¿Cómo construir una tabla de frecuencias?

¡Vamos a tomar como ejemplo un salón de clases! Imagina que eres profesor o profesora de biología de **20 estudiantes** y tienes las notas finales del semestre. Sigue estos pasos para construir tu tabla de frecuencias:

Paso 1:

Reúne los datos.

Estudiantes	Notas
1	10
2	5
3	4
4	6
5	7
6	8
7	8
8	9
9	10
10	10
11	5
12	4
13	2
14	3
15	10
16	7
17	8
18	7
19	8
20	2

Paso 2:

Crea una nueva tabla. En la primera columna, ubica las notas de 1 a 10, de menor a mayor. En la segunda columna, escribe la cantidad de veces que se repite cada nota y llama a estos datos **frecuencia absoluta**.

Notas	Frecuencia
1	0
2	2
3	1
4	2
5	2
6	1
7	3
8	4
9	1
10	4

Paso 3:

Hasta aquí tienes una **tabla de frecuencias** sencilla, pero también puedes agregarle una columna más para calcular la **frecuencia absoluta acumulada**. Sus valores se obtienen sumando los datos en diagonal.

Por ejemplo: el primer número siempre va a ser igual al primer dato de la **frecuencia absoluta**, en este caso es **cero**. Luego, para obtener el segundo dato, necesitas sumar el **cero** con el **dos**, que es el segundo número de la frecuencia absoluta y justamente, el que está ubicado de forma diagonal. Entonces: **0 + 2 = 2**.

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada
1	0	0
2	2	2
3	1	
4	2	
5	2	
6	1	
7	3	
8	4	
9	1	
10	4	

Paso 4:

Sigue sumando los números en diagonal. Ahora es el turno de **2 + 1 = 3**. Continua hasta llenar toda la columna.

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada
1	0	0
2	2	2
3	1	3
4	2	
5	2	
6	1	
7	3	
8	4	
9	1	
10	4	

Paso 5:

Una forma de verificar que la suma es correcta, es obteniendo como número final la cantidad de datos que tienes. En este caso, sería igual a **20**, porque son las notas de **20 estudiantes**. ¡Y listo!

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada
1	0	0
2	2	2
3	1	3
4	2	5
5	2	7
6	1	8
7	3	11
8	4	15
9	1	16
10	4	20

Frecuencia relativa y la frecuencia relativa absoluta

Al inicio de esta página, te explicamos que la **frecuencia relativa** se expresa en **porcentajes**. Mira cómo puedes obtenerlos a partir de los datos que ya tienes.

Paso 1:

¡Continuemos con la tabla de frecuencias del salón de clases! Añade una cuarta columna con el nombre **frecuencia relativa**. Toma cada dato de la **frecuencia absoluta** y divídelo en **20**, que es la cantidad de datos totales que tienes. Así:

0 ÷ 20 = 0 **2 ÷ 20 = 0,1** **1 ÷ 20 = 0,05**

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa
1	0	0	0
2	2	2	0,1
3	1	3	0,05
4	2	5	
5	2	7	
6	1	8	
7	3	11	
8	4	15	
9	1	16	
10	4	20	
Total	20		

Paso 2:

Realiza las divisiones hasta obtener todos los datos. Al final, la suma de esos valores debe darte **1**.

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa
1	0	0	0
2	2	2	0,1
3	1	3	0,05
4	2	5	0,1
5	2	7	0,1
6	1	8	0,05
7	3	11	0,15
8	4	15	0,2
9	1	16	0,05
10	4	20	0,2
Total	20		1

Si al sumar el resultado que obtienes es 0,98 o un número similar, no te preocupes, **puedes aproximarlo a 1**.

Paso 3:

Para la **frecuencia relativa acumulada** debes sumar los datos en diagonal, como lo hicimos para la **frecuencia absoluta acumulada**.

Entonces, el primer número siempre va a ser igual al primer dato de la frecuencia relativa, en este caso es **cero**. Luego, para obtener el segundo dato, necesitas sumar el **cero** con el **0,1**, que es el segundo número de la frecuencia relativa y justamente, el que está ubicado de forma diagonal. Así:

0 + 0,1 = 0,1 **0,1 + 0,05 = 0,15** **0,15 + 0,1 = 0,25**

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
1	0	0	0	0
2	2	2	0,1	0,1
3	1	3	0,05	0,15
4	2	5	0,1	0,25
5	2	7	0,1	
6	1	8	0,05	
7	3	11	0,15	
8	4	15	0,2	
9	1	16	0,05	
10	4	20	0,2	
Total	20		1	

Paso 4:

Suma todos los datos en diagonal hasta llenar toda la columna. El último número que obtengas **debe ser 1**.

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
1	0	0	0	0
2	2	2	0,1	0,1
3	1	3	0,05	0,15
4	2	5	0,1	0,25
5	2	7	0,1	0,35
6	1	8	0,05	0,4
7	3	11	0,15	0,55
8	4	15	0,2	0,75
9	1	16	0,05	0,08
10	4	20	0,2	1
Total	20		1	

Paso 5:

¡Ahora sí vamos a descubrir los porcentajes de la frecuencia relativa! Toma cada valor de la columna **frecuencia relativa** y multiplícalo por 100. Por ejemplo:

0 x 100 = 0 **0,1 x 100 = 10** **0,05 x 100 = 5**

Al final, la suma de esa columna debe dar 100 %.

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada	Frecuencia relativa en %
1	0	0	0	0	0%
2	2	2	0,1	0,1	10%
3	1	3	0,05	0,15	5%
4	2	5	0,1	0,25	10%
5	2	7	0,1	0,35	10%
6	1	8	0,05	0,4	5%
7	3	11	0,15	0,55	15%
8	4	15	0,2	0,75	20%
9	1	16	0,05	0,08	5%
10	4	20	0,2	1	20%
Total	20		1		100%

Paso 6:

Para terminar, calcula el porcentaje de la **frecuencia relativa acumulada en porcentajes.** Sus valores se obtienen sumando los datos en diagonal.

Por ejemplo: el primer número siempre va a ser igual al primer dato de la **frecuencia relativa en %**, es decir, a **cero por ciento**. Luego, para obtener el segundo dato, necesitas sumar el **cero** con el **10%**, que es el segundo número de la frecuencia relativa y el que está ubicado de forma diagonal. Entonces: **0 + 10 = 10**.

Continúa:

10 + 5 = 15% **15 + 10 = 25%** **25 + 10 = 35%**

Notas	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada	Frecuencia relativa en %	Frecuencia relativa acumulada %
1	0	0	0	0	0%	0%
2	2	2	0,1	0,1	10%	10%
3	1	3	0,05	0,15	5%	15%
4	2	5	0,1	0,25	10%	25%
5	2	7	0,1	0,35	10%	35%
6	1	8	0,05	0,4	5%	40%
7	3	11	0,15	0,55	15%	55%
8	4	15	0,2	0,75	20%	75%
9	1	16	0,05	0,08	5%	80%
10	4	20	0,2	1	20%	100%
Total	20		1	1	100%	100%

El último número que obtengas debe ser **100%**.

Así de fácil puedes crear tu propia tabla de frecuencias. Solo recuerda:

- Reunir tus datos y organizarlos.
- Calcular la cantidad de veces que se repite un dato para obtener la frecuencia absoluta.
- Sumar los valores diagonalmente para obtener las frecuencias acumuladas.
- La frecuencia relativa se expresa en porcentajes.

Ahora que ya sabes cómo organizar tus datos en una **tabla de frecuencias**, aprende a presentarlos en diferentes **gráficos**. ¡Ve a las siguientes páginas!

¿Qué es un gráfico de barras?

Un **diagrama de barras** es un gráfico usado para mostrar de forma resumida un grupo de datos que puede incluir **variables cualitativas y cuantitativas**.

Características de un diagrama de barras

- Se compone de columnas o barras de diferentes alturas, estas pueden ser horizontales o verticales.
- Tiene un **eje horizontal** o **eje x**, donde se ubica una variable, por lo general, cualitativa.
- Tiene un eje vertical o eje y, donde se ponen los valores que determinan la altura de las barras. A estos números se les conoce como **frecuencia**.
- El ancho de las barras y el espacio entre cada una debe ser el mismo.
- Las barras también sirven para comparar valores.

¿Cómo se construye un gráfico de barras?

Para la fiesta de cumpleaños de Daniel su mamá decidió comprar varios sabores de helado. Solo tiene dinero para tres, así que le preguntará a cada invitado cuál es su sabor favorito entre:

- Chocolate
- Vainilla
- Fresa
- Brownie
- Chicle

Paso 1:

Con las respuestas de los 50 invitados, organiza los siguientes datos en una tabla: 18 personas votaron por el helado de chocolate, 10 por vainilla, 12 por brownie, 8 por fresa y 2 por chicle.

Recuerda que a las veces que se repite un dato se le llama **Frecuencia**.

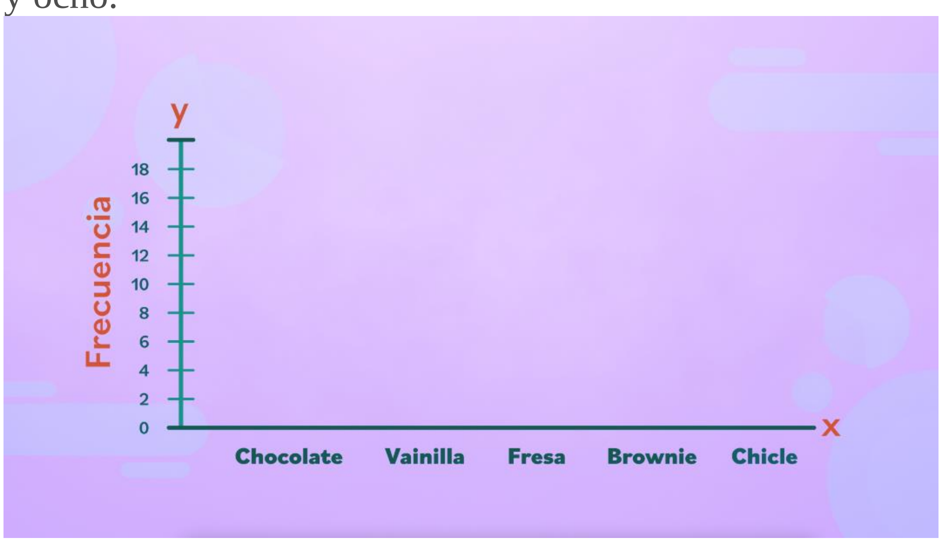
Sabor	Votos
Chocolate	18
Vainilla	10
Fresa	8
Brownie	12
Chicle	2

Paso 2:

Ubica en el **eje x** cada uno de los sabores.

Paso 3:

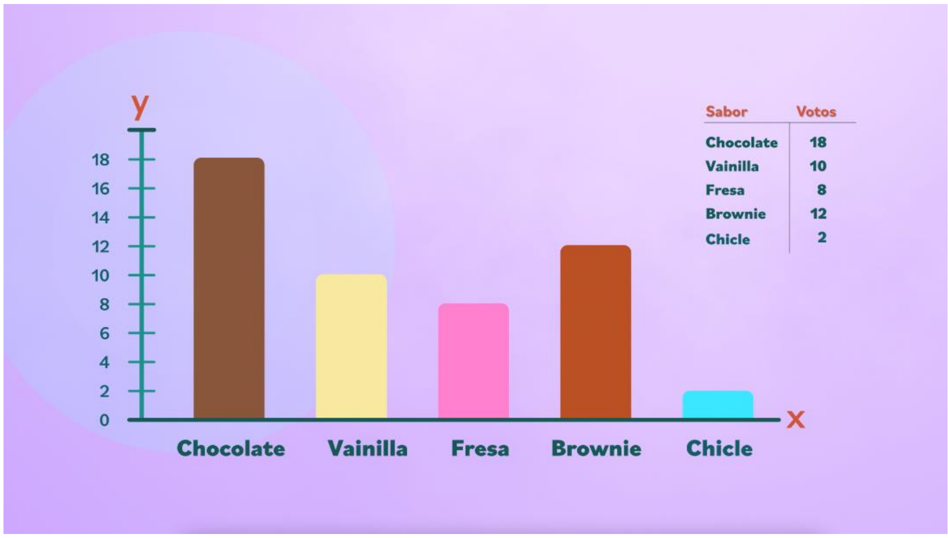
En el **eje y** coloca la cantidad de personas que votaron por cada sabor. Como el número mayor es el 18, puedes usar una frecuencia de números de dos en dos. Así: dos, cuatro, seis, ocho, diez, doce, catorce, dieciséis y diez y ocho.



La clave para elegir los números del eje y está en identificar el número más grande. Si tuvieras cifras de 200 o 355 invitados, colocar los números de dos en dos haría que la gráfica quede muy alta. En estos casos, busca una frecuencia que se acomode a lo que quieres mostrar, puede que usar números de 50 en 50, te funcione mucho mejor.

Paso 4:

Ubica las barras en dirección a donde hayas ubicado cada sabor. Su tamaño depende del número de personas que votaron. Por ejemplo, la barra de chocolate llega hasta el número 18; la de vainilla, hasta diez; y así sucesivamente hasta utilizar todos los datos de la tabla.



Paso 5:

¡Eso es todo! Con este diagrama la mamá de Daniel puede explicarle a su hijo porque los sabores de helado comprarán son: vainilla, chocolate y brownie. Después de todo, ¡son los que le gustan a la mayoría!

FORMULAS:

La fórmula para calcular la media aritmética en datos agrupados es la siguiente:

$$\bar{x} = \sum \frac{x_i \cdot f_i}{n}$$

La media se calcula sumando todos los datos y dividiendo entre el total de ellos.

¿Qué es la mediana?

La **mediana** es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor.

La **mediana** se representa por M_e

La **mediana** se puede hallar solo para **variables cuantitativas**.

Ejemplo de cálculo simple de la mediana

1. Ordenamos los datos de menor a mayor.
2. Si la serie tiene un número **impar** de medidas la mediana es la puntuación central de la misma

$$2,3,4,4,5,5,5,6,6 \qquad \mathrm{M}_e = 5$$

3. Si la serie tiene un número **par** de puntuaciones la mediana es la media entre las dos puntuaciones centrales.

7,8,9,10,11,12 $M_e = \frac{9+10}{2} = \frac{19}{2} = 9.5$

Fórmula y cálculo de la mediana para datos agrupados

La **mediana** se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas.

Es decir tenemos que buscar el intervalo en el que se encuentre.

$$M_e = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

L_i es el límite inferior de la clase donde se encuentra la mediana

$\frac{N}{2}$ es la semisuma de las frecuencias absolutas

f_i es la frecuencia absoluta de la clase mediana

F_{i-1} es la frecuencia acumulada anterior a la clase mediana

a_i es la amplitud de la clase

La **mediana** es independiente de las amplitudes de los intervalos

Fórmula para calcular la varianza

La unidad de medida de la varianza será siempre la unidad de medida correspondiente a los datos pero elevada al cuadrado. La varianza siempre es mayor o igual que cero. Al elevarse los residuos al cuadrado es matemáticamente imposible que la varianza salga negativa. Y de esa forma no puede ser menor que cero.

$$Var(X) = \frac{\sum_1^n (x_i - \bar{X})^2}{n}$$

Donde

- **X:** variable sobre la que se pretenden calcular la varianza
- **x_i:** observación número i de la variable X. i puede tomará valores entre 1 y n.
- **n:** número de observaciones.
- **$\bar{x}$:** Es la media de la variable X.

CÁLCULO DE MEDIA Y DESVIACIÓN TÍPICA PASO A PASO, (estadística descriptiva).

*El **objetivo** es hallar la media y la desviación típica de una variable discreta a partir de una tabla de frecuencias. Para ello se construirá una tabla de frecuencias y se aplicarán las fórmulas necesarias para el cálculo estadístico.*

***Prerrequisitos**: Conocer las operacones básicas, saber usar una calculadora.*

Se ha preguntado a un grupo de personas el número de veces que han ido al cine en el último trimestre. Las respuestas se recogen en la siguiente tabla:

veces:	**0**	**1**	**2**	**3**	**4**	**5**
personas:	2	20	41	26	9	2

Hallar la media y la desviación típica.

Para realizar los cálculos con mayor facilidad construiremos una tabla de frecuencias donde $\mathbf{x_i}$ será el número de veces, la variable a estudiar, y $\mathbf{f_i}$ la frecuencia con que ocurre dicha variable, es decir el número de personas que van al cine cero, una, dos... veces al año. Al lado construimos las columnas correspondientes al producto de $\mathbf{f_i}$ por $\mathbf{x_i}$, es decir $\mathbf{f_i\ x_i}$,

y la columna correspondiente al producto de la frecuencia $\mathbf{f_i}$ por el cuadrado de la variable $\mathbf{x_i}$ es decir $\mathbf{f_i\ x_i^2}$

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	2	0	0
1	20	20	20
2	41	82	164
3	26	78	234
4	9	36	144
5	2	10	50
	100	226	612

Ahora podemos **calcular la media** con facilidad dividiendo la **suma (Σ)** de todos los **f_i x_i** entre el total de personas ecuestadas (n) que es la suma de todas las **f_i**:

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{226}{100} = 2{,}26$$

Se ha ido al cine una **media de 2,26** veces.

Para calcular la **desviación típica** calculamos la raiz cuadrada de la **varianza**, que puede hacerse tal y como indica la fórmula a continuación:

$$\sigma = \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{612}{100} - 2{,}26^2} = \sqrt{1{,}0124} \approx 1{,}01$$

Así **el número medio de veces que han ido al cine en el último trimestre es 2,26 veces, con una desviación típica de 1,01 veces.**

Ejemplo paso a paso para calcular la desviación estándar

Primero necesitamos un conjunto de datos con el cual trabajar. Elijamos algo pequeño que no nos abrume por el número de datos. Este es uno bueno: 6, 2, 3, 1

Paso 1: obtener μ en $\sqrt{\frac{\sum |x-\mu|^2}{N}}$

Resumen de lo que hicimos

Descompusimos la fórmula en cinco pasos:

Paso 1: calcular la media.

Paso 2: elevar al cuadrado la distancia entre cada dato y la media.

Paso 3: obtener $\sum|x-\mu|^2$ en $\sqrt{\frac{\sum|x-\mu|^2}{N}}$

Paso 4: obtener $\frac{\sum|x-\mu|^2}{N}$ **en** $\sqrt{\frac{\sum|x-\mu|^2}{N}}$

Paso 5: calcular la desviación estándar $\sqrt{\frac{\sum|x-\mu|^2}{N}}$

¿Qué es la curtosis?

La **curtosis**, también llamada **apuntamiento**, es una medida estadística que indica el grado de concentración de una distribución alrededor de su media.

Es decir, la curtosis muestra si una distribución es escarpada o achatada. En concreto, cuanto mayor sea la curtosis de una distribución más escarpada (o apuntada) es.

En este sentido, el coeficiente de curtosis es un cálculo que se hace para cuantificar la curtosis de una distribución. Más abajo veremos cómo se calcula.

Aunque parezca contradictorio, una mayor curtosis no implica una mayor varianza, ni viceversa. Ya que la varianza es un concepto estadístico diferente a la curtosis. Si tienes dudas al respecto, puedes ver el siguiente post:

Tipos de curtosis

Hay tres **tipos de curtosis**:

- **Leptocúrtica**: la distribución es muy apuntada, es decir, los datos están muy concentrados alrededor de la media. En concreto, las distribuciones leptocúrticas se definen como aquellas distribuciones más apuntadas que la distribución normal.
- **Mesocúrtica**: la curtosis de la distribución es equivalente a la curtosis de la distribución normal. Por tanto, no se considera ni apuntada ni achatada.
- **Platicúrtica**: la distribución es muy achatada, es decir, la concentración en torno a la media es baja. Formalmente, las distribuciones platicúrticas se definen como aquellas distribuciones más achatadas que la distribución normal.

Cabe destacar que los diferentes tipos de curtosis se definen tomando como referencia la curtosis de la distribución normal.

Coeficiente de curtosis

La **fórmula del coeficiente de curtosis** es:

$$g_2 = \frac{1}{N} \cdot \frac{\sum_{i=1}^{N}(x_i - \mu)^4}{\sigma^4} - 3$$

La fórmula del coeficiente de curtosis para **datos agrupados en tablas de frecuencias**:

$$g_2 = \frac{1}{N} \cdot \frac{\sum_{i=1}^{N} f_i \cdot (x_i - \mu)^4}{\sigma^4} - 3$$

Por último, la fórmula del coeficiente de curtosis para **datos agrupados en intervalos**:

$$g_2 = \frac{1}{N} \cdot \frac{\sum_{i=1}^{N} f_i \cdot (c_i - \mu)^4}{\sigma^4} - 3$$

Donde:

- (g2)es el coeficiente de curtosis.
- (N)es el número total de datos.
- (Xi)es el dato i-ésimo de la serie.
- (μ)es la media aritmética de la distribución.
- (σ)es la desviación estándar (o desviación típica) de la distribución
- (Fi)es la frecuencia absoluta del grupo de datos i-ésimo.
- (Ci)es la marca de clase del grupo i-ésimo.

Ten en cuenta que en todas las fórmulas del coeficiente de curtosis se resta 3 porque es el valor de la curtosis de la distribución normal. De modo que el cálculo del coeficiente de curtosis se hace tomando como referencia la curtosis de la distribución normal. Por eso en ocasiones en estadística se dice que se calcula el **exceso de curtosis**.

- Si el coeficiente de curtosis es positivo, significa que la distribución es **leptocúrtica**.
- Si el coeficiente de curtosis es igual a cero, significa que la distribución es **mesocúrtica**.

- i el coeficiente de curtosis es negativo, significa que la distribución es **platicúrtica**

.