

## CAPÍTULO 2

## ORGANIZACIÓN DE DATOS

Siendo *el dato* el material que se debe procesar, es decir, la materia prima de la estadística, el primer paso es entonces la recolección de datos, para lo cual se emplean diferentes técnicas, como la entrevista personal, el cuestionario, la observación, etc.

El segundo paso es la organización y ordenamiento de los datos, lo que se hace a través de tablas, las cuales pueden ser *simples* o *con intervalos*, en ambos casos agrupando todos aquellos que corresponden a una mismo dato nominal o variable y expresando en una columna el número de veces que aparece esa variable.

*LA FRECUENCIA es el número de veces que aparece cada variable o dato nominal.*

Por ejemplo, se desea hacer una tabla que muestre las calificaciones en matemáticas de un grupo escolar. Se ve que hubieron dos alumnos que sacaron 10 de calificación, siete estudiantes sacaron 9, etc.; se dice entonces que la frecuencia del dato nominal 10 es de dos; la frecuencia de la variable 9 es siete, etc.

Una *distribución de frecuencias* es el resultado de organizar los datos recolectados en grupos, mostrando la frecuencia de cada uno. Esta puede ser simple o por intervalos.

## 2.1 DISTRIBUCIÓN DE FRECUENCIAS SIMPLE

Organizar los datos recolectados, ya sea de menor a mayor o viceversa, de manera que se muestre la frecuencia de cada uno de ellos, es hacer una *distribución de frecuencias simple*.

El primer paso es localizar el dato menor y el dato mayor dentro del conjunto de datos recolectados aún en desorden, en el caso que los datos sean de carácter numérico. Una vez conseguido lo anterior, en una primera columna se escriben todos los números que van desde el menor hasta el mayor, incluidos éstos. Luego, se cuenta cuántas veces aparece el primer valor nominal, para lo cual se aconseja ir marcando con una línea ( / ) cada vez que se cuente uno. El proceso debe repetirse para cada variable. Finalmente se cuentan el número de marcas que se hayan registrado para cada valor nominal y se procede a construir la tabla definitiva.

Ejemplo: Ordenar y construir una tabla de frecuencias simple del siguiente conjunto de datos recolectados.

24	20	32	32	29	21
21	22	33	30	27	26
23	24	20	25	26	32
28	22	29	29	33	35
31	28	32	35	33	32
27	21	33	29	25	24

Solución:

**Primer paso:** Se localizan los números más chico y más grande: son el 20 y el 35.

**Segundo paso:** Se hace una lista completa de números desde el 20 hasta el 35:

20	24	28	32
21	25	29	33
22	26	30	34
23	27	31	35

**Tercer paso:** Se cuenta cuántos datos nominales 20 aparecen y por cada uno que aparezca se pone una “rayita” (/). Se hace lo mismo para cada valor:

20	//	24	///	28	//	32	////
21	///	25	//	29	////	33	///
22	//	26	//	30	/	34	
23	/	27	//	31	/	35	//

A manera de comprobación, para tener la seguridad de que no se escapó alguno o no se contaron de más, la suma de todas las “rayitas” (/) debe ser igual al número de datos nominales del conjunto inicial. En este caso existen 36 datos nominales y 36 “rayitas”, lo que significa que el conteo fue correcto.

**Cuarto paso:** Se elabora la tabla definitiva:

<b>DATO NOMINAL</b>	<b>FRECUENCIA</b>
<i>x</i>	<i>f</i>
20	2
21	3
22	2
23	1
24	3
25	2
26	2
27	2
28	2
29	4
30	1
31	1
32	5

<b>DATO NOMINAL</b>	<b>FRECUENCIA</b>
<i>x</i>	<i>f</i>
33	4
34	0
35	2

36

Es conveniente y a veces necesario obtener el total de una columna en una tabla, lo cual se especifica como lo muestra la tabla anterior. Para evitar confundir con otro dato nominal, la suma de cualquier columna debe ponerse “afuera” de la tabla.

## CUESTIONARIO 2.1

1) Organizar el siguiente conjunto de datos en una tabla de frecuencias simple:

1	5	4	4	9	8	8	6
5	2	9	9	5	3	3	8
7	7	4	2	5	7	7	9

2) Organizar el siguiente conjunto de datos en una tabla de frecuencias simple:

5	2	2	1	9	8	3	3
5	2	9	4	6	6	3	8
8	1	1	2	4	2	4	3

3) Organizar el siguiente conjunto de datos en una tabla de frecuencias simple:

10	15	14	14	9	18	18	16
15	12	9	19	15	13	13	18
17	17	14	12	11	11	15	13
13	9	10	19				

4) Organizar los siguientes datos en una tabla de frecuencias simple:

21	25	22	24	28	28	28	22
25	28	29	29	25	20	23	28
27	27	24	22	22	28	25	25
25	21	23	27	27	27	27	27
22	22						

5) Organizar los siguientes datos en una tabla de frecuencias simple:

31	25	34	34	29	28	38	26
25	32	39	29	25	33	30	28
27	27	24	23	38	38	26	35
32	33	29	35	33	30	38	37
27	32	23	38	37	29	30	29
29	27	26					

## 2.2 DISTRIBUCIÓN DE FRECUENCIAS POR INTERVALOS

Los datos recolectados pueden también organizarse por intervalos. Por ejemplo, al realizar un censo en una ciudad, podría interesar cuántas personas tienen 0, 1 ó 2 hijos, cuántas 3, 4 ó 5 hijos, cuántas 6, 7 u 8 hijos, etc. O bien, en un grupo cuántas personas hay de 15 ó 16 años, cuántas de 17 ó 18 años, etc. Cada intervalo se llama también *clase*.

A la organización de los datos recolectados en tablas por intervalos se le llama *distribución de frecuencias por intervalos*. La característica más importante es que el ancho de cada clase o longitud del intervalo debe ser el mismo para cada intervalo.

**Ejemplo:** Ordenar y construir una tabla de frecuencias con cuatro intervalos del siguiente conjunto de datos recolectados.

24	20	32	32	29	21
21	22	33	30	27	26
23	24	20	25	26	32
28	22	29	29	33	35
31	28	32	35	33	32
27	21	33	29	25	24

**Solución:** Conviene iniciar de la misma manera que en la organización de frecuencias simple. Entonces se localizan los números más chico y más grande: son el 20 y el 35 y se hace una lista completa de números desde el 20 hasta el 35. Como los datos recolectados están a la vista, se trata de valores discontinuos o discretos. A continuación se cuentan cuántos datos nominales aparecen por cada uno y se pone una “rayita” (/), de lo que resulta:

20	//	24	///	28	//	32	////
21	///	25	//	29	////	33	///
22	//	26	//	30	/	34	
23	/	27	//	31	/	35	//

Después, se cuenta cuántos datos nominales existen dentro del conjunto. En este caso hay 16. Entonces, como hay 16 datos nominales y se piden cuatro intervalos, simplemente se dividen o se reparten, por lo que cada intervalo incluirá a cuatro datos nominales, como lo muestra la siguiente tabla:

INTERVALO	FRECUENCIA $f$
20 - 23	8
24 - 27	9
28 - 31	8
32 - 35	11

36

## CUESTIONARIO 2.2

1) Organizar el siguiente conjunto de datos en una tabla que contenga 3 intervalos:

1	5	4	4	9	8	8	6	5	2	9
9	5	3	3	8	7	7	4	2	6	9
5	8	3	5	7	3	4	2	1	1	1
5	6	3	2	4	8	8	4	6	4	7
1	9	9	5							

2) Organizar el siguiente conjunto de datos en una tabla que contenga 4 intervalos:

5	2	2	10	9	8	3	3	5	2	9
4	6	6	3	8	8	1	1	2	9	5
2	6	0	10	11	7	7	11	0	4	7
9	10	11	1	7	1	0	11	9	5	5
2	0	11	4							

3) Organizar el siguiente conjunto de datos en una tabla que contenga 5 intervalos:

10	15	14	14	10	18	18	16	15	12	10
19	15	13	13	18	17	17	14	12	15	18
13	11	12	10	11	11	14	19			

4) Organizar el siguiente conjunto de datos en una tabla que contenga 4 intervalos:

31	25	34	34	29	28	38	26	25	32	39
29	25	33	30	28	27	27	24	24	38	38
26	35	32	33	29	34	33	30	38	37	27
32	26	27	32	32	25	31	29	29	29	26
33	34	30	31							

5) Organizar el siguiente conjunto de datos en una tabla que contenga 5 intervalos:

21	25	22	24	28	28	28	22	25	28	29
29	25	20	23	28	27	27	24	22	22	22
29	21	21	21	28	20	20	25	27	25	23
20	28	26	25	21	28	24	22	22	21	20
27	27	23	21							

### 2.3 FRECUENCIAS ACUMULADAS

Muchas veces resulta de gran utilidad, una vez organizados los datos recolectados en una tabla de frecuencias, ya sea simple o por intervalos, tener información sobre la frecuencia que a partir del inicio de la tabla se tiene hasta cierto dato nominal determinado. A lo anterior de se le conoce con el nombre de *frecuencias acumuladas* ( $f_a$ ) y se añade en una columna en la misma tabla.

**Ejemplo 1:** En los datos del ejemplo de la página 68, sus frecuencias acumuladas son:

<b>DATO NOMINAL</b> $x$	<b>FRECUENCIA</b> $f$	<b>FRECUENCIAS ACUMULADAS</b> $f_a$
20	2	2
21	3	5
22	2	7
23	1	8
24	3	11
25	2	13
26	2	15
27	2	17
28	2	19
29	4	23
30	1	24
31	1	25
32	5	30
33	4	34
34	0	34
35	2	36

Obsérvese que la columna de las frecuencias acumuladas no se suma, pero debe coincidir el último valor acumulado con la suma de la columna de las frecuencias.

**Ejemplo 2:** En los datos del ejemplo de la página 71, sus frecuencias acumuladas son:

<b>INTERVALO</b>	<b>FRECUENCIA <i>f</i></b>	<b>FRECUENCIAS ACUMULADAS <i>fa</i></b>
20 - 23	8	8
24 - 27	9	17
28 - 31	8	25
32 - 35	11	36

36

### CUESTIONARIO 2.3

- 1) A las tablas del cuestionario 2.1, agregar una columna de frecuencias acumuladas.
- 2) A las tablas del cuestionario 2.2, agregar una columna de frecuencias acumuladas.