UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

Disciplina: Modelos Lineares Generalizados
Curso: Mestrado em Estatística
Código: PPGEST2217
Semestre: 2025.2
Professor: Frederico Machado Almeida

Observações:

- Questões para entregar: 2, 4, 5(d,e), 7 e 8.
- Demais questões são apenas para estudar.
- Prazo de entrega: 17/09/2025

Q01. Responda as questões abaixo:

(a) Seja A e B duas matrizes, cujo o produto é permutável. Mostre que $A^2 - B^2 = (A - B)\, C$, para alguma matriz C.

(b) Se A e B são matrizes diagonais de dimensão $n \times n$, mostre que $AB = BA$.

(c) Seja $\mathbf{X}$ uma matriz $n \times k$, particionada como $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, no qual $\mathbf{X}_1$ é $n \times k_1$ e $\mathbf{X}_2$ é $n \times k_2$.

(i) Mostre que

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} \mathbf{X}_1^\top \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{X}_2 \end{pmatrix}.$$

Qual é a dimensão de cada submatriz de $\mathbf{X}^\top \mathbf{X}$?

(ii) Seja $\mathbf{b}$ um vetor $k \times 1$ particionado como $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2)^\top$, em que $\mathbf{b}_1$ é $k_1 \times 1$ e $\mathbf{b}_2$ é $k_2 \times 1$. Mostre que

$$\left(\mathbf{X}^\top \mathbf{X}\right) \mathbf{b} = \begin{pmatrix} \mathbf{X}_1^\top \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_1^\top \mathbf{X}_2 \mathbf{b}_2 \\ \mathbf{X}_2^\top \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_2^\top \mathbf{X}_2 \mathbf{b}_2 \end{pmatrix}.$$

Q02. Mostre que $(1/n)\, \mathbf{J}$ é uma matriz idempotente, $\mathbf{I} - (1/n)\, \mathbf{J}$ é igualmente idempotente, e $[\mathbf{I} - (1/n)\, \mathbf{J}]\, (1/n)\, \mathbf{J} = \mathbf{0}$. Sendo $\mathbf{J} = (a_{ij})$, com $a_{ij} = 1$ para todo $i, j = \{1, 2, \cdots, n\}$.
Resposta: para responder essa questão, vamos usar o resultado da Definição 8 (Slides 10 da aula **GLM1-Álgebra Matricial**), tal que, $\mathbf{J_{n\times n} J_{n\times n}} = n\mathbf{J}_{n\times n}$. Assim, ($i$) para mostrar que $(1/n)\, \mathbf{J}$ é uma matriz idempotente, segue que, $[(1/n)\, \mathbf{J}]^2 = n^{-2}\mathbf{J^2} = n^{-2} n \mathbf{J} = n^{-1}\mathbf{J}$. ($ii$) Para mostrar que $\mathbf{I} - (1/n)\, \mathbf{J}$ é igualmente idempotente, segue que, $(\mathbf{I} - (1/n)\, \mathbf{J})^2 = [\mathbf{I} - (1/n)\, \mathbf{J}]\, [\mathbf{I} - (1/n)\, \mathbf{J}] = \mathbf{I}^2 - 2n^{-1}\mathbf{J} + n^{-2}\mathbf{J}^2 = \mathbf{I} - n^{-1}\mathbf{J}$, esse resultado segue da idempotência de $\mathbf{I}$ e de $n^{-1}\mathbf{J}$ mostrado em (i). Por fim, (iii) vamos provar que $[\mathbf{I} - (1/n)\, \mathbf{J}]\, (1/n)\, \mathbf{J} = \mathbf{0}$. Para tal, $[\mathbf{I} - (1/n)\, \mathbf{J}]\, (1/n)\, \mathbf{J} = n^{-1}\mathbf{J} - n^{-2}\mathbf{J}^2 = n^{-1}\mathbf{J} - n^{-1}\mathbf{J} = \mathbf{0}$.

Q03. Considere o modelo de regressão linear simples $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Assuma $e_i = y_i - \hat{y}_i$ uma realização de ϵ_i, e $c_i = \frac{1}{n} - k_i \bar{x}$, com $k_i = (x_i - \bar{x})\,/S_{xx}$. Mostre que:

(a) $\sum\limits_{i=1}^{n} y_i = \sum\limits_{i=1}^{n} \hat{y}_i$.

(b) $\sum\limits_{i=1}^{n} x_i e_i = 0$.

(c) $\sum\limits_{i=1}^{n} \hat{y}_i e_i = 0$.

(d) $\sum\limits_{i=1}^{n} c_i = 1$.

(e) $\sum\limits_{i=1}^{n} x_i c_i = 0$.

Q04. Assuma que a equação de regressão adequada para descrever as varições da resposta y_i é dada por: $y_i = \beta x_i + \epsilon_i$ (reta de regressão passando pela origem). Com $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Sob a condição de que $x_1, x_2, \cdots, x_n$ são quantidades fixas e medidas sem erro, pede-se para:

(a) Mostrar que o estimador dos mínimos quadrados de β é dado por, $\hat{\beta} = \sum\limits_{i=1}^{n} k_i y_i$, com $k_i = \frac{x_i}{\sum\limits_{i=1}^{n} x_i^2}$.

Resposta: observe que, para o modelo de regressão linear apresentado no enunciado, a soma dos quadrados dos erros é dada por: $SQE(\beta) = \sum\limits_{i=1}^{n} \epsilon_i^2 = \sum\limits_{i=1}^{n} (y_i - \beta x_i)^2$. Desta forma, o estimador dos mínimos quadrados será obtido resolvendo a seguinte equação, $\frac{d}{d\beta} SQE(\beta) = -2 \sum\limits_{i=1}^{n} (y_i - \beta x_i) x_i = 0$. Ou seja, a equação normal é dada por: $\sum\limits_{i=1}^{n} \left(y_i - \hat{\beta} x_i\right) x_i = 0 \Longleftrightarrow \hat{\beta} = \frac{\sum\limits_{i=1}^{n} y_i x_i}{\sum\limits_{i=1}^{n} x_i^2} = \sum\limits_{i=1}^{n} k_i y_i$, com $k_i = \frac{x_i}{\sum\limits_{i=1}^{n} x_i^2}$.

(b) Mostre que $\hat{\beta}$ é um estimador não-viesado para β.

Resposta: do item (a) e do enunciado, segue que, $\mathbb{E}(Y_i | x_i) = \beta x_i$. Logo, $\mathbb{E}\left(\hat{\beta}\right) = \mathbb{E}\left[\sum\limits_{i=1}^{n} k_i Y_i\right] = \sum\limits_{i=1}^{n} k_i \mathbb{E}(Y_i) = \beta \sum\limits_{i=1}^{n} k_i x_i = \beta \frac{\sum\limits_{i=1}^{n} x_i^2}{\sum\limits_{i=1}^{n} x_i^2} = \beta$. Portanto, concluímos que $\hat{\beta}$ é um estimador não-viesado para β.

(c) Encontre a variância do estimador obtido em (a), e sua respectiva distribuição amostral.

Resposta: segue da definição do modelo de regressão linear simples que, $Var(Y_i) = Var(\beta x_i + \epsilon_i) = Var(\epsilon_i) = \sigma^2$. Assim,

$$Var\left(\hat{\beta}\right) = Var\left[\sum_{i=1}^{n} k_i Y_i\right] = \sum_{i=1}^{n} k_i^2 Var(Y_i) = \sigma^2 \frac{\sum\limits_{i=1}^{n} x_i^2}{\left[\sum\limits_{i=1}^{n} x_i^2\right]^2} = \frac{\sigma^2}{\sum\limits_{i=1}^{n} x_i^2}.$$

Correção: a distribuição amostral que o enunciado faz referência é com respeito ao estimador obtido em (a). Ou seja, supondo que os erros são normais, segue igualmente que, $\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{\sum\limits_{i=1}^{n} x_i^2}\right)$. No caso da distribuição amostral do estimador de σ^2, segue que, $\frac{(n-1)SQRes}{\sigma^2} \sim \chi^2_{n-1}$, com $SQRes = \sum_{i=1}^{n}\left(y_i - \hat{\beta}x_i\right)^2$, que denota soma dos quadrados dos resíduos.

Q05. Considere o modelo de regressão linear simples $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, com $\mathbb{E}\left(\epsilon_i\right) = 0$, $\text{Var}\left(\epsilon_i\right) = \sigma^2$, e assuma que ϵ_i é não correlacionado. Pede-se para mostrar que:

(a) $\text{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\bar{x}\sigma^2/S_{xx}$.

(b) $\text{Cov}\left(\hat{\beta}_1\left(x_i - \bar{x}\right), y_j\right) = \frac{\sigma^2(x_i-\bar{x})(x_j-\bar{x})}{S_{xx}}$.

(c) $\text{Cov}\left(y_i, \hat{y}_i\right) = \sigma^2 h_{ii}$

(d) $\text{Var}\left(e_i\right) = (1 - h_{ii})\sigma^2$

(e) $\text{Cov}\left(e_i, e_j\right) = -\sigma^2 h_{ij}$, para todo $i \neq j$.

Os itens (d) e (e) indicam que, apesar da suposição de homoscedasticidade e independência dos $\epsilon_i' s$, os resíduos e_i são correlacionados, e sua variância depende dos valores de alavancagem.

Resp: (d) para responder esse item, vamos escrever o resíduo da seguinte forma: $e_i = y_i - \hat{y}_i = y_i - \sum_{j=1}^{n} h_{ij}y_j$. Desta forma, segue imediatamente que:

FORMA I:

$$
\begin{aligned}
Var\left(e_i\right) &= Var\left(Y_i - \sum_{j=1}^{n} h_{ij}Y_j\right) = Var\left(Y_i\right) + Var\left(\sum_{j=1}^{n} h_{ij}Y_j\right) - 2Cov\left(Y_i, \sum_{j=1}^{n} h_{ij}Y_j\right)\\
&= \sigma^2 + \sum_{j=1}^{n} h_{ij}^2 Var\left(Y_j\right) - 2Cov\left(Y_i, \sum_{j=1}^{n} h_{ij}Y_j\right)\\
&= \sigma^2 + \sigma^2\sum_{j=1}^{n} h_{ij}^2 - 2\underbrace{Cov\left(Y_i, h_{ii}Y_i\right)}_{para\,algum\,i=j}\\
&= \sigma^2 + \sigma^2 h_{ii} - 2h_{ii}\sigma^2\\
&= \sigma^2\left(1 - h_{ii}\right).
\end{aligned}
$$

FORMA II: vamos reescrever o i-ésimo resíduo como $e_i = y_i - \hat{y}_i = y_i - \sum_{j=1}^{n} h_{ij}y_j = y_i\left(1 - h_{ii}\right) - \sum_{i\neq j} h_{ij}y_j$. Assim, segue imediatamente que,

$$
\begin{aligned}
Var\left(e_i\right) &= Var\left(Y_i\left(1-h_{ii}\right)-\sum_{i\neq j}h_{ij}Y_j\right)\\
&= Var\left(Y_i\left(1-h_{ii}\right)\right)+Var\left(\sum_{i\neq j}h_{ij}Y_j\right)-2Cov\left(Y_i\left(1-h_{ii}\right),\sum_{i\neq j}^{n}h_{ij}Y_j\right)\\
&= \left(1-h_{ii}\right)^2Var\left(Y_i\right)+\sum_{i\neq j}h_{ij}^2Var\left(Y_j\right)-\underbrace{2Cov\left(Y_i\left(1-h_{ii}\right),\sum_{i\neq j}h_{ij}Y_j\right)}_{=0,\text{porque } i\neq j\ \text{(parte vermelha)}}\\
&= \left(1-2h_{ii}+h_{ii}^2\right)\sigma^2+\sum_{i\neq j}h_{ij}^2\sigma^2=\sigma^2\left(1-2h_{ii}\right)+\sigma^2\left(\underbrace{h_{ii}^2+\sum_{i\neq j}h_{ij}^2}_{=\sum_{j=1}^{n}h_{ij}^2}\right)\\
&= \sigma^2-2\sigma^2h_{ii}+h_{ii}\sigma^2=\sigma^2\left(1-h_{ii}\right).
\end{aligned}
$$

FORMA III: considere a notação usual do i-ésimo resíduo. isto é, $e_i=y_i-\hat{y}_i$. Desta forma, segue que,

$$
\begin{aligned}
Var\left(e_i\right) &= Var\left(Y_i-\hat{Y}_i\right)=Var\left(Y_i\right)+Var\left(\hat{Y}_i\right)-2Cov\left(Y_i,\hat{Y}_i\right)\\
&= \sigma^2+Var\left(\hat{\beta}_0+\hat{\beta}_1x_i\right)-2Cov\left(Y_i,\hat{\beta}_0+\hat{\beta}_1x_i\right)\\
&= \sigma^2+\underbrace{Var\left(\hat{\beta}_0\right)}_{(i)}+\underbrace{Var\left(\hat{\beta}_1x_i\right)}_{(ii)}+\underbrace{2Cov\left(\hat{\beta}_0,\hat{\beta}_1x_i\right)}_{(iii)}-\underbrace{2Cov\left(Y_i,\hat{\beta}_0\right)}_{(iv)}-\underbrace{2Cov\left(Y_i,\hat{\beta}_1x_i\right)}_{(v)}.
\end{aligned}
$$

Observe que, (i) e (ii) sai imediatamente. Calculando as covariâncias em (iii)-(v), e substituindo na expressão anterior obtemos: $Var\left(e_i\right)=\sigma^2\left(1-h_{ii}\right)$.

Resposta: (e) segue da definição do resíduo que, $e_i=y_i-\hat{y}_i=y_i-\sum_{j=1}^{n}h_{ij}y_j$, e $e_j=y_j-\hat{y}_j=y_j-\sum_{j=1}^{n}h_{ij}y_j$, com a suposição de que $h_{ij}=h_{ji}$. Desta forma,

FORMA I:

$$
\begin{aligned}
Cov\left(e_i,e_j\right) &= Cov\left(Y_i-\sum_{j=1}^{n}h_{ij}Y_j, Y_j-\sum_{j=1}^{n}h_{ij}Y_j\right)\\
&= Cov\left(Y_i,Y_j\right)-Cov\left(Y_i,\sum_{j=1}^{n}h_{ij}Y_j\right)-Cov\left(\sum_{j=1}^{n}h_{ij}Y_j,Y_j\right)+Cov\left(\sum_{j=1}^{n}h_{ij}Y_j,\sum_{j=1}^{n}h_{ij}Y_j\right)\\
&= \underbrace{Cov\left(Y_i,Y_j\right)}_{=0}-\underbrace{Cov\left(Y_i,h_{ii}Y_i\right)}_{para\,algum\,i=j}-\underbrace{Cov\left(h_{ij}Y_j,Y_j\right)}_{para\,algum\,j\,fixo}+Var\left(\sum_{j=1}^{n}h_{ij}Y_j\right)\\
&= 0-h_{ii}Cov\left(Y_i,Y_i\right)-h_{ij}Cov\left(Y_j,Y_j\right)+\sum_{j=1}^{n}h_{ij}^2Var\left(Y_j\right)\\
&= -h_{ii}Var\left(Y_i\right)-h_{ij}Var\left(Y_j\right)+\sigma^2\sum_{j=1}^{n}h_{ij}^2=-h_{ii}\sigma^2-h_{ij}\sigma^2+\sigma^2\underbrace{\sum_{j=1}^{n}h_{ij}^2}_{=h_{ii}}\\
&= -h_{ii}\sigma^2-h_{ij}\sigma^2+\sigma^2h_{ii}\\
&= -h_{ij}\sigma^2.
\end{aligned}
$$

FORMA II: segue da definição usual do resíduo que, $e_i=y_i-\hat{y}_i$ para o i-ésimo resíduo, e $e_j=y_j-\hat{y}_j$ para o j-ésimo resíduo. Segue então que,

$$
\begin{aligned}
Cov\left(e_i,e_j\right) &= Cov\left(Y_i-\hat{Y}_i,Y_j-\hat{Y}_j\right)\\
&= Cov\left(Y_i,Y_j\right)-Cov\left(Y_i,\hat{Y}_j\right)-Cov\left(\hat{Y}_i,Y_j\right)+Cov\left(\hat{Y}_i,\hat{Y}_j\right)\\
&= 0-Cov\left(Y_i,\hat{\beta}_0+\hat{\beta}_1x_j\right)-Cov\left(\hat{\beta}_0+\hat{\beta}_1x_i,Y_j\right)+Cov\left(\hat{\beta}_0+\hat{\beta}_1x_i,\hat{\beta}_0+\hat{\beta}_1x_j\right)\\
&= -\underbrace{Cov\left(Y_i,\hat{\beta}_0\right)}_{(i)}-\underbrace{Cov\left(Y_i,\hat{\beta}_1x_j\right)}_{(ii)}-\underbrace{Cov\left(\hat{\beta}_0,Y_j\right)}_{(iii)}-\underbrace{Cov\left(\hat{\beta}_1x_i,Y_j\right)}_{(iv)}+Var\left(\hat{\beta}_0\right)\\
&+ \underbrace{Cov\left(\hat{\beta}_0,\hat{\beta}_1x_j\right)}_{(v)}+\underbrace{Cov\left(\hat{\beta}_1x_i,\hat{\beta}_0\right)}_{(vi)}+\underbrace{Cov\left(\hat{\beta}_1x_i,\hat{\beta}_1x_j\right)}_{(vii)}
\end{aligned}
$$

Obs.: algumas das covariâncias apresentadas nos itens (i) a (vii) são obtidas facilmente, outras podem ser obtidas trocando os termos i e j. Por fim, juntando os resultados de (i) a (vii) obtemos: $Cov\left(e_i,e_j\right)=-h_{ij}\sigma^2$.

Q06. Sob as condições do modelo de regressão linear simples apresentado na **Q04**,

(a) Obtenha os estimadores de máxima verossimilhança para β_0, β_1 e σ^2.

(b) Verifique se os estimadores encontrados no em (a) são centrados nos verdadeiros valores.

(c) Do item (b), encontre a quantidade de viés presente no(s) estimador(es) que julgar ser(em) não-centrado(s). O que acontece com o viés quando o tamanho de amostra aumenta?

Q07. Considere o modelo de regressão linear simples $y_i=\beta_0+\beta_1x_i+\epsilon_i$, com $\epsilon_i\sim \mathrm{N}\left(0,\sigma^2\right)$, seja SQT, SQReg e SQE as somas dos quadrados dos totais, da regressão e dos erros, respectivamente. Mostre que:

(a) $\mathbb{E}\left(\text{SQReg}\right) = \sigma^2 + \beta_1^2 S_{xx}$.

(b) $\mathbb{E}\left(\text{SQE}\right) = (n-2)\sigma^2$.

(c) $\mathbb{E}\left(\text{SQT}\right) = \sigma^2(n-1) + \beta_1^2 S_{xx}$.

Resposta: (a) para mostrar as suposições abaixo, vamos fazer algumas simplificações para as somas dos quadrados. Como é de conhecimento, $SQT = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$, e

$$\begin{aligned} SQReg &= \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}\right)^2 \\ &= \sum_{i=1}^{n}\left(\bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1 x_i - \bar{y}\right)^2 \\ &= \hat{\beta}_1^2\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \\ &= \hat{\beta}_1^2 S_{xx}. \end{aligned}$$

Assim, segue imediatamente que,

$$\begin{aligned} \mathbb{E}\left(SQReg\right) &= \mathbb{E}\left(\hat{\beta}_1^2 S_{xx}\right) = S_{xx}\mathbb{E}\left(\hat{\beta}_1^2\right) = S_{xx}\left[Var\left(\hat{\beta}_1\right) + \mathbb{E}^2\left(\hat{\beta}_1\right)\right] \\ &= S_{xx}\left[\frac{\sigma^2}{S_{xx}} + \beta_1^2\right] = \sigma^2 + \beta_1^2 S_{xx}. \end{aligned}$$

(b) uma forma fácil de provar a suposição do item (b), segue do fato de que $SQE = SQT - SQReg$. Consequentemente, $\mathbb{E}\left(SQE\right) = \mathbb{E}\left(SQT\right) - \mathbb{E}\left(SQReg\right)$. Desta forma, seja $\mu_{y_i/x_i} = \beta_0 + \beta_1 x_i$ o valor esperado de y_i, dado x_i. Assim, $\mu_{\bar{y}|\bar{x}} = \beta_0 + \beta_1\bar{x}$. Consequentemente, segue imediatamente que,

$$\begin{aligned} \mathbb{E}\left(SQT\right) &= \sum_{i=1}^{n}\mathbb{E}\left(Y_i^2\right) - n\mathbb{E}\left(\bar{Y}^2\right) = \sum_{i=1}^{n}\left[Var\left(Y_i\right) + \mathbb{E}^2\left(Y_i\right)\right] - n\left[Var\left(\bar{Y}\right) + \mathbb{E}^2\left(\bar{Y}\right)\right] \\ &= \sum_{i=1}^{n}\left[\sigma^2 + \mu_{y_i/x_i}^2\right] - n\left[\frac{\sigma^2}{n} + \mu_{\bar{y}|\bar{x}}^2\right] = \sigma^2\left(n-1\right) + \sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i\right)^2 - n\left(\beta_0 + \beta_1\bar{x}\right)^2 \\ &= \sigma^2\left(n-1\right) + n\beta_0^2 + n\beta_0\beta_1\bar{x} + \sum_{i=1}^{n}\beta_1^2 x_i^2 - n\beta_0^2 - n\beta_0\beta_1\bar{x} - n\beta_1^2\bar{x}^2 \\ &= \sigma^2\left(n-1\right) + \beta_1^2\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] \\ &= \sigma^2\left(n-1\right) + \beta_1^2 S_{xx}. \end{aligned}$$

Por fim, temos que: $\mathbb{E}\left(SQE\right) = \sigma^2\left(n-1\right) + \beta_1^2 S_{xx} - \left[\sigma^2 + \beta_1^2 S_{xx}\right] = \sigma^2\left(n-2\right)$.

Q08. Suponha que estamos interessados em ajustar o modelo de regressão linear simples $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ porém decidiu-se reescalonar a variável regressora tomando-a como o desvio com relação a sua média, isto é, $x_i^* = x_i - \bar{x}$, com $i = 1, ..., n$.

(a) Prove que o modelo que relaciona y_i e a nova variável regressora x_i^* é $y_i = \beta_0^* + \beta_1 x_i^* + \epsilon_i$, onde $\beta_0^* = \beta_0 + \beta_1\bar{x}$.

Resposta: segue da definição do modelo de regressão linear simples que, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Como $x_i^* = x_i - \bar{x}$, segue que, $x_i = x_i^* + \bar{x}$. Logo, substituindo x_i na expressão anterior obtemos:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i = \beta_0 + \beta_1 \left(x_i^* + \bar{x}\right) + \epsilon_i \\ &= \left(\beta_0 + \beta_1 \bar{x}\right) + \beta_1 x_i^* + \epsilon_i = \beta_0^* + \beta_1 x_i^* + \epsilon_i \end{aligned}$$

(b) Prove que o estimador dos mínimos quadrados para β_0^* é igual a $\bar{y}$, e que o estimador para β_1 permanece igual ao do modelo original.

Resposta: segue do modelo em (a) $y_i = \beta_0^* + \beta_1 x_i^* + \epsilon_i$ que, $\epsilon_i = y_i - \beta_0^* + \beta_1 x_i^*$. A soma do quadrado dos erros, SQE para o modelo anterior é, $SQE = \sum_{i=1}^{n} \left(y_i - \beta_0^* + \beta_1 x_i^*\right)^2$. Derivando em relação a β_0^* e β_1 obtemos:

$$\begin{cases} \frac{dSQE}{d\beta_0^*} = -2 \sum_{i=1}^{n} \left(y_i - \beta_0^* - \beta_1 x_i^*\right) = 0 \\ \frac{dSQE}{d\beta_1} = -2 \sum_{i=1}^{n} x_i^* \left(y_i - \beta_0^* - \beta_1 x_i^*\right) = 0 \end{cases} \Longrightarrow \begin{cases} \sum_{i=1}^{n} y_i - n\beta_0^* - \beta_1 \sum_{i=1}^{n} x_i^* = 0 \\ \sum_{i=1}^{n} x_i^* y_i - \beta_0^* \sum_{i=1}^{n} \left(x_i - \bar{x}\right) - \beta_1 \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = 0 \end{cases}$$

segue da suposição de que $\sum_{i=1}^{n} \left(x_i - \bar{x}\right) = 0$. Assim, chegamos a conclusão que $\hat{\beta}_0^* = \bar{y}$. Por fim, substituindo substituindo o resultado de (1) em (2) obtemos:

$$\begin{cases} \sum_{i=1}^{n} y_i - n\beta_0^* - \beta_1 \sum_{i=1}^{n} \left(x_i - \bar{x}\right) = 0 \quad (1) \\ \sum_{i=1}^{n} \left(x_i - \bar{x}\right) y_i - \beta_0^* \sum_{i=1}^{n} \left(x_i - \bar{x}\right) - \beta_1 \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = 0 \quad (2) \end{cases} \Longrightarrow \begin{cases} \hat{\beta}_0^* = \bar{y} \\ \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \end{cases}$$

(c) Prove que ambos os modelos fornecem os mesmos valores preditos para y.

Resposta: com base os estimadores dos mínimos quadrados obtidos anteriormente, associado as suposições anteriores temos que, $\hat{\beta}_0^* = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$. Consequentemente,

$$\hat{y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1 x_i^* = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \left(x_i - \bar{x}\right) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x} = \hat{\beta}_0 + \hat{\beta}_1 x_i \equiv \hat{y}_i.$$

Q09. Faça uma simulação, gerando 50 valores para o modelo de regressão linear simples (MRLS), $y_i = -1 + 0.3x_i + \epsilon_i$, com $\epsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, 1\right)$. Assuma que $X \sim U[-1, 1]$. Ajuste o MRLS, e obtenha o diagrama de dispersão com a reta ajustada.