Estatística e Análise de Dados em Zootecnia 2025/2026

Modelo Linear

Elsa Gonçalves
Secção de Matemática, DCEB, ISA-Ulisboa

(Adaptado, Cadima J. (2021). O Modelo Linear, ISA, UlLisboa)

Regressão Linear Múltipla - Abordagem Descritiva

Quando é necessária mais do que uma variável preditora para modelar adequadamente a variável resposta de interesse.

Plano em $\mathbb{R}^3$

Qualquer plano em $\mathbb{R}^3$, no sistema $x0y0z$, tem equação

$$Ax + By + Cz + D = 0 \ .$$

No nosso contexto, e colocando:

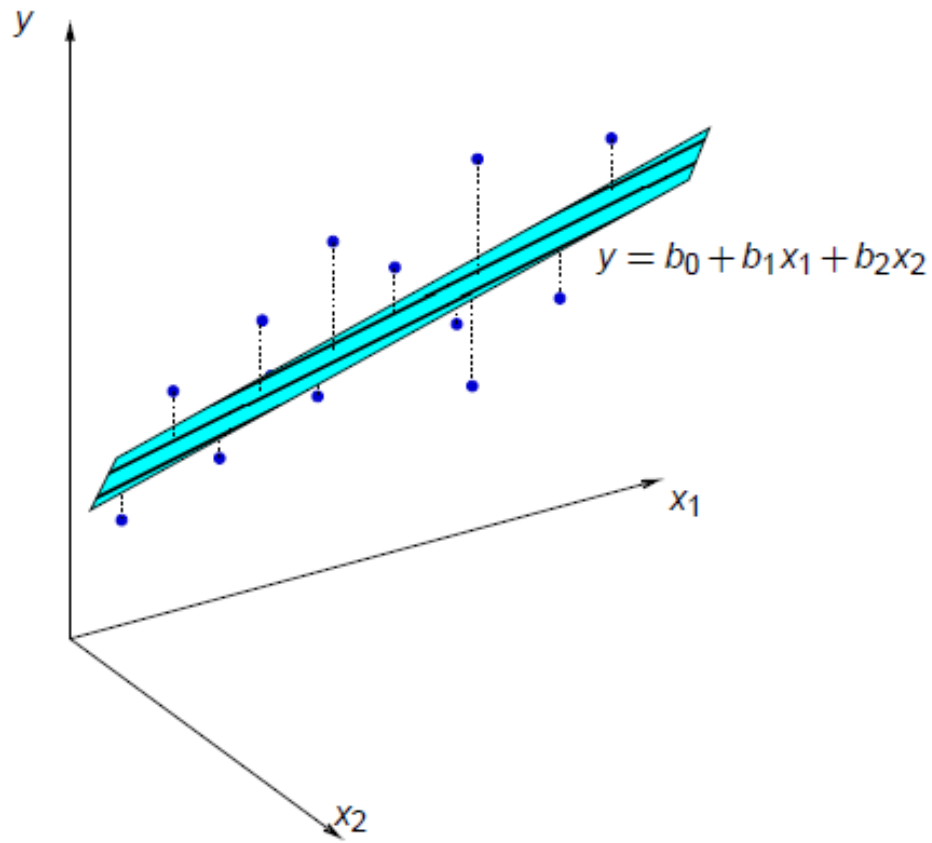
- no eixo vertical (z) a variável resposta Y;
- noutro eixo (x) um preditor X_1;
- no terceiro eixo (y) o outro preditor X_2,

A equação fica (no caso geral de planos não verticais, com $C \neq 0$):

$$\begin{aligned} Ax_1 + Bx_2 + Cy + D = 0 \quad &\Leftrightarrow \quad y = -\frac{D}{C} - \frac{A}{C}x_1 - \frac{B}{C}x_2 \\ &\Leftrightarrow \quad y = b_0 + b_1 x_1 + b_2 x_2 \end{aligned}$$

Esta equação generaliza a equação da recta, para o caso de haver dois preditores.

Regressão Múltipla - representação gráfica ($p = 2$)



$y = b_0 + b_1 x_1 + b_2 x_2$ é a equação dum plano em $\mathbb{R}^3$ $(x_1 0 x_2 0 y)$.
Pode ser ajustado pelo mesmo critério que na RLS: minimizar SQRE.

O caso geral: p preditores

Para modelar uma variável resposta Y com base numa regressão linear sobre p variáveis preditoras, $x_1, x_2, ..., x_p$, admite-se que os valores de Y oscilam em torno duma combinação linear (afim) das p variáveis preditoras:

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p \ .$$

Trata-se da equação dum hiperplano em $\mathbb{R}^{p+1}$, que define a relação de fundo entre y e os p preditores.

Tal como na Regressão Linear Simples, admite-se que dispomos de n conjuntos de observações para ajustar este hiperplano:

$$\left\{ \left(x_{1(i)}, x_{2(i)}, ... x_{p(i)}, y_i \right) \right\}_{i=1}^{n} \ .$$

Não é possível visualizar a nuvem de pontos das observações se $p > 2$.

Regressão Múltipla: o hiperplano ajustado

Admite-se que os valores de y oscilam em torno duma combinação linear (afim) das p variáveis preditoras:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p \ .$$

Trata-se da equação dum hiperplano em $\mathbb{R}^{p+1}$.

O critério utilizado para ajustar um hiperplano à nuvem de n pontos em $\mathbb{R}^{p+1}$ é o de minimizar a Soma de Quadrados dos Resíduos, ou seja, escolher os $p+1$ parâmetros $\{b_j\}_{j=0}^{p}$ que minimizem:

$$SQRE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

onde os y_i representam os valores observados da variável resposta e

$$\hat{y}_i = b_0 + b_1 x_{1(i)} + b_2 x_{2(i)} + \ldots + b_p x_{p(i)}$$

os valores ajustados pela equação do hiperplano.

Duas abordagens para a estimação dos parâmetros

Para obter os parâmetros que definem o hiperplano que melhor se ajusta às observações pode-se usar uma abordagem:

- analítica; ou
- geométrica.

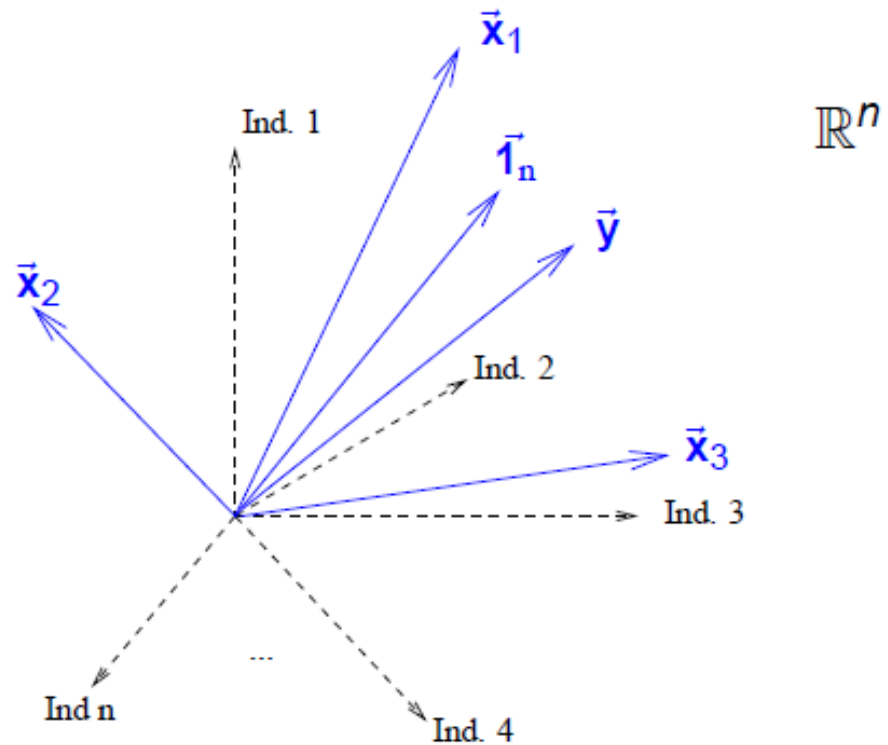
Nas duas abordagens, a notação vectorial-matricial é vantajosa.

Não existem fórmulas simples, como no caso da RLS, para cada um dos parâmetros b_j isoladamente. Mas é possível indicar uma fórmula única matricial para o conjunto dos $p+1$ parâmetros do modelo.

A representação em $\mathbb{R}^n$, o espaço das variáveis

- cada eixo corresponde a um indivíduo observado;
- cada vector corresponde a uma variável.

O vector de *n* uns, representado por $\vec{1}_n$, também é útil.



Os n valores ajustados $\hat{y}_i$ também definem um vector de $\mathbb{R}^n$, $\hat{\vec{\mathbf{y}}}$:

$$\vec{\hat{\mathbf{y}}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \dots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 x_{1_{(1)}} + b_2 x_{2_{(1)}} + \ldots + b_p x_{p_{(1)}} \\ b_0 + b_1 x_{1_{(2)}} + b_2 x_{2_{(2)}} + \ldots + b_p x_{p_{(2)}} \\ b_0 + b_1 x_{1_{(3)}} + b_2 x_{2_{(3)}} + \ldots + b_p x_{p_{(3)}} \\ \dots \\ b_0 + b_1 x_{1_{(n)}} + b_2 x_{2_{(n)}} + \ldots + b_p x_{p_{(n)}} \end{bmatrix}$$

$$= b_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_{1_{(1)}} \\ x_{1_{(2)}} \\ x_{1_{(3)}} \\ \vdots \\ x_{1_{(n)}} \end{bmatrix} + \ldots + b_p \begin{bmatrix} x_{p_{(1)}} \\ x_{p_{(2)}} \\ x_{p_{(3)}} \\ \vdots \\ x_{p_{(n)}} \end{bmatrix}$$

$$= b_0 \vec{\mathbf{1}}_n + b_1 \vec{\mathbf{x}}_1 + b_2 \vec{\mathbf{x}}_2 + \ldots + b_p \vec{\mathbf{x}}_p$$

O vector $\hat{\vec{\mathbf{y}}}$ é uma combinação linear dos vectores $\vec{\mathbf{1}}_n$, $\vec{\mathbf{x}}_1$, $\vec{\mathbf{x}}_2$, ..., $\vec{\mathbf{x}}_p$

A matriz do modelo $\mathbf{X}$

O vector $\hat{\vec{y}}$ dos valores ajustados pode também escrever-se como um produto envolvendo uma matriz $\mathbf{X}$ cujas colunas sejam os vectores $\vec{1}_n$, $\vec{x}_1$, ..., $\vec{x}_p$.

A matriz $\mathbf{X}$ do modelo

$$\mathbf{X} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{=\vec{1}_n} \underbrace{\begin{bmatrix} x_{1_{(1)}} \\ x_{1_{(2)}} \\ x_{1_{(3)}} \\ \vdots \\ x_{1_{(n)}} \end{bmatrix}}_{=\vec{x}_1} \underbrace{\begin{bmatrix} x_{2_{(1)}} \\ x_{2_{(2)}} \\ x_{2_{(3)}} \\ \vdots \\ x_{2_{(n)}} \end{bmatrix}}_{=\vec{x}_2} \cdots \underbrace{\begin{bmatrix} x_{p_{(1)}} \\ x_{p_{(2)}} \\ x_{p_{(3)}} \\ \vdots \\ x_{p_{(n)}} \end{bmatrix}}_{=\vec{x}_p}$$

A matriz do modelo $\mathbf{X}$ é de dimensão $n \times (p+1)$.

O vector $\hat{\vec{y}}$ pode ser escrito desta forma: $\hat{\vec{y}} = \mathbf{X}\vec{b}$

A matriz do modelo $\mathbf{X}$ e o seu subespaço de colunas

- O conjunto de todas as combinações lineares dum conjunto de vectores chama-se o subespaço gerado (spanned) por esses vectores.
- O subespaço gerado pelas colunas da matriz do modelo $\mathbf{X}$ chama-se subespaço das colunas (column-space) da matriz $\mathbf{X}$, $\mathscr{C}(\mathbf{X})$.
- O vector $\hat{\vec{\mathbf{y}}}$ pertence ao subespaço $\mathscr{C}(\mathbf{X})$
  (os vectores $\vec{\mathbf{1}}_n$, $\vec{\mathbf{x}}_1$, ..., $\vec{\mathbf{x}}_p$ são colunas $\mathbf{X}$ e $\hat{\vec{\mathbf{y}}} = b_0\vec{\mathbf{1}}_n + b_1\vec{\mathbf{x}}_1 + b_2\vec{\mathbf{x}}_2 + \ldots + b_p\vec{\mathbf{x}}_p$).
- $\mathscr{C}(\mathbf{X})$ é um subespaço de $\mathbb{R}^n$ ($\mathscr{C}(\mathbf{X}) \subset \mathbb{R}^n$), mas de dimensão $p+1$
  (se as colunas de $\mathbf{X}$ forem linearmente independentes, isto é, se nenhum vector se puder escrever como combinação linear dos restantes).
- Qualquer combinação linear das colunas da matriz $\mathbf{X}$, ou seja, qualquer elemento de $\mathscr{C}(\mathbf{X})$ se pode escrever como $\mathbf{X}\vec{\mathbf{a}}$, onde
  $\vec{\mathbf{a}} = (a_0, a_1, a_2, \ldots, a_p)$ é o vector dos coeficientes da combinação linear.

Os parâmetros

- Cada escolha possível de coeficientes $\vec{\mathbf{a}} = (a_0, a_1, a_2, ..., a_p)$ corresponde a um ponto/vector no subespaço $\mathscr{C}(\mathbf{X})$.
- Essa escolha de coeficientes é única caso as colunas de $\mathbf{X}$ sejam linearmente independentes, isto é, se não houver dependência linear (multicolinearidade) entre as variáveis $\vec{\mathbf{x}}_1, ..., \vec{\mathbf{x}}_p, \vec{\mathbf{1}}_n$.
- Um dos pontos/vectores do subespaço é a combinação linear dada pelo vector de coeficientes $\vec{\mathbf{b}} = (b_0, b_1, ..., b_p)$ que minimiza:

$$SQRE = \sum_{i=}^{n} e_i^2 = \sum_{i=}^{n} (y_i - \hat{y}_i)^2$$

onde os y_i são os valores observados da variável resposta e $\hat{y}_i = b_0 + b_1 x_{1(i)} + b_2 x_{2(i)} + ... + b_p x_{p(i)}$ os valores ajustados. É a combinação linear que desejamos determinar.

Como identificar esse ponto/vector?

O critério minimiza *SQRE*

O vector $\vec{\hat{\mathbf{y}}}$ que minimiza a distância ao vector de observações $\vec{\mathbf{y}}$ minimiza também o quadrado dessa distância, que é dado por:

$$dist^2(\vec{\mathbf{y}}, \vec{\hat{\mathbf{y}}}) = \|\vec{\mathbf{y}} - \vec{\hat{\mathbf{y}}}\|^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = SQRE \ .$$

Ou seja, o critério minimiza a soma de quadrados dos resíduos.

$$\frac{\partial SQRE}{\partial \vec{\boldsymbol{b}}} = \mathbf{0}$$

Os parâmetros ajustados na RL Múltipla

$$\vec{\mathbf{b}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{y}} \ .$$

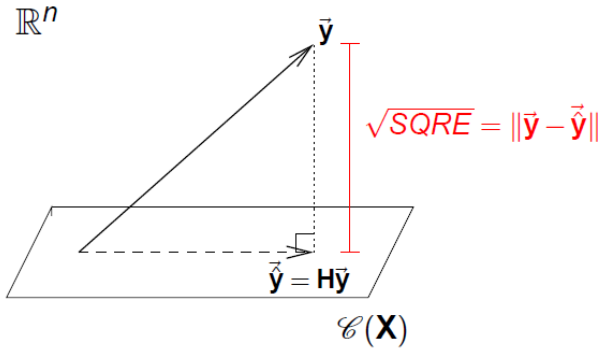
Ou, usando argumentos geométricos

- Dispomos de um vector de n observações de $\vec{\mathbf{y}}$ que está em $\mathbb{R}^n$ mas, em geral, não está no subespaço $\mathscr{C}(\mathbf{X})$.
- Queremos aproximar esse vector por outro vector, $\hat{\vec{\mathbf{y}}} = b_0\vec{\mathbf{1}}_n + b_1\vec{\mathbf{x}}_1 + \ldots + b_p\vec{\mathbf{x}}_p$, que está no subespaço $\mathscr{C}(\mathbf{X})$.
- Vamos aproximar o vector de observações $\vec{\mathbf{y}}$ pelo vector $\hat{\vec{\mathbf{y}}}$ do subespaço $\mathscr{C}(\mathbf{X})$ que está mais próximo de $\vec{\mathbf{y}}$.

SOLUÇÃO:

Tomar a projecção ortogonal de $\vec{\mathbf{y}}$ sobre $\mathscr{C}(\mathbf{X})$: $\hat{\vec{\mathbf{y}}} = \mathbf{H}\vec{\mathbf{y}}$.

SQRE na projecção ortogonal



O quadrado da distância de $\vec{\mathbf{y}}$ a $\hat{\vec{\mathbf{y}}}$ é *SQRE*, a soma dos quadrados dos resíduos.

A projecção ortogonal

A projecção ortogonal de um vector $\vec{\mathbf{y}} \in \mathbb{R}^n$ sobre o subespaço $\mathscr{C}(\mathbf{X})$ gerado pelas colunas (linearmente independentes) de $\mathbf{X}$ faz-se pré-multiplicando $\vec{\mathbf{y}}$ pela matriz de projecção ortogonal sobre $\mathscr{C}(\mathbf{X})$:

Matriz de projecção ortogonal sobre $\mathscr{C}(\mathbf{X})$

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t .$$

As matrizes de projecção ortogonal $\mathbf{P}$ sobre algum subespaço de $\mathbb{R}^n$ são as matrizes $n \times n$:

- simétricas (isto é, $\mathbf{P}^t = \mathbf{P}$); e
- idempotentes (isto é, $\mathbf{PP} = \mathbf{P}$).

A matriz $\mathbf{H}$ tem estas propriedades (verifique!).

A projecção ortogonal no contexto da RLM

No contexto duma regressão linear múltipla, tem-se:

$$\begin{aligned} & \hat{\vec{y}} = \mathbf{H}\vec{y} \\ \Leftrightarrow \quad & \hat{\vec{y}} = \mathbf{X}\underbrace{(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{y}}_{=\vec{b}} \end{aligned}$$

A combinação linear dos vectores $\vec{1}_n, \vec{x}_1, \ldots, \vec{x}_p$ que gera o vector mais próximo de $\vec{y}$ tem coeficientes dados pelos elementos do vector $\vec{b}$:

O vector de parâmetros ajustado

$$\vec{b} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{y} \; .$$

Quando $p = 1$ (RLS): $\vec{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{y} = \begin{bmatrix} \overline{y} - b_1\overline{x} \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

As três Somas de Quadrados

Na Regressão Linear Múltipla definem-se três Somas de Quadrados, de forma idêntica ao que se fez na Regressão Linear Simples:

SQRE – Soma de Quadrados dos Resíduos (já definida):

$$SQRE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 .$$

SQT – Soma de Quadrados Total:

$$SQT = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 .$$

SQR – Soma de Quadrados associada à Regressão:

$$SQR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 = \sum_{i=1}^{n} \hat{y}_i^2 - n\overline{y}^2 .$$

$$SQT = SQR + SQRE$$

Nota: Também na RL Múltipla os y observados (y_i) e os y ajustados ($\hat{y}_i$) têm a mesma média.

O Coeficiente de Determinação na Regressão Linear, $R^2 = \frac{SQR}{SQT}$

Algumas propriedades dos hiperplanos ajustados

Numa regressão linear múltipla verifica-se:

- a média dos valores observados de Y, $\{y_i\}_{i=1}^n$, é igual à média dos respectivos valores ajustados, $\{\hat{y}_i\}_{i=1}^n$.
- O hiperplano ajustado em $\mathscr{R}^{p+1}$ contém o centro de gravidade da nuvem de pontos, i.e., o ponto de coordenadas $(\overline{x}_1, \overline{x}_2, ..., \overline{x}_p, \overline{y})$:

$$\overline{y} = \overline{\hat{y}} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{(b_o + b_1 x_{1(i)} + b_2 x_{2(i)} + \cdots + b_p x_{p(i)})}_{=\hat{y}_i} = b_0 + b_1 \overline{x}_1 + b_2 \overline{x}_2 + \cdots + b_p \overline{x}_p$$

- o coeficiente b_j que multiplica o preditor X_j é a variação média em Y, associada a aumentar X_j em 1 unidade, mantendo os restantes preditores constantes.
- o valor de R^2 numa regressão múltipla não pode ser inferior ao valor de R^2 que se obteria excluindo do modelo um qualquer subconjunto de preditores. Em particular, não pode ser inferior ao R^2 das regressões lineares simples de Y sobre cada preditor individual.

Propriedades de modelos com constante aditiva

$\mathscr{C}(\mathbf{X})$ contém o vector $\vec{1}_n$ de n uns. Então $\mathbf{H}\vec{1}_n = \vec{1}_n$, pois a projecção de qualquer vector num subespaço que já o contém deixa o vector invariante. Logo:

- As médias dos valores observados e ajustados de Y são iguais:

  $$\overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \frac{1}{n}\vec{1}_n^t\vec{\hat{y}} = \frac{1}{n}\vec{1}_n^t\mathbf{H}\vec{y} = \frac{1}{n}\vec{1}_n^t\mathbf{H}^t\vec{y} = \frac{1}{n}(\mathbf{H}\vec{1}_n)^t\vec{y} = \frac{1}{n}\vec{1}_n^t\vec{y} = \overline{y}$$

- A soma dos resíduos é zero:

  $$\overline{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) = \overline{y} - \overline{\hat{y}} = 0.$$

- Em $\mathbb{R}^{p+1}$, o hiperplano ajustado contém o centro de gravidade da nuvem dos n pontos observados: $\overline{y} = b_0 + b_1\overline{x_1} + b_2\overline{x_2} + \ldots + b_p\overline{x_p}$.

  Já vimos que $\overline{y} = \overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i$. Mas $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i =$

  $$\frac{1}{n}\sum_{i=1}^{n}(b_o + b_1 x_{1(i)} + b_2 x_{2(i)} + \cdots + b_p x_{p(i)}) = b_o + b_1\overline{x_1} + b_2\overline{x_2} + \cdots + b_p\overline{x_p}$$

Os coeficientes b_j

O vector dos parâmetros ajustados pelo método dos mínimos quadrados, $\vec{\mathbf{b}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{y}}$, gera n valores ajustados:

$$\begin{aligned} \vec{\hat{\mathbf{y}}} &= \mathbf{H}\vec{\mathbf{y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{y}} = \mathbf{X}\vec{\mathbf{b}} \\ \Leftrightarrow \quad \hat{y}_i &= b_0 + b_1 x_{1(i)} + \ldots + b_p x_{p(i)} \quad , \quad \forall\, i\ . \end{aligned}$$

As unidades de medida:

- de b_0 são iguais às de y (e às de $\hat{y}$).
- dos parâmetros b_j das variáveis ($j \neq 0$) são a razão entre as unidades de y e as do preditor x_j correspondente.

Os coeficientes $\{b_j\}_{j=1}^{p}$ das variáveis preditoras interpretam-se como a diferença (média) em y, associada a aumentar o preditor x_j correspondente em uma unidade, mantendo os restantes preditores constantes.

Resíduos

As unidades de medida dos resíduos $e_i = y_i - \hat{y}_i$ são iguais às de y:

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_{1(i)} + \ldots + b_p x_{p(i)}) \quad , \quad \forall\, i$$

$$\Leftrightarrow \quad \vec{\mathbf{e}} = \vec{\mathbf{y}} - \vec{\hat{\mathbf{y}}} = \vec{\mathbf{y}} - \mathbf{H}\vec{\mathbf{y}} \, ,$$

O vector de resíduos, $\vec{\mathbf{e}}$, também pode ser obtido pré-multiplicando o vector $\vec{\mathbf{y}}$ pela matriz $\mathbf{I} - \mathbf{H}$, onde $\mathbf{I}$ é a matriz identidade $n \times n$:

$$\vec{\mathbf{e}} = \vec{\mathbf{y}} - \mathbf{H}\vec{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\vec{\mathbf{y}} \, ,$$

A regressão Linear no SAS

Por exemplo, o **proc reg** ajusta uma regressão linear.

Ilustra-se uma Regressão Linear Múltipla com um conjunto de dados famoso (iris) : os lírios de Anderson/Fisher.

analises-regressao PROC REG running

```
data iris;
input SepalLength SepalWidth PetalLength PetalWidth  Species$;
datalines;
5.1 3.5 1.4 0.2 setosa
4.9 3   1.4 0.2 setosa
4.7 3.2 1.3 0.2 setosa
4.6 3.1 1.5 0.2 setosa
5   3.6 1.4 0.2 setosa
5.4 3.9 1.7 0.4 setosa
4.6 3.4 1.4 0.3 setosa
5   3.4 1.5 0.2 setosa
```

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.

...

```
proc reg data=iris;
    model PetalWidth = SepalLength SepalWidth PetalLength/clb;
run;
```

Variável resposta

Variáveis preditoras, p=3

Ajuste-se um modelo para prever a variável resposta largura da pétala, a partir do comprimento da pétala e das duas medições das sépalas (largura e comprimento), ignorando as espécies.

A regressão Linear no SAS (cont.)

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-0.24031	0.17837	-1.35	0.1800	-0.59283	0.11221
SepalLength	1	-0.20727	0.04751	-4.36	<.0001	-0.30115	-0.11338
SepalWidth	1	0.22283	0.04894	4.55	<.0001	0.12611	0.31955
PetalLength	1	0.52408	0.02449	21.40	<.0001	0.47568	0.57249

o vector $\vec{\mathbf{b}}$ das estimativas dos $p+1$ parâmetros

$$\vec{b}=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} -0.24031 \\ -0.20727 \\ 0.22283 \\ 0.52408 \end{bmatrix}$$

O hiperplano ajustado em $\mathbb{R}^4$ $(\mathbb{R}^{p+1})$ é:

$$PW = -0.24031 - 0.20727SL + 0.22283SW + 0.52408PL$$

Modelos e submodelos

Submodelos

Dado um modelo de regressão linear múltipla, com equação

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p \ ,$$

chama-se submodelo a uma regressão linear com apenas alguns preditores.

Por exemplo, a regressão linear simples

$$Petal.Width = b_0 + b_1 Petal.Length$$

é um submodelo da regressão linear múltipla acabada de ajustar,

$$PW = b_o + b_1 Sepal.Length + b_2 Sepal.Width + b_3 Petal.Length$$

Nota: Um submodelo (S) não pode ter preditores que não façam parte do modelo completo (C). A variável resposta tem de ser a mesma.

O R^2 de submodelos

Coeficientes de Determinação de submodelos: $R_s^2 \leq R_c^2$

O R_s^2 dum submodelo não pode exceder o R_c^2 do modelo completo.
O subespaço das colunas do submodelo tem de estar contido no subespaço das colunas do modelo completo: $\mathscr{C}(\mathbf{X}_s) \subseteq \mathscr{C}(\mathbf{X}_c)$. Logo, o ângulo entre $\vec{\mathbf{y}}$ e $\hat{\vec{\mathbf{y}}}_s \in \mathscr{C}(\mathbf{X}_s)$ não pode ser menor que o ângulo entre $\vec{\mathbf{y}}$ e $\hat{\vec{\mathbf{y}}}_c \in \mathscr{C}(\mathbf{X}_c)$, pois $\hat{\vec{\mathbf{y}}}_s$ também pertence a $\mathscr{C}(\mathbf{X}_c)$.

Ainda o exemplo dos lírios

RLM

```
proc reg data=iris;
    model PetalWidth = SepalLength SepalWidth PetalLength/clb;
run;
```

R_c^2

Root MSE	0.19197	R-Square	0.9379
Dependent Mean	1.19933	Adj R-Sq	0.9366
Coeff Var	16.00615		

RLS

```
proc reg data=iris;
    model PetalWidth = PetalLength/clb;
run;
```

R_s^2

Root MSE	0.20648	R-Square	0.9271
Dependent Mean	1.19933	Adj R-Sq	0.9266
Coeff Var	17.21659		

Equações de submodelos

Os parâmetros ajustados não são iguais

A equação ajustada num submodelo não é a parte correspondente na equação ajustada do modelo.

Ainda o exemplo dos lírios

RLM

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-0.24031	0.17837	-1.35	0.1800	-0.59283	0.11221
SepalLength	1	-0.20727	0.04751	-4.36	<.0001	-0.30115	-0.11338
SepalWidth	1	0.22283	0.04894	4.55	<.0001	0.12611	0.31955
PetalLength	1	0.52408	0.02449	21.40	<.0001	0.47568	0.57249

RLS

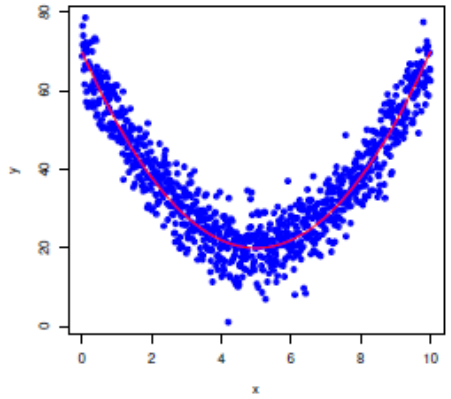
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-0.36308	0.03976	-9.13	<.0001	-0.44165	-0.28450
PetalLength	1	0.41576	0.00958	43.39	<.0001	0.39682	0.43469

Regressão Polinomial

Um caso particular de relação não-linear, mesmo que envolvendo apenas uma variável preditora e a variável resposta, pode ser facilmente tratada no âmbito duma regressão linear múltipla: o caso de relações polinomiais entre *Y* e um ou mais preditores.

Imagine-se uma relação de fundo entre uma variável resposta *Y* e uma única variável preditora *X* dada por uma parábola:



Pode ajustar-se uma qualquer parábola, com equação

$$y = b_0 + b_1 x + b_2 x^2$$

com uma regressão linear de y sobre os dois preditores $x_1 = x$ e $x_2 = x^2$

Nota: aplicável a qualquer polinómio de qualquer grau e em qualquer número de variáveis.