



## UNIDAD II DISTRIBUCIÓN DE FRECUENCIAS

### Distribución de Frecuencias:

Es un método para organizar y resumir datos en una tabla estadística, en el que se asocia cada valor de una variable con su frecuencia.

Aunque con el proceso de agrupamiento general se pierde parte del detalle original de los datos, tiene la importante ventaja de presentarlos “todos” en un sencillo cuadro que facilita el hallazgo de las relaciones que pueda haber entre ellos.

La distribución de frecuencias puede ser simple o agrupada

- **Distribución de Frecuencias Simple (No Agrupada):** Es una tabla que se construye con base en los siguientes datos: clase o variable (valores numéricos) en orden descendente o ascendente, tabulaciones y frecuencia. Cada Valor de **X** en la distribución aparece solo.

*Ejemplo:* Si tenemos el siguiente conjunto de datos:

B	C	A	B	A	D	C	B	C	B
---	---	---	---	---	---	---	---	---	---

Si **X** representa a una variable, puede usarse una distribución de frecuencias para representar este conjunto de datos, enumerando los valores **X** con sus frecuencias. Por ejemplo el valor A se presenta 2 veces en la muestra, por tanto la frecuencia de  $X=A$  es 2.

<b>X</b>	<b>f</b>
A	2
B	4
C	3
D	1
<b>N=</b>	10

- **Distribución de Frecuencias Agrupadas (acumuladas):** Cuando un gran conjunto de datos tiene muchos valores  $X$  distintos.

Los pasos para construir la distribución de frecuencia agrupada son los siguientes:

- 1)- Localizar el dato mayor y el dato menor de la serie de datos.
- 2)- Determinar el Rango ( $R$ ). El rango mide la extensión total de un conjunto de datos. Se debe restar el dato menor ( $dm$ ) del dato mayor ( $DM$ ).

$$R = DM - dm$$

3)- Establecer el número de clases ( $n_i$ ) entre las cuales vamos a distribuir los datos. El número de clases se obtiene al calcular la raíz cuadrada de  $N$ . ( $N =$  Número total de datos de la muestra). El resultado debe ser redondeado a un valor exacto o número entero. Este lineamiento se emplea para distribuciones de frecuencias con menos de 125 datos. Se suelen tomar entre 5 y 20 clases.

4)- Obtener la magnitud (amplitud) del intervalo de clases. Para ello se divide el Rango ( $R$ ) entre el número de clases ( $n_i$ ). Cuando  $i$  tiene un residuo decimal, se debe redondear al mismo número de cifras decimales que contienen los datos.

$$i = \frac{R}{n_i}$$

5)- Enumerar los límites de cada intervalo de clase. Cada clase está delimitada por el *límite inferior de la clase* ( $Li$ ) y el *límite superior de la clase* ( $Ls$ ). Se determina el primer intervalo de clase y posteriormente se van disminuyendo los límites del intervalo de clase de acuerdo al valor de la magnitud establecida previamente, así hasta alcanzar el dato mayor (o menor) de la serie de datos. (El orden puede ser descendente o ascendente). Los intervalos deben ser continuos y mutuamente excluyentes.

**Nota:** Cuando se clasifican los datos, son posibles muchas combinaciones de anchos de clase, números de clases y puntos iniciales. No existe una mejor elección. Intente algunas combinaciones diferentes y aplique su buen juicio para decidir cuál utilizar. (Pagano, p.47)

- 6)- Contar los datos en bruto de los intervalos de clase adecuados.
- 7)- Sumar las cuentas de cada intervalo para obtener la frecuencia del intervalo.

**Nota:** Revisar el ejemplo en la hoja ilustrada

❖ **Frecuencia absoluta:** ( $f_i$ ) 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 ( $N$ )

$$\sum_{i=1}^N f_i = N \qquad N = f_1 + f_2 + f_3 + \dots + f_n$$

<i>ni</i>	<i>Li - Ls</i>	<i>fi</i>
1	04 - 06	4
2	07 - 09	5
3	10 - 12	9
4	13 - 15	10
5	16 - 18	5
6	19 - 21	2
	<b>N=</b>	<b>35</b>

Después de analizar la información de los resultados, podemos responder las siguientes preguntas:

- ¿Cuántos participantes presentaron la prueba? **R = 35**
- ¿Cuál es la frecuencia de participantes que obtuvieron menos de 10 puntos? **R = 9**
- Si 10 puntos es la nota mínima aprobatoria, ¿Cuántos participantes aprobaron la prueba? **R = 26**
- ¿Cuál es el intervalo en el que se obtuvo una mayor cantidad de calificaciones? **R = Intervalo 13-15 con 10 puntuaciones.**
- ¿Cuántos participantes obtuvieron entre 10 y 15 puntos? **R = 19**

❖ **Frecuencia absoluta acumulada (  $F_i$  ):** La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Indica el número de datos que caen por debajo del límite real superior de cada intervalo.

Ejemplo de frecuencia absoluta acumulada.



$ni$	$Li - Ls$	$f_i$		$F_i$
1	04 - 06	4	4 + 0	4
2	07 - 09	5	5 + 4	9
3	10 - 12	9	9 + 9	18
4	13 - 15	10	10 + 18	28
5	16 - 18	5	5 + 28	33
6	19 - 21	2	2 + 33	35
	<b>N=</b>	<b>35</b>		

**Frecuencia relativa: ( $f_r$ )** Indica la proporción del número total de datos que aparecen en cada intervalo. Es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. La frecuencia relativa se puede expresar como una proporción o como un por ciento. La suma de las frecuencias relativas es igual a 1.

$$f_r = \frac{f_i}{N} \qquad \sum f_r = 1 \qquad f_r \times 100 = \%$$

Si la frecuencia relativa del intervalo se multiplica por 100 se obtiene el por ciento correspondiente a dicho intervalo.

$ni$	$Li - Ls$	$f_i$	<b>Frecuencia Relativa (<math>f_r</math>)</b>		
			Fracción	Proporción	Porcentaje
1	04 - 06	4	4/35	<b>0,114</b>	11,4
2	07 - 09	5	5/35	<b>0,143</b>	14,3
3	10 - 12	9	9/35	<b>0,257</b>	25,7
4	13 - 15	10	10/35	<b>0,286</b>	28,6
5	16 - 18	5	5/35	<b>0,143</b>	14,3
6	19 - 21	2	2/35	<b>0,057</b>	5,7
	<b>N=</b>	<b>35</b>		<b>1</b>	<b>100</b>

Después de analizar la tabla, se pueden responder las siguientes interrogantes:

¿Qué porcentaje de estudiantes obtuvo una calificación menor o igual a 15 puntos?  $R = 80\%$

¿Qué proporción de estudiantes obtuvo una calificación mayor a 12 puntos?  $R = 0,286 + 0,143 + 0,057 = 0,486$ .

¿Cuál es el intervalo de calificaciones con el menor porcentaje de la prueba?  $R = \text{intervalo } 19-21, \text{ con el } 5,7\% \text{ de los datos.}$

¿Cuál es la proporción de las calificaciones que se ubican entre los límites 10-12?  $R = 0,257$

- ❖ **Frecuencia Relativa Acumulada: ( $F_R$ )** Es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**. Se entiende también como la suma de las frecuencias relativas de cada uno de los valores inferiores al límite superior de cada intervalo. Se puede expresar en tantos por ciento.

$n_i$	$Li - Ls$	$f_r$		$F_R$
1	04 - 06	0,114	0,114 + 0	0,114
2	07 - 09	0,143	0,143 + 0,114	0,257
3	10 - 12	0,257	0,257 + 0,257	0,514
4	13 - 15	0,286	0,286 + 0,514	0,8
5	16 - 18	0,143	0,143 + 0,8	0,943
6	19 - 21	0,057	0,057 + 0,943	1
		1		

**Límites Reales (Fronteras de Clase, Verdaderos límites de clases):** Son los valores extremos que tiene el intervalo de clase, inferior y superior, entre los cuales van a estar los valores de los datos agrupados en ese intervalo de clase. Es el resultado de restar **0,5** (si es variable discreta) ó 0,05/0,005 (si es variable continua, según sea el caso) al *Límite Inferior* de clases y luego sumar esa misma cantidad al *Límite Superior* de clases.

$$L_{Ri} = L_i - 0,5 \text{ (0,05/0,005)}$$

$$L_{Rs} = L_s + 0,5 \text{ (0,05/0,005)}$$

<i>ni</i>	<i>Li - Ls</i>	<i>LRi</i>	<i>LRs</i>
1	04 - 06	4 - 0,5 = 3,5	6 + 0,5 = 6,5
2	07 - 09	7 - 0,5 = 6,5	9 + 0,5 = 9,5
3	10 - 12	10 - 0,5 = 9,5	12 + 0,5 = 12,5
4	13 - 15	13 - 0,5 = 12,5	15 + 0,5 = 15,5
5	16 - 18	16 - 0,5 = 15,5	18 + 0,5 = 18,5
6	19 - 21	19 - 0,5 = 18,5	21 + 0,5 = 21,5

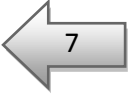
**Marca de clase (*Xi*):** Representa el valor medio de cada clase. Es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros. Se obtiene sumando los límites de clase y dividiéndolos entre dos. (Se pueden tomar los límites aparentes o los límites reales, el resultado es el mismo).

$$X_i = \frac{L_i + L_s}{2}$$

$$X_i = \frac{L_{Ri} + L_{Rs}}{2}$$

<i>ni</i>	<i>Li - Ls</i>	<i>LRi - LRs</i>	<i>Xi</i>
1	04 - 06	3,5 - 6,5	05
2	07 - 09	6,5 - 9,5	08
3	10 - 12	9,5 - 12,5	11
4	13 - 15	12,5 - 15,5	14
5	16 - 18	15,5 - 18,5	17
6	19 - 21	18,5 - 21,5	20

## Bibliografía consultada:



Howard, B. (1997). Estadística paso a paso. México: Trillas.

Johnson, R. Kuby P. (2004). *Estadística elemental. Lo esencial*. Tercera Edición. Editorial Thomson.

Pagano, R. (1999). Estadística para las ciencias del comportamiento. 5ta edición. International Thomson Editores.

Spiegel, M. (1997). *Estadística*. Segunda Edición. Editorial McGraw-Hill.

UNESR. (1985) Terminología en Estadística e Investigación. Caracas: Autor.