1^a Prova - FECD B - 2025/01

Renato Assunção

1. An agricultural researcher is studying the number of insect eggs found on individual leaves of a particular crop. For each sampled leaf, the following two features are recorded: x_1 : the leaf area (in square centimeters) and x_2 : exposure level to sunlight (coded as 0 = shaded, 1 = partial sun, 2 = full sun). Let Y_i denote the number of insect eggs observed on leaf i. The researcher assumes the counts follow a Poisson distribution with mean λ_i, and proposes the following model for the Poisson mean:
$$\lambda_i = \exp(\beta_0 + \beta_1 \cdot \text{Area}_i + \beta_2 \cdot \text{Sunlight}_i)$$
Based on a random sample of n leaves with observed values $\{(y_i, \text{Area}_i, \text{Sunlight}_i), i = 1, \ldots, n\}$, derive the MLE for $\beta_0, \beta_1, \beta_2$ under the model:
$$Y_i \sim \text{Poisson}(\lambda_i), \quad \text{with } \lambda_i = \exp(\beta_0 + \beta_1 \cdot \text{Area}_i + \beta_2 \cdot \text{Sunlight}_i)$$
Write the log-likelihood function, compute the gradient, and describe how the MLEs could be computed in practice using Newton-Raphson.

   HINT: If $Y \sim \text{Poisson}(\lambda)$ then
$$\mathbb{P}(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

---

2. Considere o modelo de regressão linear:
$$Y = X\beta + \varepsilon, \quad \text{com } \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$
onde $X \in \mathbb{R}^{n\times(p+1)}$ tem posto completo e sua primeira coluna é composta por 1's. Seja $\hat{\beta} = (X'X)^{-1}X'Y$ o estimador de mínimos quadrados e $\hat{Y} = X\hat{\beta}$ o vetor de valores ajustados. A matriz $H = X(X'X)^{-1}X'$ é chamada de matriz *hat*.

   (a) Mostre que H é idempotente e simétrica, ou seja, $H^2 = H$ e $H' = H$.

   (b) Mostre que $HX = X$ e que $(I - H)X = 0$.

---

3. **Variância dos Resíduos e Valores Ajustados** Com as mesmas definições do problema anterior, defina o vetor de resíduos como $r = Y - \hat{Y}$.

   (a) Mostre que $\text{Var}(\hat{Y}) = \sigma^2 H$ e $\text{Var}(r) = \sigma^2(I - H)$.

   (b) Mostre que os vetores $\hat{Y}$ e r são não correlacionados.

---

4. **Vício e Variância do Estimador em Regressão Linear** Considere o modelo de regressão simples:
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{com } \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.}$$
Defina $\hat{\beta}$ como o estimador de mínimos quadrados do vetor $\beta = (\beta_0, \beta_1)$.

   (a) Mostre que $\hat{\beta}$ é não-viciado para estimar β.

   (b) Derive a variância de $\hat{\beta}$.