REGRESION LINEAL SIMPLE

Ejercicio 1 Mediciones apareadas y normal bi-variada
La siguiente tabla resume los puntajes de dos pruebas de matemática de un grupo de estudiantes:

1era prueba	media: 120	desvío: 10
2da prueba	media: 130	desvío: 9
correlación: 0.6		

Asumir que los datos se distribuyen como una normal bi-variada.

1. ¿Qué nota pronosticarías para la 2da prueba a un estudiante que sacó 115 en la 1era?
2. ¿Cuál es la probabilidad de que tu pronóstico anterior sea erróneo por más de 5 puntos?

Ejercicio 2 ¿De tal palo tal astilla?
Datos de una población indican que la altura de madres e hijas se ajustan a la distribución normal bi-variada con correlación 0.5. Ambas variables tienen media 1.63m y desvío 0.05m.

1. Hallar las ecuaciones de las rectas de regresión.
2. La altura de dos amigas difiere en 5cm. ¿Cuál es tu pronóstico para la diferencia de alturas de las madres?
3. Entre las hijas que miden más que la media, ¿qué porcentaje son más bajas que sus madres?
4. Una madre mide menos que la media. ¿Cuál es tu pronóstico para la altura de su hija?

Ejercicio 3
La cantidad de cierto producto químico que se disuelve en una cantidad de agua dada, a varias temperaturas está dada por la siguiente tabla:

Y=cantidad disuelta	8	12	21	31	39	58
x=temperatura	0	10	20	30	40	60

Encontrar un intervalo de confianza para β_1 al 95 % al regresar Y sobre x.

Ejercicio 4
Los siguientes datos representan la altura x y el peso Y de 12 individuos. Aplicar un modelo de regresión lineal para predecir el peso a partir de la altura.

x (cm)	163	163	165	166	168	169	170	170	171	172	172	174
Y	60	61	62	63	65	67	68	69	70	72	71	70

Probar un TdH de que la pendiente es nula (i.e la altura no sirve para predecir el peso).

Ejercicio 5 Arena gruesa, playa honda
La siguiente tabla muestra los datos correspondientes a un estudio realizado con el objetivo de comprender la relación entre el diámetro de los granos de arena (D en mm) y la pendiente de profundidad del agua (P en grados) para $n = 9$ playas. Se asume un modelo normal bi-variado para los datos.

i	1	2	3	4	5	6	7	8	9
D	0,170	0,190	0,220	0,235	0,235	0,300	0,350	0,420	0,850
P	0,63	0,70	0,82	0,88	1,15	1,50	4,40	7,30	11,30

Algunos datos que pueden ser de utilidad en el ejercicio:

$$\sum_{i=1}^{9}(D_i - \overline{D})^2 = 0{,}355, \quad \sum_{i=1}^{9}(P_i - \overline{P})^2 = 114{,}85.$$

1. Realizar y describir un diagrama de dispersión para (D, P).
2. Calcular el coeficiente de correlación muestral r. Indicar si la correlación es débil, moderada, o fuerte.
3. Estimar la ecuación de la recta de regresión. Calcular el error cuadrático medio de predicción.
4. ¿Cuál es el porcentaje de la variabilidad explicado por la regresión?
5. ¿Cuál es la pendiente pronosticada para una playa cuyos granos de arena tienen 0,325 mm de diámetro?
6. Si el diámetro de los granos de arena de una playa es mayor que el de otra en 0,1 mm, ¿cuánto es la diferencia pronosticada entre las respectivas pendientes?

Ejercicio 6 Presión alta
El peso (en libras) y la presión arterial sistólica (en mmHg) de 26 hombres seleccionados al azar en el grupo de edad de 25 a 30 se muestran a continuación. Supongamos que el peso y la presión arterial se distribuyen de manera conjunta como una normal bi-variada.

Individuo	Peso	Presión	Individuo	Peso	Presión
1	165	130	14	172	153
2	167	133	15	159	128
3	180	150	16	168	132
4	155	128	17	174	149
5	212	151	18	183	158
6	175	146	19	215	150
7	190	150	20	195	163
8	210	140	21	180	156
9	200	148	22	143	124
10	149	125	23	240	170
11	158	133	24	235	165
12	169	135	25	192	160
13	170	150	26	187	159

1. Estimar el coeficiente de correlación.
2. Dibujar el diagrama de dispersión de los datos.
3. Hallar la recta de regresión que relaciona la presión arterial sistólica con el peso.
4. Dibujar la recta de desvíos y la recta de regresión.
5. Hacer un TdH para probar H_0: *el peso y la presión son independientes*.

Ejercicio 7
Sea (X,Y) con distribución normal y parámetros $\mu = (1,2)'$, $\text{var}(X) = \text{var}(Y) = 1$ y $\text{cov}(X,Y) = 0.8$.

1. Simular $n = 1000$ observaciones del par (X,Y) simulando primero de la distribución de X y luego de la de $Y|X = x$. Graficar las parejas (x_i, y_i) de valores simulados.

2. A partir de los valores simulados estimar los parámetros β_0, β_1 del modelo de regresión lineal simple $y_i = \beta_0 + \beta_1 x + \varepsilon_i$, por el método de mínimos cuadrados. Graficar conjuntamente las parejas $(x_1, y_1), \ldots, (x_n, y_n)$ y la recta de regresión estimada.

3. Mostrar que $\mathbb{E}(Y|X = x)$ se puede escribir en la forma $b_0 + b_1 x$ y comparar esta recta con la estimada en el punto anterior (agregarla al gráfico).

Ejercicio 8

Se hace un estudio del efecto de la temperatura (t) en la cantidad producida por cierto material (Y), en un proceso químico. Se obtienen los resultados siguientes:

t	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y	1	5	4	7	10	8	9	13	14	13	18

1. Ajustar un modelo de regresión lineal para predecir Y en función de t.

2. Probar la hipótesis de que la pendiente es nula con un nivel del 95 %.

3. Hallar un intervalo de confianza para el valor esperado de Y evaluado en $t = 3$ con un nivel del 95 %.

4. Hallar un intervalo de confianza para una futura Y evaluado en $t = 3$con un nivel del 95 %.

Ejercicio 9

1. Hallar la recta de regresión simple de la variable *velocidad* sobre la variable regresora *densidad* con los datos siguientes:

Densidad	12,7	17	66	50	87,8	81,4	75,6	66,2	81,1	62,8	77	89,6
Velocidad	62,4	50,7	17,1	25,9	12,4	13,4	13,7	17,9	13,8	17,9	15,8	12,6
Densidad	18,3	19,1	16,5	22,2	18,6	66	60,3	56	66,3	61,7	66,6	67,8
Velocidad	51,2	50,8	54,7	46,5	46,3	16,9	19,8	21,2	18,3	18	16,6	18,3

2. Calcular la estimación de σ^2 y a partir de ella las estimaciones de $s.e(\widehat{\beta_0})$ y $s.e(\widehat{\beta_1})$.

3. Escribir los intervalos de confianza para los parámetros con un nivel de confianza del 95 %.

4. Construir la tabla para la significación de la regresión y realizar dicho contraste.

5. Hallar el intervalo de predicción de la respuesta media cuando la densidad es de 50 vehículos por km con un nivel de confianza del 90 %.

6. Repetir las preguntas anteriores hallando la recta de regresión de la variable *raíz cuadrada de la velocidad* sobre la variable regresora *densidad*