Teorema de Gauss-Markov e inferencia en el modelo lineal general

Mariana Marchionni
marchionni.mariana@gmail.com

Econometría I - FCE - UNLP
www.econometria1unlp.weebly.com

Temario de la clase

1. Repaso y digresión
   - Repaso del modelo con matrices
   - Propiedades de la matriz de varianzas y covarianzas
2. Teorema de Gauss-Markov
   - Enunciado e intuición del TGM
   - Demostración
3. Inferencia en el modelo lineal general
   - Hipótesis simples
   - Hipótesis compuestas
   - Ejemplo empírico

El modelo lineal general en notación matricial

- Tenemos el modelo con K variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_K X_{Ki} + \mu_i, \; i = 1, \ldots n$$

- Y vimos que usando matrices puede escribirse como

$$Y = X\beta + \mu$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{K1} \\ 1 & X_{22} & X_{32} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{Kn} \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

donde X es $(n \times K)$, Y y μ son $(n \times 1)$, y β es $(K \times 1)$.

Definiciones

- Vector de estimadores $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$
- Vector de estimaciones de Y $(n \times 1)$: $\hat{Y} \equiv X\hat{\beta}$
- Vector de residuos o errores de estimación $(n \times 1)$: $e \equiv Y - \hat{Y} = Y - X\hat{\beta}$

MCO en matrices

- El problema de MCO:

$$\underset{\hat{\beta}}{Min}\ SRC(\hat{\beta}) \equiv \sum_{i=1}^{n} e_i^2 = e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

- Las CPO:

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0 - 2X'Y + 2X'X\hat{\beta} = 0 \Rightarrow X'X\hat{\beta} = X'Y$$

- El vector de estimadores MCO es:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

  - Para obtener $\hat{\beta}$ tuvimos que suponer que $(X'X)$ tiene inversa.

Recordemos qué es $X'X$

$$X'X = \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{Ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{Ki} \\ \sum X_{3i} & \sum X_{2i}X_{3i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{Ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{Ki} & \sum X_{2i}X_{Ki} & \sum X_{3i}X_{Ki} & \cdots & \sum X_{Ki}^2 \end{bmatrix}$$

- La existencia de $(X'X)^{-1}$ se garantiza por el supuesto de no multicolinealidad perfecta.
- Recordemos: si no hay multicolinealidad perfecta $\rho(X) = K$, lo que a su vez implica que $\rho(X'X) = K \Rightarrow \exists (X'X)^{-1}$

Los supuestos clásicos en notación matricial

1. $E(\mu) = \begin{bmatrix} E(\mu_1) \\ E(\mu_2) \\ \vdots \\ E(\mu_n) \end{bmatrix} = 0$

2. $VarCov(\mu) = \begin{bmatrix} V[\mu_1] & Cov[\mu_1, \mu_2] & \cdots & Cov[\mu_1, \mu_n] \\ Cov[\mu_2, \mu_1] & V[\mu_2] & \cdots & Cov[\mu_2, \mu_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[\mu_n, \mu_1] & Cov[\mu_n, \mu_2] & \cdots & V[\mu_n] \end{bmatrix} = \sigma^2 \mathrm{I_n}$

3. X_{nxK} no estocástica con $\rho(X) = K$

Propiedades estadísticas de los estimadores MCO

- Insesgadez: $E[\hat{\beta}] = \beta$
- Matriz de varianzas y covarianzas: $VarCov(\hat{\beta}) = \sigma^2(X'X)^{-1}$

$$= \begin{bmatrix} V[\hat{\beta}_1] & Cov[\hat{\beta}_1, \hat{\beta}_2] & \cdots & Cov[\hat{\beta}_1, \hat{\beta}_K] \\ Cov[\hat{\beta}_1, \hat{\beta}_2] & V[\hat{\beta}_2] & \cdots & Cov[\hat{\beta}_2, \hat{\beta}_K] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[\hat{\beta}_1, \hat{\beta}_K] & Cov[\hat{\beta}_2, \hat{\beta}_K] & \cdots & V[\hat{\beta}_K] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 A_{11} & \sigma^2 A_{12} & \cdots & \sigma^2 A_{1K} \\ \sigma^2 A_{21} & \sigma^2 A_{22} & \cdots & \sigma^2 A_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 A_{K1} & \sigma^2 A_{K2} & \cdots & \sigma^2 A_{KK} \end{bmatrix}$$

donde A_{kk} es el elemento en la fila k y columna k de la matriz $(X'X)^{-1}$

- El estimador de $VarCov(\hat{\beta})$ es $\widehat{VarCov}(\hat{\beta}) = S^2(X'X)^{-1}$ con $S^2 = \frac{e'e}{n-K}$

Semidefinición positiva de una matriz

Definición: matriz semidefinida positiva

Sea H una matriz $m \times m$ y sea c cualquier vector de m constantes. Se dice que H es semidefinida positiva si y sólo si $c'Hc \geq 0 \; \forall c \in \mathbb{R}^m$

- Notar:
  - que $c'Hc$ es (1×1), o sea, un número real.
  - $c'Hc$ es una **forma cuadrática.** Para ver por qué, desarrollar la expresión para $m = 2$.
- **Importante**: cualquier matriz de varianzas y covarianzas es semidefinida positiva.
  - En particular, $VarCov[\hat{\beta}]$ es semidefinida positiva.

$VarCov[\hat{\beta}]$ es semidefinición positiva...¿y qué hay con eso?

- Entonces, para cualquier vector de constantes c de dimensión $K \times 1$ se cumple que $c' VarCov[\hat{\beta}]c \geq 0$.
- La expresión $c' VarCov[\hat{\beta}]c$ es la varianza de la combinación lineal $c'\hat{\beta}$.
- Veamos el ejemplo para $K = 2$:
  - $c'\hat{\beta} = c_1\hat{\beta}_1 + c_2\hat{\beta}_2$
  - $V[c'\hat{\beta}] = c' VarCov[\hat{\beta}]c = c_1^2 V[\hat{\beta}_1] + c_2^2 V[\hat{\beta}_2] + 2c_1c_2 Cov[\hat{\beta}_1, \hat{\beta}_2]$
  - La semidefinición positiva garantiza que esta última expresión es ≥ 0.
  - Entonces todos los siguientes casos particulares son ≥ 0:
    - Si $c_1 = 1$ y $c_2 = 0 \Rightarrow V[c'\hat{\beta}] = V[\hat{\beta}_1] \geq 0$
    - Si $c_1 = 0$ y $c_2 = 1 \Rightarrow V[c'\hat{\beta}] = V[\hat{\beta}_2] \geq 0$
    - Si $c_1 = 1$ y $c_2 = 1 \Rightarrow V[c'\hat{\beta}] = V[\hat{\beta}_1] + V[\hat{\beta}_2] + 2Cov[\hat{\beta}_1, \hat{\beta}_2] \geq 0$
    - Si $c_1 = 1$ y $c_2 = -1 \Rightarrow V[c'\hat{\beta}] = V[\hat{\beta}_1] + V[\hat{\beta}_2] - 2Cov[\hat{\beta}_1, \hat{\beta}_2] \geq 0$

Teorema de Gauss-Markov (TGM)

- Bajo todos los supuestos clásicos, el estimador de MCO es el más eficiente de todos los estimadores lineales e insesgados.
- **Mejor Estimador Lineal e Insesgado (MELI).**
- **Además**: sea c un vector de K constantes arbitrarias $c_1, c_2, ...c_K$, $c'\hat{\beta}$ es el mejor estimador lineal e insesgado de $c'\beta$.
  - Es decir, la combinación lineal de los estimadores MCO es MELI para estimar la combinación lineal de los parámetros.

El enunciado formal del TGM

En el modelo lineal general $Y = X\beta + u$, si $\hat{\beta}$ es el vector de estimadores MCO del vector de parámetros β y $\tilde{\beta}$ es cualquier vector de estimadores lineales e insesgados de β, entonces, bajo los supuestos clásicos, $VarCov(\tilde{\beta}) - VarCov(\hat{\beta})$ es una matriz semidefinida positiva.

- Notar que decir que $VarCov(\tilde{\beta}) - VarCov(\hat{\beta})$ es una matriz semidefinida positiva es lo mismo que decir que la varianza de cualquiera de los $\hat{\beta}_k$ con $k = 1, 2, ...K$ (o de cualquier combinación lineal de los $\hat{\beta}_k$) es menor o igual que la varianza de cualquiera de los $\tilde{\beta}_k$ (o de la correspondiente combinación lineal).
  - Chequear esto para $K = 2$
  - ¿Por qué decimos "menor o igual"? ¿en qué caso serían iguales?

Demostración del TGM

- En https://www.youtube.com/watch?v=8HPbCEqIv8o&t=6s está disponible la explicación de esta demostración.
- Supongamos el modelo lineal, donde μ cumple con todos los supuestos clásicos.
- Consideremos un estimador $\tilde{\beta}$ que cumpla con:
  1. **Lineal:** existe una matriz $A_{K\times n}$, que depende solo de las X, tal que $\tilde{\beta} = AY$.
  2. **Insesgado:** $E(\tilde{\beta}) = \beta$
- En este contexto, el requisito de linealidad hace que:

$$E(\tilde{\beta}) = E(AY) = E(AX\beta + A\mu) = AX\beta$$

- ¿Qué requisito debe cumplir A para que $\tilde{\beta}$ además de lineal también sea insesgado?
  - tiene que ser cierto que $AX = I_k$

Demostración del TGM

- Tenemos que demostrar que $VarCov(\tilde{\beta}) - VarCov(\hat{\beta})$ es semidefinida positiva.
- Definamos $\hat{\gamma} = \tilde{\beta} - \hat{\beta}$
  - entonces, trivialmente tenemos $\tilde{\beta} = \hat{\beta} + \hat{\gamma}$
- Noten que:

$$VarCov(\tilde{\beta}) = VarCov(\hat{\beta}) + VarCov(\hat{\gamma}) + 2COV(\hat{\beta}, \hat{\gamma})$$

- Importante: si mostramos que $COV(\hat{\beta}, \hat{\gamma}) = 0$ probaremos el teorema (¿por que?).

Demostración del TGM

- Computemos la covarianza:

$$\begin{aligned} COV(\hat{\beta},\hat{\gamma}) &= E\left[(\hat{\beta}-E(\hat{\beta}))(\hat{\gamma}-E(\hat{\gamma}))'\right] \\ &= E\left[(\hat{\beta}-\beta)\hat{\gamma}'\right] \qquad \textit{¿por qué?} \end{aligned}$$

- Noten que:

$$\begin{aligned} \hat{\gamma} &= \tilde{\beta}-\hat{\beta} \\ &= AY-(X'X)^{-1}X'Y \\ &= \left[A-(X'X)^{-1}X'\right]Y \\ &= \left[A-(X'X)^{-1}X'\right](X\beta+\mu) \\ &= \left[A-(X'X)^{-1}X'\right]\mu \qquad \textit{¿por qué?} \end{aligned}$$

Demostración del TGM

- Recordando que $\hat{\beta} - \beta = (X'X)^{-1}X'\mu$, la covarianza queda:

$$
\begin{aligned}
COV(\hat{\beta}, \hat{\gamma}) &= E\left[(\hat{\beta} - \beta)\hat{\gamma}'\right] \\
&= E\left[(X'X)^{-1}X'\mu\mu'(A - (X'X)^{-1}X')'\right] \\
&= E\left[(X'X)^{-1}X'\mu\mu'(A' - X(X'X)^{-1})\right] \\
&= (X'X)^{-1}X'E(\mu\mu')\left[A' - X(X'X)^{-1}\right] \\
&= \sigma^2\left[(X'X)^{-1}X'A' - (X'X)^{-1}X'X(X'X)^{-1}\right] \\
&= 0
\end{aligned}
$$

- Por lo tanto, $VarCov(\tilde{\beta}) - VarCov(\hat{\beta}) = VarCov(\hat{\gamma})$, que por definición es semidefinida positiva.

El TGM en la práctica

- Bajo ciertas condiciones (¿cuales?), el TGM permite saber que el estimador MCO siempre le va a ganar en eficiencia a cualquier otro estimador lineal e insesgado.
- Si queremos comparar al estimador MCO con otro estimador, solo debemos chequear que se cumplan las hipótesis del teorema, es decir:
  - Se deben cumplir todos los supuestos clásicos (son condiciones necesarias y suficientes)
  - El otro estimador debe ser lineal e insesgado para β.
- De lo contrario, deberíamos comparar caso por caso.

Inferencia en el modelo lineal general

Supuesto adicional: normalidad

$$\mu \sim N(0, \sigma^2 I_n)$$

- **Resultado:** $\hat{\beta}$ tiene distribución normal multivariada (¿por qué?)

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Test de hipótesis simples

- En forma general, consideremos una prueba con la siguiente hipótesis nula:

$$H_0: \quad c'\beta - r = 0$$

  - donde c es algún vector de K constantes y r es algún escalar.
  - Notar que $H_0: \quad \sum_{j=1}^{K} c_j \beta_j - r = 0$.
  - Dependiendo de c y r, podemos considerar distintos casos particulares para H_0:
    1. Significatividad individual: $H_0: \quad \beta_j = 0$
    2. Valores particulares: $H_0: \quad \beta_j = r$
    3. Igualdad de dos coeficientes: $H_0: \quad \beta_j = \beta_h$
    4. Otros: sumas y diferencias de coeficientes, etc.

Test de hipótesis simples

- Dado que $\hat{\beta} \sim N\left(\beta, VarCov(\hat{\beta})\right)$, entonces:

$$c'\hat{\beta} - r \sim N\left(E(c'\hat{\beta} - r), V(c'\hat{\beta} - r)\right)$$

- Calculemos la esperanza y la varianza de esta combinación lineal:

$$\begin{aligned} E\left[c'\hat{\beta} - r\right] &= c'\beta - r \\ V\left[c'\hat{\beta} - r\right] &= V\left[c'\hat{\beta}\right] = c'VarCov[\hat{\beta}]c \end{aligned}$$

- Luego:

$$c'\hat{\beta} - r \sim N\left(c'\beta - r, \sigma^2 c'(X'X)^{-1}c\right)$$

Test de hipótesis simples

- Bajo la $H_0 : c'\beta - r = 0$ se cumple que:

$$\frac{c'\hat{\beta} - r}{\sqrt{\sigma^2 c'(X'X)^{-1}c}} \sim N(0,1)$$

- En la práctica, se usa S^2 en lugar de σ^2, entonces:

$$\hat{V}\left[c'\hat{\beta} - r\right] = c'\hat{VarCov}\left[\hat{\beta}\right]c = S^2 c'(X'X)^{-1}c$$

- Luego, bajo $H_0 : c'\beta - r = 0$ el estadístico de prueba es:

$$\frac{c'\hat{\beta} - r}{\sqrt{\hat{V}\left[c'\hat{\beta} - r\right]}} = \frac{c'\hat{\beta} - r}{\sqrt{S^2 c'(X'X)^{-1}c}} \sim T_{n-K}$$

- Hay que recurrir a la tabla T_{n-K} para obtener los valores críticos y/o p-valor.

- Para evaluar $H_0 : c'\beta - r = 0$ usamos el estadístico

$$\frac{c'\hat{\beta} - r}{\sqrt{\hat{V}\left[c'\hat{\beta} - r\right]}} \overset{H_0}{\sim} T_{n-K}$$

  - donde $\hat{V}(c'\hat{\beta} - r) = S^2 c'(X'X)^{-1}c$. El vector de constantes c selecciona ciertas filas y columnas de $(X'X)^{-1}$
    - llamemos A_{jk} al elemento genérico de la fila j y columna k de la matriz $(X'X)^{-1}$
    - cada elemento de la diagonal es $\hat{V}[\hat{\beta}_k] = S^2 A_{kk}$, $k = 1, ...K$ y cada elemento fuera de la diagonal es $\hat{Cov}[\hat{\beta}_j, \hat{\beta}_k] = S^2 A_{jk}$, $j \neq k$

Casos particulares de hipótesis simples

- **Caso 1: test de significatividad individual**
  Para evaluar $H_0 : \beta_k = 0$ usamos el estadístico
  $$\frac{\hat{\beta}_k}{\sqrt{\hat{V}[\hat{\beta}_k]}} = \frac{\hat{\beta}_k}{\sqrt{S^2 A_{kk}}} \overset{H_0}{\sim} T_{n-K}$$
- **Caso 2: igualdad de 2 coeficientes**
  Para evaluar $H_0 : \beta_k = \beta_l$ usamos el estadístico
  $$\frac{\hat{\beta}_k - \hat{\beta}_l}{\sqrt{\hat{V}[\hat{\beta}_k - \hat{\beta}_l]}} = \frac{\hat{\beta}_k - \hat{\beta}_l}{\sqrt{S^2[A_{kk} + A_{ll} - 2A_{kl}]}} \overset{H_0}{\sim} T_{n-K}$$
- **Caso 3: combinaciones lineales entre coeficientes**
  Para evaluar $H_0 : \beta_k + 40\beta_l = 0{,}04$ usamos el estadístico
  $$\frac{\hat{\beta}_k + 40\hat{\beta}_l - 0{,}04}{\sqrt{\hat{V}[\hat{\beta}_k + 40\hat{\beta}_l - 0{,}04]}} = \frac{\hat{\beta}_k + 40\hat{\beta}_l - 0{,}04}{\sqrt{S^2[A_{kk} + 40^2 A_{ll} + 2 \times 40 A_{kl}]}} \overset{H_0}{\sim} T_{n-K}$$

Test de significatividad global

- En el modelo de K variables,

$$Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_K X_{Ki} + \mu_i$$

- Consideremos las siguientes hipótesis:

$$\begin{aligned} H_0: &\quad \beta_2 = 0 \wedge \beta_3 = 0 \wedge ... \wedge \beta_K = 0 \\ H_1: &\quad \beta_2 \neq 0 \vee \beta_3 \neq 0 \vee ... \vee \beta_K \neq 0 \end{aligned}$$

- Para este caso, el estadístico es:

$$F = \frac{SCE/(K-1)}{SRC/(n-K)} \overset{H_0}{\sim} F_{(K-1,n-K)}$$

- ¿Qué valores esperarían para F si la H_0 fuese verdadera?
  - Rechazamos H_0 cuando F supera el valor crítico: alguna/s de las variables es relevante para explicar a Y

Significatividad de un subconjunto de variables

- En el modelo de K variables,

$$Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_K X_{Ki} + \mu_i$$

- Consideremos las siguientes hipótesis:

$$\begin{aligned} H_0: &\quad \beta_2 = 0 \wedge \beta_3 = 0 \\ H_1: &\quad \beta_2 \neq 0 \vee \beta_3 \neq 0 \end{aligned}$$

- Pensemos que estamos contrastando dos modelos:
  1. *Modelo irrestricto*: incluye las K variables explicativas.
  2. *Modelo restricto*: considera como verdadera a la H_0 (excluye a X_2 y X_3)
     - En este caso, H_0 contiene 2 restricciones, a diferencia del caso anterior que teníamos $K-1$ restricciones.

Significatividad de un subconjunto de variables

- En general, sea q el número de restricciones de significativad en la H_0. Llamemos:
  1. SRC_I a la SRC del modelo con K variables (modelo irrestricto).
  2. SRC_R a la SRC de un modelo que excluye las q variables consideradas en la H_0 (modelo restringido).
- El estadístico de prueba para contrastar H_0 es:

$$F = \frac{(SRC_R - SRC_I)/q}{SRC_I/(n-K)} \overset{H_0}{\sim} F_{(q,n-K)}$$

- Si la H_0 fuese verdadera, ¿qué valores esperarían para F?
  - Rechazamos H_0 cuando F supera el valor crítico

Ejemplo: materias aprobadas y horas de estudio

- Muestra: 1961 estudiantes de la carrera Contador Público de la UBA.
- Referencia: Sosa Escudero, Giovagnoli y Porto (2007) "*The Effects of Individual Characteristics on the Distribution of College Performances*".

```
. describe ltotmat horas horas2 edad hombre educmp

              storage  display
variable name   type   format  variable label
-------------------------------------------------------------------------------
ltotmat         float  %9.0g   Log del total materias aprobadas desde el ingreso
horas           float  %9.0g   Horas dedicadas al estudio, por semana
horas2          float  %9.0g   Horas al cuadrado
edad            float  %9.0g   Edad, en años
hombre          float  %9.0g   1 si es hombre, 0 si es mujer
educmp          float  %9.0g   Máxima educación de los padres, en años
```

- Estimación del modelo por MCO

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       horas |    .023703    .003908     6.07   0.000     .0160387    .0313673
      horas2 |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
        edad |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
      hombre |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
      educmp |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
       _cons |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- Ni la **edad** ni el **género** son individualmente significativas para explicar a la performance académica (valor-p mayores a 0.10).

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       horas |    .023703    .003908     6.07   0.000     .0160387    .0313673
      horas2 |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
        edad |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
      hombre |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
      educmp |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
       _cons |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- Un año adicional en **educación de los padres** aumenta en promedio en un 2% la cantidad de materias aprobadas, manteniendo fijos todos los demás factores.

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       horas |    .023703    .003908     6.07   0.000     .0160387    .0313673
      horas2 |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
        edad |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
      hombre |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
      educmp |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
       _cons |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- Dado todo lo demás constante, la relación entre el logaritmo del total de materias aprobadas y las horas de estudio es **no lineal** (la variable `horas2` es estadísticamente significativa al 1%). ¿Cuál es el mínimo nivel de significatividad para el que sigue siendo cierta esta afirmación?

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       horas |    .023703    .003908     6.07   0.000     .0160387    .0313673
      horas2 |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
        edad |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
      hombre |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
      educmp |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
       _cons |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- Test de significatividad global: el estadístico $F_{(5,1955)}$ sale de computar el cociente entre $Model\ SS/(K-1)$ y $Residual\ SS/(n-K)$

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       horas |    .023703    .003908     6.07   0.000     .0160387    .0313673
      horas2 |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
        edad |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
      hombre |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
      educmp |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
       _cons |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- ¿Cómo evaluarían la hipótesis de que las horas de estudio no son relevantes para explicar el total de materias aprobadas?

```
. regress ltotmat horas horas2 edad hombre educmp

      Source |       SS       df       MS              Number of obs =    1961
-------------+------------------------------           F(  5,  1955) =   38.67
       Model |  59.6725542     5  11.9345108           Prob > F      =  0.0000
    Residual |  603.300837  1955  .308593778           R-squared     =  0.0900
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  662.973391  1960   .33825173           Root MSE      =  .55551

------------------------------------------------------------------------------
     ltotmat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
?    horas   |    .023703    .003908     6.07   0.000     .0160387    .0313673
?    horas2  |  -.0001848    .000066    -2.80   0.005    -.0003143   -.0000552
       edad  |  -.0063121   .0045028    -1.40   0.161    -.0151429    .0025186
     hombre  |  -.0299594    .025489    -1.18   0.240     -.079948    .0200291
     educmp  |   .0196455   .0031429     6.25   0.000     .0134818    .0258092
      _cons  |   1.864868   .1282148    14.54   0.000     1.613416     2.11632
------------------------------------------------------------------------------
```

- Si escribimos el modelo de la siguiente manera:

$$\begin{aligned} ltotmat_i = \quad & \beta_1 + \beta_2 horas_i + \beta_3 horas_i^2 + \beta_4 edad_i + \cdots \\ & \cdots + \beta_5 hombre_i + \beta_6 educmp_i + \mu_i \end{aligned}$$

- La hipótesis nula de que las horas de estudio no afectan el total de materias aprobadas es

$$\begin{aligned} H_0: \quad & \beta_2 = 0 \wedge \beta_3 = 0 \\ H_1: \quad & \beta_2 \neq 0 \vee \beta_3 \neq 0 \end{aligned}$$

- Para contrastar estas hipótesis deberíamos estimar el modelo bajo H_0, es decir, omitiendo *horas* y su cuadrado. De allí obtener la SRC_R (SRC del modelo restringido). Y luego computar el estadístico de prueba $F_{(2,1955)}$.

- Stata lo hace por nosotros:

```
. test (horas=0) (horas2=0)

 ( 1)  horas = 0
 ( 2)  horas2 = 0

       F(  2,  1955) =   64.77
            Prob > F =    0.0000
```

- ¿Cómo plantearían la hipótesis nula de que una hora más de estudio por semana no contribuye a aumentar el número de materias?

$$\begin{aligned} ltotmat_i = \quad & \beta_1 + \beta_2 horas_i + \beta_3 horas_i^2 + \beta_4 edad_i + \cdots \\ & \cdots + \beta_5 hombre_i + \beta_6 educmp_i + \mu_i \end{aligned}$$

- Recordar que el efecto marginal estimado de las horas de estudio por semana viene dado por:

$$\frac{\partial \widehat{ltotmat}}{\partial horas} = \hat{\beta}_2 + 2\hat{\beta}_3 horas$$

- Entonces el efecto marginal depende de las horas de estudio semanales. Para un alumno que estudia 60 horas por semana la hipótesis nula sería:

$$H_0: \quad \beta_2 + 2\beta_3 \times 60 = 0$$

- ¿Es una hipótesis simple o compuesta? ¿Qué tipo de test hay que aplicar?

Bibliografía para esta clase

- Notas de clase, capítulo 2 sección 2.7 (Teorema de Gauss-Markov) y sección 2.10.
- Wooldridge, capítulo 3.5 (Teorema de Gauss-Markov) y capítulo 4.