El análisis de datos cuantitativos está referido a propiedades que se dan en las personas o grupos sociales, en modalidades o en magnitudes diferentes. El lector recordará que tales propiedades reciben el nombre de *variables*. Específicamente, tal análisis busca la determinación de expresiones estadísticas con las cuales es posible caracterizar a uno o más grupos. Esta búsqueda, de manera general, está predefinida en los objetivos de la investigación. Decimos en general, porque, en algunos casos, el investigador busca en los datos rasgos distintivos que no habían sido expresados en esos objetivos. Lo cierto es que, ya sea una u otra de esas situaciones, podemos decir que las principales direcciones del análisis buscan la determinación de estadísticas como las siguientes:

1. Razones, proporciones, porcentajes y tasas.
2. Distribuciones de frecuencias.
3. Medidas de tendencia central, de dispersión y de concentración.
4. Medidas de asociación y de correlación.

A la presentación detallada de tales medidas está dedicada esta parte del libro.

12

Razones, proporciones, porcentajes y tasas

Las *razones*, las *proporciones*, los *porcentajes* y las *tasas* son medidas estadísticas sencillas que se utilizan para describir la distribución de las personas de un colectivo, en una o más categorías de una variable cualitativa o nominal. Tales medidas tienen especial uso en la comparación de resultados entre esas categorías, como también en la caracterización global de un colectivo.

RAZONES

Las *razones* son cocientes que resultan de comparar el número de personas que quedan ubicadas en una de las categorías de una variable cualitativa dicotómica, con el número de personas que queda en la otra categoría de esa variable. Si, por ejemplo, en una encuesta se encuentra que 120 personas están de acuerdo con la pregunta "¿Está usted de acuerdo con el divorcio?", y 80 están en desacuerdo, la razón entre esos números es:

$$120/80$$

Ese valor se expresa diciendo que la razón de las personas que están de acuerdo con el divorcio respecto de las que están en desacuerdo es de 120 a 80, o, simplificando, de 3 a 2. Al hacer la división, la razón sería igual a 1.5.

Si la anterior relación se hubiese calculado en sentido inverso, se habría obtenido una razón de:

$$80/120, \text{ o sea, de } 2 \text{ a } 3\ (= 0.67)$$

En casos como el anterior, si la división de los valores diera un resultado menor que 1, conviene multiplicar el resultado por 100 (o por 1000, según el número de decimales obtenidos). Conforme a esta convención, la razón entre las personas en desacuerdo y de acuerdo, en el ejemplo dado anteriormente, sería de 80:120 = 0.6666 × 100 = 66.7%. Cuando se utiliza cualquier potencia de 10 en los cálculos, conviene señalarlo de manera explícita.

Como decíamos, las razones tienen especial utilización al comparar la variable del caso entre dos o más grupos. Si hubiese ocurrido que la encuesta citada hubiera sido aplicada en otro lugar y respecto de la pregunta sobre el divorcio hubiese dado una razón de 1.2, diríamos que aquí hay menor diferencia entre los que están de acuerdo con el divorcio y los que están en desacuerdo, ya que en el caso anterior la razón era de 1.5. Igual comparación podría hacerse en un mismo lugar en momentos diferentes para mostrar si ha habido un cambio en la actitud frente al divorcio (o en cualquier otra situación enfocada por una investigación).

Conviene hacer notar que la razón no depende del número absoluto de personas consideradas en el cálculo. Así, si los números de las respectivas categorías de la variable cualitativa (actitud frente al divorcio) hubiesen sido 12 y 8, la razón habría sido la misma: 1.5. Esto significa que las razones son medidas relativas que permiten hacer comparaciones similares con otros grupos, con independencia de sus tamaños.

PROPORCIONES

Las *proporciones*, de manera similar a las razones, son medidas que sintetizan la distribución de las personas en las categorías que pueda tener la variable cualitativa que se considere en el análisis. En este caso, el cálculo se realiza entre el número de unidades o personas en la categoría p y el total de las personas del grupo.

Para utilizar el ejemplo dado al presentar las razones, si las personas que están de acuerdo con el divorcio son 120 y el total del grupo es de 200 (120 + 80), entonces la proporción de personas que tienen esa actitud, en el nivel del grupo, es de:

120/200 = 0.6 (o 0.60, según quiera presentarse el resultado)

La proporción de las personas en desacuerdo es:

80/200 = 0.4 (o 0.40)

En general, si p es el número de personas de una de las categorías de una variable dicotómica, y q es el número de personas de la otra categoría, entonces:

$$p + q = 1$$

En el ejemplo: 0.6 + 0.4 = 1.

A partir del conocimiento del valor de una razón se puede calcular la proporción correspondiente. Si la razón entre quienes están de acuerdo con el divorcio fuera de 7/4 (siete de acuerdo y cuatro en desacuerdo), la proporción respectiva sería 7/11, o sea, 0.63.

Cuando la variable cualitativa tiene dos categorías, es indiferente calcular una razón o una proporción. Sin embargo, cuando esa variable tiene más de dos categorías, es conveniente utilizar proporciones. En la tabla que sigue se muestran las distribuciones de la variable "tipo de educación", completa e incompleta, de 845 personas de las localidades de Amira y de Belicia. La comparación de las cifras absolutas correspondientes a los tipos de educación en ambas localidades hace difícil decir cuál de las dos tiene una distribución de la educación más equitativa o en cuál de ellas existe un mayor nivel de educación superior, pues las localidades difieren bastante en el tamaño de la población que se compara.

Proporción de personas de 30 años y más en tres tipos de educación, en las localidades de Amira y Belicia, 1995

Tipo de educación	*Amira*	*p*	*Belicia*	*p*
Básica	480	0.57	284	0.54
Media	320	0.38	190	0.36
Superior	45	0.05	51	0.10
Total	845	1.00	525	1.00

El cálculo de proporciones facilita mostrar la distribución entre las diversas categorías de una variable dentro de un mismo grupo. Por ejemplo, en el caso de la localidad de Amira, se puede constatar la baja proporción de personas con educación superior (0.05). Con esta cifra, la proporción de personas con educación superior en Belicia es el doble del subgrupo correspondiente en la localidad de Amira (0.10).

PORCENTAJES

Un *porcentaje* es una proporción multiplicada por 100. Los porcentajes cumplen la misma función comparativa que esa medida. De acuerdo con lo expresado, el cuadro anterior (utilizado para el cálculo de proporciones) quedaría así:

Proporción de personas de 30 años y más en tres tipos de educación, en las localidades de Amira y Belicia, 1995

Tipo de educación	*Amira* (%)	*Belicia* (%)
Básica	57.0	54.0
Media	38.0	36.0
Superior	5.0	10.0
Total	100.0 (845)	100.0 (525)

En el ejemplo, los porcentajes se han derivado directamente de los correspondientes cálculos de las proporciones, en los cuales se hizo una aproximación de los decimales para presentarlos con dos dígitos. Si el cálculo original se hubiese hecho con tres dígitos, entonces la proporción de personas con educación básica en Amira habría sido de 0.568, y el correspondiente porcentaje, de 56.8. Las necesidades de precisión deberán considerarse en cada caso.

Cuando se calculan porcentajes, es necesario tener en cuenta la recomendación que nos dice que siempre debe indicarse la base respecto de la cual se hizo el cálculo (en el ejemplo, 845 y 525).

TASAS

La tasa es un tipo especial de razón en la cual el numerador indica el número de un cierto suceso que ocurre durante un determinado periodo, y el denominador es el número de sucesos con los cuales el primero está relacionado. Generalmente, el cociente obtenido se multiplica por 100 o por 1000.

Veamos unos ejemplos de tasas.

1. Tasa de atención en salud en el municipio A:

$$\frac{\text{Total de horas de atención en salud}}{\text{Total de la población del municipio}}$$

   Las horas de atención en salud pueden referirse al total mensual de horas de prestación por el médico o médicos de los servicios de salud del municipio, de las enfermeras, de los auxiliares de salud o de otro personal de salud, de manera separada o por categoría profesional.

2. Tasa de participación en la fuerza de trabajo:

$$\frac{\text{Núm. de personas que trabajan}}{\text{Población total}} \times 100$$

3. Tasa de retención escolar. Es la proporción o porcentaje de alumnos que terminan un cierto ciclo escolar respecto del número de esa cohorte al iniciar el ciclo. Así, si en 1985, 200 alumnos comenzaron sus estudios de un ciclo que terminó en 1990 con 112 alumnos, la tasa de retención es de:

$$112/200 = 0.56 \text{ (o 56 \%)}$$

Otra razón modificada es el llamado *porcentaje de cambio*, que expresa la magnitud del cambio de una variable entre dos periodos. Por ejemplo, si en, digamos, 1990, el número de alumnos matriculados en la carrera de trabajo social era de 1250 (f_1) y en 1999 era de 2540 (f_2), el porcentaje del cambio de la matrícula en esa carrera se calcula con la siguiente fórmula:

$$\text{Porcentaje de cambio: } \frac{f_2 - f_1}{f_1} \times 100$$

Introduciendo los datos del ejemplo:

Porcentaje de cambio de la matrícula en la carrera

$$A = \frac{2540 - 1250}{1250} \times 100 = 103$$

Los cálculos muestran que la matrícula experimentó un aumento de 3 puntos porcentuales entre 1990 y 1999.

13

Distribuciones de frecuencias

A diferencia de las variables cualitativas que vimos en el capítulo anterior, en las cuales las unidades de un colectivo se distribuyen en dos categorías (en el caso de variables dicotómicas) o en unas pocas categorías (como sucede en las variables politómicas), esta situación cambia cuando se trata de describir y analizar la distribución de casos en los múltiples valores que puede tomar una variable cuantitativa que, en el caso de una variable teórica, es un número infinito de valores.

Para los objetivos enunciados de describir y analizar los datos recogidos para una investigación, se utilizan tres tipos principales de su distribución: distribución de frecuencias absolutas, distribución de porcentajes y distribuciones acumuladas de frecuencias o de porcentajes.

DISTRIBUCIONES DE FRECUENCIAS ABSOLUTAS

Las *distribuciones de frecuencias* muestran las frecuencias con las cuales se presentan los valores de una variable, ordenadas según el tamaño de esos valores. Cuando la variable es discreta (es decir, cuando entre dos valores de ella no hay otro valor intermedio), no hay mayor problema en el ordenamiento de sus valores; en tal caso, lo primero es elegir un número de intervalos, que no debería ser mayor que 8, con el fin de favorecer la comparación entre los grupos. Por ejemplo, supongamos que tenemos las escolaridades de 150 personas con los siguientes números de años de estudio:

12, 14, 11, 13, 8, 7 y 9

Para hacer la distribución de frecuencias, procedemos a elegir un número de intervalos, como dijimos. Si se eligen cinco intervalos (también denominados *intervalos de clases*), dividimos la diferencia entre los valores extremos (14 − 7 = 7) entre el número de intervalos elegido (7/5 = 1.4). Como el resultado es inferior a 1.5 (cifra que permitiría elevar ese valor a 2), bajamos al número 1. Entonces los intervalos son:

7-8; 9-10; 11-12; 13-14 y 15-16

A continuación se procede a anotar el número de personas incluidas en cada intervalo, según su nivel educativo. Para simplificar la presentación, supongamos que se da la siguiente distribución de frecuencias:

Niveles de escolaridad	*Frecuencias (f)*
7-8	77
9-10	45
11-12	19
13-14	9
15-16	0
Total	150

La distribución muestra que el mayor número de personas tiene escolaridades de siete y ocho años, seguido por las personas que tienen nueve y 10 años. En el grupo de mayor escolaridad sólo se tienen nueve personas. (En la tabla se colocó el intervalo 15-16 que fue definido por los cinco intervalos elegidos al comienzo. En ese intervalo no se tiene ninguna persona. Cuando sucede esto, no se coloca el intervalo, pero conviene decir la razón por la cual no figura en la tabla.)

Cuando la variable es *continua*, es decir, variable en la cual entre dos números sucesivos de ella puede darse otro valor (por ejemplo, entre 4 y 5 puede darse el valor 4.86), la elaboración de la distribución de frecuencias es un poco más complicada, como veremos a continuación.

Supongamos que se tienen los valores que se indican enseguida, correspondientes a tasas de deserción de 33 escuelas de la localidad de Salvadores:

17.8, 6.9, 21.2, 28.5, 33.2, 12.5, 18.2, 23.8, 28.5, 7.5, 14.2, 17.5, 21.9, 26.8, 24.5, 8.5, 18.3, 22.7, 23.5, 22.8, 8.7, 12.4, 19.5, 27.8, 13.8, 16.4, 23.4, 25.8, 11.0, 18.1, 21.7, 32.1 y 23.2

Para organizar esos datos con propósitos de análisis, es necesario realizar las operaciones siguientes:

1. *Seleccionar el número de categorías en las cuales se distribuirán los datos.* Un número entre 5 y 8, recordemos, se considera habitualmente como conveniente. Sin embargo, ese número dependerá del interés que tenga el investigador en mostrar con mayores o menores detalles la distribución de los datos.

2. *Determinar el tamaño del intervalo de clase.* Si bien se han propuesto fórmulas sencillas para determinar ese tamaño, como lo hicimos en el ejemplo anterior (diferencia entre el tamaño del dato mayor y el tamaño del dato menor, dividida entre el número de intervalos que se desea tener), en general es preferible utilizar tamaños pequeños, pues, si es conveniente para un análisis posterior, podemos reagrupar los intervalos en otros de mayor amplitud (por ejemplo, reagrupar los datos en intervalos de 10 puntos, en lugar de intervalos de cinco puntos).

3. *Formar los intervalos.* Si, para los datos del ejemplo, elegimos un intervalo de 5, el límite inferior del primer intervalo podría ser 5.0 (que es inferior al menor de los datos: 6.9), y el límite superior, 9.9; el segundo quedaría entre 10.0 y 14.9, etcétera.

La tabla quedaría como sigue:

Distribución de las frecuencias de las tasas de deserción de 33 escuelas de la localidad de Salvadores

Intervalos	*Frecuencias*
5.0-9.9	4
10.0-14.9	5
15.0-19.9	7
20.0-24.9	10
25.0-29.9	5
30.0-34.9	2
Total	33

Como se puede ver en la tabla, los límites de los intervalos se han elegido de modo que no se confundan con los que le siguen. Sin embargo, si se considera que en el cálculo de las tasas de deserción (o de los datos que estamos agrupando) se ha podido proceder a redondear las cifras decimales, entonces queda la pregunta acerca de qué sucede con tasas que, como ejemplos, tenían los valores 19.95 y 19.94. La respuesta general es que, en casos como ésos, si el valor del segundo decimal es igual o mayor que 5, hace subir al primer decimal al dígito inmediato superior, y si el valor es menor que 5, el primer decimal no cambia el valor. Según esta convención, la tasa de 19.95 queda en el intervalo de 20.0-24.9 y la tasa de 19.94 queda en el intervalo de 15.0-19.9. Los valores de 19.95 y 19.94 constituyen los *límites reales* de los intervalos de clase.

DISTRIBUCIONES DE PORCENTAJES

Las *distribuciones de porcentajes* están constituidas por los porcentajes de las frecuencias absolutas calculadas respecto del total de ellas. En general, los porcentajes permiten una mejor comparación de la distribución relativa de los casos en el colectivo. En el ejemplo anterior, la distribución de frecuencias tomaría esta forma:

Distribución porcentual de 33 escuelas de la localidad de Salvadores, según sus tasas de deserción

Tasas	*Frecuencias* (*f*)	*Porcentajes* (%)
5.0-9.9	4	12.1
10.0-14.9	5	15.2
15.0-19.9	7	21.2
20.0-24.9	10	30.3
25.0-29.9	5	15.2
30.0-34.9	2	6.0
	33	100.0 (33)

En la tabla de una distribución de porcentajes no es necesario colocar las columnas de frecuencias absolutas, pero sí debe indicarse la base de su cálculo. En el ejemplo, el número 33, que se coloca al pie de 100%.

La utilidad de las distribuciones de porcentajes para comparaciones se puede ver en la tabla que sigue, tomada de una investigación sobre el proceso de "elitización" de las universidades chilenas durante el periodo 1976-1981 (publicada en Guillermo Briones, "La educación superior en el modelo de la economía neoliberal", en *Las transformaciones educacionales bajo el régimen militar*, vol. 2, PIIE, Santiago, 1984, pp. 237-354).

Composición social de los estudiantes seleccionados para ingresar a las universidades chilenas, 1976 y 1981

Nivel educativo del padre	*1976*	*1981*	*1976* (%)		*1981* (%)	
Sin estudios	208	139	0.8		0.5	
Primaria incompleta	3.058	3.186	11.9	29.7%	12.6	26.1%
Primaria completa	4.387	3.296	17.0		13.0	
Enseñanza media incompleta	5.652	4.484	21.9		17.7	
Enseñanza media completa	6.446	7.276	25.0		28.7	
Universitaria incompleta	1.514	1.479	5.9		5.8	
Egresado de universidad	546	586	2.1		2.3	18.1%
Universitaria completa	3.201	4.010	12.4	14.5%	15.8	
Estudios militares	739	881	2.9		3.5	
Total	25.751	25.337	100.0		100.0	

NOTA: En este y en los cuadros que siguen no se consideran los alumnos seleccionados para las carreras que se indican a continuación, por carecer de información comparable en ambos años o por tener sólo la condición de preseleccionados: carreras de Universidad Católica de Valparaíso; Psicología, Pedagogía en Artes Plásticas, Música, Idiomas, Educación Diferencial, Ingeniería de la Universidad Técnica Federico Santa María; Bioquímica de la Universidad Católica.

FUENTE: Comisión Coordinadora del Proceso de Admisión de Alumnos a las Universidades Chilenas.

DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS

En algunos casos, los datos de una investigación se presentan de manera tal que se pueda mostrar el número de casos que son mayores o menores que un cierto valor. La presentación se puede hacer a partir de distribuciones de frecuencias absolutas o de frecuencias porcentuales, como se ve en la tabla siguiente:

Distribuciones de frecuencias acumuladas de consultas mensuales de salud, de una muestra de 138 personas

Número de consultas	*Frecuencias (%)*		*Frecuencias acumuladas*		*Porcentaje acumulado*	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1-2	15	10.9	15	138	10.9	100.0
3-4	22	15.9	37	123	26.8	89.1
5-6	36	26.1	73	101	52.9	73.2
7-8	40	29.0	113	65	81.9	47.1
9-10	15	10.9	128	25	92.8	18.1
11-12	10	7.2	138	10	100.0	7.2

La columna 6 de la tabla presenta los porcentajes acumulados hacia abajo, y cada cifra indica el porcentaje de casos que queda dentro y por encima del intervalo de clase que le corresponde. Así, la cifra 52.9 significa que ese porcentaje de personas han solicitado atención de salud entre una y seis veces. En la columna 7, los porcentajes se han acumulado hacia arriba, y cada cifra indica el porcentaje de casos que quedan dentro y por debajo del intervalo de clase correspondiente. Así, la cifra 18.1 indica que ese porcentaje de personas han hecho entre nueve y 12 consultas.

CRUCE DE VARIABLES. REGLA DE ZEISEL

En los ejemplos dados de distribuciones de frecuencias vimos cómo las personas se distribuían en las categorías o intervalos de una sola variable. Una forma distinta de distribución se da cuando se utilizan dos o tres variables con dos o más categorías, cada una de ellas para comparar, precisamente, cómo se distribuyen en las subcategorías que resultan por el cruce de ellas. Un cuadro en el cual las dos variables tienen el mismo número de categorías se denomina *cuadro de* $n \times n$ (por ejemplo, de 3×3). Un cuadro en el cual se han cruzado dos variables con distinto número de categorías se denomina de $m \times n$. La multiplicación da el número de celdas del cuadro.

Lo que queremos decir hasta aquí se podrá apreciar directamente en el siguiente ejemplo. Supongamos que hemos aplicado una prueba de autoestima a una muestra de personas que tienen distintos niveles de escolaridad. Supongamos, además, que cada una de esas variables ha sido categorizada en tres niveles. Al cruzarlas y calcular los respectivos porcentajes, obtendríamos un cuadro de doble entrada, como el siguiente:

	Escolaridad		
Autoestima	*Alta* (%)	*Media* (%)	*Baja* (%)
Alta	61.3	51.2	26.3
Media	27.2	29.3	52.6
Baja	11.4	19.5	21.1
Total	100.0 (44)	100.0 (84)	100.0 (114)

Las cifras del cuadro muestran que una mayor proporción de personas con alta escolaridad tiene mayores niveles de autoestima (61.3%); siguen las personas con escolaridad media (51.2%), y luego, las personas con baja escolaridad (21.1%).

Los cuadros de doble entrada tienen dos tipos de marginales o de totales: los que corresponden a las filas y los que corresponden a las columnas. Por lo mismo, en algunas ocasiones se produce una cierta confusión para determinar en qué sentido deben calcularse los porcentajes que se van a comparar. Las respuestas a esta duda están dadas por la llamada *regla de Zeisel*, que dice: los porcentajes deben calcularse en la dirección de la variable independiente. Es decir, las bases de los porcentajes son los números correspondientes a cada una de las categorías de esa variable. En el ejemplo dado, la variable independiente es la escolaridad, ya que esta característica podría producir variaciones en la autoestima. Las bases de cálculo son, por consecuencia, los números 44, 84 y 114. (En el ejemplo no dimos las frecuencias absolutas correspondientes a cada una de las celdas del cuadro de porcentajes.)

REPRESENTACIONES GRÁFICAS

Las diversas técnicas que se utilizan en la presentación de datos permiten a muchas personas captar mejor los datos numéricos obtenidos en una investigación. De esas técnicas, que toman numerosas formas, vamos a presentar sólo las que expresan, por medio de gráficos, distribuciones de frecuencias y que, según el caso, reciben los nombres de histograma, polígono de frecuencias y ojiva.

Histograma

El *histograma* es la representación gráfica de una distribución de frecuencias absolutas o de porcentajes, en la cual las frecuencias o los porcentajes correspondientes a cada categoría o intervalo de clase se representan mediante una barra o rectángulo, cuya altura es proporcional al número de casos (se indica en el eje de las *y*) y su base tiene una longitud arbitraria (que se marca en el eje de las *x*) que es igual para todos los intervalos si éstos tienen igual tamaño.

Debe tenerse en cuenta que el eje de las *y* siempre comienza en el origen o en el punto cero, pero el eje de las *x* comienza, convencionalmente, con el intervalo que tiene la menor frecuencia. En el caso de variables nominales, las barras que representan proporcionalmente las frecuencias suelen presentarse separadas para indicar que corresponden a categorías diferentes (por ejemplo, si las frecuencias del caso correspondieran a un cierto número de personas que respondieran "Sí" y "No" a una pregunta de un cuestionario). La representación gráfica se conoce comúnmente con el nombre de *diagrama de barras* (*véase* fig. 13.1). En el caso de variables ordinales, el procedimiento de representaciones es similar al que se realiza con una variable continua.

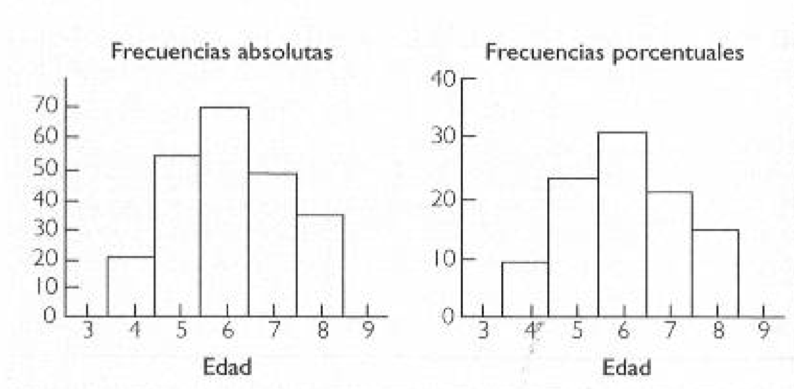

Fig. 13.1. Histogramas de las frecuencias absolutas y porcentuales de las edades de 223 niños.

Polígonos de frecuencias

Son figuras que resultan de unir los puntos medios de las bases superiores de las barras o rectángulos del respectivo histograma. Los rectángulos no aparecen en el gráfico final.

Al unir los puntos mencionados, es necesario que el polígono inicie en el punto medio del primer rectángulo y llegue hasta el punto medio del último rectángulo del histograma. La prolongación errada hasta el punto cero de los intervalos del histograma implica una falsa representación de los datos que sirven de base a la representación gráfica.

Cuando se comparan polígonos de dos o más grupos, es necesario convertir las frecuencias absolutas de las correspondientes distribuciones en frecuencias relativas o porcentuales (*véase* fig. 13.2).

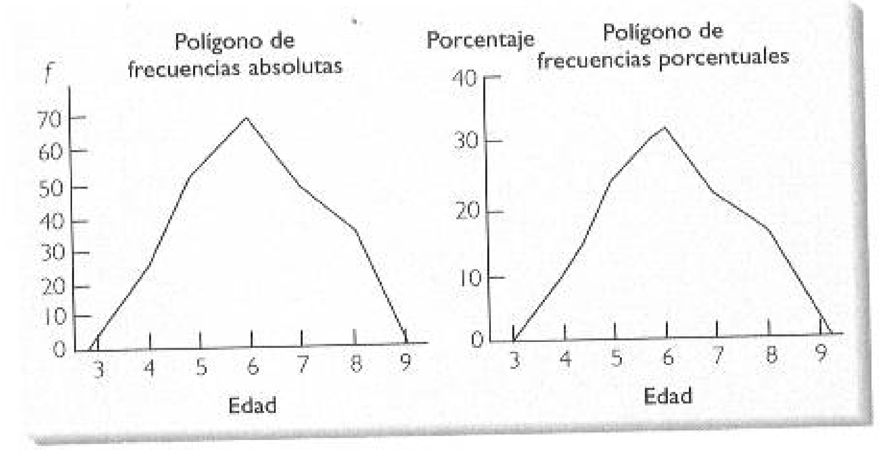

Fig. 13.2

Ojiva

La *ojiva* es la representación gráfica de las frecuencias absolutas acumuladas o de los respectivos porcentajes acumulados. El procedimiento de construcción es, en ambos casos, similar al que se utiliza en el histograma.

Cualquiera de las dos ojivas permite apreciar en forma gráfica el número de casos o de porcentajes que quedan por encima o por debajo de un determinado valor de la variable (*véase* fig. 13.3).

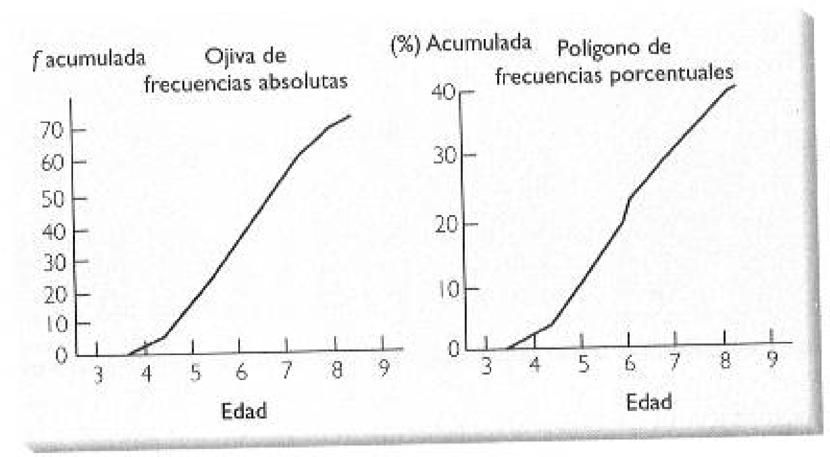

Fig. 13.3

Otras formas de representación gráfica

Existen, asimismo, muchas otras maneras de representar visualmente las frecuencias que tienen los valores de las variables que investigamos (*véase* fig. 13.4).

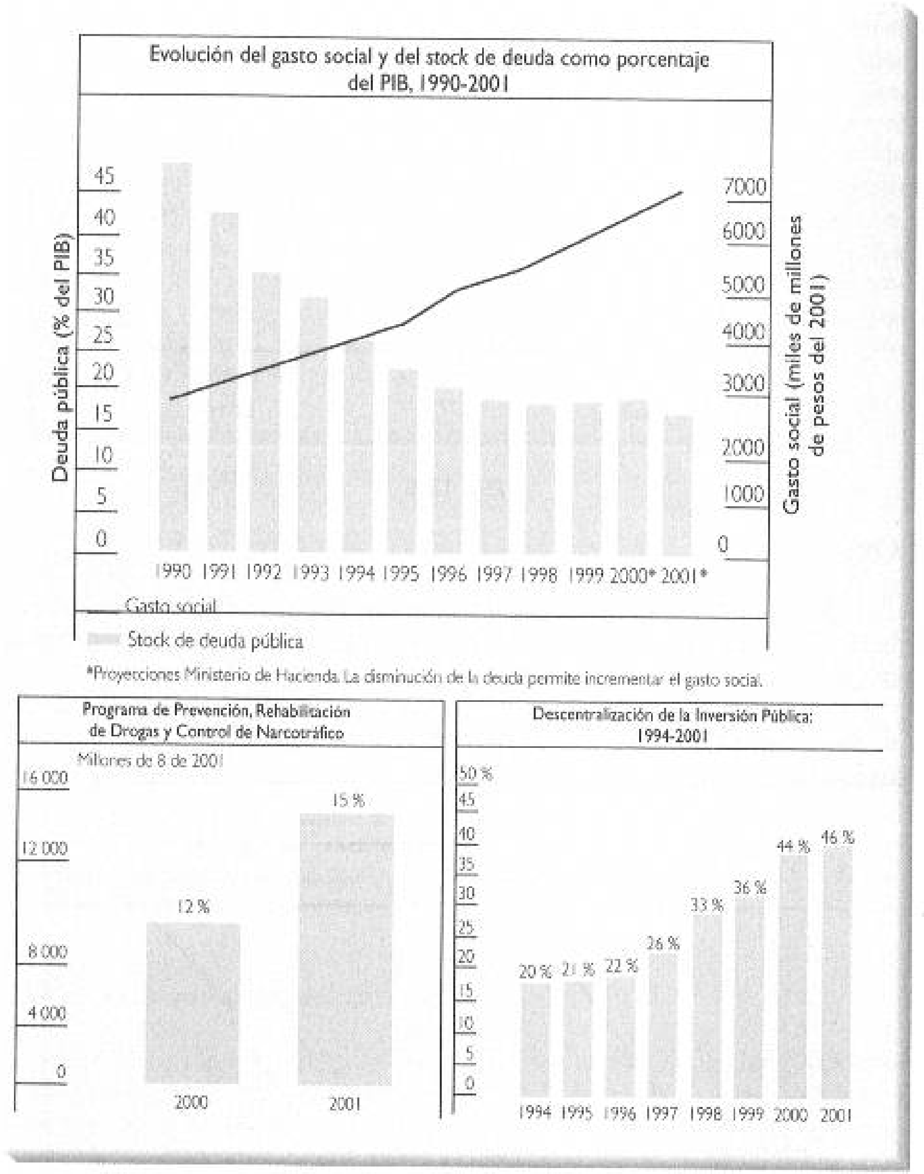

Fig. 13.4. Otras formas de representación gráfica de frecuencias.

SERIES DE TIEMPO

Una *serie de tiempo* es un conjunto de datos numéricos que se recolectan en intervalos sucesivos de tiempo. En el análisis de una serie de tiempo se consideran los siguientes cuatro componentes:

- La tendencia (T).
- La fluctuación cíclica (C).
- La variación estacional (E).
- Un movimiento irregular (I).

De acuerdo con esta distinción, se acepta que el valor de una serie temporal puede expresarse en función de la tendencia que se dé en ella, de su fluctuación cíclica, de su variación estacional o de un movimiento irregular.

La *tendencia* se define como aquella parte de la serie de tiempo que muestra un movimiento suave o regular en un periodo largo. La *variación estacional* es la componente que muestra una forma o patrón regular de comportamiento durante ciertos subperiodos del periodo total considerado. La *variación cíclica* se muestra en los datos como movimientos alrededor de la tendencia y, finalmente, la componente *irregular* es la parte de la serie que se explica por fenómenos poco frecuentes que la afectan, como podría ser una catástrofe natural (terremoto, huelga, etc.) o también podría presentarse por baja calidad de los datos.

A continuación veremos dos métodos que se utilizan en la determinación de la tendencia y de una serie de tiempo.

Cálculo de la tendencia

Los métodos más utilizados son: el método de los semipromedios y el método de los promedios móviles.

Método de los semipromedios

Consiste en dividir la serie temporal en dos partes, calcular luego la media aritmética de cada parte, ubicar esos dos valores en la serie original y, finalmente, unir esos dos puntos mediante una recta. Tal recta es la expresión gráfica de la tendencia buscada. Veamos estos cálculos con un ejemplo. Supongamos que tenemos la serie siguiente, que muestra las variaciones de la población en edad escolar que termina el último año de la enseñanza básica en cada uno de los años que se indica.

Año	x	Población de escolares que termina la educación básica (en miles): y	
1993	0	50.4	
1994	1	55.2	
1995	2	62.4	$\bar{x}_0 = 56.0$
1996	3	68.4	
1997	4	74.4	
1998	5	79.2	$\bar{x}_1 = 74.0$

Como dijimos, en primer lugar se calcula el promedio de los alumnos egresados en los años 1993, 1994 y 1995, correspondientes a la primera mitad de la serie, que es igual a 56.0; luego se procede con un cálculo similar en la segunda mitad, correspondiente a los años 1996, 1997 y 1998, lo que da una media aritmética de 74.0. Enseguida se realiza –si no se ha hecho antes– la representación gráfica de la serie original, y en tal representación se hace pasar una línea entre los puntos 56.0 y 74.0, ubicados en el centro de los semi-intervalos de tiempo (1994 y 1997). Tal línea se denomina *línea de los semipromedios*; su prolongación hasta cortar el eje de los números de egresados indica la tendencia de la serie. Esa tendencia muestra que, por ejemplo, en el año 2000, el número de egresados sería de unos 108 000 (*véase* fig. 13.5).

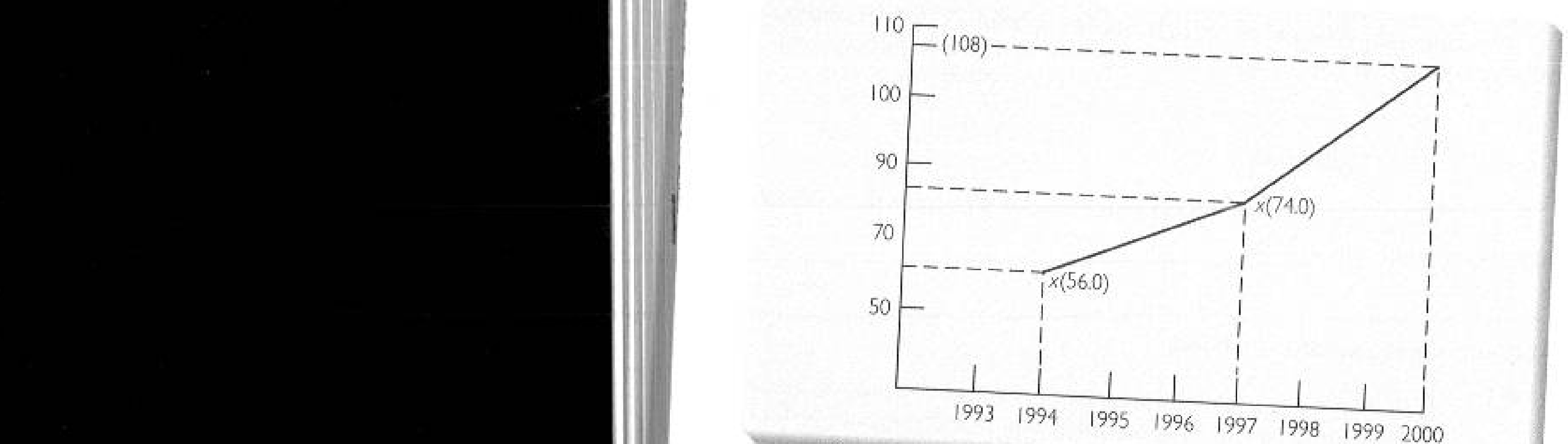

Fig. 13.5

Cuando la serie temporal se compone de un número impar de periodos, se puede eliminar uno de ellos para obtener un número par de valores.

Método de los promedios móviles

Este método permite eliminar en mayor magnitud las variaciones de una serie, debidas a componentes distintos de la tendencia. Por eso se le utiliza, en general, para suavizar los datos de una serie temporal. Comprende las siguientes etapas:

1. Se determinan en la serie periodos de tres o cinco años, según convenga mejor a los datos de la serie total.
2. Se calculan los totales de la serie formada por los datos del primer periodo elegido (vamos a suponer periodos de tres años).
3. El segundo periodo elegido comienza en el segundo dato del primer periodo. El tercero y siguientes se calculan de la misma manera (esto es, tomando siempre el segundo dato de la serie anterior como origen de la serie siguiente).
4. En cada uno de los periodos calculados según el procedimiento señalado anteriormente, se calculan los promedios dividiendo el total entre el número de años del periodo (en el ejemplo, 3).

Veamos un ejemplo parcial de este procedimiento. Supongamos que tenemos la serie de la población femenina matriculada en la educación básica de una cierta región.

Año	Población femenina	Serie
1993	13 580	13 580 + 15 430 + 18 260 = 47 270
1994	15 430	47 270 : 3 = 15 757
1995	18 260	15430 + 18 260 + 21 530 = 55 220
1996	21 530	55 220 : 3 = 18 406
1997	25 720	etcétera
1998	29 800	

Como en el caso anterior, la unión de los puntos dados por los promedios móviles da la recta que, al cortar el eje de la variable de la serie, indica su tendencia. Sin embargo, este procedimiento tiene dos inconvenientes principales: 1. se pierden los años del extremo inferior y del extremo superior de la serie, y 2. no se puede calcular la forma matemática de la línea recta.