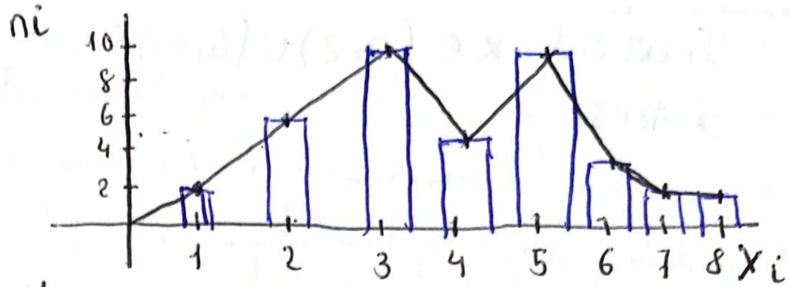
Problemas: T1.

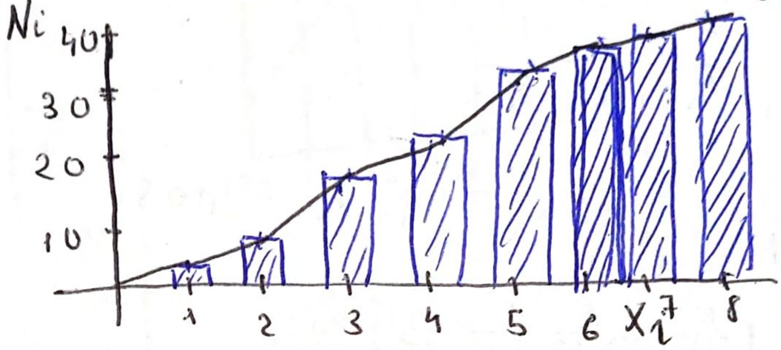
Nº niños : 2 6 10 5 10 3 2 2
Nº metros : 1 2 3 4 5 6 7 8

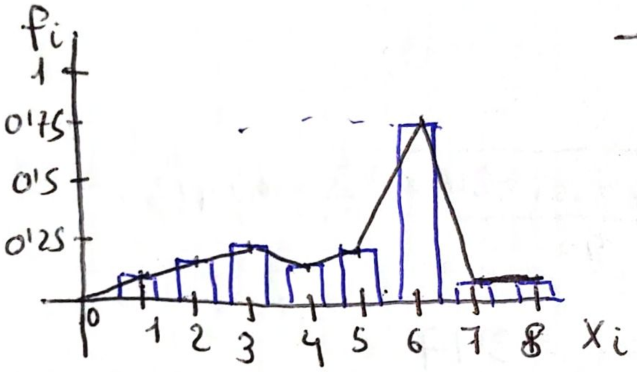
a)

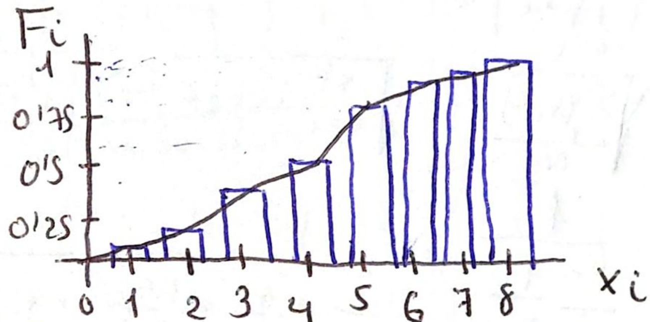
X	n	f	N	F
1	2	0'05	2	0'05
2	6	0'15	8	0'20
3	10	0'25	18	0'45
4	5	0'125	23	0'575
5	10	0'25	33	0'825
6	3	0'075	36	0'9
7	2	0'05	38	0'95
8	2	0'05	40	1

$\sum_{i=1}^{k} n_i = N = 40$









b) Moda es una d. bimodal, $Md = \{3, 5\}$

Med $N/2 = 20$, el valor no corresponde a la tabla N_i, por lo que tomamos $N_i = 23$
$x_i = 4$

Si consideramos x_i una var. continua, interpolamos entre $x=3$ y $x=4$.

$$\frac{N_{i+1} - N_i}{x_{i+1} - x_i} = \frac{qN - N_i}{P_q - x_i} \rightarrow \frac{23-18}{4-3} = \frac{20-18}{P_{0'5} - 3} \Rightarrow P_{0'5} = 3'4.$$

Cuartiles, $P_{0'25}\ (N/4 = 10) = 3$ $\quad P_{0'75}\ (N \cdot 3/4 = 30) = 5$ como d. discreta

Como agrup. de datos, $P_{0'25}(N/4 = 10) = 2'2$. $\quad P_{0'75}\ (N \cdot 3/4 = 30) = 4'7$

Deciles, $P_{0'1}\ (N/10 = 4) = 2$ . . . D. discretas

Agrup. de datos. $P_{0'1}\ (N/10 = 4) \Rightarrow \frac{8-2}{2-1} = \frac{4-2}{P_{0'1} - 1} \Rightarrow P_{0'1} = 1'5$.

c) $RI = P_{0'75} - P_{0'25}$ Discreto, $RI = 2$, Agrupado, $RI = 2'5$.

dato atípico leve $1'5\,RI$ en $P_{0'75}$ o $P_{0'25} - 1'5\,RI$

$P_{0'75} + 1'5\,RI = 5 + 3 = 8$

$P_{0'25} - 1'5\,RI = 0$

No tenemos datos atípicos leves, ya que ninguno $x > 0$ ni $x < 0$. Evidentemente tampoco existen datos atípicos extremos.

En caso de aparecer no deben eliminarse, debemos estudiar su origen si se presentan. Errores de medida, reevaluar incertidumbres, ...

d) $\overline{x_g}$, $\overline{x}$, $\overline{x_q}$, $\overline{x_a}$.

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \sum_{i=1}^{k} f_i x_i = 4'05.$$

$$\overline{x}_g = \left(\prod_{i=1}^{N} x_i^{n_i}\right)^{\frac{1}{N}} = 3'76.$$

$$\overline{x_q} = \sqrt{\frac{\sum_{i=1}^{N}(x_i^2)}{N}} = \sqrt{\sum_{i=1}^{k} f_i x_i^2} = \sqrt{\frac{1^2 \cdot 2 + 2^2 \cdot 6 + 3^2 \cdot 10 \dots}{40}} = 4'416.$$

$$\overline{x_a} = \frac{1}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{1}{\frac{2}{40} + \frac{6}{40 \cdot 2} + \frac{10}{40 \cdot 3} + \dots + \frac{2}{40 \cdot 6}} = 3'17.$$

$$\overline{x_a} \leq \overline{x_g} \leq \overline{x} \leq \overline{x_q}$$

e) Dispersión

$\longrightarrow$ Varianza. $s^2 = \sum_{i=1}^{k} f_i (x_i - \overline{x})^2 = 0'05(1-4'05)^2 + 0'15(2-4'05)^2 \dots = \underline{3'09}$

$\longrightarrow$ Desv. típica $s = \sqrt{s^2} = 1'76$

Coef. Var. Pearson $CV = \frac{s}{|\overline{x}|} = 0'43.$

f) $m_1(0) = \sum_{i=1}^{n} f_i (x_i - 0)^1 = \sum_{i=1}^{n} f_i x_i = \overline{x} = 4'05$

$m_2(0) = \sum_{i=1}^{k} f_i (x_i - 0)^2 = \sum_{i=1}^{k} f_i (x_i)^2 = 19'5$

$m_3(0) = \sum_{i=1}^{n} f_i (x_i - 0)^3 = \sum_{i=1}^{k} f_i (x_i)^3 = 106'2.$

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

$m_2(\bar{x}) = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 3'0975$

$m_3(\bar{x}) = \sum_{i=1}^{k} f_i (x_i - \bar{x})^3 = 2'1353$

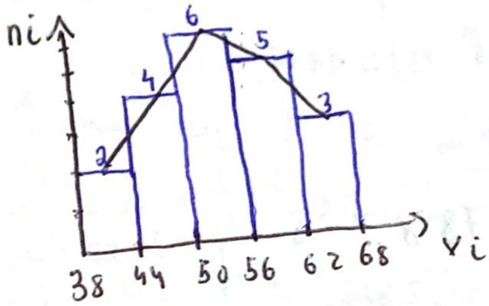
h) $A_F = \frac{1}{S^3} \sum_{i=1}^{k} f_i (x_i - \bar{x})^3 = \frac{m_3(\bar{x})}{S^3} = 0'3917$

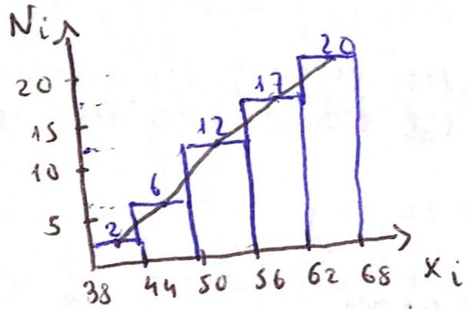
$g = \beta_2 = \frac{1}{S^4} \sum_{i=1}^{k} f (x_i - \bar{x})^4 = \frac{m_4(\bar{x})}{S^4} = 2'5563.$

1.2) a) $\sum dat = 545$ div. en 5 m. de clase $\frac{max - min}{n^o \text{ marcas clase}}$ = long. interv.

Intervalo	$\bar{x}$	n_i	f_i	N_i	F_i
[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

$\sum_{i=0}^{N} n_i = 20$





b) $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \sum_{i=1}^{k} f_i x_i = 53'9$

$\bar{x}_g = \left(\prod_{i=1}^{k} x_i^{n_i} \right)^{\frac{1}{N}} = 53'411.$

$S^2 = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 51'39$ = varianza = $m_2(\bar{x})$

$S = \sqrt{S} = 7'16$ = desviación típica

Coef. var. Pearson = $CV = \frac{S}{|\bar{x}|} = 0'13$

$A_F = \frac{1}{S^3} \left(\sum_{i=1}^{k} f_i (x_i - \bar{x})^3 \right) = \frac{m_3(\bar{x})}{S^3} = -0'1148$ lo que nos muestra el coef. de asim. de Pearson-Fisher es una asim. negativa

$g = \beta_2 = \frac{1}{S^4} \sum_{i=1}^{k} f_i (x_i - \bar{x})^4 = \frac{m_4(\bar{x})}{S^4} = 2'1466 < 3 \Rightarrow$ es una distrib platicúrtica.

La concentración del exp. es $\bar{x} = 53'09$ $S_A(\bar{x}) = \frac{S}{\sqrt{n}} = \frac{7'16}{\sqrt{20}} = 1'6.$

Si no se agrupasen los datos.

$\bar{x} = 53'3$ $S_A(\bar{x}) = 1'67$

(1.2) c) Existen dat. atípicos. ¿Qué medida de centralización es más resistente a estos datos?

↓

2 a mediana.

$P_{0'25}\ (N/4 = 5) \Rightarrow \frac{6-2}{47-41} = \frac{5-2}{P_{0'25}-41} \Rightarrow P_{0'25} = \underline{45'5}$

$Md = P_{0'5} = 51$ (53 si consideramos incluir los datos)

$P_{0'75} = (3N/4 = 15) \Rightarrow \frac{17-12}{59-53} = \frac{15-12}{P_{0'75}-53} \Rightarrow P_{0'75} = 56'6$

$RI = (P_{0'75} - P_{0'25}) = 11'1.$

Los dat atípicos son los que superan $P_{0'75} + 1'5\,RI$ o son a $P_{0'25} - 1'5\,RI$

$P_{0'75} + 1'5 \cdot RI = 73'25$

$P_{0'25} - 1'5\,RI = 28'85$ } En nuestro caso los datos no son atípicos.

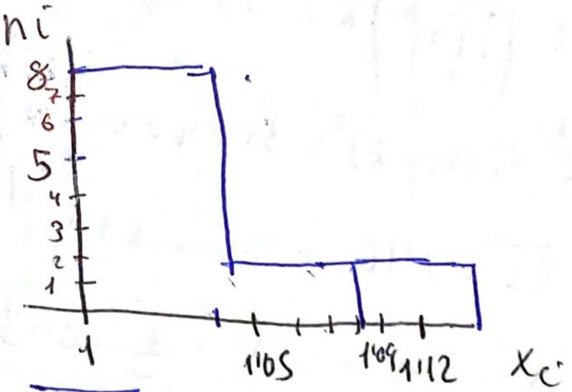
(1.3) m(g) 1'1 1'2 1'1 0'9 1'0 1'0 0'9 1'1 1'0 0'9

V(cm3) 1'1 1'1 1'1 0'9 1'0 1'0 0'8 1'1 1'0 0'9.

a) Histograma de $\rho = \frac{m}{V}$.

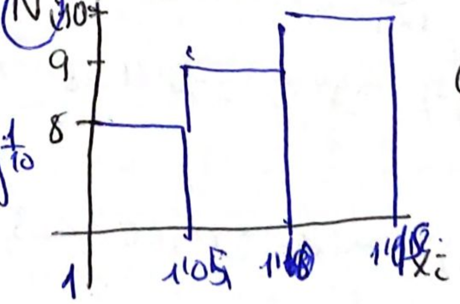
$\rho = \{(1), (1'09), (1), (1), (1), (1), (1'125), (1), (1), (1)\}$

(ρ) X	ni	fi	Ni	Fi
1	8	0'8	8	0'8
1'09	1	0'1	9	0'9
1'125	1	0'1	10	1



$\Sigma n = 10$

b) $\bar{x} = \sum_{i=1}^{k} f_i x_i = 1'0215$



$\bar{x}_g = \left(\prod_{i=1}^{k} x_i^{n_i}\right)^{\frac{1}{N}} = \left(1^8 \cdot 1'09 \cdot 1'125\right)^{\frac{1}{10}} = \underline{1'01}$

Concentración de densidad

$\bar{x} = 1'0215$

$S_A(\bar{x}) = \frac{0'043}{\sqrt{10}} = 0'014$

$s^2 = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 0'00191.$

$AF = \frac{1}{s^3} \sum_{i=1}^{k} f_i (x_i - \bar{x})^3 = 1'586$

Pearson-Fisher > 0, asimetría positiva

$s = \sqrt{s^2} = 0'043$

$CV = \frac{s}{|\bar{x}|} = \frac{0'043}{1'0215} = 0'0427$

$g = \beta_2 = \frac{1}{s^4} \sum_{i=1}^{k} f_i (x_i - \bar{x})^4 = 3'69$

$g > 3$, coef de apuntamiento nos dice que la distrib es leptocúrtica.

1.4) estaturas (cm): 160'0 172'4 168'0 167'0 175'0 179'0 180'0 198'0
164'0 166'0 174'0 177'0 182'5 185'0 191'0 173'5.

a) Es una variable aparentemente continua, donde los datos están influenciados por la precisión del instrumento de medida. Como tenemos 16 datos, el número de grupos óptimos será $\sqrt{16} = 4$ grupos / marcas de clase.

Valor min = 160'0   Valor max = 198'0. $\frac{198'0 - 160'0}{4} = \frac{38}{4} = 9'5$

Intervalo	x → marca de clase	n	f_i
[160, 170)	165	5	0'312
[170, 180)	175	6	0'375
[180, 190)	185	3	0'1875
[190, 200)	195	2	0'125

$\sum n_i = 16 = N$

b) $\bar{x}$, $\bar{x}_g$, $\bar{x}_a$

$$\bar{x} = \sum_{i=1}^{k} f_i x_i = \frac{1}{N} \sum_{i=1}^{k} n_i x_i = 176'25.$$

$$\bar{x}_g = \sqrt[16]{\prod_{i=1}^{k} x_i^{n_i}} = \left(165^5 \cdot 175^6 \cdot 185^3 \cdot 195^2\right)^{1/16} = 175'975$$

$$\bar{x}_a = \frac{1}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = 175'705.$$

$$\bar{x}_a \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_q$$

c) Med y DMe.

Med = $P_{0'5}$

$$\frac{N_{i+1} - N_i}{x_{i+1} - x_i} = \frac{pN - N_i}{P_{0'5} - x_i}$$

$$\frac{11 - 5}{175 - 165} = \frac{0'5(16) - 5}{P_{0'5} - 165}.$$

$\underline{P_{0'5} = 170}$

$$D_{Me} = \sum_{i=1}^{k} f_i (x_i - Me) = 9'375$$

d) Coef de var. media.

$$CV_{MMe} = \frac{DM_{Me}}{|Me|} = \frac{9'375}{170} = 0'055$$

e) $CV = \frac{S}{|\bar{x}|} = 0'058.$

$$A_F = \frac{1}{S^3} \sum_{i=1}^{k} f_i (x_i - \bar{x})^3 = \frac{m_3(\bar{x})}{S^3} = -0'516 < 0$$, asim. negativa.

$$g = \beta_2 = \frac{1}{S^4} \sum_{i=1}^{k} f_i (x_i - \bar{x})^4 = \frac{m_4(\bar{x})}{S^4} = 2'22 < 3$$

→ distrib. platicúrtica

1.5) (g/mol)

134'56	134'89	133'99	133'56	135'03	134'65	135'10	137'20
134'25	134'78	134'29	133'62	135'23	134'99	135'56	136'65
134'34	134'32	133'05	134'78	133'25	133'01	132'76	132'85
134'66	134'89	133'20	133'82	135'67	136'02	133'78	133'98

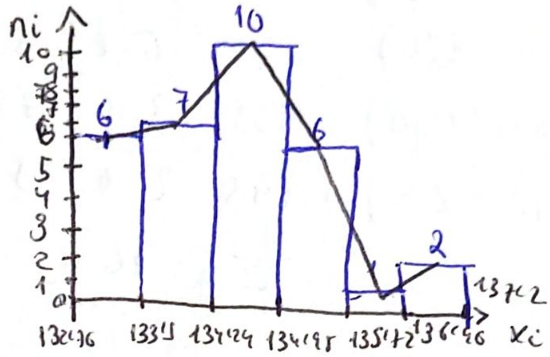
a) Histogramas frec abs. y rel. ¿Cuántas y qué clases hay?

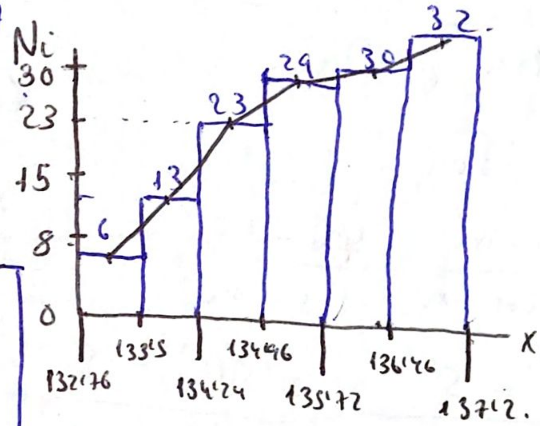
32 datos $\rightarrow$ 5'65 = 6 clases.

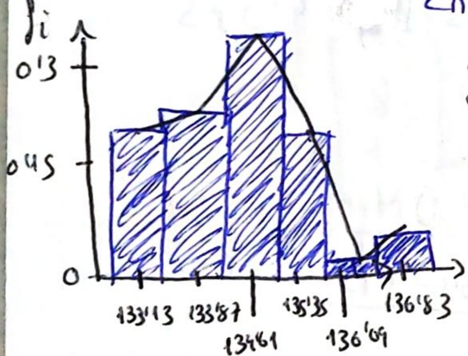
$$\frac{\text{Max}(x) - \text{Min}(x)}{\text{nº clases}} = \frac{137'20 - 132'76}{6} = 0'74$$

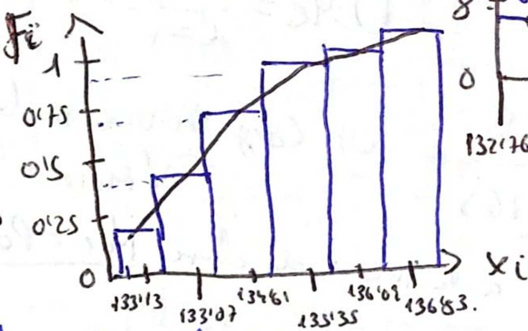
Intervalo	x (Marca de clase)	n_i	f_i	N_i	F_i
[132'76, 133'5)	133'13	6	0'1875	6	0'1875
[133'5, 134'24)	133'87	7	0'2187	13	0'4062
[134'24, 134'98)	134'61	10	0'312	23	0'7182
[134'98, 135'72)	135'35	6	0'1875	29	0'9057
[135'72, 136'46)	136'09	1	0'031	30	0'9367
[136'46, 137'2]	136'83	2	0'062	32	1'0

$\Sigma n = 32$









b) el valor adecuado será la media de los datos.

$\bar{x} = \sum_{i=1}^{k} f_i x_i = 135'65$

c) $S_A(\bar{x}) = \frac{S_x}{\sqrt{N}}$ $S_x = \sqrt{s^2_x} = \sqrt{\sum_{i=i}^{k} f_i (x_i - \bar{x})^2} = 1'41$

desv. típica de la media. desv. típica de la distrib.

$S_A(\bar{x}) = 0'24$.

Dispersión de la muestra $\bar{x} = 135'65$

$S_A(\bar{x}) = 0'24$.

d) $P_{0'25}$ ($N/4 = 8$) = 133'34.

$P_{0'75}$ ($3N/4 = 24$) = 134'73.

$RI = P_{0'75} - P_{0'25} = 1'39$.

1'5RI = 2'09.

1'5RI por encima de $P_{0'75}$ = 136'82.

1'5RI por debajo de $P_{0'25}$ = 131'25. $\Downarrow$ Hay dat atip leves

1.5) e) $m_1(0) = \sum_{i=1}^{k} f_i (x_i - 0) = \bar{x} = 135'65$.

$m_2(0) = \sum_{i=1}^{n} f_i (x_i - 0)^2 = 1806611$.

$m_3(0) = \sum_{i=1}^{k} f_i (x_i - 0)^3 = 2'43 \cdot 10^6$

$m_4(0) = \sum_{i=1}^{k} f_i (x_i - 0)^4 = 197'73 \cdot 10^6$

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

$m_2(\bar{x}) = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 =$

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

$m_4(\bar{x}) = \sum_{i=1}^{k} f_i (x_i - \bar{x})^4 =$

f) $A_F = \frac{1}{S^3} \sum_{i=1}^{k} f_i (x_i - \bar{x})^3 = 0'58 (>0) \Rightarrow$ asim positiva

Curtosis $\rightarrow g = \beta_2 = \frac{1}{S^4} \sum_{i=1}^{k} f_i (x_i - \bar{x})^4 = \frac{m_4(\bar{x})}{S^4} = 3'48 > 3 \Rightarrow$ distrib leptocúrtica.

1.6) Demuestrese que una var. Z construida con n_1 datos de X y n_2 datos de Y, tiene por media: $\bar{z} = \frac{n_1}{n_1+n_2} \bar{x} + \frac{n_2}{n_1+n_2} \bar{y}$.

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

1.7)

X	1	2	3	4	5	$\bar{x} = 3$
Y	0'34	0'70	1'08	1'43	1'70	$\bar{y} = 1'05$

$$Cov(x,y) = m_{11}(\bar{x}, \bar{y}) = \sum_{i=1}^{k} \sum_{j=1}^{p} f_{ij} (x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 0'69$$

Correl. positiva

1.8) dos variables X e Y. Calcúlese. $cov[\underbrace{(x+y), (x-y)}_{A}]$

$$COV(A) = m_{11}\left(\overline{(x+y)}, \overline{(x-y)}\right) = \frac{1}{n} \sum_{i=1}^{n} \left((x_i + y_i) - \overline{(x+y)}\right)\left((x_i - y_i) - \overline{(x-y)}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} [(x_i - \bar{x}) + (y_i - \bar{y})] [(x_i - \bar{x}) - (y_i - \bar{y})] = (x_i - \bar{x})^2 - (y_i - \bar{y})^2$$

1.12) $\vec{X} = (X, Y)$

Calcúlese:

$X \backslash Y$	1	2	4	6	n_{xi}
1	2	0	0	1	3
3	3	1	0	1	5
5	0	1	0	5	6
n_{yi}	5	2	0	7	

- $\sum_i \sum_j x_{ij} = 14$
- $f_{ij} = \frac{n_{ij}}{N} = \frac{n_{ij}}{14}$
- Distrib marginales n_{xi} n_{yj}, $\bar{x}$, $\bar{y}$.

$X \backslash Y$	1	2	4	6
1	2/14	0	0	1/14
3	3/14	1/14	0	1/14
5	0	1/14	0	5/14

a) $M_{1,0}(0,0)$
$M_{0,1}(0,0)$.

b) $M_{1,1}(\bar{x}, \bar{y})$

~~$\bar{x} = 4'67$~~
~~$\bar{y} = 3'5$.~~

$$\bar{x} = \frac{\sum_{i=1}^{k} (n_{xi}) \cdot x_i}{N} = \frac{(3 \cdot 1) + 5 \cdot 3 + 6 \cdot 5}{14} = 3'43.$$

$$M_{r,s}(c,d) = \sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij} (x_i - c)^r (y_j - d)^s$$

$$\bar{y} = \frac{\sum_{i=1}^{k} n_{yi} \cdot y_i}{N} = \frac{5 \cdot 1 + 2 \cdot 2 + 7 \cdot 6}{14} = 3'64.$$

a) $M_{1,0}(0,0) = \sum_{i=1} \sum_{j=1} f_{ij} (x_i)$
$= \sum_{i=1}^{k} f_{xi}(x_i) = \frac{1}{N} \sum_{i=1}^{k} n_i (x_i)$.
$= 3'43$.

b) $\mathrm{COV}(x, y) = M_{1,1}(\bar{x}, \bar{y})$

$$M_{1,1}(\bar{x}, \bar{y}) = \sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij} (x_i - \bar{x})(y_j - \bar{y}) = 1'867$$

Correlación positiva entre variables

$M_{0,1}(0,0)$
$$\sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij} (y_j) = \sum_{i=1}^{k} f_{xi}(x_i) = 3'64$$

1.13)

(Precio kg en cts)	X	21	19	29	36	31	29	37	31	33	35
(miles de kg)	Y	100	140	120	110	200	200	110	160	160	200

a)

marcas de clase

Intervalos x	x	intervalo y	y
[15, 25)	20	[100, 135)	117'5
[25, 35)	30	[135, 170)	152'5
[35, 45)	40	[170, 205)	187'5

$X \backslash Y$	117'5	152'5	187'5	n_{xi}
20	1	1	0	2
30	1	2	2	5
40	2	0	1	3
n_{yi}	4	3	3	N = 10

b) Distrib. marginales.

c) distrib de $X / Y = 200$

$n(x / y=200) = n_{ij}$
$n(20/200) = 0$
$n(30/200) = 2$
$n(40/200) = 1$

$$f(x_i / y = y_j) = \frac{n_{ij}}{n_{yj}}$$

$f(20/200) = 0$
$f(30/200) = 2/3$
$f(40/200) = 1/3$

1.13) d) Media, Mediana y moda.

$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = 31'0$. $Me_x = 30$ $P_{0'25\,x} = 30$ $P_{0'75\,x} = 40$

$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = 156'0$ $Me_y = 117'5$ $P_{0'25\,y} = 117'5$ $P_{0'75\,y} = 187'5$

$Mod(X) = 30$
$Mod(Y) = 117'5$

X	n_{xi}	N_i	F_i	Y	n_{yj}	N_j	F_j
20	2	2	0'2	117'5	4	4	0'4
30	5	7	0'7	152'5	3	7	0'7
40	3	10	1	187'5	3	10	1

e) Recorrido, desv. típica, desv. media, coef var. Pearson X e Y.

$S^2_x = \sum_{i=1}^{3} f_{xi}(x_i - \bar{x})^2 = 49$

$S_x = 7$

$S^2_y = \sum_{j=1}^{3} f_{yj}(y_j - \bar{y})^2 = 1104$

$S_y = 33'23$.

$C.V_x = \frac{S_x}{|\bar{x}|} = \frac{7}{31} = 0'226$

$DM\bar{x} = \sum_{i=1}^{3} f_{xi}|x_i - \bar{x}| \rightarrow 5'4$.

$CV_y = \frac{S_y}{|\bar{y}|} = \frac{33'23}{156'0} = 0'213$.

$DM\bar{y} = \sum_{i=1}^{3} f_{yj}|y_j - \bar{y}| \rightarrow 28'8$.

f) Ap g) $g = \beta_2$. h) $Cv_x > Cv_y$

$\Downarrow$

X es menos representativa

$Ap_x = \frac{\bar{x} - \overset{\text{Moda}}{Md_x}}{S_x} = \frac{31-30}{7} = \frac{1}{7} = 0'14$

$Ap_y = \frac{\bar{y} - Md_y}{S_y} = \frac{156 - 117'5}{33'23} = 1'16$

i) % años donde el precio fue inferior a 0'27€

∫ intervalo $[15, 25) = 20\%$.

$g_x = \frac{1}{S^4}\sum_{i=1}^{h} f_i(x_i - \bar{x})^4 = \frac{m_4(\bar{x})}{S_x^4} = 2'03$

$g_y = \frac{1}{S^4}\sum_{j=1}^{k} f_j(y_j - \bar{y})^4 = \frac{m_4(\bar{y})}{S_y^4} = 1'47$.

j) Regresión $Y = a_0 + b_0 x$ y matriz cov.

$b_0 = \frac{\overline{xy}}{\overline{x^2} - (\bar{x})^2} = \frac{cov(x,y)}{S_x^2} =$

$a_0 = \bar{y} - \frac{\overline{xy}}{\overline{x^2} - (\bar{x})^2}\bar{x} = \bar{y} - \frac{cov(x,y)}{S_x^2}\bar{x} =$

$$M = \begin{pmatrix} S^2_x & cov(x,y) \\ cov(y,x) & S^2_y \end{pmatrix}$$

$r = \frac{cov(x,y)}{S_x S_y} = 0'076$.