



# **Econometría I**

**Variables cualitativas en el modelo de regresión**

**Miguel Jerez y Sonia Sotoca**

**Universidad Complutense de Madrid**

**Diciembre 2008**

# Índice

- Ideas básicas
- Interpretación
- Formas computables
- Cruces entre variables
- Contrastes de cambio estructural

## Ideas básicas

- A menudo los modelos de regresión incluyen variables binarias (0-1), que reflejan una característica presente en parte de la muestra
- A estas variables también se las denomina “variables dummy” o “variables cualitativas”
- Los coeficientes de las variables cualitativas pueden interpretarse como medias, o diferencias de medias
- A partir de un mismo modelo básico de regresión con variables cualitativas pueden obtenerse varias formas computables
- Todas las formas computables que se derivan de un mismo modelo básico son equivalentes, lo único que cambia entre una y otra es la interpretación de los parámetros
- Las variables cualitativas de un modelo de regresión pueden cruzarse multiplicativamente con otras variables del modelo
- Combinando: (a) estos resultados sobre variables cualitativas y (b) el contraste de hipótesis a partir de modelos libres y restringidos, resulta fácil construir un test para la hipótesis de cambio estructural

# Interpretación

- En un modelo de regresión, los coeficientes de las variables cualitativas pueden interpretarse como medias, o diferencias de medias
- Supongamos una regresión del peso de los niños al nacer frente a las variables SMOKE y PARITY

Dependent Variable: BWT Method: Least Squares Date: 11/20/06 Time: 12:54 Sample: 1 1174 Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.500040	0.020040	168.3552	0.0000
SMOKE	-0.263155	0.030109	-8.740110	0.0000
PARITY	-0.054642	0.033398	-1.630076	0.1021
R-squared	0.063049	Mean dependent var	3.386698	
Adjusted R-squared	0.061449	S.D. dependent var	0.519603	
S.E. of regression	0.503385	Akaike info criterion	1.467631	
Sum squared resid	296.7278	Schwarz criterion	1.480582	
Log likelihood	-858.4995	F-statistic	39.39950	
Durbin-Watson stat	2.047068	Prob(F-statistic)	0.000000	

El término constante es el peso esperado si la madre no es primípara ni fuma

Los otros coeficientes son diferenciales de peso esperado si la madre fuma o si es primípara

# Formas computables

- Un mismo modelo básico de regresión con variables cualitativas da lugar a varias formas computables
- Todas las formas computables que se derivan de un mismo modelo básico son equivalentes, lo único que cambia entre una y otra es la interpretación de los parámetros

Dependent Variable: BWT				
Method: Least Squares				
Date: 11/20/06 Time: 12:54				
Sample: 1 1174				
Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.503919	0.020813	168.3552	0.0000
SMOKE	-0.263155	0.030109	-8.740110	0.0000
PARITY	-0.054642	0.033398	-1.636076	0.1021
R squared	0.063049	Mean dependent var	3.386698	
Adjusted R-squared	0.061449	S.D. dependent var	0.519603	
S.E. of regression	0.503385	Akaike info criterion	1.467631	
Sum squared resid	296.7278	Schwarz criterion	1.480582	
Log likelihood	-858.4995	F-statistic	39.39950	
Durbin-Watson stat	2.047060	Prob(F-statistic)	0.000000	

Dependent Variable: BWT				
Method: Least Squares				
Date: 11/20/06 Time: 12:53				
Sample: 1 1174				
Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
1-SMOKE	3.503919	0.020813	168.3552	0.0000
SMOKE	3.240764	0.025016	129.5495	0.0000
PARITY	-0.054642	0.033398	-1.636076	0.1021
R-squared	0.063049	Mean dependent var	3.386698	
Adjusted R-squared	0.061449	S.D. dependent var	0.519603	
S.E. of regression	0.503385	Akaike info criterion	1.467631	
Sum squared resid	296.7278	Schwarz criterion	1.480582	
Log likelihood	-858.4995	Durbin-Watson stat	2.047068	

Los cuadros muestran dos formas computables derivadas del mismo modelo básico. Los valores de la verosimilitud, suma de cuadrados y los criterios de ajuste son los mismos. Asimismo, los parámetros de cualquiera de estos modelos pueden calcularse a partir de los del otro

## Cruces entre variables

- Las variables cualitativas de un modelo de regresión pueden cruzarse multiplicativamente con otras variables del modelo

Dependent Variable: BWT Method: Least Squares Date: 11/20/06 Time: 12:53 Sample: 1 1174 Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.509174	0.021974	159.6989	0.0000
SMOKE	-0.276501	0.035017	-7.896088	0.0000
PARITY	-0.074417	0.042626	-1.745799	0.0811
SMOKE*PARITY	0.051247	0.068628	0.746825	0.4553
R-squared	0.063496	Mean dependent var	3.386698	
Adjusted R-squared	0.061094	S.D. dependent var	0.519603	
S.E. of regression	0.503481	Akaike info criterion	1.468858	
Sum squared resid	296.5865	Schwarz criterion	1.486126	
Log likelihood	-858.2197	F-statistic	26.44233	
Durbin-Watson stat	2.045012	Prob(F-statistic)	0.000000	

En este ejemplo, la nueva variable permite que el peso esperado del hijo de una madre primípara fumadora sea distinto del de:

- una madre primípara no fumadora, o
- de una madre fumadora no primípara, o
- de una madre no primípara y no fumadora

## Contrastes de cambio estructural (I): Contraste

Combinando los resultados previos sobre variables cualitativas y (b) el contraste de hipótesis a partir de modelos libres y restringidos, resulta fácil construir un test para la hipótesis de cambio estructural

- Consideremos el “modelo restringido”:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- ... y el “modelo libre”:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

- A partir de la estimación MCO de ambos modelos, el estadístico:

$$F = \frac{n - k}{k} \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}_1^T \hat{\boldsymbol{\varepsilon}}_1 - \hat{\boldsymbol{\varepsilon}}_2^T \hat{\boldsymbol{\varepsilon}}_2}{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}} \underset{H_0}{\sim} F_{k, n-k}$$

permite contrastar la hipótesis nula de ausencia de cambio estructural. Este contraste se conoce como “Test de Chow”

## Contrastes de cambio estructural (II): Ejemplo

Dependent Variable: BWT Method: Least Squares Date: 11/28/07 Time: 10:54 Sample: 1 1174 Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.391297	0.410860	-5.820216	0.0000
GESTATION	0.013033	0.000849	15.34776	0.0000
PARITY	-0.089083	0.031101	-2.864298	0.0043
HEIGHT	1.180172	0.235549	5.010295	0.0000
WEIGHT	0.004183	0.001610	2.598072	0.0095
R-squared	0.208471	Mean dependent var	3.386698	
Adjusted R-squared	0.205763	S.D. dependent var	0.519603	
S.E. of regression	0.463070	Akaike info criterion	1.302374	
Sum squared resid	250.6734	Schwarz criterion	1.323959	
Log likelihood	-759.4934	F-statistic	76.97229	
Durbin-Watson stat	2.051684	Prob(F-statistic)	0.000000	

Dependent Variable: BWT Method: Least Squares Date: 11/28/07 Time: 10:51 Sample: 1 1174 Included observations: 1174				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
1-SMOKE	-1.922726	0.513492	-3.744412	0.0002
SMOKE	-3.221600	0.627319	-5.135506	0.0000
GESTATION*(1-SMOK	0.010486	0.001021	10.27342	0.0000
GESTATION*SMOKE	0.016504	0.001388	11.89450	0.0000
PARITY*(1-SMOKE)	-0.103552	0.038381	-2.698000	0.0071
PARITY*SMOKE	-0.075619	0.048227	-1.567999	0.1172
HEIGHT*(1-SMOKE)	1.442713	0.295709	4.878821	0.0000
HEIGHT*SMOKE	1.050331	0.357209	2.940387	0.0033
WEIGHT*(1-SMOKE)	0.002708	0.001952	1.386854	0.1658
WEIGHT*SMOKE	0.002972	0.002599	1.143405	0.2531
R-squared	0.266177	Mean dependent var	3.386698	
Adjusted R-squared	0.260503	S.D. dependent var	0.519603	
S.E. of regression	0.446828	Akaike info criterion	1.235194	
Sum squared resid	232.3983	Schwarz criterion	1.278364	
Log likelihood	-715.0587	Durbin-Watson stat	2.055058	

- El análisis sobre el impacto de una madre fumadora sobre el peso de un niño al nacer puede plantearse como un problema de cambio estructural
- El modelo en el listado de la izquierda pone en relación el peso con el período de gestación, si la madre es primípara o no, la altura y el peso de la madre. Implícitamente se impone la restricción de que el que la madre fume o no es irrelevante
- El modelo del listado de la derecha permite que todos los parámetros del modelo anterior sean distintos para las madres fumadoras y no fumadoras
- El estadístico de cambio estructural sería:  $((1174-5)/5) * (250.6734 - 232.3983) / 250.6734 = 17.0450$
- Por lo que la nula se rechaza al 5% de significación (valor crítico=2.124) y al 1% (valor crítico=3.017). Alternativamente, el  $p$ -valor del test anterior es 0