

Regresión múltiple

Gabriel V. Montes-Rojas

Regresión múltiple

Tomemos el ejemplo de retornos a la educación en una ecuación de Mincer. Como se imaginan, la educación no es el único determinante de los salarios. Otras variables que pueden afectar salarios son:

- Edad
- Experiencia
- Habilidad (¿empresarial?)
- Sexo (¿por?)
- Raza/nacionalidad (¿por?)

Por ejemplo, un modelo más cercano a la realidad es:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

¿Cambia la interpretación de los coeficientes?

En el modelo

$$wage = \gamma_0 + \gamma_1 educ + e,$$

si omitimos $exper$, y si $educ$ y $exper$ están relacionadas, o sea $exper(educ)$, entonces,

$$\begin{aligned}\gamma_1 &= \frac{\partial E[wage|educ]}{\partial educ} \\ &= \frac{\partial E[wage|educ, exper]}{\partial educ} + \frac{\partial E[wage|educ, exper]}{\partial exper} \times \frac{\partial exper}{\partial educ}\end{aligned}$$

¿Cuál es el problema? Si no controlamos por $exper$, estaríamos estimando un efecto de $educ$ que no es el que queremos. Es el problema de variables omitidas que veremos más adelante. Aparece el problema de la **causalidad**.

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

$$\beta_1 = \frac{\partial E[wage | educ, exper]}{\partial educ}$$

Ahora β_1 es el efecto de *educ* sobre *wage*, **manteniendo** *exper* **constante**. Incluyendo *exper* en el modelo, podemos medir el efecto de *educ* **sin confundirlo con el efecto de *exper***.

Nota: En Economía esto se usa mucho: *ceteris paribus*, dejando todo lo demás constante.

Regresión múltiple

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_K x_K + u$$

Tenemos K variables explicativas. Asumimos que $x_1 \equiv 1$ para que todos los modelos tengan un intercepto (excepto cuando se indique).

En notación matricial tenemos $\mathbf{x} = (x_1, x_2, \dots, x_K)$ que es un vector $1 \times K$ y $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_K)'$ es un vector $K \times 1$, entonces

$$y = \mathbf{x}\boldsymbol{\beta} + u$$

Entonces para cada observación i tenemos

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$$

Regresión múltiple

- Para el caso de muchos regresores también tenemos la misma interpretación de los coeficientes de la regresión:

$$\beta_j = \frac{\text{Cov}(y, \tilde{x}_j)}{\text{Var}(\tilde{x}_j)}, j = 2, 3, \dots, K,$$

donde \tilde{x}_j es el residuo de la regresión de x_j en todos los demás, \mathbf{x}_{-j} .

Prueba: $\text{Cov}(y, \tilde{x}_j) = \text{Cov}(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_K x_K + u, \tilde{x}_j)$. Por uno de los supuestos de MCO, dado que \tilde{x}_j es una función de los regresores, entonces no está correlacionado con u . Además, tampoco está correlacionado con \mathbf{x}_{-j} . Entonces, $\text{Cov}(y, \tilde{x}_j) = \text{Cov}(\beta_j x_j, \tilde{x}_j) = \beta_j \text{Cov}(x_j, \tilde{x}_j) = \beta_j \text{Var}(\tilde{x}_j)$.

Álgebra de MCO

El estimador de MCO es

$$\hat{\beta}_{MCO} = \underset{\mathbf{b} \in \mathbb{R}^K}{\operatorname{argmin}} \sum_{i=1}^N (y_i - b_1 - b_2 x_{2i} - \dots - b_K x_{Ki})^2$$

Para la minimización tomamos derivadas con respecto a $\mathbf{b} = (b_1, b_2, \dots, b_K)$, lo cual nos da las siguientes condiciones de primer orden:

$$\sum_{i=1}^N x_{ji} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_K x_{Ki}) = 0, j = 1, 2, \dots, K.$$

Al igual que con la regresión simple podemos pensar la solución como un método de momentos:

Momentos en la población
$E[x_j u] = E[x_j (y - \beta_1 - \beta_2 x_2 - \dots - \beta_K x_K)] = 0$
$j = 1, 2, \dots, K$
Momentos en la muestra (CPO de la minimización)
$N^{-1} \sum_{i=1}^N x_{ji} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_K x_{Ki}) = 0$
$j = 1, 2, \dots, K$

En ambos casos tenemos un sistema de ecuaciones con K ecuaciones y K parámetros.

Álgebra de MCO

$$\hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

donde:

- \mathbf{X} es una matriz $N \times K$, que contiene en cada una de las N filas la observación i y en cada una de las K columnas los parámetros. Nota: $x_{1i} = 1$ para todo $i = 1, 2, \dots, N$;
- \mathbf{y} es un vector $N \times 1$, que contiene la variable dependiente;
- ... así $(\mathbf{X}'\mathbf{X})$ es una matriz $K \times K$; ' representa la transpuesta de una matriz;
- ... $(\mathbf{X}'\mathbf{X})^{-1}$ es una matriz $K \times K$, la inversa de $(\mathbf{X}'\mathbf{X})$;
- ... $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ es un vector $K \times 1$.

Álgebra de MCO

En detalle, las matrices y vectores son:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{21} & \dots & x_{K1} \\ 1 & x_{22} & \dots & x_{K2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2N} & \dots & x_{KN} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix},$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum_{i=1}^N x_{2i} & \dots & \sum_{i=1}^N x_{Ki} \\ \sum_{i=1}^N x_{2i} & \sum_{i=1}^N x_{2i}^2 & \dots & \sum_{i=1}^N x_{2i}x_{Ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^N x_{Ki} & \sum_{i=1}^N x_{Ki}x_{2i} & \dots & \sum_{i=1}^N x_{Ki}^2 \end{bmatrix}, \mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{2i}y_i \\ \vdots \\ \sum_{i=1}^N x_{Ki}y_i \end{bmatrix}.$$

La expresión para $(\mathbf{X}'\mathbf{X})^{-1}$ requiere repasar cómo se calcula la inversa de una matriz.

Ejercicio: Resolver analíticamente para $K = 2$ (regresión simple) y $K = 3$ usando álgebra matricial.

Álgebra de MCO

$$\hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Prueba: Planteemos el problema de minimización como

$$\begin{aligned}\hat{\beta}_{MCO} &= \underset{\mathbf{b} \in \mathbb{R}^K}{\operatorname{argmin}} \mathbf{u}(\mathbf{b})' \mathbf{u}(\mathbf{b}) = \underset{\mathbf{b} \in \mathbb{R}^K}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \underset{\mathbf{b} \in \mathbb{R}^K}{\operatorname{argmin}} [\mathbf{y}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - 2\mathbf{b}'\mathbf{X}'\mathbf{y}].\end{aligned}$$

Estamos definiendo $\mathbf{u}(\mathbf{b}) \equiv \mathbf{y} - \mathbf{X}\mathbf{b}$. Tenemos que tomar derivadas con respecto a $\mathbf{b} \in \mathbb{R}^K$, un vector $K \times 1$. La solución es un vector $K \times 1$ de condiciones de primer orden (hay reglas específicas para derivar vectores y matrices). Entonces, $2\mathbf{X}'\mathbf{X}\mathbf{b} - 2\mathbf{X}'\mathbf{y} = \mathbf{0}_K$, donde $\mathbf{0}_K$ es un vector $K \times 1$ de ceros. Finalmente, $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$, $\Rightarrow \hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Hay que chequear también las condiciones de segundo orden (para mínimo). Tomando la segunda derivada tenemos, $2\mathbf{X}'\mathbf{X}$ que es una matriz positiva semidefinida (chequear).

Álgebra de MCO

Otra forma de verlo es que las condiciones de momento se pueden expresar como:

$$\mathbf{X}'\mathbf{u}(\hat{\beta}_{MCO}) = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}_{MCO}) = \mathbf{0}_K.$$

Entonces, $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\beta}_{MCO}$. Este es un sistema de ecuaciones lineales, no homogéneo. Para que tenga solución los coeficientes de las ecuaciones, dados por $\mathbf{X}'\mathbf{X}$, no pueden ser linealmente dependientes. O sea el determinante no puede ser cero. Entonces,

$$\Rightarrow \hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Teorema Gauss-Markov

- **Supuesto 1: Lineal en parámetros** La variable dependiente y se relaciona con \mathbf{x} por una función lineal, $y = \beta_1 + \beta_2 x_2 + \dots + \beta_K x_K + u$.
- **Supuesto 2: Muestreo aleatorio** $\{(y_i, x_{1i}, x_{2i}, \dots, x_{Ki})\}$, $i = 1, 2, \dots, N$ es una muestra aleatoria del modelo del Supuesto 1.
- **Supuesto 3: Ausencia de colinealidad perfecta en \mathbf{X}** Para esto necesitamos que $(\mathbf{X}'\mathbf{X})$ sea no singular o $\text{rango}(\mathbf{X}'\mathbf{X}) = K$ (notar que esto implica que $K \leq N$). Condición necesaria y suficiente para esto es que no haya una relación lineal exacta entre los regresores (no confundir con multicolinealidad en general).
- **Supuesto 4: Media condicional cero** $E[\mathbf{u}|\mathbf{X}] = 0$

MCO es insesgado $E[\hat{\beta}_j|\mathbf{X}] = \beta_j$, $j = 1, 2, \dots, K$ o $E[\hat{\beta}|\mathbf{X}] = \beta$ donde β es el vector de todos los parámetros.

Prueba: $E[\hat{\beta}_{MCO}|\mathbf{X}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})|\mathbf{X}] = E[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\beta|\mathbf{X}] + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}|\mathbf{X}] = \beta$.

Especificar donde se usa cada supuesto explícitamente.

Teorema Gauss-Markov

- **Supuesto 5: Homocedasticidad** $\text{Var}[\mathbf{u}|\mathbf{X}] = \sigma^2 \mathbf{I}_N$

Teorema Gauss-Markov: Bajo los Supuestos 1-5, los estimadores MCO $(\hat{\beta}_1, \dots, \hat{\beta}_K)$ son los mejores estimadores lineales insesgados de $(\beta_1, \dots, \beta_K)$. Note: MEJOR significa mínima varianza dentro de la familia de estimadores lineales insesgados. En inglés es BLUE, best linear unbiased estimator.

Varianza de MCO

En notación matricial tenemos $Var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Prueba:

$$Var(\hat{\beta}|\mathbf{X}) = Var[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var[\mathbf{y}|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Notar que $Var(\mathbf{y}|\mathbf{X}) = Var(\mathbf{X}\beta + \mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_N$. Por otro lado $\Omega = Var(\mathbf{u}|\mathbf{X}) = E[\mathbf{u}\mathbf{u}'|\mathbf{X}] + E[\mathbf{u}|\mathbf{X}]E[\mathbf{u}'|\mathbf{X}] = E[\mathbf{u}\mathbf{u}'|\mathbf{X}]$ (¿por qué?). Ω juega un rol central para analizar la varianza. Con los supuestos de Gauss-Markov,

$$\Omega = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

En general, para los estimadores MCO podemos escribir la varianza como una forma sandwich,

$$Var(\hat{\beta}|\mathbf{X}) = Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \\ = (\mathbf{X}'\mathbf{X})^{-1}E[\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}|\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{u}|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Teorema Gauss-Markov

Ahora podemos probar el Teorema de Gauss-Markov.

Un estimador insesgado requiere que sea una combinación lineal de \mathbf{y} .

Definamos $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}$ donde \mathbf{D} es una matriz no nula $K \times N$, no estocástica. Ahora definamos $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$ como otro estimador.

$$\begin{aligned} E[\tilde{\boldsymbol{\beta}}] &= E[\mathbf{C}\mathbf{y}] = E[\mathbf{C}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})] = \boldsymbol{\beta} + E[\mathbf{D}\mathbf{X}\boldsymbol{\beta}] + E[\mathbf{C}\mathbf{u}] \\ &= (\mathbf{I}_K + \mathbf{D}\mathbf{X})\boldsymbol{\beta}. \end{aligned}$$

Para que sea insesgado debemos tener $\mathbf{D}\mathbf{X} = \mathbf{0}_{K \times K}$. Notar que $E[\mathbf{C}\mathbf{u}] = \mathbf{0}_{K \times 1}$. Ahora calculemos la varianza, siempre condicional en \mathbf{X} ,

$$\begin{aligned} \text{Var}[\tilde{\boldsymbol{\beta}}|\mathbf{X}] &= \text{Var}[\mathbf{C}\mathbf{y}|\mathbf{X}] = \mathbf{C}\text{Var}(\mathbf{y}|\mathbf{X})\mathbf{C}' = \sigma^2\mathbf{C}\mathbf{C}' = \sigma^2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D})' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}' = \text{Var}[\hat{\boldsymbol{\beta}}_{MCO}|\mathbf{X}] + \sigma^2\mathbf{D}\mathbf{D}'. \end{aligned}$$

Dado que $\mathbf{D}\mathbf{D}'$ es una matriz positiva semidefinida, tenemos el resultado:

$\text{Var}[\tilde{\boldsymbol{\beta}}|\mathbf{X}] - \text{Var}[\hat{\boldsymbol{\beta}}_{MCO}|\mathbf{X}]$ es una matriz positiva semidefinida.

Notar que solo MCO no tiene forma sandwich: $(\mathbf{X}'\mathbf{X})^{-1} \times \text{ALGO} \times (\mathbf{X}'\mathbf{X})^{-1}$. Esto es un punto central para detectar eficiencia.

Análisis de varianza

Definiciones...

Suma de Cuadrados Totales (SCT): $\sum_{i=1}^N (y_i - \bar{y})^2$

⇒ Variación total en y ; $\widehat{Var}(y) = N^{-1} \sum_{i=1}^N (y_i - \bar{y})^2$

Suma de Cuadrados Explicados (SCE): $\sum_{i=1}^N (\hat{y}_i - \bar{y})^2$

⇒ Variación en y explicada por el modelo

Suma de Cuadrados Residuales (SCR): $\sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N \hat{u}_i^2$

⇒ Variación total en u

Se puede probar que: $SCT = SCE + SCR$

R^2

Definición: El R – cuadrado o R^2 de una regresión es la fracción de la variación en y que es explicada por \mathbf{X} , el modelo lineal propuesto.

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = SCE / SCT = 1 - SCR / SCT$$

$$0 \leq R^2 \leq 1$$

Precaución

Cuantas más variables explicativas se usan en el modelo, mayor va ser el R^2 . ¿Por qué? Las X s ayudan a explicar la variación en y siempre, aún cuando sean irrelevantes.

Nunca hay que juzgar un modelo en base al R^2 .

Los valores de R^2 dependen del tipo de problema y de la experiencia...

R^2 ajustado

Definición: \bar{R}^2 o R^2 ajustado es un estadístico como el R^2 pero donde se **penaliza** por la inclusión de variables.

$$R^2 = 1 - \frac{SCR/(N - K)}{SCT/(N - 1)}$$

¡Ahora perdemos la interpretación de $0 \leq \bar{R}^2 \leq 1$!

Nota: \bar{R}^2 se incrementa sólo si la variable adicional tiene un valor t mayor a uno en valor absoluto. Se puede usar como criterio de selección.

Varianza de MCO

Teorema: Bajo los Supuestos 1-5 de Gauss Markov,

$$\text{Var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{SCT_j(1 - R_j^2)}, \quad j = 2, \dots, K$$

donde $SCT_j = \sum_{i=1}^N (x_{ji} - \bar{x}_j)^2$ es la variación total en x_j y R_j^2 es el R-cuadrado de una regresión de x_j en todas las otras variables (incluyendo el intercepto) $\{1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_K\}$.

Contrastes de hipótesis simples

Tomemos la **hipótesis nula** $H_0 : \beta_j = \beta_{j0}, j = 1, \dots, K$
contra la **hipótesis alternativa** $H_A : \beta_j \neq \beta_{j0}$

(si $H_0 : \beta_j = 0$ es verdad entonces no hay relación lineal entre y con x_j , **luego de controlar por el efecto de las otras variables.**)

Contrastes de hipótesis simples

- Bajo $H_0 : \beta_j = \beta_{j0}$ y asumiendo que $u \sim \text{Normal}(0, \sigma^2)$, tiene distribución normal estándar, tenemos que

$$(\hat{\beta}_j - \beta_{j0}) / \widehat{se}(\hat{\beta}_j) \sim t_{N-K}$$

donde $se(\cdot)$ es error estándar (raíz cuadrada de la varianza) y t_{N-K} es una distribución t con grados de libertad $N - K$.

- ¿De dónde viene $N - K$? De que tenemos N observaciones y estimamos K parámetros. La variable aleatoria t - *Student* tiene colas mas anchas cuanto menos grados de libertad tiene. Por otro lado cuando los grados de libertad tienden a infinito, la distribución t se vuelve igual a la normal/gaussiana.
- Para obtener $\text{Var}(\hat{\beta}|\mathbf{X})$ necesitamos un estimador de σ^2 , la varianza del error, la cual también necesita ser estimada.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \hat{u}_i^2}{N - K}$$

Se puede probar que $E(\hat{\sigma}^2|\mathbf{x}) = \sigma^2$.

Contrastes de hipótesis múltiples

- Se pueden contrastar hipótesis como $H_0 : \beta_1 = 0, \beta_2 = 0$,
- o como $H_0 : \beta_1 = \beta_2$.
- Estas hipótesis se contrastan con test F.

Contrastes de hipótesis múltiples

En general, si tenemos un modelo con K variables independientes, pero queremos testear por Q restricciones lineales (no perfectamente colineales entre sí), se define ur (unrestricted model) modelo sin restricciones (K variables)
 r (restricted model) modelo con las restricciones (el modelo estimado satisfaciendo todas las Q restricciones)

Entonces:

$$F = \frac{(R_{ur}^2 - R_r^2)/Q}{(1 - R_{ur}^2)/(N - K)} \sim F(Q, N - K)$$

Contraste para la significancia del modelo

Supongamos que nos interesa: $H_0 : \beta_2 = \dots = \beta_K = 0$

En este caso $Q = K - 1$, el número de restricciones es igual al número de variables explicativas (excepto la constante).

Then:

$$F = \frac{(R^2)/K - 1}{(1 - R^2)/(N - K)} \sim F(K - 1, N - K)$$

Ejemplo: Efecto de fumar sobre el peso de los recién nacidos

Consideremos la siguiente regresión:

$$\text{bwght} = \beta_0 + \beta_1 \text{cigs} + \beta_2 \text{parity} + \beta_3 \text{faminc} \\ + \beta_4 \text{motheduc} + \beta_5 \text{fatheduc} + u$$

donde

bwght: birth weight, in pounds;

cigs: average number of cigarettes the mother smoked per day during pregnancy;

parity: birth order of the child;

faminc: annual family income;

motheduc: years of schooling of the mother;

fatheduc: years of schooling of the father.

Ejemplo: Efecto de fumar sobre el peso de los recién nacidos

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/bwght, clear
reg bwght cigs parity faminc motheduc fatheduc
```

STATA

bwght	Coef.	Std. Err.	t	$P > t $
cigs	-.5959362	.1103479	-5.401	0.000
parity	1.787603	.6594055	2.711	0.007
faminc	.0560414	.0365616	1.533	0.126
motheduc	-.3704503	.3198551	-1.158	0.247
fatheduc	.4723944	.2826433	1.671	0.095
cons	114.5243	3.728453	30.716	0.000

<http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge4.html>

Ejemplo: Efecto de fumar sobre el peso de los recién nacidos

Si quisieramos hacer un contraste de $H_0 : \beta_5 = 0, \beta_6 = 0$

```
test motheduc fatheduc
```

(1) motheduc = 0.0

(2) fatheduc = 0.0

F(2, 1185) = 1.44

Prob > F = 0.2380

Ejemplo: Efecto de fumar sobre el peso de los recién nacidos

El estadístico F se puede construir a mano:

```
reg bwght cigs parity faminc motheduc fatheduc
scalar R2ur=e(r2) (guarda  $R_{ur}^2$ )
reg bwght cigs parity faminc if fatheduc~= .
scalar R2r=e(r2) (guarda  $R_r^2$ )
```

$$F = \frac{(R_{ur}^2 - R_r^2)/Q}{(1 - R_{ur}^2)/(N - Q)} \sim F(Q, N - K)$$

```
scalar F=(R2ur-R2r)/2/(1-R2ur)*(e(N)-5-1) (estadístico F)
scalar pvalueF=Ftail(2,e(N)-5-1,F)(obtiene el p-valor)
display "F statistic : " F
display "p-value : " pvalueF
```

Ejemplo: Efecto de fumar sobre el peso de los recién nacidos

¿Cómo podemos testear por $H_0 : \beta_5 = \beta_6$?

```
test motheduc=fatheduc
```

```
...
```

```
reg bwght cigs parity faminc motheduc fatheduc
```

```
scalar R2ur=e(r2) (guarda  $R_{ur}^2$ )
```

```
gen mfeduc=matheduc+fatheduc
```

```
reg bwght cigs parity faminc mfeduc (guarda  $R_r^2$ )
```

```
scalar R2r=e(r2) (keep the  $R_r^2$ )
```

```
scalar F=(R2ur-R2r)/1/(1-R2ur)*(e(N)-5-1) (estadístico F)
```

```
scalar pvalueF=Ftail(2,e(N)-5-1,F)(obtiene el p-valor)
```

```
display "F statistic : " F
```

```
display "p-value : " pvalueF
```