

La *grandària de la mostra*, que es simbolitza amb una N , és de 25. Coincideix amb la suma del recompte de les dades.

El recompte dels valors es diu *freqüència absoluta* de la variable i es simbolitza amb f_A , o simplement f .

Podem ampliar aquesta taula, trobant altres columnes:

- Freqüències relatives: és igual a la freqüència absoluta dividit pel total d'objectes. Es simbolitza amb f_R .

$$f_R = f_A/N$$

- Percentatge d'aparició: és igual a la freqüència relativa entre el total de dades per 100.

$$\% = f_R \cdot 100$$

- Freqüències absolutes acumulades, F_A , igual a la suma de les freqüències absolutes fins a un valor donat.

Si aplicam això a les nostres dades de llibres, obtenim:

Nombre de llibres	f_A	f_R	%	F_A
1	4	$4/25 = 0,16$	16%	4
2	7	$7/25 = 0,28$	28%	$4 + 7 = 11$
3	9	$9/25 = 0,36$	36%	$4 + 7 + 9 = 20$
4	5	$5/25 = 0,20$	20%	$4 + 7 + 9 + 5 = 24$
	24	1,00	100 %	

Table 1 Taula de freqüències *expandida*

CAS DISCRET AMB *moltes* DADES

Suposem ara que tenim una variable estadística discreta però amb una gamma de valors molt gran, és a dir, en principi amb un elevat nombre de valors diferents. Per exemple, si feim una enquesta on es demana l'edat, aleshores podem tractar aquesta variable estadística com a discreta (els enquestats repondran que tenen, per exemple, 39 anys, però no 39,5). El nombre de valors possibles de l'edat és molt gran: va de 0 (recent nats) fins a una fita superior hipotètica de 130 anys¹. Per tant, si féssim una taula de freqüències, podríem tenir fins a 130 valors

¹ La persona més vella del món és Jeanne Calment amb 122 anys, encara que aquest record no està verificat. El rècord d'edat verificat és de Misao Okawa, de 116 anys.

diferents. Això fa les dades immanejables. Per aquest motiu s'agrupen les dades en intervals.

En aquest exemple, podríem agrupar les dades en intervals d'amplitud 10, representant les dècades.² D'aquesta manera, obtenim 13 intervals: $[0, 10)$, $[10, 20)$, $[20, 30)$, $[30, 40)$, $[40, 50)$, $[50, 60)$, $[60, 70)$, $[70, 80)$, $[80, 90)$, $[90, 100)$, $[100, 110)$, $[110, 120)$, $[120, 130]$. En cada interval, el parèntesi significa que no prenem l'extrem i el claudàtor significa que l'extrem està inclòs. Així l'interval $[40, 50)$ representa a totes les persones de 40 anys fins a 50, incloses les de 40 anys però excloses les de 50.

D'aquesta manera, tendrem una taula de freqüències, com la següent, que en comptes de valors tendrà intervals:

Edat	Recompte
$[0, 10)$	4.859
$[10, 20)$	4.391
$[20, 30)$	5.337
$[30, 40)$	7.750
$[40, 50)$	7.517
$[50, 60)$	6.079
$[60, 70)$	4.771
$[70, 80)$	3.440
$[80, 90)$	2.160
$[90, 100)$	379
$[100, 110)$	12
$[110, 120)$	0
$[120, 130]$	0
	46.695

Table 2 Taula de freqüències³

Noteu que el darrer interval inclou 130 perquè hem de poder prendre el màxim teòric a qualque interval.

A partir d'aquí, el càlcul de la freqüència absoluta acumulada i la freqüència relativa es realitza de la mateixa manera que en el cas anterior de la variable discreta (amb poques dades).

Edat	f	f_r	F_A
[0, 10)	4.859	0,10	b
[10, 20)	4.391	0,09	b
[20, 30)	5.337	0,11	b
[30, 40)	7.750	0,17	b
[40, 50)	7.517	0,16	b
[50, 60)	6.079	0,13	b
[60, 70)	4.771	0,10	b
[70, 80)	3.440	0,07	b
[80, 90)	2.160	0,05	b
[90, 100)	379	0,01	b
[100, 110)	12	0,00	b
[110, 120)	0	0,00	b
[120, 130]	0	0,00	b
	46.695	0,99	

Table 3 Taula de freqüències ampliada

CAS CONTINU

TODO

Exemple 1. S'agrupen les dades per intervals iguals. Almenys hem de tenir 6 intervals. Tot l'altre és igual

Exercici 1. Demanem a 20 persones sobre el nombre de vegades que havien viatjat a l'estranger i el resultat ha estat aquest (figura 4):

3	5	4	4	2	3	3	3	5	2
6	1	2	3	3	6	5	4	4	3

Table 4 Viatges a l'estranger

- De quin tipus de variable es tracte?
- Quin és el tamany de la població? I de la mostra?
- Organitza les dades. Fes-ne un recompte

² El nombre d'intervals a collir... TODO.

³ Les dades són una extrapolació a partir de les dades de l' [INE](#) de l'any 2013.