$$\frac{\quad}{6} + \frac{\quad}{5} = \frac{\quad}{11}$$ CALIFICACION:

**PARTE 1** (6 puntos)

Una empresa inmobiliaria desea conocer los determinantes del precio de la vivienda en una ciudad de tamaño medio. Para ello recoge información sobre las siguientes variables de las últimas 1000 ventas de casas realizadas: precio de venta en miles de euros ($P$), tamaño en metros cuadrados útiles ($T$), antigüedad en años desde que fue construida ($A$), situación ($S_i = 1$ si la vivienda i-ésima está en el centro), garaje ($G_i = 1$ si la vivienda í-ésima tiene garage). En la siguiente tabla se muestran algunos de estos datos.

| Precio | Tamaño | Antigüedad | Situación | $S_i$ | Garage | $G_i$ |
|---|---|---|---|---|---|---|
| 205,45 | 217,94 | 6 | no | | no | |
| 197,24 | 230,39 | 10 | no | | sí | |
| 134,31 | 190,16 | 17 | no | | no | |
| 198,65 | 209,67 | 7 | no | | sí | |
| 337,44 | 270,21 | 2 | sí | | no | |
| 224,90 | 190,53 | 31 | sí | | no | |
| 287,33 | 218, 56 | 12 | sí | | sí | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Se especifica el siguiente modelo de regresión para el precio de las viviendas:

$$(1) \qquad P_i = \beta_0 + \beta_1 T_i + \beta_2 A_i + \beta_3 S_i + \beta_4 G_i + u_i \qquad i = 1, 2, \ldots, 1000$$

donde:

$$S_i = \Bigg\{ \qquad\qquad G_i = \Bigg\{$$

1. Rellena las definiciones de las variables $S_i$ y $G_i$ y completa la tabla de datos anterior con los valores para las variables $S_i$ y $G_i$.

$$S_i = \begin{cases} 1 & \text{si la vivienda esta en el centro} \\ 0 & \text{en otro caso} \end{cases} \qquad G_i = \begin{cases} 1 & \text{si la vivienda tiene garage} \\ 0 & \text{en otro caso} \end{cases}$$

| Precio | Tamaño | Antigüedad | Situación | $S_i$ | Garage | $G_i$ |
|---|---|---|---|---|---|---|
| 205,45 | 217,94 | 6 | no | 0 | no | 0 |
| 197,24 | 230,39 | 10 | no | 0 | sí | 1 |
| 134,31 | 190,16 | 17 | no | 0 | no | 0 |
| 198,65 | 209,67 | 7 | no | 0 | sí | 1 |
| 337,44 | 270,21 | 2 | sí | 1 | no | 0 |
| 224,90 | 190,53 | 31 | sí | 1 | no | 0 |
| 287,33 | 218, 56 | 12 | sí | 1 | sí | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

2. Interpreta los coeficientes $\beta_2$ y $\beta_4$ del modelo (1).

   $\beta_2$ : variación en el precio de venta de una vivienda por cada año de antigüedad adicional, manteniéndose fijas el resto de las variables explicativas (tamaño, situación, garage).

   $\beta_4$ : diferencia esperada en el precio de venta de una vivienda con garage, respecto a viviendas sin garage, manteniéndose fijas el resto de las variables explicativas (tamaño, antigüedad, situación).

3. Los resultados de la estimación del modelo (1) con las 1000 observaciones son

$$\underset{(desv.)}{\hat{P}_i} = \underset{(4,29)}{7,03} + \underset{(0,018)}{0,90}\ T_i - \underset{(0,05)}{0,19}\ A_i + \underset{(0,97)}{60,19}\ S_i + \underset{(1,20)}{4,28}\ G_i \qquad t = 1, 2, ..., 1000.$$

$$SCR = 234216,5 \quad \sum (P_i - \bar{P})^2 = 1778409,26$$

   ¿Cuánto está dispuesto a pagar un comprador por un metro cuadrado adicional de superficie, manteniéndose el resto de las variables constantes?

   El coeficiente estimado para la variable $T_i$ ($\hat{\beta}_1 = 0,90$) recoge este efecto, es decir, indica el incremento estimado en el precio de venta de una vivienda por cada metro cuadrado adicional, manteniéndose fijas el resto de las variables explicativas.

   Dado que la variable precio está medida en miles de euros, se estima que, ceteris paribus, un comprador está dispuesto a pagar 900 euros por cada metro cuadrado adicional.

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

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

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

$$F = \frac{0,8683/4}{(1 - 0,8683)/(1000 - 5)} = 1640,01 > \mathcal{F}_{0,05}(4, 995) = 2,37$$

   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.

5. ¿Qué valor toma el coeficiente de determinación para la estimación del modelo (1)? Interpreta su significado.

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{234216,5}{1778409,26} = 0,8683$$

   El 86,83 % de la variabilidad muestral de los precios de venta de las viviendas viene explicada por las variaciones de las cuatro variables explicativas del modelo de regresión (tamaño, antigúedad, situación y garage) en términos lineales.

6. Contrasta la hipótesis de que, ceteris paribus, las casas del centro de la ciudad valen más que las de las afueras.

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

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

$$\mathrm{t} = \frac{60,19 - 0}{0,97} = 62,05 > t_{0,05}(995) = 1,645$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, las casas del centro valen más que las de las afueras.

**PARTE 2** (5 puntos)

Algunos expertos en el mercado inmobiliario sostienen la hipótesis de que el valor del metro cuadrado adicional va a ser superior en el centro de la ciudad que en los barrios periféricos. Para recoger este efecto se plantea el siguiente modelo:

$$(2) \qquad P_i = \beta_0 + \beta_1\, T_i + \beta_2\, A_i + \beta_3\, S_i + \beta_4\, G_i + \beta_5\,(S_i\, \times\, T_i) + u_i \qquad i = 1, 2, \ldots, 1000$$

1. Explica la especificación de este modelo y por qué recoge el efecto que buscamos: diferenciar el valor del metro cuadrado adicional por zonas de la ciudad. ¿Cuál es el cambio esperado en el precio de una vivienda por cada metro cuadrado adicional?

   En el modelo (1) el coeficiente $\beta_1$ recoge el efecto neto de un $m^2$ adicional en el precio de venta de la vivienda. El valor de este $m^2$ adicional era el mismo independientemente de la situación de la vivienda, de si tiene garage o no y de la antigüedad de la vivienda. El modelo (1) es un modelo lineal en todas las variables, por lo tanto, los efectos netos de incrementos unitarios en las variables explicativas sobre la variable precio son constantes.

   El modelo (2) propone un modelo para la determinación del precio que tiene en cuenta un factor explicativo más: la diferencia de precio del metro cuadrado adicional en las dos zonas de la ciudad consideradas. Se quiere especificar un modelo en el que el efecto del incremento de 1 $m^2$ de superficie sobre el precio de la vivienda ya no sea independiente de la situación de la vivienda, sino que tome un valor diferente para las viviendas del centro y otro valor para las de las afueras. Es decir, el efecto neto de las dos variables, Tamaño y Situación, no se ha de medir de forma independiente, sino que el efecto de una variable debe depender del valor de la otra. Para introducir este efecto no lineal de interacción entre las dos variables definimos la siguiente variable artifical:

$$S_i\, \times\, T_i = \begin{cases} T_i & \text{si la vivienda esta en el centro} \\ 0 & \text{en otro caso} \end{cases}$$

   Ahora, el modelo (2) proporciona la siguiente información sobre el precio esperado de las casas:

$$\begin{array}{lrcl} \text{Centro} & E[P_i|T_i, A_i, G_i, S_i = 1] & = & (\beta_0 + \beta_3) + (\beta_1 + \beta_5)\, T_i + \beta_2\, A_i + \beta_4\, G_i \\ \text{Afueras} & E[P_i|T_i, A_i, G_i, S_i = 0] & = & \beta_0 + \beta_1\, T_i + \beta_2\, A_i + \beta_4\, G_i \end{array}$$

   Es decir, la situación de la vivienda además de influir en el precio esperado a través del término independiente, influye también en el efecto del incremento unitario de superficie de la vivienda sobre el precio.

   El cambio esperado en el precio de una vivienda por cada metro cuadrado adicional viene dado:

$$\frac{\partial E(P_i)}{\partial T_i} = \beta_1 + \beta_5\, S_i \quad \Rightarrow \quad \begin{cases} \beta_1 & \text{si la vivienda esta en las afueras} \\ \beta_1 + \beta_5 & \text{si la vivienda esta en el centro} \end{cases}$$

2. Los resultados de la estimación del modelo (2) son los siguientes:

$$\hat{P}_i = 24,14 + 0,82\,T_i - 0,19\,A_i + 28,37\,S_i + 4,29\,G_i + 0,14\,(S_i \times T_i) \qquad i = 1, 2, \ldots, 1000$$

`Matriz de covarianzas de los estimadores`

| | C | Tamaño | Antigüedad | Situación | Garage | Sit $\times$ Tam |
|---|---|---|---|---|---|---|
| C | 38,36 | -0,16 | -0,027 | -37,96 | -0,27 | 0,16 |
| Tamaño | | 0,0007 | 0,0000073 | 0,16 | 0,000029 | -0,00069 |
| Antigüedad | | | 0,0026 | -0,0056 | -0,0017 | 0,00003 |
| Situación | | | | 70,78 | -0,048 | -0,298 |
| Garage | | | | | 1,432 | 0,0001 |
| Sit $\times$ Tam | | | | | | 0,0012 |

¿Tienen los coeficientes estimados el signo esperado?

En principio, sí. Es de esperar que un incremento de un $m^2$ en la superficie de la vivienda tenga un efecto neto positivo en el precio de venta (0,82 para las viviendas de las afueras y (0,82 + 0,14) para las viviendas del centro), mientras que un año adicional de antigüedad tenga un efecto neto negativo (-0,19). Por otro lado, es de esperar que las viviendas con garage valgan más que las que no tienen garage (manteniéndose el resto de las variables fijas)

3. Contrasta la hipótesis planteada por los expertos inmobiliarios.

$$H_0: \quad \beta_5 = 0 \qquad\qquad H_a: \beta_5 > 0$$

$$\mathrm{t} = \frac{\hat{\beta}_5 - \beta_5^0}{\hat{\sigma}_{\hat{\beta}_5}} \sim t(T - (k+1))$$

$$\mathrm{t} = \frac{0,14 - 0}{0,04} = 3,81 > t_{0,05}(994) = 1,645$$

Se rechaza la hipótesis nula con un nivel de significación del 5 %. Por lo tanto, el metro cuadrado adicional vale más en el centro que en las afueras.

4. Predice por punto el precio de venta de una casa de 226 $m^2$, de reciente construcción (3 años), situada en las afueras y con garage. La inmobiliaria ha fijado un precio de 300000 euros por esta casa. Según tus resultados, ¿crees que este precio responde a las características de la vivienda?

Una vez estimado el modelo de regresión lineal propuesto para la determinación de los precios de las viviendas, lo podemos utilizar para predecir:

$$\begin{aligned} \hat{P}_0 &= 24,14 + 0,82\,T_0 - 0,19\,A_0 + 28,37\,S_0 + 4,29\,G_0 + 0,14\,(S_0 \times T_0) \\ &= 24,14 + 0,82\,226 - 0,19\,3 + 4,29\,1 = 213,18 \text{ mil euros} \end{aligned}$$

Según el modelo, se estima que una casa de esas características vale 213180 euros.

Construimos un intervalo de confianza del 95 % para nuestra predicción: si 300000 euros está dentro del intervalo este precio respondería a las características de la vivienda.

$$\text{Intervalo de predicción:} \qquad \left[\hat{P}_0 \pm t_{0,025}(T - (k+1))\,\sqrt{\hat{V}(\hat{u}_0)}\right]$$

$$
\begin{aligned}
\hat{u}_0 &= P_0 - \hat{P}_0 = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)\, T_0 + (\beta_2 - \hat{\beta}_2)\, A_0 + (\beta_3 - \hat{\beta}_3)\, S_0 + (\beta_4 - \hat{\beta}_4)\, G_0 \\
&+ (\beta_5 - \hat{\beta}_5)\, (S_0 T_0) + u_0 = (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)\, T_0 + (\beta_2 - \hat{\beta}_2)\, A_0 + (\beta_4 - \hat{\beta}_4)\, G_0 + u_0 \\
\hat{V}(\hat{u}_0) &= \hat{V}(\hat{\beta}_0) + T_0^2\, \hat{V}(\hat{\beta}_1) + A_0^2\, \hat{V}(\hat{\beta}_2) + G_0^2\, \hat{V}(\hat{\beta}_4) + 2T_0 Cov(\hat{\beta}_0\, \hat{\beta}_1) + 2A_0 Cov(\hat{\beta}_0\, \hat{\beta}_2) \\
&+ 2G_0 Cov(\hat{\beta}_0\, \hat{\beta}_3) + 2T_0 A_0 Cov(\hat{\beta}_1\, \hat{\beta}_2) + 2T_0 G_0 Cov(\hat{\beta}_1\, \hat{\beta}_4) + 2G_0 A_0 Cov(\hat{\beta}_2\, \hat{\beta}_4) + \hat{\sigma}_u^2 = \\
&= 38,36 + 226^2\, (0,00069) + 3^2\, (0,0026) + 1^2\, (1,432) + (2)\, (226)\, (-0,1609) + \\
&+ (2)\, (3)\, (-0,0268) + (2)\, (-0,2698) + (2)\, (226)\, (3)\, (0,0000073) + (2)\, (226)\, (0,000029) + \\
&+ (2)\, (3)\, (-0,00173) + 232,25 = 0,1663 + 232,25 = 232,41
\end{aligned}
$$

$$
\left[\hat{P}_0 \pm t_{0,025}(T-(k+1))\, \sqrt{\hat{V}(\hat{u}_0)}\right] = \left[213,18 \pm 1,96\, \times\, \sqrt{232,41}\right] = [183,30 \quad 243,06]
$$

Como se puede observar el precio pedido por la inmobiliaria, 300 mil euros, no pertenece al intervalo de predicción, por lo tanto, según nuestro modelo, el precio es muy alto.

5. Dados los resultados obtenidos en la estimación MCO del modelo (2) y en los contrastes realizados, ¿qué propiedades tienen los estimadores MCO del modelo (1)? ¿Son válidos los contrastes sobre los coeficientes que realizaste en el modelo (1)? ¿por qué?

La diferencia que existe entre el modelo (1) y el modelo (2) es que este último incluye un factor explicativo más en el modelo que permite que el valor del metro cuadrado adicional sea diferente según las zonas de la ciudad. Dado el resultado del contraste realizado en el apartado 3 en el que se rechaza la hipótesis de que este efecto no sea significativo, podemos concluir que el modelo (1) está mal especificado, en el sentido, de que se ha omitido un factor relevante, más específicamente, la forma funcional no estaba bien especificada.

Cuando se omite una variable relevante en el modelo o la forma funcional no es la adecuada, ya no se puede sostener que se cumpla el supuesto del modelo de regresión lineal general: $E[u] = 0$. Los estimadores MCO del modelo (1) van a ser, en general, sesgados y no van a ser ELIO, es decir, no son los estimadores lineales insesgados de varianza mínima.

Los contrastes realizados sobre el modelo (1) se hicieron bajo el supuesto de que este modelo cumplía todos los supuestos del modelo lineal general. Como esto no es así, ya que no se cumple que $E[u] = 0$, los estimadores MCO son sesgados, no están centrados en el verdadero valor y su varianza no vendría dada por la fórmula habitual MCO. Como los estimadores MCO no siguen las distribuciones habituales en las que se basan las distribuciones de los estadísticos que utilizamos para contrastar hipótesis nulas sobre la significatividad de las variables explicativas del modelo, estas distribuciones ya no son válidas y, por lo tanto, los contrastes realizados sobre el modelo (1) no son válidos y las conclusiones que se obtienen pueden ser erróneas.