Estadística Descriptiva e Introducción a la Probabilidad

Daniel Alconchel Vázquez
Nars El Farissi
Mario García Márquez
Alberto Diaz Cencillo
Javier Garrues Apecechea

5 de abril de 2020

1. Se han lanzado dos dados varias veces, obteniendo los resultados que se presentan en la siguiente table, 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

a) Construir la tabla de frecuencias.

b) Calcular las puntuaciones medias obtenidas con cada dado y ver cuales son más homogéneas.

c) ¿Qué resultado del segundo dado esmás frecuente cuando en el primero se obtiene 3?

d) Calcular la puntuación del 50 % de las puntuaciones demás bajas obtenidas con el primer dado si con el segundo se ha obtenido un 2 o un 5 .

Este ejercicio representa, para este tema, una pérdida de tiempo innecesario; por lo que, de momento, lo dejaremos apartado.

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
48	3	2	2	1	0	0
51	2	3	4	2	2	1
54	1	3	6	8	5	1
57	0	0	1	2	8	3
60	0	0	0	2	4	4

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

b) Calcular el porcentaje de individuos que pesan menos de 55 Kg y miden 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?

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

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

En una población de tamaño $n = 70$ personas se ha observado dos variables estadísticas, X = peso de los individuos en kilogramos e Y = altura de los individuos en cm , que han presentado $k = 5$ y $p = 6$ modalidades distintas respectivamente, con distribución conjunta $(x_i, y_j), n_{ij}$ con $i = 1, ..., 5$ y $j = 1, ..., 6$

Vamos a comenzar completando la tabla para operaciones futuras:

X\Y	160	162	164	166	168	170	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$
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	54,1714286	
$n_{.j}y_j$	960	1296	2132	2490	3192	1530	165,714286		
$n_{.j}y_j^2$	153600	209952	349648	413340	536256	260100			

a) Calculamos el peso y la altura media:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{5} x_i n_{i.} = 54,171428 \qquad (1)$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{6} y_j n_{.j} = 165,714286 \tag{2}$$

Para ver que medida es más representativa, calculamos las desviaciones típicas de X e Y y, a continuación, calculamos sus respectivos coeficientes de varación de Pearson:

$$\sigma_x = +\sqrt{\sigma_x} = +\sqrt{m_{20} - m_{10}^2} = 3,5816Kg \tag{3}$$

$$\sigma_y = +\sqrt{\sigma_y} = +\sqrt{m_{02} - m_{01}^2} = 2,8661cm \tag{4}$$

Ahora calculamos los coeficientes de variación de Pearson:

$$CV(x) = \frac{\sigma_x}{|\overline{x}|} = 0,0661 \tag{5}$$

$$CV(y) = \frac{\sigma_y}{|\overline{y}|} = 0,0173 \tag{6}$$

Como $CV(y) < CV(x)$, concluimos con que la media de Y es más representativa.

b) Mirando la tabla, vemos que el número de individuos que pesan menos de 55 Kg y miden más de 165 cm son 20. Luego:

$$\% = \frac{20 * 100}{70} = 28,5714\,\% \tag{7}$$

de los individuos pesan menos de 55 Kg y miden más de 165 cm

c)
El número de individuos que miden más de 165 cm y pesan más de 52 Kg son 37. El numero total de personas que miden más de 165 cm son 43, luego:

$$\% = \frac{37 * 100}{43} = 86,04565\,\% \tag{8}$$

de las personas que miden más de 165 cm pesan más de 55 Kg

d) La altura más frecuente es 168 cm.

e)
Comenzamos haciendo la tabla de las distribuciones condicionadas:

Tabla de x condicionada a y = 164:

X\Y	n_{i3}	$n_{i3}x_i$	$n_{i3}x_i^2$
48	2	96	4608
51	4	204	10404
54	6	324	17496
57	1	57	3249
60	0	0	0
	13	52,3846154	2750,53846

Tabla de x condicionada a y = 168:

X\Y	n_{i5}	$n_{i5}x_i$	$n_{i5}x_i^2$
48	0	0	0
51	2	102	5202
54	5	270	14580
57	8	456	25992
60	4	240	14400
	19	56,2105263	3167,05263

Comenzamos calculando las medias de las distribuciones condicionadas:

$$\overline{x/y_{=164}} = 52,3846154Kg \tag{9}$$

$$\overline{x/y_{=168}} = 56,2105263Kg \tag{10}$$

Calculamos ahora las desviaciones típicas:

$$\sigma_{x/y_{=164}} = 2,527950265Kg \tag{11}$$

$$\sigma_{x/y_{=168}} = 2,725685763Kg \tag{12}$$

Por último, calculamos el coeficiente de variación:

$$CV(x/y_{y=164}) = \frac{\sigma_{x/y_{=164}}}{\overline{x/y_{=164}}} = 0,048257494 \tag{13}$$

$$CV(x/y_{y=168}) = \frac{\sigma_{x/y_{=168}}}{\overline{x/y_{=168}}} = 0,048490664 \tag{14}$$

Luego, tenemos que$CV(x/y_{y=164}) < CV(x/y_{y=168})$, pero la diferencia es tan pequeña que pdríamos decir que las dos son igual de representativas.

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 siguientesn resultados:

X\Y	1	2	3	4
1	7	0	0	0
2	10	2	0	0
3	11	5	1	0
4	10	6	6	0
5	8	6	4	2
6	1	2	3	1
7	1	0	0	1
8	0	0	1	1

a) Calcular la recta de regresión de Y sobre X

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

En una población de tamaño n = 89 familias se han observado dos variables estadísticas X = número de miembros de cada familia e Y = número depersonas activas, las cuales han presentado,respectivamente, k = 8 y p = 4 modalidades, con distribución de frecuencia conjunta $(x_i y_j), n_{ij}$ i = 1,...,8 j=1,...4

X\Y	1	2	3	4	$n_{i.}$	$x_i n_{i.}$	$x_i^2 n_{i.}$	$\sum_{j=1}^{p} n_{ij} y_j$	$x_i \sum_{j=1}^{p} n_{ij} y_j$
1	7	0	0	0	7	7	7	7	7
2	10	2	0	0	12	24	48	14	28
3	11	5	1	0	17	51	153	24	72
4	10	6	6	0	22	88	352	40	160
5	8	6	4	2	20	100	500	40	200
6	1	2	3	1	7	42	252	18	108
7	1	0	0	1	2	14	98	5	35
8	0	0	1	1	2	16	128	7	56
n.j	48	21	15	5	89	342	1538		666
yjn.j	48	42	45	20	155				
yj2n.j	48	84	135	80	347				

Para calcular la recta de regresión de Y sobre X necesitamos las siguientes medidas:

$$\overline{x} = \frac{342}{89} = 3,842696629 individuos \tag{15}$$

$$\overline{y} = \frac{155}{89} = 1,741573034 individuosactivos \tag{16}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 2,514581492 individuos^2 \tag{17}$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i yj - \overline{x} * \overline{y} = 0,790809241 \tag{18}$$

Ahora que tenemos todo lo necesario calculamos la recta:

$$y = ax + b \tag{19}$$

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = 0,314489407 \tag{20}$$

$$b = \overline{y} - a\overline{x} = 0,533085651 \tag{21}$$

Luego, la recta de regresión de Y sobre X es $y = 0,314489407 * x + 0,533085651$

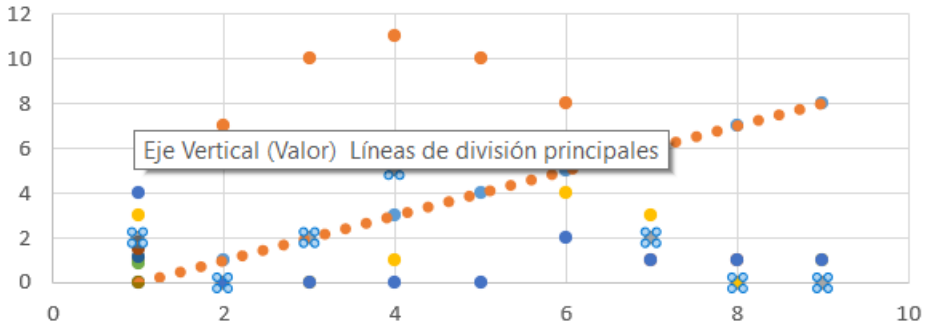
Para ver la bondad de la recta, calculamos el coeficiente de determinación lineal r^2:

$$\sigma y^2 = m_{02} - m_{01}^2 = 0,865799773 \tag{22}$$

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x \sigma y} = 0,287250167 \tag{23}$$

Este resultado nos indica que la recta de regresión de Y sobre X explica menos del 30 % de la variabilidad de Y, luego no es adecuado suponer una relación lineal entre las variables.

En el siguiente gráfico podemos ver como están distribuidos los datos respecto a la recta de regresión


12
10
8
6
4
2
0
0
2
4
6
8
10
Eje Vertical (Valor) Líneas de división principales

4.Se ha realizado un estudio sobre la tensión de vapor de agua (Y, en ml. de Hg.) a distintas temperaturas de (X, en ºC). Efectuadas 21 medidas diferentes, los resultados son:

X\Y	(0.5, 1.5]	(1.5, 2.5]	(2.5, 5.5]
(1,15]	4	2	0
(15, 25]	1	4	2
(25, 30]	0	3	5

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

En una población de tamaño n = 21 medidas se ha observado dos variables estadísticas, X = temperatura en ºC e Y = vapor de agua en ml de Hg, las cuales han presentado k = 3, p = 3 modalidades distintas, con distribución de frecuencia conjunta $(x_i y_j), n_{ij}$ i = 1,...,3 j=1,...3

Comenzamos ampliando nuestra tabla:

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

Siguiendo el procedimiento del ejercicio anterior, calculamos la recta de regresión de Y sobre X:

$$\overline{x} = \frac{408}{21} = 19,42857143 \check{z} C \tag{24}$$

$$\overline{y} = \frac{51}{21} = 2,428571429 ml de Hg \tag{25}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 62,24489796 \check{z} C^2 \tag{26}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 1,387755102 ml.Hg^2 \tag{27}$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i yj - \overline{x} * \overline{y} = 6,102040816 \tag{28}$$

Ahora que tenemos todo lo necesario calculamos la recta:

$$y = ax + b \tag{29}$$

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = 0,098032787 \tag{30}$$

$$b = \overline{y} - a\overline{x} = 0,523934426 \tag{31}$$

Luego, la recta de regresión de Y sobre X es $y = 0,098032787 * x + 0,523934426$

Para ver la bondad de la recta, calculamos el coeficiente de determinación lineal r^2:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x \sigma y} = 0,431055931 \tag{32}$$

Este resultado nos indica que la recta de regresión de Y sobre X explica menos del 45 % de la variabilidad de Y, luego no es adecuado suponer una relación lineal entre las variables(la bondad de la función es baja). Pese a esto, el coeficiente de correlación lineal $r = \sqrt{r^2} = 0,6565$ nos indica que hay bastante grado de correlación lineal directa entre las variables.

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
10	2	4	6	10	8
20	1	2	3	5	4
30	3	6	9	15	12
40	4	8	12	20	16

X\Y	1	2	3
-1	0	1	0
0	1	0	1
1	0	1	0

En el primer caso no tiene sentido calcular la curva de regresión, ya que las variables son independientes. Para verlo basta con comprobar:

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
nj.	10	20	30	50	40	

$$\frac{n_{i1}}{n_{.i}} = ... = \frac{n_{ij}}{n_{.j}} \; i = 1, ..., k \; j = 1, ..., p \tag{33}$$

$$\frac{1}{10} = \frac{2}{20} = \frac{3}{30} = \frac{5}{50} = \frac{4}{40} \tag{34}$$

$$\frac{3}{10} = \frac{6}{20} = \frac{9}{30} = \frac{15}{50} = \frac{20}{40} \tag{35}$$

$$\frac{4}{10} = \frac{8}{20} = \frac{12}{30} = \frac{20}{50} = \frac{16}{40} \tag{36}$$

Luego, las variables son independientes, por lo que $\sigma_{xy} = 0$

En el caso de la segunda distribución, no se da la condición de independencia al coexistir frecuencias nulas y no nulas.

X\Y	1	2	3	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i \sum_{j=1}^{p} n_{ij}y_j$
-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
$n_{j.}y_j$	1	4	3				
$n_{j.}y_j^2$	1	8	9				

Vamos a calcular la curva de regresión de tipo 1. Para ello, tendremos en cuenta que dicha curva pasa por los puntos $(x_i, \overline{y}_i)$ i = 1,...,k

$$Punto1 = (-1, \frac{2*1}{1}) = (-1, 2) \tag{37}$$

$$Punto2 = (0, \frac{1*1+3*1}{2}) = (0, 2) \tag{38}$$

$$Punto3 = (1, \frac{2*1}{1}) = (1, 2) \tag{39}$$

Luego, la curva de regresión pasa por dichos puntos.

6. Dada la siguiente distribución bidimensional:

X\Y	1	2	3	4
10	1	3	0	0
12	0	1	4	3
14	2	0	0	2
16	4	0	0	0

a)¿Son estadísticamente independientes?

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

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

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

Para el apartado a, basta con ver que coexisten frecuencias nulas y no nulas, por lo que no se cumplirá la condición de independencia estadística.
La condición es:

$$\frac{n_{i1}}{n_{.i}} = ... = \frac{n_{ij}}{n_{.j}} \; i = 1, ..., k \; j = 1, ..., p \tag{40}$$

Para calcular la curva de regresión de Y/X, debemos hallar los puntos por los que pasa, los cuales son, $(x_i, \overline{y}_i)$ i = 1,...,k

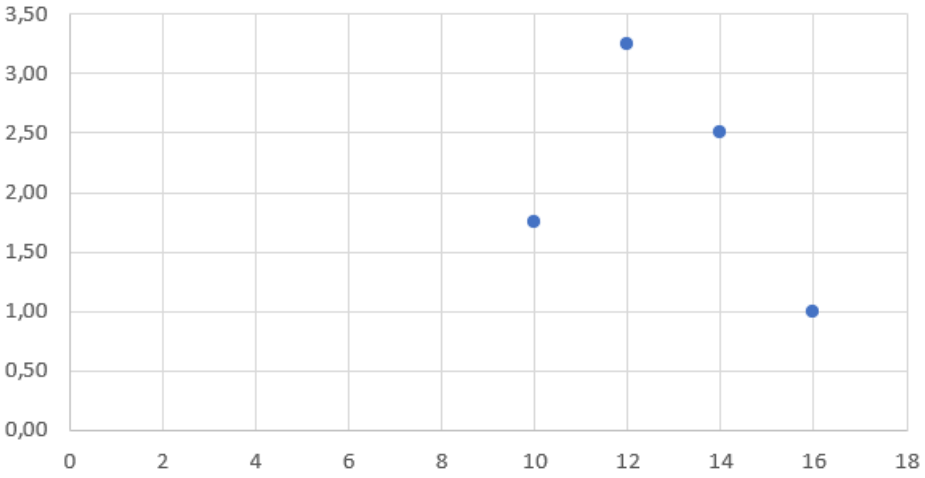
$$Punto1 = (10, \frac{3*2+1*1}{4}) = (-1, \frac{7}{4}) \tag{41}$$

$$Punto2 = (12, \frac{1*2+3*4+4*3}{8}) = (12, \frac{13}{4}) \tag{42}$$

$$Punto3 = (1, \frac{2*1+2*4}{4}) = (14, \frac{10}{4}) \tag{43}$$

$$Punto4 = (16, \frac{4*1}{4}) = (16, 1) \tag{44}$$

Luego, la curva de regresión de Y sobre X será aquella que pase por dichos puntos:



Para calcular la curva de regresión de X sobre Y calculamos los puntos $(y_i, \overline{x}_i)$ i = 1,...,k

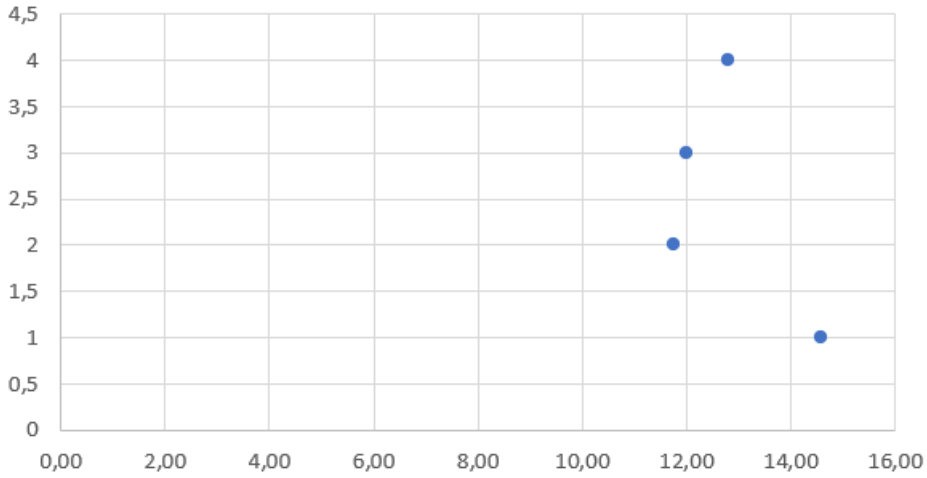
$$Punto1 = (1, \frac{1*10+2*14+4*16}{7}) = (1, \frac{102}{7}) \tag{45}$$

$$Punto2 = (2, \frac{10*3+12*1}{4}) = (2, \frac{42}{4}) \tag{46}$$

$$Punto3 = (3, \frac{4*12}{4}) = (3, 12) \tag{47}$$

$$Punto4 = (4, \frac{12*3+14*2}{5}) = (4, \frac{64}{5}) \tag{48}$$

Luego, la curva de regresión de X sobre Y será la que pase por dichos puntos:



Calculamos ahora para caso el coeficiente de correlación η^2. Para ello realizamos los siguientes cálculos:

X\Y	1	2	3	4	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i \sum_{j=1}^{p} n_{ij}y_j$
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
$n_{.j}y_j$	7	8	12	20	47			
$n_{.j}y_j^2$	7	16	36	80	139			

$$\overline{x} = \frac{256}{20} = 12,8 \tag{49}$$

$$\overline{y} = \frac{27}{20} = 2,35 \tag{50}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 4,16 \tag{51}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 1,4275 \tag{52}$$

Realizamos,además, las dos siguientes tablas:

Luego, tenemos que:

x_i	y_i	n_{ij}	f(x_i)	r_{ij}	r_{ij}^2	$n_{ij}r_{ij}^2$
10	1	1	1,75	-0,75	0,56	0,5625
10	2	3	1,75	0,25	0,06	0,1875
12	2	1	3,25	-1,25	1,56	1,5625
12	3	4	3,25	-0,25	0,06	0,25
12	4	3	3,25	0,75	0,56	1,6875
14	1	2	2,50	-1,50	2,25	4,5
14	4	2	2,50	1,50	2,25	4,5
16	1	4	1,00	0,00	0,00	0
						13,25

y_i	x_i	n_{ij}	f(y_i)	r_{ij}	r_{ij}^2	$n_{ij}r_{ij}^2$
1	10	1	14,57	-4,57	20,90	20,8979592
1	14	2	14,57	-0,57	0,33	0,65306122
1	16	4	14,57	1,43	2,04	8,16326531
2	10	3	10,50	-0,50	0,25	0,75
2	12	1	10,50	1,50	2,25	2,25
3	12	4	12,00	0,00	0,00	0
4	12	3	12,80	-0,80	0,64	1,92
4	14	2	12,80	1,20	1,44	2,88
						37,5142857

$$\sigma_{ry}^2 = \frac{13,25}{20} = 0,6625 \tag{53}$$

$$\sigma_{rx}^2 = \frac{37,51}{20} = 1,8755 \tag{54}$$

Con todo esto calculamos el coeficiente de correlación:

$$\eta_{\frac{y}{x}}^2 = 1 - \frac{\sigma_{ry}}{\sigma_y^2} = 0,5359 \tag{55}$$

$$\eta_{\frac{x}{y}}^2 = 1 - \frac{\sigma_{rx}}{\sigma_x^2} = 0,549 \tag{56}$$

Como vemos, en ambos casos la bondad es inferior al 55 %, por lo que la curva no es muy precisa para realizar estimaciones.

Porúltimo, calculamos el coeficiente de correlación lineal:

$$\sigma_{xy} = -0,78 \tag{57}$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = -0,32 \tag{58}$$

Como vemos las variables tienen baja correlación lineal e inversa.
Calculamos las expresiones de las rectas
Recta de Y sobre X:

$$y = ax + b \tag{59}$$

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = -0,1875 \tag{60}$$

$$b = \overline{y} - a\overline{x} = 4,75 \tag{61}$$

Luego, la recta de regresión de Y sobre X es $y = -0,1875x + 4,75$
Calculamos ahora la recta de regresión de X sobre Y

$$x = ay + b \tag{62}$$

$$a = \frac{\sigma_{xy}}{\sigma_y^2} = -0,55 \tag{63}$$

$$b = \overline{x} - a\overline{y} = 14 \tag{64}$$

Luego, la recta de regresión de X sobre Y es $x = -0,55y + 14$

7. Para cada una de las siguientes 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?

b) Calcular las curvas de regresión y comentar los resultados

En la distribución 1, X depende funcionalmente de Y, ya que para cada valor de x existe un único valor de y, es decir, existe un único elemento no nulo en cada fila, pero Y no depende funcinalmente de X, ya que el la columna dos hay 2 elementos no nulos.

En la distribución 2, X depende funcionalmente de Y e Y depende funcionalmente de X, ya que en cada fila y cada columna exste un único elemento no nulo, luego, a cada valor x le corresponde un único valor de y y viceversa.

En la distribución 3, Y depende funcionalmente de X, ya que en cada columna existe un único elemento no nulo, pero X no depende funcionalmente de Y, ya que en la fila 1 hay más de un elemento no nulo.

En la distribución 1, como X depende funcionalmente de Y no hay que calcular la curva de regresión, pero si tenemos que calcular la de X sobre Y. Para ello calculamos los puntos de la forma $(\overline{x_j}, y_j)$ j=1, ..., p.

$$Punto1 = (\frac{1*2}{2}, 10) = (1, 10) \tag{65}$$

$$Punto2 = (\frac{1*2+4*1}{3}, 15) = (2, 15) \tag{66}$$

$$Punto3 = (\frac{3*3}{3}, 20) = (3, 3) \tag{67}$$

Por tanto, la curva de regresión será aquella que pase por dichos puntos.

Para la distribución 2 no es necesario calcular la curva de regresión, ya que ambas variables dependen funcionalmente una de la otra.

Para la distribución 3 no es necesario calcular la curva de regresión X sobre Y, ya que Y depende funcionalmente de X, pero si hay que calcular la de Y sobre X, para lo que hallamos los puntos de la forma $(x_i, \overline{y_i})$ i = 1,...,k

$$Punto1 = (1, \frac{3*15+25*1}{4}) = (1, \frac{35}{2}) \tag{68}$$

$$Punto2 = (2, \frac{20*1}{1}) = (2, 20) \tag{69}$$

$$Punto3 = (3, \frac{2*10}{2}) = (3, 10) \tag{70}$$

Luego, la curva de regresión será la que pasa por dichos puntos.

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
1	1	2	0	0
2	1	2	3	1
3	0	1	2	6
4	0	0	2	3

a) Determinar las rectas de regresión

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

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

En una población de tamaño n = 24 se ha observado dos variables estadísticas, X = número de balanzas e Y = número de dependientes, las cuales han presentado k = 4, p = 4 modalidades distintas, con distribución de frecuencia conjunta $(x_i y_j), n_{ij}$ i = 1,...,4 j=1,...4

Comenzamos completando la tabla:

X\Y	1	2	3	4	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i \sum_{j=1}^{p} n_{ij}y_j$
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	20	80	72
$n_{.j}$	2	5	7	10	24	64	192	209
$n_{.j}y_j$	2	10	21	40	73			
$n_{.j}y_j^2$	2	20	63	160	245			

Calculemos todos los datos necesarios para rectas de regresión:

$$\overline{x} = \frac{64}{24} = 2,666666667 \tag{71}$$

$$\overline{y} = \frac{73}{24} = 3,041666667 \tag{72}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 0,888888889 \tag{73}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 0,956597222 \tag{74}$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i yj - \overline{x} * \overline{y} = 0,597222222 \tag{75}$$

Ahora que tenemos todo lo necesario calculamos las rectas de regresión:
Recta de Y sobre X:

$$y = ax + b \tag{76}$$

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = 0,671875 \tag{77}$$

$$b = \overline{y} - a\overline{x} = 1,25 \tag{78}$$

Luego, la recta de regresión de Y sobre X es $y = 0,671875x + 1,25$
Calculamos ahora la recta de regresión de X sobre Y

$$x = ay + b \tag{79}$$

$$a = \frac{\sigma_{xy}}{\sigma_y^2} = 0,624319419 \tag{80}$$

$$b = \overline{x} - a\overline{y} = 1,376814882 \tag{81}$$

Luego, la recta de regresión de X sobre Y es $x = 0,624319419y + 1,376814882$

Para ver si es apropiado suponer una relación lineal calculamos el coeficiente de determinación lineal:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x \sigma y} = 0,41946461 \tag{82}$$

Como vemos, el coeficiente de determinación lineal es, aproximadamente, $0,42$, lo que nos indica que menos del $42\,\%$ de la variabilidad de Y está explicada en X y viceversa, por lo que la bondad de la funcón es baja. Pese a esto, el coeficiente de correlación lineal $r = \sqrt{r^2} = 0,648$ nos indica que existe bastante correlación linal directa entre las variables.l
Dado y=6, veamos el pronóstico del número de dependientes:

$$x = 0,624319419 * 6 + 1,376814882 = 5,122731397 \tag{83}$$

Esta predicción no es fiable por dos motivos. Primero y más importante, estamos haciendo pronósticos fuera de los límites de la variable explicativa y, segundo, el coeficiente de determinación lineal era bastante bajo.

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

X\Y	(10, 20]	(20, 25]	(25, 30]	(30, 35]	(35, 40]
(15, 18]	3	2	3	0	0
(18, 21]	0	4	2	2	0
(21, 24]	0	7	10	6	1
(24, 27]	0	0	2	5	3

Estudia la interdependencia lineal entre ambas variables

En una población de tamaño n = 50 matrimonios se ha observado dos variables estadísticas, X = edad del hombre en el matrimonio e Y = edad de la mujer en el matrimonio, las cuales han presentado k = 4, p = 4 modalidades distintas, con distribución de frecuencia conjunta $(x_i y_j), n_{ij}$ i = 1,...,4 j=1,...4

Comenzamos completando la tabla:

X\Y	(10, 20]	(20, 25]	(25, 30]	(30, 35]	(35, 40]	c_i		$n_{i.}$	$n_i c_i$	$n_i c_i^2$	$x_i \sum_{j=1}^{p} n_i$
(15, 18]	3	2	3	0	0	16,5		8	132	2178	2846
(18, 21]	0	4	2	2	0	19,5		8	156	3042	4
(21, 24]	0	7	10	6	1	22,5		24	540	12150	1496
(24, 27]	0	0	2	5	3	25,5		10	255	6502,5	8
$n_{.j}$	3	13	17	13	4			50	1083	23872,5	30318
c_j	15	22,5	27,5	32,5	37,5						
$n_{.j} c_j$	45	292,5	467,5	422,5	150	1377,5					
$n_{.j} c_j^2$	675	6581,25	12856,25	13731,25	5625	39468,75					

Comenzamos calculando el coeficiente de correlación lineal para saber si es adecuado suponer una relación lineal, para ello:

$$\overline{x} = \frac{1083}{50} = 21,66 \tag{84}$$

$$\overline{y} = \frac{1377,5}{50} = 27,55 \tag{85}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 8,2944 \tag{86}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 30,3725 \tag{87}$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} x_i yj - \overline{x} * \overline{y} = 9,642 \tag{88}$$

Luego, el coeficiente de determinación lineal es:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x \sigma y} = 0,369036003 \tag{89}$$

Como vemos el coeficiente de determinación lineal es de 0,36, lo que significa que solo menos del 36 % de la variabilidad de la variable explicada está explicada en la variable explicativa, por lo que no tiene sentido suponer utilizar dicha recta para hacer estimaciones. Pese a ello, existe un cierto grado de correlación lineal directa entre las variables, el cual, viene indicado por $r = \sqrt{r^2} = 0,6$

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

$$x + 4y = 1 \tag{90}$$

$$x + 5y = 2 \tag{91}$$

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

$$y = \frac{-1}{5}x + \frac{2}{5}x = -4y + 1 \tag{92}$$

(Las pendientes deben tener el mismo signo o no serían rectas de regresión)
Como sabemos que:

$$a = \frac{\sigma_{xy}}{\sigma_x^2}, a' = \frac{\sigma_{xy}}{\sigma_y^2} \implies a * a' = \frac{\sigma_{xy}^2}{\sigma x^2 \sigma_y^2} = r^2 = \frac{4}{5} \implies r = \sqrt{r^2} = \frac{-2}{\sqrt{5}} \tag{93}$$

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:

xi	3	5	8	9
ni.	5	1	2	1

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

Sabemos que $x = ay + b$, de donde $a = \frac{\sigma_{xy}}{\sigma_y^2}$ y $b = \overline{x} - a\overline{y}$.

Tenemos que:

$$y = 5x + 20 \implies a' = 5 = \frac{\sigma_{xy}}{\sigma_x^2} \tag{94}$$

$$b = 20 = \overline{y} - a'\overline{x} \tag{95}$$

Necesitamos calcular los siguientes datos:

$$m_{02} = \frac{1}{n}\sum y_i^2 n_{.j} = 360 \tag{96}$$

$$m_{01} = \overline{y} \tag{97}$$

$$\overline{x} = \frac{1}{9}\sum x_i n_{i.} = 5 \tag{98}$$

$$y - \overline{y} = 5(x - \overline{x}) \implies \overline{y} - 5\overline{x} = -20 \implies \overline{y} = 5 \tag{99}$$

$$\sigma_y^2 = m_{02} - m_{01}^2 = 360 - 25 = 335 \tag{100}$$

$$\sigma_x^2 = m_{02} - m_{10}^2 = \frac{1}{9}\sum x_i 2n_{i.} - 25 = 6 \tag{101}$$

$$\frac{\sigma_{xy}}{\sigma_y^2} = \frac{\sigma_{xy}}{\sigma_x^2}\frac{\sigma x^2}{\sigma_y^2} = 5\frac{6}{335} = 0,0896 \tag{102}$$

$$x - \overline{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \overline{y}) \implies x = 0,0896y + 4,552 \tag{103}$$

Calculamos la bondad de la función

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y 2} = 0,448 \tag{104}$$

Solo el 44,8 % de la variabilidad de X está explicada en Y, por lo que no es bueno suponer una relación lineal.

De las estadísticas de "Tiempo de vuelo y consumo de combustilbe" 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:

$$\sum y_i = 219,719, \sum y_i^2 = 2396,504, \sum x_i y_i = 349,486, \sum x_i = 31,470, \tag{105}$$

$$\sum x_i^2 = 51,75, \sum x_i^2 y_i = 633,993, \sum x_i^4 = 182,997, \sum x_i^3 = 93,6 \tag{106}$$

12. 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 del programa de vuelos compuesto de 100 vuelos de media hora, 200 de una hora y 100 de dos horas?¿Es fiable esta estimación?

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)?

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

Comenzamos con el apartado a. Para ello calculamos:

$$\overline{x} = \frac{31,470}{24} = 1,311 \tag{107}$$

$$\overline{y} = \frac{219,719}{24} = 9,155 \tag{108}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = \frac{51,075}{24} - 1,311^2 = 0,409 \tag{109}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = \frac{2396,504}{24} - 9,155^2 = 16,04 \tag{110}$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} x_i yj - \overline{x} * \overline{y} = \frac{349,486}{24} - 1,311 * 9,155 = 2,5597 \tag{111}$$

Luego, para calcular la recta $y = ax + b$, calculamos:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = 6,2584 \tag{112}$$

$$b = \overline{y} - a\overline{x} = 0,95 \tag{113}$$

Luego, la recta de regresión de Y sobre X es $y = 6,2584x + 0,95$

Calculamos el coeficiente de determinación lineal para ver la bondad de la función:

$$\eta^2 = r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y 2} = 0,99 \tag{114}$$

El 99 % de la variabilidad de Y está explicado en X, por lo que, si las estimaciones que vamos a hacer a continuación están dentro de los límites de la variable explicativa, estas serán muy fiables.

Para 100 vuelos de media hora:

$100 * (6,2584 * 0,5 + 0,95) = 407,8$ miles de libras

Para 200 vuelos de una hora:

$200 * (6,2584 * 1 + 0,95) = 1441,7$ miles de libras

Para 100 vuelos de dos horas:

$100 * (6,2584 * 2 + 0,95) = 1346,68$ miles de libras.

Ahora, hacemos el ajuste al modelo $Y = a + bX + cX^2$. Para ello hay que resolver el siguiente sistema:

$$m_{01} = a + bm_{10} + cm_{20} \tag{115}$$

$$m_{11} = am_{10} + bm_{20} + cm_{30} \tag{116}$$

$$m_{21} = am_{20} + b_m 30 + cm_{40} \tag{117}$$

De este sistema se obtiene la siguiente solución: $a = 0,86$ $b = 6,43$ y $c = -0,069 \implies Y = -0,069 * X^2 + 6,43 * X + 0,86$

Usando este pronóstico las estimaciones son las siguientes:

Para 100 vuelos de media hora:

$100 * (-0,069 * 0,5^2 + 6,43 * 0,5 + 0,86) = 405,775$ miles de libras

Para 200 vuelos de una hora:

$100 * (-0,069 * 1^2 + 6,43 * 1 + 0,86) = 1444,2$ miles de libras

Para 100 vuelos de dos horas:

$100 * (-0,069 * 2^2 + 6,43 * 2 + 0,86) = 1346,4$ miles de libras

Por último, no podemos calcular la bondad de la función (η^2), ya que el problema no nos proporciona los valores de x_i. Debido a esto, lo único que podemos afirmar es que la bondad de la parábola será mayor o igual a la de la recta; por lo que en principio, es mejor suponer el segundo modelo.

13. La curva de Engel, que exprsa 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 dats, 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

Comenzamos haciendo un cambio de variable:

$$X' = \frac{1}{X}$$

X'	0,1	0,08	0,05	0,04
Y	50	90	160	180

Calculamos la función de regresión de Y'sobre X':

X'\Y	50	90	160	180	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i \sum_{j=1}^{p} n_{ij}y_j$
0,1	1	0	0	0	1	0,1	0,01	5
0,08	0	1	0	0	1	0,08	0,0064	7,2
0,05	0	0	1	0	1	0,05	0,0025	8
0,04	0	0	0	1	1	0,04	0,0016	7,2
$n_{.j}$	1	1	1	1	4	0,27	0,0205	27,4
$n_{.j}y_j$	50	90	160	180	480			
$n_{.j}y_j^2$	2500	8100	25600	32400	68600			

$$\overline{x}' = \frac{0,27}{4} = 0,0675 \tag{118}$$

$$\overline{y}' = \frac{480}{4} = 120 \tag{119}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 0,00056875 \tag{120}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 2750 \tag{121}$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i yj - \overline{x} * \overline{y} = -1,25 \tag{122}$$

Luego, para calcular la recta $y' = ax' + b$, calculamos:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = -2197,802198 \tag{123}$$

$$b = \overline{y} - a\overline{x} = 268,3516484 \tag{124}$$

Luego, la recta de regresión de Y sobre X es $y = -2197,8 * x' + 268,35$
Deshaciendo el cambio de variable y reordenando la función $\Longrightarrow$
$y = -2197,8 * \frac{1}{x} + 268,35$

Para calcular la bondad realizamos la siguiente tabla:

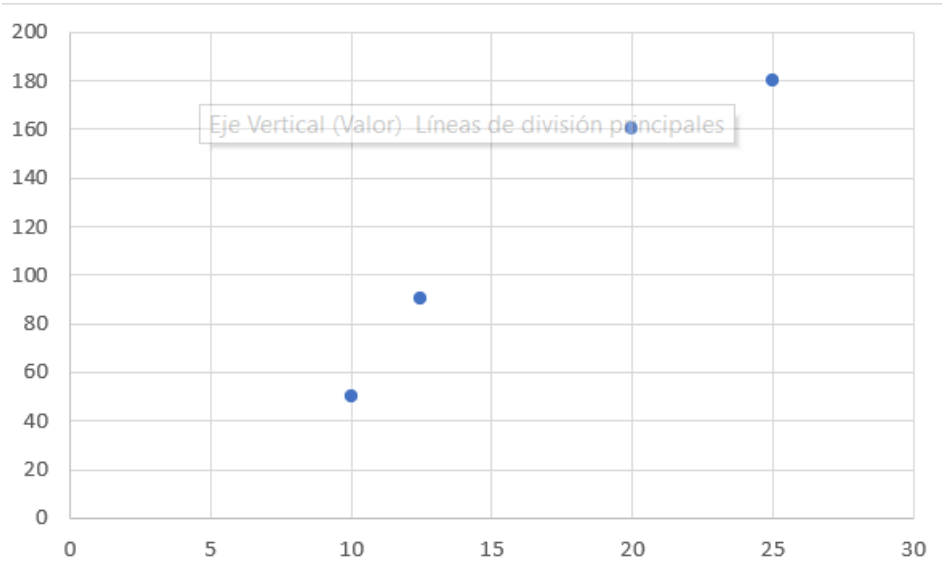
x_i	y_i	n_{ij}	f(x_i)	r_{ij}	r_{ij}^2	$n_{ij}r_{ij}^2$
10	50	48,57	1	1,43	2,0449	2,0449
12,5	90	92,526	1	-2,526	6,380676	6,380676
20	160	158,46	1	1,54	2,3716	2,3716
25	180	180,438	1	-0,438	0,191844	0,191844
						10,98902

$$\sigma_{r_y}^2 = \frac{10,98902}{4} = 2,747255 \tag{125}$$

Luego, tenemos que:

$$\eta_{\frac{y}{x}}^2 = 1 - \frac{\sigma_{r_y}^2}{\sigma_y^2} = 0,999000998 \tag{126}$$

Como vemos la bondad es prácticamente del 100 %. De hecho, si nos fijamos en la distribución, vemos que ambas variables son funcionalmete dependientes, ya que en cada fila y en cada columna existe un único elemento no nulo.Si nos fijamos antentamente en la disposición de los puntos, vemos como,por dichos puntos, puede pasar una rama de una hipérbola equilátera.


200
180
160
140
120
100
80
60
40
20
0
0
5
10
15
20
25
30
Eje Vertical (Valor) Líneas de división principales

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:

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

Explicar el comportamiento de Y por X mediante:

a) Relación lineal.

b) Hipérbola equilátera.

c) Curva potencial.

d) Curva exponencial.

¿Qué ajuste es más adecuado?

Comenzamos completando la tabla:

X\Y	30	50	70	80	120	140	$n_{i.}$	$n_i c_i$	$n_i c_i^2$	$x_i \sum_{j=1}^{p} n_{ij} y_j$
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	2058	9930
$n_{.j} y_j$	30	50	70	80	120	140	490			
$n_{.j} y_j^2$	900	2500	4900	6400	14400	19600	48700			

Comenzemos con la relación lineal:

$$\overline{x} = 16,66666667 \tag{127}$$

$$\overline{y} = 81,66666667 \tag{128}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 65,22222222 \tag{129}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 1447,222222 \tag{130}$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} x_i yj - \overline{x} * \overline{y} = 293,8888889 \tag{131}$$

Luego, para calcular la recta $y = ax + b$, calculamos:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = 4,5 \tag{132}$$

$$b = \overline{y} - a\overline{x} = 6,567 \tag{133}$$

Luego, la recta de regresión de Y sobre X es $y = 4,5 * x - 6,567$

Calculamos el coeficiente de determinación lineal para ver la bondad de la función:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma : x^2 \sigma_y^2} = 0,915030393 \tag{134}$$

El 90 % de la variabilidad de Y está explicada en X, por lo que la bondad de la función es alta.

Calculamos ahora la hipérbola equilatera

Hacemos el siguiente cambio de variable X'= $\frac{1}{X}$, luego, nuestra tabla ahora será

X'\Y	30	50	70	80	120	140	$n_{i.}$	$n_i c_i$	$n_i c_i^2$	$x_i \sum_{j=1}^{p} n_{ij} y_j$
0,11111111	1	0	0	0	0	0	1	0,11111111	0,01234568	3,33333333
0,1	0	1	0	0	0	0	1	0,1	0,01	5
0,08333333	0	0	1	0	0	0	1	0,08333333	0,00694444	5,83333333
0,06666667	0	0	0	1	0	0	1	0,06666667	0,00444444	5,33333333
0,04545455	0	0	0	0	1	0	1	0,04545455	0,00206612	5,45454545
0,03125	0	0	0	0	0	1	1	0,03125	0,00097656	4,375
$n_{.j}$	1	1	1	1	1	1	6	0,43781566	0,03677725	29,3295455
$n_{.j} y_j$	30	50	70	80	120	140	490			
$n_{.j} y_j^2$	900	2500	4900	6400	14400	19600	48700			

$$\overline{x}' = 0,072969276 \tag{135}$$

$$\overline{y}' = 81,66666667 \tag{136}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 0,000805026 \tag{137}$$

$$\sigma y^2 = m_{02} - m_{01}^2 = 1447,222222 \tag{138}$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} x_i yj - \overline{x} * \overline{y} = -1,070899972 \tag{139}$$

Luego, para calcular la recta $y' = ax' + b$, calculamos:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = -1330,26795 \tag{140}$$

$$b = \overline{y} - a\overline{x} = 178,735356 \tag{141}$$

Luego, la hipérbola equilatera será $y = -1330,26795 * x' + 178,74 \implies y = -1330,26794\frac{1}{x} + 178,74$

Calculamos la razón de correlación para ver la bondad de la función:

x_i	y_i	n_{ij}	f(x_i)	r_{ij}	r_{ij}^2	$n_{ij}r_{ij}^2$
9	30	1	30,9325556	-0,93255556	0,86965986	7,82693878
10	50	1	45,7133	4,2867	18,3757969	183,757969
12	70	1	67,8844167	2,11558333	4,47569284	53,7083141
15	80	1	90,0555333	-10,0555333	101,113751	1516,70626
22	120	1	118,273318	1,72668182	2,9814301	65,5914622
32	140	1	137,169156	2,83084375	8,01367634	256,437643
						2084,02859

$$\eta^2 = \frac{\sigma_{ey}^2}{\sigma_y^2} = 0,984357389 \tag{142}$$

Calculamos ahora el modelo de curva potencial:

Hacemos un nuevo cambio de variable $\implies y = bx^a \implies log(y) = a * log(x) + log(b) \implies V = log(y), B = log(b), W = log(x)$

Luego, tenemos que:

W\V	3,40119738	3,91202301	4,24849524	4,38202663	4,78749174	4,94164242	$n_{i.}$
2,19722458	1	0	0	0	0	0	1
2,30258509	0	1	0	0	0	0	1
2,48490665	0	0	1	0	0	0	1
2,7080502	0	0	0	1	0	0	1
3,09104245	0	0	0	0	1	0	1
3,4657359	0	0	0	0	0	1	1
$n_{.j}$	1	1	1	1	1	1	6
$n_{.j}y_j$	3,40119738	3,91202301	4,24849524	4,38202663	4,78749174	4,94164242	25,6728764
$n_{.j}y_j^2$	11,5681436	15,303924	18,0497118	19,2021574	22,9200772	24,4198298	111,463844

$$\overline{w} = 2,70825748 \tag{143}$$

$$\overline{v} = 4,278812738 \tag{144}$$

$n_i c_i$	$n_i c_i^2$	$x_i \sum_{j=1}^{p} n_{ij} y_j$
2,19722458	4,82779584	7,47319448
2,30258509	5,30189811	9,00776586
2,48490665	6,17476106	10,5571141
2,7080502	7,33353589	11,8667481
3,09104245	9,55454345	14,7983402
3,4657359	12,0113253	17,1264276
16,2495449	45,2038597	70,8295903

$$\sigma_w^2 = m_{20} - m_{10}^2 = 0,199318041 \tag{145}$$

$$\sigma v^2 = m_{02} - m_{01}^2 = 0,269068867 \tag{146}$$

$$\sigma_{wv} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} x_i yj - \overline{w} * \overline{v} = 0,216805116 \tag{147}$$

Luego, para calcular la recta $v = aw + b$, calculamos:

$$a = \frac{\sigma_{wv}}{\sigma_w^2} = 1,087734532 \tag{148}$$

$$b = \overline{v} - a\overline{w} = 1,332947556 \tag{149}$$

Luego, la curva pontencial será $v = 1,087734532 * w + 1,332947556 \implies$
$log(y) = 1,087734532 * log(x) + 1,332947556 \implies$
$y = x^{1,087734532} * 3,792204666$

Calculamos ahora la bondad de la función:

x_i	y_i	n_{ij}	f(x_i)	r_{ij}	r_{ij}^2	$n_{ij} r_{ij}^2$
9	30	1	41,3860805	-11,3860805	129,642828	129,642828
10	50	1	46,4115743	3,58842569	12,8767989	12,8767989
12	70	1	56,5919262	13,4080738	179,776444	179,776444
15	80	1	72,1384561	7,86154387	61,8038721	61,8038721
22	120	1	109,41863	10,5813695	111,965381	111,965381
32	140	1	164,473288	-24,4732884	598,941844	598,941844
						1095,00717

$$\eta^2 = \frac{\sigma_{ey}^2}{\sigma_y^2} = 0,873895528 \tag{150}$$

Por último, calculamos la curva exponencial, para lo que hacemos el siguiente cambio de variable:

$y = ba^x \implies log(y) = log(b) + x * log(a) \implies V = log(y), B = log(b), A = log(a)$

X\V	3,40119738	3,91202301	4,24849524	4,38202663	4,78749174	4,94164242	$n_{i.}$
9	1	0	0	0	0	0	1
10	0	1	0	0	0	0	1
12	0	0	1	0	0	0	1
15	0	0	0	1	0	0	1
22	0	0	0	0	1	0	1
32	0	0	0	0	0	1	1
$n_{.j}$	1	1	1	1	1	1	6
$n_{.j}y_j$	3,40119738	3,91202301	4,24849524	4,38202663	4,78749174	4,94164242	25,6728764
$n_{.j}y_j^2$	11,5681436	15,303924	18,0497118	19,2021574	22,9200772	24,4198298	111,463844

$n_i c_i$	$n_i c_i^2$	$x_i \sum_{j=1}^{p} n_{ij} y_j$
9	81	30,6107764
10	100	39,1202301
12	144	50,9819429
15	225	65,7303995
22	484	105,324818
32	1024	158,132558
100	2058	449,900725

$$\overline{x} = 16,66666667 \tag{151}$$

$$\overline{v} = 4,278812738 \tag{152}$$

$$\sigma_x^2 = m_{20} - m_{10}^2 = 65,22222222 \tag{153}$$

$$\sigma v^2 = m_{02} - m_{01}^2 = 0,269068867 \tag{154}$$

$$\sigma_{xv} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij} x_i yj - \overline{x} * \overline{v} = 3,669908493 \tag{155}$$

Luego, para calcular la recta $v = aw + b$, calculamos:

$$a = \frac{\sigma_{xv}}{\sigma_x^2} = 0,056267762 \tag{156}$$

$$b = \overline{v} - a\overline{x} = 3,341016701 \tag{157}$$

Luego, la curva pontencial será $v = 0,056267762 * x + 3,341016701 \implies y = 1,057915^x * 28,2191267$

Calculamos, por último, la bondad del ajuste

x_i	y_i	n_{ij}	f(x_i)	r_{ij}	r_{ij}^2	$n_{ij}r_{ij}^2$
9	30	1	46,8382375	-16,8382375	283,526241	283,526241
10	50	1	49,550874	0,44912601	0,20171417	0,20171417
12	70	1	55,4565527	14,5434473	211,511861	211,511861
15	80	1	65,6606525	14,3393475	205,616888	205,616888
22	120	1	97,377948	22,622052	511,757236	511,757236
32	140	1	170,989077	-30,9890771	960,322901	960,322901
						2172,93684

$$\eta^2 = \frac{\sigma_{ey}^2}{\sigma_y^2} = 0,749757753 \tag{158}$$

Conclusión: El ajuste más adecuado es el hiperbólico equilátero