Modelos de Regressão Múltipla - Parte I

Erica Castilho Rodrigues

4 de Outubro de 2016

Regressão Linear Simples na Forma Matricial

ANOVA em forma matricial

Introdução

- Quando há apenas uma variável explicativa X, temos um problema de regressão linear simples

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

- Quando há pelo menos duas variáveis explicativas

$$X_1, X_2, ..., X_p$$

temos um problema de regressão linear múltipla:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i + \cdots + \beta_p X_i + \epsilon_i$$

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

- Estudamos a influência ou explicação conjunta das variáveis X' na variável Y.
- A interpretação de β_0 muda?

- Estudamos a influência ou explicação conjunta das variáveis X' na variável Y.
- A interpretação de β_0 muda? Não.
- Qual a interpretação?

- Estudamos a influência ou explicação conjunta das variáveis X' na variável Y.
- A interpretação de β_0 muda? Não.
- Qual a interpretação?
- β_0 é o valor esperado de Y quando todas variáveis são iguais a zero.
- Se alguma das variáveis não puder assumir valor zero, β_0 perde essa interpretação.
- Mesmo assim ele deve continuar na equação.

- A interpretação dos coeficientes

$$\beta_1, \beta_2, \ldots, \beta_p$$

  é feita separadamente para cada variável, mantendo as demais fixas.
- Por exemplo, se $p = 2$.
- β_1 representa a variação esperada em Y quando X_1 aumenta em uma unidade e X_2 é mantida fixa.
- Qual interpretação do β_2?

- A interpretação dos coeficientes

$$\beta_1, \beta_2, \ldots, \beta_p$$

  é feita separadamente para cada variável, mantendo as demais fixas.
- Por exemplo, se $p = 2$.
- β_1 representa a variação esperada em Y quando X_1 aumenta em uma unidade e X_2 é mantida fixa.
- Qual interpretação do β_2?
- β_2 representa a variação esperada em Y quando X_2 aumenta em uma unidade e X_1 é mantida fixa.

Exemplo

- Fonte:CRISP- Informativo no 1, dez/2001, pg. 12.
- Deseja-se analisar incidência de roubos nas unidades de planejamento de Belo Horizonte.
- Quais variáveis podem explicar?

Exemplo

- Fonte:CRISP- Informativo no 1, dez/2001, pg. 12.
- Deseja-se analisar incidência de roubos nas unidades de planejamento de Belo Horizonte.
- Quais variáveis podem explicar?
- Renda, porcentagem de área comercial, índice de proteção social.
- Variável Resposta:

  Y=logaritmo do número de roubos nasunidades de planejamento.
- Precisamos da transformação pois a variável Y é uma contagem, logo não segue distribuição normal.

Exemplo (continuação)

- As variáveis explicativas são as seguintes:

$$X_1 = \{\text{ renda média do chefe da família (dólares)}\}$$

$$X_2 = \{\text{ porcentagem de área comercial }\}$$

$$X_3 = \{\text{ Índice de Proteção Social }\}$$

- O modelo ajustado é o seguinte

$$\ln(roubo) = 3.9 + 0.00086(renda) + 0.244(\%\text{ de comércio}) - 0.529(IPS)$$

Exemplo (continuação)

- Vamos interpretar os parâmetros do modelo.
- Nesse caso β_0 não tem sentido prático.
- Como a variável resposta está na escala logarítimica, a interpretação modificia um pouco.
- Vejamos qual interpretação de β_1.
- Quais variáveis mantemos fixas?

Exemplo (continuação)

- Vamos interpretar os parâmetros do modelo.
- Nesse caso β_0 não tem sentido prático.
- Como a variável resposta está na escala logarítimica, a interpretação modificia um pouco.
- Vejamos qual interpretação de β_1.
- Quais variáveis mantemos fixas?
- % de área comercial e IPS.

Exemplo (continuação)

- Suponha que a renda média do chefe de família aumenta em 1 dólar.
- Qual aumento esperado no $ln(roubo)$?

Exemplo (continuação)

- Suponha que a renda média do chefe de família aumenta em 1 dólar.
- Qual aumento esperado no *ln*(*roubo*)? 0,00086.
- E se a renda aumenta em 100 dólares?

Exemplo (continuação)

- Suponha que a renda média do chefe de família aumenta em 1 dólar.
- Qual aumento esperado no *ln*(*roubo*)? 0,00086.
- E se a renda aumenta em 100 dólares? 0,086

$$\ln(roubo) = 3.9 + 0.00086(renda) + 0.244(\%\,\text{de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) = 3.9 + 0.00086(renda + 100) + 0.244(\%\,\text{de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) - \ln(roubo) =$$

Exemplo (continuação)

- Suponha que a renda média do chefe de família aumenta em 1 dólar.
- Qual aumento esperado no *ln*(*roubo*)? 0,00086.
- E se a renda aumenta em 100 dólares? 0,086

$$\ln(roubo) = 3.9 + 0.00086(renda) + 0.244(\%\text{ de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) = 3.9 + 0.00086(renda + 100) + 0.244(\%\text{ de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) - \ln(roubo) = 0.00086(renda + 100) - 0.00086(renda)$$

$$=$$

Exemplo (continuação)

- Suponha que a renda média do chefe de família aumenta em 1 dólar.
- Qual aumento esperado no $ln(roubo)$? 0,00086.
- E se a renda aumenta em 100 dólares? 0,086

$$\ln(roubo) = 3.9 + 0.00086(renda) + 0.244(\%\,\text{de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) = 3.9 + 0.00086(renda + 100) + 0.244(\%\,\text{de comércio}) - 0.529(IPS)$$

$$\ln(roubo^*) - \ln(roubo) = 0.00086(renda + 100) - 0.00086(renda)$$

$$= 0.00086(100) = 0.086$$

ou seja

$$\ln(roubo^*) = \ln(roubo) + 0.086$$

Exemplo (continuação)

- Vamos tirar o logaritmo dos dois lados da equação

$$\ln(roubo^*) = \ln(roubo) + 0.086$$

ficamos com

Exemplo (continuação)

- Vamos tirar o logaritmo dos dois lados da equação

$$\ln(roubo^{*}) = \ln(roubo) + 0.086$$

ficamos com

$$roubo^{*} = \exp\{\ln(roubo) + 0.086\} =$$

Exemplo (continuação)

- Vamos tirar o logaritmo dos dois lados da equação

$$\ln(roubo^*) = \ln(roubo) + 0.086$$

ficamos com

$$roubo^* = \exp\{\ln(roubo) + 0.086\} = roubo(\exp\{0.086\})$$

como $\exp\{0.086\} = 1.0898$ então

$$roubo^* = 1.0898(roubo)\,.$$

- Qual a variação percentual no número de roubos quando a renda aumeenta em 100 dólares **e as demais variáveis são mantidas fixas**?

Exemplo (continuação)

- Vamos tirar o logaritmo dos dois lados da equação

$$\ln(roubo^*) = \ln(roubo) + 0.086$$

ficamos com

$$roubo^* = \exp\{\ln(roubo) + 0.086\} = roubo(\exp\{0.086\})$$

como $\exp\{0.086\} = 1.0898$ então

$$roubo^* = 1.0898(roubo)\,.$$

- Qual a variação percentual no número de roubos quando a renda aumeenta em 100 dólares **e as demais variáveis são mantidas fixas**?
- O roubo aumenta em 8.98%.

Exemplo (continuação)

- Qual interperetação de β_2?
- Matemos fixas quais variáveis?

Exemplo (continuação)

- Qual interperetação de β_2?
- Matemos fixas quais variáveis?
- Renda e IPS.
- Suponha que a porcentagem de área comercial aumenta em uma unidade (um ponto percentual).
- Qual aumento esperado no $\ln(roubo)$?

Exemplo (continuação)

- Qual interperetação de β_2?
- Matemos fixas quais variáveis?
- Renda e IPS.
- Suponha que a porcentagem de área comercial aumenta em uma unidade (um ponto percentual).
- Qual aumento esperado no $\ln(roubo)$?0.244
- O que isso significa?

Exemplo (continuação)

- Qual interperetação de β_2?
- Matemos fixas quais variáveis?
- Renda e IPS.
- Suponha que a porcentagem de área comercial aumenta em uma unidade (um ponto percentual).
- Qual aumento esperado no $\ln(roubo)$?0.244
- O que isso significa?

$$roubo^* = (\exp\{0.244\})roubo$$

como $(\exp\{0.244\}) = 1.276$ temos que

$$roubo^* = 1.276 roubo$$

Exemplo (continuação)

- O que isso significa?

Exemplo (continuação)

- O que isso significa?
- Espera-se um aumento de 27.6% no número de roubos quando a ocupação compercial aumenta em um ponto percentual e as demais variáveis são mantidas fixas
- Vejamos a intrepretação de β_3.
- Qual variação esperada no $\ln(roubo)$?

Exemplo (continuação)

- O que isso significa?
- Espera-se um aumento de 27.6% no número de roubos quando a ocupação compercial aumenta em um ponto percentual e as demais variáveis são mantidas fixas
- Vejamos a intrepretação de β_3.
- Qual variação esperada no $\ln(roubo)$?-0.529
- O que isso significa?

Exemplo (continuação)

- O que isso significa?
- Espera-se um aumento de 27.6% no número de roubos quando a ocupação compercial aumenta em um ponto percentual e as demais variáveis são mantidas fixas
- Vejamos a intrepretação de β_3.
- Qual variação esperada no ln(*roubo*)?-0.529
- O que isso significa?

$$roubo^* = (\exp\{-0.529\})roubo$$

como $(\exp\{-0.529\}) = 0.948$ temos que

$$roubo^* = 0.948 roubo$$

Exemplo (continuação)

- O que isso significa?

Exemplo (continuação)

- O que isso significa?
- Espera-se uma redução de 5.2% no número de roubos quando o IPS aumenta em uma unidade e as demais variáveis são mantidas fixas
- Observe que uma redução de 5.2% equivale a multiplicar po 0.948.

Regressão Linear Simples na Forma Matricial

- Podemos escrever o modelo de regressão linear simples de forma matricial.
- A solução obtida pode ser usada para qualquer modelo de regressão linear.
- Pode ser usada para um modelo de Regressão Linear Múltipla

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i + \cdots + \beta_p X_i + \epsilon_i$$

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

- O modelo de Regressão Linear Simples é dado por

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{para} \quad i = 1, 2, \ldots, n$$

ou seja

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n \end{aligned}$$

- Como podemos reescrever esse sistema usando matrizes? (fazer no quadro)

- Como podemos reescrever esse sistema usando matrizes? (fazer no quadro)

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \beta_1 X_1 \\ \beta_1 X_2 \\ \vdots \\ \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

onde

Y: vetor $n \times 1$ de observações da variável resposta

X: matriz $n \times 2$ de observações da variável explicativa

$\boldsymbol{\beta}$: vetor 2×1 de parâmetros

$\boldsymbol{\epsilon}$: vetor $n \times 1$ dos erros

- Vejamos como ficam os estimadores de mínimos quadrados.
- Queremos minimizar a soma dos quadrados dos resíduos

$$SQE = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i e_i^2$$

- Podemos reescrever da seguinte maneira

$$SQE = \sum_i e_i^2 = \begin{bmatrix} \epsilon_1 & e_2 & \dots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{e}'\mathbf{e}$$

onde

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix} = \mathbf{Y} - \mathbf{X}\beta$$

- Temos então que

$$SQE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

(fazer no quadro)

- Temos então que

$$SQE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

(fazer no quadro)

$$= (\mathbf{Y}' - (\mathbf{X}\beta)')(\mathbf{Y} - \mathbf{X}\beta) = (\mathbf{Y}' - \beta'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\beta)$$

$$= \mathbf{Y}'\mathbf{Y} - \beta'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta$$

observe que como β tem dimensão 2×1 e $\mathbf{X}$ tem dimensão $n \times 2$ então

$$\beta'\mathbf{X}'\mathbf{Y} = \beta'_{1\times 2}\mathbf{X}'_{2\times n}\mathbf{Y}_{n\times 1} = \left(\beta'\mathbf{X}'\right)_{1\times n}\mathbf{Y}_{n\times 1} = \left(\beta'\mathbf{X}'\mathbf{Y}\right)_{1\times 1}$$

- Portanto $\beta'\mathbf{X}'\mathbf{Y}$ é um escalar e

$$\beta'\mathbf{X}'\mathbf{Y} = \left(\beta'\mathbf{X}'\mathbf{Y}\right)' = \mathbf{Y}'\mathbf{X}\beta$$

- Então a SQE fica

$$
\begin{aligned}
SQE &= \mathbf{Y}'\mathbf{Y} - \beta'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \\
&= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta
\end{aligned}
$$

- Queremos encontrar o vetor β que minimiza a SQE.
- Vamos derivar SQE em relação a β.
- Vamos usar os seguintes resultados

$$\frac{\partial \mathbf{b}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{b}'\mathbf{A})'$$

se **A** é simétrica

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

- Temos que

$$\frac{\partial \mathbf{Y}'\mathbf{X}\beta}{\partial \beta} = (\mathbf{Y}'\mathbf{X})' = \mathbf{X}'\mathbf{Y}$$

$$\frac{\partial \beta'\mathbf{X}'\mathbf{X}\beta}{\partial \beta} = 2\mathbf{X}'\mathbf{X}\beta$$

- Portanto

$$\frac{\partial SQE}{\partial \beta} = \frac{\partial \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta}{\partial \beta} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta\,.$$

- Igualando a derivada a zero ficamos com

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta = 0$$

implica que

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y}$$

se $\mathbf{X}'\mathbf{X}$ for inversível, podemos multiplicar por $(\mathbf{X}'\mathbf{X})^{-1}$ dos dois lados

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\,.$$

- Concluímos asssim que o estimador de mínimos quadrados de β é dado por

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

- Esse resultado é válido para p variáveis explicativas.
- No caso da regressão linear simples o vetor β é 2×1.
- No caso genérico com p variáveis β é $(p+1) \times 1$.
- Aparece $p+1$ por causa do β_0.

- Vamos mostrar que para o caso da regressão linear simples os resultados coincidem com aqueles obtidos anteriormente.
- A Regressão Linear Simples é um caso particular da Regressão Linear Múltipla.
- Vamos relembrar alguns resultados sobre matrizes.
- Seja **M** uma matriz 2×2

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

- A inversa de **M** é dada por

- A inversa de **M** é dada por

$$\mathbf{M}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

onde

$$D = \det(\mathbf{M}) = ad - bc\,.$$

- A matriz $\mathbf{X}'\mathbf{X}$ é dada por

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum_i X_i \\ \sum_i X_i & \sum_i X_i^2 \end{bmatrix}$$

temos que

$$\det(\mathbf{X}\mathbf{X}') = n \sum_i X_i^2 - (\sum_i X_i)^2$$

observer que

$$n \sum_i (X_i - \overline{X})^2 = n \left(\sum_i X_i^2 - 2\overline{X} \sum_i X_i + \overline{X}^2 \right)$$

$$= n \sum_i X_i^2 - 2n(\sum_i X_i)/n(\sum_i X_i) + n^2 \frac{(\sum_i X_i)^2}{n^2}$$

$$= n \sum_i X_i^2 - (\sum_i X_i)^2$$

- E portanto

$$\det(\mathbf{X}'\mathbf{X}) = n\sum_i (X_i - \overline{X})^2$$

- Como

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_i X_i \\ \sum_i X_i & \sum_i X_i^2 \end{bmatrix}$$

temos que

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \sum_i X_i^2 & -\sum_i X_i \\ -\sum_i X_i & n \end{bmatrix}$$

- Temos ainda que

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

- Portanto

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

$$= \frac{1}{n\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \sum_i X_i^2 & -\sum_i X_i \\ -\sum_i X_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

$$= \frac{1}{n\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i \\ -\sum_i X_i \sum_i Y_i + n \sum_i X_i Y_i \end{bmatrix}$$

- Vamos verificar o primeiro termo do vetor

$$\frac{\sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i}{n\sum_i (X_i - \overline{X})^2}$$

somando e subtraindo $n\overline{X}^2 \sum_i Y_i$

$$\frac{\sum_i X_i^2 \sum_i Y_i - n\overline{X}^2 \sum_i Y_i + n\overline{X}^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i}{n\sum_i (X_i - \overline{X})^2}$$

$$= \frac{\sum_i Y_i (\sum_i X_i^2 - n\overline{X}^2) + n\frac{(\sum_i X_i)^2}{n^2} \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i}{n\sum_i (X_i - \overline{X})^2}$$

- Como
$$\sum_i X_i^2 - n\overline{X}^2 = \sum_i (X_i - \overline{X})^2$$
ficamos com
$$= \frac{\sum_i Y_i}{n} \frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2} + \frac{\sum_i X_i \left(\frac{\sum_i X_i \sum_i Y_i}{n} - \sum_i X_i Y_i \right)}{n \sum_i (X_i - \overline{X})^2}$$
$$= \overline{Y} - \overline{X} \frac{\left(\frac{-\sum_i X_i \sum_i Y_i}{n} + \sum_i X_i Y_i \right)}{\sum_i (X_i - \overline{X})^2}$$
já vimos que
$$\sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{n} = \sum_i (X_i - \overline{X})(Y_i - \overline{Y})$$

- Portanto ficamos com

$$\overline{Y} - \overline{X}\frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}$$

mas

$$\hat{\beta}_1 = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}$$

- Mostramos assim que o primeiro elemento do vetor $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ é dado por

$$\overline{Y} - \overline{X}\hat{\beta}_1$$

que é justamente a expressão que tínhamos para o $\hat{\beta}_0$ no caso da Regressão Linear Simples.

- Vejamosa agora como fica o segundo elemento de

$$\hat{\beta} = \frac{1}{n\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i \\ -\sum_i X_i \sum_i Y_i + n\sum_i X_i Y_i \end{bmatrix}.$$

- Temos que

$$\frac{-\sum_i X_i \sum_i Y_i + n\sum_i X_i Y_i}{n\sum_i (X_i - \overline{X})^2}$$

$$= \frac{n\left(\sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{n}\right)}{n\sum_i (X_i - \overline{X})^2}$$

já vimos que

$$\sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{n} = \sum_i (X_i - \overline{X})(Y_i - \overline{Y})$$

- Ficamos então com

$$\frac{n\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{n\sum_i (X_i - \overline{X})^2} = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}$$

que é a expressão que encontramos para $\hat{\beta}_1$ anteriormente.

- Motrasmo assim que

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

- Portanto as expressões obtidas para o caso genérico condizem com aquelas que foram obtidas para o modelo de Regressão Linear Simples.

ANOVA em forma matricial

- As somas de quadrados também podem ser escritas em forma matricial.
- A Soma de Quadrados Totais pode ser escrita como

$$SQT = \sum_i (Y_i - \overline{Y})^2 = \sum_i Y_i^2 - n\overline{Y}^2$$

mas

$$\sum_i Y_i^2 = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \mathbf{Y}'\mathbf{Y}$$

portanto a SQT pode ser escrita como

$$SQT = \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$$

- A Soma de Quadrados da Regressão pode ser escrita como

$$SQR = \sum_i (\hat{Y}_i - \overline{Y})^2 = \sum_i \hat{Y}_i^2 - 2\overline{Y} \sum_i \hat{Y}_i + n\overline{Y}^2$$

mas sabemos que

$$\sum_i e_i = \sum_i (\hat{Y}_i - Y_i) = 0 \Rightarrow \sum_i \hat{Y}_i = \sum_i Y_i$$

logo

$$SQR = \sum_i \hat{Y}_i^2 - 2\overline{Y} \sum_i \hat{Y}_i + n\overline{Y}^2 = \sum_i \hat{Y}_i - 2\overline{Y} \sum_i Y_i + n\overline{Y}^2$$

$$= \sum_i \hat{Y}_i^2 - 2n\overline{Y} + n\overline{Y}^2 = \sum_i \hat{Y}_i^2 - n\overline{Y}^2 .$$

- Temos ainda que
$$\sum_i \hat{Y}_i^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$$
mas
$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} \Rightarrow \hat{\mathbf{Y}}' = \hat{\beta}'\mathbf{X}'$$
e podemos escrever ainda
$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$$
substituindo $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}$ ficamos com
$$\hat{\mathbf{Y}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

- Substituindo

$$\hat{\mathbf{Y}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \quad \text{e} \quad \hat{\mathbf{Y}}' = \hat{\beta}'\mathbf{X}'$$

em

$$\sum_i \hat{Y}_i^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$$

ficamos com

$$\sum_i \hat{Y}_i^2 = \hat{\beta}'\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \hat{\beta}'\mathbf{X}'\mathbf{Y}\,.$$

- Portanto

$$SQR = \sum_i \hat{Y}_i^2 - n\overline{Y}^2 = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2\,.$$

- A Soma de Quadrados do Resíduo é dada por

$$SQE = SQT - SQR\,.$$

- Mas vimos que

$$SQT = \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2 \quad SQR = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2\,.$$

- Logo

$$\begin{aligned} SQE &= \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2 - \left(\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2\right) \\ &= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}\,. \end{aligned}$$

- A tabela ANOVA pode ser escrita matricialmente da seguinte forma

Fonte de Variação	Graus de Liberdade	Soma de Quadrados
Regressão	1	$\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$
Resíduo	$n-2$	$\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$
Total	$n-1$	$\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$

Tabela: Tabela ANOVA

- Vejamos agora como fica a variância e covariância de $\hat{\boldsymbol{\beta}}$ em notação matricial.
- Vimos que

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_i X_i^2}{n \sum_i (X_i - \overline{X})^2}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

- Temos ainda que

$$
\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\overline{Y} - \hat{\beta}_1 \overline{X}, \hat{\beta}_1) \\
&= Cov(\overline{Y}, \hat{\beta}_1) + Cov(-\hat{\beta}_1 \overline{X}, \hat{\beta}_1) \\
&= Cov(\overline{Y}, \hat{\beta}_1) - \overline{X} Cov(\hat{\beta}_1, \hat{\beta}_1)
\end{aligned}
$$

já vimos que

$$Cov(\overline{Y}, \hat{\beta}_1) = 0$$

além disso

$$Cov(\hat{\beta}_1, \hat{\beta}_1) = Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

portanto

$$
\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\overline{Y}, \hat{\beta}_1) - \overline{X} Cov(\hat{\beta}_1, \hat{\beta}_1) \\
&= 0 - \overline{X} \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} = -\overline{X} \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} .
\end{aligned}
$$

- A matriz de variâncias e covariâncias do vetor $(\hat{\beta}_0, \hat{\beta}_1)'$ é dada por

$$Var\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) \end{bmatrix}$$

$$\begin{bmatrix} \frac{\sigma^2 \sum_i X_i^2}{n\sum_i (X_i - \overline{X})^2} & -\overline{X}\frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} \\ -\overline{X}\frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} & \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} \end{bmatrix}$$

$$= \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}\begin{bmatrix} \frac{\sum_i X_i^2}{n} & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}$$

- Vamos agora chegar nesse mesmo resultado, porém usando a notação matricial do vetor $(\hat{\beta}_0, \hat{\beta}_1)'$.
- Vimos que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

- Vamos usar o seguinte resultado

$$Var(\mathbf{AY}) = \mathbf{A}\,Var(\mathbf{Y})\mathbf{A}'$$

onde **A** é uma matriz de constantes e **Y** é um vetor aleatório.

- Observe que esse resultdo é o caso geral de

$$Var(aY) = a^2 Var(Y)\,.$$

- Vimos que

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

e $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ é uma matriz de constantes.

- Logo

$$Var(\hat{\beta}) = Var(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y})$$

$$= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] Var(\mathbf{Y}) \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' .$$

- Mas sabemos que os Y_i's são i.i.d. com variância constante σ^2
- Portanto a matriz de variâncias e covariâncias do vetor **Y** é dada por

$$Var(\mathbf{Y}) = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

- E portanto

$$
\begin{aligned}
Var(\hat{\beta}) &= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] Var(\mathbf{Y}) \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' \\
&= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] \sigma^2\mathbf{I}_n \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' \\
&= \sigma^2 \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' \\
&= \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} .
\end{aligned}
$$

- Já vimos que

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{n\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \sum_i X_i^2 & -\sum_i X_i \\ -\sum_i X_i & n \end{bmatrix} \\
&= \frac{1}{\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \frac{\sum_i X_i^2}{n} & -\frac{\sum_i X_i}{n} \\ -\frac{\sum_i X_i}{n} & \frac{n}{n} \end{bmatrix} \\
&= \frac{1}{\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \frac{\sum_i X_i^2}{n} & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}
\end{aligned}
$$

- Portanto
$$Var(\hat{\beta}) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$
$$= \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \frac{\sum_i X_i^2}{n} & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}$$
que foi exatamente o mesmo resultado que obtivemos para o caso da Regressão Linear Simples.
- Observe que o resultado $Var(\hat{\beta}) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$ vale para um número genérico que variáveis.
- Estamos mostrando aqui que para o caso particular $p = 1$ o resultado coincide com o que obtivemos anteriormente.