

Relación de problemas 4

Regresión y Correlación

1. El departamento comercial de una empresa se plantea si resultan rentables los gastos en publicidad de un producto. Los datos de los que dispone son:

Beneficios	Gasto en publicidad
3	0.3
3.5	1
2.8	0.7
3	1
3.3	1.2
4	2
3.7	1.1

Calcular la recta de regresión de los beneficios en función de los gastos en publicidad.

2. Se sabe que el consumo de pan expresado en función de la renta per cápita adopta la forma de una función lineal. Se dispone de los siguientes datos, referidos a diferentes países:

kg de pan	Renta
40	14.8
41	15.1
57	13.5
71	7.3
61	12
42	14
69	7.2

- a) Ajustar la recta de regresión.
 - b) ¿En qué medida el consumo de pan viene explicado por los valores de la renta?
3. Una inmobiliaria observa que durante el mes de agosto el número de apartamentos alquilados varía según el precio semanal de los mismos, con los siguientes valores:

Precio	Apartamentos
100-160	9350
160-250	5120
250-370	3640
370-490	1790
490-1000	800

- a) Si se supone una relación lineal entre el precio y los alquileres, ¿cuántos apartamentos se alquilarán a un precio de 480?
- b) ¿En qué medida el número de apartamentos alquilados depende linealmente del precio?
4. Las notas de física (X) y matemáticas (Y) de 10 alumnos fueron:

X	9	7	3	6	7	5	10	8	2	5
Y	8	5	4	2	9	6	10	9	1	5

Suponiendo una relación lineal entre las notas de física y de matemáticas:

- a) Para un alumno que tenga un 4 en física, ¿cuál será su nota en matemáticas?
- b) Para un alumno que tenga un 7 en matemáticas, ¿cuál será su nota en física?
5. A partir de los siguientes datos:

$$\bar{x} = 5; \quad \bar{y} = 8; \quad Cov(x, y) = 15; \quad Var(y) = 20; \quad r^2 = 0,9$$

obtener la varianza de X y las rectas de regresión de Y/X y X/Y .

6. Un empresario sostiene que existe una relación lineal entre las deudas y los recursos de su empresa. Para comprobarlo toma una muestra de 12 elementos, siendo Y las deudas y X los recursos (medidos ambos en millones de euros). Los datos que obtiene son:

$$\sum y_i = 1200; \quad \sum (y_i - \bar{y})^2 = 1800; \quad \sum x_i = 84; \quad \sum (x_i - \bar{x})^2 = 85; \quad r_{xy} = 0,82$$

Estimar cuál sería la deuda de una empresa con unos recursos que asciendan a 8.2 millones.

7. Se eligen 50 matrimonios y se les pregunta la edad al casarse:

(X, Y)	15-20	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

donde X es la edad del hombre e Y es la edad de la mujer. Calcular:

- a) Rectas de regresión de Y/X y X/Y .
- b) Si una mujer se casó con 22 años, ¿qué edad tendría su pareja cuando se casaron? Y si la mujer se casó a los 24 años, ¿cuál sería la edad de su pareja cuando se casaron?
8. Las calificaciones obtenidas por 10 alumnos en el examen parcial y final de estadística son:

X - Parcial	5	8	6	9	2	3	1	2	4	7
Y - Final	6	6	5	8	1	4	2	1	5	8

- a) Calcular el coeficiente de correlación lineal de Pearson.
- b) La recta de regresión de Y sobre X y la recta de regresión de X sobre Y .
9. 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 X e Y estadísticamente independientes?
 b) Calcular la recta de regresión de Y sobre X .

10. Dada la siguiente distribución bidimensional:

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

Cuantificar el grado de relación lineal entre las variables.

11. Dada la siguiente distribución bidimensional:

X/Y	2	4	5	7
1	4	5	7	9
2	2	4	8	12
3	1	3	4	6
4	1	2	2	3

- a) Calcular la covarianza y el coeficiente de correlación entre X e Y .
 b) Calcular la recta de regresión de X sobre Y y la de Y sobre X .
 c) Representar gráficamente las dos rectas en el mismo eje. ¿Se cortan en algún punto?
 d) Para un valor de la variable $X = 5$, ¿cuánto valdrá Y según la recta de regresión?
 e) Para un valor de la variable $Y = 6$, ¿cuánto vale X según la recta de regresión?

12. Un psicólogo afirma que a medida que los niños crecen dan menos respuestas inadecuadas en cierta situación experimental. Los datos que ha tomado son:

Edad	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
Respuestas	11	12	10	13	11	9	10	7	12	8	7	3	6	5	5

Se pide:

- a) Determinar la validez de esta conclusión.
 b) Si un niño de 10 años y medio participa en el experimento, ¿cuál es el número de respuestas inadecuadas que se puede predecir para él?

13. La siguiente tabla muestra los valores de la edad (X) y la presión sanguínea (Y) de 10 mujeres:

X	56	42	72	36	63	47	55	49	38	42
Y	148	126	159	118	149	130	151	142	114	141

Dar una predicción lineal para la presión sanguínea de una mujer de 51 años. Dar una medida de la bondad de dicha predicción.

14. El precio Y en euros y la antigüedad X en años de siete motocicletas de cierto modelo se recogen en la siguiente tabla:

X	2	3	4	5	6	7	8
Y	415	360	312	270	234	204	180

- a) Calcular la recta de regresión de Y sobre X .
- b) Obtener el precio estimado de una motocicleta de ese modelo con 9 años de antigüedad. ¿Es buena esta predicción?
15. Dada la variable bidimensional discreta (X, Y) con la siguiente distribución de frecuencias relativas:

(X, Y)	0.5	1.5
1	0.2	0.1
2	0.1	0.3
3	0.1	0.2

Calcular el coeficiente de correlación de las variables X e Y .

16. Un hipermercado ha decidido ampliar el negocio. Decide estudiar de forma exhaustiva el número de cajas registradoras que va a instalar, para evitar grandes colas. Para ello, se obtienen los siguientes datos procedentes de otros establecimientos similares acerca del número de cajas registradoras (X) y del tiempo medio de espera (Y).

X	10	12	14	16	18	20
Y	59	51	42	32	26	22

Bajo el supuesto de que el tiempo medio de espera depende linealmente del número de cajas registradoras, se pretende saber

- a) ¿Cómo varía el tiempo medio de espera por cada unidad de caja adicional?
- b) Si se instalaran 17 cajas registradoras, ¿cuál sería el tiempo medio de espera? ¿Es fiable esta predicción?
17. Las rectas de regresión entre dos variables son:

$$8x + 2y = 1 \quad \text{y} \quad 16x + 9y - 1 = 0$$

calcular las medias de X e Y y el coeficiente de correlación lineal. ¿Cuál es la recta de Y sobre X y cuál la de X sobre Y ?

18. Dados los siguientes pares de rectas, determinar si se puede asumir que sean rectas de regresión y en su caso calcular r
- a) $x + y = 125$
 $16x + 25y = 2675$
- b) $2x - 4y = 50$
 $6x + 5y = 75$
- c) $4x - 8y = 70$
 $2x - 4y = 50$

19. Dada una distribución bidimensional (X, Y) , donde $\bar{x} = 9$ y la recta de regresión de Y sobre X viene dada por la ecuación $y = -\frac{7}{18}x + \frac{13}{2}$, obtener los datos que faltan en la tabla:

x_i	y_j	n_{ij}
8	y_1	2
12	2	4
x_3	y_3	n_{33}
14	1	4
		16

Determinar la recta de regresión de X sobre Y , así como el coeficiente de determinación y comentar el resultado.

20. Sea la distribución unidimensional

x_i	3	5	8	9
n_i	5	1	2	1

que es una distribución marginal de la bidimensional (X, Y) , de la que se conoce $\sum_j n_j y_j^2 = 3240$, y la ecuación $y = 5x - 20$.

- Determinar la recta de regresión de $X|Y$.
 - Estudiar la bondad del ajuste lineal.
 - Obtener la varianza de la variable dependiente, así como la varianza residual.
21. Dadas las rectas

$$y = 1 + 2x \quad x = 10 + 5y$$

justificar si es posible que sean las rectas de regresión de $Y|X$ y de $X|Y$ respectivamente, de una misma distribución bidimensional.

22. De una distribución bidimensional (X, Y) se conoce que la media marginal de X es 1, la media marginal de Y es 2 y $r^2 = 0,9$. Si la recta de regresión de $Y|X$ es $y = a - 3x$, obtener el valor de a y el coeficiente de correlación lineal.
23. Un gestor financiero sostiene que existe una relación lineal directa entre las deudas totales de la empresa (Y) y sus recursos propios (X). Para comprobar la validez de aquella creencia considera una muestra piloto de 12 elementos, obteniendo los resultados siguientes:

$$\sum_{j=1}^{12} y_j = 1200; \quad \sum_{i=1}^{12} x_i = 84; \quad \sum_{j=1}^{12} (y_j - \bar{y})^2 = 10800; \quad \sum_{i=1}^{12} (x_i - \bar{x})^2 = 85; \quad r_{XY} = 0,82$$

Se pide:

- Obtener la recta de regresión de Y sobre X .
- Estimar cuál podría ser el exigible de una empresa de condiciones similares a las de la muestra con un capital de 82 millones de euros.

24. Se supone una distribución bidimensional cuyas rectas de regresión son:

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

Calcular el coeficiente de correlación.

25. Se ha estimado la siguiente recta de regresión: $y = 5 + 3x$.

- a) Obtener la recta de X sobre Y sabiendo que $r = 1$.
- b) ¿Sería posible en este modelo que r fuera negativo?

26. En la estimación de un modelo de regresión lineal se obtuvo:

$$\bar{x} = 5; \quad \bar{y} = 8; \quad \sigma_{X,Y} = 15; \quad \sigma_Y^2 = 20; \quad r^2 = 0,9$$

- a) Calcular la varianza de X .
- b) Calcular las rectas de regresión.

27. A partir de un conjunto de datos sobre las variables X e Y se ha calculado la regresión de Y sobre X , obteniéndose los siguientes resultados:

$$y = 10 + 0,45x; \quad r^2 = 0,9; \quad \bar{x} = 20$$

Obtener la recta de regresión de X sobre Y .

28. De una variable estadística bidimensional (X, Y) con frecuencias unitarias se conoce que:

- los valores x_i de la variable (las modalidades) son: 1, 2, 3, 4, 5.
- la recta de regresión de Y sobre X es: $x - 3y + 8 = 0$.
- el coeficiente de correlación es $r = 0,8$.

Se pide:

- a) Hallar \bar{x} y σ_X .
 - b) El coeficiente de regresión de Y sobre X .
 - c) Calcular \bar{y} y σ_Y .
 - d) La ecuación de la recta de regresión de X sobre Y .
29. Si el coeficiente de correlación lineal entre dos variables es -1, decidir si las siguientes afirmaciones son verdaderas o falsas y por qué:
- a) La covarianza es negativa.
 - b) La recta de regresión es decreciente.
 - c) El ajuste de la recta de regresión es perfecto.
 - d) Al calcular la pendiente de la recta de regresión, saldrá un valor negativo.
 - e) El coeficiente de determinación es 1.
 - f) Las variables no son independientes.
30. Para los datos del ejercicio 1, tomando como variable X los beneficios y como Y los gastos, calcular la relación de Y sobre X mediante:
- a) Una hipérbola equilátera.

- b) Una curva potencial.
 c) Una curva exponencial.
 d) ¿Cuál de estos ajustes es mejor?
31. Sobre el gasto en libros (Y) y la renta disponible mensual (X) se dispone de la información referente a 6 familias:

Y	0.3	0.5	0.6	0.9	1	1.4
X	6	7	8	10	15	21

Explicar el comportamiento de Y por X mediante:

- a) Una relación lineal.
 b) Una hipérbola equilátera.
 c) Una curva potencial.
 d) Una curva exponencial.
 e) ¿Qué ajuste es el más adecuado?
32. En una empresa se toma una muestra de 100 trabajadores con la finalidad de estudiar si hay relación entre su edad (X) y los días que están de baja en el año (Y). Se obtienen los siguientes resultados:

(X, Y)	$[0,20]$	$(20,40]$	$(40, 60]$
$[18,30]$	28	2	0
$(30,40]$	26	15	4
$(40,50]$	6	14	5

- a) ¿Cuál es la edad más frecuente de los trabajadores?
 b) ¿Es simétrica la distribución del número de días de baja de los trabajadores?
 c) Ajustar mediante un modelo exponencial los días de baja en función de la edad.
 d) Realizar el mismo ajuste considerando un modelo lineal.