

TEMA 5

EL MODELO LINEAL GENERAL (I)

I.- PLANTEAMIENTO DEL MODELO LINEAL GENERAL.

El modelo lineal general es una extensión del modelo lineal simple e incluye a este, es decir, $MLS \in MLG$.

Lo que pretende este modelo, es explicar el comportamiento de la variable (Y_t) mediante un conjunto de variables ($X_{2t} \ X_{3t} \ \dots \ X_{kt}$) a partir de los datos del periodo $t= 1,2,\dots,n$.

Para estimar un modelo partimos de una serie de observaciones tanto de la variable endógena o explicada (Y_t), como de las exógenas o explicativas ($X_{2t} \ X_{3t} \ \dots \ X_{kt}$). Supongamos que estas observaciones vienen dadas en una tabla como la siguiente:

$$\begin{array}{c|ccccc}
 & Y_t & X_{2,t} & X_{3,t} & \dots & X_{k,t} \\
 \hline
 1 & Y_1 & X_{2,1} & X_{3,1} & \dots & X_{k,1} \\
 2 & Y_2 & X_{2,2} & X_{3,2} & \dots & X_{k,2} \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 n & Y_n & X_{2,n} & X_{3,n} & \dots & X_{k,n}
 \end{array}
 \quad \bar{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}
 \quad \bar{x}_j = \begin{pmatrix} X_{j,1} \\ X_{j,2} \\ \vdots \\ X_{j,n} \end{pmatrix}$$

$$\bar{y} = \beta_1 + \beta_2 \bar{x}_2 + \beta_3 \bar{x}_3 + \dots + \beta_k \bar{x}_k + \bar{u}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} X_{2,1} \\ X_{2,2} \\ \vdots \\ X_{2,n} \end{pmatrix} + \beta_3 \begin{pmatrix} X_{3,1} \\ X_{3,2} \\ \vdots \\ X_{3,n} \end{pmatrix} + \dots + \beta_k \begin{pmatrix} X_{k,1} \\ X_{k,2} \\ \vdots \\ X_{k,n} \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$$Y_1 = \beta_1 + \beta_2 X_{2,1} + \beta_3 X_{3,1} + \dots + \beta_k X_{k,1} + u_1$$

y en general

$$Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \dots + \beta_k X_{k,i} + u_i \quad \text{para } i = 1,2,\dots,n$$

$Y_j \Rightarrow$ Variable endógena

$X_{kj} \Rightarrow$ Variables explicativas o exógenas

$\beta_k \Rightarrow$ Parámetros del modelo, llamados parámetros de posición.

$\bar{u} \Rightarrow$ Perturbación aleatoria.

Lo que pretendemos estimar son los parámetros de posición.

La perturbación aleatoria no es un parámetro que pertenezca al modelo, sino que surge de la necesidad de realizar un ajuste, es decir, una vez hechas todas las sustituciones de β_k, x_j y de y_i , no se produce la igualdad, y necesitamos de una perturbación para la realización de dicho ajuste. Estos desajustes se deben a que los datos:

- Tienen pequeñas imperfecciones
- Proceden del ser humano, su comportamiento, y este no se puede encuadrar en ninguna fórmula.

En el modelo vamos a exigir que las perturbaciones económicas se comporten de un determinado modo, de forma que con la definición de éstas, tenemos el modelo definido.

La perturbación económica tiene que ser de la forma

$\bar{u} \rightarrow N(\bar{0}, \sigma^2 I_n)$, siendo σ^2 , un parámetro de dispersión que mide la varianza de la perturbación económica. Por ser la matriz de covarianzas $\sigma^2 I_n$, obligamos a que todos los u_i , tengan la misma varianza.

EXPRESIÓN MATRICIAL DEL MODELO LINEAL GENERAL.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1} = \begin{pmatrix} 1 & X_{2,1} & X_{3,1} & \cdots & X_{k,1} \\ 1 & X_{2,2} & X_{3,2} & \cdots & X_{k,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2,n} & X_{3,n} & \cdots & X_{k,n} \end{pmatrix}_{n \times k} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k \times 1} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}_{n \times 1}$$

EXPRESIÓN MATRICIAL SIMPLIFICADA DEL MODELO LINEAL GENERAL.

$$\bar{y} = X\bar{\beta} + \bar{u}$$

HIPOTESIS INICIALES DEL MODELO

1)

$$E[u_i] = 0 \quad \forall i$$

2)

$$\begin{cases} E[u_i^2] = \text{var}(u_i) = \sigma^2 \\ E[u_i \cdot u_j] = 0 \quad i \neq j \\ E[u_i \cdot u_j] = \sigma^2 \quad i = j \end{cases}$$

- 3) X , la matriz de información, es una matriz no estocástica, es decir, que sus componentes han de ser números.
- 4) El número de observaciones ha de ser estrictamente mayor que k , es decir $n > k$ y el rango de la matriz de información (X), k , es decir, $rg(X) = k$.

El modelo se puede estimar sin que se cumplan estas hipótesis:

$$\begin{array}{ccc}
 \text{MLG} & \supset & \text{MLGN} \\
 \cup & & \cup \\
 \text{MLS} & \supset & \text{MLSN}
 \end{array}$$

ESTIMACIÓN MÍNIMO CUADRÁTICA. PROPIEDADES

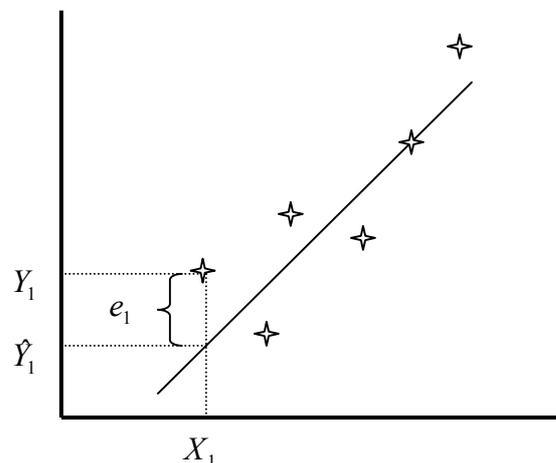
Supongamos que hayamos estimado el modelo, por tanto hemos conseguido unos valores $\hat{\beta}_j$, que componen las coordenadas del vector de estimadores $\hat{\beta}$. Con este vector $\hat{\beta}$, podemos obtener los valores estimados \hat{y} realizando la siguiente operación (véase el modelo inicial)

$$\hat{y} = X\hat{\beta}$$

Una vez que tenemos los valores estimados, podemos calcular los residuos (e_j), es decir, la diferencia entre los valores observados y los valores estimados con el modelo.

$$\bar{e} = \begin{pmatrix} e_1 \\ e \\ \cdot \\ \cdot \\ e_n \end{pmatrix} = \begin{pmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{pmatrix}$$

GRÁFICAMENTE



EL MODELO DE MÍNIMOS CUADRADOS ORDINARIOS: MCO

El modelo de mínimos cuadrados ordinarios consiste en estimar el vector de parámetros ($\hat{\beta}$) de modo tal que cumpla una determinada propiedad, consistente en hacer que la suma de los cuadrados del error sea mínima, es decir, minimizar la expresión: $\sum e_i^2$.

En el caso de que hubiésemos estimado el vector de parámetros por el método de mínimos cuadrados ordinarios, sería del siguiente modo:

$$\begin{aligned} \sum e_i^2 &= \bar{e}'\bar{e} = (\bar{y} - X\hat{\beta})'(\bar{y} - X\hat{\beta}) = (\bar{y}' - \hat{\beta}'X')(\bar{y} - X\hat{\beta}) = \\ &= \bar{y}'\bar{y} - \underbrace{\bar{y}'_{1 \times n} X_{n \times k} \hat{\beta}_{k \times 1}}_{\substack{\text{una es la traspuesta de la otra y como son} \\ \text{números, son la misma expresión.}}} - \underbrace{\hat{\beta}'_{1 \times k} X'_{k \times n} \bar{y}_{n \times 1}}_{\substack{\text{una es la traspuesta de la otra y como son} \\ \text{números, son la misma expresión.}}} + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

$$\begin{aligned} \sum e_i^2 &= \bar{y}'\bar{y} - 2\hat{\beta}'X'\bar{y} + \hat{\beta}'X'X\hat{\beta} \\ \frac{\partial \sum e_i^2}{\partial \hat{\beta}} &= -2X'\bar{y} + 2(X'X)\hat{\beta} = 0 \end{aligned}$$

⇓

$$X'\bar{y} = (X'X)\hat{\beta} \quad (1)$$

$$\begin{aligned} \sum e_i^2 &= \bar{y}'\bar{y} - 2\bar{y}'X\hat{\beta}' + \hat{\beta}'X'X\hat{\beta} \\ \frac{\partial \sum e_i^2}{\partial \hat{\beta}} &= -2\bar{y}'X + 2\hat{\beta}'(X'X) = 0 \end{aligned}$$

⇓

$$\bar{y}'X = \hat{\beta}'(X'X) \quad (2)$$

Como podemos observar, la expresión 1, es la traspuesta de la expresión 2, y ambas son las ecuaciones normales.

Si pudiésemos asegurar que $|X'X| \neq 0$, entonces podríamos decir que:

$$\left\{ \begin{array}{l} (X'X)^{-1} X'\bar{y} = (X'X)^{-1} (X'X)\hat{\beta} \Rightarrow \hat{\beta} = (X'X)^{-1} X'\bar{y} \\ \bar{y}'X(X'X)^{-1} = \hat{\beta}'X'X(X'X)^{-1} \Rightarrow \hat{\beta}' = \bar{y}'X(X'X)^{-1} \end{array} \right.$$

Para ello debemos asegurar que el rango de la matriz sea k, pero esto no supone un problema ya que es una de las exigencias de partida del modelo.

PROPIEDADES DEL ESTIMADOR MÍNIMO CUADRÁTICO A PARTIR DE LAS ECUACIONES NORMALES.

Desarrollo de las ecuaciones normales:

$$(X'X)\hat{\beta} = X'y$$

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & X_{2,3} & \cdots & X_{2,n} \\ X_{3,1} & X_{3,2} & X_{3,3} & \cdots & X_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{k,1} & X_{k,2} & X_{k,3} & \cdots & X_{k,n} \end{pmatrix} \begin{pmatrix} 1 & X_{2,1} & X_{3,1} & \cdots & X_{k,1} \\ 1 & X_{2,2} & X_{3,2} & \cdots & X_{k,2} \\ 1 & X_{2,3} & X_{3,3} & \cdots & X_{k,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2,n} & X_{3,n} & \cdots & X_{k,n} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{2,1} & X_{2,2} & X_{2,3} & \cdots & X_{2,n} \\ X_{3,1} & X_{3,2} & X_{3,3} & \cdots & X_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{k,1} & X_{k,2} & X_{k,3} & \cdots & X_{k,n} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} =$$

$$\begin{pmatrix} n & \sum_{i=1}^n X_{2,i} & \sum_{i=1}^n X_{3,i} & \cdots & \sum_{i=1}^n X_{k,i} \\ \sum_{i=1}^n X_{2,i} & \sum_{i=1}^n X_{2,i}^2 & \sum_{i=1}^n X_{2,i}X_{3,i} & \cdots & \sum_{i=1}^n X_{2,i}X_{k,i} \\ \sum_{i=1}^n X_{3,i} & \sum_{i=1}^n X_{2,i}X_{3,i} & \sum_{i=1}^n X_{3,i}^2 & \cdots & \sum_{i=1}^n X_{3,i}X_{k,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{k,i} & \sum_{i=1}^n X_{2,i}X_{k,i} & \sum_{i=1}^n X_{3,i}X_{k,i} & \cdots & \sum_{i=1}^n X_{k,i}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{2,i}Y_i \\ \sum_{i=1}^n X_{3,i}Y_i \\ \vdots \\ \sum_{i=1}^n X_{k,i}Y_i \end{pmatrix}$$

1ª Propiedad.-

La suma de los residuos mínimo cuadráticos es cero, analíticamente:

$$e_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2,t} - \hat{\beta}_3 X_{3,t} - \cdots - \hat{\beta}_k X_{k,t}$$

$$\sum_{t=1}^n e_t = \sum_{t=1}^n (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2,t} - \hat{\beta}_3 X_{3,t} - \cdots - \hat{\beta}_k X_{k,t}) =$$

$$\sum_{t=1}^n Y_t - \underbrace{\sum_{t=1}^n \hat{\beta}_1}_{=n\hat{\beta}_1} - \hat{\beta}_2 \sum_{t=1}^n X_{2,t} - \cdots - \hat{\beta}_k \sum_{t=1}^n X_{k,t} = 0$$

2ª Propiedad.-

La suma de los valores observados es igual a la suma de los valores estimados:

$$\text{Si } \sum_{t=1}^n e_t = 0 \Rightarrow \sum_{t=1}^n (Y_t - \hat{Y}_t) = 0$$

$$\Downarrow$$

$$\sum_{t=1}^n Y_t = \sum_{t=1}^n \hat{Y}_t \Rightarrow \frac{\sum_{t=1}^n Y_t}{n} = \frac{\sum_{t=1}^n \hat{Y}_t}{n}$$

$$\Downarrow$$

$$\bar{Y} = \bar{\hat{Y}}$$

3ª Propiedad

El hiperplano de la regresión pasa por un punto muy concreto, que es el punto cuyas coordenadas son

$$n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n X_{2,i} + \hat{\beta}_3 \sum_{i=1}^n X_{3,i} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{k,i} = \sum_{i=1}^n Y_i$$

$$\Downarrow$$

$$\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \dots + \hat{\beta}_k \bar{X}_k = \bar{Y}$$

4ª Propiedad.-

Los momentos de 2º orden entre cada regresor y los residuos son igual a 0. Esto quiere decir que:

$$X' \bar{e} = 0$$

DEMOSTRACIÓN:

$$X \cdot \bar{e} = X' \left(\bar{y} - X \hat{\beta} \right) = X \bar{y} - X X' \underbrace{(X X')^{-1} X \bar{y}}_{=\hat{\beta}} = X \bar{y} - X \bar{y} = \bar{0}$$

5ª Propiedad.-

Los momentos de 2º orden entre los valores estimados para \hat{y} y los residuos son 0.

$$\hat{y}' \bar{e} = 0$$

DEMOSTRACIÓN:

$$\hat{y}' \cdot \bar{e} = \left(X \hat{\beta} \right)' \bar{e} = \hat{\beta}' \underbrace{X' \bar{e}}_{=0 \text{ por la propiedad 4)} = \hat{\beta}' \bar{0} = \bar{0}$$

PROPIEDADES ESTOCÁSTICAS DEL ESTIMADOR MCO

En primer lugar hemos de considerar que el estimador es en realidad una variable aleatoria, al serlo la perturbación aleatoria

1)

Se trata de un estimador lineal respecto de los valores observados. (trivial)

2)

El estimador de mínimos cuadrados ordinarios es un estimador insesgado, es decir, que el valor esperado coincide con el parámetro que queremos estimar.

$$\begin{aligned}\hat{\beta} &= [(X'X)^{-1} X'] \bar{y}, \quad \bar{y} = X\bar{\beta} + \bar{u} \\ \hat{\beta} &= [(X'X)^{-1} X'] (X\bar{\beta} + \bar{u}) = (X'X)^{-1} (X'X)\bar{\beta} + (X'X)^{-1} X'\bar{u} \\ \hat{\beta} &= \bar{\beta} + (X'X)^{-1} X'\bar{u}\end{aligned}$$

como \bar{u} es una variable aleatoria, $\hat{\beta}$ también tiene que serlo por ser una combinación lineal de \bar{u} .

$$\begin{aligned}E[\hat{\beta}] &= \underbrace{\beta + (X'X)^{-1} X'}_{\substack{\text{son números, por eso} \\ \text{salen fuera de la } E[\]}} \underbrace{E[\bar{u}]}_{=0} = \beta + \underbrace{(X'X)^{-1} X \cdot \bar{0}}_{=0} = \beta \\ &\Downarrow \\ E[\hat{\beta}] &= \beta\end{aligned}$$

$$\begin{aligned}Cov(\hat{\beta}) &= E\left[\left(\hat{\beta} - E[\hat{\beta}]\right)\left(\hat{\beta} - E[\hat{\beta}]\right)'\right] = E\left[\left(\hat{\beta} - \bar{\beta}\right)\left(\hat{\beta} - \bar{\beta}\right)'\right] = \\ &= E\left[\left((X'X)^{-1} X'\bar{u}\right)\left((X'X)^{-1} X'\bar{u}\right)'\right] = E\left[\left((X'X)^{-1} X'\bar{u}\right)\left(\bar{u}'X(X'X)^{-1}\right)\right] = \\ &= E\left[\left((X'X)^{-1} X'\bar{u}\bar{u}'X(X'X)^{-1}\right)\right] = (X'X)^{-1} X' \underbrace{E[\bar{u}\bar{u}']}_{=\sigma^2 I} X (X'X)^{-1} = \\ &= (X'X)^{-1} X' \sigma^2 I X (X'X)^{-1} = \sigma^2 (X'X)^{-1} \underbrace{X'X(X'X)^{-1}}_{=I} = \sigma^2 (X'X)^{-1} \\ &\Downarrow \\ Cov(\hat{\beta}) &= \sigma^2 (X'X)^{-1} \Rightarrow \begin{cases} Var(\hat{\beta}_i) = \sigma^2 a_{i,i} \\ Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a_{i,j} \end{cases}\end{aligned}$$

Vamos ahora a buscar el estimador de la varianza

Lógicamente la única pista con datos numéricos nos la va a proporcionar los residuos, por esta razón vamos a calcular:

$$\begin{aligned}SCR &= \sum e_i = \bar{e}'\bar{e} \\ \bar{y} &= X\bar{\beta} + \bar{u} \\ \hat{\bar{y}} &= X\hat{\beta}\end{aligned}$$

$$\begin{aligned}
\bar{e} &= \bar{y} - X\hat{\beta} = \bar{y} - \hat{\bar{y}} = (X\bar{\beta} + \bar{u}) - X(X'X)^{-1} X'\bar{y} = \\
&= (X\bar{\beta} + \bar{u}) - X(X'X)^{-1} X'(X\bar{\beta} + \bar{u}) = \\
&= X\bar{\beta} + \bar{u} - \underbrace{X(X'X)^{-1} X'X}_{=I} \bar{\beta} - X(X'X)^{-1} X'\bar{u} = \\
&= X\bar{\beta} + \bar{u} - X\bar{\beta} - X(X'X)^{-1} X'\bar{u} = \\
&= \bar{u} - X(X'X)^{-1} X'\bar{u} = \\
&= (\mathbf{I} - X(X'X)^{-1} X')\bar{u} = M\bar{u} \\
&\Rightarrow M = (\mathbf{I} - X(X'X)^{-1} X')
\end{aligned}$$

Tenemos que demostrar que la matriz M cumple las siguientes características: simétrica, idempotente y semidefinida positiva.

SIMÉTRICA:

$$M' = \mathbf{I}' - (X(X'X)^{-1} X')' = \mathbf{I} - (X(X'X)^{-1} X') = M$$

IDEMPOTENTE:

$$\begin{aligned}
MM &= (\mathbf{I} - X(X'X)^{-1} X')(\mathbf{I} - X(X'X)^{-1} X') = \\
&= \mathbf{I} - X(X'X)^{-1} X' - X(X'X)^{-1} X' + X(X'X)^{-1} \underbrace{X'X(X'X)^{-1}}_{=I} X' = \\
&= \mathbf{I} - X(X'X)^{-1} X' - X(X'X)^{-1} X' + X(X'X)^{-1} X' = \\
&= \mathbf{I} - X(X'X)^{-1} X' = M
\end{aligned}$$

SEMIDEFINIDA POSITIVA:

$$\bar{e}' = \bar{u}'M' = \bar{u}'M \Rightarrow \bar{e}'\bar{e} = \bar{u}' \underbrace{MM}_{\substack{=M \\ \text{por ser} \\ \text{idempotente}}} \bar{u} = \bar{u}'M\bar{u} = \sum_{i=1}^n e_i^2 \geq 0$$

$\Rightarrow M$ es semidefinida positiva.

$$\begin{aligned} \bar{e}'\bar{e} &= \bar{u}'M\bar{u} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \\ &= \sum_{i=1}^n m_{i,i} u_i^2 + \sum_{i \neq j} m_{i,j} u_i u_j \\ E[\bar{e}'\bar{e}] &= E \left[\sum_{i=1}^n m_{i,i} u_i^2 + \sum_{i \neq j} m_{i,j} u_i u_j \right] = \sum_{i=1}^n m_{i,i} E(u_i)^2 + \sum_{i \neq j} m_{i,j} E(u_i u_j) \\ &= \sum_{i=1}^n m_{i,i} \sigma^2 = \sigma^2 \sum_{i=1}^n m_{i,i} = \sigma^2 \text{TRAZA}(M) \xrightarrow{(1)} = \bar{e}'\bar{e} = \sigma^2(n-k) \end{aligned}$$

(1)

$$\begin{aligned} \text{TRAZA}(M) &= \text{TRAZA}(\mathbf{I}_n - X_{n \times k} (X'X)^{-1}_{k \times k} X'_{k \times n}) = \\ &= \text{TRAZA}(\mathbf{I}_n) - \text{TRAZA}(X_{n \times k} (X'X)^{-1}_{k \times k} X'_{k \times n}) = \\ &= \text{TRAZA}(\mathbf{I}_n) - \text{TRAZA}(\underbrace{X'_{k \times n} X_{n \times k} (X'X)^{-1}_{k \times k}}_{=\mathbf{I}_k}) = \\ &= \text{TRAZA}(\mathbf{I}_n) - \text{TRAZA}(\mathbf{I}_k) = n - k \end{aligned}$$

Con todo esto podemos concluir que: $\hat{\sigma}^2 = \frac{\bar{e}'\bar{e}}{n-k}$, es el estimador de la varianza y que presenta las siguientes CARACTERÍSTICAS:

INSESGADO:

$$E[\hat{\sigma}^2] = E \left[\frac{\bar{e}'\bar{e}}{n-k} \right] = \frac{1}{n-k} E[\bar{e}'\bar{e}] = \frac{1}{n-k} (n-k)\sigma^2 = \sigma^2$$

NO LINEAL:

El estimador de la varianza no es lineal ya que la expresión no se obtiene de forma lineal. $e_i^2 = (y_i - \hat{y}_i)^2$

Este método de estimar la varianza resulta muy largo y complejo, por lo que buscaremos un método más sencillo de hacer esta estimación. Y como lo difícil es calcular la SCR, vamos a tratar de agilizar dicho proceso:

$$\begin{aligned}
\bar{e}'\bar{e} &= (\bar{y} - X\hat{\beta})'(\bar{y} - X\hat{\beta}) = (\bar{y}' - \hat{\beta}'X')(\bar{y} - X\hat{\beta}) = \\
&= \bar{y}'\bar{y} - \bar{y}'X\hat{\beta} - \hat{\beta}'X'\bar{y} + \hat{\beta}'X'X\hat{\beta} = \\
&= \bar{y}'\bar{y} - \bar{y}'X\hat{\beta} - \hat{\beta}'X'\bar{y} + \hat{\beta}'X'X(X'X)^{-1}X'\bar{y} = \\
&= \bar{y}'\bar{y} - \bar{y}'X\hat{\beta} - \hat{\beta}'X'\bar{y} + \hat{\beta}'X'\bar{y} = \\
&= \bar{y}'\bar{y} - \bar{y}'X\hat{\beta} = \\
&= \bar{y}'\bar{y} - 2\hat{\beta}'X'\bar{y} + \hat{\beta}'X'\bar{y} = \\
&= \bar{y}'\bar{y} - \hat{\beta}'X'\bar{y} = \bar{e}'\bar{e}
\end{aligned}$$

TEOREMA DE GAUSS-MARCOV

Vamos a calcular un estimador de varianza mínima al cual vamos a llamar estimador óptimo.

Nos encontramos con el problema de que el estimador $\hat{\beta}$ es un vector y la varianza de un vector aleatorio es una matriz para las cuales no hay establecida una relación de orden. Por tanto, para comparar la varianza tenemos que irnos a la varianza de los componentes del vector $\bar{\beta}$ y para ello vamos a recurrir a una combinación lineal.

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} \quad \bar{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, c_i \in \mathfrak{R}$$

$$\bar{c}'\bar{\beta} = (c_1 \quad c_2 \quad \cdots \quad c_n) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = c_1\beta_1 + c_2\beta_2 + \cdots + c_n\beta_n$$

Podemos demostrar que: $\bar{c}'\hat{\beta}$ es un estimador lineal e insesgado.

Le linealidad es evidente:

$$\bar{c}'\hat{\beta} = (c_1 \quad c_2 \quad \cdots \quad c_n) \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = c_1\hat{\beta}_1 + c_2\hat{\beta}_2 + \cdots + c_n\hat{\beta}_n$$

INSESGADEZ:

$$\begin{aligned}
E[\bar{c}'\hat{\beta}] &= \bar{c}'E[\hat{\beta}] = \bar{c}'\bar{\beta} \\
Var[\bar{c}'\hat{\beta}] &= E\left[\left[(\bar{c}'\hat{\beta}) - E[\bar{c}'\hat{\beta}]\right]\left[(\bar{c}'\hat{\beta}) - E[\bar{c}'\hat{\beta}]\right]'\right] = \\
&\left(\text{como}\left((\bar{c}'\hat{\beta}) - E[\bar{c}'\hat{\beta}]\right) = \bar{c}'\hat{\beta} - \bar{c}'\bar{\beta} = \bar{c}'(\hat{\beta} - \bar{\beta})\right) \\
&= E\left[\bar{c}'(\hat{\beta} - \bar{\beta})(\hat{\beta} - \bar{\beta})'\bar{c}\right] = \bar{c}'E\left[(\hat{\beta} - \bar{\beta})(\hat{\beta} - \bar{\beta})'\right]\bar{c} = \\
&= \bar{c}'Cov(\hat{\beta})\bar{c} = \sigma^2\bar{c}'(X'X)^{-1}\bar{c} \\
\Rightarrow Var[\bar{c}'\hat{\beta}] &= \sigma^2\bar{c}'(X'X)^{-1}\bar{c}
\end{aligned}$$

Si ahora elegimos un estimador cualquiera, tal como \hat{b} , se cumple que:

$$\hat{b} = \bar{a}'\bar{y} \Leftarrow \text{por ser lineal}$$

$$\text{Por ser insesgado, se debe cumplir que: } E[\hat{b}] = E[\bar{a}'\bar{y}] = \bar{c}'\bar{\beta}$$

$$E[\bar{a}'\bar{y}] = E[\bar{a}'(X\bar{\beta} + \bar{u})] = E[\bar{a}'X\bar{\beta}] + \underbrace{\bar{a}'E[\bar{u}]}_{=0} = \bar{a}'X\bar{\beta}$$

para que sea insesgado $\bar{a}'X\bar{\beta}$ tiene que ser igual a $\bar{c}'\bar{\beta}$, por tanto

$$\bar{a}'X\bar{\beta} = \bar{c}'\bar{\beta} \Rightarrow \bar{a}'X = \bar{c}' \text{ y en consecuencia :}$$

$$\left. \begin{aligned}
\hat{b} &= \bar{a}'(X\bar{\beta} + \bar{u}) \\
\hat{b} &= \bar{a}'X\bar{\beta} + \bar{a}'\bar{u} \\
\hat{b} &= \bar{c}'\bar{\beta} + \bar{a}'\bar{u} \\
\hat{b} - E[\hat{b}] &= \bar{c}'\bar{\beta} + \bar{a}'\bar{u} - \bar{c}'\bar{\beta} \\
\hat{b} - E[\hat{b}] &= \bar{a}'\bar{u}
\end{aligned} \right\} \Rightarrow Var(\hat{b}) = E[\bar{a}'\bar{u}\bar{u}'\bar{a}] = \bar{a}'E[\bar{u}\bar{u}']\bar{a} = \sigma^2\bar{a}'\bar{1}\bar{a}$$

Y si restamos las expresiones $Var(\hat{b})$ y $Var(\bar{c}'\hat{\beta})$, tenemos que:

$$Var(\hat{b}) = E[\bar{a}'\bar{u}\bar{u}'\bar{a}] = \bar{a}'E[\bar{u}\bar{u}']\bar{a} = \sigma^2\bar{a}'\mathbf{1}\bar{a}$$

-

$$Var(\bar{c}'\hat{\beta}) = \sigma^2\bar{a}'X(X'X)^{-1}X'\bar{a}$$

=

$$\begin{aligned} Var(\hat{b}) - Var(\bar{c}'\hat{\beta}) &= \sigma^2\bar{a}'\mathbf{1}\bar{a} - \sigma^2\bar{a}'X(X'X)^{-1}X'\bar{a} = \\ &= \sigma^2(\bar{a}'\mathbf{1}\bar{a} - \bar{a}'X(X'X)^{-1}X'\bar{a}) = \sigma^2\bar{a}'(\mathbf{I} - X(X'X)^{-1}X')\bar{a} = \\ &= \sigma^2\bar{a}'M\bar{a} \geq 0 \end{aligned}$$

Luego si $Var(\hat{b}) - Var(\bar{c}'\hat{\beta})$ siempre es mayor o igual que cero, podemos afirmar que $Var(\hat{b})$ siempre será mayor, o como mucho igual a $Var(\bar{c}'\hat{\beta})$, y así queda demostrado que $(\bar{c}'\hat{\beta})$ es el estimador de mínima varianza.

Como el vector c es un vector cualquiera, elegimos como tal el vector

$$\bar{c} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \bar{c}'\bar{\beta} = (1 \quad 0 \quad \dots \quad 0) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \beta_1,$$

y generalizando,

$$\bar{c}'\bar{\beta} = \begin{pmatrix} 0 & \dots & 0 & \overset{i}{1} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \beta_i \quad i = 1, 2, \dots, k$$

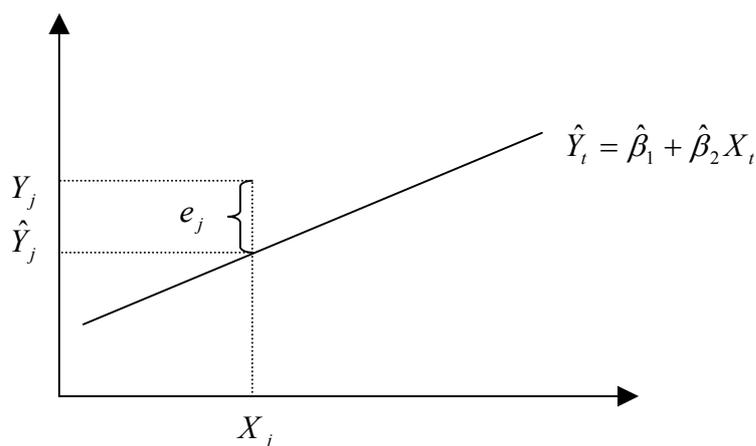
$$\bar{c}'\hat{\beta} = \begin{pmatrix} 0 & \dots & 0 & \overset{i}{1} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \hat{\beta}_i \quad i = 1, 2, \dots, k$$

Con esto queda demostrado que para $\beta_i \quad i = 1, 2, \dots, k$, el estimador ELIO, es decir, estimador lineal insesgado y óptimo es $\hat{\beta}_i$.

COEFICIENTE DE DETERMINACION.-

Para medir la adecuación o no de un ajuste, se suele utilizar la siguiente descomposición:

$$SCT = SCR + SCE$$



Podemos comprobar:

$$\begin{aligned} (Y_j - \bar{Y})^2 &= (\hat{Y}_j - \bar{Y} + e_j)^2 \\ \overbrace{\sum_{j=1}^n (Y_j - \bar{Y})^2}^{SCT} &= \sum_{j=1}^n (\hat{Y}_j - \bar{Y} + e_j)^2 = \\ &= \underbrace{\sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2}_{SCE} + \underbrace{\sum_{j=1}^n e_j^2}_{SCR} + 2 \sum_{j=1}^n (\hat{Y}_j - \bar{Y}) \cdot e_j \end{aligned}$$

Para que nuestra descomposición se cumpla tenemos que demostrar que

$$\sum_{j=1}^n (\hat{Y}_j - \bar{Y}) \cdot e_j = 0, \text{ y para ello, que } \sum_{j=1}^n \hat{Y}_j e_j - \bar{Y} \underbrace{\sum_{j=1}^n e_j}_{=0} = 0 \Rightarrow \sum_{j=1}^n \hat{Y}_j e_j = 0 \Rightarrow \hat{y}'\bar{e} = 0, \text{ y}$$

esta última expresión se da siempre ya que es la 5ª propiedad de las que se deducen de las ecuaciones normales

La descomposición, por tanto, queda del siguiente modo:

$$\overbrace{\sum_{j=1}^n (y_j - \bar{y})^2}^{SCT} = \overbrace{\sum_{j=1}^n (\hat{y}_j - \bar{y})^2}_{SCE} + \overbrace{\sum_{j=1}^n e_j^2}_{SCR}$$

El ajuste será tanto mejor cuanto mayor sea la SCE, y menor sea la SCR, ya que cuando la SCR es nula el ajuste es perfecto.

Todo esto da lugar a definir el coeficiente de determinación $R^2 = \frac{SCE}{SCT}$, cuyo valor oscila entre cero y la unidad, es decir, $0 \leq R^2 \leq 1$. Cuanto mas cerca de la unidad esté el valor del coeficiente, mejor será el ajuste de la regresión.

Podemos a continuación obtener una serie de expresiones que nos ayudaran a realizar los cálculos de forma simplificada:

$$\begin{aligned} \sum (Y_i - \bar{Y})^2 &= \sum (Y_i^2 + \bar{Y}^2 + 2\bar{Y}Y_i) = \sum Y_i^2 + n\bar{Y}^2 + 2\bar{Y} \underbrace{\sum Y_i}_{n\bar{Y}} = \\ &= \sum Y_i^2 + n\bar{Y}^2 - 2n\bar{Y}^2 = \sum Y_i^2 - n\bar{Y}^2 = \bar{y}'\bar{y} - n \left(\frac{\sum Y_i}{n} \right)^2 = \end{aligned}$$

$$SCT = \bar{y}'\bar{y} - \frac{(\sum y_i)^2}{n}$$

-

$$SCR = \bar{y}'\bar{y} - \hat{\beta}'X'\bar{y}$$

=

$$SCE = \hat{\beta}'X'\bar{y} - \frac{(\sum Y_i)^2}{n}$$

$$R^2 = \frac{\hat{\beta}'X'\bar{y} - \frac{(\sum Y_i)^2}{n}}{\bar{y}'\bar{y} - \frac{(\sum Y_i)^2}{n}}$$

Existe una inconveniente a la hora de utilizar el R^2 como medida de la bondad del ajuste de la regresión, y es el hecho de que este sea manipulable, si observamos su formación, podemos detectar que cada vez que añadimos un parámetro a la regresión le damos a ésta una mayor libertad y en consecuencia el coeficiente de determinación muestra un mejor ajuste aunque realmente no sea así en el contexto del problema, ya que la variable introducida puede no tener ninguna relación con la variable explicada

Este hecho podemos demostrarlo fácilmente:

$$\hat{\sigma}^2 = \frac{SCR}{n-k} \Rightarrow SCR = (n-k)\hat{\sigma}^2 \Rightarrow R^2 = 1 - \frac{(n-k)\hat{\sigma}^2}{SCT}, \text{ en esta última expresión}$$

podemos observar que cuando aumentamos k, es decir, el número de parámetros, el coeficiente de determinación aumenta.

Por otro lado podemos escribir:

$$\bar{e} = \bar{y} - \hat{y} = \bar{y} - X\hat{\beta} = \bar{y} - X(X'X)^{-1}X'\bar{y}$$

SI aumentamos el numero de parámetros k, hasta que este tome el valor del numero de observaciones n, la matriz X, será una matriz cuadrada y tendrá inversa, la expresión anterior se convertirá en:

$$\bar{e} = \bar{y} - XX^{-1}X'\bar{y} = \bar{y} - \bar{y} = \bar{0}$$

Y entonces se cumplirá que $SCR = 0$

Para evitar esto, vamos a definir el coeficiente de determinación corregido: \bar{R}^2 , este coeficiente difiere del coeficiente de determinación en que en lugar de usar las varianzas, usamos las cuasivarianzas, lo cual hace que dicho coeficiente no sea manipulable, y, además, nos permite comparar los de varias regresiones.

Podemos comprobar que el coeficiente de determinación se puede expresar como:

$$R^2 = 1 - \frac{SCR}{SCT}$$

Si dividimos los miembros del cociente por sus grados de libertad obtenemos:

$$\bar{R}^2 = 1 - \frac{\frac{SCR}{n-k}}{\frac{SCT}{n-1}} = \frac{1-k}{n-k} + \frac{n-1}{n-k} R^2$$

El coeficiente de determinación corregido nos servirá para comparar modelos que utilicen distintos números de variables explicativas, elegiremos aquel cuyo coeficiente de determinación corregido sea mayor:

HIPÓTESIS DE NORMALIDAD Y ESTIMACIÓN MÁXIMO VEROSÍMIL.

FUNDAMENTO TEÓRICO:

- Debemos de suponer que la perturbación aleatoria sigue una distribución normal multivariante
- No conocemos los parámetros de la distribución asociada a la variable aleatoria \bar{y} , solamente una muestra (los valores observados), y entre todos los posibles parámetros asociados a la distribución de \bar{y} , vamos a elegir aquellos que cumplan la condición de que el resultado conocido, es decir la muestra, sea, en teoría, el resultado con mayor probabilidad..

$$\bar{u} \rightarrow N(\bar{0}, \sigma^2 \mathbf{I})$$

$$E[\bar{u}] = \bar{0}$$

$$E[u_i] = 0 \quad \forall i$$

$$E[\bar{u}'\bar{u}] = \sigma^2 \mathbf{I} \Rightarrow \begin{cases} E[u_i u_j] = 0 & i \neq j \\ E[u_i^2] = \sigma^2 \end{cases}$$

El motivo por el cual hemos elegido la distribución normal es porque no existen desarrollos para otras distribuciones.

$$f(\bar{u}) = \frac{1}{(\sqrt{2\Pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\bar{u}'\bar{u}} = \frac{1}{(2\Pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\bar{u}'\bar{u}}$$

Realicemos ahora el siguiente cambio de variable:

$$\bar{y} = X\bar{\beta} + \bar{u} \Rightarrow \begin{cases} \bar{u} = \bar{y} - X\bar{\beta} \\ \bar{u}' = \bar{y}' - \bar{\beta}'X' \end{cases}$$

JACOBIANO:

$$\left| \frac{\partial \bar{y}}{\partial \bar{u}} \right| = \begin{vmatrix} \frac{\partial Y_1}{\partial u_1} & \frac{\partial Y_1}{\partial u_2} & \dots & \frac{\partial Y_1}{\partial u_n} \\ \frac{\partial Y_2}{\partial u_1} & \frac{\partial Y_2}{\partial u_2} & \dots & \frac{\partial Y_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Y_n}{\partial u_1} & \frac{\partial Y_n}{\partial u_2} & \dots & \frac{\partial Y_n}{\partial u_n} \end{vmatrix} = \frac{\partial(Y_1 \ Y_2 \ \dots \ Y_n)}{\partial(u_1 \ u_2 \ \dots \ u_n)} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Queremos maximizar la siguiente expresión:

$$f(Y_1 \ Y_2 \ \dots \ Y_n) = \frac{1}{(2\Pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(\bar{y}' - \bar{\beta}'X')(\bar{y} - X\bar{\beta})}$$

Para llevar a cabo la maximización vamos a tomar logaritmos, ya que si una función tiene un máximo en un punto, el logaritmo de la misma tiene a su vez un máximo en el mismo punto.

$$\begin{aligned} \ln f(y_1 \ y_2 \ \dots \ y_n) &= -\ln(2\Pi\sigma^2)^{n/2} - \frac{1}{2\sigma^2}(\bar{y}' - \bar{\beta}'X')(\bar{y} - X\bar{\beta}) = \\ &= -\ln(2\Pi\sigma^2)^{n/2} - \frac{1}{2\sigma^2}(\bar{y}'\bar{y} - \bar{y}'X\bar{\beta} - \bar{\beta}'X'\bar{y} + \bar{\beta}'X'X\bar{\beta}) = \\ &= -\ln(2\Pi\sigma^2)^{n/2} - \frac{1}{2\sigma^2}(\bar{y}'\bar{y} - 2\bar{\beta}'X'\bar{y} + \bar{\beta}'X'X\bar{\beta}) \end{aligned}$$

Para encontrar el máximo de la función debemos derivar la expresión anterior e igualarla a 0.

$$\begin{aligned} \frac{\partial f(Y_1 \ Y_2 \ \dots \ Y_n)}{\partial \bar{\beta}'} &= 0 \Rightarrow -\frac{1}{2\sigma^2}(-2X'\bar{y} + 2(X'X)\bar{\beta}) = 0 \Rightarrow (X'X)\bar{\beta} = X'\bar{y} \\ \Rightarrow \tilde{\beta}_{MV} &= (X'X)^{-1}X'\bar{y} \text{ es el estimador de máxima verosimilitud.} \end{aligned}$$

$$\tilde{\beta}_{MV} = (X'X)^{-1}X'\bar{y} = \hat{\beta}_{MCO}$$

Para obtener el estimador de la varianza, vamos a derivar el logaritmo con respecto a la misma, es decir:

$$\frac{\partial f(Y_1 \ Y_2 \ \dots \ Y_n)}{\partial \sigma^2} = \frac{n}{2} \frac{1}{2\Pi\sigma^2} 2\Pi + \frac{1}{2} \frac{1}{(\sigma^2)^2} \bar{e}'\bar{e} = 0$$

$$\Rightarrow \tilde{\sigma}_{MV}^2 = \frac{\bar{e}'\bar{e}}{n} \neq \hat{\sigma}_{MCO}^2 = \frac{\bar{e}'\bar{e}}{n-k}$$

$$E[\tilde{\sigma}_{MV}^2] = E\left[\frac{\bar{e}'\bar{e}}{n}\right] = \frac{E[\bar{e}'\bar{e}]}{n} = \frac{(n-k)\sigma^2}{n} = \sigma^2 - \frac{k}{n}\sigma^2$$

Este estimador no es insesgado, pero si estudiamos sus características asintóticas, podemos observar que:

$$\lim_{n \rightarrow \infty} (\tilde{\sigma}_{MV}^2) = \lim_{n \rightarrow \infty} \left(\sigma^2 - \frac{k}{n}\sigma^2 \right) = \sigma^2$$

OTRAS CONCLUSIONES DE INTERES

Si hemos hecho la hipótesis de normalidad, entonces:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'y = (X'X)^{-1} X'(X\bar{\beta} + \bar{u}) \Rightarrow \\ \Rightarrow \hat{\beta} &= \bar{\beta} + (X'X)^{-1} X'\bar{u} \text{ y } \bar{u} \rightarrow N(\bar{0}, \sigma^2 I) \\ \Rightarrow \hat{\beta} &\rightarrow \text{Normal multivariante} \end{aligned}$$

$$\begin{aligned} E[\hat{\beta}] &= \bar{\beta} \text{ y } Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} \\ &\Downarrow \\ \hat{\beta} &\rightarrow N(\bar{\beta}, \sigma^2 (X'X)^{-1}) \end{aligned}$$

y esto es así debido a que:

$$\begin{aligned} \hat{\beta}_1 &\rightarrow N(\beta_1, \sigma^2 a_{1,1}) \\ &\vdots \\ \hat{\beta}_i &\rightarrow N(\beta_i, \sigma^2 a_{i,i}) \end{aligned}$$

$$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}$$