Inferencia

Modelo de Regresión Lineal por mínimos cuadrados ordinarios

1. Distribuciones muestrales de estimadores

Teorema de Gauss Markow

- **LINEALIDAD** (MLR1) → Relación lineal entre parámetros MLR
- **INDEPENDENCIA OBSERVACIONES** (MLR2) → Muestreo aleatorio $cov(u_i, u_j | X) = 0 \ \forall i \neq j$
- **COLINEALIDAD IMPERFECTA** (MLR3) → $X_{m \times n}$ de rango completo, $n \geq k+1$
  *Que no sean combinación lineal perfecta
- **MEDIA CONDICIONAL CERO** (MLR4) → $E(u | x_1, \dots x_k) = 0$, $E(u) = u'u = 0$, ORTOGONALIDAD
- **HOMOCEDASTICIDAD** (MLR5) → $Var(U | X_1, \dots, X_k) = \sigma^2 \cdot I_n$

$E(\hat{\beta} | X) = \beta$ **INSESGAMIENTO** + $Var(\hat{\beta} | X) = \sigma^2 I_n$ **EFICIENCIA** = **ESTIMADORES INSESGADOS DE VARIANZA MÍNIMA**

*No solo lineales

Modelo lineal clásico CML → **Normalidad** (MLR6) → $u \sim N(0, \sigma^2)$

- Nos hablan de momentos de la distribución muestral de los estimadores → Pero la distribución depende del término de error u

MLR 6

▶ NORMALIDAD MLR6
DE LOS ERRORES

El error poblacional u es independiente de los regresores $x_1, \dots, x_K$ y se distribuye en forma normal con media cero y varianza σ^2

$$u \sim N(0, \sigma^2)$$

- $E(u | x_1, \dots, x_K) = E(u) = 0$ RLM4
- $Var(u | x_1, \dots, x_K) = Var(u) = \sigma^2$ RLM5

→ Modelo lineal Clásico (CLM)
- MLR 1 : MLR 5 + MLR6
  Gauss Markow + $\underbrace{\text{Normalidad}}_{\text{f.d.p}}$

- MCO son estimadores insesgados de mínima varianza
  (no es necesario que sean lineales en y_j)

$$y | x_1, \dots, x_K = y | X \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K, \sigma^2)$$

$$\hat{\beta}_j | X \sim Normal(\beta_j, Var(\hat{\beta}_j | X))$$

estandarizando

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} | X = N(0,1)$$

Problemas
- Variables que no distribuyen normal: Ingresos, N° libros, N° educacional
  Pero en análisis asintótico se soluciona (Muestras grandes)

Demostración

$$\hat{\beta}_j = \beta_j + \underbrace{\overbrace{\sum \frac{\hat{u}_{ij}}{\sum \hat{u}_{ij}^2}}^{\text{Constante}}}_{\text{independiente de } X} \cdot u_i$$

$$\hat{\beta}_j = \beta_j + c \cdot u_i \implies \hat{\beta}_j \text{ es combinación lineal de errores } \mu_i$$

→ MLR6 + (y MLR2), los errores son V.A iid Normal $(0, \sigma^2)$

→ La suma de V.A iid ~ Normal, también se distribuye normal

→ $\hat{\beta} | X$ también distribuye normal con media β_j y $Var(\beta_j | x)$

2. Prueba de hipótesis

- Nunca sabemos los valores exactos de β_j → PRUEBAS DE HIPÓTESIS
  - T-student (único parámetro)
  - F (múltiples parámetros)

T STUDENT – ÚNICO PARÁMETRO

- $$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} | X \sim t_{n-k-1} = t_{gl}$$
  se: σ^2 desconocido, $\hat{\sigma}^2$ muestral

test dos colas con $H_0: \beta_j = 0$

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

$\hat{\beta}_j$ → coeficientes STATA
$se(\hat{\beta}_j)$ → std. error en STATA

- $H_0: \beta_j = 0$
  $H_1: \beta_j \neq 0$
  } El efecto parcial de x_j sobre y, después de controlar por el resto de X

→ Hipótesis es acerca de parámetros poblacionales

$|t_{\hat{\beta}}|$ muy grande
- LEFT TAILED (muy negativas)
- RIGHT TAILED (muy positivas)

t_c = en el percentil $\underbrace{1-\alpha}_{\text{NIVEL SIG}}$ en $\underbrace{n-k-1}_{\text{GL}}$

$t_{\hat{\beta}}$ compara t_c → zona de rechazo a un nivel de confianza

- p-value

$$P(|T| > |t|)$$

- Fuerza o debilidad de evidencia empírica sobre H_0
- p-value se compara nivel de confianza (mínimo)
  - altamente significativa 0.01
  - significativa 0.05
  - marginalmente significativa 0.1

STATA
1. reg
2. eststo reg
3. esstab

- Intervalos de confianza

$$\hat{\beta}_j \pm t_c \cdot se(\hat{\beta})$$

} En sucesivas muestras aleatorias, computando $\hat{\beta}_j \cdot \hat{\beta}_j$, el valor poblacional β_j se encontraría en el intervalo $(\hat{\beta}, \hat{\beta})$ en un 95% muestras

STATA

- t_c = percentil 97.5 en distribución t_{n-k-1}
- Con un 95% confianza sabemos si el estimador puntual está en el intervalo

ECONÓMICAMENTE → $\frac{\log}{\log}$ = ELASTICIDADES → log(wage) : (como %)

PRUEBA F

- Prueba para un grupo de variables (q) que pertenecen a K

Modelo no restringido

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + u$$

Modelo restringido

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{K-q} x_{K-q} + u$$

HIPÓTESIS

- Probaremos que q de estas K variables **no** tienen un efecto parcial en y

$$H_0 : \beta_{K-q+1} = 0, \dots, \beta_K = 0$$

q restricciones de exclusión

H_1 : que H_0 no es verdadero

restricción es más mejor que 6.

STATA
test var1 var2 var3

- $SSR_{NR} \leq SSR_R$

En MCO, al agregar más variables $\uparrow R^2$ (o $\downarrow SSR$)

¿Es estadísticamente significativa la diferencia?

$$F \equiv \frac{SSR_r - SSR_{NR}}{SSR_{NR}/(n-K-1)}$$

- Rechazamos H_0

$F_{obs} > F_{teorico}$

Parámetros
- q
- $gl = n-K-1$
- IC

→ q (en conjunto) son estadísticamente significativas sobre y

- Contextos de uso de F
  - Alta correlación de predictores
    - Si x_1 e x_2 ↑ correlación, ↑ $sd(\beta)$, ↓ t.
    - EJ: x_1 educación padre, x_2 educación madre, y = rendimiento académico

Variables que comparten varianza conjunta para explicar y

- Chequear corr entre variables
- Multicolinealidad es menos relevante en pruebas conjuntas (F)
- F útil para pruebas de exclusión conjunta

Relación F con $t_{student}$

Para un solo parámetro

t^2_{n-K-1} tiene una distribución $F_{(1,n-K-1)}$

OJO F no se utiliza siempre
- t + flexible para pruebas de un parámetro
- F (-) poder para determinar H_0.
  Rechazar H_0 cuando es falsa → Poder = 1 − P(Error tipo II | H_0)
  z. rechazo pequeña
  error de no rechazar H_0 cuando H_0 es falsa
  * Lo contrario a nivel de significancia
- F nos habla de conjuntos. Algunos parámetros pueden no ser sig. individualmente, pero sí en conjunto

Relación entre R y F

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad SSR = \sum_{i=1}^{n} (\hat{y} - y_i)^2 = \sum_{i=1}^{n} \hat{u}_i^2 \qquad R^2 = \frac{SSE}{SST} = 1 - SSR$$

$$SSR_r = SST(1 - R^2_r) \qquad SSR_{NR} = SST(1 - R^2_{nr})$$

$$F = \frac{(SSR_r - SSR_{nr})/SST}{SSR_r/SST} \times \frac{(n-K-1)}{q} \rightarrow \frac{(R^2_{NR} - R^2_R)}{(1 - R^2_R)} \times \frac{(n-K-1)}{q}$$

F-ESTADÍSTICO GLOBAL

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_K = 0$$

- Modelo r: $y = \beta_0 + u$
- $R^2 = 0$ → nada de la variación de y es explicada pues no tenemos predictores

$$F = \frac{R^2_{NR}}{(1 - R^2_{NR})} \times \frac{(n-K-1)}{K}$$

* Nota que no es q.

SIGNIFICANCIA CONJUNTA DE TODOS LOS REGRESORES

EN STATA

$$H_0 : \beta_1 = 1, \ \beta_2 = 0, \ \beta_3 = 0$$

STATA
reg var1 var2 var3
test var2 var3 (los no significativos)
test var = 1 (H_0)

test (var1=1) (var2=0) (var3=0)
si F pequeño, para no rechazar H_0

3. INFERENCIA EN TÉRMINOS MATRICIALES

- Testear múltiples hipótesis acerca de parámetros

1. $H_0: R\beta = r$ vs $H_1: R\beta \neq r$

R matriz $q \times (K+1)$

r vector $q \times 1$ con $q \leq K+1$ restricciones lineales

2. $$\begin{bmatrix} r_{1,0} & \cdots & r_{1,K} \\ \vdots & \ddots & \vdots \\ r_{q,0} & \cdots & r_{q,K} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} r_1 \\ \vdots \\ r_q \end{bmatrix}$$

- Hipótesis simple $H_0: \beta_1 = 1$
  $R = [0 \ 1 \ 0 \cdots 0]$, $r = 1$
- Hipótesis igualdad $H_0: \beta_1 = \beta_2$ $\quad \beta_1 - \beta_2 = 0$
  $R = [0 \ 1 \ -1 \cdots 0]$ $\quad r = 0$
- Hipótesis conjunta $H_0: \beta_0 = 2, \ \beta_1 = 1 - \beta_2$ $\quad \beta_1 + \beta_2 = 1$.
  $R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \end{bmatrix}$ $\quad r = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

3. TEST F

$$\hat{F} = (R\hat{\beta} - r)'(\hat{\sigma}^2 R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/q \sim F(q, n-K-1)$$

Demostración

E.2-E.4

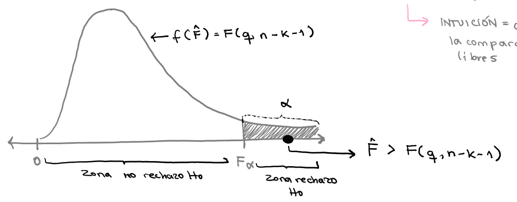
- Por MLR6 sabemos que la normalidad de errores, condicional en X, implica

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

3. TEST F

$$\hat{F} = (R\hat{\beta} - r)' \left(\hat{\sigma}^2 R(X'X)^{-1} R' \right)^{-1} (R\hat{\beta} - r) / q \sim F(q, n-k-1)$$

INTUICIÓN = ¿ $\hat{\beta}$? es el que nos da la comparación con parámetros libres



B. T-STUDENT

Recordemos que $t^2_{n-k-1} \sim F_{(1, n-k-1)}$ 1 parámetro

$$\frac{\hat{\beta}_1 - \hat{\beta}_2}{SE(\hat{\beta}_1 - \hat{\beta}_2)} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2C(\hat{\beta}_1, \hat{\beta}_2)}} \sim t_{n-k-1}$$

$H_0: \ \hat{\beta}_1 = \hat{\beta}_2 \quad o \quad \boxed{\hat{\beta}_1 - \hat{\beta}_2 = 0}$

DEMOSTRACIÓN

E.2 - E.4

- Por MLR6 sabemos que la normalidad de errores, condicional en X, implica

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

Entonces

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

$$(R\hat{\beta} - R\beta) \sim N(0, \sigma^2 R(X'X)^{-1} R')$$

Bajo Ho podemos demostrar que

$$(R\hat{\beta} - r)' [\sigma^2 R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r) \sim X^2(q)$$

Como σ^2 es desconocido, usamos su estimador y se obtiene (1) con distribución F de Fisher.