

Ejercicio 1: Variable discreta. Datos con frecuencias. Tabla de doble entrada

En una clase compuesta por 30 alumnos se ha hecho un estudio sobre el número de horas diarias de estudio X y el número de suspensos Y, obteniéndose los siguientes resultados: (2,0) (2,2) (0,5) (2,1), (1,2) (2,1) (3,1) (4,0) (0,4) (2,2) (2,1) (2,1) (4,0) (3,1) (2,4) (2,1) (1,2) (2,1) (2,0) (3,0) (3,2) (2,2) (2,2) (2,1) (0,5) (1,3) (2,2) (2,1) (1,3) (1,4).

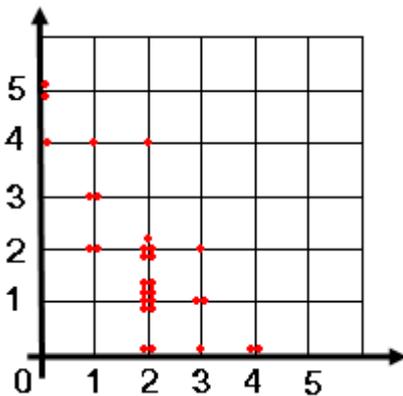
a) Construye la tabla estadística bidimensional de doble entrada y las tablas de distribución marginales

Y \ X	0	1	2	3	4	Totales
0	0	0	2	1	2	5
1	0	0	8	2	0	10
2	0	2	5	1	0	8
3	0	2	0	0	0	2
4	1	1	1	0	0	3
5	2	0	0	0	0	2
Totales	3	5	16	4	2	30

x_i	f_i
0	3
1	5
2	16
3	4
4	2
	30

y_i	f_i
0	5
1	10
2	8
3	2
4	3
5	2
	30

b) Realiza el diagrama de dispersión



c) Calcula media y desviación típica de cada variable

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
0	3	0	0
1	5	5	5
2	16	32	64
3	4	12	36
4	2	8	32
	30	57	137

y_i	f_i	$y_i \cdot f_i$	$y_i^2 \cdot f_i$
0	5	0	0
1	10	10	10
2	8	16	32
3	2	6	18
4	3	12	48
5	2	10	50
	30	54	158

$$\bar{x} = \frac{57}{30} = 1,9$$

$$\bar{y} = \frac{54}{30} = 1,8$$

$$\sigma_x = \sqrt{\frac{137}{30} - 1,9^2} = 0,978$$

$$\sigma_y = \sqrt{\frac{158}{30} - 1,8^2} = 1,424$$

d) Analiza la correlación lineal calculando el coeficiente de correlación de Pearson

x_i	2	3	4	2	3	1	2	3	1	0	1	2	0	
y_i	0	0	0	1	1	2	2	2	3	4	4	4	5	Suma
f_i	2	1	2	8	2	2	5	1	2	1	1	1	2	30
$x_i \cdot y_i \cdot f_i$	0	0	0	16	6	4	20	6	6	0	4	8	0	70

$$\sigma_{xy} = \frac{70}{30} - 1,9 \cdot 1,8 = -1,08666 \quad r = \frac{-1,08666}{0,978 \cdot 1,424} = -0,78 \text{ La dependencia es moderadamente alta y negativa}$$

e) Calcula las rectas de regresión

$$\text{De Y sobre X: } y - 1,8 = \frac{-1,08666}{0,95666}(x - 1,9) \quad \Rightarrow \quad y = -1,136x + 3,958$$

$$\text{De X sobre Y: } x - 1,9 = \frac{-1,08666}{2,02666}(y - 1,8) \quad \Rightarrow \quad x = -0,536y + 2,865$$

f) Estima el número de suspensos de un alumno que ha estudiado media hora

$$x=0,5 \quad \Rightarrow \quad y = -1,136 \cdot 0,5 + 3,958 = 3,39 \text{ Entre tres y cuatro suspensos}$$

g) Estima las horas de estudio diarias de un alumno con cinco suspensos

$$y=5 \quad \Rightarrow \quad x = -0,536 \cdot 5 + 2,865 = 0,185 \text{ horas (unos doce minutos)}$$

Ejercicio 2: Variable discreta. Datos sin frecuencia. Tabla simple

Una empresa dedicada a la elaboración y venta de ropa para jóvenes ha realizado gastos en publicidad y ha obtenido las ventas que figuran en la siguiente tabla. Llamamos X a la variable "gastos en publicidad" e Y a "beneficios de ventas". Los datos vienen expresados en miles de euros y se refieren a los últimos diez años

X	7,5	8	8,5	10	10,5	12	13	14	15	18
Y	200	205	230	240	250	270	280	300	310	325

a) Calcula el coeficiente de correlación. Analiza la dependencia de ambas variables

x_i	x_i^2	y_i	y_i^2	$x_i \cdot y_i$
7,5	56,25	200	40000	1500
8	64	205	42025	1640
8,5	72,25	230	52900	1955
10	100	240	57600	2400
10,5	110,25	250	62500	2625
12	144	270	72900	3240
13	169	280	78400	3640
14	196	300	90000	4200
15	225	310	96100	4650
18	324	325	105625	5850
116,5	1460,75	2610	698050	31700

$$\bar{x} = \frac{116,5}{10} = 11,65 \quad \sigma_x = \sqrt{\frac{1460,75}{10} - 11,65^2} = 3,22$$

$$\bar{y} = \frac{2610}{10} = 261 \quad \sigma_y = \sqrt{\frac{698050}{10} - 261^2} = 41,04$$

$$\sigma_{xy} = \frac{31700}{10} - 11,65 \cdot 261 = 129,35$$

$$r = \frac{129,35}{3,22 \cdot 41,04} = 0,98$$

La dependencia es muy alta y positiva

b) Calcula las rectas de regresión

$$\text{De Y sobre X: } y - 261 = \frac{129,35}{3,22^2} (x - 11,65) \quad \Rightarrow \quad y = 12,49x + 115,44$$

$$\text{De X sobre Y: } x - 11,65 = \frac{129,35}{41,04^2} (y - 261) \quad \Rightarrow \quad x = 0,0768y - 8,3944$$

c) La empresa decide invertir el próximo año 25000 € en publicidad. Si se mantiene la misma tendencia de los años anteriores ¿cuál será el volumen de ventas esperado?

$$x=25 \quad \Rightarrow \quad y = 12,49 \cdot 25 + 115,44 = 427,69 \quad \text{se esperan unas ventas de 427690 €}$$

d) Si la empresa desea lograr 500000 € de ventas, ¿Cuánto debe invertir en publicidad?

$$y=500 \quad \Rightarrow \quad x = 0,0768 \cdot 500 - 8,3944 = 30,0056 \quad \text{se debe invertir 30005,6 € en publicidad}$$

Ejercicio 3: Variable continua. Datos con frecuencia. Tabla doble entrada

Las notas en Matemáticas (X) y Lengua (Y) de de 30 alumnos de 1º de Bachillerato son:

Y \ X	[2,5, 5)	[5, 6)	[6, 7)	[7, 8,5)	[8,5, 10]	
[2,5, 5)	4	2	1	0	0	7
[5, 6)	2	4	2	1	0	9
[6, 7)	1	2	3	0	0	6
[7, 8,5)	0	1	2	1	2	6
[8,5, 10]	0	0	0	1	1	2
	7	9	8	3	3	30

a) Calcula media y desviación típica de cada variable y el coeficiente de correlación

x_i	y_i	f_i	$x_i \cdot f_i$	$y_i \cdot f_i$	$x_i^2 \cdot f_i$	$y_i^2 \cdot f_i$	$x_i \cdot y_i \cdot f_i$
3,75	3,75	4	15	15	56,25	56,25	56,25
3,75	5,5	2	7,5	11	28,125	60,5	41,25
3,75	6,5	1	3,75	6,5	14,0625	42,25	24,375
5,5	3,75	2	11	7,5	60,5	28,125	41,25
5,5	5,5	4	22	22	121	121	121
5,5	6,5	2	11	13	60,5	84,5	71,5
5,5	7,75	1	5,5	7,75	30,25	60,0625	42,625
6,5	3,75	1	6,5	3,75	42,25	14,0625	24,375
6,5	5,5	2	13	11	84,5	60,5	71,5
6,5	6,5	3	19,5	19,5	126,75	126,75	126,75
6,5	7,75	2	13	15,5	84,5	120,125	100,75
7,75	5,5	1	7,75	5,5	60,0625	30,25	42,625
7,75	7,75	1	7,75	7,75	60,0625	60,0625	60,0625
7,75	9,25	1	7,75	9,25	60,0625	85,5625	71,6875
9,25	7,75	2	18,5	15,5	171,125	120,125	143,375
9,25	9,25	1	9,25	9,25	85,5625	85,5625	85,5625
			178,75	179,75	1145,5625	1155,6875	1124,9375

$$\bar{x} = \frac{178,75}{30} = 5,96 \quad \sigma_x = \sqrt{\frac{1145,5625}{30} - 5,96^2} = 1,63$$

$$\bar{y} = \frac{179,75}{30} = 5,99 \quad \sigma_y = \sqrt{\frac{1155,6875}{30} - 5,99^2} = 1,63$$

$$\sigma_{xy} = \frac{1124,9375}{30} - 5,96 \cdot 5,99 = 1,8$$

$$r = \frac{1,8}{1,63 \cdot 1,63} = 0,68 \quad (r^2 \cdot 100 = 46,24\% \text{ de fiabilidad}) \text{ correlación positiva y débil}$$

b) Calcula las rectas de regresión

$$\text{De Y sobre X: } y - 5,99 = \frac{1,8}{1,63^2} (x - 5,96) \quad \Leftrightarrow \quad y = 0,67x + 1,95$$

$$\text{De X sobre Y: } x - 5,96 = \frac{1,8}{1,63^2} (y - 5,99) \quad \Leftrightarrow \quad x = 0,67y + 1,9$$

c) Estima la nota en Lengua de un alumno que saca un 7,3 en Matemáticas

$$x = 7,3 \quad y = 0,67 \cdot 7,3 + 1,95 = 6,84$$

d) Estima la nota en Matemáticas de un alumno que ha sacado un 3,1 en Lengua

$$y = 3,1 \quad x = 0,67 \cdot 3,1 + 1,9 = 3,97$$

12) Se ha observado una variable estadística bidimensional y se han obtenido los datos de la tabla adjunta.

- Calcula la covarianza
- Obtén e interpreta el coeficiente de correlación lineal
- Determina la recta de regresión de y sobre x

	X	100	50	25	
Y					
	14	1	1	0	2
	18	2	3	0	5
	22	0	1	2	3
		3	5	2	10

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
100	3	300	30000
50	5	250	12500
25	2	50	1250
	10	600	43750

$$\bar{x} = \frac{600}{10} = 60 \quad \sigma_x = \sqrt{\frac{43750}{10} - 60^2} = 27,83$$

y_i	f_i	$y_i \cdot f_i$	$y_i^2 \cdot f_i$
14	2	28	392
18	5	90	1620
22	3	66	1452
	10	184	3464

$$\bar{y} = \frac{184}{10} = 18,4 \quad \sigma_y = \sqrt{\frac{3464}{10} - 18,4^2} = 2,8$$

x_i	100	50	100	50	50	25	
y_i	14	14	18	18	22	22	Suma
f_i	1	1	2	3	1	2	10
$x_i \cdot y_i \cdot f_i$	1400	700	3600	2700	1100	1100	10600

$$a) \sigma_{xy} = \frac{10600}{10} - 60 \cdot 18,4 = -44$$

$$b) r = \frac{-44}{27,83 \cdot 2,8} = -0,56 \text{ Correl. débil y negativa}$$

$$c) \text{ Recta de regresión de Y sobre X: } y - 18,4 = \frac{-44}{774,51} (x - 60)$$

13) Una compañía discográfica ha recopilado la siguiente información sobre el número de conciertos dados, durante el verano, por 15 grupos musicales y las ventas de discos de estos grupos (expresadas en miles de CDs) obteniendo los siguientes datos.

- Calcula el número medio de CDs vendidos
- ¿Cómo es el grado de dependencia lineal del número de conciertos dado por el grupo con respecto al número de discos que ha vendido?
- Obtén la recta de regresión que explica la dependencia anterior
- Si un grupo musical ha vendido 18000 CDs ¿Qué número de conciertos es previsible que dé?

	Conciertos	10-30	30-40	40-80	
CDs					
	1-5	3	0	0	3
	5-10	1	4	1	6
	10-20	0	1	5	6
		4	5	6	15

Como los datos están agrupados en intervalos, de cada uno de ellos hay que calcular la marca de clase, el punto medio del intervalo y esas marcas de clase son los valores de las variables. Llamamos X al número de CDs vendidos (en miles) e Y al número de conciertos

a)

x_i	x_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
3	3	9	27
7,5	6	45	337,5
15	6	90	1350
	15	144	1714,5

$$\bar{x} = \frac{144}{15} = 9,6 \quad \sigma_x = \sqrt{\frac{1714,5}{15} - 9,6^2} = 4,705$$

y_i	f_i	$y_i \cdot f_i$	$y_i^2 \cdot f_i$
20	4	80	1600
35	5	175	6125
60	6	360	21600
	15	615	29325

$$\bar{y} = \frac{615}{15} = 41 \quad \sigma_y = \sqrt{\frac{29325}{15} - 41^2} = 16,55$$

x_i	3	7,5	7,5	7,5	15	15	
y_i	20	20	35	60	35	60	Suma
f_i	3	1	4	1	1	5	15
$x_i \cdot y_i \cdot f_i$	180	150	1050	450	525	4500	6855

$$b) \sigma_{xy} = \frac{6855}{15} - 9,6 \cdot 41 = 63,4$$

$$r = \frac{63,4}{4,705 \cdot 16,55} = 0,814 \text{ La dependencia es moderada y positiva}$$

$$c) \text{ La recta de regresión es: } y - 41 = \frac{63,4}{4,705^2} (x - 9,6) \quad y - 41 = 2,8636 \cdot (x - 9,6)$$

$$d) \text{ Si } x = 18 \quad y - 41 = 2,8636 \cdot (18 - 9,6) \quad \text{c} \quad y = 65,05 \text{ Han dado 65 conciertos}$$

14) La estadística de ingresos de determinadas empresas, en miles de euros, y de empleados, en miles es la siguiente:

Ingresos (X)	5,7	3,8	1,9	1	1
Empleados (Y)	16	29	17	6	9

a) Estudia la correlación existente entre ambas variables. Como la frecuencia de los datos es uno, evitamos multiplicar por f_i y todas las operaciones pueden ir en una sola tabla, teniendo en cuenta que el número de datos es 5:

						Suma
x_i	5,7	3,8	1,9	1	1	13,4
y_i	16	29	17	6	9	77
x_i^2	32,49	14,44	3,61	1	1	52,54
y_i^2	256	841	289	36	81	1503
$x_i \cdot y_i$	91,2	110,2	32,3	6	9	248,7

$$\bar{x} = \frac{13,4}{5} = 2,68 \quad \sigma_x = \sqrt{\frac{52,54}{5} - 2,68^2} = 1,98$$

$$\bar{y} = \frac{77}{5} = 15,4 \quad \sigma_y = \sqrt{\frac{1503}{5} - 15,4^2} = 7,96$$

$$\sigma_{xy} = \frac{248,7}{5} - 2,68 \cdot 15,4 = 8,47 \quad r = \frac{8,47}{1,98 \cdot 7,96} = 0,54$$

la correlación es débil y positiva

b) La recta de regresión es $y - 15,4 = \frac{8,47}{3,92}(x - 2,68)$

15) Dada la distribución bidimensional (X,Y):

X	5	6,5	8	4	3
Y	4,5	7	7,5	5	3,5

a) Calcula el coeficiente de correlación lineal, interpretando el resultado. Como la frecuencia de los datos es uno, evitamos multiplicar por f_i

						Suma
x_i	5	6,5	8	4	3	26,5
y_i	4,5	7	7,5	5	3,5	27,5
x_i^2	25	42,25	64	16	9	156,3
y_i^2	20,25	49	56,3	25	12,3	162,8
$x_i \cdot y_i$	22,5	45,5	60	20	10,5	158,5

$$\bar{x} = \frac{26,5}{5} = 5,3 \quad \sigma_x = \sqrt{\frac{156,3}{5} - 5,3^2} = 1,78$$

$$\bar{y} = \frac{27,5}{5} = 5,5 \quad \sigma_y = \sqrt{\frac{162,8}{5} - 5,5^2} = 1,52$$

$$\sigma_{xy} = \frac{158,5}{5} - 5,3 \cdot 5,5 = 2,55 \quad r = \frac{2,55}{1,78 \cdot 1,52} = 0,94$$

Correlación muy alta

y positiva

b) Las rectas de regresión son: De Y sobre X: $y - 5,5 = \frac{2,55}{3,17}(x - 5,3)$; de X sobre Y: $x - 5,3 = \frac{2,55}{2,31}(y - 5,5)$

c) Ambas rectas se cortan en el punto (5,3, 5,5) que es el punto de coordenadas (\bar{x}, \bar{y})

16) Se observaron las edades de cinco niños y sus pesos respectivos, obteniéndose los siguientes resultados:

Edad en años (X)	2	4,5	6	7,2	8
Peso en Kg (Y)	15	19	25	33	34

a) Calcula el coeficiente de correlación lineal, y rectas de regresión. Como la frecuencia de los datos es uno, evitamos multiplicar por f_i

						Suma
x_i	2	4,5	6	7,2	8	27,7
y_i	15	19	25	33	34	126
x_i^2	4	20,25	36	51,84	64	176,1
y_i^2	225	361	625	1089	1156	3456
$x_i \cdot y_i$	30	85,5	150	237,6	272	775,1

$$\bar{x} = \frac{27,7}{5} = 5,54 \quad \sigma_x = \sqrt{\frac{176,1}{5} - 5,54^2} = 2,13$$

$$\bar{y} = \frac{126}{5} = 25,2 \quad \sigma_y = \sqrt{\frac{3456}{5} - 25,2^2} = 7,49$$

$$\sigma_{xy} = \frac{775,1}{5} - 5,54 \cdot 25,2 = 15,42 \quad r = \frac{15,42}{2,13 \cdot 7,49} = 0,97$$

Correlación muy alta y positiva

De Y sobre X: $y - 25,2 = \frac{15,42}{4,54}(x - 5,54)$ De X sobre Y: $x - 5,54 = \frac{15,41}{51,12}(y - 25,2)$

b) Un chico de $x=5$ años pesa aproximadamente: $y = 25,2 + \frac{15,42}{4,54}(5 - 5,54) = 23,36$ Kg. Es buena predicción pues r es alto y 5 está entre los datos

c) La edad que corresponde a un peso de $y=36$ es: $x = 5,54 + \frac{15,41}{51,12}(36 - 25,2) = 8,79$ años. Es peor predicción pues 36 se sale un poco del rango de los datos

