

Clase: Regresión Múltiple

Orlando Joaqui Barandica
orlando.joaqui@javerianacali.edu.co

Pontificia Universidad Javeriana Cali

2023

Tabla de Contenido

1 Variables independientes cualitativas

2 Modelos con interacción

Variables independientes cualitativas

Para utilizar una variable cualitativa (Ej: Género: Masculino, Femenino) en el análisis de regresión, se emplea un esquema de variables ficticias. En el cual cada una de las dos condiciones posibles se codifica con un 0 o un 1.

Variable ficticia: variable en la que solo existen dos resultados posibles. Para el análisis uno de los resultados se codifica con un 1, y el otro con un 0.

Variables independientes cualitativas

Para utilizar una variable cualitativa (Ej: Género: Masculino, Femenino) en el análisis de regresión, se emplea un esquema de variables ficticias. En el cual cada una de las dos condiciones posibles se codifica con un 0 o un 1.

Variable ficticia: variable en la que solo existen dos resultados posibles. Para el análisis uno de los resultados se codifica con un 1, y el otro con un 0.

- Ejemplo: Se quiere estimar el salario de un ejecutivo con base en los años de su experiencia laboral y si él o ella se graduó o no de la universidad. Graduación de la universidad solo puede adoptar una de dos condiciones: si o no.

$$\text{Salario} = \beta_0 + \beta_1 \text{Años} + \beta_2 \text{Graduación} + e \quad (1)$$

Variables independientes cualitativas

Es muy común realizar modelos de regresión donde las variables predictoras que se utilicen sean variables de tipo cuantitativo, es decir manejando una escala bien definida (Ej: Temperatura, Ingresos económicos, Edad, etc). Pero a veces se necesita modelar incluyendo variables de tipo cualitativo donde se manejen atributos categóricos (Ej: Sexo, Color, Marca, etc), por eso es necesario revisar como manipular estas variables dentro de la modelación.

Heckman S.A. vende casas en la costa este de USA. Una de las preguntas más frecuentes de los compradores potenciales es: *Si compramos esta casa, ¿Cuánto gastaremos en calefacción durante el invierno?*. Heckman S.A. desarrolló directrices respecto de los costos de calefacción de casas unifamiliares. Se considera que tres variables se relacionan con dichos costos: 1) Temperatura externa media. 2) Pulgadas de aislamiento en el ático. 3) La casa tiene o no garaje (1: Si, 0: No). 4) Antigüedad del calentador (Años).

¿Cuál es el efecto de la variable garaje?

$$\hat{Y} = 393,666 - 3,96X_1 - 11,334X_2 + 77,432X_3 \quad (2)$$

Para mostrar el efecto suponga que se tienen exactamente dos casas iguales, una al lado de la otra, una con garaje y la otra no. Las dos casas tienen tres pulgadas de aislamiento y la temperatura media de 20 grados.

¿Cuál es el costo estimado para las dos casas?

Variables independientes cualitativas

$$\hat{Y} = 393,666 - 3,96X_1 - 11,334X_2 + 77,432X_3 \quad (2)$$

Para mostrar el efecto suponga que se tienen exactamente dos casas iguales, una al lado de la otra, una con garaje y la otra no. Las dos casas tienen tres pulgadas de aislamiento y la temperatura media de 20 grados.

¿Cuál es el costo estimado para las dos casas?

Casa1:

$$\hat{Y} = 393,666 - 3,96(20) - 11,334(3) + 77,432(0) = 280,404 \quad (3)$$

Casa 2:

$$\hat{Y} = 393,666 - 3,96(20) - 11,334(3) + 77,432(1) = 357,836 \quad (4)$$

Niveles variable indicadora

Del modelo anterior puede detallar que:

Suponiendo un modelo de primer orden se tiene:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e \quad (5)$$

Dónde X_2 es una variable ficticia o llamada también variable indicadora o variable dummy.

$$X_2 = \begin{cases} 0 & \text{Cumple la característica A} \\ 1 & \text{Cumple la característica B} \end{cases} \quad (6)$$

Niveles variable indicadora

Del modelo anterior puede detallar que:

Suponiendo un modelo de primer orden se tiene:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e \quad (5)$$

Dónde X_2 es una variable ficticia o llamada también variable indicadora o variable dummy.

$$X_2 = \begin{cases} 0 & \text{Cumple la característica A} \\ 1 & \text{Cumple la característica B} \end{cases} \quad (6)$$

Cuando $X_2 = 0$

$$Y = \beta_0 + \beta_1 X_1 + e \quad (7)$$

Cuando $X_2 = 1$

$$Y = (\beta_0 + \beta_2) + \beta_1 X_1 + e \quad (8)$$

Pero.. ¿y si X_2 tiene otra categoría?

Pero.. ¿y si X_2 tiene otra categoría?

Se puede generalizar este método a factores cualitativos con cualquier cantidad de niveles...

$$X_2 = \begin{cases} X_2 & X_3 \\ 0 & 0 & \text{Cumple la característica A} \\ 1 & 0 & \text{Cumple la característica B} \\ 0 & 1 & \text{Cumple la característica B} \end{cases} \quad (9)$$

Pero.. ¿y si X_2 tiene otra categoría?

Se puede generalizar este método a factores cualitativos con cualquier cantidad de niveles...

$$X_2 = \begin{cases} X_2 & X_3 \\ 0 & 0 & \text{Cumple la característica A} \\ 1 & 0 & \text{Cumple la característica B} \\ 0 & 1 & \text{Cumple la característica B} \end{cases} \quad (9)$$

En general, una variable cualitativa con α niveles se representa con $\alpha - 1$ variables indicadoras y cada una asume los valores 0 y 1.

Y el modelo de regresión es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e \quad (10)$$

Niveles variable indicadora

Una empresa eléctrica está investigando el efecto que tiene el tamaño de la vivienda familiar y el tipo de acondicionamiento de aire que se usa en ella, sobre el consumo total de electricidad durante los meses calurosos. Sea Y el consumo eléctrico total (en kw/h), durante el periodo septiembre a junio, y X_1 el tamaño de la casa (m^2). Hay cuatro tipos de acondicionamiento de aire:

- Sin acondicionamiento
- Unidades de ventana
- Bomba térmica
- Acondicionamiento central

Plantee las el esquema de variables indicadoras. Defina el modelo de regresión, estableciendo para cada categoría como quedaría el modelo de regresión.

Cabría esperar que el consumo eléctrico medio aumentará al aumentar el tamaño de la casa, pero el consumo será distinto para una casa con ventanas a una casa con acondicionamiento central, dado que será más eficiente una casa grande con acondicionamiento central que una casa grande con ventanas.

Cabría esperar que el consumo eléctrico medio aumentará al aumentar el tamaño de la casa, pero el consumo será distinto para una casa con ventanas a una casa con acondicionamiento central, dado que será más eficiente una casa grande con acondicionamiento central que una casa grande con ventanas.

Por lo tanto, se observa que existe una interacción entre la variable tamaño de casa y clase de acondicionamiento y debería ser tomada en cuenta en el modelo.

Cabría esperar que el consumo eléctrico medio aumentará al aumentar el tamaño de la casa, pero el consumo será distinto para una casa con ventanas a una casa con acondicionamiento central, dado que será más eficiente una casa grande con acondicionamiento central que una casa grande con ventanas.

Por lo tanto, se observa que existe una interacción entre la variable tamaño de casa y clase de acondicionamiento y debería ser tomada en cuenta en el modelo.

El cual se define como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_2 + \beta_6 X_1 X_3 + \beta_7 X_1 X_4 + e \quad (11)$$

Modelos con interacción

Defina ahora como quedarían los modelos para cada sistema de acondicionamiento:

Modelos con interacción

Defina ahora como quedarían los modelos para cada sistema de acondicionamiento:

$$Y = \beta_0 + \beta_1 X_1 + e \quad (\text{Sin acondicionamiento de aire})$$

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_5) X_1 + e \quad (\text{Unidades de ventana})$$

$$Y = (\beta_0 + \beta_3) + (\beta_1 + \beta_6) X_1 + e \quad (\text{Bomba térmica})$$

$$Y = (\beta_0 + \beta_4) + (\beta_1 + \beta_7) X_1 + e \quad (\text{Acondicionamiento central})$$

Observe que cada clase de sistema de acondicionamiento de aire, puede tener un intercepto diferente con una pendiente diferente.

Según la base de datos suministrada ¿Hay alguna interacción entre la temperatura externa y la cantidad de aislamiento? Si ambas crecen, ¿será mayor el efecto en el costo de calefacción que la suma de los ahorros derivados por una temperatura más cálida y los de mayor aislamiento, por separado?

Para el caso de estudio de Calimuebles realice la estimación del mejor modelo de regresión utilizando todas las variables de la base de datos. Evalúe posibles efectos de interacción.