

Econometria

Incorporando variables categóricas como explicativas.

Regresión Múltiple

Pasquini, Ricardo

Facultad Ciencias Empresariales Austral
Universidad Austral

April 27, 2024

Agregando una Variable Dummy Independiente

- Definición de una variable dummy y su papel en el análisis de regresión.
- Ejemplo: Agregar una variable dummy para capturar el efecto de una variable categórica (por ejemplo, género, región, etc.) en la variable dependiente.
- Interpretación del coeficiente: Cómo el coeficiente de una variable dummy representa la diferencia en la variable dependiente entre la categoría de referencia y la categoría dummy.

Efecto de Incluir una Variable Dummy para el Sexo

- **Modelo Inicial:**

$$\text{Ingreso} = \beta_0 + \beta_1 \text{educación} + \varepsilon$$

- El coeficiente β_1 representa el cambio en el ingreso por cada unidad de cambio en la educación.

Efecto de Incluir una Variable Dummy para el Sexo

- **Modelo con Variable Dummy incluyendo el Sexo:**

$$\text{Ingreso} = \beta_0 + \beta_1 \text{educación} + \beta_2 d\text{sexo} + \varepsilon$$

- Este modelo incluye una variable dummy ($d\text{sexo}$) donde $d\text{sexo} = 1$ para hombres y $d\text{sexo} = 0$ para mujeres.
- Para mujeres ($d\text{sexo} = 0$), el modelo es equivalente a:

$$\text{Ingreso} = \beta_0 + \beta_1 \text{educación} + \varepsilon$$

- Para hombres ($d\text{sexo} = 1$), el modelo es equivalente a:

$$\text{Ingreso} = (\beta_0 + \beta_2) + \beta_1 \text{educación} + \varepsilon$$

- Es decir, el modelo equivale a un modelo simple con un intercepto $(\beta_0 + \beta_2)$ para hombres, reflejando la diferencia en el ingreso promedio.
- β_2 representa la diferencia en el ingreso promedio entre géneros, manteniendo la educación constante.

Variables Dummy para Múltiples Categorías

- Podemos incorporar una variable cualitativa con múltiples categorías si convertimos cada una de sus categorías en dummies.
- Ejemplo: Supongamos una variable llamada *Nivel de Educación* con tres categorías: *Bachillerato*, *Licenciatura* y *Posgrado*.
- En principio cada categoría podría traducirse en una dummy, por ejemplo, $DBachillerato=1$ si la persona completó el Bachillerato y 0 de otro modo.
- Notar que no es necesario definir variables para todas las categorías, sino solo para todas menos 1.
- La razón es que si ya señalamos a todas menos 1, qué observaciones pertenecen a la última categoría ya está implícito, por lo que sería redundante.

Incorporando Información Ordinal

- Las variables categóricas ordinales son aquellas cuyas categorías, por su misma definición o naturaleza, tienen un orden implícito.
- Ejemplo: Considere los niveles de educación en una encuesta donde los datos relevados son del tipo: "Primario incompleto", "Primario completo", "Secundario incompleto", etc.
- Ejemplo 2: Considere el rating de crédito de un bono por S&P: "AAA", "AA", "A", "BBB", etc.
- En estos casos aplica el procedimiento general de utilizar una dummy para cada categoría.
- Seleccionamos una categoría de base para la comparación dejando la dummy afuera de la regresión.
- Notar que en muchos casos prácticos podemos encontrar variables que son numéricas pero que en realidad conviene pensar como categóricas. Ejemplo: "1", "2", "3" para primario, secundario, terciario, etc.
- Tratar una variable como numérica implica asumir que el efecto entre categorías es constante.

- Wooldridge ch.7 Multiple Regression Analysis with Qualitative Information