

Ejercicios 1:

1. Se realizó un estudio para determinar si ciertas medidas de la fuerza estática del brazo influyen en las características de “levantamiento dinámico” de un individuo. Veinticinco individuos se sometieron a pruebas de fuerza y luego se les pidió que hicieran una prueba de levantamiento de peso, en el que el peso se elevaba en forma dinámica por encima de la cabeza. A continuación se presentan los datos.

Individual	Fuerza del brazo, x	Levantamiento dinámico, y
1	17.3	71.7
2	19.3	48.3
3	19.5	88.3
4	19.7	75.0
5	22.9	91.7
6	23.1	100.0
7	26.4	73.3
8	26.8	65.0
9	27.6	75.0
10	28.1	88.3
11	28.2	68.3
12	28.7	96.7
13	29.0	76.7
14	29.6	78.3
15	29.9	60.0
16	29.9	71.7
17	30.3	85.0
18	31.3	85.0
19	36.0	88.3
20	39.5	100.0
21	40.4	100.0
22	44.3	100.0
23	44.6	91.7
24	50.4	100.0
25	55.9	71.7

- a. Realice el diagrama de dispersión entre las variables, obtenga e interprete las estadísticas descriptivas de las mismas.
- b. Calcule el coeficiente de correlación lineal y realice la prueba de hipótesis correspondiente para validar si existe o no relación lineal entre las variables, con una significancia del 5%.
- c. Estime los coeficientes del modelo de regresión lineal simple por MCO.
- d. Obtenga la estimación puntual para una fuerza del brazo de 30.
- e. Grafique los residuales en comparación con la variable independiente y comente los resultados.

2. Doce marcas de shampoo de venta en México han compartido información acerca de sus ventas y del monto en inversión publicitaria durante 2023. Los datos anuales de ambas variables se presentan a continuación:

Marca	Ventas (Millones de lts.)	Inversión Publicitaria en Redes Sociales (Millones de pesos)
A	2.4	6
B	1.6	2
C	2.3	5
D	1.5	1
E	3.2	4
F	2.5	7
G	1.8	4
H	1.8	3
I	3.5	8
J	3.4	11
K	1.5	2
L	3.2	12

- Realice el diagrama de dispersión entre las variables, obtenga e interprete las estadísticas descriptivas de las mismas.
 - Calcule el intervalo de confianza al 95% para coeficiente de correlación lineal
 - Estime los coeficientes del modelo de regresión lineal simple por MCO.
 - Utilizando un nivel de significancia del 5%, pruebe la hipótesis estadística de que por cada 500mil pesos adicionales de inversión anual en redes sociales, se espera un incremento en las ventas anuales de shampoo mayor a 50 mil litros.
3. El siguiente juego de datos describe la producción de esencia floral en una comunidad en Francia. La variable independiente (X) se refiere a la cantidad de flores procesadas para extraer su esencia por cada productor. La variable dependiente (Y) es una medida del aceite extra ído en onzas por la cantidad de flores (en miles de unidades) procesadas.

Tabla 1: Datos de la Producción de esencia floral.

obs	cons	acres
1	1	1.71
2	1.08	1.52
3	1.15	1.29
4	1.15	3.09
5	1.2	2.21
6	1.3	2.26
7	1.37	2.4
8	1.37	2.1
9	1.43	1.96
10	1.46	2.09
11	1.52	2.02
12	1.57	1.31
13	1.65	2.17
14	1.65	2.28
15	1.65	2.41
16	1.66	2.23
17	1.87	3.04
18	2.03	2.06
19	2.05	2.73
20	2.3	2.36

- Grafique los datos en un diagrama de dispersión.
- ¿Cree que exista una relación entre la producción de esencia y la cantidad de flores procesadas? ¿Es esta positiva o negativa?
- Por medio del ajuste de una línea recta, verifique que:
 $b_0 = 1.38$, $b_1 = 0.52$, $S^2 = 0.206$

- d. Construya la tabla de análisis de varianza (anova) y realice la prueba de significancia para la regresión. ¿Es esta significativa o no?
 - e. Encuentre el error estándar de la pendiente y un intervalo de confianza al 95%.
 - f. ¿Qué porcentaje de la variabilidad total de la respuesta es explicado por el modelo? (Sugerencia: Recordar bondad de ajuste).
 - g. Determine un intervalo de confianza para la respuesta media al 95% cuando el número de flores procesadas flores es igual a 1.25.
 - h. Determine un intervalo de predicción al 95% cuando el número de flores procesadas flores es igual a 1.95.
4. Considere el ejemplo de servicio de TV por cable descritos en la Tabla 2.

Encuesta de televisión por cable

Una empresa de televisión por cable encargó a un bufete un estudio de mercado para conocer el perfil de los clientes potenciales de una zona residencial formada por dos colonias. Las colonias constan de 12 y 25 manzanas con un total de 236 y 605 hogares, respectivamente. Mediante muestreo probabilístico (no discutido aquí) se seleccionó una muestra de ocho manzanas y cinco hogares por manzana. En cada hogar seleccionado se recabaron varias respuestas de las que presentamos solamente algunas de éstas.

Tabla 2 **Encuesta de televisión por cable**

Variable	Descripción
1. Colonia	Colonia a la que pertenece el hogar de la zona residencial
2. Manzana	Número de manzana a la que pertenece el hogar
3. Adultos	Número de adultos por hogar
4. Niños	Número de niños menores de 12 años por hogar
5. Teles	Número de televisores por hogar
6. Tipo	Tipo de televisor que posee: blanco y negro (B), color (C), ambos (A)
7. Tvtot	Suma del número de horas frente al televisor en la semana de todos los miembros de la familia
8. Renta	Cantidad máxima de renta que el jefe del hogar estaría dispuesto a pagar al mes por servicio de TV por cable (múltiplos de \$5)
9. Valor	Valor catastral del hogar (m\$). La respuesta se usa para dar idea aproximada del ingreso familiar

- a. Ajuste por mínimos cuadrados un modelo de regresión lineal simple para la respuesta renta, con el valor catastral (en miles de pesos) como variable independiente. Calcule los coeficientes y el error estándar de la regresión, y grafique los residuales contra el regresor x_i (gráfica de dispersión).
- b. ¿Cuál es la significancia de la regresión? (Valor-p del estadístico F.)

- c. Repita los incisos anteriores pero sin considerar los 2 casos donde $y = 0$.
¿Consideraría los nuevos coeficientes estadísticamente iguales a los anteriores? Comente.
- d. Compare los coeficientes de determinación R^2 en ambos casos. Comente.
5. Determine cuáles de entre los siguientes modelos son lineales en los parámetros, en las variables, o en ambos ¿Cuáles de estos modelos son modelos de regresión lineal?

a) $Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + \epsilon_i$

b) $Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$

c) $Y_i = \beta_0 X_i^{\beta_1} + \epsilon_i$

d) $\ln Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

e) $\ln Y_i = \ln \beta_0 + \beta_1 \ln X_i + \epsilon_i$

f) $Y_i = \beta_0 + \beta_1^3 X_i + \epsilon_i$