Estadística II
Ejercicios Tema 4

1. Los siguientes datos muestran la estatura (en cm.) y el peso (en Kg.) para una muestra de cinco alumnos de una clase:

estatura (cm.)	peso (Kg.)
154	60
158	62
162	61
171	66
176	84

   a) Calcula e interpreta los coeficientes de covarianza y de correlación lineal muestrales entre la estatura y el peso.

   b) Calcula estimadores puntuales para los parámetros (pendiente y constante) de la recta de regresión del peso en función de la estatura, así como para la varianza del error de la respuesta. ¿En cuántos Kg. aumenta el peso, en promedio, por cada 10 cm. adicionales de estatura?

   c) Calcula estimadores por intervalos al 10 % de confianza para los parámetros (pendiente y constante) de la recta de regresión del peso en función de la estatura, así como para la varianza del error de la respuesta.

   d) ¿Aportan los datos evidencia significativa al 10 % para concluir que el peso depende linealmente de la estatura? Plantea y resuelve el contraste de hipótesis correspondiente, y acota su p-valor. ¿Para qué niveles de significación puedes asegurar que el peso depende de la estatura?

   e) Supongamos que un alumno mide 174 cm. A partir de los datos dados, ¿cuál es el peso medio estimado de los alumnos que miden 174 cm? Da un intervalo de confianza al 95 % para el peso medio de los alumnos que miden 174 cm. Supongamos que cierto alumno de la clase mide 174 cm. Calcula un intervalo de predicción al 95 % de confianza para su peso.

2. 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

   a) Estimar por Mínimos Cuadrados un modelo de regresión simple para explicar el número de ejemplares vendidos en términos del PIB.

   b) 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?

   c) ¿Cuál será la venta de prensa que se podría predecir para una comunidad cuyo PIB por habitante fuese de 15000 euros?

   d) Si para una región cualquiera el valor del PIB aumentase en 2500 euros, ¿cómo cabría esperar que variase la venta de prensa diaria?

3. Se pretende estimar la relación entre el número de habitantes (x) de cada ciudad de España (medidos en millones) y el número de ventas de ejemplares (y) de un cierto libro (medidos en miles de unidades) en dicha ciudad. Una muestra tomada sobre cinco ciudades arroja los siguientes datos:

$$\bar{x} = 1,\ \bar{y} = 22,\ \sum_{i=1}^{5} x_i^2 = 5{,}98,\ \sum_{i=1}^{5} y_i^2 = 3118, \sum_{i=1}^{5} x_i y_i = 136.$$

a) Construye un modelo de regresión lineal simple que modele las ventas del libro en función del número de habitantes de cada ciudad. Interpreta los coeficientes del modelo desde un punto de vista económico.

b) Si la población de Bilbao es aproximadamente de 350.000 habitantes, efectúa una predicción de cuántos libros se venderán en 2011 en dicha ciudad. ¿Es razonable, a un 95 % de confianza, suponer que se venderán 12.000 libros? Puede utilizarse que $\sum_{i=1}^{5} e_i^2 = 8{,}204$, donde los e_i son los residuos de la regresión.

c) Contrasta mediante el p-valor si los habitantes producen un efecto lineal significativo en las ventas. Interpreta los resultados.

d) ¿Cómo cambiará el modelo de regresión si decidimos medir las ventas de libros en unidades y el número de habitantes en miles?

4. 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

(a) Calcular el coeficiente de correlación entre ambas variables, existe una relación de tipo lineal entre ellas?

(b) Estimar un modelo de regresión lineal que permita estimar la cuantía mensual de la factura telefónica en términos del tiempo de conexión.

(c) 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.

(d) ¿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?

5. 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?

6. 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 número 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.

7. 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
Amoníaco 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).

8. 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 $\bar{x} = 3$.

9. 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.

10. Una empresa desea investigar la relación entre el número 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:

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

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.

11. A partir de una muestra de pares de datos, $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, se han obtenido los siguientes resultados:

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

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.

12. Los siguientes datos corresponden a una muestra de 10 alumnos universitarios seleccionados al azar donde la variable X representa el número medio de hijos de sus abuelos y la variable Y el número de hijos de sus padres:

X :	6	4	3	4	6,5	2	4,5	3	5	1
Y :	4	3	4	4	8	1	4	5	4	2

Se pide:

a) Calcular el coeficiente de correlación entre las dos variables. Interpretar el resultado.

b) Construir la recta de regresión que explique la variable Y en función de los valores de X e interpretar su significado. ¿Cuáles son los valores de los parámetros del modelo de regresión lineal simple? ?Cuál es el significado de la estimación del coeficiente de regresión (coeficiente de X)?

c) Hacer un contraste a un nivel de significación del 1 % para ver si la relación entre las dos variables es significativa.

13. Los médicos están interesados en estudiar la relación entre la dosis de un medicamento y el tiempo que necesita un paciente para recuperarse.

La siguiente tabla muestra, para una muestra de 5 pacientes, las dosis administradas (en gramos), x, y los tiempos de recuperación (en horas), y. Estos pacientes tenían características similares excepto por la dosis del medicamento que se les administró:

Dosis (gr)	1,2	1,0	1,5	1,2	1,4
Tiempo de recuperación (horas)	25	40	10	27	16

Datos de interés:

$$\sum_{i=1}^{5} x_i = 6{,}3, \quad \sum_{i=1}^{5} x_i^2 = 8{,}09, \quad \sum_{i=1}^{5} x_i y_i = 139{,}8$$
$$\sum_{i=1}^{5} y_i = 118, \quad \sum_{i=1}^{5} y_i^2 = 3310, \quad \sum_{i=1}^{5} e_i^2 = 6{,}42105$$

a) Calcular e interpretar el coeficiente de correlación lineal entre ambas variables.

b) Realizar un ajuste lineal que exprese el tiempo de recuperación en función de la dosis administrada.

c) Contrastar la hipótesis de que el tiempo de recuperación depende linealmente de la dosis suministrada.

d) Determinar cuál será el tiempo medio de recuperación previsto para aquellos pacientes similares a los que se les administren 135 centigramos de medicamento. Obtener un intervalo de confianza al 95 % para dicha predicción.

e) Determinar cuál será el tiempo de recuperación para un paciente al que se le administren 135 centigramos de medicamento. Obtener un intervalo de confianza al 95 % para dicha predicción.

f) Obtener la ecuación de la recta de regresión si expresásemos la dosis en centigramos y el tiempo de recuperación en minutos.

g) Después de responder a los apartados anteriores, completa la tabla que se obtendría como resultado de llevar a cabo este análisis utilizando software (completa los valores indicados con una interrogación):

```
Regression Analysis - Linear model: Y = a + b*X
-------------------------------------------------------------------------------------
Dependent variable: T_Recuperacion
Independent variable: Dosis
-------------------------------------------------------------------------------------
                              Standard              T
Parameter       Estimate        Error            Statistic         P-Value
-------------------------------------------------------------------------------------
Intercept          ?           4,7732             20,3659          0,0003
Slope              ?             ?                   ?                ?
-------------------------------------------------------------------------------------
```