



DISTRIBUCIÓN DE FRECUENCIAS

Organización de la serie de DATOS

- SIN AGRUPAR: POCOS DATOS DIFERENTES

(I pequeño comparado con N).

- AGRUPADA en intervalos: MUCHOS DATOS DIFERENTES (I grande).

Cómo agrupar en intervalos muchos datos diferentes:

- Observar valor mínimo x_m y valor máximo x_M .

- Recorrido de la variable (amplitud total): $Re = x_M - x_m$

- Número de intervalos (k):

$$\begin{cases} \text{Sturges } k = \frac{\ln N}{\ln 2} + 1 \\ k = \sqrt{N} \text{ (N no muy grande)} \end{cases}$$

- Amplitud de intervalos $c = \frac{Re}{k}$



Ejemplo 1. El departamento de prevención de riesgos laborales de una gran empresa de la construcción ha recogido información sobre el número de accidentes laborales diarios con baja laboral que se han producido durante los 44 días siguientes a la aplicación de nuevas normas de seguridad, obteniendo los siguientes resultados:

X: Número de accidentes diarios 44 días)

2	1	0	3	3	4	4	3	7	4	4
1	0	4	2	4	0	2	2	4	3	2
0	3	0	3	5	1	5	0	0	3	0
7	5	4	5	3	9	3	10	3	0	9

Obtener:

1. La tabla estadística o distribución de frecuencias.
2. Calcúlese la media, la moda y los cuantiles de orden 0,25, 0,50 (mediana) y 0,75 (cuartiles).



Solución:

DISTRIBUCIÓN DE FRECUENCIAS

x	frecuencia n_i	porcentaje $f_i \times 100$	frec.acum N_i	porcentaje ac. $F_i \times 100$
0	9	20,5	9	20,5
1	3	6,8	12	27,3
2	5	11,4	17	38,6
3	10	22,7	27	61,4
4	8	18,2	35	79,5
5	4	9,1	39	88,6
7	2	4,5	41	93,2
9	2	4,5	43	97,7
10	1	2,3	44	100,0
Total	44	100,0		

MODA

CUANTILES

MEDIANA



Ejemplo 2. Las tasas de variación interanual de las finanzas de 64 empresas del sector Multimedia en el año 2011 aparecen en la siguiente serie de datos:

1,3	4,1	3,4	5,6	1,2	5,7	2,2	5,9	2,6	5,1	3,9	4,4	5,3	4,5	4,8	5,1
2,5	1,8	5,7	7,5	2,4	6,8	2,6	6,2	3,2	7,0	4,8	1,3	3,8	4,7	2,8	6,1
4,3	3,6	3,3	6,7	3,1	1,1	3,8	7,1	4,1	2,3	5,4	3,2	4,6	4,2	4,9	5,6
5,2	2,7	3,1	1,5	4,6	2,5	4,6	4,8	5,8	6,2	6,2	3,5	5,6	5,2	4,5	4,2

1. Forma una distribución de frecuencias con 7 intervalos.
2. Intervalo modal e intervalo mediano.
3. Observa los cuartiles sobre la distribución sin agrupar.



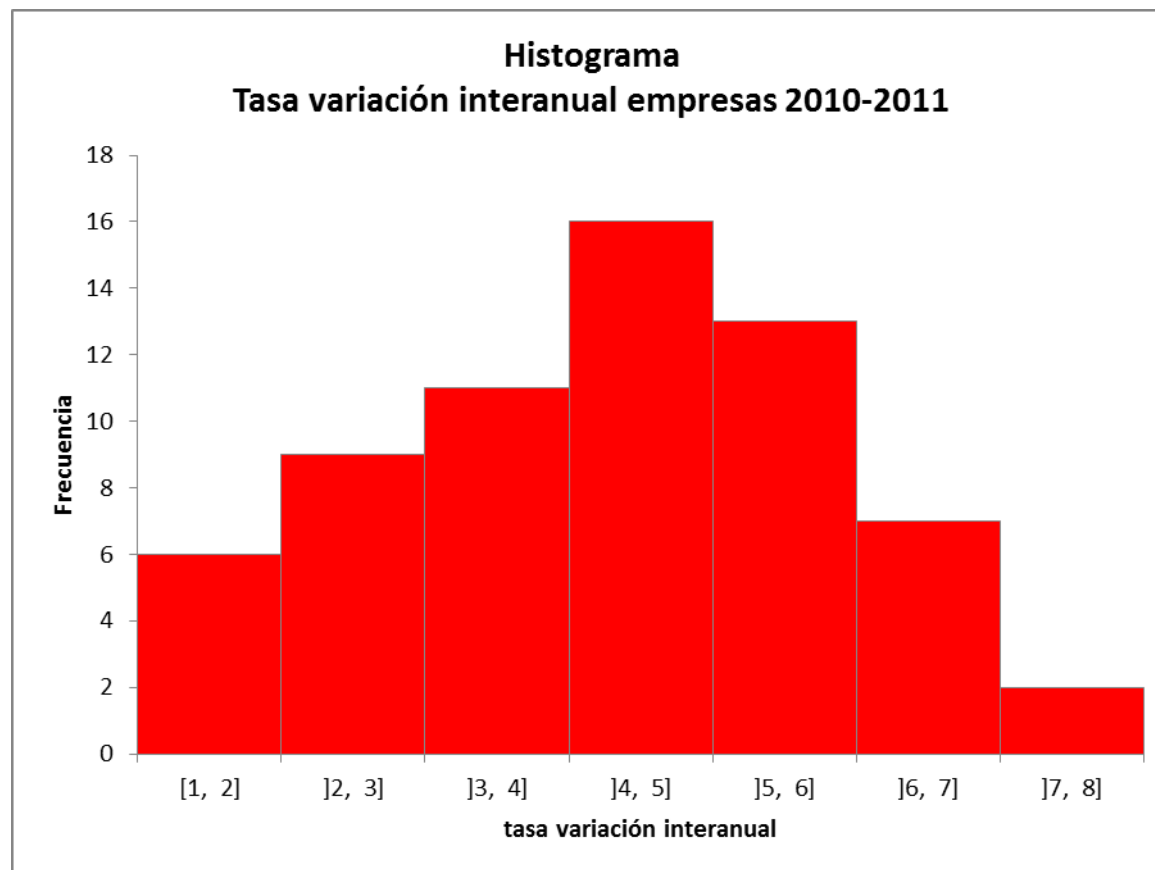
Solución:

1. DISTRIBUCIÓN AGRUPADA EN INTERVALOS

<i>Intervalos X: Tasa de var.</i>	<i>Frecuencia n_i nº empresas</i>	<i>Porcentaje $f_i \times 100$</i>	<i>Porcentaje acum. $F_i \times 100$</i>
[1, 2]	6	9,4%	9,4%
]2, 3]	9	14,1%	23,4%
]3, 4]	11	17,2%	40,6%
]4, 5]	16	25,0%	65,6%
]5, 6]	13	20,3%	85,9%
]6, 7]	7	10,9%	96,9%
]7, 8]	2	3,1%	100,0%
Totales	64	100%	

2. Intervalo modal:

Intervalo mediano:



Como todos los intervalos tienen la misma amplitud (que es lo más habitual), la altura de los rectángulos coincide con la frecuencia (no es necesario calcular la densidad de frecuencia

$$h_i = \frac{n_i}{c_i}$$



3. EJEMPLO 2. DISTRIBUCIÓN SIN AGRUPAR

<i>Tasa var</i>	<i>frecuencia</i>	<i>porcentaje</i>	<i>Porcentaje acum.</i>
1,1	1	1,56%	1,56%
1,2	1	1,56%	3,13%
1,3	2	3,13%	6,25%
1,5	1	1,56%	7,81%
1,8	1	1,56%	9,38%
2,2	1	1,56%	10,94%
2,3	1	1,56%	12,50%
2,4	1	1,56%	14,06%
2,5	2	3,13%	17,19%
2,6	2	3,13%	20,31%
2,7	1	1,56%	21,88%
2,8	1	1,56%	23,44%
3,1	2	3,13%	26,56%
3,2	2	3,13%	29,69%
3,3	1	1,56%	31,25%
3,4	1	1,56%	32,81%
3,5	1	1,56%	34,38%
3,6	1	1,56%	35,94%
3,8	2	3,13%	39,06%
3,9	1	1,56%	40,63%
4,1	2	3,13%	43,75%
4,2	2	3,13%	46,88%
4,3	1	1,56%	48,44%
4,4	1	1,56%	50,00%

<i>Tasa var</i>	<i>frecuencia</i>	<i>porcentaje</i>	<i>Porcentaje acum.</i>
4,5	2	3,13%	53,13%
4,6	3	4,69%	57,81%
4,7	1	1,56%	59,38%
4,8	3	4,69%	64,06%
4,9	1	1,56%	65,63%
5,1	2	3,13%	68,75%
5,2	2	3,13%	71,88%
5,3	1	1,56%	73,44%
5,4	1	1,56%	75,00%
5,6	3	4,69%	79,69%
5,7	2	3,13%	82,81%
5,8	1	1,56%	84,38%
5,9	1	1,56%	85,94%
6,1	1	1,56%	87,50%
6,2	3	4,69%	92,19%
6,7	1	1,56%	93,75%
6,8	1	1,56%	95,31%
7,0	1	1,56%	96,88%
7,1	1	1,56%	98,44%
7,5	1	1,56%	100,00%
Total general	64	100,00%	

CUARTILES:



Ejemplo 3. Calcúlese la MODA de los siguientes conjuntos de datos y distribuciones:

a) $X: 2, 2, 3, 2, 6, 7, 6$

b) Intervalos	Frecuencia	Intervalos	Frecuencia
0,0 – 1,5	6	2 – 3	10
1,5 – 2,5	12	3 – 4	40
2,5 – 3,0	12	4 – 5	8

Solución: a) $Mo = 2$ b) $Mo \in (2, 5; 3)$ c) $Mo \in (3; 4)$



CONSIDERACIONES SOBRE LA MEDIANA Y LA MEDIA:

1. La mediana es más estable que la media al no verse afectada por valores extremos. Por ejemplo:

$$1, 2, 3, 4, 5, 6, 7 \Rightarrow \bar{x} = 4 \text{ y } Me = 4$$

$$1, 2, 3, 4, 5, 6, 49 \Rightarrow \bar{x} = 10 \text{ y } Me = 4$$

2. Si una variable X tiene una distribución simétrica, la mediana y la media coinciden:

$$\bar{x} = Me.$$

La mediana y la media son medidas de posición central. La moda también es una medida de posición central en distribuciones unimodales y campaniformes.



Ejemplo 4. Se ha contabilizado el número de días de baja, durante un trimestre, de los trabajadores de dos empresas obteniéndose los siguientes resultados:

DATOS empresa A

0	2	1	1	3
2	0	1	5	2
2	3	3	2	1
4	2	2	1	3

DATOS empresa B

0	1	1	2	9	1	1	1
0	0	1	1	0	9	1	9
1	1	0	1	1	1	1	1
9	9	1	0	1	1	1	0
0	0	1	1	9	0	2	1

Vamos a DESCRIBIR las dos series de datos y a comparar el número de días de baja en las dos empresas. Para ello se determinará:

Distribución de frecuencias. Media aritmética. **Varianza y desviación típica.**

**SOLUCIÓN:**

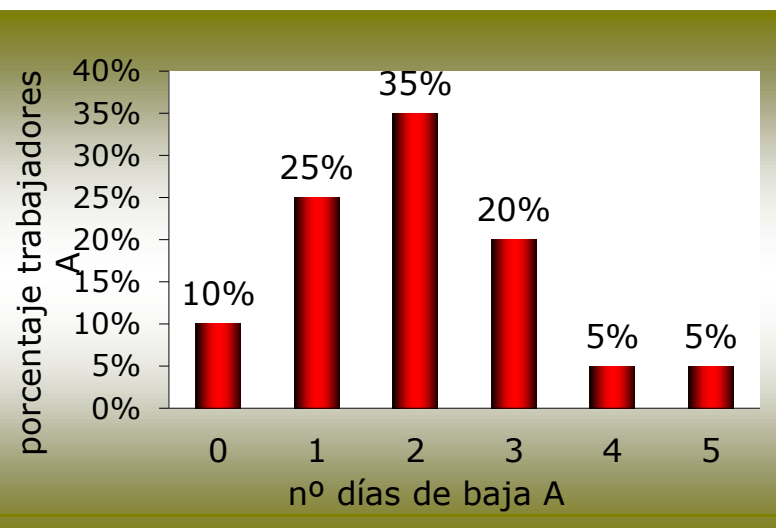
X: nº días baja
empresa A

nº trabajadores

Valores x	Frecuencia n_i	Porcentaje $f_i \times 100$	Porcent.acum $F_i \times 100$
0	2	10%	10%
1	5	25%	35%
2	7	35%	70%
3	4	20%	90%
4	1	5%	95%
5	1	5%	100%
Total	20	100%	

DISTRIBUCIÓN DE
FRECUENCIAS

GRÁFICO DE BARRAS





nº dias baja
empresa B

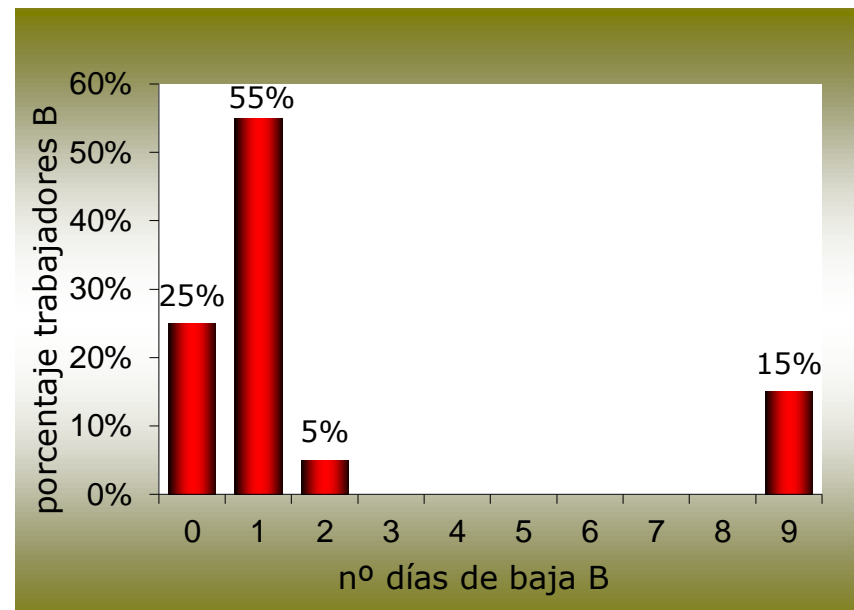
nº trabajadores

DISTRIBUCIÓN
DE FRECUENCIAS

Valores x	Frecuencia n_i	Porcentaje $f_i \times 100$	Porcent.acum $F_i \times 100$
0	10	25%	25%
1	22	55%	80%
2	2	5%	85%
9	6	15%	100%
Total	40	100%	

GRÁFICO DE BARRAS

Para comparar gráficos de diferentes variables es conveniente hacerlo en términos relativos (porcentajes o frecuencias relativas)





CÁLCULO MEDIA Y VARIANZA

nº dias

baja A

nº trabajadores

valores	Frecuencia n_i	Porcentaje $f_i \times 100$	Porcent.acum $F_i \times 100$	cal.media	cal.varianza
0	2	10	10	0	8
1	5	25	35	5	5
2	7	35	70	14	0
3	4	20	90	12	4
4	1	5	95	4	4
5	1	5	100	5	9
Total	20	100		40	30

nº dias

baja B

nº trabajadores

valores	Frecuencia n_i	Porcentaje $f_i \times 100$	Porcent.acum $F_i \times 100$	cal.media	cal.varianza
0	10	25,00	25,00	0	40
1	22	55,00	80,00	22	22
2	2	5,00	85,00	4	0
9	6	15,00	100,00	54	294
Total	40	100,00		80	356



media	2,00
varianza	1,5000
desv.tip.	1,22

RESUMEN DATOS EMPRESA A

media	2,00
varianza	8,9000
desv.tip.	2,98

RESUMEN DATOS EMPRESA B

Cuando la unidad de medida es la misma en las variables y las medias son iguales, las varianzas y desviaciones típicas se pueden comparar directamente.

En caso contrario se ha de calcular el coeficiente de Variación de Pearson para comparar la dispersión (dispersión relativa).