Problemas

1.1.

Número de niños	2	6	10	5	10	3	2	2
Número de metros	1	2	3	4	5	6	7	8

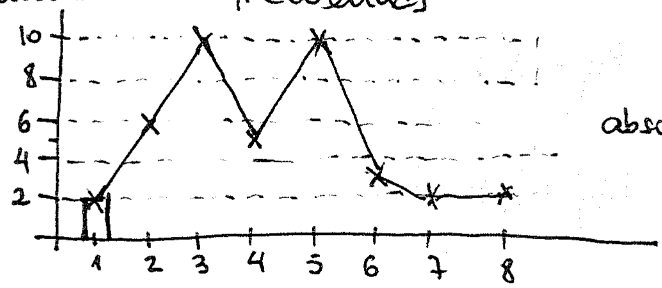
Veamos cuántos niños abarca el estudio

$$N = 2+6+10+5+10+3+2+2 = 40$$

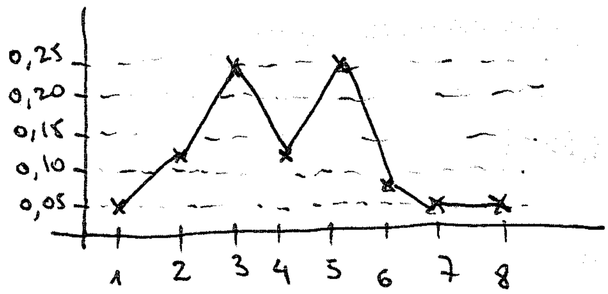
a) ~~Frec relativas~~.

Longitud	1	2	3	4	5	6	7	8
n_i	2	6	10	5	10	3	2	2
f_i	0,05	0,15	0,25	0,125	0,25	0,075	0,05	0,05
F_i	0,05	0,20	0,45	0,575	0,825	0,9	0,95	1,0

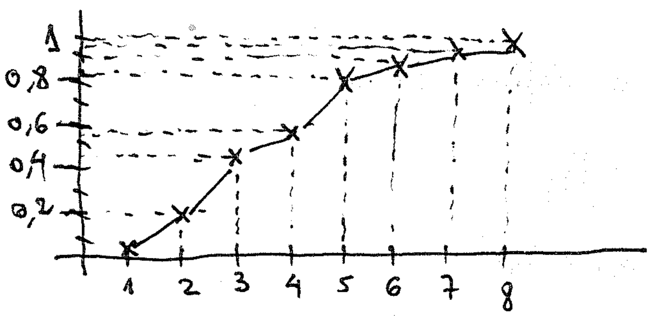
Diagramas de frecuencias



absolutas



relativas



relativas acumuladas

b) Moda, mediana, cuartiles y deciles * (omitimos aquí las unidades pero se entiende que son m)

i) se trata de una distribución bimodal con valores para la variable aleatoria longitud de 3 y 5.

$$Md \equiv \{3,5\}$$

ii) La mediana se puede considerar como aquella que corresponde a datos agrupados, por lo que realizando una interpolación lineal tendríamos

$$0{,}45 + (x-3) \cdot 0{,}125 = 0{,}5$$

$$x = 3 + \frac{0{,}5 - 0{,}45}{0{,}125} = 3{,}4$$

Por lo que $Me = 3{,}4$

iii) Cuartiles $P_{1/4}$ y $P_{3/4}$, observamos que

$$\left.\begin{array}{l} \left.\begin{array}{l} F(2) = 0{,}2 \\ F(3) = 0{,}45 \end{array}\right\} \Rightarrow P_{1/4} = 2{,}2 \\ \left.\begin{array}{l} F(4) = 0{,}575 \\ F(5) = 0{,}825 \end{array}\right\} \Rightarrow P_{3/4} = 4{,}7 \end{array}\right\}$$

iv) Deciles $P_{1/10}$ y $P_{9/10}$, observemos que

$$\left.\begin{array}{l} \left.\begin{array}{l} F(1) = 0{,}05 \\ F(2) = 0{,}2 \end{array}\right\} \Rightarrow P_{1/10} = 1{,}33 \\ F(6) = 0{,}9 \Rightarrow P_{9/10} = 6 \end{array}\right\}$$

c) El rango intercuartílico se define como

$$R_I = P_{3/4} - P_{1/4}$$

$$R_I = 4{,}7 - 2{,}2 = 2{,}5$$

- Se considera un dato atípico leve al que se encuentra a más de $1.5\,R_I$ por encima o por debajo de $P_{3/4}$ y $P_{1/4}$ respectivamente. Un dato atípico extremo es el que está a más de tres veces de estos cuartiles.

- Como $P_{1/4} = 2{,}2$ no hay datos inferiores a este cuartil con distancia superior a R_I. Por otro parte los datos de mayor longitud recorrida se corresponden al valor 8 que cumple

$$8 - P_{3/4} = 8 - 4{,}7 = 3{,}3 = 1.32 \cdot R_I$$

por lo tanto tampoco encontramos datos atípicos leves ni extremos.

- En caso de que los encontraramos, estos sucesos serían revisados para evaluar su naturaleza (error, fluctuación estadística, etc). Es importante advertir que no podemos rechazar de modo automático los datos atípicos.

d) Media aritmética, geométrica, cuadrática y armónica

$$\bar{x}_a = \sum_{i=1}^{8} f_i \cdot x_i = 4.05$$

$$\bar{x}_g = \left[\prod_{i=1}^{8} x_i^{n_i}\right]^{1/N} \qquad \log(\bar{x}_g) = \sum_{i=1}^{8} f_i \cdot \log(x_i)$$

$$\bar{x}_g = 3.6325$$

$$\bar{x}_q = \sqrt{\sum_{i=1}^{8} f_i \cdot x_i^2} = 4.416$$

$$\bar{x}_a = \frac{1}{\sum_{i=1}^{8} \frac{f_i}{x_i}} = 3.17$$

e) Analizar la dispersión de la distribución: varianza, desviación típica y coeficiente de variación de Pearson.

$$s^2 = \sum_{i=1}^{8} f_i (x_i - \bar{x})^2 = 3.0975$$

$$s = 1.76$$

$$CV = \frac{s}{|\bar{x}|} = \frac{1.76}{4.05} = 0.4346$$

f) Momentos respecto al origen de 1er, 2° y 3er orden

$$m_1(0) = \sum_{i=1}^{8} f_i \cdot x_i = \bar{x} = 4{,}05$$

$$m_2(0) = \sum_{i=1}^{8} f_i \, x_i^2 = 19{,}5$$

$$m_3(0) = \sum_{i=1}^{8} f_i \, x_i^3 = 106{,}2$$

g) Momentos centrales de orden 1° y 3°

$$m_1(\bar{x}) = \sum_{i=1}^{8} f_i (x_i - \bar{x}) = 0$$

$$m_2(\bar{x}) = s^2 = 3.0975$$

$$m_3(\bar{x}) = \sum_{i=1}^{8} f_i (x_i - \bar{x})^3 = 2.1353$$

$$m_2(\bar{x}) = m_2(0) - [m_1(0)]^2 = 19.5 - (4.05)^2 = 3.0975$$

$$m_3(\bar{x}) = m_3(0) - 3\, m_2(0) \cdot m_1(0) + 2[m_1(0)]^3 = 2.1353$$

h) Estúdiese la asimetría y curtosis

$$A_P = \frac{\bar{x} - Md}{s}$$ ⇒ no se puede aplicar a multimodales

$$A_{BY} = \frac{P_{3/4} + P_{1/4} - 2Me}{P_{3/4} - P_{1/4}} = \frac{4{,}7 + 2{,}2 - 2 \cdot 3{,}4}{4{,}7 - 2{,}2} = 0.04$$

$$A_F = \frac{m_3(\bar{x})}{s^3} = \frac{1}{s^3} \sum_{i=1}^{8} f_i (x_i - \bar{x})^3$$

$$A_F = 0.3917$$

- Por tanto, ésta es una distribución con una cierta asimetría positiva.

- La curtosis está definida como

$$g = \frac{m_4(\bar{x})}{s^4} = \frac{1}{s^4} \sum_{i=1}^{8} f_i (x_i - \bar{x})^4 = 2.5563$$

Se trata de una distribución con $g<3$ por lo tanto es platicúrtica.

1.2. En un determinado experimento se mide la concentración de una sustancia en mmol/l obteniéndose 20 medidas entre 38 y 67.
Considerando el nº total de medidas consideraremos unas 5 clases de agrupamiento

a)
$[38, 44) \longrightarrow \{38, 41\}$
$[44, 50) \longrightarrow \{49, 47, 46, 46\}$
$[50, 56) \longrightarrow \{52, 51, 50, 52, 55, 54\}$
$[56, 62) \longrightarrow \{61, 59, 57, 56, 57\}$
$[62, 68) \longrightarrow \{63, 67, 65\}$

Intervalo	x	ni	fi	Ni	Fi
[38, 44)	41	2	0.1	2	0.1
[44, 50)	47	4	0.2	6	0.3
[50, 56)	53	6	0.3	12	0.6
[56, 62)	59	5	0.25	17	0.85
[62, 68)	65	3	0.15	20	1

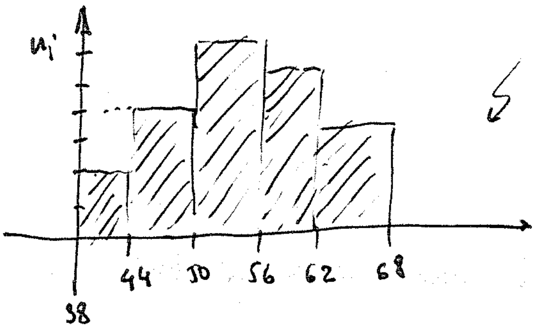


Diagrama de frecuencias absolutas.

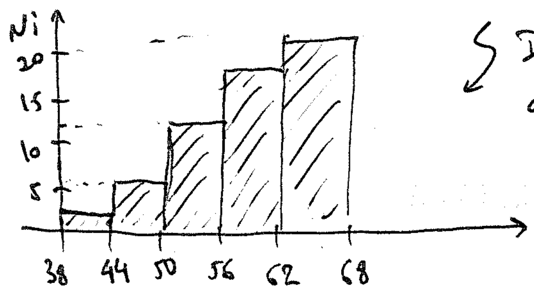


Diagrama de frecuencias absolutas acumuladas

b) $\bar{x} = \sum_{i=1}^{5} f_i x_i = 53.9$

$\bar{x}_g = \left[\prod_{i=1}^{8} x_i^{n_i}\right]^{1/5} = 53.411$

$s^2 = \sum_{i=1}^{5} f_i (x_i - \bar{x})^2 = 51.39 \qquad s = 7.1687$

$CV = \frac{s}{|\bar{x}|} = 0.133$

$A_p = \frac{\bar{x} - Md}{s} = \frac{53.9 - 53}{7.1687} = 0.1255$

$A_F = \frac{m_3(\bar{x})}{s^3} = -0.1148$ ligera asimetría negativa

$g = \frac{m_4(\bar{x})}{s^4} = 2.1466$ platicúrtica

Normalmente habría que considerar

$\bar{x} \pm s(\bar{x}) \Rightarrow 53.9 \pm \frac{7.17}{\sqrt{20}}$

53.9 ± 1.6

P4

Si no agrupamos los datos podremos calcular

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{20} x_i = 53.3$$

$$s^2 = \frac{1}{N} \sum_{i=1}^{20} (x_i - \bar{x})^2 = 55.91 \quad ; \quad s = 7.48$$

c) Tendremos que calcular $P^{1/4}$ y $P^{3/4}$ considerando las marcas de clase.

$N/4 = 5 \qquad 3/4\, N = 15$

$R_I = 11.1$

$P^{1/4} = 45.5 \qquad P^{3/4} = 56.6$

Datos atípicos $1.5 * R_I = 16.65$

$$\left.\begin{array}{l} P^{1/4} - 1.5\, R_I = 28.85 \\ P^{3/4} + 1.5\, R_I = 73.25 \end{array}\right|$$

- No encontramos datos atípicos.

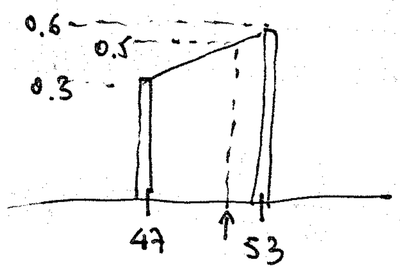
Mediana. Si la calculamos de manera individual

$\{38, 41, 46, 46, 47, 49, 50, 51, 52, 52, 54, 55, 56, 57, 57, 59, 61, 63, 65, 67\}$

$53 \equiv M_e$

Si la calculamos a partir de las frecuencias relativas acumuladas,

$M_e^* = 51$



1.6 Si tomamos una variable Z mezclando n_1 valores de la variable x y n_2 de la variable Y, la media es

$$\bar{z} = \frac{n_1}{n_1+n_2}\bar{x} + \frac{n_2}{n_1+n_2}\bar{y}$$

$$\bar{z} = \sum_{i=1}^{n_1+n_2} z_i \frac{1}{n_1+n_2} = \frac{1}{n_1+n_2}\left[\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i\right]$$

$$\bar{z} = \frac{1}{n_1+n_2}\left[n_1\bar{x} + n_2\bar{y}\right]$$

Y si consideramos la varianza?

$$s^2(z) = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}(z_i-\bar{z})^2 =$$

$$= \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}\left[z_i - \frac{n_1}{n_1+n_2}\bar{x} - \frac{n_2}{n_1+n_2}\bar{y}\right]^2 =$$

$$= \frac{1}{n_1+n_2}\left\{\sum_{i=1}^{n_1}\left(x_i - \frac{n_1}{n_1+n_2}\bar{x} - \frac{n_2}{n_1+n_2}\bar{y}\right)^2 + \right.$$

$$\left. + \sum_{i=1}^{n_2}\left(y_i - \frac{n_1}{n_1+n_2}\bar{x} - \frac{n_2}{n_1+n_2}\bar{y}\right)^2\right\}$$

1.7 Consideremos la distribución de datos

x	1	2	3	4	5
y	0.34	0.70	1.08	1.43	1.70

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$\bar{x} = 3$ $\quad \bar{y} = 1.05$

$s_x^2 = 2$ $\quad s_y^2 = 0.2389$

$\mathrm{cov}(x,y) = 0.69$

(1.8) $s^2 = s_x^2 = s_y^2$

$$\text{cov}\left[(x+y),(x-y)\right] = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i+y_i)-(\bar{x}+\bar{y})\right]\left[(x_i-y_i)-(\bar{x}-\bar{y})\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\bar{x})+(y_i-\bar{y})\right]\left[(x_i-\bar{x})-(y_i-\bar{y})\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\bar{x})^2-(y_i-\bar{y})^2\right] = s_x^2 - s_y^2 = 0$$

(1.9) Probar que si se obtiene una variable q como función de dos variables aleatorias si $(x-\bar{x}), (y-\bar{y})$ es lo bastante pequeño, entonces

$$s_q^2 = \left(\frac{\partial q}{\partial x}\right)^2_{\bar{x},\bar{y}} \cdot s_x^2 + \left(\frac{\partial q}{\partial y}\right)^2_{\bar{x},\bar{y}} s_y^2 + 2\left(\frac{\partial q}{\partial x}\right)_{\bar{x}\bar{y}}\left(\frac{\partial q}{\partial y}\right)_{\bar{x}\bar{y}} \text{cov}(x,y)$$

A primer orden en el desarrollo en serie de Taylor podremos escribir

$$q(x,y) - q(\bar{x},\bar{y}) = \left(\frac{\partial q}{\partial x}\right)\cdot(x-\bar{x}) + \left(\frac{\partial q}{\partial y}\right)(y-\bar{y}) + O^2(x-\bar{x}, y-\bar{y})$$

considerando el valor medio

$$\bar{q} = \sum_{i=1}^{k} f_i\, q(x_i,y_i) = \sum_{i=1}^{k} f_i\left[q(\bar{x},\bar{y}) + \left(\frac{\partial q}{\partial x}\right)(x_i-\bar{x}) + \left(\frac{\partial q}{\partial y}\right)(y_i-\bar{y}) + \ldots\right]$$

$$\bar{q} = q(\bar{x},\bar{y}) + \left(\frac{\partial q}{\partial x}\right)\underbrace{\sum_{i=1}^{k} f_i(x_i-\bar{x})}_{\simeq 0} + \left(\frac{\partial q}{\partial y}\right)\underbrace{\sum_{i=1}^{k} f_i(y_i-\bar{y})}_{\simeq 0} + \ldots$$

$$\bar{q} \simeq q(\bar{x},\bar{y})$$

Por lo tanto

$$s_q^2 = \sum_{i=1}^{K} f_i \left[q(x_i,y_i) - \bar{q}\right]^2 \simeq$$

$$\simeq \sum_{i=1}^{K} f_i \left[q(x_i,y_i) - q(\bar{x},\bar{y})\right]^2 \simeq$$

$$\simeq \sum_{i=1}^{K} f_i \left[\left(\frac{\partial q}{\partial x}\right)(x_i-\bar{x}) + \left(\frac{\partial q}{\partial y}\right)(y_i-\bar{y})\right]^2$$

Por ello

$$s_q^2 \simeq \sum_{i=1}^{K} f_i \left[\left(\frac{\partial q}{\partial x}\right)^2 (x_i-\bar{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 (y_i-\bar{y})^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)(x_i-\bar{x})(y_i-\bar{y})\right]$$

$$s_q^2 \simeq \left(\frac{\partial q}{\partial x}\right)^2 \sum_{i=1}^{K} f_i (x_i-\bar{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sum_{i=1}^{K} f_i (y_i-\bar{y})^2 +$$

$$+ 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right) \cdot \sum_{i=1}^{K} f_i (x_i-\bar{x})(y_i-\bar{y}) =$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 s_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right) \mathrm{cov}(x,y)$$

1.10 consideramos la función $y = kx \quad k \in \mathbb{Z}^+$

a) $\bar{y} = k \cdot \bar{x}$

$$s_y^2 = \sum_{i=1}^{P} f_i (y_i-\bar{y})^2 = \sum_{i=1}^{P} f_i (k \cdot x_i - k\bar{x})^2 =$$

$$= k^2 \sum_{i=1}^{P} f_i (x_i-\bar{x})^2 = k^2 \cdot s_x^2$$

$s_y = k \cdot s_x \Rightarrow$ ¿Cómo es la función de distribución de la y? ¡Es la misma que la de x!

si calculamos la

$$\operatorname{cov}\left[(x+y),(x-y)\right] = \operatorname{cov}\left[(1+k)x,(1-k)x\right] =$$

$$= \sum_{i=1}^{p} (1+k)(x_i-\bar{x})(1-k)(x_i-\bar{x}) = (1-k^2)s_x^2$$

Igualmente según lo que hemos visto antes

$$\operatorname{cov}\left[(x+y),(x-y)\right] = s_x^2 - s_y^2 = s_x^2 - k^2 s_x^2 = (1-k^2)s_x^2$$

b) En cambio si obtenemos y como suma de k valores de la misma variable aleatoria tomados al azar

$$y = \sum_{i=1}^{k} x_i$$

se obtiene fácilmente que

$$\bar{y} = \sum_{i=1}^{k} \bar{x}_i = k \cdot \bar{x}$$ ya que las distribuciones son iguales

si calculamos la varianza,

$$s_y^2 = \sum_{i=1}^{p} f_i\left[y_i-\bar{y}\right]^2 = \sum_{i=1}^{p} f_i\left[\sum_{j=1}^{k} x_j^i - k\bar{x}\right]^2 =$$

$$= \sum_{i=1}^{p} f_i\left[\sum_{j=1}^{k}(x_j^i-\bar{x})\right]^2 \equiv$$

$$\equiv \sum_{i=1}^{p} f_i\left[\sum_{j=1}^{k}(x_j^i-\bar{x})^2 + 2\sum_{l<m}(x_l^i-\bar{x})(x_m^i-\bar{x})\right]$$

$$\equiv \sum_{j=1}^{k}\sum_{i=1}^{p} f_i(x_j^i-\bar{x})^2 + 2\sum_{l<m}\underbrace{\sum_{i=1}^{p} f_i(x_l^i-\bar{x})(x_m^i-\bar{x})}$$

se entenderá mejor si realizamos n sorteos para la variable y,

$$\left.\begin{aligned} y_1 &= x_1^1 + x_2^1 + \dots + x_k^1 \\ y_2 &= x_1^2 + x_2^2 + \dots + x_k^2 \\ &\vdots \\ y_n &= x_1^n + x_2^n + \dots + x_k^n \end{aligned}\right\}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} x_j^i = \sum_{j=1}^{k}\underbrace{\left[\frac{1}{n}\sum_{i=1}^{n} x_j^i\right]}_{\bar{x}} = k\cdot\bar{x}$$

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{k} x_j^i - k\cdot\bar{x}\right]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{k}(x_j^i - \bar{x})\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{k}(x_j^i-\bar{x})^2 + \right.$$

$$\left. + 2\sum_{l<m}(x_l^i - \bar{x})(x_m^i - \bar{x})\right] =$$

$$= \sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(x_j^i - \bar{x})^2 + 2\underbrace{\sum_{l<m}\frac{1}{n}\sum_{i=1}^{n}(x_l^i - \bar{x})(x_m^i - \bar{x})}_{=0} =$$

$$= k\cdot s_x^2$$

¡La distribución de probabilidad para y no es la misma que la distribución de probabilidad para x!

$$\operatorname{cov}\left[(x_i+y),(x_i-y)\right] = s_{x_i}^2 - s_y^2 = s_x^2 - k\, s_x^2 = (1-k)s_x^2$$

1.12

Dada la variable aleatoria bidimensional

X \ Y	1	2	4	6	n_{x_i}
1	2	0	0	1	3
3	3	1	0	1	5
5	0	1	0	5	6
n_{y_j}	5	2	0	7	

i) $\sum_{i=1}^{3}\sum_{j=1}^{4} n_{ij} = (2+0+0+1)+(3+1+0+1)+(0+1+0+5) = 14$

$= \sum_{i=1}^{3} n_{x_i} = 3+5+6 =$

$= \sum_{j=1}^{4} n_{y_j} = 5+2+0+7$

ii) Las frecuencias relativas $f_{ij} = \frac{n_{ij}}{n}$

X \ Y	1	2	4	6	f_{x_i}
1	$1/7$	0	0	$1/14$	$3/14$
3	$3/14$	$1/14$	0	$1/14$	$5/14$
5	0	$1/14$	0	$5/14$	$3/7$
f_{y_j}	$5/14$	$1/7$	0	$1/2$	

iii) La media marginal $\bar{X}$ se obtiene como

$\bar{X} = \sum_{i=1}^{3} f_{x_i} \cdot x_i = 1 \cdot 3/14 + 3 \cdot 5/14 + 5 \cdot 3/7 = 3.4286$

$\bar{Y} = \sum_{j=1}^{4} f_{y_j}\, y_j = 1 \cdot 5/14 + 2 \cdot 1/7 + 4 \cdot 0 + 6 \cdot 1/2 = 3.6429$

iv)

$$m_{r,s}(c,d) = \sum_{i,j} f_{ij}\,(x_i-c)^r\,(y_j-d)^s$$

$$M_{1,0}(0,0) = \sum_{i=1}^{3}\sum_{j=1}^{4} f_{ij}\, x_i^1 \cdot y_j^0 = \sum_{i=1}^{3} x_i\, f_{x_i} = 3.4286$$

$$M_{0,1}(0,0) = \sum_{i=1}^{3}\sum_{j=1}^{4} f_{ij}\, x_i^0\, y_j^1 = \sum_{j=1}^{4} y_j\, f_{y_j} = 3.6429$$

$$M_{1,1}(\bar{x},\bar{y}) = \mathrm{cov}(x,y) = \sum_{i=1}^{3}\sum_{j=1}^{4} f_{ij}\,(x_i-\bar{x})(y_j-\bar{y})$$

$$\begin{aligned}\mathrm{cov}(x,y) = {} & \tfrac{1}{7}(1-3.4286)\cdot(1-3.6429) + \\ & + \tfrac{1}{14}(1-3.4286)\cdot(6-3.6429) + \\ & + \tfrac{3}{14}(3-3.4286)\cdot(1-3.6429) + \\ & + \tfrac{1}{14}(3-3.4286)\cdot(2-3.6429) + \\ & + \tfrac{1}{14}(3-3.4286)\cdot(6-3.6429) + \\ & + \tfrac{1}{14}\cdot(5-3.4286)\cdot(2-3.6429) + \\ & + \tfrac{5}{14}(5-3.4286)\cdot(6-3.6429) = 1.8673\end{aligned}$$

La covarianza nos da una idea del grado de correlación estadística de las variables, como se observa en este caso la covarianza es positiva.