

(7)

(7)

Nombre: Raúl Bortolero Hernández

Y: \$ del bono recibido [M\$] (i)

X: Antigüedad laboral en la empresa [años] (j)

Y/X	[0;5)	[5;10)	[10;15)	[15;20]	TOTAL	m_x
[3;3,8)	22	27	21	0	70	3,4
[3,8;4,6)	15	21	13	0	49	4,2
[4,6;5,4)	55	12	15	23	105	5
[5,4;6,2)	0	0	25	45	70	5,8
[6,2;7]	0	14	14	27	55	6,6
TOTAL	92	74	88	95	349	
m_y	2,5	7,5	12,5	17,5		

a) [0;5)

j=1

m_x	m_i	f_i	N_i	F_i
3,4	22	22/92	22	22/92
4,2	15	15/92	37	37/92
5	55	55/92	92	1
5,8	0	0	92	1
6,6	0	0	92	1
TOTAL	92			

En primer lugar se asumirá que cada clase de "Antigüedad laboral" es un estrato & como parámetro consideramos a una clase. (Se podría hacer de cualquiera de los 2 formas)

$$\bar{Y}_1 = \sum_{i=1}^5 f_i \cdot m_i = \frac{22}{92} \cdot 3,4 + \frac{15}{92} \cdot 4,2 + \frac{55}{92} \cdot 5 + 0 \cdot 5,8 + 0 \cdot 6,6$$

$$\Rightarrow \boxed{\bar{Y}_1 = 4,487}$$

$$s_{y1}^2 = \left(\sum_{i=1}^5 f_i \cdot m_i^2 \right) - \bar{Y}_1^2 = \left(\frac{22}{92} (3,4)^2 + \frac{15}{92} (4,2)^2 + \frac{55}{92} (5)^2 + 0 \cdot (5,8)^2 + 0 \cdot (6,6)^2 \right) - (4,487)^2$$

$$s_{y1}^2 = 0,453 \text{ (Varianza)}$$

Desv. STD $S = \sqrt{s^2} \Rightarrow s_{y1} = \sqrt{0,453} \Rightarrow \boxed{s_{y1} = 0,673}$

• [5;10)
j=2

m_x	m_i	f_i
3,4	27	27/74
4,2	21	21/74
5	12	12/74
5,8	0	0
6,6	14	14/74
TOTAL	74	

$$s_{y2} = \sum_{i=1}^5 f_i \cdot m_i$$

$$\Rightarrow \bar{Y}_2 = \frac{27}{74} \cdot 3,4 + \frac{21}{74} \cdot 4,2 + \frac{12}{74} \cdot 5 + 0 \cdot 5,8 + \frac{14}{74} \cdot 6,6$$

$$\Rightarrow \boxed{\bar{Y}_2 = 4,492}$$

2

$$S_2^2 = \left(\sum_{i=1}^5 f_i \cdot m_i^2 \right) - \bar{Y}_2^2 = \left(\frac{27}{74} \cdot (3,4)^2 + \frac{13}{74} \cdot (4,2)^2 + \frac{12}{74} \cdot (5)^2 + 0 \cdot (5,8)^2 + \frac{14}{74} \cdot (6,6)^2 \right) - (4,492)^2$$

$$\Rightarrow S_2^2 = 1,342$$

$$D. STD \Rightarrow S_2 = \sqrt{S_2^2} = \sqrt{1,342} = 1,158$$

• $[10;15]$
 $f=3$

m_i	m_i	f_i
3,4	27	$27/88$
4,2	13	$13/88$
5	15	$15/88$
5,8	25	$25/88$
6,6	14	$14/88$
TOTAL	88	

$$N. \bar{Y}_3 = \sum_{i=1}^5 f_i \cdot m_i$$

$$\bar{Y}_3 = \frac{27}{88} \cdot 3,4 + \frac{13}{88} \cdot 4,2 + \frac{15}{88} \cdot 5 + \frac{25}{88} \cdot 5,8 + \frac{14}{88} \cdot 6,6 = 4,982$$

Varianza $S_3^2 = \left(\sum_{i=1}^5 f_i \cdot m_i^2 \right) - \bar{Y}_3^2$

$$= \left(\frac{27}{88} \cdot (3,4)^2 + \frac{13}{88} \cdot (4,2)^2 + \frac{15}{88} \cdot (5)^2 + \frac{25}{88} \cdot (5,8)^2 + \frac{14}{88} \cdot (6,6)^2 \right) - (4,982)^2$$

$$S_3^2 = 1,294 \Rightarrow D. STD \rightarrow S_3 = \sqrt{S_3^2} = 1,138$$

• $[15;20]$
 $f=4$

m_i	m_i	f_i
3,4	0	0
4,2	0	0
5	23	$23/95$
5,8	45	$45/95$
6,6	27	$27/95$
TOTAL	95	

$$\bar{Y}_4 = \sum_{i=1}^5 f_i \cdot m_i$$

$$\Rightarrow \bar{Y}_4 = 0 \cdot 3,4 + 0 \cdot 4,2 + \frac{23}{95} \cdot 5 + \frac{45}{95} \cdot 5,8 + \frac{27}{95} \cdot 6,6 = 5,834$$

Varianza $S_4^2 = \left(\sum_{i=1}^5 f_i \cdot m_i^2 \right) - \bar{Y}_4^2$

$$\Rightarrow S_4^2 = \left(0 \cdot (3,4)^2 + 0 \cdot (4,2)^2 + \frac{23}{95} \cdot (5)^2 + \frac{45}{95} \cdot (5,8)^2 + \frac{27}{95} \cdot (6,6)^2 \right) - (5,834)^2$$

$$S_4^2 = 0,336 \Rightarrow DESV. STD \rightarrow S_4 = \sqrt{S_4^2} = 0,579$$

$\therefore$ El rango que presenta un promedio mayor en el loro cubito es el de $[15;20]$ con $\bar{X} = 5,834$ [M\$]
y a la vez es el que presenta la menor Desv. STD con $S = 0,579$

(3)

b) Al igual que en el punto anterior, se supone como que cada clase de "Antigüedad laboral" es un estrato

1) Calcular la Varianza TOTAL dejando al caso como numerador, es decir, considerando la variable "Antigüedad" (X) y no sea la suma de todos los f_i .

m_i	m_i	f_i
3,4	70	70/349
4,2	49	49/349
5	105	105/349
5,8	70	70/349
6,6	55	55/349
Total (m)	349	

k : Cant de Clases donde $k=5$

m_i : Cantidad de "estrato" donde $m=4$

$$S^2_{TOTAL} = \left(\sum_{i=1}^k f_i \cdot m_i^2 \right) - \bar{Y}_{TOTAL}^2 \quad \text{donde} \quad \bar{Y}_{TOTAL} = \sum_{i=1}^k f_i \cdot m_i$$

$$\Rightarrow \bar{Y}_{TOTAL} = \sum_{i=1}^5 f_i \cdot m_i = \frac{70}{349} \cdot 3,4 + \frac{49}{349} \cdot 4,2 + \frac{105}{349} \cdot 5 + \frac{70}{349} \cdot 5,8 + \frac{55}{349} \cdot 6,6$$

$$\Rightarrow \bar{Y}_{TOTAL} = 4,979$$

$$\Rightarrow S^2_{TOTAL} = \left(\sum_{i=1}^5 f_i \cdot m_i^2 \right) - \bar{Y}_{TOTAL}^2$$

$$= \left[\frac{70}{349} \cdot (3,4)^2 + \frac{49}{349} \cdot (4,2)^2 + \frac{105}{349} \cdot (5)^2 + \frac{70}{349} \cdot (5,8)^2 + \frac{55}{349} \cdot (6,6)^2 \right] - (4,979)^2$$

$$S^2_{TOTAL} = 1,135 \quad \rightarrow \text{Varianza TOTAL (es igual a la Varianza marginal } S_y^2)$$

$$\text{Considerando que } S^2_{TOTAL} = \underbrace{\sum_{i=1}^m W_i S_i^2}_{\text{Varianza INTRA}} + \underbrace{\sum_{i=1}^m W_i (\bar{Y}_i - \bar{Y}_{TOTAL})^2}_{\text{Varianza INTER}}$$

Ahora calcular las varianzas INTRA; donde las varianzas de cada estrato están calculadas anteriormente y un peso de cada estrato son:

$$W_1 = \frac{92}{349}, \quad W_2 = \frac{74}{349}, \quad W_3 = \frac{88}{349}, \quad W_4 = \frac{95}{349}$$

y sus Varianzas

$$S_1^2 = 0,453, \quad S_2^2 = 1,342, \quad S_3^2 = 1,274, \quad S_4^2 = 0,336$$

$$\textcircled{4} \quad \therefore \text{VAR. INTA} = \sum_{j=1}^m W_j S_j^2$$

$$= \left(\frac{92}{349} \cdot 0,453 + \frac{74}{349} \cdot 1,342 + \frac{88}{349} \cdot 1,294 + \frac{95}{349} \cdot 0,336 \right)$$

$$\text{VAR. INTA} = 0,822$$

$$\text{y si } \text{VAR. TOTAL} = \text{VAR. INTA} + \text{VAR. INTER.}$$

$$\Rightarrow \text{VAR. INTER} = \text{VAR. TOTAL} - \text{VAR. INTA}$$

$$= 1,135 - 0,822$$

$$\boxed{\text{VAR. INTER} = 0,313} \quad \checkmark$$

$$\% \text{ Var INTER} = \frac{\text{VAR. INTER}}{\text{VAR. TOTAL}} \cdot 100$$

$$= \frac{0,313}{1,135} \cdot 100 = 27,58\% \quad \checkmark$$

$\therefore$ el 27,58% de la varianza total es explicada por la variabilidad entre los rangos de antigüedad LABORAL.

- 5) c) Para generar un modelo de regresión lineal, primero definamos las variables donde:

Variable predictora o independiente sea X , "antigüedad laboral" y
Variable respuesta o a ser explicada sea Y , "Monto del bono recibido"

$$\text{donde } \boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i} \quad \text{y} \quad \boxed{\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 &= \frac{\text{COV}(X, Y)}{S_X^2} \end{aligned}}$$

La Marginal de Y se obtiene

X	m_{ij}	$m_{j\cdot}$	$f_{j\cdot}$
[0;5)	2,5	92	92/349
[5;10)	7,5	74	74/349
[10;15)	12,5	88	88/349
[15;20]	17,5	95	95/349
TOTAL (m)		349	

$$\bar{X} = \sum_{j=1}^4 f_{j\cdot} \cdot m_j = \left(\frac{92}{349} \cdot 2,5 + \frac{74}{349} \cdot 7,5 + \frac{88}{349} \cdot 12,5 + \frac{95}{349} \cdot 17,5 \right)$$

$$\Rightarrow \boxed{\bar{X} = 10,165} \quad \text{Promedio Marginal de } X$$

La Varianza Marginal de X (S_X^2) es: $S_X^2 = \sum_{j=1}^4 f_{j\cdot} (m_j - \bar{X})^2$ o también

$$S_X^2 = \left(\sum_{j=1}^4 f_{j\cdot} \cdot m_j^2 \right) - \bar{X}^2 = \left(\frac{92}{349} \cdot (2,5)^2 + \frac{74}{349} \cdot (7,5)^2 + \frac{88}{349} \cdot (12,5)^2 + \frac{95}{349} \cdot (17,5)^2 \right) - (10,165)^2$$

$$\boxed{S_X^2 = 33,014} \quad \text{Varianza Marginal de } X$$

$$\text{La } \text{COV}(X, Y) = \left[\sum_{i=1}^5 \sum_{j=1}^4 f_{ij} \cdot m_i \cdot m_j \right] - \bar{X} \cdot \bar{Y}$$

$$\text{donde } \bar{Y}_{\text{TOTAL}} = \bar{Y} = 4,979 \quad (\text{obtenido anteriormente})$$

$$\text{La } f_{ij} = \frac{m_{ij}}{m} \quad \text{y } m = 349$$

m_i / m_j	2,5	7,5	12,5	17,5
2,5	22	27	21	0
7,5	15	21	13	0
12,5	55	12	15	23
17,5	0	0	25	45
22,5	0	14	14	27

6)

$$\Rightarrow \text{COV}(X,Y) = \frac{1}{n} \left[\sum_{i=1}^5 \sum_{j=1}^4 m_{ij} \cdot m_{i.} \cdot m_{.j} \right] - \bar{X} \cdot \bar{Y}$$

$$= \frac{1}{349} \left[22 \cdot 3,4 \cdot 2,5 + 27 \cdot 3,4 \cdot 7,5 + 27 \cdot 3,4 \cdot 12,5 + 0 \cdot 3,4 \cdot 17,5 + \right. \\ 15 \cdot 4,2 \cdot 2,5 + 27 \cdot 4,2 \cdot 7,5 + 13 \cdot 4,2 \cdot 12,5 + 0 \cdot 4,2 \cdot 17,5 + \\ 55 \cdot 5 \cdot 2,5 + 12 \cdot 5 \cdot 7,5 + 15 \cdot 5 \cdot 12,5 + 23 \cdot 5 \cdot 17,5 + \\ 0 \cdot 5,8 \cdot 2,5 + 0 \cdot 5,8 \cdot 7,5 + 25 \cdot 5,8 \cdot 12,5 + 45 \cdot 5,8 \cdot 17,5 + \\ \left. 0 \cdot 6,6 \cdot 2,5 + 14 \cdot 6,6 \cdot 7,5 + 14 \cdot 6,6 \cdot 12,5 + 27 \cdot 6,6 \cdot 17,5 \right] \\ - (10,165) \cdot (4,979)$$

$$\text{COV}(X,Y) = \frac{1}{349} [18703,5] - (10,165)(4,979)$$

$$\Rightarrow \boxed{\text{COV}(X,Y) = 2,98} \quad \text{COVARIANZA}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\text{COV}(X,Y)}{S_x^2} = \frac{2,98}{33,014} = 0,09$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 4,979 - (0,09)(10,165)$$

$$\boxed{\hat{\beta}_0 = 4,06}$$

$$\Rightarrow \boxed{Y_x = 4,06 + 0,09 X_x}$$

$$\text{Si } X_1 = 7 \Rightarrow \boxed{Y_1 = 4,69}$$

$\therefore$ si un trabajador lleva 7 años debe recibir 4,69 \$M

Otro si $X_2 = 25$, No se puede determinar porque la predicción debe estar dentro de la "nube" de puntos con la que se hizo la regresión y $X_2 = 25$ está fuera

$\therefore$ No se puede estimar (No sería correcto)

7) a) Para determinar la calidad del modelo se debe calcular el coeficiente de determinación del modelo de regresión lineal (R^2)

$$\text{donde } R^2 = \frac{[\text{COV}(X, Y)]^2}{S_x^2 S_y^2}, \text{ ya que } \text{COV}(X, Y) = 2,98$$

$$S_x^2 = 33,014 \quad (\text{obtenido en c)})$$

$$S_y^2 = 1,135 \quad (\text{obtenido en b)})$$

$$\Rightarrow R^2 = \frac{(2,98)^2}{(33,014)(1,135)} = 0,24$$

Considerando que $R^2 \in [0, 1]$ donde si:

$R^2 = 0 \rightarrow$ No es confiable la estimación y no explica el modelo.

$R^2 = 1 \rightarrow$ Es totalmente confiable y explica perfectamente el modelo.

$\Rightarrow$ Si $R^2 = 0,24 \Rightarrow$ quiere decir que la estimación para 7 años es poco confiable ya que es muy bajo su coeficiente de determinación (está más cercano a "0" que a "1")