Exercícios - Estatística e Delineamento Experimental - 2024-25

AVISO:

Muitos dos exercícios utilizam conjuntos dados disponibilizados na página *web* da disciplina, na secção *Materiais de apoio*, subsecção *Dados*.

1 Regressão Linear - abordagem descritiva

1.1 Exercícios práticos

1. Com base nos dados do Instituto Nacional de Estatística (INE), foi criado um ficheiro em formato CSV (*Comma separated values*) chamado `cereais.csv` contendo a evolução da superfície agrícola utilizada anualmente na produção de cereais para grão (variável `area`, em km^2) em Portugal, no período de 1986 a 2011 (variável `ano`). O ficheiro encontra-se disponível na página *web* da disciplina e deve ser descarregado e guardado na directoria onde se localiza a sessão de trabalho do `R`. Seguidamente o seu conteúdo pode ser lido para a sessão do `R` através do comando:

```
> Cereais <- read.csv("cereais.csv")
```

   (a) Construa uma nuvem de pontos de superfície agrícola *vs.* ano e comente.

   (b) A partir do gráfico obtido na alínea anterior, sugira um valor para o coeficiente de correlação entre superfície agrícola e ano. Depois, utilize os comandos do `R` para calcular esse mesmo coeficiente de correlação. Comente o seu significado.

   (c) Ajuste uma recta de regressão de superfície agrícola utilizada sobre anos. Discuta o significado dos parâmetros da recta ajustada, no contexto do problema sob estudo.

   (d) Comente a qualidade da recta obtida, calculando o respectivo coeficiente de determinação e interpretando o valor obtido.

   (e) Trace a recta de regressão ajustada em cima da nuvem de pontos e comente.

   (f) Calcule a Soma de Quadrados Total (SQT), a partir do cálculo da variância amostral de y.

   (g) Calcule o valor da Soma de Quadrados da Regressão (SQR).

   (h) Calcule a Soma de Quadrados dos Resíduos (SQRE), directamente a partir dos resíduos, e verifique numericamente a relação fundamental da Regressão Linear: SQT=SQR+SQRE.

   (i) Altere as unidades de medida da variável `area`, de km^2 para hectares (`area` $\rightarrow$ `area`$\times 100$). Ajuste novamente a regressão, após efectuar esta alteração. O que aconteceu aos parâmetros estimados e ao coeficiente de determinação R^2? Comente.

   (j) De novo a partir dos dados originais, transforme a variável `ano` num contador dos anos do estudo (`ano` $\rightarrow$ `ano`-1985). Ajuste novamente a regressão, após efectuar esta alteração. O que aconteceu aos parâmetros estimados e ao coeficiente de determinação R^2? Comente.

2. O ficheiro `azeite.xls`, disponível na página *web* da disciplina, é um ficheiro de tipo folha de cálculo que contém dados relativos à produção de azeite em Portugal no período 1995-2010, disponibilizados pelo Instituto Nacional de Estatística (www.ine.pt). As colunas "Azeitona" e "Azeite" correspondem à produção de azeitona oleificada (em t) e azeite (em hl), respectivamente.

(a) Abra o ficheiro `azeite.xls` com um programa do tipo *Office* e guarde a folha de cálculo num ficheiro de texto de nome `azeite.txt`, utilizando o *Save as* com a opção *Ficheiro de Texto*. Coloque esse ficheiro na pasta de trabalho do R.

(b) Numa sessão do R, guarde os dados do ficheiro `azeite.txt` (criado na alínea anterior) numa *data frame* de nome `azeite`. Pode usar o comando:

```
> azeite <- read.table("azeite.txt", header=TRUE)
```

(c) Crie a nuvem de pontos relacionando as produções de Azeite (eixo vertical, variável y) e Azeitona (eixo horizontal, variável x).

(d) Com base na nuvem de pontos, sugira um valor para o coeficiente de correlação entre as duas variáveis. Avalie a sua sugestão calculando o valor de r_{xy}. Comente o valor obtido.

(e) Calcule as estimativas de mínimos quadrados para os parâmetros da recta de regressão, e comente o seu significado.

(f) Calcule a precisão da recta de regressão estimada de y sobre x e comente o valor obtido.

3. O programa `R` tem um grande número de pacotes adicionais disponíveis. Um desses pacotes adicionais designa-se `MASS`. Pode ser carregado mediante o comando `library(MASS)`.

Considere o conjunto de dados `Animals`, disponível no referido módulo `MASS`, onde se listam pesos médios dos cérebros (em g) e dos corpos (em kg) para 28 espécies de animais terrestres. Pretende-se estudar uma relação entre pesos do cérebro (variável resposta, y) e pesos do corpo (variável preditora, x).

(a) Construa uma nuvem de pontos de pesos do corpo (eixo horizontal) e pesos do cérebro (eixo vertical). Calcule o coeficiente de correlação correspondente e comente.

(b) Construa nuvens de pontos com as seguintes transformações de uma ou ambas as variáveis:

i. $\ln(y)$ vs. x;
ii. y vs. $\ln(x)$;
iii. $\ln(y)$ vs. $\ln(x)$.

(c) Considere uma relação linear entre $\ln(y)$ e $\ln(x)$. Explicite a relação de base correspondente entre as variáveis originais (não logaritmizadas). Comente.

Nas alíneas seguintes considere sempre os *dados logaritmizados*.

(d) Calcule os coeficientes de correlação e de determinação associados à relação entre $\ln(x)$ e $\ln(y)$. Interprete os valores obtidos. Como se explica que o Coeficiente de Determinação não seja particularmente elevado, sendo evidente a partir da nuvem de pontos que existe uma boa relação linear entre log-peso do corpo e log-peso do cérebro para a generalidade das espécies?

(e) Ajuste a recta de regressão de log-peso do cérebro sobre log-peso do corpo (utilizando a totalidade das observações). Trace essa recta sobre a nuvem de pontos e comente.

(f) Considere agora a estimativa para o declive da recta, $b_1 = 0.49599$. Qual o significado biológico deste valor, quer na relação entre variáveis logaritmizadas, quer na relação entre as variáveis originais (não logaritmizadas)?

(g) Considere a nuvem de pontos das variáveis logaritmizadas. Identifique os três pontos que se destacam na parte inferior direita da nuvem. (NOTA: explore o comando `identify` do R). Comente.

(h) Ajuste a recta de regressão de log-peso do cérebro sobre log-peso do corpo considerando apenas os dados (logaritmizados) respeitantes a espécies que *não sejam de dinossáurios*. Trace essa recta sobre a nuvem de pontos e comente. Em particular, como se explica a elevação considerável no valor do

coeficiente de determinação? (NOTA: Aproveite a nuvem de pontos anterior, com a totalidade das espécies, para melhor compreender o efeito da exclusão das três espécies de dinossáurios sobre a recta ajustada).

4. Num estudo sobre poluição numa grande cidade, foram efectuadas medições, em 116 dias, da quantidade de ozono no ar (em partes por mil milhões) às 14h00 e da temperatura máxima (em °C) no respectivo dia. Essas observações encontram-se num ficheiro em formato `csv` de nome `ozono.csv`, que se encontra disponível na página *web* da disciplina e pode ser descarregado para a directoria de trabalho da sessão do `R`. Seguidamente, o conteúdo desse ficheiro pode ser lido para dentro da sessão do `R` e armazenado num objecto de nome `ozono`, através do comando `read.csv`:

```
> ozono <- read.csv("ozono.csv")
```

   (a) Construa a nuvem de pontos de ozono (eixo vertical) *vs.* temperatura máxima (eixo horizontal).

   (b) Tendo em conta a curvatura observada no gráfico, foi sugerido o ajustamento dum modelo exponencial, da forma $y = a\,\mathrm{e}^{b\,x}$.

      i. Construa a nuvem de pontos com as transformações adequadas para verificar se o modelo exponencial é, efectivamente, uma boa opção.

      ii. Ajuste o modelo *linearizado* recorrendo ao comando `lm` do `R`. Determine o respectivo coeficiente de determinação e comente.

      iii. Interprete os parâmetros da recta que ajustou, directamente em termos do modelo exponencial.

      iv. Indique, justificando, qual o teor médio de ozono (em partes por mil milhões) estimado pelo modelo ajustado, para um dia em que a temperatura máxima seja de 25°C.

   (c) Considere novamente a nuvem de pontos original. Trace a curva exponencial correspondente ao ajustamento efectuado na alínea anterior.

5. Num estudo sobre reacções enzimáticas, procura-se analisar a "velocidade" da reacção em células tratadas com puromicina. Para diferentes concentrações do substrato (variável *conc*), medidas em partes por milhão (ppm), registou-se o número de emissões radioactivas por minuto, e a partir destas calculou-se a taxa inicial ou "velocidade" da reacção, em contagens/minuto/minuto (variável *taxa*). Os resultados obtidos são dados na tabela seguinte e encontram-se *nas duas primeiras colunas* e nas *doze primeiras linhas* da *data frame* `Puromycin` do `R`, com as designações `conc` e `rate`, respectivamente (são as linhas a que corresponde o nível `treated` no factor `state`):

conc	0.02	0.02	0.06	0.06	0.11	0.11	0.22	0.22	0.56	0.56	1.10	1.10
taxa	76	47	97	107	123	139	159	152	191	201	207	200

   (a) Construa a nuvem de pontos das taxas da reacção (eixo vertical) *vs.* concentrações do substrato (eixo horizontal).

   (b) Tendo em conta a curvatura observada no gráfico, admite-se que o modelo de Michaelis-Menten é adequado à descrição da relação referida, e decide-se usar este modelo com a seguinte parametrização (onde y representa a *taxa* e x a concentração *conc*),

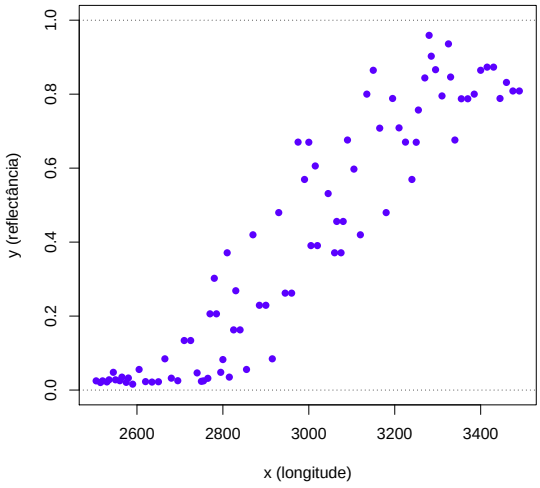
$$y = \frac{ax}{b+x} \qquad (a > 0, b > 0 \quad \text{e} \quad x > 0)\ .$$

      i. Mostre que o modelo referido pode ser linearizado, indicando a relação linearizada e as transformações de variáveis necessárias.

      ii. Ajuste o modelo linearizado que escolheu na alínea anterior, através do comando `lm` do `R`.

iii. Estime os parâmetros a e b na relação original no modelo de Michaelis-Menten. Como interpreta o valor estimado do parâmetro a? Trace a curva correspondente ao ajustamento na nuvem de pontos original para melhor compreender o seu significado. Comente o resultado.

6. No estudo de imagens obtidas por detecção remota, é importante a *reflectância*, que mede a proporção de radiação incidente que é reflectida pela superfície estudada. Por forma a determinar a localização de uma estrutura no terreno, que separa uma zona de baixa reflectância (valores próximos de 0) de uma região de muito alta reflectância (valores próximos de 1) foram realizadas medições da posição (variável x correspondente à longitude, em metros) e de reflectância (variável y, adimensional) para um conjunto de 85 pixels de uma imagem obtida por detecção remota. Os valores observados são indicados no gráfico.



As médias e variâncias das observações de cada variável observada, bem como o respectivo coeficiente de correlação, são:

$$\bar{x} = 2966.882 \qquad s_x^2 = 84859.51 \qquad \bar{y} = 0.4010485 \qquad s_y^2 = 0.1077003 \qquad r_{xy} = 0.9326$$

(a) Ajuste a recta de regressão linear de reflectância sobre longitude. Qual a proporção da variabilidade total da reflectância que é explicada por esta regressão?

(b) Tendo em conta a nuvem de pontos e o facto de as reflectâncias apenas tomarem valores no intervalo $]0, 1[$, é sugerida uma relação logística entre x e y, de equação $y = \frac{1}{1+e^{-(c+dx)}}$. Mostre que essa relação pode ser linearizada através da transformação $y^* = \ln\left(\frac{y}{1-y}\right)$, indicando a relação dos parâmetros da recta obtida com os parâmetros da equação da logística.

(c) Considere agora a recta de regressão *do modelo linearizado*, ou seja, de y^* sobre x. Os parâmetros do modelo ajustado são $b_0 = -20.50$ e $b_1 = 0.006629$. O Quadrado Médio Residual obtido é $QMRE = 0.6081371$. Calcule o valor do coeficiente de determinação do modelo linearizado.

7. No repositório de dados (`http://archive.ics.uci.edu/ml/`) da Universidade da Califórnia, Irvine, encontra-se um conjunto de dados "Wine recognition data" (fonte: Forina, M. et al, *PARVUS - An Extendible Package for Data Exploration, Classification and Correlation.* Institute of Pharmaceutical and

Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy) que contém os resultados da análise química de vinhos de três castas de uma determinada região de Itália. Esse ficheiro encontra-se agora também disponível na página *web* da disciplina em formato `txt` de nome `vinhos.txt`. As 14 colunas da tabela de dados correspondem respectivamente às variáveis casta (factor V1 com 3 níveis, que será ignorado neste Exercício), teor alcoólico (V2), teor de ácido málico (V3), cinzas (V4), alcalinidade das cinzas (V5), teor de magnésio (V6), índice de fenóis totais (V7), teor de flavonóides (V8), teor de outros fenóis (V9), teor de proantocianidinas (V10), intensidade de cor (V11), matiz (V12), razão de densidades ópticas em duas frequências, OD280/OD315, (V13) e teor de prolina (V14). Há interesse em modelar o teor de flavonóides (variável `V8`), um antioxidante de medição difícil e dispendiosa.

(a) Numa sessão do R, guarde os dados do ficheiro `vinhos.txt` numa *data frame* de nome `vinhos`. De seguida, exclua da tabela de dados a primeira coluna (um factor que indica a casta) substituindo a *data frame* anterior, através do comando `vinhos<-vinhos[,-1]`.

(b) Execute o comando `plot(vinhos)` e obtenha a matriz de correlações entre as variáveis sob estudo. Comente os resultados.

(c) Ajuste a recta de regressão do teor de flavonóides (`V8`) sobre o teor alcoólico (`V2`). Calcule o coeficiente de determinação e determine o valor das três Somas de Quadrados associadas a esta regressão.

(d) A partir da matriz de correlações entre as variáveis sob estudo, diga qual a melhor recta de regressão simples para prever o teor de flavonóides (variável `V8`). Para a regressão linear simples que escolher, determine o coeficiente de determinação e realize a correspondente decomposição da soma dos quadrados total.

(e) A variável preditora utilizada na alínea anterior também não é simples de medir, tal como sucede com as variáveis `V9` e `V10`. Foi sugerido procurar um modelo de regressão linear múltipla para a variável resposta teor de flavonóides (`V8`) que não utiliza esses preditores. Foi proposto um modelo com cinco variáveis preditoras: `V4`, `V5`, `V11`, `V12` e `V13`. Ajuste este modelo, e comente o respectivo coeficiente de determinação, comparando-o com o R^2 do modelo da alínea anterior. O comando do `R` para ajustar esta regressão linear múltipla é:

```
> lm(V8 ~ V4 + V5 + V11 + V12 + V13 , data=vinhos)
```

(f) Ajuste uma regressão linear múltipla do teor de flavonóides (variável `V8`) sobre todas as restantes variáveis com o comando `summary(lm(V8 ~ . , data=vinhos))`.

i. Use o valor do coeficiente de determinação obtido com esse comando para determinar a decomposição da soma dos quadrados totais. Comente os resultados.

ii. Compare os coeficientes estimados das variáveis preditoras com os correspondentes coeficientes das variáveis preditoras presentes nos modelos anteriores. Comente.

8. Considere novamente o estudo sobre framboesas descrito no exercício 4 dos *Conceitos Introdutórios de Estatística e do Programa R*. Os dados encontram-se no ficheiro `brix.txt`.

(a) Pretende-se modelar o teor de *Brix* a partir das restantes variáveis observadas. Escreva a equação de base do modelo de regressão linear múltipla com *Brix* como variável resposta e as restantes variáveis como preditoras. Quantos parâmetros tem este modelo?

(b) Determine o valor das estimativas dos parâmetros do modelo indicado na alínea anterior,

i. com o comando do `R` para ajustar essa regressão linear múltipla;

ii. obtendo o vector $\vec{\mathbf{b}}$ dos parâmetros ajustados, através da sua fórmula, $\vec{\mathbf{b}} = (\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t\vec{\mathbf{y}})$, onde $\vec{\mathbf{y}}$ é o vector das observações da variável resposta (utilize o comando `model.matrix` do `R` para construir a matriz $\mathbf{X}$ do modelo).

(c) Discuta o significado biológico do coeficiente ajustado da variável *Peso*. Quais são as unidades de medida desta estimativa?

(d) Discuta o significado da ordenada na origem b_0 resultante do ajustamento. Comente.

(e) Execute os procedimentos necessários para calcular o coeficiente de determinação da regressão linear múltipla ajustada e comparar esse coeficiente de determinação com os coeficientes de determinação associados às regressões lineares simples (com a mesma variável resposta). Comente.

(f) É sugerido eliminar o preditor *Peso*. Escreva a equação de base desse submodelo de regressão linear múltipla. Sem fazer quaiquer cálculos, diga qual o intervalo de valores possível para o coeficiente de determinação desse submodelo. O que poderá dizer sobre os valores das estimativas dos parâmetros desse submodelo?

1.2 Exercícios teóricos

1. Demonstre as seguintes relações algébricas:

   (a) $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, para qualquer conjunto de n valores, $\{x_i\}_{i=1}^n$, de média $\overline{x}$.

   (b) $(n-1)\text{cov}_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i = \sum_{i=1}^{n}(y_i - \overline{y})x_i$, para quaisquer n pares de valores, $\{(x_i, y_i)\}_{i=1}^n$, de médias $\overline{x}$ e $\overline{y}$, respectivamente.

   (c) $(n-1)\text{s}^2{}_x = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i - \overline{x})x_i$, para qualquer conjunto de n valores, $\{x_i\}_{i=1}^n$, de média $\overline{x}$.

2. Considere uma regressão linear simples, ajustada com n pares de observações $\{(x_i, y_i)\}_{i=1}^n$. Mostre que:

   (a) A igualdade da média dos valores observados e da média dos valores ajustados de y.

   (b) A média dos resíduos ($e_i = y_i - \hat{y}_i$) é nula.

   (c) As três Somas de Quadrados da regressão são múltiplos de variâncias:

$$\begin{aligned} SQT &= \sum_{i=1}^{n}(y_i - \overline{y})^2 = (n-1)\cdot s_y^2 \\ SQR &= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = (n-1)\cdot s_{\hat{y}}^2 \\ SQRE &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = (n-1)\cdot s_e^2 \,, \end{aligned}$$

   onde $s_\star^2$ indica a variância amostral das quantidades representadas por $\star$.

   (d) $SQR = b_1^2 \cdot (n-1) \cdot s_x^2$, onde $(n-1)\cdot s_x^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2$.

   (e) $SQT = SQR + SQRE$.

   (f) O declive da recta de regressão de y sobre x pode-se escrever em termos do desvio padrão de cada variável e do coeficiente de correlação entre as duas variáveis, sendo dado por:

$$b_1 = r_{xy} \cdot \frac{s_y}{s_x} \,.$$

(g) O coeficiente de determinação R^2 é igual ao quadrado do coeficiente de correlação entre as observações da variável preditora x e da variável resposta y.

(h) O quadrado do coeficiente de correlação entre os n valores observados y_i e os n correspondentes valores ajustados, $\hat{y}_i$, é também igual ao coeficiente de determinação: $(r_{y\hat{y}})^2 = R^2$.

3. Considere uma regressão linear simples ajustada com n pares de observações $\{(x_i, y_i)\}_{i=1}^n$. Considere ainda a notação utilizada nas aulas (em que $\mathbf{X}$ indica uma matriz com duas colunas: uma coluna de n uns, e uma coluna com os n valores x_i da variável preditora X; e $\vec{\mathbf{y}}$ indica um vector com os n valores da variável resposta). Mostre que:

(a) $$\mathbf{X}^t\vec{\mathbf{y}} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = \begin{bmatrix} n\overline{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = \begin{bmatrix} n\overline{y} \\ (n-1)\, cov_{xy} + n\overline{x}\overline{y} \end{bmatrix}.$$

(b) $$\mathbf{X}^t\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\,\overline{x} \\ n\,\overline{x} & (n-1)s_x^2 + n\,\overline{x}^2 \end{bmatrix}.$$

(c) $$(\mathbf{X}^t\mathbf{X})^{-1} = \frac{1}{n(n-1)\cdot s_x^2} \begin{bmatrix} (n-1)\cdot s_x^2 + n\overline{x}^2 & -n\,\overline{x} \\ -n\,\overline{x} & n \end{bmatrix}.$$

(d) Deduza a partir do facto que $\vec{\mathbf{b}} = (\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t\vec{\mathbf{y}})$, as fórmulas para b_0 e b_1 da recta de regressão.

4. (a) Mostre, a partir da sua definição, que a matriz de projecção ortogonal $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ numa regressão linear múltipla é idempotente ($\mathbf{H}\,\mathbf{H} = \mathbf{H}$) e simétrica ($\mathbf{H}^t = \mathbf{H}$).

(b) Mostre que a projecção ortogonal sobre o subespaço das colunas da matriz $\mathbf{X}$, $\mathcal{C}(\mathbf{X})$, de qualquer vector pertencente a esse mesmo espaço ($\mathbf{X}\vec{\mathbf{a}} \in \mathcal{C}(\mathbf{X})$) deixa esse vector invariante.

(c) Mostre, a partir da expressão do vector dos valores ajustados de Y, $\hat{\vec{\mathbf{y}}} = \mathbf{H}\vec{\mathbf{y}}$ que, também numa regressão linear múltipla, a média amostral valores observados de Y, $\{y_i\}_{i=1}^n$, é igual à média amostral dos valores ajustados $\{\hat{y}_i\}_{i=1}^n$.

5. **[Opcional]** Considere uma regressão linear múltipla.

(a) Considere o vector $\vec{\mathbf{1}}_n \in \mathbb{R}^n$, constituido por n uns. Construa a matriz $\mathbf{P} = \vec{\mathbf{1}}_n(\vec{\mathbf{1}}_n^t\,\vec{\mathbf{1}}_n)^{-1}\vec{\mathbf{1}}_n^t$ de projecção ortogonal sobre o subespaço $\mathcal{C}(\vec{\mathbf{1}}_n) \subset \mathbb{R}^n$ gerado pelo vector $\vec{\mathbf{1}}_n$. Mostre que a matriz $\mathbf{P}$ é simétrica e idempotente.

(b) Mostre que se verificam as seguintes igualdades:

$$\begin{aligned} SQT &= \|\vec{\mathbf{y}} - \mathbf{P}\vec{\mathbf{y}}\|^2 = \vec{\mathbf{y}}^t(\mathbf{I}-\mathbf{P})\vec{\mathbf{y}} \\ SQR &= \|\mathbf{H}\vec{\mathbf{y}} - \mathbf{P}\vec{\mathbf{y}}\|^2 = \vec{\mathbf{y}}^t(\mathbf{H}-\mathbf{P})\vec{\mathbf{y}} \\ SQRE &= \|\vec{\mathbf{y}} - \mathbf{H}\vec{\mathbf{y}}\|^2 = \vec{\mathbf{y}}^t(\mathbf{I}-\mathbf{H})\vec{\mathbf{y}} \end{aligned}$$

onde $\vec{\mathbf{y}}$ indica o vector de observações da variável resposta, $\mathbf{H}$ é a matriz de projecção ortogonal sobre o subespaço $\mathcal{C}(\mathbf{X})$ gerado pelas colunas da matriz $\mathbf{X}$ e $\mathbf{P}$ é a matriz de projecção ortogonal sobre o subespaço $\mathcal{C}(\vec{\mathbf{1}}_n)$ gerado pelo vector dos n uns, $\vec{\mathbf{1}}_n$.

(c) Mostre, algebricamente, que $SQT = SQR + SQRE$.