

En una encuesta realizada a cuarenta familias se ha tratado de cuantificar la incidencia de la renta anual disponible en millones de pesos ( $X_1$ ), de las variables ficticias, Urbano: variable que toma el valor de 1 si la familia vive en una zona urbana, y 0 si vive en una zona rural ( $X_2$ ), ESTMED: variable que toma el valor de 1 si el sustentador principal ha alcanzado el nivel de estudios medios, y 0 si no ha alcanzado dicho nivel o ha alcanzado el nivel superior ( $X_3$ ), ESTSUP: variable que toma el valor de 1 si el sustentador principal ha alcanzado estudios superiores, y 0 si no ha alcanzado ( $X_4$ ), y de las variables Proporción: proporción, expresada en tanto por uno, de miembros del hogar con menos de 5 años ( $X_5$ ), TAMHOG: tamaño del hogar, es decir, número de miembros que hay en el hogar ( $X_6$ ), sobre el gasto anual en productos lácteos en millones de pesos ( $Y$ ). Para evitar distorsiones debidas al diferente tamaño de los hogares tanto el consumo como la renta se han expresado en términos per cápita.

- 1) Estime la siguiente ecuación de regresión lineal múltiple,

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon_i$$

e interprete los parámetros estimados  $\beta_1$  y  $\beta_2$ .

- 2) A partir de la matriz de correlación de todas las variables explicativas, i) ¿qué se puede decir sobre la multicolinealidad? ii) Ahora calcule el determinante de la matriz de correlaciones, ¿qué se puede interpretar?
- 3) Calcule el VIF para cada variable regresora y determine en cada caso si la variable  $X_i$ ,  $i = 1, 2, 3, 4$  está altamente correlacionada con el resto de las  $X$ .
- 4) Halle e interprete un intervalo de confianza del 95% para  $\beta_1$ .
- 5) Halle e interprete un intervalo de confianza del 95% para  $\beta_4$ .
- 6) Contrastar la significancia individual de todas las variables explicativas. ¿Está de acuerdo con la especificación del modelo? ¿Consideraría eliminar alguna de las variables? Si es así, ¿cuáles?
- 7) Contrastar la significancia global del modelo, considerando un alpha de 0.05. Interprete.
- 8) Realice el siguiente contraste  $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_6 = 0$ , considerando un alpha de 0.05. Interprete.

- 9) Si tiene una o más variables explicativas no significativas, elimine la variable con mayor PROB y estime de nuevo el modelo. Contraste de nuevo la significancia individual de las variables explicativas. Si de nuevo tiene una o más variables explicativas no significativas, elimine la variable con mayor PROB y estime de nuevo el modelo. Así sucesivamente hasta que en el modelo le queden solamente las variables que sean significativas. ¿Cómo es el nuevo coeficiente de determinación ajustado comparado con el del numeral (1)? Explique.
- 10) Utilizando el modelo estimado en (9), encuentre un intervalo de confianza del 94% para el pronóstico del gasto anual promedio esperado en productos lácteos, cuando una familia tiene una renta anual disponible de 2745 millones de pesos y la proporción de miembros del hogar con menos de 5 años es de 0.2. También encuentre un intervalo de confianza del 95% para el pronóstico del gasto anual promedio esperado en productos lácteos, cuando una familia tiene una renta anual disponible de 1380 millones de pesos y la proporción de miembros del hogar con menos de 5 años es de 0.33. Interprete sus resultados.
- 11) Verifique si se cumplen los supuestos del modelo de regresión estimado en el literal (9). ¿Es adecuado el modelo de regresión lineal múltiple?
- 12) Para el modelo estimado en (9), obtenga los residuales estandarizados. ¿Se presentan residuales estandarizados en valor absoluto mayores que 2.5?, si es así, defina una variable dummy por cada valor atípico y estime de nuevo el modelo incorporando la(s) variable(s) dummy. Verifique si se cumplen los supuestos de dicho modelo.
- 13) Estime los modelos **LOG-LOG, LOG-LIN y LIN-LOG**, con las variables regresoras significativas que le quedaron en el numeral (9) y la(s) variable(s) dummy que encontró en el numeral (12). Verifique si se cumplen los supuestos de cada uno de dichos modelos de regresión.
- 14) Estime el siguiente modelo de regresión lineal múltiple,  $(LnY)$ ,  $(LnX_1)$ ,  $(X_5^2)$  y la o las dummy que encontró. Es decir, estime el siguiente modelo,

$$Ln(Y_i) = \beta_0 + \beta_1 Ln(X_1) + \beta_2 X_5^2 + \dots + \varepsilon_i \quad (*)$$

En los puntos suspensivos de la ecuación (\*) van la o las dummy que encontró. Valide si se cumplen los supuestos de dicho modelo.

- 15)** Estime el modelo **Raíz cuadrada de Y, contra el cuadrado de las variables regresoras significativas que le quedaron en el numeral (9)** y la(s) variable(s) dummy que encontró en el numeral (12). Valide los supuestos.
- 16)** Entre los modelos que validan todos los supuestos, seleccione el mejor. Usando el mejor modelo, encuentre un intervalo de confianza del 94% para el pronóstico del gasto anual promedio esperado en productos lácteos, cuando una familia tiene una renta anual disponible de 2745 millones de pesos y la proporción de miembros del hogar con menos de 5 años es de 0.2. También encuentre un intervalo de confianza del 94% para el pronóstico del gasto anual promedio esperado en productos lácteos, cuando una familia tiene una renta anual disponible de 1380 millones de pesos y la proporción de miembros del hogar con menos de 5 años es de 0.33. Interprete sus resultados. *Compare con los resultados obtenidos en el numeral (10).*