Problema 1

Suponga los siguientes conjuntos de datos:

$$x = \{x_1, \dots, x_9\}, \quad y = \{y_1, \dots, y_9\}$$

Los cuales corresponden a concentraciones de dos minerales en cierto depósito. Se obtuvo que $\bar{x} = 5, \bar{y} = 3, \sum_{i=1}^{9}(x_i - \bar{x})^2 = 350$, y $\sum_{i=1}^{9}(y_i - \bar{y})^2 = 248$. ;

1. Obtenga el coeficiente de correlación entre x y y.
2. Mediante R, obtenga todos los estadísticos mencionados en el problema sabiendo que los datos vienen dados por:

$$\boxed{x = \{0, 5, 0, 10, 0, 15, 0, 15, 0\}, \quad y = \{0, 0, 0, 10, 0, 15, 0, 2, 0\}}$$

Solución:

Recuerdo:

$Cov(\vec{x}, \vec{y})$; $\vec{x}, \vec{y} \in \mathbb{R}^n$; $x_1, \dots, x_n$, $y_1, \dots, y_n$ datos

$$Cov(\vec{x}, \vec{y}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad \boxed{\text{Cov. muestral}}$$

cov(·,·) define un producto interno.

$$(x + y) \bullet z = x \bullet z + y \bullet z$$

$Cov(\alpha x + y, z)$; $\alpha \in \mathbb{R}$

$$= \alpha\, Cov(x, z) + cov(y, z)$$

$$Cov(x, y) = \frac{1}{n-1} \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}_{S_{xy}}$$

$$Cov(x, y) = \frac{1}{n-1} S_{xy}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad ; \quad S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad ; \quad S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\boxed{Cor(x, y) = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n-1} \sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}}$$

$$= \frac{Cov(x, y)}{\sqrt{S_x^2 \cdot S_y^2}}$$

$$= \frac{Cov(x, y)}{S_x \cdot S_y}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y}$$

$\bar{x} = \frac{1}{n} \sum x_i$

$\Rightarrow \sum x_i = n\bar{x}$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + n \bar{x} \bar{y}$$

$$= \sum x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y}$$

$$= \sum x_i y_i - n \bar{x} \bar{y} \quad //$$

$\bar{x} = \dots$

$\bar{y} = \dots$

$\sum (x_i - \dots$

Problema 2

En un esfuerzo por obtener el máximo rendimiento en una reacción química, un experto analiza los valores de las siguientes variables:

- T: Temperatura (en °C)
- P: Porcentaje de material convertido al producto deseado.

Los datos, para una muestra de tamaño 020, se resumen en la siguiente tabla:

	T/P	$40-50$	$50-60$	$60-70$	$70-80$	Total
165	$160-170$	3	1	0	0	4
175	$170-180$	0	3	5	0	8
185	$180-190$	0	2	3	3	8
	Total	3	6	8	3	20

1. Calcule los promedios y varianzas marginales de cada variable. ¿Cuál variable es más homogénea?
2. Calcule el porcentaje promedio de material convertido, dado que la temperatura es superior a 170°C.
3. ¿Existe evidencia de asociación lineal entre las variables?

1.

Temp.	mcT_i	f_i
	165	4
	175	8
	185	8

Porc.	mcP	f_i
	45	3
	55	6
	65	8
	75	3

$$\text{mean}(\text{Temp}) = \frac{1}{20}\sum_{i=1}^{3} mcT_i \cdot f_i$$

$$= \frac{1}{20}\{164\cdot 4 + 175\ 8 + 185\cdot 8\}$$

Temp.	mcT_i	f_i
	165	4
	175	8
	185	8

Porc.	mcP	f_i
	45	3
	55	6
	65	8
	75	3

$$\text{Var}(\text{Temp}) = \frac{1}{n-1}\sum_{i=1}^{3}(x_i - \bar{x})^2 \cdot f_i$$

$$= \frac{1}{19}\left((165-\bar{x})^2\cdot 4 + (175-\bar{x})^2\cdot 8 + (185-\bar{x})^2\cdot 8\right)$$

$$= \dots$$

Viendo los datos en color amarillo tenemos que esos son los que nos sirven, por tanto:

mcP	f_i
45	0
55	5
65	8
75	3

$\rightsquigarrow$ mean $= 63.75$

$\rho \in [-1,1]$

si $\rho \approx 1$

$\Rightarrow$ Relación linear entre X e Y

si $\rho \approx 0$

$\Rightarrow$ No hay c.l.

si $\rho \approx -1$

existe relación inversa a la $\rho\approx 1$

3. Cor

$\Rightarrow$ Cov

Problema 3

Los meteoritos se pueden clasificar de acuerdo a su composición y procedencia; de ahí es que existen aquéllos denominados Condritas (C), Acondritas (A), Metálicos (M) y Siderolíticos (S). Se seleccionó al azar 22 caídas de meteoritos documentadas en el norte chileno y argentino para estudiarlos en detalle, observando en ellos: tipo de meteorito, diámetro (cms.), pero (kg.) y número de fragmentos recuperados: Información adicional; considere D_i, P_i y F_i como el diametro, el peso y cantidad de fragmentos de l i-ésimo

Tipo	C	A	A	C	S	C	A	S	S	M	A
Diámetro	22,44	19,85	17,66	21,25	40,55	33,25	23,15	26,55	36,55	23,65	14,95
Peso	11,6	16,8	14,4	12,3	2,1	6,1	13,5	9,1	3,8	14	18
Fragmentos	2	3	3	2	1	2	2	2	1	1	2
Tipo	A	M	C	M	C	A	A	C	S	M	A
Diámetro	22,9	30,95	24,7	32,15	24,95	29,05	22,95	28,35	36,55	27,45	24,1
Peso	9,1	4,5	7,7	4,8	7,7	9,3	8,9	6,6	4,1	6,6	10
Fragmentos	2	1	2	1	1	3	3	2	1	1	1

meteorito, respectivamente. Luego

$$\sum_{i=1}^{22} D_i = 583{,}95 \quad \sum_{i=1}^{22} (D_i-\bar{D})^2 = 863{,}6107 \quad \sum_{i=1}^{22} (D_i-\bar{D})(P_i-\bar{P}) = -526{,}6715$$
$$\sum_{i=1}^{22} P_i = 201 \quad \sum_{i=1}^{22} (P_i-\bar{P})^2 = 394{,}4709 \quad \sum_{i=1}^{22} (D_i-\bar{D})(F_i-\bar{F}) = -54{,}67409$$
$$\sum_{i=1}^{22} F_i = 39 \quad \sum_{i=1}^{22} (F_i-\bar{F})^2 = 11{,}86364 \quad \sum_{i=1}^{22} (F_i-\bar{F})(P_i-\bar{P}) = 37{,}48182$$

1. Para **cada una de las variables** entregadas, clasifique si son cualitativas o cuantitativas, y discretas o continuas según corresponda.

2. Para la variable 'tipo de meteorito' construya la tabla de frecuencias (con todas las frecuencias admisibles). Calcule una medida de centralidad para la misma variable. ¿Qué se puede decir respecto a la simetría de la distribución de frecuencias construidas?

3. Construya la gráfica boxplot para el Diámetro de los meteoritos y comente la forma de la distribución del Diámetro en base a la gráfica obtenida.

4. Analice la veracidad del siguiente comentario: El Peso de los meteoritos es relativamente menos variable que el Diámetro de los meteoritos. Fundamente su respuesta mediante un indicador apropiado.

Tipo	f_i	F_i	fr_i	Fr_{Ai}
C	6	6	6/22	6/22
A	8	14	8/22	14/22
M	4	18	4/22	18/22
S	4	22	4/22	1

Medida de centralidad.

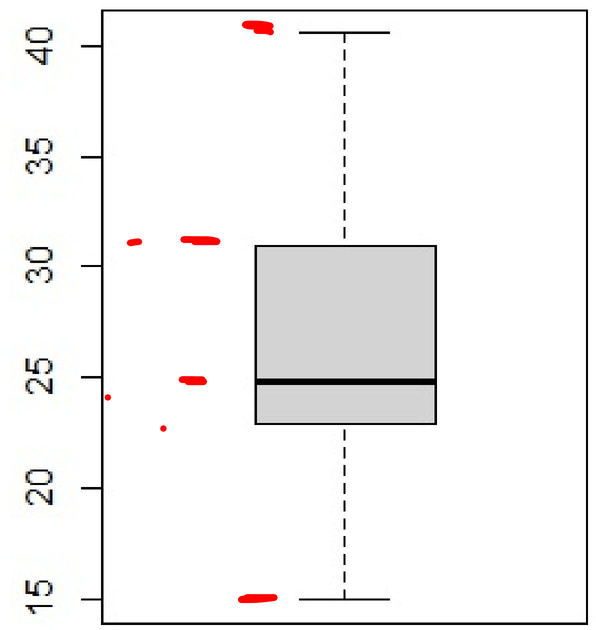
La moda es (A)

3. Obs: $\vec{D}$ vector de datos.

$\boxed{P_{25} \; ; \; P_{75} \; ; \; M_e}$

```
> quantile(diam)
     0%     25%     50%     75%    100%
14.9500 22.9125 24.8250 30.4750 40.5500
```

IQR



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

$CV_{diam} = \frac{S_x}{\overline{diam}}$

$= 0{,}24$

$CV_{peso} = 0{,}47$