APÊNDICE C - MODELO DE REGRESSÃO LINEAR EM FORMA MATRICIAL

1-O MODELO DE REGRESSÃO LINEAR SIMPLES

O objetivo deste capítulo é escrever em forma matricial o modelo de regressão linear que vocês já viram antes. Isto torna a notação mais fácil, e nos permite tratar o problema de diversas variáveis explicativas (regressão múltipla) de uma maneira eficiente, e também permite que todo o processo de estimação e inferência seja escrito de uma maneira compacta.

Considere inicialmente o modelo de **regressão linear simples** dado pelas equações:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2.x_1 + e_1 \\ y_2 &= \beta_1 + \beta_2.x_2 + e_2 \\ &\ldots\ldots \\ y_n &= \beta_1 + \beta_2.x_n + e_n \end{aligned} \qquad (1)$$

Em primeiro lugar, o que as equações (1) significam?

- De um ponto de vista puramente operacional, e sem saber nada de estatística, temos um conjunto de n pares de pontos (x1, y1), (x2, y2), ..., (xn, yn) e queremos passar por estes pontos a "melhor reta" segundo algum critério.
- O critério utilizado para encontrar esta reta é o de mínimos quadrados ordinários (MQO)
- "Encontrar" a reta significa obter estimativas para os coeficientes desconhecidos $\beta 1$ e $\beta 2$ que aparecem nas equações (1) acima.

Mas, até agora não falamos nada do modelo estatístico que está por trás desta estimação. No fim das contas, uma pergunta relevante é: por que eu decido encontrar esta "melhor" reta por mínimos quadrados, e quais as implicações disso?

O modelo subjacente às equações (1) supõe que os erros e_i são independentes e identicamente distribuídos (iid) com média zero e variância constante σ^2. Supor a normalidade dos erros também é uma ótima idéia, pois me permite construir intervalos de confiança e fazer testes de hipóteses, mas as duas primeiras hipóteses são as verdadeiramente cruciais.

Sob as hipóteses de que os erros têm médias zero e variâncias constantes, então o que podemos dizer sobre a média e a variância dos yi's?

Aqui supomos (e isso é importante!), que os x's são fixos (não aleatórios). Então para todos os i=1,2...n:

$E(y_i) = \beta_1 + \beta_2.x_i$

$VAR(y_i) = VAR(e_i) = \sigma^2$

Quantos parâmetros existem no modelo representado pelas equações (1)? Três, a saber: β_1, β_2 e σ^2, ou seja, os parâmetros que definem a reta e a variância dos erros.

O próximo passo é reescrever as equações (1) em forma matricial. Veremos que isso não é muito difícil.

Sejam:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ .... \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \multicolumn{2}{c}{........} \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \qquad (2)$$

Note que Y é um vetor n x 1, X (a matriz de dados) é n x 2, β (o vetor de parâmetros desconhecidos da reta) é 2 x 1, assim como ε.

Então, a partir das definições (2) podemos reescrever as equações do sistema (1) como:

$$\mathbf{Y = X.\ \beta + \varepsilon} \qquad (3)$$

A equação (3) é a forma matricial do sistema (1), e pode ser facilmente estendida para o caso de mais de uma variável explicativa, bastando alterar as dimensões de X e β.

E como funciona a estimação num modelo de regressão simples (1) (ou (3))? Você deve se lembrar do **método de mínimos quadrados ordinários (MQO)**, em que o objetivo é encontrar a reta que minimiza a soma dos quadrados dos **resíduos**.

Mas, o que são mesmo os resíduos? São as diferenças entre os yi's reais e os valores que os estimam usando a reta obtida por MQO. Especificamente, o i-ésimo resíduo é:

$$y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2.x_i \quad \text{onde } \hat{\beta}_1 \text{ e } \hat{\beta}_2 \text{ são os coeficientes da reta estimados por MQO} \qquad (4)$$

Já que o objetivo do método é minimizar a soma dos quadrados dos resíduos, buscamos $\hat{\beta}_1$ e $\hat{\beta}_2$ tais que:

$$RSS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \quad (5)$$

Seja mínima. A quantidade RSS é chamada de **soma de quadrado dos resíduos**.

Note que a expressão (5) acima é bastante geral, não se aplica apenas ao caso de uma variável explicativa. O que estamos dizendo é que, no método de MQO, o objetivo é encontrar os coeficientes (que aparecem "dentro" da expressão dos resíduos) de tal forma que a gente minimize RSS.

Reescrevendo (5) para o caso particular do modelo de regressão simples leva a:

$$RSS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 - \hat{\beta}_2 . x_i \right)^2 = \sum_{i=1}^{n} \left(y_i^{\,2} + \hat{\beta}_1^{\,2} + 2.\hat{\beta}_1 . \hat{\beta}_2 . x_i + \hat{\beta}_2^{\,2} . x_i^{\,2} - 2.y_i . \left(\hat{\beta}_1 + \hat{\beta}_2 . x_i \right) \right) \quad (6)$$

E como minimizar (6)?

A condição de primeira ordem é derivar em relação aos $\hat{\beta}_i$ e igualar as derivadas a zero. A rigor você deveria verificar também o sinal da segunda derivada....

- Derivada com relação a $\hat{\beta}_1$

$$\begin{aligned} &\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \Rightarrow \sum_{i=1}^{n} \left(2\hat{\beta}_1 + 2\hat{\beta}_2 . x_i - 2y_i \right) = 0 \Rightarrow 2n.\hat{\beta}_1 + 2\hat{\beta}_2 \sum_{i=1}^{n} x_i - 2.\sum_{i=1}^{n} y_i = \\ &\Rightarrow 2n\hat{\beta}_1 + 2n\hat{\beta}_2 . \bar{x} - 2.n.\bar{y} = 0 \Rightarrow \bar{y} = \hat{\beta}_1 + \hat{\beta}_2 . \bar{x} \end{aligned} \quad (7)$$

  Ou seja, a reta ajustada passa pelo ponto médio $(\bar{x}, \bar{y})$.

- Derivada com relação a $\hat{\beta}_2$

$$\begin{aligned} &\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 \Rightarrow \sum_{i=1}^{n} \left(2\hat{\beta}_1 . x_i + 2\hat{\beta}_2 . x_i^2 - 2y_i . x_i \right) = 0 \Rightarrow \sum_{i=1}^{n} \left\{ 2x_i \left(\hat{\beta}_1 - y_i \right) + 2\hat{\beta}_2 . x_i^2 \right\} \\ &\Rightarrow \sum_{i=1}^{n} \left\{ x_i \left(\hat{\beta}_1 - y_i \right) + \hat{\beta}_2 . x_i^2 \right\} = 0 \Rightarrow n\hat{\beta}_1 . \bar{x} - \sum_{i=1}^{n} x_i . y_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0 \end{aligned}$$

Substituindo da equação (7) leva a:

$$n\hat{\beta}_1.\bar{x} - \sum_{i=1}^{n} x_i.y_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0 \Rightarrow n\left(\bar{y} - \hat{\beta}_2.\bar{x}\right)\bar{x} - \sum_{i=1}^{n} x_i.y_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0$$

$$n.\bar{x}.\bar{y} - n\hat{\beta}_2.\bar{x}^2 - \sum_{i=1}^{n} x_i.y_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \hat{\beta}_2 \left\{ \sum_{i=1}^{n} x_i^2 - n.\bar{x}^2 \right\} = \sum_{i=1}^{n} x_i.y_i - n.\bar{x}.\bar{y} \qquad (8)$$

$$\Rightarrow \hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i.y_i - n.\bar{x}.\bar{y}}{\sum_{i=1}^{n} x_i^2 - n.\bar{x}^2} = \frac{\sum_{i=1}^{n} x_i.y_i - n.\bar{x}.\bar{y}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Sejam:

$$SXY = \sum_{i=1}^{n} x_i.y_i - n.\bar{x}.\bar{y} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \text{ e } \mathrm{SXX} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (9)$$

Então, os estimadores MQO no caso da regressão linear simples podem ser escritos como:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2.\bar{x} \quad \text{e} \quad \hat{\beta}_2 = \frac{SXY}{SXX} \qquad (10)$$

Importante

Note que, sob as nossas hipóteses, os x's são fixos e assim as únivas variáveis aleatórias são os yi's (e os εi's). Pelas equações (10) note que $\hat{\beta}_2 = \frac{SXY}{SXX}$ é uma função LINEAR (embora complicada) dos y's. O mesmo se pode dizer de $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2.\bar{x}$, ele também é uma função LINEAR dos y's.

2 – O Modelo de Regressão Linear Múltipla

Suponha agora que, ao invés de n pares de pontos, temos n (k-1)-uplas de pontos, ou seja, cada medição consiste em uma variável de "saída" y_i e (k-1) variáveis explicativas x1i, x2i, ..., x(k-1)i.

Qual a "cara" do nosso modelo? Basta estender o sistema de equações (2) considerando agora a existência de k colunas na matriz do modelo X (a 1ª. coluna corresponde ao termo constante e ao coeficiente β1). Teremos uma representação matricial análoga à (2), como mostrado abaixo.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots\cdot \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{1,1} \cdots & x_{1,(k-1)} \\ 1 & x_{2,1} \cdots & x_{2,(k-1)} \\ \cdots\cdots\cdots & & \\ 1 & x_{n,1} \cdots & x_{n,(k-1)} \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{pmatrix} \tag{11}$$

Note que Y é n x 1, X é n x k, β é k x 1 e ε é n x 1. Assim, em forma matricial, o modelo de regressão linear múltipla tem exatamente a forma dada por (3).

Nota – a interpretação da equação estimada

Tão (ou mais importante) que encontrar os coeficientes que relacionam os xij e yi é interpretar o significado destes coeficientes. As estimativas dos coeficientes num modelo de regressão múltipla devem ser pensadas como tendo efeitos parciais, ou seja, "mantido tudo o mais constante". É a famosa hipótese "ceteris paribus" em Economia.

De onde vem isso? Para simplificar, considere um modelo com apenas duas variáveis explicativas:

$$y_i = \beta_1 + \beta_2 . x_{1i} + \beta_3 . x_{2.i} + \varepsilon_i$$

Onde os erros têm as hipóteses usuais de média zero, variância constante e são independentes (ou pelo menos descorrelatados). Nestas condições usamos o método de MQO para estimar os coeficientes do modelo, e encontramos neste caso particular:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 . x_{1i} + \hat{\beta}_3 . x_{2.i} \tag{12}$$

O intercepto $\hat{\beta}_1$ é o valor previsto de y_i quando ambas as variáveis explicativas são zero. Agora pense em termos de variações nos x's e no y. Podemos reescrever (12) como:

$$\Delta\hat{y}_i = \hat{\beta}_2 . \Delta x_{1i} + \hat{\beta}_3 . \Delta x_{2.i} \tag{13}$$

Esta expressão nos dá a variação prevista em y a partir das variações em x1 e x2. Em particular, se mantivermos x2 constante, e assim $\Delta x_{2,i} = 0$ segue que $\Delta\hat{y}_i = \hat{\beta}_2 . \Delta x_{1i}$. Analogamente, se x1 é fixo, a variação em y é o coeficiente de x2 vezes a variação de x2.

3 – Premissas do Modelo Linear Clássico

A questão principal agora é: quais são as hipóteses "desejáveis" para trabalharmos com um modelo como (13)? Quais são as implicações destas hipóteses.

Antes de enumerar as hipóteses do modelo linear clássico dado pelo sistema de equações (13), convém lembrar mais uma vez as hipóteses empregadas no modelo de regressão linear simples e as consequências destas.

No modelo de **regressão linear simples** supomos que:

1- Os erros do modelo têm média zero e variância constante; $E(e_i) = 0$ e $VAR(e_i) = \sigma^2$. A implicação imediata disso, dada a hipótese 3 abaixo, é que os **yi's também têm variância constante**, e sua média (condicional aos x's) é a própria equação de regressão que se quer estimar.

2- Os erros são independentes (ou, no mínimo, descorrelatados) e assim $E(e_i e_j) = E(e_i)E(e_j) = 0$.

3- Os x's são fixos (não estocásticos).

4- Se desejarmos inferir sobre a distribuição dos coeficientes estimados por MQO, geralmente supomos a Normalidade dos erros. Sob normalidade, independência e correlação zero são hipóteses equivalentes, e assim podemos escrever e_i iid $N(0, \sigma^2)$ para i=1,2,...,n.

Quais as implicações destas hipóteses?

- Os estimadores MQO são, sob as hipóteses anteriores:
    - LINEARES
    - NÃO TENDENCIOSOS
    - DENTRE OS ESTIMADORES LINEARES, ELES TÊM VARIÂNCIA MÍNIMA, ou seja, são os MELHORES estimadores lineares não tendenciosos (são BLUE = best linear unbiased estimators). Esta é a propriedade de **Gauss-Markov**.

E por que isso é importante?

- O fato de um estimador ser linear torna o seu cálculo fácil. Por pior que ele "pareça", sabe-se que ele tem a forma $\hat{\beta}_j = \sum_{l=1}^{n} k_l y_l + k_0$.
- O fato de um estimador ser não tendencioso é muito conveniente – seu valor esperado é exatamente o parâmetro desconhecido que ele pretende estimar.
- O fato dos estimadores MQO, sob as condições descritas acima, serem os melhores estimadores lineares não tendenciosos nos diz que, dentre os estimadores "fáceis" de calcular e com uma propriedade desejável (não tendenciosidade), os que encontramos são os mais eficientes, pois têm a menor variância, ou seja, a eles está associada a menor incerteza dentre todos os estimadores na mesma categoria.

Como adaptar estas hipóteses para o modelo de regressão múltipla escrito em forma matricial?

1- O vetor de erros tem média zero, i.e. E(ε) = 0 onde e é o vetor n x 1 que aparece nas definições (11).

2- A matriz de variância-covariância de ε é a matriz n x n COV = σ^2.I, onde I é a matriz identidade de dimensão n. Em outras palavras, todos os elementos da diagonal desta matriz (que são as variâncias dos erros) são iguais a σ^2 e todos os elementos fora da diagonal (que são as covariâncias entre os erros) são zero.

3- A matriz de dados X é uma matriz n x k não estocástica, ou seja, seus elementos NÃO são variáveis aleatórias.

4- O posto da matriz X é k = min(n, k) – supõe-se que existem menos variáveis (k) que observações (n), e portanto os parâmetros do modelo podem ser estimados. Esta condição indica que NÃO EXISTE COLINEARIDADE EXATA entre as variáveis explicativas, ou seja, NÃO EXISTE MULTICOLINEARIDADE entre as colunas de X. Em outras palavras, não podemos escrever qualquer coluna de X como uma combinação linear de outras colunas.

5- O vetor de erros tem distribuição Normal multivaridada de dimensão n com vetor de médias 0 e matriz de variância-covariância COV = σ^2.I.

4-ESTIMADORES MQO NO MODELO DE REGRESSÃO MÚLTIPLA

Considere o modelo de regressão múltipla escrito na forma matricial (11). Podemos escrevê-lo também na forma (3). Como obter os estimadores MQO?

Por construção, o método MQO consiste na minimização da soma do quadrado dos resíduos.

Sejam: $\hat{y} = X.\hat{\beta}$ o vetor de valores ajustados pelo modelo (de dimensão n x 1) e $\hat{\varepsilon} = y - \hat{y}$ o vetor de resíduos, também de dimensão n x 1.

A soma do quadrado dos resíduos é dada por:

$$RSS = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n).\begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \ldots \\ \hat{\varepsilon}_n \end{pmatrix} = \hat{\varepsilon}^t \hat{\varepsilon} = (y - \hat{y})^t (y - \hat{y}) = (y - X\hat{\beta})^t (y - X\hat{\beta}) \quad (14)$$

Mas, é importante usar agora uma propriedade da transposta de matrizes: para quaisquer matrizes C e D (com as dimensões corretas), então $(C.D)^t = D^t.C^t$. Usando esta propriedade acima:

$$RSS = \left(y - X\hat{\beta}\right)^t \left(y - X\hat{\beta}\right) = y^t y - \left(X\hat{\beta}\right)^t y - y^t.\left(X\hat{\beta}\right) + \left(X\hat{\beta}\right)^t \left(X\hat{\beta}\right) =$$
$$= y^t y - 2\hat{\beta}^t X^t y + \hat{\beta}^t X^t X\hat{\beta}$$

Note que, acima, ao calcular o termo intermediário, notamos que $\hat{\beta}^t X^t y$ é um escalar (e portanto igual ao seu transposto $y^t.\left(X\hat{\beta}\right)$.

Derivando a soma do quadrado dos resíduos e igualando a zero leva a (veja os detalhes no quadro sobre derivadas de matrizes):

$$\left(X^t X\right).\hat{\beta} = X^t.y \qquad (15)$$

Multiplicando os dois lados da equação (15) por $(X^tX)^{-1}$ leva a:

$$\left(X^t X\right)^{-1}.\left(X^t X\right).\hat{\beta} = \left(X^t X\right)^{-1}.X^t.y \Rightarrow \hat{\beta} = \left(X^t X\right)^{-1}.X^t.y \qquad (16)$$

A equação (16) fornece o estimador MQO em forma matricial. Note que a matriz (X^tX) tem dimensão k x k, assim como sua inversa, X^t em dimensão k x n e y tem dimensão n x 1. Assim, $\hat{\beta}$ calculado em (16) tem dimensão k x 1 como apropriado.

Nota explicativa – derivadas de matrizes (Gujarati, p.745)

Regra 1

Sejam a um vetor coluna n x 1 de números e x um vetor coluna n x 1 de variáveis. Então $a^t.x$ é um escalar e:

$$\frac{\partial(a^t.x)}{\partial x} = a = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}$$

Regra 2

Seja x um vetor n x 1 de variáveis e A uma matriz n x n de números. Então $x^t.A.x$ é um escalar e:

$$\frac{\partial(x^t Ax)}{\partial x} = 2.A.x = 2\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = 2\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{pmatrix}$$

Que é um vetor coluna n x 1 ou:

$$\frac{\partial(x^t Ax)}{\partial x} = 2.x^t A = 2(x_1, x_2, \dots, x_n)\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} =$$

$$= 2(a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n, \quad a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \quad \dots \quad a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)$$

Que é um vetor linha 1 x n.

Exemplo 4.1. (O estimador MQO da regressão simples em forma matricial)

O modelo de regressão linear simples em forma matricial é dado por (2). Note que:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots\dots & \\ 1 & x_n \end{pmatrix} \quad \text{e então } X^t = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

Logo:

$$X^tX = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots\dots & \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

$$X^ty = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

A inversa de X^tX é:

$$\left(X^tX\right)^{-1} = \frac{1}{SXX} \begin{pmatrix} \dfrac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ onde } SXX = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (17)$$

Para casa: verifique que a matriz acima é realmente a inversa de X^tX.

Logo, os estimadores MQO são dados em forma matricial por:

$$\hat{\beta} = \left(X^tX\right)^{-1}.X^t.y = \frac{1}{SXX} \begin{pmatrix} \sum x_i^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} =$$

$$= \frac{1}{SXX} \begin{pmatrix} \sum x_i^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = \frac{1}{SXX} \begin{pmatrix} \left(\dfrac{\sum x_i^2}{n}\right).n\bar{y} - \bar{x}\sum x_i y_i \\ -n\bar{x}\bar{y} + \sum x_i y_i \end{pmatrix} =$$

$$= \frac{1}{SXX} \begin{pmatrix} \bar{y}\sum x_i^2 - \bar{x}\sum x_i y_i \\ \sum x_i y_i - n\bar{x}\bar{y} \end{pmatrix} = \frac{1}{SXX} \begin{pmatrix} \bar{y}\left(\sum x_i^2 - n\bar{x}^2 + n\bar{x}^2\right) - \bar{x}\sum x_i y_i \\ SXY \end{pmatrix} =$$

$$= \frac{1}{SXX} \begin{pmatrix} \bar{y}\left(\sum x_i^2 - n\bar{x}^2\right) - \bar{x}\sum x_i y_i + n\bar{y}\bar{x}^2 \\ SXY \end{pmatrix} = \frac{1}{SXX} \begin{pmatrix} \bar{y}(SXX) - \bar{x}\left(\sum x_i y_i - n\bar{y}\bar{x}\right) \\ SXY \end{pmatrix} =$$

$$= \frac{1}{SXX} \begin{pmatrix} \bar{y}(SXX) - \bar{x}(SXY) \\ SXY \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{x}\dfrac{SXY}{SXX} \\ \dfrac{SXY}{SXX} \end{pmatrix}$$

Que são exatamente os estimadores MQO dados pela equação (10).

O seguinte resultado é muito útil em inúmeras demonstrações que envolvem estimadores MQO, e assim merece ser enunciado separadamente.

Resultado 4.2.

Sejam $\hat{\beta}$ e $\hat{\varepsilon}$ os vetores de parâmetros e resíduos, respectivamente, de um modelo de regressão múltipla. Então:

$$X^t\hat{\varepsilon} = X^t(y-\hat{y}) = X^t\left(y - X.\hat{\beta}\right) = X^t\left(y - X.\left(X^tX\right)^{-1}X^ty\right) = X^ty - X^tX\left(X^tX\right)^{-1}X^ty =$$
$$= X^ty - IX^ty = X^ty - X^ty = 0$$

A soma do quadrado dos resíduos, RSS, pode ser escrita como:

$$RSS = \hat{\varepsilon}^t\hat{\varepsilon} = \left(y - X\hat{\beta}\right)^t\left(y - X\hat{\beta}\right) = y^ty - \left(X\hat{\beta}\right)^t y - y^t.\left(X\hat{\beta}\right) + \left(X\hat{\beta}\right)^t\left(X\hat{\beta}\right) =$$
$$= y^ty - 2\hat{\beta}^tX^ty + \hat{\beta}^tX^tX\hat{\beta} = y^ty - 2\hat{\beta}^tX^ty + \hat{\beta}^tX^tX\left(X^tX\right)^{-1}X^ty =$$
$$= y^ty - 2\hat{\beta}^tX^ty + \hat{\beta}^tX^ty = y^ty - \hat{\beta}^tX^ty$$

Teorema 4.3.

Os estimadores MQO $\hat{\beta}$ são não tendenciosos.

Demonstração

Pela equação (16): $\hat{\beta} = \left(X^tX\right)^{-1}.X^t.y$ e devemos lembrar que X é uma matriz fixa (não aleatória). Então:

$$E\left(\hat{\beta}\right) = \left(X^tX\right)^{-1}.X^t.E(y) = \left(X^tX\right)^{-1}.X^t.E(X\beta + \varepsilon) = \left(X^tX\right)^{-1}.X^tX.\beta + \left(X^tX\right)^{-1}.X^tE(\varepsilon) =$$
$$= \beta + \left(X^tX\right)^{-1}.X^t(0) = \beta$$

Pois os erros têm média nula por hipótese.

Teorema 4.4.

A matriz de variância-covariância doss estimadores MQO $\hat{\beta}$ é dada por:

$$V = \sigma^2\left(X^tX\right)^{-1} \quad \text{onde } \sigma^2 = \mathrm{VAR}(\varepsilon_i) \text{ para } i = 1,2,\ldots,n$$

Demonstração

$$VAR(\hat{\beta}) = E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^t\right]$$

Mas, note que:

$$\hat{\beta}-\beta=\left(X^{t}X\right)^{-1}X^{t}Y-\beta=\left(X^{t}X\right)^{-1}X^{t}\left(X\beta+\varepsilon\right)-\beta=\beta+\left(X^{t}X\right)^{-1}X^{t}\varepsilon-\beta=$$
$$=\left(X^{t}X\right)^{-1}X^{t}\varepsilon$$

Seja $A=\left(X^{t}X\right)^{-1}X^{t}$

Então:

$$VAR(\hat{\beta})=E\left[(A\varepsilon)(A\varepsilon)^{t}\right]=E\left[A\varepsilon.\varepsilon^{t}A^{t}\right]=AE\left[\varepsilon.\varepsilon^{t}\right]A^{t}=\sigma^{2}AIA^{t}=\sigma^{2}AA^{t}$$

Mas:

$$AA^{t}=\left(X^{t}X\right)^{-1}X^{t}\left[\left(X^{t}X\right)^{-1}X^{t}\right]^{t}=\left(X^{t}X\right)^{-1}X^{t}.X\left(X^{t}X\right)^{-1}=\left(X^{t}X\right)^{-1}$$

Finalmente:

$$VAR(\hat{\beta})=\sigma^{2}AA^{t}=\sigma^{2}\left(X^{t}X\right)^{-1}$$