

**Martes
28
de junio**

Segundo de Secundaria Matemáticas

Los histogramas III

Aprendizaje esperado: recolecta, registra y lee datos en histogramas, polígonos de frecuencia y gráficas de línea.

Énfasis: recolectar y registrar datos en histogramas.

¿Qué vamos a aprender?

En esta sesión aprenderás a recolectar y registrar datos en histogramas.

¿Qué hacemos?

¿Sabías qué, para registrar información económica, política, científica, etcétera, se utilizan varios tipos de gráficas, entre ellas los histogramas, que nos permiten leer e interpretar dicha información?

Analiza la siguiente situación.

Los Salarios

Se registraron los salarios, en pesos, de 65 empleados de una empresa.

Se requiere hacer un análisis de estos datos.

¿Cómo podría realizarse de manera eficiente dicho análisis?

Una forma adecuada de registrar los datos para permitir su lectura y análisis es presentarlos en gráficas.

Sin embargo, previo a la presentación en gráficas, es conveniente organizar la información en una tabla.

Salarios (\$)	f
(50-60)	8
(60-70)	10
(70-80)	16
(80-90)	14
(90-100)	10
(100-110)	5
(110-120)	2
Total	65

Dato menor = 50
Dato mayor = 120

En la primera columna de esta tabla, titulada “Salarios”, se presenta, en pesos, los sueldos agrupados en 7 intervalos. En la segunda columna, encabezada con la letra “f” se encuentran las frecuencias con que cada salario aparece. Así, por ejemplo, en el segundo intervalo, que corresponde a los salarios entre 60 y 70 pesos, corresponde a la frecuencia 10 y significa que 10 empleados perciben salarios entre 60 y 70 pesos.

Para organizar los datos en una tabla como la anterior, se requiere saber la frecuencia con que aparece cada dato. La frecuencia es el número de veces que aparece un dato en el conjunto del cual forma parte.

Ejemplo:

55, 120, 110, 91, 120, 60, 73, 75, 110.

Por ejemplo, se han escrito sólo 9 datos de los salarios, en pesos, de la empresa. El dato 120, así como el dato 110, aparecen sólo 2 veces cada uno, por lo tanto, se dice que la frecuencia de cada dato es 2. Cuando se habla de datos, se hace referencia a la variable estadística que se está estudiando, en este caso serían los salarios, pero los datos pueden ser un sin número de variables, por ejemplo: Estaturas, pesos, calificaciones.

Ahora se analizará cómo se construyó la tabla de distribución de frecuencias de los salarios de los empleados de la empresa. Para ello, uno de los primeros pasos es calcular el rango. Se conoce como rango al valor que surge de restar el dato mayor menos el dato menor del conjunto, por eso el rango se define como la diferencia entre el dato mayor y el dato menor. El rango se simboliza con la letra erre mayúscula. Como el salario más alto de los empleados de la empresa es de 120 pesos y el que menor, 50 pesos, la diferencia entre ellos es 70, así que el rango es igual a 70.

Dato mayor = 120 Dato menor = 50

$$R = \text{dato mayor} - \text{dato menor}$$

$$R = 120 - 50 = 70$$

Este valor significa que la variable estadística, que en este caso la variable “salarios”, recorre 70 valores, desde el dato menor hasta el dato mayor. Pero no quiere decir que haya al menos alguno de todos esos 70 valores. Por ejemplo, supóngase que se ordenaron los salarios de menor a mayor y que se presentan únicamente los 5 primeros y los 5 últimos datos.

Ejemplo:

50, 50, 51, 52, 54...116 117,117, 120.

En los 5 primeros datos, el valor de la variable estadística “salarios” no aparece para el valor 53, y en los últimos, no aparece para los valores 118 y 119. Esto significa que el valor de la variable estadística para esos datos es cero. Pero, aun así, la variable estadística “salarios” recorre esos valores.

Una vez que se conoce el rango, se procede a calcular el número de intervalos que se presentarán en la tabla. Un intervalo es un grupo o clase de datos en los que se organiza la información estadística; también se les llama clases. Un ejemplo de intervalos son los que aparecen como 50 a 60 y 60 a 70, que quieren decir que el grupo de los salarios entre 50 y 60 pesos se encuentran registrados en el primer intervalo y los de 60 a 70 se encuentran en el segundo intervalo. Pero, ¿Cuántos intervalos se deben registrar y de qué tamaño? En los ejemplos el tamaño de los intervalos es de 10, porque si sumas 10 a 50, resulta 60, así mismo, si sumas 10 a 60, resulta 70.

El número de intervalos depende de la cantidad de datos. Este número se puede calcular obteniendo la raíz cuadrada de “n”, siendo “n” la cantidad de datos que se están organizando.

$$\sqrt{n}$$

$n = \text{cantidad de datos}$

Los datos corresponden a los salarios de los 65 empleados, por lo tanto, “n”, que es la cantidad de datos, es igual a 65. Para obtener el número de intervalos se saca la raíz cuadrada de 65.

$$\sqrt{65} = 8.06$$

Número de intervalos (NI) = 8

Tenemos que se deben construir 8 intervalos, grupos o clases, pero ¿De qué tamaño deben ser los intervalos?

Para obtener el tamaño de los intervalos, se divide el rango entre el número de intervalos. El rango que se obtuvo es 70. Dividiendo rango entre el número de intervalos, que es 8.

$$I = \frac{R}{NI} = \frac{70}{8} = 8.75 \sim 9$$

Entonces el tamaño del intervalo puede ser 9, sin embargo, los matemáticos afirman que la regla es no utilizar más de 20 intervalos ni menos de 5, en una distribución de frecuencias. Por lo que nos conviene que el número de intervalos no sea 8 ó 9, sino 7. ¿Por qué tomar esta decisión? Si dividimos 70 entre 7, resulta 10 y si tomamos 10 como el tamaño de los intervalos se facilitan la construcción de los mismos, además, no es incorrecto tomar 7 intervalos, ya que este número está entre 5 y 20, como es matemáticamente correcto.

$$I = \frac{R}{NI} = \frac{70}{7} = 10$$

En la primera columna de la tabla se presentan los siete intervalos, donde el primer número del intervalo se conoce como límite inferior y el segundo, como límite superior.

Salarios (en pesos)
50-60
60-70
70-80
80-90
90-100
100-110
110-120
Total

The diagram shows the relationship between intervals. A vertical arrow points from 'Límite inferior' to '60 - 70'. A horizontal arrow points from '50 - 60' to '60 - 70'. A diagonal arrow points from 'Límite superior' to '60 - 70'.

Su construcción sigue el siguiente procedimiento: Como el dato menor es 50, este número es el límite inferior del primer intervalo, a éste se le suman 10, que es el tamaño del intervalo, de lo que resulta 60, que es el límite superior de ese mismo intervalo. Los demás intervalos se construyen de la misma forma hasta llegar a 120, que es el límite superior del último intervalo y que coincide con el dato mayor del conjunto de datos. Observen que el límite superior del primer intervalo se convierte en el límite inferior del segundo, como se indica con las flechas rojas y negras, y así con todos los demás intervalos.

Para completar la tabla, en la segunda fila se colocan las frecuencias correspondientes a cada uno de los intervalos.

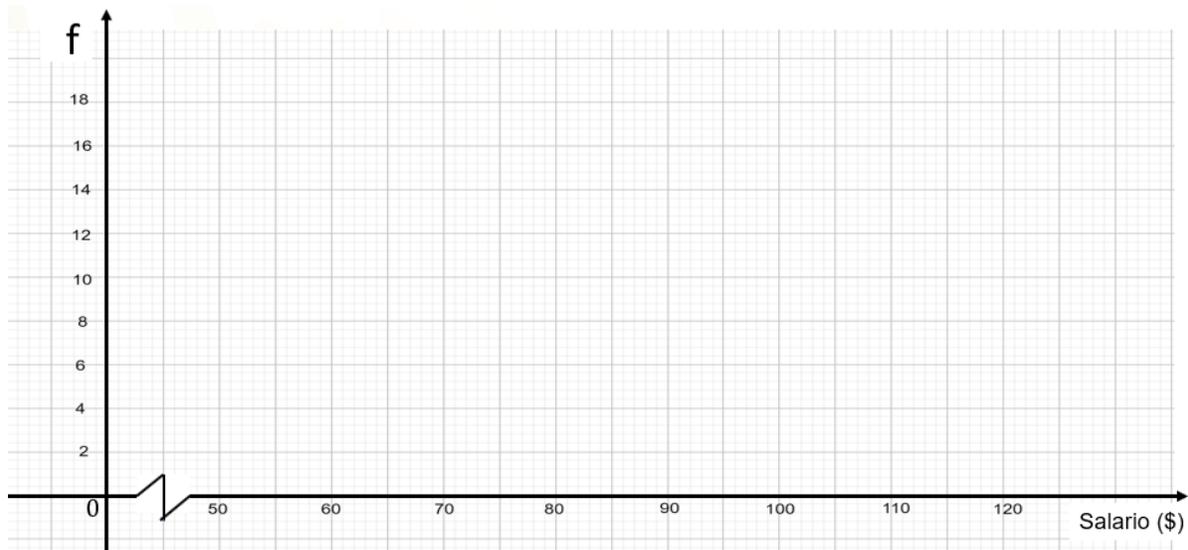
Salarios (en pesos)	f
[50-60]	8
(60-70]	10
(70-80]	16
(80-90]	14
(90-100]	10
(100-110]	5
(110-120]	2
Total	65

Entonces, por ejemplo, el intervalo 50-60, contiene los salarios de los empleados que ganan entre 50 y 60 pesos, incluyendo los salarios 50 y 60 pesos. Como se puede ver, el número 60 vuelve aparecer en el siguiente intervalo 60- 70, sin embargo, como ya se contó en el intervalo anterior, el salario 60 pesos ya no se debe contabilizar en este intervalo. Esto mismo sucede en los otros intervalos, con el límite inferior de cada uno, exceptuando al primer intervalo.

A esta forma de organizar la tabla se le conoce como distribución de frecuencias con datos agrupados, porque los datos se agrupan en intervalos.

Una distribución de frecuencias de datos agrupados, además de poder presentarse en una tabla, se puede presentar en una gráfica. En este caso la vamos a presentar en un histograma. Un histograma es una gráfica que consiste en una serie de barras que tienen sus bases sobre el eje "x" y con una altura igual a la frecuencia de los datos que representan esas barras.

Para la construcción del histograma, primero se trazan dos ejes cartesianos. El eje de las "x" se etiqueta con el nombre de la variable estadística, en este caso "Salario en Pesos". El eje de las "y" se etiqueta con la letra "f", y se elige una escala conveniente de manera que se puedan observar todas las frecuencias, es decir, es el eje de las frecuencias. Los números que corresponden a los intervalos y que, al mismo tiempo, es donde se colocan las bases de las barras del histograma, están colocados a la misma distancia; con las frecuencias sucede lo mismo.



Una vez ubicados los intervalos en el eje horizontal y las frecuencias en el eje vertical, se dibujan las barras cuya base representa el ancho de cada intervalo y su altura es la frecuencia correspondiente a cada intervalo.



En el histograma, como todos los intervalos están a la misma distancia, todas las barras deben tener la misma anchura o amplitud; también se debe cumplir que las barras deben estar pegadas unas a otras, pero ¿Por qué?

Para contestar esta pregunta, se hará una comparación de un histograma con una gráfica de barras.

La gráfica de barras representa los puntos que obtuvieron 6 equipos: A, B, C, D, E y F en un campeonato de futbol ¿Cuál es la diferencia entre esta gráfica y el histograma? En la gráfica de barras en realidad no es importante la anchura de las barras, aún

representadas con líneas, proporcionan la misma información, ya que no se trata de datos agrupados.

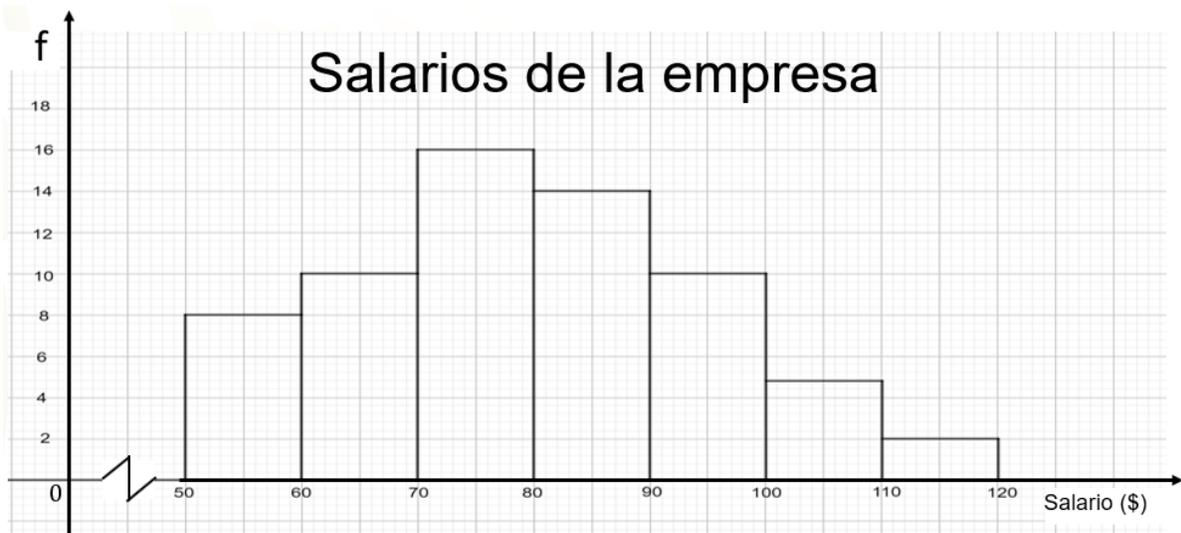


Como en un histograma se representan datos agrupados en intervalos, las barras deben ir pegadas, ya que el límite superior de un intervalo, es el límite inferior del siguiente. Como esto no sucede en la gráfica de barras, las barras no se trazan pegadas unas a otras.

La gráfica de barras se utiliza para representar datos cualitativos o atributos, así, en el eje horizontal de esta gráfica, no se representan datos cuantitativos. También se usan para representar datos cuantitativos, siempre y cuando no sean datos agrupados. Por ejemplo, las calificaciones trimestrales que van de 5 a 10, se pueden representar en una gráfica de barras, ya que no tiene caso agrupar los datos, pues la variable estadística solo recorre 6 valores: 5, 6, 7, 8, 9, y 10.

¿Será conveniente representar en una gráfica de barras los salarios de la empresa, donde la variable estadística recorre 70 valores?

Después de haber visto la diferencia entre el histograma y la gráfica de barras, analizarás de la información que se puede leer en el histograma.



La barra más alta indica que los salarios más frecuentes o que más empleados tienen en esta empresa son los que están entre 70 y 80 pesos, así mismo, que sólo dos empleados ganan entre 110 y 120 pesos, representado esto por la última barra del histograma. En la primera barra se puede leer que 8 empleados tienen el sueldo más bajo y que este sueldo se ubica entre 50 y 60 pesos.

Para continuar ejercitando la construcción de histogramas, analiza ahora otra situación.

Ahora te presentamos una situación referida a la emisión diaria de óxido de azufre, en toneladas, de una planta industrial, que son arrojadas al ambiente. Los datos aparecen desordenados.

La contaminación

Emisión diaria de óxido de azufre (en toneladas) de una planta industrial

9	25	29	19	27
29	19	16	14	20
18	14	19	11	18
19	11	23	18	15
23	9	25	23	14

Se trata de registrar estos datos en un histograma ¿Qué piensas que se puede hacer para construir ese histograma?

Una manera de iniciar es ordenar los datos, puede ser de mayor a menor o de menor a mayor.

Los datos ordenados

9	9	11	11	14
14	14	15	16	18
18	18	19	19	19
19	20	23	23	23
25	25	27	29	29

Dato mayor: 29. Dato menor: 9

en este caso se han ordenado de menor a mayor. De la tabla se puede observar fácilmente que el dato mayor es 29 y el dato menor es 9. ¿Recuerdas para qué se pueden utilizar estos dos datos?

Estos datos se requieren para calcular el rango. Recapitula los pasos a seguir para registrar este conjunto de datos en un histograma.

Los pasos son los siguientes:

- Se obtiene el rango.
- Se calcula el número de intervalos que llevará la distribución de frecuencias.
- Se determina el tamaño del intervalo.
- Se registra la frecuencia con que aparecen los datos en cada intervalo.
- Se procede a trazar los ejes, en el eje horizontal se ubican los intervalos, y en el vertical, las frecuencias; sobre el eje horizontal se determinan los intervalos o clases, que representan la base de las barras del histograma.
- Se trazan las barras del histograma con altura igual a la frecuencia del intervalo que representa.

El rango y el número de intervalos

1. El rango:

$$R = 29 - 9 = 20$$

2. El número de intervalos.

$$n = 25$$

$$NI = \sqrt{25} = 5$$

Los primeros pasos son calcular el rango y obtener el número de intervalos. El rango se obtiene de la diferencia entre el dato mayor y el dato menor.

1. El rango:

$$R = 29 - 9 = 20$$

Este valor es el rango, lo que significa que la variable estadística “toneladas de óxido de azufre” recorre 20 valores. El número de intervalos se puede calcular con la raíz cuadrada del número de datos, que en este caso es 25.

2. El número de intervalos.

$$n = 25$$
$$NI = \sqrt{25} = 5$$

Es importante no confundir el recorrido o rango de los datos con el número de datos. En este caso son 25 datos porque se registraron en la tabla la emisión de toneladas de óxido de azufre durante 25 días, el rango es 20 porque es la diferencia entre el dato mayor y menor.

Sabiendo el número de intervalos, calculamos el tamaño de cada intervalo, que también se conoce como el ancho del intervalo. Éste se obtiene dividiendo el rango entre el número de intervalos, en este caso dividiremos 20 entre 5.

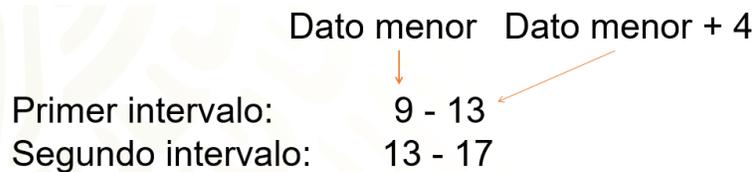
3. El tamaño del intervalo

$$I = \frac{R}{NI}$$
$$I = \frac{20}{5} = 4$$

El resultado es 4, que es el tamaño o el ancho que se utilizará para formar los intervalos y significa que la variable estadística “Toneladas de óxido de azufre” tomará 4 valores en cada intervalo.

Los intervalos se forman empezando por el dato menor; en este caso es 9, que será el límite inferior del primer intervalo, al cual se le sumarán 4, que es el tamaño que deben tener los intervalos. El resultado obtenido 13, que será el límite superior de ese intervalo; así, formamos el primer intervalo que es 9 a 13.

En el segundo intervalo, el límite superior del primer intervalo se convierte en el límite inferior, y al sumarle a éste, 4, formamos ahora el intervalo 13 a 17.



Los demás intervalos se obtienen de la misma manera: Tomando como límite inferior del siguiente intervalo, el límite superior del anterior y sumándole 4.

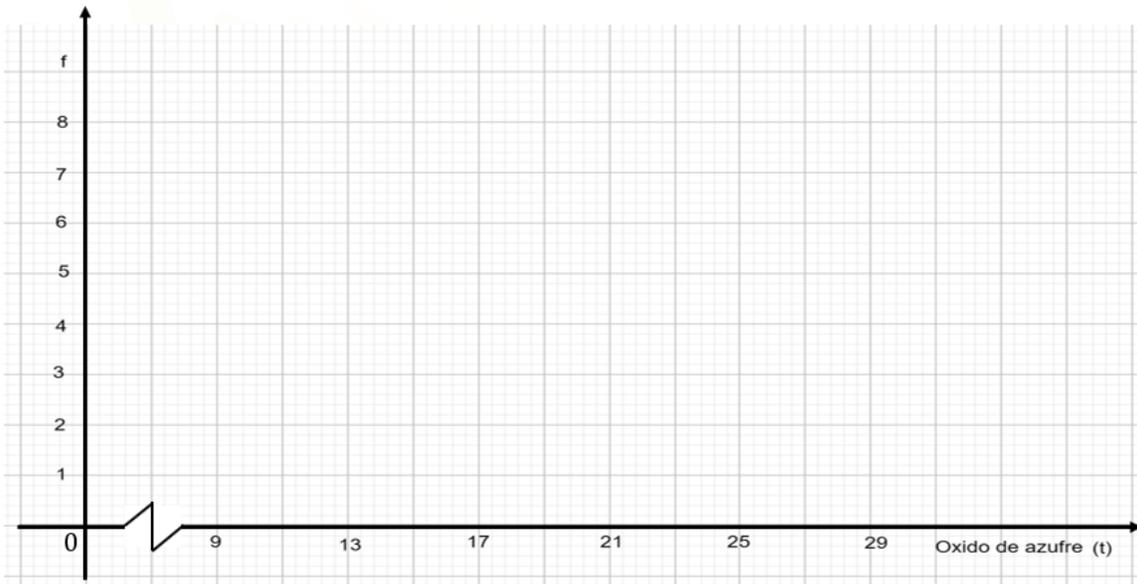
A continuación, se muestra la tabla con los datos ordenados, que ya construimos anteriormente y la tabla donde se muestran los 5 intervalos que se van a utilizar para presentar la distribución de frecuencias en el histograma. Se observa también que el límite superior del último intervalo coincide con el dato mayor, que es 29.

9	9	11	11	14	Óxido de azufre (t)	f
14	14	15	16	18	(9 – 13)	4
18	18	19	19	19	(13 – 17)	5
19	20	23	23	23	(17 – 21)	8
25	25	27	29	29	(21 – 25)	5
					(25 – 29)	3
					Total	25

El número 4 que se registra en la columna de las frecuencias, simbolizado con la letra "f", de la segunda columna en la tabla de la derecha, se refiere a que hay 4 datos entre 9 y 13. Así, se cuenta el número de veces que se repiten los datos que pertenecen a cada intervalo, este número es la frecuencia.

Ahora pon atención en el intervalo 21 – 25, ahí ya se contaron los datos que están entre 21 y 25 y te puedes percatar que el número 25 se contó para este intervalo, pero el 25 vuelve aparecer como límite inferior del último intervalo, así que, como este número ya se contó para el intervalo anterior, ya no se cuenta para el último intervalo, por ello en este solo se registran 3 datos, el 27 y los dos 29. Este hecho se indica con un corchete cuando se cuenta el dato. Para indicar que un dato es un límite, pero no se incluye en el intervalo, se coloca un paréntesis. Como en el caso del 25, donde se incluye para el intervalo de 21 a 25, pero no se incluye en el intervalo de 25 a 29.

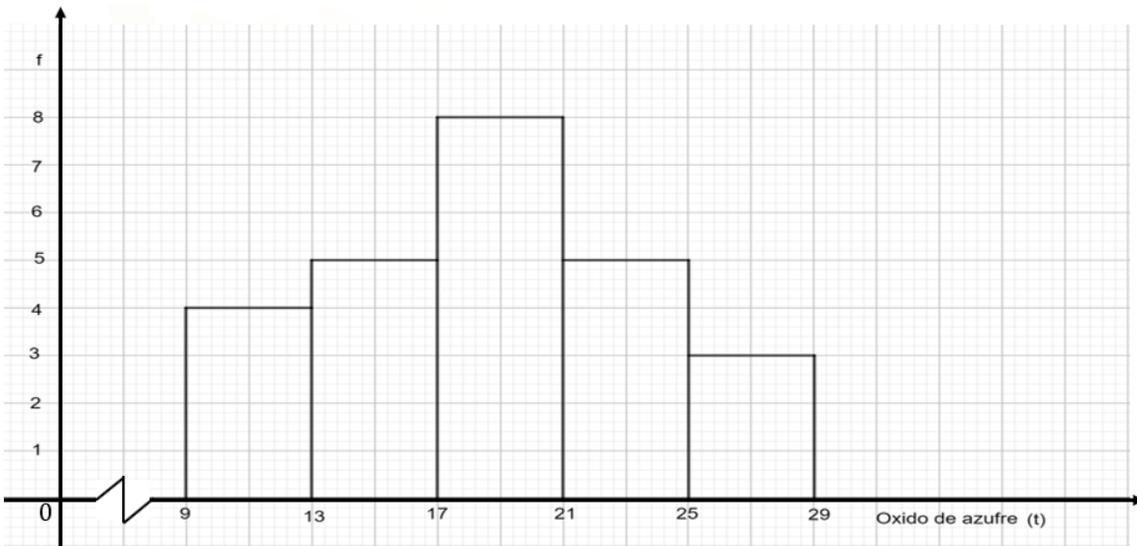
Ahora, se trazan los ejes. En el eje vertical se colocan las frecuencias y en el horizontal, los intervalos.



Observa que el eje vertical se rotula con la letra “f” de frecuencia y el horizontal, con la frase “óxido de azufre” entre paréntesis “t”, que representa la unidad de medida “toneladas”, porque los intervalos se refieren precisamente a la cantidad de toneladas de óxido de azufre que la planta industrial despidе al ambiente.

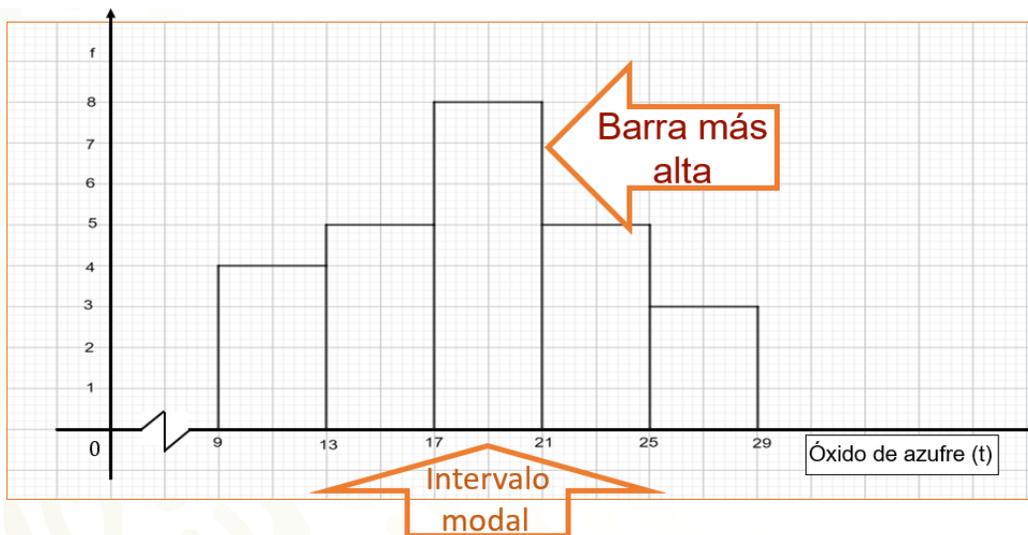
El límite superior de cada uno de los intervalos es el límite inferior del siguiente intervalo, por esa razón se coloca solo una vez el número correspondiente a los límites y no dos veces.

Cuando ya se cuenta con los ejes, se trazan las barras desde el eje horizontal, todas con el mismo ancho, una a continuación de la otra, pero con diferente altura, la cual debe corresponder a la frecuencia con que se presentan los datos de cada intervalo.



Al registro de la información en el histograma se conoce como distribución de frecuencias.

Algo que nos dice el histograma de forma inmediata es el intervalo donde se registran con mayor frecuencia los datos, porque corresponde a la barra más alta. A este intervalo se le llama intervalo modal, recordando que, en un conjunto de datos, la moda corresponde al dato de mayor frecuencia. Como, en este caso, se registran intervalos de datos, entonces lo que tiene mayor frecuencia es un intervalo, el intervalo modal.



A las frecuencias registradas en las tablas anteriores se les conoce como frecuencias absolutas, porque se presentan en números absolutos. Estos números no proporcionan una idea precisa acerca de que tan grandes o pequeños son con respecto al conjunto total de datos. Por ello se incluye una columna en tabla con las letras “f” subíndice “r”, que significan frecuencia relativa.

Óxido de azufre (t)	f	f _r
(9 – 13)	4	
(13 – 17)	5	
(17 – 21)	8	
(21 – 25)	5	
(25 – 29)	3	
Total	25	

La frecuencia relativa es el cociente de la frecuencia de cada intervalo entre el número total de datos y se utiliza para expresar la frecuencia en porcentaje. A continuación, calcularás las frecuencias relativas de los intervalos de la situación que se están trabajando, y las expresarás en porcentaje.

La frecuencia relativa es igual a la frecuencia absoluta, expresada en la fórmula con la letra “f”, entre el número total de datos, expresado con la letra “N” mayúscula.

$$f_r = \frac{f}{N}$$

Entonces, para calcular la frecuencia relativa del primer intervalo, cuya frecuencia absoluta es 4, procedemos a dividir 4 entre 25.

Óxido de azufre (t)	f	f _r	f _r (%)
(9 – 13)	4	0.16	16

$$f_r = \frac{4}{25} = 0.16$$

Esta manera de expresar la frecuencia relativa significa, en este caso, que 4 de 25 días, la planta industrial arrojó entre 9 y 13 toneladas de óxido de azufre.

Para obtener la expresión porcentual de la frecuencia relativa, se multiplica el cociente anterior, 0.16, por 100 que representa el 16%.

En porcentaje

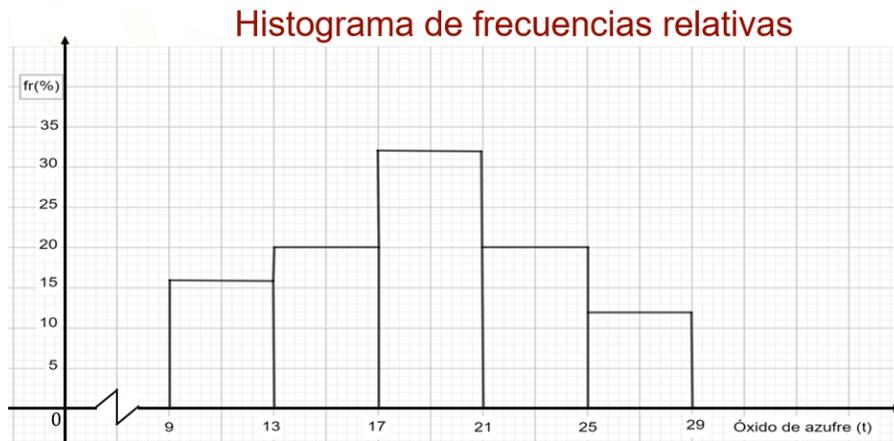
$$0.16 \times 100 = 16\%$$

En la tabla se incluyen dos columnas más, una para la frecuencia relativa y otra para los porcentajes, en que se expresa dicha frecuencia.

Óxido de azufre (t)	f	f _r	f _r (%)
(9 – 13)	4	0.16	16
(13 – 17)	5	0.2	20
(17 – 21)	8	0.32	32
(21 – 25)	5	0.2	20
(25 – 29)	3	0.12	12
Total	25	1	100

Nota que la suma de las frecuencias absolutas es 25, que es el número total de datos. Ahora bien, la suma de las frecuencias relativas es igual a 1, es decir que los 25 datos representan una unidad, y, por último, la suma de los porcentajes es 100, significa que los 25 datos corresponden al 100%, es decir, al total.

Para trazar el histograma de frecuencias relativas en porcentajes se usa un procedimiento similar al utilizado para frecuencias absolutas, lo que cambia es que en el eje vertical se deben ubicar los porcentajes, en una escala adecuada, los cuales expresan dichas frecuencias. Además, se rotula el eje vertical con las letras “f” subíndice “r” y, entre paréntesis, el símbolo de porcentaje.



En el eje horizontal se observan los mismos datos que en el histograma para las frecuencias absolutas. Así, puedes ver, entre otros hechos, que la planta industrial arroja el 32 por ciento de las veces entre 17 y 21 toneladas de óxido de azufre, y que esto corresponde al mayor porcentaje.

Recuerda enriquecer tu conocimiento sobre este tema consultando en otras fuentes, como tus libros de texto, páginas electrónicas o enciclopedias.

¡Buen trabajo!

Gracias por tu esfuerzo.