Exercícios: revisão A1

Disciplina: Modelagem Estatística
Instrutor: Luiz Max Carvalho
Monitor: Ezequiel Braga

Abril/2025

Questões

1. **(Weighted regression)** Em muitas situações faz sentido ponderar alguns pontos de dados mais do que outros ao ajustar um modelo de regressão, o que é feito usando mínimos quadrados ponderados. Para isso, considere que estamos interessados em minimizar $\sum_{i=1}^{n} w_i(y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2$, para $\boldsymbol{X} \in \mathbb{R}^{n\times(p+1)}$, com a primeira coluna de $1's$; $\boldsymbol{Y} \in \mathbb{R}^n$; $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$; e $w_i > 0,\ i = 1, \ldots, n$.

   (a) Escreva o problema de minimização na forma matricial, usando $\boldsymbol{X}$, $\boldsymbol{Y}$ e $\boldsymbol{W} = \text{diag}(\boldsymbol{w})$, com $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^T$.

   Resolução: O problema de minimização que buscamos resolver inicialmente é:

$$\min_{\beta} \sum_{i=1}^{n} w_i(y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2.$$

   Que pode ser reescrito como:

$$\min_{\beta} \, (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

   (b) Determine o estimador de mínimos quadrados para $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{\text{wls}}$.

   Resolução: Neste caso, seguimos o procedimento de otimização. Para $L(\beta) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$, sabemos que:

$$\hat{\boldsymbol{\beta}}_{\text{wls}} = \arg\min_{\beta} \; L(\beta).$$

   Seguimos:

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} L(\boldsymbol{\beta}) = -2\boldsymbol{X}^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

   Como L é estritamente convexa (verificar), sabemos que $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} L(\hat{\boldsymbol{\beta}}_{\text{wls}}) = 0$ e, portanto:

$$
\begin{aligned}
-2\boldsymbol{X}^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{wls}}) &= 0 \\
\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{wls}} &= 0 \\
\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{wls}} &= \boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{Y} \\
\hat{\boldsymbol{\beta}}_{\text{wls}} &= (\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{Y}
\end{aligned}
$$

(c) O modelo acima pode ser usado quando a premissa de homocedasticidade é violada, isto é, os erros não possuem a mesma variância. Assim, suponha que $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, com $\boldsymbol{\epsilon} \sim \text{MVN}(0, \boldsymbol{V})$ e $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ tal que $\boldsymbol{V}_{ij} = \begin{cases} \sigma_i^2, & \text{se } i = j; \\ 0, & \text{c.c.} \end{cases}$.

   i. Calcule a log-verossimilhança do modelo.
   ii. Determine o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ e deduza a matriz de pesos $\boldsymbol{W}$.

Resolução: Temos

$$l(\boldsymbol{\beta}) \propto -\frac{1}{2}(Y - \boldsymbol{X}\boldsymbol{\beta})^\top V^{-1}(Y - \boldsymbol{X}\boldsymbol{\beta}).$$

Derivando e usando que, para B simétrica, $\dfrac{\partial}{\partial s}(x - As)^T B(x - As) = -2A^\top B(x - As)$,

$$\nabla_{\boldsymbol{\beta}} l(\boldsymbol{\beta}) = \boldsymbol{X}^\top V^{-1}(Y - \boldsymbol{X}\boldsymbol{\beta}).$$

Igualando a 0, concluímos que $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top V^{-1}Y$ e, portanto, $W = V^{-1}$.

2. O modelo linear (de regressão) é um dos cavalos de batalha da Estatística, sendo aplicado em problemas de Finanças, Medicina e Engenharia. Vamos agora estudar como utilizar as propriedades deste modelo para desenhar experimentos com garantias matemáticas de desempenho e obter estimadores de quantidades de interesse.

   a) Uma prática comum em regressão é a de **centrar** a variável independente (covariável), isto é subtrair a média; isto facilita a interpretação do intercepto e também simplifica alguns cálculos importantes. Mostre que no caso com a covariável centrada, $\hat{\beta}_0$ e $\hat{\beta}_1$ são independentes;
   b) Mais uma vez considerando o caso centrado, mostre como obter o número de observações n que faz com que a variância do estimador de máxima verossimilhança do intercepto seja menor que $v > 0$;
   c) Mostre como obter um estimador não-viesado da quantidade $\theta = a\beta_0 + b\beta_1 + c$, com $a, b, c \neq 0$, e encontre o seu erro quadrático médio.

d) Quando $x_{\text{pred}} = \bar{x}$, mostre como obter o número de observações n necessário para que o intervalo de predição de $100(1-\alpha_0)\%$ para a variável-resposta (Y) tenha largura menor ou igual a $l > 0$ com probabilidade pelo menos γ.

Dicas:(i) A expressão dependerá *também* da variância dos resíduos, σ^2 e (ii) Você não precisa calcular n, apenas mostrar o procedimento para obtê-lo.

Resolução: Em uma regressão linear simples, temos:

$$\hat{\beta}_0 \sim \text{Normal}\left(\beta_0, \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{s_x^2}\right)\right),$$
$$\hat{\beta}_1 \sim \text{Normal}\left(\beta_1, \frac{\sigma^2}{s_x^2}\right),$$
$$\text{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\bar{x}\sigma^2}{s_x^2},$$

onde $s_x = \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ e $\hat{\beta}_0$ e $\hat{\beta}_1$ são os estimadores de máxima verossimilhança dos coeficientes.

Para resolver a) basta perceber que quando substituímos a covariável original X por $X' = X - \bar{x}$ temos $\bar{x}' = 0$ e portanto $\text{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\bar{x}'\sigma^2}{s_x^2} = 0$. Para afirmarmos que $\hat{\beta}_0$ e $\hat{\beta}_1$ são independentes é preciso lembrar que estes estimadores têm distribuição conjunta Normal bivariada; quando a covariância é zero, sabemos que são independentes. A resposta de b) vem mais uma vez utilizando a dica dada. Vemos que no caso centrado a variância de $\hat{\beta}_0$ é σ^2/n. Desta forma, precisamos apenas encontrar n tal que $\sigma^2/n < v$, isto é $n > \sigma^2/v$. Como sabemos que os estimadores dos coeficientes são não-viesados (trabalhado em aula, presente nas dicas), podemos encontrar $\hat{\theta} = a\hat{\beta}_0 + b\hat{\beta}_1 + c$ como nosso estimador não-viesado de θ. O EQM de tal estimador é a sua variância:

$$\begin{aligned}
E[(\hat{\theta}-\theta)^2] = \text{Var}(\hat{\theta}) &= a^2\,\text{Var}(\hat{\beta}_0) + b^2\,\text{Var}(\hat{\beta}_1) - ab\,\text{Cov}(\hat{\beta}_0, \hat{\beta}_1),\\
&= a^2\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{s_x^2}\right) + b^2\frac{\sigma^2}{s_x^2} + ab\frac{\bar{x}\sigma^2}{s_x^2},\\
&= \sigma^2\left(\frac{a^2}{n}+\frac{a^2\bar{x}^2}{s_x^2}+\frac{b^2}{s_x^2}+\frac{ab\bar{x}}{s_x^2}\right).
\end{aligned}$$

Por fim, vamos responder d). Note que a expressão necessária aqui é a do intervalo de predição:

$$\hat{Y} \pm c(n,\alpha_0)\cdot\hat{\sigma}'_r\cdot\sqrt{\left[1+\frac{1}{n}+\frac{(x_{\text{pred}}-\bar{x})^2}{s_x^2}\right]},$$

onde

$$c(n,\alpha_0) := T^{-1}\left(1-\frac{\alpha_0}{2}; n-2\right),$$

e

$$\hat{\sigma}'_r := \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)^2}{n-2}}.$$

Quando $x_{\text{pred}} = \bar{x}$ a expressão se reduz um pouco e podemos deduzir que a largura do intervalo é

$$\hat{l} = 2 \cdot c(n, \alpha_0) \cdot \hat{\sigma}'_r \sqrt{\left[1 + \frac{1}{n}\right]}.$$

Desejamos, portanto, encontrar n tal que

$$\Pr\left(\hat{l} < l\right) \geq \gamma,$$

$$\Pr\left(\hat{\sigma}'_r < \frac{l}{2 \cdot c(n, \alpha_0) \cdot \sqrt{\left[1 + \frac{1}{n}\right]}}\right) \geq \gamma,$$

isto é conseguimos reduzir nossa afirmação probabilística a uma afirmação com respeito à f.d.a. (ou CDF) de $\hat{\sigma}'_r$. Para completar nossos cálculos só precisamos nos lembrar que $n\hat{\sigma}'_r/\sigma^2$ tem distribuição qui-quadrado com $n-2$ graus de liberdade (De Groot, Teorema 11.3.2) e, portanto,

$$\Pr\left(\hat{\sigma}'_r \leq a\right) = F_{\chi}\left(\frac{\sigma^2}{n} a; n-2\right).$$

3. **(Regressão polinomial)** Para $i = 1, 2, \ldots, n$, considere o modelo

$$Y_i = aX_i + bX_i^2 + cX_i^3 + d + \varepsilon_i, \tag{1}$$

onde ε_i são variáveis aleatórias com média 0 e variância constante e X_i são constantes observadas.

a) O modelo em questão é linear?

Resolução: Sim, pois é linear nos parâmetros que estão sendo estimados. É um caso especial de regressão linear.

b) Como você estimaria os parâmetros desse modelo? Escreva a função de perda.

Resolução: Podemos estimar por mínimos quadrados, isto é,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left[\right] Y_i - \left(aX_i + bX_i^2 + cX_i^3 + d\right)\big]^2,$$

para $\boldsymbol{\beta} = (d, a, b, c)^\top$.

c) É possível obter um estimador não-viesado de $\boldsymbol{\beta} = (d, a, b, c)^\top$?

Resolução: Sim. Reescrevendo o problema para $\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$ e $\boldsymbol{\varepsilon} = \sigma^2 I_n$, temos $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, cuja solução dos mínimos quadrados é $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. Como vimos na regressão linear, este estimador é não-viesado.

d) Qual a interpretação de d? E a de b?

Resolução: d é o intercepto como na regressão linear. Ele representa a mudança esperada em $\boldsymbol{Y}_i$ quando $\boldsymbol{X}_i = 0$. Já b mede a inclinação da curva, isto é, a mudança esperada na inclinação quanto ela muda quando $\boldsymbol{X}_i$ cresce.

e) (Desafio) Mostre que o estimador de mínimos quadrados está bem definido se e somente se todos os X_i são distintos.

Resolução: Precisamos mostrar que $\boldsymbol{X}^\top \boldsymbol{X}$ é invertível. A ideia é usar que $\operatorname{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = \operatorname{rank}(\boldsymbol{X})$. Assim, precisamos mostrar que $\operatorname{rank}(\boldsymbol{X}) = p+1$ (nesse caso, $p=3$). Antes de tudo, precisamos ter $n \geq p+1$ para ser possível satisfazer o rank desejado. Considere $\boldsymbol{Z}$ o bloco de $\boldsymbol{X}$ formado pelas primeiras $p+1$ linhas, isto é,

$$\boldsymbol{Z} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ 1 & x_4 & x_4^2 & x_4^3 \end{bmatrix}.$$

Note que $\boldsymbol{Z}$ é uma matriz de Vandermonde, cujo determinante é conhecido por ser

$$\det(\boldsymbol{Z}) = \prod_{0 \leq i < j \leq p+1} (x_j - x_i).$$

Logo, $\det(\boldsymbol{Z}) \neq 0 \Leftrightarrow x_i \neq x_j$, $\forall i, j = 1, \ldots, p+1$, ou seja, $\operatorname{rank}(\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{Z}) = p+1 \Leftrightarrow x_i \neq x_j$, $\forall i, j = 1, \ldots, p+1$, o que conclui a prova.

4. **(Família exponencial)** Considere as distribuições de probabilidade abaixo. Usando as propriedades da família exponencial, para cada uma, mostre:

   i) A estatística suficiente e sua esperança

   ii) O estimador de máxima verossimilhança.

   (a) `Pareto`. Para $y > 1$ e $\theta > 0$ a densidade vale

   $$f(y;\theta) = \theta y^{-\theta-1}.$$

   (b) `Binomial negativa`. Para $y \in \mathbb{N} \cup \{0\}$ e $\theta, \phi > 0$ a função de massa vale

   $$f(y;\theta) = \binom{y+\phi-1}{y} \left(\frac{\theta}{\theta+\phi}\right)^y \left(\frac{\phi}{\theta+\phi}\right)^\phi,$$

   com ϕ fixo e conhecido.

Resolução: Tome uma família exponencial em forma canônica, isto é, $f_{\boldsymbol{X}}(x \mid \boldsymbol{\eta}) = h(x)\exp\{\boldsymbol{\eta}^\top T(x) - A(\boldsymbol{\eta})\}$, onde fica evidente que, para uma amostra aleatória $X_1, X_2, \ldots, X_n$, $T(\boldsymbol{X}) = \sum_{i=1}^n T(X_i)$ é estatística suficiente, pelo teorema da fatorização. Fica de exercício verificar que $\nabla A(\boldsymbol{\eta}) = \mathbb{E}_{\boldsymbol{\eta}}[T(\boldsymbol{X})] =: \mu(\boldsymbol{\eta})$ e que $\widehat{\boldsymbol{\eta}}_{\text{MLE}} = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^n T(X_i)\right)$. A partir disso, basta escrever as distribuições acima na forma canônica e obter as respostas. Para a Pareto$(\theta, 1)$, $A(\eta) = -\log(-1-\eta)$ e $T(x) = \log(x)$, o que nos leva a $\mu(\eta) = -\frac{1}{1+\eta}$ e $\widehat{\eta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^n \log(X_i)} - 1$. Logo, $\mathbb{E}[T(\boldsymbol{X})] = -\frac{n}{1+\eta} = \frac{n}{\theta}$. Agora, note que $\eta = -\theta - 1$, ou seja, $\widehat{\theta}_{\text{MLE}} = -\widehat{\eta}_{\text{MLE}} - 1 = \frac{n}{\sum_{i=1}^n \log(X_i)}$ (pela invariância do MLE). Para a binomial negativa, temos $A(\eta) = -\phi \log(1 - \exp(\eta))$, $T(x) = x$ e $\eta = \log\left(\frac{\theta}{\theta+\phi}\right)$, o que nos leva a $\mu(\eta) = \frac{\phi \exp(\eta)}{1-\exp(\eta)}$ e $\widehat{\eta}_{\text{MLE}} = \log\left(\frac{\bar{X}}{\bar{X}+\phi}\right)$. Como $\eta = \log\left(\frac{\theta}{\theta+\phi}\right)$, $\widehat{\eta}_{\text{MLE}} = \bar{X}$. Além disso, $\mathbb{E}[T(\boldsymbol{X})] = \frac{n\phi \exp(\eta)}{1-\exp(\eta)} = n\theta$.