reales
De 18 a 20	10	19	17,5 - 20,5
21 a 25	14	23	20,5 - 25,5
26 a 35	23	30,5	25,5 - 35,5
36 a 45	20	40,5	35,5 - 45,5
46 a 60	29	53	45,5 - 60,5
61 a 75	24	68	60,5 - 75,5
<b>TOTAL</b>	<b>N=120</b>		

ERROR DE AGRUPAMIENTO: son las alteraciones que se producen al realizar determinados agrupamientos y conduce a la variación de N.

Ejemplo de Error de Agrupamiento:

**Modelo I**

Intervalos	fi	Xi
1	5	5
2	2	4
3	1	3
4	2	8
5	0	0
6	3	18
<b>TOTAL</b>	<b>N=13</b>	<b>38 = fXi</b> Nº de casos totales de cada categoría

\* En el modelo I agrupamos por la unidad

**Modelo II**

Clase	Xi	fi	fXi
1 a 2	1,5	7	10,5
3 a 4	3,5	3	10,5
5 a 6	5,5	3	16,5
		<b>N = 13</b>	<b>37,5 = fXi</b>

\* En el modelo 2 agrupamos de dos en dos

En el ejemplo, el número de casos en cada categoría no coinciden.

La diferencia entre ambos números de debe a que hemos calculado los totales parciales fXi utilizando el valor medio de cada categoría.



La diferencia entre los dos totales (38 y 37,5) es lo que se llama **ERROR DE AGRUPACIÓN** y se produce porque los puntos medios de cada clase o categoría no representa convenientemente el valor de los casos que se engloban en cada categoría

SPIEGEL (1975) establece dos reglas para formar distribuciones y minimizar el error de agrupamiento:

1. Determinar el mayor y el menor entre los datos disponibles y así encontrar el rango (diferencia entre el mayor y el menor de los datos).
2. Dividir el rango en un número conveniente de intervalos de idéntico tamaño y si no es posible, utilizaremos intervalos de diferentes tamaños o abiertos. El número de intervalos es de 5 a 20 dependiendo de los datos de partida y los intervalos se elegirán de forma que los puntos intermedios coincidan con los datos realmente observados.

#### 1.3.2.2. Distribución Porcentual.

Para calcular un porcentaje es preciso calcular una proporción previamente. La proporción de casos en una categoría dada es igual al nº de casos en la categoría dividido por el total de casos en la distribución. El valor de una proporción no puede ser superior a la unidad y la suma de las proporciones en todas las categorías será la unidad.

Los porcentajes se obtienen a partir de las proporciones multiplicando por 100 de ahí que también se denominen por ciento.

Los porcentajes suponen estandarizar según el tamaño. Es calcular el número total de casos en una categoría si el número total de casos fuera 100 y si la proporción en cada categoría no se alterase. Del mismo modo que la suma de los porcentajes de una distribución es igual a la unidad, la suma de sus porcentajes será 100.

Si en lugar de utilizar valores absolutos en una distribución de frecuencias, utilizamos porcentajes, tendremos una **DISTRIBUCIÓN PORCENTUAL**. Esto tiene una serie de ventajas: por un lado facilita la comparación y por otro lado evita una fuente errores.



Intervalos	fi	% A	% B
De 18 a 20	10	8,33	8,3
21 a 25	14	11,66	11,7
26 a 35	23	19,16	19,2
36 a 45	20	16,66	16,7
46 a 60	29	24,16	24,2
61 a 75	24	20,00	20,0
<b>TOTAL</b>	<b>N=120</b>	<b>99,97</b>	<b>100,1</b>

Para calcular el porcentaje en cada categoría, se ha dividido cada  $f_i$  por  $N$  y se ha multiplicado por 100. Observar que el porcentaje (A) no es exactamente 100, debido a que sólo hemos tomado dos cifras decimales y no hemos redondeado el porcentaje resultante. Es aconsejable utilizar solo una cifra decimal. Si utilizamos una cifra, nos pasamos pero redondeando posteriormente el total, tendríamos el valor 100.

Hay que señalar que para calcular porcentajes, el valor de  $N$  debe de ser lo suficientemente elevado (Blalock, 1960), señala el número 50 como el mínimo aproximado de casos que ha de contar una distribución para poder calcular los porcentajes.

No siempre es conveniente usar porcentajes. Por ejemplo, si tenemos dos empresas, la empresa A que pasa a vender de 1 millón a 2 millones, lo que significa que ha incrementado sus ventas un 100% y la empresa B que pasa a vender de 4 a 7 millones, con un incremento del 75%. Visto de este modo parece que el incremento sea mayor en la empresa A que en la B. Si en vez de fijarnos en los porcentajes nos fijáramos en el volumen real de ventas, la empresa A pasa de vender 1 millón a 2 millones, hay un incremento sólo de 1 millón, mientras la empresa B que vendía 4 millones pasa a vender 7 millones, el incremento es de 3 millones. De este modo parece que en la empresa B aumentan más las venta que en la empresa A. Lo mismo se podría decir de la utilización de porcentajes cuando superan o exceden demasiado de 100. Cuando por ejemplo decimos que la población ha aumentado en un periodo un 1200 por 100. En estos casos sería mejor decir que la población ha aumentado 12 veces en relación al periodo anterior.

### 1.3.2.3. Distribución Acumulada.

Se forma al indicar para cada categoría el número (o porcentaje) de casos que queda por debajo del límite real superior de dicha categoría.



Comenzaremos a acumular desde las categoría de orden inferior e ir así acumulando hasta N o 100 por 100, según se trate respectivamente, de una distribución de frecuencias o una distribución porcentual.

Intervalos	fi	%	Frecuencia Acumulada	Porcentaje Acumulado
De 18 a 20	10	8,3	10	8,3
21 a 25	14	11,7	24	20,0
26 a 35	23	19,2	47	39,1
36 a 45	20	16,7	67	55,8
46 a 60	29	24,2	96	80,0
61 a 75	24	20,0	120	100
<b>TOTAL</b>	<b>N=120</b>			

Las distribuciones acumuladas son útiles en la comparación cuando se desea comparar la forma en que los casos se distribuyen a lo largo de una escala. Permiten en una sola medida ofrecer los casos que se encuentran por debajo o por encima de unos niveles determinados.

### 1.3.3. Percentiles.

El percentil es el valor por debajo del cual queda un porcentaje determinado de casos y se representa con la letra “P” (Pi oscila entre los valores 1 y 100). Ejemplo, el percentil 20 o  $P_{20}$  deja por debajo de su valor un 20 por 100 de casos.

El valor que divide a los datos en dos partes iguales es el percentil  $P_{50}$  y se llama también MEDIANA.

Cuando dividimos los datos en cuatro partes iguales tenemos lo que llamamos los CUARTILES y viene representado por:  $Q_1$ ,  $Q_2$  (MEDIANA),  $Q_3$  y  $Q_4$ .

Cuando divido los datos en diez partes iguales tenemos lo que llamamos los DECILES y viene representado por  $D_1, D_2, \dots, D_{10}$ .

El  $P_{50}$  es igual  $Q_2$  y esto es lo mismo que hablar de la MEDIANA:  $P_{50} = Q_2 = \text{MEDIANA}$ .

Los percentiles pueden ser positivo o negativos y no se refieren a un dato sino al valor que deja un porcentaje.



Vamos a ver de que forma se puede calcular un percentil. Para calcular el rango del percentil de una puntuación determinada utilizaremos la siguiente formula:

Xi
3
5
9
11
15
17
22

$$\text{Rango de percentil de una puntuación dada} = \frac{\text{Lugar que ocupa en la distribución}}{\text{n}^\circ \text{ de valores}}$$

$$\text{Rango de percentil del valor 15} = \frac{5}{7} \times 100 = 71,43 / \text{Percentil 70}$$

$$\text{Rango de percentil del valor 5} = \frac{2}{7} \times 100 = 28,57 / \text{Percentil 30}$$

O podemos realizarlo al revés, ósea, calcular el valor que se encuentra en el percentil 50:

$$\text{Valor que buscamos} = \text{Percentil en tantos por uno} \times \text{n}^\circ \text{ de valores}$$

$$\text{Valor que buscamos} = 0,50 \times 7 = 3,5 \text{ redondeando es } 4 \text{ (lugar que ocupa)}$$

$$P_{50} = 11$$

$$P_{70} = 0,70 \times 7 = 4,9 \text{ redondeando es } 5 \text{ (lugar que ocupa)}$$

$$P_{70} = 15$$

$$P_{30} = 0,30 \times 7 = 2,1 \text{ redondeando es } 2 \text{ (lugar que ocupa)}$$

$$P_{30} = 5$$

Es útil para comparar unas distribuciones, unos grupos específicos, situados en un lugar específico de las distribuciones, con otro grupos situados en el mismo o diferente lugar.

El **cálculo del percentil** cuando los datos vienen organizados en intervalos se calcula de la siguiente forma:

$$P = \text{lim inf} + \frac{ai}{fi} (P - F)$$

Donde “lim inf” corresponde al limite inferior donde se encuentra el valor que corresponde al percentil que vamos a calcular, “ai” es la amplitud del intervalo que estamos utilizando, “fi” es la frecuencia relativa de ese mismo intervalo, “P” es el valor del percentil que estamos calculando y “F” es la frecuencia acumulada del intervalo anterior.



Ejemplo del calculo del percentil cuando los datos vienen ordenados por intervalos:

Intervalos	fi	Frecuencia Acumulada
5-7	1	10
8-10	2	3
11-13	5	8
14-16	8	16
17-19	11	27
20-22	14	41
23-25	18	59
26-28	13	72
29-31	10	82
32-34	9	91
35-37	4	95
38-40	2	97
41,43	1	98
<b>TOTAL</b>	<b>N=98</b>	

$P_{75} = 28,95$  (indicated by a red dashed line pointing to the interval 26-28) and  $P_{75} = 73,5$  (indicated by a red dashed line pointing to the interval 29-31).

Vamos a calcular el percentil 75 ( $P_{75}$ ) de esta distribución: Primero, calculamos el valor que se encuentra en el percentil 75 ( $P_{75}$ ) según la formula anterior:

$$P_{75} = 0,75 \times 98 = 73,5$$

$$P_{75} = 73,5$$

Segundo, busco el intervalo donde se encuentra la puntuación que hemos obtenido (73,5), en este caso, busco en la frecuencia absoluta, donde se encuentra el valor 73,5 y obtengo el intervalo al que corresponde ese valor:



$$P = \text{lim inf} + \frac{ai}{fi} (P - F)$$

Y tercero, sustituyo en la ecuación los diferentes términos y obtengo el valor que busco.

$$P = 28,5 + \frac{(31,5 - 28,5)}{10} (73,5 - 72)$$

$$P_{75} = 28,95$$

### 1.3.4. Razón.

La razón de un número A (numerador) a otro número B (denominador) se define como A dividido por B.

Ejemplo: En un parlamento la fuerzas políticas se reparten del siguiente modo: 160 de izquierda, 150 de derecha y 80 regionalistas. La razón de diputados de izquierda a los



de derecha será  $160/150$ . Los diputados de izquierda y regionalistas a los de derecha, será  $(160+80)/150$ , aunque, como en este caso, se podrían reducir las cifras eliminando el último cero  $(16+8)/15 = 24/15$ .

Las proporciones y porcentajes son un tipo especial de razón, donde en el denominador tenemos el número total de casos y el numerador una fracción dada de dicho número, en el caso de la proporción. Y esa misma fracción del número multiplicada por 100 en el caso del porcentaje.

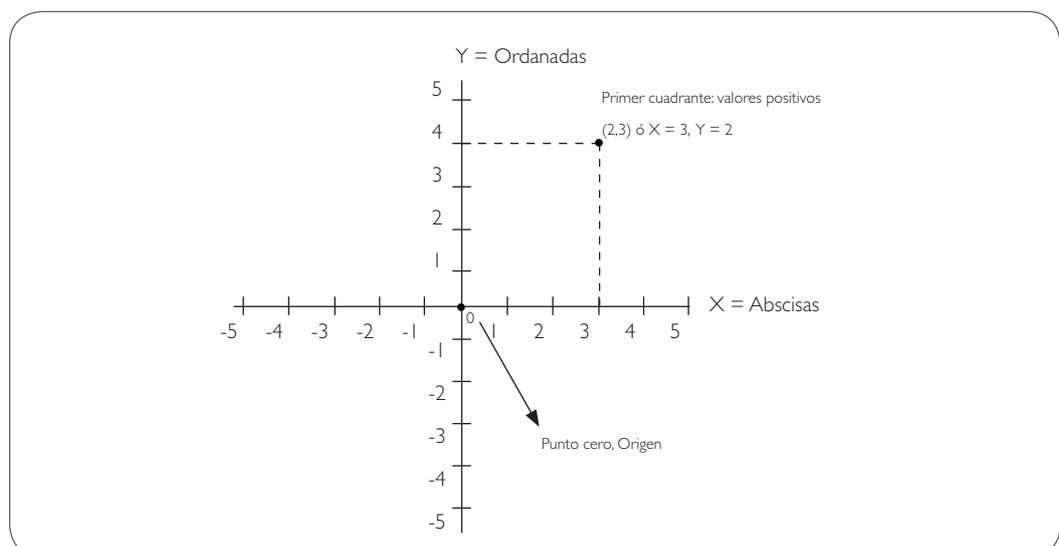
A diferencia de la proporción, la razón puede ser mayor que la unidad, es el caso en demografía para comparar el número de hombre y mujeres. En el número total de hombre respecto al de mujeres, se suele decir que hay más mujeres que hombre y se habla por ejemplo de una razón de 94 hombres por cada 100 mujeres (en el numerador tendríamos el número de hombres y en denominador el número de mujeres).

Cuando utilizamos bases mayores que 100, por ejemplo 1000, 10000 o un millón, hablamos entonces de TASA. Se utilizan las tasa cuando el uso de porcentajes arroja números decimales.

La TASA de CRECIMIENTO RELATIVO (otro tipo de razón) se utiliza cuando interviene o queremos tener en cuenta el tiempo o la evolución en el tiempo

#### 1.4. Técnicas de representación gráfica.

Las representaciones gráficas se basan en la representación sobre dos ejes (Sistema de Coordenadas Cartesianas) que consiste en dos líneas o dos dimensiones perpendiculares. La línea o eje Y (vertical) eje de ordenadas y la línea o eje X (horizontal) eje de abscisas (Désartes).



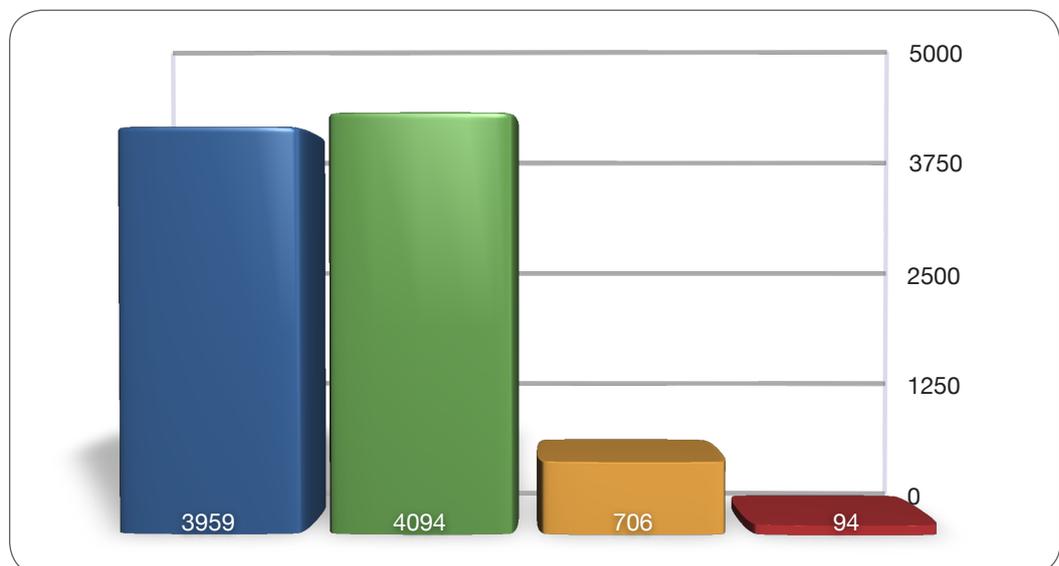


### 1.4.1. Histograma.

El histograma consiste en una representación de frecuencias o porcentual. Cada frecuencia (o porcentaje) se representa mediante un rectángulo que tiene su base sobre el eje horizontal (eje X), con centro, el punto medio de la categoría, la anchura es igual al tamaño del intervalo de dicha categoría y la altura será proporcional a la frecuencia (o porcentaje) de casos que se incluyen en la categoría. El área del rectángulo, será proporcional a la correspondiente frecuencia (absoluta o relativa).

Ejemplo: la distribución de familias según el número de miembros.

Nº de miembros	Punto medio	Frecuencias (en miles)
0,0 - 3	1,5	3959
3,1 - 6	4,5	4094
6,1 - 9	7,5	706
9,1 - 12	10,5	94
		<b>8853</b>



Los histogramas se pueden utilizar con ciertas variaciones según que los datos que forman la distribución de frecuencias se encuentre medidos a nivel nominal, ordinal o de intervalo.

Si es nominal, los rectángulos se separan para que se sepan que son categorías diferentes unas de otras.



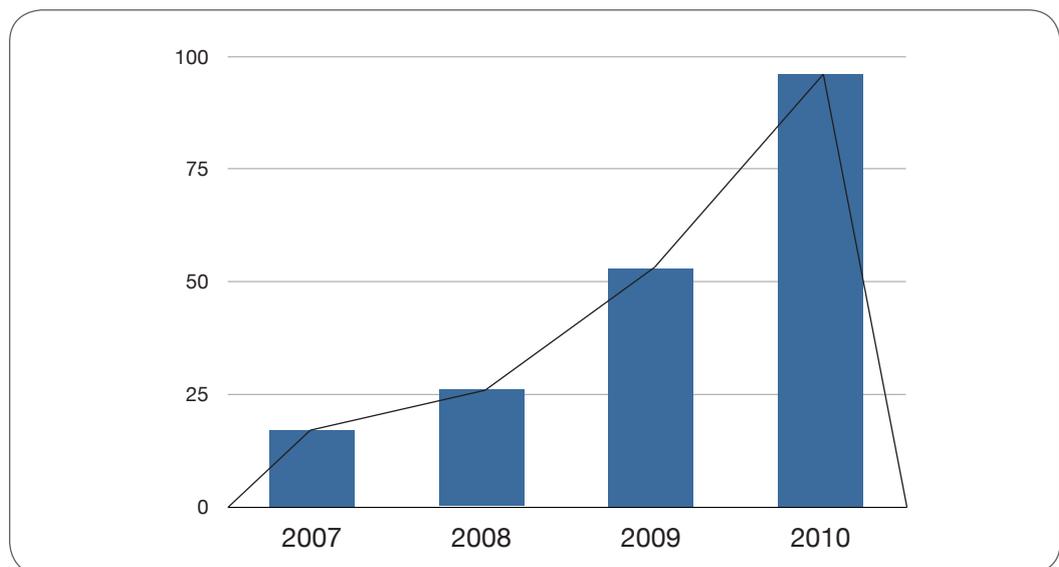
Si se trata de categorías ordinales (si no se distinguen distancias iguales) se suelen separar, aunque a veces se mantiene juntas por estética.

Y por último, normalmente se utilizan los histogramas con variables con un nivel de medida de intervalo y los rectángulos se representan juntos.

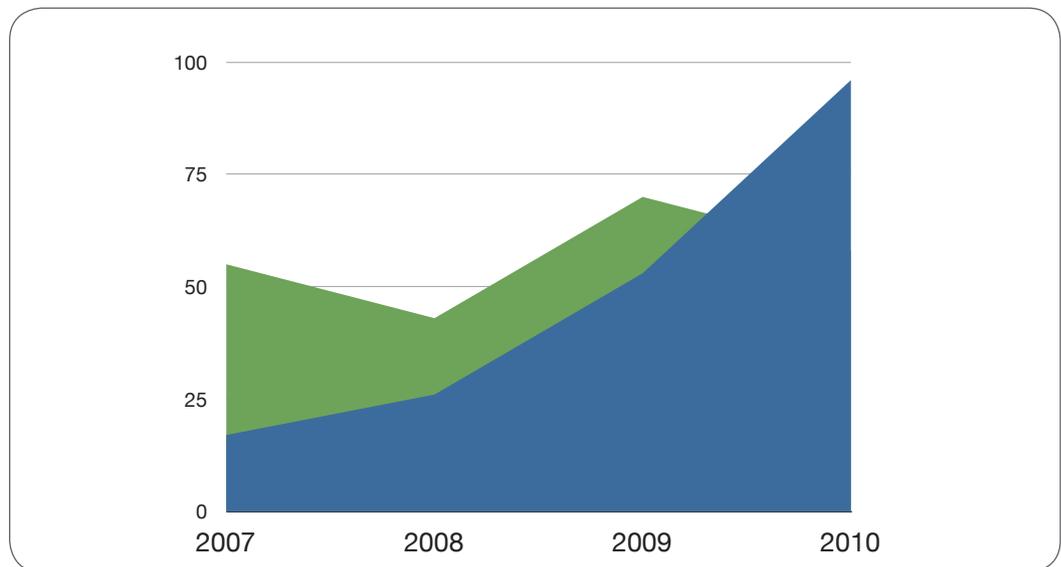
#### 1.4.2. Polígonos.

Es una figura que se cierra al unir los puntos medios de cada intervalo a una altura proporcional a la frecuencia (o porcentaje) de dicho intervalo.

La unión de esos puntos constituye un polígono de frecuencias. El área que queda debajo es igual al área contenida dentro del correspondiente histograma, es igual al área de la suma de los rectángulos.



Al superponer dos polígonos de frecuencias, veremos las coincidencias y diferencias entre dos distribuciones distintas, permitiendo si es oportuno establecer comparaciones entre las dos distribuciones. Calculando el área de coincidencia entre dos distribuciones distintas podremos establecer el porcentaje de igualdad entre las dos, pudiendo, además, establecer algún tipo de explicación del suceso dependiendo de la información de las distribuciones que estemos utilizando.



### 1.4.3. Ojivas.

La representación de una Ojiva, no es más que la representación de un polígono de frecuencias acumuladas.

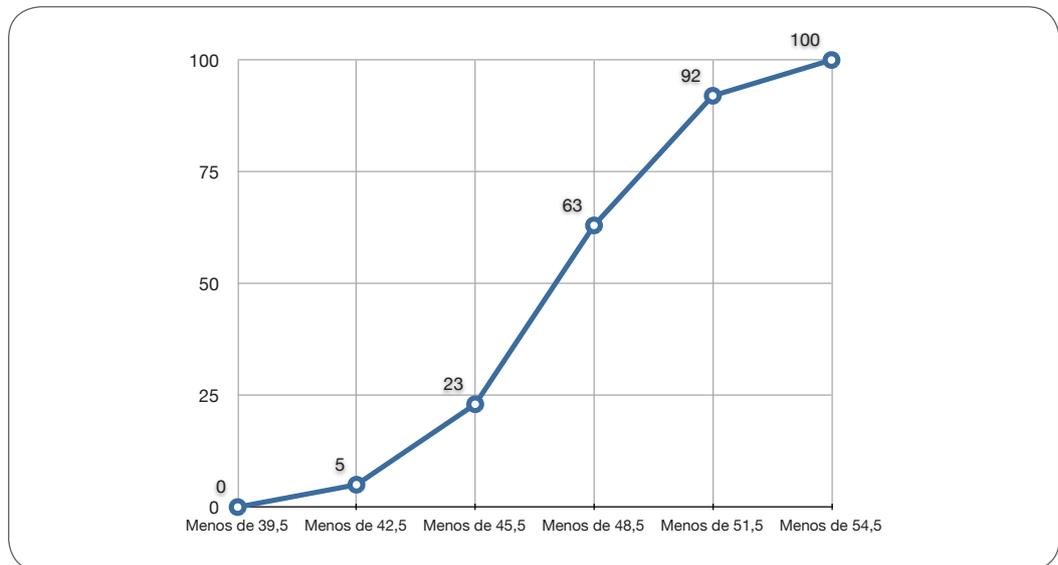
El primer punto de dicho polígono vendrá dado sobre el límite real inferior del primer intervalo. A continuación, en la vertical sobre el límite real superior de cada intervalo, y a una altura proporcional a la frecuencia (o porcentaje) acumulada de dicho intervalo, dibujaremos el primer punto. Uniendo cada uno de los puntos de dos en dos consecutivamente se obtiene el polígono de frecuencias acumuladas u ojiva.

Distribución de frecuencias

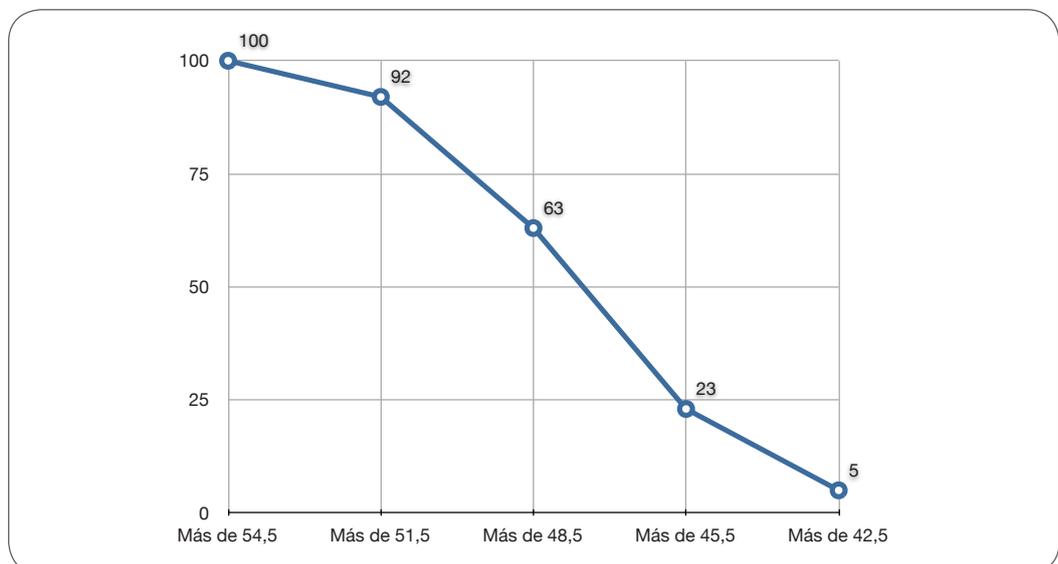
Ingresos	Nº trabajadores
40 a 42	5
43 a 45	18
46 a 48	40
49 a 51	29
52 a 54	8
<b>Total</b>	<b>100</b>

Distribución de frecuencias acumulada

Ingresos	Nº trabajadores
Menos de 39,5	0
Menos de 42,5	5
Menos de 45,5	23
Menos de 48,5	63
Menos de 51,5	92
Menos de 54,5	100



Se puede representar una distribución de frecuencias acumulada en dos sentidos. Teniendo en cuenta que los valores son “menos de”, como el ejemplo que acabamos de ver, o la misma distribución con el sentido cambiado “ más que” cuya representación será a la inversa.



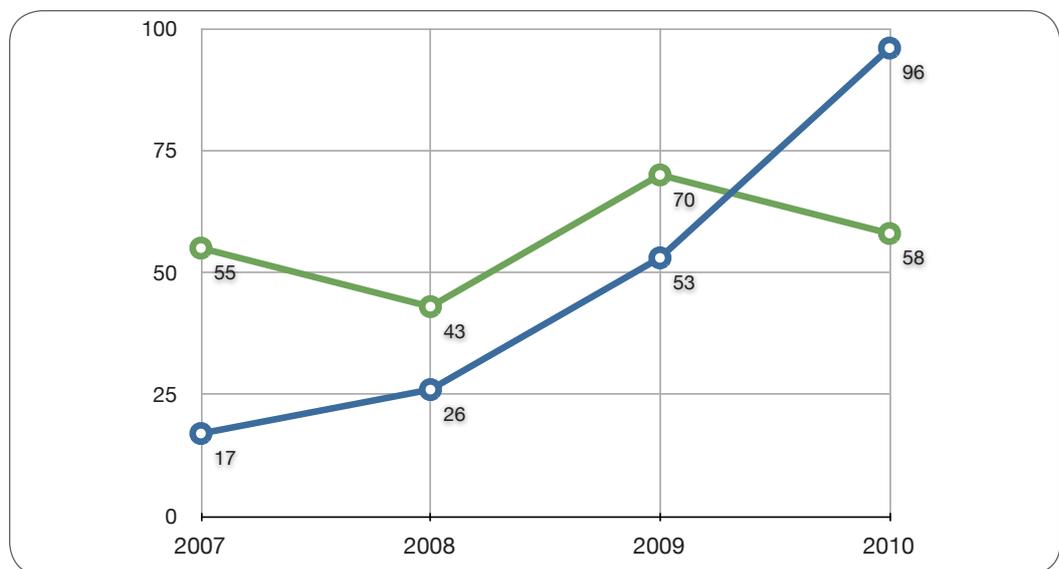
#### 1.4.4. Línea de Grafos.

Esta representación, muestra el valor de una variable dependiente (representada en el eje Y) para cada valor de las diferentes categorías de otra variable, normalmente utilizada como variable independiente (que se representa a lo largo del eje X).



La línea no se cierra sobre el eje X, como se hace en la representación de los polígonos, por lo que el área que queda por debajo de la línea de grafos no tiene ningún significado especial, como ocurre con los histogramas y polígonos.

En este tipo de gráficos, lo que interesa, es más bien la forma que adopta, la pendiente que toma al crecer o decrecer la línea y, en el caso de representar sobre el mismo cuadrante dos o más líneas, las semejanzas y diferencias que presentan.

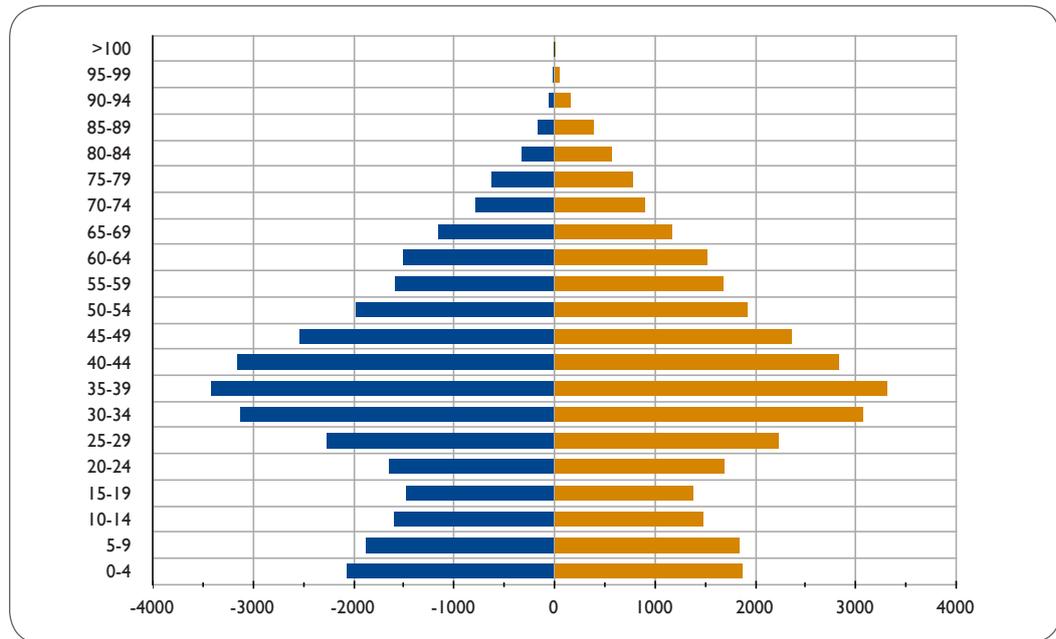


#### 1.4.5. Otras representaciones gráficas.

##### 1.4.5.1. Pirámide de población o estructura demográfica.

La estructura demográfica de una población es su distribución por edad y sexo. Esta distribución suele representarse en un gráfico de barras horizontales (histograma) denominado pirámide de población, en la que quedan reflejados las proporciones respecto al total de la población, de hombres (a la izquierda del gráfico) y mujeres (a la derecha del gráfico), y los diferentes grupos de edad (de cinco en cinco años), representados en forma barras.

La distribución por sexos es desigual en la base y en la cumbre. Mientras que en la base existen más hombres que mujeres (estadísticamente se ha comprobado que nacen 105 hombres por cada 100 mujeres), la esperanza de vida de las mujeres es más alta, por lo que en los tramos intermedios se equilibran y en los tramos altos superan el número de mujeres al de hombres.



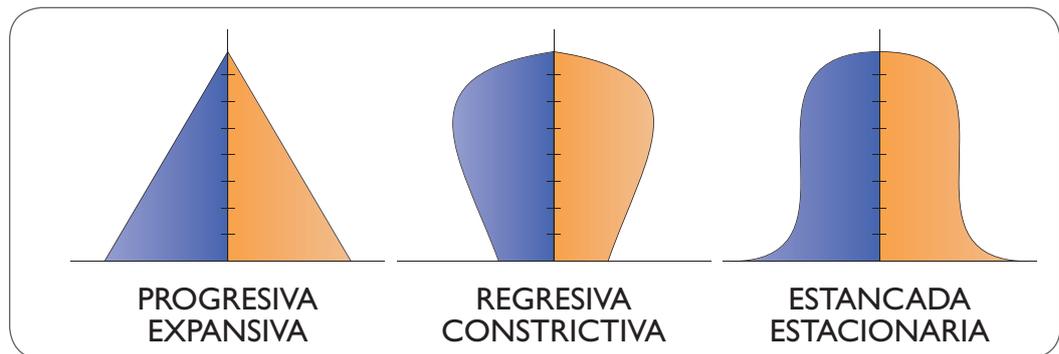
En el **perfil** de la pirámide se refleja la historia demográfica reciente de la población estudiada, así las muescas o incisos nos indica pérdidas de población en ese estrato, que se pueden deber a efectos de las guerras (siempre será mayor en la parte masculina), emigraciones, epidemias, etc. También pueden aparecer estratos más alargados de lo que hubiera sido previsible, lo que nos indicará que ese lugar ha sido destino de migraciones.

La **forma** de la pirámide viene marcada por la distribución por edad y nos indicará el grado de desarrollo del lugar cuya población representamos. Se suelen clasificar tres grandes grupos de edad: niños y jóvenes (hasta los catorce años), adultos (entre los quince y sesenta y cuatro años) y los ancianos, (mayores de 65 años). La mayor o menor proporción de cada grupo dibujará tres formas básicas de pirámides.

1. **Pirámide progresiva** (expansiva), que tiene forma de pagoda, debido a que tiene un gran contingente de población joven en la base, que va desapareciendo rápidamente según avanzan los grupos de edad, en cuya cumbre quedan muy pocos efectivos. Son por tanto las típicas de países subdesarrollados, que presentan unas altas tasas de natalidad, y una esperanza de vida muy baja debido a la alta mortalidad. Son poblaciones muy jóvenes, por lo que presentan un alto crecimiento.
2. **Pirámide regresiva** (constrictiva), tiene forma de bulbo, debido a que en la base existe menos población que en los tramos intermedios, mientras que en la cumbre existe un número importante de efectivos. Son las típicas de los países desarrollados, en los que la natalidad ha descendido rápidamente, y sin embargo las tasas de mortalidad llevan mucho tiempo controladas, siendo la esperanza de vida cada vez mayor. Son poblaciones muy envejecidas, en las que no se garantiza el relevo generacional.

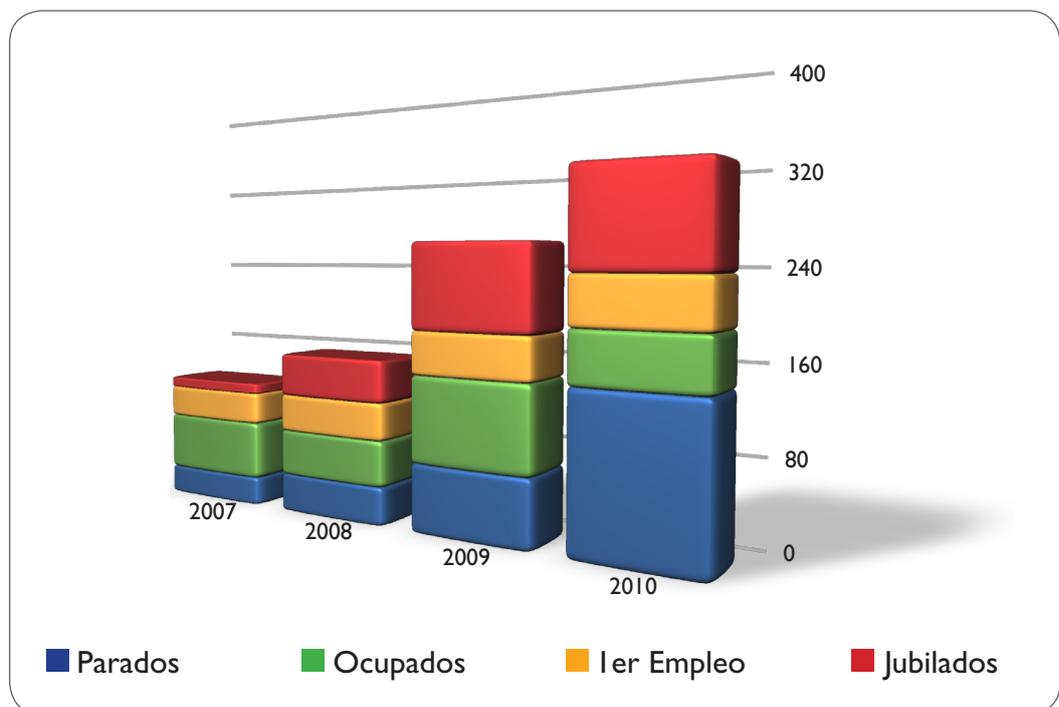


3. **Pirámide Estancada** (estacionaria), o en forma de campana, debido a que los tramos intermedios de edades tienen los mismos efectivos que la base, existiendo una reducción importante hacia la cumbre. Son las típicas de los países en vías de desarrollo, en los que se ha controlado la mortalidad y se ven los primeros indicios de control de la natalidad reciente. Se puede considerar como el paso intermedio desde una pirámide progresiva a otra regresiva.



#### 1.4.5.2. Gráfico Rectangular. Histograma Acumulado.

Es una variación del histograma, utilizado bien para representar variables nominales o bien para destacar categorías específicas de variables. En el ejemplo se representa la estructura ocupacional en España en distintos años, es una simple representación de la variable nominal "ocupación" a lo largo de varios años.

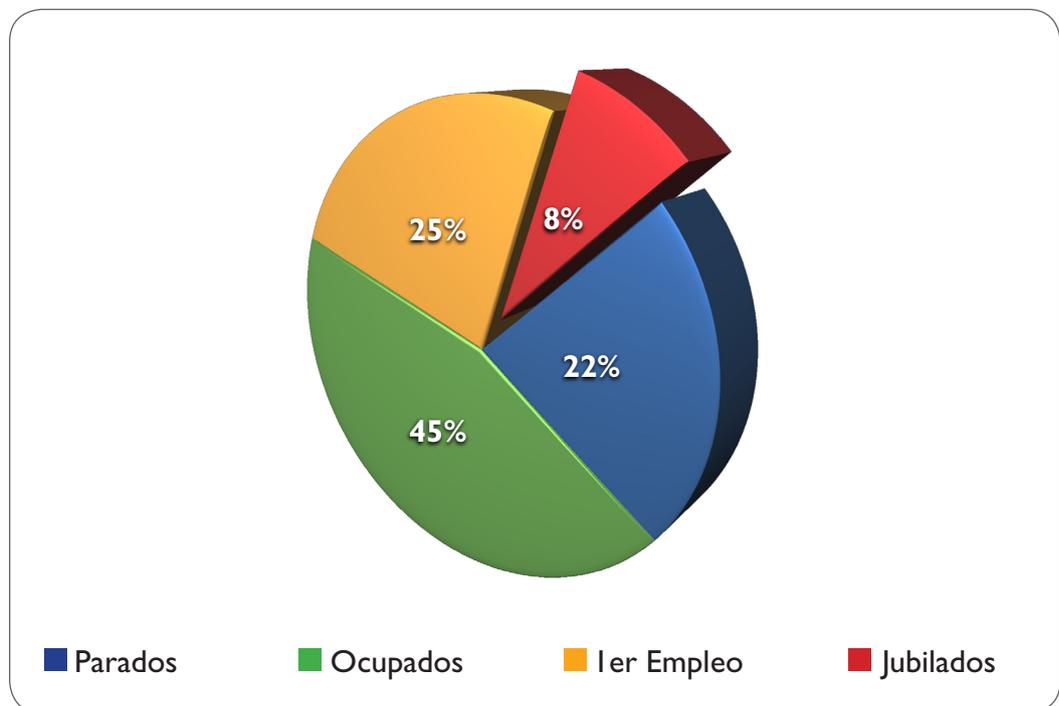




### 1.4.5.3. Gráfico de Sectores o Diagrama de Tarta.

Es un gráfico en donde un círculo representa el total de una característica, en este caso y como en el ejemplo anterior; la estructura ocupacional en España en la suma de los cuatro años examinados. Las áreas representan los porcentajes totales de los cuatro años de cada una de las categorías de la estructura ocupacional. El círculo se divide en sectores cuyo ángulo refleja el porcentaje del total para cada categoría.

Existen muchas más representaciones gráficas pero muchas de ellas no son más que combinaciones o versiones diferentes de las que ya hemos vistos en estos ejemplos, si bien es verdad, éstas son las más representativas.



## BIBLIOGRAFÍA

- DE MIGUEL, Amado: *Manual de estructura social en España*, Madrid, Tecnos, 1974.
- DE MIGUEL, Amado: *La pirámide social española*, Barcelona, Ariel, 1977.
- GARCÍA FERRANDO, M.: *Socioestadística. Introducción a la estadística en sociología*, Madrid, Alianza Editorial, 1989.
- GONZÁLEZ BLASCO, Pedro: *El investigador científico en España*, Madrid, C.I.S., 1980.
- SPIEGEL, Murray E.: *Estadística*, Madrid, Alianza Ediciones Colina, S.A., 1975.