

ESTADÍSTICA II
 EJERCICIOS TEMA 4
 CURSO 2010/11

1. La siguiente tabla muestra información sobre la venta en 1998 de prensa diaria escrita en ejemplares diarios vendidos por cada mil habitantes para 8 comunidades autónomas españolas relacionándola con su producción económica basada en el Producto Interior Bruto (PIB) por habitante en miles de euros (Fuente: INE. Anuario Estadístico).

PIB	8,3	9,7	10,7	11,7	12,4	15,4	16,3	17,2
Ejemplares	57,4	106,8	104,4	131,9	144,6	146,4	177,4	186,9

- Estimar por Mínimos Cuadrados un modelo de regresión simple para explicar el número de ejemplares vendidos en términos del PIB.
 - Construir un intervalo de confianza al 95 % para la pendiente de la recta de regresión y contrastar la hipótesis de que dicho parámetro toma valor cero, ¿puede afirmarse que el número de ejemplares vendidos depende linealmente del PIB?
 - ¿Cuál sería la venta de prensa que se podría predecir para una comunidad cuyo PIB por habitante fuese de 15000 euros?
 - Si para una región cualquiera el valor del PIB aumentase en 2500 euros, ¿cómo cabría esperar que variase la venta de prensa diaria?
2. Uno de los administradores de una empresa argumenta que el uso de Internet es la principal causa del gasto en la factura telefónica. Para corroborar esta afirmación, se toman datos en distintos departamentos del gasto telefónico mensual en euros y los tiempos de conexión en minutos.

Cuantía de la factura telefónica	55	100	118	120	142
Tiempo de conexión	200	500	700	800	1000

- Calcular el coeficiente de correlación entre ambas variables, ¿existe una relación de tipo lineal entre ellas?
 - Estimar un modelo de regresión lineal que permita estimar la cuantía mensual de la factura telefónica en términos del tiempo de conexión.
 - De acuerdo a esta relación lineal, ¿cuál sería la cuantía de la factura telefónica de un departamento que no se conectase a Internet? Calcular un intervalo de confianza al 95 % para dicha estimación.
 - ¿Cuál sería el gasto telefónico estimado según esta relación lineal si el tiempo de conexión a Internet de un departamento fuera de 2000 minutos? ¿Le parece aceptable tal predicción?
3. Con el objetivo de estudiar la relación lineal entre el precio de los automóviles y el número de unidades vendidas, se procedió a recoger datos sobre tales magnitudes durante el pasado mes en una determinada región. Los resultados obtenidos fueron los siguientes:

Precio (miles de euros)	Cantidad vendida (unidades mensuales)
7,5	450
9	425
10,5	400
12	350
14	325
16	300
18	290
20,5	280
23,5	260
27	200

- a) Calcular e interpretar el coeficiente de correlación lineal entre ambas variables.
- b) Una empresa radicada en la región tiene previsto para el mes próximo aumentar el precio de su modelo más vendido en 500 euros. Si suponemos como válida la relación lineal entre las dos variables analizadas para los datos del pasado mes, ¿cómo afectaría este hecho a las ventas de dicho modelo?
- c) Si expresamos el precio en euros y las cantidades vendidas en 10^2 unidades, ¿cuál sería el modelo lineal que explica las ventas en función del precio? ¿Y el coeficiente de correlación?
4. Una gasolinera ha recogido información acerca de su recaudación diaria durante una semana, así como del número de clientes que acudieron a la misma en cada día:

Recaudación (10^3 euros)	1,5	10	8	3	5	15	2
Número de clientes (10^2)	3	6	5	3,5	4	8	3,2

- a) Realizar un ajuste lineal que exprese la recaudación en función del n° de clientes.
- b) Determinar cuál sería la recaudación media prevista para aquellos días en los que lleguen a la gasolinera 720 clientes. Obtener un intervalo de confianza al 95 % para dicha predicción.
- c) Determinar cuál sería la recaudación prevista para un día en el que lleguen a la gasolinera 720 clientes. Obtener un intervalo de confianza al 95 % para dicha predicción.
- d) Obtener la recta de regresión si expresásemos la recaudación en euros y el número de clientes en unidades.
5. Un distribuidor de productos de droguería y limpieza reparte sus productos entre los comercios del sector de todos los pueblos de una comarca. Entre los artículos que distribuye, ha seleccionado una muestra de 6 de ellos, que son los que considera más importantes, debido a su demanda por parte de los comerciantes y del público en general. De esos 6 artículos se tienen los datos correspondientes al último mes de Abril correspondientes al precio unitario del artículo en euros y al volumen de ventas correspondientes a dicho artículo en miles de euros, que son los siguientes:

	Precio	Volumen
Lejía	0,65	2,50
Amoniaco perfumado	0,85	1,68
Estropajo fibra verde	1,25	4,23
Detergente concentrado	2,6	5,69
Suavizante	2,1	5,17
Detergente lavavajillas	2,5	5,50

- a) Construir la recta de regresión del volumen de ventas con respecto al precio unitario para el mes de Abril.
- b) Contrastar la hipótesis de que el volumen de ventas depende linealmente del precio de los artículos.
- c) Suponiendo que los precios durante el mes de Mayo no varíen con respecto al mes anterior, obtener una predicción para el volumen de ventas en el mes de Mayo de un artículo cuyo precio unitario es de 1 euro. Dar una medida de la fiabilidad de dicha predicción mediante un intervalo de confianza al 95 %.
- d) El distribuidor está convencido de que el volumen de ventas sería el mismo si en el mes de Mayo se sube en 5 céntimos el precio de cada artículo. Obtener la ecuación de la recta de regresión en este caso y una predicción para las ventas en el mes de Mayo de un artículo cuyo precio unitario sea de 1 euro (incluyendo la subida de 5 céntimos).
6. Una aplicación importante del análisis de regresión en contabilidad es para estimar costos en función del volumen de ventas. A partir de una muestra de 8 pares de datos sobre volumen de ventas y costos, se obtiene mediante el método de mínimos cuadrados la recta de regresión de los costos con respecto al volumen de ventas, obteniéndose los siguientes residuos:

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
10	-3	2	1	1	-1	-2	

- a) Calcular el residuo número 8
 - b) Calcular la varianza residual.
 - c) Calcular un intervalo de confianza al 95 % para la varianza de los errores.
 - d) Si el modelo de regresión lineal construido es $\hat{y} = 2 + 5x$ y la varianza de las observaciones de la variable independiente es $s_X^2 = 100$, analizar si es significativa la relación lineal entre ambas variables.
 - e) Cuál es la media de las observaciones de la variable dependiente, \bar{y} , si la media de las observaciones de la variable independiente es $\bar{x} = 3$.
7. Se desea realizar un estudio entre el personal administrativo de una empresa que permita estimar el número de errores diarios cometidos por cada trabajador (variable Y) en función del número de años contratados en la empresa (variable X). Con esta finalidad se controló el trabajo de 10 empleados seleccionados al azar obteniéndose la siguiente información:

$$\sum_{i=1}^{10} x_i^2 = 296; \quad \sum_{i=1}^{10} x_i = 50; \quad \sum_{i=1}^{10} y_i^2 = 1036; \quad \sum_{i=1}^{10} y_i = 86; \quad \sum_{i=1}^{10} x_i y_i = 331$$

- a) Calcular e interpretar el coeficiente de correlación lineal entre ambas variables.
 - b) Estimar la recta de regresión correspondiente indicando el significado de cada coeficiente.
 - c) Estimar la cantidad promedio de errores que cometerá un trabajador que lleve cuatro años trabajando en la empresa.
8. Una empresa desea investigar la relación entre la cantidad de días que faltan sin permiso los empleados por año (variable Y) y la distancia en kilómetros desde su hogar a su trabajo (variable X). Al analizar una muestra de empleados se obtuvieron los siguientes resultados:

$$\text{cov}(x, y) = 5,4; \quad \text{cor}(x, y) = 0,7838; \quad \bar{x} = 25; \quad s_x^2 = 12$$

Además, para una distancia de 20 kilómetros del hogar al trabajo, se ha estimado que el número medio de días que faltan sin permiso es de 4.

- a) Estimar un modelo de regresión lineal.
 - b) En término medio, ¿cuál será el número de días sin permiso que faltará un trabajador que viva a 18 kilómetros del lugar de trabajo?
 - c) Responder a los apartados anteriores expresando la distancia en metros.
9. A partir de una muestra de pares de datos, $\{(x_1, y_1), \dots, (x_n, y_n)\}$, se han obtenido los siguientes resultados:

$$\text{cor}(x, y) = 0,9; \quad \bar{x} = 5; \quad s_x^2 = 1,44; \quad \bar{y} = 10; \quad s_y^2 = 4,41$$

- a) Determinar la recta de regresión de Y sobre X .
- b) ¿Cuál sería la recta de regresión si $\text{cor}(x, y) = 0$?
- c) Responder a los apartados anteriores si el valor de cada una de las observaciones de la variable respuesta, y_i , aumenta en 2 unidades.