

Distribuciones de frecuencias

CAPÍTULO 2

DATOS SUELTOS

Se les llama *datos sueltos* a los datos recolectados que no han sido organizados numéricamente. Un ejemplo es el conjunto de las estaturas de 100 estudiantes hombres, obtenidas del registro universitario, que está ordenado en forma alfabética.

ORDENACIÓN

Una *ordenación* es un conjunto de datos numéricos en orden creciente o decreciente de magnitud. A la diferencia entre el número mayor y el menor se le conoce como *rango* de los datos. Por ejemplo, si la estatura mayor de los 100 estudiantes es 74 pulg y la menor es 60, el rango es $74 - 60 = 14$ pulg.

DISTRIBUCIONES DE FRECUENCIAS

Si se reúnen grandes cantidades de datos sueltos es útil distribuirlos en *clases* o *categorías*, y determinar el número de individuos que pertenecen a cada categoría, a lo que se le llama *frecuencia de clase*. A una disposición tabular de los datos por clases, con sus correspondientes frecuencias de clase, se le conoce como *distribución de frecuencias* o *tabla de frecuencias*. La tabla 2-1 es una distribución de frecuencias de las estaturas de 100 estudiantes hombres de la universidad XYZ.

La primera clase (o categoría), por ejemplo, comprende las estaturas entre 60 y 62 pulg y se indica con el rango 60-62. Como hay cinco estudiantes en esta clase, la correspondiente frecuencia de clase es 5.

A los datos así organizados y reunidos en clases, como en la anterior distribución de frecuencias, se les llama *datos agrupados*. Aunque el proceso de agrupamiento generalmente quita detalles originales de los datos, es muy ventajosa pues proporciona una visión amplia y clara, además de que se obtienen relaciones evidentes.

Tabla 2-1 Estaturas de 100 estudiantes hombres de la universidad XYZ

Estatura (pulg)	Número de estudiantes
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
Total	100

INTERVALOS DE CLASE Y LÍMITES DE CLASE

El símbolo que define una clase, como el 60-62 de la tabla 2-1, se llama *intervalo de clase*. A los números 60 y 62 se les conoce como *límites de clase*; el número más pequeño (60) es el *límite inferior de clase*, mientras que el número más grande (62) es el *límite superior de clase*. Se acostumbra usar los términos *clase* e *intervalo de clase* indistintamente, aunque el intervalo de clase es en realidad un símbolo de la clase.

A un intervalo de clase que, por lo menos teóricamente, no tiene límite de clase inferior o límite de clase superior se le llama *intervalo de clase abierto*. Por ejemplo, en grupos de edades de individuos, el intervalo de clase “65 años o más” es un intervalo de clase abierto.

FRONTERAS DE CLASE

Si se miden estaturas con exactitud de 1 pulg, en teoría el intervalo de clase 60-62 incluye todas las medidas desde 59.5000 hasta 62.5000 pulg. Estos números, indicados brevemente por los números exactos 59.5 y 62.5 se llaman *fronteras de clase* o *límites verdaderos de clase*; el número menor (59.5) es la *frontera inferior de clase* y el número mayor (62.5), la *frontera superior de clase*.

En la práctica, las fronteras de clase se obtienen promediando los límites superior e inferior de un intervalo de clase.

Algunas veces se emplean las fronteras para simbolizar clases. Por ejemplo, las clases en la primera columna de la tabla 2-1 se podrían indicar mediante 59.5-62.5, 62.5-65.5, etcétera. Para evitar ambigüedad al usar tal notación, las fronteras de clase no deben coincidir con observaciones posibles. Si una observación fuera 62.5 no sería posible decidir si pertenece al intervalo de clase 59.5-62.5 o al 62.5-65.5.

TAMAÑO O AMPLITUD DE UN INTERVALO DE CLASE

El *tamaño* o la *amplitud de un intervalo de clase* es la diferencia entre las fronteras de clase inferior y superior y se le conoce como *amplitud*, *tamaño* o *longitud de clase*. Si todos los intervalos de una distribución de frecuencia son de la misma amplitud, a esta amplitud común se le denota por c . En tal caso, c es igual a la diferencia entre dos límites inferiores (o superiores) de clases sucesivas. Para los datos de la tabla 2-1, por ejemplo, la amplitud del intervalo de clase es $c = 62.5 - 59.5 = 65.5 - 62.5 = 3$.

MARCA DE CLASE

La *marca de clase*, que es el punto medio del intervalo de clase, se obtiene promediando los límites inferior y superior de clase. De este modo, la marca de clase del intervalo 60-62 es $(60 + 62)/2 = 61$. A la marca de clase también se le denomina *punto medio de la clase*.

Para efectos de análisis matemáticos subsiguientes, se asume que todas las observaciones pertenecientes a un mismo intervalo de clase coinciden con la marca de clase; de esta manera, todas las estaturas en el intervalo de clase 60-62 pulg se considerarán de 61 pulg.

FIGURA 2-2

REGLAS GENERALES PARA CONSTRUIR DISTRIBUCIONES DE FRECUENCIA

1. Determinar el número mayor y el menor en los datos sueltos con el fin de especificar el rango (la diferencia entre ambos).
2. Dividir el rango en un número adecuado de intervalos de clase del mismo tamaño. Si esto no es posible, usar intervalos de clase de distintos tamaños o intervalos de clase abiertos (véase problema 2.12). Se suelen tomar entre 5 y 20 intervalos de clase, según los datos. Los intervalos de clase se eligen también de modo tal que las marcas de clase (o puntos medios) coincidan con los datos realmente observados. Ello tiende a disminuir el llamado *error de agrupamiento* que se produce en análisis matemáticos posteriores. No obstante, las fronteras de clase no debieran coincidir con los datos realmente observados.
3. Determinar el número de observaciones que corresponden a cada intervalo de clase; es decir, hallar las frecuencias de clase. Esto se logra mejor con una *hoja de recuentos* o *registro de marcas* (véase el problema 2.8).

HISTOGRAMAS Y POLÍGONOS DE FRECUENCIAS

Los histogramas y los polígonos de frecuencias son dos representaciones gráficas de las distribuciones de frecuencias.

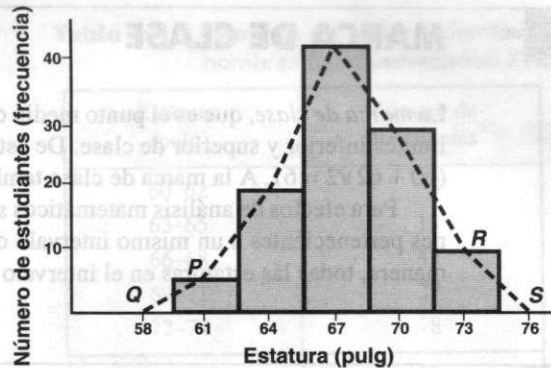
1. Un *histograma* o *histograma de frecuencias* consiste en un conjunto de rectángulos que tienen *a)* sus bases en el eje *X* horizontal, sus centros en las marcas de clase y longitudes iguales a los tamaños de los intervalos de clase, y *b)* áreas proporcionales a las frecuencias de clase.

Si todos los intervalos de clase son del mismo tamaño, las alturas de los rectángulos son proporcionales a las frecuencias de clase. Entonces se acostumbra tomar las alturas numéricamente iguales a las frecuencias de clase. Si los intervalos de clase no son todos del mismo tamaño, hay que ajustar las alturas (véase el problema 2.13).

2. Un *polígono de frecuencia* es una gráfica de línea de las frecuencias de clase dibujada con respecto a la marca de clase. Puede obtenerse uniendo los puntos medios de las partes superiores de los rectángulos del histograma.

El histograma y el polígono de frecuencias correspondientes a la distribución de frecuencias de las estaturas de la tabla 2-1 se muestran en los mismos ejes de la figura 2-1. Suelen añadirse las longitudes *PQ* y *RS* a las siguientes marcas de clase menor y mayor, las cuales tienen frecuencias de clase iguales a cero. En tal caso, la suma de las áreas de los rectángulos del histograma es igual al área total limitada por el polígono de frecuencias y el eje *X* (véase el problema 2.11).

FIGURA 2-1



DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

La *frecuencia relativa de una clase* es su frecuencia dividida entre la frecuencia total de todas las clases y se expresa generalmente como un porcentaje. Por ejemplo, la frecuencia relativa de la clase 66-68 en la tabla 2-1 es $42/100 = 42\%$. Es claro que la suma de todas las frecuencias relativas de las clases es 1, es decir, 100 por ciento.

Si se sustituyen las frecuencias de la tabla 2-1 por las correspondientes frecuencias relativas, a la tabla resultante se le llama *distribución de frecuencias relativas*, *distribución de porcentajes* o *tabla de frecuencias relativas*.

Una representación gráfica de distribuciones de frecuencias relativas se puede obtener, a partir del histograma o del polígono de frecuencias, con sólo cambiar la escala vertical de frecuencias a frecuencias relativas y manteniendo exactamente el mismo diagrama. Las gráficas resultantes se denominan *histogramas de frecuencias relativas* (o *histogramas de porcentajes*) y *polígonos de frecuencias relativas* (o *polígonos de porcentajes*), en ese orden.

DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS Y OJIVAS

La frecuencia total de todos los valores menores que la frontera de clase superior de un intervalo de clase dado se conoce como *frecuencia acumulada* hasta ese intervalo de clase, inclusive. Por ejemplo, la frecuencia acumulada, incluyendo hasta el intervalo de clase 66-68 de la tabla 2-1, es $5 + 18 + 42 = 65$, lo que significa que 65 estudiantes tienen estaturas por debajo de 68.5 pulg.

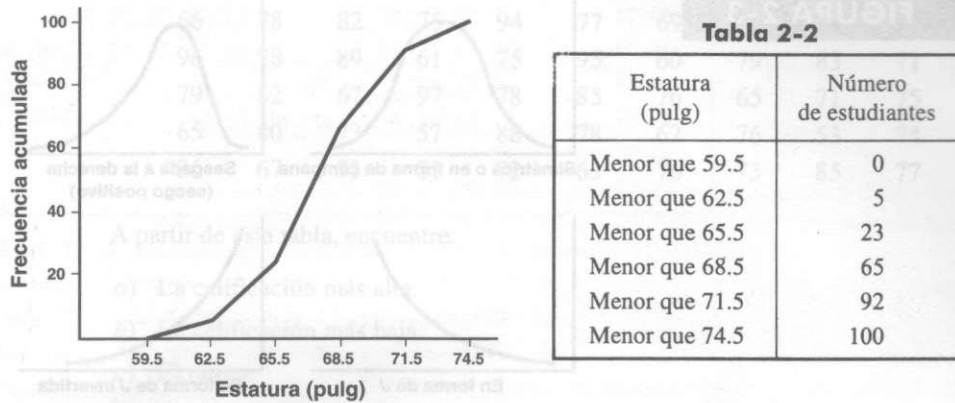
Una tabla que presenta tales frecuencias acumuladas se llama una *distribución de frecuencias acumuladas*, *tabla de frecuencias acumuladas* o, brevemente, una *distribución acumulada*. En la tabla 2-2 se muestra una distribución acumulada para la distribución de las estaturas de los estudiantes de la tabla 2-1.

Una gráfica que recoja las frecuencias acumuladas por debajo de cualquiera de las fronteras de clase superiores respecto de dicha frontera es denominada un *polígono de frecuencias acumuladas* u *ojiva*; en la figura 2-2 se ilustra esta gráfica para la distribución de estaturas de los estudiantes de la tabla 2-1.

Para ciertos propósitos, es deseable considerar una distribución de frecuencias acumuladas de todos los valores mayores o iguales que la frontera de clase inferior de cada intervalo de clase. Como de este modo se obtienen estaturas de 59.5 pulg o más, de 62.5 pulg o más, etcétera, se le suele conocer como una *distribución acumulada "o más"*,

mientras que la antes considerada es una *distribución acumulada "menor que"*. Es fácil deducir una de otra (véase el problema 2.15). Las correspondientes ojivas se conocen, entonces, como ojivas "o más" y "menores que". Siempre que se hable de distribuciones acumuladas u ojivas, sin especificar el tipo, se estará mencionando el caso "menor que".

FIGURA 2-2



DISTRIBUCIONES DE FRECUENCIAS RELATIVAS ACUMULADAS Y OJIVAS DE PORCENTAJES

La *frecuencia relativa acumulada* o *frecuencia acumulada en porcentajes* es la frecuencia acumulada dividida entre la frecuencia total. Así, la frecuencia relativa acumulada de estaturas menores que 68.5 pulg es $65/100 = 65\%$, lo que significa que 65% de los estudiantes mide menos de 68.5 pulg.

Si se usan frecuencias relativas acumuladas en la tabla 2-2 y en la figura 2-2, en vez de frecuencias acumuladas, los resultados se llaman *distribuciones de frecuencias relativas acumuladas* (o *distribuciones acumuladas en porcentajes*) y *polígonos de frecuencias relativas acumuladas* (u *ojivas en porcentajes*), respectivamente.

CURVAS DE FRECUENCIA Y OJIVAS SUAVIZADAS

Los datos recolectados suelen considerarse como pertenecientes a la muestra tomada de una población grande. Ya que son posibles muchas observaciones sobre esa población, en teoría es posible (para datos continuos) escoger intervalos de clase muy pequeños y, aún así, tener un número razonable de observaciones en cada clase. De esta manera, cabe esperar que el polígono de frecuencias o el polígono de frecuencias relativas correspondiente a una población grande contenga tantos pequeños segmentos de línea que se aproxime a curvas a las que se les conoce como *curva de frecuencias* o *curva de frecuencias relativas*, respectivamente.

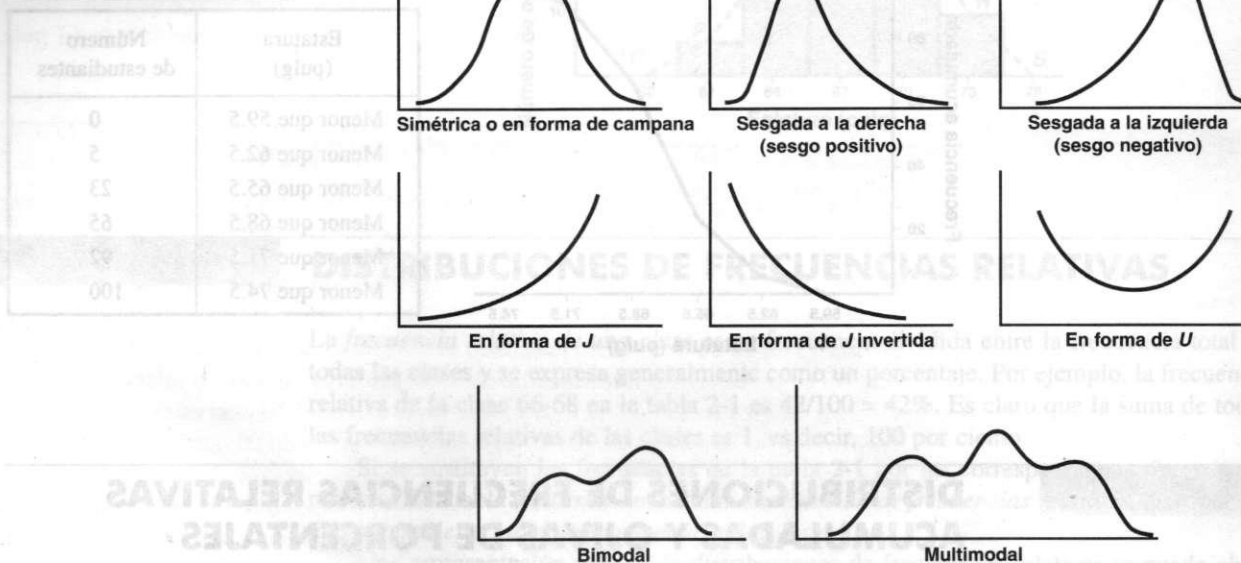
Conviene esperar aproximaciones a dichas curvas teóricas si se suavizan los polígonos de frecuencias o los polígonos de frecuencias relativas de la muestra; la aproximación se mejora conforme crece el tamaño de la muestra. Por esa razón, una curva de frecuencias se designa a veces como *polígono de frecuencia suavizado*.

De forma análoga, al suavizar los polígonos de frecuencias acumuladas u ojivas se obtienen *ojivas suavizadas*. Suele ser más fácil suavizar una ojiva que un polígono de frecuencias (véase el problema 2.18).

TIPOS DE CURVAS DE FRECUENCIAS

Las curvas de frecuencias que aparecen, en la práctica adoptan ciertas formas características, como se ilustra en la figura 2-3.

FIGURA 2-3



1. Las curvas de frecuencias *simétricas o en forma de campana* se caracterizan porque las observaciones equidistantes del máximo central tienen la misma frecuencia. Un ejemplo importante es la *curva normal*.
2. En las curvas de frecuencias *moderadamente asimétricas, o sesgadas*, la cola de la curva a un lado del máximo central es más larga que al otro lado. Si la cola más larga está a la derecha, se dice que la curva es *sesgada a la derecha* o que tiene *asimetría positiva*. En caso contrario, se afirma que la curva es *sesgada a la izquierda* o que es de *asimetría negativa*.
3. En una curva *en forma de J* o de *J invertida* hay un máximo en un extremo.
4. Una curva de frecuencias *en forma de U* tiene máximos en ambos extremos.
5. Una curva de frecuencias *bimodal* tiene dos máximos.
6. Una curva de frecuencias *multimodal* tiene más de dos máximos.