Regresión 2015.

Práctica de introducción a regresión.

Con el fin de estudiar la participación en la fuerza laboral (variable dependiente) de las familias urbanas pobres, se obtuvo una muestra de segmentos censales de un censo de población. Los datos se dan a continuación.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Area | % en fuerza laboral | Ingreso familiar medio | Tamaño familiar medio | Tasa de desempleo |
| 137 | 64.3 | 1998 | 2.95 | 4.4 |
| 139 | 45.4 | 1114 | 3.4 | 3.4 |
| 141 | 26.6 | 1942 | 3.72 | 1.1 |
| 142 | 87.5 | 1998 | 4.43 | 3.1 |
| 143 | 71.3 | 2026 | 3.82 | 7.7 |
| 145 | 82.4 | 1853 | 3.9 | 5 |
| 147 | 26.3 | 1666 | 3.32 | 6.2 |
| 149 | 61.6 | 1434 | 3.8 | 5.4 |
| 151 | 52.9 | 1513 | 3.49 | 12.2 |
| 153 | 64.7 | 2008 | 3.85 | 4.8 |
| 155 | 64.9 | 1704 | 4.69 | 2.9 |
| 157 | 70.5 | 1525 | 3.89 | 4.8 |
| 159 | 87.2 | 1842 | 3.53 | 3.9 |
| 161 | 81.2 | 1735 | 4.96 | 7.2 |
| 163 | 67.9 | 1639 | 3.68 | 3.6 |

Notas: 1. Área es el identificador del segmento censal

2. El ingreso familiar medio está medido en dólares corrientes.

3. Tasa de desempleo está medida en puntos porcentuales.

Si el objetivo es plantear un modelo en el que se trate de predecir el porcentaje de participación en el mercado laboral, conteste las siguientes preguntas:

1. Usando las operaciones matriciales, encuentre los estimadores mínimo-cuadráticos para los coeficientes de regresión.
2. Utilizando el comando de lm, estime el modelo de regresión.
3. Interprete todos los coeficientes (beta0, beta1, beta2 y beta3)
4. Interprete el coeficiente de determinación múltiple de la “salida” del lm
5. Estime una tabla de Análisis de Variancia con el modelo, y calcule el coeficiente de determinación múltiple a partir de los resultados de dicha tabla.
6. Calcule los coeficientes de determinación parcial de cada variable condicional a que las otras dos variables ya fueron incorporadas.
7. Con operaciones matriciales, calcule la matriz de variancia-covariancia de los coeficientes de regresión.
8. Calcule el valor ajustado para el área No. 163.
9. Usando las funciones ya programadas de R, calcule un intervalo de confianza al 95% para el porcentaje de participación laboral promedio de todas aquellas áreas que compartan las características del área 163.
10. Usando las funciones ya programadas de R, calcule el intervalo de confianza al 95% para el porcentaje de participación predicho para el área 163.
11. Usando operaciones matriciales, calcule el intervalo de confianza al 95% para el porcentaje solicitado en el punto anterior.
12. Reproduzca con funciones de R el valor t y el p-value para el contraste de hipótesis de que el coeficiente del tamaño familiar medio (Beta2) sea igual a 0.
13. Haga la prueba de hipótesis de que los coeficientes de regresión del tamaño familiar medio y de la tasa de desempleo (beta2 y beta3) son iguales a cero simultáneamente.
14. Calcule los terciles a la variable ingreso familiar medio. Con base en dichos terciles, cree una variable categórica que sea igual a 1 si el ingreso es menor al primer tercil, que sea igual a 2 si el ingreso está entre el primer y segundo tercil, y que sea igual a 3 si el ingreso es mayor al segundo tercil. Convierta esta variable a tipo factor y llámele tercilesingreso. Para efectos de interpretación, llame a las categorías 1: Ingreso muy bajo, 2: Ingreso bajo, 3: Ingreso moderado.
15. Estime con el comando lm un modelo de regresión en el que las variables independientes sean el tamaño familiar medio, la tasa de desempleo y la variable tercilesingreso (pista: automáticamente R generará variables dummy).
16. Interprete los coeficientes de regresión para las variables dummy generadas automáticamente al estimar un modelo con tercilesingreso.
17. Trate de averiguar cómo puede en R cambiar la categoría de referencia y estime el mismo modelo del punto anterior, pero poniendo de categoría de referencia a la categoría 2 (ingreso bajo). Interprete los coeficientes para las dumies generadas con tercilesingreso.
18. Estime con lm un modelo en el que prediga la participación en la fuerza laboral en función del tamaño familiar medio, la tasa de desempleo, los terciles de ingreso, y las interacciones de los terciles de ingreso con el tamaño familiar medio.
19. Independientemente de si las interacciones son significativas, escriba las 3 ecuaciones implícitas (o sea, las ecuaciones para cada tercil de ingreso), implícitas en el modelo del punto anterior.