

ORGANIZACIÓN DE LOS DATOS.

La toma de datos es una de las partes de mayor importancia en el desarrollo de una investigación. Así los datos obtenidos mediante un primer proceso reciben el nombre de datos sin tratar o en bruto. Los datos en bruto son largas listas de números que no son de gran utilidad y no brindan al investigador la información que requiere si antes no se tratan. Los datos sin tratar se les debe sintetizar o resumir de manera que sea posible interpretarlos, entenderlos y utilizarlos. La manera de organizar los datos es mediante **tablas de distribución de frecuencias**.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS.

En estadística existen investigaciones relacionadas con los precios de los productos de la dieta diaria, la estatura y el peso de un grupo de individuos, los salarios de los empleados, los grados de temperatura del medio ambiente, las calificaciones de los estudiantes, etc., que pueden adquirir diferentes valores gracias a una unidad apropiada, que recibe el nombre de variable. La representación numérica de las variables se denomina dato estadístico.

Las tablas de distribución de frecuencia es una disposición tabular de datos estadísticos, ordenados ascendente o descendientemente, con la frecuencia (**fi**) de cada dato. Las distribuciones de frecuencias pueden ser para *datos no agrupados* y para *datos agrupados* o de intervalos de clase.

TABLA DE DISTRIBUCIÓN DE FRECUENCIA PARA DATOS NO AGRUPADOS.

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto sin que se haya hecho ninguna modificación al tamaño de las unidades originales. En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

La distribución de frecuencias no agrupadas o tabla con datos no agrupados se emplea si las variables toman un número pequeño de valores o la variable es discreta (siempre se asocia con valores enteros).

Tipos de frecuencias.

Frecuencia absoluta (fi).

La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico.

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

$$f_{i_1} + f_{i_2} + f_{i_3} \dots f_{i_n} = N$$

Frecuencia relativa (fri).

La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos.

Se puede expresar en tantos por ciento y se representa por **fri**.

$$fri = \frac{fi}{N} \quad \text{La suma de las frecuencias relativas es igual a 1.}$$

Frecuencia acumulada (fai).

La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

Frecuencia relativa acumulada (frai).

La 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

$$frai = \frac{fai}{N}$$

Ejemplo resuelto.

Los siguientes datos corresponden al número de libros leídos por un grupo de 30 alumnos, en un lapso de 2 años. Elaborar la tabla de distribución de frecuencias.

3	2	1	4	5	3	2	1	3	1
2	3	5	1	2	2	1	3	4	2
3	4	0	1	2	2	0	1	2	3

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

N° de libros leídos.	Frecuencia a fi	Frecuencia relativa fri	Frecuencia acumulada fai	Frecuencia relativa acumulada frai
0	2	$\frac{2}{30} = 0.066$	2	$\frac{2}{30} = 0.066$
1	7	$\frac{7}{30} = 0.23$	2+7=9	$\frac{9}{30} = 0.3$
2	9	$\frac{9}{30} = 0.3$	9+9=18	$\frac{18}{30} = 0.66$
3	7	$\frac{7}{30} = 0.23$	18+7=25	$\frac{25}{30} = 0.83$
4	3	$\frac{3}{30} = 0.1$	25+3=28	$\frac{28}{30} = 0.93$
5	2	$\frac{2}{30} = 0.066$	28+2=30	$\frac{30}{30} = 1$
Total	30	1	30	

Nota. Los valores de la frecuencia relativa y la frecuencia relativa acumulada al multiplicarse por 100%, nos daría el porcentaje que representa la variable que se analiza.

¿Qué porcentaje de alumnos leyeron 5 libros? $0.066 \times 100\% = 6.6\%$

¿Qué porcentaje de alumnos leyeron menos de 4 libros? $0.83 \times 100\% = 83\%$

TABLA DE DISTRIBUCIÓN DE FRECUENCIA DE CLASE O DE DATOS AGRUPADOS.

Es aquella distribución en la que la disposición tabular de los datos estadísticos se encuentra ordenada en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clase. No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o datos no agrupados; sin embargo, se sugiere que cuando el número total de datos (N) es igual o superior 50 y además el rango o recorrido de la serie de datos es mayor de 20, entonces, se utilizará *la distribución de frecuencia para datos agrupados*, también se utilizará este tipo de distribución cuando se requiera elaborar gráficos lineales como el histograma, el polígono de frecuencia o la ojiva.

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 le puede asociar con un números racionales e irracionales).

La organización de los datos generalmente implica el arreglo de las observaciones en clases. Al arreglo de los datos para expresar la frecuencia de ocurrencia de las observaciones en cada una de estas clases se conoce como distribución de frecuencias. La construcción de una tabla de distribución de frecuencias requiere en primer lugar de la selección de los **intervalos de clase**.

Aun cuando la selección de los intervalos de clase es un arte y depende de los datos involucrados, los siguientes pasos resultaran útiles:

Paso 1.

Ordenar los datos de menor a mayor para su clasificación.

Paso 2.

Calcular el **rango (R)** de los datos, es decir, la longitud del **intervalo (I)** que los contiene:

$$R = X_{\text{máx}} - X_{\text{min}} \quad I = [X_{\text{min}}, X_{\text{máx}}]$$

R=Rango.

$X_{\text{máx}}$ = Dato de mayor valor numérico.

X_{min} =Dato de menor valor numérico.

I= Intervalo numérico.

Paso 3.

Definir el **número de clases (Nc)**, el cual no deberá ser tan pequeño (menos de 6) o tan grande (más de 20) que la verdadera naturaleza de la distribución sea imposible visualizar.

A continuación se debe elegir un número específico de clases o categorías en las que se deben clasificar los datos, la elección del número de clases o categorías es de manera arbitraria, sin embargo deberá elegirse un número suficiente de clases de modo que los datos no queden amontonados, pero tampoco se elijen demasiados puesto que las tablas de distribución de frecuencias sería difícil de manejar. Existen dos métodos para determinar el número de clases, los cuales son los más utilizados:

1er. Método.

Método de la raíz, el cual consiste en extraer la raíz cuadrada del tamaño de la muestra y redondear el resultado así obtenido al entero mayor.

$$Nc = \sqrt{n}$$

Nc=Número de clases.

n= Tamaño de la muestra.

2do. Método.

Método de **Sturges**. Este método puede darnos una aproximación razonable para determinar el número de clases; éste se obtiene con el siguiente modelo matemático.

$$Nc = 1 + 3.3 (\log n)$$

Nc= Número de clases.

Log= Logaritmo de base 10.

n= Tamaño de la muestra.

Paso 4.

El siguiente paso consiste en determinar la **Amplitud de clase (Ac)**, es conveniente que el tamaño de las clases sea un número entero, por tanto es necesario redondear el cociente al entero más cercano.

$$Ac = \frac{R}{Nc}$$

Ac= Amplitud de clase.

R=Rango.

Nc = Número de clases.

Paso 5.

A continuación clasifiquemos los datos en cada una de las clases definiendo el **límite inferior (Li) y el límite superior (Ls)** de cada clase. En el caso en que se llegue a presentar un exceso o faltante con respecto al dato de mayor valor numérico (Xmáx) se deberá repartir el exceso o faltante de la manera más equitativa posible entre los extremos del intervalo.

Paso 6.

Definir los límites reales de clase de acuerdo a la siguiente forma:

Límite real inferior (Lri)= Li - 0.5

Límite real superior (Lrs)= Ls + 0.5

Paso 7.

Definir las **frecuencias de cada clase (fi)**, los cuales son el número de datos que quedan incluidos en cada intervalo de clase.

Paso 8.

Definir las **frecuencias relativas de cada clase (fri)**, éstas son porcentajes de los datos que hay en cada intervalo de clase con respecto al tamaño de la muestra, por lo que para obtenerlas se divide la frecuencia de cada intervalo de clase entre el tamaño de la muestra.

$$fri = \frac{fi}{n}$$

fri=Frecuencia relativa de la clase i.

fi= Frecuencia de la clase i.

n=Tamaño de la muestra.

Paso 9.

Determinar la **marca de clase (mi)**, es el punto medio de la amplitud de clase.

$$mi = \frac{Li + Ls}{2}$$

mi= marca de clase de la clase i.

Paso 10.

Definir la **frecuencias acumuladas (fai)**, se pueden obtener sencillamente si se suman las frecuencias de los intervalos de clase precedentes.

Paso 11.

Definir la frecuencia relativa acumulada (frai), Se obtiene dividiendo cada una de las frecuencias acumuladas entre el tamaño de la muestra (n)

$$frai = \frac{fai}{n}$$

frai=Frecuencia relativa acumulada de la clase i.

fai= Frecuencia acumulada de la clase i.

n=Tamaño de la muestra.

Ejemplo resuelto 1.

Los siguientes datos corresponden al número de clientes que acudieron al “CAFÉ INTERNET INN” durante 40 días.

30 35 15 21 18 32 39 20 19 20
 20 34 13 13 20 35 30 17 30 31
 10 32 22 14 30 36 23 14 20 34
 20 28 20 16 31 38 22 12 28 13

Para elaborar la tabla de distribución de frecuencia, seguir los siguientes pasos.

Paso 1. Ordenar de menor a mayor.

10 12 13 13 13 14 14 15 16 17 18 19 20 20 20 20 20 20 20 21 22 22 23 28 28
 30 30 30 30 31 31 32 32 34 34 35 35 36 38 39.

Paso2.
 Calcular el rango
 $R = X_{\text{máx}} - X_{\text{min}}$
 $R = 39 - 10$
 $R = 29$

Paso 3.
 Calcular el número de clases
 $Nc = \sqrt{n}$
 $Nc = \sqrt{40}$
 $Nc = 6.32 = 6$

Paso 4.
 Calcular la amplitud de clase.

$$Ac = \frac{R}{Nc}$$

$$Ac = \frac{29}{6} = 4.8 = 5$$

Paso 5 al paso 11.

Tabla de distribución de frecuencias

Clase	Límites de clase		Límites reales de clase		Frecuencia	Frecuencia relativa	marca de clase	frecuencia acumulada	frecuencia relativa acumulada
N°	Li	Ls	Lri	Lrs	fi	fri	mi	fai	frai
1	10	14	9.5	14.5	7	0.175	12	7	0.175
2	15	19	14.5	19.5	5	0.125	17	12	0.3
3	20	24	19.5	24.5	11	0.275	22	23	0.575
4	25	29	24.5	29.5	2	0.05	27	25	0.625
5	30	34	29.5	34.5	10	0.25	32	35	0.875
6	35	39	34.5	39.5	5	0.125	37	40	1

Como llenar la tabla.

Para los límites de clase.

Para la clase 1

$$Li = X_{\min} = 10$$

$Ls = X_{\min} + (Ac-1)$ esto es porque el intervalo debe corresponder a la Ac.

$$Ls = 10 + 4 = 14$$

Para la clase 2

$$Li = Ls \text{ de la clase anterior} + 1$$

$$Li = 14 + 1 = 15$$

$$Ls = Li \text{ de la clase 2} + (Ac-1)$$

$$Ls = 15 + 4 = 19$$

Así sucesivamente para las clases posteriores.

Para los límites reales de clase.

Para la clase 1

$$Lri = \text{Límite real inferior} = Li - 0.5$$

$$Lri = 10 - 0.5 = 9.5$$

$$Lrs = \text{Límite real superior} = Ls + 0.5$$

$$Lrs = 14 + 0.5 = 14.5$$

Así sucesivamente para las clases posteriores.

Para la frecuencia de cada clase.

f_i = Número de datos que quedan incluidos en cada intervalo de clase.

Para la clase 1.

Para saber cuántos datos quedan incluidos en el intervalo $I = [10, 14]$, se deberá contar en los datos ordenados, para esta primera clase $f_i = 7$, que corresponde a los datos 10, 12, 13, 13, 14 y 14

Para la clase 2.

Para saber cuántos datos quedan incluidos en el intervalo $I = [15, 19]$, se deberá contar en los datos ordenados, para esta primera clase $f_i = 5$, que corresponde a los datos 15, 16, 17, 18 y 19.

Así sucesivamente para las clases posteriores.

Para la frecuencia relativa.

Para la clase 1.

$$fri = \frac{f_i}{n}$$

$$fri = \frac{7}{40} = 0.175$$

Para la clase 2.

$$fri = \frac{f_i}{n}$$

$$fri = \frac{5}{40} = 0.125$$

Así sucesivamente para las clases posteriores.

Para el cálculo de la marca de clase (mi).

$$mi = \frac{Li + Ls}{2}$$

Para la clase 1

$$mi = \frac{10 + 14}{2} = \frac{24}{2} = 12$$

Para la clase 2.

$$mi = \frac{15 + 19}{2} = \frac{34}{2} = 17$$

Así sucesivamente para las clases posteriores.

Para la frecuencia acumulada (fai).

Se pueden obtener sencillamente si se suman las frecuencias de los intervalos de clase precedentes.

Para la clase 1.

Como no hay clases precedentes fai= la frecuencia de la clase 1.

$$fai=7$$

Para la clase 2.

Como ya es la segunda clase, sumamos la frecuencia de la clase 2 más la frecuencia de las clases precedentes.

$$fai= 5+7=12$$

Para la clase 3.

$$fai= 11+5+7=23$$

Así sucesivamente para las clases posteriores.

Para la frecuencia relativa acumulada.

$$frai = \frac{fai}{n}$$

Para la clase 1.

$$frai = \frac{7}{40} = 0.175$$

Para la clase 2.

$$frai = \frac{12}{40} = 0.3$$

Así sucesivamente para las clases posteriores

Ejemplo resuelto 2.

Los siguientes datos corresponden al número de goles anotados en 40 partidos de fútbol rápido en la liga local.

3	8	2	3	8	4	10	15	4	5
7	12	6	1	9	6	4	4	7	13
8	18	9	2	11	9	7	2	14	2
1	24	10	4	2	7	6	6	1	8

En base a estos datos se construyó la siguiente tabla de distribución de frecuencia.

Tabla de distribución de frecuencias

Clase	Límites de clase		Límites reales de clase		Frecuencia	Frecuencia relativa	marca de clase	frecuencia acumulada	frecuencia relativa acumulada
N°	Li	Ls	Lri	Lrs	fi	fri	mi	fai	frai
1	1	4	0.5	4.5	15	0.375	2.5	15	0.375
2	5	8	4.5	8.5	13	0.325	6.5	28	0.7
3	9	12	8.5	12.5	7	0.175	10.5	35	0.875
4	13	16	12.5	16.5	3	0.075	14.5	38	0.95
5	17	20	16.5	20.5	1	0.025	18.5	39	0.975
6	21	24	20.5	24.5	1	0.025	22.5	40	1

40

En base a la tabla anterior contestar lo siguiente:

1.- ¿Cuál fue el menor número de goles que se anotó en la liga durante los 40 días?

La respuesta a esto está en nuestro menor dato, que es de 1 gol anotado.

2- ¿En cuántos partidos se anotó entre 1 y 8 goles?

La respuesta es en 28 partidos, lo cual corresponde a las frecuencias de clases 1 y 2.

3.- ¿Qué porcentaje de partidos se anotó entre 5 y 8 goles?

La respuesta se encuentra en la frecuencia relativa de la clase 2 y corresponde a $0.325 \times 100\% = 32.5\%$

4.- ¿Qué porcentaje de los partidos se anotó 12 goles o menos?

La respuesta se encuentra en la frecuencia relativa acumulada desde la clase 1 hasta la clase 3, esta es de $0.875 \times 100\% = 87.5\%$

5.- ¿Qué porcentaje de los partidos se anotó más de 12 goles?

Solo en las clases 4, 5 y 6 se anotaron más de 12 goles, y el porcentaje se obtiene sumando las frecuencias relativas de dichas clases, siendo $0.075 + 0.025 + 0.025 = 0.125 \times 100\% = 12.5\%$

6.- ¿En la mayoría de los partidos entre cuántos goles se anotaron?

La respuesta es entre 1 y 4 goles que corresponden a 15 partidos.