$$\frac{}{6} + \frac{}{10} + \frac{}{3} = \frac{}{19} \qquad \text{CALIFICACION:}$$

VENTAS.GDT

Una empresa que produce una marca de detergente líquido desea contar con un modelo para planificar su producción, estimar las necesidades de materias primas y de almacenamiento y predecir la demanda de la botella de tamaño grande de detergente. La empresa cree que las ventas de su producto (en miles de botellas), V, vienen determinadas entre otros factores por las variables siguientes:

- P: precio de venta de fábrica en euros.
- PC: precio medio de los productos sustitutivos existentes en el mercado en euros.
- GP: gastos en publicidad en miles de euros.

Además la empresa cree que sus ventas crecen en el tiempo de forma constante.

Se cuenta con datos trimestrales desde el primer trimestre de 2001 hasta el segundo trimestre de 2008 de las variables antes mencionadas.

PARTE 1 (6 puntos)

1. Con los datos disponibles, estima por MCO el siguiente modelo de regresión lineal y escribe la recta de regresión muestral:

$$(1) \quad V_t = \beta_0 + \beta_1 P_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 t + u_t \qquad t = 2001:1,\ldots,2008:2$$

$$\hat{V}_t = \underset{(14,46)}{35,48} - \underset{(3,76)}{9,74}\ P_t + \underset{(1,73)}{2,94}\ PC_t - \underset{(0,75)}{0,87}\ GP_t + \underset{(0,03)}{0,32}\ t \qquad t = 2001:1,\ldots,2008:2$$

2. ¿Para qué se incluye la variable t, tiempo, en el modelo, es decir, qué factor determinante de las ventas trata de recoger?

   Con la variable tiempo se quiere recoger el comportamiento creciente constante de las ventas, independiente del resto de las variables explicativas, que la empresa sospecha que se da.

3. Explica qué valores toma la variable tiempo en un modelo de regresión lineal e ilústralo rellenando la siguiente tabla de datos para esta variable explicativa t:

Observación	P_t	PC_t	GP_t	t
2001:1	3,85	3,80	5,50	1
2001:2	3,75	4,00	6,75	2
2001:3	3,70	4,30	7,25	3
2001:4	3,70	3,70	5,50	4
2002.1	3,60	3,85	7,00	5
2002.2	3,60	3,80	6,50	6
⋮	⋮	⋮	⋮	⋮

La variable tiempo es una variable que no tiene valores en sí misma sino que se asignan libremente: es una especie de variable ficticia. La variable tiempo trata de recoger la evolución cronológica de la variable endógena Ventas, por lo que el único requisito que se debe cumplir a la hora de asignar valores a esta variable es que sean correlativos y respeten la sucesión temporal de la serie de interés. Así, es lo mismo asignar los valores, $t = 1, 2, 3, \ldots$ o $t = 23, 24, 25, \ldots$, etc. Lo más habitual es utilizar la primera opción, pero cualquier otra serie ordenada de números es igualmente válida para recoger el efecto de la evolución temporal de la serie. La opción es la que, por defecto, utiliza Gretl al generar la variable de tendencia temporal.

4. Interpreta los coeficientes estimados que acompañan a las variables explicativas del modelo. ¿Tienen los signos esperados?

   - $\hat{\beta}_1$: el incremento estimado en el volumen de ventas (en miles de botellas) cuando el precio de venta aumenta en 1 euro manteniéndose el resto de las características fijas.
   - $\hat{\beta}_2$: el incremento estimado en el volumen de ventas (en miles de botellas) cuando el precio de venta de los productos sustitutivos aumenta en 1 euro, manteniéndose el resto de las características fijas.
   - $\hat{\beta}_3$: el incremento estimado en el volumen de ventas (en miles de botellas) cuando la empresa gasta mil euros más en publicidad manteniéndose el resto de las características fijas.
   - $\hat{\beta}_4$: el incremento estimado en el volumen de ventas (en miles de botellas) de un trimestre al siguiente, manteniéndose el resto de las características fijas. Por lo tanto, las ventas crecen en el tiempo de forma constante.

   En un modelo como el presente, se espera que las ventas de un producto disminuyan al aumentar su precio, $\hat{\beta}_1 < 0$, aumenten si el precio de los bienes sustitutivos sube, $\hat{\beta}_2 > 0$, y como la empresa cree que las ventas crecen en el tiempo, $\hat{\beta}_4 > 0$. Por lo tanto, estos tres coeficientes estimados tienen los signos esperados. Sin embargo, las empresas se gastan dinero en publicidad con el fin de aumentar las ventas por lo que se espera que $\hat{\beta}_3 > 0$, lo que no sucede y es un resultado que lleva a sospechar que algo puede estar pasando con el modelo.

5. Contrasta a un nivel de significación del 5 % si los gastos en publicidad influyen en las ventas.

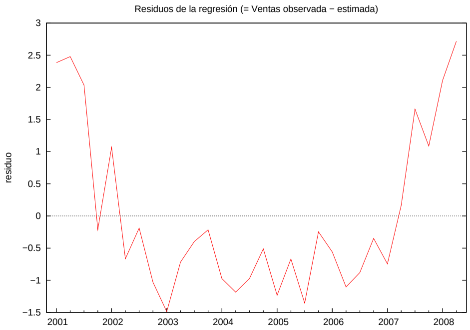
$$\begin{aligned} H_0 : &\quad \beta_3 = 0 \\ H_a : &\quad \beta_3 \neq 0 \end{aligned}$$

$$\mathrm{t} = \frac{\hat{\beta}_3 - \beta_3^0}{\hat{\sigma}_{\hat{\beta}_3}} \sim t(T - (k+1))$$

$$\mathrm{t} = \frac{-0,87 - 0}{0,75} = -1,16 \qquad |t| = |-1,16| = 1,16 < t_{0,025}(30-5) = 2,060$$

   NO se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, no se rechaza la hipótesis de que los gastos en publicidad no influyen en las ventas.

6. Representa gráficamente los residuos del modelo y coméntalos. ¿Crees que el modelo propuesto cumple los supuestos básicos del modelo de regresión lineal general?



El gráfico de los residuos muestra que éstos no oscilan en torno a cero como sería de esperar sino que tienen un patrón claro de comportamiento: primero decrecen, luego se mantienen a un nivel más o menos constante y luego crecen. Es decir, se observan rachas de residuos: al principio de la muestra todos positivos, luego todos negativos y al final vuelven a ser positivos. Por lo tanto, se puede concluir que los residuos no oscilan de forma aleatoria en torno a cero, lo que puede indicar que las perturbaciones tampoco se comportan de esta manera. Puede ser debido a que se haya omitido alguna variable relevante en el modelo, a que la forma funcional no esté bien especificada o a que las perturbaciones estén autocorrelacionadas.

PARTE 2 (10 puntos)

Teniendo en cuenta las conclusiones obtenidas en la Parte 1, el experto de la empresa estima un modelo de regresión alternativo para la demanda del limpiador:

$$(2) \quad V_t = \beta_0 + \beta_1 P_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 t + \beta_5 t^2 + u_t \qquad t = 2001:1, \ldots, 2008:2$$

1. ¿Cómo ha solucionado el experto de la empresa el problema que tenía el modelo de la Parte 1? ¿Cumple este modelo el supuesto de que el modelo de regresión sea lineal? ¿por qué?

   El experto ha solucionado el problema del modelo de la Parte 1 suponiendo que se debe a la mala especificación de la forma funcional e introduciendo no linealidad en el modelo a través del cuadrado de la variable tiempo. En este modelo alternativo la variable tiempo entra con una forma funcional cuadrática no lineal tratando de recoger ese comportamiento cuadrático no lineal que se observaba en los residuos del modelo estimado en la Parte 1.

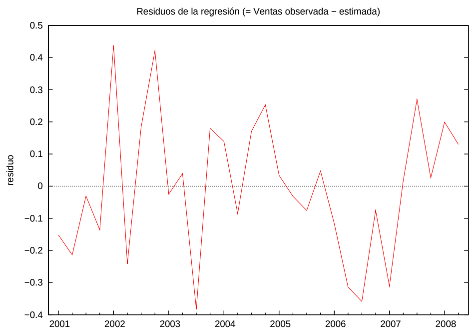
   Ahora bien, el modelo de regresión planteado sí cumple el supuesto de linealidad porque este supuesto exige linealidad en los coeficientes de la regresión (que se cumple) y no en las variables.

2. Estima el modelo por MCO. Escribe la recta de regresión muestral.

$$\hat{V}_t = \underset{(2,65)}{6,92} - \underset{(0,69)}{2,15}\ P_t + \underset{(0,30)}{1,55}\ PC_t + \underset{(0,14)}{0,55}\ GP_t - \underset{(0,023)}{0,32}\ t + \underset{(0,0007)}{0,02}\ t^2$$

3. Representa gráficamente los residuos y coméntalos.

   El gráfico de los residuos muestra que éstos oscilan en torno a cero sin un patrón claro de comportamiento. Por lo tanto, parece que el cambio en la forma funcional de la tendencia ha recogido la estructura cuadrática que se observaban en los residuos del modelo (1).



4. ¿Cuál es el efecto neto estimado de la variable tiempo en las ventas, manteniéndose el resto de las variables fijas? ¿Es constante?

$$\frac{\partial \widehat{E(V_t)}}{\partial\, t} = \hat{\beta}_4 + 2\ \hat{\beta}_5\ t = -0,32 + 2\ (0,02)\ t$$

Por lo tanto, el incremento promedio estimado en las ventas por trimestre no es constante sino que depende del momento de tiempo en que nos encontremos. Así, para t=1 (2001:1):

$$\frac{\partial \widehat{E(V_t)}}{\partial\, t} = \hat{\beta}_4 + 2\ \hat{\beta}_4\ t = -0,32 + 0,04\ t = -0,32 + 0,04 = -0,26$$

para t=30 (2008:4):

$$\frac{\partial \widehat{E(V_t)}}{\partial\, t} = \hat{\beta}_4 + 2\ \hat{\beta}_4\ t = -0,32 + 0,04\ t = -0,32 + 0,04\ \times\ 30 = 0,88$$

5. ¿Son las variables explicativas del modelo conjuntamente significativas?

$$H_0: \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$
$$H_a: \quad \exists!\ \beta_j \neq 0, \quad j = 1,2,3,4,5$$

$$F = \frac{R^2/q}{(1-R^2)/(T-(k+1))} \sim \mathcal{F}(q, T-(k+1))$$

$$F = \frac{0,9955/5}{(1-0,9955)/(30-6)} = 1106,11 > \mathcal{F}_{0{,}05}(5,24) = 2,62$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, rechazamos la hipótesis de que las variables explicativas del modelo no son conjuntamente significativas.

6. ¿Son las variables explicativas del modelo individualmente significativas?

Para las variables precio, precio de los bienes sustitutivos y gastos de publicidad (α=5 %):

$$H_0: \quad \beta_i = 0 \qquad\qquad \text{t} = \frac{\hat{\beta}_i - 0}{\hat{\sigma}_{\hat{\beta}_i}} \sim t(30-6)$$

$$H_a: \quad \beta_i \neq 0$$

$i = 1 \qquad \text{t} = -3,11 \qquad |t| = 3,11 > t_{0,025}(24) = 2,06 \Rightarrow$ la variable P es significativa

$i = 2 \qquad \text{t} = 5,196 \qquad |t| = 5,196 > t_{0,025}(24) = 2,06 \Rightarrow$ la variable PC es significativa

$i = 3 \qquad \text{t} = 4,05 \qquad |t| = 4,05 > t_{0,025}(24) = 2,06 \Rightarrow$ la variable GP es significativa

Para la variable tiempo ($\alpha = 5\,\%$):

$$H_0: \quad \beta_4 = \beta_5 = 0$$
$$H_a: \quad \beta_4 \neq 0 \,\text{y/o}\, \beta_5 \neq 0$$

El estadístico de contraste es:

$$F = \frac{(SCR_R - SCR_{NR})/2}{SCR_{NR}/(30-6)} \sim \mathcal{F}(2, 30-6)$$

donde

- Modelo No restringido: Modelo (2).
- Modelo Restringido:

$$V_t = \beta_0 + \beta_1\, P_t + \beta_2\, PC_t + \beta_3\, GP_t + u_t$$

Una vez estimados ambos modelos con la muestra disponible, se obtienen las correspondientes sumas de cuadrados de los residuos y se puede computar el estadístico F muestral.

$$F = \frac{(266,81 - 1,3343)/2}{1,3343/(30-6)} = 2387,45 \qquad \Rightarrow \qquad F = 2387 > \mathcal{F}_{0,05}(2,24) = 3,40$$

Por lo que se concluye que el tiempo (la tendencia cuadrática) es una variable explicativa significativa.

7. Contrasta si la tendencia adecuada para el modelo de ventas es lineal o cuadrática.

$$H_0: \quad \beta_5 = 0$$
$$H_a: \quad \beta_3 \neq 0$$

$$\text{t} = \frac{\hat{\beta}_5 - 0}{\hat{\sigma}_{\hat{\beta}_5}} \sim t(30-6)$$

$$\text{t} = \frac{0,02 - 0}{0,0007} = 28,57 \qquad |t| = 28,57 > t_{0,025}(30-6) = 2,06$$

Se rechaza la hipótesis nula con un nivel de significación del $5\,\%$. Por lo tanto, la tendencia adecuada es la cuadrática.

8. Contrasta si los gastos de publicidad influyen positivamente en la demanda del producto considerado.

$$H_0: \quad \beta_3 = 0$$
$$H_a: \quad \beta_3 > 0$$

$$\mathrm{t} = \frac{\hat{\beta}_3 - \beta_3^0}{\hat{\sigma}_{\hat{\beta}_3}} \sim t(30-6)$$

$$\mathrm{t} = \frac{0,55-0}{0,136} = 4,05 \qquad t = 4,05 > t_{0,05}(30-6) = 1,711$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, se rechaza la hipótesis de que los gastos en publicidad no influyen positivamente en las ventas.

9. Comprueba la sospecha del experto de la empresa, de que la empresa podría contrarrestar una política de precios de la competencia consistente en bajar los precios en 1 euro, respondiendo con la misma política, es decir, bajando a su vez el precio de venta en 1 euro.

$$H_0: \quad \beta_2 + \beta_3 = 0$$
$$H_a: \quad \beta_2 + \beta_3 \neq 0$$

$$\mathrm{t} = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)^0}{\hat{\sigma}_{\hat{\beta}_2 + \hat{\beta}_3}} \sim t(T-(k+1))$$

dado que:

$$E[\hat{\beta}_2 + \hat{\beta}_3] = \beta_2 + \beta_3$$

y

$$\hat{V}(\hat{\beta}_2 + \hat{\beta}_3) = \hat{V}(\hat{\beta}_2) + \hat{V}(\hat{\beta}_3) + 2\,\widehat{cov}(\hat{\beta}_2 + \hat{\beta}_3) = 0,48 + 0,09 + 2 \times (-0,11) = 0,35$$

$$\mathrm{t} = \frac{(-2,15+1,55)-0}{\sqrt{0,35}} = -1,01 \qquad |t| = |-1,01| = 1,01 < t_{0,025}(30-6) = 2,06$$

No se rechaza la hipótesis nula con un nivel de significación del 5 %. Es decir, no se rechaza la hipótesis de que la empresa puede compensar el efecto sobre sus ventas de las políticas de precios de la competencia aplicando las mismas políticas de precios.

10. Predice por punto las ventas futuras del producto para cada uno de los trimestres del año 2009 bajo el supuesto de que los valores de las variables explicativas son los siguientes:

Año 2009	P_t	PC_t	GP_t
Trimestre 1	3,65	4,10	7,0
Trimestre 2	3,60	4,10	6,8
Trimestre 3	3,50	4,10	7,0
Trimestre 4	3,55	4,10	7,2

¿Cual sería la predicción para el total de las ventas del año 2009?

Las predicciones por punto óptimas se obtienen utilizando la recta de regresión muestral:

$$\hat{V}_{09:i} = 6,92 - 2,15\,P_t + 1,55\,PC_t + 0,55\,GP_t - 0,32\,t + 0,02\,t^2$$

$$\begin{aligned}
\hat{V}_{09:1} &= 6,92 - 2,15\,(3,65) + 1,55\,(4,10) + 0,55\,(7,0) - 0,32\,(33) + 0,02\,(33^2) = 20,50 \\
\hat{V}_{09:2} &= 6,92 - 2,15\,(3,60) + 1,55\,(4,10) + 0,55\,(6,8) - 0,32\,(34) + 0,02\,(34^2) = 21,52 \\
\hat{V}_{09:3} &= 6,92 - 2,15\,(3,50) + 1,55\,(4,10) + 0,55\,(7,0) - 0,32\,(35) + 0,02\,(35^2) = 22,90 \\
\hat{V}_{09:4} &= 6,92 - 2,15\,(3,55) + 1,55\,(4,10) + 0,55\,(7,2) - 0,32\,(36) + 0,02\,(36^2) = 24,00
\end{aligned}$$

La predicción por punto óptima para las ventas anuales viene dada por la suma de las predicciones óptimas para los cuatro trimestres del año 2009:

$$\hat{V}_{2009} = \hat{V}_{09:1} + \hat{V}_{09:2} + \hat{V}_{09:3} + \hat{V}_{09:4} = 88,92$$

PARTE 3 (3 puntos)

Por último, el experto de la empresa se plantea, que dado que los datos son trimestrales podría existir comportamiento estacional en las ventas.

1. Especifica y estima un modelo de regresión lineal para las ventas que incluya también el componente estacional.

   El comportamiento estacional de la variable endógena es un efecto cualitativo que supone que cada trimestre las ventas tienen un comportamiento diferente. Esta variable se incluye en el modelo utilizando variables artificiales o variables ficticias. Se define una variable ficticia por cada trimestre, para poder diferenciar el comportamiento cada uno de ellos:

$$T_t^i = \begin{cases} 1 & \text{si} \quad t \quad \epsilon \quad \text{trimestre i-ésimo} \\ 0 & \text{en otro caso} \end{cases} \qquad\qquad i = 1, 2, 3, 4$$

   La variable explicativa estacionalidad se incluye en el modelo a través de (4-1) variables ficticias para evitar problemas de colinealidad perfecta en la matriz X. De esta forma, tomamos un trimestre como referencia de comportamiento y analizamos lo que ocurre con los otros tres en relación a él.

   Si especificamos el modelo dejando fuera el cuarto trimestre que queda como referencia, se tiene:

$$(3) \quad V_t = \beta_0 + \gamma_1 T_t^1 + \gamma_2 T_t^2 + \gamma_3 T_t^3 + \beta_1 P_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 t + \beta_5 t^2 + u_t$$

   La recta de regresión muestral es:

$$\begin{aligned} \hat{V}_t \;=\; & \underset{(2,67)}{6,49} - \underset{(0,12)}{0,098}\, T_t^1 - \underset{(0,13)}{0,235}\, T_t^2 - \underset{(0,14)}{0,17}\; T_t^3 - \underset{(0,69)}{2,17}\; P_t + \underset{(0,31)}{1,67}\; PC_t + \\ + \; & \underset{(0,14)}{0,58}\; GP_t - \underset{(0,02)}{0,33}\; t + \underset{(0,0007)}{0,02}\; t^2 \end{aligned}$$

2. Interpreta los coeficientes del modelo referidos al efecto estacional.

   - β_0 : el valor esperado de las ventas (en miles de botellas) para el cuarto trimestre cuando las variables precio, precio del consumidor, gastos en publicidad y tiempo toman el valor cero.
   - γ_1 : la diferencia esperada en las ventas (en miles de botellas) entre el primer y el cuarto trimestre cuando el resto de las variables (precio, precio bienes sustitutivos, publicidad y tiempo) se mantienen fijas.
   - γ_2 : la diferencia esperada en las ventas (en miles de botellas) entre el segundo y el cuarto trimestre cuando el resto de las variables se mantienen fijas.
   - γ_3 : la diferencia esperada en las ventas (en miles de botellas) entre el tercer y el cuarto trimestre cuando el resto de las variables se mantienen fijas.

3. Contrasta si la estacionalidad es una variable explicativa significativa, explicando detalladamente el mecanismo de contraste (hipótesis nula, estadístico, distribución, regla de decisión).

   Para contrastar la hipótesis nula de que no existe estacionalidad en las ventas de detergente, se realiza el siguiente contraste:

$$\begin{aligned} H_0 :& \quad \gamma_1 = \gamma_2 = \gamma_3 = 0 \quad \text{(no hay comportamiento estacional)} \\ H_a :& \quad \exists!\; \gamma_j \neq 0, \quad j = 1, 2, 3 \end{aligned}$$

El estadístico de contraste es:

$$F = \frac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/(T-(k+1))} \sim \mathcal{F}(q, T-(k+1))$$

donde el modelo no restringido es el modelo (3) y el modelo restringido es el modelo (2).

Una vez estimados ambos modelos con la muestra disponible, se obtienen las correspondientes sumas de cuadrados de los residuos y se puede computar el estadístico F muestral.

$$F = \frac{(1,3343 - 1,1524)/3}{1,1524/(30-9)} = 1,11$$

Regla de decisión:

$$F = 1,11 < \mathcal{F}_{0,05}(3, 30-9) = 3,07$$

No se rechaza la hipótesis nula con un nivel de significación del 5 %, es decir, no se rechaza la hipótesis de que la variable ventas no tiene comportamiento estacional.