

TABLAS PARA VARIABLES CUANTITATIVAS

La **organización de datos** depende del tipo de variable estadística a ser analizada. Considere las siguientes definiciones.

Clase

Una “clase” o categoría es uno de los posibles atributos de una variable cualitativa a la que una UA puede pertenecer. Las clases deben ser excluyentes o disjuntas, ya que de esta forma no existe ambigüedad en la clasificación de las UAs.

Marca e intervalo de clase

Una “clase” se denomina “marca de clase” cuando la variable es discreta. En el caso de variables continuas, la “clase” se denomina “intervalo de clase” y el punto medio del intervalo se llama “marca de clase”.

TABLAS DE FRECUENCIA

Se desea estudiar una variable agrupada en k clases excluyentes, c_1, c_2, \dots, c_k , para variables cualitativas, o k marcas de clases excluyentes, x_1, x_2, \dots, x_k , para variables cuantitativas.

Frecuencia absoluta

Corresponde al número de UAs que pertenecen a la clase c_i y se denota por n_i , para $i = 1, \dots, k$, donde

$$\sum_{i=1}^k n_i = n.$$

Frecuencia acumulada

Corresponde al número acumulado de UAs que pertenecen a las clases c_1, c_2, \dots, c_i y está dada por $N_i = \sum_{j=1}^i n_j$, para $i = 1, \dots, k$. Así,

$$N_1 = n_1 \text{ y } N_k = \sum_{i=1}^k n_i = n.$$

Frecuencia relativa

Corresponde a la proporción de UAs que pertenecen a la clase c_i y está dada por $f_i = n_i/n$, para $i = 1, \dots, k$, donde

$$\sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^k n_i = 1.$$

Si multiplicamos la frecuencia relativa por 100, obtenemos la frecuencia porcentual.

Frecuencia relativa acumulada

Corresponde a la proporción acumulada de UAs que pertenecen a las clases c_1, c_2, \dots, c_i y está dada por $F_i = \sum_{j=1}^i f_j$, para $i = 1, \dots, k$. Así,

$$F_1 = f_1 \text{ y } F_k = \sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^k n_i = 1.$$

TABLAS PARA VARIABLES CUANTITATIVAS

Si la variable en estudio es cuantitativa continua (o discreta con alta variabilidad), las clases de la tabla anterior se reemplazan por “intervalos de clases” tal como en la tabla de abajo. En la siguiente página se detalla un método para calcular el número de intervalos de clase.

Intervalos de clase	Marca de clase (x_i)	Frec. abs. (n_i)	Frec. rel. (f_i)	Frec. rel. acum. (F_i)
$[\text{mín}\{x_i\}; \text{mín}\{x_i\} + a[$	x_1	n_1	f_1	F_1
$[\text{mín}\{x_i\} + a; \text{mín}\{x_i\} + 2a[$	x_2	n_2	f_2	F_2
\vdots	\vdots	\vdots	\vdots	\vdots
$[\text{mín}\{x_i\} + (k - 1)a; \text{máx}\{x_i\}]$	x_k	n_k	f_k	$F_k = 1$
Total		n	1	

Donde la i -ésima marca de clase (x_i) corresponde al promedio de los límites del i -ésimo intervalo, para $i = 1, \dots, k$.

Algoritmo para calcular intervalos de clase:

Paso 1: contar el número n de datos.

Paso 2: calcular el rango $r = \text{máx}\{x_i\} - \text{mín}\{x_i\}$, donde $\text{mín}\{x_i\}$ y $\text{máx}\{x_i\}$ corresponden al mínimo y al máximo de los datos.

Paso 3: escoger el número k de intervalos de clase tomando la parte entera de $k = 1 + 3,3 \log_{10}(n)$, donde $\log_{10}(\cdot)$ es el logaritmo en base 10 (regla de Sturges). También, el investigador puede elegir k usando un criterio propio adecuado.

Paso 4: calcular la amplitud de los intervalos como $a = r/k$.

Paso 5: tomar como límite inferior el valor $\text{mín}\{x_i\}$ y como límite superior el valor $\text{mín}\{x_i\} + a$ para determinar los extremos del primer intervalo de clase.

Paso 6: sumar a al límite inferior para obtener las restantes intervalos de clases, donde el límite inferior de las sucesivas clases corresponderá al límite superior de la clase anterior.

EJEMPLO: COEFICIENTE INTELECTUAL (CI)

Ejemplo

Se realizó un test de conocimientos a 30 estudiantes pertenecientes a una misma carrera de la UCM. Los puntajes del test son los siguientes.

8,70	9,20	9,30	9,60	9,90	10,10	10,20	10,30	10,40	10,40
10,50	10,90	11,40	11,40	11,50	11,60	11,80	11,90	12,30	12,30
12,40	12,70	12,80	13,00	13,10	13,60	13,80	14,50	14,70	15,80

- Identifique la UA, la variable, su tipo y escala de medición.
- Construya una tabla estadística que incluya intervalos de clase, frecuencias absoluta (n_i), relativa (f_i), relativa porcentual (h_i), relativa acumulada (F_i) y relativa acumulada porcentual (H_i).
- Interprete n_3 , h_4 y H_2 .

SOLUCIÓN DEL EJEMPLO ANTERIOR

- El mínimo y el máximo están dados por $\min\{x_i\} = 8,70$ y $\max\{x_i\} = 15,80$.
- El rango es $r = 7,1$.
- El número de intervalos de clase es $k = 5$, pues $k = 1 + 3,3 \log_{10}(30) = 5,87$.
- Por lo tanto, la amplitud de los intervalos será $a = (7,1)/5 = 1,42$. Entonces el primer intervalo parte desde $8,70$ hasta $8,70 + 1,42 = 10,12$ (sin considerar este valor); el siguiente intervalo parte desde $10,12$ hasta $10,12 + 1,42 = 11,54$ (sin considerar este valor); y así sucesivamente.
- El último intervalo va desde $14,38$ hasta $15,80$ (considerando ambos valores).

SOLUCIÓN DEL EJEMPLO ANTERIOR

La tabla de frecuencias es la siguiente.

Intervalo de clase	Marca de clase (x_i)	Frecuencia (n_i)	Porcentaje (h_i)	% acumulado (H_i)
[8, 70; 10, 12[9,41	6	20,0	20,0
[10, 12; 11, 54[10,83	9	30,0	50,0
[11, 54; 12, 96[12,25	8	26,7	76,7
[12, 96; 14, 38[13,67	4	13,3	90,0
[14, 38; 15, 80]	15,09	3	10,0	100,0
Total		30	100	

¿Cómo se interpretan los elementos de esta tabla?

Ejemplo

Sobre un grupo de $n = 21$ personas se realizan las siguientes observaciones de sus pesos, medidos en kilogramos:

$X \sim x_1, x_2, \dots, x_{21}$						
58	42	51	54	40	39	49
56	58	57	59	63	58	66
70	72	71	69	70	68	64

Agrupar los datos en una tabla estadística.

Ejemplo

Calcular los datos que faltan en la siguiente tabla:

$l_{i-1} - l_i$	n_i	f_i	N_i
0 - 10	60	f_1	60
10 - 20	n_2	0,4	N_2
20 - 30	30	f_3	170
30 - 100	n_4	0,1	N_4
100 - 200	n_5	f_5	200
	n		