Primeira avaliação (A1)

Disciplina: Modelagem Estatística
Instrutor: Luiz Max Carvalho
Monitor: Isaque Pim

15 de Abril de 2023

- O tempo para realização da prova é de 3 horas;
- Leia a prova toda com calma antes de começar a responder;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo ou figura geométrica de sua preferência;
- A prova vale 100 pontos. A pontuação restante é contada como bônus;
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;
- Você tem direito a trazer **uma folha de "cola"** tamanho A4 frente e verso, que deverá ser entregue junto com as respostas da prova.

1. Todo dia ela faz tudo sempre igual.

Palmirinha se levanta todos dias às seis horas da manhã para ir ao estúdio gravar seus programas de culinária. Preocupada com o trânsito, ela resolve documentar quanto tempo (em minutos) ela leva para chegar ao trabalho (y) dependendo da hora em que sai de casa, x, medida em 'minutos depois das 6h'. Os dados obtidos foram

x	0	10	20	30	40	50	60
y	16	27	28	39	39	48	51

Muito esperta, Palmirinha logo ajusta uma regressão linear por mínimos quadrados, obtendo o seguinte gráfico:

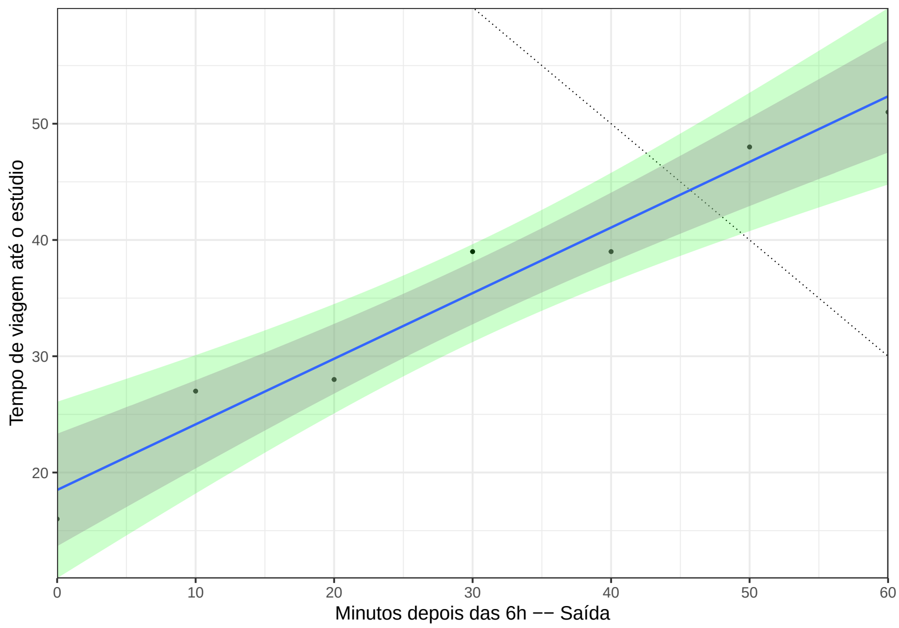

Figura 1: Tempo de viagem até o estúdio de acordo com a hora da saída de casa.

a) (10 pontos) O neto de Palmirinha, Vítor Pereira, muito espevitado, olha os dados e exclama: "Vovó, nem precisa estimar o intercepto! É só olhar os dados! O intercepto é 16.". Vítor está errado, é claro. Mas porque será que ele pensa assim? Qual a estimativa correta do intercepto?

b) (10 pontos) Palmirinha já foi avisada pela chefia de que não deve chegar depois das 7:30h. Mostre como ela pode determinar a hora máxima que pode sair de casa para que não chegue atrasada, em média.

c) (10 pontos) Nossa heroína se preocupa que fazer seus cálculos tendo em vista a média do tempo de chegada e não o tempo de chegada em si pode criar problemas. Mostre à Palmirinha como determinar a hora de sair de casa para que ela chegue antes do horário limite com 99% de probabilidade.

d) (20 pontos) Sendo uma bayesiana *old school*, Palmirinha decide fazer um ajuste bayesiano do modelo linear aos dados apresentados. Como gosta de viver perigosamente, ela decide pela seguinte especificação de modelo e prioris:

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i, \\
\varepsilon_i &\sim \text{Normal}(0, \sigma^2), \\
\pi_B(\beta_0, \beta_1) &\propto 1, \\
\pi_S(\sigma^2) &\propto \frac{1}{\sigma^2},
\end{aligned}$$

isto é, Palmirinha escolheu prioris **impróprias** para as quantidades desconhecidas do modelo. Calcule as esperanças *a posteriori* $E_p[\beta_0 \mid \boldsymbol{y}; \boldsymbol{x}]$ e $E_p[\beta_1 \mid \boldsymbol{y}; \boldsymbol{x}]$ e discuta sua relação com os estimadores de mínimos quadrados obtidos anteriormente.

Conceitos trabalhados: regressão linear; estimativas pontuais e intervalares; predição; análise bayesiana; prioris impróprias. **Nível de dificuldade**: fácil.

Resolução: O intercepto é o ponto em que a reta $\alpha + \beta x$ *intercepta* o eixo das ordenadas. Acontece que esta reta modela $E[Y]$ e não Y; desta forma, não podemos inferir o intercepto *apenas* a partir da observação de um certo par $(0, y)$. Vamos lembrar de algumas fórmulas úteis:

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},
\end{aligned}$$

onde $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$ e $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$. Utilizando estas fórmulas chegamos a $\hat{\beta}_0 = 18.5$ e $\hat{\beta}_1 = 0.56$. Para responder à questão b), notamos que queremos $x^\star$ tal que

$$\begin{aligned}
x^\star + \hat{y}(x^\star) &< 90, \\
x^\star + \hat{\beta}_0 + \hat{\beta}_1 x^\star &< 90, \\
x^\star &< \frac{90 - \hat{\beta}_0}{1 + \hat{\beta}_1},
\end{aligned}$$

que dá $x^\star = 45.71$. Isso significa que se Palmirinha sair até as 6h46, não deve se atrasar, <u>em média</u>. Esta inferência, conquanto útil, ainda sofre da falta de uma garantia probabilística que ofereça algum controle à Palmirinha. No item c) portanto, ela busca $x^{\star\star}$ tal que

$$\Pr\left(\tilde{Y}(x^{\star\star}) \leq 90 - x^{\star\star}\right) > 0.99.$$

Sabendo que $\tilde{Y}(x^{\star\star}) \sim \text{Normal}\left(\hat{\beta}_0 + \hat{\beta}_1 x^{\star\star}, \sigma^2\right)$ e substituindo σ^2 pelo seu estimador não-viesado ($\hat{\sigma^2} = 2.76$), podemos começar o nosso problema de otimização notando que queremos encontrar $x^{\star\star}$ tal que

$$\Phi\left(\frac{90 - \hat{\beta}_0 + (\hat{\beta}_1 - 1)x^{\star\star}}{\sqrt{\hat{\sigma^2}}}\right) > 0.99,$$

onde Φ é f.d.a. de uma normal padrão. Resolvendo o problema numericamente, encontramos $x^{\star\star} < 41.6$ minutos. Poderíamos obter uma resposta aproximada calculando

$$\frac{90 - \hat{\beta}_0 - 3\sqrt{\hat{\sigma^2}}}{1 + \hat{\beta}_1} = 40.41,$$

que certamente atende ao que foi pedido. Poderíamos também aproveitar o gráfico fornecido por Palmirinha para fazer algumas dessas inferências. Do gráfico é possível deduzir que o intercepto está entre 18 e 19; que o ponto em que a parte superior da banda de predição verde (99%) intercepta a linha pontilhada (que é a reta $y = -x + 90$). Agora, vamos responder ao item d)[1]. Para começar, vamos definir $\mu_{x_i}(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_i$ e escrever a verossimilhança:

$$f_{\boldsymbol{x}}(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu_{x_i}(\boldsymbol{\beta}))^2\right).$$

Para facilitar a notação, vamos definir $S_X := \sum_{i=1}^n x_i$, $S_{X,Y} := \sum_{i=1}^n x_i y_i$, $Q_X := \sum_{i=1}^n x_i^2$, $Q_Y := \sum_{i=1}^n y_i^2$, o que nos permite expandir a verossimilhança e escrever

$$\begin{aligned} f_{\boldsymbol{x}}(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^n \times \\ \exp\left(-\frac{1}{2\sigma^2}\left\{Q_Y - 2S_Y\beta_0 - 2S_{X,Y}\beta_1 + n\beta_0^2 + 2S_X\beta_0\beta_1 + Q_X\beta_1^2\right\}\right). \end{aligned}$$

Multiplicando pela priori $\pi(\beta_0, \beta_1, \sigma_2) \propto [\sigma^2]^{-1}$, e coletando todos os termos que dependem de cada parâmetro, conseguimos deduzir que a posteriori é da forma

$$\begin{aligned} p(\beta_0, \beta_1, \sigma^2 \mid \boldsymbol{y}) &\propto, \\ &\underbrace{\propto \exp\left(\frac{2(S_Y\beta_0 + S_{X,Y}\beta_1) - n\beta_0^2 - 2S_X\beta_0\beta_1 - Q_X\beta_1^2}{2\sigma^2}\right)}_{\text{Normal}_2(\boldsymbol{\mu}_B, \sigma^2\boldsymbol{A})} \times \\ &\underbrace{(\sigma^2)^{-n/2-1}\exp\left(-\frac{Q_Y}{2\sigma^2}\right)}_{\text{Gama-Inversa}(\frac{n}{2}, \frac{Q_Y}{2})}, \\ &= p_{\boldsymbol{x}}(\beta_0, \beta_1 \mid \boldsymbol{y}, \sigma^2) p_S(\sigma^2 \mid \boldsymbol{y}). \end{aligned} \tag{1}$$

Para deduzir a forma de $p_{\boldsymbol{x}}(\beta_0, \beta_1 \mid \boldsymbol{y}, \sigma^2)$, considere que a f.d.p. de uma normal bivariada com vetor de médias $\boldsymbol{m} = (\mu_X, \mu_Y)$ e matriz de variância-covariância

$$\boldsymbol{S} = \begin{bmatrix} \sigma_X^2 & \rho\sqrt{\sigma_X^2\sigma_Y^2} \\ \rho\sqrt{\sigma_X^2\sigma_Y^2} & \sigma_Y^2 \end{bmatrix},$$

[1]Vamos adaptar a resposta à questão 3 da A1/2021 de Inferência Estatística: `https://github.com/maxbiostat/Statistical_Inference_BSc/blob/master/provas/PDF/A12021_solucoes.pdf`

é

$$f_{XY}(x, y \mid \boldsymbol{m}, \boldsymbol{S}) \propto \exp\left(-\frac{1}{2(1-\rho)^2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sqrt{\sigma_X^2\sigma_Y^2}}\right]\right).$$

Igualando termos e rearranjando, chegamos a

$$\mu_0 = \frac{S_Y}{n} - \mu_1\frac{S_X}{n},$$
$$\mu_1 = \frac{S_{X,Y}}{Q_X + \left(\frac{S_X}{n}\right)^2[1-2S_X]},$$
$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{n} + \frac{(S_X)^2}{n^2Q_X} & -\frac{S_X}{nQ_X} \\ -\frac{S_X}{nQ_X} & \frac{1}{Q_X} \end{bmatrix}.$$

Usando as propriedades da normal multivariada[2], sabemos que $E_p[\beta_0 \mid \boldsymbol{y}, \sigma^2] = E_p[\beta_0 \mid \boldsymbol{y}] = \mu_0$ e $E_p[\beta_0 \mid \boldsymbol{y}, \sigma^2] = E_p[\beta_1 \mid \boldsymbol{y}] = \mu_1$, chegando à brilhante conclusão de que a média *a posteriori* dos coeficientes coincide com os estimadores de máxima verossimilhança no caso de prioris impróprias, como visto em aula. ■

Comentário A questão a) pode parecer boba, mas ela fala sobre uma distinção fundamental em ciência: inferências *cruas*, baseadas nos dados apenas e inferências baseadas em modelos (*model-based*); o modelo de regressão pega informação emprestada de todos os pontos na amostra para estimar o que acontece para $x = 0$. Nas questões b) e c), vimos como utilizar o modelo ajustado para fazer inferências sobre quantidades de interesse, como por exemplo o valor do preditor (x) que leva a algum desfecho (*outcome*) de interesse. Por fim, fizemos uma análise algebricamente diferente do caso bayesiano que foi feito em aula, especificamente para o caso da regressão linear simples (um preditor). Note que a priori imprópria que utilizamos aqui pode ser vista como um limite particular da Normal-Inversa-Gamma, que é própria. Interessante, não?

2. Facts and logi(sti)c

Um dos tipos mais comuns de dados são os dados binários, da forma $Y_i \in \{0, 1\}$. Suponha que $\boldsymbol{X}$ é uma matriz $n \times P$ de covariáveis e $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ são observações binárias da variável dependente. Para este tipo de dado, em geral estamos interessados em modelar a probabilidade $p_i := \Pr(Y_i = 1; \boldsymbol{X}_i)$. Considere o modelo

$$Y_i \sim \text{Bernoulli}(p_i),$$
$$p_i := \Pr(Y_i = 1; \boldsymbol{X}_i) = \frac{\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right)}.$$

a) (20 pontos) Mostre que a *log chance* (em inglês, *log-odds*) vale

$$\log\left(\frac{\Pr(Y_i = 1; \boldsymbol{X}_i)}{\Pr(Y_i = 0; \boldsymbol{X}_i)}\right) = \boldsymbol{X}_i^T\boldsymbol{\beta}.$$

[2] Em particular que a média da i-ésima coordenada é μ_i e que a média não depende da (matriz de) covariância.

b) (20 pontos) Gostaríamos de obter estimativas $\hat{\boldsymbol{\beta}}$ para os coeficientes. Para tanto, podemos tentar maximizar a (log) verossimilhança dos dados observados sob o modelo em questão. No desenrolar dos cálculos no entanto, veremos que a equação

$$\nabla l(\boldsymbol{\beta}) = 0$$

não tem solução em forma fechada. Para encontrar a raiz – isto é, o EMV $\hat{\boldsymbol{\beta}}$ – precisamos de um método aproximado. O Método de Newton-Raphson envolve fazer updates da forma

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left[\boldsymbol{H}(\boldsymbol{\beta}^{(t)})\right]^{-1} \nabla l(\boldsymbol{\beta}^{(t)}),$$

para $t = 1, 2, \ldots$ a partir de um chute inicial $\boldsymbol{\beta}^{(0)}$.

Um critério de parada comum é

$$\frac{||\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}||}{||\boldsymbol{\beta}^{(t)}||} \leq \varepsilon.$$

(a) Calcule a log-verossimilhança em função de $\boldsymbol{\beta}$, bem como seu gradiente ($\nabla l(\boldsymbol{\beta}^{(t)})$) e matriz hessiana ($\boldsymbol{H}(\boldsymbol{\beta}^{(t)})$);

Dica: Se quiser, pode considerar apenas um preditor mais o intercepto, isto é, $\boldsymbol{\beta} = (\beta_0, \beta_1)$.

(b) Argumente que a log-verossimilhança é estritamente côncava e que, portanto, o método de Newton-Raphson converge para um máximo global;

c) (20 pontos) Suponha que queremos estudar a probabilidade de ter um ataque do coração: $Y_i = 1$ se o i-ésimo indivíduo sofreu um ataque do coração e $Y_i = 0$ caso contrário. Suponha ainda que X_j é uma covariável binária que codifica se o indivíduo torce para o Clube de Regatas do Flamengo (CRF): $X_{ij} = 1$ se o i-ésimo indivíduo torce para o CRF e $X_{ij} = 0$ caso contrário. Mostre como obter um <u>intervalo de confiança</u> aproximado de 68% para a razão de chance (*odds ratio*), OR_j:

$$\mathrm{OR}_j = \frac{\Pr(Y_i = 1; X_{ij} = 1)}{\Pr(Y_i = 0; X_{ij} = 1)} \Bigg/ \frac{\Pr(Y_i = 1; X_{ij} = 0)}{\Pr(Y_i = 0; X_{ij} = 0)}.$$

Dica: Expresse OR_j em função dos coeficientes e assuma que é possível obter uma estimativa do erro-padrão se_j para cada coeficiente.

Conceitos trabalhados: regressão logística; modelos lineares generalizados; estimação. **Nível de dificuldade**: médio.

Resolução: Para facilitar a notação, vamos escrever $p_i = \Pr(Y_i = 1; X_i)$. Notando que $\Pr(Y_i = 0; X_i) = 1 - p_i = \frac{1}{1+\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})}$, temos

$$\begin{aligned}
\log\left(\frac{\Pr(Y_i = 1; \boldsymbol{X}_i)}{\Pr(Y_i = 0; \boldsymbol{X}_i)}\right) &= \log\left(\frac{p_i}{1-p_i}\right), \\
&= \log\left(\left[1 + \exp(\boldsymbol{X}_i^T\boldsymbol{\beta})\right] \frac{\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}_i^T\boldsymbol{\beta})}\right), \\
&= \boldsymbol{X}_i^T\boldsymbol{\beta}.
\end{aligned}$$

Agora vamos fazer os subitens (a) e (b) do item b). A verossimilhança pode ser escrita como

$$f_{\boldsymbol{X}}\left(\boldsymbol{\beta};\boldsymbol{Y}\right)=\prod_{i=1}^{n}p_i(\boldsymbol{\beta})^{Y_i}\left(1-p_i(\boldsymbol{\beta})\right)^{1-Y_i},$$

porque assumimos que os Y_i são condicionalmente independentes uma vez observada a matriz de desenho $\boldsymbol{X}$. Tomando o logaritmo, temos

$$\begin{aligned} l(\beta) &:= \log\left(f_{\boldsymbol{X}}\left(\boldsymbol{\beta}\right);\boldsymbol{Y}\right)=\sum_{i=1}^{n}\log\left(p_i(\boldsymbol{\beta})^{Y_i}\left(1-p_i(\boldsymbol{\beta})\right)^{1-Y_i}\right), \\ &=\sum_{i=1}^{n}Y_i\log\left(\frac{p_i(\boldsymbol{\beta})}{1-p_i(\boldsymbol{\beta})}\right)=\sum_{i=1}^{n}Y_i\cdot g\left(p_i(\boldsymbol{\beta})\right). \end{aligned}$$

Usando a regra da cadeia, obtemos

$$\begin{aligned} \frac{\partial l}{\partial\boldsymbol{\beta}} &= \sum_{i=1}^{n}Y_ip_i'(\boldsymbol{\beta})g'(p_i(\boldsymbol{\beta})), \\ \frac{\partial^2 l}{\partial\boldsymbol{\beta}^2}=\boldsymbol{H}\left(\boldsymbol{\beta}\right) &= \sum_{i=1}^{n}Y_i\left\{\left[p_i'(\boldsymbol{\beta})\right]^2g''\left(p_i(\boldsymbol{\beta})\right)+p_i''(\boldsymbol{\beta})g'(p_i(\boldsymbol{\beta}))\right\}. \end{aligned}$$

Podemos então escrever

$$\begin{aligned} \frac{\partial l}{\partial\beta_j} &= \sum_{i=1}^{n}Y_ip_i'(\beta_j)g'(p_i(\beta_j)), \\ \frac{\partial^2 l}{\beta_j\beta_k} &= -\sum_{i=1}^{n}\left\{X_{ij}X_{ik}p_i(\boldsymbol{\beta})\left(1-p_i(\boldsymbol{\beta})\right)\right\}. \end{aligned}$$

A matriz hessiana é diagonal, e suas entradas diagonais são todas negativas. Sendo a hessiana negativa-definida, a função em questão é côncava, e Newton-Raphson deve convergir para um máximo global. Para responder c), vamos primeiro calcular

$$\begin{aligned} \mathrm{OR}_j &= \frac{p_i(1)}{1-p_i(1)}\cdot\frac{1-p_i(0)}{p_i(0)}, \\ &= \exp\left(\sum_{k\neq j}^{P}\beta_kX_{ik}+\beta_j\cdot 1\right)\exp\left(-\sum_{k\neq j}^{P}\beta_kX_{ik}+\beta_j\cdot 0\right), \\ &= \exp\left(\beta_j\right). \end{aligned}$$

Para construir um intervalo de confiança para β_j, podemos fazer $\hat{\beta}_j\pm Z\mathrm{se}_j$. Fazendo $Z=1$, temos um intervalo aproximado de 68%. Para completar a questão, só é preciso notar que $\exp(\cdot)$ é uma transformação bijetiva, então podemos simplesmente mapear o intervalo de confiança para β_j em

$$\mathrm{OR}_j\in\left(\exp\left(\hat{\beta}_j-\mathrm{se}_j\right),\exp\left(\hat{\beta}_j+\mathrm{se}_j\right)\right). \tag{2}$$

■

Comentário: Nesta questão obtivemos vários resultados importantes a respeito da regressão logística, que é o GLM binomial com link *logit* (canônico).

Obtivemos as quantidades necessárias para fazer estimação por máxima verossimilhança usando o método de Newton-Raphson – e outros métodos baseados em gradientes. Por fim, vimos construir intervalos de confiança aproximados para uma quantidade de interesse científico direto, que é a razão de chances, OR_j. Para determinar o efeito de uma covariável, podemos inspecionar o intervalo de confiança para OR_j e verificar se ele inclui $a \in \mathbb{R}$. Quanto deve valer a? O que acontece se a é menor que a cota inferior do intervalo? E se for maior que a cota superior? Pense sobre isso.

3. Found in translation

Tome $\boldsymbol{Y}$ um vetor $n \times 1$ de variáveis dependentes com $Y_i \in \mathbb{R}$ e $\boldsymbol{X}$ uma matriz ~~$n \times P$ de covariáveis.~~ $n \times (P+1)$ de P covariáveis mais uma coluna de uns $X_0 = (1, 1, \ldots, 1)^T$ representando o intercepto. Considere a seguinte transformação das covariáveis:

$$Z_j = c_{j0} + \sum_{k=1} c_{jk} X_k,$$

formando a matriz $\boldsymbol{Z}$, que é $n \times (P+1)$.

Agora suponha que vamos ajustar dois modelos de regressão linear múltipla:

I. $\boldsymbol{Y} = \boldsymbol{X}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, onde $\boldsymbol{\varepsilon}$ é um vetor $n \times 1$ de erros independentes com média 0 e variância σ^2.

II. $\boldsymbol{Y} = \boldsymbol{Z}^T \boldsymbol{\alpha} + \tilde{\boldsymbol{\varepsilon}}$, onde $\tilde{\boldsymbol{\varepsilon}}$ é um vetor $n \times 1$ de erros independentes com média 0 e variância τ^2.

a) (10 pontos) Mostre que $\boldsymbol{Z}$ pode ser escrita como

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{t},$$

onde $\boldsymbol{t}$ é uma matriz $(P+1) \times (P+1)$.

b) (10 pontos) Mostre que a *hat matrix* nos dois modelos é idêntica se $\boldsymbol{t}$ for inversível.

c) (20 pontos) Mostre que se $\boldsymbol{t}$ for inversível,

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{t}^{-1} \widehat{\boldsymbol{\beta}},$$

onde $\hat{\boldsymbol{\beta}}$ é o estimador de mínimos quadrados de $\boldsymbol{\beta}$. Comente sobre o que esse resultado implica sobre a interpretação dos coeficientes sob transformações lineares da matriz de desenho.

Conceitos trabalhados: transformações lineares; matriz chapéu; estimador de mínimos quadrados.
Nível de dificuldade: fácil.
Resolução: Para o item a), note que cada coluna de $\boldsymbol{Z}$ é uma combinação linear das colunas de $\boldsymbol{X}$. É fácil ver que, definindo o vetor $c_j = (c_{j,0}, c_{j,1}, \ldots, c_{j,P+1})^T$, temos

$$Z_j = \boldsymbol{X} c_j,$$

e portanto

$$Z = \boldsymbol{X} \begin{bmatrix} | & | & | & | \\ c_1 & c_2 & \dots & c_{P+1} \\ | & | & | & | \end{bmatrix} = \boldsymbol{X}\boldsymbol{t}.$$

Basta então tomar as colunas de $\boldsymbol{t}$ como sendo os vetores $\boldsymbol{c}_j$. Para o item b) podemos escrever a matriz chapéu do modelo II como

$$H_{II} = \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T.$$

Substituindo o resultado obtido no item a) obtemos

$$\begin{aligned} H_{II} &= \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T, \\ &= (\boldsymbol{X}\boldsymbol{t})\left((\boldsymbol{X}\boldsymbol{t})^T(\boldsymbol{X}\boldsymbol{t})\right)^{-1}(\boldsymbol{X}\boldsymbol{t})^T, \\ &= \boldsymbol{X}\boldsymbol{t}\left(\boldsymbol{t}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{t}\right)^{-1}\boldsymbol{t}^T\boldsymbol{X}^T, \\ &= \boldsymbol{X}\boldsymbol{t}\boldsymbol{t}^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\boldsymbol{t}^T\right)^{-1}\boldsymbol{t}^T\boldsymbol{X}^T, \\ &= \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T = H_I. \end{aligned}$$

Logo

$$H_{II} = H_I = H.$$

Para o item c) sabemos que o estimador de mínimos quadrados para o modelo de regressão múltipla como em II toma forma

$$\hat{\boldsymbol{\alpha}} = \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Y}.$$

Substituindo o resultado do item a) na equação acima obtemos:

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Y}, \\ &= \left((\boldsymbol{X}\boldsymbol{t})^T(\boldsymbol{X}\boldsymbol{t})\right)^{-1}(\boldsymbol{X}\boldsymbol{t})^T\boldsymbol{Y}, \\ &= \left(\boldsymbol{t}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{t}\right)^{-1}\boldsymbol{t}^T\boldsymbol{X}^T\boldsymbol{Y}, \\ &= \boldsymbol{t}^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{t}^{T^{-1}}\boldsymbol{t}^T\boldsymbol{X}^T\boldsymbol{Y}, \\ &= \boldsymbol{t}^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \\ &= \boldsymbol{t}^{-1}\hat{\boldsymbol{\beta}}. \end{aligned}$$

Podemos resolver também usando o resultado da matriz chapéu:

$$\boldsymbol{Z}\hat{\boldsymbol{\alpha}} = H_{II}\boldsymbol{Y} = H\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}.$$

Substituindo o resultado do item a) no lado esquerdo da equação

$$\boldsymbol{X}\boldsymbol{t}\hat{\boldsymbol{\alpha}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \implies \boldsymbol{t}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\beta}} \implies \hat{\boldsymbol{\alpha}} = \boldsymbol{t}^{-1}\hat{\boldsymbol{\beta}}.$$

■

Comentário: Nesta questão vimos que uma das benesses de um modelo ser linear nos parâmetros é que é possível recuperar as estimativas dos coeficientes uma vez que se saiba qual transformação (linear) foi aplicada na matriz de desenho. Este tipo de procedimento é útil pois transformações da matriz de desenho podem ser utilizadas para melhorar a interpretabilidade e a estabilidade numérica das estimativas; com o resultado que acabamos de derivar, podemos ajustar o modelo apenas uma vez e estudar como as estimativas mudariam sem precisar refazer o procedimento de estimação – que pode ser custoso no caso de milhões de pontos de dados, por exemplo.