Uso de Variables Ficticias (*Dummy*)

- En Econometría no sólo se usan variables que son fácilmente cuantificables (por ejemplo, precio, ingreso, cantidad demandada, etc.), sino también variables que son esencialmente cualitativas. Ejemplos de éstas son sexo, raza, religión, nacionalidad, etc.

- Estas variables cualitativas son de carácter dicotómico o binario. Por ello, es fácil expresarlas como variables que puedan tomar el valor de 1 ó 0. Por ejemplo, si una persona tiene educación universitaria, la variable toma el valor de 1, si no tiene educación universitaria, toma el valor de 0.

- Las variables dicotómicas reciben el mismo tratamiento que las demás variables del modelo de regresión. Para ilustrar este punto veremos diversos ejemplos.

I Ejemplos de Modelos con Variables Ficticias

- 1.1. **Modelo de Variable Ficticia con Dos Categorías**

Se estima el modelo:

$Y_i = \beta_1 + \beta_2 D_i + u_i$,

donde:

Y_i es la nota promedio de cada alumno en el curso de Macro II.
D_i es una variable cualitativa que toma el valor de 1 si el alumno ha llevado Macro I, y toma el valor de 0 si no lo ha hecho.

Matriz de datos:

Y	Intercepto	D
Y_1	1	1
Y_2	1	0
Y_3	1	1
...	...	...
Y_n	1	1

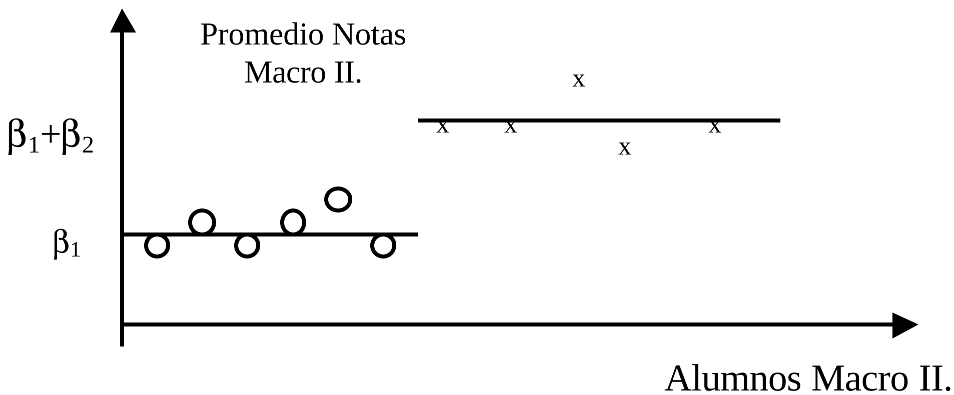
✓ Este modelo se puede estimar por MICO.

- ✓ $E(Y_i|D_i=0)=\beta_1$, es el valor esperado de la nota de los alumnos que NO han llevado Macro I.

- ✓ $E(Y_i|D_i=1)=\beta_1 + \beta_2$, es el valor esperado de la nota de los alumnos que han llevado Macro I.

- ✓ Si β_2 es significativo estadísticamente y mayor que 0, entonces el hecho de haber cursado Macro I antes de Macro II sí tiene importancia.

- ✓ Gráficamente esto se puede ver de la siguiente manera:



- **1.2 Modelo con Variables Ficticias y Cuantitativas**

Se estima el modelo:

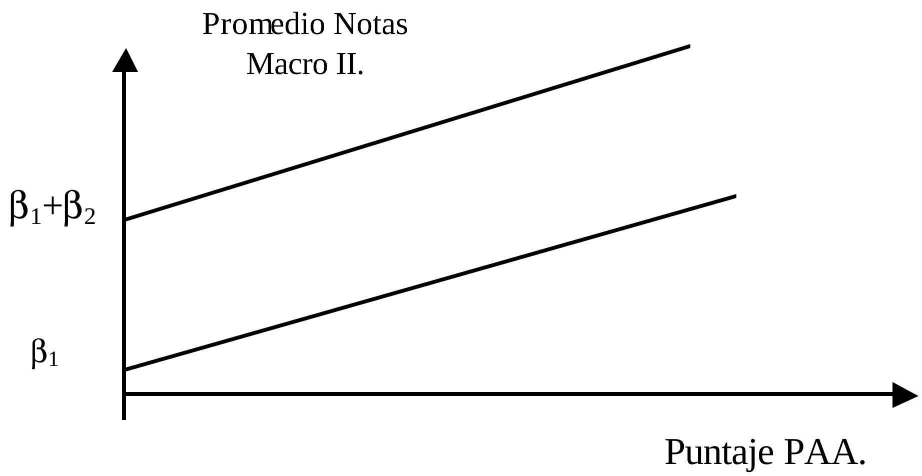
$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + u_i.$$

donde Y_i = Nota promedio en Macro II;
D_i = 1 si llevó Macro I, 0 si no;
X_i = Puntaje en la Prueba de Aptitud Académica (PAA).

- Si $\beta_3>0$, el puntaje de la PAA tiene un efecto positivo sobre el promedio de notas en Macro II, cualquiera sea éste. Si además el parámetro β_2 resulta positivo y estadísticamente significativo, haber cursado Macro I tiene también un efecto positivo sobre la nota promedio en Macro II.

- La inclusión de la variable cualitativa permite distinguir entre dos grupos de alumnos. El parámetro β_2 se denomina "intercepto diferencial".

  ✓ Gráficamente:



- La interpretación de los resultados de la regresión es la siguiente:

✓ Valor esperado de la nota de los alumnos que han llevado Macro I (dado su puntaje en la prueba de aptitud):

$$E(Y_i|D_i=1, X_i)=\beta_1 + \beta_2 + \beta_3 X_i.$$

✓ Valor esperado de la nota de los alumnos que NO han llevado Macro I (dado su puntaje en la prueba de aptitud):

$$E(Y_i|D_i=0, X_i)=\beta_1 + \beta_3 X_i.$$

- **1.3 Variables Ficticias con Más de Dos Categorías**

- Una variable *dummy* con dos categorías es, por ejemplo, "sexo" (masculino o femenino). Existen, sin embargo, variables ficticias con más de dos categorías como, por ejemplo, “nacionalidad” (chino, japonés, coreano, etc.). Veamos el siguiente caso:

Ejemplo

En un curso del magíster en economía hay 2 alumnos chilenos, 1 inglés y 1 sueco que cursaron su pregrado en sus respectivos países. Se piensa que la nacionalidad puede incidir en la nota final del curso (Y). Por lo tanto, se postula el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 D_{i2} + \beta_3 D_{i3} + \beta_4 D_{i4} + u_i,$$

donde,

D_2= 1 si es chileno; 0 si no lo es.

D_3= 1 si es inglés; 0 si no lo es.

D_4= 1 si es sueco; 0 si no lo es.

✓ Veamos la matriz de datos:

Nacionalidad	Y	Intercepto	D_2	D_3	D_4
Sueco	Y_1	1	0	0	1
Inglés	Y_2	1	0	1	0
Chileno	Y_3	1	1	0	0
Chileno	Y_4	1	1	0	0

✓ En este caso, la suma de las *dummy* es igual al intercepto ($D_2+D_3+D_4=1$, $\forall i$, i=1,..., 4). Es decir, una de las columnas de la matriz X es una combinación lineal de las columnas restantes. Ello implica que la matriz $X'X$ es singular y que, por tanto, no se puede estimar el modelo. Esta es la llamada "trampa de las variables *dummy*".

- **Importante***:* Si una variable cualitativa tiene m categorías, introdúzcase sólo m-1 variables *dummy*. En el ejemplo anterior, podríamos suprimir la categoría "sueco", por ejemplo, y estimar el modelo:

$$Y_i = \beta_1 + \beta_2 D_{i2} + \beta_3 D_{i3} + u_{i.}$$

II La Prueba de Estabilidad Estructural de un Modelo de Regresión

> Las variables ficticias pueden utilizarse para realizar un test de estabilidad de los parámetros poblacionales, como una alternativa al test de Chow. Esto es, estas permiten determinar si ha habido un cambio en los parámetros del modelo, ya sea en el intercepto o en la pendiente.

✓ Por ejemplo, se estima un modelo de consumo para Chile para los años 1960-1997, suponiendo un cambio estructural en 1975:

$$C_t = \beta_1 + \beta_2 D_t + \beta_3 I_t + \beta_4 D_t I_t + u_t, \qquad t=1960, 1961,.., 1997$$

donde C_t = Consumo Privado, I_t = Ingreso Disponible

Definamos:

$$D_t = \begin{cases} 0 \text{ para } t < 1975 & \text{(pre-cambio, 1960-1974).} \\ 1 \text{ para } t \geq 1975 & \text{(post-cambio, 1975-1997).} \end{cases}$$

De ello,

$$E(C_t|I_t, D=0) = \beta_1 + \beta_3 I_t \qquad \text{pre-cambio, 1975-1997}$$

$$E(C_t|I_t, D=1) = (\beta_1+\beta_2) + (\beta_3+\beta_4)I_t \qquad \text{post-cambio, 1975-1997}$$

Matriz de Datos

C	Intercepto	D	I	DxI
C_{60}	1	0	I_{60}	0
C_{61}	1	0	I_{61}	0
...	...	...	...	...
C_{74}	1	0	I_{74}	0
C_{75}	1	1	I_{75}	I_{75}
C_{76}	1	1	I_{76}	I_{76}
...	...	...	...	...
C_{97}	1	1	I_{97}	I_{97}

✓ Hay <u>cuatro casos posibles</u>:

1. **b_2 y b_4 no son significativos**: No hay cambio estructural, es decir, las regresiones para los dos períodos son idénticas. ¿Cómo contrastar esta hipótesis? Con un test F:

   H_0: $\beta_2=\beta_4=0$

   Si no hay cambio estructural, no rechazamos H_0.

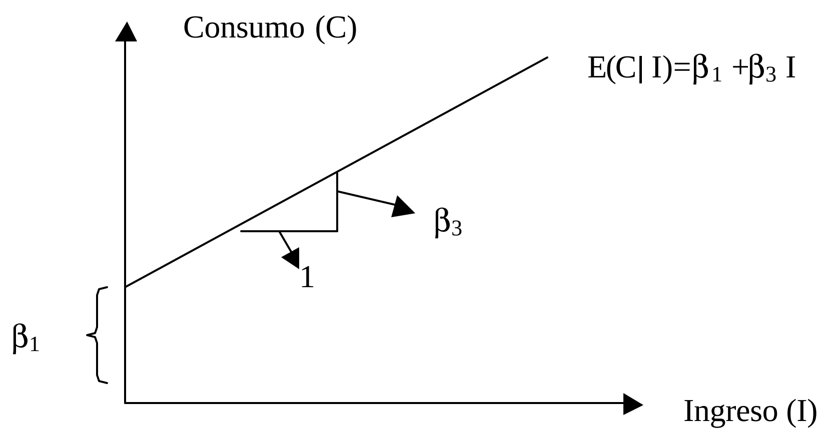
2. **b_2 es significativo, pero b_4 no es significativo**: Hay dos regresiones paralelas, pero con distintos interceptos. ¿Cómo contrastar esta hipótesis? Mirando los test t individuales de cada estimador.

3. **b_2 no es significativo, pero b_4 sí es significativo**: Hay dos regresiones concurrentes (esto es, igual intercepto) con pendientes diferentes. ¿Cómo contrastar esta hipótesis? Mirando los test t individuales de cada estimador.

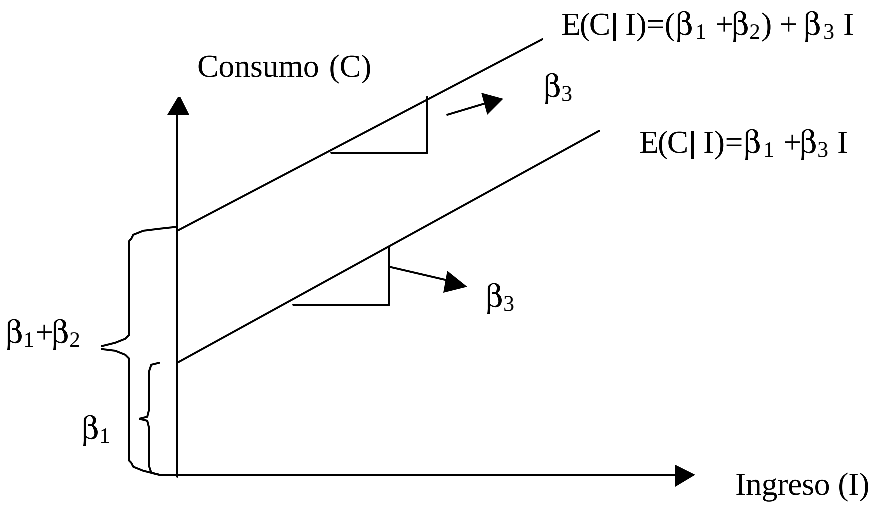
4. **b_2 y b_4 son estadísticamente significativos**: Hay dos regresiones distintas. Es decir, todos los parámetros han cambiado entre uno y otro período. En este caso, la hipótesis H_0: $\beta_2=\beta_4=0$ es rechazada y, además, β_2 y β_4 son **individualmente** estadísticamente distintos de cero.

   Gráficamente (y suponiendo β_2, β_4 >0),

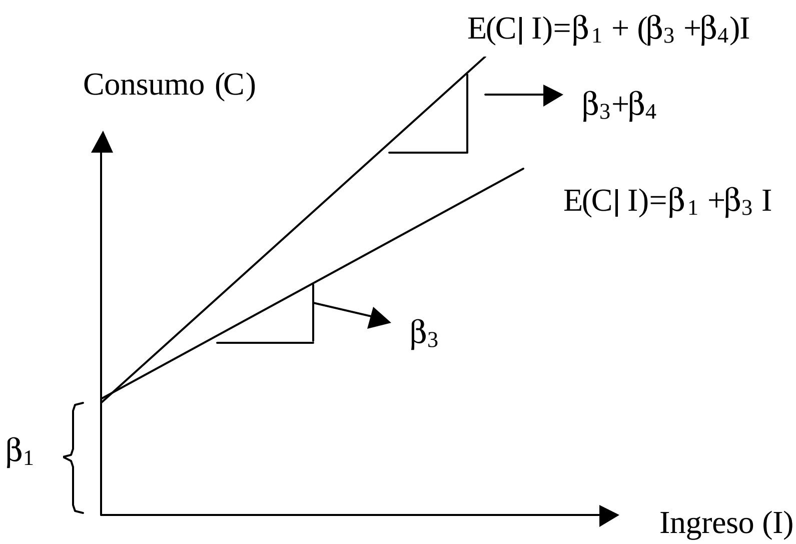
   Caso 1: Ausencia de Cambio Estructural



Caso 2: Cambio en intercepto, igual pendiente



Caso 3: Igual intercepto, distinta pendiente



Caso 4: Distinto Intercepto y Pendiente