

# Estadística Descriptiva I

## Distribuciones unidimensionales

PROBLEMAS TEMA 2  
Ideas para resolución

Mar Angulo Martínez

Curso 2023-2024

- **Problema 1**

<u>Consumos en <math>\frac{\text{kw}}{\text{h}}</math></u> $I_i$	Nº hogares $n_i$	Ni	xi.ni
[1, 3)	1	1	2
[3, 5)	1	2	4
[5, 7)	11	13	66
[7, 9)	21	34	168
[9, 11)	25	59	250
[11, 13)	17	76	204
[13, 15)	9	85	126
[15, 17)	4	89	64
[17, 19)	1	90	18
	<b>n=90</b>		<b>902</b>

a)  $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{902}{90} = 10,02$  miles de kw/h es el consumo medio de esos 90 hogares

- Como todos los intervalos tienen la misma amplitud, la Md está en el intervalo de mayor frecuencia  $n_i$
- La moda está en el intervalo [9, 11)

- $Md = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} c_i = 9 + \frac{17}{17+21} 2 = 9,895$  kw/h es el consumo más habitual

b) La compañía decide hacer una oferta al 10 % de los hogares que más consumen y aplicar una exención de 2 meses al 3 % de los que tienen los consumos más elevados.

¿Con qué consumos se consigue ser beneficiario de la oferta y de la exención?

$$\frac{90n}{100} = 81 \longrightarrow P_{90} \in [13,15)$$

$$P_{90} = L_{i-1} + \frac{\frac{90n}{100} - N_{i-1}}{n_i} c_i = 13 + \frac{81-76}{9} * 2 = 14,11 \text{ kw/h es el consumo mínimo para acceder a la oferta}$$

$$\frac{97n}{100} = 87,3$$

$$P_{90} \in [15,17)$$

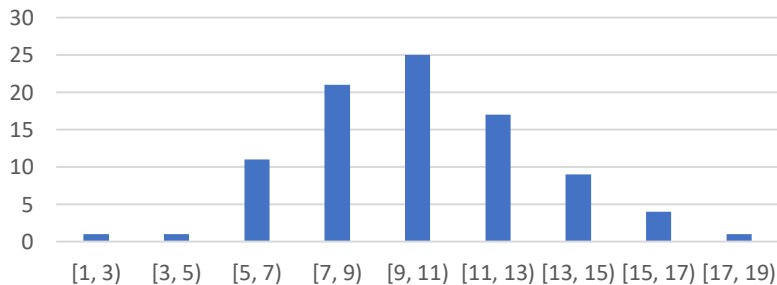
$$P_{97} = L_{i-1} + \frac{\frac{97n}{100} - N_{i-1}}{n_i} c_i = 15 + \frac{87,3 - 85}{4} * 2 = 16,15 \text{ kw/h}$$

es el consumo mínimo para acceder a la exención

c) Como hemos calculado ya media y moda, utilizamos coeficiente de asimetría de Pearson

$$A_p = \frac{\bar{x} - Md}{s} = \frac{10,02 - 9,895}{s} > 0 \rightarrow \text{Asimétrica hacia la derecha (ligerísima)}$$

Título del gráfico



e) Consumos entre 7 y 15 kw/h:  
Son  $21 + 25 + 17 + 9 = 72$  hogares  
Representan un 80% de la muestra

- **Problema 2**

<u>Pesos</u> <u>(grs)</u>	Nº productos $n_i$	Ni	xi.ni	$x_i^2.n_i$
70	4	4	280	19.600
74	9	13	666	49.284
78	16	29	1248	97.344
82	30	59	2460	201.720
86	44	103	3784	325.424
90	36	139	3240	291.600
94	20	159	1880	176.720
98	12	171	1176	115.248
102	6	177	612	62.424
	<b>n=177</b>		<b>15.346</b>	<b>1.339.364</b>

- $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{15.346}{177} = 86,7$  grs es el peso medio de los 177 productos de la muestra
- Mediana  $\frac{n}{2} = 88,5 \longrightarrow Me = 86$
- Moda = 86 (es el valor al que corresponde la mayor  $n_i$ )

$$A_p = \frac{\bar{x} - Md}{s} = \frac{86,7 - 86}{s} > 0 \longrightarrow \text{Asimétrica hacia la derecha}$$

$$b) s^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{1.339.364}{177} - (86,7)^2 = 50,14$$

$$S = +\sqrt{s^2} = 7,08 \quad CV_p = \frac{7,08}{86,7} = 0,0817 \quad (\text{del } 8,17\%)$$

la media es muy representativa, puesto que la dispersión es muy pequeña.

- C) ¿Qué % de productos tienen un peso mayor de 78 y menor que 86? ¿y como mínimo de 78 y como mucho de 86?
- Hay 30 productos con un peso mayor de 78 y menor de 86 grs: representan un 16,95% de la muestra
- Hay 90 productos con un peso mínimo de 78 y máximo de 86 grs: representan un 50,85% de la muestra

d) Si en las ventas se desechan el 15% de productos de menor peso; ¿a partir de qué peso un producto de esta muestra se considera apto para la venta?

Tenemos que calcular el  $P_{15}$   $\frac{15n}{100} = 26,55 \longrightarrow P_{15} = 78$

Los productos con menos de 78 grs se consideran no aptos para la venta

### ■ Problema 3

Nº plazas	Nº establecimientos	Ni	xi.ni	$x_i^2 \cdot n_i$
[0,100)	25	25	1.250	62.500
[100,200)	37	62	5.550	832.500
[200,300)	12	74	3.000	750.000
[300,400)	22	96	7.700	2.695.000
[400,500)	21	117	9.450	4.252.500
	<b>n=117</b>		<b>26.950</b>	<b>8.592.500</b>



- a)  $22+21= 43$  establecimientos tienen al menos 300 plazas  
b)  $37+12+22 = 71$  establecimientos que representan  $71/117=60,68\%$

c) **P85:**

- 1) calculamos  $85n/100 = 99,45-----\rightarrow$  **P85** está en  $[400,500)$
- 2) **P85**  $= L_{i-1} + \frac{\frac{85n}{100} - N_{i-1}}{n_i} c_i = 400 + \frac{99,45-96}{21} * 100 = 416,43$  plazas

***Un hotel con 417 plazas está en el 15% de los que más plazas ofertan***

- d)  $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{26.950}{117} = 230,34$  plazas
- la Md está en  $[100,200)$
- $Md = L_i - 1 + \frac{n_{i+1}}{n_{i+1}+n_{i-1}} c_i = 100 + \frac{12}{12+25} 100 = 132.43$
- $Md = L_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 100 + \frac{37-25}{(37-25)+(37-12)} 100 = 132.43$
- ***El número de plazas más frecuente en esa muestra de hoteles es de 132***

■ **e) Nos piden el Percentil 35**

- 1) calculamos  $35n/100 = 40,95$ -----→ **P35** está en [100,200)
- 2)  $P_{35} = L_{i-1} + \frac{\frac{35n}{100} - N_{i-1}}{n_i} c_i = 100 + \frac{40,95 - 25}{37} * 100 = 100,43$  plazas
- ***Un hotel con 101 plazas supera en capacidad al 35% de los establecimientos de la muestra. Entra por tanto en el plan previsto***

f) Calculamos la varianza  $s^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{8.592.500}{117} - (230,34)^2 = 20383,65$   
 $s = 142,77$  plazas

$$A_P = \frac{\bar{x} - Md}{s} = \frac{230,34 - 132,43}{142,77} = 0,686 > 0 \longrightarrow \text{Asimétrica hacia la derecha}$$

## • Problema 4

<u>Nº</u> <u>incumplimientos</u> $I_i$	Nº transductores $n_i$	$N_i$	$x_i \cdot n_i$	$x_i^2 \cdot n_i$	$k_i$
[0, 2)	9	9	9	9	4,5
[2, 6)	15	24	60	240	3,75
[6, 8)	17	41	119	833	8,5
[8, 10)	6	47	54	486	3
[10, 15)	3	50	37,5	468,75	0,6
	<b>n=50</b>		<b>279,5</b>	<b>2036,75</b>	

$\bar{x} = \frac{279,5}{50} = 5,59$  es el nº medio de incumplimientos de ese conjunto de transductores

b) Tenemos que calcular el  $P_{85}$

$$\frac{85n}{100} = 42,5 \longrightarrow P_{85} \in [8,10)$$

$$\mathbf{P85} = L_{i-1} + \frac{\frac{85n}{100} - N_{i-1}}{n_i} c_i = 8 + \frac{42,5 - 41}{6} * 2 = 8,5 \text{ incumplimientos}$$

c) Si a través de un proceso mecánico se consigue reducir los incumplimientos de cada transductor en un 15%, ¿Cómo quedan modificados la media y el coeficiente de variación? (Indicar cuáles son los nuevos valores)

Los nuevos valores serían  $y_i = 0,85 x_i$

Entonces  $\bar{y} = 0,85 \bar{x} = 4,75$ ;  $s_y^2 = (0,85)^2 s_x^2 \longrightarrow s_y = 0,85 s_x \longrightarrow CV_y = \frac{s_y}{\bar{y}} = \frac{s_x}{\bar{x}} = 0,551$

$$s^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{2036,75}{50} - 5,59^2 = 9,487$$

$$S = +\sqrt{s^2} = +\sqrt{9,487} = 3,08 \quad CV_p = \frac{3,08}{5,59} = 0,551$$

d) la Md está en [6, 8) que es el intervalo de mayor densidad de frecuencia ( de mayor  $k_i$ )

▪  $Md = L_{i-1} + \frac{k_{i+1}}{k_{i+1}+k_{i-1}} c_i = 6 + \frac{3}{3+3,75} 2 = 6,89$  incumplimientos es lo más habitual

e) *Para analizar la simetría podemos utilizar*  $A_P = \frac{\bar{x} - Md}{s} = \frac{5,59 - 6,89}{3,08} = -0,422$

La distribución presenta una cierta asimetría hacia la izquierda

### ■ Problema 5

xi	ni	$x_i^2 \cdot n_i$
350	1	122.500
408	1	166.464
540	1	291.600
555	1	308.025
575	1	330.625
590	1	348.100
608	1	369.664
679	1	461.041
815	1	664.225
1285	1	1.651.225

- $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{6.405}{10} = 640,5$  m. dólares
- $Me = \frac{575+590}{2} = 582,5$  m. dólares

### ■ Cambios de origen y escala

b)  $\bar{x} + 40 = 680,5$  m dólares       $Me_y = Me + 40 = 622,5$  m. dólares

c)  $\bar{x}/3 = 213,5$  m dólares       $Me_y = Me/3 = 207,5$  m. Dólares

$$s^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{4.713.469}{10} - 640,5^2 = 61.106,75$$

$$s = +\sqrt{s^2} = +\sqrt{61.106,75} = 247,2$$

$$CV_p = \frac{247,2}{640,5} = 0,386$$

e) ¿Cómo se ve afectada la dispersión de los datos en cada uno de los supuestos anteriores?

- Si  $y_i = x_i + 40$

$$s_y^2 = s_x^2 = 61.106,75 \longrightarrow S_y = S_x = 247,2$$

$$CV_y = \frac{s_y}{\bar{x} + 40} = \frac{247,2}{680,5} = 0,3633$$

- Si  $t_i = x_i/3$

$$s_t^2 = \frac{s_x^2}{9} = \frac{61.106,75}{9} = 6.789,64$$

$$S_t = s_x/3 = 82,4$$

$$CV_t = \frac{s_t}{\bar{t}} = \frac{s_x/3}{\bar{x}/3} = \frac{s_x}{\bar{x}} = CV_x = 0,386$$

## • Problema 6

<u>Nº artículos</u> $I_i$	Nº autores $n_i$	$N_i$	$x_i.n_i$	$x_i^2.n_i$	$k_i$
[1, 4)	16	16	40	100	5,33
[4, 6)	12	28	60	300	6
[6, 8)	5	33	35	245	2,5
[8, 10)	7	40	63	567	3,5
[10, 15)	2	42	25	312,5	0,4
	<b>n=42</b>		<b>223</b>	<b>1524,5</b>	

a) Han publicado menos de 10 artículos 40 autores de un total de 42: un 95,24% de ellos.



b)  $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{223}{42} = 5,31$  artículos es el nº medio de publicaciones de esos 42 autores

c) Los valores que delimitan esos grupos son los  $P_{85}$  y  $P_5$  respectivamente

$$\frac{85n}{100} = 35,7 \longrightarrow P_{85} \in [8,10)$$

$$P_{85} = L_{i-1} + \frac{\frac{85n}{100} - N_{i-1}}{n_i} c_i = 8 + \frac{35,7 - 33}{7} * 2 = 8,77 \text{ artículos}$$

$$\frac{5n}{100} = 2,1 \quad \longrightarrow \quad P_5 \in [1,4)$$

$$P_5 = L_{i-1} + \frac{\frac{5n}{100} - N_{i-1}}{n_i} c_i = 1 + \frac{2,1-0}{16} * 3 = 1,394 \text{ artículos}$$

**Por tanto, participarán en el debate los autores que publican menos de 1,394 artículos o más de 8,77 artículos: en definitiva, los autores que publican 1 ó bien 9 y más artículos**

$$e) \quad s^2 = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{1524,5}{42} - (5,31)^2 = 8,1039$$

$$S = +\sqrt{s^2} = +\sqrt{8,1039} = 2,847 \quad CV_p = \frac{2,847}{5,31} = 0,5362$$

Si cada autor publica 3 artículos más, tenemos nuevos valores  $y_i = x_i + 3$

$$\bar{y} = \bar{x} + 3 = 8,31 \text{ artículos} \quad s_y^2 = s_x^2 \quad CV_y = \frac{s_y}{\bar{y}} = \frac{s_x}{\bar{x}+3} = \frac{2,847}{5,31+3} = 0,3426$$

## ■ Problema 7

$$a) \bar{x} = \frac{3.310 \cdot 20 + 1.450 \cdot 50 + 1.870 \cdot 130}{200} = \frac{381.800}{200} = 1.909 \text{ euros}$$

$$s_e^2 = \frac{n_1 s_1^2 + n_2 s_2^2 + n_3 s_3^2}{n} = \frac{20(920)^2 + 50(125)^2 + 130(420)^2}{200} = 203.206,25 \quad s = 450,78$$

$$CV = \frac{450,78}{1.909} = 0,2361$$

b) Para comparar dispersiones utilizamos el CV de Pearson:

$$\begin{aligned} CV_1 &= \frac{920}{3310} & CV_2 &= \frac{125}{1450} & CV_3 &= \frac{420}{1870} \\ CV_1 &= 0,278 & CV_2 &= 0,0862 & CV_3 &= 0,225 \end{aligned}$$

*Hay por tanto mayor homogeneidad salarial en la categoría de administrativos  
Dispersión significativamente menor en la categoría de administrativos*

Si se proponen dos alternativas de incrementos salariales

a. Elevación de todos los salarios un 5%

b. Incremento lineal de 250 euros

Analizar cuál de las medidas disminuye más la dispersión relativa inicial de los salarios de la empresa

c) Si  $y_i = 1,05 * x_i$

Los nuevos CV serán  $\frac{1,05 * s_x}{1,05 * \bar{x}}$  ---  $\rightarrow$  se mantendrán iguales

Si  $y_i = x_i + 250$

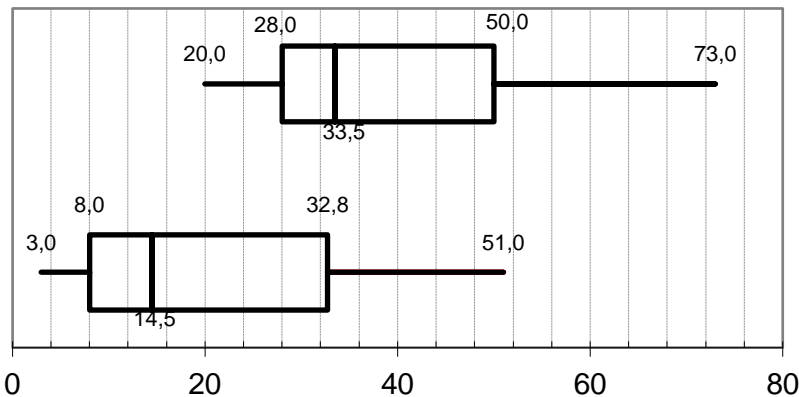
Los nuevos CV serán  $\frac{s_x}{\bar{x} + 250}$  la dispersión relativa se reduce

## Problema 9

DOS VARIABLES

Mínimo	Q1	Mediana	Q3	Máximo
3,00	8,00	14,50	32,75	51,00
20,00	28,00	33,50	50,00	73,00

Gráfico de caja y bigote para dos variables



**X:** edad de los espectadores de una serie sobre videojuegos  
**Y:** edad de los espectadores de una serie sobre política

- Identificar los siguientes valores para cada grupo: Mediana,  $Q_1$ ,  $Q_3$ , valor mínimo, valor máximo, rango y rango intercuartílico.
- ¿En qué variable es mayor la mediana?
- ¿En qué variable es mayor el rango intercuartílico?
- Interpretar el  $Q_3(x)$  y  $Me(y)$

- ¿Qué conclusiones puedes extraer acerca de la dispersión y acerca de la simetría de ambas variables?
- ¿Entre qué dos valores se encuentra en cada caso el 50% central de la distribución?
- ¿Qué edad podría tener una persona que está en el 10% de los más mayores que siguen la serie sobre videojuegos?
- ¿Qué edad podría tener un espectador que está entre el 15% de los más jóvenes de los que siguen la serie sobre política?

## • Problema 10

- $n = 153$        $\bar{x} = 135.39$        $Me = 135.40$        $s = 4.59$
- Valor mínimo = 122.20    Valor máximo = 147.70     $Q_1 = 132.95$        $Q_3 = 138.25$

Calculamos el Rango Intercuartílico =  $138,25 - 132,95 = 5,3$

Y los Lím Inf =  $132,95 - 1,5 * 5,3 = 125$

Y Lím Sup =  $138,25 + 1,5 * 5,3 = 146,2$

Asimétrica hacia la derecha     $Q_3 - Me = 2,85$      $Me - Q_1 = 2,45$

CV = 0,034 ligerísima dispersión →

Media muy representativa

Los valores máx y mín son valores atípicos    ¿son extremos?

