

Cuando sobre una población hemos realizado una encuesta o cualquier registro para conocer los valores que toman las variables, nos encontramos ante una gran cantidad de datos que debemos organizar. La mejor forma de organizar esta información es mediante tablas que llamaremos tablas de frecuencias.

Definición: la tabla de frecuencias es una ordenación, en forma de tabla, de los datos estadísticos, asignado a cada dato su frecuencia correspondiente

Frecuencia absoluta

Definición: la frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa por f_i .

Supongamos una variable estadística X , constituida por N valores, $X_1, X_2, X_3 \dots, X_N$ procedentes de la observación de una determinada característica sobre una población o muestra compuesta por N individuos. Y supondremos que toma k valores distintos que denotamos por $x_1, x_2, x_3 \dots, x_k$ (con mayúscula todos los datos de la población y con minúscula los que son distintos y forman las modalidades o valores de la variable).

La frecuencia absoluta f_i es el número de veces que se repite el valor x_i . Estas frecuencias se disponen en forma de tabla, con la siguiente estructura.

| Valores | Frecuencia absoluta |
|---------|---------------------|
| x_1 | f_1 |
| x_2 | f_2 |
| \dots | |
| | |
| x_k | f_k |
| TOTAL | $N =$ |

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

$$f_1 + f_2 + f_3 + \dots + f_k = N$$

Para indicar de manera resumida estas sumas se utiliza la letra griega Σ (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^k f_i = N$$

Ejemplo para una variable discreta

A 100 alumnos de un colegio se les pregunta la edad, obteniendo los siguientes valores:

12 13 12 12 13 14 13 13 13 12 13 14 13 15 14 13 13 13 14 14
 14 15 12 15 14 15 15 16 14 16 12 14 14 14 18 15 16 16 13 15
 16 14 15 17 15 16 18 16 16 16 12 14 13 13 16 13 12 13 13 13
 14 15 15 13 14 17 17 13 14 14 14 14 17 15 13 14 13 14 15 17
 13 14 13 14 14 14 16 17 14 14 15 15 18 13 16 15 13 12 17 17

Vamos a indicar en la columna “valores” los valores distintos que la variable, y en la columna “Frecuencia absoluta” el número de que se repite cada uno de ellos.

| Valores | Frecuencia absoluta |
|---------|---------------------|
| 12 | 9 |
| 13 | 25 |
| 14 | 27 |
| 15 | 16 |
| 16 | 12 |
| 17 | 8 |
| 18 | 3 |

Frecuencia relativa

Definición: la frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se representa por fr_i , aunque algunos autores la representan con h_i o n_i

La expresión para calcular la frecuencia relativa es:

$$fr_i = \frac{f_i}{N}$$

La suma de las frecuencias relativas es igual a 1 .

$$\sum_{i=1}^n fr_i = 1$$

La frecuencia relativa también se puede expresar en forma de porcentajes:

$$p_i = 100 \cdot fr_i$$

Frecuencia acumulada

Definición: la frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por F_i .

Así, la frecuencia acumulada viene dada por la expresión:

$$F_k = \sum_{i=1}^k f_i$$

$F_1 = f_1$ $F_2 = f_1 + f_2$ y sucesivamente hasta llegar a $F_k = f_1 + f_2 + \dots + f_k$

Por tanto, la frecuencia acumulada del último dato, será siempre igual al número total de datos, $F_n = N$

Si completamos la tabla anterior con todas estas frecuencias, la tabla quedaría así:

| Valores | Frecuencia absoluta | Frecuencia relativa | Porcentajes | Frecuencias acumuladas |
|----------|------------------------|-------------------------|--------------------------|--------------------------|
| x_1 | f_1 | $fr_1 = \frac{f_1}{N}$ | $p_1 = 100 \cdot fr_1$ | $F_1 = f_1$ |
| x_2 | f_2 | $fr_2 = \frac{f_2}{N}$ | $p_2 = 100 \cdot fr_2$ | $F_2 = f_1 + f_2$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| x_k | f_k | $fr_k = \frac{f_k}{N}$ | $p_k = 100 \cdot fr_k$ | $F_k = \sum_{i=1}^k f_i$ |
| TOTAL | $N = \sum_{i=1}^k f_i$ | $\sum_{i=1}^k fr_i = 1$ | $\sum_{i=1}^k p_i = 100$ | |

Ejemplo para una variable discreta

Completamos la tabla de frecuencias del ejemplo anterior correspondiente al recuento de las edades de los 100 alumnos de un centro con los valores de las frecuencias relativas, porcentuales y acumuladas.

Vamos a hacer el recuento de los datos y presentarlos en una tabla de frecuencias:

| Valores | Frecuencia absoluta | Frecuencia relativa | Porcentajes | Frecuencias acumuladas |
|---------|---------------------|---------------------|-------------|------------------------|
| 12 | 9 | $9/100 = 0,09$ | 9 % | 9 |
| 13 | 25 | $25/100 = 0,25$ | 25 % | 34 |
| 14 | 27 | $27/100 = 0,27$ | 27 % | 61 |
| 15 | 16 | $16/100 = 0,16$ | 16 % | 77 |
| 16 | 12 | $12/100 = 0,12$ | 12 % | 89 |
| 17 | 8 | $8/100 = 0,08$ | 8 % | 97 |
| 18 | 3 | $3/100 = 0,03$ | 3 % | 100 |
| TOTAL | N = 100 | 1 | 100 | |

Agrupación en intervalos.

Cuando tenemos una variable que presenta una gran cantidad de datos agrupamos los valores en intervalos para realizar el recuento más fácilmente, convirtiendo así la variable en una variable continua. Los valores se agrupan usualmente en intervalos de la forma $(a, b]$.

Para establecer el número adecuado de intervalos hay varios métodos entre los que destacan: la fórmula de Sturges y la raíz del número de datos.

- Número de intervalos $k = \sqrt{N}$, donde N es el número total de datos.
- Fórmula de Sturges: número de intervalos $k = 1 + 3,3 \cdot \log N$, donde N es el número total de datos.

Para poca cantidad de datos, aproximadamente menos de 50, la raíz cuadrada es sencilla de calcular haciendo uso de la calculadora, pero cuando este valor es grande el número de intervalos también aumenta.

Cuando ya hemos determinado el número de intervalos los construimos. Generalmente los intervalos serán de la forma $(a_{i-1}, a_i]$ y, para construir la tabla de frecuencias, a cada uno de ellos se le asocia un valor representativo, denominado marca de clase, que se denota x_i , y que usualmente es el punto medio del intervalo, es decir:

$$x_i = \frac{a_{i-1} + a_i}{2}$$

Ejemplo para una variable continua

A los 100 alumnos del colegio anterior también se les pregunta el peso, obteniendo las siguientes respuesta en 96 casos (4 no respondieron):

| | | | | | | | | | |
|------|------|------|------|------|------|------|-------|------|------|
| 34,5 | 35,2 | 36,1 | 37,0 | 37,9 | 38,5 | 38,5 | 39,1 | 39,6 | 40,0 |
| 40,4 | 40,4 | 40,5 | 40,8 | 40,9 | 41,1 | 45,0 | 45,2 | 46,0 | 47,3 |
| 47,7 | 47,8 | 48,0 | 48,2 | 48,3 | 48,3 | 48,7 | 49,0 | 49,1 | 49,1 |
| 49,2 | 50,3 | 50,5 | 50,5 | 50,6 | 50,9 | 52,3 | 52,8 | 52,9 | 53,0 |
| 53,3 | 53,5 | 54,0 | 54,2 | 54,9 | 55,1 | 55,3 | 55,3 | | |
| 55,4 | 55,6 | 55,8 | 55,8 | 55,8 | 56,0 | 56,2 | 56,4 | 57,4 | 58,1 |
| 58,0 | 58,9 | 58,9 | 59,0 | 59,3 | 59,3 | 60,1 | 60,4 | 60,5 | 60,5 |
| 60,7 | 62,5 | 62,7 | 63,0 | 63,1 | 63,2 | 63,8 | 64,6 | 65,0 | 65,0 |
| 65,0 | 65,5 | 65,6 | 65,7 | 65,8 | 68,2 | 68,4 | 69,6 | 70,1 | 70,3 |
| 72,5 | 72,5 | 73,0 | 79,0 | 80,4 | 80,7 | 85,8 | 108,4 | | |

| Intervalos | Marca de clase | Frecuencia absoluta | Frecuencia relativa | Porcentajes | Frecuencias acumuladas |
|------------|----------------|---------------------|---------------------|-------------|------------------------|
| (30, 40] | 35 | 9 | $9/96 = 0,09$ | 9 % | 9 |
| (40, 50] | 45 | 22 | $22/96 = 0,23$ | 23 % | 31 |
| (50, 60] | 55 | 33 | $33/96 = 0,34$ | 34 % | 64 |
| (60, 70] | 65 | 22 | $22/96 = 0,23$ | 23 % | 86 |
| (70, 80] | 75 | 6 | $6/96 = 0,06$ | 6 % | 92 |
| (80, 90] | 85 | 3 | $3/96 = 0,03$ | 3 % | 95 |
| (90, 100] | 95 | 0 | $0/96 = 0$ | 0 % | 95 |
| (100, 110] | 105 | 1 | $1/96 = 0,01$ | 1 % | 96 |
| TOTAL | | N = 96 | 1 | 100 % | |

Vamos a calcular el número de intervalos aplicando la fórmula de Sturges: $k = 1 + 3,3 \cdot \log 96 = 7,541 \approx 8$ intervalos mayor valor-menor valor = $108 - 33 = 75$. $75/8 = 9,38 \approx 10$ amplitud de los intervalos.