RELACIÓN DE PROBLEMAS 3

- Regresión y correlación.

1. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \frac{1}{2} \qquad 0 < x < y < 2$$

   Calcular:

   a) Coeficiente de correlación lineal entre X e Y.
   b) Rectas de regresión.
   c) Curvas de regresión y razones de correlación.
   d) Error cuadrático medio asociado a cada una de las funciones de regresión.

2. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \mathrm{e}^{-y} \qquad 0 < x < y < \infty$$

   Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

3. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \frac{1}{2} \qquad (x, y) \in \square\{(0, 1), (1, 0), (-1, 0), (0, -1)\}$$

   Calcular las rectas de regresión, el coeficiente de correlación lineal, las curvas de regresión, las razones de correlación y errores cuadráticos medios asociados. Representar las distintas funciones de regresión e interpretar los resultados.

4. Se elige una palabra al azar de la siguiente frase: "IT IS TOO GOOD TO BE TRUE". Sea L la longitud de la palabra seleccionada y X el número de "O". Calcular

   - Mejor predicción por mínimos cuadrados de L a partir de X y el ECM cometido en la predicción.
   - Mejor predicción por mínimos cuadrados de L a partir de un valor concreto X $(X = x)$ y el ECM cometido en cada predicción.
   - Mejor predicción por mínimos cuadrados de L sin observar X y el ECM cometido en la predicción.

5. Sean X e Y variables aleatorias con función de densidad conjunta

$$f(x,y) = \begin{cases} x+y & 0 < x < 1, 0 < y < 1 \\ \\ 0 & \text{en otro caso} \end{cases}$$

   a) Calcular la mejor predicción mínimo cuadrática de Y a partir de X y el error cuadrático medio asociado a la predicción.

   b) Si se observa $X = 1/2$, ¿qué predicción de Y tiene menor error cuadrático medio? ¿Cuál es el valor de dicho ECM?

   c) Supóngase que una persona debe pagar un costo C por la oportunidad de observar el valor de X antes de predecir el valor de Y o puede simplemente predecir el valor de Y sin observar primero el valor de X. Si la persona considera que su pérdida total es el costo C más el ECM de su predicción, ¿cuál es el valor máximo de C que debería estar dispuesta a pagar?

6. El vector aleatorio (X,Y) está uniformemente distribuido en la región limitada por las curvas $y = x^2$, $y = 0$, $x = 1$. Se desea predecir la variable aleatoria Y por medio de:

   a) Una constante.

   b) Una función lineal de X.

   c) Una función arbitraria de X.

   Para cada uno de los apartados, calcular la mejor predicción en el sentido de menor error cuadrático medio y dicho error.

7. Sea X la calificación obtenida por una persona en una prueba de aptitudes matemáticas e Y en una prueba de aptitudes musicales. Supongamos que (X,Y) tiene función de densidad

$$f(x,y) = \frac{2}{5}(2x+3y) \qquad 0 < x < 1, 0 < y < 1$$

   a) Si se selecciona una persona al azar, ¿qué predicción de su calificación en música tiene menor ECM? ¿Cuál es dicho error?

   b) Dar la medida del grado de dependencia de la calificación en música respecto de la calificación en matemáticas.

   c) Si la calificación en matemáticas de una persona seleccionada ha resultado ser 0.8, ¿cuál es la mejor predicción para su calificación en música? ¿Cuál es el ECM de esa predicción?

   d) Supongamos que el coste de la observación de la variable X para la predicción de Y es de 0.01, y que la pérdida total para el observador es el ECM de la aproximación más el coste de la observación, mientras que la pérdida obtenida al predecir Y sin observar X es el ECM. ¿Debe el observador observar X para predecir Y?

8. Sea (X,Y) un vector aleatorio con función de densidad

$$f(x,y) = 1 \qquad (x,y) \in \triangle \{(0,0),(2,0),(1,1)\}$$

   Se desea predecir la variable Y por medio de

   a) Una constante.

b) Una función lineal de X.

c) Una función arbitraria de X.

Para cada apartado, calcular la mejor predicción de Y en el sentido de menor ECM y dicho error.

Si la observación de X antes de predecir Y supone un coste C y la pérdida total en la predicción es el coste de la observación más el ECM de la predicción. ¿Qué cantidad máxima se estaría dispuesto a pagar si se predice Y mediante una función lineal de X? ¿Y si se predice Y mediante una función arbitraria de X?

9. Sea (X, Y) un vector aleatorio distribuido uniformemente en el paralelogramo de vértices $(0, 0)$, $(2, 0), (3, 1)$ y $(1, 1)$. ¿Qué es más fiable predecir Y por X o X por Y? ¿Qué error cuadrático medio se comete con cada predicción?

10. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \mathrm{e}^{-x-y} \qquad x, y > 0$$

y sean $U = X^2$ y $V = X + Y$. Calcular las curvas y rectas de regresión de U sobre V y de V sobre U y los errores cuadráticos medios asociados. Obtener el coeficiente de correlación lineal y las razones de correlación. Interpretar los resultados.

11. Sea (X, Y) un vector aleatorio con rectas de regresión

$$\begin{aligned} x + 4y &= 1 \\ x + 5y &= 2 \end{aligned}$$

a) ¿Cuál es la recta de regresión de Y sobre X?

b) Calcular el coeficiente de correlación lineal y la proporción de varianza de cada variable que queda explicada por la regresión lineal.

12. Sea (X,Y) un vector aleatorio con función masa de probabilidad

a)

X \ Y	10	15	20
1	0	2/7	0
2	1/7	0	0
3	0	0	3/7
4	0	1/7	0

b)

X \ Y	10	15	20
1	0	2/6	0
2	1/6	0	0
3	0	0	3/6

c)

X \ Y	10	15	20	25
1	0	3/7	0	1/7
2	0	0	1/7	0
3	2/7	0	0	0

Para cada apartado, analizar la dependencia entre las variables y calcular las curvas de regresión. Comentar los resultados.

13. Sea X la variable aleatoria número de balanzas e Y el número de dependientes en los puntos de venta de un mercado. Si la función masa de probabilidad de (X, Y) viene dada por

X \ Y	1	2	3	4
1	1/24	2/24	0	0
2	1/24	2/24	3/24	1/24
3	0	1/24	2/24	6/24
4	0	0	2/24	3/24

a) Determinar las rectas de regresión.

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

14. Dada la función masa de probabilidad del vector aleatorio (X, Y)

X \ Y	0	1	2	3
0	0,2	0,2	0,05	0
1	0,1	0,1	0,1	0,05
2	0	0,05	0,05	0,1

a) Calcular las rectas de regresión mínimo cuadráticas de Y sobre X y de X sobre Y y medir la bondad del ajuste. Determinar el valor de y para $x = 5$ mediante la recta correspondiente.

b) Calcular las curvas de regresión de Y sobre X y de X sobre Y. Medir la bondad de cada uno de los ajustes.

15. Si las funciones masa de probabilidad de los vectores (X, Y) y (X^*, Y) son las dadas, ¿a partir de que variable, X o X^*, se obtiene una mejor predicción mínimo cuadrática de la variable Y? ¿Por qué?

X \ Y	0	1	2
0	1/5	0	0
2	0	1/5	0
3	1/5	0	0
4	0	0	2/5

Y / X^*	0	1	2
0	1/5	0	1/5
2	0	1/5	0
3	1/5	0	0
4	0	0	1/5