TEMA 7 - ESTATÍSTICA

- **Poboación**: individuos que se van estudar.
  - → Mostra: subconxunto máis pequeno da poboación.

- Variable Estatística: cualidade a estudar.

Variables
- Cualitativas: non medibles numéricamente.
- Cuantitativas: pódense medir
  - Discretas: toman valores numerables
  - Continuas: o valor pode ser calquera número dentro dun intervalo.

Estudo Estatístico → Táboa Frecuencias.

X = variable a estudar

x_i = valores que toma a variable

f_i = frecuencia absoluta

F_i = " " acumulada

h_i = " relativa

H_i = " " acumulada

N = tamaño da poboación

x_i	f_i	F_i	h_i	H_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
1	4	4	4/17	4/17	4	4
2	7	11	7/17	11/17	14	28
3	1	12	1/17	12/17	3	9
4	5	17	5/17	17/17	20	80
	17		17/17 = 1		41	121

→ Mediana é 2

→ Moda é 2

$$\bar{X} = \frac{\sum x_i \cdot f_i}{N} = \frac{41}{17} = 2{,}4 \text{ pers.}$$

$$\sigma^2 = \frac{121}{17} - 2{,}4^2 = 1{,}35 \text{ pers.}^2$$

$$\sigma = \sqrt{1{,}35} = 1{,}16 \text{ persoas}$$

- Medidas de centralización:

- Media : $\bar{X} = \dfrac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_n \cdot f_n}{N} = \dfrac{\sum_{i=1}^{n} x_i f_i}{N}$

- Mediana: se ordenamos todos os datos, o dato do medio.

- Moda: o valor máis repetido.

- Medidas de dispersión:

- Rango: $X_{max} - X_{min}$

- Varianza:

$$Var = S^2 = \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i}{N} = \frac{\sum_{i=1}^{n} x_i^2 \cdot f_i}{N} - \bar{X}^2$$

- Desviación típica:

$$S = \sigma = \sqrt{Varianza}$$

- Coeficiente de Variación: $CV = \frac{\sigma}{\bar{X}} \longrightarrow CV(\%)$

– Medidas de Posición:

- Cuartiles : C_i / Q_i

  Dividen en 4 ós datos.

  C_1 agrupa o 25% dos datos

  C_2 .. . 50% -- --

- Deciles : D_i

  Dividen en 10 ós datos.

- Percentil : P_i

  Dividen en 100 ós datos.

1. A seguinte táboa da información sobre cantos anos tardan os alumnos do instituto en cambiar de móbil:

xi	fi	Fi	hi (%)	Hi (%)	xi·fi	xi²·fi
0	56	56	11,2	11,2	0	0
1	89	145	17,8	29	89	89
2	168	313	33,6	62,6	336	672
3	136	449	27,2	89,8	408	1224
4	35	484	7	96,8	140	560
5	16	500	3,2	100	80	400
Σ	500		100		1053	2945

Calcula: ~~media~~, ~~mediana~~, ~~moda~~, ~~rango~~, ~~varianza~~, ~~desviación típica~~, ~~CV~~, cuartil 3, decil 7 e percentil 12.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N} = \frac{1053}{500} = 2{,}106 \text{ anos}$$

Mediana $= 2 = C_2 = D_5 = P_{50}$

Moda $= 2$

Rango $= 5 - 0 = 5$

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 \cdot f_i}{N} - \bar{X}^2 = \frac{2945}{500} - 2{,}106^2 = 1{,}4547\ldots \rightarrow \sigma = \sqrt{1{,}45\ldots} = 1{,}206$$

$$CV = \frac{\sigma}{\bar{X}} = \frac{1{,}206}{2{,}106} = 0{,}57 = 57\,\%$$

C_3 : Posición $\rightarrow \frac{3}{4} \cdot 500 = 375$

$C_3 = 3$

D_7 : Agrupa o 70% dos datos

$D_7 = 3$

P_{12} : Agrupa o 12% dos datos

$P_{12} = 1$

2. Nunha enquisa sobre o prezo da vivenda na cidade de Vigo obtemos os seguintes datos:

Clases	ci	fi	Fi	hi(%)	Hi(%)	ci·fi	ci²·fi
[300,400)	350	42	42	4,2	4,2	14700	5145000
[400,500)	450	55	97	5,5	9,7	24750	11137500
[500,600)	550	135	232	13,5	23,2	74250	408375000
[600,700)	650	300	532	30	53,2	195000	126750000
[700,800)	750	283	815	28,3	81,5	212250	159187500
[800,900)	850	110	925	11	92,5	93500	79475000
[900,1000)	950	75	1000	7,5	100	71250	67687500
		1000		100		685700	490220000

Calcula: media, mediana, moda, rango, varianza, desviación típica, CV, cuartil 2, decil 3 e percentil 72.

$D_3 = 650$ $P_{72} = 750$

$$\bar{c} = \bar{x} = \frac{\sum_{i=1}^{n} c_i \cdot f_i}{N} = \frac{685700}{1000} = 685,7 €$$

$$Me = 650 € = C_2 = D_5 = P_{50}$$

$$Mo = 650 €$$

$$Rango = 950 - 350 = 600 €$$

$$\sigma^2 = \frac{\sum c_i^2 \cdot f_i}{N} - \bar{c}^2 = \frac{490220\cancel{000}}{1\cancel{000}} - 685,7^2 =$$

$$= 20035.51 €^2$$

$$\sigma = \sqrt{\sigma^2} = 141,55 €$$

$$CV = \frac{\sigma}{\bar{c}} = \frac{141,55}{685,7} = 0,2064 \longrightarrow CV = 20,64\%$$

Clases e Marcas de Clase

Cando os valores que pode tomar a variable son moi numerosos, resulta moi útil agrupalos en intervalos que chamamos clases.

O seu valor intermedio chámase marca de clase (c_i) e é o que usaremos para os cálculos.

4 $0 \longrightarrow 14$

$$\text{Rango} = 14 - 0 = 14$$

$$\frac{14+1}{5} = \frac{15}{5} = 3$$

$[0, 2]$

$[3, 5]$

$[6, 8]$

Variables Estatísticas Bidimensionais

Estúdanse 2 variables á vez, normalmente para comprobar se existe relación entre elas.

$$f_{ij} : \text{frecuencia conxunta} \begin{cases} i : \text{variable } X \\ j : \text{variable } Y \end{cases}$$

9 En una clase se van a comparar las notas obtenidas en las asignaturas de Inglés (X) y Filosofía (Y) para analizar su comportamiento. Se han recogido estos datos:

(8, 10), (6, 5), (3, 6), (7, 9), (9, 9), (10, 8), (6, 6), (5, 5), (5, 7), (4, 4), (9, 7), (5, 3), (8, 6), (7, 6), (6, 7), (6, 2), (4, 9), (6, 8), (9, 5), (5, 6), (10,10), (9, 8), (7, 7), (5, 4), (4, 6), (3, 2), (3, 5), (7, 5), (5, 6), (8, 9), (9, 6), (5, 7)

a) Construye la tabla de distribución conjunta.

b) Determina las distribuciones marginales.

c) Calcula la media de la variable X y la de la variable Y.

d) Halla la distribución condicionada de la variable X sabiendo que Y toma el valor ~~5~~ [5,6]. ¿Qué indica?

e) Determina la distribución condicionada de la variable Y cuando X toma el valor 4. ¿Qué indica?

X \ Y	[1,2]	[3,4]	d) [5,6]	[7,8]	[9,10]	
[0,2]	0	0	0	0	0	0
[3,4]	1	1	3	0	1	6
[5,6]	1	2	5	4	0	12
[7,8]	0	0	3	1	3	7
[9,10]	0	0	2	3	2	7
	2	3	13	8	6	32

Distribucións Marxinais

X_i	f_i	$x_i f_i$	$x_i^2 f_i$
[1,2]	0	0	0
[3,4]	6	21	73,5
[5,6]	12	66	363
[7,8]	7	52,5	404,25
[9,10]	7	66,5	631,75
	32	206	

$$\bar{X} = \frac{206}{32} = 6{,}4375$$

Y_j	f_j	$y_j f_j$	$y_j^2 \cdot f_j$
[1,2]	2	3	4,5
[3,4]	3	10,5	36,75
[5,6]	13	71,5	393,25
[7,8]	8	60	450
[9,10]	6	57	541,5
	32	202	

$$\bar{Y} = \frac{202}{32} = 6{,}3125$$

d)

$X \mid Y = Y_3$	f_i
[1,2]	0
[3,4]	3
[5,6]	5
[7,8]	3
[9,10]	2

- Distribucións Marxinais:
  Estudo de cada variable por separado, coa súa táboa propia.
- Distribucións Condicionadas:
  Estudo dunha variable condicionada a que a outra variable tome un valor concreto.

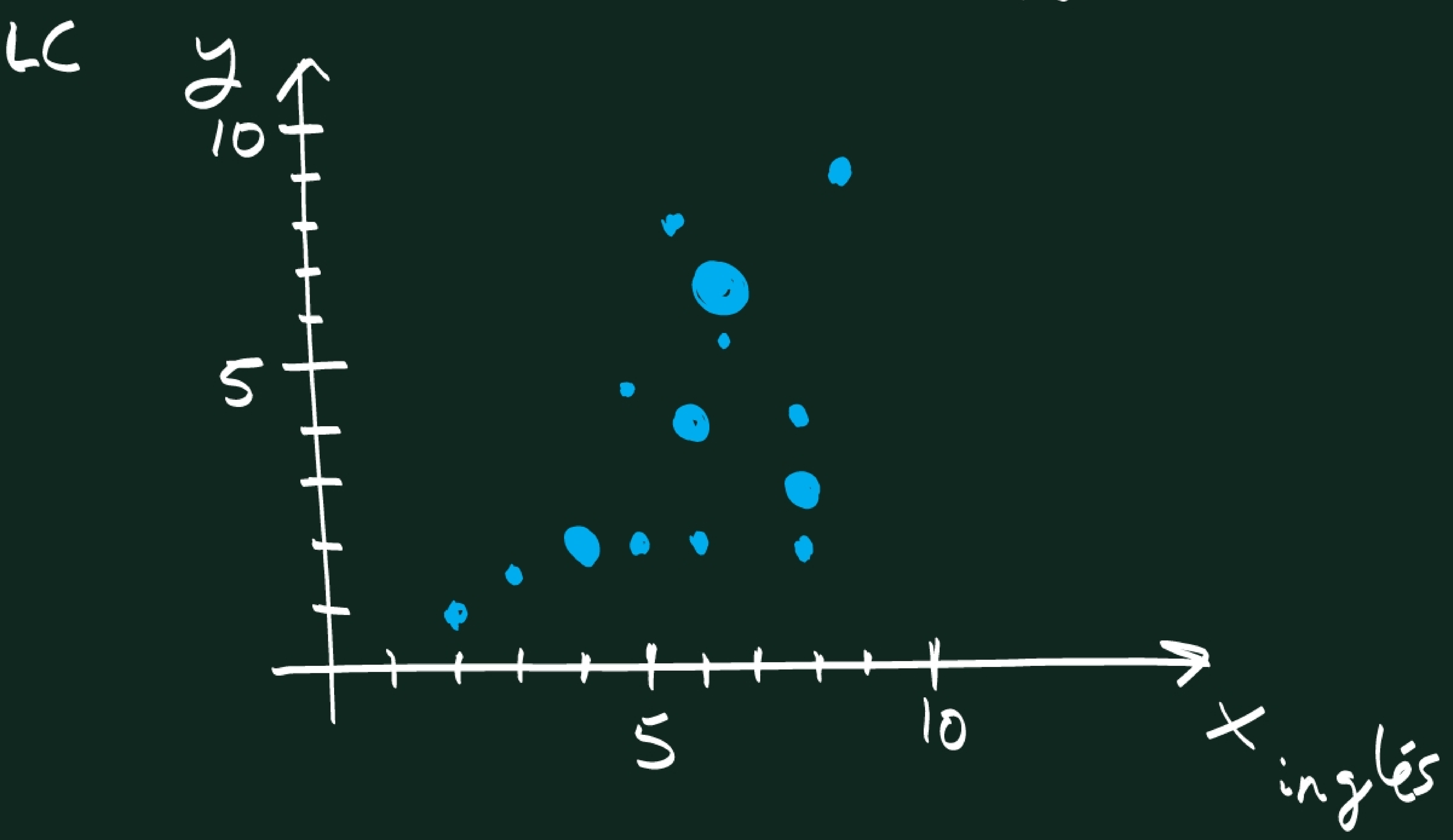
Covarianza:

Varianza conxunta das dúas variables estudadas.

$$S_{xy} = \sigma_{xy} = \frac{\sum x_i \cdot y_j \cdot f_{ij}}{N} - \bar{x} \cdot \bar{y}$$

$\sigma_{xy} > 0$: correlación positiva $\uparrow / \uparrow$

$\sigma_{xy} < 0$: correlación negativa $\uparrow / \downarrow$



- Se as variable son independentes : $\sigma_{xy} = 0$
- Se $\sigma_{xy} = 0$; as variables non teñen por que ser independentes.

Regresión Lineal

$r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ $\rightarrow$ r: coeficiente de correlación lineal.

$0 < r < 1$: correlación directa $\uparrow / \uparrow$

$-1 < r < 0$: correlación inversa $\uparrow / \downarrow$

Canto máis se aproxime "r" a 1 ou -1 máis forte é a correlación.

- Rectas de Regresión Lineal:

De Y sobre X: $y - \bar{y} = \dfrac{\sigma_{xy}}{\sigma_x^2} \cdot (x - \bar{x})$

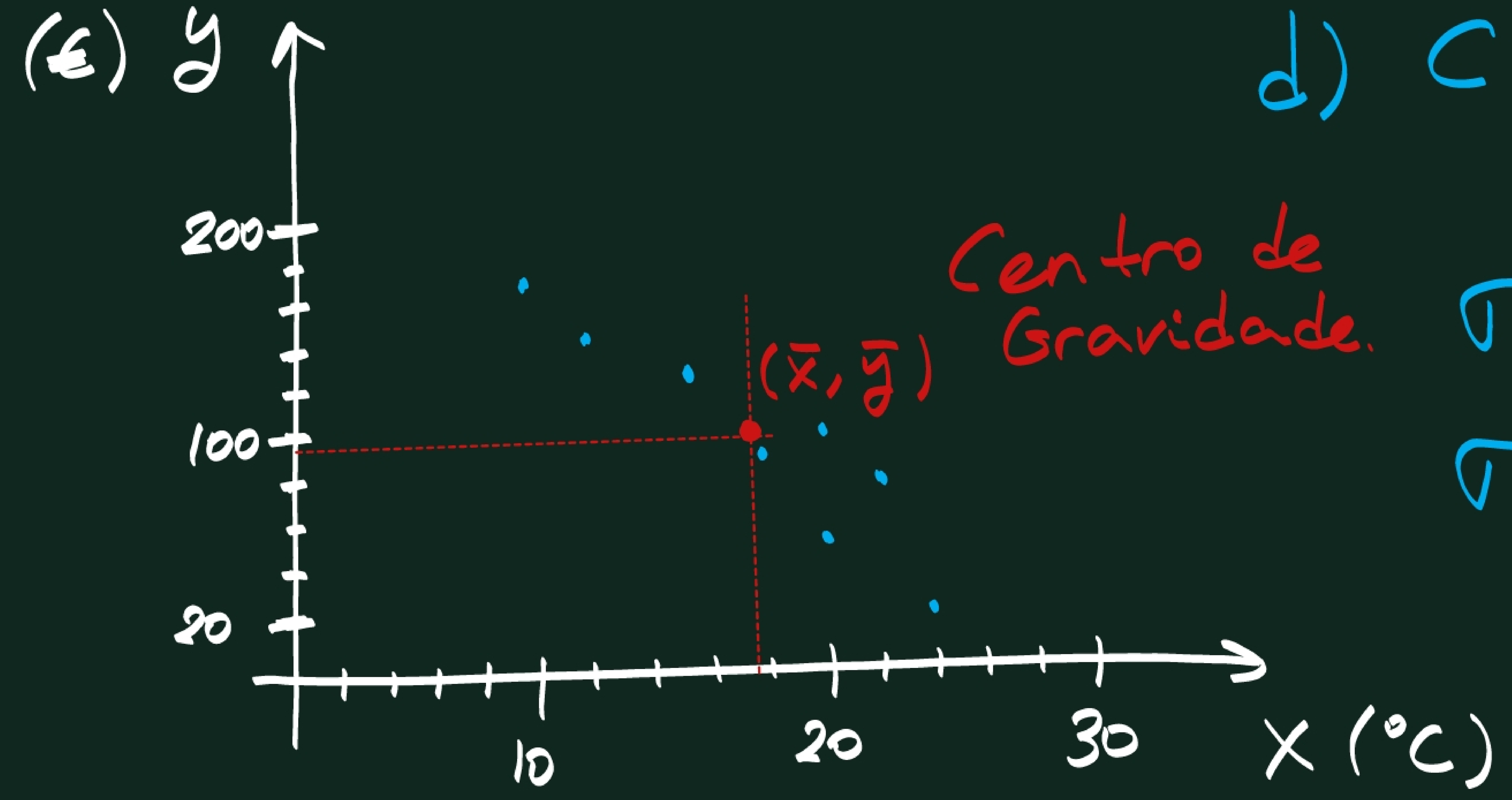
De X sobre Y: $x - \bar{x} = \dfrac{\sigma_{xy}}{\sigma_y^2} \cdot (y - \bar{y})$

$R^2 = \dfrac{\sigma_{xy}}{\sigma_x^2} \cdot \dfrac{\sigma_{xy}}{\sigma_y^2} = \dfrac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = r^2$ $\longrightarrow$ Coeficiente de Determinación

1)

x_i	y_i	x_i^2	y_i^2
22	75	484	5625
12	140	144	19600
24	20	576	400
10	170	100	28900
20	52	400	2704
18	90	324	8100
20	100	400	10000
16	125	256	15625
142	772	2684	90954

a)



b) Hai correlación inversa, e parece forte.

c) $\bar{X} = \frac{142}{8} = 17{,}75\,°C$ $\quad \bar{Y} = \frac{772}{8} = 96{,}5\,€$

d) $CV = \frac{\sigma}{\bar{x}}$

$\sigma = \sqrt{\sigma^2}$

$\sigma^2 = \frac{\sum x_i^2 f_i}{N} - \bar{x}^2$

d)

$\sigma_x^2 = \frac{2684}{8} - 17{,}75^2 =$

$= 20{,}4375$

$\sigma_x = \sqrt{\sigma_x^2} = 4{,}52$

$CV_x = \frac{4{,}52}{17{,}75} = 0{,}2547$

$\sigma_y^2 = \frac{90954}{8} - 96{,}5^2 =$

$= 2057$

$\sigma_y = \sqrt{2057} = 45{,}35$

$CV = \frac{45{,}35}{96{,}5} = 0{,}47$

O gasto en calefacción presenta maior variabilidade.

13

x	y	x_i^2	y_i^2	$x_i \cdot y_i$
2	1	4	1	2
2	2	4	4	4
3	2	9	4	6
3	3	9	9	9
4	2	16	4	8
4	4	16	16	16
4	5	16	25	20
4	6	16	36	24
5	4	25	16	20
5	6	25	36	30
36	35	140	151	139

a) $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot (x - \bar{x})$

$$\sigma_{xy} = \frac{\sum x_i y_i \cancel{f_i}}{N} - \bar{x} \cdot \bar{y}$$