

Regressão Linear Múltipla

II.2. Regressão Linear Múltipla

Por vezes, é necessário **mais do que uma variável preditora** para modelar adequadamente a variável resposta de interesse.

Exemplo: teor de antocianinas

Num estudo sobre uma população experimental de clones da casta Tinta Francisca, realizado no Tabuaço em 2003, foram medidos os valores das seguintes variáveis para 24 videiras:

- **teor de antocianinas** (variável **antoci**, em mg/dm^3);
- **fenóis totais** (variável **fentot**);
- **pH** (variável **pH**).

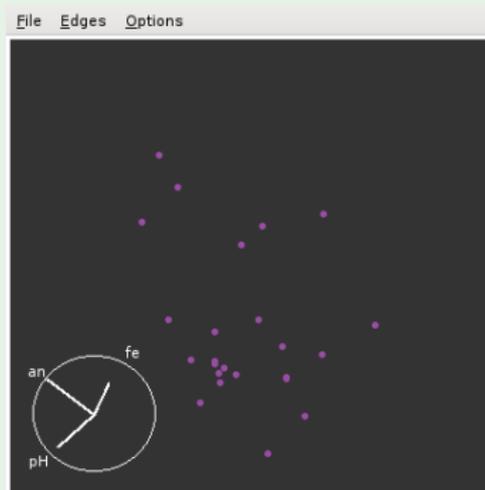
Há interesse em estudar a relação entre o teor de antocianinas (variável resposta) e o teor de fenóis totais e pH.

A nuvem de pontos - uma perspectiva

Exemplo antocianinas (cont.)

$n=24$ observações em 3 variáveis definem uma nuvem de 24 pontos em \mathbb{R}^3 . A visualização de nuvens de pontos em \mathbb{R}^3 exige *software* específico.

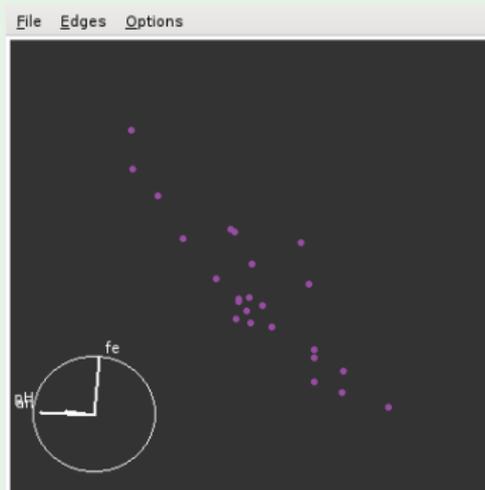
Neste ângulo de visão, a nuvem de pontos não parece ter relação especial:



A nuvem de pontos - outra perspectiva

Exemplo antocianinas (cont.)

Noutro ângulo de visão vê-se que os pontos se dispersam aproximadamente em torno de **um plano**:



Plano em \mathbb{R}^3

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

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

No nosso contexto, e colocando:

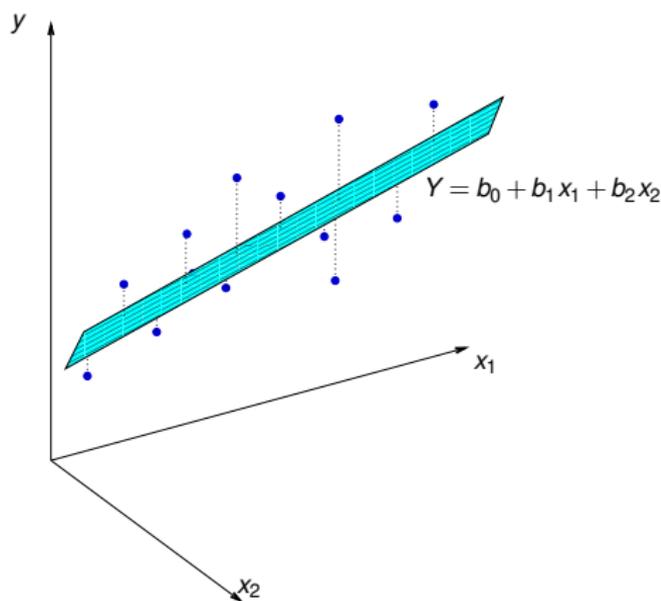
- no eixo vertical (z) a variável resposta Y ;
- noutro eixo (x) um preditor X_1 ;
- no terceiro eixo (y) o outro preditor X_2 ,

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

$$\begin{aligned} Ax_1 + Bx_2 + Cy + D = 0 &\Leftrightarrow y = -\frac{D}{C} - \frac{A}{C}x_1 - \frac{B}{C}x_2 \\ &\Leftrightarrow y = b_0 + b_1x_1 + b_2x_2 \end{aligned}$$

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 ($x_1 \geq 0, x_2 \geq 0, y \geq 0$).
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, \dots, x_p , admite-se que os valores de Y oscilam em torno duma combinação linear (afim) das p variáveis preditoras:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p .$$

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:

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

O caso geral: valores ajustados e resíduos

O hiperplano $y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$ serve para obter **valores ajustados** \hat{y}_i , para cada uma das n observações:

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

Definem-se de forma igual os **resíduos** associados a cada observação:

$$e_i = y_i - \hat{y}_i$$

A **Soma de Quadrados dos Resíduos** também se define de forma idêntica:

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

Os **parâmetros** b_j do hiperplano ajustado deverão **minimizar SQRE**.

As dificuldades na representação gráfica

A **representação gráfica usual** da nuvem de n pontos observados **exige $p + 1$ eixos**: um para Y e um para cada um dos p preditores.

Para $p > 2$, são necessários mais de três eixos e **a visualização torna-se impossível**.

As características fundamentais dessas representações são:

- $p + 1$ eixos – um para cada **variável** em questão.
- n pontos – um para cada **indivíduo (unidade experimental)** observado.
- Tem-se uma **nuvem de n pontos num espaço $(p + 1)$ -dimensional**.

Na regressão linear múltipla admite-se que os pontos se dispõem em torno dum **hiperplano em \mathbb{R}^{p+1}** , de equação $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$.

Outra representação gráfica

A representação gráfica em \mathbb{R}^{p+1} das n observações de Y e as p variáveis preditoras não é a única possível.

Há **outra representação possível** dos dados, que **casa conceitos geométricos e conceitos estatísticos** e é útil na determinação dos parâmetros ajustados.

As n observações de Y definem um **vector em \mathbb{R}^n** :

$$\vec{y} = (y_1, y_2, y_3, \dots, y_n).$$

Da mesma forma, as n observações de cada variável preditora definem um **vector de \mathbb{R}^n** .

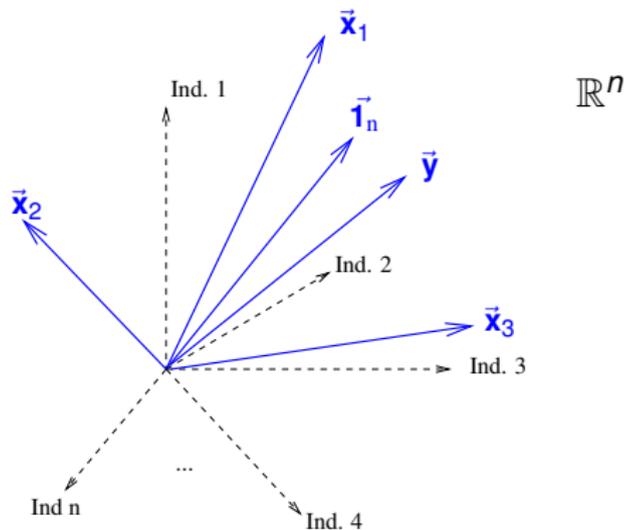
$$\vec{x}_j = (x_{j(1)}, x_{j(2)}, x_{j(3)}, \dots, x_{j(n)}) \quad (j = 1, 2, \dots, p).$$

Podemos representar todas as variáveis por vectores **em \mathbb{R}^n** .

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.



Vantagens da representação gráfica alternativa

Os n valores ajustados \hat{y}_i também definem um vector de \mathbb{R}^n que é uma combinação linear dos vectores $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \dots, \vec{\mathbf{x}}_p$:

$$\begin{aligned}\vec{\hat{y}} &= \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 x_{1(1)} + b_2 x_{2(1)} + \dots + b_p x_{p(1)} \\ b_0 + b_1 x_{1(2)} + b_2 x_{2(2)} + \dots + b_p x_{p(2)} \\ b_0 + b_1 x_{1(3)} + b_2 x_{2(3)} + \dots + b_p x_{p(3)} \\ \dots \\ b_0 + b_1 x_{1(n)} + b_2 x_{2(n)} + \dots + b_p x_{p(n)} \end{bmatrix} \\ &= b_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_{1(1)} \\ x_{1(2)} \\ x_{1(3)} \\ \vdots \\ x_{1(n)} \end{bmatrix} + \dots + b_p \begin{bmatrix} x_{p(1)} \\ x_{p(2)} \\ x_{p(3)} \\ \vdots \\ x_{p(n)} \end{bmatrix} \\ &= b_0 \vec{\mathbf{1}}_n + b_1 \vec{\mathbf{x}}_1 + b_2 \vec{\mathbf{x}}_2 + \dots + b_p \vec{\mathbf{x}}_p\end{aligned}$$

$\vec{\hat{y}}$ é uma combinação linear dos $p+1$ vectors $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \dots, \vec{\mathbf{x}}_p$.

A matriz do modelo \mathbf{X}

- O conjunto de **todas** as combinações lineares dos vectores $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_p$ chama-se o **subespaço gerado** por esses vectores (lembrar UC **Álgebra Linear** dos 1os. ciclos do ISA).
- O vector $\vec{\mathbf{y}}$ pertence ao subespaço gerado pelos vectores $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_p$.
- Define-se a **matriz do modelo \mathbf{X}** como a matriz cujas **colunas** são os vectores $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_p$.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \cdots & x_{p(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \cdots & x_{p(2)} \\ 1 & x_{1(3)} & x_{2(3)} & \cdots & x_{p(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1(n)} & x_{2(n)} & \cdots & x_{p(n)} \end{bmatrix}.$$

É uma matriz de dimensão $n \times (p + 1)$.

O subespaço de colunas de \mathbf{X}

- O subespaço gerado pelas colunas da matriz do modelo \mathbf{X} chama-se o **subespaço das colunas de \mathbf{X}** e representa-se por $\mathcal{C}(\mathbf{X})$.
- O vector $\vec{\hat{y}}$ pertence ao subespaço $\mathcal{C}(\mathbf{X})$
(as colunas de \mathbf{X} são os vectores $\vec{\mathbf{1}}_n, \vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_p$).
- $\mathcal{C}(\mathbf{X})$ é um subespaço de \mathbb{R}^n ($\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^n$), 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 $\mathcal{C}(\mathbf{X})$** se pode escrever como $\mathbf{X}\vec{\mathbf{a}}$, onde $\vec{\mathbf{a}} = (a_0, a_1, a_2, \dots, a_p)$ é o vector dos coeficientes da combinação linear.

Um produto matricial $\mathbf{X}\vec{a}$

$$\begin{aligned}\mathbf{X}\vec{a} &= \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \cdots & x_{p(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \cdots & x_{p(2)} \\ 1 & x_{1(3)} & x_{2(3)} & \cdots & x_{p(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1(n)} & x_{2(n)} & \cdots & x_{p(n)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \\ &= \begin{bmatrix} a_0 + a_1 x_{1(1)} + a_2 x_{2(1)} + \dots + a_p x_{p(1)} \\ a_0 + a_1 x_{1(2)} + a_2 x_{2(2)} + \dots + a_p x_{p(2)} \\ a_0 + a_1 x_{1(3)} + a_2 x_{2(3)} + \dots + a_p x_{p(3)} \\ \dots \\ a_0 + a_1 x_{1(n)} + a_2 x_{2(n)} + \dots + a_p x_{p(n)} \end{bmatrix} \\ &= a_0 \vec{\mathbf{1}}_n + a_1 \vec{\mathbf{x}}_1 + a_2 \vec{\mathbf{x}}_2 + \dots + a_p \vec{\mathbf{x}}_p \in \mathcal{C}(\mathbf{X}).\end{aligned}$$

Os parâmetros

- Cada escolha possível de coeficientes $\vec{\mathbf{a}} = (a_0, a_1, a_2, \dots, a_p)$ corresponde a um ponto/vector no subespaço $\mathcal{C}(\mathbf{X})$.
- 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 $\vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_p, \vec{\mathbf{1}}_n$.
- Um dos pontos/vectores do subespaço é a combinação linear dada pelo vector de coeficientes $\vec{\mathbf{b}} = (b_0, b_1, \dots, b_p)$ que minimiza:

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

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

Como identificar esse ponto/vector?

Geometria

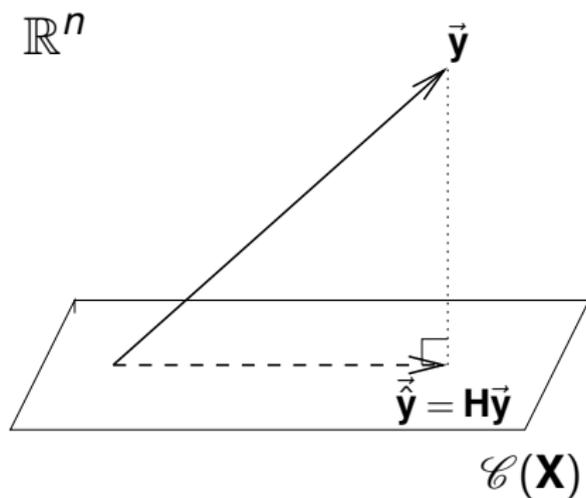
Vamos usar argumentos geométricos.

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

SOLUÇÃO:

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

A projecção ortogonal de \vec{y} sobre $\mathcal{C}(\mathbf{X})$



O vector de $\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^n$ mais próximo dum vector $\vec{y} \in \mathbb{R}^n$ é o vector $\hat{\vec{y}}$ que resulta de projectar ortogonalmente \vec{y} sobre $\mathcal{C}(\mathbf{X})$.

O critério minimiza *SQRE*

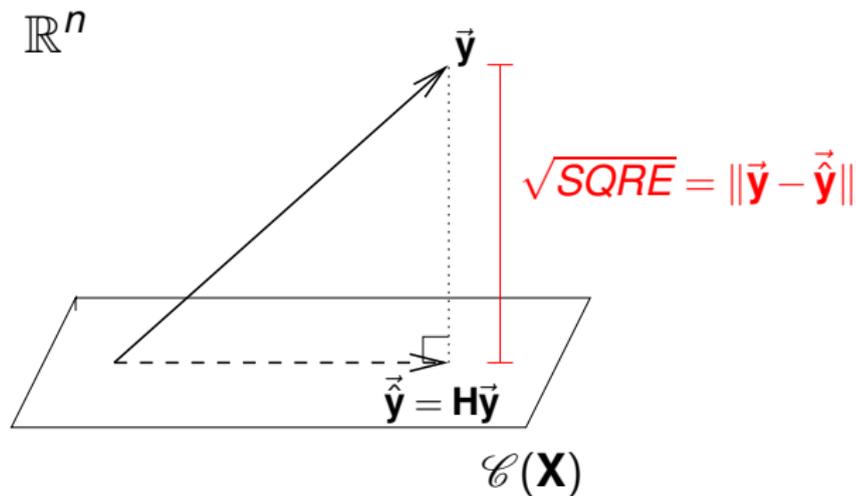
O vector $\vec{\hat{y}}$ que minimiza a distância ao vector de observações \vec{y} minimiza também o **quadrado dessa distância**, que é dado por:

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

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

Trata-se do **mesmo critério** que foi usado na Regressão Linear Simples.

SQRE na projecção ortogonal



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

A projecção ortogonal

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

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

Logo, temos:

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

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

Os parâmetros ajustados na RL Múltipla

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

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:

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

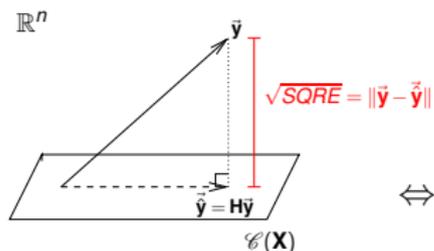
SQR – Soma de Quadrados associada à Regressão:

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

Nota: Também aqui os y observados (y_i) e os y ajustados (\hat{y}_i) têm a mesma média (ver Exercício RLM 4).

Pitágoras e a Regressão

O **Teorema de Pitágoras** aplica-se em qualquer espaço euclidiano \mathbb{R}^n .
Aplicado ao triângulo rectângulo do acetato 179 tem-se:



$$\begin{aligned}\|\vec{y}\|^2 &= \|\vec{\hat{y}}\|^2 + \|\vec{y} - \vec{\hat{y}}\|^2 \\ \Leftrightarrow \sum_{i=1}^n y_i^2 &= \sum_{i=1}^n \hat{y}_i^2 + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{= SQRE} \\ \Leftrightarrow \sum_{i=1}^n y_i^2 - n\bar{y}^2 &= \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 + SQRE \\ \Leftrightarrow SQT &= SQR + SQRE\end{aligned}$$

Revisitando Pitágoras

A relação fundamental da Regressão Linear ($SQT = SQR + SQRE$) resulta da aplicação do Teorema de Pitágoras. Mas foi necessário introduzir subtrair $n\bar{y}^2$. Um outro triângulo rectângulo é estatisticamente mais interessante.

Seja \vec{y}^c o **vector centrado** das observações da variável resposta, isto é, o vector **de elemento genérico** $y_i - \bar{y}$. Este vector obtém-se subtraindo a \vec{y} o vector que repete n vezes o escalar \bar{y} :

$$\vec{y}^c = \vec{y} - \bar{y}\mathbf{1}_n = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})^t.$$

A norma deste vector é $\|\vec{y}^c\| = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{SQT}$.

Revisitando Pitágoras (cont.)

A projecção ortogonal do vector \vec{y}^c sobre o subespaço $\mathcal{C}(\mathbf{X})$ gera o vector:

$$\begin{aligned}\mathbf{H}\vec{y}^c &= \mathbf{H}(\vec{y} - \bar{y}\vec{1}_n) \\ \Leftrightarrow \mathbf{H}\vec{y}^c &= \mathbf{H}\vec{y} - \bar{y}\mathbf{H}\vec{1}_n \\ \Leftrightarrow \mathbf{H}\vec{y}^c &= \hat{\vec{y}} - \bar{y}\vec{1}_n\end{aligned}$$

já que $\mathbf{H}\vec{1}_n = \vec{1}_n$, pois o vector $\vec{1}_n$ pertence ao subespaço $\mathcal{C}(\mathbf{X})$, logo fica invariante quando projectado nesse mesmo subespaço (Exercício RLM 4b).

O vector $\mathbf{H}\vec{y}^c$ tem elemento genérico $\hat{y}_i - \bar{y}$, e a sua norma é

$$\|\mathbf{H}\vec{y}^c\| = \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} = \sqrt{SQR}.$$

Revisitando Pitágoras (cont.)

A distância entre o vector \vec{y}^c e a sua projecção ortogonal sobre $\mathcal{C}(\mathbf{X})$ continua a ser \sqrt{SQRE} :

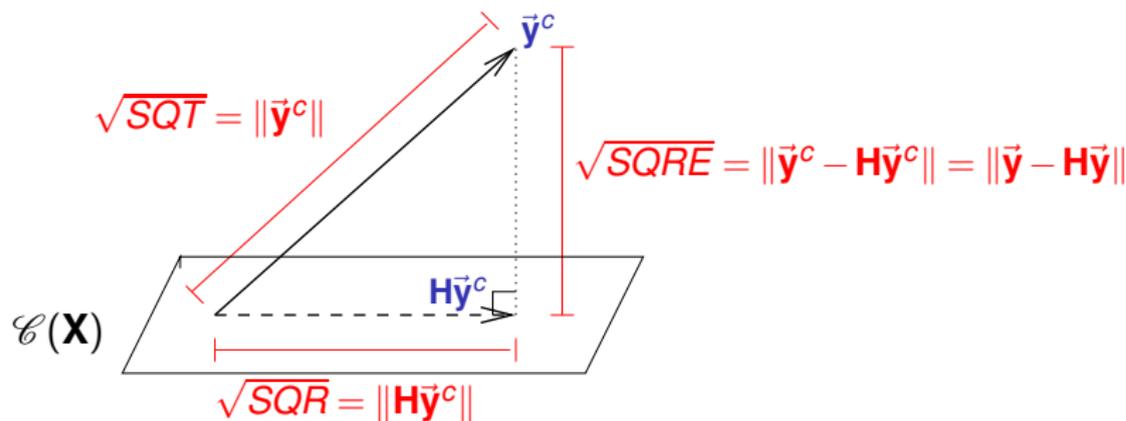
$$\begin{aligned}\vec{y}^c - \mathbf{H}\vec{y}^c &= (\vec{y} - \bar{y}\vec{1}_n) - (\vec{\hat{y}} - \bar{y}\vec{1}_n) \\ \Leftrightarrow \vec{y}^c - \mathbf{H}\vec{y}^c &= \vec{y} - \vec{\hat{y}}\end{aligned}$$

pelo que

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

Revisitando Pitágoras (cont.)

\mathbb{R}^n



A fórmula fundamental da Regressão Linear, $SQT = SQR + SQRE$, é uma aplicação directa do Teorema de Pitágoras ao triângulo definido por \vec{y}^c e a sua projecção ortogonal sobre $\mathcal{C}(\mathbf{X})$.

Pitágoras e o Coeficiente de Determinação

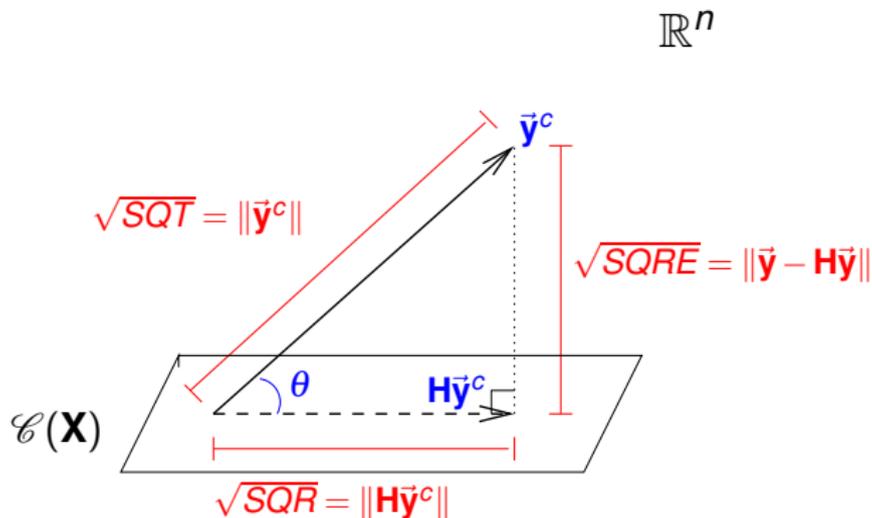
Torna-se evidente outra relação importante entre a geometria e a estatística da Regressão Linear.

O **coeficiente de determinação** $R^2 = \frac{SQR}{SQT}$ é o quadrado do cosseno do ângulo entre o vector centrado das observações da variável resposta, \vec{y}^c , e a sua projecção ortogonal sobre o subespaço $\mathcal{C}(\mathbf{X})$:

$$\cos^2(\theta) = \frac{SQR}{SQT} = R^2,$$

onde θ é o ângulo entre os vectores \vec{y}^c e $\mathbf{H}\vec{y}^c$.

Pitágoras e o Coeficiente de Determinação (cont.)



O Coeficiente de Determinação na Regressão Linear, $R^2 = \frac{SQR}{SQT}$, é o quadrado do cosseno do ângulo entre \vec{y}^c e $H\vec{y}^c$.

Propriedades do Coeficiente de Determinação

A abordagem geométrica confirma que, também na Regressão Linear Múltipla, são válidas as propriedades (já conhecidas da Regressão Linear Simples) do Coeficiente de Determinação:

- R^2 toma valores entre 0 e 1.
- Quanto mais próximo de 1 estiver R^2 , menor o ângulo θ , e portanto melhor será a correspondência entre o vector (centrado) das observações, \vec{y}^c , e o seu ajustamento em $\mathcal{C}(\mathbf{X})$.
- Se $R^2 \approx 0$, o vector \vec{y}^c é quase perpendicular ao subespaço $\mathcal{C}(\mathbf{X})$ onde se pretende aproximá-lo, e a projecção vai quase anular todas os elementos do vector projectado. **O resultado será de má qualidade:** perde-se quase toda a variabilidade nos valores de y .

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, $\{\hat{y}_i\}_{i=1}^n$ (ver Exercício RLM 4c).
- O hiperplano ajustado em \mathcal{R}^{p+1} contém o centro de gravidade da nuvem de pontos, i.e., o ponto de coordenadas $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p, \bar{y})$:

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

- o coeficiente b_j que multiplica o preditor X_j é a variação média em Y , associada a aumentar a 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.

A Regressão Múltipla no

Uma Regressão Múltipla no  estuda-se através do mesmo comando `lm` usado para a regressão linear simples. A indicação de qual a variável resposta y e quais as variáveis preditoras x_1, \dots, x_p faz-se de forma semelhante à da RLS.

Por exemplo, se a variável resposta se chama y e existirem três preditores de nome x_1 , x_2 e x_3 , a fórmula que indica a relação será:

$$y \sim x_1 + x_2 + x_3$$

O comando correspondente no R será:

```
> lm ( y ~ x1 + x2 + x3 , data=dados)
```

O resultado produzido por este comando será o **vector das estimativas dos $p+1$ parâmetros do modelo, b_0, b_1, \dots, b_p .**

A Regressão Múltipla no (cont.)

O exemplo dos lírios (*iris*, Ex. 8 e 15)

Pretende-se modelar a variável resposta largura da pétala, não apenas com base no comprimento da pétala, mas também das duas medições (largura e comprimento) das sépalas.

```
> iris2.lm <- lm(Petal.Width ~ Petal.Length + Sepal.Length +  
+ Sepal.Width , data=iris)
```

```
> iris2.lm
```

```
(...)
```

Coefficients:

(Intercept)	Petal.Length	Sepal.Length	Sepal.Width
-0.2403	0.5241	-0.2073	0.2228

O hiperplano ajustado é:

$$PW = -0.2403 + 0.5241 PL - 0.2073 SL + 0.2228 SW$$

O coeficiente de determinação é $R^2 = 0.9379$, ligeiramente maior que o valor $R^2 = 0.9271$ do modelo RLS (acetato 135).