CE075 Análise de Dados Longitudinais: Lista 1

Questão 1. Suponha que você esteja planejando um estudo clínico aleatorizado cujo objetivo é reter usuários de droga em programas de reabilitação. Existem dois tipos de tratamentos. Identifique uma resposta que leve a análises transversais, longitudinais e de sobrevivência (tempo até a ocorrência de um evento).

Questão 2. A tabela a seguir mostra uma série de taxas de incidência transversais (por 100.000) do câncer Y por idade e ano de calendário.

	Ano de calendário							
Idade	1950	1955	1960	1965	1970	1975	1980	1985
20-24	10	15	22	30	33	37	41	44
25-29	8	17	20	24	29	38	40	43
30-34	5	12	22	25	28	35	42	45
35-39	3	12	15	26	30	32	39	42
40-44	2	10	17	19	28	32	39	42
45-49	2	12	15	18	21	33	40	42
50-54	2	10	16	20	25	32	42	44
55-59	2	15	17	19	22	27	43	44

a) Após observar as taxas de incidência por idade em um dado ano, se concluiu que "aumento na idade não está relacionado com um aumento na incidência de Y podendo até estar relacionada com uma diminuição na incidência". Você concorda com esta observação? Justifique sua resposta.

b) Quais são os objetivos de examinar taxas nas coortes de nascimento (análise longitudinal) *versus* taxas transversais (análise transversal)?

Questão 3. Considere um estudo longitudinal que tem por objetivo comparar as taxas de crescimento de meninos e meninas avaliados nas idades de 8, 10, 12 e 14 anos. Para fins de análise o tempo é padronizado na idade média, ou seja, 11 anos. O verdadeiro perfil de crescimento na população pode ser descrito através do seguinte modelo:

$$\begin{aligned} Y_{ij} = &\beta_0 + \beta_1 \times Sexo_i + \beta_2 \times Tempo_j + \beta_3 \times Sexo_i \times Tempo_j + \varepsilon_{ij}, \\ &i = 1, 2, \ldots, N, \quad j = -3, -1, 1, 3, \end{aligned}$$

com $E(\varepsilon_{ij}) = 0$ e $Var(\varepsilon_{ij}) = 2\ \forall j$. Considere a variável *Sexo* como binária assumindo os valores zero para menina e um para menino. Os verdadeiros valores de $\boldsymbol{\beta}$ são $\boldsymbol{\beta} = (22, 2.30, 0.50, 0.30)^T$. Assim, admitimos que existe diferença entre os sexos quanto às taxas de crescimento e, ainda, meninos e meninas crescem em ritmos diferentes.

Considere as estruturas de correlação a seguir:

a) Simetria composta com $\rho = 0.8$;

b) Autorregressivo com $\rho = 0.90$;

c) Não estruturada com

$$Corr(Y_i) = \begin{bmatrix} 1.00 & 0.55 & 0.81 & 0.40 \\ 0.55 & 1.00 & 0.60 & 0.75 \\ 0.81 & 0.60 & 1.00 & 0.79 \\ 0.40 & 0.75 & 0.79 & 1.00 \end{bmatrix}$$

Com o objetivo de comparar os estimadores GLS e GEE faça um estudo de simulação gerando 1000 amostras de tamanho $N = 100$ cada (50 crianças em cada grupo) do modelo acima. Pede-se:

1. Para cada estrutura de correlação obtenha os estimadores GLS e GEE ajustando as estruturas *independente*, *simetria composta*, *ar1* e *não estruturada* aos dados gerados.

2. Compare o vício das estimativas, erro padrão e cobertura empírica para os oito modelos ajustados.

3. O que se pode dizer sobre o impacto da má especificação da estrutura de correlação quando se considera o estimador GLS? A robustez do método GEE foi verificada?

Questão 4. Considere os seguintes modelos:

1) $Y_{ij} = \beta_0 + \beta x_{ij} + \varepsilon_{ij}$, (Efeito Transversal igual a Longitudinal);

2) $Y_{ij} = \beta_0 + \beta_T x_{i1} + \beta_L(x_{ij} - x_{i1}) + \varepsilon_{ij}$, (Efeitos Longitudinal e Transversal diferentes).

$i = 1, \ldots, N,\ j = 1, \ldots, n,\ E(\varepsilon_{ij}) = 0.$

a) Encontre o estimador de mínimos quadrados de β para o primeiro modelo.

b) Se o modelo verdadeiro é o segundo, encontre $E(\hat{\beta})$, para $\hat{\beta}$ encontrado em (a).

c) Considere os seguintes dados referentes a cinco indivíduos avaliados em três ocasiões (1, 5 e 10 dias). Ajuste os dois modelos, interprete e compare os coeficientes estimados.

Ind.	Y_{i1}	Y_{i2}	Y_{i3}	x_{i1}	x_{i2}	x_{i3}
1	10	12	16	1	3	5
2	9	10	14	2	4	6
3	8	11	15	3	5	7
4	6	8	12	4	6	8
5	7	10	15	5	7	9

Questão 5. Seleciona-se uma amostra aleatória de N pessoas, sendo $N/2$ homens e $N/2$ mulheres. Em cada indivíduo, realizam-se duas medidas: uma no início e outra no final do experimento. Seja Y_{it} a resposta do i-ésimo indivíduo no tempo t ($t = 0, 1$). Como o objetivo do estudo é avaliar o efeito de grupo (sexo) na resposta, decidiu-se usar o modelo $Y_{it} = \beta_0 + \beta_1 x_i + \varepsilon_{it}$, em que $x_i = 0$ (homem) e 1 (mulher). Considere que $Var(\varepsilon_{it}) = \sigma^2$ que $Corr(\varepsilon_{i0}, \varepsilon_{i1}) = \rho$. Mostre que:

(a) O estimador de mínimos quadrados de β_1 é dado por:

$$\hat{\beta}_1 = \frac{\sum_i x_i(y_{i0} + y_{i1}) - \sum_i (1 - x_i)(y_{i0} + y_{i1})}{N};$$

(b) e a variância é dada por:

$$Var(\hat{\beta}_1) = \frac{2\sigma^2}{N}(1 + \rho).$$

Questão 6. Faça uma análise exploratória dos dados referentes aos ratos em terapia hormonal. Os 50 ratos foram divididos em três grupos (controle, dose baixa e dose alta) e foram avaliados em 7 ocasiões (50, 60, 70, 80, 90, 100 e 110 dias). A resposta medida foi a distância cranio-facial no raio-x (em pixels) e os objetivos são: (1) comparar os grupos e (2) avaliar as mudanças temporais. O banco de dados está disponível em https://docs.ufpr.br/ jlpadilha/CE075/Datasets/rats01.xlsx.

Questão 7. No Exercício (6) foi realizada uma análise exploratória dos dados referentes aos ratos em terapia hormonal. A partir dos resultados exploratórios proponha um ou mais modelos para estes dados. Ajuste os modelos e interprete os resultados. Existem diferenças entre os grupos?