

## Tema 4 : Tabulación de datos

La tabulación consiste en presentar los datos estadísticos en forma de tablas o cuadros.

### --Partes de una tabla

- TITULO de la tabla, que debe ser preciso y conciso
- CONTENIDO, con
  - la *fila de encabezamiento o cabecera* (títulos de las columnas)
  - la *columna matriz*, con las modalidades o clases de la variable
  - *columnas de parámetros*
- NOTAS EXPLICATIVAS (opcional), como fuente de los datos, abreviaturas, etc.

### --Forma de tabular

#### VARIABLES CUALITATIVAS

pueden representarse :

- la frecuencia absoluta (símbolo : **f** ó **n**), que es el nº de veces que aparece cada modalidad (resultado del recuento). La frecuencia total, de todas las modalidades juntas, se representa por **N**.
- la frecuencia relativa (**fr**) o proporción se obtiene dividiendo la frecuencia de cada modalidad entre el total de datos.  $fr = f / N$  . Los valores posibles oscilan entre 0 y 1. Suele expresarse con 3 decimales. La suma de todas las fr tiene que dar 1 ó un número muy cercano al 1, si ha habido redondeos.
- el porcentaje (**P** o **%**), que es la frecuencia relativa multiplicada por 100.  $P = fr * 100$  ó  $\% = (f*100)/N$  . Suele expresarse con 3 dígitos. La suma de todos los porcentajes debe dar 100 o un número muy próximo, si ha habido redondeos.
- las frecuencia acumuladas ( **$\Sigma f$**  ó  **$\Sigma n$**  ) que se obtienen sumando la frecuencia de cada modalidad a las frecuencias ya acumuladas anteriormente. En la primera modalidad no hay nada acumulado de antes y por tanto su frecuencia acumulada será su misma frecuencia. La última modalidad tiene que dar una frecuencia acumulada igual a N.
- las frecuencias relativas acumuladas y los porcentajes acumulados se obtienen de forma similar
- En las variables nominales las modalidades pueden ponerse en el orden que se quiera, pero en las ordinales hay que respetar el orden lógico.

Ejemplo:

Residencia Sanitaria S. S. de Castellón Ingresos en Pediatría. Marzo 1980						
Sección	f	fr	%	$\Sigma f$	$\Sigma fr$	$\Sigma \%$
Neonatología	25	0,125	12,5	25	0,125	12,5
Lactantes	95	0,475	47,5	120	0,6	60
Preescolares	80	0,400	40	200	1	100
Total	200	1	100			

En la tabla definitiva no se presentan todos estos parámetros, sino los más adecuados en cada caso concreto. Casi siempre **f** y/o **%** . Sólo el porcentaje, sin que conste **N**, no es correcto. En este ejemplo bastaría con **f** y **%** .

## VARIABLES CUANTITATIVAS

Los datos se agrupan según la frecuencia de los valores. Es lo que se denomina *Distribución de frecuencias*. La forma de tabular depende del nº de datos.

----Si son pocos (la mayoría de autores pone el tope en 30) , se hace una tabla simple de forma similar a lo visto para las variables CL. Cada dato equivale a una modalidad. Al final nos quedaremos con la f de cada número y si se prefiere también con el %. Los números se ordenan de menor a mayor o de mayor a menor. La tabla puede hacerse en sentido vertical u horizontal.

Ejemplo: Si  $x = (4, 1, 7, 2, 2, 9, 7, 2, 2, 9, 7, 1, 4)$

x	1	2	4	7	9
f	2	4	2	3	2

o bien

x	f
1	2
2	4
4	2
7	3
9	2

----Si son muchos se agrupan en clases, que son intervalos sucesivos de valores. Los datos se asignan a la clase que les corresponde y se cuentan los datos de cada clase, que está representada por el punto medio o centro de clase (pm ó c).

Esta agrupación es **arbitraria** con dos condiciones esenciales: que las clases sean mutuamente excluyentes y que todos los datos puedan ser asignados a una clase. Ahora bien, la experiencia ha ido introduciendo una serie de normas, que permiten hacer esta agrupación de la forma más racional posible.

Yo recomendaría los siguientes pasos:

- 1) calcular el **RECORRIDO ( R )**, (a veces mal llamado Rango)  
= (límite real superior del dato mayor – límite real inferior del dato menor)  
O si se prefiere: = (valor tabulado máximo – valor tabulado mínimo) + 1
- 2) calcular el **Nº DE CLASES (NC)**.

Es función de N (tamaño de la muestra) y no hay reglas fijas.

En general: “entre 4 y 20” .

Ayudas:  $NC = 1 + 3,32 \cdot \log N$  ó  $1 + 1,44 \cdot \ln N$

O la siguiente tabla: N 8 16 32 64 128 256 etc.

-----  
NC 4 5 6 7 8 9 etc.

De entrada nos quedamos con 2 ó 3 opciones

- 3) calcular la **AMPLITUD** de las clases ó INTERVALO (i) :  $i = R / NC$   
Si **i** no es número entero, se redondea al número entero superior para que  $NC \cdot i \geq R$  y así queden englobados todos los datos  
Como probamos con 2 ó 3 opciones, conviene elegir una i que sea impar, pues así el punto medio de la clase (pm ó c) tendrá una cifra menos.

**En principio todas las clases deben tener la misma amplitud.**

- 4) Ver si hay **SOBRAS**, que son la diferencia entre  $NC \cdot i$  y R. Se reparten lo mejor posible entre ambos extremos de la distribución fijando así los límites definitivos de la tabla.

5) Construir el esquema de la tabla, poniendo **columnas** de

- CLASES ó LIMITES TABULADOS
- LIMITES REALES
- PUNTO MEDIO (pm ó c)
- FRECUENCIA ( f ó n)
- FRECUENCIA RELATIVA ( fr)
- PORCENTAJE ( P ó %)
- FRECUENCIAS ACUMULADAS (  $\Sigma f$  ó  $\Sigma n$ )
- FRECUENCIAS RELATIVAS ACUMULADAS (  $\Sigma fr$ )
- PORCENTAJES ACUMULADOS (  $\Sigma \%$ )

6) Hacer el **RECuento** de datos y rellenar las casillas correspondientes

7) Escribir la **TABLA DEFINITIVA**. Son obligadas las clases y la frecuencia absoluta, pudiendo añadir otros parámetros, si se considera que mejoran la información. Una tabla excesivamente prolija resulta más difícil de leer. Por tanto la norma es: poner todo lo necesario, pero no más de lo necesario.

*Es recomendable probar con al menos 2 tablas y elegir la que quede mejor.*

Algunos de éstos parámetros son los mismos que se han visto para las variables CL. Otros precisan una aclaración:

Los **límites de las clases** son los valores inferior y superior de cada clase. (Límite inferior y límite superior). Hay que distinguir entre los **límites tabulados (LT)** y los **límites reales (LR)**. Los límites tabulados son los datos originales que abren y cierran una clase. Los límites reales son el límite real inferior del primer valor (LRI) y el límite real superior del último (LRS).

El **punto medio o centro** de la clase (pm ó c) representa a la clase cuando se hacen operaciones matemáticas. Es la media de los límites. Da lo mismo tomar los límites reales que los tabulados, ya que ambos dan el mismo resultado.

En una distribución con todas las clases de la misma amplitud las diferencias entre los puntos medios, los límites inferiores y los límites superiores de dos clases consecutivas valen lo mismo y son igual a la amplitud de la clase (i). Esto facilita la construcción de la tabla.

Una **clase** es **abierta** cuando carece de un límite. Sólo pueden ser abiertas la primera clase (p.e.  $<10$  ; no tiene límite inferior) y la última (p.e.  $>100$  ; no tiene límite superior). No deben usarse, a no ser que no haya otro remedio.

## EJEMPLO:

Tabular los 70 valores siguientes:

### DATOS ORIGINALES (N = 70)

40 55 19 51 62 15 20 44 60 60 45 15 21 31 13 44 41 43 51 35 50 33 25 16 61  
14 14 59 59 59 20 23 25 29 29 59 58 54 50 49 39 27 37 23 24 58 27 28 57 32  
32 34 57 56 35 35 54 36 43 46 52 50 49 42 43 46 40 39 31 48

### PASOS DE LA TABULACION

-dato mayor: 62, cuyo LRS es 62,5

-dato menor: 13, cuyo LRI es 12,5

-recorrido (R):  $62,5 - 12,5 = 50$  ó  $(62 - 13) + 1 = 50$

-nº de clases (NC): 7 u 8

-amplitud (i):

-si  $NC = 7$ ,  $i = 50/7 = 7,1 \rightarrow 8$  (par)

-si  $NC = 8$ ,  $i = 50/8 = 6,2 \rightarrow 7$  (impar)

-nos quedamos pues con  $NC = 8$  de amplitud 7, que es impar

-sobras:  $(8 \cdot 7) - 50 = 6$ , que repartimos así: 3 abajo y 3 arriba

la 1ª clase empezará en 10 (13-3)

la última terminará con el 65 (62+3)

--ya se puede construir el esquema de la tabla (clases, LR y punto medio) y proceder al recuento de los datos que corresponden a cada clase, para completar las otras columnas

Clases (Límites tabulados)	Límites reales	punto medio c	f	fr	%	$\Sigma f$	$\Sigma fr$	$\Sigma \%$
10 – 16	9,5 – 16,5	13	6	0,09	8,57	6	0,09	8,57
17 – 23	16,5 – 23,5	20	6	0,09	8,57	12	0,17	17,1
24 – 30	23,5 – 30,5	27	8	0,11	11,4	20	0,29	28,6
31 – 37	30,5 – 37,5	34	11	0,16	15,7	31	0,44	44,3
38 – 44	37,5 – 44,5	41	11	0,16	15,7	42	0,60	60,0
45 – 51	44,5 – 51,5	48	11	0,16	15,7	53	0,76	75,7
52 – 58	51,5 – 58,5	55	9	0,13	12,9	62	0,89	88,6
59 - 65	58,5 – 65,5	62	8	0,11	11,4	70	1,00	100
Suma			70	1,01	99,94			

\*\*\*Esta no es la única tabla posible, aunque probablemente sea la mejor.

Podríamos hacerla con 7 clases de amplitud 8; sobras: 6 . Clases: 10 – 17 ; 18 – 25; ...; 58 - 65

O bien 6 clases de amplitud 9. Sobras: 4 . Clases: 11- 19 ; 20 – 28; ...; 56 - 64

o bien 10 clases de amplitud 5 . Sin sobras. Clases: 13 –22 ; 23 – 32 ; .... ; 53 - 62

\*\*\*En la tabla definitiva no suelen ponerse los LR. Las clases y la frecuencia están prácticamente siempre. Según la naturaleza de la variable puede ser conveniente añadir algún otro parámetro, que contribuya a una información mejor y más clara.

\*\*\* En la página siguiente puede verse la tabla y el gráfico que elabora automáticamente mi programa de Excel, Exceltabla.xls, a partir de los 70 datos del ejemplo anterior, introducidos en la columna A.

Tabla e histograma del ejemplo de la página 4-4 que hace “Exceltabla”

Lim.Tab.Inf.	LimTab.Sup.	pm	f	%	$\Sigma f$	$\Sigma \%$
10	16	13	6	8,6	6	8,6
17	23	20	6	8,6	12	17,1
24	30	27	8	11,4	20	28,6
31	37	34	11	15,7	31	44,3
38	44	41	11	15,7	42	60,0
45	51	48	11	15,7	53	75,7
52	58	55	9	12,9	62	88,6
59	65	62	8	11,4	70	100,0
			0			

<b>Datos origin.:</b>	SESGO	-0,196	MODA	59,00
	CURTOSIS	-1,105	p3	14,07
	MEDIA GEO	36,53	p10	19,90
	MEDIANA	40,50	p25	28,25
	MEDIA	39,59	p75	51,00
	DS	14,41	p90	59,00
	VARIANZA	207,58	p97	60,00

