ESTADÍSTICA

Trimestre 25-O

UNIDAD I. Introducción a la estadística.

1.1 Definiciones básicas y conceptos

1.2 Escalas de medición fundamentales.

1.3 Medidas de tendencia central y de dispersión.

1.4 Distribución de frecuencias.

1.5 Distribución normal y distribuciones muestrales.

Introducción

Estadística significa ciencia del Estado, y proviene del término alemán Statistik.

¿Por qué la ciencia del Estado?

- Porque en sus orígenes la estadística se utilizaba exclusivamente con fines estatales, en el sentido de que los gobiernos de las distintas naciones tenían (y tienen) la necesidad, por razones de organización, de conocer las características de su población para gestionar el pago de impuestos, el reclutamiento de soldados, el reparto de tierras o bienes, la prestación de servicios públicos etc. Esta necesidad llevó a los gobernantes a establecer sistemas para recoger y procesar de alguna manera la información obtenida, es decir, a hacer estadísticas sobre la población.

El papel de la Estadística es extraer información de los datos para adquirir el conocimiento necesario para tomar decisiones.

$$\left.\begin{array}{l}\text{Información buena}\\ \text{Información suficiente}\\ \text{Procesamiento correcto}\end{array}\right\} \Rightarrow \quad \text{buena decisión}$$

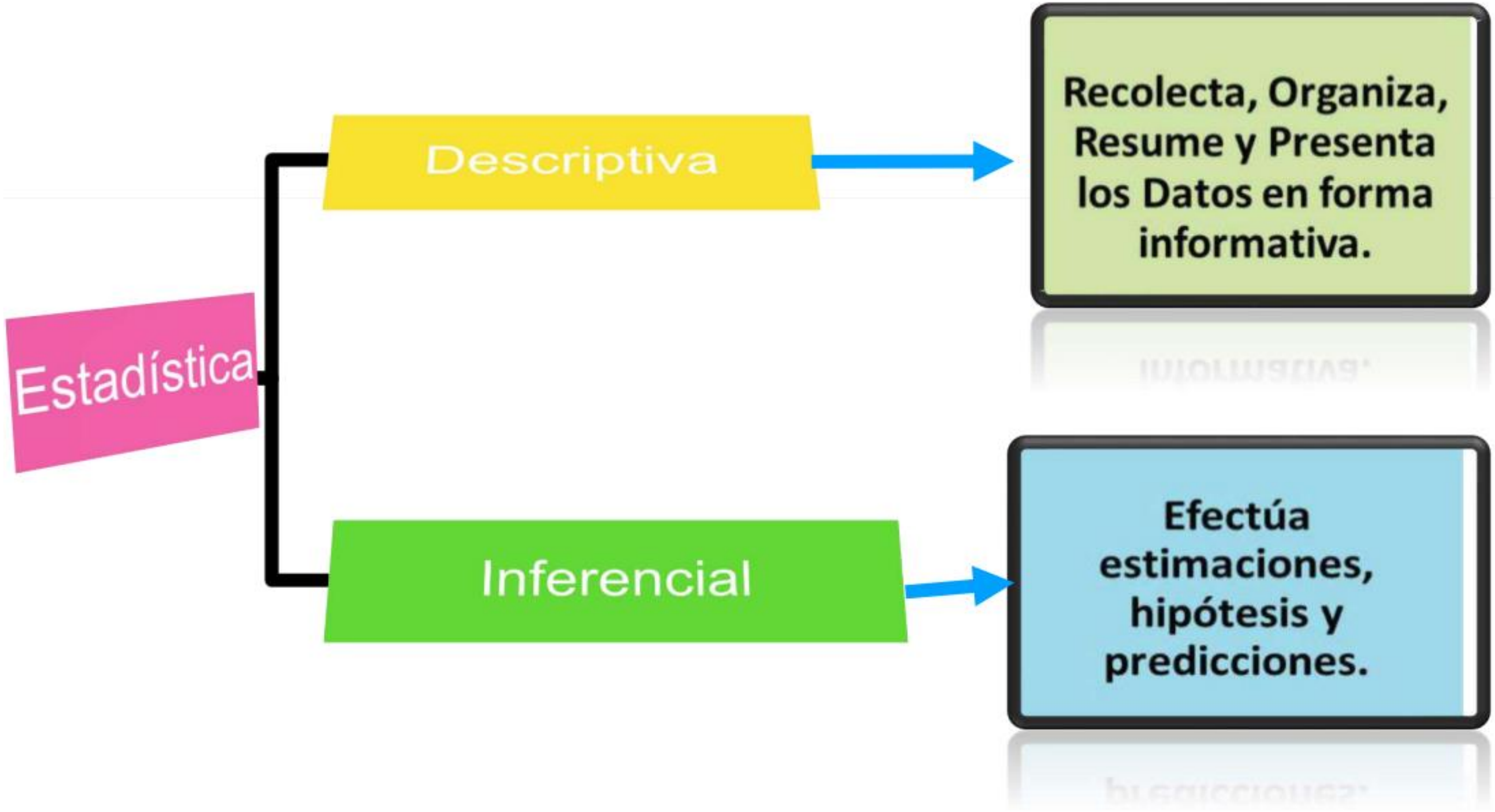
Estadística

Descriptiva

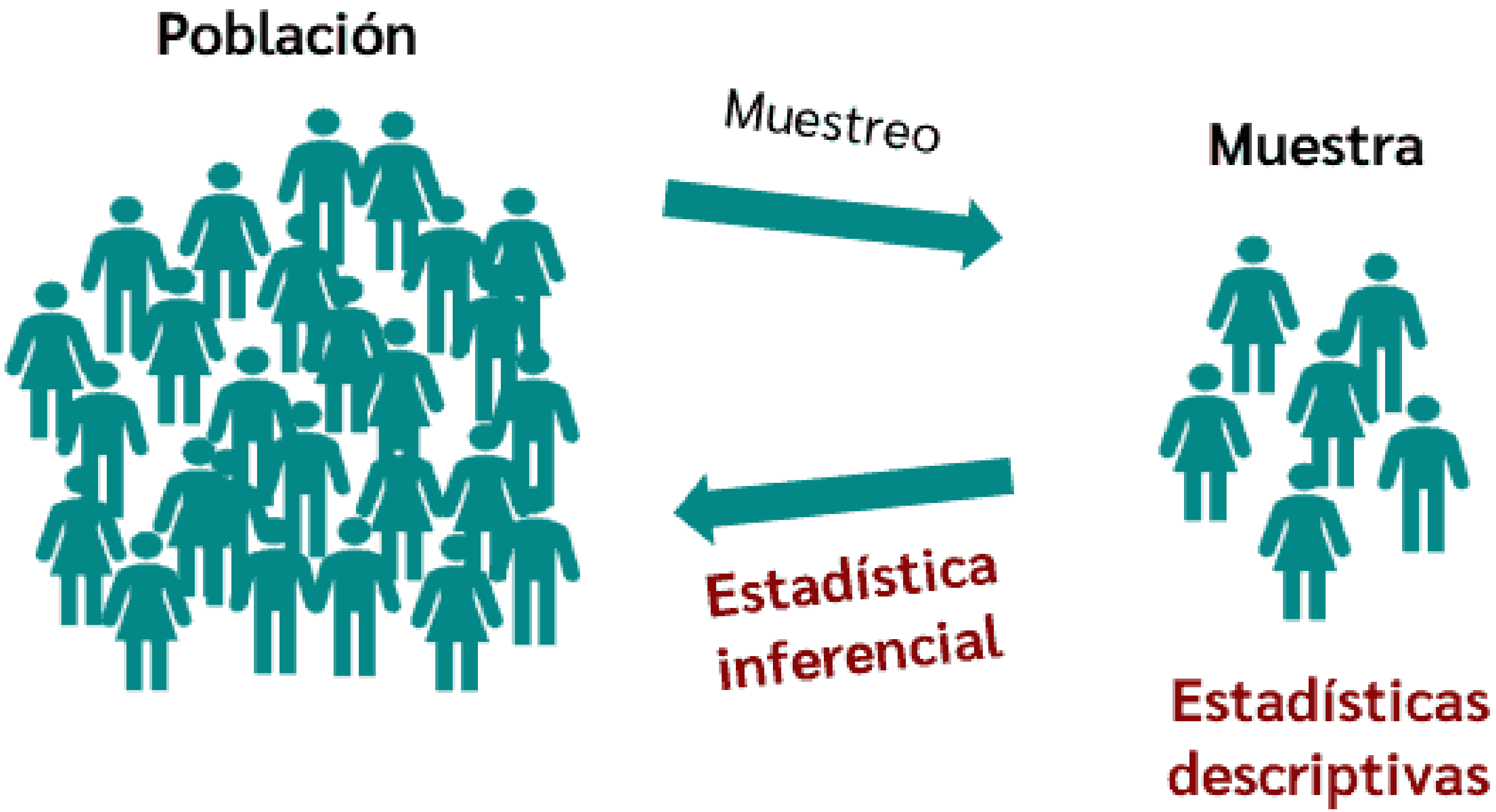
Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

Inferencial

Efectúa estimaciones, toma de decisiones, formula predicciones apoyándose en cálculo de probabilidades y a partir de datos muestrales.



Estadística
Descriptiva
Recolecta, Organiza, Resume y Presenta los Datos en forma informativa.
Inferencial
Efectúa estimaciones, hipótesis y predicciones.


Población
Muestreo
Muestra
Estadística inferencial
Estadísticas descriptivas

Áreas de aplicación de la estadística

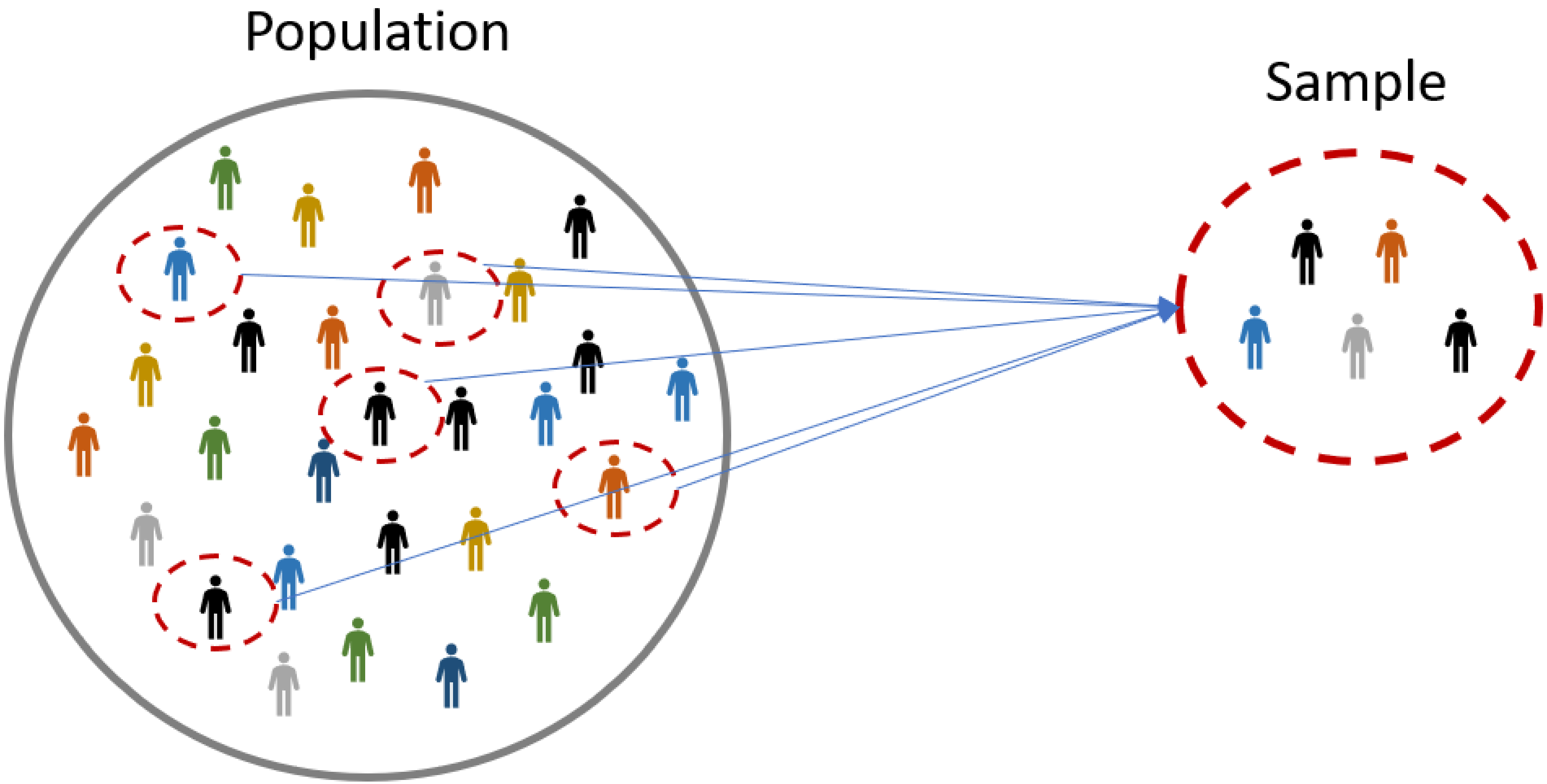
- La estadística se utiliza en todas las áreas del conocimiento, ya sean humanísticas, técnicas, científicas, laborales, deportivas, etc. Esto es, actualmente resulta difícil indicar algún área o ciencia que no utilice la Estadística.

- La estadística se aplica en la Ingeniera, Medicina, Psicología, Economía, Geografía, Física, Química, Agronomía, Administración, Biología, Ecología, Antropología, Historia, Contaduría, Planeación, Política, etc., y aunque los problemas de cada área o ciencia son diferentes, las técnicas que se utilizan para el análisis estadístico son las mismas, debido a que se trabaja en la mayoría de los casos con datos numéricos

- La estadística es fundamental para un Químico Farmacéutico Biólogo (QFB) ya que proporciona herramientas para analizar datos, hacer inferencias sobre poblaciones y tomar decisiones basadas en evidencia científica. Las estadísticas descriptivas resumen datos, mientras que las inferenciales permiten hacer predicciones y probar hipótesis.
- Para un QFB, esto se aplica al diseño de experimentos, al control de calidad, a la interpretación de resultados de laboratorio y a la evaluación de la eficacia de medicamentos.

Conceptos básicos

- Una **población** es el conjunto de todos los elementos a los que se somete a un estudio estadístico.
- Un **individuo** o **unidad estadística** es cada uno de los elementos que componen la población.
- Una **muestra** es un subconjunto representativo de la población de referencia.
- Una **variable** es una característica que varía de individuo en individuo.
- Los **datos** son los valores de la variable en estudio.


Population
Sample

Variables estadísticas

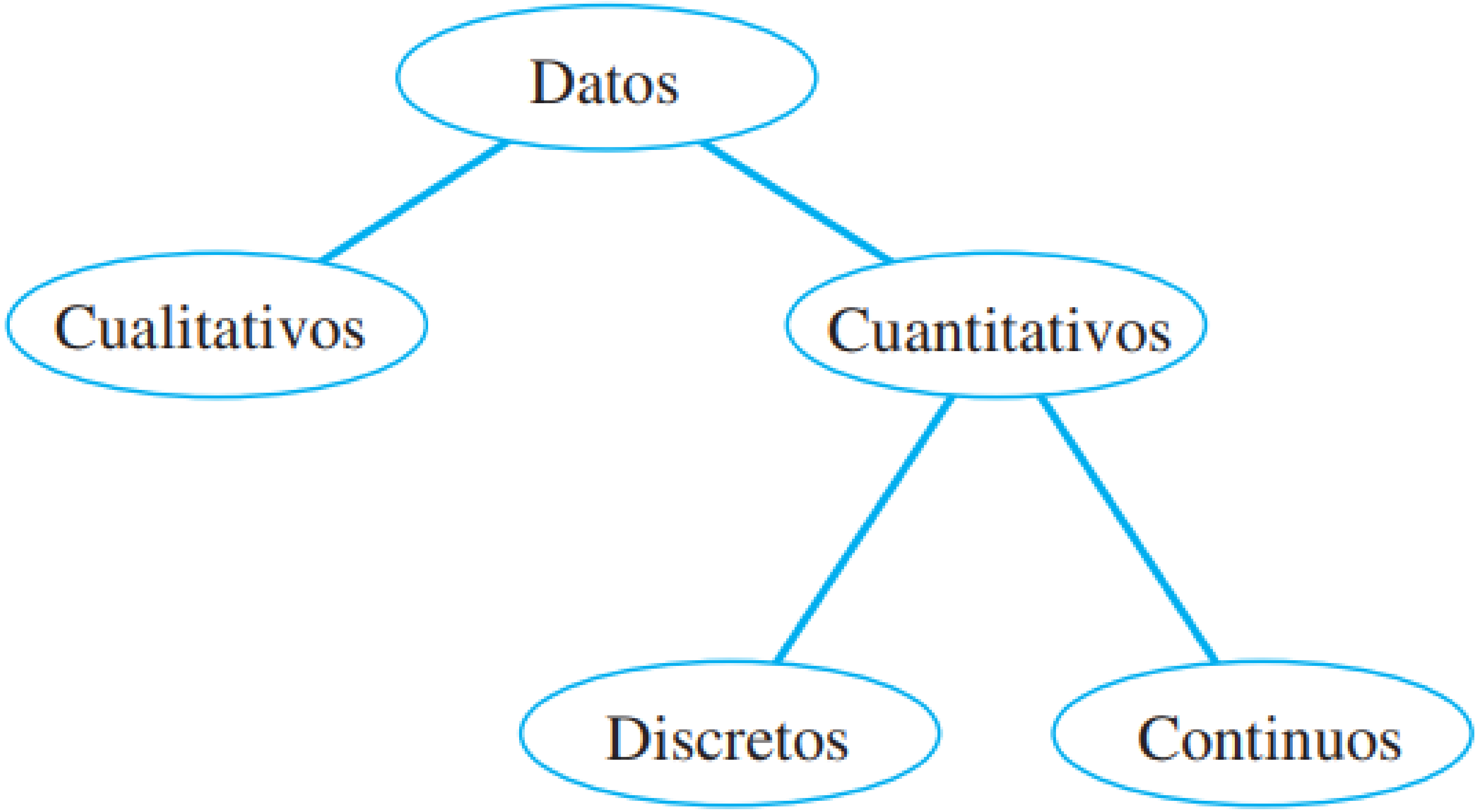
Si

- todos opináramos igual respecto a la elección de un determinado candidato
- todos tuviésemos el mismo peso y altura
- todos reaccionáramos de idéntica forma ante un medicamento
- los productos resultantes de un proceso industrial fuesen idénticos
- al repetir un experimento se obtuviera siempre el mismo resultado

no serían necesarios los análisis estadísticos

Podremos distinguir dos tipos de **variabilidad:**

1. Variabilidad debido a diferencias entre individuos respeto de alguna característica.
2. Variabilidad debida a errores de medición.


Datos
Cualitativos
Cuantitativos
Discretos
Continuos

Errores en el Proceso de Medición

En los trabajos científicos se manejan dos tipos de números:

- números exactos (cuyos valores se conocen con exactitud) y
- números inexactos (cuyos valores tienen cierta incertidumbre).

Los números que se obtienen midiendo siempre son inexactos. Siempre hay limitaciones inherentes en el equipo empleado para medir cantidades (errores de equipo), y hay diferencias en la forma en que diferentes personas realizan la misma medición (errores humanos).

Precisión y exactitud

La distribución de los dardos en un blanco ilustra la diferencia entre exactitud y precisión

Buena exactitud
Buena precisión

Mala exactitud
Buena precisión

Mala exactitud
Mala precisión

La precisión es una medida de la concordancia de mediciones individuales entre sí.

La exactitud se refiere a qué tanto las mediciones individuales se acercan al valor correcto, o “verdadero”.

Muestreo

En ocasiones es posible estudiar toda la población de interés; esto se denomina Censo.

Pero muchas veces no es posible:

- No es posible analizar la totalidad de un tanque lleno de leche (contenido de grasa)
- No es posible analizar toda el agua de un río (contaminante)
- Muchas técnicas de análisis son destructivas y por lo tanto no pueden aplicarse a la totalidad de un objeto.

El proceso de selección de los elementos que compondrán una muestra se conoce como **muestreo**.

- Para que una muestra refleje información fidedigna sobre la población global debe ser representativa de la misma, lo que significa que debe reproducir a pequeña escala la variabilidad de la población.
- El objetivo es obtener una muestra representativa de la población.

EJEMPLO. Como parte del control de calidad se desea obtener el peso promedio de las pastillas de un gran lote. Elegimos 10 y pesamos cada una de ellas.

- **Población**: todas las pastillas del lote
- **Variable**: peso de una pastilla cualquiera del lote
- **Valores poblacionales** de la variable peso: peso de cada una de las pastillas del lote.
- **Datos**: el peso de cada una de las 10 pastillas

--- pero….. **¿Cómo elegimos las 10 pastillas del lote?**

Mediante un procedimiento de **MUESTREO ALEATORIO**.

Decimos que nuestros datos son una **muestra aleatoria**.

- **MUESTREO ALEATORIO**: procedimiento de selección por el cual todos los miembros de la población tienen la misma posibilidad de ser elegidos.

Determinación del tamaño muestral

- Una de las preguntas más interesantes que surge inmediatamente es: **¿cuántos individuos es necesario tomar en la muestra para tener un conocimiento aproximado pero suficiente de la población?**

- La respuesta depende de varios factores, como la variabilidad de la población o la fiabilidad deseada para las extrapolaciones que se hagan hacia la población.

- En general, cuantos más individuos haya en la muestra, más fiables serán las conclusiones sobre la población, pero también será más lento y costoso el estudio.

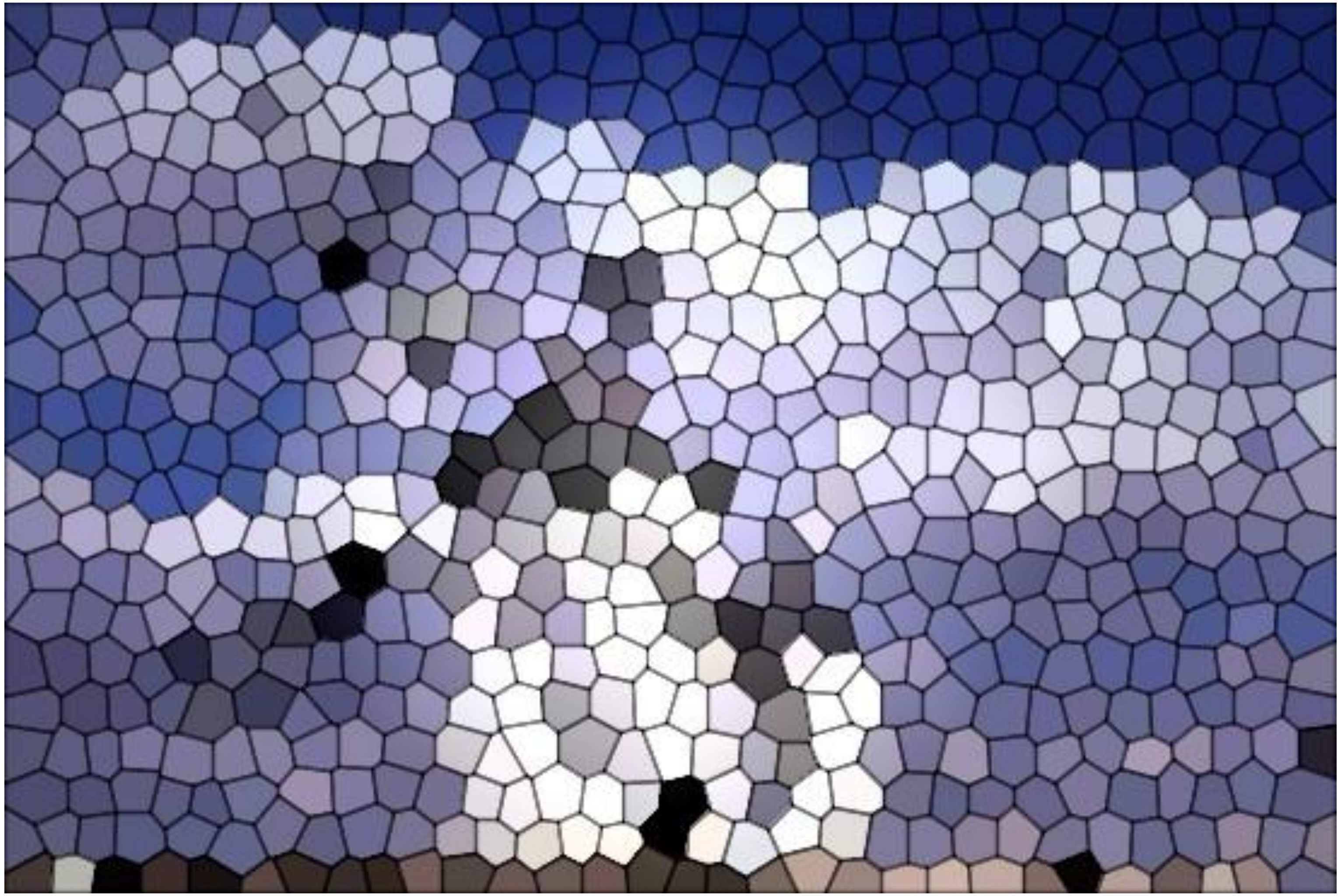
- Para entender a qué nos referimos cuando hablamos de un tamaño muestral suficiente para comprender lo que ocurre en la población, podemos utilizar el siguiente símil en que se trata de comprender el motivo que representa una fotografía.

- Una fotografía digital está formada por multitud de pequeños puntitos llamados pixels que se dispone en una enorme tabla de filas y columnas (cuantas más filas y columnas haya se habla de que la foto tiene más resolución). Aquí la población estaría formada por todos y cada uno de los píxeles que forman la foto. Por otro lado cada pixel tiene un color y es la variedad de colores a lo largo de los pixels la que permite formar la imagen de la fotografía.

- ¿Cuántos píxeles debemos tomar en una muestra para averiguar la imagen de la foto?

- La respuesta depende de la variabilidad de colores en la foto. Si todos los pixels de la foto son del mismo color, entonces un sólo pixel basta para desvelar la imagen. Pero, si la foto tiene mucha variabilidad de colores, necesitaremos muchos más pixels en la muestra para descubrir el motivo de la foto.

La imagen siguiente contiene una muestra pequeña de píxeles de una foto.
¿Puedes averiguar el motivo de la foto?



¡Con una muestra pequeña es difícil averiguar el contenido de la imagen!

La siguiente imagen contiene una muestra mayor de píxeles.
¿Eres capaz de adivinar el motivo de la foto ahora?

¡Con una muestra mayor es posible desvelar el motivo de la foto!

Y aquí está la población completa.

Lo importante es que **¡No es necesario conocer todos los píxeles para averiguar la imagen!**

Tipos de razonamiento

- Habitualmente realizaremos el estudio de la población a partir de muestras y luego trataremos de extrapolar lo observado en la muestra al resto de la población. A este tipo de razonamiento que saca conclusiones desde la muestra hacia la población se le conoce como **razonamiento inductivo**.

- Características de la **deducción**: Si las premisas son ciertas, garantiza la certeza de las conclusiones (es decir, si algo se cumple en la población, también se cumple en la muestra). Sin embargo, ¡no aporta conocimiento nuevo!

- Características de la **inducción**: No garantiza la certeza de las conclusiones (si algo se cumple en la muestra, puede que no se cumpla en la población, así que ¡cuidado con las extrapolaciones!), pero ¡es la única forma de generar conocimiento nuevo!

- La estadística se apoya fundamentalmente en el razonamiento inductivo ya que utiliza la información obtenida a partir de muestras para sacar conclusiones sobre las poblaciones.

- A diferencia del razonamiento deductivo que va de lo general a lo particular, o de la población a la muestra, el razonamiento inductivo no garantiza la certeza de las conclusiones, por lo que debemos ser cuidadosos a la hora de generalizar sobre la población lo observado en al muestra, ya que si la muestra no es representativa de la población o contiene sesgos, las conclusiones pueden ser erróneas.

Modalidades de muestreo

Existen muchas técnicas de muestreo pero se pueden agrupar en dos categorías:

- **Muestreo Aleatorio**: Elección aleatoria de los individuos de la muestra. Todos tienen la misma probabilidad de ser elegidos (equiprobabilidad).

- **Muestreo No Aleatorio**: Los individuos se eligen de forma no aleatoria. Algunos individuos tienen más probabilidad de ser seleccionados que otros.

Sólo las técnicas aleatorias evitan el sesgo de selección, y por tanto, garantizan la representatividad de la muestra extraída, y en consecuencia la validez de las conclusiones.

Las técnicas no aleatorias no sirven para hacer generalizaciones, ya que no garantizan la representatividad de la muestra. Sin embargo, son menos costosas y pueden utilizarse en estudios exploratorios.

Muestreo aleatorio simple

Dentro de las modalidades de muestreo aleatorio, el tipo más conocido es el muestreo aleatorio simple, caracterizado por:

- Todos los individuos de la población tienen la misma probabilidad de ser elegidos para la muestra.
- La selección de individuos es con reemplazamiento, es decir, cada individuo seleccionado es devuelto a la población antes de seleccionar al siguiente (y por tanto no se altera la población de partida).
- Las sucesivas selecciones de un individuo son independientes.

La única forma de realizar un muestreo aleatorio es asignar un número a cada individuo de la población (censo) y realizar un sorteo aleatorio.

Variables estadísticas

Todo estudio estadístico comienza por la identificación de las características que interesa estudiar en la población y que se medirán en los individuos de la muestra.

Una **variable estadística** es una propiedad o característica medida en los individuos de la población.

Los **datos** son los valores observados en las variables estadísticas.

Estas características pueden ser de distintos tipos de acuerdo a su naturaleza y su escala:

Variables cualitativas o atributos: Miden cualidades no numéricas. Pueden ser:

- **Nominales**: No existe un orden entre las categorías.

  Ejemplo: El color de pelo o el sexo.

- **Ordinales**: Existe un orden entre las categorías.

  Ejemplo: El nivel de estudios o la gravedad de una enfermedad.

Variables cuantitativas: Miden cantidades numéricas. Pueden ser:

- **Discretas**: Toman valores numéricos aislados (habitualmente números enteros).

  Ejemplo: El número de hijos o el número de coches en una familia.

- **Continuas**: Pueden tomar cualquier valor en un intervalo real.

  Ejemplo: El peso o la estatura.

Variables cualitativas

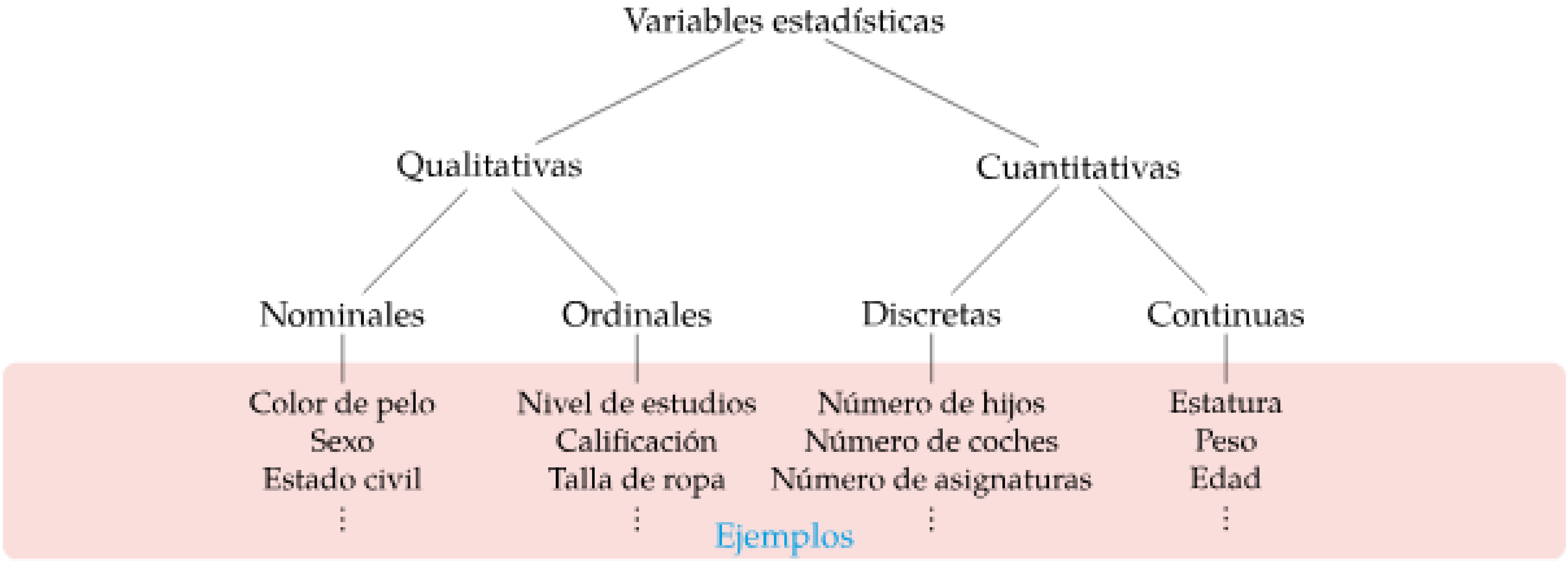
Nombre	Color de pelo	Color de ojos
Kiara	Café	Azul
Simba	Gris	Café
Rocky	Gris	Gris
Nala	Blanco	Amarillo
Bruno	Café	Café

Variables cuantitativas

25 cm

52 kg


Tipos de Variables
Cualitativas
(categóricas)
Nominales
A B C D
Clasificar
Ordinales
I II III IV
Jerarquizar
Binarias
alto - bajo
frio - caliente
Escoger
Cuantitativas
(numéricas)
Discretas
1 2 3 4
Contar
Continuas
1.5 7.4 0.5
Medir


Variables estadísticas
Qualitativas
Cuantitativas
Nominales
Ordinales
Discretas
Continuas
Color de pelo
Sexo
Estado civil
Nivel de estudios
Calificación
Talla de ropa
Número de hijos
Número de coches
Número de asignaturas
Estatura
Peso
Edad
Ejemplos

Elección del tipo de variable más apropiado

En ocasiones una característica puede medirse mediante variables de distinto tipo.

Ejemplo. Si una persona fuma o no podría medirse de diferentes formas:

Fuma: si/no. (Nominal)

Nivel de fumador: No fuma / ocasional / moderado / bastante / empedernido. (Ordinal)

Número de cigarros diarios: 0,1,2,… (Discreta)

En estos casos es preferible usar variables cuantitativas a cualitativas. Dentro de las cuantitativas es preferible usar las continuas a las discretas y dentro de las cualitativas es preferible usar ordinales a nominales pues aportan más información.

Cantidad de información de los tipos de variables estadísticas.

De acuerdo al papel que juegan en el estudio las variables también pueden clasificarse como:

- **Variables independientes**: Variables que supuestamente no dependen de otras variables en el estudio. Habitualmente son las variables manipuladas en el experimento para ver su efecto en las variables dependientes. Se conocen también como variables predictivas.

- **Variables dependientes**: Variables que supuestamente dependen de otras variables en el estudio. No son manipuladas en el experimento y también se conocen como variables respuesta.

Ejemplo. En un estudio sobre el rendimiento de los alumnos de un curso, la inteligencia de los alumnos y el número de horas de estudio diarias serían variables independientes y la nota del curso sería una variable dependiente.

Ejemplo. Trabaje en colaboración para determinar el tipo de datos correcto (cuantitativo o cualitativo). Indique si los datos cuantitativos son continuos o discretos.
Pista: Los datos que son discretos suelen empezar con las palabras "el número de".

a. el número de pares de zapatos que tiene
b. el tipo de automóvil que conduce
c. la distancia que hay desde su casa hasta la tienda de comestibles más cercana
d. el número de clases que se imparten por año escolar.
e. el tipo de calculadora que utiliza
f. pesos de luchadores de sumo
g. número de respuestas correctas en un cuestionario
h. Calificaciones de IQ (esto puede provocar alguna discusión).

Solución
Los ítems a, d y g son cuantitativamente discretos; los ítems c, f y h son cuantitativamente continuos; los ítems b y e son cualitativos o categóricos.

Tipos de estudios estadísticos

Dependiendo de si se manipulan las variables independientes existen dos tipos de estudios:

Experimentales: Cuando las variables independientes son manipuladas para ver el efecto que producen en las variables dependientes.

- Ejemplo. En un estudio sobre el rendimiento de los estudiantes en un test, el profesor manipula la metodología de estudio para crear dos o más grupos con metodologías de estudio distintas.

No experimentales: Cuando las variables independientes no son manipuladas. Esto no significa que sea imposible hacerlo, sino que es difícil o poco ético hacerlo.

- Ejemplo. En un estudio un investigador puede estar interesado en el efecto de fumar sobre el cáncer de pulmón. Aunque es posible, no sería ético pedirle a los pacientes que fumasen para ver el efecto que tiene sobre sus pulmones. En este caso, el investigador podría estudiar dos grupos de pacientes, uno con cáncer de pulmón y otro sin cáncer, y observar en cada grupo cuántos fuman o no.

Tipos de estudios estadísticos

Los estudios experimentales permiten identificar causas y efectos entre las variables del estudio, mientras que los no experimentales sólo permiten identificar relaciones de asociación entre las variables.

Presentación de los datos mediante tablas

- Las variables a estudiar se medirán en cada uno de los individuos de la muestra, obteniendo un conjunto de datos que suele organizarse en forma de matriz que se conoce como tabla de datos.

- En esta tabla cada columna contiene la información de una variable y cada fila la información de un individuo.

Ejemplo. La siguiente tabla contiene información de las variables Nombre, Edad, Sexo, Peso y Altura de una muestra de 6 personas.

Nombre	Edad	Sexo	Peso(Kg)	Altura(cm)
José Luis Martínez	18	H	85	179
Rosa Díaz	32	M	65	173
Javier García	24	H	71	181
Carmen López	35	M	65	170
Marisa López	46	M	51	158
Antonio Ruiz	68	H	66	174

Datos agrupados

- son datos que se agrupan en intervalos para facilitar su análisis.
- Se agrupan los datos en clases de acuerdo a ciertas características

Ejemplo

Un ejemplo de datos agrupados sería el siguiente, donde hemos resumido la información sobre los ingresos mensuales de un grupo de personas:

Ingresos mensuales	**Frecuencia**
[1.500-2.500]	120
(2.500-3.500]	210
(3.500-4.500]	300
(4.500-5.500]	250
(5.500-6.500]	400
(6.500-7.500]	510
(7.500-8.500]	420
(8.500-9.500]	416
(9.500-10.500]	100

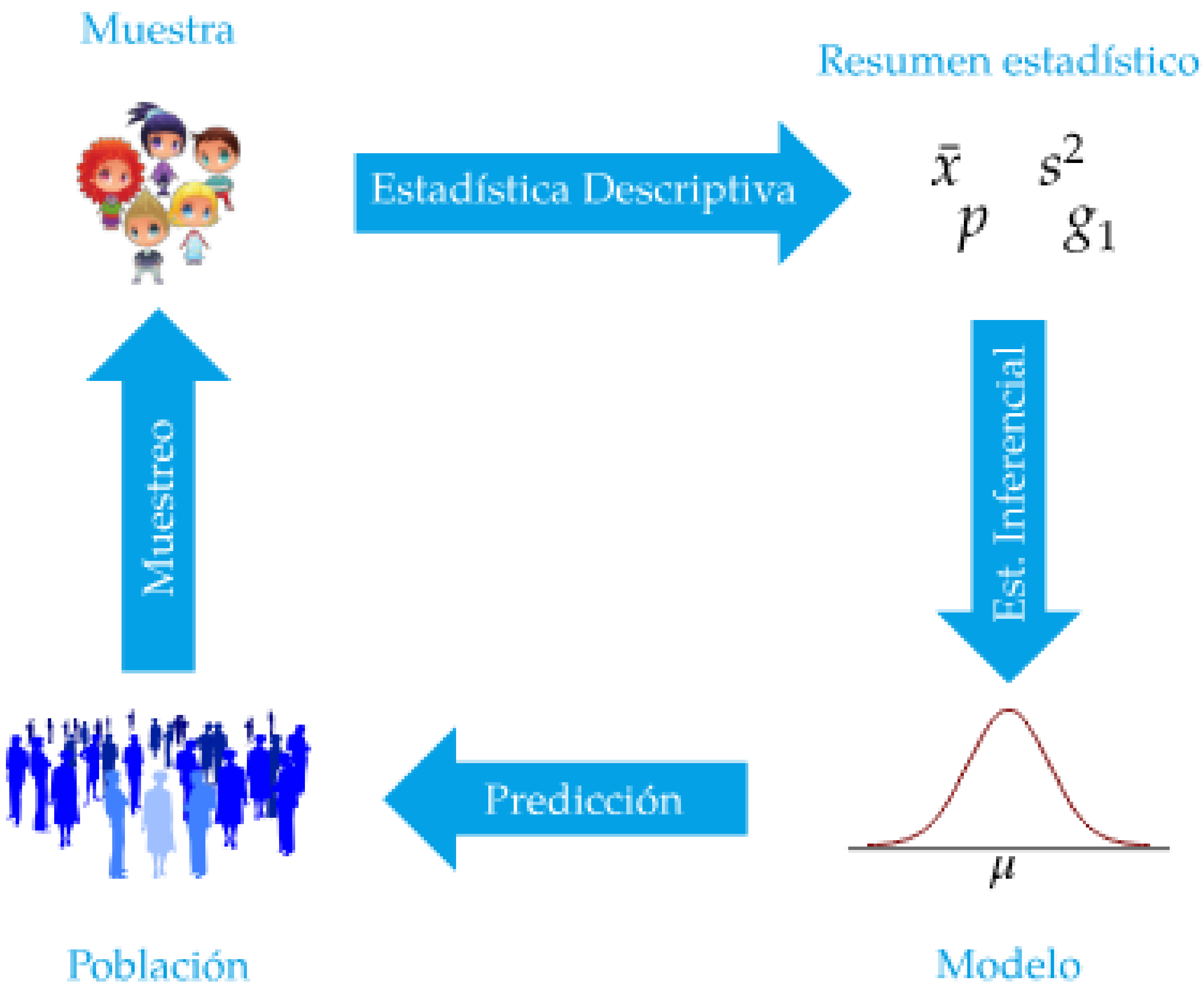
En la tabla podemos observar que, por ejemplo, 210 personas de la muestra tienen ingresos mensuales de entre 2.500 y 3.500 euros.

Fases del análisis estadístico

Normalmente un estudio estadístico pasa por las siguientes etapas:

1. El estudio comienza por el diseño previo del mismo en el que se establezcan los objetivos del mismo, la población, las variables que se medirán y el tamaño muestral requerido.
2. A continuación se seleccionará una muestra representativa del tamaño establecido y se medirán las variables en los individuos de la muestra obteniendo la tabla de datos. De esto se encarga el Muestreo.
3. El siguiente paso consiste en describir y resumir la información que contiene la muestra. De esto se encarga la Estadística Descriptiva.
4. La información obtenida es proyectada sobre un modelo matemático que intenta explicar el comportamiento de la población y el modelo se valida. De todo esto se encarga la Estadística Inferencial.
5. Finalmente, el modelo validado nos permite hacer predicciones y sacar conclusiones sobre la población de partida con cierta confianza.

El ciclo estadístico



Estadística Descriptiva

La estadística descriptiva es la parte de la estadística encargada de representar, analizar y resumir la información contenida en la muestra.

Tras el proceso de muestreo, es la siguiente etapa de todo estudio estadístico y suele consistir en:

1. Clasificar, agrupar y ordenar los datos de la muestra.
2. Tabular y representar gráficamente los datos de acuerdo a sus frecuencias.
3. Calcular medidas que resuman la información que contiene la muestra (estadísticos muestrales).

Interpretación:

No tiene poder inferencial, por lo que nunca deben sacarse conclusiones sobre la población a partir de las medidas resumen que aporta la Estadística Descriptiva.

Construcción de tablas de frecuencia

- Definición de distribución de frecuencias
- Construcción de una tabla de frecuencias
- Tipos de frecuencias
- Distribución de frecuencias agrupadas
- Construcción de una tabla de frecuencis con datos agrupados

Clasificación de la muestra

Consiste colocar juntos los valores iguales y ordenarlos si existe un orden entre ellos.

Recuento de frecuencias

Distribución

El **patrón de variación** de una variable es llamado distribución. La distribución registra los valores numéricos de la variable y cuán frecuentemente ocurre cada valor.

Distribución de frecuencias

El estudio de una variable estadística comienza por medir la variable en los individuos de la muestra y clasificar los valores obtenidos.

Existen dos formas de clasificar estos valores:

1. **Sin agrupar**: Ordenar todos los valores obtenidos en la muestra de menor a mayor. Se utiliza con atributos y variables discretas con pocos valores diferentes.

2. **Agrupados**: Agrupar los valores en clases (intervalos) y ordenar dichas clases de menor a mayor. Se utiliza con variables continuas y con variables discretas con muchos valores diferentes.

Tabla de frecuencias

- **Frecuencia absoluta**: es la cantidad de veces que aparece el valor en el estudio. La sumatoria de las frecuencias absolutas es igual al número de datos.
- **Frecuencia acumulada**: es el acumulado o suma de las frecuencias absolutas, indica cuantos datos se van contando hasta ese momento.
- **Frecuencia relativa**: es la fracción o proporción de elementos que pertenecen a una clase o categoría. Se calcula dividiendo la frecuencia absoluta entre el número de datos del estudio.
- **Frecuencia relativa acumulada**: es la proporción de datos respecto al total que se han reportado hasta ese momento. Es la suma de las frecuencias relativas, y se puede calcular también dividiendo la frecuencia acumulada entre el número de datos del estudio.
- **Frecuencia porcentual**: es el porcentaje de elementos que pertenecen a una clase o categoría. Se puede calcular multiplicando la frecuencia relativa por 100%.
- **Frecuencia porcentual acumulada**: es el porcentaje de datos respecto al total que se han reportado hasta ese momento. Se puede calcular multiplicando la frecuencia relativa acumulada por 100%.

¿Qué es una tabla de frecuencias?

- La **distribución de frecuencias** o **tabla de frecuencias** es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

* Puedes usar las tablas de frecuencias para ordenar variables cuantitativas o cualitativas.

Tipos de frecuencias

- Frecuencias absolutas
- Frecuencias relativas
- Frecuencia acumuladas

Frecuencia absoluta

- Es el número de veces que aparece un determinado valor en un estudio estadístico.
- Se representa por f_i, aunque otros autores la representan como n_i.

- La suma de las frecuencias absolutas es igual al número total de datos, que se representa por N.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

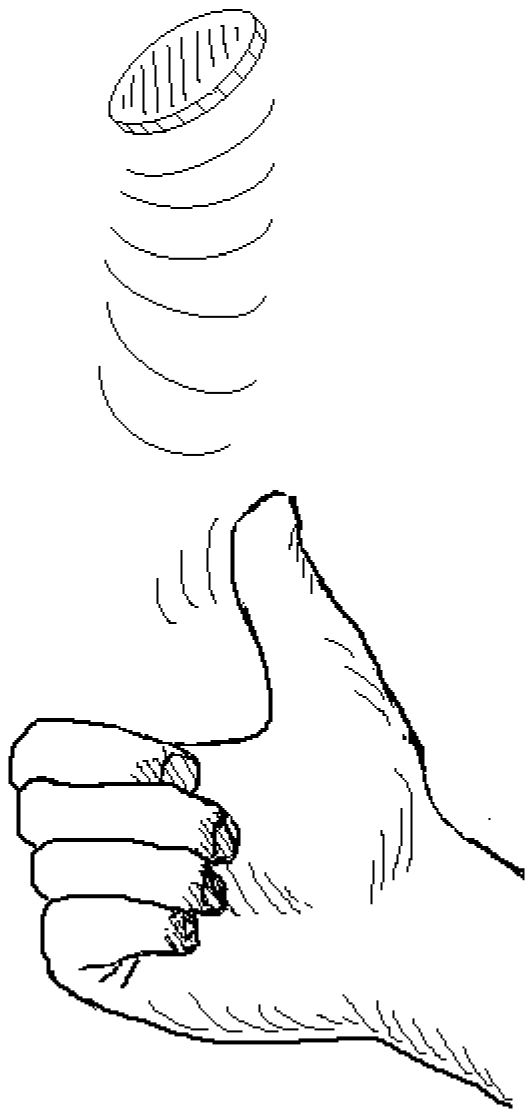
- Para indicar resumidamente estas sumas se utiliza la letra griega $\sum$ (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^{i=n} f_i = N$$

Ejemplo. Al tirar una moneda 50 veces, salen 35 caras.

$$f_{cara} = 35$$

$$f_{cruz} = 15$$



- La suma de las frecuencias absolutas es igual al número total de datos:

$$N = 35 + 15 = 50$$

Frecuencia relativa

- Es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos.
- Se puede expresar en tanto por ciento y se representa por ***n_i***.

$$n_i = \frac{f_i}{N}$$

- La frecuencia relativa es un número comprendido entre 0 y 1.
- La suma de las frecuencias relativas es igual a 1.

$$n_{cara} = \frac{35}{50}$$

$$n_{cruz} = \frac{15}{50}$$

$$n_{cara} + n_{cruz} = \frac{35}{50} + \frac{15}{50} = 1$$

Frecuencia acumulada

- Es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

- Se representa por F_i.

Frecuencia relativa acumulada

- Es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos.

- Se puede expresar en tantos por ciento.

Ejemplo: Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

$$32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32,$$
$$31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29$$

x_i	f_i	F_i	n_i	N_i

x_i	f_i	F_i	n_i	N_i
27	1	1	0.032	0.032
28	2	3	0.065	0.097
29	6	9	0.194	0.290
30	7	16	0.226	0.516
31	8	24	0.258	0.774
32	3	27	0.097	0.871
33	3	30	0.097	0.968
34	1	31	0.032	1
	31		1	

Este tipo de tablas de frecuencias se utilizan con variables discretas.

Distribución de frecuencias agrupadas

- La distribución de frecuencias agrupadas o tabla con datos agrupados se emplea si las variables toman un número grande de valores o la variable es continua.

- Se agrupan los valores en intervalos que tengan la misma amplitud denominados clases. A cada clase se le asigna su frecuencia correspondiente.

Límites de la clase

- Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase.

Amplitud de la clase

- es la diferencia entre el límite superior e inferior de la clase.

$$a_i = Límite\ superior\ de\ la\ clase - Límite\ inferior\ de\ la\ clase$$

Marca de clase

- es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros.

$$c_i = \frac{Límite\ superior\ de\ la\ clase + Límite\ inferior\ de\ la\ clase}{2}$$

Construcción de una tabla de datos agrupados

$$3,\ 15,\ 24,\ 28,\ 33,\ 35,\ 38,\ 42,\ 43,\ 38,\ 36,\ 34,\ 29,\ 25,\ 17,\ 7,\ 34,\ 36,\ 39,\ 44,$$
$$31,\ 26,\ 20,\ 11,\ 13,\ 22,\ 27,\ 47,\ 39,\ 37,\ 34,\ 32,\ 35,\ 28,\ 38,\ 41,\ 48,\ 15,\ 32,\ 13$$

1. Se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

2. Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos que queremos establecer. Es conveniente que el número de intervalos oscile entre 6 y 15. En este caso, 48 – 3 = 45, incrementamos el número hasta 50,
$50 \div 5 = 10$ intervalos.

Intervalo	c_i	f_i	F_i	n_i	N_i
[0,5)	2.5	1	1	0.025	0.025
[5,10)	7.5	1	2	0.025	0.050
[10,15)	12.5	3	5	0.075	0.125
[15,20)	17.5	3	8	0.075	0.200
[20,25)	22.5	3	11	0.075	0.275
[25,30)	27.5	6	17	0.150	0.425
[30,35)	32.5	7	24	0.175	0.600
[35,40)	37.5	10	34	0.250	0.850
[40,45)	42.5	4	38	0.100	0.950
[45,50)	47.5	2	40	0.050	1
		40		1	

EJERCICIO COLABORATIVO

- Pida a sus compañeros de clase que anoten el tiempo promedio (en horas, redondeado a la media hora más cercana) que duermen por noche. Registren los datos. A continuación, cree un gráfico sencillo (llamado **diagrama de puntos**) de los datos.
- Un diagrama de puntos consiste en una línea numérica y puntos (o pequeños círculos) colocados sobre la línea numérica.
- Por ejemplo, considere los siguientes datos:

5; 5,5; 6; 6; 6; 6,5; 6,5; 6,5; 6,5; 7; 7; 8; 8; 9

El diagrama de puntos para estos datos sería el siguiente:

Gráficas de estadística

- Diagrama de barras
- Diagrama de sectores
- Histograma
- Polígonos de frecuencia

Diagrama de sectores

- Un diagrama de sectores se puede utilizar para todo tipo de variables, pero se usa frecuentemente para las variables cualitativas.

- Los datos se representan en un círculo, de modo que el ángulo de cada sector es proporcional a la frecuencia absoluta correspondiente.

$$\alpha = \frac{360^\circ}{N} \cdot f_i$$

Ejemplo. En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 9 juegan al fútbol y el resto no practica ningún deporte.

$$\alpha_1 = \frac{360°}{30} \cdot 12 = 144°$$

$$\alpha_3 = \frac{360°}{30} \cdot 9 = 108°$$

$$\alpha_2 = \frac{360°}{30} \cdot 3 = 36°$$

$$\alpha_4 = \frac{360°}{30} \cdot 6 = 72°$$

Deporte	Alumnos	Ángulo
Baloncesto	12	144°
Natación	3	36°
Fútbol	9	108°
Sin deporte	6	72°
Total	30	360°

Diagrama de barras

- Un diagrama de barras se utiliza para presentar datos cualitativos o datos cuantitativos de tipo discreto.

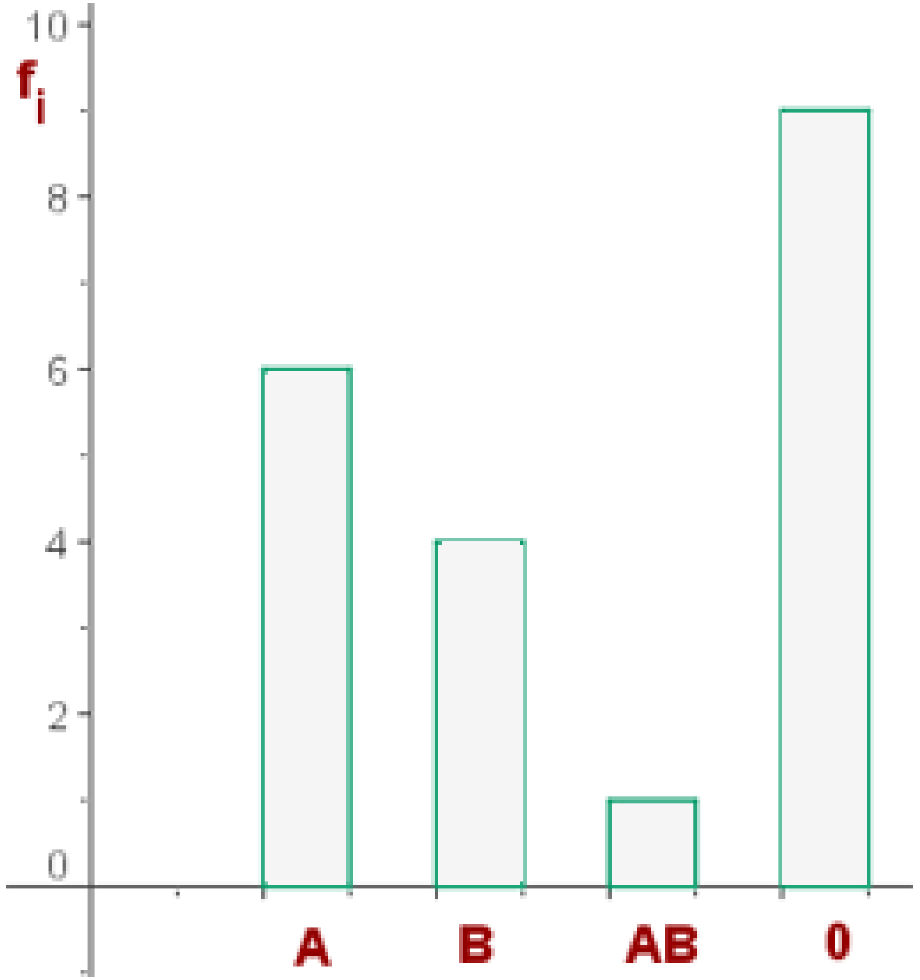
- Se representan sobre unos ejes de coordenadas, en el eje de abscisas se colocan los valores de la variable, y sobre el eje de ordenadas las frecuencias absolutas o relativas o acumuladas.

- Los datos se representan mediante barras de una altura proporcional a la frecuencia.

Ejemplo

Un estudio hecho al conjunto de los 20 alumnos de una clase para determinar su grupo sanguíneo ha dado el siguiente resultado:

Grupo sanguíneo	f_i
A	6
B	4
AB	1
0	9
	20



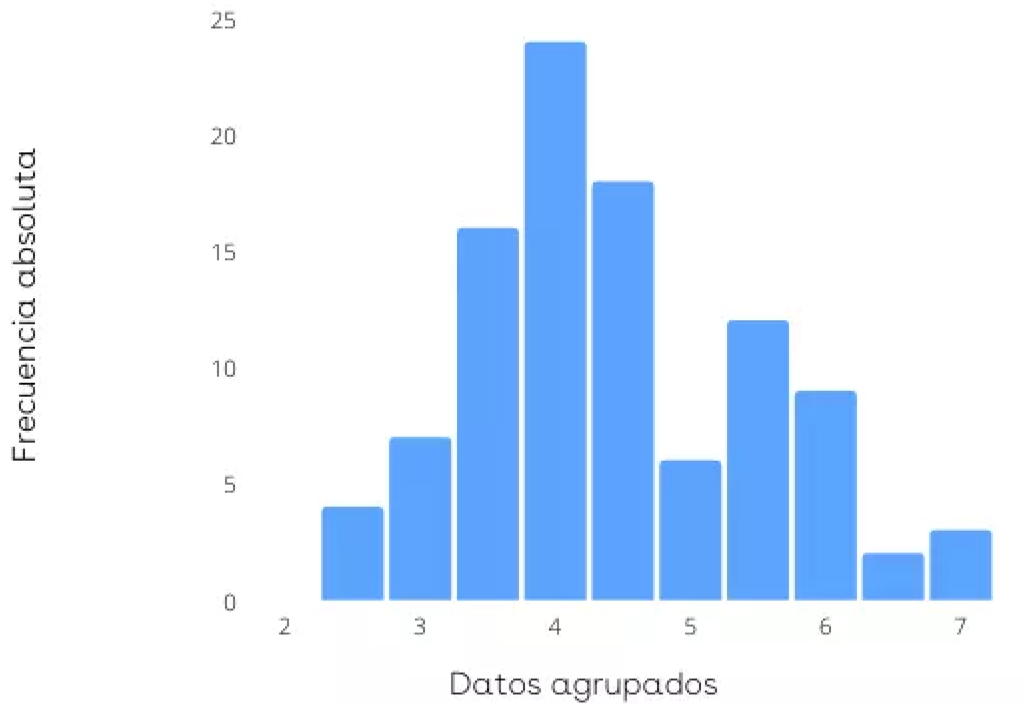
Histograma

- Un histograma es una representación gráfica de una variable en forma de barras.

- Se utilizan para variables continuas o para variables discretas, con un gran número de datos, y que se han agrupado en clases.

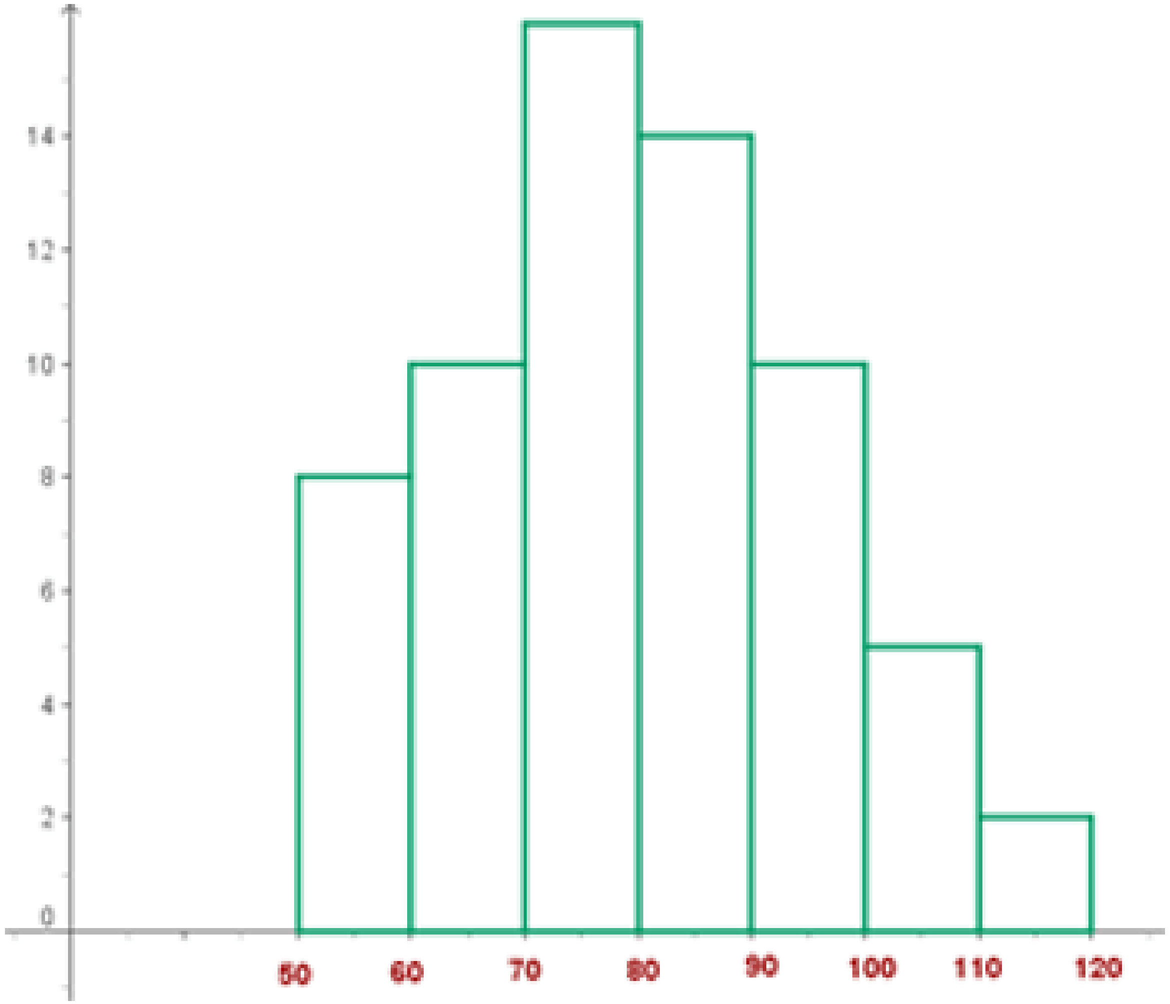
- En el eje abscisas se construyen unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia absoluta de cada intervalo.

- La superficie de cada barra es proporcional a la frecuencia de los valores representados.

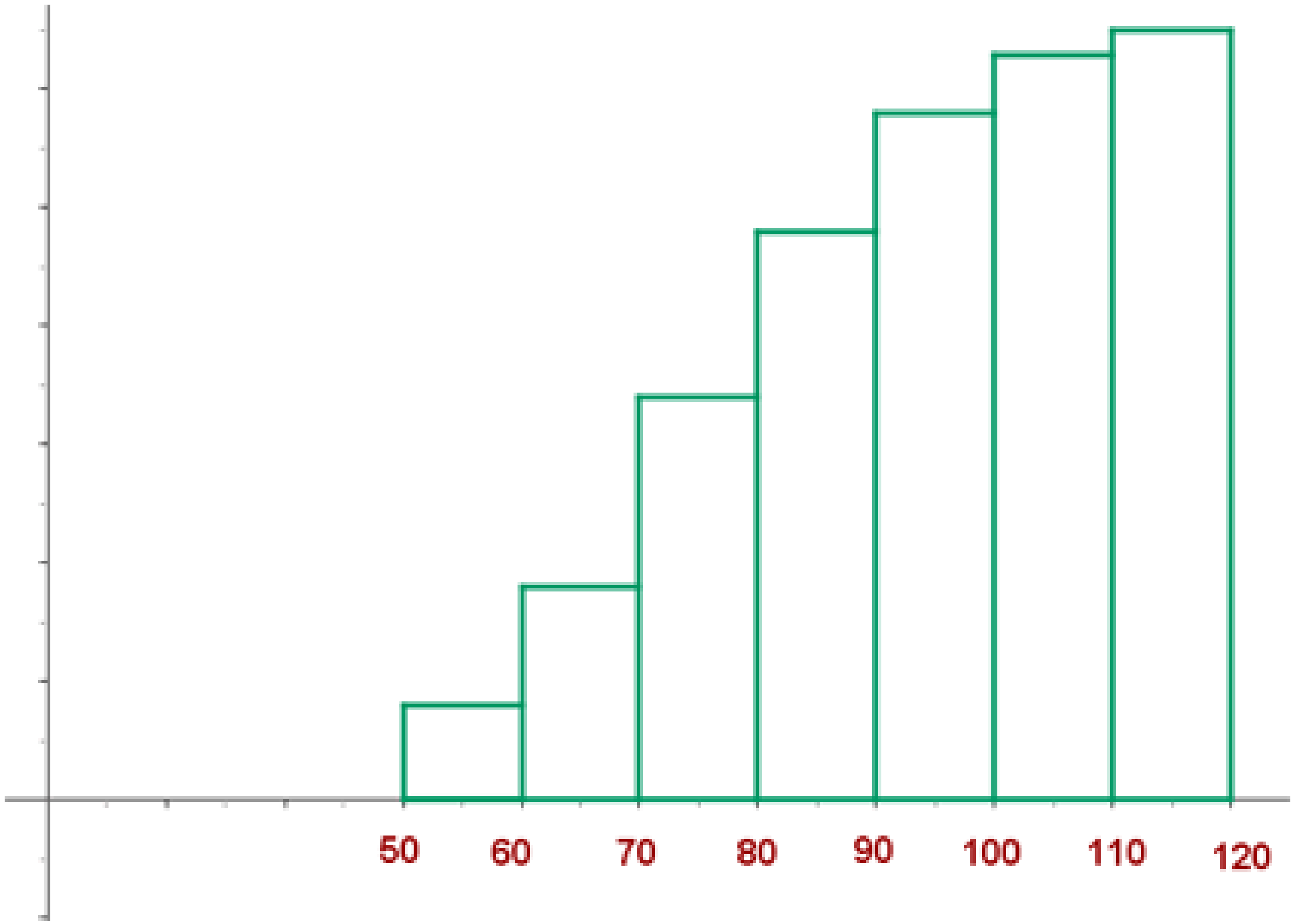


Ejemplo. El peso de 65 personas adultas viene dado por la siguiente tabla:

Intervalo	c_i	f_i	F_i
[50,60)	55	8	8
[60,70)	65	10	18
[70,80)	75	16	34
[80,90)	85	14	48
[90,100)	95	10	58
[100,110)	105	5	63
[110,120)	115	2	65
		65	



Histograma de frecuencias acumuladas



Histogramas con intervalos de amplitud diferente

- Para construir un histograma con intervalo de amplitud diferente tenemos que calcular las alturas de los rectángulos del histograma.

$$h_i = \frac{f_i}{a_i}$$

$\boldsymbol{h_i}$ es la altura del intervalo

f_i es la frecuencia del intervalo

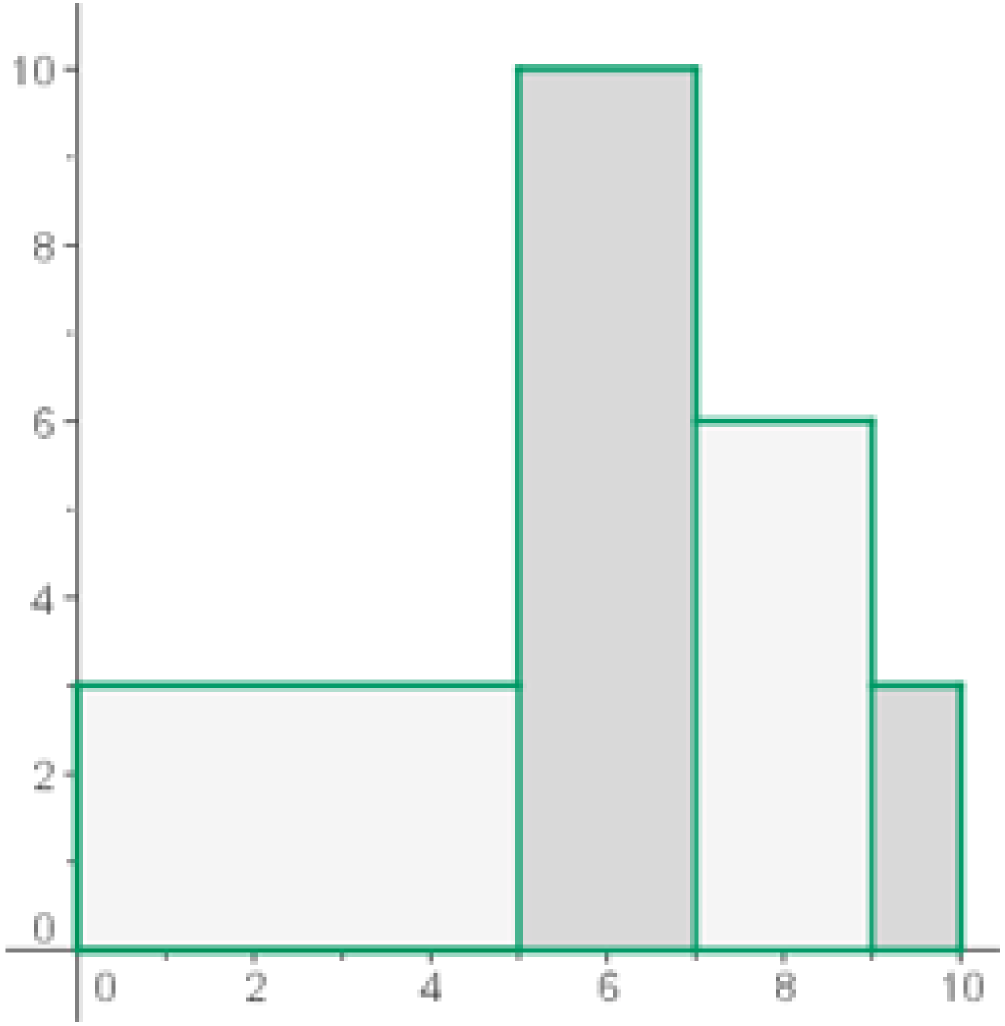
a_i es la amplitud del intervalo

Ejemplo

En la siguiente tabla se muestra las calificaciones (reprobado, aprobado, notable y sobresaliente) obtenidas pr un grupo de 50 alumnos.

Intervalo	a_i	f_i	h_i
[0,5)	5	15	3
[5,7)	2	20	10
[7,9)	2	12	6
[9,10)	1	3	3
		50	

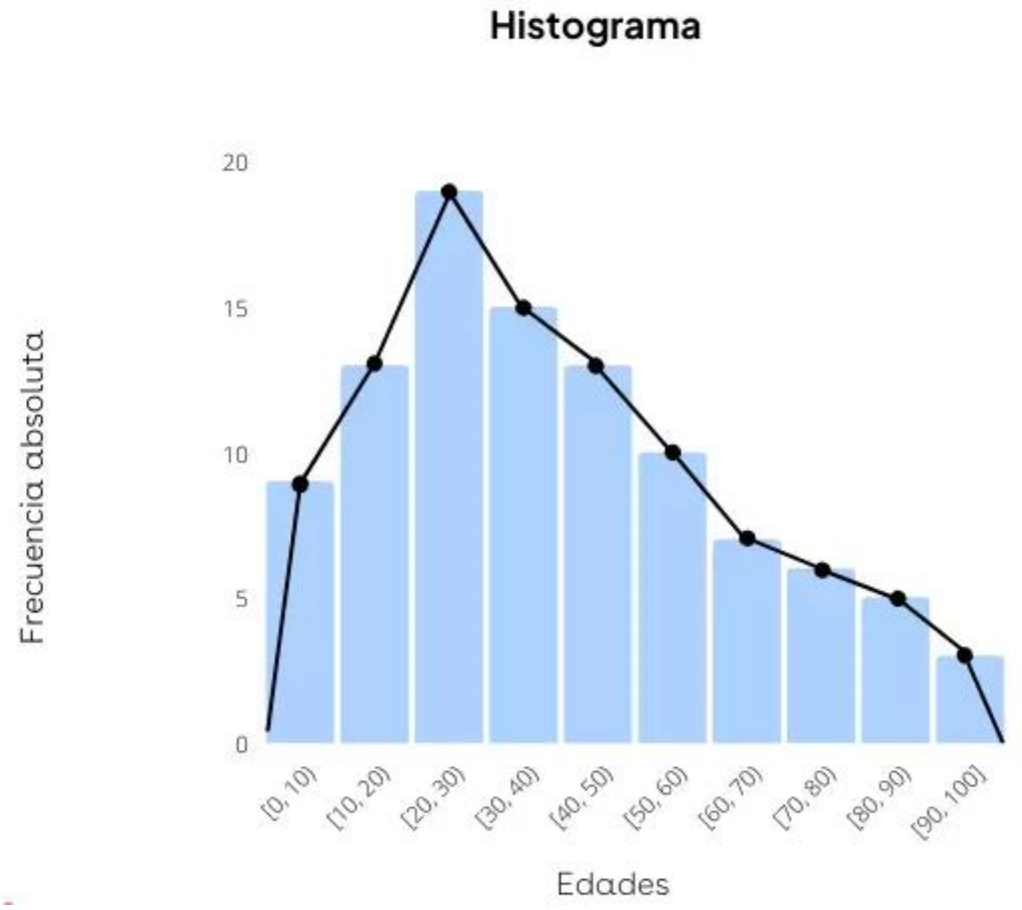
a_i es la amplitud del intervalo



Polígonos de frecuencia

- Un polígono de frecuencias se forma uniendo los extremos de las barras mediante segmentos.

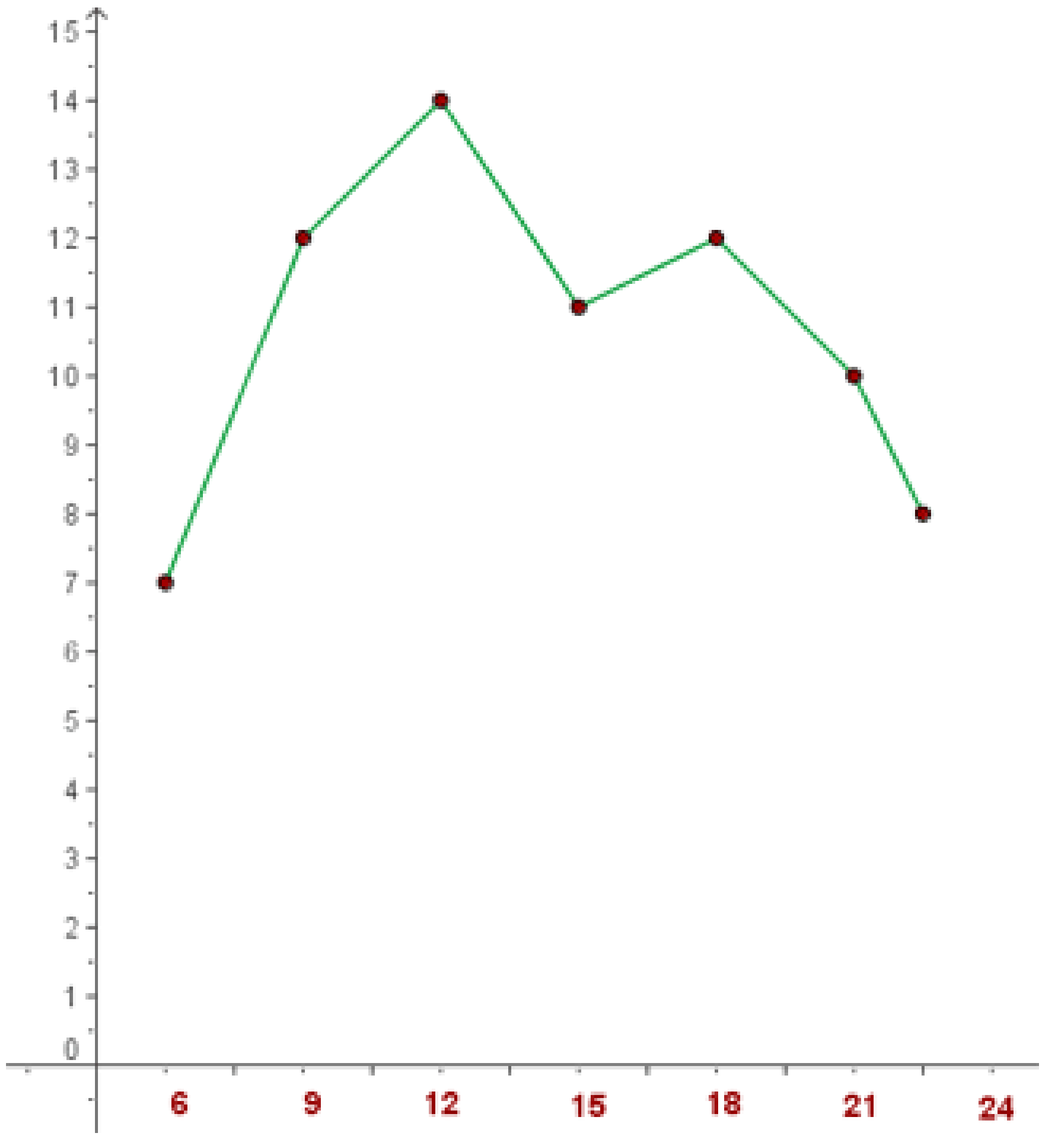
- También se puede realizar trazando los puntos que representan las frecuencias y uniéndolos mediante segmentos.



Ejemplo:

Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

Hora	Temperatura
6	7º
9	12°
12	14°
15	11°
18	12°
21	10°
24	8°

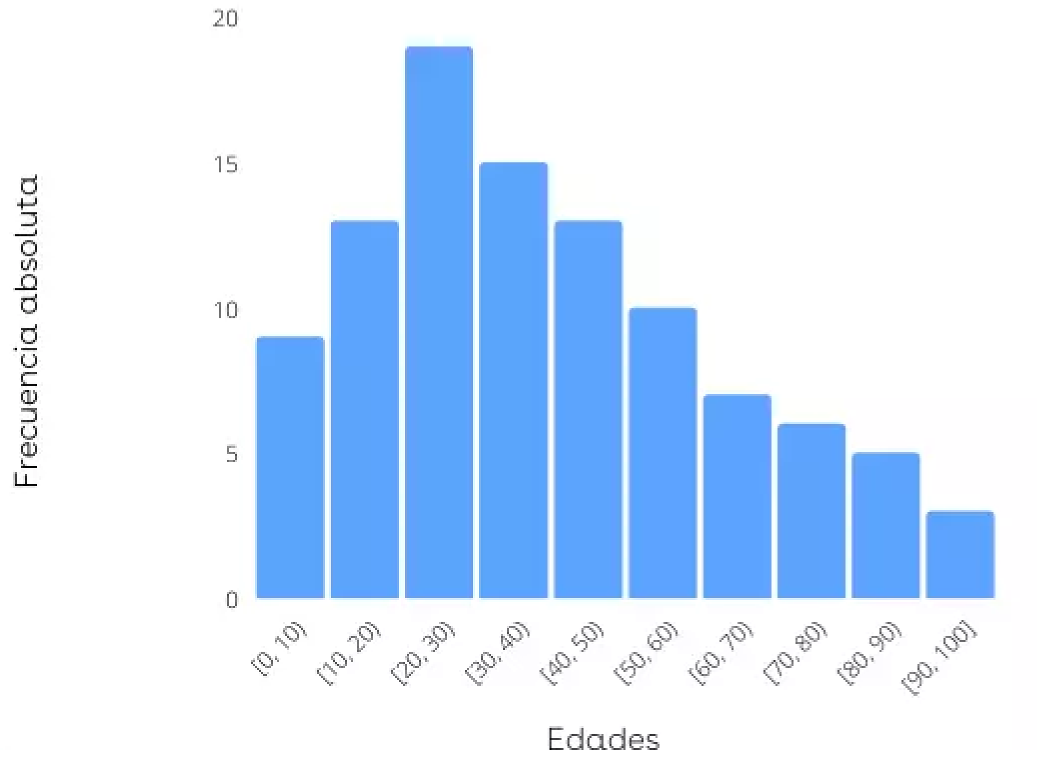


Ejemplo. Consideremos los siguientes datos

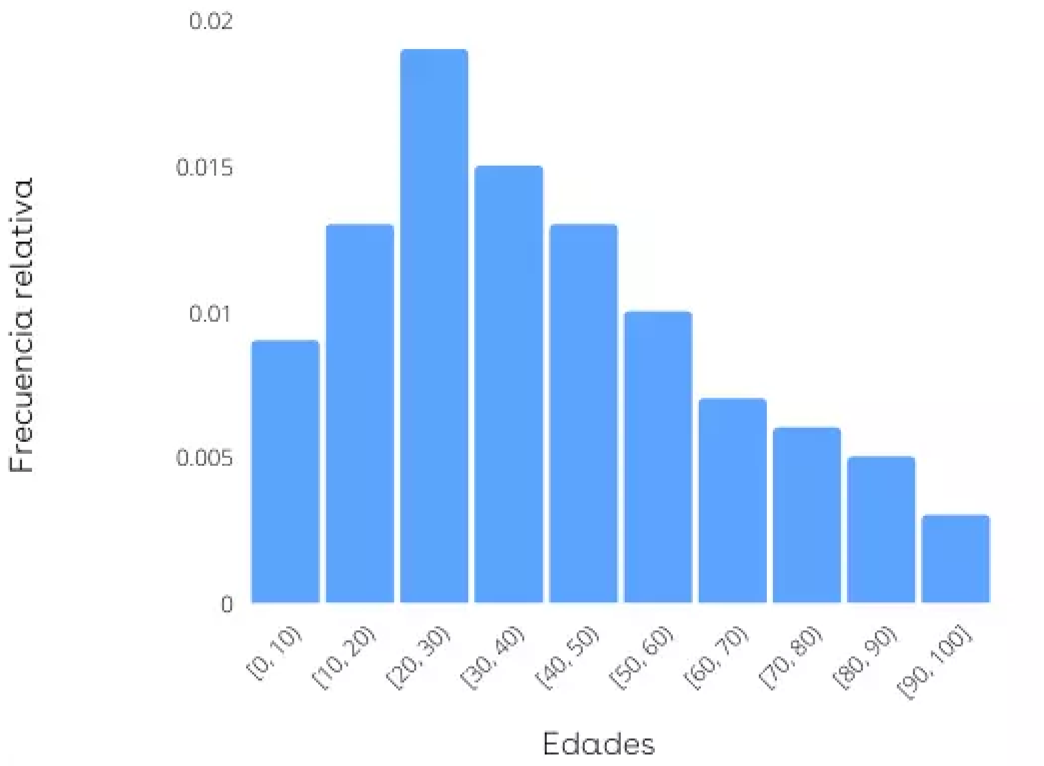
Edad	Amplitud (a_i)	Personas (f_i)
[0,10)	10	9
[10,20)	10	13
[20,30)	10	19
[30,40)	10	15
[40,50)	10	13
[50,60)	10	10
[60,70)	10	7
[70,80)	10	6
[80,90)	10	5
[90,100)	10	3
		100

Construye:

- histograma de frecuencias absolutas
- histograma de frecuencias relativas

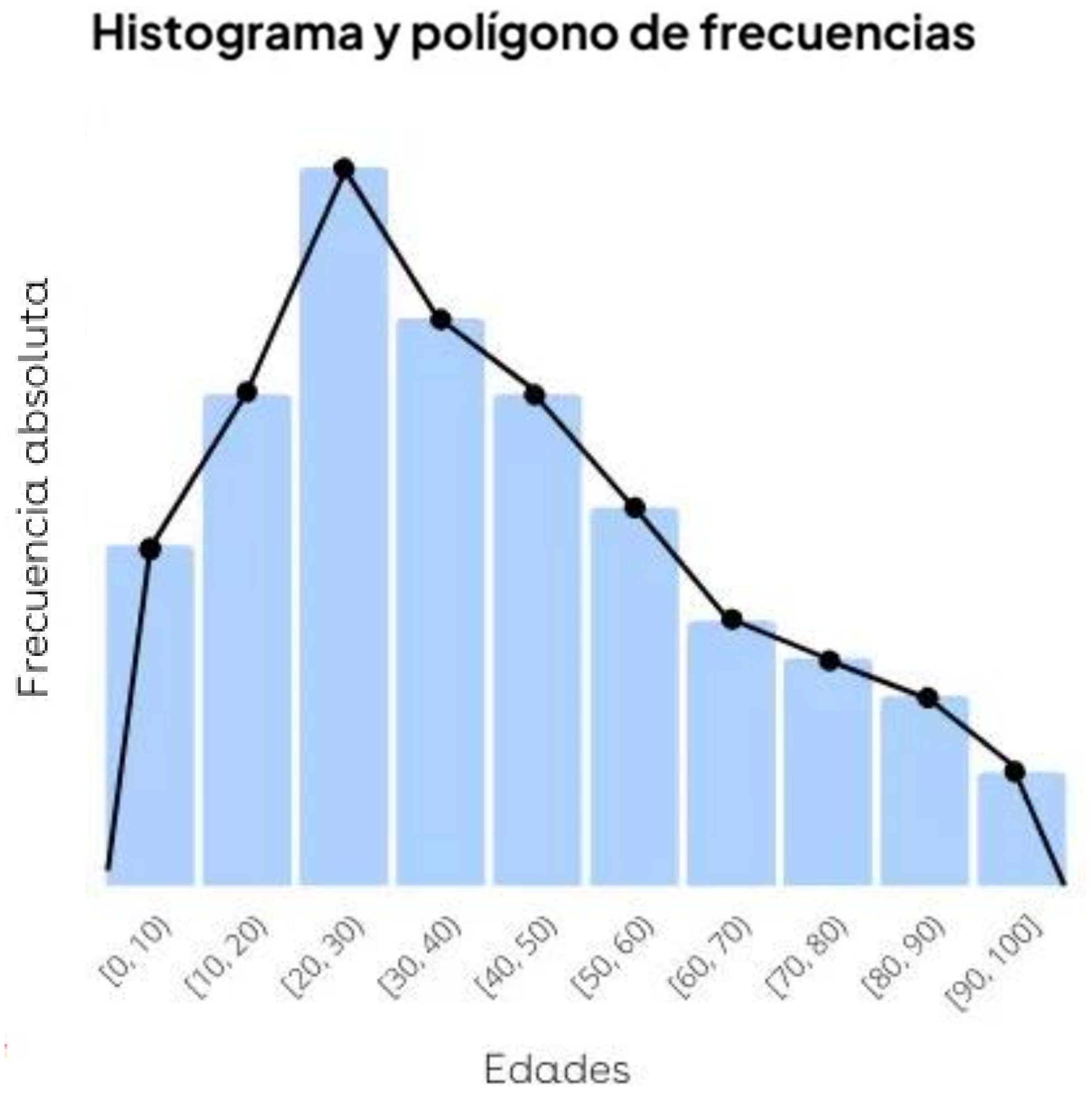


Histograma de frecuencias absolutas

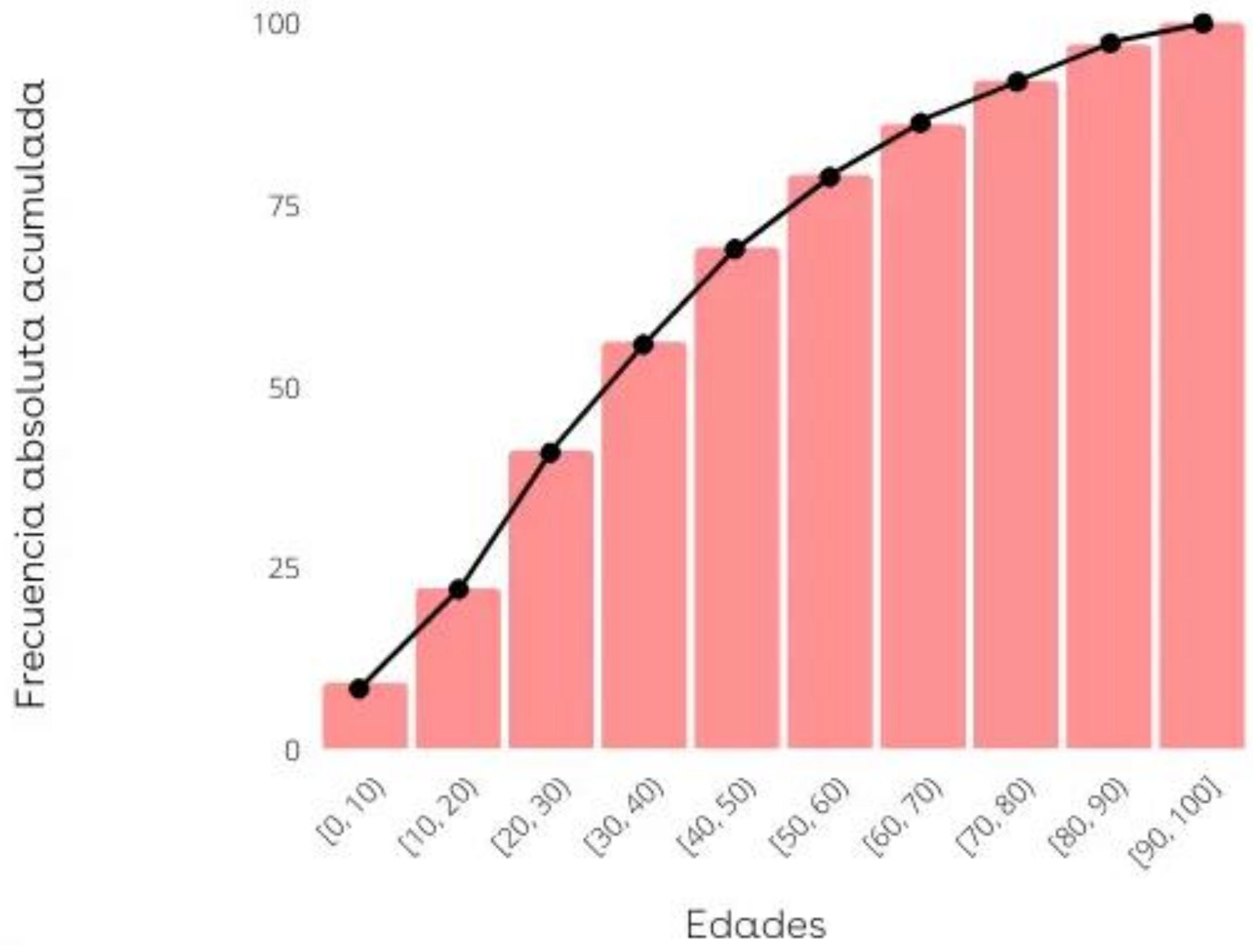


Histograma de frecuencias relativas

Edad	Personas (f_i)	F_i	c_i
[0,10)	9	9	5
[10,20)	13	22	15
[20,30)	19	41	25
[30,40)	15	56	35
[40,50)	13	69	45
[50,60)	10	79	55
[60,70)	7	86	65
[70,80)	6	92	75
[80,90)	5	97	85
[90,100)	3	100	95
	100		



Histograma y polígono de frecuencias acumuladas



Ejemplo. Consideremos una agrupación distinta de los datos de los ejemplos anteriores

Edad	Amplitud (a_i)	Personas (f_i)	F_i	c_i	h_i *relativa*
[0,10)	10	9	0.09	5	0.009
[10,20)	10	13	0.22	15	0.013
[20,30)	10	19	0.41	25	0.019
[30,60)	**30**	**15**	**0.56**	**35**	**0.012667**
[60,70)	10	13	0.69	45	0.007
[70,80)	10	10	0.79	55	0.006
[80,90)	10	7	0.86	65	0.005
[90,100)	10	6	0.92	75	0.003
		100	1		

Las alturas relativas se calculan dividiendo la frecuencia relativa entre la amplitud:

$$h_i = \frac{F_i}{a_i}$$

→ frecuencia relativa

→ amplitud

$$h_1 = \frac{0.09}{10} = 0.009$$

$$h_2 = \frac{0.13}{10} = 0.013$$

$$h_3 = \frac{0.19}{10} = 0.019$$

$$h_4 = \frac{0.38}{30} = 0.012667$$

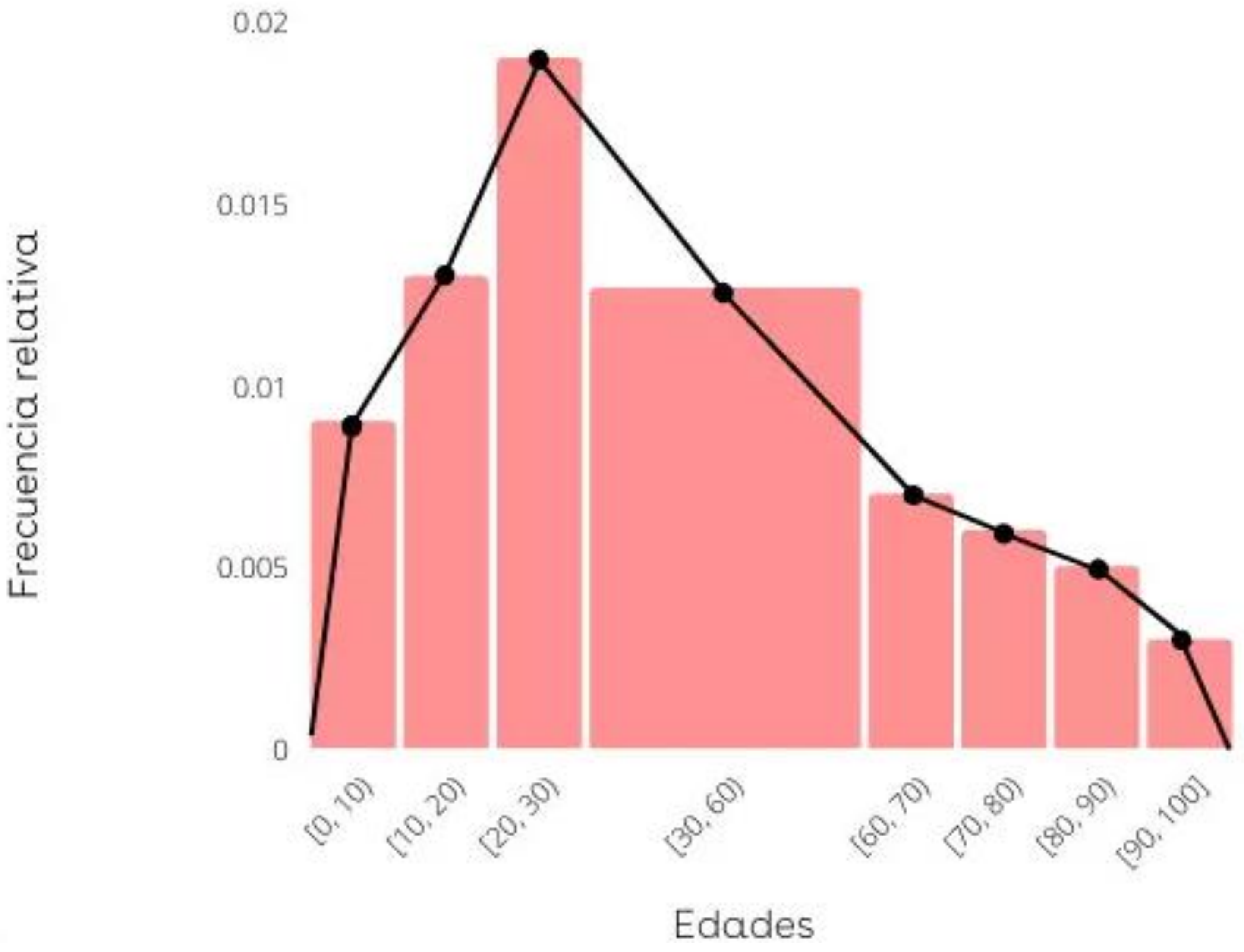
$$h_5 = \frac{0.07}{10} = 0.007$$

$$h_6 = \frac{0.06}{10} = 0.006$$

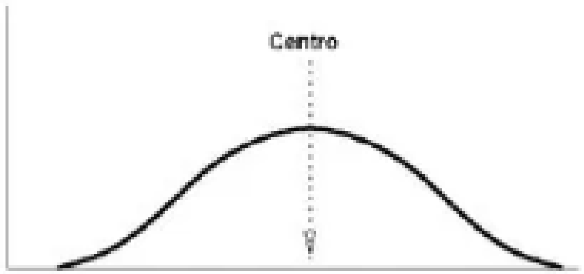
$$h_7 = \frac{0.05}{10} = 0.005$$

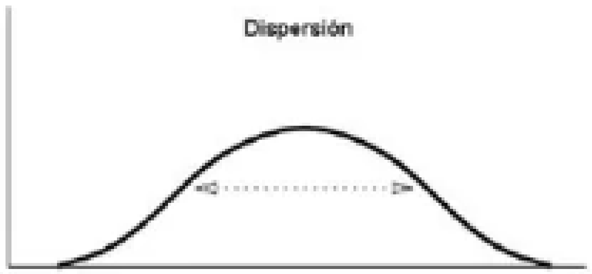
$$h_8 = \frac{0.03}{10} = 0.003$$

Histograma y polígono de frecuencias relativas



Medidas estadísticas



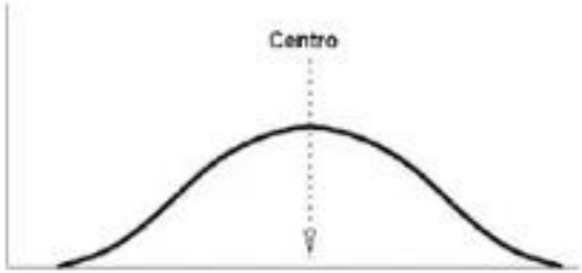


Medidas estadísticas

Tendencia central:

Indican valores con respecto a los que los datos parecen agruparse

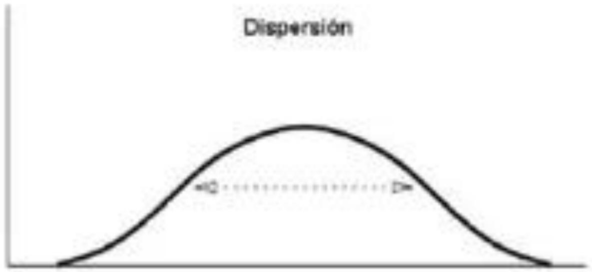
- Media
- Mediana
- Moda



Dispersión

Indican el nivel de concentración de los datos con respecto a la medida

- Rango
- Varianza
- Desv Estándar
- Coef. de Variación



Medidas de tendencia central

Media

- es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de datos.

Mediana

- es el valor medio de una secuencia ordenada de datos

Moda

- es el dato que más veces se repite

Media

- Es el valor promedio del grupo de datos, es decir, la cifra que se obtiene al sumar todos los datos y dividir el resultado entre la cantidad de los mismos.

$$Media = \frac{Suma\ de\ valores}{Cantidad\ de\ valores}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \cdots\ x_n}{N}$$

Calcular el valor medio

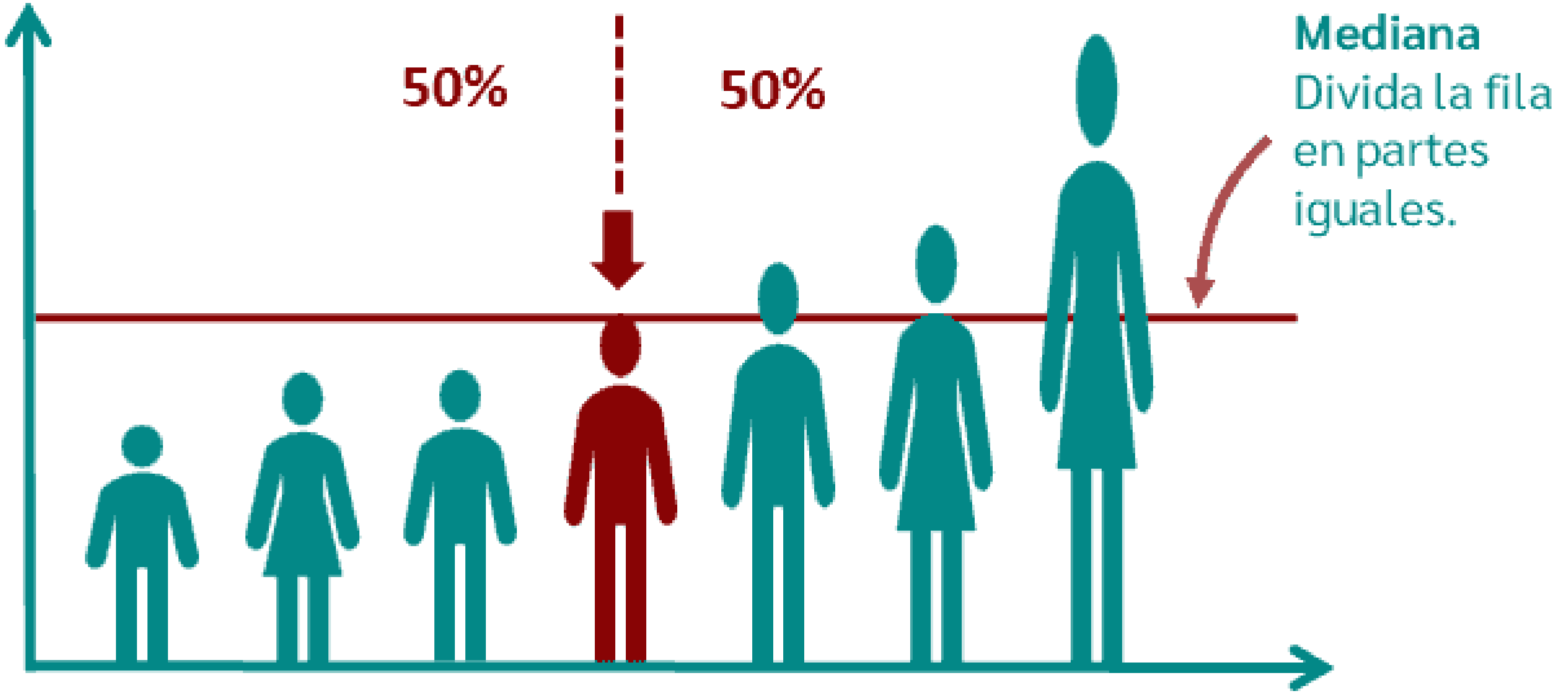
- Se ha preguntado a un grupo de 5 estudiantes de estadística cuántas tazas de café beben a la semana. El resultado es: 21, 25, 10, 8 y 11 tazas. La media es, por tanto, 15.

	1	2	3	4	5
	21	25	10	8	11

$$\frac{21 + 25 + 10 + 8 + 11}{5} = 15$$

Mediana

- La mediana es el valor medio de una secuencia ordenada de datos. Si no hay empates, la mitad de las observaciones serán menores y la otra mitad serán mayores.
    - Calcular la mediana es mucho más fácil porque es justo el valor central, es decir, el que se encuentra en la mitad de la lista.



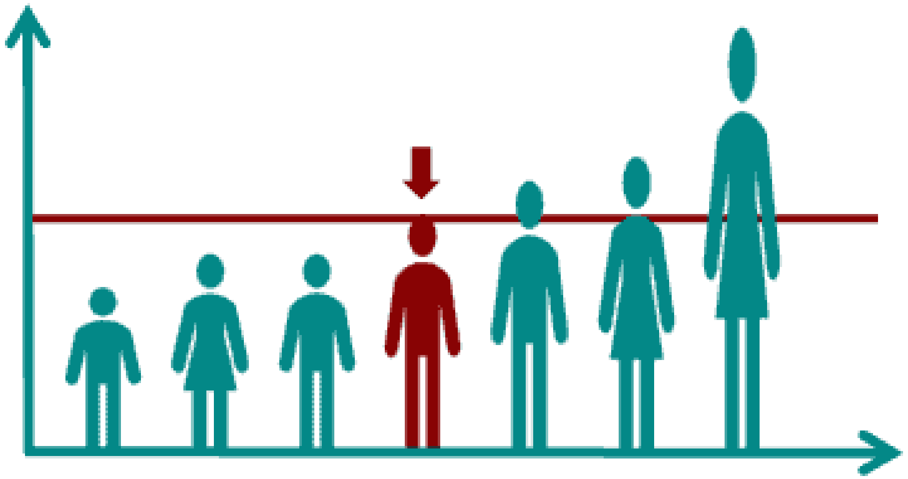
Procedimiento para calcular la mediana

1. Ordene los datos
2. Utiliza la expresión:

$$\tilde{X} = \begin{cases} X_{(n+1)/2} \; si\ n\ es\ impar \\ \frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) si\ n\ es\ par \end{cases}$$

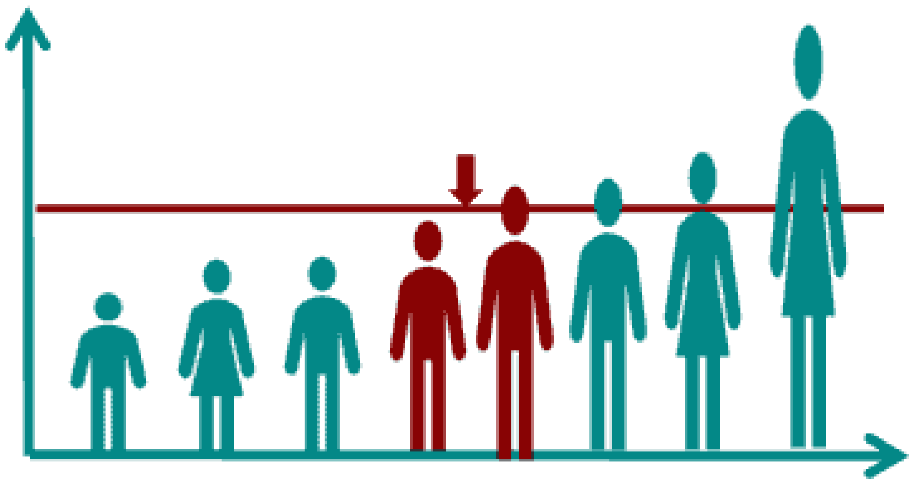
Número impar

La mediana es un valor que se da realmente.



Número par

Es la media de los dos valores medios.



Ejemplo. Tenemos el siguiente conjunto de datos:

2.0, 2.5, 5.3, 4.2, 3.8

Calcular la mediana.

1. Ordenamos los datos: 2.0, 2.5, 3.8, 4.2, 5.3
2. *N* es impar, *N* =5.
3. Utilizamos la expresión: $\tilde{x} = x_{(n+1)/2} = x_{(5+1)/2} = x_3$
4. La mediana es el número ubicado en la tercera posición: 3.8

Ejemplo. Calcular la mediana del siguiente conjunto de datos:

480, 485, 485, 490, 490, 495, 495, 500, 500, 505

480, 485, 485, 490, $\boxed{490, 495,}$ 495, 500, 500, 505

1. Los datos están ordenados.

2. *N* es par, *N* =10.

3. Utilizamos la expresión:

$$\tilde{x} = \frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right)$$

$$\tilde{x} = \frac{1}{2}(x_5 + x_6)$$

$$Mediana = \frac{1}{2}(490 + 495) = 492.5$$

Ejemplo. Calcular la mediana del siguiente conjunto de datos:

460, 460, 470, 500, 500, 500, 500, 500, 500, 500

460, 460, 470, 500, $\boxed{500, 500,}$ 500, 500, 500, 500

1. Los datos están ordenados.

2. *N* es par, *N* =10.

3. Utilizamos la expresión:

$$\tilde{x} = \frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right)$$

$$\tilde{x} = \frac{1}{2}(x_5 + x_6)$$

$$Mediana = \frac{1}{2}(500 + 500) = 500$$

Ejemplo. Una toma de datos de un experimento arroja los siguientes resultados. Calcular la mediana.

x_i	*t(s) ±0.1*
x_1	0.48
x_2	0.50
x_3	0.52
x_4	**0.55**
x_5	0.56
x_6	0.58
x_7	0.60
x_8	0.63
x_9	0.66
x_{10}	0.69

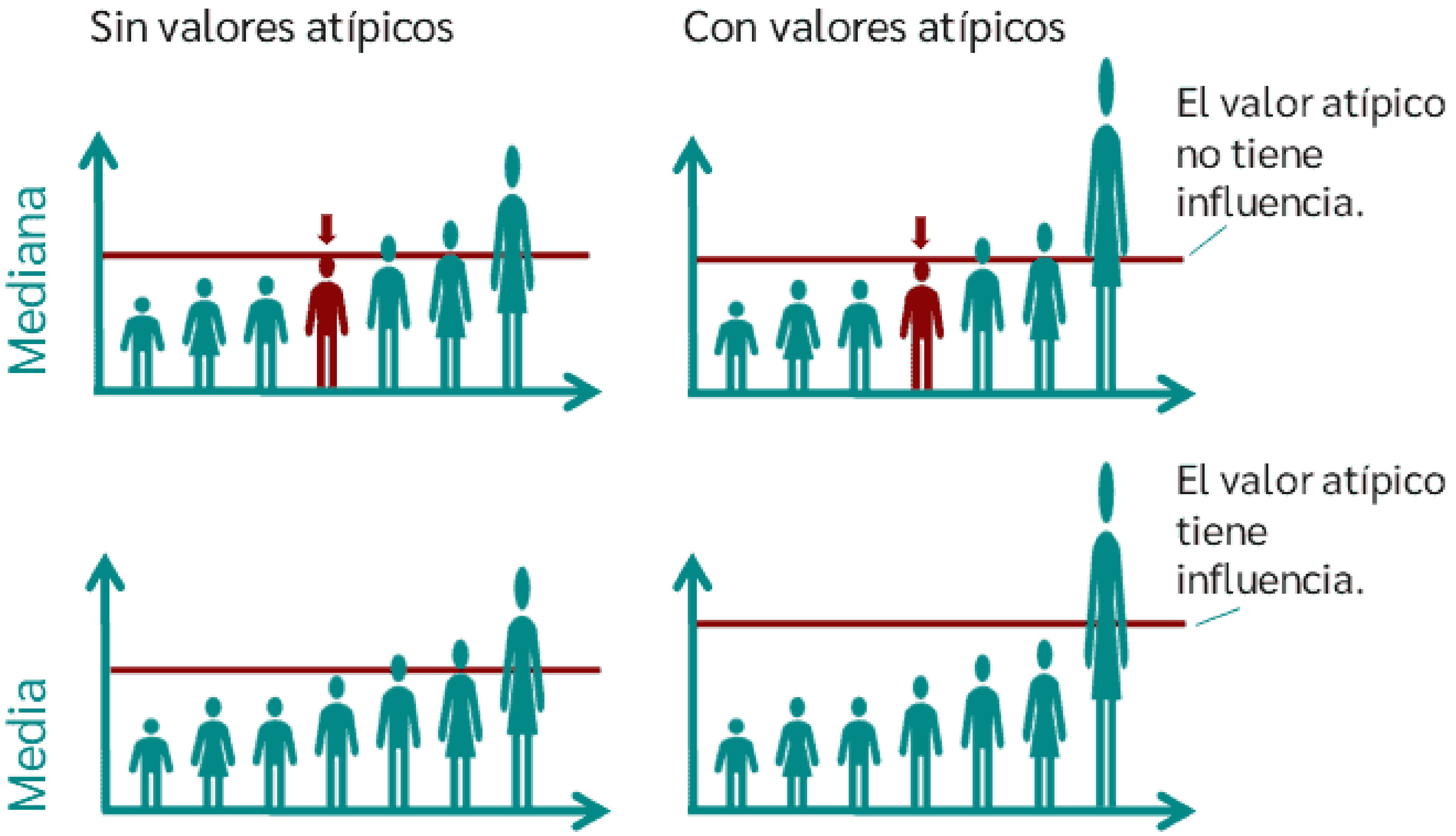
x_i	$t(s) \pm 0.1$
x_1	0.48
x_2	0.50
x_3	0.52
x_4	0.55
x_5	0.56
x_6	0.58
x_7	0.60
x_8	0.63
x_9	0.66
x_{10}	0.69

1. Los datos están ordenados.
2. *N* es par, *N* = 10.
3. Utilizamos la expresión:

$$\frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) = \frac{1}{2}\left(X_{\frac{10}{2}} + X_{\frac{10}{2}+1}\right) = \frac{1}{2}(X_5 + X_6) = 0.57$$

Media frente a mediana

- En comparación con la media, la mediana es mucho más robusta frente a la dispersión.
- Un valor atípico no suele influir en la mediana, pero tiene una influencia más o menos grande en la media.



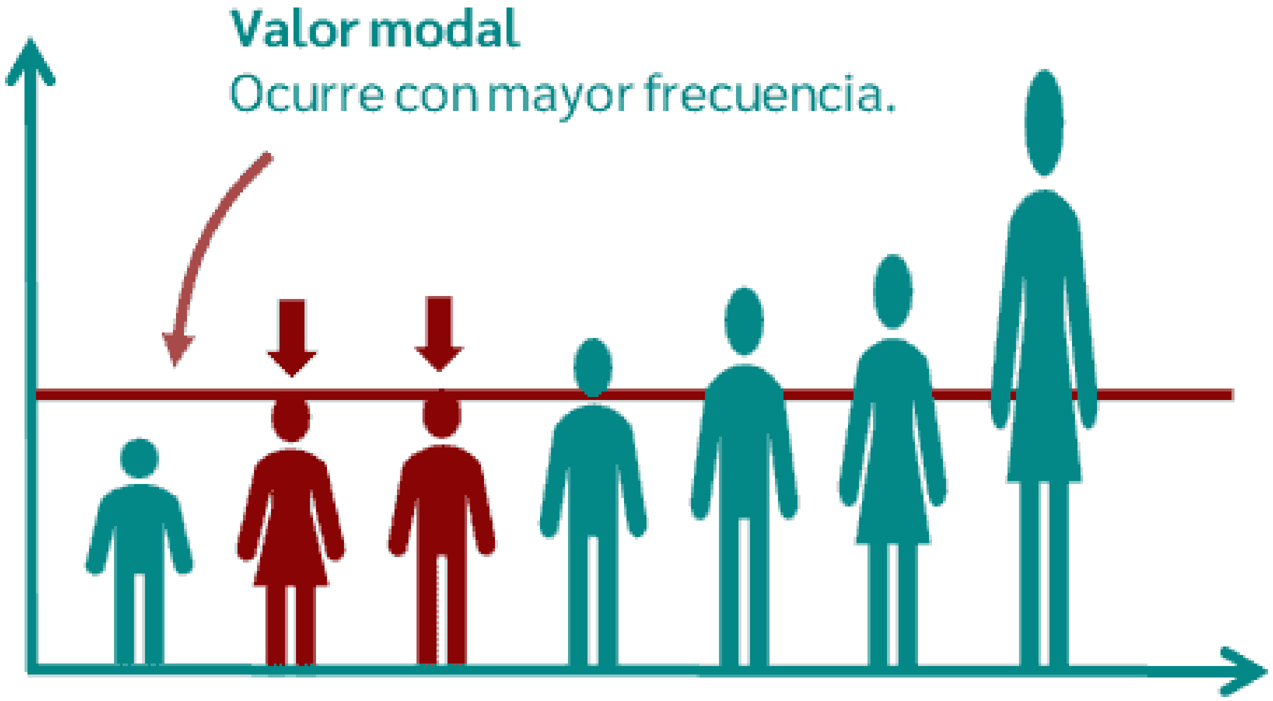
Moda

- La moda de un conjunto de datos es el dato que más veces se repite, es decir, aquel que tiene mayor frecuencia absoluta.

- En caso de existir dos valores de la variable que tengan la mayor frecuencia absoluta, habría dos modas.

- Si no se repite ningún valor, no existe moda.

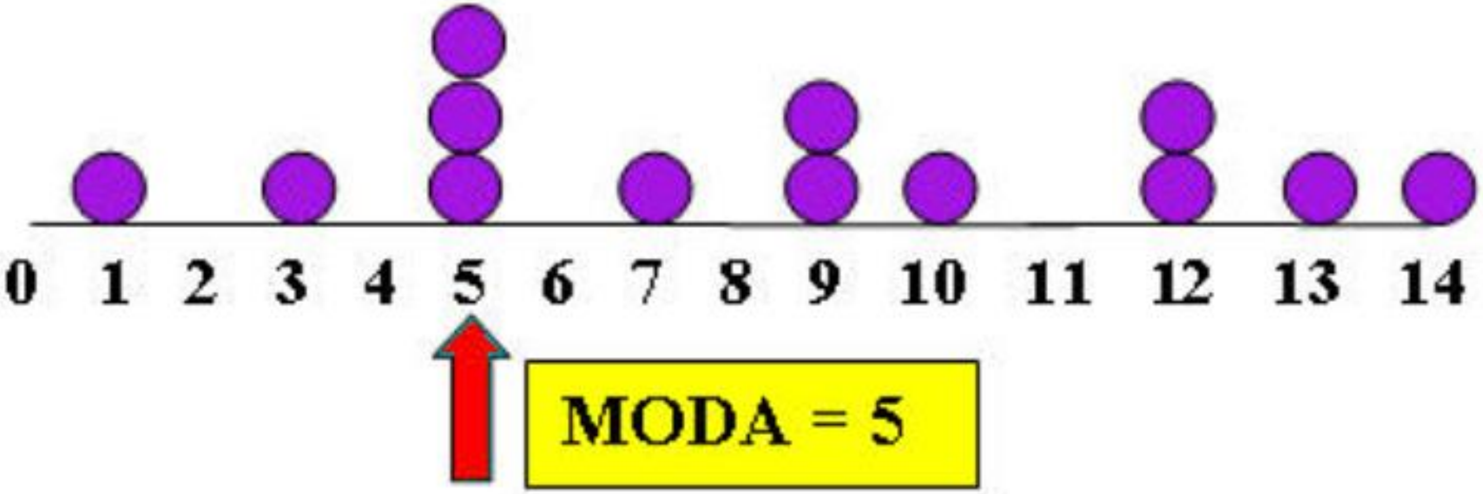
- A diferencia de la media aritmética, la moda estadística no se ve afectada por la ocurrencia de los valores extremos.

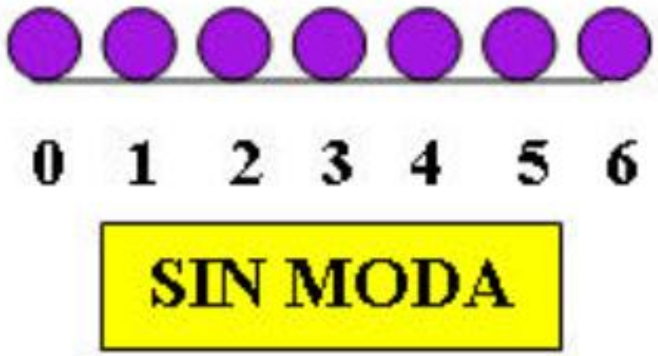

Valor modal
Ocurre con mayor frecuencia.

Calcular la moda

Ejemplo: En una muestra de 70 directivos de Berlín, 20 conducen un Daimler, 25 un BMW, 10 un VW y 15 un Audi. La marca de coches BMW es la más común. Por tanto, la Moda es "BMW".

Car brand	Daimler	BMW	VW	Audi
Frequency	20	25	10	15


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
MODA = 5


0 1 2 3 4 5 6
SIN MODA

Cálculo de la moda para datos agrupados

Caso 1: Cuando todos los intervalos tienen la misma amplitud.

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

L_i es el límite inferior de la clase modal

f_i es la frecuencia absoluta de la clase modal

f_{i-1} es la frecuencia absoluta inmediatamente inferior a la clase modal

f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal

a_i es la amplitud de la clase

- También se utiliza otra fórmula de la moda que da un valor aproximado de ésta:

$$M_o = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot a_i$$

Ejemplo:

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

Intervalo	*Frecuencia absoluta* f_i
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

- Los intervalos tienen la misma amplitud: $a_i = 3$

$$a_i = Límite\ superior\ de\ la\ clase - Límite\ inferior\ de\ la\ clase$$

Intervalo	*Frecuencia absoluta* f_i
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

1. Buscamos el intervalo donde se encuentra la moda, que será el intervalo que tenga la mayor frecuencia absoluta f_i.

   La **clase modal** es: (66,69)

Aplicaremos la fórmula para el cálculo de la moda para datos agrupados:

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

Intervalo	*Frecuencia absoluta f_i*
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

$Límite\ inferior = 66$

$f_i = 42$

$f_{i-1} = 18$

$f_{i+1} = 27$

$a_i = 3$

$$M_o = 66 + \frac{(42-18)}{(42-18)+(42-27)} \cdot 3 = 67.846$$

$$M_o = 66 + \frac{27}{18+27} \cdot 3 = 67.8$$

Cálculo de la moda para datos agrupados

Caso 2: Cuando los intervalos tienen amplitudes distintas.

1. En primer lugar tenemos que hallar las alturas. $h_i = \frac{f_i}{a_i}$

2. La clase modal es la que tiene mayor altura.

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

3. La fórmula de la moda aproximada cuando existen distintas amplitudes es:

$$M_o = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

Ejemplo:

En la siguiente tabla se muestra las calificaciones obtenidas por un grupo de 50 alumnos. Calcular la moda.

Intervalo	*Frecuencia absoluta* f_i
[0,5)	15
[5,7)	20
[7,9)	12
[9,10)	3

0,1,2,3,4 Amplitud: 5

5,6 Amplitud: 2

7,8 Amplitud: 2

9 Amplitud: 1

- Creamos una nueva columna con las alturas, dividiendo las frecuencias absolutas entre las amplitudes de los intervalos correspondientes:

$$h_1 = \frac{15}{5} = 3$$

$$h_1 = \frac{20}{2} = 10$$

$$h_3 = \frac{12}{2} = 6$$

$$h_4 = \frac{3}{1} = 3$$

	Frecuencia absoluta f_i	*Altura h_i*
[0,5)	15	3
[5,7)	20	10
[7,9)	12	6
[9,10)	3	3
	50	

- La **clase modal** es [5,7) porque es la que tiene mayor altura

Aplicaremos la fórmula para el cálculo de la moda para datos agrupados con intervalos de amplitudes distintas:

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

	Frecuencia absoluta f_i	*Altura h_i*
[0,5)	15	3
[5,7)	20	10
[7,9)	12	6
[9,10)	3	3
	50	

$$Límite\ inferior = 5$$

$$h_i = 10$$

$$f_{i-1} = 3$$

$$f_{i+1} = 6$$

$$a_i = 2$$

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

$$M_o = 5 + \frac{(10-3)}{(10-3)+(10-6)} \cdot 2 = 6.27$$

$$M_o = 5 + \frac{6}{3+6} \cdot 2 = 6.33$$

Ventajas y desventajas de la media, la mediana y la moda

… la cuestión es …

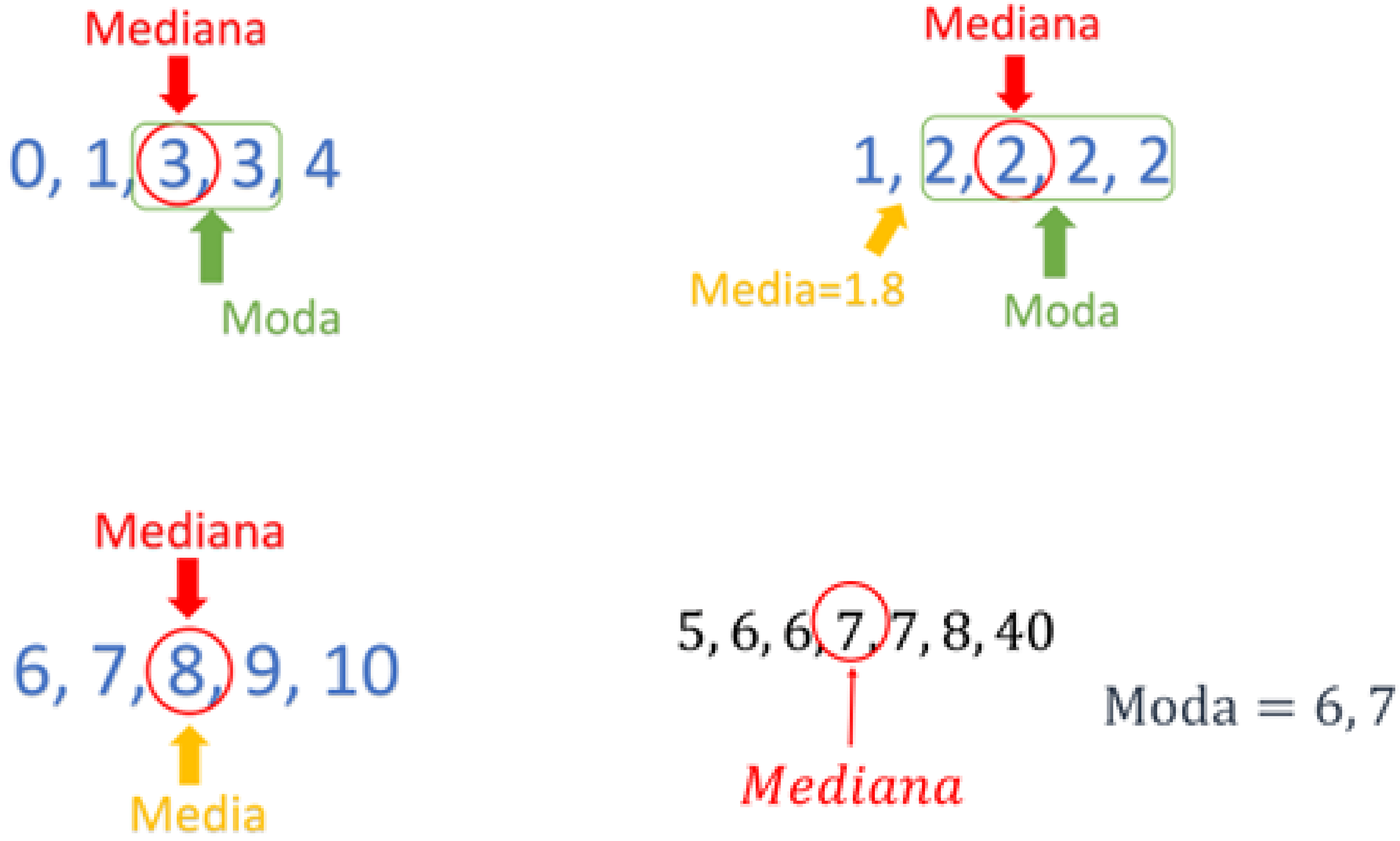
¿cuál de las medidas de tendencia central hay que utilizar?

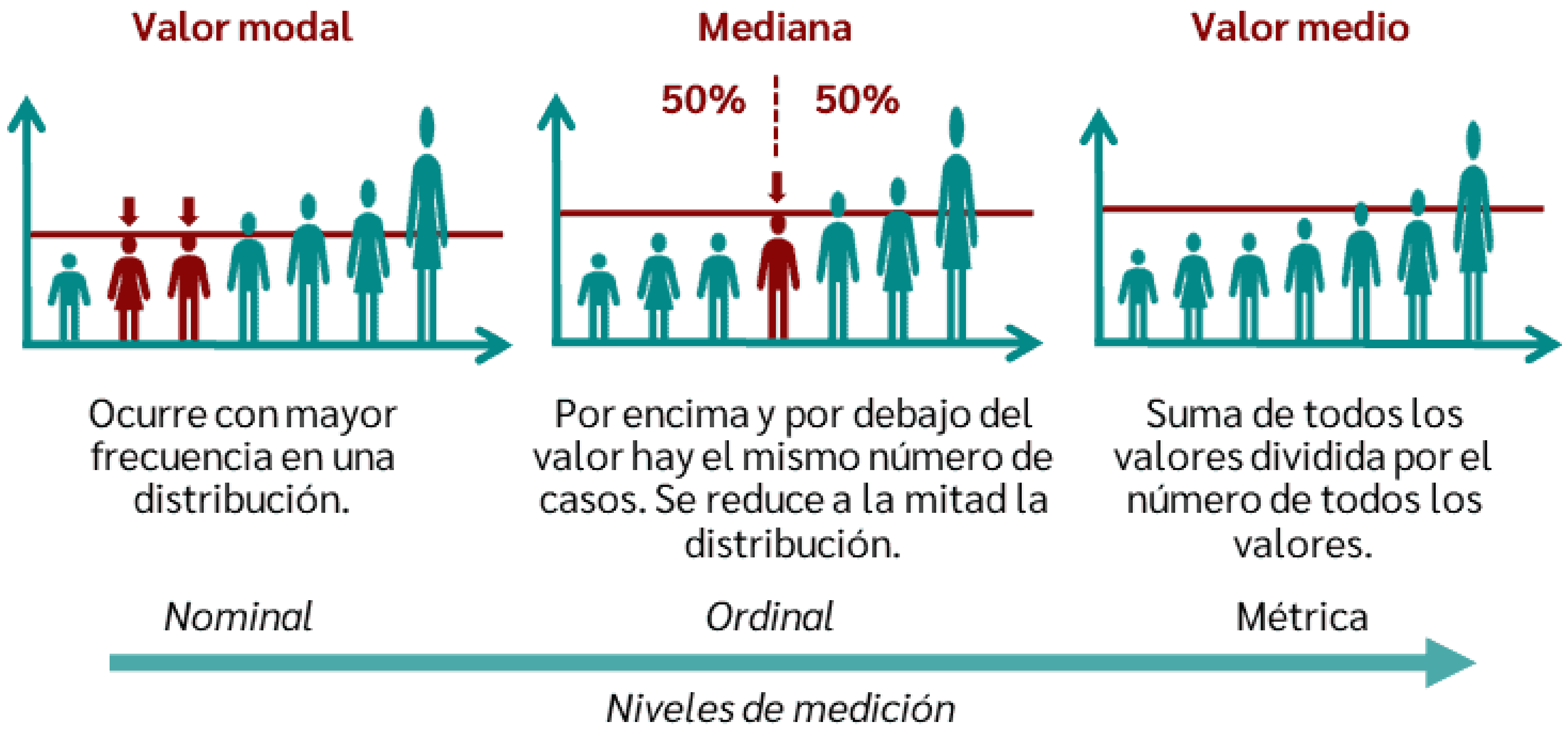
- Si la distribución es simétrica, la media y la mediana son iguales, y si la distribución es simétrica y unimodal, las tres medidas son iguales.

- **Media**: El valor medio es, con mucho, el más utilizado. Las desventajas de la media son que es sensible a los valores atípicos. **Los datos deben tener un nivel de escala métrica**.

- **Mediana**: La gran ventaja de la mediana es que es muy robusta frente a los valores atípicos y que **los datos sólo tienen que tener una escala ordinaria**.

- **Moda**: La moda es el valor que se da con mayor frecuencia, lo que tiene la ventaja de que el valor se da realmente. Además, la moda también puede calcularse para los datos que no pueden ordenarse y, por tanto, tienen un nivel de **escala nominal**. La desventaja es que la moda no tiene en cuenta los demás datos existentes.


Mediana
0, 1, 3, 3, 4
Moda
Mediana
1, 2, 2, 2, 2
Media=1.8
Moda
Mediana
6, 7, 8, 9, 10
Media
5, 6, 6, 7, 7, 8, 40
Moda = 6, 7
Mediana


Valor modal
Mediana
Valor medio
50%
50%
Ocurre con mayor frecuencia en una distribución.
Por encima y por debajo del valor hay el mismo número de casos. Se reduce a la mitad la distribución.
Suma de todos los valores dividida por el número de todos los valores.
Nominal
Ordinal
Métrica
Niveles de medición

Ejemplo. Investigando sobre el precio de una lata atún del mismo contenido y marca. Para ello, visitaron cinco tiendas y registramos lo siguiente.

Tienda	1	2	3	4	5
Precio de la lata de atún	$14.90	$16.25	$14.90	$15.90	$16.75

¿qué medida de tendencia central representa mejor el costo de una lata de atún en estas tiendas?

$$\bar{x} = \frac{14.90 + 16.25 + 14.90 + 15.90 + 16.75}{5} = \frac{78.70}{5} = 15.74$$

El precio promedio es de $15.74

Mediana

14.90, 14.90, $\boxed{15.90}$, 16.25, 16.75

La mediana es $15.90

Moda

La moda es $14.90

Ejemplo.

A un conjunto de 5 números cuya media es 7.31 se le añaden los números 4.47 y 10.15. ¿Cuál es la media del nuevo conjunto de números?

Ejemplo.

A un conjunto de 5 números cuya media es 7.31 se le añaden los números 4.47 y 10.15. ¿Cuál es la media del nuevo conjunto de números?

Solución

La media del conjunto de los 5 números es: $\overline{x} = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = 7.31$

Entonces: $x_1 + x_2 + x_3 + x_4 + x_5 = 7.31 \cdot 5$

La media de los 7 números es: $\overline{x}' = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5 + 4.47 + 10.15}{7}$

Que es lo mismo que:

$$\overline{x}' = \frac{7.31 \cdot 5 + 4.47 + 10.15}{7} = 7.31$$

Ejemplo.

Un dentista observa el número de caries en cada uno de los 100 niños de cierto colegio. La información obtenida aparece resumida en la siguiente tabla:

# de caries x_i	f_i	n_i
0	25	0.25
1	20	0.2
2	x	z
3	15	0.15
4	y	0.05

- Completar la tabla obteniendo los valores x, y, z.
- Hacer un diagrama de sectores.
- Calcular el número medio de caries.

- La suma de las frecuencias relativas debe ser igual a 1:

$$0.25 + 0.2 + z + 0.15 + 0.05 = 1$$

$$z = 0.35$$

# de caries x_i	f_i	n_i
0	25	0.25
1	20	0.2
2	x	0.35
3	15	0.15
4	y	0.05
	100	1

- La frecuencia relativa de un dato es igual a su frecuencia absoluta dividida entre 100, que es la suma de las frecuencias absolutas.

$$\frac{x}{100} = 0.35$$

$$x = 35$$

$$\frac{y}{100} = 0.05$$

$$y = 5$$

# de caries x_i	f_i	n_i
0	25	0.25
1	20	0.2
2	35	0.35
3	15	0.15
4	5	0.05
	100	1

Calculamos los grados que corresponden a una unidad de frecuencia absoluta

$$x = \frac{360}{100} = 3.6°$$

Calculamos los grados que corresponden a cada frecuencia absoluta.

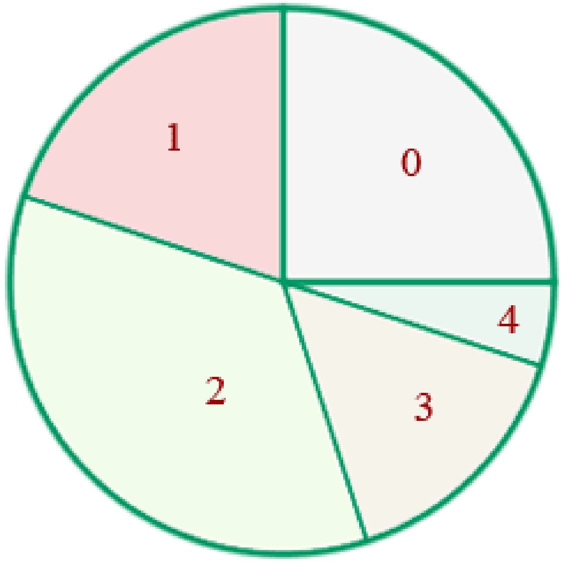
$$25 \cdot 3.6° = 90°$$

$$20 \cdot 3.6° = 72°$$

$$35 \cdot 3.6° = 126°$$

$$15 \cdot 3.6° = 54°$$

$$5 \cdot 3.6° = 18°$$



# de caries x_i	f_i	n_i	$f_i * x_i$
0	25	0.25	0
1	20	0.2	20
2	35	0.35	70
3	15	0.15	45
4	5	0.05	20
	100	1	155

Media aritmética $\overline{x} = \frac{155}{100} = 1.55$

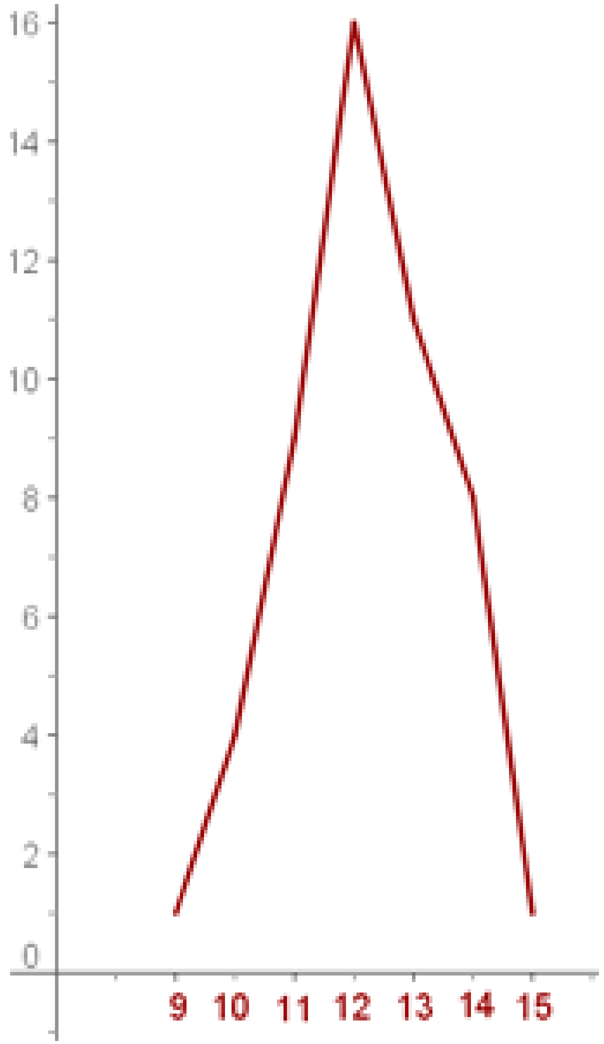
Ejemplo.

Un pediatra obtuvo la siguiente tabla sobre los meses de edad de niños de su consulta en el momento de andar por primera vez:

Meses	*Niños*
9	1
10	4
11	9
12	16
13	11
14	8
15	1

- Dibujar el polígono de frecuencias.
- Calcular la moda, la mediana, la media y la varianza

Polígono de frecuencias



Meses	*Niños*
9	1
10	4
11	9
12	16
13	11
14	8
15	1

Ayuda virtual

Calculo de media, mediana y moda con ayuda de Excel

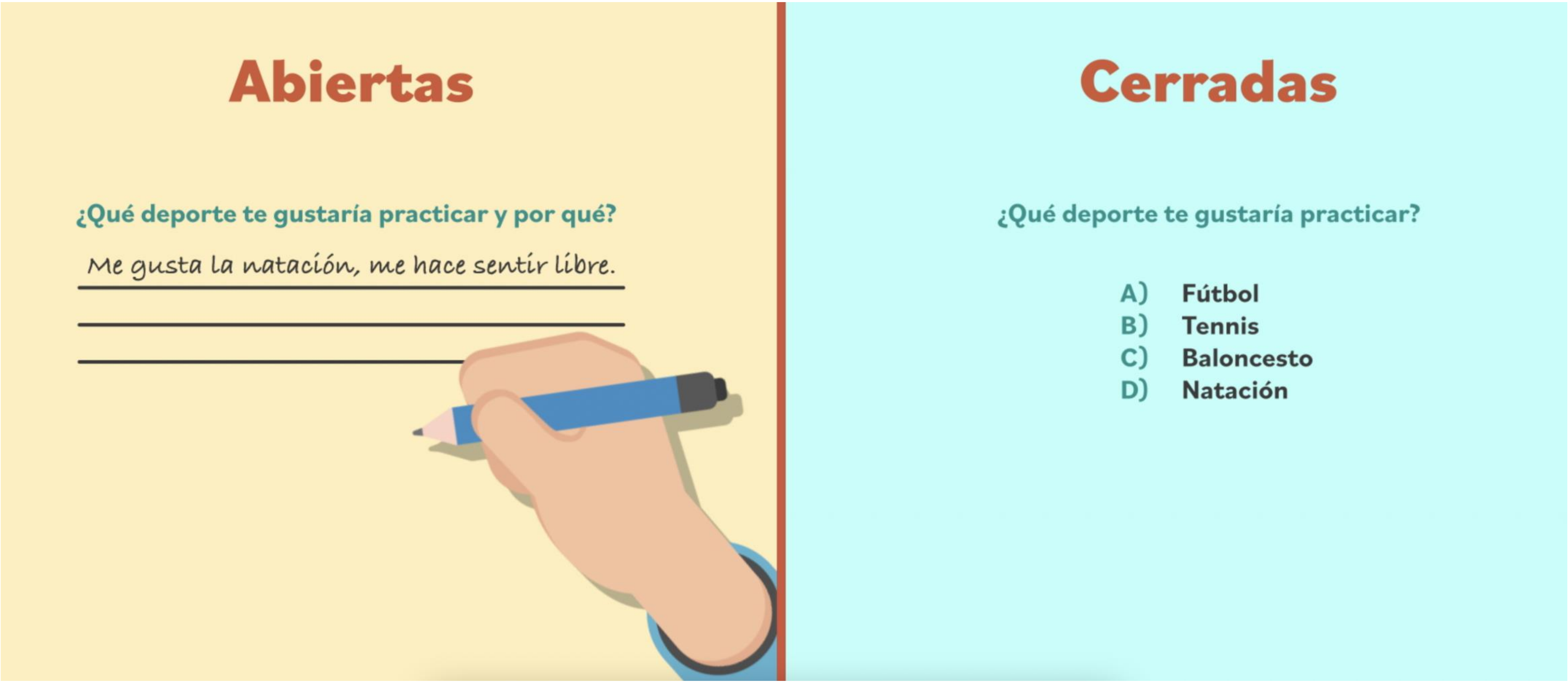
https://www.youtube.com/watch?v=1vBGPXx5t8w

La encuesta

Las encuestas son cuestionarios diseñados para que las personas respondan una serie de preguntas que buscan obtener información o conocer su opinión sobre un tema.

- El objetivo de las respuestas es convertir la información en datos que puedan ser analizados para luego, presentar los resultados en un informe.

Preguntas abiertas y cerradas



Ejemplo 1:

Se le pidió a un grupo de personas que indiquen su color favorito, y se obtuvieron los siguientes resultados:

negro	azul	amarillo	rojo	azul
azul	rojo	negro	amarillo	rojo
rojo	amarillo	amarillo	azul	rojo
negro	azul	rojo	negro	amarillo

Con los resultados obtenidos, elaborar una tabla de frecuencias.

En la primera columna, colocamos los valores de nuestra variable, en la segunda la frecuencia absoluta, luego la frecuencia acumulada, seguida por la frecuencia relativa, y finalmente la frecuencia relativa acumulada.

Color	Frecuencia absoluta	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Negro	4	4	0,20	0,20
Azul	5	9	0,25	0,45
Amarillo	5	14	0,25	0,70
Rojo	6	20	0,30	1
Total	20		1	

Ejemplo 2:

En una tienda de autos, se registra la cantidad de autos Toyota vendidos en cada día del mes de Septiembre.

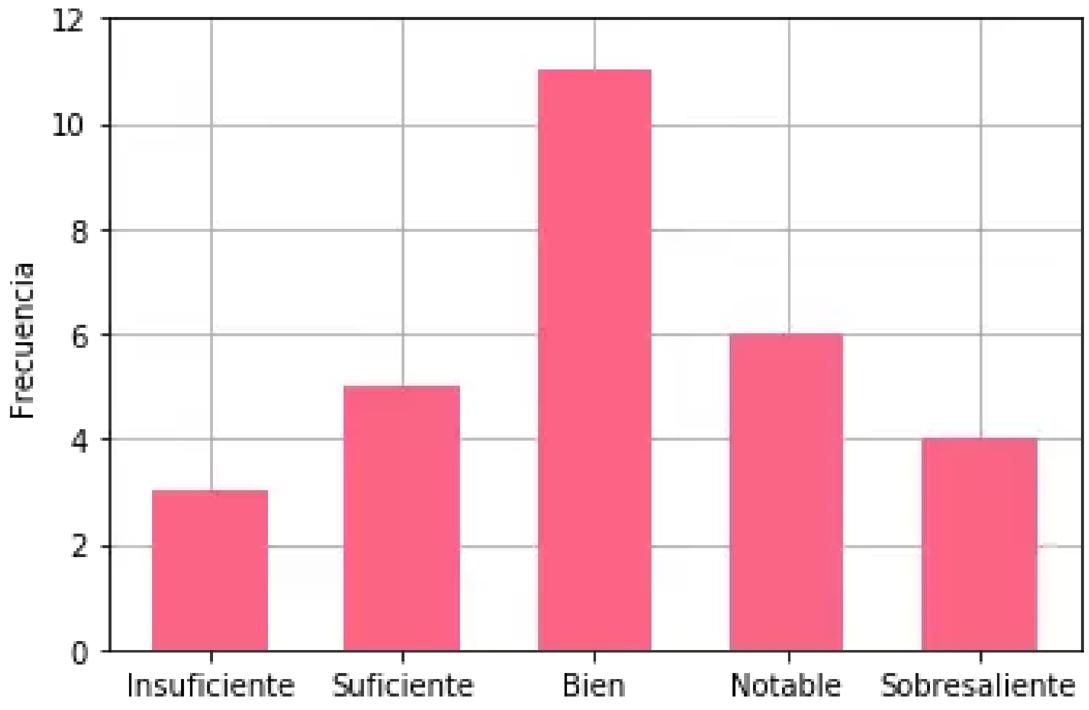
0; 1; 2; 1; 2; 0; 3; 2; 4; 0; 4; 2; 1; 0; 3; 0; 0; 3; 4; 2; 0; 1; 1; 3; 0; 1; 2; 1; 2; 3

Con los datos obtenidos, elaborar una tabla de frecuencias.

Autos vendidos	Frecuencia absoluta	Frecuencia acumulada	Frecuencia relativa	Frec. relativa acumulada	Frecuencia porcentual	Frec. porcentual acumulada
0	8	8	0,267	0,267	26,7%	26,7%
1	7	15	0,233	0,500	23,3%	50,0%
2	7	22	0,233	0,733	23,3%	73,3%
3	5	27	0,167	0,900	16,7%	90,0%
4	3	30	0,100	1	10,0%	100%
Total	30		1		100%	

Ejemplo.

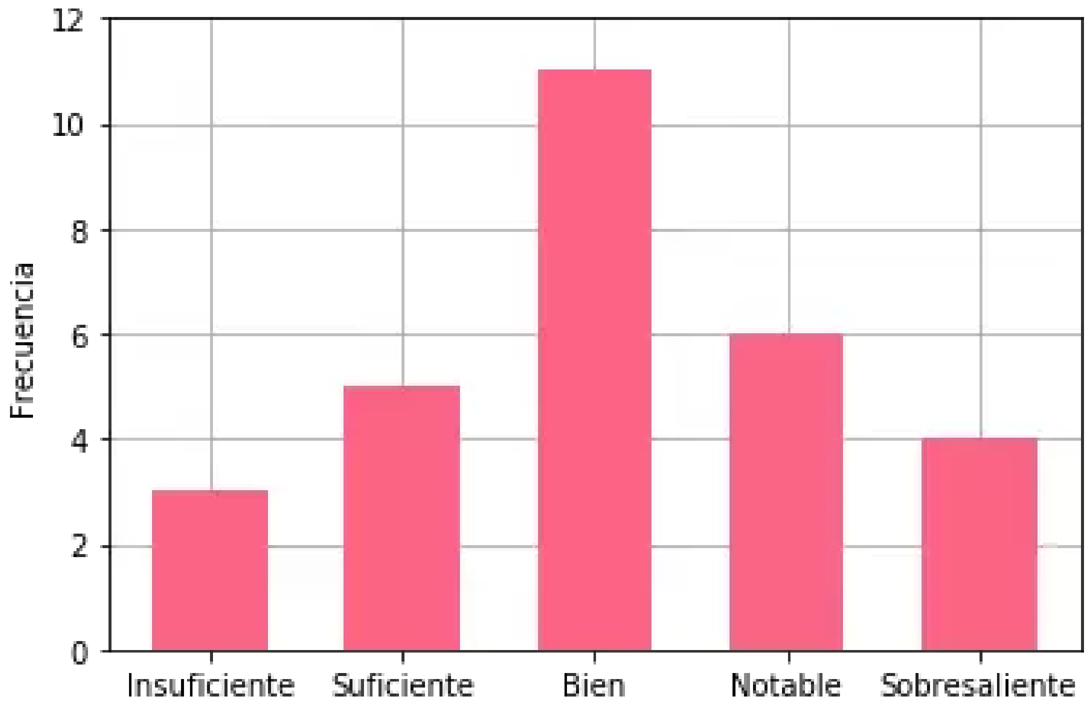
El siguiente diagrama de barras contabiliza las notas de los alumnos de una clase. Completa la tabla y responde a las preguntas:



1. ¿Qué nota es la más común?
2. ¿Cuántos estudiantes han reprobado la asignatura?
3. ¿Cuántos estudiantes han aprobado la asignatura?
4. ¿Cuántos estudiantes hay en la clase?

Ejemplo.

El siguiente diagrama de barras contabiliza las notas de los alumnos de una clase. Completa la tabla y responde a las preguntas:

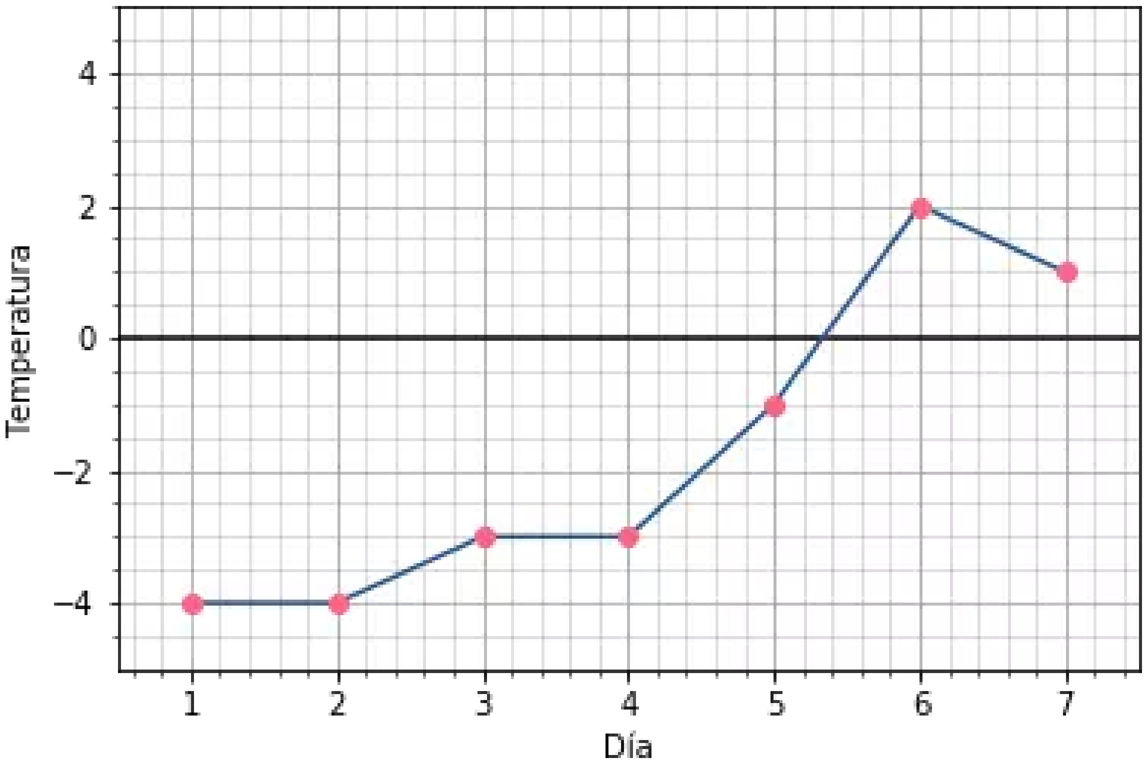


Nota	Frecuencia
Insuficiente	3
Suficiente	5
Bien	11
Notable	6
Sobresaliente	4

1. ¿Qué nota es la más común? **Bien**
2. ¿Cuántos estudiantes han reprobado la asignatura? **3**
3. ¿Cuántos estudiantes han aprobado la asignatura? **26**
4. ¿Cuántos estudiantes hay en la clase? **29**

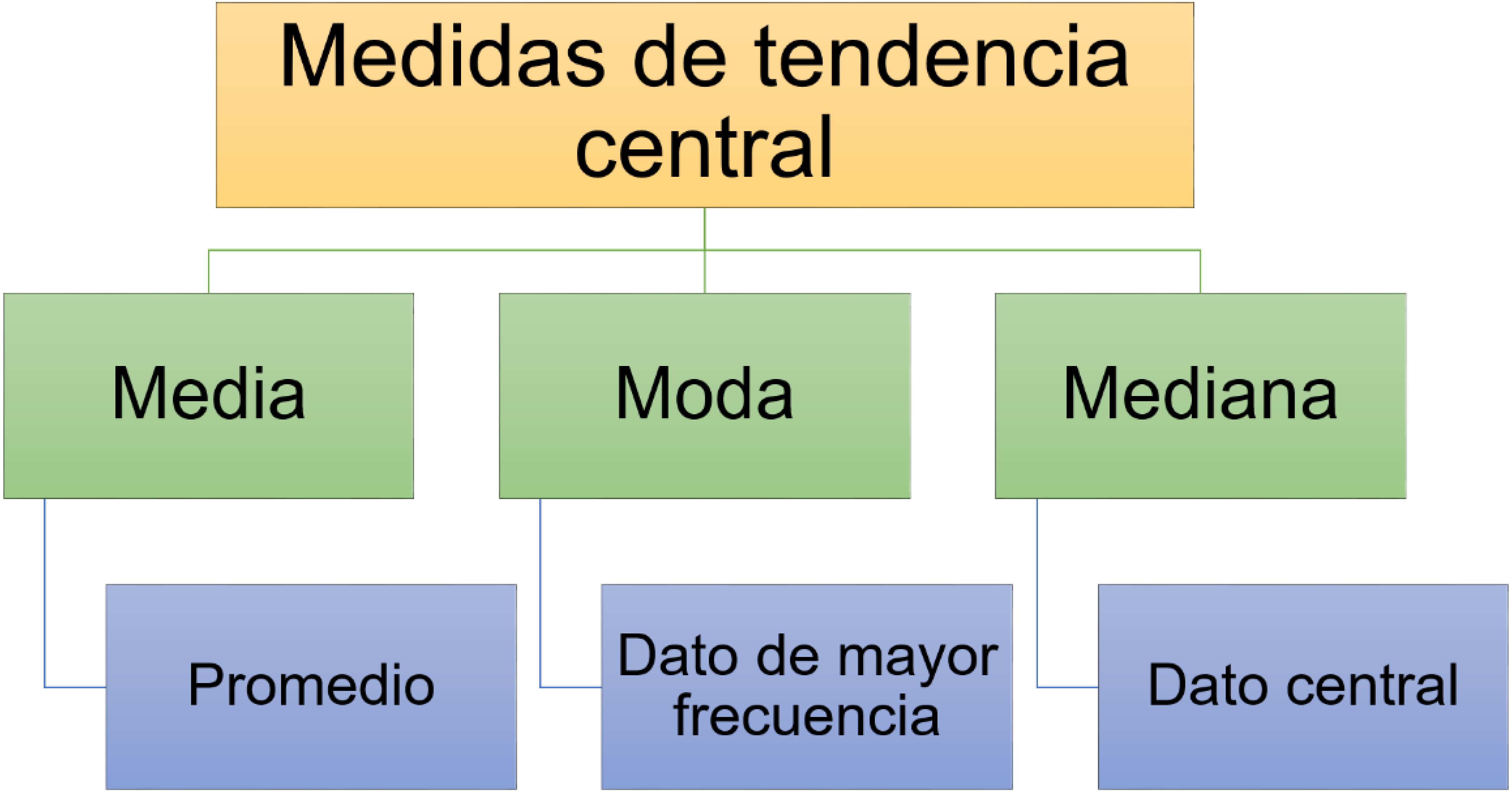
Ejemplo.

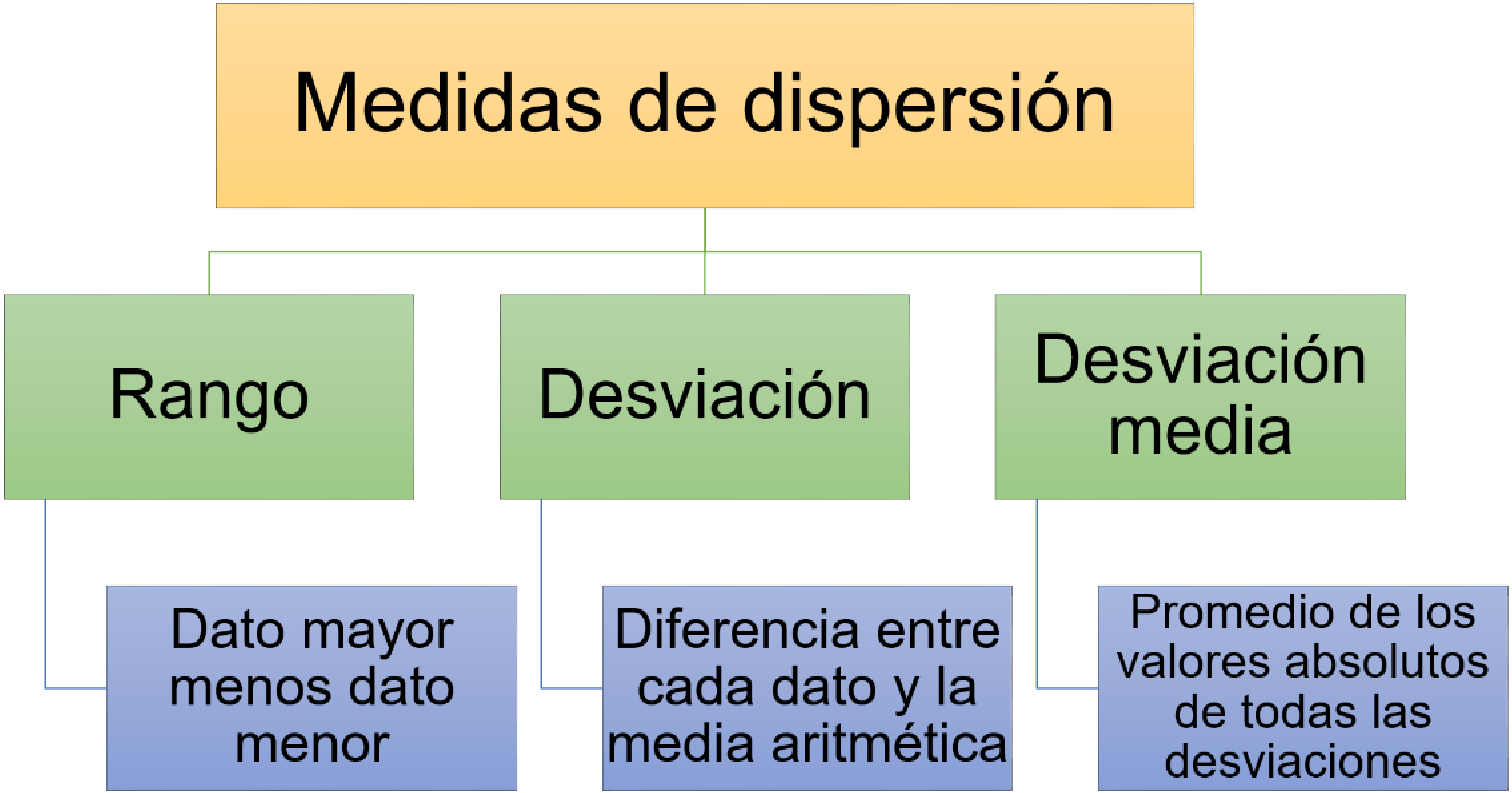
El siguiente polígono de frecuencia muestra la temperatura media diaria en una ciudad a lo largo de los siete días de una semana. Completa la tabla y responde las preguntas:



Día	Temperatura
1	
2	
3	
4	
5	
6	
7	

- ¿Qué día hizo menos frío?
- ¿La mayoría de los días, la temperatura fue bajo cero o sobre cero?
- ¿Cuál fue la temperatura los dos primeros días?


Medidas de tendencia central
Media
Moda
Mediana
Promedio
Dato de mayor frecuencia
Dato central


Medidas de dispersión
Rango
Desviación
Desviación media
Dato mayor menos dato menor
Diferencia entre cada dato y la media aritmética
Promedio de los valores absolutos de todas las desviaciones

Medidas de dispersión

Rango o recorrido

- es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

Desviación media

- es la diferencia entre cada valor de la variable estadística y la media aritmética.

Varianza

- es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

Desviación típica

- es la raíz cuadrada de la varianza.

Medidas de dispersión

- Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

Rango o recorrido

El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

Desviación media

La **desviación respecto a la media** es la diferencia entre cada valor de la variable estadística y la media aritmética.

$$\mathrm{D}_i = x_i - \overline{x}$$

La **desviación media** es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

$$\begin{aligned} D_{\overline{x}} &= \frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \cdots + |x_n - \overline{x}|}{N} \\ &= \frac{\displaystyle\sum_{i=1}^{n} |x_i - \overline{x}|}{N} \end{aligned}$$

Ejemplo: Calcular la desviación media de la distribución:

$$9,\ 3,\ 8,\ 8,\ 9,\ 8,\ 9,\ 18$$

Ejemplo: Calcular la desviación media de la distribución:

$$9, 3, 8, 8, 9, 8, 9, 18$$

Primero calculamos la media:

$$\overline{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

Calculamos la distribución media:

$$\begin{aligned} D_{\overline{x}} &= \frac{|9-9|+|3-9|+|8-9|+|8-9|+|9-9|+|8-9|+|9-9|+|18-9|}{8} \\ &= 2.25 \end{aligned}$$

Desviación media para datos agrupados

- Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

$$D_{\overline{x}} = \frac{|x_1 - \overline{x}| \cdot f_1 + |x_2 - \overline{x}| \cdot f_2 + \cdots + |x_n - \overline{x}| \cdot f_n}{N}$$

$$= \frac{\displaystyle\sum_{i=1}^{n} |x_i - \overline{x}| \cdot f_i}{N}$$

Ejemplo: Calcular la desviación media de la distribución:

Intervalo	x_i	a_i	f_i	$x_i \cdot f_i$	$—x_i - \overline{x}\|$	$—x_i - \overline{x}\| \cdot f_i$
[10,15)	12.5	5	3	37.5	9.286	27.858
[15,20)	17.5	5	5	87.5	4.286	21.43
[20,25)	22.5	5	7	157.5	0.714	4.998
[25,30)	27.5	5	4	110	5.714	22.856
[30,35)	32.5	5	2	65	10.174	21.428
			21	457.5		98.57

21 datos

Realizamos la suma de las últimas cuatro columnas y calculamos la media

$$\overline{x} = \frac{457.5}{21} = 21.786$$

La desviación media es $$\mathrm{D}_{\overline{x}} = \frac{98.57}{21} = 4.69$$

Varianza

- La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

La varianza se representa por σ^2

$$\sigma^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{N}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{N}$$

Varizanza para datos agrupados

$$\begin{aligned}\sigma^2 &= \frac{(x_1-\overline{x})^2\cdot f_1+(x_2-\overline{x})^2\cdot f_2+\cdots+(x_n-\overline{x})^2\cdot f_n}{N}\\ &= \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\cdot f_i}{N}\end{aligned}$$

Para simplificar el cálculo de la varianza vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\begin{aligned}\sigma^2 &= \frac{x_1^2+x_2^2+\cdots+x_n^2}{N}-\overline{x}^2\\ &= \sum_{i=1}^{n}\frac{x_i^2}{N}-\overline{x}^2\end{aligned}$$

Ejemplo: Calcular la varianza de la distribución:

$$9,\ 3,\ 8,\ 8,\ 9,\ 8,\ 9,\ 18$$

Primero calculamos la media:

$$\overline{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

Con el valor de la media, ya podemos encontrarla varianza:

$$\begin{aligned} \sigma^2 &= \frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8} \\ &= 15 \end{aligned}$$

Ejemplo: Calcular la varianza de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10,20)	15	1	15	225
[20,30)	25	8	200	5000
[30,40)	35	10	350	12.250
[40,50)	45	9	405	18.225
[50,60)	55	8	440	24.200
[60,70)	65	4	260	16.900
[70,80)	75	2	150	11.250
		42	1.820	88.050

Realizamos las operaciones correspondientes para obtener la tabla.

Realizamos la suma de las últimas tres columnas y calculamos la media

$$\overline{x} = \frac{1820}{42} = 43.33$$

La varianza es

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Desviación típica

La desviación típica es la raíz cuadrada de la varianza.

La desviación típica se representa por σ

$$\sigma = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{N}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1 - \overline{x})^2 \cdot f_1 + (x_2 - \overline{x})^2 \cdot f_2 + \cdots + (x_n - \overline{x})^2 \cdot f_n}{N}}$$

$$= \sqrt{\frac{\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\begin{aligned}\sigma &= \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{N} - \overline{x}^2} \\ &= \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{N} - \overline{x}^2}\end{aligned}$$

Desviación típica para datos agrupados

$$\begin{aligned}\sigma &= \sqrt{\frac{x_1^2 \cdot f_1 + x_2^2 \cdot f_2 + \cdots + x_n^2 \cdot f_n}{N} - \overline{x}^2} \\ &= \sqrt{\sum_{i=1}^{n} \frac{x_i^2 \cdot f_i}{N} - \overline{x}^2}\end{aligned}$$

Coeficiente de variación

El coeficiente de variación es la relación entre la desviación típica de una muestra y su media.

$$\mathrm{C.V} = \frac{\sigma}{\overline{x}} \cdot 100$$

Puntuaciones típicas

Las puntuaciones típicas son el resultado de dividir las puntuaciones diferenciales entre la desviación típica. Este proceso se llama tipificación.

$$\mathrm{z} = \frac{x_i - \overline{x}}{\sigma}$$

Diagrama de barras

- Un diagrama de barras se utiliza para representar datos cualitativos o datos cuantitativos de tipo discreto.

- Se representan sobre unos ejes de coordenadas, en el eje de abscisas se colocan los valores de la variable, y sobre el eje de ordenadas las frecuencias absolutas o relativas o acumuladas.

- Los datos se representan mediante barras de una altura proporcional a la frecuencia.

Ejemplo:

Un estudio hecho al conjunto de los 20 alumnos de una clase para determinar su grupo sanguíneo ha dado el siguiente resultado:

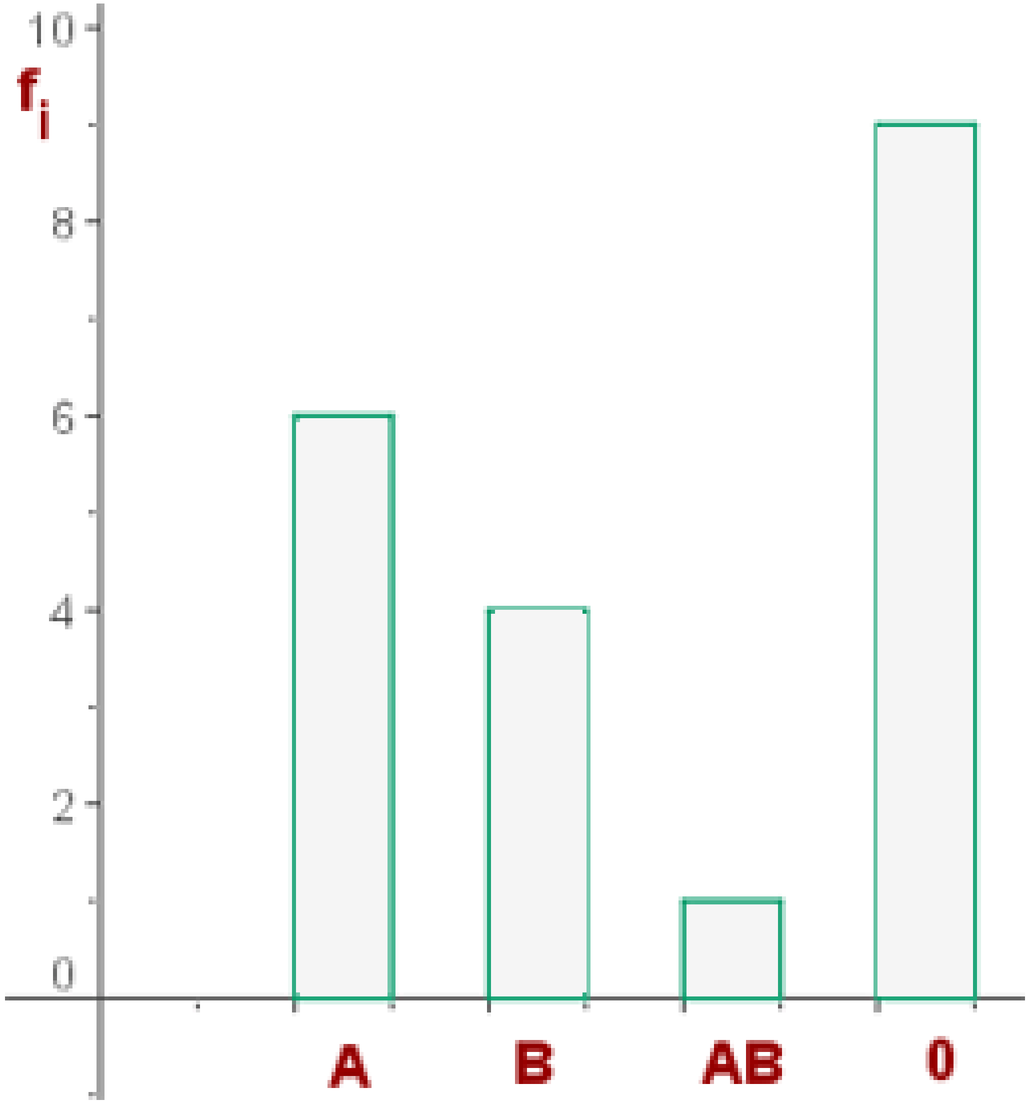
Grupo sanguíneo	f_i
A	6
B	4
AB	1
0	9
	20

Representa lo anterior mediante un diagrama de barras.

1. En el eje de las abscisas colocamos los elemento del grupo sanguíneo y en el eje de las ordenadas la frecuencias de cada elemento.

2. Representamos cada elemento del grupo sanguíneo con una barra. Todas las barras deben tener el mismo ancho.

3. Obtenemos la siguiente gráfica



Polígonos de frecuencia

- Un polígono de frecuencias se forma uniendo los extremos de las barras mediante segmentos.

- También se puede realizar trazando los puntos que representan las frecuencias y uniéndolos mediante segmentos.

Ejemplo:

Las temperaturas en un día de otoño de una ciudad han sufrido las siguientes variaciones:

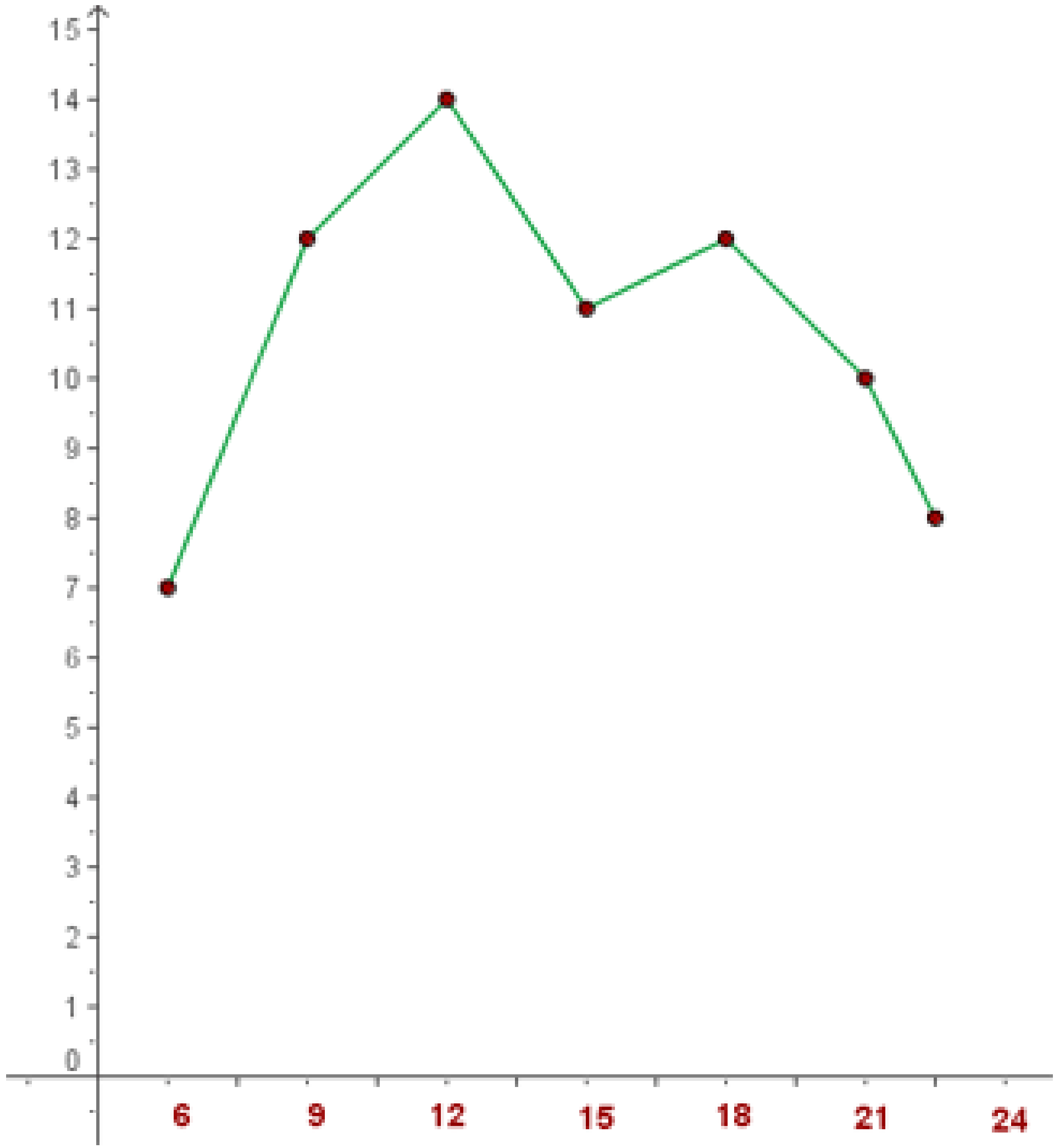
Hora	Temperatura
6	7º
9	12°
12	14°
15	11°
18	12°
21	10°
24	8°

Representa lo anterior mediante un polígono de frecuencias.

1. En el eje de las abscisas colocamos las horas y en el eje de las ordenadas las temperaturas.

2. Trazamos para cada hora de la tabla su respectiva frecuencia.

3. Obtenemos la siguiente gráfica



Ejemplo. Una tienda en línea registra el tiempo que tarda la empresa de correos en hacer llegar su mercadería a los clientes. Los tiempos en días registrados son los siguientes:

2	7	10	16	19
22	6	25	5	20
13	32	13	29	18
20	13	6	12	35