Homework 01

Instruções:

1. Este Homework (HW) é uma atividade individual;
2. O HW precisará conter o **nome completo** e a **matrícula** do discente;
3. O HW deverá ser enviado via `Google Classroom` **até as 09h51 do dia 06/05/2026** em **um único arquivo** e em **PDF** (*Portable Document Format*);
4. **Não serão considerados** Homeworks com péssima qualidade visual ou em desacordo com as instruções anteriores.

Exercícios.

1. Sejam $Y = \alpha + \beta X + \varepsilon$, em que α e β são constantes, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ e X uma variável aleatória, com X e ε independentes. Calcule a covariância entre Y e X quando X tem distribuição:

   (a) uniforme no intervalo (a, b).

   (b) normal de média μ e variância σ^2.

   (c) exponencial λ, em que $\mathbb{E}(X) = 1/\lambda$.

2. A partir da forma matricial da soma de quadrados,

$$Q = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

   prove que o estimador de mínimos quadrados de $\boldsymbol{\beta}$:

   (a) é dado por $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$;

   (b) é não viesado;

   (c) e têm a menor variância entre todos os estimadores lineares não viesados.

3. Suponha que o interesse é encontrar o estimador de mínimos quadrados de $\boldsymbol{\beta}$ no modelo de regressão linear $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ sujeito à restrição $\boldsymbol{T}\boldsymbol{\beta} = \boldsymbol{c}$. Mostre que este estimador é dado por

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{T}^\top[\boldsymbol{T}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{T}^\top]^{-1}(\boldsymbol{c} - \boldsymbol{T}\widehat{\boldsymbol{\beta}}),$$

em que $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$.

4. Prove que os estimadores de máxima verossimilhança de β_1 e β_2, do modelo de regressão normal linear simples,

$$Y_\ell = \beta_1 + \beta_2 x_\ell + \varepsilon_\ell,$$

em que $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\ell = 1, 2, \ldots, n$, podem ser escritos da seguinte forma:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{x},\ \hat{\beta}_2 = \frac{\sum_{\ell=1}^{n}(x_\ell - \bar{x})(Y_\ell - \bar{Y})}{\sum_{\ell=1}^{n}(x_\ell - \bar{x})^2}.$$

5. Prove que as matrizes abaixo são simétricas e idempotentes:

   (a) $\boldsymbol{A} = \frac{1}{n}\boldsymbol{J} = \boldsymbol{1}_n(\boldsymbol{1}_n^\top\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^\top$;

   (b) $\boldsymbol{B} = \boldsymbol{I}_n - \boldsymbol{A}$.

6. Considere o modelo de regressão normal linear com intercepto $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, em que $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

   (a) Determine $\mathbb{E}(\text{QMT})$ para o caso geral, em função de $\boldsymbol{\beta}$ e σ^2.

   (b) Mostre que, sob a hipótese nula $\mathcal{H} : \beta_2 = \beta_3 = \cdots = \beta_p = 0$, o QMT é um estimador não-viesado para σ^2.

7. Considere o modelo de regressão normal linear com intercepto $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, em que $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. Sob a hipótese nula $\mathcal{H} : \beta_2 = \beta_3 = \cdots = \beta_p = 0$, prove que:

   (a) $\text{SQReg}/\sigma^2 \sim \chi^2_{p-1}$;

(b) $\text{SQRes}/\sigma^2 \sim \chi^2_{n-p}$;

(c) $\text{QMReg}/\text{QMRes} \sim F(p-1, n-p)$.

8. Mostre que a inclusão de uma nova variável explicativa no modelo de regressão linear aumenta o valor do coeficiente de determinação ajustado (R^2_{a}) se, e somente se, a estatística F parcial associada ao teste de significância dessa variável for maior que 1.

9. Tomando como referências Buse (1982) e Montoril and Souza (2013), descreva as características, semelhanças e diferenças das estatísticas da razão de verossimilhança, Wald, escore e gradiente. Discuta como essas estatísticas se relacionam no contexto do modelo de regressão linear normal e comente sobre a facilidade computacional de cada uma em relação à necessidade de estimar o modelo sob a hipótese nula ou alternativa.

10. Seja o seguinte modelo de regressão normal linear simples,

$$Y_\ell = 1 + x_{\ell 2} + \varepsilon_\ell,$$

em que $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\ell = 1, 2, \ldots, n$. A partir de uma simulação de Monte Carlo, com 5.000 réplicas, avalie o comportamento do viés absoluto e do erro quadrático médio dos estimadores, nos seguintes cenários:

- $x_{\ell 2} \sim$ uniforme no intervalo $(0, 1)$.
- $x_{\ell 2} \sim$ uniforme no intervalo $(1, 2)$.
- $x_{\ell 2} \sim$ normal de média 0 e variância 1.
- $x_{\ell 2} \sim$ exponencial 1.

Para $\sigma^2 \in \{4, 12\}$ e $n \in \{10, 20, 40, 80, 160\}$. Comente os resultados obtidos, dentro de cada cenário e de maneira global.

Observações:

i. Para um mesmo σ^2 e n, a variável preditora $\boldsymbol{x}_2$ não se altera dentro das 5.000 réplicas.

ii. Utilizar os três últimos dígitos da matrícula (ignorando possíveis letras) como semente aleatória.

iii. Em caso de utilização de um código diferente do disponibilizado pelo professor, anexá-lo.

Referências

Buse, A. (1982). The likelihood ratio, wald and lagrange multiplier tests: an expository note. *The American Statistician 36*(3), 153–157.

Montoril, M. H. and E. A. Souza (2013). Estatística gradiente: propriedades e aplicações. *Revista Brasileira de Biometria 31*(1), 43–60.