Modelo Lineal y MCO

Walter Sosa-Escudero

Universidad de San Andres y CONICET

Ejemplo: Financiamiento de la obra publica

- Cuanto estamos dispuestos a pagar por el aire puro?
- Problema clasico en finanzas publicas: *free riding*.
- Idea: Se paga mas por vivir en lugares menos contaminados.
- Estrategia: comparar valores de casas en lugares con distinto nivel de contaminacion.
- Problema?

Modelo lineal

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + u_i \qquad i = 1, \ldots, n$$

- Relacion lineal no exacta entre Y y $X_2, \ldots, X_K$.
- Y_i, X_{ki}, observables, β_j, no.
- u_i, 'error', no observable. Error?

Objetivo: recuperar β_j's. Recuperar = estimar?. Estimar 'bien'.

Notacion

En realidad

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_{Ki} X_{Ki} + u_i, \quad i = 1, \ldots, n$$

es un sistema de n ecuaciones

$$\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \ldots + \beta_K X_{K1} + u_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \ldots + \beta_K X_{K2} + u_2 \\
\ldots & \quad \ldots \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \ldots + \beta_K X_{Kn} + u_n
\end{aligned}$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & & X_{K2} \\ \vdots & & \ddots & \vdots \\ X_{1n} & & & X_{Kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{bmatrix}$$

$$Y = X\beta + u$$

Definiciones

$$Y = X\beta + u$$

- $\hat{\beta}$, estimador de β.
- $\hat{Y} \equiv X\hat{\beta}$
- $e \equiv Y - \hat{Y} = Y - X\hat{\beta}$

Idea: $\hat{\beta}$ lo mas parecido a β, o Y a $\hat{Y}$, o e a cero.

Minimos cuadrados

Propuesta: $\hat{\beta}$ tal que

$$\hat{\beta} = \min_{\hat{\beta}} \sum_{i=1}^{n} e_i^2$$

Solucion:

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

Estimador de minimos cuadrados ordinarios (MCO)

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

$$\begin{aligned}\sum e_i^2 = e'e &= (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) \\ &= Y'Y - \tilde{\beta}'X'Y - Y'X\tilde{\beta} + \tilde{\beta}'X'X\tilde{\beta} \\ &= Y'Y - 2\tilde{\beta}'X'Y + \tilde{\beta}'X'X\tilde{\beta},\end{aligned}$$

ya que $-\tilde{\beta}'X'Y$ es un escalar, trivialmente igual a su traspuesta, $-Y'X\tilde{\beta}$, lo que conduce al resultado de la tercera linea. $e'e$ es estrictamente convexa como funcion de $\hat{\beta}$ si $\rho(X) = K$ y diferenciable con respecto a $\tilde{\beta}$, entonces, las condiciones de primer orden son necesarias y suficientes para un unico minimo global:

$$\frac{\partial e'e}{\partial \tilde{\beta}} = 0$$

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

un sistema de K ecuaciones lineales con K incognitas ($\tilde{\beta}$). Resolviendo para $\tilde{\beta}$ obtenemos:

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

Nota: Recordar que $\rho(X) = K$ implica $\rho(X'X) = K$ de modo que $(X'X)$ es invertible.
Comentario: Hay una forma de demostrar esto sin usar derivadas. Si alguien se anima, le cuento.

Supuestos clasicos

$\hat{\beta}$ es un buen estimador para β? Necesitamos mas estructura para esta comparacion: supuestos clasicos

1. $Y = X\beta + u$ (modelo lineal)
2. X aleatoria.
3. $E(u|X) = 0$ (exogeneidad).
4. $V(u_i|X) = \sigma^2$ (homocedasticidad)
5. $Cov(u_i, u_j|X) = 0$ (no correlacion serial)
6. Las variables X no pueden tener relaciones lineales exactas (no multicolinealidad, $\rho(X) = K$)

Supuestos *pedagogicos*. Discusion. X no aleatoria?

Esperanzas condicionales

Y, X, VA's

$$E(Y|X) = \int_{-\infty}^{+\infty} y \ f_{Y|X}(y|X) \ dy$$

- Si X no esta fija en un valor, $E(Y|X)$ es una *funcion* de X: una VA.
- Ley de Esperanzas Iteradas (LEI): $E(Y) = E(E(Y|X))$.
- $E[g(X)Y|X] = g(X) \ E(Y|X)$
- $E[g(X)Y] = E[g(X)E(Y|X)]$

Propiedades de muestra finita

Bajo los supuestos clasicos:

1. $\hat{\beta}$ es *lineal*: $(\hat{\beta} = AY)$.
2. $\hat{\beta}$ es *insesgado:* $E(\hat{\beta}) = \beta$.
3. $V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$
4. *Teorema de Gauss/Markov:* $\hat{\beta}$ es el estimador lineal insesgado de menor varianza.

Lineal

Lineal: existe $A_{n\times K}$ con rango K y que depende solo de X tal que $\hat{\beta} = AY$

Prueba: $A \equiv (X'X)^{-1}X'$ satisface todos los requisitos.

$\hat{\beta}$ 'hereda' la aleatoriedad de u linealmente a traves de Y. Tecnicamente relevante. Conceptualmente?

Insesgadez

Recordar $E(\hat{\beta}) = E(E(\hat{\beta}|X))$, por LIE.

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X\beta + u) \\
&= \beta + (X'X)^{-1}X'u \\
E(\hat{\beta}|X) &= \beta + E\left[(X'X)^{-1}X'u|X\right] \\
&= \beta + (X'X)^{-1}X'E\left[u|X\right] \\
&= \beta
\end{aligned}
$$

Entonces: $E(\hat{\beta}) = E(E(\hat{\beta}|X)) = E(\beta) = \beta$

Cual es 'el' supuesto? Heterocedasticidad, normalidad? Que supuestos se usan?

Varianza: $V(\hat{\beta}|X) = \sigma^2 (X'X)^{-1}$.

$$\begin{aligned} V(\hat{\beta}|X) &= V(\hat{\beta}-\beta|X) \quad (\beta \text{ no es aleatoria}) \\ &= V\left[(X'X)^{-1}X'u|X\right] \quad \text{(de la prueba anterior...)} \\ &= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}|X\right] \\ &= (X'X)^{-1}X'E(uu'|X)X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \quad \text{(homocedasticidad, no corr serial)} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Recordatorios: 1) $V(u) \equiv E\big((u-E(u)(u-E(u)')\big)$. Si $E(u)=0$, $V(u)=E(uu')$. 2) $V(u)_{ij} = Cov(u_i, u_j)$. Entonces homocedasticidad y no correlacion lineal implican $V(u) = E(uu') = \sigma^2 I_n$

Gauss/Markov (caso X no aleatoria: $E(u|X) = E(u) = 0$): Sea $\tilde{\beta}$ cualquier estimador lineal e insesgado.

$\tilde{\beta}$ lineal: existe $A_{K\times n}$ que depende solo de X, con rango K, tal que $\tilde{\beta} = AY$.
Bajo los supuestos clasicos

$$E(\tilde{\beta}) = E(AY) = E\left(A(X\beta + u)\right) = AX\beta \tag{1}$$

$\tilde{\beta}$ insesgado:

$$E(\tilde{\beta}) = \beta \tag{2}$$

$\tilde{\beta}$ lineal e insesgado: (1) y (2) simultaneamente, requiere $AX = I$.

Trivialmente, $\tilde{\beta} = \hat{\beta} + \tilde{\beta} - \hat{\beta} = \hat{\beta} + \hat{\gamma}$, con $\hat{\gamma} \equiv \tilde{\beta} - \hat{\beta}$.

Notar que: $V(\tilde{\beta}) = V(\hat{\beta}) + V(\hat{\gamma})$ sii $Cov(\hat{\beta}, \hat{\gamma}) = 0$.

Entonces, si podemos probar que $Cov(\hat{\beta}, \hat{\gamma}) = 0$, obtendremos nuestro resultado. Por que?

Recordar: $Cov(\hat{\beta}, \hat{\gamma}) \equiv E\big(\hat{\beta} - E(\hat{\beta}))(\hat{\gamma} - E(\hat{\gamma}))'\big)$

Trivialmente $E(\hat{\gamma}) = 0$ (Por que?). Como ademas $\hat{\beta}$ es insesgado, $Cov(\hat{\beta}, \hat{\gamma}) = E[(\hat{\beta} - \beta)\hat{\gamma}']$

Notar que

$$\begin{aligned}
\hat{\gamma} &= AY - (X'X)^{-1}X'Y \\
&= (A - (X'X)^{-1}X')Y \\
&= (A - (X'X)^{-1}X')(X\beta + u) \\
&= (A - (X'X)^{-1}X')u \quad \text{(dado que } AX = I\text{)}
\end{aligned}$$

Reemplazando:

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

donde hemos usado $V(u) = E(uu') = \sigma^2 I_n$, y, nuevamente, $AX = I$. Entonces, de acuerdo a nuestros resultados anteriores:

$$V(\tilde{\beta}) - V(\hat{\beta}) = V(\hat{\gamma}),$$

que es positiva semidefinida, por construccion (es una varianza!).

Interpretacion basica

$Y = X\beta + u$. Bajo exogeneidad:

$$E(Y|X) = X\beta$$

Bajo diferenciabilidad y si las X no estan funcionalmente relacionadas:

$$\frac{\partial E(Y|X)}{\partial X_j} = \beta_j$$

Interpretacion causal? Mas adelante....

Inferencia

- $H_0 : \beta_j = 0$ (significatividad)?, Intervalo de confianza?
- Necesitamos la distribucion de $\hat{\beta}$.
- Supuesto adicional: u_i conjuntamente normales.
- *Resultado:* Dado X, $\hat{\beta}$ es normal. Por que?

Bajo supuestos clasicos: $\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$.

Tests de significatitivad: $H_0 : \beta_j = 0$

Si $V(\hat{\beta}|X) \equiv \sigma^2(X'X)^{-1}$. Entonces

$$V(\hat{\beta}_j) = \sigma^2 D_{jj}$$

con $D \equiv (X'X)^{-1}$. Entonces

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 D_{jj})$$

Bajo $H_0 : \beta_j = 0$:

$$z_j \equiv \frac{\hat{\beta}_j}{\sqrt{\sigma^2 D_{jj}}} \sim N(0,1)$$

Idea: rechazar H_0 si z_j es muy grande. Problema: σ^2 no observable

Estimemos σ^2 usando $S^2 \equiv \sum e_i^2/(n-K)$

Resultado: bajo los supuestos clasicos mas normalidad y cuando $H_0 : \beta_j = 0$

$$t_j \equiv \frac{\hat{\beta}_j}{\sqrt{S^2 D_{jj}}} \sim t_{n-K}$$

Prueba: Tres resultados prelminares:

- $e = Mu$ con $M \equiv I_n - (X'X)^{-1}X'$ (demostrar como ejercicio). M simetrica e idempotente ($M = M'M$).
- Si $Z \sim N(0,1)$ y $Y \sim \chi^2(m)$, Z y Y independientes, entonces

$$\frac{Z}{\sqrt{Y/m}} \sim t_m$$

- Si $Z \sim N(0, I_m)$, entonces $Z'AZ \sim \chi^2(\rho(A))$.

Recordar que $z_j \equiv \hat{\beta}_j/\sqrt{\sigma^2 D_{jj}}$. Notar que:

$$t_j = \frac{\hat{\beta}_j}{\sqrt{S^2 D_{jj}}} = \frac{z_j}{\sqrt{S^2/\sigma^2}} = \frac{z_j}{\sqrt{\frac{e'e/(n-K)}{\sigma^2}}}$$

Vimos que $z_j \sim N(0,1)$. Si podemos mostrar que $e'e/\sigma^2 \sim \chi^2(n-K)$ e independencia, estamos.

1. Numerador χ^2: De $e = Mu$, $e'e = u'MM'u = u'Mu$ dado que M es idempotente. Bajo normalidad $u|X \sim N(0, \sigma^2 I_n)$, entonces, por los resultados anterioes $u'Mu/\sigma^2 \sim \chi^2(n-K)$ dado que $\rho(M) = n - K$.
2. Independencia: $\hat{\beta} = \beta + (X'X)^{-1}X'u$ y $e = Mu$ son funciones lineales de u, entonces, son conjuntamente normales, dado X. Ahora bien,

$$\begin{aligned} Cov(\hat{\beta}, e) &= E[(\hat{\beta} - \beta)e'] \\ &= E[(X'X)^{-1}X'uu'M] \\ &= (X'X)^{-1}X'E(uu|X)M \\ &= \sigma^2(X'X)^{-1}X'M = 0 \end{aligned}$$

usando la Ley de Esperanzas Iteradas, los supuestos clasicos y que $X'M = 0$. Bajo normalidad conjunta, covarianza nula equivale a independencia.

Kit basico de regresion

1. $\hat{\beta} = (X'X)^{-1}X'Y$
2. $\hat{V}(\hat{\beta}_j) = \sigma^2 D_{jj}$
3. $t_j = \hat{\beta}_j / \sqrt{\sigma^2 D_{jj}} \sim t_{n-K}$

- Estimaciones puntuales, desvio estandar, intervalos de confianza, significatividad, p-valores, etc.
- Supuestos clasicos mas normalidad
- Insesgadez, Gauss/Markov
- Inferencia valida

Ejemplo: Financiamiento de la obra publica

- Un clasico: Harrison, D. and Rubinfeld, D., 1978. Hedonic prices and the demand for clean air, *Journal of Environmental Economics and Management*, 5, 81-102
- Cuanto estamos dispuestos a pagar por el aire puro?
- Problema clasico en finanzas publicas: *free riding*.
- Idea: Se paga mas por vivir en lugares menos contaminados.
- Estrategia: comparar valores de casas en lugares con distinto nivel de contaminacion.
- Problema?

- Modelo hedonico.
- Recuperar la contribucion de cada caracteristica al precio total.
- Regresion: aisla.

- Datos: Harrison y Rubinfeld (1978). 506 viviendas.
- Variable explicada: VALUE, valor promedio de casas ocupadas en Boston (miles de dolares).
- Variable relevante: NITOX concentracion of oxido de nitrogeno (partes per million, concentracion annual promedio).
- Controles: CRIME, ROOMS, AGE, DIST, ACCESS, TAX, PRATIO.

Variable	Obs	Mean	Std. Dev.	Min	Max
value	506	22.53281	9.197104	5	50
crime	506	3.613525	8.601545	.0063	88.9762
nitox	506	.5546951	.1158777	.385	.871
rooms	506	6.284634	.7026172	3.561	8.78
age	506	68.5749	28.14886	2.9	100
dist	506	3.795043	2.10571	1.1296	12.1265
access	506	9.549407	8.707259	1	24
tax	506	408.2372	168.5371	187	711
ptratio	506	18.45553	2.164946	12.6	22

Source	SS	df	MS
Model	28064.0746	8	3508.00932
Residual	14652.221	497	29.48133
Total	42716.2956	505	84.586724

Number of obs	=	506
F(8, 497)	=	118.99
Prob > F	=	0.0000
R-squared	=	0.6570
Adj R-squared	=	0.6515
Root MSE	=	5.4297

value	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
crime	-.1834488	.0364887	-5.03	0.000	-.25514	-.1117576
nitox	-22.81088	4.160742	-5.48	0.000	-30.98569	-14.63607
rooms	6.371512	.3923866	16.24	0.000	5.600571	7.142453
age	-.0477499	.0141018	-3.39	0.001	-.0754564	-.0200434
dist	-1.335269	.2001468	-6.67	0.000	-1.728507	-.942031
access	.272282	.072276	3.77	0.000	.1302777	.4142863
tax	-.0125921	.0037702	-3.34	0.001	-.0199995	-.0051847
ptratio	-1.176787	.1394154	-8.44	0.000	-1.450703	-.9028705
_cons	28.40667	5.365948	5.29	0.000	17.86393	38.9494

- Politica: reducir 5% la contaminacin. Incremento en valor de vivienda?
- $22810 \times 0.05x = 1140.5x$
- Barrio con contaminacion promedio (0.55ppm): incrementa el valor en 627 dolares.
- Si el costo de reducir la contaminacion en 5 es mayor que $627N$: inviable. N, cantidad de viviendas del barrio.