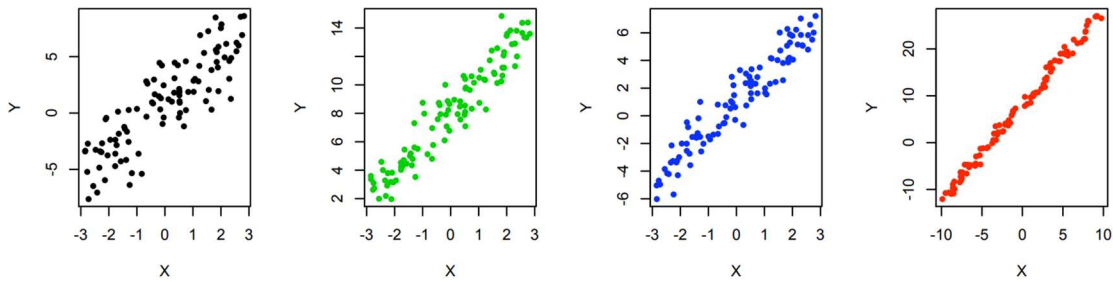
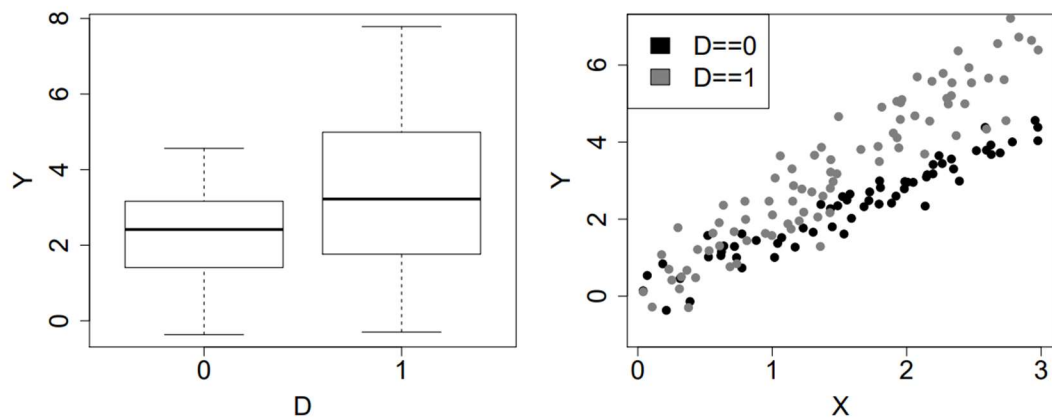


1. Abajo hay 4 diagramas de dispersión de una variable de salida y versus un predictor x seguido por 4 salidas resúmenes nombradas A, B, C y D. Empareje cada grafico con su salida correspondiente.



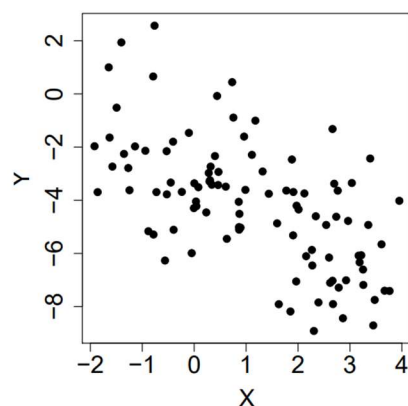
Dataset	intercept	slope	residual standard error	SSR/SST
A	$b_0 = 8.1, s_{b_0} = 0.11$	$b_1 = 2.1, s_{b_1} = 0.066$	$s = 1.08$	$R^2 = 0.90$
B	$b_0 = 8.0, s_{b_0} = 0.10$	$b_1 = 2.0, s_{b_1} = 0.017$	$s = 1.01$	$R^2 = 0.99$
C	$b_0 = 1.0, s_{b_0} = 0.10$	$b_1 = 2.0, s_{b_1} = 0.060$	$s = 0.97$	$R^2 = 0.93$
D	$b_0 = 0.9, s_{b_0} = 0.20$	$b_1 = 1.9, s_{b_1} = 0.120$	$s = 2.09$	$R^2 = 0.71$

2. A continuación, se encuentra el plot de dos variables continuas Y y X , adicionalmente en el grafico se han diferenciado dos grupos usando una variable binaria D .



Que puedes concluir de la inclusión de la variable binaria basada en los gráficos.

- 3.Cuál de las siguientes rectas de regresión se ajusta mejor a los datos ploteados.



- (i) $b_0 = 0, b_1 = -1$
- (ii) $b_0 = -3, b_1 = 1$
- (iii) $b_0 = -5, b_1 = 2$
- (iv) $b_0 = -3, b_1 = -1$
- (v) $b_0 = 0, b_1 = -3$

4. Cual seria el R^2 si estimaras un modelo de regresión con solo el intercepto?

- a. 1
- b. 0

- c. 0.5
 - d. No podemos saberlo a menos que se proporcione el tamaño de la muestra.
 - e. No podemos decirlo incluso si se da el tamaño de la muestra.
5. El siguiente código genera valores de ventas de un negocio de comercio electrónico.

```
import numpy as np
import pandas as pd

# Establecemos una semilla para que los resultados sean reproducibles
np.random.seed(42)

# Generamos las ventas
fechas = pd.date_range(start="2022-01-01", end="2022-12-31", freq='D')
ventas = np.random.normal(loc=500, scale=100, size=len(fechas)) # ventas diarias con media de 500 y desviación estándar de 100

df_ventas = pd.DataFrame({
    'Fecha': fechas,
    'Ventas': ventas
})

df_ventas.head()
```

- I. Calcule la media, mediana y moda de las ventas diarias.
 - II. ¿Cuál es la desviación estándar de las ventas?
 - III. Identifique el mes con las ventas promedio más altas y el mes con las ventas promedio más bajas.
 - IV. Basándose en el histograma de las ventas diarias, ¿cómo se distribuyen las ventas?
(Recuerde no haremos códigos en nuestro parcial, tendríamos la salida, así que la pregunta mas pertinente es la última, las otras partes del punto le permiten seguir practicando Python y familiarizarse con su uso)
6. Imagina que eres el analista de un popular e-commerce y se te ha proporcionado un conjunto de datos de una encuesta de satisfacción del cliente. Los clientes calificaron su experiencia de compra en una escala de 1 a 10. Tras analizar 1,000 respuestas, obtuviste los siguientes resultados:
- a. Media de calificaciones: 6.5
 - b. Mediana de calificaciones: 7
 - c. Moda de calificaciones: 8
 - d. Desviación estándar de calificaciones: 2.1
 - e. Adicionalmente, basándote en un histograma, notaste que las calificaciones entre 7 y 9 eran las más frecuentes, pero también había un número significativo de calificaciones en el rango de 1 a 4.

Preguntas basadas en este escenario:

- I. A pesar de que la calificación más frecuente (moda) es 8, ¿qué te sugiere el hecho de que la media sea 6.5 respecto a la experiencia general de los clientes?
 - II. Considerando que la desviación estándar es 2.1, ¿qué puedes inferir sobre la variabilidad o consistencia de las respuestas de los clientes?
 - III. Si la empresa establece que cualquier calificación por debajo de 6 es insatisfactoria, y sabiendo que hay un número significativo de respuestas en el rango de 1 a 4, ¿qué acciones estratégicas podrías recomendar?
 - IV. ¿Cómo podrías usar esta información para mejorar la experiencia del cliente y, potencialmente, aumentar las calificaciones en futuras encuestas?
7. En un estudio realizado por una empresa de telecomunicaciones, se encontró que:
- a. La probabilidad de que un cliente nuevo use más de 5GB de datos en su primer mes es del 0.4.

- b. De aquellos que usaron más de 5GB en su primer mes, el 70% decidió renovar su contrato al final del año.
- c. De aquellos que usaron 5GB o menos, sólo el 50% decidió renovar su contrato al final del año.

Pregunta:

Un cliente seleccionado al azar renovó su contrato al final del año. ¿Cuál es la probabilidad de que haya usado más de 5GB en su primer mes?

8. Una empresa de comercio electrónico lanzó una campaña de publicidad para aumentar las ventas. El 12% de 100 clientes expuestos a la campaña hicieron una compra. Sin embargo, históricamente, el 10% de los clientes hacen una compra sin ver la publicidad. Dado un valor p de 0.07, ¿la campaña de publicidad tuvo un impacto significativo en las ventas? Justifique su respuesta con base en el valor p y un nivel de significancia de 0.05.
9. Una fábrica de snacks introduce un nuevo sabor y quiere determinar si es preferido en comparación con el sabor original. Realizan pruebas de sabor con diferentes grupos y obtienen los siguientes resultados:
 - a. Muestra de 50 personas: valor p = 0.30
 - b. Muestra de 500 personas: valor p = 0.10
 - c. Muestra de 5,000 personas: valor p = 0.03

¿Cómo impacta el tamaño de la muestra en la determinación de si el nuevo sabor es preferido sobre el original?

10. Observando el gráfico, ¿qué puede decir sobre la diferencia entre las medias de los Grupos A y B.

