
Universidad de San Andrés – Departamento de Economía – Econometría – Semestre de otoño

MULTICOLINEALIDAD

Mariana Marchionni

mariana@depeco.econo.unlp.edu.ar

Ejemplo empírico:

la relación entre la educación de los hijos y la educación de los padres

- El enfoque económico de la decisión de estudiar se basa fundamentalmente en el trabajo de Becker (1965), “A Theory of the Allocation of Time”, Economic Journal, 75: 493-517.
- La educación surge como el resultado de un problema de maximización de la utilidad. El nivel de educación óptima es aquél que iguala los costos marginales de educarse (directos, indirectos, costos de oportunidad) con los beneficios marginales (ingresos futuros).
- Al menos para los hijos pequeños, son generalmente los padres los que toman la decisión de si mandar o no a un hijo a la escuela. Y para los hijos mayores, es posible que los padres sigan teniendo todavía alguna influencia sobre esa decisión.

- Con mercados de capitales perfectos, los recursos de la familia no serían relevantes en la decisión de cuántos años mandar a la escuela a un chico.
- Pero con restricciones de liquidez, deberíamos observar que hijos de familias más ricas se educan más.
- Si padres más educados son potencialmente más ricos, con restricciones de liquidez deberíamos observar que los hijos de padres más educados se educan más.
- Hay “otros factores” que reflejan restricciones psicológicas, culturales y sociales que influyen sobre la manera en que las familias evalúan la decisión de cuánto deben educarse sus hijos. La educación de los padres también puede reflejar este tipo de restricciones, además de las estrictamente económicas.
- Objetivo. Estudiar la relación entre la educación de los padres y la de los hijos.
- Muestra (datos ficticios). 500 individuos de más de 30 años. De cada uno sabemos si es hombre o mujer y tenemos información sobre los años de educación formal propios (*edu*) y de sus padres (*edupadre* y *edumadre*).

Composición de la muestra

```
summ hombre mujer edu edupadre edumadre
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
hombre	500	.5	.5005008	0	1
mujer	500	.5	.5005008	0	1
edu	500	9.222	3.7455	0	19
edupadre	500	7.366	4.398948	0	15
edumadre	500	7.86	4.384387	0	16

Un 50% (0.5) de los individuos de la muestra son mujeres.
Recordar: el promedio de una variable binaria que toma sólo valores 0 y 1 es la proporción de unos en la muestra

Caso 1: La trampa de la variable binaria

```
. regress edu edumadre edupadre mujer hombre
```

edu	Coef.	Std. Err.	t	P> t
edumadre	.563601	.1906244	2.96	0.003
edupadre	.1312789	.1899992	0.69	0.490
mujer	1.031988	.1899127	5.43	0.000
hombre	(dropped)			
_cons	3.309102	.2357071	14.04	0.000

Al momento de la estimación, el programa elimina automáticamente una de las variables linealmente dependientes

- NO puede estimarse el modelo por MCO si entre dos o más regresores hay multicolinealidad perfecta.
- Si una variable cualitativa tiene S categorías, usamos $(S-1)$ variables binarias en el modelo.

Caso 2: una característica común a toda la muestra

Estimamos el modelo anterior solamente para las mujeres

```
regress edu edumadre edupadre mujer if mujer==1
```

edu	Coef.	Std. Err.	t	P> t
edumadre	.7261334	.2690986	2.70	0.007
edupadre	-.0490358	.2663003	-0.18	0.854
mujer	(dropped)			
_cons	4.396034	.3030608	14.51	0.000

El programa elimina automáticamente la variable dummy *mujer*

- Este es otro caso de multicolinealidad perfecta
- La variable dummy *mujer* es perfectamente colineal (o linealmente dependiente) con qué otra variable?

Caso 3: multicolinealidad alta (no perfecta)

```
regress edu edumadre edupadre mujer
```

edu	Coef.	Std. Err.	t	P> t
edumadre	.563601	.1906244	2.96	0.003
edupadre	.1312789	.1899992	0.69	0.490
mujer	1.031988	.1899127	5.43	0.000
_cons	3.309102	.2357071	14.04	0.000

- La educación de la madre (*edumadre*) es muy significativa para explicar la educación de los hijos
- Pero la educación del padre (*edupadre*) no parece ser relevante.

Estimemos otra vez el modelo excluyendo la variable *edumadre*

```
regress edu edupadre mujer
```

edu	Coef.	Std. Err.	t	P> t
edupadre	.6893938	.0217549	31.69	0.000
mujer	1.007606	.1912056	5.27	0.000
_cons	3.640122	.2090288	17.41	0.000

- Ahora la educación del padre aparece como muy significativa para explicar la educación de los hijos
- ¿A qué se debe este drástico cambio en los resultados? Generalmente, este tipo de comportamiento es indicio de una fuerte (pero no perfecta) relación lineal entre las variables del modelo.

- ¿Cuál es es grado de relación lineal entre *edupadre* y *edumadre*

corr	edupadre	edumadre
		edupadre edumadre
-----	+	-----
edupadre		1.0000
edumadre		0.9935 1.0000

En este ejemplo vemos cómo la no significatividad estadística de una variable (*edupadre*) es consecuencia de la incapacidad de los estimadores de “identificar” o “diferenciar” el efecto de cada una de las variables altamente correlacionadas (*edupadre* y *edumadre*).

Una solución sería construir una medida resumen o agregada de la educación de los padres. Por ejemplo, educación promedio: $\text{edupromedio} = (\text{edupadre} + \text{edumadre})/2$

```
regress edu edu_promedio mujer
```

	Coef.	Std. Err.	t	P> t
edu				
edu_promedio	.6941675	.0216643	32.04	0.000
mujer	1.022515	.1897863	5.39	0.000
_cons	3.426046	.212175	16.15	0.000