

Estadística Básica: Distribuciones Unidimensionales

Prof. Carlos M. Abrisqueta Valcárcel

Curso 2024/2025

Curso de Especialización en Big Data e Inteligencia Artificial
Programación IA

Tema 2: Distribuciones de frecuencias unidimensionales

1. Tabulación de datos. Distribuciones de frecuencias unidimensionales
2. Representaciones gráficas
3. Medidas de posición
4. Medidas de dispersión: absolutas y relativas
5. Medidas de forma: asimetría y curtosis
6. Medidas de concentración: Índice de Gini y Curva de Lorenz

1. Tabulación de datos. Distribuciones de frecuencias unidimensionales

Se dispone de N observaciones de la variable X que toma los valores $\{x_1, x_2, \dots, x_r\}$, valores ordenados de menor a mayor.

x_i	n_i	f_i	N_i	F_i
x_1	n_1	f_1	N_1	F_1
x_2	n_2	f_2	N_2	F_2
\vdots	\vdots	\vdots	\vdots	\vdots
x_r	n_r	f_r	N_r	F_r
	N	1		

Frecuencia absoluta: n_i

Frecuencia relativa: $f_i = \frac{n_i}{N}$

Frecuencia absoluta acumulada: $N_i = \sum_{j=1}^i n_j = n_1 + n_2 + \dots + n_i = N_{i-1} + n_i$

Frecuencia relativa acumulada: $F_i = \sum_{j=1}^i f_j = f_1 + f_2 + \dots + f_i = F_{i-1} + f_i$

Distribución de frecuencias: $\{x_i, n_i\}_{i=1,2,\dots,r}$.

El objetivo de la tabulación es resumir la información.

2. Tabulación de datos cualitativos (modalidades no ordenables)

Ejemplo:

X = “Marca de vaqueros preferida”, $N = 30$

A C A D A
 B D A C B
 A C B B B
 C A A C C
 A A B C D
 D A C A A

x_i	n_i	f_i	p_i
A	12	0,4000	40 %
B	6	0,2000	20 %
C	8	0,2667	26,67 %
D	4	0,1333	13,33 %
	30	1	100 %

3. Tabulación de datos cualitativos (modalidades ordenables)

Ejemplo:

X = “Calificación en Introducción a la Estadística”, $N = 24$

suspenso *aprobado* *aprobado* *suspenso*
notable *aprobado* *notable* *aprobado*
suspenso *aprobado* *sobresaliente* *notable*
suspenso *aprobado* *aprobado* *suspenso*
sobresaliente *suspenso* *aprobado* *suspenso*
suspenso *suspenso* *suspenso* *notable*

x_i	n_i	f_i	p_i	N_i	F_i	P_i
<i>Suspenso</i>	10	0,4167	41,67 %	10	0,4167	41,67 %
<i>Aprobado</i>	8	0,3333	33,33 %	18	0,75	75 %
<i>Notable</i>	4	0,1667	16,67 %	22	0,9167	91,67 %
<i>Sobresaliente</i>	2	0,0833	8,33 %	24	1	100 %
	24	1	100 %			

4. Tabulación de datos cuantitativos (no agrupando en intervalos)

Ejemplo:

X = “Número de hijos en las familias de los alumnos”, $N = 25$

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

x_i	n_i	f_i	p_i	N_i	F_i	P_i
1	2	0,08	8 %	2	0,08	8 %
2	6	0,24	24 %	8	0,32	32 %
3	10	0,40	40 %	18	0,72	72 %
4	6	0,24	24 %	24	0,96	96 %
5	1	0,04	4 %	25	1	100 %
	25	1	100 %			

5. Tabulación de datos cuantitativos (agrupando en intervalos)

Ejemplo:

$X = \text{"Peso en kg."}$

68,2 87,8 85 57,5 68,2 73 71,5 77,5 94
 75,2 77,5 78,3 81,5 64 88,5 59,6 73
 62,5 85,9 83,6 78,1 61,2 85,2 78,3 61,5

<i>Peso</i>	n_i
57,5	1
59,6	1
61,2	1
61,5	1
62,5	1
64	1
68,2	2
71,5	1
73	2
75,2	1
77,5	2
78,1	1
78,3	2
81,5	1
83,6	1
85	1
85,2	1
85,9	1
87,8	1
88,5	1
94	1
	25

Criterios para agrupar en intervalos:

- En cuanto al número de intervalos, se considera adecuado hacer entre 5 y 15 intervalos, en general, aproximadamente \sqrt{N}
- En cuanto a la amplitud, constante o no, dependerá de las necesidades de la persona que solicita la información o de la naturaleza de la propia variable.
- El primer intervalo no necesariamente debe tener como extremo inferior el menor valor de la variable, ni el último intervalo como extremo superior el mayor valor de la variable.

$(L_{i-1}, L_i]$	n_i	x_i	f_i	p_i	N_i	F_i	P_i	a_i	d_i
[55, 65]	6	60	0,24	24 %	6	0,24	24 %	10	0,6
(65, 75]	5	70	0,2	20 %	11	0,44	44 %	10	0,5
(75, 85]	9	80	0,36	36 %	20	0,80	80 %	10	0,9
(85, 95]	5	90	0,20	20 %	25	1,00	100 %	10	0,5
	25		1	100 %					

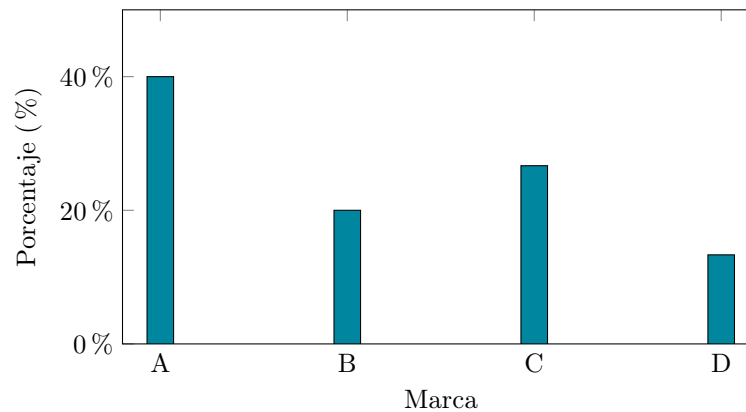
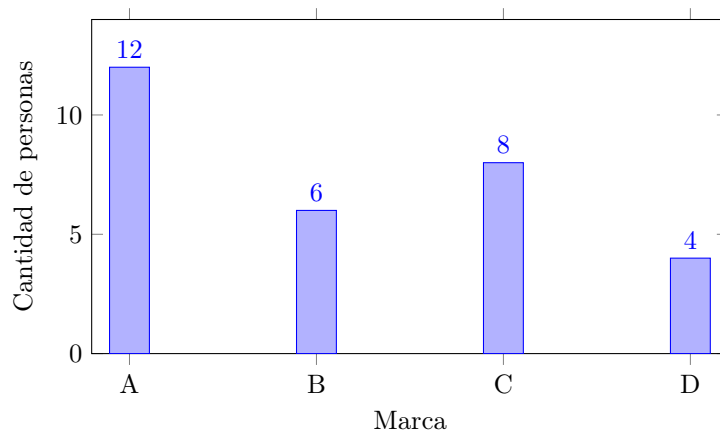
6. Representaciones gráficas

6.1. Gráficos para atributos

6.1.1. Diagrama de rectángulos

x_i	n_i	f_i
A	12	0,4000
B	6	0,2000
C	8	0,2667
D	4	0,1333
	30	1

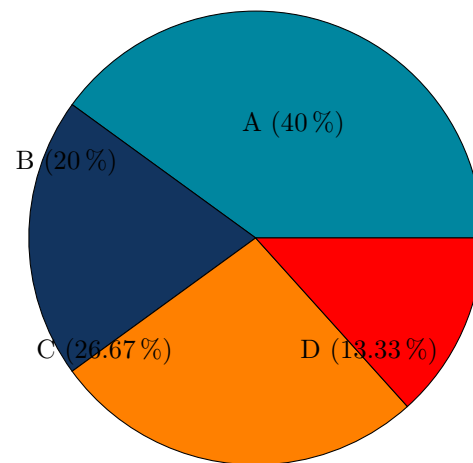
Marca de vaqueros preferida



6.1.2. Diagrama de sectores

x_i	n_i	f_i
A	12	0,4000
B	6	0,2000
C	8	0,2667
D	4	0,1333
	30	1

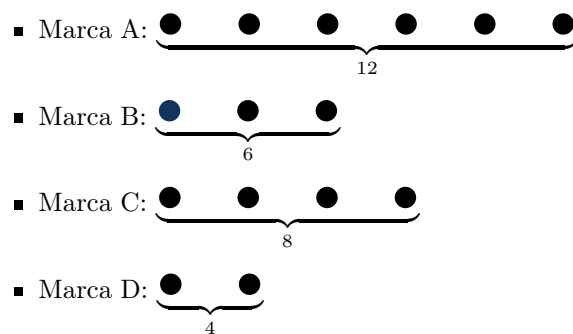
Marca de vaqueros preferida



6.1.3. Pictograma

x_i	n_i	f_i
A	12	0,4000
B	6	0,2000
C	8	0,2667
D	4	0,1333
	30	1

Marca de vaqueros preferida

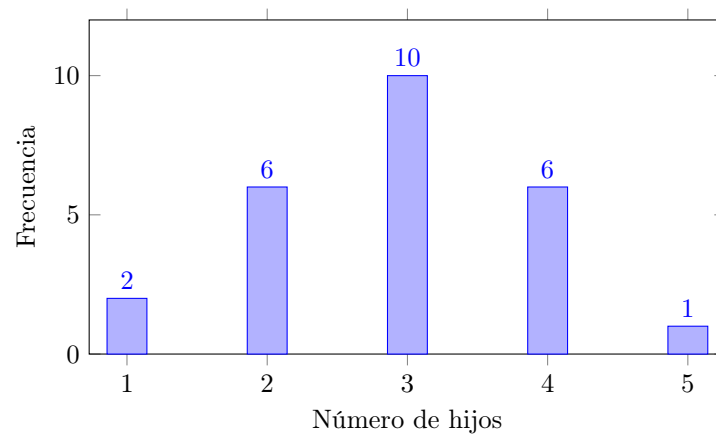


6.2. Gráficos para variables

6.2.1. Distribuciones con datos no agrupados en intervalos

Diagrama de barras

x_i	n_i	f_i
1	2	0,08
2	6	0,24
3	10	0,40
4	6	0,24
5	1	0,04
	25	1



Poligonal de frecuencias

x_i	n_i	f_i
1	2	0,08
2	6	0,24
3	10	0,40
4	6	0,24
5	1	0,04
	25	1

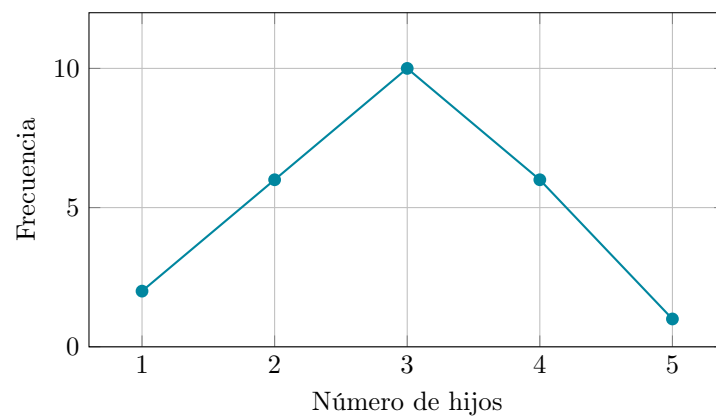
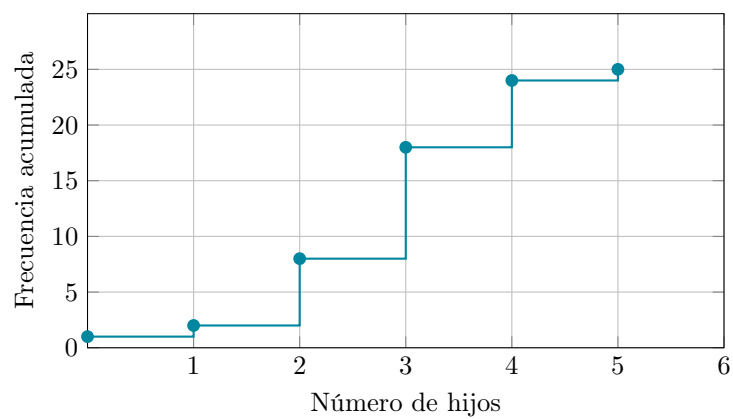


Diagrama acumulativo de frecuencias

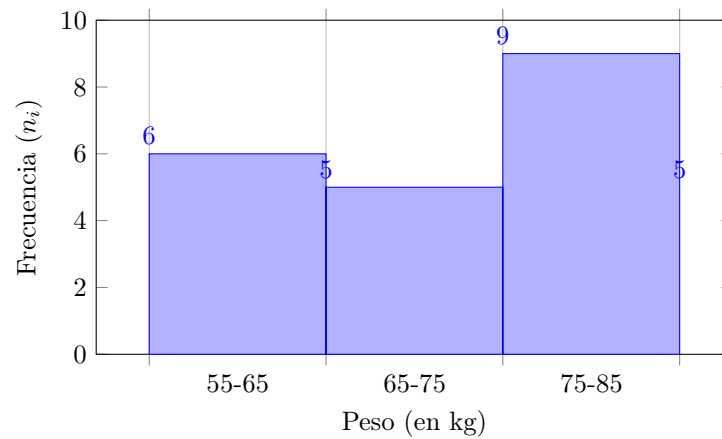
x_i	n_i	N_i	F_i
1	2	2	0,08
2	6	8	0,32
3	10	18	0,72
4	6	24	0,96
5	1	25	1
	25	1	



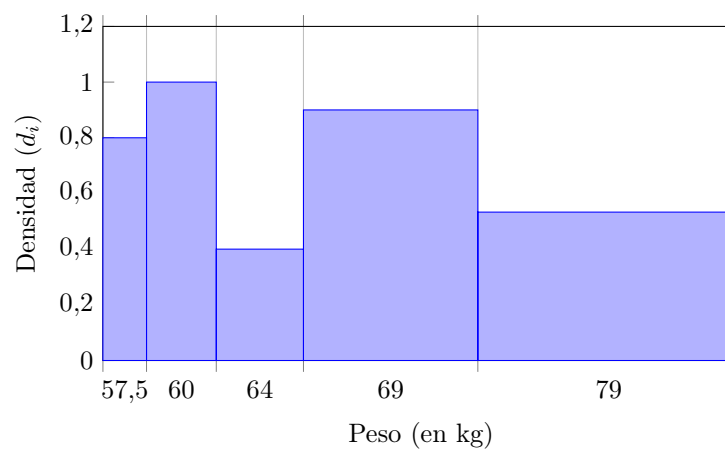
6.2.2. Distribuciones con datos agrupados en intervalos

Histograma

$(L_{i-1}, L_i]$	n_i	a_i
55 – 65	6	10
65 – 75	5	10
75 – 85	9	10
85 – 95	5	10
	25	

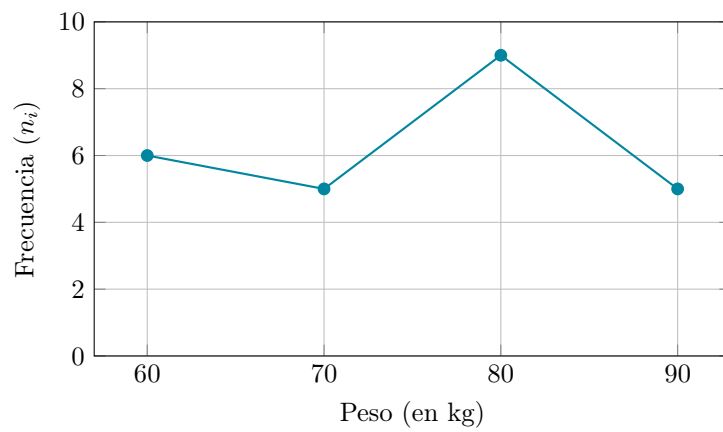


$(L_{i-1}, L_i]$	n_i	a_i	d_i
57,5 – 60	2	2,5	0,8
60 – 64	4	4	1
64 – 69	2	5	0,4
69 – 79	9	10	0,9
79 – 94	8	15	0,533
	25		



Polígono de frecuencias

$(L_{i-1}, L_i]$	x_i	n_i	f_i	a_i
55 – 65	60	6	0,24	10
65 – 75	70	5	0,20	10
75 – 85	80	9	0,36	10
85 – 95	90	5	0,20	10
		25	1	



Polígono acumulativo de frecuencias

$(L_{i-1}, L_i]$	n_i	N_i	F_i
55 – 65	6	6	0,24
65 – 75	5	11	0,44
75 – 85	9	20	0,80
85 – 95	5	25	1
	25		

