UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
MAT02035 - Modelos para dados correlacionados

Avaliação 01

Exercício 1. (1 ponto) Apresente as principais características de um estudo longitudinal.

Exercício 2. (1 ponto) Explique por que dados longitudinais podem ser vistos como dados agrupados. Dê um exemplo de outro tipo de estudo que pode fornecer dados agrupados.

Exercício 3. (1 ponto) Descreva as diferenças (incluindo possíveis vantagens e desvantagens) entre estudos longitudinais e estudos transversais.

Exercício 4. (1 ponto) Para um estudo longitudinal, descreva as diferenças entre os delineamentos balanceado e desbalanceado. Dê um exemplo para cada um dos delineamentos.

Exercício 5. (1 ponto) Quais as consequências de se ignorar (na análise) a correlação presente entre os dados de medidas repetidas?

Exercício 6. (2 pontos) O estudo *Treatment of Lead-Exposed Children* (TLC) foi um estudo aleatorizado e controlado por placebo de *succimer* (um agente quelante) em crianças com níveis de chumbo no sangue de 20 a 44 microgramas/dL. Lembre-se de que os dados consistem em quatro medições repetidas dos níveis de chumbo no sangue obtidos na linha de base (ou semana 0), semana 1, semana 4 e semana 6 em 100 crianças que foram aleatoriamente designadas para tratamento de quelação com *succimer* ou placebo. Para este conjunto de problemas, focamos apenas nas 50 crianças designadas para tratamento de quelação com *succimer*.

Os dados brutos são armazenados em um arquivo externo: `lead.dta`. Cada linha do conjunto de dados contém as 5 variáveis a seguir: `ID`, `Y1`, `Y2`, `Y3`, `Y4`.

a. Leia os dados do arquivo externo e calcule as médias da amostra, os desvios padrão e as variâncias dos níveis de chumbo no sangue em cada ocasião.
b. Construa um gráfico de tempo dos níveis médios de chumbo no sangue versus tempo (em semanas). Descreva as características gerais da tendência temporal.
c. Calcule as matrizes de covariância e correlação 4×4 para as quatro medidas repetidas dos níveis de chumbo no sangue.
d. Verifique se os elementos diagonais da matriz de covariâncias são as variâncias comparando com a estatística descritiva obtida no item (a).

Exercício 7. (3 pontos) Seja $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$ um vetor de respostas de medidas repetidas, X_i uma matriz $n_i \times p$ de covariáveis associadas ao vetor de respostas Y_i, para $i = 1, \ldots, N$, em que N é o número de indivíduos em um certo estudo. Considere o seguinte modelo:

$$Y_i = X_i\beta + e_i, i = 1, \ldots, N,$$

em que $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ é um vetor de coeficientes de regressão desconhecidos, e $e_i = (e_{i1}, e_{i2}, \ldots, e_{in_i})'$ é um vetor de erros aleatórios. Ainda, suponha que $e_i \sim N_{n_i}(0, \Sigma_i)$, em que Σ_i é uma matriz de covariâncias.

a. Qual a distribuição condicional de $Y_i|X_i$?
b. Escreva a função de verossimilhança de β (considere Σ_i conhecido).
c. Utilizando o método da máxima verossimilhança, encontre a expressão do estimador de β (considere Σ_i conhecido). (Apresente o desenvolvimento)
d. Demonstre que o estimador de máxima verossimilhança $\widehat{\beta}$ é não-enviesado para β. Em palavras, explique as implicações práticas deste resultado.
e. Suponha que este modelo foi ajustado para um certo conjunto de dados, considerando três covariáveis (portanto, $\beta = (\beta_1, \beta_2, \beta_3)'$) e que $\widehat{\beta}_2 = 1.4$. A matriz de covariância estimada de $\widehat{\beta}$ é apresentada a seguir:

$$\widehat{\text{Cov}}(\widehat{\beta}) = \begin{bmatrix} 0.50 & -0.50 & -0.12 \\ -0.50 & 1.01 & 0.12 \\ -0.12 & 0.12 & 0.63 \end{bmatrix}.$$

Apresente um intervalo de confiança (IC) de 95% para β_2. Justifique este método de construção do IC.