

RELACIÓN DE PROBLEMAS 2

Daniel Zufri Quesada, Daniel Pérez Ruiz, David Suárez González, Pablo Nieto López, Alejandro Moreno Guerrero (Grupo B4)

1º DGIIM UGR

EJERCICIO 1

Se han lanzado dos dados varias veces, obteniendo los resultados que se presentan en la siguiente tabla, donde X designa el resultado del primer dado e Y el resultado del segundo:

X	1	2	2	3	5	4	1	3	3	4	1	2	5	4	3	4	4	5	3	1	6	5	4	6
Y	2	3	1	4	3	2	6	4	1	6	6	5	1	2	5	1	1	2	6	6	2	1	2	5

- Construir la tabla de frecuencias.
- Calcular las puntuaciones medias obtenidas con cada dado y ver cuáles son más homogéneas.
- ¿Qué resultado del segundo dado es más frecuente cuando en el primero se obtiene un 3?
- Calcular la puntuación máxima del 50 % de las puntuaciones más bajas obtenidas con el primer dado si con el segundo se ha obtenido un 2 o un 5.

ESTE LO HACES TÚ.

EJERCICIO 2

Medidos los pesos, X (en Kg), y las alturas, Y (en cm), a un grupo de individuos, se han obtenido los siguientes resultados:

X/Y	160	162	164	166	168	170	ni.	$x_i n_i$	$x_i^2 n_i$
48	3	2	2	1	0	0	8	384	18432
51	2	3	4	2	2	1	14	714	36414
54	1	3	6	8	5	1	24	1296	69984
57	0	0	1	2	8	3	14	798	45486
60	0	0	0	2	4	4	10	600	36000
n.j	6	8	13	15	19	9	70	3792	
$y_j n_j$	960	1296	2132	2490	3192	1530	11600		
$y_j^2 n_j$	153600	209952	349648	413340	536256	260100			

- a) Calcular el peso medio y la altura media y decir cuál es más representativo.**

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^5 n_i \cdot x_i = \frac{384 + 714 + 1296 + 798 + 600}{70} = 54,1714 \text{ Kg}$$

$$\bar{y} = \frac{1}{N} \sum_{j=1}^6 n_j \cdot y_j = \frac{960 + 1296 + 2132 + 2490 + 3192 + 1530}{70} = 165,714 \text{ cm}$$

- Para ver qué media es más representativa, calculamos las desviaciones típicas de X e Y, y después el coeficiente de variación:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{4,3527 + 2,0116 + 0,0101 + 1,6002 + 4,8532} \\ = \sqrt{12,8278} = 3,5816 \text{ kg}$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{3,7318 + 2,7592 + 1,0076 + 0,0163 + 0,7464} = \sqrt{8,2612} \\ = 2,8742 \text{ cm}$$

- Ahora calculamos los respectivos coeficientes de variación:

○ $C.V.(X) = \frac{3,5816}{54,1714} = 0,0661$

○ $C.V.(Y) = \frac{2,8742}{165,714} = 0,0173$

- Como $C.V.(Y) < C.V.(X)$, concluimos que la media de Y es más representativa que la de X.

b) Calcular el porcentaje de individuos que pesan menos de 55 Kg y miden más de 165 cm.

- Total de individuos que cumplen las características mencionadas: **20 personas.**
- Si $X\% \cdot 70 = 20 \Rightarrow X = \frac{20 \cdot 100}{70} = 28,57143 \sim 28,57$
- Por tanto, el 28'57% de las personas de la población estudiada pesan menos de 55 Kg y más de 165 cm.

c) Entre los que miden más de 165 cm, ¿cuál es el porcentaje de los que pesan más de 52 Kg?

- Ahora tenemos un subconjunto de la población total, cuya suma es **43 personas**, que son las que miden más de 165 cm.
- Total de individuos que cumplen las características mencionadas: **37 personas.**
- Si $X\% \cdot 43 = 37 \Rightarrow X = \frac{37 \cdot 100}{43} = 86,0465$
- Por tanto, el 86,0465% de las personas de la población estudiada pesan más de 52 Kg.

d) ¿Cuál es la altura más frecuente entre los individuos cuyo peso oscila entre 51 y 57 Kg?

- La altura más frecuente es 168 cm.

e) ¿Qué peso medio es más representativo, el de los individuos que miden 164 cm o el de los que miden 168 cm.

➤ Calculamos las medias de X condicionados a $y_1 = 164$ cm y a $y_2 = 168$ cm.

$$\overline{x/y_1} = \frac{1}{N_1} \cdot \sum_{i=1}^5 n_{i.} \cdot x_i = \frac{96 + 204 + 324 + 57}{13} = 52,3846 Kg$$

$$\overline{x/y_2} = \frac{1}{N_2} \cdot \sum_{i=1}^5 n_{i.} \cdot x_i = \frac{102 + 270 + 456 + 240}{19} = 56,2105 Kg$$

➤ Ahora calcularemos el coeficiente de variación de Pearson:

$$\sigma_{x_1}^2 = \frac{1}{13} \cdot (4608 + 10404 + 17496 + 3249) - 52,3846^2 = 6,3921 Kg^2$$

$$\sigma_{x_2}^2 = \frac{1}{19} \cdot (5202 + 14580 + 25992 + 14400) - 56,2105^2 = 7,4323 Kg^2$$

$$C.V.(X_1) = \frac{6,3921}{52,3846} = 0,0483$$

$$C.V.(X_2) = \frac{7,4323}{56,2105} = 0,0485$$

➤ $C.V.(X_1) < C.V.(X_2)$, pero la diferencia es tan pequeña que podemos decir que ambos pesos son igual de representativos en sus respectivas distribuciones.

EJERCICIO 3

En una encuesta de familias sobre el número de individuos que la componen (X) y el número de personas activas en ellas (Y) se han obtenido los siguientes resultados:

X/Y	1	2	3	4	ni.	$x_i n_{i.}$	$x_i^2 n_{i.}$	$x_i \sum_j n_{ij} y_j$
1	7	0	0	0	7	7	7	7
2	10	2	0	0	12	24	48	28
3	11	5	1	0	17	51	153	72
4	10	6	6	0	22	88	352	160
5	8	6	4	2	20	100	500	200
6	1	2	3	1	7	42	252	108
7	1	0	0	1	2	14	98	35
8	0	0	1	1	2	16	128	56
n.j	48	21	15	5	89	342	1538	666
$y_j n_{.j}$	48	42	45	20	155			
$y_j^2 n_{.j}$	48	84	135	80	347			

a) Calcular la recta de regresión de Y sobre X.

- CÁLCULO DE LAS MEDIAS DE “X” Y DE “Y”:

$$\bar{x} = \frac{7 + 24 + 51 + 88 + 100 + 42 + 14 + 16}{89} = 3,8427 \text{ individuos}$$

$$\bar{y} = \frac{48 + 42 + 45 + 20}{89} = 1,7416 \text{ activos}$$

- CÁLCULO DE LA VARIANZA DE “X” Y DE “Y”:

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^k n_i \cdot x_i^2 - \bar{x}^2 = \frac{1538}{89} - 3,8427^2 = 2,5146 \text{ individuos}^2$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum_{j=1}^p n_j \cdot y_j^2 - \bar{y}^2 = \frac{347}{89} - 1,7416^2 = 0,8657 \text{ personas activas}^2$$

- CÁLCULO DE LA COVARIANZA:

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i y_j - \bar{x} \bar{y} = \frac{1}{89} \cdot 666 - 6,6924 = 0,7907$$

- CÁLCULO DE LOS COEFICIENTES “a” y “b”:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0,7907}{2,5146} = 0,3144$$

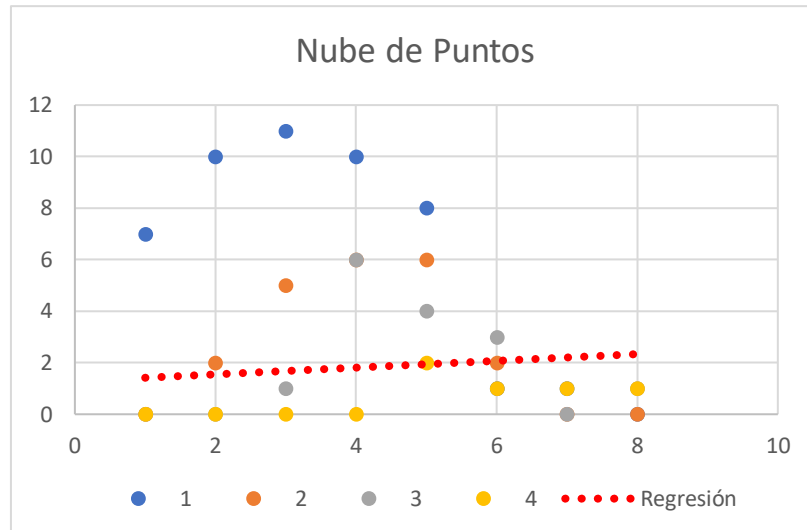
$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = 1,7416 - 0,3144 \cdot 3,8427 = 0,5335$$

$$y = 0,3144x + 0,5335$$

- b) ¿Es adecuado suponer una relación lineal para explicar el comportamiento de Y a partir de X?**

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{0,7907^2}{2,5146 \cdot 0,8657} = 0,1527$$

La respuesta es **NO**.



EJERCICIO 4

Se realiza un estudio sobre la tensión de vapor de agua (Y , en ml. de Hg.) a distintas temperaturas (X, en °C). Efectuadas 21 medidas, los resultados son:

X/Y		(0.5, 1.5]	(1.5, 2.5]	(2.5, 5.5]	ni.	$x_i n_{i.}$	$x_i^2 n_{i.}$	$x_i \sum_j n_{ij} y_j$
c_j / d_j		1	2	4				
(1,15]	8	4	2	0	6	48	384	64
(15,25]	20	1	4	2	7	140	2800	340
(25,30]	27,5	0	3	5	8	220	6050	715
n.j		5	9	7	21	408	9234	1119
$y_j n_{.j}$		5	18	28	51			
$y_j^2 n_{.j}$		5	36	112	153			

Explicar el comportamiento de la tensión de vapor en términos de la temperatura mediante una función lineal. ¿Es adecuado asumir este tipo de relación?

- Calculemos las medias de X y de Y:

$$\bar{x} = \frac{48 + 140 + 220}{21} = 19,4286 \text{ } ^\circ\text{C}$$

$$\bar{y} = \frac{5 + 18 + 28}{21} = 2,4286 \text{ Hg}$$

- Ahora, calculemos las varianzas y la covarianza:

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^k n_{i.} x_i^2 - \bar{x}^2 = \frac{9234}{21} - 19,4286^2 = 62,2438 \text{ } ^\circ\text{C}^2$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum_{j=1}^p n_{.j} y_j^2 - \bar{y}^2 = \frac{153}{21} - 2,4286^2 = 1,3876 \text{ Hg}^2$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i y_j - \bar{x} \bar{y} = \frac{1}{21} \cdot 1119 - 47,1843 = 6,1014$$

➤ Calculemos los coeficientes de la recta de regresión:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{6,1014}{62,2438} = 0,098$$

$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = 2,4286 - 0,098 \cdot 19,4286 = 0,5241$$

$$y = 0,0980x + 0,5239$$

➤ Calculemos el coeficiente de correlación lineal:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{7,1667^2}{72,5568 \cdot 1,5833} = 0,4471$$

La función explica menos del 50% de los casos: la relación que describe la función no es muy buena.

EJERCICIO 5

Estudiar la dependencia o independencia de las variables en cada una de las siguientes distribuciones. Dar, en cada caso, las curvas de regresión y la covarianza de las dos variables.

X/Y	1	2	3	4	5	ni.
10	2	4	6	10	8	30
20	1	2	3	5	4	15
30	3	6	9	15	12	45
40	4	8	12	20	16	60
n.j	10	20	30	50	40	150

➤ En esta primera distribución, no tiene sentido realizar el cálculo de la curva de regresión puesto que, ambas variables son estadísticamente independientes. Para verlo veamos que:

$$\frac{n_{i1}}{n_{.1}} = \frac{n_{i2}}{n_{.2}} = \dots = \frac{n_{ij}}{n_{.j}} = \dots = \frac{n_{ip}}{n_{.p}} \quad \forall i = 1, 2, \dots, k.$$

- *FILA 1* : $\frac{2}{10} = \frac{4}{20} = \frac{6}{30} = \frac{10}{50} = \frac{8}{40}$
- *FILA 2* : $\frac{1}{10} = \frac{2}{20} = \frac{3}{30} = \frac{5}{50} = \frac{4}{40}$
- *FILA 3* : $\frac{3}{10} = \frac{6}{20} = \frac{9}{30} = \frac{15}{50} = \frac{12}{40}$
- *FILA 4* : $\frac{4}{10} = \frac{8}{20} = \frac{12}{30} = \frac{20}{50} = \frac{16}{40}$

- Efectivamente, se cumple la condición de independencia de variables, por tanto, X e Y son variables estadísticamente independientes.
- La covarianza de estas variables es 0, por ser variables independientes.

$$\sigma_{xy} = 0,0000$$

X/Y	1	2	3	ni.	$x_i n_i$	$x_i^2 n_i$	$x_i \sum_j n_{ij} y_j$
-1	0	1	0	1	-1	1	-2
0	1	0	1	2	0	0	0
1	0	1	0	1	1	1	2
n.j	1	2	1	4	0	2	0
$y_j n_{.j}$	1	4	3	8			
$y_j^2 n_{.j}$	1	8	9	18			

- En la siguiente distribución, las variables estadísticas observadas X e Y son estadísticamente dependientes, puesto que no se cumple la condición para que sean independientes:

- *FILA 1*: $\frac{0}{1} \neq \frac{1}{2} \neq \frac{0}{1}$

- Aquí se ve que no se verifica, y por tanto, son estadísticamente dependientes. Además, no podemos hablar de dependencia funcional porque hay un caso en el que a un determinado valor x_i le corresponde más de un valor y_j de Y.

- Vamos a calcular la covarianza de ambas variables:

$$\sigma_{xy} = \frac{-2 + 0 + 2}{4} = 0$$

- Observación: Es bastante interesante el hecho de que la covarianza nos de 0, sabiendo que las variables son estadísticamente dependientes. Esto prueba que el recíproco de la premisa: "Si las variables son estadísticamente independientes, entonces su covarianza es 0.", **NO ES CIERTO**.

Ahora vamos a calcular la **curva de regresión de tipo 1**:

- Sabemos que pasa por los puntos: (x_i, \bar{y}_i) ; $i = 1, \dots, k$

○ Punto 1: (-1,2)

○ Punto 2: $(0, \bar{y}_2)$

$$y_2 = \frac{1+3}{2} = 2$$

○ Punto 3: (1, 2)

➤ Por tanto, la curva de regresión de Y/X que pasa por los puntos: (-1,2); (0,2); (1,2)

EJERCICIO 6

Dada la siguiente distribución bidimensional:

X/Y	1	2	3	4	n _{i.}	$x_i n_{i.}$	$x_i^2 n_{i.}$	$x_i \sum_j n_{ij} y_j$
10	1	3	0	0	4	40	400	70
12	0	1	4	3	8	96	1152	312
14	2	0	0	2	4	56	784	140
16	4	0	0	0	4	64	1024	64
n. _j	7	4	4	5	20	256	3360	586
$y_j n_{.j}$	7	8	12	20	47			
$y_j^2 n_{.j}$	7	16	36	80	139			

a) ¿Son estadísticamente independientes X e Y?

➤ No son estadísticamente independientes puesto que no se verifica la condición para que se produzca, por ejemplo:

▪ *FILA 2* : $\frac{0}{7} \neq \frac{1}{4} \neq \frac{4}{7} \neq \frac{3}{5}$

➤ Por tanto, las variables X e Y son estadísticamente dependientes. No podemos decir que sean dependientes de manera funcional puesto que para el valor x_1 le corresponden más de un valor y_j .

b) Calcular y representar las curvas de regresión de X/Y e Y/X.

➤ Para empezar, calcularemos la curva de regresión de Y/X, que se define como la curva que pasa por los puntos (x_i, \bar{y}_i) ; $i = 1, \dots, k$

➤ Calculemos los puntos:

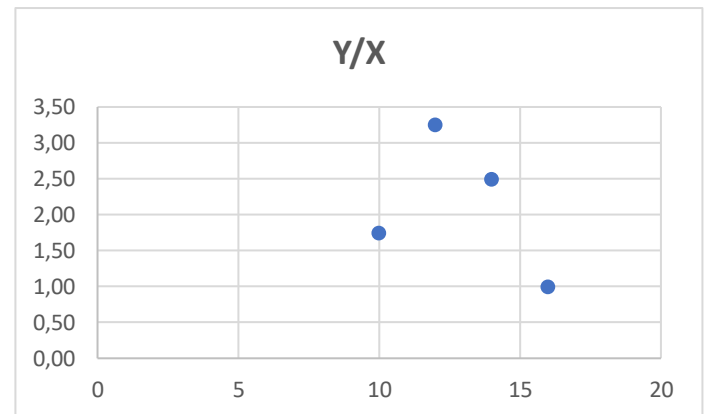
○ Punto 1: [x = 10]

$$\bar{y}_1 = \frac{1+6}{4} = \frac{7}{4}$$

$$\begin{aligned} \circ \text{ Punto 2: } [x = 12] \\ \bar{y}_2 = \frac{2+12+12}{8} = \frac{13}{4} \end{aligned}$$

$$\begin{aligned} \circ \text{ Punto 3: } [x = 14] \\ \bar{y}_3 = \frac{2+8}{4} = \frac{5}{2} \end{aligned}$$

$$\begin{aligned} \circ \text{ Punto 4: } [x = 16] \\ \bar{y}_4 = \frac{4}{4} = 1 \end{aligned}$$



Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(10, 7/4); (12, 13/4); (14, 5/2); (16, 1)\}$$

➤ Ahora calcularemos la curva de regresión de X/Y, que se define como la curva que pasa por los puntos $(y_i, \bar{x}_i); i = 1, \dots, k$.

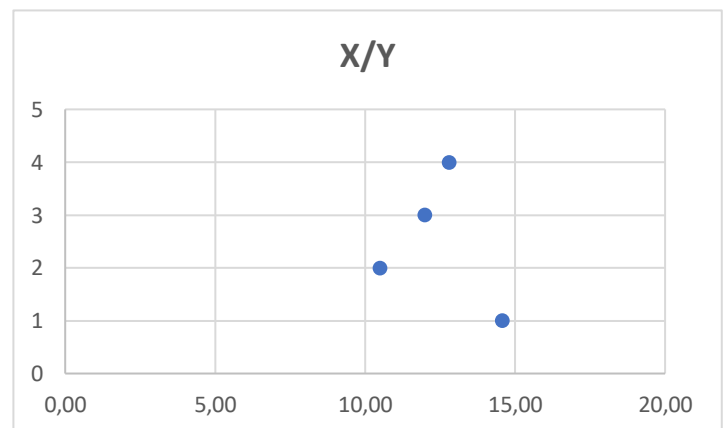
➤ Calculemos los puntos:

$$\begin{aligned} \circ \text{ Punto 1: } [y = 1] \\ \bar{x}_1 = \frac{10+28+64}{7} = \frac{102}{7} \end{aligned}$$

$$\begin{aligned} \circ \text{ Punto 2: } [y = 2] \\ \bar{x}_2 = \frac{30+12}{4} = \frac{21}{2} \end{aligned}$$

$$\begin{aligned} \circ \text{ Punto 3: } [y = 3] \\ \bar{x}_3 = \frac{48}{4} = 12 \end{aligned}$$

$$\begin{aligned} \circ \text{ Punto 4: } [y = 4] \\ \bar{x}_4 = \frac{36+28}{5} = \frac{64}{5} \end{aligned}$$



Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(102/7, 1); (21/2, 2); (12, 3); (64/5, 4)\}$$

c) **Cuantificar el grado en que cada variable es explicada por la otra mediante la correspondiente curva de regresión.**

- Para ello, vamos a calcular el valor de la correlación lineal de Y/X:

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2}$$

$$\sigma_{ry} = \frac{1}{N} \cdot \sum_{i=1}^p \sum_{j=1}^k n_{ij} (y_i - f(x_i))^2 = \frac{0.75 + 3.5 + 4.75}{20} = 0.45$$

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = 1 - 0.4316 = 0.5683$$

- Para ello, vamos a calcular el valor de la correlación lineal de X/Y:

$$\eta_{X/Y}^2 = 1 - \frac{\sigma_{rx}^2}{\sigma_x^2}$$

$$\sigma_{ry} = \frac{1}{N} \cdot \sum_{i=1}^p \sum_{j=1}^k n_{ij} (y_i - f(x_i))^2 = \frac{29.71 + 3 + 4.8}{20} = 1.8757$$

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = 1 - 0.4508 = 0.5491$$

d) **¿Están X e Y correladas linealmente? Dar las expresiones de las rectas de regresión.**

$$\bar{x} = \frac{40 + 96 + 56 + 64}{20} = 12,8$$

$$\bar{y} = \frac{7 + 8 + 12 + 20}{20} = 2,35$$

$$\sigma_{xy} = \frac{586}{20} - 12,8 \cdot 2,35 = -0,78$$

$$\sigma_x^2 = \frac{3360}{20} - 12,8^2 = 4,16$$

$$\sigma_y^2 = \frac{139}{20} - 2,35^2 = 1,4275$$

➤ Calculamos la recta de regresión Y/X:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{-0,78}{4,16} = -0,1875$$

$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = 2,35 + 0,1875 \cdot 12,8 = 4,75$$

$$y = -0,1875x + 4,75$$

➤ Calculamos la recta de regresión X/Y:

$$a = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{-0,78}{1,4275} = -0,5464$$

$$b = \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y} = 12,8 + 0,5464 \cdot 2,35 = 4,75$$

$$x = -0,5464y + 14,0840$$

➤ Calculamos r^2 y r :

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{-0,78^2}{4,16 \cdot 1,4275} = 0,1025$$

$$r = -\sqrt{r^2} = -0,3202$$

Observamos que **X e Y no están correladas linealmente en absoluto.**

EJERCICIO 7

Para cada una de las distribuciones:

X/Y	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3
4	0	1	0

X/Y	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3

X/Y	10	15	20	25
-----	----	----	----	----

1	0	3	0	1
2	0	0	1	0
3	2	0	0	0

a) ¿Dependen funcionalmente X de Y o Y de X?

En la distribución A, X depende funcionalmente de Y ya que para cada valor de x solo existe un único valor de Y.

En la distribución B, X depende funcionalmente de Y e Y depende funcionalmente de X ya que para cada x existe un único y; para cada x un único valor de y.

En la última distribución solamente Y depende funcionalmente de X ya que para cada valor de y se corresponde un único valor de x.

b) Calcular las curvas de regresión y comentar los resultados.

DISTRIBUCIÓN A

➤ Para empezar, calcularemos la curva de regresión de Y/X, que se define como la curva que pasa por los puntos (x_i, \bar{y}_i) ; $i = 1, \dots, k$

➤ Calculemos los puntos:

○ Punto 1: $[x = 1]$

$$\bar{y}_1 = \frac{30}{2} = 15$$

○ Punto 2: $[x = 2]$

$$\bar{y}_2 = \frac{10}{1} = 10$$

○ Punto 3: $[x = 3]$

$$\bar{y}_3 = \frac{60}{3}$$

○ Punto 4: $[x = 4]$

$$\bar{y}_4 = \frac{15}{1} = 15$$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(1, 15/2); (2, 10); (3, 60/3); (4, 15)\}$$

➤ Ahora calcularemos la curva de regresión de X/Y, que se define como la curva que pasa por los puntos $(\bar{x}_j, (y_j))$; $j = 1, \dots, p$

➤ Calculemos los puntos:

○ Punto 1: $[y = 10]$
$$\bar{x}_1 = \frac{2}{1} = 2$$

○ Punto 2: $[y = 15]$
$$\bar{x}_2 = \frac{2 + 4}{3} = 2$$

○ Punto 3: $[y = 20]$
$$\bar{x}_3 = \frac{9}{3} = 3$$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(2, 10); (2, 15); (3, 20)\}$$

DISTRIBUCIÓN B

➤ Calcularemos la curva de regresión de Y/X, que se define como la curva que pasa por los puntos (x_i, \bar{y}_i) ; $i = 1, \dots, k$

➤ Calculemos los puntos:

○ Punto 1: $[x = 1]$
$$\bar{y}_1 = \frac{30}{2} = 15$$

○ Punto 2: $[x = 2]$
$$\bar{y}_2 = \frac{10}{2} = 5$$

○ Punto 3: $[x = 3]$
$$\bar{y}_3 = \frac{60}{3} = 20$$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(1, 15); (2, 5); (3, 20)\}$$

➤ Ahora calcularemos la curva de regresión de X/Y, que se define como la curva que pasa por los puntos $(\bar{x}_j, (y_j))$; $j = 1, \dots, p$

➤ Calculemos los puntos:

○ Punto 1: $[y = 10]$
 $\overline{x_1} = \frac{2}{1} = 2$

○ Punto 2: $[y = 15]$
 $\overline{x_2} = \frac{2}{2} = 1$

○ Punto 3: $[y = 20]$
 $\overline{x_3} = \frac{9}{3} = 3$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(2, 10); (1, 15); (3, 20)\}$$

DISTRIBUCIÓN C

➤ Vamos a calcular la curva de regresión de Y/X, que se define como la curva que pasa por los puntos $(x_i, \overline{y_i})$; $i = 1, \dots, k$

➤ Calculemos los puntos:

○ Punto 1: $[x = 1]$
 $\overline{y_1} = \frac{70}{4} = \frac{35}{2}$

○ Punto 2: $[x = 2]$
 $\overline{y_2} = \frac{20}{1} = 20$

○ Punto 3: $[x = 3]$
 $\overline{y_3} = \frac{20}{2} = 10$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(1, 35/2); (2, 20); (3, 10)\}$$

➤ Ahora calcularemos la curva de regresión de X/Y, que se define como la curva que pasa por los puntos $(\overline{x_j}, (y_j))$; $j = 1, \dots, p$

➤ Calculemos los puntos:

- Punto 1: [y = 10]

$$\bar{x}_1 = \frac{6}{2} = 3$$

- Punto 2: [y = 15]

$$\bar{x}_2 = \frac{3}{3} = 1$$

- Punto 3: [y = 20]

$$\bar{x}_3 = \frac{2}{1} = 2$$

- Punto 4: [y = 25]

$$\bar{x}_4 = \frac{1}{1} = 1$$

Por tanto, la curva de regresión de Y/X es la que pasa por los puntos:

$$\{(3, 10); (1, 15); (2, 20); (1, 25)\}$$

EJERCICIO 8

De una muestra de 24 puestos de venta en un mercado de abastos se ha recogido información sobre el número de balanzas (X) y el número de dependientes (Y). Los resultados aparecen en la siguiente tabla:

X/Y	1	2	3	4	ni.	$x_i n_i$	$x_i^2 n_i$	$x_i \sum_j n_{ij} y_j$
1	1	2	0	0	3	3	3	5
2	1	2	3	1	7	14	28	36
3	0	1	2	6	9	27	81	96
4	0	0	2	3	5	24	96	72
n.j	2	5	7	10	24	68	208	
$y_j n_j$	2	10	21	40	244			
$y_j^2 n_j$	2	20	63	160	245			
$y_j \sum_i n_{ij} x_i$	3	18	60	128				209

a) *Determinar las rectas de regresión.*

$$\bar{x} = \frac{3 + 14 + 27 + 20}{24} = 2,6667 \text{ balanzas}$$

$$\bar{y} = \frac{2 + 10 + 21 + 40}{24} = 3,0417 \text{ dependientes}$$

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^k n_i \cdot x_i^2 - \bar{x}^2 = \frac{208}{24} - 2,6667^2 = 1,5554$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum_{j=1}^p n_j \cdot y_j^2 - \bar{y}^2 = \frac{245}{24} - 3,0417^2 = 0,9147$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i y_j - \bar{x} \bar{y} = \frac{1}{24} \cdot 209 - 8,113 = 0,5953$$

Por tanto:

➤ Para Y/X:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0,5953}{1,5554} = 0,3827$$

$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = 3,0417 - \frac{0,5953}{1,5554} \cdot 2,6667 = 2,0211$$

$$y = 0,3827x + 2,0211$$

➤ Para X/Y:

$$a = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{0,5953}{0,9147} = 0,6508$$

$$b = \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y} = 2,6667 - \frac{0,5953}{0,9147} \cdot 3,0417 = 0,6871$$

$$x = 0,6508y + 0,6871$$

a) *¿Es apropiado suponer que existe una relación lineal entre las variables?*

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{0,3544}{1,5554 \cdot 0,9147} = 0,2491$$

Obtenemos un coeficiente de correlación lineal muy bajo, ergo la respuesta es **NO**.

b) *Predecir, a partir de los resultados, el número de balanzas que puede esperarse en un puesto con seis dependientes. ¿Es fiable esta predicción?*

$$x = 0,6508y + 0,6871 = 0,6508 \cdot 6 + 0,6871 = 4,5919 \approx 5 \text{ balanzas}$$

EJERCICIO 9

Se eligen 50 matrimonios al azar y se les pregunta la edad de ambos al contraer matrimonio. Los resultados se recogen en la siguiente tabla, en la que X denota la edad del hombre e Y la de la mujer:

	y_j	15	22,5	27,5	32,5	37,5				
x_i	X/Y	(10,20]	(20,25]	(25,30]	(30,35]	(35,40]	ni.	$x_i n_i$	$x_i^2 n_i$	$x_i \sum_j n_{ij} y_j$
16,5	(15,18]	3	2	3	0	0	8	132	2178	2846.25
19,5	(18,21]	0	4	2	2	0	8	156	3042	4095
22,5	(21,24]	0	7	10	6	1	24	540	12150	14962.5
25,5	(24,27]	0	0	2	5	3	10	255	6502.5	8415
	n.j	3	13	17	13	4	50	1083	23872.5	
	$y_j n_j$	45	292.5	467.5	422.5	150	1377.5			
	$y_j^2 n_j$	675	6581.25	12856.25	13731.25	5625	39468.75			
	$y_j \sum_i n_{ij} x_i$	742.5	6041.25	10023.75	9798.75	3712.5				30318.75

Estudiar la interdependencia lineal entre ambas variables.

$$\bar{x} = \frac{132 + 156 + 540 + 255}{50} = 21.66 \text{ años}$$

$$\bar{y} = \frac{45 + 292.5 + 467.5 + 422.5 + 150}{50} = 27.55 \text{ años}$$

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 = \frac{23872.5}{50} - 21.66^2 = 8.2944$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum_{j=1}^p n_j y_j^2 - \bar{y}^2 = \frac{39468.75}{50} - 27.55^2 = 30.3725$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i y_j - \bar{x} \bar{y} = \frac{1}{50} \cdot 30318.75 - 21.66 \cdot 27.55 = 9.627$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{9.627}{2.88 \cdot 5.5111} = 0.6065$$

Por tanto, concluimos afirmando que la correlación es positiva y el grado de dependencia es elevado.

EJERCICIO 10

Calcular el coeficiente de correlación lineal de dos variables cuyas rectas de regresión son:

$$x + 4y = 1 \quad x + 5y = 2$$

- Supongamos que la primera ecuación es para Y/X y la segunda para X/Y:

$$y = \frac{-1}{4}x + \frac{1}{4} \quad x = -5y + 2$$

Sabemos que las pendientes tienen que tener el mismo signo, pues en caso contrario no serían rectas de regresión.

$$a = \frac{\sigma_{xy}}{\sigma_x^2}, a' = \frac{\sigma_{xy}}{\sigma_y^2} \Rightarrow a \cdot a' = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = r^2 = \frac{5}{4} > 1$$

Este caso es imposible, pues $0 \leq r^2 \leq 1$.

- Supongamos el otro caso:

$$x = -4y + 1 \quad y = -\frac{1}{5}x + \frac{2}{5}$$

$$r^2 = \frac{4}{5} \Rightarrow r = \sqrt{r^2} = \frac{-2}{\sqrt{5}}$$

EJERCICIO 11

Consideremos una distribución bidimensional en la que la recta de regresión de Y sobre X es $y = 5x - 20$, y $\sum y_j^2 n_{.j} = 3240$. Supongamos, además, que la distribución marginal de X es:

x_i	3	5	8	9
$n_{i.}$	5	1	2	1

Determinar la recta de regresión de X sobre Y, y la bondad de los ajustes lineales.

- Calculamos los momentos necesarios:

$$m_{02} = (1/9) \cdot \sum y_i^2 n_{.j} = 360$$

$$m_{01} = \bar{y}$$

- Obtenemos las medias:

$$\bar{x} = (1/9) \cdot \sum x_i n_{i.} = 5$$

$$y - \bar{y} = 5(x - \bar{x}) \rightarrow \bar{y} - 5\bar{x} = -20 \rightarrow \bar{y} = 5$$

- Cálculo de las varianzas:

$$\sigma_y^2 = m_{02} - m_{01}^2 = 360 - 25 = 335$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = (1/9) \cdot \sum x_i^2 n_i - 25 = 6$$

○ Obtención de la recta:

$$\frac{\sigma_{xy}}{\sigma_y^2} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot \frac{\sigma_x^2}{\sigma_y^2} = 5 \cdot \frac{6}{335} = 0,0896$$

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}) \rightarrow x = 0,0896 \cdot y + 4,552$$

La bondad de los ajustes lineales viene dada por la razón de correlación lineal:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = 5 \cdot 0,0896 = 0,448$$

EJERCICIO 12

De las estadísticas de "Tiempos de vuelo y consumos de combustible" de una compañía aérea, se han obtenido datos relativos a 24 trayectos distintos realizados por el avión DC-9. A partir de estos datos se han obtenido las siguientes medidas:

$$\begin{aligned} \sum y_i &= 219,719 & \sum y_i^2 &= 2396,504 & \sum x_i y_i &= 349,486 & \sum x_i &= 31,470 \\ \sum x_i^2 &= 51,075 & \sum x_i^2 y_i &= 633,993 & \sum x_i^4 &= 182,977 & \sum x_i^3 &= 93,6 \end{aligned}$$

La variable Y expresa el consumo total de combustible, en miles de libras, correspondiente a un vuelo de duración X (el tiempo se expresa en horas, y se utilizan como unidades de orden inferior fracciones decimales de la hora).

a) Ajustar un modelo del tipo $Y = aX + b$. ¿Qué consumo total se estimaría para un programa de vuelos compuesto de 100 vuelos de media hora, 200 de una hora y 100 de dos horas? ¿Es fiable esta estimación?

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{m_{11} - m_{10}m_{01}}{m_{20} - m_{10}^2} = \frac{\frac{349,486}{24} - \frac{219,719 \cdot 31,470}{24 \cdot 24}}{\frac{51,075}{24} - \frac{31,470^2}{24^2}} = 6,2568$$

$$b = \bar{y} - a \cdot \bar{x} = m_{01} - a \cdot m_{10} = \frac{219,719}{24} - 6,2568 \cdot \frac{31,470}{24} = 0,9507$$

Por tanto:

$$Y = 6,2568 \cdot X + 0,9507$$

Para 100 vuelos de media hora se estimaría un consumo de:

$$100 \cdot (6,2568 \cdot 0,5 + 0,9507) = 407,91 \text{ miles de libras de combustible}$$

Para 200 vuelos de una hora se estimaría un consumo de:

$$200 \cdot (6,2568 \cdot 1 + 0,9507) = 1441,5 \text{ miles de libras de combustible}$$

Para 100 vuelos de dos horas se estimaría un consumo de:

$$100 \cdot (6,2568 \cdot 2 + 0,9507) = 1346,43 \text{ miles de libras de combustible}$$

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{(m_{11} - m_{10}m_{01})^2}{(m_{20} - m_{10}^2) \cdot (m_{20} - m_{10}^2)} = 0,9975$$

Como la razón de correlación es muy próxima a 1, estas estimaciones son fiables

b) Ajustar un modelo del tipo $Y = a + bX + cX^2$. ¿Qué consumo total se estimaría para el mismo programa de vuelos del apartado a)?

➤ Hay que resolver el siguiente sistema de ecuaciones:

$$\begin{aligned} m_{01} &= a + bm_{10} + cm_{20} \\ m_{11} &= am_{10} + bm_{20} + cm_{30} \\ m_{21} &= am_{20} + bm_{30} + cm_{40} \end{aligned}$$

➤ Se obtiene la siguiente solución:

$$\begin{aligned} a &= 0.8055 \\ b &= 6.5321 \\ c &= -0.1014 \end{aligned}$$

Por tanto:

$$Y = -0.1014 \cdot X^2 + 6.5321 \cdot X + 0.8055$$

▪ Para 100 vuelos de media hora se estimaría un consumo de:

$$100 \cdot (-0.1014 \cdot 0.5^2 + 6.5321 \cdot 0.5 + 0.8055) = \mathbf{404.62 \text{ miles de libras de combustible}}$$

▪ Para 200 vuelos de una hora se estimaría un consumo de:

$$200 \cdot (-0.1014 \cdot 1^2 + 6.5321 \cdot 1 + 0.8055) = \mathbf{1447.24 \text{ miles de libras de combustible}}$$

▪ Para 100 vuelos de dos horas se estimaría un consumo de:

$$100 \cdot (-0.1014 \cdot 2^2 + 6.5321 \cdot 2 + 0.8055) = \mathbf{1346.41 \text{ miles de libras de combustible}}$$

c) ¿Cuál de los dos modelos se ajusta mejor? Razonar la respuesta.

Para el segundo caso se tiene lo siguiente:

$$\sigma_{ey}^2 = \sum_{i=1}^{24} \sum_{j=1}^{24} f_{ij} (-0,1014x_i^2 + 6,5321x_i + 0,8055 - \bar{y})^2 =$$

$$(1/24) \sum_{i=1}^{24} (0,0103x_i^4 - 11,3248x_i^3 + 42,5049x_i^2 + 10,5232x_i^2 + 0,2028x_i^2\bar{y} - 13,0642x_i\bar{y} - 1,611\bar{y} + \bar{y}^2 + 0,6488)$$

$$r^2 = \frac{\sigma_{ey}^2}{\sigma_y^2} = 0,9976$$

Como las razones de correlación son similares, **ambos modelos se ajustan prácticamente igual de bien.**

EJERCICIO 13

La curva de Engel, que expresa el gasto en un determinado bien en función de la renta, adopta en ocasiones la forma de una hipérbola equilátera. Ajustar dicha curva a los siguientes datos, en los que X denota la renta en miles de euros e Y el gasto en euros. Cuantificar la bondad del ajuste:

X	10	12.5	20	25
Y	50	90	160	180

Efectuamos los siguientes cambios de variable:

$$X' = X^2 \quad Y' = Y^2$$

X'	100	156,25	400	625
Y'	2500	8100	25600	32400

➤ Para hallar la función de regresión lineal de Y' sobre X':

$$y' - \bar{y}' = \frac{\sigma_{x'y'}}{\sigma_{x'}^2} (x' - \bar{x}')$$

$$\bar{x} = 320,3125$$

$$\bar{y} = 17150$$

$$\sigma_{xy} = (1/4) \sum_{i=1}^4 x_i y_i - \bar{x} \bar{y} = 2508046,875$$

$$\sigma_x^2 = (1/4) \sum_{i=1}^4 x_i^2 - \bar{x}^2 = 43659,6680$$

$$y' = 57,4454 \cdot x' - 1250,4781$$

- Deshaciendo los cambios de variable y reordenando:

$$x^2 - \frac{y^2}{57,4454} = 21,7681$$

EJERCICIO 14

Se dispone de la siguiente información referente al gasto en espectáculos (Y , en euros) y la renta disponible mensual (X, en cientos de euros) de 6 familias:

Y	30	50	70	80	120	140
X	9	10	12	15	22	32

X/Y	30	50	70	80	120	140	ni.	$x_i n_i$	$x_i^2 n_i$	$x_i \sum_j n_{ij} y_j$
9	1	0	0	0	0	0	1	9	81	270
10	0	1	0	0	0	0	1	10	100	500
12	0	0	1	0	0	0	1	12	144	840
15	0	0	0	1	0	0	1	15	225	1200
22	0	0	0	0	1	0	1	22	484	2640
32	0	0	0	0	0	1	1	32	1024	4480
n.j	1	1	1	1	1	1	6	100	48700	9930
$y_j n_j$	30	50	70	80	120	140	490			
$y_j^2 n_j$	900	2500	4900	6400	14400	19600	48700			

Explicar el comportamiento de Y por X mediante:

a) Relación lineal.

- Calculemos la media de X, de Y, y sus respectivas varianzas, así como la covarianza:

$$\bar{y} = \frac{1}{N} \cdot \sum_{i=1}^6 n_i \cdot y_i = \frac{30 + 50 + 70 + 80 + 120 + 140}{6} = 81,6667 \text{ euros}$$

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^6 n_i \cdot x_i = \frac{100}{6} = 16,6667 \text{ cientos de euros}$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_i \cdot y_i^2 - \bar{y}^2 = \frac{48700}{6} - 81,6667^2 = 1447,2222 \text{ euros}^2$$

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_i \cdot x_i^2 - \bar{x}^2 = \frac{2058}{6} - 16,6667^2 = 65,2222 \text{ cientos de euros}^2$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i y_j - \bar{x}\bar{y} = \frac{1}{6} \cdot 9930 - 1361,1143 = 293.8857$$

- Ahora que tenemos todo, calculemos los coeficientes “a” y “b”:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{293,8857}{65.2222} = 4,5060$$

$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = 81.6667 - 4,5060 \cdot 16.6667 = 6.5673$$

- Por tanto, la recta de regresión de Y sobre X es:

$$y = 4.5060x + 6.5673$$

- Para ver cómo de bueno es el ajuste, calculamos la razón de correlación lineal:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = 0.9150$$

b) Hipérbola equilátera.

- Llamemos $Z = 1/X$

Y	30	50	70	80	120	140
Z	0.1111	0.1	0.0833	0.0667	0.0455	0.0313

$$\bar{z} = \frac{1}{N} \cdot \sum_{i=1}^6 n_{i.} \cdot z_i = 0.07297 \text{ cientos de euros}$$

$$\sigma_z^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_{i.} z_i^2 - \bar{z}^2 = \frac{0.0368}{6} - 0.07297^2 = 0,0008 \text{ cientos de euros}^2$$

$$\sigma_{zy} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} z_i y_j - \bar{z}\bar{y} = 4.8883 - 5.9592 = -1,0709$$

- Ahora que tenemos todo, calculemos los coeficientes “a” y “b”:

$$a = \frac{\sigma_{zy}}{\sigma_z^2} = \frac{-1,0709}{0.0008} = -1330,2679$$

$$b = \bar{y} - \frac{\sigma_{zy}}{\sigma_z^2} \cdot \bar{z} = 81.6667 + 1330.2679 \cdot 0.07297 = 178.7354$$

- Por tanto, el ajuste hiperbólico de Y sobre X es:

$$y = -1330.2679z + 178.7354 = \frac{-1330.2679}{x} + 178.7354$$

- Si queremos saber cómo de bueno es este ajuste, veamos el valor de la razón de correlación:

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = \frac{\sigma_{ey}^2}{\sigma_y^2}$$

$$\sigma_{ey} = \frac{1}{N} \cdot \sum_{i=1}^p \sum_{j=1}^k n_{ij} (\hat{y}_j - \bar{y})^2 =$$

$$= \frac{2574.4303 + 1292.9842 + 190.0802 + 70.2944 + 1339.7047 + 3080.0089}{6} = 1424.5838$$

- Valor de la razón de la correlación:

$$\eta_{Y/X}^2 = \frac{1424.5838}{1447.2222} = 0.9844$$

c) Curva potencial.

$$y = bx^a \Leftrightarrow \log(y) = a \cdot \log(x) + \log(b)$$

$$V = \log(y) ; B = \log(b) ; W = \log(x)$$

V	3.4011	3.912	4.2484	4.382	4.7874	4.9416
W	2.1972	2.3025	2.4849	2.708	3.091	3.4657

$$\bar{V} = \frac{1}{N} \cdot \sum_{i=1}^6 n_{i.} \cdot V_i = 4.27875 \text{ euros}$$

$$\bar{W} = \frac{1}{N} \cdot \sum_{i=1}^6 n_{i.} \cdot W_i = 2.70822 \text{ cientos de euros}$$

$$\sigma_V^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_{i.} V_i^2 - \bar{V}^2 = 0.2671 \text{ euros}^2$$

$$\sigma_W^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_{i.} W_i^2 - \bar{W}^2 = 0.1993 \text{ cientos de euros}^2$$

$$\sigma_{VW} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} V_i W_j - \bar{V} \bar{W} = 0.2168$$

- Ahora que tenemos todo, calculemos los coeficientes “a” y “b”:

$$a = \frac{\sigma_{VW}}{\sigma_W^2} = 1.0877$$

$$B = \bar{V} - \frac{\sigma_{VW}}{\sigma_W^2} \cdot \bar{W} = 1.3329$$

➤ Por tanto, el ajuste potencial de Y sobre X es:

$$V = 1,3329 + 1,0877 \cdot W \Leftrightarrow y = 3.792 \cdot x^{1,0887}$$

➤ Si queremos saber cómo de bueno es este ajuste, veamos el valor de la razón de correlación:

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2}$$

$$\begin{aligned} \sigma_{ry} &= \frac{1}{N} \cdot \sum_{i=1}^p \sum_{j=1}^k n_{ij} (y_i - f(x_i))^2 = \\ &= \frac{131.6006 + 12.1637 + 176.2310 + 58.9301 + 105.2731 + 625.7823}{6} = 184.9968 \end{aligned}$$

➤ Valor de la razón de la correlación:

$$\eta_{Y/X}^2 = 1 - \frac{184.9968}{1447.2222} = 1 - 0.1278 = 0.8722$$

d) Curva exponencial.

$$y = ba^x \Leftrightarrow \log(y) = \log(b) + x \cdot \log(a)$$

$$V = \log(y); B = \log(b); A = \log(a)$$

V	3.4011	3.912	4.2484	4.382	4.7874	4.9416
x	9	10	12	15	22	32

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^6 n_{i.} \cdot x_i = \frac{100}{6} = 16,6667 \text{ cientos de euros}$$

$$\bar{V} = \frac{1}{N} \cdot \sum_{i=1}^6 n_{i.} \cdot V_i = 4.27875 \text{ euros}$$

$$\sigma_V^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_{i.} V_i^2 - \bar{V}^2 = 0.2671 \text{ euros}^2$$

$$\sigma_x^2 = \frac{1}{n} \cdot \sum_{i=1}^6 n_i \cdot x_i^2 - \bar{x}^2 = \frac{2058}{6} - 16.6667^2 = 65,2222 \text{ cientos de euros}^2$$

$$\sigma_{xV} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^p n_{ij} x_i V_j - \bar{V} \bar{x} = 3.6700$$

- Ahora que tenemos todo, calculemos los coeficientes “a” y “b”:

$$A = \frac{\sigma_{Vx}}{\sigma_x^2} = 0.0563$$

$$B = \bar{V} - \frac{\sigma_{Vx}}{\sigma_x^2} \cdot \bar{x} = 3.3409$$

$$V = 3.3409 + 0.0563 \cdot x \Leftrightarrow y = 28.24445 \cdot 1.0579^x$$

- Si queremos saber cómo de bueno es este ajuste, veamos el valor de la razón de correlación:

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2}$$

$$\begin{aligned} \sigma_{ry} &= \frac{1}{N} \cdot \sum_{i=1}^p \sum_{j=1}^k n_{ij} (y_i - f(x_i))^2 = \\ &= \frac{284.7416 + 0.1695 + 210.3406 + 204.3299 + 509.1822 + 965.0271}{6} = 362.2985 \end{aligned}$$

- Valor de la razón de la correlación:

$$\eta_{Y/X}^2 = 1 - \frac{362.2985}{1447.2222} = 1 - 0.2503 = 0.7497$$

¿Qué ajuste es más adecuado?

- El más adecuado es el ajuste hiperbólico equilátero.