

8 Unidad

Organización y análisis de datos estadísticos

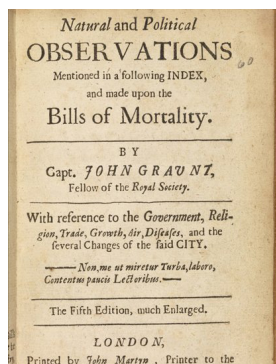


Imagen del trabajo realizado por John Graunt en 1662.

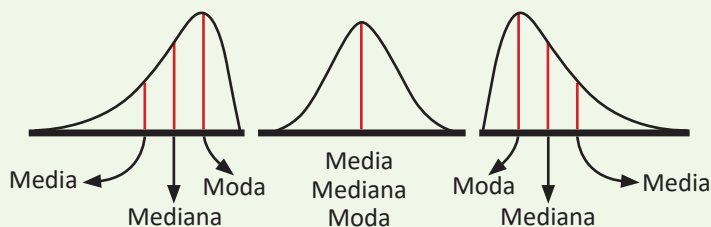
La estadística tuvo sus orígenes en actividades que estaban relacionadas directamente con la organización política, jurídica y administrativa de distintas civilizaciones, entre las que se pueden mencionar los egipcios, los babilónicos y los romanos; para ello, los funcionarios públicos tenían la obligación de anotar nacimientos, matrimonios y defunciones, sin olvidar los recuentos periódicos del ganado y de las riquezas que poseían los territorios conquistados; pero John Graunt (1620 - 1674) fue el primero que puso las bases de una estadística científica, realizando un trabajo a partir de las tablas de mortalidad de la ciudad de Londres.

La estadística juega un papel importante en distintas áreas; por ejemplo, en el sector educativo, económico, tecnológico, social y de la salud, proporcionando herramientas metodológicas que facilitan la recolección, comparación y análisis de datos, con el fin de generar modelos para hacer predicciones y facilitar la toma de decisiones. Por ejemplo, las redes sociales, se nutren de un continuo análisis estadístico en el desarrollo de sus aplicaciones internas.

Gráfico 2. Evolución del presupuesto del MINSAL como proporción del PIB, (2005 - 2011)



Ministerio de Salud, Estudio de exclusión social en salud, diciembre 2011.



Relaciones entre las medidas de tendencia central.

En el desarrollo de los contenidos de esta unidad aprenderás a organizar datos en tablas, representarlos mediante gráficas y determinar valores representativos que se conocen como medidas de tendencia central; así como algunas de sus propiedades y aplicaciones en situaciones cotidianas.

1.1 Agrupación de datos



Doña Carmen es la propietaria de la sala de belleza El Buen Gusto, para atender a todos sus clientes ha tenido que abrir dos sucursales A y B. A continuación se muestran los datos del registro de la edad de 30 clientes atendidos el día de la secretaria, en la sucursal A.

24	23	28	30	29	31
27	26	24	30	32	29
21	22	27	33	30	24
24	21	26	24	30	39
24	22	22	24	21	20

1. Clasifica los datos en 5 grupos de 4 en 4, inicia en 20 y termina en 40.
2. ¿En qué grupo se concentró el mayor número de clientes?
3. ¿Qué cantidad de clientes tiene la mayor edad registrada?
4. ¿Qué cantidad de clientes tiene una edad inferior a 32 años?



1. Para clasificar los datos, primero se crean los grupos, como deben iniciar en 20 e ir de 4 en 4, entonces los grupos serán de 20 a 24, de 24 a 28, de 28 a 32, de 32 a 36 y de 36 a 40. Para facilitar la clasificación de los datos, se van escribiendo en una tabla, siguiendo el orden en que aparecen, tal como se muestra en la tabla de la derecha.

2. En el grupo de edades de 24 a 28 años queda concentrado el mayor número de clientes (igual a 11).

3. La mayor edad de los clientes que se atendieron corresponde a 39 años (que es el dato mayor).

4. Para determinar la cantidad de clientes que se atendió y que tienen una edad inferior a 32 años, se cuentan los que quedan en los primeros 3 grupos: $8 + 11 + 8 = 27$, es decir, 27 clientes tienen una edad inferior a 32 años.

	24			
	24			
	24			
20	26	30		
21	24	30		
22	24	29		
22	27	30		
21	24	31		
22	26	29		
21	27	30	33	
23	24	28	32	39
De 20 a 24	De 24 a 28	De 28 a 32	De 32 a 36	De 36 a 40



Para organizar una serie de datos en grupos, se realiza lo siguiente:

1. Se definen las categorías considerando el número de grupos a crear y los límites a considerar.
2. Se colocan los datos uno a uno, en el grupo al que pertenecen, teniendo cuidado que en cada grupo deben quedar los datos, cuyo valor es igual o mayor al del límite menor, pero que sean menores que el límite mayor, tal como se muestra en el ejemplo desarrollado, por ejemplo: en el grupo 1 de 20 a 24 quedarán todos los datos que son iguales o mayores a 20, pero menores que 24; lo que significa que los datos cuyo valor es 24 quedan en el siguiente grupo.



A continuación se muestran los datos del registro de la edad de 30 clientes atendidos el día de la secretaria, en la sucursal B de la sala de belleza El Buen Gusto, de doña Carmen.

20	22	24	22	30
27	34	35	29	28
24	21	20	23	26
23	26	20	29	36
28	29	24	23	34
24	21	20	36	24

1. Clasifica los datos en 5 grupos de 4 en 4, inicia en 20 y termina en 40.
2. ¿En qué grupo se concentró el mayor número de clientes?
3. ¿Qué cantidad de clientes tiene la mayor edad registrada?
4. ¿Qué cantidad de clientes tiene una edad igual o mayor a 32 años?

1.2 Tabla de frecuencias



Retomando los datos de los 30 clientes atendidos el día de la secretaria, en la sucursal A de la sala de belleza:

1. Organiza los grupos en una tabla.
2. Determina el total de datos de cada grupo y anota el resultado.

24				
24				
24				
20	26	30		
21	24	30		
22	24	29		
22	27	30		
21	24	31		
22	26	29		
21	27	30	33	
23	24	28	32	39
De 20 a 24	De 24 a 28	De 28 a 32	De 32 a 36	De 36 a 40



1. Para organizar los datos en grupos, se elabora una tabla y en la primera columna se colocan los grupos.
2. Al realizar el conteo de los datos que quedan en cada grupo y colocar el resultado en la tabla, se obtiene la columna 2.

Edades	Número de clientes
20 - 24	8
24 - 28	11
28 - 32	8
32 - 36	2
36 - 40	1
Total	30



La tabla en la que se organizan los grupos de datos de una serie, tal como el ejemplo desarrollado, se llama **tabla de distribución de frecuencias** y a cada grupo de datos formado se le llama **clases**, de donde se puede decir que los datos del ejemplo han sido organizados en 5 clases. Además, al total de datos que corresponde a cada clase se le llama **frecuencia**.

- Por tanto, para organizar una serie de datos en una tabla de distribución de frecuencias, es necesario:
- Organizar los datos en tantas clases como sea necesario.
 - Realizar el conteo de los datos que pertenecen a cada clase para determinar la frecuencia.



Mario y Carlos se reúnen todas las tardes para jugar baloncesto, durante el último mes han llevado un registro de los tiempos jugados por cada uno, cuyos datos se muestran a continuación:

Mario

10	13			
11	15		21	
12	14		21	
11	13	16	20	
12	15	17	19	
11	13	16	20	
10	14	17	19	23
12	13	18	20	22
De 10 a 13	De 13 a 16	De 16 a 19	De 19 a 22	De 22 a 24

Carlos

				24
			21	22
		17	21	24
	13	18	20	22
11	13	16	19	23
11	14	16	19	22
11	13	18	19	22
10	15	17	20	23
De 10 a 13	De 13 a 16	De 16 a 19	De 19 a 22	De 22 a 25

1. Ordena los datos en una tabla de distribución de frecuencias.
2. Completa los resultados y responde, ¿quién es el que ha jugado mayor tiempo?

1.3 Elementos de la tabla de frecuencias



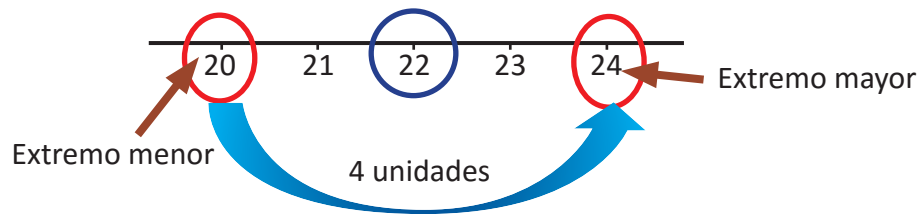
La tabla contiene el registro de la edad de 30 clientes atendidos el día de la secretaria, en la sucursal A de la sala de belleza El Buen Gusto de doña Carmen, analiza los datos y responde:

Edades	Número de clientes
	<i>f</i>
20 - 24	8
24 - 28	11
28 - 32	8
32 - 36	2
36 - 40	1
Total	30

1. Determina el tamaño de cada clase.
2. Calcula el valor del número que está en el centro de cada clase.
3. ¿Cuál es la frecuencia de la clase cuyo valor que está en medio es 30?

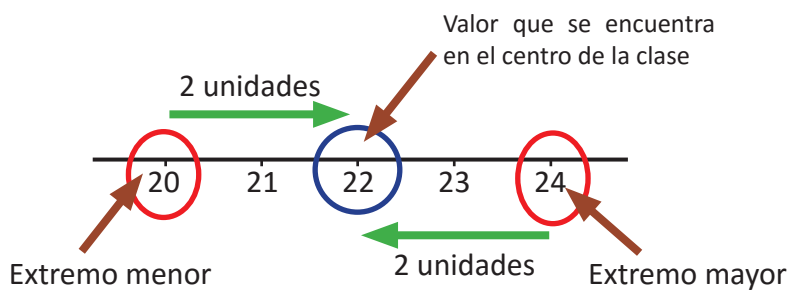


1. Al analizar el tamaño de la primera clase, puede observarse que es igual a 4 unidades. Por ejemplo:



Se puede calcular restando los extremos de cada clase: $24 - 20 = 4$.

2. Al observar la primera clase, se puede obtener el valor del número que está en el centro de la clase, gráficamente contando igual cantidad de unidades desde cada uno de los extremos, tal como se muestra a continuación:



Edades	Número de clientes	Valor medio
	<i>f</i>	
20 - 24	8	22
24 - 28	11	26
28 - 32	8	30
32 - 36	2	34
36 - 40	1	38
Total	30	

Se puede calcular sumando los extremos y dividiendo por dos:

$$\frac{24 + 20}{2} = \frac{44}{2} = 22$$

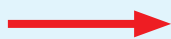
3. Al observar la tabla puede verse que la clase, cuyo valor que está en medio es 30, es la clase de 28 a 32, y tiene una frecuencia de 8.



Al tamaño de una clase se le llama **ancho de clases** y a los valores extremos **límite de clases**; por ejemplo, para la primera clase, de 20 a 24, los límites de clase son 20 y 24, se tiene que

Límite inferior = extremo menor = 20

Límite superior = extremo mayor = 24



Ancho de clase = $24 - 20 = 4$.

Para calcular el ancho de una clase cualquiera se utiliza la ecuación:

$$\text{Ancho de clase} = \text{límite superior} - \text{límite inferior}$$

El número que está en el centro de cada clase se llama **punto medio** y se determina mediante la ecuación:

$$\text{Punto medio} = \frac{\text{límite superior} + \text{límite inferior}}{2}$$



1. La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal B de la sala de belleza El Buen Gusto, de doña Carmen, analiza los datos y responde:

- Determina el ancho de las clases.
- Calcula el punto medio de cada clase.
- ¿Cuál es la frecuencia de la clase cuyo punto medio es 26?

Edades	Número de clientes
	<i>f</i>
20 - 24	11
24 - 28	8
28 - 32	6
32 - 36	3
36 - 40	2
Total	30

2. La tabla de frecuencia muestra resultados de la toma de presión sanguínea sistólica a 100 hombres adultos saludables, analiza los datos y responde:

- Determina el ancho de las clases.
- Calcula el punto medio de las clases de la distribución.
- ¿Cuántos hombres tienen una presión de 125 mmHg en promedio?

Presión sanguínea (mmHg)	Total de hombres
	<i>f</i>
100 - 110	4
110 - 120	18
120 - 130	38
130 - 140	55
140 - 150	17
150 - 160	6
Total	138

3. Investiga la edad de tus compañeras y compañeros de grado, con los datos recopilados realiza lo siguiente:

- Identifica el dato menor y el dato mayor, luego organízalos en 5 grupos.
- Organiza los datos en una tabla de frecuencias.
- Determina los límites de clases y las respectivas frecuencias.
- Calcula el punto medio de cada clase.

1.4 Gráficas estadísticas



La tabla contiene el registro de la edad de 30 clientes atendidos el día de la secretaria, en la sucursal A de la sala de belleza El Buen Gusto, realiza lo siguiente:

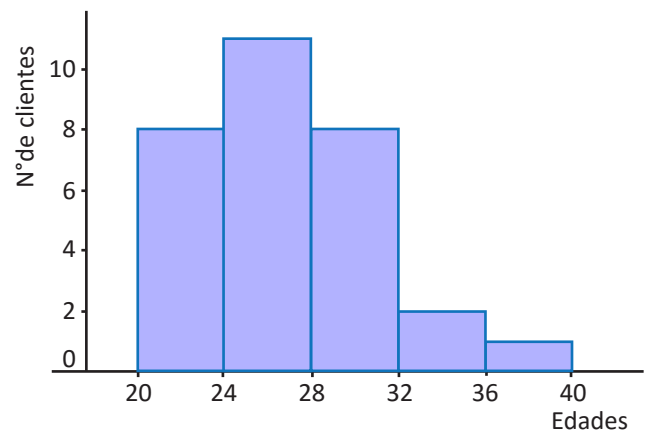
1. Representa mediante rectángulos las clases con las respectivas frecuencias.
2. Qué características tiene la gráfica que muestra la distribución de los clientes atendidos en la sucursal A de la sala de belleza.
3. Grafica el punto medio y la frecuencia de cada clase como pares ordenados.
4. Une con segmentos de recta los puntos graficados en el numeral anterior.

Edades	f
20 - 24	8
24 - 28	11
28 - 32	8
32 - 36	2
36 - 40	1
Total	30



1. Para representar mediante rectángulos las clases, con las respectivas frecuencias, se realiza lo siguiente:

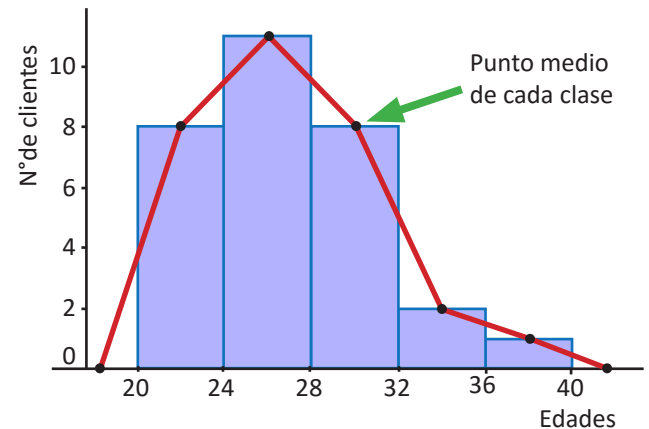
- Colocar en el eje horizontal los límites de las clases.
- En el eje vertical se coloca el número de clientes que corresponde a la frecuencia de cada clase.
- Sobre el ancho de clases, se levantan rectángulos cuya altura coincida con la frecuencia de cada clase.



2. Al observar la gráfica puede verse que los primeros rectángulos son más altos, lo que indica que la mayor cantidad de clientes que se atendió tiene una edad entre 20 y 32 años. Además, como el límite superior de una clase es igual al inferior de la siguiente, los rectángulos quedan pegaditos, uno a continuación del otro.

3. Al ampliar una clase inferior y una superior con frecuencia cero, y graficar como pares ordenados el punto medio con las respectivas frecuencias, se obtienen puntos ubicados en el centro de la parte superior de cada rectángulo.

4. Al unir los puntos se obtiene una línea poligonal abierta que inicia en el punto medio de la primera clase y termina en el punto medio de la última clase, tal como se muestra en la gráfica de la derecha.





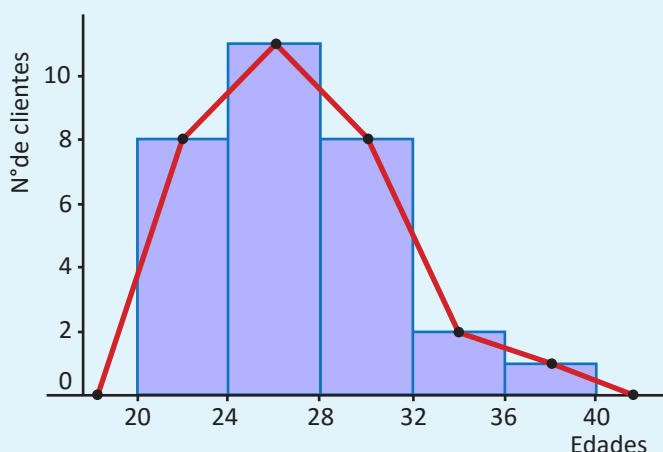
La gráfica que se obtiene al representar las clases con sus respectivas frecuencias se le llama **histograma** y para elaborarlo se realiza lo siguiente:

- Se coloca en el eje horizontal los límites de las clases.
- En el eje vertical se coloca la frecuencia, se busca una escala adecuada, considerando los valores de la frecuencia de la distribución de los datos.
- Se levantan rectángulos cuya base coincide con el ancho de clases y la altura con la frecuencia de la respectiva clase.

Al observar el histograma se puede encontrar que

- Tiene forma parecida a la de una montaña y la parte más alta indica dónde se encuentra concentrado el mayor número de datos.
- Los rectángulos que forman el histograma tienen un área proporcional a la frecuencia de su clase.

En algunos casos es importante resaltar la forma de la distribución de los datos, en ese caso, se coloca un punto en el punto medio del lado superior de cada rectángulo, se unen con segmentos de recta los puntos identificados; luego, el extremo izquierdo se conecta con el punto medio de una clase imaginaria anterior a la menor, con frecuencia cero y el extremo derecho se conecta con el punto medio de una clase imaginaria posterior a la mayor, también con frecuencia cero. A la gráfica que se obtiene se le llama **polígono de frecuencia**.



La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal B de la sala de belleza El Buen Gusto, con este registro realiza lo siguiente:

1. Representa los datos mediante un histograma.
2. ¿Qué características tiene la gráfica que muestra la distribución de las edades de los 30 clientes atendidos en la sucursal B de la sala de belleza?
3. Grafica el polígono de frecuencia a partir del histograma.

Edades	f
20 - 24	11
24 - 28	8
28 - 32	6
32 - 36	3
36 - 40	2
Total	30

1.5 Uso del polígono de frecuencias



En el octavo grado de un centro escolar se aplicó una prueba a las dos secciones, los resultados se muestran en la tabla de la derecha. Con la información realiza lo siguiente:

1. ¿Es posible comparar los resultados obtenidos en las dos secciones? En caso de no ser posible, analiza una manera que permita comparar los resultados de las dos secciones.
2. Calcula el porcentaje de alumnos cuyos resultados quedan agrupados en cada clase.
3. Representa los datos en un polígono de frecuencia.

Puntajes	Sección (A)	Sección (B)
	f_A	f_B
0 - 20	3	5
20 - 40	5	8
40 - 60	12	17
60 - 80	6	10
80 - 100	4	5
Total	30	45



1. Como el número de estudiantes de las dos secciones es distinto, no tiene sentido comparar las frecuencias, por ejemplo, en la clase de 40 a 60, quedan comprendidos los resultados de 12/30 estudiantes en la sección A; mientras que en la sección B son 17/45 estudiantes.

Como no es posible comparar las frecuencias, entonces se puede calcular la razón de la frecuencia de cada clase entre el total de la frecuencia en lugar de utilizar la frecuencia por sí sola; por ejemplo, para la primera clase de 0 a 20 se tiene $\frac{3}{30} = 0.10$ para la sección A y $\frac{5}{45} = 0.11$ para la sección B, al calcular para todas las clases se obtienen los datos de la siguiente tabla.

Puntajes	Sección (A)	Sección (B)
0 - 20	0.10	0.11
20 - 40	0.17	0.18
40 - 60	0.40	0.38
60 - 80	0.20	0.22
80 - 100	0.13	0.11
Total	1.00	1.00

2. Llamando x al porcentaje de una clase y considerando que el total de la frecuencia corresponde al 100% de los estudiantes de una sección, entonces para calcular el porcentaje de la primer clase se tiene:

Para la sección A:

$$3/30 = x/100\%, \text{ de donde se obtiene que } x = \frac{3}{30} \times 100\%.$$

Para la sección B:

$$5/45 = x/100\%, \text{ de donde se obtiene que } x = \frac{5}{45} \times 100\%.$$

Al comparar los resultados con el numeral anterior, se observa que el porcentaje de una clase se puede obtener multiplicando por 100% la razón entre la frecuencia y el total de la frecuencia; es decir, se puede calcular los porcentajes multiplicando únicamente por 100% los resultados de la tabla anterior, por ejemplo para la segunda clase de 20 a 40, se tiene:

$$\text{Sección A: } 0.17 \times 100\% = 17\%.$$

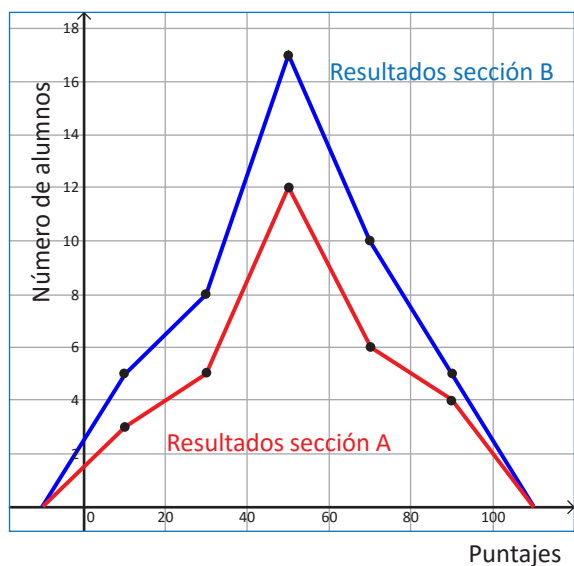
$$\text{Sección B: } 0.18 \times 100\% = 18\%.$$

Puntajes	Sección (A)	Sección (B)
0 - 20	10	11
20 - 40	17	18
40 - 60	40	38
60 - 80	20	22
80 - 100	13	11
Total	100%	100%

Así, sucesivamente, se determinan los porcentajes de las clases restantes, obteniendo los resultados de la tabla.

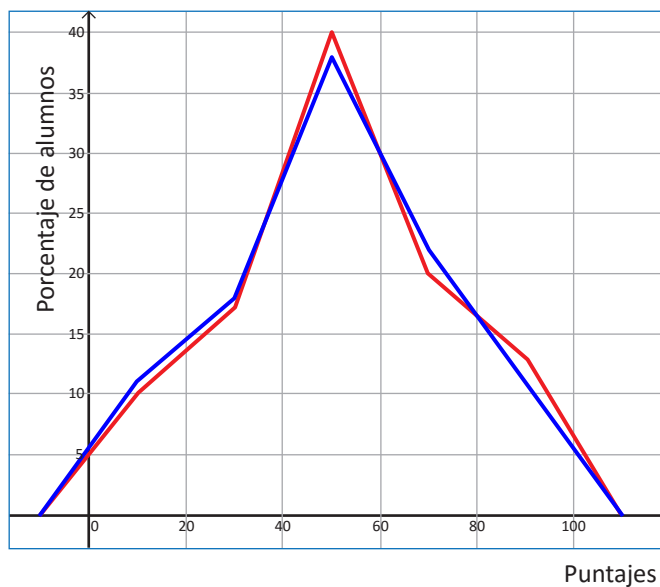
3. Al representar los resultados de cada una de las secciones mediante un polígono de frecuencias en un mismo plano, se obtiene la gráfica 1, en la cual no se pueden realizar comparaciones por tener distinto número de datos, pero si en lugar de las frecuencias se toman los porcentajes, entonces se puede hacer una comparación gráfica entre los resultados de las dos secciones (ver gráfica 2).

1.



Gráfica 1

2.



Gráfica 2



La comparación de datos estadísticos generalmente no se puede realizar directamente con las frecuencias de cada clase, en estos casos, es necesario calcular la razón entre la frecuencia de cada clase y el total de la frecuencia; tal como se hizo en el ejemplo anterior, $\frac{\text{frecuencia}}{\text{total de frecuencias}}$, a este cociente se le llama **frecuencia relativa (f_r)**. Considerando que el total de las frecuencias es igual al número de datos (n), entonces $f_r = \frac{\text{frecuencia}}{\text{total de frecuencias}} = \frac{f}{n}$.

Al producto que se obtiene al multiplicar la frecuencia relativa por 100 se le llama **frecuencia relativa porcentual ($f_r\%$)**, es decir que $f_r\% = \frac{\text{frecuencia}}{\text{total de frecuencias}} \times 100 = \frac{f}{n} \times 100$ se utiliza para determinar los porcentajes de datos que corresponden a cada clase de la distribución; para facilitar el análisis y/o comparación de una o más series de datos.



Miguel tiene una finca y para recolectar el café organizó a los trabajadores en dos cuadrillas, en la tabla se muestra el registro de la cantidad de café recolectada en un día específico por cada una de las cuadrillas. Con la información realiza lo siguiente:

1. Calcula las frecuencias relativas y las frecuencias relativas porcentuales.
2. Representa los datos en un polígono de frecuencia relativa porcentual.
3. ¿En cuál cuadrilla los trabajadores tuvieron mejores resultados?

Arrobas de café	Cuadrilla 1	Cuadrilla 2
0 - 3	1	2
3 - 6	2	4
6 - 9	3	7
9 - 12	5	8
12 - 15	6	10
15 - 18	5	7
18 - 21	2	5
21 - 24	1	2
Total	25	45

1.6 Interpretación de datos estadísticos



En el Instituto Nacional Buena Vista realizaron el examen de admisión para el año próximo, los resultados se muestran en la tabla. Analiza y realiza los respectivos cálculos, luego responde:

1. ¿Qué porcentaje de estudiantes obtuvo un puntaje inferior a 40?
2. ¿Qué porcentaje de los alumnos obtuvo un puntaje mayor o igual a 70?
3. Si únicamente se aceptarán a los que obtuvieron al menos 50 puntos de la prueba, ¿cuántos de los estudiantes evaluados serán aceptados?

Puntajes	Número de alumnos/as
0 - 10	1
10 - 20	6
20 - 30	10
30 - 40	16
40 - 50	22
50 - 60	25
60 - 70	12
70 - 80	9
80 - 90	7
90 - 100	2
Total	110



Primero es necesario calcular los porcentajes de cada clase, por ejemplo:



- Para la clase 1: $f_r\% = \frac{1}{110} \times 100\% = 0.9\%$
- Para la clase 2: $f_r\% = \frac{6}{110} \times 100\% = 5.5\%$

1. El porcentaje de estudiantes que obtuvo un puntaje inferior a 40 se determina sumando los porcentajes de las clases que corresponden a los respectivos puntajes: $0.9 + 5.5 + 9.1 + 14.5 = 30\%$.
2. Para determinar el porcentaje de estudiantes que obtuvo un puntaje mayor o igual a 70, se procede de igual manera que el caso anterior, sumando los respectivos porcentajes: $8.2 + 6.4 + 1.8 = 16.4\%$, que se puede aproximar a 16%.
3. Si se aceptarán a los estudiantes que hayan obtenido al menos 50 puntos de la prueba, para determinar el total de estudiantes, se suman las frecuencias de las respectivas clases: $25 + 12 + 9 + 7 + 2 = 55$; por tanto, serán aceptados únicamente 55 estudiantes.

Puntajes	Porcentaje de alumnos/as
0 - 10	0.9
10 - 20	5.5
20 - 30	9.1
30 - 40	14.5
40 - 50	20.0
50 - 60	22.7
60 - 70	10.9
70 - 80	8.2
80 - 90	6.4
90 - 100	1.8
Total	100%



En una clase de educación física se ha cronometrado el tiempo, en segundos, que tarda cada estudiante en recorrer la pista de la cancha de fútbol.



1. ¿Qué porcentaje de estudiantes hizo un tiempo inferior a 10 segundos?
2. ¿Qué porcentaje de estudiantes hizo un tiempo mayor o igual a 12 segundos?
3. Si se seleccionará al 50% de los estudiantes considerando los que tienen mayor velocidad, ¿cuál es el tiempo máximo que se aceptará?

Tiempo en segundos	Número de alumnos/as
8 - 9	11
9 - 10	10
10 - 11	8
11 - 12	5
12 - 13	3
13 - 14	2
14 - 15	1
Total	40

2.1 Sentido de las medidas de tendencia central



Los datos corresponden al registro del total de clientes atendidos en las dos sucursales de una panadería.

Sucursal A					
14	23	38	40	19	31
49	26	24	30	32	

Sucursal B					
10	22	24	20	30	57
34	46	29	28	24	21

1. Ordena la cantidad de clientes atendidos en ambas sucursales de menor a mayor.
2. Identifica la cantidad mínima y la cantidad máxima de clientes atendidos en ambas sucursales.
3. Determina el valor de la mediana de los datos de las dos sucursales.
4. ¿Cuál es el valor de la moda de los datos sobre la cantidad de clientes atendidos en las dos sucursales de la panadería?
5. Calcula la media aritmética de los datos de las dos sucursales de la panadería.



1. Al ordenar los datos de los clientes atendidos en las dos sucursales de la panadería, se tiene:

Sucursal A: 14, 19, 23, 24, 26, 30, 31, 32, 38, 40, 49.

Sucursal B: 10, 20, 21, 22, 24, 24, 28, 29, 30, 34, 46, 57.

2. La cantidad menor y la cantidad mayor de clientes atendidos en ambas sucursales es:

Sucursal A

Cantidad menor: 14

Cantidad mayor: 49

La cantidad de clientes atendidos en sucursal A oscila entre 14 y 49.

Sucursal B

Cantidad menor: 10

Cantidad mayor: 57

La cantidad de clientes atendidos en la sucursal B oscila entre 10 y 57.

3. Como la mediana es el dato que ocupa la posición central en la serie de datos, entonces, para cada una de las series se tiene:

Sucursal A

14, 19, 23, 24, 26, **30**, 31, 32, 38, 40, 49

Como son 11 datos, entonces la mediana es el dato que ocupa la posición central, es decir, la posición 6, por tanto: mediana = 30.

Sucursal B

10, 20, 21, 22, 24, **24, 28**, 29, 30, 34, 46, 57

Como tiene 12 datos, entonces se toman los dos valores centrales y se calcula el punto medio entre ambos, es decir: mediana = $\frac{24 + 28}{2} = 26$.

4. Al observar las dos series de datos se puede concluir que

- En la sucursal A, todos los datos aparecen una sola vez, por tanto, no tiene moda.
- En la sucursal B, el número 24 aparece 2 veces, entonces: moda = 24.

5. Para calcular la media aritmética es necesario sumar todos los datos y dividir el resultado entre el número de datos, tal como se aprendió en educación básica; *Media aritmética* = $\frac{\text{Suma de todos los datos}}{\text{número de datos}}$ entonces, para los datos de las dos sucursales, se tiene:

Sucursal A: 14, 23, 38, 40, 19, 31, 49, 26, 24, 30, 32.

$$\text{Media aritmética} = \frac{14 + 23 + 38 + 40 + 19 + 31 + 49 + 26 + 24 + 30 + 32}{11} = \frac{326}{11} = 29.6.$$

Sucursal B: 10, 22, 24, 20, 30, 57, 34, 46, 29, 28, 24, 21.

$$\text{Media aritmética} = \frac{10 + 22 + 24 + 20 + 30 + 57 + 34 + 46 + 29 + 28 + 24 + 21}{12} = \frac{345}{12} = 28.8.$$



Tal como se aprendió en educación básica, se pueden calcular valores representativos que pueden describir una serie de datos, los cuales se han calculado en el ejemplo anterior y se detallan a continuación:

La mediana es el valor que ocupa la posición central en una serie de datos, cuando ya han sido ordenados de menor a mayor. Para determinar el valor de la mediana, se consideran los siguientes casos:

a) Cuando el número de datos n es impar, la mediana es el dato x que ocupa la posición central. En este caso, para determinar la posición central se utiliza la fórmula $\frac{n+1}{2}$, para el ejemplo anterior de la sucursal A, $n = 11$, entonces la posición de la mediana es: $\frac{11+1}{2} = \frac{12}{2} = 6$.

b) Cuando el número de datos n es par, la mediana es el número que se encuentra entre los datos centrales, pues al determinar la posición de la mediana, se obtiene un valor que no corresponde a la posición de ningún dato de la serie, por ejemplo, para el caso de la sucursal B, $n = 12$, entonces, al determinar la posición de la mediana, $\frac{12+1}{2} = \frac{13}{2} = 6.5$, lo que indica que la mediana es el número que está entre el dato 6 y el dato 7. En este caso, la mediana = *Punto medio de los dos datos centrales*.

La moda es el valor que aparece la mayor cantidad de veces en una serie, es decir, la moda es el dato que tiene la mayor frecuencia. En casos en que todos los datos aparecen igual cantidad de veces, se dice que la serie no tiene moda o que carece de moda.

La media aritmética (μ) es el número que resulta de dividir la suma de todos los datos x entre el número de datos n y que se conoce también como **promedio**. Media aritmética = $\frac{\text{Suma de todos los } x}{n}$.



Las siguientes series de datos corresponden a las ventas expresadas en dólares, de los últimos 15 días, de las dos sucursales de la minitienda La Esquina:

Sucursal 1: 125, 35, 50, 40, 80, 100, 70, 50, 125, 75, 80, 90, 80, 80, 35.

Sucursal 2: 100, 75, 50, 80, 60, 40, 70, 75, 140, 90, 75, 70, 150, 50, 90.

Con los datos de cada sucursal realiza lo siguiente:

1. Ordena los datos de menor a mayor.
2. Identifica el mínimo y el máximo.
3. Determina la mediana.
4. Identifica el valor de la moda.
5. Calcula la media aritmética.
6. ¿Es posible determinar cuál sucursal genera mayores ingresos?

2.2 Media aritmética



¿Cómo se puede determinar la media aritmética de una serie de datos organizada en una distribución de frecuencias?

La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal A de la sala de belleza El Buen Gusto. Realiza lo siguiente:

1. Calcula el punto medio de cada clase.
2. Multiplica cada punto medio por la respectiva frecuencia.
3. Suma los resultados obtenidos en el numeral dos, luego divide el total obtenido entre el número de datos.
4. Compara el resultado obtenido en el numeral 3, con la media aritmética de los datos de la clase 1 de esta unidad.

Edades	Número de clientes
	f
20 - 24	8
24 - 28	11
28 - 32	8
32 - 36	2
36 - 40	1
Total	30

Revisa la clase 1 de la unidad.



1. Como ya se aprendió a calcular el punto medio en una distribución de frecuencias en clases anteriores, entonces para calcularlos se aplica la fórmula: $punto\ medio = \frac{límite\ superior + límite\ inferior}{2}$, y se le agrega otra columna a la tabla anterior para escribir los resultados. Por ejemplo, para la clase 1:

$$Pm = \frac{24 + 20}{2} = 22$$

Edades	Número de clientes	Punto medio (Pm)
	f	
20 - 24	8	22
24 - 28	11	26
28 - 32	8	30
32 - 36	2	34
36 - 40	1	38
Total	30	

2. Se multiplica el punto medio de cada clase por la respectiva frecuencia, y se le agrega una nueva columna a la tabla para agregar los resultados, por ejemplo para la clase 1 se tiene:

$$Punto\ medio \times frecuencia = Pm \times f = 22 \times 8 = 176$$

Edades	Número de clientes	Punto medio (Pm)	$Pm \times f$
	f		
20 - 24	8	22	176
24 - 28	11	26	286
28 - 32	8	30	240
32 - 36	2	34	68
36 - 40	1	38	38
Total	30		808

3. Se suman los resultados obtenidos en el numeral anterior, luego se divide el resultado entre el número de datos.

$$\frac{\text{Suma de todos los productos de } P_m \times f}{\text{Número de datos}} = \frac{808}{30} = 26.9$$

Generalmente, en las empresas, cuando el objetivo no es el cálculo, sino el análisis de datos, el promedio o media aritmética se puede determinar mediante el uso de un programa informático como Excel, Calc o GeoGebra.

Por ejemplo, se puede determinar el promedio de la edad de clientes atendidos en la sucursal A de la sala de belleza El Buen Gusto.

Se digitan los datos que corresponden a la edad de los 30 clientes de la sucursal A de la sala de belleza "El Buen Gusto", y se calcula la media aritmética, obteniéndose $\mu = 26.2$.

24	23	28	30	29	31
27	26	24	30	32	29
21	22	27	33	30	24
24	21	26	24	30	39
24	22	22	24	21	20

4. El resultado del numeral 3 es 26.9 y la media aritmética de los datos sin organizar en distribución de frecuencias calculado mediante el uso de un programa informático es 26.2, tal como se muestra arriba, la diferencia entre los dos valores es pequeña, por lo que se puede utilizar cualquiera de las dos formas para calcular el promedio de las edades.



Para determinar la media aritmética de una serie de datos organizados en una distribución de frecuencias, se utiliza la ecuación: $\text{Media aritmética} = \frac{\text{Suma de todos los productos de } P_m \times f}{\text{Número de datos}}$, tal como se muestra en el ejemplo desarrollado.



- La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal B, de la sala de belleza El Buen Gusto, realiza lo siguiente:
 - Completa la tabla.
 - Calcula la media aritmética.
- Compara la media aritmética de las dos sucursales, ¿en cuál de las dos es mayor la edad promedio de los clientes atendidos?

Edades	Número de clientes (f)	Punto medio (P_m)	$f \times P_m$
20 - 24	11		
24 - 28	8		
28 - 32	6		
32 - 36	3		
36 - 40	2		
Total	30		

2.3 Propiedades de la media aritmética



Analiza la siguiente situación, luego realiza lo que se pide en cada caso.

La empresa A tiene 25 empleados y les paga un salario promedio de 350 dólares, mientras que la empresa B tiene únicamente 15 empleados con un salario promedio de 600 dólares.

1. Calcula el monto mensual que invierte cada una de las empresas en el pago de los empleados.
2. Si la empresa B realiza un aumento general de 50 dólares, ¿cuál es el nuevo salario promedio?
3. Los empleados de la empresa A piden un aumento para el próximo año, el dueño de la empresa les presenta dos opciones, ¿calcula el nuevo salario promedio en cada caso?, ¿cuál opción recomendarías a los empleados?, ¿por qué?
 - a) Un aumento general de 65 dólares.
 - b) Un aumento del 20% sobre el salario actual.



Para resolver las situaciones planteadas se considerará la información proporcionada en el problema.

<p>EMPRESA A</p> <p>Salario promedio = media aritmética = \$350</p>

<p>EMPRESA B</p> <p>Salario promedio = media aritmética = \$600</p>

1. El monto mensual se determina multiplicando el salario promedio por la cantidad de empleados que tiene cada empresa.

$\text{Monto mensual} = 350 \times 25 = 8\,750$

$\text{Monto mensual} = 600 \times 15 = 9\,000$

2. Si la empresa aumenta 50 dólares en el salario a todos los empleados, entonces estará invirtiendo cada mes un total de

$9\,000 + 15(50) = 9\,000 + 750 = 9\,750$, al dividir el total entre el número de empleados se tiene:
 $\frac{9\,750}{15} = 650$; por tanto, el nuevo salario mensual será de 650 dólares; observa que es la media que se tenía más los 50 dólares.

3. Al determinar el nuevo salario considerando las dos propuestas, se tiene:

Opción 1: aumento general de 65 dólares

Salario promedio actual: 350
Monto mensual: 8 750
Nuevo gasto mensual: $8\,750 + 65 \times 25 = 10\,375$
Nuevo salario promedio: $\frac{10\,375}{25} = 415$

Opción 2: aumento del 20% sobre el salario actual

Salario promedio actual: 350
Monto mensual: 8 750
Nuevo gasto mensual: $8\,750 + 20\%(8\,750) = 10\,500$
Nuevo salario promedio: $\frac{10\,500}{25} = 420$

Justificación:

- Recomendaría la primera opción, pues aunque el salario promedio es 5 dólares menos que la segunda opción, todos recibirán igual cantidad y es más justo, ya que en el caso de la segunda opción recibirán mayor aumento los que tengan el salario más alto, mientras que los que ganan menos tendrán un menor aumento.



A partir de la definición de la media aritmética $\mu = \frac{\text{Suma de todos los datos } (x)}{n}$, se obtiene que la suma de los datos de una serie es igual a n veces la media aritmética; es decir, $n\mu = \text{Suma de todos los datos } x$. La media aritmética posee algunas propiedades, entre las cuales se tienen:

- Si a todos los valores de la variable se les suma una misma cantidad, la media aritmética queda aumentada en dicha cantidad. Por ejemplo, la serie 3, 4, 5, 4, 9; tiene $\mu = 5$, si a cada dato se le suma 2, se obtiene la serie 5, 6, 7, 6, 11; cuya media es $\mu = 5 + 2 = 7$.
- Si todos los valores de la variable se multiplican por una misma constante, la media aritmética queda multiplicada por dicha constante. Por ejemplo, la serie 3, 4, 5, 4, 9; tiene $\mu = 5$, si cada dato se multiplica por 2, se obtiene la serie 6, 8, 10, 8, 18 cuya media es $\mu = 5(2) = 10$.



Analiza la siguiente situación, luego realiza lo que se pide en cada caso.

1. Durante un mes, el Dr. Martínez llevó un registro del pago realizado por sus pacientes en cada cita a la que asistieron, al final realizó el cálculo y obtuvo un pago promedio de 75 dólares; para el próximo mes ha pensado poner una promoción y tiene las dos propuestas siguientes:

- a) Descuento del 10% sobre el costo total al momento de realizar el pago.
- b) Descuento de 10 dólares sobre el monto a pagar.

Calcula el valor medio de pago en cada caso, ¿cuál opción crees que beneficia más a los pacientes?, ¿por qué?

2. En un supermercado, cada cajera/o al final del turno entrega una venta promedio de \$3,500.00. Con el objetivo de mejorar las ventas, el administrador propone a todos los cajeros las siguientes opciones:

- a) Aumentar las ventas en un 10% sobre el total que entregan en este momento.
- b) Aumentar 300 dólares más de la meta establecida en ese momento.

Calcula el valor medio de venta en cada caso, ¿cuál opción crees que beneficia a la empresa? Justifica tu respuesta.

3. El salario promedio de 3 técnicos es de \$900.00, y el salario promedio de otros 7 técnicos es de \$1,050.00. ¿Cuál es el salario promedio de los 10 técnicos?

4. Un conductor estuvo yendo dos horas a una velocidad promedio de 120 km/hora, la hora siguiente viajó a una velocidad de 90 km/hora. Calcula la velocidad media a la que viajó durante toda la carrera.

2.4 Mediana y moda



¿Cómo se puede determinar la mediana y la moda de una serie de datos organizada en una distribución de frecuencias?

La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal A, de la sala de belleza El Buen Gusto, realiza lo siguiente:

Edades	Número de clientes
20 - 24	8
24 - 28	11
28 - 32	8
32 - 36	2
36 - 40	1
Total	30

1. Identifica la clase donde se encuentra la mediana.
2. Calcula el punto medio de la clase donde se encuentra la mediana.
3. Identifica la clase donde se encuentra la mayor frecuencia.
4. Calcula el punto medio de la clase que tiene la mayor frecuencia.



1. Como la mediana es el dato que ocupa la posición central, entonces es necesario identificar la clase en la que se encuentra el dato central, para ello se suman las frecuencias hasta obtener la mitad del total. Como el total de datos es 30, la mitad es 15, entonces la clase en que se encuentra la mediana es la segunda; pues $8 + 11 = 19$.

2. El punto medio de la segunda clase es $Pm = \frac{24+28}{2} = \frac{52}{2} = 26$.

3. La clase que tiene la mayor frecuencia, para esta distribución es la segunda, de 24 a 28.

4. El punto medio de la clase que tiene la mayor frecuencia es 26.

Edades	Número de clientes	Datos acumulados
20 - 24	8	8
24 - 28	11	19
28 - 32	8	27
32 - 36	2	29
36 - 40	1	30
Total	30	

El punto medio de la clase donde se encuentra la mediana corresponde aproximadamente al dato que ocupa la posición central de la serie, es decir corresponde al valor de la mediana; mientras que el punto medio de la clase de mayor frecuencia corresponde aproximadamente al valor de la moda.



Cuando se tiene una distribución de frecuencias, existen distintos métodos para determinar el valor de la mediana y la moda, en este caso se ha considerado únicamente el método que se conoce como **aproximado**, donde

Para determinar la mediana:

- Se identifica la clase donde queda ubicado el dato que ocupa la posición central $\frac{n}{2}$ **clase mediana**.
- El valor aproximado de la mediana será el punto medio de la clase mediana.

Para determinar la moda:

- Se identifica la clase cuya frecuencia sea mayor **clase modal**.
- El valor aproximado de la moda es el valor medio de la clase modal.



1. La tabla contiene el registro de las edades de 30 clientes atendidos el día de la secretaria, en la sucursal B, de la sala de belleza El Buen Gusto, realiza lo siguiente:

- a) Determina la moda.
- b) Determina la mediana.

2. Construye el polígono de frecuencia de la distribución de datos e identifica el valor de la moda.

Edades	Número de clientes
20 - 24	11
24 - 28	8
28 - 32	6
32 - 36	3
36 - 40	2
Total	30

2.5 Propiedades de las medidas de tendencia central



Compara las siguientes series de datos A, B y C.

A: 3, 4, 5, 5, 7, 8, 10

B: 3, 4, 5, 5, 7, 8, 20

C: 9, 12, 15, 15, 21, 24, 30

1. Determina la moda, mediana y media aritmética para cada serie.
2. Al cambiar el número 10 de la serie A por el 20 en la serie B, ¿qué sucede con el valor de cada una de las siguientes medidas de tendencia central?
3. Al multiplicar por 3 los datos de la serie A, se genera la serie de datos C, ¿qué sucede con los valores de cada una de las siguientes medidas de tendencia central?



1. Para determinar las medidas de tendencia central de cada serie, se trabajan por separado:

Para la serie A: 3, 4, 5, 5, 7, 8, 10 Moda = 5 Mediana = 5 $\mu = \frac{3+4+5+5+7+8+10}{7} = 6$

Para la serie B: 3, 4, 5, 5, 7, 8, 20 Moda = 5 Mediana = 5 $\mu = \frac{3+4+5+5+7+8+20}{7} = 7.4$

Para la serie C: 9, 12, 15, 15, 21, 24, 30 Moda = 15 Mediana = 15 $\mu = \frac{9+12+15+15+21+24+30}{7} = 18$

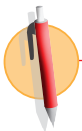
2. Al comparar los valores de la moda, mediana y media aritmética de la serie A con los de la serie B, se puede observar que el valor de la moda y de la mediana, se mantienen, pero el de la media aritmética aumenta.
3. Al comparar los valores de la moda, mediana y media aritmética de la serie A con la serie C, se puede observar que quedan multiplicados por 3. Por ejemplo la moda y la mediana para la serie A es 5 y para la serie C es 15, la media para la serie A es 6 y para la serie C es 18.



Características y usos de las medidas de tendencia central:

La moda, mediana y media aritmética son llamadas medidas de tendencia central, debido a que cuando los datos se ordenan de menor a mayor o viceversa, estas tienden a quedar ubicadas en el centro de la serie.

- **La moda y la mediana** se pueden utilizar para series de datos cualitativos (no numéricos) y cuantitativos (numéricos), además no se ven afectadas por los valores extremos de una serie de datos; tal como se muestra en el numeral 2 del ejemplo anterior.
- **La media aritmética** se utiliza únicamente para series de datos cuantitativos (numéricos); aunque la media es confiable en el sentido de que toma en cuenta todos los valores del conjunto de datos, puede verse afectada por valores extremos que no son representativos del resto de los datos, tal como se muestra en el numeral 2 del ejemplo anterior.



Determina la moda, mediana y media aritmética para cada una de las siguientes series de datos:

A) 0, 1, 1, 2, 2, 3, 3, 3, 4, 5

B) 0, 1, 1, 2, 2, 3, 3, 3, 4, 9

C) 0, 2, 2, 4, 4, 6, 6, 6, 8, 10

D) 0, 5, 5, 10, 10, 15, 15, 15, 20, 25

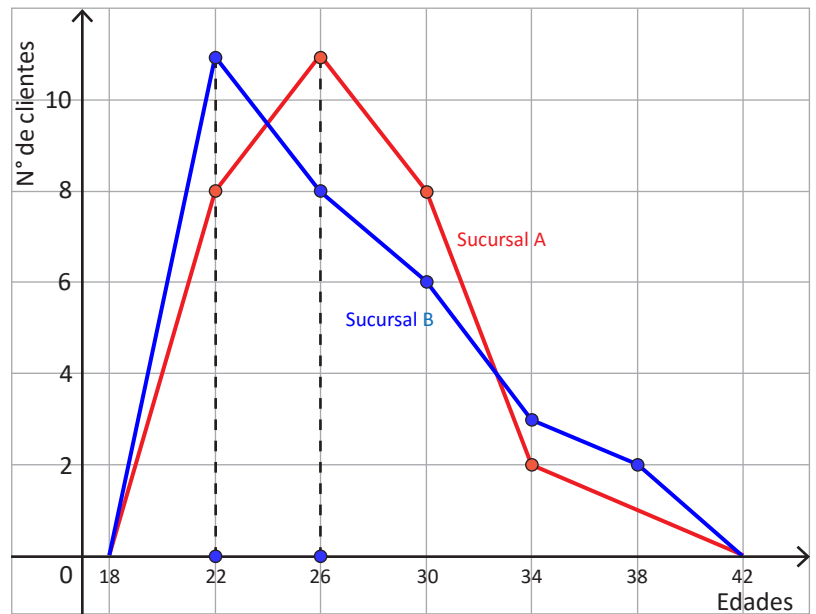
1. Compara los resultados obtenidos, ¿qué concluyes?
2. ¿Qué sucedería con las tres medidas de tendencia central si a cada dato de la serie A se le suma 6?

2.8 Relación entre media, moda y mediana

P

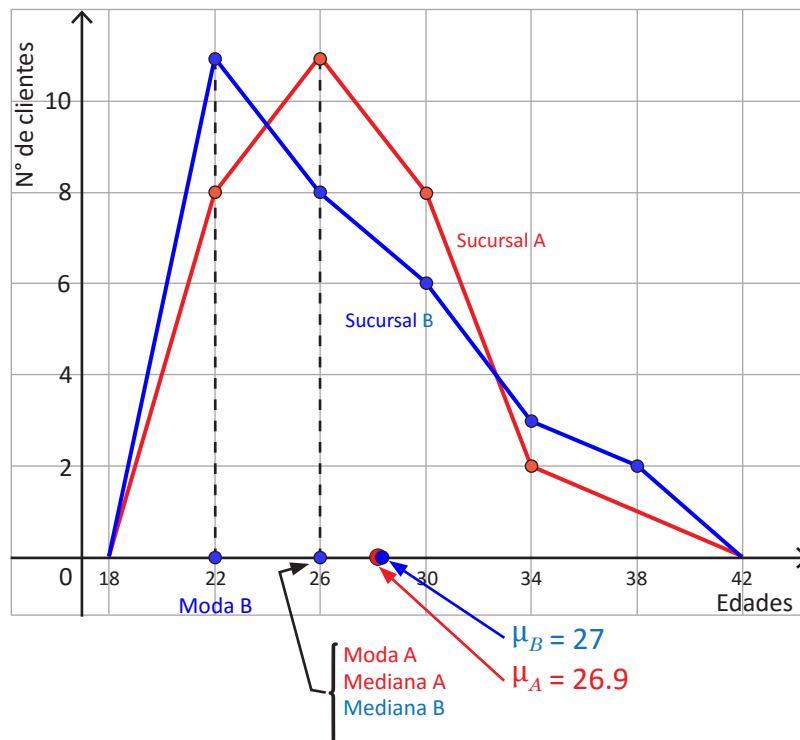
La gráfica corresponde al registro de las edades de 30 clientes atendidos el día de la secretaria en cada una de las dos sucursales A y B de la sala de belleza El Buen Gusto, realiza lo siguiente:

1. Traza una línea vertical para identificar el valor de la moda.
2. Compara los valores de la moda, mediana y media aritmética, luego identifica qué posición le corresponde a la media y la mediana respecto a la moda, para cada distribución (estos valores han sido calculados en clases anteriores).



S

1. En la gráfica se muestra la línea vertical que se traza desde el punto más alto del polígono hacia la recta horizontal o eje x , el punto donde corta al eje x , es el valor aproximado de la moda. Para la sucursal A es 26 y para la sucursal B es 22.



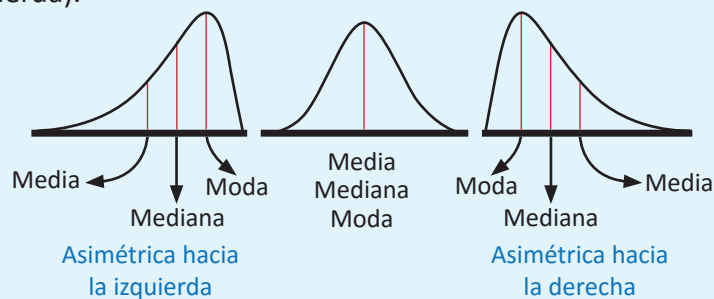
2. Al comparar los valores de la moda, mediana y media aritmética, se tiene:

- Para la distribución correspondiente a la sucursal A, moda = 26, mediana = 26 y $\mu = 26.9$, es decir que la moda y la mediana tienen igual valor y la media aritmética es mayor.
- Para la distribución correspondiente a la sucursal B, moda = 22, mediana = 26 y $\mu = 27$, es decir que la moda es menor que la mediana y esta a su vez es menor que la media aritmética.



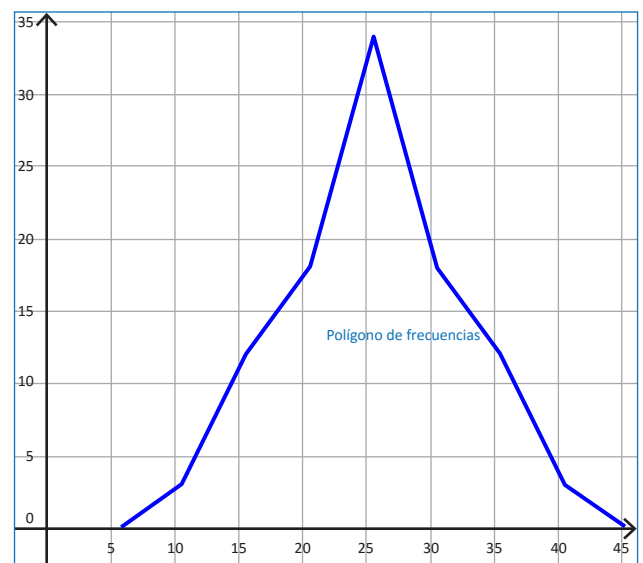
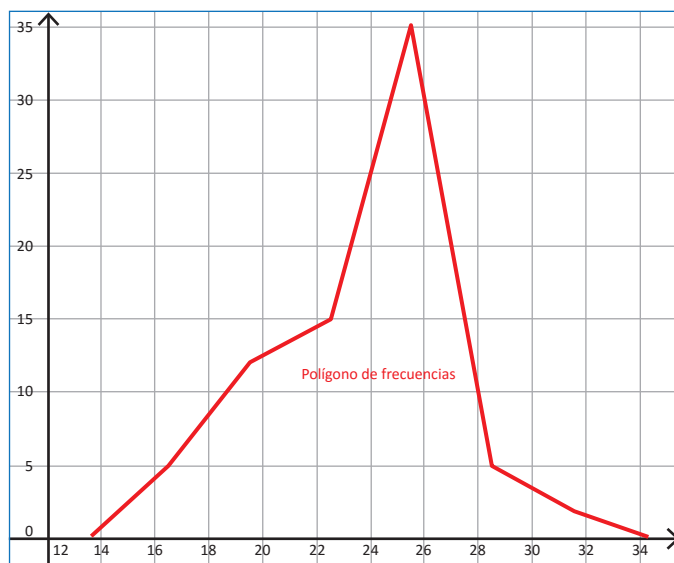
Para una distribución de frecuencias, la forma del gráfico depende de la relación que existe entre el valor de la moda, mediana y media aritmética, es decir:

- Si en una distribución de frecuencias, la moda, mediana y media aritmética tienen igual valor, se dice que es una distribución simétrica.
- Si en una distribución de frecuencias, la moda, mediana y media aritmética tienen la siguiente relación $\text{media} > \text{mediana} > \text{moda}$, se dice que la distribución es asimétrica o con cola a la derecha (sesgada a la derecha).
- Si en una distribución de frecuencias, la moda, mediana y media aritmética tienen la siguiente relación: $\text{media} < \text{mediana} < \text{moda}$, se dice que la distribución es asimétrica o con cola a la izquierda (sesgada a la izquierda).



1. Observa la forma de las siguientes gráficas, las cuales corresponden a una distribución de datos, luego realiza lo siguiente para cada caso:

- Identifica el valor aproximado de la moda.
- Determina la relación entre media, moda y mediana a partir de la forma del gráfico.



2. La distribución de los resultados de la PAES 2016 en un complejo educativo tiene los siguientes valores representativos: media aritmética 7.7, moda 6.5 y mediana 7 puntos.

- A partir de la relación entre los tres valores representativos, describe el tipo de distribución que corresponde a los resultados de la PAES, para ese complejo educativo.
- Elabora un boceto de la distribución representando los valores dados.

PAES, es la Prueba de Aprendizaje y Aptitudes que el Ministerio de Educación aplica a estudiantes egresados de Educación Media del sector público y privado.