Teoría Econométrica I - EAE- 250-A

El modelo de regresión lineal

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Agosto - Septiembre 2021

Esperanza condicional

- La **esperanza condicional de Y (escalar) dado X (vector en $\Re^k$)** es un mapping o función escalar que se escribe:

$$\mathbb{E}(Y|X) : \Re^k \to \Re$$

- Representa el primer momento de la distribución condicional de Y en X
- Dado que X no es fijo, la esperanza condicional se convierte en una variable aleatoria que es función de X

  $\Rightarrow$ De hecho estamos refiriendo a $E(Y|X = X_0)$ $\forall X_0$ realizaciones de X

$$\mathbb{E}(Y|X) = \begin{cases} E(Y|X = X_1) & \text{con probabilidad} \quad P(X = X_1) \\ E(Y|X = X_2) & \text{con probabilidad} \quad P(X = X_2) \\ \vdots & \vdots \end{cases},$$

Con VA continuas

- **Ejemplo VA discretas:**
  - Y: rendimiento en el curso de Econometría I (4,0 ; 5,0; 5,5; 6,0)
  - X: altura (1,30; 1,50; 1,70; 1,90)
- La relación entre las distintas alturas y el rendimiento econometría puede o no entregar una relación **causal**
- Pero si tenemos la distribución conjunta podemos obtener la $E(NotaEcI|Altura)$
- **VA continuas:**
- La esperanza condicional se escribe simplemente de la siguiente manera si la variable aleatoria es continua:

$$\mathbb{E}(Y|X) = \int_{-\infty}^{\infty} \frac{yf(x,y)}{f(x)} dy$$

Otros momentos de interés

(1) **Efecto Parcial** de X_j sobre $\mathbb{E}(Y|X)$:

$$\frac{\partial \mathbb{E}_(Y|X)}{\partial X_j}$$

El efecto parcial es marginal si queremos efectos de cambios mayores, "ΔX_j"

$$\frac{\Delta \mathbb{E}_(Y|X)}{\Delta X_j} \approx \frac{\partial \mathbb{E}_(Y|X)}{\partial X_j} \Delta X_j$$

(2) **Elasticidad:** cambios porcentuales

$$\xi_{\mathbb{E}_(Y|X), X} = \frac{\partial \mathbb{E}_(Y|X)}{\partial X_j} \frac{X_j}{\mathbb{E}_(Y|X)}$$

- Esta expresión también puede ser escrita como:

$$\xi_{\mathbb{E}_(Y|X), X} = \frac{\partial \ln \mathbb{E}_(Y|X)}{\partial \ln X_j}$$

Otros objetos de interés

- Notar que:

$$\xi_{\mathbb{E}_(Y|X),X} = \frac{\partial \ln \mathbb{E}_(Y|X)}{\partial \ln X_j} \neq \frac{\partial \mathbb{E}_((\ln Y|X))}{\partial \ln X_j}$$

- **Salvo en este caso particular:**

(i) Supongamos un modelo expresado en logaritmo (su variable dependiente):

$$\ln(y) = g(x) + \mu$$

donde μ es independiente de X con media 0. Por simplicidad supongamos que x es escalar.

(ii) Tomamos exponenciales:

$$y = exp(g(x)) * exp(\mu)$$

(iii) Por lo tanto,

$$\mathbb{E}[y|x] = exp(g(x)) * \mathbb{E}(exp(\mu)|x)$$

La elasticidad queda:

$$\begin{aligned} \frac{\partial \mathbb{E}_(Y|X)}{\partial X_j} \frac{X_j}{\mathbb{E}_(Y|X)} \xi &= \cancel{exp(g(x))} g'(x) \cancel{\mathbb{E}(exp(\mu))} \frac{x}{\cancel{exp(g(x))} \cancel{\mathbb{E}(exp(\mu))}} \\ \xi &= g'(x)x \end{aligned}$$

Otros objetos de interés

- Ahora el modelo en logaritmos:

$$
\begin{aligned}
\mathbb{E}[\ln(y)|x] &= g(x) \\
\mathbb{E}[\ln(y)|x] &= g(\exp(\ln(x))) \\
\frac{\partial \mathbb{E}[\ln(y)|x]}{\partial \ln(x)} &= g'(x)\exp(\ln(x)) \\
\xi &= g'(x)x
\end{aligned}
$$

Otros objetos de interés

(3) **Semi Elasticidad:** Se define como

$$Semielasticidad = \frac{\partial \mathbb{E}(Y|X)}{\partial X_j} \frac{1}{\mathbb{E}(Y|X)}$$

- Un ejemplo de semielasticidad la podemos encontrar en el efecto marginal de la escolaridad en el ingreso, ya que la variable escolaridad está medida en niveles, y la de ingreso en logaritmos.

Ley de esperanzas iteradas

- **Ley de Esperanzas Iteradas:** En su versión más sencilla esta enuncia que:

$$\mathbb{E}_x(\mathbb{E}(y|x)) = \mathbb{E}(y)$$

- **Demostración** (en variable continua):

$$\begin{aligned}
\mathbb{E}_x(\mathbb{E}(y|x)) &= \int \left[\int y \frac{f(x,y)}{f(x)} dy \right] f(x)dx \\
\mathbb{E}_x(\mathbb{E}(y|x)) &= \int \int y \frac{f(x,y)}{f(x)} f(x)dxdy \\
\mathbb{E}_x(\mathbb{E}(y|x)) &= \int y \left[\int f(x,y)dx \right] dy \\
\mathbb{E}_x(\mathbb{E}(y|x)) &= \int yf(y)dy \text{ porque } \int f(x,y)dx = f(y) \\
\mathbb{E}_x(\mathbb{E}(y|x)) &= \mathbb{E}_(y)
\end{aligned}$$

Notación

- Para una muestra de n observaciones el modelo (poblacional) es:

$$Y = X\beta + U$$

con

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k-1} \\ \vdots & \ddots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk-1} \end{pmatrix},$$

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

- k denota al número de variables explicativas (incluyendo la constante)

Supuestos

1. $\mathbb{E}(Y|X) = X\beta$ **Linealidad en los parámetros**
2. $V(Y) = \mathbb{E}([(y - \mathbb{E}(y))(y - \mathbb{E}(y))'] = \sigma^2 I$ **Homocedasticidad y no autocorrelación** en el término de error
3. Observaciones provenientes de una **muestra aleatoria**

   o Regresores eterminísticos (No estocásticos) $\Rightarrow$ muy de momento!!

   En este ultimo caso: $\mathbb{E}(Y|X) = \mathbb{E}(Y) = X\beta$
4. **Condición de Identificación** Matriz X de Rango Completo(mínimo de columnas o filas linealmente independiente). No puede haber multicolinealidad perfecta. En otras palabras, que $X'X$ sea invertible.
5. **Independencia /Ortogonalidad**: $\mathbb{E}(U|X) = 0$

Paradigma Estadístico

- El paradigma estadístico dice que una variable aleatoria la podemos descomponer en su esperanza condicional más un residuo.
- Este residuo cumple una propiedad conocida:

$$y = \mathbb{E}[y|x] + \varepsilon$$

  donde $\mathbb{E}(\varepsilon|x) = 0$.
- Luego los supuestos los podemos expresar en función de ε

$$\begin{aligned} \varepsilon &= Y - \mathbb{E}(Y|X) \\ \varepsilon &= Y - X\beta \end{aligned}$$

- En esta notación los supuestos son:
  1. Linearidad en los parámetros
  2. Esperanza lineal: $\mathbb{E}(\varepsilon|X) = \mathbb{E}|\mathbb{X} = 0$
  3. $V(\varepsilon) = \sigma^2 I_n$
  4. Muestra Aleatoria (vs. regresores deterministicos)
  5. Rango completo de $X(k)$

Derivación del estimador de MCO/OLS

- El estimador de MCO es aquel que minimiza la suma de los errores al cuadrado;

$$\underset{\beta}{\operatorname{argmin}}\, S = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$

- Se podría haber elegido otra métrica!

$\Rightarrow$ Ejemplo:

$$\underset{\beta}{\operatorname{argmin}}\, S^* = \sum_{i=1}^{n} |\varepsilon_i|$$

$\Rightarrow$ Se conoce como Least Absolute Deviations (LAD)

$\Rightarrow$ Pasa por las medianas y es insensible a ouliers, pero propiedades difíciles de demostrar

Optimización

- Recuerde que escribimos el criterio de minimización y lo derivamos con respecto a β

$$S = (Y - X\beta)'(Y - X\beta) = Y'Y - 2Y'X\beta + \beta'X'X\beta$$

$\Rightarrow$ Usamos las reglas de derivación que puede ver en el apéndice del Wooldridge/Greene

- Si A es una matriz simétrica y **x** un vector y los productos A**x** y **x**'A**x** son conformables, tenemos que:

$$\frac{\partial \mathrm{A}\mathbf{x}}{\partial \mathbf{x}} = \mathrm{A}'$$

$$\frac{\partial \mathbf{x}'\mathrm{A}\mathbf{x}}{\partial \mathbf{x}} = (A' + A)\mathbf{x} = 2\mathrm{A}\mathbf{x}$$

Optimización

- Aplicamos cálculo diferencial matricial a S tenemos la ecuación normal:

$$
\begin{aligned}
\mathbf{0} &= \mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} x_i' x_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} x_i' yi\right).
\end{aligned}
$$

- Donde x_i es un vector de $1 \times k$ con la i-ésima observación para los k regresores.
- Estas son las dos representaciones típicas del estimador OLS de β.

Derivación desde la estadística

- Es posible encontrar un estimador por el Principio de la Analogía o Método de Momentos explotando el supuesto de identificación $\mathbb{E}(\epsilon|x) = 0$

$$y = \mathbf{x}\beta + \epsilon$$

  con $\mathbf{x}$ vector de $1 \times k$ (una sola observación)
- El supuesto de identificación de MCO implica que

$$\mathbb{E}(\mathbf{x}'\epsilon|\mathbf{x}) = 0$$

- β puede ser expresado en momentos de las variables observables:

$$\mathbf{x}'y = \mathbf{x}'\mathbf{x}\beta + \mathbf{x}'\epsilon$$

  tomando valor esperado tenemos que:

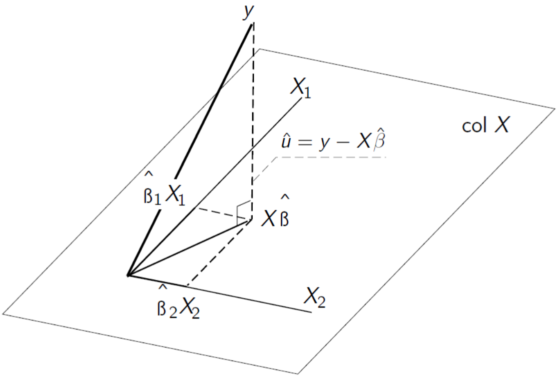
$$\beta = \mathbb{E}(\mathbf{x}'\mathbf{x}|\mathbf{x})^{-1}\mathbb{E}(\mathbf{x}'y|\mathbf{x})$$

- El *analogy principle* (Goldberger (1968), Manski (1988)) implica

$$\hat{\beta}_{MM} = \hat{\beta}_{MCO} = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'\mathbf{x}_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'y_i\right] \equiv (X'X)^{-1}X'Y$$

Interpretación geométrica del estimador MCO

- Se define $P = X(X^\top X)^{-1}X^\top$, matriz de proyección en el espacio generado por las columnas de X.
- Se define $M = I - P$, matriz de proyección en el espacio ortogonal al espacio generado por las columnas de X.
- $\hat{Y} = X\hat{\beta} = X(X^\top X)^{-1}X^\top Y = PY$
- $\hat{U} = Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X^\top X)^{-1}X^\top Y = Y - PY = MY$
- $Y = \hat{Y} + \hat{U} = PY + MY$ es la descomposición del vector y en dos espacios ortogonales.
- Las matrices P y M verifican dos propiedades:
    - Son simétricas: $P^\top = P, M^\top = M$
    - Son idempotentes: $P^2 = P, M^2 = M$



y
X_1
col X
$\hat{u} = y - X\hat{\beta}$
$\hat{\beta}_1 X_1$
$X\hat{\beta}$
$\hat{\beta}_2 X_2$
X_2

Bondad de Ajuste

- Si la primera columna de **X** es una vector columna "ι" igual a 1, se tiene una medida resumen para la "bondad de ajuste" de los valores predichos de la siguiente identidad:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

$$\begin{aligned} SST &= SSR + SSE \\ 1 &= \frac{SSR}{SST} + \frac{SSE}{SST} \end{aligned}$$

- Donde
  - SST: suma la variación total del modelo (suma cuadrados totales)
  - SSR: suma de los residuos al cuadrado
  - SSE: Suma explicada por el modelo

  La medida de bondad de ajuste es

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

En términos matriciales

- Matricialmente,

$$
\begin{aligned}
R^2 &\equiv 1 - \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{(\mathbf{y} - \overline{y}\iota)'(\mathbf{y} - \overline{y}\iota)} \\
&= 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{(\mathbf{y} - \overline{y}\iota)'(\mathbf{y} - \overline{y}\iota)}
\end{aligned}
$$

- donde $\overline{y}$ es el promedio muestral (escalar) de la variable dependiente,

$$
\overline{y} \equiv \frac{1}{N}\sum_i Y_i
$$

- y ι es un vector de unos de dimensión N.

Sobre el R cuadrado:

1. El coeficiente de determinación es siempre menor a 1. Ello porque $SSR \leq SST$ y por lo tanto $\frac{SSR}{SST} \leq 1$.
2. El análisis de varianza anterior fue derivado bajo el supuesto que el modelo incluía una constante. También tenemos $R^2 \geq 0$.
3. Al agregar regresores al modelo, el R^2 nunca decrecerá (se mantendrá constante o aumentará).
4. No es claro cuan bueno sea como predictor de ajuste.

También está el R^2 ajustado :

- El R^2 ajustado, $\bar{R}^2$ se define:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k)}{SST/(n-1)}$$

Regresión Particionada

- Al particionar la matriz de regresores podemos escribir X como la *concatenazión horizontal* de dos sub-matrices

$$X \equiv [X_1 \; X_2]$$

- Donde X_1 es de $n \times k_1$ y X_2 de $n \times k_2$ con $k_1 + k_2 = k$ lo que, junto con una versión particionada del estimador OLS

$$\hat{\beta} \equiv \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

- Así, el modelo de regresión lineal se puede escribir de la siguiente manera

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

Regresión Particionada

- Esta notación hace posible que podamos derivar una relación entre el subvector $\hat{\beta}_1$ de la regresión "larga" (de y sobre X_1 y X_2) y los coeficientes de la regresión "corta" (de y sobre X_1),

$$\hat{\beta}_1^* \equiv (X_1'X_1)^{-1}X_1'Y$$

- donde $\hat{\beta}_1^*$ solamente usa la submatriz X_1 de regresores.

$$\hat{\beta}_1^* = \hat{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2$$

- donde la relación anterior viene de reemplazar $Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\epsilon}$ en el coeficiente de la regresión "corta".
- Si $X_1'X_2 \neq 0$ y $\hat{\beta}_2 \neq 0$ el término $(X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2$ corresponde al *sesgo por omisión de variables relevantes*

Teorema de Frisch-Waugh-Lovell

- Sea P_i la matrix de proyección en el subespacio generado por X_i , y $M_i = I - P_i$ la matrix que proyecta en el subespacio ortogonal al generado por X_i

$$P_i \equiv X_i(X_i'X_i)^{-1}X_i' \; ; M_i \equiv I - (X_i(X_i'X_i)^{-1}X_i')$$

- **Teorema FWL:** Los coeficientes de la regresión larga pueden ser escritos como:

$$\hat{\beta}_1 = (X_1'M_2'M_2X_1)^{-1}X_1'M_2'M_2Y$$

- Nos dice que podemos hallar los estimadores en dos etapas (regresando X_1 en X_2 y Y en X_2)
- Solían considerarse separadas la constante y una tendencia deterministica y la interpretación era la de Y sin tendencia temporal
- Utilidad cuando invertir matrices mu grandes era un problema computacional importante

Momentos de los estimadores OLS

- Las reglas para el cálculo de la media (vector) y matriz de varianzas y covarianzas de una función lineal **Ay** de un vector aleatorio **y** (con **A** no estocástica) son:

$$\mathbb{E}[\mathbf{Ay}] = \mathbf{A}\mathbb{E}[\mathbf{y}]$$
$$V[\mathbf{Ay}] = \mathbf{A}V[\mathbf{y}]\mathbf{A}'$$

- $\Rightarrow$ Recordamos: cuando condicionamos en una V.A, podemos tratar esa V.A como "no estocástica" (en esa condición)
- Aplicando estas reglas al estimador OLS $\hat{\beta}$, bajo los supuestos estándar, tenemos que

$$\begin{aligned}\mathbb{E}[\hat{\beta}|X] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{Y}|X] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= \beta\end{aligned}$$

- esto significa que el estimador OLS $\hat{\beta}$ es insesgado.

Varianza del estimador OLS

- Por otro lado, tenemos que

$$\begin{aligned} V[\hat{\beta}|X] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' V[\mathbf{Y}|X]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\sigma^2\mathbf{I}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- De forma similar es posible demostrar que s^2 es un estimador insesgado de σ^2, siendo s^2

$$s^2 \equiv \frac{1}{N-K}\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)$$

- Tarea: demostrar que $\mathbb{E}[s^2] = \sigma^2$. Hint: La demostración implica intercambiar el operador traza y esperanza convenientemente además del uso de propiedades como $tr(AB) = tr(BA)$, etc.

Gauss–Markov

- Bajo los supuestos vistos, los estimadores OLS son MELI o (BLUE)
- El término Mejor hace referencia a el más eficiente, de menor varianza
- Esto lo demuestra el teorema de Gauss–Markov:

1. Sea $\tilde{\beta} = m + MY$ un estimador lineal de Y
2. Buscamos entre los estimadores insesgados, entonces:$\mathbb{E}(\tilde{\beta}) = \beta$
3. Entonces

$$\begin{aligned}\mathbb{E}(\tilde{\beta}|X) &= m + M\mathbb{E}(Y|X) \\ &= M\mathbb{E}(X\beta + \varepsilon|X) \\ &= m + MX\beta\end{aligned}$$

4. : Esto implica que:

$$m = 0 \text{ y que } MX = I_k$$

$\Rightarrow$ Notar que pasa cuando definimos $M = (X'X)^{-1}X'$

5. WLOG, podemos definir $M = (X'X)^{-1}X' + C$

Gauss–Markov

6. Tenemos entonces:

$$\begin{aligned} MX &= I_k \\ &\Rightarrow ((X'X)^{-1}X' + C)X = I_k \\ &\Rightarrow I_k + CX = I_k \\ &\Rightarrow CX = 0 \end{aligned}$$

7. Podemos escribir $\tilde{\beta}$ en términos de β para computar luego su varianza:

$$\begin{aligned} \tilde{\beta} &= MY \\ &= M(X\beta + \varepsilon) \\ &\Rightarrow \tilde{\beta} - \beta = M\varepsilon \end{aligned}$$

8. Entonces computamos la varianza condicional en X

$$\begin{aligned} Var(\tilde{\beta}|X) &= \mathbb{E}((\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'|X) \\ &= \mathbb{E}(M\varepsilon(M\varepsilon)'|X) \\ &= \mathbb{E}(M\varepsilon\varepsilon'M'|X) \end{aligned}$$

Gauss–Markov

Seguimos...

$$\begin{aligned} Var(\tilde{\beta}|X) &= M\mathbb{E}(\varepsilon\varepsilon'|X)M' \\ &= M\sigma^2 I_n M' \\ &= \sigma^2 MM' \end{aligned}$$

9. Descomponemos MM'

$$\begin{aligned} MM' &= ((X'X)^{-1}X' + C)((X'X)^{-1}X' + C)' \\ &= (X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'C' + CX(X'X)^{-1} + CC' \\ &= (X'X)^{-1} + CC' \end{aligned}$$

10. Y CC' es una matriz de productos cruzados por tanto semidefinida positiva

El Modelo Normal de Regresión Lineal

- Hasta aquí no hemos asumido una distribución para el término de error
- La ventaja de asumir normalidad es que podemos hacer inferencia exacta, es decir, tendremos estadísticos cuya distribución no depende de aproximaciones asintóticas
- Van a tener“buenas propiedades” incluso en muestra finita
- Para lograr esto, debemos aumentar en un supuesto el modelo anterior. Este modelo fue propuesto por Gauss en 1809 varios años después de que Legendre propusiera el estimador de MCO.
- **Supuesto 6:** **Normalidad**. El vector Y (o equivalentemente $\epsilon = Y - X\beta$) tiene una distribución normal multivariada.
  Luego, bajo los supuestos 1-6, la distribución del vector y es

$$y \sim \mathcal{N}(X\beta, \sigma^2 I)$$

y

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

Estimador OLS también es el estimador de máxima verosimilitud

- Bajo los supuestos anteriores Gauss en 1809 derivó el estimador MCO del modelo como un estimador de Máxima Verosimilitud para este modelo.
- Es fácil darse cuenta que

$$L = f(y_1, y_2, \ldots, y_n; X, \sigma^2, \beta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp^{-\frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}}$$

con lo cual, el estimador $\hat{\beta}_{MV}$ se obtiene maximizando:

$$\begin{aligned}\max_{\beta,\sigma^2} \ln(L) &= \max_{\beta,\sigma^2} \ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp^{-\frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}}\right) \\ &= \max_{\beta,\sigma^2}\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}\right) \quad (1)\end{aligned}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\hat{\sigma}^2} X'(Y - X\hat{\beta}) = 0 \Longrightarrow \hat{\beta}_{\mathbf{MV}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- Luego, tenemos que $\hat{\beta}_{MV} = \hat{\beta}_{MCO} = \hat{\beta}$ tiene una distribución normal

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

Inferencia en el modelo normal de regresión

- El resultado anterior tiene implicancias directas. Cualquier combinación lineal de $\hat{\beta}$ tendrá una distribución normal. Así,

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

  donde $\theta = R\beta$

- El supuesto de normalidad también implica:

$$\frac{(N-K)s^2}{\sigma^2} \sim \chi^2_{N-K}$$

- Con los resultados anteriores podemos hacer inferencia exacta. Si R tiene sólo una fila (es una hipótesis lineal) tenemos que

$$\frac{\hat{\theta} - \theta}{\sqrt{s^2 R(X'X)^{-1}R'}} \sim t_{N-K}$$

Inferencia sobre más de un parámetro a la vez

- Si R tiene más de una fila, entonces $R(X'X)^{-1}R'$ no es escalar y podemos fabricar un test F. Si R tiene r filas, tenemos que

$$(\hat{\theta}-\theta)'[s^2R(X'X)^{-1}R']^{-1}(\hat{\theta}-\theta)/r \sim F_{r,N-K}$$

- Este resultado nos sirve para testear hipótesis del tipo $H_0 : R\beta = \theta_0$ y $H_a : R\beta \neq \theta_0$ donde sólo debemos reemplazar el parámetro θ por el hipotético θ_0

Significación conjunta de todos los parametros (menos la constante)

- Un caso especial en donde la matriz $\mathbf{X}$ esté particionada de la forma $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$ y donde $\mathbf{X}_1$ es un vector columna de unos ($\mathbf{X}_1 = \iota$)
- Con la correspondiente partición de β, permite testear la hipótesis nula $H_0 : \beta_2 = \mathbf{0}$ usando el estadístico R^2.
- Bajo los supuestos vistos anteriormente y la hipótesis nula recién enunciado tenemos:

$$\frac{N-K}{N-1} \frac{R^2}{1-R^2} \sim F_{K-1,N-K}$$

- En el caso del modelo de regresión lineal tenemos que existe una relación monotónica entre el R^2 y el estadístico F para testear que todos los coeficientes (menos el intercepto) son cero

Intervalos de confianza

- Tenemos que

$$\frac{\hat{\theta} - \theta}{\sqrt{s^2 R(X'X)^{-1} R'}} \sim t_{N-K}$$

- Entonces podemos construir un IC del $(1-\alpha)\%$, es decir de $\alpha\%$ de significancia para θ usando los valores de tabla de la distribución t:

$$1 - \alpha = Pr\left[t_{N-K}^{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{s^2 R(X'X)^{-1} R'}} \leq t_{N-K}^{1-\alpha/2}\right]$$

$$= Pr\left[-t_{N-K}^{1-\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{s^2 R(X'X)^{-1} R'}} \leq t_{N-K}^{1-\alpha/2}\right]$$

$$= Pr\left[\hat{\theta} - t_{N-K}^{1-\alpha/2}\sqrt{s^2 R(X'X)^{-1} R'} \leq \theta \leq \hat{\theta} + t_{N-K}^{1-\alpha/2}\sqrt{s^2 R(X'X)^{-1} R'}\right]$$

- Tendremos asi:

$$\theta \in \left[\hat{\theta} \pm t_{N-K}^{1-\alpha/2} \times \sqrt{s^2 R(X'X)^{-1} R'}\right]$$

IC para test sobre más de un parámetro

- También podemos invertir un test F.
- Tendremos "regiones de confianza" (un poco más complicado) y
- Para el caso que la matriz R tenga sólo 2 filas tiene una representación gráfica en R^2. Las regiones de confianza serán elipses. Para ello sólo debemos notar lo siguiente:

$$Pr[(\hat{\theta}-\theta)'[s^2R(X'X)^{-1}R']^{-1}(\hat{\theta}-\theta)/r \leq F_{r,N-K}^{1-\alpha}] = 1-\alpha$$

luego resolvemos para el argumento y nos quedará la ecuación de una elipse.

Ejemplo regiones de confianza

- **EJEMPLO:** Considere el modelo $y = \beta_1 x_1 + \beta_2 x_2 + u$. Se quiere testear simultáneamente que $\beta_1 = r_1$ y $\beta_2 = r_2$. Note que

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Si,

$$X'X = \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix}$$

desarrollando el argumento de la región de confianza llegamos a una ecuación de la elipse:

$$\frac{1}{s^2 r}\left[(\beta_1 - r_1)^2 \phi_1 + (\beta_2 - r_2)(\beta_1 - r_1)(\phi_2 + \phi_3) + (\beta_2 - r_2)^2 \phi_4 = F^{1-\alpha}_{r,n-k}\right]$$

Notación

Figure: Regiones de Confianza Elípticas

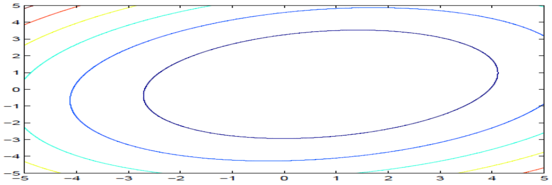
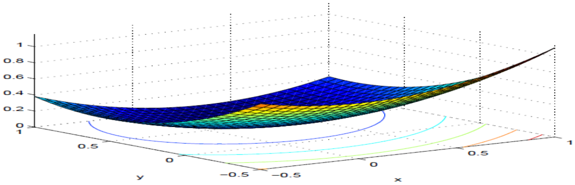


Figure: Regiones de Confianza Elípticas3D



Desviaciones de los Supuestos Clásicos

1. **(No Normalidad)** Si **y** no es multinormalmente distribuida entonces la distribución **exacta** del estimador MCO (normalidad para $\hat{\beta}$ y chi-cuadrado para s^2) ya no aplica.

   — Afortunadamente la teoría asintótica nos dice que $\hat{\beta}$ se distribuye aproximadamente normal y que este error de aproximación se va hacia cero cuando el tamaño muestral aumenta.

   Si los supuestos estándar cumplen con las condiciones que los teoremas del límite y Slutsky requieren, entonces el estimador OLS se distribuye aproximadamente normal

$$\hat{\beta} \overset{A}{\sim} \left(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

Desviaciones de los Supuestos Clásicos

2. **(Multicolinealidad)** Si la matriz $\mathbf{X}$ no es de rango (columna) completo entonces la matriz $\mathbf{X}'\mathbf{X}$ no es invertible y el verdadero vector de parámetros β no puede ser identificado a partir de los datos observados
3. **(Matriz de Covarianzas No Escalar)** Cuando la matriz de covarianzas de $\mathbf{y}$ (o ϵ) no es proporcional a una matriz identidad - $V(\mathbf{y}) \equiv \Sigma \neq \sigma^2\mathbf{I}$ para cualquier σ^2- entonces el estimador OLS clásico, aun cuando es lineal, deja de ser el "mejor" en su tipo
4. **(Regresores Endógenos)** Si falla el supuesto de que la esperanza de $\mathbf{y}$ dado $\mathbf{X}$ no es una combinación lineal de $\mathbf{X}$, se tiene la complicación más grave del caso OLS clásico.