Fundamentos de Econometría

Ignacio Lobato

ITAM

Regresión Multivariada

- Se busca estudiar la relación de la variable dependiente Y con un vector de variables explicativas $X_2, \ldots, X_k$
- Vector k-variado $(Y, X_2, \ldots, X_k)$ con función de densidad conjunta $f(y, x_2, \ldots, x_k)$
- Función de distribución condicional de Y dado $X_2, \ldots, X_k$:

$$g(y|x_2, \ldots, x_k) = \frac{f(y, x_2, \ldots, x_k)}{f_1(x_2, \ldots, x_k)}$$

- Donde $f_1(x_2, \ldots, x_k) = \int f(y, x_2, \ldots, x_k)\, dy$

CEF

- La CEF de Y dado $X_2, \ldots, X_k$ es:

$$E\left(Y|X_2, \ldots, X_k\right) = \int yg\left(y|x_2, \ldots, x_k\right) dy$$

- La CEF es el mejor predictor de Y dadas las X's en el sentido de minimizar la esperanza de los errores $U = \left(Y - h\left(\bar{X}\right)\right)$ al cuadrado:

$$\min_h E\left(Y - h\left(\bar{X}\right)\right)^2 = E\left(U^2\right)$$

MPL Muestral

- Se busca la mejor combinación lineal de $X_2, \ldots, X_k$ para predecir Y
- Se minimiza de acuerdo al criterio: $\phi(c_1, \ldots, c_k) = \sum_{i=1}^{N} u_i^2$, donde

$$u_i = y_i - (c_1 + c_2 x_{i2} + \ldots + c_k x_{ik})$$

- Derivando respecto a c_j:

$$\frac{\partial \phi}{\partial c_j} = \sum_i \left(\frac{\partial u_i^2}{\partial c_j} \right) = \sum_i 2u_i \left(\frac{\partial u_i}{\partial c_j} \right) = -2 \sum_i x_{ij} u_i$$

- Condiciones de primer orden: $\sum_i x_{ij} u_i = 0$

- De las condiciones de primer orden se tiene un sistema de k ecuaciones lineales en $c_1, \ldots, c_k$
- Por simplicidad, adoptamos notación matricial: $\underset{nxk}{X} = (x_1, \ldots, x_k)$

$$\underset{kx1}{c} = (c_1, \ldots, c_k)', \quad \phi(\underset{kx1}{c}) = u'u$$

$$\underset{nx1}{u} = Y - Xc$$

- Condiciones de primer orden: X'u=0
- c^* denota el vector solución

$$X'(Y - Xc^*) = 0, \quad o\ bien,\ X'Xc^* = X'Y$$

- La matriz $\underset{kxk}{Q} = X'X$ es simétrica y tiene como columnas las sumas de cuadrados y productos cruzados de los regresores

Full-rank case

- Q es no singular, invertible, $|Q| \neq 0$
- Sucede cuando $rango(X) = k$
- $Qc^* = X'Y$ tiene solución única: $\underset{kx1}{b} = Q^{-1}X'Y = AY$
- Fitted-value vector: $\underset{nx1}{\hat{Y}} = Xb = XAY = NY$
- Residuales: $\underset{nx1}{e} = Y - \hat{Y} = (I - N)\,Y = MY$
- Matriz de proyección: $\underset{nxn}{N} = XA = XQ^{-1}X'$
- $\underset{nxn}{M} = I - N = I - XQ^{-1}X'$
- Matrices M y N son idempotentes:

$$MM = M, \quad NN = N$$

- M y N son ortogonales: $MN = 0$
- Propiedades de las matrices Q, M, N y A en p. 155 de Goldberger

Regresión Clásica

- Se tienen n variables aleatorias $(Y_1, \ldots, Y_n)$ con función de densidad conjunta $f(y_1, \ldots y_n)$
- $E(y_i) = \mu_i, \; V(y_i) = \sigma_i^2 = \sigma_{ii}, \; C(y_h, y_i) = \sigma_{hi} = \sigma_{ih}$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \sum = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \sigma_{hi} & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}$$

- Definimos $\epsilon = Y - \mu$, la desviación de y de su media
- $E(\epsilon) = 0, \quad E(\epsilon\epsilon') = \sum = V(Y) = V(\epsilon)$
- **Def:** la matriz de covarianzas entre 2 vectores aleatorios $\underset{mx1}{x}$ y $\underset{nx1}{z}$ esta dada por:

$$\underset{mxn}{C(x, z)} = E\left\{[x - E(x)][z - E(z)]'\right\}$$

- El elemento (h, i) esta dado por $C(x_h, z_i)$

- **Def:** $z = g + h'Y$ es una función escalar lineal, donde g es un escalar y h es un vector de constantes de nx1

$$E(z) = g + h'\mu, \quad V(z) = h'E\left(\epsilon\epsilon'\right)h = h'\sum h$$

- **Def:** $z = g + HY$ es una función vectorial lineal, donde g es un vector kx1 y H es una matriz kxn de constantes

$$E(z) = g + H\mu, \quad V(z) = HE\left(\epsilon\epsilon'\right)H' = H\sum H'$$

Esperanza de Cuadrados

- **Teorema:** si Y es un vector aleatorio de dimensión nx1:

$$E\left(YY'\right) = \mu\mu' + \sum$$

- Escribimos:

$$YY' = (\mu + \epsilon)(\mu + \epsilon)' = \mu\mu' + \mu\epsilon' + \epsilon\mu' + \epsilon\epsilon'$$

- Obteniendo esperanza:

$$E\left(YY'\right) = \mu\mu' + \sum$$

dado que $E(\epsilon) = 0$ y μ es constante

Suma de Cuadrados

- **Teorema:** para un vector aleatorio Y de dimensión nx1:

$$E\left(Y'Y\right) = tr\left(\sum\right) + \mu'\mu$$

- Recordemos que para 2 matrices cuadradas A y B, $tr\left(AB\right) = tr\left(BA\right)$
- Un escalar es igual a su traza, escribimos:

$$Y'Y = tr\left(Y'Y\right) = tr\left(YY'\right)$$

- Obteniendo esperanza y utilizando el teorema de suma de cuadrados:

$$E\left(Y'Y\right) = E\left(tr\left(YY'\right)\right) = tr\left(E\left(YY'\right)\right) = tr\left(\sum + \mu\mu'\right)$$

- Utilizando el operador lineal de la traza:

$$tr\left(\sum + \mu\mu'\right) = tr\left(\sum\right) + tr\left(\mu\mu'\right) = tr\left(\sum\right) + \mu'\mu$$

- Entonces: $E\left(Y'Y\right) = tr\left(\sum\right) + \mu'\mu$

Forma Cuadrática

- **Teorema:** sea $Y'TY$, donde Y es un vector aleatorio de dimensión nx1 y T es una matriz nxn de constantes, entonces:

$$E\left(Y'TY\right) = tr\left(T\sum\right) + \mu' T\mu$$

- Escribimos $Y'TY = tr\left(TYY'\right)$
- Calculando esperanzas:

$$E\left(Y'TY\right) = E\left(tr\left(TYY'\right)\right) = tr\left(TE\left(YY'\right)\right)$$

- Del teorema de suma de cuadrados:

$$tr\left(TE\left(YY'\right)\right) = tr\left(T\left(\sum + \mu\mu'\right)\right) = tr\left(T\sum\right) + tr\left(T\mu\mu'\right)$$

- Por lo tanto:

$$E\left(Y'TY\right) = tr\left(T\sum\right) + tr\left(T\mu\mu'\right) = tr\left(T\sum\right) + \mu' T\mu$$

Par de Vectores de Funciones Lineales

- **Teorema:** para un par de vectores de funciones lineales $z_1 = g_1 + H_1 Y$ y $z_2 = g_2 + H_2 Y$, donde g_1 es vector de $m_1 x 1$, H_1 es matriz de $m_1 x n$, g_2 es vector de $m_2 x 1$ y H_2 es matriz de $m_2 x n$:

$$C(z_1, z_2) = H_1 \sum H_2'$$

- Sea $z_1^* = z_1 - E(z_1) = H_1 \epsilon$ y $z_2^* = z_2 - E(z_2) = H_2 \epsilon$
- Multiplicando, $z_1^* z_2^{*'} = H_1 \epsilon \epsilon' H_2'$, entonces:

$$C(z_1, z_2) = E\left(z_1^* z_2^{*'}\right) = H_1 E\left(\epsilon \epsilon'\right) H_2' = H_1 \sum H_2$$

Regresión Clásica: Supuestos

- Para un vector aleatorio $\underset{nx1}{Y}$ y una matriz $\underset{nxk}{X} = (x_1, \ldots, x_k)$:
- $E(Y) = X'\beta$
- $V(Y) = \sigma^2 I_n$
- X es no estocástica
- $rango(X) = k$

Estimación de β

- Para estimar β, del principio de analogía y dado que $\mathit{rango}\,(X) = k$, se propone $b = AY$

$$E\,(b) = E\,(AY) = AE\,(Y) = AX\beta = \beta$$

$$V\,(b) = V\,(AY) = AV\,(Y)\,A' = A\sigma^2 IA' = \sigma^2 Q^{-1}$$

Teorema de Gauss-Markov

- En el modelo clásico de regresión, b es el estimador lineal insesgado de β con menor varianza:

$$V\left(b^{*}\right) \geq V\left(b\right) \; para\; todo\; b^{*} \; tal\; que\; E\left(b^{*}\right) = \beta$$

Gauss-Markov

- b* es una función lineal de Y, puede escribirse como CY, donde $\underset{kxn}{C}$ es alguna matriz no estocástica
  - $E(b^*) = CE(Y) = CE(Y) = CX\beta$
  - $V(b^*) = CV(Y)C' = \sigma^2 CC'$
- Sea $C = A + D$, entonces:
  - $CX = AX + DX = I + DX$
  - $CC' = (A + D)(A + D)' = AA' + DD' + AD' + DA'$
- Para que b^* sea insesgado, debe ser que $CX = I$, lo que implica que $DX = 0$ y $DXQ^{-1} = 0$, es decir, $DA' = AD' = 0$

Gauss-Markov

$$V\left(b*\right) = \sigma^2 \left(AA' + DD'\right) = V\left(b\right) + \sigma^2 DD'$$

- La matriz DD' es positiva semidefinida
- Por lo tanto:
$$V\left(b^*\right) \geq V\left(b\right)$$

Estimación de σ^2 y V(b)

- $e = MY$, por lo tanto $E(e) = ME(Y) = MX\beta = 0$ y $V(e) = MV(Y)M' = \sigma^2 M$
- Del teorema de suma de cuadrados: $E(e'e) = tr(V(e)) + E(e)' E(e) = \sigma^2 tr(M)$
- $M = I_n - N$, por lo tanto $tr(M) = n - k$
- Dado lo anterior y el principio de analogía, se propone $\hat{\sigma}^2 = \frac{e'e}{n-k}$ para estimar σ^2
- $\hat{\sigma}^2$ es insesgado:

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{e'e}{n-k}\right) = \frac{\sigma^2 (n-k)}{n-k} = \sigma^2$$

- Finalmente, se usa $\hat{V}(b) = \hat{\sigma}^2 Q^{-1}$ para estimar $V(b)$

Ejemplo

- Sea $(Y, x_2, \ldots, x_k)$ un vector aleatorio con función de densidad conjunta $f(Y, x_2, \ldots, x_k)$
- $E(Y) = \mu_y, \quad V(Y) = \sigma^2, \quad C(x_h, x_j) = \sigma_{hj}$
- Se asume función de esperanza condicional lineal y varianza constante:

$$E(Y|x_2, \ldots, x_k) = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

$$V(Y|x_2, \ldots, x_k) = \sigma^2$$

- Tomando $x = (1, x_2, \ldots, x_k)'$, y $\beta = (\beta_1, \ldots, \beta_k)'$, se expresa lo anterior como $E(Y|x) = x'\beta$ y $V(Y|x) = \sigma^2$

Estimación de Funciones Lineales de β

- Se desea estimar $\theta = h'\beta$, h es un vector de constantes de dimensión kx1
- Del principio de analogía, se utiliza $\hat{\theta}_n = h'b$
- $\hat{\theta}_n$ es una función lineal de y: $\hat{\theta}_n = h'b = h'AY = w'Y$
- $\hat{\theta}_n$ es insesgado:

$$E\left(\hat{\theta}_n\right) = h'E\left(b\right) = h'\beta = \theta$$

- De las propiedades de la varianza:

$$V\left(\hat{\theta}_n\right) = h'V\left(b\right)h = \sigma^2 h'Q^{-1}h$$

- Para cualquier otro estimador de θ : $t^* = w^* Y$, donde w* es un vector de constantes de nx1:

$$E(t^*) = w^{*'} X\beta, \quad V(t^*) = \sigma^2 w^{*'} w^*$$

- t^* es insesgado $\Leftrightarrow w^{*'} X = h'$
- En ese caso, $h' Q^{-1} h = w^{*'} X Q^{-1} X' w^* = w^{*'} N w^*$
- Reescribimos entonces $V\left(\hat{\theta}_n\right) = \sigma^2 h' Q^{-1} h = \sigma^2 w^{*'} N w^*$

$$V(t^*) - V\left(\hat{\theta}_n\right) = \sigma^2 w^{*'} M w^* \geq 0$$

- $\hat{\theta}_n$ es el estimador lineal insesgado de menor varianza
- El error estándar de $\hat{\theta}_n$ es $\hat{\sigma}_t = \hat{\sigma}\sqrt{(h' Q^{-1} h)}$

Estimación Puntual

- Para estimar $\mu_i = E(y_i) = x_i'\beta$, del principio de analogía se utiliza $\hat{y}_i = x_i'b$

$$E(\hat{y}_i) = \mu_i, \quad V(\hat{y}_i) = \sigma^2 x_i' Q^{-1} x_i$$

- Se desea estimar $\mu_0 = x_0'\beta$, donde x_0 es algún vector de dimensión $kx1$
- Del principio de analogía, $\hat{\mu}_0 = x_0'b$

$$E(\hat{\mu}_0) = \mu_0, \quad V(\hat{\mu}_0) = \sigma^2 x_0' Q^{-1} x_0$$

- El error estándar de $\hat{\mu}_0$ es $\hat{\sigma}\sqrt{(x_0' Q^{-1} x_0)}$

Predicción

- El objetivo es predecir el valor de y_0
- Si β fuera conocido, la predicción sería $\mu_0 = x_0'\beta$, y el error $\epsilon_0 = y_0 - \mu_0$
$$E(\epsilon_0) = 0, \quad V(\epsilon_0) = E\left(\epsilon_0^2\right) = \sigma^2$$
- En general β no es conocido, pero es estimado con b
- El predictor será $\hat{\mu}_0 = x_0'b$, con error de predicción $u = y_0 - \hat{\mu}_0$
$$E(u) = 0, \quad V(u) = V(y_0) + V(\hat{\mu}_0) - 2C(y_0, \hat{\mu}_0)$$
$$= \sigma^2 \left(1 + x_0' Q^{-1} x_0\right)$$
- $C(y_0, \hat{\mu}_0) = 0$ porque y_0 es independiente de la muestra
- El error estándar de $\hat{\mu}_0$ es $\hat{\sigma}\sqrt{(1 + x_0' Q^{-1} x_0)}$

Bondad del Ajuste

- Se reporta una medida para observar que tan bien ajusta el modelo lineal
- Recordemos que:

$$Y = \hat{Y} + e \Rightarrow Y'Y = \left(\hat{Y} + e\right)'\left(\hat{Y} + e\right) = \hat{Y}'\hat{Y} + e'e$$

algebraicamente:

$$\sum_i y_i^2 = \sum_i \hat{y}_i^2 + \sum_i e_i^2$$

- También tenemos:

$$\sum_i y_i = \sum_i \hat{y}_i + \sum_i e_i \Rightarrow \bar{Y} = \bar{\hat{Y}} + \bar{e}$$

Coeficiente de Determinación

- Si $\bar{e} = 0$, $\bar{Y} = \bar{\hat{Y}} \Rightarrow n\bar{Y}^2 = n\bar{\hat{Y}}^2$, y restando este resultado a la ecuación anterior:

$$\sum_i \left(y_i - \bar{Y}\right)^2 = \sum_i \left(\hat{y}_i - \bar{Y}\right)^2 + \sum_i e_i^2$$

- Definimos ahora el coeficiente de determinación:

$$R_n^2 = \frac{\sum_i \left(\hat{y}_i - \bar{Y}\right)^2}{\sum_i \left(y_i - \bar{Y}\right)^2} = 1 - \frac{\sum_i e_i^2}{\sum_i \left(y_i - \bar{Y}\right)^2}$$

$0 \leq R_n^2 \leq 1$

- R_n^2 mide la proporción de la varianza de Y explicada por el predictor lineal

Coeficiente de Determinación

- Intuitivamente, R_n^2 mide que tan bien ajustados están los datos a la regresión
- Si $R_n^2 = 1$, el ajuste es perfecto, todas las y's pertenecen a una función lineal de las x's:

$$R_n^2 = 1 \Leftrightarrow \sum_i e_i^2 = 0 \Leftrightarrow e'e = 0 \Leftrightarrow e = 0 \Leftrightarrow Y = X\beta$$

- Si $R_n^2 = 0$, el mejor predictor lineal es horizontal:

$$R_n^2 = 0 \Leftrightarrow \sum_i \left(\hat{y}_i - \bar{Y}\right)^2 = 0 \Leftrightarrow \hat{y}_i = \bar{Y}$$

Coeficiente de Determinación (apuntes)

- Solo habrá coeficiente de determinación cuando exista intercepto, es decir:
  - Una columna de X debe ser un vector de 1's
  - Existe una combinación lineal de columnas de X igual a un vector de 1's
- R_n^2 suele aumentar cuando se agregan variables explicativas

Coeficiente de Determinación Ajustado

- Se introduce el coeficiente de determinación ajustado ${\bar{R}_n}^2$
- $$1 - {\bar{R}_n}^2 = \frac{(n-1)\left(1 - R_n^2\right)}{n-k}$$
- Despejando ${\bar{R}_n}^2$:

$$\bar{R}_n^2 = 1 - \frac{\sum_i e_i^2 / (n-k)}{\sum_i \left(y_i - \bar{Y}\right)^2 / (n-1)}$$

- $\bar{R}_n^2$ aumenta la proporción no explicada y disminuye la proporción explicada