Capítulo 7

DISTRIBUCIONES DE FRECUENCIAS

Sumario

Objetivos específicos

Al finalizar el estudio del capítulo, el estudiante será capaz de:

1. Justificar la necesidad de la distribución de frecuencias en el análisis estadístico.
2. Distinguir e interpretar correctamente los elementos, criterios y principios que entran en la elaboración de una distribución de frecuencias.
3. Construir correctamente distribuciones de frecuencias para variables discretas y continuas, siguiendo los principios fundamentales que se enuncian para su elaboración.
4. Seleccionar y construir las representaciones gráficas más apropiadas para los distintos tipos de distribuciones de frecuencias.

Resumen

En este tema se señala la necesidad de resumir la información para su mejor compresión, análisis y transmisión, se presentan las distribuciones de frecuencias como una técnica para lograr los objetivos anteriores. Se indica el procedimiento por seguir para su construcción y se exponen también los tipos de gráficos para la apropiada representación de las distintas distribuciones de frecuencias.

7.1. NECESIDAD DE RESUMIR LA INFORMACIÓN CUANTITATIVA: LA DISTRIBUCIÓN DE FRECUENCIAS

Los datos estadísticos provienen de las operaciones de contar o de medir, y pueden obtenerse si se anota el número de elementos correspondientes a cada una de las categorías definidas o también ser el fruto de una operación más compleja, como la medición de la intensidad o magnitud de una característica. Esto da origen, como ya se expuso en el capítulo 1, a dos tipos de variables: cualitativas (o atributos) y cuantitativas, dentro de estas últimas se distinguen las continuas y las discretas.

En los dos capítulos inmediatamente anteriores, se ha señalado la necesidad de clasificar los datos y los requisitos que debe cumplir esta labor; también, se explicaron los procedimientos apropiados para construir y representar series cualitativas mediante cuadros y gráficos. Ahora, el interés se centrará en las distribuciones de frecuencias, estas son clasificaciones que se refieren a variables cuantitativas, continuas o discretas y constituyen un instrumento muy útil en el trabajo estadístico.[1]

En el análisis e interpretación de los datos estadísticos correspondientes a variables continuas y discretas, resulta muy valioso disponer de elementos descriptivos que den información acerca de tres aspectos:

a) La **forma** o patrón de distribución de los datos.

b) La **posición** de la distribución, o sea, alrededor de qué valor se tienden a concentrar los datos (valores centrales).

1. Puede hablarse también de distribuciones de frecuencias cualitativas, sin embargo, el término se empleará aquí únicamente para hacer referencia a clasificaciones que involucran variables continuas o discretas.

c) La **dispersión** de los datos alrededor de los valores centrales o promedios (variabilidad).

Esta información puede lograrse fácilmente cuando el conjunto de interés contiene pocos datos: basta ordenarlos de acuerdo con su magnitud y con un simple examen se podría conocer cuál es la forma de su distribución (a), el valor mayor y el menor, la amplitud general (o sea, la diferencia entre esos dos valores extremos) que da idea de su variabilidad (c); la existencia de concentraciones alrededor de algunos valores (b), además de otros detalles similares. Cuando los datos son relativamente numerosos, sin embargo, lo anterior resulta insuficiente y se hace difícil apreciar las regularidades que existen en el conjunto, realizar análisis o sacar conclusiones. Por ello, debe recurrirse a agruparlos en una distribución de frecuencias, esta puede definirse como una ordenación o arreglo en clases o categorías que muestra, para cada una de ellas, el número de elementos contenidos o su frecuencia.

Un ejemplo del uso de la distribución de frecuencias aparece en el informe de una encuesta realizada hace unos años por una consultora privada, la cual abarcó un cierto número de puertos del Golfo de México y estudió una muestra estadística de 1200 trabajadores portuarios. Se indagó, entre otras cosas, el monto del salario percibido en el mes previo al estudio, el cual para fines prácticos fue expresado en dólares.[2] La información individual suministrada por cada uno de los trabajadores entrevistados fue agrupada y resumida, con fines de análisis e interpretación, en la distribución de frecuencias que se incluye seguidamente. En ella, las clases aparecen en la primera columna, las frecuencias absolutas en la segunda y las frecuencias relativas, expresadas en porcentajes, en la tercera.

No es difícil imaginar lo complicado que resultaría manejar 1200 datos y tratar de sacar de ellos, tal como se recogieron, conclusiones acerca de la distribución de los salarios, los más comunes, su variabilidad, entre otros. Tampoco resulta complejo apreciar cómo –después de agruparlos en una distribución de frecuencias– es posible lograr una serie de detalles acerca de esos aspectos mencionados. Así, por ejemplo, los datos agrupados revelan, entre otras cosas, que la mayor frecuencia corresponde a la categoría (clase) 500 a 999 dólares, la cual contiene un total de 366 entrevistados y, conforme se consideran salarios más elevados, las frecuencias tienden a disminuir, de tal manera que solo 13 ganan 4000 dólares o más. También se aprecia que solo 104 de los 1200 entrevistados ganan menos de 500 dólares.

2. El estudio incluyó trabajadores que se encargan directamente de los procesos de carga-descarga y de actividades de apoyo, así como a los supervisores y otro personal directamente responsable de la realización de esas actividades, pero excluyó a otro como el encargado de asuntos administrativos, financieros y contables.

Cuadro 7.1

SUELDO MENSUAL DE UNA MUESTRA DE 1200 TRABAJADORES PORTUARIOS DE UN GRUPO SELECCIONADO DE PUERTOS DEL GOLFO DE MÉXICO
(En dólares)

SALARIO EN DÓLARES (clases)	NÚMERO DE ENTREVISTADOS (frecuencia absoluta)	PORCENTAJE DEL TOTAL (frecuencia relativa)
Menos de 500	104	8,7
500-999	366	30,5
1000-1499	300	25,0
1500-1999	178	14,8
2000-2499	99	8,3
2500-2999	68	5,7
3000-3499	47	3,9
3500-3999	25	2,1
4000-4499	6	0,5
4500 o más	7	0,6
TOTAL	**1200**	**100**

Fuente: evaluación de la situación laboral y salarial de los trabajadores portuarios del Golfo de México. Documento confidencial

Se logra información más útil si se revisa la columna de las frecuencias relativas (%). Es fácil determinar que un 30,5% de los entrevistados tienen un salario mensual entre 500 y 1000 dólares y que un poco más del 70% devengan salarios entre 500 y 2000 dólares. Se aprecia, además, que los entrevistados con salarios menores de 500 dólares representan solo un 8,7% del total y aproximadamente 1% gana 4000 dólares o más. El resto del presente capítulo se dedica a tratar sobre las distribuciones de frecuencias. Lo referente a medidas de posición y variabilidad se considerará en los dos siguientes.

7.2. DISTRIBUCIÓN DE FRECUENCIAS DE VARIABLES DISCRETAS

Para ilustrar la construcción de distribuciones de frecuencias para variables discretas, considere la información del cuadro 7.2, que corresponde al número de hermanos de los 32 alumnos de un colegio rural de la provincia de San José.

Cuadro 7.2

NÚMERO DE HERMANOS DE LOS 32 ALUMNOS DE UN COLEGIO RURAL

ESTUDIANTE	Nº HERMANOS	ESTUDIANTE	Nº HERMANOS	ESTUDIANTE	Nº HERMANOS
AA	3	NL	1	KL	2
RA	2	PE	2	OT	1
AD	2	LT	5	UA	1
MM	3	NR	2	FJ	2
RQ	2	EK	1	MN	3
RG	3	DG	2	BD	2
FG	0	JT	2	VC	1
RJ	3	GA	2	CF	2
JZ	4	IM	1	EZ	3
JF	2	TG	3	SL	2
WA	6	PG	2		

Un análisis más cómodo y eficaz puede lograrse al agrupar las respuestas en una distribución, tal como se indica en el cuadro 7.3. La frecuencia para cada número de hermanos se logra al marcar una línea vertical (|) cada vez que aparece una observación que corresponde a ese número (columna 2). Para facilitar la totalización posterior, cada quinta observación se marca con una línea diagonal que cruza las anteriores, así:

Cuadro 7.3

RESPUESTAS AGRUPADAS DEL NÚMERO DE HERMANOS
DE LOS 32 ALUMNOS DE UN COLEGIO RURAL

Nº HERMANOS	RECUENTO	FRECUENCIA ABSOLUTA (fi)	FRECUENCIA RELATIVA (fr)
(1)	(2)	(3)	(4)
0	I	1	0,03
1	卌 I	6	0,19
2	卌 卌 IIII	15	0,47
3	卌 II	7	0,22
4	I	1	0,03
5	I	1	0,03
6	I	1	0,03
		32	1,00

La distribución muestra (columna 3) que el número más frecuente de hermanos es 2 y que la mayoría de los alumnos (28) tiene entre 1 y 3 hermanos.

Si se divide la frecuencia absoluta de cada clase entre el total de observaciones (32 en este caso), se obtienen las frecuencias relativas (columna 4). Estas permiten conocer qué proporción de los alumnos tiene un número dado de hermanos. Por ejemplo: 0,19, que es la frecuencia relativa correspondiente a la segunda clase, indica que un 19% de los alumnos tiene un hermano.

Las distribuciones de variables discretas pueden presentarse por medio de un gráfico de bastones o de un gráfico de barras verticales, al utilizar las frecuencias absolutas o las relativas. Seguidamente se ilustran, con este ejemplo del número de hermanos, esos dos tipos de gráficos, empleando las frecuencias absolutas.

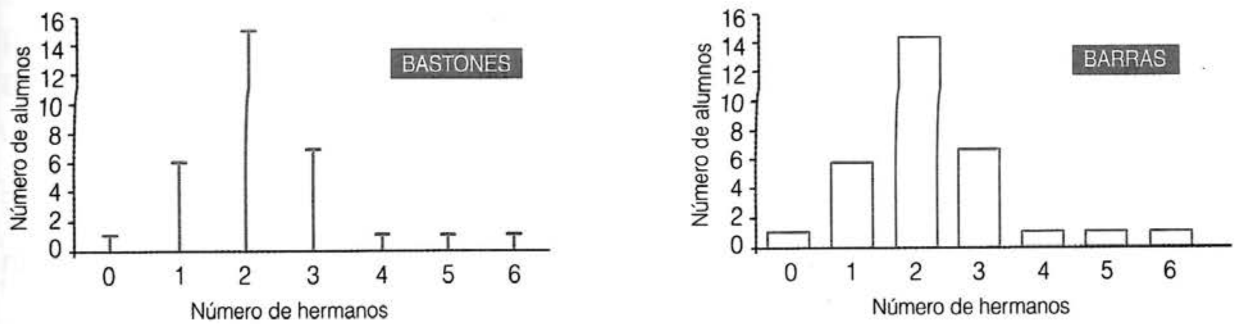

Figura 7.1. Gráficos que muestran los estudiantes de un colegio rural de la provincia de San José, según número de hermanos, año 2009

Para ciertos propósitos resulta útil acumular las frecuencias absolutas o las relativas. Esto puede hacerse sumándolas hacia abajo, para obtener una distribución acumulada "menos de", o hacia arriba, lo que produce una distribución acumulada "más de" (ver cuadro 7.4). Así, por ejemplo, si se suman las frecuencias absolutas de las clases primera, segunda y tercera, se obtiene 1 + 6 + 15 = 22, este valor representa la frecuencia acumulada "menos de" de la tercera clase. Si se realiza la misma operación con las frecuencias relativas, se obtiene 0,69, quiere decir que 22 estudiantes tienen 2 hermanos o menos (recuérdese que se sumaron las categorías 1, 2 y 3); aunque también sería correcto decir que 22 estudiantes tienen menos de 3 hermanos. La interpretación del 0,69 es similar; 69% de los estudiantes poseen menos de 3 hermanos.

Si la suma se realiza de abajo hacia arriba, al partir de la última clase se obtiene una frecuencia acumulada "más de". Los números 10 y 0,31 de la cuarta clase son ejemplos de ellas. Estos valores se obtienen de la siguiente forma: el número 10, al sumar las frecuencias absolutas de las clases sétima, sexta, quinta y cuarta, o sea, 1 + 1 + 1 + 7 = 10 y el número 0,31 sumando 0,03 + 0,03 + 0,03 + 0,22 = 0,31. Su interpretación se deriva de la forma de cálculo. El número 10 indica que 10 estudiantes tienen 3 hermanos o más, aunque también sería correcto decir que 10 estudiantes tienen más de 2 hermanos. El número 0,31 se interpreta en forma similar: 31% de los estudiantes tiene 3 hermanos o más.

Cuadro 7.4

DISTRIBUCIÓN ACUMULADA

Nº HERMANOS	f	fr	ACUMULADA Absoluta	"MENOS DE" Relativa	ACUMULADA Absoluta	"MÁS DE" Relativa
0	1	0,03	1	0,03	32	1,00
1	6	0,19	7	0,22	31	0,97
2	15	0,47	22	0,69	25	0,78
3	7	0,22	29	0,91	10	0,31
4	1	0,03	30	0,94	3	0,09
5	1	0,03	31	0,97	2	0,06
6	1	0,03	32	1,00	1	0,03
	32	1,00				

Ambos tipos de frecuencias acumuladas, la "menos de" y la "más de" son útiles; sin embargo, la más empleada en la práctica es la primera.

La representación gráfica de las frecuencias acumuladas discretas –absolutas o relativas– puede hacerse con gráficos de bastones o barras verticales. Sin embargo, cuando las clases son numerosas, la construcción de las barras puede resultar lenta y tediosa, y el gráfico final destaca poco los espacios entre ellas y más bien tiende a dar una impresión de continuidad. Por este motivo, algunas personas prefieren usar la curva simple para representar las acumuladas discretas y otras, el tipo de gráficos que se indican a continuación (gráficos de la figura 7.2).

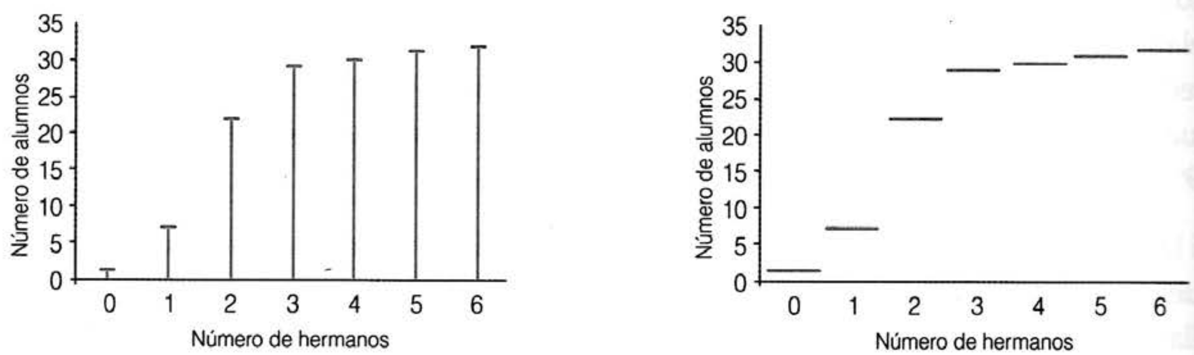

Figura 7.2. Gráficos con las frecuencias acumuladas "menos de" para el número de hermanos de los estudiantes de un colegio rural de la provincia de San José, año 2009

7.3. LA MEDICIÓN DE LAS VARIABLES CONTINUAS Y EL PROBLEMA DEL REDONDEO

Desde el punto de vista teórico, una variable continua puede ser medida con la exactitud que se quiera. En la práctica, sin embargo, depende de la sensibilidad o precisión del instrumento empleado para hacerla. Así, por ejemplo, una persona puede pesarse en la balanza de una farmacia o supermercado y obtener su peso en kilogramos (kg) o quizás en décimas de kilogramo, o puede ser pesada en un laboratorio con un aparato que permite obtenerlo con cinco cifras decimales, o sea en cienmilésimas de kilogramo. Para efectos prácticos, sin embargo, esta elevada exactitud puede no ser necesaria y, por ello, el individuo o quien lo pesa se conforma con el peso en kilogramo o en décimas de kilogramo.

En la realidad, entonces, lo usual es que las mediciones de las *variables continuas* se expresen redondeadas en cierto tipo de unidades, ya sea porque el instrumento disponible no permita ir más allá o porque quienes utilizarán los datos no requieren registrarlos con tanta exactitud como admite el instrumento de medición.

Esto lleva al tema de las formas de redondeo que se emplean. La costumbre es el redondeo usual, o sea, "a *la unidad más próxima*". Sin embargo, este no es el único procedimiento posible, los números también pueden redondearse "*hacia arriba*" o "*hacia abajo*". Seguidamente, se comenta la naturaleza de cada uno de esos procedimientos.

7.3.1. Redondeo a la unidad más próxima

Este método tiene muchas ventajas prácticas y es el más comúnmente usado. Para explicar cómo funciona, considere un ejemplo. Suponga que cuatro amigos se pesan y la balanza arroja los siguientes valores, en kilogramos:

64,38 63,50 69,67 66,50

Si quiere redondear estos pesos a kilogramos completos, con el método de la *unidad más próxima*, ¿cómo debe proceder?

En primer término, note que en este caso la unidad de interés es el kilogramo, al redondear se quiere expresar cada uno de los pesos en un número redondeado entero, dado en kilogramos completos, con el criterio de la unidad más próxima.

Para esto se debe analizar la magnitud del número a la derecha de la coma, el que se quiere eliminar, el cual está expresado en centésimas de kilogramo, o sea en centésimas de la unidad en la que interesa expresar el número redondeado. Para hacerlo a la unidad más próxima, y obtener el valor resultante, se determina si ese número de centésimas es menor o mayor que la mitad de la unidad de interés (kilogramo).

Si la fracción es menor de 0,50, la parte entera permanece sin cambio. Así, 64,38 se redondea a 64 (número que está más cerca de 64 que de 65).

Si la fracción es mayor de 0,50, la parte entera se aumenta en una unidad. Así, 69,67 se redondea a 70 (número más cercano a 70 que a 69).

Ahora bien, si la fracción es exactamente 0,50, el número está a igual distancia del entero inferior y del superior. En este caso no hay un criterio de proximidad definido e inequívoco para hacer el redondeo. ¿Entonces, cómo debe procederse para hacer el redondeo?

Para solucionar esta situación, se adopta la siguiente convención:

a) Si el dígito precedente a la fracción que se desea eliminar es *par*, el número entero permanece sin cambio. Así, 66,50 se redondea a 66.

b) Si el dígito precedente es *impar*, el número entero debe aumentarse en una unidad. Entonces, 63,50 se redondea a 64.

Note que la convención anterior evita cometer un error acumulativo de redondeo, el cual se produciría si el dígito que precede a un 5 "exacto" se deja siempre sin cambio o siempre se incrementa en una unidad.

REGLAS PARA EL REDONDEO "A LA UNIDAD MÁS PRÓXIMA"

- Si el primer dígito de la parte del número que se eliminará en el redondeo es menor que 5, el dígito precedente permanece sin cambio.
- Si la parte del número que serás eliminada en el redondeo es mayor que 5, el dígito precedente se incrementa en una unidad.
- Si el primer dígito de la parte por eliminar en el redondeo es "exactamente" 5, o sea un 5 seguido *únicamente* de ceros, el dígito precedente se aumenta en una unidad si es *impar* y se deja sin cambio si es *par*. Para estos efectos, el cero es considerado como par.

Seguidamente, se presentan algunas ilustraciones de la aplicación del método del redondeo "a la unidad más próxima".

Ejemplo 1

Se tiene el valor de 113 549 y se quiere redondear a cientos. En este caso, deben ser eliminadas las dos últimas cifras 4 y 9, equivalente a 0,49 cientos, que es menor que 0,50 cientos, por lo tanto el dígito precedente 5 se mantiene sin cambio. Valor redondeado: 1135 *cientos, que equivale a 113 500.*

◇◇

Ejemplo 2

Se tiene el número 48,501 y se quiere redondear eliminando los tres decimales. La parte que debe ser suprimida es 0,501 unidades, la cual es mayor que 0,500, por lo tanto, el dígito 8 se aumenta en una unidad. Valor redondeado: 49 *unidades*.

◇◇

Ejemplo 3

Se tienen los números 475 y 485 y se quieren redondear eliminando el último dígito. La parte que debe ser suprimida es exactamente media decena y para ello, en el primer número, el dígito precedente 7 se cambia a 8 y el 8 del segundo número se deja sin cambio. Valores redondeados: 48 *decenas* en ambos casos.

◇◇

EJEMPLO 4

Se quiere expresar en dos cifras, solamente, el valor numérico 20 500 000 (20 millones 500 mil). El dígito que precede al 5 es un cero (par). Valor redondeado: 20 *millones*.

◇◇

7.3.2. Redondeo hacia abajo

En este procedimiento, se conserva el último dígito que interesa y el resto del número se elimina. Así, por ejemplo, si se tiene 48,371 y se quiere prescindir de los tres decimales bajo el criterio de redondeo hacia abajo, el número redondeado es 48. Igualmente, si se tiene 48,999 el número redondeado será también 48.

Un ejemplo real de este tipo de redondeo se da cuando se pide a una persona que indique su edad en "años cumplidos" o al "último cumpleaños". En este caso, la persona indica el número de años completos que tiene, omitiendo los meses y los días. Con este criterio, una persona que tiene 21 años y 2 días, dice 21, y también dice 21 otra con 21 años, 11 meses, 29 días y 23 horas, aunque esté a punto de celebrar sus 22 años.

7.3.3. Redondeo hacia arriba

Bajo este procedimiento, el último dígito que se desea conservar siempre se incrementa en una unidad, excepto si va seguido únicamente de ceros. Ejemplos: 48,001 = 49; 48,371 = 49; 48,999 = 49; pero 48,000 se mantiene 48.

7.3.4. Algunos ejemplos de redondeos

Para concluir, resulta útil ilustrar el redondeo de varios números con los tres diferentes métodos que se acaban de explicar. En el primer bloque, el redondeo es a miles y el segundo a milésimas, según se muestra en el cuadro 7.5.

Cuadro 7.5
REDONDEO CON TRES MÉTODOS DIFERENTES

	REDONDEO USUAL	REDONDEO HACIA ABAJO	REDONDEO HACIA ARRIBA
24 351	24 000	24 000	25 000
24 500	24 000	24 000	25 000
24 892	25 000	24 000	25 000
25 000	25 000	25 000	25 000
25 001	25 000	25 000	26 000
25 383	25 000	25 000	26 000
25 500	26 000	25 000	26 000
25 776	26 000	25 000	26 000
0,00723	0,007	0,007	0,008
0,00749	0,007	0,007	0,008
0,00750	0,008	0,007	0,008
0,00799	0,008	0,007	0,008
0,00800	0,008	0,008	0,008
0,00801	0,008	0,008	0,009
0,00850	0,008	0,008	0,009

7.4. DISTRIBUCIÓN DE FRECUENCIAS DE VARIABLES CONTINUAS

Por lo general, las distribuciones de frecuencias para variables discretas son de muy fácil construcción; sin embargo, cuando se trata de variables continuas aparecen ciertas complicaciones y dificultades, ya que han de considerarse detalles o aspectos que deben ser tratados con especial atención y cuidado, tales como el número de clases por hacer y su amplitud, pautas por seguir al fijar los límites, entre otros. Las explicaciones siguientes se basan en un conjunto de datos correspondientes a una variable continua típica: el peso.

Como ilustración, considérense los siguientes datos referentes a los pesos, en kilogramos, de 60 estudiantes hombres, de último año de dos colegios privados.

Cuadro 7.6

PESOS EN KILOGRAMOS DE 60 ESTUDIANTES HOMBRES

63	71	53	66	63	88
53	84	70	62	71	70
60	67	72	60	52	67
75	55	61	52	77	64
64	60	56	45	61	62
64	57	61	56	68	65
55	53	52	73	80	84
61	86	87	67	55	75
79	57	75	62	62	56
68	64	63	60	65	62

Así como se presentan los datos, resulta difícil sacar alguna conclusión, por lo tanto, como no son muchos, puede iniciarse el análisis ordenándolos de menor a mayor.

Cuadro 7.7

PESOS EN KILOGRAMOS DE 60 ESTUDIANTES HOMBRES
ORDENADOS DE MENOR A MAYOR

45	56	61	63	67	75
52	56	61	64	68	75
52	56	61	64	68	77
52	57	62	64	70	79
53	57	62	64	70	80
53	60	62	65	71	84
53	60	62	65	71	84
55	60	62	66	72	86
55	60	63	67	73	87
55	61	63	67	75	88

El valor mayor es 88 y el menor es 45 y la amplitud general o recorrido es la diferencia entre ellos: 88 – 45 = 43. No existen concentraciones importantes alrededor de ningún valor específico y puede notarse que la mayoría de los valores está entre 55 y 70.

Una mejor idea acerca de los datos se obtiene a través de una distribución de frecuencias. Su construcción es relativamente sencilla si se tiene alguna experiencia y si se toman en cuenta ciertas reglas generales. Las principales dificultades consisten en decidir dos cosas: a) ¿dónde empezar la primera clase?, y b) ¿qué intervalo o amplitud usar para las clases? O, lo que es lo mismo, ¿cuántas clases definir? Una vez tomadas decisiones básicas, todos los demás pasos son relativamente simples y rutinarios.

El número de clases y su amplitud están íntimamente ligados, ya que cuando se decide emplear una determinada amplitud de clase, de hecho se fija el número de ellas y viceversa. Ahora bien, el número de clases que se haga, en un caso concreto, depende de la cantidad de observaciones disponibles y de la amplitud general (diferencia entre el valor mayor y el menor del conjunto). Si el número de observaciones es grande, el de clases debe ser elevado, siempre que no sea excesivo. Si los datos son pocos, no conviene hacer muchas clases, porque la distribución no mostrará una configuración definida, pero sí una forma irregular donde aparezcan clases con muy pocas observaciones y hasta algunas completamente vacías.

Además, en una situación como esta, el detalle es tan grande que se incumple uno de los fines básicos perseguidos al construir distribuciones de frecuencias: el de resumir la información. Por otra parte, hacer muy pocas clases, o definirlas muy amplias, conduce al problema contrario, es decir, a una pérdida excesiva de detalle. En el cuadro 7.8 se ilustran esas posibles situaciones con los datos de peso de los estudiantes, la primera emplea clases de 2 kg y la segunda de 15 kg.

Lo ideal es buscar un término medio, que brinde un balance óptimo entre la necesidad de resumir la información, por un lado, y la de mantener el detalle necesario para apreciar las características principales de los datos, por el otro. En general, se usa como un elemento de orientación la regla de que el número de clases no debe ser menor de 6 ni mayor de 15, aunque en circunstancias especiales puede hacerse un número de clases superior o inferior a estos límites citados.

En la práctica, el procedimiento usual es partir de la amplitud general y luego probar con diferentes números de clases o diferentes amplitudes hasta llegar a un número de clases y a una amplitud cómoda y adecuada. Se trata de que las clases sean todas de igual amplitud y, por comodidad, se prefiere que el intervalo sea 5, 10 o un múltiplo de estos números.

Cuadro 7.8

PESOS EN KILOGRAMOS DE 60 ESTUDIANTES HOMBRES
SEGÚN AMPLITUD DEL NÚMERO DE CLASE

(poca amplitud de clase)					(mucha amplitud de clase)	
Clase	Frecuencia	Clase	Frecuencia		Clase	Frecuencia
45-46	1	67-68	5		45-59	15
47-48	0	69-70	2		60-73	34
49-50	0	71-72	3		74-88	11
51-52	3	73-74	1			
53-54	3	75-76	3			
55-56	6	77-78	1			
57-58	2	79-80	2			
59-60	4	81-82	0			
61-62	9	83-84	2			
63-64	7	85-86	1			
65-66	3	87-88	2			
			60			60

Para el presente caso, en el que la amplitud general es de 43, al dividir entre números posibles de clases se obtiene:

$$\frac{43}{6} = 7,16 \text{ de amplitud;}$$
$$\frac{43}{8} = 5,38 \text{ de amplitud;}$$
$$\frac{43}{10} = 4,30 \text{ de amplitud.}$$

Esto sugiere que el intervalo de clase más apropiado es de 5 kg y, por tanto, el número de clases debe ser 9. Si se escogiera una amplitud de 10, por parecer más cómoda, el número de clases sería de 5, el cual tiene el inconveniente de que hace perder mucho detalle; si se hicieran 6 clases, la amplitud sería 7 kg, valor poco cómodo para los cálculos.

Respecto al punto de partida, coincide que 45 es el menor valor del conjunto y el hecho de que no se observan concentraciones importantes en números determinados, este valor 45 kg parece apropiado. Si se parte de 45 kg y se usa 5 como intervalo de clase, se tienen las siguientes clases:

$$45-50,$$
$$50-55,$$
$$55-60,$$
entre otras.

Sin embargo, este tipo de indicación de los intervalos de clase sería a todas luces ambiguo, ya que no se sabría, con respecto a los valores que coinciden con los límites de clase 50, 55, 60, entre otros, donde incluirlos. Para evitar esto, en la práctica, lo que se hace es indicar los límites en la forma siguiente, la cual elimina la ambigüedad citada:

45 – 49,
50 – 54,
55 – 59,
entre otras.

Después de aclarar algunos de los aspectos más significativos en relación con la construcción de distribuciones de frecuencias, puede procederse ahora a obtener la correspondiente a los datos de peso que se han comentando. Al igual que en el caso de las variables discretas, esto se hace con cada uno de los valores observados y se marca una línea vertical frente a la clase a la cual pertenece. Así, por ejemplo, el valor 57 que corresponde al peso del primer estudiante origina una marca frente a la clase 55-59 (tercera). Note, de nuevo, la ventaja de marcar siempre cada quinta línea en forma transversal cruzando a las otras cuatro (𝍸).

La distribución revela que la gran mayoría de los estudiantes se concentra en las tres categorías que van de 55 a 69, las cuales abarcan 36 de las 60 *observaciones*. También se aprecia que las frecuencias se reducen conforme se consideran las clases más alejadas del centro.

Cuadro 7.9

PESOS EN KILOGRAMOS DE 60 ESTUDIANTES HOMBRES SEGÚN CLASE Y FRECUENCIA

CLASES (peso en kg)	RECUENTO	FRECUENCIA
45-49	𝍷	1
50-54	𝍸 𝍷	6
55-59	𝍸 𝍷𝍷𝍷	8
60-64	𝍸 𝍸 𝍸 𝍸	20
65-69	𝍸 𝍷𝍷𝍷	8
70-74	𝍸 𝍷	6
75-79	𝍸	5
80-84	𝍷𝍷𝍷	3
85-89	𝍷𝍷𝍷	3
		60

Para terminar este punto, conviene señalar que algunas veces, ya sea por la naturaleza propia de los datos o por errores en el proceso para obtener de la información, se producen concentraciones muy marcadas alrededor de ciertos valores. Un caso típico de este fenómeno lo constituye la tendencia de la gente a redondear los pesos, edades, entre otros, en números terminados en 0 y en 5. Cuando esta situación se presenta, debe tratarse de que los puntos de concentración queden en el centro de la clase o punto medio. Así, por ejemplo, si los pesos que se han considerado mostraran esa tendencia a concentrarse en 0 y en 5, los intervalos de clase recomendables serían aquellos que se centren en 45, 50, 55, 60, y que originan las clases 43-47, 48-52, 53-57; con este procedimiento, los errores que se originan en la tendencia a informar los pesos redondeados, en números terminados en 0 y 5, tienen poco efecto, pues estudiantes que, por ejemplo, informan un peso de 55 kg, cuando realmente pesan 56, 57, 53 o 54 quedan dentro de la misma clase, 53-57, en la cual habrían quedado si hubieran informado el peso correctamente.

7.5. LÍMITES REALES Y LÍMITES INDICADOS, INTERVALO DE CLASE Y PUNTO MEDIO

Antes de seguir adelante, es útil aclarar algunos conceptos que son de gran importancia en el manejo y análisis de las distribuciones de frecuencias.

7.5.1. Límites de clase

Son los valores que definen una clase separándola de la anterior y de la posterior. Los límites deben ser tales que definan clases que sean *exhaustivas* –permitan clasificar a todas las observaciones en alguna de ellas– y *mutuamente excluyentes* –no permitan que una observación quede incluida en más de una clase–.

Con respecto a los límites de clase, es conveniente distinguir entre los *límites indicados* y los *límites reales* o verdaderos. Los *indicados* son los que aparecen en la distribución, y los *reales* son aquellos que señalan la verdadera extensión de la *clase*, o sea, los valores de la característica de interés que abarca realmente. La distribución es importante porque si bien es cierto que con frecuencia los límites indicados son precisamente los reales, también es común que eso no suceda.

En el caso del peso de los estudiantes, los límites indicados son 45-49, 50-54, 55-59, entre otros; pero como el peso es una variable continua, obviamente esos límites no señalan la verdadera extensión de las clases. Si estos fueran los límites reales, se tendría una discontinuidad entre 49 y 50, otra entre 54 y 55 y así sucesivamente, además se daría la impresión de que no existen pesos entre 49 y 50 kg, entre 54 y 55 kg, etc. Como los pesos fueron redondeados al kilogramo más próximo, un estudiante que aparece con 54 kg puede haber arrojado cualquier peso entre 53,5 y 54,5; igualmente, el que aparece con 55 kg puede tener alguno entre 54,5 y 55,5.

Por este motivo, para obtener los límites reales del peso de los estudiantes a partir de los indicados, y en cualquier otro caso de variables continuas redondeadas por el sistema usual –debe ajustarse el efecto del redondeo sobre los límites indicados– se agrega media unidad al límite superior y se resta media unidad al inferior, como se ilustra seguidamente:

Indicados: 45 49 50 54 55 59 60

Reales: 44,5 49,5 54,5 59,5

Entonces, los límites reales correspondientes a los indicados en el caso del peso son:

LÍMITES INDICADOS	LÍMITES REALES
45 – 49	44,5 – 49,5
50 – 54	49,5 – 54,5
55 – 59	54,5 – 59,5
etc.	etc.

El lector podría pensar que al colocar el mismo valor, por ejemplo, 49,5 como límite superior de una clase y como inferior de la siguiente, se plantea una ambigüedad, ya que no se sabe en qué clase debe situarse 49,5; en rigor, esto es correcto, pero como los datos que se clasifican (los pesos) vienen redondeados en kilogramos enteros, en la práctica no se presenta esa ambigüedad, ya que todos los pesos pueden clasificarse inequívocamente. Si los datos estuvieran redondeados a décimas de kilogramo, entonces se fijan los límites reales en centésimas de kilogramo, según el procedimiento antes indicado.

Lo anterior señala una clara relación de los límites reales con el criterio de redondeo utilizado en los datos y llama la atención a la necesidad de tomar en cuenta esta circunstancia al derivar los reales a partir de los indicados. Para ilustrar este punto, considérese una distribución de frecuencias de personas según la edad, en años, con los siguientes límites indicados: 10-14, 15-19, 20-24, entre otros. ¿Cuáles son los límites reales de esta distribución?

Depende del criterio con el que fueron redondeadas las edades clasificadas, tal como puede observarse a continuación.

Cuadro 7.10

DISTRIBUCIÓN DE FRECUENCIAS DE PERSONAS SEGÚN LA EDAD, EN AÑOS

LÍMITES INDICADOS	MÉTODO USUAL (Cumpleaños más cercano)	HACIA ABAJO (Edad cumplida)	HACIA ARRIBA (Próximo cumpleaños)
10 – 14	9,5 – 14,5	10 a menos de 15	más de 9 a 14
15 – 19	14,5 – 19,5	15 a menos de 20	más de 14 a 19
20 – 24	19,5 – 24,5	20 a menos de 25	más de 19 a 24
etc.	etc.	etc.	etc.

En el caso del redondeo hacia abajo, los límites reales del segundo intervalo van de 15 a menos de 20, porque una persona que declara tener 15 años realmente puede tener cualquier edad exacta entre 15 y 16 años, sin incluir el 16; y la que declara 19 puede tener cualquier edad entre 19 y 20, sin incluir esta última. La persona, al indicar su edad, la reduce a unidades completas, desprecia la fracción del año. En los censos decenales de población, se investiga en esta forma por medio de la pregunta "¿Cuál es su edad en años cumplidos?"

Respecto al redondeo hacia abajo, la persona indica siempre la edad siguiente a la que tiene, excepto cuando está cumpliendo años, cuando indicará precisamente la que cumple. Por ello, los límites, por ejemplo, van de más de 9 a 14 años, ya que quienes informan de 10 años están realmente entre 9 y 10, sin incluir 9 y los que informan 14, entre 13 y 14 sin incluir 13.[3]

7.5.2. Intervalo de clase

Indica la *amplitud de clase*; se calcula a partir de la diferencia entre el *límite real superior y el límite real inferior* de una clase. Como las más usadas son las distribuciones en clases de igual amplitud, en la práctica, el intervalo es usualmente uniforme en una distribución. Sin embargo, en circunstancias especiales, como en el caso del ingreso, por ejemplo, la naturaleza de la variable hace que se usen clases desiguales y, consecuentemente, la amplitud varía. En el ejemplo del peso de los estudiantes comentado, la distribución tiene clases de amplitud uniforme igual a 5.

3. En Costa Rica, en el pasado, fue frecuente que los niños y los padres, cuando se preguntaba la edad del niño, la dieran redondeando hacia arriba y respondiendo "Entrado en…". Así, si una niña tenía 14 años cumplidos, se decía "entrada en 15".

7.5.3. Punto medio

Se le da este nombre al *valor central de la clase*. Se obtiene al calcular el *promedio* de los *límites reales*, pero puede hacerse también sumando la mitad del intervalo de la clase al límite real inferior. Para el caso de los pesos de los estudiantes que se han considerado, por ejemplo, el punto medio de la primera puede obtenerse de dos formas diferentes:

$$\frac{44,5+49,5}{2}=\frac{94}{2}$$

$$44,5-\frac{5}{2}=44,5+2,5=47.$$

Además, como todas las clases son de igual amplitud, los puntos medios de la clase restante pueden obtenerse al sumar repetidamente el intervalo al punto medio de la clase anterior 47 + 5 = 52, 52 + 5 = 57, así sucesivamente. Si se sigue este procedimiento, es conveniente calcular de nuevo el punto medio de la última clase, con la definición, para asegurarse de no cometer algún error en este proceso.

El punto medio cumple una función muy importante en la distribución de frecuencias y es la de representar a la clase cuando se desea realizar ciertos cálculos. Al agrupar los datos en una distribución de frecuencias, se gana en cuanto a claridad para el análisis e interpretación de los datos y en comodidad para su manejo, pero se pierde en detalle. Se conoce que en una clase hay un determinado número de observaciones (la frecuencia), pero no se sabe cuál es el valor específico de cada una de esas observaciones; esta fue la concesión que se hizo al agrupar los datos: perder la información individual.

Para los cálculos posteriores, ante la ausencia de información para cada una de las observaciones, se acostumbra suponer que el punto medio es el valor más adecuado para representarlas.

De esta manera, si en una clase hay por ejemplo 10 observaciones, para los efectos correspondientes se supone que todas son iguales al punto medio; de esta función, que cumple el punto medio, como valor representativo de la clase, se deriva su importancia.

Si examina el cuadro 7.11, se observa que a la clase 79,5 – 84,5 le corresponde el punto medio 82 y la frecuencia absoluta respectiva es 3; esto quiere decir que tres alumnos tienen un peso entre 79,5 y 84,5 kg. Ahora bien, para efectos prácticos de cálculos con la distribución de frecuencias, suponga que a cada uno de esos tres alumnos le corresponde un peso de 82, cuando en realidad pesan 80, 84 y 84 kg, como puede observarse al revisar los datos de los pesos de los estudiantes presentados en los cuadros 7.6 y 7.7. Esta aproximación evidentemente introduce cierto error, ya que, por ejemplo, la suma verdadera de los tres pesos es 248 kg, mientras que 3 · 69,5 = 238,5. Sin embargo, se sabe que las diferencias producidas en un sentido, en una clase, también ocurren en el inverso en otra, haciendo que los errores se compensen y globalmente el error sea despreciable.

7.5.4. Clases abiertas

Algunas veces se tienen datos que se apartan mucho de la mayoría, hacia arriba o hacia abajo, y querer incluirlos dentro de una distribución con clases de igual amplitud obligaría a tener clases intermedias vacías. Para evitarlo se recurre a las clases abiertas, las cuales se ubican al principio o al final de la distribución. En el cuadro 7.1, hay una clase abierta al principio: "menos de 500" y otra al final, "4500 y más".

Las clases abiertas resuelven problemas especiales de clasificación, pero al desconocerse el límite inferior y el superior de las clases extremas, no permiten el cálculo del punto medio ni de los intervalos de clase y obligan, por lo tanto, a usar supuestos en cierta medida arbitrarios cuando se representan y analizan los datos. Por ello, es mejor evitar su uso, excepto cuando sea totalmente necesario.

7.6. FRECUENCIAS ABSOLUTAS Y RELATIVAS, SIMPLES Y ACUMULADAS

Cuando se estudiaron las distribuciones de variables discretas, se definieron e interpretaron las frecuencias absolutas y relativas, las simples y las acumuladas. Sin embargo, como las variables continuas presentan ciertos detalles que hacen su análisis más complejo y cuidadoso, es conveniente hacer una revisión de esos conceptos e ilustrarlos luego con los datos del peso de los estudiantes.

7.6.1. Frecuencia absoluta

Se define como el número de elementos u observaciones pertenecientes a una misma clase. Normalmente, se le llama frecuencia.

7.6.2. Frecuencia relativa

La frecuencia relativa de una clase se obtiene al dividir la frecuencia absoluta por el número total de observaciones. Indica la importancia relativa de la clase. Es muy conveniente expresarla en términos de porcentajes, es decir, con el cociente: frecuencia absoluta/número de observaciones, multiplicado por 100. Las frecuencias relativas, precisamente por estar en términos relativos, facilitan mucho el análisis de los datos y son en especial útiles cuando se tienen que comparar varias distribuciones de frecuencias basadas en diferentes números de observaciones.

7.6.3. Frecuencias acumuladas

Para ciertos fines resulta útil conocer, además de la frecuencia simple de una clase, el número de observaciones que son mayores o menores que uno de sus límites, esto es, la frecuencia acumulada. Para obtener las frecuencias acumuladas, se procede a sumar las frecuencias absolutas (o relativas) en sentido ascendente o descendente, según se quieran acumular "hacia arriba" o "hacia abajo". Las frecuencias acumuladas resultan útiles, como se verá más adelante, para el cálculo de la mediana y de las otras medidas denominadas percentiles y, en general, para responder a preguntas como: ¿cuántos alumnos sacaron menos de 7 en el examen? ¿Qué proporción de los empleados gana más de 500 000 colones? ¿Cuántos estudiantes pesan menos de 60 kg?

Seguidamente se presenta la distribución de frecuencias correspondientes al peso de los 60 estudiantes; como puede apreciarse, se dan los límites reales y se incluyen, además de varios tipos de frecuencias, los puntos medios.

Cuadro 7.11

CLASIFICACIÓN DE LOS 60 ESTUDIANTES DE DOS COLEGIOS DE ACUERDO CON SU PESO EN KILOGRAMOS

CLASES	PUNTOS MEDIOS (xi)	FRECUENCIA Absoluta (fi)	FRECUENCIA Relativa (fr)	ACUMULADA "menos de" Absoluta	ACUMULADA "menos de" Relativa	ACUMULADA "más de" Absoluta	ACUMULADA "más de" Relativa
44,5 – 49,5	47	1	0,017	1	0,017	60	1,000
49,5 – 54,5	52	6	0,100	7	0,117	59	0,983
54,5 – 59,5	57	8	0,133	15	0,250	53	0,883
59,5 – 64,5	62	20	0,333	35	0,583	45	0,750
64,5 – 69,5	67	8	0,133	43	0,717	25	0,417
69,5 – 74,5	72	6	0,100	49	0,817	17	0,283
74,5 – 79,5	77	5	0,083	54	0,900	11	0,183
79,5 – 84,5	82	3	0,050	57	0,950	6	0,100
84,5 – 89,5	87	3	0,050	60	1,000	3	0,050
TOTAL		60	1,000				

Es conveniente interpretar algunos de los valores de la distribución anterior. Esto eliminará cualquier duda sobre el significado de las diferentes columnas y dará, además, una idea de la utilidad de los diferentes tipos de frecuencias. Por ejemplo: en la quinta clase, el número 8, en la columna de las frecuencias, indica que 8 estudiantes tienen pesos entre 64,5 y 69,5 kg; el número 0,133 que un 13,3% pesa entre 64,5 y 69,5 kg; los números 43 y 0,717 indican, respectivamente, que 43 o el 71,7% pesa menos de 69,5 kg. El 25 y el

0,417 correspondientes a las acumuladas "más de" indican que 25, o sea 41,7% del total, pesan más de 64,5 kg.

Finalmente, para terminar estos comentarios sobre construcción de las distribuciones de frecuencias, es útil resumir algunas reglas básicas que conviene tener presentes.

- Si las observaciones no son muchas, puede resultar innecesario construir una distribución de frecuencias y será suficiente ordenar los datos por magnitud creciente.
- Las clases deben ser exhaustivas y mutuamente excluyentes.
- Debe procurarse, como regla general, que el número de clases no sea menor que 6 ni mayor a 15.
- Siempre que sea posible, evite las clases de diferente amplitud y también las clases abiertas.
- Si hay valores alrededor de los cuales existen concentraciones de los datos, es recomendable que se tornen como puntos medios.

7.7. LA REPRESENTACIÓN GRÁFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS: HISTOGRAMAS, POLÍGONOS Y OJIVAS

El patrón de variación de los datos puede apreciarse mejor si se representa gráficamente la distribución de frecuencias. Esto no solo contribuye a un mejor análisis de los datos, sino que, también, facilita la comprensión del fenómeno considerado por un mayor número de personas.

En la presente sección se considerarán los tipos de gráficos que se utilizan para representar las distribuciones de frecuencias: ***histogramas*** y ***polígonos de frecuencias*** para las simples, y ***ojivas*** para las acumuladas.

7.7.1. Histograma

Un histograma es un gráfico de barras verticales en el que, a diferencia de los descritos anteriormente (capítulo 6), las barras no guardan separación entre sí, además pueden tener diferente anchura.

Para construir un histograma, se define una escala horizontal apropiada y en ella se marcan los límites reales de todas las clases de la distribución que se quiere representar.

La escala no necesita empezar en cero, pero sí un intervalo de clase antes del límite inferior de la clase más baja. Las frecuencias se representan en la escala vertical, la cual debe empezar en cero, no tener "cortes" o interrupciones y ser lo suficientemente amplia para incluir la mayor de las frecuencias de la distribución. La combinación de las dos escalas debe producir un rectángulo en el que se dé una relación aproximada de 1,5 a 1 entre la base y la altura, tal como se recomendó en el capítulo 6.

Definidas las escalas, se procede a dibujar, sobre el intervalo correspondiente a cada clase, una barra (rectángulo) cuya área sea proporcional a la frecuencia de esa clase. Esto significa que si una clase tiene una frecuencia igual a 1, y se le asigna una barra de área de 20 mm^2, a una clase cuya frecuencia sea 3 le corresponderá una barra de área de 60 mm^2.

Para ilustrar la construcción del histograma, considere la distribución del peso de los 60 estudiantes hombres, comentada y expuesta en el cuadro 7.8.

Como se trata de una distribución que tiene clases de igual amplitud, la construcción de barras de área proporcional a la frecuencia de la clase se reduce simplemente a cuidar que la altura de la barra sea igual a la frecuencia, ya que al tener todas las mismas bases, las áreas serán automáticamente proporcionales a la frecuencia. La observación del gráfico de la figura 7.3, en el que aparece el histograma para la distribución de los pesos de los estudiantes, por ejemplo, muestra que la barra que corresponde a la sexta clase, que tiene una frecuencia de 3, es tres veces más alta y tiene el triple del área de la barra de la quinta clase, cuya frecuencia absoluta es 1. De igual forma, la barra correspondiente a la segunda clase (frecuencia de 8) tiene una altura y un área que es el doble de la correspondiente a la primera clase (frecuencia de 4).

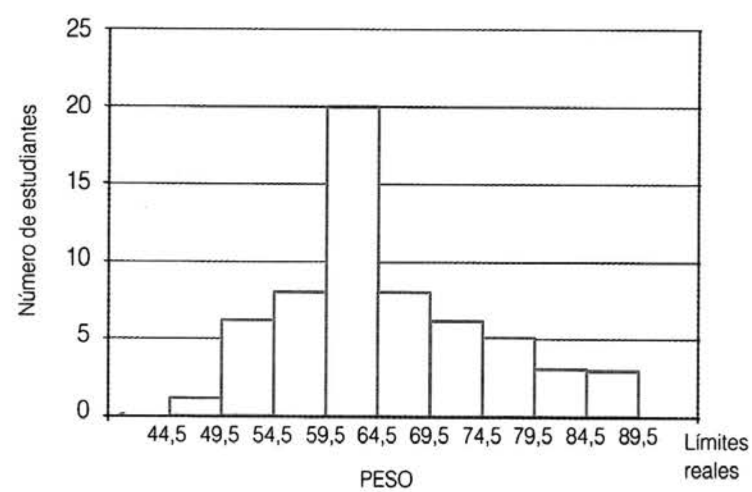

Figura 7.3. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios privados

7.7.2. Polígono de frecuencias

Una forma alternativa de representar la distribución de frecuencias consiste en marcar, sobre cada clase, un punto; se toma como *abscisa* el punto medio de la clase y como ordenada la frecuencia. Esos puntos se unen luego con secciones de rectas y la figura resultante se denomina *polígono de frecuencias*. Como ilustración, se incluye el polígono correspondiente a los datos del peso de los 32 estudiantes de un colegio (gráfico de la figura 7.4).

El área bajo el polígono debe ser igual al área comprendida bajo el histograma; para lograr esto, corrientemente el polígono se prolonga, tal como puede apreciarse en el gráfico 7.6, se procede como si existiera una clase adicional al principio y otra al final, ambas con frecuencia cero.

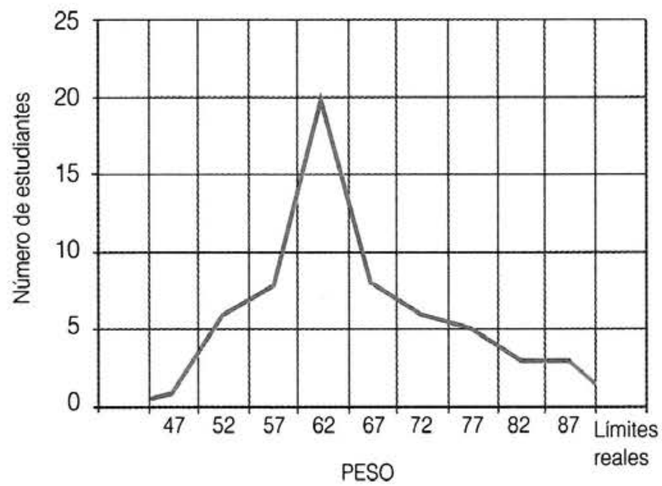

Figura 7.4. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios

La correspondencia entre las áreas bajo el polígono y el histograma se aprecia con mayor claridad en el gráfico de la figura 7.5. El área sombreada indica las partes del histograma y del polígono donde se produce la compensación.

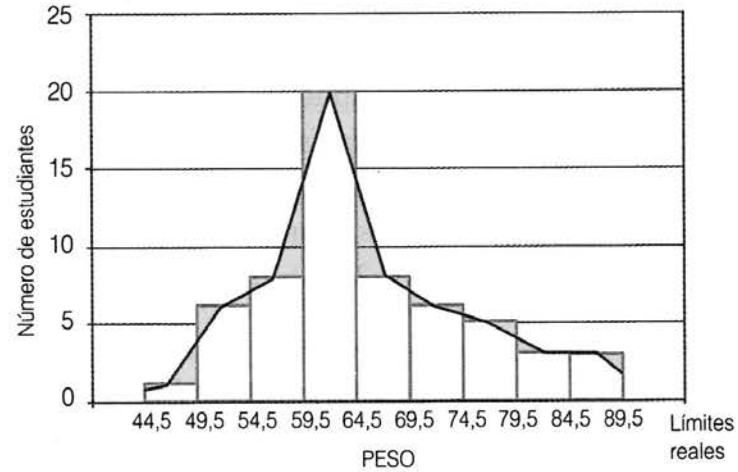

Figura 7.5. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios

7.7.3. Representación de distribuciones con intervalos desiguales

Cuando las clases son de igual amplitud, la construcción del histograma no plantea ningún problema especial, así resulta fácil cumplir con el principio de que "a cada clase debe corresponder una barra de área proporcional a su frecuencia"; para esto, lo único que debe ser considerado es que la altura de la barra sea igual a la frecuencia de la clase. Este procedimiento, sin embargo, no es correcto cuando se tienen intervalos desiguales, pues de aplicarse produciría un histograma que distorsionaría la verdadera distribución de los datos.

Para ilustrar este punto de la representación de distribuciones con intervalos desiguales, considere la distribución del peso de los 60 estudiantes de un colegio y suponga que es modificada, uniéndose en una sola clase las dos primeras y en otra la cuarta, quinta y sexta. Hecho esto, la nueva distribución de los pesos quedaría como se presenta en el cuadro 7.12.

Cuadro 7.12

DISTRIBUCIÓN CON INTERVALOS DESIGUALES DEL PESO DE 60 ESTUDIANTES DE UN COLEGIO

CLASES	FRECUENCIA	INTERVALO DE CLASE
44,5 – 54,5	7	10
54,5 – 59,5	8	5
59,5 – 64,5	20	5
64,5 – 69,5	8	5
69,5 – 84,5	14	15
84,5 – 89,5	3	5

Si esta distribución se representa, dándole a cada clase una barra de altura igual a su frecuencia, se obtiene el siguiente histograma:

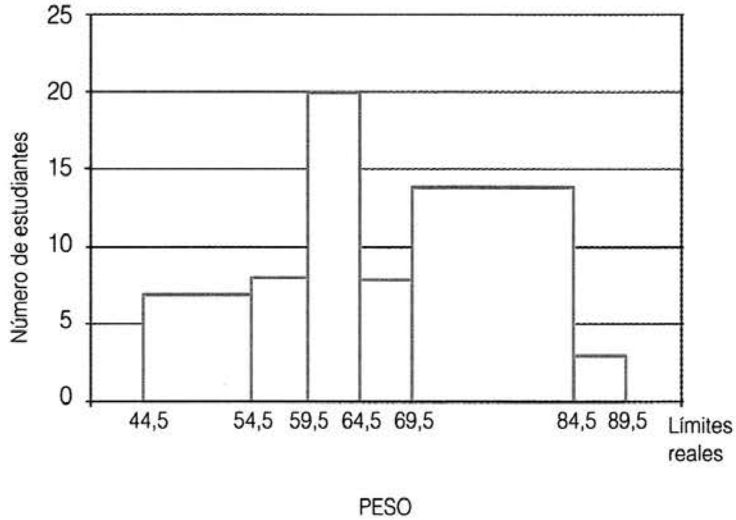

Figura 7.6. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios privados

No cabe duda que el gráfico de la figura 7.6 produce una idea engañosa de la distribución de los pesos de los estudiantes. En primer lugar, se obtiene la impresión de que entre 69,5 y 84,5 kg se encuentra casi la mitad de las observaciones, y no solo 14, como sucede realmente; en segundo lugar, el histograma da la idea de que entre 44,5 y 54,5 (primera clase) hay casi el doble de estudiantes que entre 54,5 y 59,5. Estas equivocaciones se originan en la construcción del gráfico, cuando no se respeta el principio tantas veces aludido de que cada clase debe tener una barra proporcional a su frecuencia; y el lector, al analizar el histograma, lo que hace es comparar en forma intuitiva las áreas de las barras y no sus alturas.

El problema antes citado se resuelve al modificar la altura de las barras, para que el área sea proporcional a la frecuencia. Una forma es calcular la *densidad de frecuencia*, o sea, dividir la frecuencia de la clase entre el intervalo y representar esos valores. Pero si todos los intervalos de clase son múltiplos del intervalo más pequeño, resulta más conveniente y cómodo tomar este como referencia y ajustar solo aquellas clases cuya amplitud es mayor. En la distribución que se considera, por ejemplo, la quinta clase tiene amplitud 15 y la primera, 10; la amplitud de todas las demás es de 5. Es apropiado, entonces, tomar el intervalo de 5 como referencia y ajustar sólo las frecuencias de la primera y quinta clases. Esto se hace al dividir entre 3 la frecuencia de la quinta, porque su amplitud de clase es triple $\left(\frac{14}{3} = 4,7\right)$ y entre 2 la primera $\left(\frac{7}{2} = 4,7\right)$. Hecho esto, se procede a construir el histograma en la forma usual (gráfico de la figura 7.7). Del examen del histograma se aprecia cómo, al ajustar las frecuencias de la cuarta y sexta clase, se obtiene una visión correcta de la distribución de los pesos de los 60 estudiantes.

Es importante advertir que, en estos casos de representación de distribuciones con intervalos desiguales, deben ajustarse las frecuencias y que bajo ningún concepto debe intentarse modificar la escala horizontal donde se marcaron los límites reales.

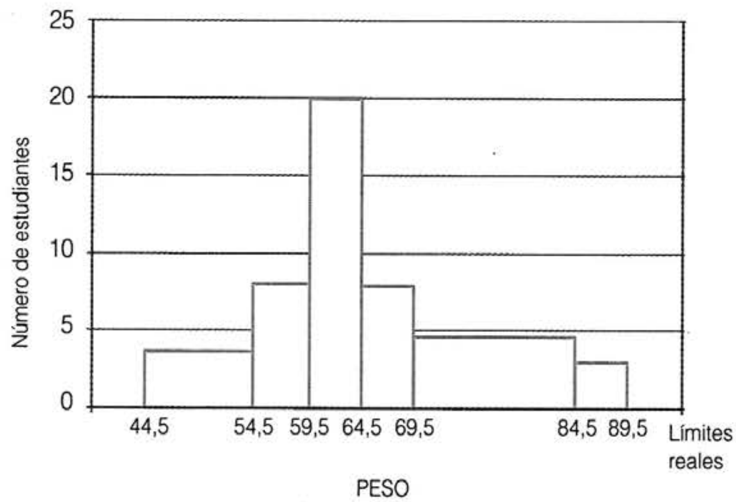

Figura 7.7. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios (frecuencias ajustadas para corregir efecto de clases desiguales)

7.7.4. Las "ojivas" o polígonos de frecuencias acumuladas

Las frecuencias acumuladas se representan usualmente mediante polígonos. En su construcción, se siguen los mismos procedimientos y criterios explicados para el caso de los polígonos de frecuencias simples; la única excepción es que la ordenada no se levanta sobre el punto medio de la clase, sino sobre el límite superior o sobre el inferior, según se tenga una frecuencia acumulada "menos de" o "más de". Esto se hace porque, debido al procedimiento de acumulación, la frecuencia "menos de", para una cierta clase, incluye todas las frecuencias *menores que* el límite superior de esa clase; y la acumulada "más de", todas las frecuencias *mayores que* el límite inferior de la clase.

Para ilustrar la técnica, se empleará la acumulada "menos de" correspondiente a la distribución de los pesos de los 60 estudiantes. Los datos se incluyen seguidamente y la representación se hace en el gráfico de la figura 7.8. Note que, como la frecuencia acumulada "menos de" se obtiene al sumar la frecuencia absoluta de la clase en consideración y la de todas las anteriores, y por ello se relaciona con el límite superior de la clase, es posible expresarla en una forma resumida como se muestra a la derecha de la distribución. Otra observación importante es que para dibujar la parte inicial del polígono (45,5 a 49,5) se supone, como es correcto, que la frecuencia acumulada correspondiente a 45,5 es 0.

examen del histograma se aprecia cómo, al ajustar las frecuencias de la cuarta y sexta clase, se obtiene una visión correcta de la distribución de los pesos de los 60 estudiantes.

Es importante advertir que, en estos casos de representación de distribuciones con intervalos desiguales, deben ajustarse las frecuencias y que bajo ningún concepto debe intentarse modificar la escala horizontal donde se marcaron los límites reales.

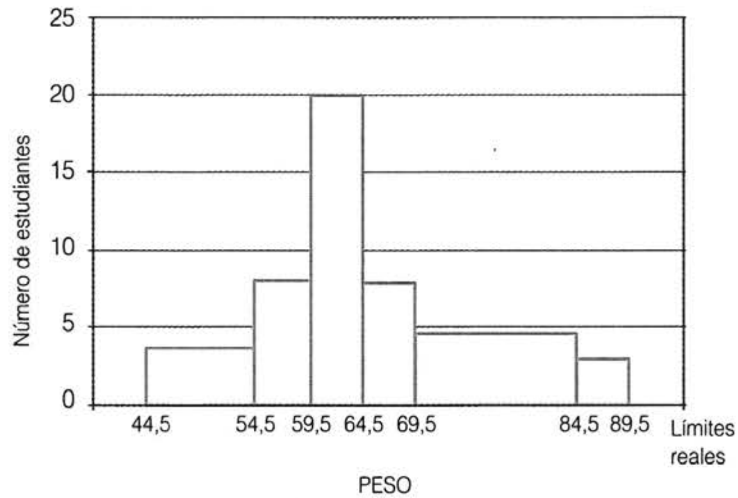

Figura 7.7. Gráfico que muestra el peso en kilogramos de 60 estudiantes de dos colegios (frecuencias ajustadas para corregir efecto de clases desiguales)

7.7.4. Las "ojivas" o polígonos de frecuencias acumuladas

Las frecuencias acumuladas se representan usualmente mediante polígonos. En su construcción, se siguen los mismos procedimientos y criterios explicados para el caso de los polígonos de frecuencias simples; la única excepción es que la ordenada no se levanta sobre el punto medio de la clase, sino sobre el límite superior o sobre el inferior, según se tenga una frecuencia acumulada "menos de" o "más de". Esto se hace porque, debido al procedimiento de acumulación, la frecuencia "menos de", para una cierta clase, incluye todas las frecuencias *menores que* el límite superior de esa clase; y la acumulada "más de", todas las frecuencias *mayores que* el límite inferior de la clase.

Para ilustrar la técnica, se empleará la acumulada "menos de" correspondiente a la distribución de los pesos de los 60 estudiantes. Los datos se incluyen seguidamente y la representación se hace en el gráfico de la figura 7.8. Note que, como la frecuencia acumulada "menos de" se obtiene al sumar la frecuencia absoluta de la clase en consideración y la de todas las anteriores, y por ello se relaciona con el límite superior de la clase, es posible expresarla en una forma resumida como se muestra a la derecha de la distribución. Otra observación importante es que para dibujar la parte inicial del polígono (45,5 a 49,5) se supone, como es correcto, que la frecuencia acumulada correspondiente a 45,5 es 0.

CLASES	Frecuencias absoluta	Frecuencias acumuladas "menos de"
44,5 – 49,5	1	1
49,5 – 54,5	6	7
54,5 – 59,5	8	15
59,5 – 64,5	20	35
64,5 – 69,5	8	43
69,5 – 74,5	6	49
74,5 – 79,5	5	54
79,5 – 84,5	3	57
84,5 – 89,5	3	60
TOTAL	60	

Puede expresarse como:

Menos de 44,5	0
Menos de 49,5	1
Menos de 54,5	7
Menos de 59,5	15
Menos de 64,5	35
Menos de 69,5	43
Menos de 74,5	49
Menos de 79,5	54
Menos de 84,5	57
Menos de 89,5	60

Como puede observarse el polígono resultante, al representar las frecuencias acumuladas "menos de", tiene una forma de S alargada y, por ello, se le denomina corrientemente "ojiva". Este mismo nombre se le aplica también al gráfico de las frecuencias acumuladas "más de".

Una característica de las ojivas, que quizás haya sido notada por el lector pero conveniente de señalar, es que, como están ligadas a los límites de clase, no se ven afectadas por el hecho de que la distribución tenga clases desiguales.

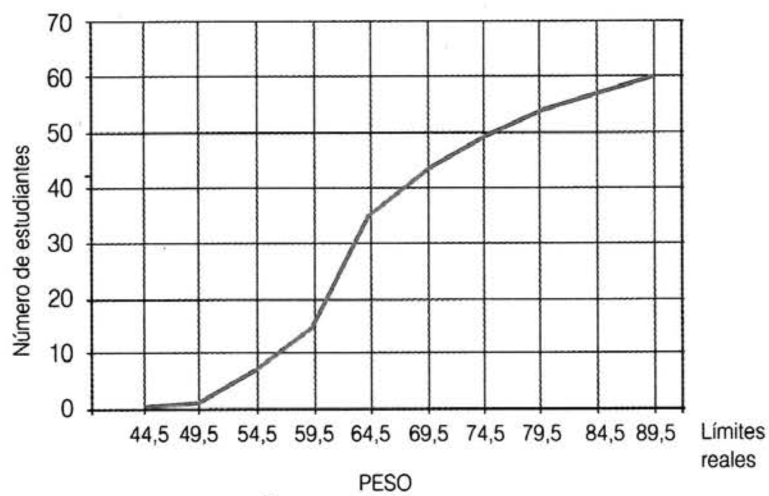

Figura 7.8. Gráfico que muestra la frecuencia acumulada "menos de" correspondiente al peso en kilogramos de 60 estudiantes de dos colegios

EJERCICIOS DE AUTOEVALUACIÓN

1. Para el análisis de un conjunto de datos estadísticos, ¿de qué depende que se utilice una distribución de frecuencias?
2. Señale tres propiedades de los conjuntos de datos cuyo conocimiento se facilita mucho construyendo una distribución de frecuencias.
3. De acuerdo con los tres criterios de redondeo, proceda a redondear las cifras siguientes de manera que pierdan los dos últimos dígitos.

Datos	Redondeo usual	Redondeo hacia abajo	Redondeo hacia arriba
42,542			
24,35			
8,7532			
4,278			
23,839			
1485,01			
31,457			
317 547			
253,12			

4. Defina los siguientes conceptos:
   a) Clase
   b) Intervalo de clase
   c) Punto medio de clase
   d) Frecuencia absoluta
   e) Frecuencia relativa
   f) fFrecuencias acumuladas
5. Señale la importancia del punto medio en una distribución de frecuencias.

6. Diga qué tipo de frecuencias (absoluta, relativa, simple o acumulada) se utilizó para hacer cada una de las siguientes afirmaciones:
   a) Un 26% de los estudiantes tiene notas entre 65 y 70.
   b) Un 40% de los estudiantes tiene notas inferiores a 80.
   c) 29 alumnos pasaron el curso (obtuvieron una nota igual o mayor a 70).
   d) 3 alumnos obtuvieron notas entre 85 y 90.
   e) Un 40% de los alumnos reprobó el curso.

7. Se tiene la siguiente distribución de frecuencias en la que no aparece toda la información:

CLASES	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA "más de"	PUNTOS MEDIOS
17 – 24				21
25 – 28	4			27
29 – 32		0,20		31
33 – 36			17	35
37 – 40			9	39
41 – 52	4			47
TOTAL	30			

   a) Rellene correctamente los espacios vacíos
   b) Obtenga los límites reales
   c) Obtenga los intervalos de clase

8. Una muestra de 74 familias de un cantón rural de San José fue investigada para conocer el consumo mensual de energía eléctrica. Seguidamente se dan las cifras en kWh, correspondientes al mes de diciembre del 2008.

274	191	139	389	189	87
186	194	126	322	224	207
49	101	140	167	151	76
249	174	117	295	201	105
76	420	232	357	123	269
315	175	243	76	130	388
194	125	215	246	71	139
113	27	70	121	161	205
121	178	277	348	93	262
263	128				

a) Construya una distribución de frecuencias con intervalos de clase: 20 – 59, 60 – 99, 20 – 59, 100 – 139, etc., según el cuadro siguiente:

CLASES	PUNTOS MEDIOS (xi)	FRECUENCIA		ACUMULADA "menos de"		ACUMULADA "más de"	
		Absoluta (fi)	Relativa (fr)	Absoluta	Relativa	Absoluta	Relativa

b) Responda a las siguientes preguntas:

1. ¿Cuántas familias consumen menos de 300 kWh mensuales en energía eléctrica?
2. ¿Cuántas familias consumen más de 100 kWh?
3. ¿Cuántas familias consumen entre 100 y 220 kWh?

9. Un examen parcial realizado en un curso universitario arrojó las siguientes notas:

5,5 6,1 2,2 8,8 7,7 6,0 7,7 4,0 5,7 4,4 5,7 5,0

4,1 5,1 5,7 7,5 9,8 5,4 7,7 6,0 8,1 4,4 4,1 7,5

6,2 5,5 5,4 6,2 4,8 5,7 2,7 3,8 6,7 7,3 5,5 4,8

4,8 8,4 6,1 6,4 6,8 8,3 6,9 6,2 2,7 3,4 2,3 7,0

a) Construya una distribución de frecuencias, con intervalos de clase iguales, de amplitud un punto (suponga que las notas fueron redondeadas a la décima más próxima, o sea, siguiendo el método usual).

b) Calcule los puntos medios y las frecuencias relativas.

c) Obtenga las frecuencias acumuladas "menos de", represéntelas gráficamente y luego estime, usando el gráfico, cuántos alumnos tuvieron una nota inferior a 7 y cuántos una nota superior a 4,5.

d) Construya un histograma empleando las frecuencias relativas.

e) Represente en un mismo gráfico las ojivas "más de" y "menos de" (utilice las frecuencias relativas) y, con base en este, diga cuál fue la nota "mediana"; es decir, una nota tal que aproximadamente la mitad de los alumnos tiene una nota menor y la otra mitad una superior a ella.

10. Con gran frecuencia se presenta el problema práctico de comparar distribuciones de la misma característica, pero correspondientes a diferentes momentos o lugares, con el fin de apreciar si se han dado cambios o si existen discrepancias. Como ejemplo, considere los datos sobre suicidios, clasificados según edad, que se incluyen abajo. Suponga que se desea determinar, gráficamente, si se han dado cambios en el patrón del fenómeno según la edad, entre 1998 y el 2008. Para ello, usted deberá:

    a) Construir un gráfico que permita investigar el fenómeno.

    b) Analizar el gráfico y escribir las conclusiones respectivas.

    Antes de iniciar la construcción del gráfico esté seguro de comprender claramente lo que se quiere y no olvide que el propósito básico es comparar las dos series de datos.

GRUPOS DE EDADES	SUICIDIOS TOTALES REGISTRADOS	
	1998	2008
5 – 14	6	3
15 – 24	48	82
25 – 34	63	67
35 – 44	53	74
45 – 54	28	60
55 – 64	14	25
65 y más	11	14
TOTAL	223	325

Fuente: Instituto Nacional de Estadística y Censos (INEC). Estadísticas Vitales de 1998 y 2008.

RESPUESTA A LOS EJERCICIOS DE AUTOEVALUACIÓN

1. El que se utilice una distribución de frecuencias para el análisis de un conjunto de datos depende de la cantidad de los datos, esto quiere decir que cuando son muy numerosos lo más adecuado para su análisis es agruparlos en una distribución de frecuencias.

2. Las propiedades acerca de las cuales la distribución de frecuencias proporciona información son:

   a) La forma o patrón de distribución de los datos.

   b) La posición de la distribución, o sea, alrededor de qué valor se tienden a concentrar los datos (valores centrales).

   c) La dispersión de los datos alrededor de los valores centrales o promedios (variabilidad).

3.

Datos	Redondeo usual	Redondeo hacia abajo	Redondeo hacia arriba
42,542	42,5	42,5	42,6
24,35	24,4	24,3	24,4
8,7532	8,75	8,75	8,76
4,278	4,3	4,2	4,3
23,839	23,8	23,8	23,9
1 485,01	1485	1485	1486
31,457	31,5	31,4	31,5
317 547	3175	3175	3176
253,12	253	253	254

4. a) *Clases:* agrupaciones de los distintos valores que toma la variable. Las clases deben ser *exhaustivas*, es decir, todos los valores deben estar incluidos en alguna clase, y *mutuamente excluyentes*, a saber, un valor no debe pertenecer a más de una clase.

   b) *Intervalo de clase:* indica la amplitud de la clase y la diferencia entre el límite real superior y el límite real inferior de la clase.

   c) *Punto medio de clase:* es el valor central de la clase y se obtiene al hacer el promedio de los dos límites reales de la clase.

d) *Frecuencia absoluta:* número de observaciones pertenecientes a una misma clase.

e) *Frecuencia relativa:* la frecuencia relativa de una clase se obtiene dividiendo la frecuencia absoluta por el total de observaciones; por tanto, indica la importancia relativa de cada clase dentro de la distribución.

f) *Frecuencia acumulada:* es la suma ascendente o descendente de las frecuencias absolutas o relativas, según se quiera acumular hacia arriba o hacia abajo, respectivamente.

5. La importancia del punto medio en una distribución de frecuencias es la de representar a la clase cuando se requiera realizar algunos cálculos para análisis posteriores.

6. a) Frecuencia relativa

   b) Frecuencia relativa acumulada "menos de"

   c) Frecuencia absoluta acumulada "más de"

   d) Frecuencia absoluta

   e) Frecuencia relativa

7. a)

CLASES	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA	FREC ACUMULADA "MÁS DE"	PUNTOS MEDIOS
17 – 24	3	0,10	30	21
25 – 28	4	0,13	27	27
29 – 32	6	0,20	23	31
33 – 36	8	0,27	17	35
37 – 40	5	0,17	9	39
41 – 52	4	0,13	4	47
TOTAL	30	1,00		

b)

LÍMITES REALES	INTERVALO DE CLASE
17 a menos de 25	8
25 a menos de 29	4
29 a menos de 33	4
33 a menos de 37	4
37 a menos de 41	4
41 a menos de 53	12

8. a)

CONSUMO MENSUAL DE ENERGÍA ELÉCTRICA DE 74 FAMILIAS DE UN CANTÓN RURAL DE LA PROVINCIA DE SAN JOSÉ

Clases	Puntos medios	Frecuencias		Frecuencias acumuladas "menos de"		Frecuencias acumuladas "más de"	
		Absolutas	Relativas				
	x_i	f_i	f_r	Absoluta	Relativa	Absoluta	Relativa
19,5 – 59,5	40	3	0,04	3	0,04	74	1
59,5 – 99,5	80	11	0,15	14	0,19	71	0,96
99,5 – 139,5	120	15	0,20	29	0,39	60	0,81
139,5 – 179,5	160	11	0,15	40	0,54	45	0,61
179,5 – 219,5	200	10	0,14	50	0,68	34	0,46
219,5 – 259,5	240	9	0,12	59	0,80	24	0,32
259,5 – 299,5	280	6	0,08	65	0,88	15	0,20
299,5 – 339,5	320	3	0,04	68	0,92	9	0,12
339,5 – 379,5	360	2	0,03	70	0,95	6	0,08
379,5 – 419,5	400	4	0,05	74	1,00	4	0,05
		74	1				

b) 1. 65 familias consumen menos de 300 kWh mensuales en energía eléctrica.

2. 60 familias consumen más de 100 kWh mensuales en energía eléctrica.

3. 36 familias consumen entre 100 y 220 kWh mensuales.

9.

a) Clases	Frecuencias fi	b) Puntos medios	Frecuencias relativas	c) Frecuencias "menos de"	Frecuencias "más de"
1,95 – 2,95	4	2,45	0,08	4	52
2,95 – 3,95	2	3,45	0,04	6	58
3,95 – 4,95	8	4,45	0,15	14	46
4,95 – 5,95	13	5,45	0,25	27	38
5,95 – 6,95	11	6,45	0,21	38	25
6,95 – 7,95	8	7,45	0,15	46	14
7,95 – 8,95	5	8,45	0,1	51	6
8,95 – 9,95	1	9,45	0,02	52	1
TOTAL	52		1,00		

d) Construcción de un histograma para representar las frecuencias relativas

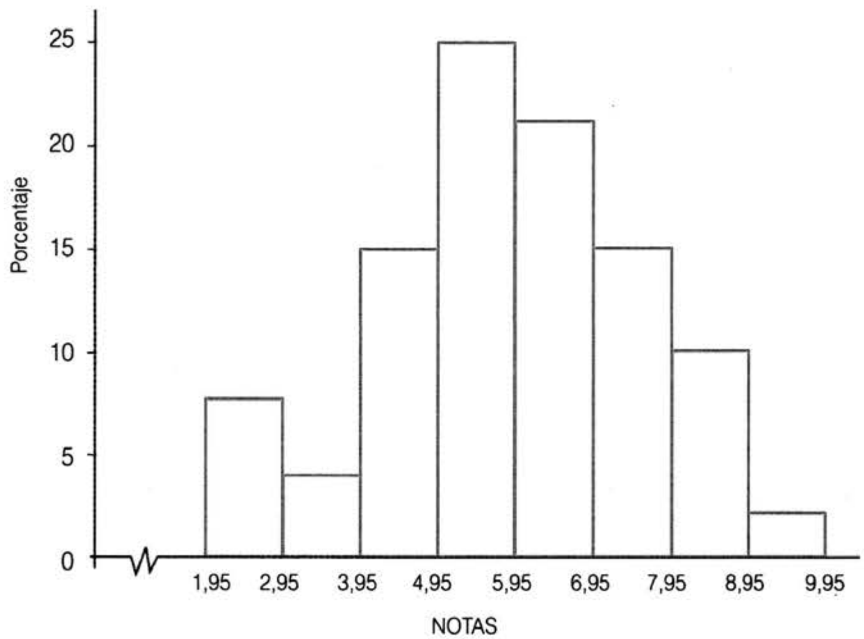

Figura 7.9. Distribución de las notas de un examen parcial de un curso universitario (porcentajes del total de alumnos)

e)

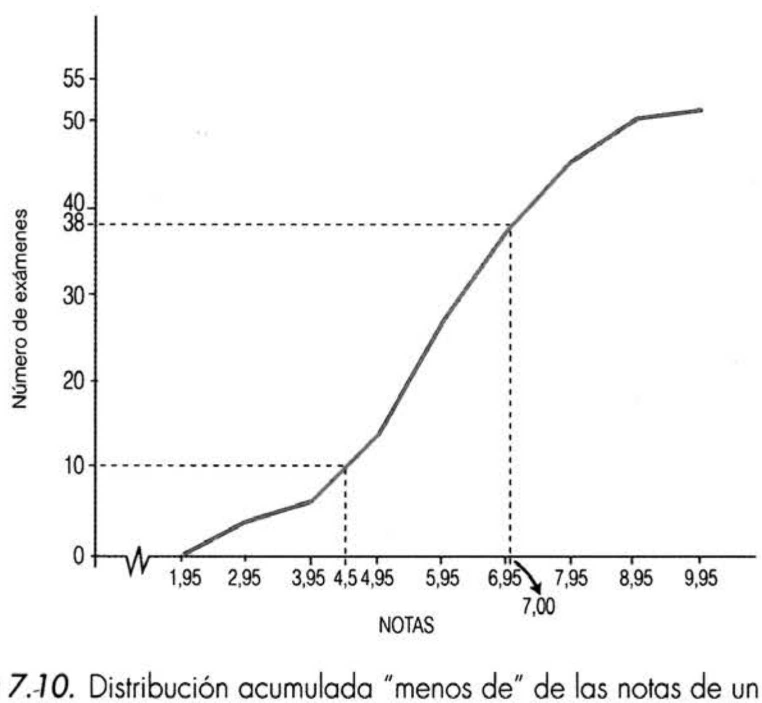

Figura 7.10. Distribución acumulada "menos de" de las notas de un examen parcial de un curso universitario

Aproximadamente 38 alumnos tienen una nota inferior a 7,00 y aproximadamente 42 tienen una superior a 4,50.

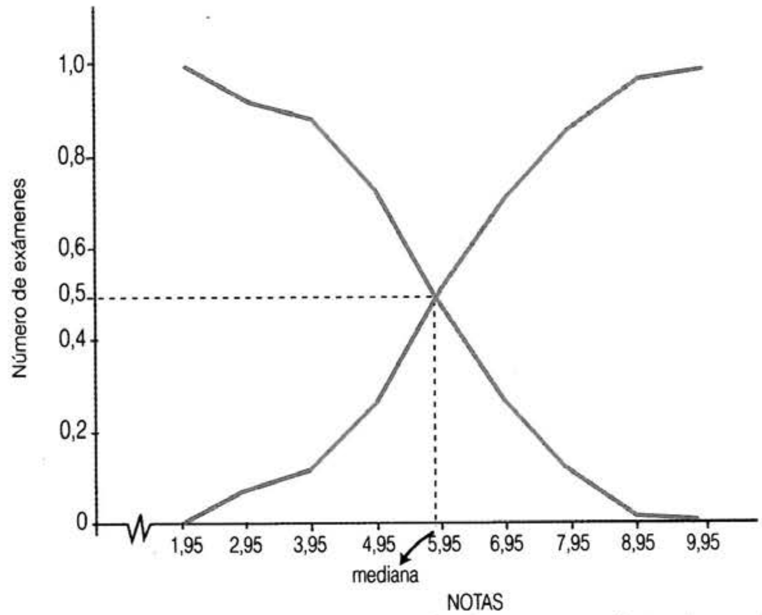

Figura 7.11. Distribuciones acumuladas "menos de" y "más de" de las notas de un examen parcial de un curso universitario

El valor aproximado de la mediana es 5,8.

10. a) Gráfico. Para comparar las distribuciones del número de suicidios registrados según la edad, se deben utilizar las frecuencias relativas, ya que el número absoluto de suicidios en el años 2008 es casi un 50% mayor que en 1998 (325 contra 223). Esto hace que los polígonos basados en las frecuencias absolutas queden muy separados y la comparación resulte limitada.

Grupos edades	Frecuencia relativa	
	1998	2008
5 –14	0,03	0,01
15 –24	0,22	0,25
25 –34	0,28	0,24
35 –44	0,20	0,17
45 –54	0,13	0,17
55 –64	0,06	0,10
65 y más	0,09	0,06
TOTAL	1,00	1,00

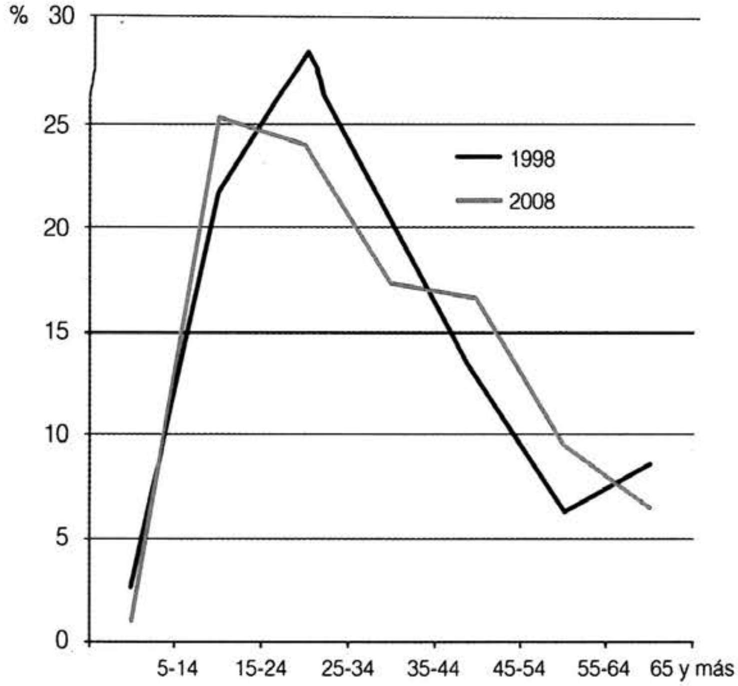

Figura 7.12. Suicidios según edad, años 1998 y 2008.

Fuente: INEC, Estadísticas Vitales de 1998 y 2008

b) Puede observarse que el patrón que siguen los suicidios según la edad es muy similar en ambos años: ocurren pocos antes de los 15 años y en las edades avanzadas, la gran mayoría se concentra entre los 15 y los 45 años. Por otra parte, la comparación de las distribuciones parece sugerir que la edad a la que ocurren la mayoría de los suicidios ha disminuido; así, mientras en 1998 la clase modal (aquella con la frecuencia más alta) era 25-34, en el 2008 es la 15-24. Sin embargo, el examen de los grupos de edades 45-54 y 55-64 señala una mayor proporción de suicidios en el año 2008 en esas edades. Una conclusión más precisa requeriría usar las medidas de posición, que se verán en el próximo capítulo, además el empleo de datos para años intermedios, no solo para aumentar el número de casos bajo análisis, sino también para ver si es posible, con el uso de varios años, definir una serie de tiempo y determinar si existe realmente o no una tendencia a que los suicidios se den a edades más tempranas.