descriptiva incluye dos operaciones fundamentales. La primera de ellas se refiere a la organización y ordenación de los datos o medidas obtenidos en algún tipo de distribución, mientras que la segunda de dichas operaciones se refiere al tratamiento aritmético de dichos datos, bien sea por medio de la resta o sustracción o bien por medio de la división. Tal como destacan Loether y McTavish (1974, 43), la idea de la división, esto es, la creación de una relación entre un número (numerador) y otro número (denominador), es uno de los temas organizadores básicos tanto de la estadística descriptiva como de la estadística inferencial. Desde el punto de vista de la relevancia que tiene para la investigación sociológica la creación de tales relaciones, el problema consiste en saber qué es lo que hay que dividir entre qué, y la respuesta, normalmente, vendrá dada por el esquema teórico en el que se enmarque la investigación.

A continuación vamos a presentar un breve panorama de las operaciones básicas de comparación, utilizando para ello ejemplos prácticos de carácter sociológico.

2.3.1. *La organización de los datos*

Una lista de datos que no esté organizada según un criterio determinado suele ser de poca utilidad para el investigador interesado en realizar algún tipo de comparación. Una vez se hayan obtenido los datos que estimamos relevantes para realizar el análisis deseado es conveniente ordenarlos según algún criterio, bien sea de mayor a menor o de otra forma, con el fin de que se pueda obtener el máximo de información posible de los datos. La ordenación permitirá observar con mayor facilidad la distribución de los datos y el lugar dónde termina un grupo y comienza otro en relación a otros grupos.

Supongamos que estamos estudiando la población extranjera de origen europeo residente en España. La primera información que necesitamos reunir es la referente al número y origen de esta población extranjera. El *Anuario Estadístico* del Instituto Nacional de Estadística ofrece los datos que se recogen en la tabla 1 sobre la nacionalidad de origen y el número de extranjeros que han residido en España en 1979.

Los países se presentan ordenados en el *Anuario* por orden alfabético y así los hemos transcrito en la tabla 1. Una ordenación de este tipo puede que no resulte la más interesante para ofrecer de una forma relevante la información. Cabría pensar en realizar otra ordenación de los países según el número de personas que tienen residiendo en España. Así, tendríamos encabezando la lista a Portugal, con 21.801 personas, seguido de Alemania, con 18.144, y en el otro extremo estarían los países con menor número de residentes, que son la URSS, con 22, y Rumanía, con 26.

TABLA 1

Extranjeros europeos residentes en España, según nacionalidades (1979)

Nacionalidad	*Número*	*Nacionalidad*	*Número*
Alemania	18.144	Italia	9.192
Austria	1.145	Noruega	749
Bélgica	3.764	Países Bajos	4.784
Dinamarca	2.009	Polonia	88
Finlandia	892	Portugal	21.801
Francia	14.891	Rumanía	26
Gran Bretaña	17.330	Suecia	3.229
Grecia	425	Suiza	3.576
Hungría	36	U.R.S.S.	22
Irlanda	376	Yugoslavia	33

FUENTE: *Anuario Estadístico de España*, Madrid, I.N.E., 1979.

Ahora bien, teniendo en cuenta que el número de extranjeros que residen en un país dependerá, entre otros factores, del tipo de relaciones que guarden los países entre sí, es decir, de su proximidad política, cultural y económica, aparte de su proximidad geográfica, cabe realizar una ordenación de los países en grupos regionales, tal como se presenta en la tabla 2.

Con la nueva agrupación se obtiene, a primera vista, una ordenación más significativa de los datos. Así, se observa que el grupo más amplio de extranjeros, con 70.889 personas, proviene de los países de la Europa occidental, con los que España mantiene unos estrechos contactos de todo tipo. Este elevado grupo de europeos occidentales contrasta con el pequeño grupo de extranjeros que provienen de los países europeos socialistas, sólo 205, con los que España mantiene unas relaciones mucho más escasas y distanciadas. El resto de los países lo hemos distribuido entre países mediterráneos, que incluye a Grecia e Italia, con 9.617 residentes en España —la mayoría de ellos italianos—, y Portugal, que, como país vecino, lo hemos mantenido en una categoría aparte, además de tener el máximo número de residentes extranjeros en España.

La agrupación realizada, pues, ha permitido realizar una comparación con la que se pueden analizar de forma más relevante los datos originales. El marco teórico en el que se inscriba el análisis cuantitativo debe ser, en toda investigación empírica, el criterio básico que se ha de seguir para agrupar los datos y poder realizar una comparación significativa.

TABLA 2

Extranjeros europeos residentes en España, según áreas regionales (1979)

Resto Europa Occidental		*Península Ibérica*		*Países mediterráneos*		*Países socialistas*	
Alemania	18.144	Portugal	21.801	Grecia	425	Polonia	88
Inglaterra	17.330			Italia	9.192	Hungría	36
Francia	14.891	TOTAL	21.801			Yugoslavia	33
Holanda	4.784			TOTAL	9.617	Rumania	26
Bélgica	3.764					URSS	22
Suiza	3.576						
Suecia	3.229					TOTAL	205
Dinamarca	2.009						
Austria	1.145						
Finlandia	892						
Noruega	749						
Irlanda	376						
TOTAL	70.889						

FUENTE: I.N.E., *op. cit.*, elaboración propia.

2.3.2. *Distribuciones*

Con el fin de obtener una organización más resumida y operativa de los datos, se utilizan tres tipos de distribuciones: *a)* la distribución de frecuencias; *b)* la distribución porcentual; *c)* la distribución acumulada.

2.3.2.1. *Distribución de frecuencias*

Cuando se está manejando un número amplio de datos, resulta conveniente distribuirlos en *clases* o *categorías* y determinar el número de casos que pertenece a cada clase. Este número se denomina *frecuencia de clase*, y se simboliza por medio de la letra f o f_i, en donde i se refiere a la clase i de la variable ordenada. El número total de casos es igual, por tanto, a la suma de la columna de las frecuencias, y se simboliza por la letra N, o bien como Σf_i, en donde Σ, que es la letra griega sigma, simboliza la suma de todas las frecuencias de clase.

El número de clases o categorías que se seleccionan vendrá determinado por las necesidades de la investigación. Supongamos que tenemos un grupo de 120 individuos adultos mayores de dieciocho años y menores de setenta y cinco años y queremos distribuirlos según su edad. Una distribución útil puede ser la siguiente:

Edad (años)	f_i
De 18 a 20	10
De 21 a 25	14
De 26 a 35	23
De 36 a 45	20
De 46 a 60	29
De 61 a 75	24
	N = 120

La primera clase o categoría de edad es «de 18 a 20 años», y a ella pertenecen 10 individuos, es decir, que la frecuencia de esta clase o categoría es 10.

Los datos, tal como han sido ordenados y resumidos en la distribución de frecuencia anterior, se suelen denominar *datos agrupados*. Aunque con el proceso de agrupamiento se pierde algo de la información que contienen los datos originales —por ejemplo, en la categoría «18 a 20» no sabemos cuántos individuos tienen dieciocho, diecinueve o veinte años—, sin embargo, ofrece la gran ventaja de presentar todos los datos de una forma sencilla en un pequeño cuadro, lo que facilita, evidentemente, su estudio.

Continuando con la terminología que se utiliza en la distribución de frecuencia, denominamos *intervalo de clase o categoría* al símbolo que define una clase o categoría; por ejemplo, la clase «de 21 a 25» la simbolizamos como 21-25. Los números extremos de cada clase o categoría, en este caso 21 y 25, se denominan *límites de clase*, siendo el mayor de ellos el *límite superior* y el menor el *límite inferior*. Los términos clase o categoría e intervalo de clase o categoría, que, al menos teóricamente, no tienen límite superior e inferior, se conocen como *intervalo de clase o categoría abierto*. Así, podemos escribir la anterior distribución de frecuencias dejando abierta la categoría «menos de 21» y «más de 60»:

Edad (años)	*f*
Menos de 21	10
De 21 a 25	14
De 26 a 35	23
De 36 a 45	20
De 46 a 60	29
Más de 60	24
	N = 120

Si las edades se registran con una aproximación de meses, el intervalo de la categoría 21-25 incluye, teóricamente, todos los individuos con edades que van desde 20,5 a 25,5 años. Estos números se conocen con la denominación de *límites reales o verdaderos de clase o categoría*, siendo el menor de ellos el *límite real inferior* y el mayor de ellos el *límite real superior*. En la práctica, los límites reales de clase o categoría se obtienen sumando al límite superior de un intervalo de clase o categoría el límite inferior del intervalo contiguo superior y dividiendo a continuación por dos.

Los límites reales se pueden utilizar igualmente para simbolizar las clases o categorías. Así, las diversas categorías del ejemplo anterior podrían indicarse por 17,5-20,5, 20,5-25,5, 25,5-35,5, etc. No obstante, esto se hace raramente, ya que con dicha simbolización se introduce un elemento perturbador por su ambigüedad, ya que los límites reales no coincidirán siempre con las observaciones reales. Por ejemplo, para la edad 25,5 no se puede saber si pertenece al intervalo de la categoría 20,5-25,5 o a la 25,5-35,5. Por esta razón resulta aconsejable utilizar intervalos cuyos límites sean mutuamente excluyentes para las diversas clases o categorías.

El tamaño o amplitud de la clase o categoría es la diferencia entre los límites reales que forman cada clase o categoría, y se conoce como *amplitud, tamaño o longitud* de clase o categoría, según los autores. El tamaño de cada categoría puede ser idéntico o diferente. En la distribución de frecuencia de edades utilizada anteriormente aparecen cuatro tamaños de categoría diferentes, una de tres, otra de cinco, dos de diez y otras dos de quince años, como se observa a continuación: $20{,}5 - 17{,}5 = 3$; $25{,}5 - 20{,}5 = 5$; $35{,}5 - 25{,}5 = 10$; $45{,}5 - 35{,}5 = 10$; $60{,}5 - 45{,}5 = 15$; $75{,}5 - 60{,}5 = 15$.

El punto medio del intervalo de clase o categoría se obtiene sumando los límites inferior y superior de la clase o categoría y dividiendo a continuación por dos. También se denomina *punto medio de la clase o categoría*, y se simboliza por X_i. Así, el punto medio del intervalo 21-25 es $(21+25)/2 = 23$. En los cálculos estadísticos ulteriores, las observaciones pertenecientes a un intervalo de categoría dado se supone que son coincidentes con el punto medio de la categoría. Así, todas las edades del intervalo de la categoría 21-25 se considerarán como de edad de veintitrés años. La anterior distribución de frecuencias según categorías de edad se puede escribir del siguiente modo, incluyendo límites reales y puntos medios:

Edad (años)	f_i	*Punto medio* X_i	*Límites reales*
De 18 a 20	10	19	17,5 a 20,5
De 21 a 25	14	23	20,5 a 25,5
De 26 a 35	23	30,5	25,5 a 35,5
De 36 a 45	20	40,5	35,5 a 45,5
De 46 a 60	29	53	45,5 a 60,5
De 61 a 75	24	68	60,5 a 75,5
	N = 120		

Tal como se ha señalado anteriormente, el agrupamiento de datos no sólo reporta ventajas, tales como las de resumir y permitir un manejo más fácil de la información, como también presenta algún inconveniente, siendo el principal lo que se denomina *error de agrupamiento*. Con este término nos referimos a las alteraciones que se producen al realizar determinados agrupamientos, lo que conduce a la variación de *N*. Veamos a través de un ejemplo la aparición de este tipo de error:

a_1)

X_i	f_i	fX_i
1	5	5
2	2	4
3	1	3
4	2	8
5	0	0
6	3	18
	N=13	$38=fX_i$

a_2)

Clase	X_i	f_i	fX_i
1 a 2	1,5	7	10,5
3 a 4	3,5	3	10,5
5 a 6	5,5	3	16,5
		N=17	$37{,}5=fX_i$

Hemos partido de 13 puntuaciones correspondientes a una distribución de 6 categorías cuyo tamaño es la unidad. En la tercera columna (fX_i) aparecen los números de casos totales dentro de cada categoría. La suma de estos totales parciales, fX_i, es igual a 38. Pero si ahora agrupamos los mismos datos en tres categorías cuya anchura sea 2, en lugar de 1, tal como aparece en el apartado a_2) del cuadro, la columna de frecuencias totaliza el mismo número que en el caso anterior, pero no ocurre así con la columna del total de casos en cada categoría, cuya suma ya no es 38, sino 37,5. La diferencia entre ambos números se debe a que hemos calculado los totales parciales fX_i utilizando el valor medio de cada categoría. Precisamente la diferencia entre 38 y 37,5 es lo que se llama error de agrupamiento, y se produce porque los puntos medios de cada clase o categoría en el ejemplo no representan convenientemente el valor de los casos que se engloban en cada categoría.

Por tanto, al agrupar los datos, las categorías se calcularán con sumo cuidado, de forma que los valores medios de cada una de ellas refleje de la forma más exacta posible el valor de los casos en la categoría. Spiegel (1975, pág. 28) ofrece las dos siguientes reglas para formar las distribuciones de frecuencias y minimizar el error de agrupamiento: 1) determinar el mayor y el menor entre los datos registrados 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 clase de idéntico tamaño. Si ello no fuera posible, será preciso utilizar intervalos de clase de diferente tamaño e intervalos abiertos. El número de intervalos se pone generalmente entre 5 y 20, dependiendo de los datos de partida. Los intervalos se elegirán de forma que los puntos medios coincidan con datos realmente observados.

2.3.2.2. *Distribución porcentual*

Para calcular un porcentaje es preciso calcular previamente una proporción. La *proporción* de casos en una categoría dada es igual al número de casos en la categoría dividido por el número total de casos en la distribución. En una distribución de frecuencias de cinco categorías, en la que el número de casos en cada categoría fuese N_i y el número total de casos fuese N, la proporción de casos en cada categoría será N_i/N. Obviamente, el valor de una proporción no puede ser mayor que la unidad. Dado que

$$N_1+N_2+N_3+N_4+N_5=N$$

se tiene:

$$\frac{N_1}{N}+\frac{N_2}{N}+\frac{N_3}{N}+\frac{N_4}{N}+\frac{N_5}{N}=\frac{N}{N}=1$$

Por tanto, si se suman las proporciones de casos en todas las categorías, el resultado será la unidad. Se trata de una propiedad fundamental de las proporciones, y se puede generalizar a cualquier número de categorías.

Los *porcentajes* se obtienen a partir de las proporciones simplemente multiplicando por 100; de ahí que también se denominen *por ciento*. Al utilizar porcentajes, lo que se hace realmente es estandarizar según el tamaño, ya que se calcula el número de casos que habría 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 las proporciones de una distribución dada es igual a la unidad, la suma de sus porcentajes será 100.

Si en lugar de los valores absolutos en una distribución de frecuencias se utilizan los correspondientes porcentajes, tendremos una *distribución porcentual*, que presenta algunas ventajas sobre la primera. Sobre todo, facilita la comparación, aparte de evitar una fuente importante de error. El porcentaje, que es en realidad una razón simple, se entiende fácilmente porque, tal como señalan Loether y McTavish (1974, pág. 54), tendemos en nuestra cultura a pensar en términos de partes de 100.

La distribución por edades anterior se puede escribir en términos de porcentajes del siguiente modo:

Edad (años)	f_i	%
De 18 a 20	10	8,33
De 21 a 25	14	11,66
De 26 a 35	23	19,16
De 36 a 45	20	16,66
De 46 a 60	29	24,16
De 61 a 75	24	20,00
	N=120	99,97

Para calcular el porcentaje de cada categoría se ha dividido cada f_i por N y se ha multiplicado por 100. Obsérvese que la suma de los porcentajes no es exactamente 100,0, debido a que sólo hemos tomado dos cifras decimales y no hemos redondeado el porcentaje resultante. Es aconsejable utilizar una sola cifra decimal, redondeándola de forma que si el número de la centésima es menor de 5 se mantiene el valor de la décima, pero si el número de la centésima es 5 o superior a 5 se incrementa en una unidad la cifra de las décimas. Realizando esta operación de redondeamiento, la anterior distribución porcentual quedaría del siguiente modo:

Edad (años)	%
De 18 a 20	8,3
De 21 a 25	11,7
De 26 a 35	19,2
De 36 a 45	16,7
De 46 a 60	24,2
De 61 a 75	20,0
TOTAL	100,1
	(120)

Ahora, la suma porcentual es 100,1, es decir, una décima superior a 100, por efecto de la operación de redondeamiento. Obsérvese también que el número que representa los casos totales N se ha puesto, entre paréntesis, debajo del 100. Esta práctica es habitual en la presentación de las tablas de distribuciones porcentuales, porque de este modo se indica la base real sobre la que se ha calculado el porcentaje.

Resulta conveniente señalar que, para calcular porcentajes, el valor de N ha de ser suficientemente elevado. Blalock (1960, pág. 28) señala el número 50 como el mínimo aproximado de casos que ha de contar una distribución para poder calcular los porcentajes. Si el número de casos es bastante inferior a 50, resulta más adecuado ofrecer el número real de casos en cada categoría en lugar de los porcentajes.

No siempre puede estar indicado la utilización de porcentajes para realizar comparaciones significativas y, en tal caso, convendrá operar con las cifras absolutas. Zeisel, que ha escrito quizá los capítulos más didácticos en el campo de la metodología de las ciencias sociales sobre el uso de los porcentajes, utiliza el siguiente ejemplo. Supongamos que queremos comparar dos empresas en términos de la variación anual de sus ventas. Supongamos que la empresa *A* ha aumentado en el último año su volumen de ventas de 1 a 2 millones de pesetas, lo que significa un aumento del 100 por 100; mientras que la empresa *B* ha pasado en sus ventas de 4 a 7 millones de pesetas, lo que significa un aumento del 75 por 100. Si comparamos las empresas *A* y *B* según sus cifras absolutas, *B* aventaja claramente a *A*, ya que sus ventas experimentaron un incremento de 3 millones de pesetas, mientras que la segunda experimentó una subida de sólo 1 millón. Sin embargo, si comparamos las dos empresas según sus incrementos porcentuales o relativos, la empresa *A*, con el 100 por 100, claramente supera a la empresa *B*, que sólo aumentó el 75 por 100. Para Zeisel, en caso de duda sobre la forma en que deben realizarse las comparaciones, «la consideración más general es presentar el aumento de forma que determine tan exactamente como sea posible el concepto que deseamos medir» (Zeisel, 1962, pág. 27).

Tampoco recomienda Zeisel el uso de porcentajes que excedan considerablemente de 100. Decir, por ejemplo, que los visitantes extranjeros en España aumentaron en la década de los sesenta un 1.200 por 100 sobre el número de visitantes en la década de los cincuenta puede producir una cifra impresionante, pero estadísticamente es un recurso muy pobre; resulta más correcto decir que el número de visitantes aumentó 12 veces en relación al período anterior.

2.3.2.3. *Distribución acumulada*

Una distribución acumulada se forma al indicar para cada categoría el número (o porcentaje) de casos que quedan por debajo del límite real superior de dicha categoría. Normalmente, se sigue la convención de crear distribuciones acumuladas, comenzando a acumular desde las categorías 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. Para el caso de la distribución por edad que venimos utilizando, las dos distribuciones acumuladas quedarían del siguiente modo:

Edad (años)	*Frecuencia*	*Porcentaje*	*Frecuencia acumulada*	*Porcentaje acumulado*
De 18 a 20	10	8,3	10	8,3
De 21 a 25	14	11,7	24	20,0
De 26 a 35	23	19,2	47	39,2
De 36 a 45	20	16,7	67	55,9
De 46 a 60	29	24,2	96	80,1
De 61 a 75	24	20,0	120	100,1
	N 120			

Así, para la categoría de 36 a 45 años, la frecuencia acumulada de 67, o el porcentaje de acumulado de 55,9 por 100, indican que el número o porcentaje de individuos con esa edad o menos es el que se indica.

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. Así, por ejemplo, al comparar los niveles de ingresos familiares en hogares españoles cuyo cabeza de familia pertenece a la clase social alta y media alta, o a la clase obrera, se obtienen los siguientes resultados:

TABLA 3

Distribuciones porcentuales acumuladas de los ingresos familiares, por clase social

Cantidad de pesetas mensuales	*Clase social alta y media-alta* (1)	*Clase social obrera* (2)	*Frecuencia acumulada* (1)	*Frecuencia acumulada* (2)
Más de 50.000	4	—	4	—
De 30.501 a 50.000	7	1	11	1
De 20.501 a 30.500	22	1	33	2
De 14.501 a 20.500	19	4	52	6
De 12.501 a 14.500	8	5	60	11
De 10.501 a 12.500	9	8	69	19
Menos de 10.000	31	81	100	100
	100	100		
	(279)	(1.126)		

FUENTE: FOESSA, 1970, pág. 563. Elaboración propia.

Las distribuciones acumuladas permiten una comparación más clara de las tremendas diferencias que, en materia de ingresos familiares, existían en los hogares españoles en el momento de realizar el estudio (finales de los años sesenta). Mientras que en los hogares cuyo cabeza de familia se identificaba con las clases sociales más altas el 69 por 100 disfrutaba de unos ingresos superiores a 10.500 pesetas, tal porcentaje era tan sólo del 19 por 100 en los hogares de familias obreras. De esta forma, vemos cómo los porcentajes acumulados permiten en una sola medida ofrecer los casos que se encuentran por debajo o por encima de unos niveles determinados.

2.3.3. *Percentiles*

El valor por debajo del cual queda un porcentaje determinado de casos es un *percentil*, y podemos representarlo por P_i, siendo i un valor que oscila entre 1 y 100. Así, 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, P_{50}, se llama también *mediana*. Por extensión, se puede hablar de aquellos valores que dividen a los datos en cuatro partes iguales. Estos valores, que podemos representar por Q_1, Q_2 y Q_3, se llaman primero, segundo y tercer *cuartil*, respectivamente; el valor de Q_2 es el valor que divide a los datos en dos partes iguales, y que se denomina mediana. De igual modo, los valores que dividen los datos en diez partes iguales se denominan *deciles*, y podemos representarlos por D_1, D_2, ..., D_9. Los resultados de muchas evaluaciones (tests) se presentan en forma de percentiles —el porcentaje de individuos que, en un determinado test, ha obtenido una puntuación igual o superior a un valor concreto.

Dos problemas de cálculo se presentan en relación a los percentiles. Cuando se desea calcular el rango de percentil de una puntuación determinada, hay que utilizar la siguiente fórmula:

$$\text{Rango de percentil de una puntuación dada} = \frac{\text{Lugar que ocupa la puntuación en la distribución}}{N} \times 100$$

En la distribución siguiente:

3
5
9
11
15
17
22

el rango de percentil del valor «11» será $= \dfrac{4}{7} \times 100 = 57{,}1$, ya que el valor 11 ocupa el cuarto lugar y $N=7$.

De manera inversa, se puede calcular el valor o puntuación correspondiente a un rango de percentil dado. Para ello se multiplicará el percentil por N y, a continuación, se buscará en la distribución el lugar que corresponde al número así calculado. Por ejemplo, en la distribución anterior, al percentil 70 le corresponde la puntuación 15, que ocupa el quinto lugar, ya que 5 es lo que resulta de redondear el número 4,9, que se obtiene al multiplicar 0,70 por N, que en este caso es 7.

El uso de percentiles resulta muy apropiado cuando se desea comparar, dadas una serie de distribuciones, unos grupos específicos, situados en un lugar dado de las distribuciones, con otros grupos situados en el mismo o diferente lugar. Murillo Ferrol hace un buen uso de esta lógica de la comparación cuando, al estudiar la distribución de las rentas en Andalucía, señala y denuncia que el incremento absoluto del volumen de las rentas no ha venido acompañado de un proceso de mejora en la redistribución de tales rentas entre todas las clases de población, ya que los pobres, mayoritarios, continúan percibiendo una proporción pequeña de las rentas, mientras que los ricos, que son muy pocos, reciben la parte más amplia de los ingresos. En la tabla 4 hemos reproducido las comparaciones porcentuales que realiza Murillo Ferrol, en base a la proporción de ingresos que corresponde al 20 por 100 más pobre de los hogares, al 20 por 100 más favorable y al 5 por 100 último de los más favorecidos. Esta utilización de los percentiles sirve mejor que otro algoritmo para evidenciar el desequilibrio existente en la distribución de las rentas. Así, y observando con más detenimiento los resultados de la tabla 4, se aprecia que, para el conjunto nacional, el 20 por 100 más pobre de la población recibe tan sólo el 6,8 por 100 de los ingresos totales, frente al 45,2 por 100 de ingresos que recibe el 20 por 100 más favorecido, o el 19,4 por 100 de ingresos que recibe el 5 por 100 más favorecido. Para las ocho provincias andaluzas, los resultados comparativos son parecidos a los de la media nacional, lo que revela un sistema de distribución económico muy injusto.

TABLA 4

Ingresos que corresponden a determinados grupos de la población

Provincias	*% de ingresos que corresponde al 20 % más pobre de los hogares*	*% de ingresos que corresponde al 20 % más favorecido*	*% de ingresos que corresponde al 5 % último de los más favorecidos*
Almería	8,0	43,3	19,5
Cádiz	8,2	41,9	15,3
Córdoba	8,1	43,8	20,8
Granada	7,8	47,5	20,8
Huelva	8,5	42,0	17,1
Jaén	7,7	46,2	24,0
Málaga	8,5	38,0	15,1
Sevilla	7,5	46,4	20,7
España	6,8	45,2	19,4

FUENTE: MURILLO FERROL, F., «La distribución de la renta en Andalucía», *Anales de Sociología*, 4, 1968, pág. 40.

2.3.4. *Razón*

La *razón* * de un número *A* a otro número *B* se define como *A* dividido por *B*. La cantidad que precede a la palabra clave «a» se coloca en

* Algunos tratadistas utilizan la palabra inglesa *ratio* para referirse al término razón, aunque resulta conveniente emplear este último.

el numerador, mientras que el número que le sigue va al denominador. Así, si en un parlamento hay 160 diputados de izquierdas, 150 diputados de derechas y 80 diputados regionalistas, la razón de los diputados de izquierda a los diputados de derecha será 160/150, mientras que la razón de los diputados de izquierda y regionalistas a los diputados de derecha será (160+80)/150. Se puede, pues, utilizar un número compuesto tanto en el numerador como en el denominador, aunque el resultado se suele expresar de la forma numérica más simple posible, en este caso como 24/15.

Las proporciones y los porcentajes son un tipo especial de razón, en donde el denominador es 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. Pero, a diferencia de la proporción, la razón puede ser mayor que la unidad, como ocurre en el ejemplo de los diputados empleado anteriormente.

Las razones se suelen expresar también en términos de cualquier base que resulte conveniente para nuestros objetivos descriptivos. Una razón muy empleada en demografía es la de sexos, que se define como el número de varones de una población determinada dividido por el número de mujeres. Dado que el número total de varones es menor que el número de mujeres (aunque nacen más niños que niñas, la tasa de mortalidad masculina es mayor que la tasa de mortalidad femenina, por lo que entre la población adulta es mayor el número de mujeres que el de varones), la razón de los sexos será un número decimal, por lo que convencionalmente se suele multiplicar por 100, con lo que una razón de sexos de 94 indicará que hay 94 varones por cada 100 mujeres.

Cuando se utilizan bases mayores que 100, tales como 1.000, 10.000 o un millón, tenemos las *tasas*, que son otro tipo de razón. Las tasas se emplean cuando el uso de porcentajes arroja números decimales. Las tasas también se utilizan abundantemente en demografía y, en general, cuando se quiere disponer de indicadores sencillos referentes a la población general. Así, una tasa bruta de natalidad de 30 por 1.000 significa que se han producido 30 nacimientos por cada 1.000 habitantes.

Las *tasas de crecimiento relativo* son otro tipo muy utilizado de razón. Para calcular la tasa de crecimiento en un período de tiempo dado, se toma el incremento real durante el período y se divide por el tamaño que había al comienzo del período. Así, si la renta *per capita* en un país determinado ha pasado, en el período 1960-1970, de 1.500 a 2.000 dólares, la tasa de crecimiento relativo de la renta *per capita* será:

$$\frac{2.000-1.500}{1.500}=\frac{500}{1.500}=0{,}33$$

o, si se quiere expresar en términos porcentuales, del 33 por 100. Naturalmente, cuando la cantidad al final del período sea más reducida que al comienzo, la tasa resultante será negativa por reflejar un decrecimiento. También se pueden expresar las tasas de crecimiento relativo en relación a 1.000, 10.000 u otra cantidad que resulte conveniente con fines descriptivos y analíticos. En general, la tasa de crecimiento relativo se puede expresar como $(b - a/a) \cdot k$, siendo a y b las cantidades al principio y al final del período, respectivamente, y k la base que se decida utilizar, y que normalmente será una constante con ceros, del tipo de 1.000, 10.000, un millón, etc.

2.4. TÉCNICAS BÁSICAS DE REPRESENTACIÓN GRÁFICA

Los resultados de las investigaciones estadísticas se suelen representar muchas veces gráficamente, con el fin de obtener un panorama más intuitivo y directo de los mismos. Aunque son muchos los recursos gráficos que los sociólogos utilizan para ofrecer una visión directa y simple de sus investigaciones cuantitativas, aquí vamos a referirnos, para comenzar, a las representaciones gráficas utilizadas en estadística para el caso de las distribuciones de frecuencia. Los *histogramas*, los *polígonos* y las *ojivas* son tales representaciones, y junto con la *línea de grafos*, constituyen los procedimientos gráficos básicos más utilizados en el campo de la estadística.

En toda representación gráfica se encuentra subyacente la idea de un *sistema de referencias* o *sistema de coordenadas*. El sistema de coordenadas más usual en las representaciones gráficas consiste en dos líneas, o «dimensiones», perpendiculares que forman el sistema de Coordenadas Cartesianas —en honor del filósofo René Déscartes (1596-1650), que fue el primero en combinar el álgebra con el análisis gráfico—. Como es sabido, la línea o eje vertical se llama *ordenada* o eje de las Y, y la línea o eje horizontal se denomina *abscisa* o eje de las X. Ambos ejes dividen al plano en cuatro *cuadrantes*, y el punto donde se cruzan ambos ejes se denomina *origen* o *punto cero*, ya que las escalas numéricas parten del origen en las cuatro direcciones. Las puntuaciones que parten del origen hacia arriba por el eje Y y, a la derecha, por el eje X son positivas, mientras que las puntuaciones que parten del origen hacia abajo por el eje X y, a la izquierda, por el eje Y son negativas. Dado que la mayoría de las mediciones en sociología se realizan en escalas que parten desde cero sólo en la dirección positiva, el cuadrante primero es el que se suele necesitar preferentemente, por lo que en las representaciones gráficas se omiten con frecuencia el resto de los cuadrantes y sólo se representa el primer cuadrante (ver fig. 1).

2.4.1. *Histogramas*

Un histograma, o histograma de frecuencias, consiste en la representación de una distribución de frecuencias o porcentual, en la que la fre-