

ESCUELA UNIVERSITARIA DE ESTUDIOS EMPRESARIALES  
ECONOMÍA APLICADA I



# ESTADÍSTICA

DIPLOMATURA EN CIENCIAS EMPRESARIALES  
(PRIMER CURSO)

RELACIÓN DE EJERCICIOS Y PROBLEMAS SOBRE  
REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE

CURSO 2006-07

1. Sobre un conjunto de conductores se ha realizado una encuesta para analizar su edad (Y) y el número de accidentes que han sufrido (X). A partir de la misma, se obtuvieron los siguientes resultados:

X/Y	(20,30]	(30,40]	(40,50]	(50,60]	(60,70]
0	74	82	78	72	7
1	7	6	5	6	5
2	3	2	2	1	1

A partir de estos datos, se le pide que determine para esta distribución las curvas de regresión de Y sobre X y de X sobre Y.

2. Para la economía española, disponemos de los datos anuales redondeados sobre consumo final de los hogares a precios corrientes (Y) y renta nacional disponible neta (X), tomados de la Contabilidad Nacional de España base 1995 del INE , para el período 1995-2002, ambos expresados en miles de millones de euros:

Año	1995	1996	1997	1998	1999	2000	2001	2002
$Y_t$	258'6	273'6	289'7	308'9	331'0	355'0	377'1	400'4
$X_t$	381'7	402'2	426'5	454'3	486'5	520'2	553'3	590'0

Considerando que el consumo se puede expresar como función lineal de la renta ( $Y_t = a + b \cdot X_t$ ), determine:

- a) Los parámetros a y b de la recta de regresión.
  - b) La varianza de la variable consumo y su descomposición en varianza explicada y no explicada por el modelo.
  - c) El coeficiente de determinación de dicha regresión.
  - d) La predicción del valor que tomará el consumo para una renta de 650.000 millones de euros.
3. Se supone que se puede establecer cierta relación lineal entre las exportaciones de un país y la producción interna de dicho país. En el caso de España, tenemos los datos anuales (expresados en miles de millones de pesetas) para tales variables correspondientes al quinquenio 1992-96 en la siguiente tabla:

Años	Producción	Exportaciones
1992	52.654	10.420
1993	53.972	11.841
1994	57.383	14.443
1995	61.829	16.732
1996	65.381	18.760

A partir de tal información, y considerando como válida dicha relación lineal, se pide:

- a) Si la producción para el año 1997 fue de 2.210.6100 millones de pesetas, ¿cual sería la predicción de las exportaciones para este año? ¿Qué grado de precisión tendría dicha predicción?

- b) Si sabemos que las exportaciones para 1997 fueron de 69.045.704 millones de pesetas, ¿cuál sería la producción interna aproximada para ese año? ¿Qué grado de confianza daría usted a esta predicción?
- c) ¿Qué tanto por ciento de la varianza de las exportaciones no vienen explicadas por la producción interna, y se debe a otro tipo de variables?

4. De una distribución de dos variables se conocen los siguientes datos:

$$r_{xy} = 0.9; s_x = 1.2; s_y = 2.1; \bar{x} = 5; \bar{y} = 10$$

A partir de los mismos, obténganse las rectas de regresión mínimo cuadráticas de X sobre Y y de Y sobre X.

5. Para un mismo grupo de observaciones de las variables X e Y, hemos obtenido las dos rectas de regresión siguientes:

$$3x + 2y = 26$$

$$6x + 2y = 32$$

En función de las mismas, responda a las siguientes cuestiones:

- a) ¿Qué valores tomarían las medias de X e Y?
- b) Represente gráficamente ambas rectas de regresión.
- c) Determine el valor del coeficiente de correlación lineal  $r_{xy}$ .
- d) ¿Porqué la regresión de Y sobre X y la de X sobre Y no coinciden?
6. A partir de un conjunto de valores de las variables X e Y, se ha determinado la regresión de Y sobre X, obteniéndose la siguiente recta:

$$Y^* = 10 + 0.45X; \quad R^2 = 0.9; \quad \bar{x} = 20$$

Se pide que, a partir de la definición de la anterior recta, determine los parámetros de la recta de regresión de X sobre Y.

7. Se han observado, en varios modelos de automóviles, los datos de potencia del motor (X), en caballos, y la aceleración (Y), medida en el número de segundos necesarios para acelerar de 0 a 100 Km./h. La tabla adjunta refleja los valores obtenidos.

<b>X</b>	50	75	90	100	120	150
<b>Y</b>	15	12	10.5	10	9	8

A partir de tales datos, se ha decidido expresar la aceleración en función de la potencia mediante el ajuste de una función potencial (mediante el correspondiente cambio logarítmico). Bajo esta hipótesis:

- a) Determine la función de ajuste y el ECM conseguido.
- b) Si aumenta la potencia de un motor en un 10%, ¿en qué porcentaje repercutirá dicho aumento en la aceleración prevista? ¿Depende ello de la potencia que tenga el motor en cuestión?

8. En un nuevo proceso artesanal de fabricación de cierto artículo que está implantado, se ha considerado que era interesante ir anotando periódicamente el tiempo medio (medido en minutos) que se utiliza para realizar una pieza (variable Y) y el número de días desde que empezó dicho proceso de fabricación (variable X). Con ello, se pretende analizar cómo los operarios van adaptándose al nuevo proceso, mejorando paulatinamente su ritmo de producción conforme van adquiriendo más experiencia en él. A partir de las cifras recogidas, que aparecen en la tabla adjunta, se decide ajustar una función exponencial que explique el tiempo de fabricación en función del número de días que se lleva trabajando con ese método.

<b>X</b>	10	20	30	40	50	60	70
<b>Y</b>	35	28	23	20	18	15	13

Desde el correspondiente ajuste propuesto, se pide que determine:

- ¿Qué tiempo se predeciría para la fabricación del artículo cuando se lleven 100 días?
  - ¿Qué tiempo transcurriría hasta que el tiempo de fabricación que se prediga sea de 10 minutos?
  - ¿Qué porcentaje de tiempo se reduce por cada día que pasa?
9. La empresa COLOBONA S. A .L. ha trabajado hasta ahora con la hipótesis de que las ventas de un período dependen linealmente de los gastos de publicidad efectuados en el período anterior. En este momento, le solicitan a usted la realización de un análisis que ponga de manifiesto si la hipótesis, hasta ahora mantenida, se puede seguir sosteniendo en función de los datos que le suministran.

<b>AÑOS</b>	<b>GASTOS</b>	<b>VENTAS</b>
1987	21	...
1988	22	19
1989	25	20
1990	26	22
1991	27	23
1992	29	24
1993	30	26

En el informe final de su análisis, deberá responder a las siguientes preguntas:

- ¿Se incrementarán las ventas del período siguiente al aumentar los gastos en publicidad del período actual?
- ¿Es adecuado suponer que el ajuste entre estas variables es efectivamente lineal teniendo en cuenta los valores de las variables? Ajuste el modelo lineal e interprete los coeficientes del mismo. ¿Qué porcentaje de la varianza de las ventas no son explicadas por las variaciones de los gastos en publicidad?
- ¿Cuál será la predicción de las ventas para 1994? ¿Qué precisión tendrá ese pronóstico?
- Si para el año 1994 se piensa incrementar los gastos de publicidad en un 10%, ¿qué incremento relativo cabría esperar para las ventas de 1995 con respecto a las de 1994, según el modelo ajustado?

10. Un estudiante de la Escuela Universitaria de Estudios Empresariales de la Universidad de Sevilla, para poder pagarse sus estudios, debe trabajar como camarero en un bar de copas de su localidad. A este establecimiento, suelen acudir todos los jóvenes de la zona. Este año, con los conocimientos aprendidos, decide por fin estudiar la relación existente entre la cantidad de sal de las galletas saladas y el consumo de bebidas, ya que es costumbre dar al cliente este aperitivo cuando pide una consumición. Se sabe que las galletas no pueden tener una concentración de sal superior a 3'5 gramos por cada 1000 galletas y, por ello, decide ir variando a partir de 1 gramo la concentración de 0'5 en 0'5 gramos cada semana e ir anotando el incremento en caja semanalmente, obteniendo la siguiente tabla:

Gramos de sal por 1000 galletas	Ingresos (pesetas)
1	140300
1,5	150000
2	165000
2,5	175000
3	200000

A partir de tales cifras, se quiere conocer:

- ¿Considera justificado el planteamiento de un modelo lineal para expresar la relación entre las variables?
  - Si el propietario desea unos ingresos de 160.000 pesetas, ¿qué cantidad de sal debería aportar por cada 1000 galletas? Si aporta el máximo permitido de sal, ¿cuál sería el ingreso en caja? Explicar cuál de las dos predicciones le merece mayor confianza.
  - ¿Cuál sería la variación porcentual de los ingresos cuando la cantidad de sal aumenta en un 1% sobre el último valor de la tabla? Si aumentamos en 1gr. la sal por cada 1000 galletas, ¿cuánto variarán los ingresos?
11. Cierta empresa ha analizado estadísticamente la relación existente entre dos de las magnitudes que más le preocupan: los costes totales (Y) y los costes variables (X), expresados en unidades monetarias y considerados ambos a corto plazo. Analizados los datos se comprueba que:
- Si se construyeran las rectas de regresión de Y sobre X y de X sobre Y ambas rectas se cortarían en el punto de coordenadas (6,4).
  - Si para un coste variable de 12 unidades monetarias se incrementa el mismo en un 1%, el incremento que experimentarían los costes totales, bajo el modelo lineal de Y sobre X, sería del 0'8 %.
  - Si se incrementan en una unidad monetaria los costes totales, los costes variables aumentarían 1'8225 unidades monetarias bajo el modelo lineal de X sobre Y.
  - La dispersión relativa de los costes variables (X), medida en términos de su coeficiente de variación, ha sido del 50 %.
- Bajo estos supuestos, se pide que determine las rectas de regresión de Y sobre X y de X sobre Y y mida las bondades de ambos ajustes.
12. La dirección de una empresa ha llevado a cabo un análisis del coste salarial mensual S (datos en  $10^4$  pesetas) de sus 115 empleados en relación con un determinado indicador de productividad P, habiendo resultado la siguiente tabla de frecuencias conjunta:

Intervalos de S	Intervalos de P			
	6-8	8-10	10-12	12-14
10-12	28	2	0	0
12-15	3	31	1	0
15-18	2	19	2	0
18-22	0	6	6	0
22-28	0	0	8	1
28-38	0	0	0	6

A partir de estos datos, se pretende conocer lo siguiente:

- ¿En qué medida puede pensarse que el coste salarial explica mediante una relación lineal el indicador de la productividad?
- Supuesto que se ha calculado esta relación lineal, ¿para qué coste salarial se verifica que un incremento del 1% se traduce en un incremento del 0'5% del indicador de la productividad?

- Cierto distribuidor de productos de droguería y limpieza tiene su centro de distribución en una importante ciudad de nuestra provincia. Desde dicho centro reparte sus productos entre los comercios del sector de todos los pueblos de la comarca. Entre los artículos que distribuye, se ha seleccionado una muestra de 7 de ellos, que son los que se consideran más importantes, debido a su demanda por parte de los comerciantes y del público en general. De esos 7 artículos se tienen los datos correspondientes al mes de Abril último. Estos datos, que corresponden al precio unitario del artículo en pesetas (variable X) y al volumen de ventas correspondientes a dicho artículo en  $10^5$  pesetas (variable Y), son los siguientes:

- Si se construyera la recta de regresión que explicara el volumen de ventas en función del precio unitario, el porcentaje de varianza explicado sería de un 65'61%.
- Con dicha recta, para un artículo cuyo precio unitario fuese de 100 pesetas, el pronóstico que se efectuaría de ventas sería de 750.000 pesetas. Además, si dicho precio lo incrementásemos un 1%, las ventas crecerían un 0'333% (1/3%).
- Para un precio unitario igual al precio medio, se pronosticarían unas ventas de 850.000 pesetas.
- Para los datos de este mes de Abril, se conoce que la desviación típica de las ventas fue de 200.000 pesetas.

Se dispone, además, de los datos de ventas (en  $10^5$  pesetas) de esos mismos artículos correspondientes al mes de mayo:

Lejía	6'1
Amoníaco perfumado	8'2
Estropajo fibra verde	7'3
Detergente concentrado	14'1
Suavizante	12'3
Detergente lavavajillas	11'2
Limpiacristales	9'1

Teniendo en cuenta que los precios de esos artículos durante el mes de Mayo no han variado con respecto al mes anterior, y suponiendo que se mantiene la misma correlación entre precio y ventas que el mes de Abril, se pide:

- a) Hallar el precio medio de esos artículos y el volumen medio de ventas para el mes de Abril.
  - b) Construir la recta de regresión del volumen de ventas con respecto al precio unitario para el mes de Abril.
  - c) Construir la recta de regresión del volumen de ventas con respecto al precio unitario para el mes de Mayo.
  - d) Pronóstico de las ventas para el mes de Mayo de un artículo cuyo precio unitario sea de 100 pesetas. ¿Qué fiabilidad le merece dicho pronóstico?
14. En un determinado hotel, el responsable de la piscina del mismo debe añadir periódicamente un compuesto de cloro al agua para mantenerla en buenas condiciones. Dicha persona ha observado la relación existente entre el número de días que dura el efecto del producto (variable  $X_1$ ) y los gramos de cloro empleado (variable  $X_2$ ), obteniendo los siguientes resultados:
- $s_{12}=5,4$  días·gramo;  $s_2^2 = 12$  gramos<sup>2</sup>
  - El porcentaje de varianza explicada por la regresión lineal de  $X_1$  sobre  $X_2$  sería del 78'387%.
  - A partir de la regresión lineal de  $X_1$  sobre  $X_2$ , el valor estimado para 21 gramos de cloro sería de 4 días.
  - $\bar{x} = 25$  gramos
- A partir de esta información, determine ambas rectas de regresión y, en función de ellas, calcule qué cantidad de cloro habría que utilizar para que los efectos del producto durasen 7 días.
15. La factura mensual del gasto telefónico de una pequeña empresa se ha incrementado notablemente en los últimos meses. Los estudios realizados por el administrador de la misma argumentan que el mayor uso de Internet dentro de la misma es la principal causa del mayor gasto en teléfono, lo que ha hecho que se estudie la posibilidad de acogerse a alguno de los múltiples bonos o tarifas especiales que ofrecen las compañías, lo que hasta la fecha todavía no se llevó a cabo. Las últimas cifras mensuales no hacen sino confirmar esta relación:

Mes	Enero	Febrero	Marzo	Abril	Mayo
<b>Cuantía de la factura telefónica (€)</b>	55	100	118	120	142
<b>Tiempo de conexión (minutos)</b>	200	500	700	800	1000

De acuerdo con la información anterior, responda a las siguientes preguntas:

- a) Suponiendo la existencia de una relación de tipo lineal entre tiempo de conexión y gasto telefónico, ¿qué porcentaje de las variaciones en la cuantía de la factura telefónica no podrían ser explicadas linealmente por el tiempo de conexión a Internet dentro de la compañía?
- b) ¿Cuál sería la cuantía de la factura telefónica de la compañía de acuerdo a esta relación lineal si no se conectase a Internet en la empresa?
- c) ¿Cuál sería el gasto telefónico estimado según esta relación lineal si el tiempo de conexión a Internet fuera de 2000 minutos? ¿Le parece aceptable tal predicción? Razone su respuesta.
- d) Se considera que un incremento del 20% en el tiempo de conexión a Internet respecto al realizado en el mes de mayo conllevaría a que la factura telefónica se elevase de forma extraordinaria. ¿Cuál sería el incremento relativo en la misma si ello se produjese? Razone su respuesta.

16. En una muestra de familias se han analizado las variables ahorro anual (Y) y renta anual (X), medidas ambas en miles de euros. Los datos obtenidos han sido los siguientes:

<b>Ahorro (Y)</b>	1'9	1'8	2'0	2'1	1'9	2'0	2'2	2'3	2'7	3'0
<b>Renta (X)</b>	20'5	20'8	21'2	21'7	22'1	22'3	22'2	22'6	23'1	23'5

A partir de tales datos, se pide:

- Obtener el modelo lineal que explica el ahorro de las familias en función de su renta.
  - ¿Qué familia aumentaría en un mayor porcentaje su ahorro si su renta se viese incrementada en un 5%, la familia que tiene la menor renta de entre todas o la que posee la mayor renta?
  - ¿Cuál será el incremento absoluto del ahorro cuando una familia aumente su renta anual en 500 €?
  - ¿Qué porcentaje de varianza de la variable ahorro queda explicado por la variable renta a través del modelo lineal planteado?
17. 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:

<b>Precio (miles de €)</b>	<b>Cantidad vendida (unidades mensuales)</b>
7'5	450
9	425
10'5	400
12	350
14	325
16	300
18	290
20'5	280
23'5	260
27	200

En función a los datos recabados para esa región y mes:

- Una empresa radicada en la región tiene previsto para el mes próximo aumentar el precio de su modelo más vendido en 500 €. 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?
  - Si el modelo más caro de la tabla anterior se abaratase para el mes próximo un 3%, ¿cómo variarían las ventas de dicho modelo?
  - Obtenga la descomposición de la varianza total de las cantidades vendidas en varianza explicada y varianza no explicada por el modelo lineal y, a partir de ella, determine el coeficiente de determinación.
  - Si expresamos el precio en € 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 determinación de tal modelo?
18. Para 8 comunidades autónomas españolas, se ha recogido información sobre la venta en 1998 de prensa diaria escrita (variable Y, en ejemplares diarios vendidos por cada mil habitantes),



relacionándola con su producción económica, en concreto con el Producto Interior Bruto por habitante (variable X, en  $10^3$  €). Las cifras de ambas variables se recogen en la tabla siguiente:

<b>X</b>	8'3	9'7	10'7	11'7	12'4	15'4	16'3	17'2
<b>Y</b>	57'4	106'8	104'4	131'9	144'6	146'4	177'4	186'9

Fuentes: INE. Anuario Estadístico (datos adaptados y redondeados)

Con esta información, y teniendo en cuenta que  $\bar{x} = 12'7125 \cdot 10^3$  €,  $s_x^2 = 9'2436 \cdot (10^3 \text{ €})^2$ ,  $\bar{y} = 131'9750$  ejemplares y  $s_y^2 = 1550'4069$  ejemplares<sup>2</sup>, se pregunta:

- ¿Puede afirmarse que existe realmente una relación de tipo lineal entre las dos variables consideradas? Cuantifique y justifique la respuesta. ¿Qué porcentaje de la variabilidad de Y no viene explicada por su relación lineal con X?
- ¿Cuál sería la venta de prensa que se podría predecir para una comunidad cuyo PIB por habitante fuese de 25000 €? Y si se supiese que la venta diaria de prensa es de 150 ejemplares por cada mil habitantes, ¿qué PIB se podría predecir? ¿Cuál de las dos predicciones merece mayor confianza y por qué?
- Si para 1999, el PIB de la región más rica aumenta un 1%, ¿en qué porcentaje se estima que crecería la venta de ejemplares de prensa? Y si para una región cualquiera el valor del PIB aumenta en 2500 €, ¿cómo cabría esperar que variase la venta de prensa diaria?
- En la regresión de Y sobre X, calcule la suma de los cuadrados de los residuos correspondientes y el coeficiente de variación de las predicciones  $Y^*$  para los valores observados de X, pero sin calcular ni los residuos ni las predicciones. Exponga claramente en qué se ha basado para obtener los resultados.

19. Para un conjunto de personas que están siguiendo una dieta de adelgazamiento, se han recogido datos sobre el peso perdido desde el inicio de la misma (variable Y, en Kg.) y el tiempo que llevan siguiendo la dieta (variable X, en semanas), los cuales se muestran en la siguiente tabla:

<b>Y</b>	2'4	5'4	5'6	8'4	10'6	13'5	15	15
<b>X</b>	3	5	6	8	11	13	15	16

A partir de esta información, responda a las siguientes cuestiones:

- Estime el modelo lineal que explica el peso perdido en función del tiempo que se lleva siguiendo la dieta e interprete los parámetros.
- Para el modelo estimado en el apartado anterior, descomponga la varianza total como suma de la explicada y la no explicada por el mismo y obtenga, a partir de tal descomposición, el coeficiente de determinación.
- Según el modelo considerado, ¿qué peso esperaría perder una persona que siga la dieta durante 2 meses (8 semanas)? ¿Y una persona que esté dispuesta a seguir la dieta durante dos años (108 semanas)? ¿Qué fiabilidad le otorga a cada una de las estimaciones anteriores?

20. Conocemos, para una gasolinera situada en Sevilla, la información acerca de su recaudación durante las últimas 7 semanas, así como del número de clientes que acudieron a la misma durante estos períodos:

<b>Recaudación (<math>10^3</math> €)</b>	1'5	10	8	3	5	15	2
<b>Nº de clientes (<math>10^2</math>)</b>	3	6	5	3'5	4	8	3'2

A partir de tal información, responda razonadamente a las siguientes cuestiones:

- De acuerdo con los datos anteriores, y a partir de un ajuste lineal que exprese la recaudación en función del nº de clientes, determine cuál sería la recaudación prevista si llegasen a la gasolinera 720 clientes. ¿Qué fiabilidad otorgaría a dicha predicción?
- ¿Qué variación experimenta la recaudación por cada 10 clientes más? ¿Y por una disminución del 3% en los clientes sobre la media?
- Si expresásemos la recaudación en € y el número de clientes en unidades, ¿cuál sería la expresión de la recta de regresión? ¿Se modificaría la bondad del ajuste?

21. Una juguetería ha examinado la evolución reciente de las ventas de su muñeco “Qco” (Y, en millones de euros) junto con los gastos de publicidad de ese muñeco (X, en millones de euros), obteniéndose los siguientes resultados:

<b>Año</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>
<b>Y</b>	126	135	156	156	150	150	180
<b>X</b>	20	25	30	32'5	35	32	34

Además se sabe que:

$$\bar{x} = 29'79; \bar{y} = 150'43; s_x^2 = 25'1327; s_y^2 = 253'1020; s_{xy} = 64'3776$$

A partir de esta información, responda a las siguientes cuestiones:

- ¿Se puede afirmar que al aumentar los gastos en publicidad se incrementarán las ventas? Obtenga los parámetros del ajuste lineal que explique las ventas en función de los gastos. Interprete dichos coeficientes e indique la bondad del ajuste realizado.
- Si para el año 2003 aumentásemos los gastos de publicidad en un 1%, ¿en qué porcentaje se espera que variasen las ventas, según el modelo lineal?
- Suponiendo que entre X e Y existe la relación  $Y^* = A \cdot X^b$ , calcular A y b. ¿Qué utilizaría para medir la bondad de dicho ajuste?
- Si para el años 2003 aumentamos los gastos de publicidad en un 1%, ¿en qué porcentaje se espera que varíe las ventas, según el modelo del apartado c)?

## Distribuciones tridimensionales

22. No es algo desconocido el hecho de que las entradas de turistas extranjeros en España no han hecho sino crecer de forma continuada durante las últimas décadas, coadyuvando a este hecho las peculiares características climatológicas, económicas y culturales de nuestro país. A partir de esta idea de principio, se ha querido ahondar en el tema, analizando si la riqueza del turista y la distancia de su país al nuestro son factores vinculantes a la hora de tomar la decisión de visitarnos. Para ello, hemos recogido en la siguiente tabla las entradas de turistas procedentes de 5 países diversos para un determinado ejercicio económico, junto con su PIB per cápita (en miles de euros) y la distancia en kilómetros desde cada uno de los países considerados al nuestro:

País	Nº turistas entrados	PIB per cápita ( $10^3$ €)	Distancia (Kms.)
A	1800000	3'1	1500
B	2500000	4'2	1200
C	700000	2'7	3100
D	1200000	3'0	2500
E	1900000	4'0	1800

A partir de tal información, se le pide que:

- Determine la relación lineal que explicaría las entradas de turistas en función del PIB per cápita y de la distancia.
  - ¿Es dicho modelo suficientemente explicativo?
  - ¿Cuál de las dos variables explicativas está más correlacionada con la entrada de turistas?
23. La siguiente tabla proporciona los valores para un determinado año de las variables  $X_1$  "Saldo de imposiciones",  $X_2$  "Renta Familiar Disponible" y  $X_3$  "Renta per cápita", para 10 provincias españolas:

Provincias	$X_1$	$X_2$	$X_3$
Alicante	27	86	96
Almería	7	29	71
Baleares	24	61	120
Málaga	20	67	77
Soria	8	45	80
Tarragona	5	9	89
Tenerife	17	43	107
Toledo	9	37	81
Zamora	3	17	65
Zaragoza	35	71	102

Desde tales cifras, determine:

- a) La matriz de varianzas-covarianzas y la matriz de correlaciones de la distribución.
- b) La relación lineal que explique  $X_1$  mediante  $X_2$  y  $X_3$ .
- c) El coeficiente de correlación múltiple entre  $X_1$  y  $X_2$ ,  $X_3$  y los coeficientes de correlación parcial de  $X_1$  con  $X_2$  y de  $X_1$  con  $X_3$ . Interprete cada uno de tales coeficientes.

24. Porfirio Labrador es un joven agricultor de la provincia de Sevilla, que adquirió conocimientos de estadística en su formación universitaria en esta escuela. Tiene una finca de su propiedad, dedicada fundamentalmente a la producción de trigo duro especial para la fabricación de pastas, dadas las importantes subvenciones que recibe de la Comunidad Europea. Dada la influencia que supone que ejerce la cantidad de lluvia caída sobre la producción, utiliza un pluviómetro para medir ésta, mes a mes. En el cuadro adjunto aparece recogida la producción de trigo duro en Quintales métricos por hectárea ( $X_1$ ) y la cantidad de lluvia caída en su finca, medida en litros por metro cuadrado ( $X_2$ ), durante las cinco últimas campañas:

CAMPAÑA	$X_1$	$X_2$
89/90	80	120
90/91	95	145
91/92	83	108
92/93	75	90
93/94	72	85

- a) Este agricultor, que es pesimista como la mayoría de los de su sector, piensa que la pertinaz sequía va a continuar para la próxima campaña 94/95, y que la cantidad de lluvia para esa campaña va a ser la misma que la recogida en la campaña anterior. Desde este supuesto, ¿qué predicción haría para la producción de la campaña venidera? ¿Qué fiabilidad otorgaría a tal predicción? Si pensara que la cantidad de lluvia recogida aumentase en un 1%, ¿cuál sería dicha predicción?
- b) La cantidad de abono (urea y nitrato) que utiliza, medida en quintales métricos por hectárea ( $X_3$ ), factor que influye también en la producción alcanzada, se ha constatado que es prácticamente la misma en cada campaña, siendo la media de las últimas 5 campañas de 3 Qm./Ha. con un coeficiente de variación del 2'9814%. La recta de regresión de  $X_1$  sobre  $X_3$ , obtenida con los datos de las últimas cinco campañas es  $X_1^* = 51 + 10 \cdot X_3$ , siendo la covarianza entre  $X_2$  y  $X_3$  de 0'04. Si el Sr. Labrador utilizase durante la próxima campaña 3 Qm./Ha., y bajo el supuesto de que la cantidad de lluvia caída no se modificase respecto de la recogida en la campaña 93/94, ¿qué producción se alcanzaría en la campaña 94/95, considerando las dos variables explicativas? ¿Qué fiabilidad le otorgaría a tal predicción?

25. Un estudiante de la Universidad de Detroit observó, en su estancia desde el primer trimestre de 1995 hasta el cuarto trimestre de 1998, los importantes beneficios que habían obtenido las floristerías de la ciudad y pensó que ello era debido a la gran demanda de rosas que existía en la misma durante cualquier época del año. Después de analizar con cierto detenimiento el fenómeno, recabó información acerca de las variables cantidad de rosas vendidas ( $X_1$ ), expresada en docenas por trimestre, precio medio de las rosas ( $X_2$ ), expresada en dólares por docena y trimestre, e ingreso medio disponible semanal ( $X_3$ ), expresado en dólares por trimestre. A partir de los datos proporcionados por un mayorista del área metropolitana de Detroit para los 16 trimestres considerados, realizó una serie de cálculos previos que arrojaron los siguientes valores:

$$\begin{aligned}\sum x_{1i} &= 121; \sum x_{2i} = 52; \sum x_{3i} = 2887 \\ \sum x_{1i}^2 &= 981; \sum x_{2i}^2 = 174; \sum x_{3i}^2 = 522407 \\ \sum x_{1i} \cdot x_{2i} &= 380; \sum x_{1i} \cdot x_{3i} = 21705; \sum x_{2i} \cdot x_{3i} = 9420\end{aligned}$$

Desde estas cifras, se pretende analizar la influencia lineal que, sobre la demanda de rosas, ejercen las otras dos variables consideradas. Para ello, responda a las siguientes cuestiones:

- ¿Cuál sería la ecuación matemática que defina la correspondiente regresión múltiple?
- A partir de la anterior relación, determine la contribución marginal del precio medio de las rosas y del ingreso medio familiar disponible sobre la demanda de rosas. Cuantifique y razone su respuesta.
- ¿Se puede afirmar que la subida del precio de las rosas conduce a una menor demanda de las mismas? Razone su respuesta.
- ¿Qué porcentaje de la variabilidad de la demanda puede explicarse por la influencia lineal del precio medio y del ingreso familiar?
- Si para el segundo semestre de 1999 se previese una reducción del precio medio de 0'5 dólares/docena, respecto al último trimestre de 1998, no experimentando el ingreso familiar disponible variación alguna en dicho período, ¿cuál sería la variación absoluta que se esperaría en la demanda de rosas en ese período?

26. En una determinada economía se conocen los datos relativos a tres indicadores, expresados como tasa de variación (porcentaje respecto al período inicial): número de empresas ( $X_1$ ), empleo ( $X_2$ ) e inflación ( $X_3$ ). Acerca de ellos, se tiene determinada información:

- ü El valor medio de la inflación es de 4'057, mientras que para el empleo, el valor medio es de 2'9 y la varianza de 2'2714. Además, el número medio de empresas se situó en 4'1.
- ü La inflación explica el 89'41% de la varianza del número de empresas, mientras que los datos de empleo explican el 94'97% de la varianza del número de empresas, entendiendo que estas explicaciones son a través de relaciones lineales y directas.
- ü El grado de relación lineal que se aprecia entre la inflación y el empleo, medido a través de la covarianza, es de 1'7985, lo que indica una relación directa. Por otra parte, el coeficiente de correlación entre estas variables se sitúa en 0'95385.
- ü El incremento que experimenta la tasa de inflación cuando la tasa de variación del número de empresas aumenta un punto porcentual es de 0'6126. Por otra parte, si la tasa de inflación aumenta en un punto, la tasa de variación del número de empresas se incrementa, a su vez, en 1'4602 puntos porcentuales.

Con estos datos, se desea conocer:

- ¿Cuál de las tres variables está mejor representada por su media aritmética?
- ¿Cuál de las variables, individualmente sin el efecto de las otras, explica mejor (en el sentido de una relación lineal) las tasas de variación del número de empresas, la inflación o el empleo?
- Si se incrementa la tasa de inflación en un punto, sin que el empleo cambie, ¿en cuánto variará la tasa de variación del número de empresas?
- ¿Qué valor alcanzará la tasa de variación del número de empresas cuando las tasas de inflación y empleo se sitúan en la mitad de sus valores medios. ¿Cómo mediría el grado de credibilidad o fiabilidad del valor resultante?

27. Una determinada empresa se ha planteado investigar si resultan rentables los gastos en publicidad para sus productos, utilizando para ello las cifras de gastos y beneficios (en miles de millones de pesetas), proporcionados por algunas empresas de su sector:

Empresa	A	B	C	D	E	F	G
Gastos en publicidad ( $X_2$ )	0'3	1	0'7	1	1'2	2	1'1
Beneficios ( $X_1$ )	1'3	3'5	2'8	3	3'3	4	3'7

En todo caso, conoce que una variable que también podría coadyuvar sobre tales beneficios sería el número de horas extraordinarias que mensualmente trabajan los empleados ( $X_3$ ). A partir de los datos recogidos para tales empresas, conocemos también que:

$$\hat{\sigma}_{23}=0'58$$

$$\hat{\sigma}_{13} \text{ La recta de regresión estimada de } X_1 \text{ sobre } X_3: X_1^* = -0'18 + 0'37 \cdot X_3, \text{ con } R_{13}^2 = 0'697$$

Desde esta información, ¿podría indicar qué variable ejerce más influencia sobre el beneficio anual de las dos consideradas? Razone su respuesta.

28. Una determinada cooperativa agrícola dedicada a la producción y comercialización de fresas se plantea hacer un estudio para explicar el volumen de sus ventas ( $X_1$ ), expresadas en millones de pesetas. Para ello, a partir de los datos semestrales obtenidos desde la creación de la cooperativa, se plantea un modelo lineal usando como variables explicativas el gasto en publicidad ( $X_2$ ), expresado en millones de pesetas, y el número de supermercados que comercializan sus productos ( $X_3$ ). Seguidamente recogemos la información de la que dispone la empresa:

Semestre	$X_2$	$X_3$	$X_1$
2º de 1994	1'5	15	13
1º de 1995	1'7	17	41
2º de 1995	2	19	16
1º de 1996	2'3	20	47
2º de 1996	2'5	23	19
1º de 1997	3	25	55
2º de 1997	3'5	26	22
1º de 1998	4	29	63
2º de 1998	4'3	31	25
1º de 1999	4'5	34	65

Desde tales cifras, responda a las siguientes cuestiones:

- Determine el vector de medias, la matriz de varianzas y covarianzas y la matriz de correlación de la distribución. A partir de tal información, ¿qué variable considera que está mejor representada por su media? ¿Entre qué dos variables existe una mayor relación lineal?
- Determine el modelo de regresión planteado e interprete sus coeficientes, indicando a su vez cuál es la capacidad explicativa del mismo.
- A partir de los resultados obtenidos en el apartado anterior, ¿sugeriría usted que pueden existir otras variables no tenidas en cuenta aquí y de importancia relevante para la explicación de las ventas?

Observando los datos, nos damos cuenta que las ventas correspondientes al primer semestre son siempre mucho mayores que las del segundo semestre del mismo año. Para recoger este aspecto,

se decidió introducir una nueva variable explicativa  $X_4$ , cuyo valor sería uno si el dato corresponde al primer semestre, y cero en caso contrario. Utilizando esta variable adicional:

- d) Determine el nuevo modelo de regresión planteado e interprete sus coeficientes, indicando la capacidad explicativa del mismo.
- e) Bajo este modelo, ¿en cuánto se incrementarían las ventas en el segundo semestre de 1999 (con respecto al segundo semestre de 1998) si los gastos en publicidad se incrementasen en 500000 pesetas y se espera que dos nuevos supermercados comercialicen los productos de la cooperativa? ¿Y si los incrementos fuesen del 2% en ambas variables explicativas?

29. La Asociación de Consumidores y Usuarios de una determinada ciudad andaluza está analizando la relación existente entre tres variables que consideran íntimamente interconectadas y de influencia en la economía de la ciudad, como son: la renta bruta ( $X_1$ ), la renta disponible ( $X_2$ ) y el nivel de consumo medio de los ciudadanos ( $X_3$ ), expresadas todas en miles de unidades monetarias. Para ello, seleccionó una muestra de 100 hogares de la ciudad y obtuvo los siguientes datos:

- Ü Los valores medios de las variables enunciadas se situaron, respectivamente, 45'35; 32'06 y 16'42.
- Ü La varianza de  $X_1$  resultó ser de 43'98 y la de  $X_3$  6'28.
- Ü La covarianza entre  $X_1$  y  $X_2$  fue de 25'25; mientras que la existente entre  $X_1$  y  $X_3$  alcanzó un valor de 16'41.
- Ü Si se establece la relación lineal entre el consumo y la renta disponible, la pendiente de tal recta de regresión sería de 0'64; logrando explicar la renta disponible el 99'32% de las variaciones del consumo.

A partir de la información suministrada, responda a las siguientes cuestiones:

- a) ¿Cuál sería la ecuación del plano de regresión que explica el consumo en función de la renta bruta y la renta disponible, sabiendo que el determinante de la matriz de varianzas y covarianzas es igual a 0'4214? Interprete los parámetros del modelo y valore razonadamente su fiabilidad.
- b) ¿Qué variable cree que tiene realmente más influencia (en términos lineales) sobre el consumo, la renta bruta o la disponible? Cuantifique su respuesta.
- c) Si se produce un incremento de la renta bruta en 1000 u.m. en un determinado período, ¿qué incidencia tendría este hecho en el consumo estimado, supuesto que la renta disponible se mantiene constante?

30. Se han recogido datos sobre  $X_1$  = "Número de habitantes (en miles)",  $X_2$  = "Número de lectores habituales de periódicos por cada 100 habitantes" y  $X_3$  = "Renta por habitante en unidades monetarias" en varios municipios de Andalucía. De los datos disponibles se deduce que:

$$\bar{x}_1 = 10'542; \bar{x}_2 = 24; \bar{x}_3 = 84; s_1 = 12'342; s_2 = 3'2; s_3 = 14; r_{12} = 0'25; r_{13} = 0'33; r_{23} = 0'19$$

A partir de tal información:

- a) ¿Qué grado de ajuste se conseguirá al expresar la renta por habitante como función lineal del número de habitantes y de la proporción de lectores de periódicos?
- b) Según la anterior función lineal, ¿qué variación ocurrirá en el pronóstico de  $X_3$  al disminuir el número de habitantes en 1.500 personas (manteniéndose constante  $X_2$ )?

- c) ¿En qué grado y cuánta variaría la proporción de lectura de periódicos al incrementarse la renta, suponiendo constante el tamaño de la población del municipio? Cuantifique la respuesta.

31. En una conocida empresa hortofructícola se decidió dedicar hace cinco años un pequeño terreno a la producción de fresas, habiendo obtenido unos resultados altamente satisfactorios durante ese período. Intentando profundizar en las cifras, se ha descubierto una relación muy significativa en términos lineales entre la producción obtenida (en miles de toneladas de fresas), el volumen de agua utilizado (en miles de litros) y la cantidad de abono específico utilizado (en kilos), que hacen ser muy optimistas a los propietarios de la empresa para el futuro, toda vez que ambas variables pueden ser controlables.

Para el pasado quinquenio los datos correspondientes a las tres variables arrojaron las siguientes cifras:

Año	Producción ( $X_1$ )	Agua ( $X_2$ )	Abono ( $X_3$ )
1998	120	100	10
1999	150	130	12
2000	200	175	14
2001	270	240	17
2002	350	310	20
Medidas	Producción ( $X_1$ )	Agua ( $X_2$ )	Abono ( $X_3$ )
Media	218	191	14'6
Varianza	6936	5764	12'64

Se conoce además la siguiente información:

- ü La recta de regresión de la producción sobre la cantidad de abono utilizado ( $X_1^* = -123 + 23'354 \cdot X_3$ ) consigue explicar el 99'4% de las variaciones en la producción.
- ü La covarianza entre las variables agua y abono utilizado es igual a 269'4.
- ü Para la matriz de varianzas y covarianzas:  $c_{11} = 280'6$ ;  $c_{22} = 528$ ;  $c_{33} = 11420$ ;  $|C| = 104$

A partir de este conjunto de datos, se le pide que responda a las siguientes cuestiones:

- a) Calcule la matriz de varianzas y covarianzas de las tres variables consideradas.
- b) Si se hubiera incrementado la cantidad de agua utilizada en 40.000 litros y el abono en 5 Kg. para 2003 respecto de las cantidades de 2002, ¿cómo habría variado la producción de fresas? ¿Qué fiabilidad le merece tal predicción?
- c) ¿Cómo habría variado la producción de fresas para 2003 si sólo se hubiera incrementado por razones presupuestarias la cantidad de agua utilizada en 20.000 litros, permaneciendo fija la cantidad de abono utilizado? ¿Qué fiabilidad le merece tal predicción?
- d) ¿Qué influye más en la producción de fresas, el volumen de agua utilizado o la cantidad de abono utilizado? Razone su respuesta.

32. Se dispone de los siguientes datos para 6 familias sobre gasto anual en libros de texto ( $X_1$ ), número de miembros ( $X_2$ ) y renta anual ( $X_3$ ):



Gasto anual (10 <sup>3</sup> u.m.)	Número de miembros	Renta anual (10 <sup>6</sup> u.m.)
20	1	1
30	2	2
45	3	3
35	2	2'5
30	2	1'8
40	2	3'2

Adicionalmente, se conoce también que

- Ü El determinante de la matriz de varianzas y covarianzas  $C$  es 0'050.
- Ü El adjunto  $C_{11}$  es igual a 0'075.
- Ü El adjunto  $C_{22}$  es igual a 2'651.
- Ü El adjunto  $C_{33}$  es igual a 3'914.
- Ü Las varianzas correspondientes a las tres variables son iguales a 63'889; 0'333 y 0'559 respectivamente.

A partir de tal información, responda a las siguientes cuestiones:

- a) Calcule la matriz de varianzas y covarianzas de las tres variables consideradas.
- b) Determine los parámetros de la función lineal que explica el gasto anual en libros a partir del número de miembros y de la renta anual.
- c) ¿Qué variación cabría esperarse en el gasto ante un aumento de la renta en un millón de u.m., permaneciendo constante el número de miembros de la familia?
- d) Determine si el modelo es adecuado para realizar predicciones. ¿Qué proporción de la varianza del gasto anual en libros explica el modelo?
- e) ¿Podría afirmarse que, en términos estrictos, existe una fuerte relación lineal entre el gasto anual y el tamaño familiar? Razone su respuesta.

33. Como es bien sabido, se puede establecer una dependencia entre el consumo de una familia (variable  $X_1$ , en miles ptas./mes) y su renta (variable  $X_2$ , en miles ptas./mes). Con el objeto de concretar tal relación en una determinada economía, se recogieron datos referentes a 200 familias, obteniéndose los siguientes resultados:

$X_1 \backslash X_2$	50 – 100	100 - 150	150 - 200	200 - 250	250-350
0 - 50	14	5	2	0	0
50 – 150	2	19	62	3	1
150 - 300	0	0	1	62	3
300 - 350	0	0	0	0	26

A partir de ellos, se desea conocer, suponiendo que existe una relación lineal entre ambas variables,

- a) ¿Cuál sería el consumo esperado para una renta nula?
- b) ¿Cuál sería la propensión marginal al consumo de estas familias?
- c) ¿Qué porcentaje de varianza de  $X_1$  no vendría explicada por  $X_2$ ?
- d) Con objeto de disminuir el porcentaje del apartado anterior se consideró una tercera variable,  $X_3$ , el número de hijos de la familia. Constrúyase el modelo de regresión lineal múltiple para explicar  $X_1$ , si sabemos que, a partir de dicho modelo, si una familia aumenta

en 2 el número de hijos, manteniéndose la misma renta, el consumo se incrementaría en 29.176 ptas. Además la correlación parcial entre  $X_1$  y  $X_3$  resultó ser de 0'3327; y la covarianza entre dichas variables es de 16 ( $10^3$  ptas. por hijo).

- e) Calcule el coeficiente de determinación del plano de regresión de  $X_1$  sobre  $X_2$  y  $X_3$ .

34. Un hipermercado ha decidido ampliar el negocio y decide estudiar de forma exhaustiva el número de cajas registradoras que va a instalar, para evitar grandes colas. Para ello, se obtuvieron los siguientes datos procedentes de otros establecimientos similares acerca del número de cajas registradoras (variable  $X_2$ ) y del tiempo medio de espera (variable  $X_1$ )

<b>Número de cajas registradoras</b>	10	12	14	16	18	20
<b>Tiempo medio de espera (minutos)</b>	59	51	42	32	26	22

Bajo el supuesto de que el tiempo de espera medio depende linealmente del número de cajas registradoras, se pretende saber:

- ¿Cómo variaría el tiempo de espera por cada unidad de caja adicional?
- Si se instalarán 22 cajas registradoras, ¿cuál sería el tiempo medio de espera? ¿Es fiable dicho pronóstico?
- Posteriormente, se pensó que podría ser conveniente para explicar el tiempo de espera tener en cuenta además el número de clientes que acuden diariamente a estos establecimientos (variable  $X_3$ ), por lo que se estudio el tema, obteniéndose un coeficiente de correlación lineal entre ambas variables de 0'9343, siendo el número medio de clientes para los establecimientos anteriores de 850 personas, con un coeficiente de variación del 10'74%. Además, la covarianza entre el número de cajas y el número de clientes resultó ser de -300. Calcule los parámetros del modelo lineal que explica el tiempo medio de espera, en función del número de cajas y del número de clientes que acude directamente.

35. Del Instituto Nacional de Estadística se ha obtenido la siguiente información del año 2001 referida a 8 de las comunidades autónomas españolas, respecto al consumo de agua en los hogares, donde las variables  $X_1$  = litros por habitante y día y  $X_2$  = precio del  $m^3$  de agua en céntimos de euro y  $X_3$  = media de habitantes por hogar, siendo los datos los siguientes:

$X_1$	183	171	114	125	174	200	134	163
$X_2$	64	59	145	166	53	48	95	50
$X_3$	4'1	4'2	3'9	3'3	3	2'6	4'2	3'5

Desde esta información, responda a las siguientes cuestiones:

- ¿Qué porcentaje de la variabilidad en el consumo de agua no viene explicado por su precio?
- ¿Cuánto se incrementa en términos absolutos el consumo de agua cuando se aumenta en 1 céntimo el precio del  $m^3$ ?
- ¿Qué consumo de agua se podría predecir si el precio del agua fuese de 3 euros por cada  $m^3$ ? ¿Qué grado de confianza le merece tal predicción?
- Si el gobierno de la comunidad en la que el agua es más barata (0.48 €) decide subir su precio un 1 %, ¿en qué porcentaje disminuiría el consumo de agua en dicha comunidad?

- e) Calcule la desviación típica de los residuos que se obtendrían al explicar el consumo de agua a través del precio para las 8 comunidades cuyos datos están disponibles ¿Cuál sería el coeficiente de variación de Pearson de tales residuos?
- f) Calcule la expresión lineal mínimo cuadrática que explica  $X_1$  en función de la variable  $X_1'$ , que sería el Precio del  $m^3$  de agua en euros.
- g) Obtenga el modelo de regresión lineal mínimo cuadrático que explica  $X_1$  en función de las otras dos variables. ¿Qué predicción del consumo tiene lugar si el precio del agua es de 1.5 € por  $m^3$  y el número de habitantes por hogar es de 4.2?
- h) ¿Qué influye más en la variación absoluta del consumo de agua: la subida de 1 céntimo de euro en el precio del  $m^3$  sin alterar el valor de  $X_3$  o el aumento en una unidad de  $X_3$  sin que se altere el precio del  $m^3$ ? Justifique claramente la respuesta.
- i) ¿En cuánto disminuye la proporción de la varianza de  $X_1$  explicada por el modelo del apartado a) si se utiliza  $X_2$  como única variable explicativa del comportamiento de  $X_1$ ?

36. Para montar una empresa dedicada a la venta de artículos deportivos, un empresario analiza la información correspondiente al beneficio (variable  $X_1$ , expresada en miles de euros) e inversión (variable  $X_2$ , expresada en miles de euros) de 30 empresas del sector que ya están en marcha, obteniendo la siguiente información:

$X_1 \backslash X_2$	0-2	2-5	5-10
6-10	10	2	0
10-14	1	4	1
14-20	0	2	10

- a) Si suponemos una relación lineal entre beneficios e inversión, determine el beneficio que se obtendría en el caso de que la inversión se situase en el valor que más se da y cuantifique la fiabilidad de esta predicción.

Para aumentar la información, el empresario antes citado analiza el volumen de ventas (variable  $X_3$ , expresada en miles de euros) de las mismas 30 empresas. Obtiene que el volumen medio de ventas de éstas fue de 25000 €, con una dispersión del 32%, medida a través de su coeficiente de variación. Se sabe además que las variaciones del beneficio se deben en un 90% a variaciones en las ventas según el correspondiente modelo lineal entre las dos variables y se conoce también la existencia de una correlación directa entre las ventas y la inversión, que determina un coeficiente de correlación lineal simple entre ambas variables igual a 0.89. A partir de estos datos, responda a las siguientes cuestiones:

- b) Calcule, si decide invertir 8000 € y espera unas ventas de 23000, cuál sería el beneficio previsible y la fiabilidad que otorgaría a tal predicción.
- c) Determine qué variable en sentido estricto, la inversión o el volumen de ventas, ejerce más influencia sobre los beneficios.

37. Una empresa está analizando las ventas de un perfume que comercializa. Para ello dispone de las ventas ( $X_1$ , en miles de unidades), del precio de los perfumes ( $X_2$ , en euros por unidad) y del precio medio de sus competidores más directos ( $X_3$ , en euros por unidad), durante los últimos veinte años. Del análisis de la información recabada, obtenemos los siguientes datos:

$$\begin{aligned}\sum_i x_{1i} &= 2430 & \sum_i x_{1i}^2 &= 250418 & \sum_i x_{1i} \cdot x_{2i} &= 15055'5 & \sum_i x_{1i} \cdot x_{3i} &= 13428'4 \\ \sum_i x_{2i} &= 144'9 & & & \sum_i x_{2i}^2 &= 937'83 & \sum_i x_{2i} \cdot x_{3i} &= 835'41 \\ \sum_i x_{3i} &= 128'3 & & & & & \sum_i x_{3i}^2 &= 752'21\end{aligned}$$

$$\begin{array}{ccccc}\text{Matriz de Var-Cov} & 182'521 & 16'016 & 18'251 \\ & 16'016 & 2'625 & 2'533 \\ & 18'251 & 2'533 & 2'764\end{array}$$

A partir de estas cifras, responda a las siguientes cuestiones:

- Estime la función lineal que explica la demanda en función del precio del perfume. ¿Son los resultados razonables desde el punto de vista del grado de explicación?
- Según esta función, si aumentamos en un euro el precio del perfume, ¿cómo variará su demanda?
- ¿En qué porcentaje variaría la demanda del perfume si su precio se encareciese un 5%, pasando de 6 € a 6'30 €?
- Estime la función lineal que explica la demanda del perfume, considerando tanto su precio como el medio de sus competidores. Según esta función, ¿cómo afectaría a la demanda del perfume el encarecimiento en su precio en un euro, permaneciendo constante el precio de la competencia?
- ¿Cuál de los dos precios guardaría mayor relación lineal, en términos estrictos, con la demanda del perfume, su precio o el de los competidores?

38. Los últimos estudios sobre el comportamiento del sector del transporte ferroviario de pasajeros en un país de la Unión Europea muestran un comportamiento muy favorable de éste, que hacen ser optimistas a las empresas del mismo sobre las posibilidades de futuro. Para confirmar tal perspectiva, se le ofrece información de los últimos diez años para ese país acerca de los ingresos obtenidos (variable X1, expresada en millones de euros), del número de viajeros que utilizaron dicho transporte (variable X2, expresada en miles de personas) y de los costes por consumo eléctrico y carburantes (variable X3, expresada en millones de euros). Así mismo, también conocemos la matriz de varianzas y covarianzas C de esa distribución:

Año	X1	X2	X3
1995	610	193,98	10,45
1996	650	190,284	10,59
1997	680	194,265	11,03
1998	700	181,891	10,92
1999	780	274,35	11,41
2000	800	316,327	11,78
2001	850	358,615	11,82
2002	820	339,416	11,17
2003	840	351,53	11,44
2004	930	365,503	11,74
<b>Media</b>	<b>766</b>	<b>276,616</b>	<b>11,24</b>

$$C = \begin{pmatrix} 9324 & 6795'843 & 40'03 \\ 6795'843 & 5579'922 & 29'265 \\ 40'03 & 29'265 & 0'215 \end{pmatrix}$$

A partir de tal información, responda a las siguientes cuestiones:

- Si definimos la relación lineal que explica los ingresos en función exclusivamente del número de viajeros, ¿cuál sería el porcentaje de varianza explicada por la regresión? Desde este modelo, ¿en cuántos millones de euros variarían los ingresos si se aumenta el número de pasajeros en 20.000 en 2005 respecto del año anterior? ¿Qué fiabilidad le otorgaría a dicha predicción? Razone su respuesta.
- Considerando la misma relación lineal del apartado anterior, si se supusiese que en 2005 el número de viajeros se incrementa en un 10% respecto del año anterior, ¿en qué porcentaje variarían los ingresos? Razone su respuesta.
- Si suponemos que los ingresos obtenidos (X1) dependen linealmente del número de pasajeros (X2) y de los costes por consumo eléctrico y carburantes (X3), a través del correspondiente plano de regresión, ¿en cuánto variarían los ingresos para este modelo con respecto al inicial? (1 punto)
- Si consideramos el modelo de regresión múltiple, ¿en cuántos millones de euros variarían los ingresos en 2005 si se aumenta en 20.000 el número de viajeros respecto de 2004, suponiendo que los costes no varían?
- A partir de la distribución conjunta definida, ¿qué variable considera que ejerce una mayor influencia sobre los ingresos, el número de viajeros o los costes por consumo eléctrico y carburantes, eliminando la influencia de la tercera variable? Justifique la respuesta.

39. La Comisión de Control de un Plan de Pensiones de una determinada organización está interesada en saber qué factores influyen en la rentabilidad de su cartera de valores. Del estudio previo realizado, parece razonable pensar que una de las magnitudes que podrían influir en tal rentabilidad sería la cantidad de renta variable dentro de la composición de la cartera del Plan. Con idea de ahondar en esta idea inicial, seleccionó una muestra de 40 Planes de parecidas características al de su organización y recogió los valores de los porcentajes de rentabilidad obtenida a medio plazo (variable X1) y el porcentaje de renta variable de la cartera (variable X2), obteniéndose los siguientes resultados:

X1/X2	0-10	10-25	25-35
0	2	2	0
1'5	4	2	1
2	0	5	5
2'25	3	1	1
3	2	0	8
3'65	0	0	4

A partir de estos datos, se desea saber:

- a) ¿En cuánto variaría la rentabilidad de la cartera si aumentamos cinco puntos porcentuales el porcentaje de renta variable, supuesta una relación lineal entre ambas variables? ¿Qué fiabilidad le merece esta respuesta?

Ante los resultados obtenidos, la Comisión de Control se está planteando la posibilidad de considerar también como variable que influye en la rentabilidad, el porcentaje de renta fija de que disponen esas mismas empresas en sus respectivas carteras de valores (variable  $X_3$ ). Para plantear el correspondiente análisis, recoge la siguiente información complementaria sobre los 40 Planes de referencia:

- El porcentaje medio de renta fija es de 16'45, siendo su varianza de 178.
- Si aumentamos en diez puntos el porcentaje de renta fija, bajo el supuesto de una relación de tipo lineal simple, el de renta variable aumenta en 0'862 puntos.
- Si disminuimos el porcentaje de renta fija en un punto, bajo el supuesto de una relación de tipo lineal simple, la rentabilidad disminuye 0'057 puntos.

A partir de esta información se desea conocer:

- b) ¿En cuanto variaría la rentabilidad estimada para la cartera si el porcentaje de renta fija aumentase en 1'5 unidades, permaneciendo constante el porcentaje de renta variable? ¿Qué grado de confianza daría a esta respuesta?
- c) ¿Qué variable ejerce más influencia sobre la rentabilidad, el porcentaje de renta fija o el de renta variable? Razone su respuesta.