

Universidade do Minho
Escola de Ciências

Elsa Pombo Soares

**Estruturas de Correlação Paramétricas
e Semi-Paramétricas para Modelos
Longitudinais**



Universidade do Minho
Escola de Ciências

Elsa Pombo Soares

**Estruturas de Correlação Paramétricas
e Semi-Paramétricas para Modelos
Longitudinais**

Dissertação de Mestrado
Mestrado em Estatística para Ciência de Dados

Trabalho efetuado sob a orientação da
**Professora Doutora Inês Pereira Silva Cunha
de Sousa**

janeiro de 2023

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-CompartilhaIgual

CC BY-NC-SA

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Agradecimentos

Ao meu pai, Berto, e à minha mãe, Lúcia, pela proteção, segurança e por me oferecerem oportunidades e recursos para continuar a estudar.

Ao meu irmão, Carlos, por estar sempre lá, por compartilhar memórias e experiências e por ser uma parte importante da minha vida.

Ao meu incrível namorado Vítor por toda a paciência que tem comigo e por tornar a minha vida mais feliz.

Aos meus amigos de infância por estarem sempre prontos para tudo, a qualquer hora.

Aos meus amigos de Braga, em especial, ao meu grande amigo Hugo por todo o apoio emocional e académico.

À professora Inês Sousa por toda a orientação durante o desenvolvimento desta dissertação e por me oferecer oportunidade de crescimento.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Resumo

Modelos para dados longitudinais são modelos estatísticos para a análise de dados repetidos ao longo do tempo para diferentes indivíduos. Estes modelos, para além de incorporarem uma modelação de regressão para a média da população, permitem modelar a estrutura de correlação existente nos dados. A modelação da estrutura de correlação é fundamental, uma vez que dados repetidos ao longo do tempo para vários indivíduos pressupõem a existência de correlação temporal entre as medidas e a correlação entre observações de um mesmo indivíduo.

Em muitos casos, assume-se uma estrutura de correlação baseada em conhecimentos prévios ou intuição, mas, a escolha incorreta desta pode levar a estimativas tendenciosas ou ineficientes. Neste sentido, foi desenvolvido um trabalho teórico destes modelos de correlação paramétricos e semi-paramétricos e apresentada uma ferramenta usual, intitulada de variograma, que permite que a estrutura seja selecionada com base na evidência empírica.

Além disso, esta dissertação apresenta a teoria de modelos conjuntos para dados longitudinais e de sobrevivência onde passará pelas diferentes metodologias existentes e apresentará exemplos de dados reais para contextualizar esses modelos. O Modelo Gaussiano Transformado é uma abordagem particular, totalmente paramétrica, destes modelos, e a sua estrutura de correlação é especificada através da decomposição de três componentes de correlação. Aqui descreve-se quais os principais pressupostos por detrás desta estrutura, e como estes podem ser utilizados de forma apurada para extrair informação relevante dos dados. Por fim, é realizado um estudo de simulação para as possíveis estruturas de correlação.

Palavras Chave: Modelo Longitudinal, Modelos lineares, Estrutura de Correlação, Modelo Gaussiano Transformado, Correlação Conjunta, Variograma.

Abstract

Longitudinal data models are statistical models for analyzing data repeated over time for different individuals. These models, in addition to incorporating regression modeling for the population mean, allow the correlation structure in the data to be modeled. Modeling the correlation structure is crucial, since data repeated over time for several individuals assumes the existence of temporal correlation between measurements and correlation between observations from the same individual.

In many cases, a correlation structure is assumed based on prior knowledge or intuition, but choosing this incorrectly can lead to biased or inefficient estimates. In this sense, a theoretical work of these parametric and semi-parametric correlation models was developed and a usual tool, entitled variogram, was presented, which allows the structure to be selected based on empirical evidence.

In addition, this dissertation presents the theory of joint models for longitudinal and survival data where it will go through the different methodologies that exist and present real data examples to put these models into context. The Transformed Gaussian Model is a particular approach, fully parametric, to these models, and its correlation structure is specified by decomposing three correlation components. Here it is described what the main assumptions behind this structure are, and how these can be used in a refined way to extract relevant information from the data. Finally, a simulation study is conducted for the possible correlation structures.

Palavras Chave: Longitudinal model, Linear Models, Correlation Structure, Transformed Gaussian Model, Joint Correlation, Variogram.

Conteúdo

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Contextualização e Objetivos | 1 |
| 1.2 | Estrutura da Dissertação | 3 |
| 2 | Conceitos de Álgebra Linear | 5 |
| 2.1 | Valores e Vetores Próprios | 5 |
| 2.2 | Norma Vetorial e Norma Matricial | 7 |
| 2.3 | Matrizes Simétricas Definidas Positivas | 10 |
| 3 | Metodologia Estatística para Dados Longitudinais | 11 |
| 3.1 | Modelos para Dados Longitudinais | 14 |
| 3.1.1 | Modelo Linear Geral | 16 |
| 3.1.2 | Modelo Linear Misto | 24 |
| 3.1.3 | Modelo Linear Misto Geral | 26 |
| 3.2 | Modelos para a Estrutura de Correlação | 28 |
| 3.2.1 | Estruturas de Correlação para dados balanceados | 28 |
| 3.2.2 | Estruturas de Correlação Serial para tempo contínuo | 33 |
| 3.2.3 | Função Variograma | 34 |
| 3.3 | Ajuste e Seleção do Modelo | 40 |
| 3.4 | Modelos Conjuntos | 42 |
| 4 | Modelo Gaussiano Transformado | 50 |
| 4.1 | Função Likelihood | 51 |
| 4.2 | Estudo de Estrutura de Correlação | 52 |
| 4.3 | Estudo de Simulação | 58 |
| 5 | Conclusões e Trabalho Futuro | 61 |
| | Referências Bibliográficas | 62 |

Lista de Figuras

| | | |
|-----|---|----|
| 3.1 | Função de correlação exponencial e gaussiana para $\phi = 5, 10$ | 34 |
| 3.2 | Variogramas teóricos assumindo $\sigma^2 = 1$ para o modelo de correlação exponencial e gaussiano, em $\phi = 5$ e 10 | 36 |
| 3.3 | Variogramas teóricos assumindo $\tau^2 = 0.2$ e $\sigma^2 = 1$ para o modelo de correlação exponencial e gaussiano em $\phi = 5$ e 10 | 37 |
| 3.4 | Variogramas teóricos assumindo $\tau^2 = 0.2$, $\sigma^2 = 1$ e $\nu^2 = 0.2$ para o modelo de correlação exponencial e gaussiano em $\phi = 5$ e 10 | 38 |
| 4.1 | Função de correlação exponencial e gaussiana para $\rho = 1$ 35 7. | 55 |
| 4.2 | Função de correlação exponencial e gaussiana dos logaritmos dos tempos para $\rho = 1$ 35 7. | 57 |

Capítulo 1

Introdução

1.1 Contextualização e Objetivos

Muitos estudos clínicos e epidemiológicos geram dados longitudinais e/ou de sobrevivência. Os primeiros são obtidos quando uma resposta é medida repetidamente num grupo de unidades e os de sobrevivência envolvem a modelagem do tempo de duração esperado até a ocorrência de um ou mais eventos de interesse.

Nos últimos anos tem-se assistido a um forte desenvolvimento de metodologias estatísticas para analisar estes tipos de dados em separado, as quais desempenham um papel preponderante nas mais variadas áreas da ciência. Contudo, esta abordagem pode ser desajustada quando a variável longitudinal está correlacionada com o estado de saúde do paciente.

De forma a resolver esta questão, o principal objetivo desta dissertação é o de apresentar e estudar um modelo conjunto particular de dados longitudinais e de sobrevivência, denominado Modelo Gaussiano Transformado (7). Este é descrito de forma totalmente paramétrica e, para além de incorporar uma modelação de regressão para a média da população, permite modelar a estrutura de correlação presente nos dados.

No que diz respeito à análise e interpretação da componente da parte da média é, essencialmente, o estudo que se utiliza nos modelos de regressão linear usuais. Relativamente à modelação da estrutura de correlação, é necessário ter em consideração que dados repetidos ao longo do tempo para vários indivíduos pressupõem a existência de correlação temporal entre as medidas, mas também a correlação entre observações do mesmo indivíduo. Como poderemos, posteriormente, constatar, existe uma interdependência entre a resposta média e a estrutura de covariância pelo que, de facto, a especificação incorreta da estrutura para a média influencia a estrutura de covariância, devendo esta ser escolhida com base nos resíduos do modelo retirando a parte da média.

Neste trabalho estamos, fundamentalmente, interessados na modelação e análise de pos-

síveis estruturas de correlação paramétricas e semi-paramétricas a serem incorporadas neste modelo, estabelecendo relações entre a componente da estrutura de correlação das medidas repetidas ao longo do tempo e a componente da estrutura de correlação conjunta.

Na literatura estatística, existem grandes contribuições para estruturas de correlações longitudinais, ao contrário do que acontece na componente conjunta. Neste sentido, para além de fazermos uma descrição das estruturas de correlação disponíveis para as medidas repetidas tais como, o modelo de correlação saturado, a correlação de simetria composta e estruturas de correlação serial quer para tempos discretos, quer para tempos contínuos, são também propostas novas estruturas admissíveis para a estrutura de correlação conjunta, cuja interpretação é bastante intuitiva. A ideia principal é caracterizá-la como sendo uma função de correlação que, além de depender de um conjunto finito de parâmetros, necessita dos tempos em que as observações são medidas, tendo o último tempo um papel fundamental.

Além disso, na maioria dos casos, especificar, através de uma base de dados real, qual poderá ser uma estrutura de correlação adequada a incorporar no modelo não é uma tarefa fácil. Neste sentido, é também sugerido uma “espécie” de variograma empírico que, à semelhança com o variograma usual, permite explorar, descrever e interpretar qual poderá ser a verdadeira estrutura de correlação conjunta existente nos dados.

Deste modo, através de um estudo teórico, obtém-se propriedades específicas da estrutura de correlação total, incluindo as três componentes de correlação: a do processo longitudinal; a do processo de sobrevivência e a estrutura de correlação conjunta. De seguida, é realizado um estudo de simulação em que supomos, à partida, a forma como os dados estão correlacionados e averiguamos que tipo de informações ou propriedades podemos extrair sobre os dados.

De forma a uniformizar a notação aqui utilizada, matrizes serão representadas por letras maiúsculas em negrito (por exemplo, \mathbf{A} , \mathbf{V} , $\mathbf{\Sigma}$) e vetores por letras minúsculas em negrito (por exemplo, \mathbf{g} , $\mathbf{\beta}$, $\mathbf{\gamma}$). O conjunto de matrizes com m linhas e n colunas, cujos elementos são reais ou de entradas reais, será denotado por $\mathbb{R}^{m \times n}$ e quando se trata de vetores ($n = 1$) representamos simplesmente por \mathbb{R}^m .

Poderá ser relevante identificarmos unicamente a ordem da matriz, pelo que escreveremos $(m \times n)$. $A^{m \times n}$ representa a matriz de dimensão $(m \times n)$, a qual pode ser expressa como $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ em que \mathbf{a}_j denota a j -ésima coluna, ou seja,

$$\mathbf{A}^{m \times n} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \quad \mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

Por vezes é útil utilizar a notação (i, j) ou a_{ij} para designar o elemento da matriz \mathbf{A}

situado no cruzamento da i -ésima linha com a j -ésima coluna. Além disso, denote-se $\mathbf{0}^n$ o vetor nulo de ordem n e, se não for necessário especificar a ordem do vetor nulo representamos simplesmente por $\mathbf{0}$.

1.2 Estrutura da Dissertação

Esta dissertação encontra-se dividida em cinco capítulos: o primeiro consiste numa breve introdução onde se pretende identificar e elucidar a relevância deste trabalho; o segundo e o terceiro capítulos destinam-se, principalmente, à exposição de conceitos e resultados já conhecidos na literatura de modo a facilitar a compreensão dos resultados obtidos no quarto capítulo. Por último, o quinto apresenta uma breve conclusão das ideias principais referidas ao longo dos capítulos.

De forma mais detalhada, no segundo capítulo são abordados conceitos e estruturas fundamentais da matemática no ramo da Álgebra Linear, nomeadamente tópicos da teoria de matrizes, tais como, valores e vetores próprios, raio espectral, norma vetorial, norma matricial, matrizes definidas positivas e as suas propriedades, etc. Nesta dissertação, o reconhecimento de teoremas que interligam estes conceitos são informações de grande interesse, devido ao extenso uso de matrizes simétricas definidas positivas para definir estruturas imprescindíveis na utilização de modelos estatísticos.

O terceiro capítulo é dedicado a metodologias estatísticas para medidas repetidas, nos quais são abordados modelos longitudinais e modelos conjuntos. Nos Modelos de Dados Longitudinais é apresentada uma visão geral das ideias subjacentes à análise de dados longitudinais gaussianos, nomeadamente o modelo linear geral com erros correlacionados e uma extensão destes, os modelos lineares gerais mistos. Os modelos lineares gerais mistos incorporam as três fontes de variação aleatória: efeitos aleatórios, correlação em série e erros de medida. A estimação dos parâmetros para estes modelos é baseada no método de máxima verosimilhança (ML) e no método de máxima verosimilhança restrita (REML). Estes parâmetros, de interesse direto, admitem características importantes para a realização de testes de hipóteses, nomeadamente, são estimadores consistentes e a sua distribuição assintótica é gaussiana multivariada.

Algumas estruturas de correlação tradicionais para o processo de medição, bem como as suas características, são abordadas no segundo subcapítulo, quer para estudos não balanceados, quer para estudos observacionais. É, também, apresentada uma ferramenta usual, intitulada de variograma, que nos permite selecionar a estrutura de correlação com base na evidência empírica.

Na subsecção de modelos conjuntos começamos por apresentar uma introdução e explicar

os conceitos básicos da análise de sobrevivência. Sucintamente, descrevemos também a sua utilidade bem como funções e parâmetros importantes neste tipo de abordagens. Além disso, se o interesse da inferência está na associação entre a variável resposta e o processo de sobrevivência, os dois devem ser modelados em conjunto, tendo em consideração parâmetros que representem a sua correlação. Neste sentido, é também apresentada uma revisão das principais famílias de modelos conjuntos, propostos na literatura, para o processo estocástico longitudinal e o tempo até ao evento de interesse.

O quarto capítulo envereda no estudo de um caso particular destes modelos conjuntos, denominado Modelo Gaussiano Transformado (7). Descrevemos quais os principais pressupostos por detrás deste modelo, e como estes podem ser utilizados de forma apurada para extrair informação relevante dos dados. O objetivo principal desta secção é o estudo de estruturas de correlação conjunta consoante as condições em que a utilização deste modelo é apropriado na prática estatística diária. Para isso, é realizado um estudo de simulação onde são ilustrados exemplos de estruturas de correlação propostas e estabelecidas as respetivas propriedades comparando-as com as características que possam existir em dados reais. Todos os conceitos e objetivos apresentados neste capítulo exigem noções básicas de álgebra linear.

Capítulo 2

Conceitos de Álgebra Linear

Esta secção é dedicada à exposição de alguns tópicos da teoria de matrizes considerados importantes para o desenvolvimento desta dissertação.

Inicialmente são apresentadas algumas definições gerais de Álgebra Linear, nomeadamente, noção de valores e vetores próprios, espectro e raio espectral de uma matriz, norma vetorial e norma matricial e as suas propriedades.

Além disso, abordam-se resultados específicos para matrizes simétricas definidas positivas, que serão necessários para o desenvolvimento de tópicos abordados posteriormente.

2.1 Valores e Vetores Próprios

Podemos definir os conceitos algébricos de valores e vetores próprios da seguinte forma (Thulasidas, 2021(32)):

Definição 2.1.1. *Sejam $\mathbf{A} \in \mathbb{R}^{n \times n}$ e $\mathbf{s} \in \mathbb{R}^n \neq \mathbf{0}^n$. Diz-se que \mathbf{s} é um vetor próprio de \mathbf{A} se $\mathbf{A}\mathbf{s} = \lambda\mathbf{s}$ com valor próprio λ .*

Algumas notas sobre valores e vetores próprios:

- $\mathbf{A} \in \mathbb{R}^{n \times n}$ não implica que $\lambda \in \mathbb{R}$. Em outras palavras, podemos ter uma matriz quadrada \mathbf{A} que pertence ao corpo dos reais e os seus valores próprios não. Assim como nem todas as matrizes $\mathbf{A} \in \mathbb{R}^{n \times n}$ têm vetores próprios reais;
- Quando $\lambda = 0$ temos $\mathbf{A}\mathbf{s} = \mathbf{0}$, mas $\mathbf{s} \neq \mathbf{0}^n$;
- Se \mathbf{s} é um vetor próprio de \mathbf{A} associado ao valor próprio λ , então qualquer versão escalonada dele ($r\mathbf{s}$) está associado ao mesmo valor próprio λ ,

$$\mathbf{A}(r\mathbf{s}) = r\mathbf{A}\mathbf{s} = r\lambda\mathbf{s} = \lambda(r\mathbf{s}).$$

Assim, a partir das suas definições, pretende-se encontrar valores e vetores próprios de uma matriz $\mathbf{A} \in \mathbb{R}^{n \times n}$ tal que

$$\mathbf{A}\mathbf{s} = \lambda\mathbf{s} \implies (\mathbf{A} - \lambda\mathbf{I})\mathbf{s} = \mathbf{0}^n \underset{s \neq \mathbf{0}^n}{\implies} |\mathbf{A} - \lambda\mathbf{I}| = 0.$$

O desenvolvimento do determinante em $|\mathbf{A} - \lambda\mathbf{I}| = 0$ origina um polinómio de grau n em λ , a que chamamos **polinómio caraterístico de \mathbf{A}** , cujas raízes são os valores próprios de \mathbf{A} . Ao conjunto de todos os valores próprios de \mathbf{A} chamamos **espectro de \mathbf{A}** e representamos por $\sigma(\mathbf{A})$.

Tendo em conta o Teorema Fundamental da Álgebra, tem-se que toda a matriz quadrada de ordem n tem n valores próprios em \mathbb{C} (contando com as multiplicidades). A multiplicidade algébrica de λ é o número de vezes que λ aparece como raiz do polinómio caraterístico e a multiplicidade geométrica diz respeito à dimensão do subespaço vetorial $Nuc(\lambda\mathbf{I} - \mathbf{A})$.

Além disso, os valores e vetores próprios fornecem perceções sobre a estrutura da matriz e possuem propriedades relacionadas às propriedades da própria matriz. Estamos, agora, em condições de apresentar algumas dessas propriedades relevantes.

1. Sejam $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ e $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ os valores próprios de \mathbf{A} , então

$$\sum_{i=1}^n \lambda_i = \text{traço}(\mathbf{A}) = \sum_{i=1}^n a_{ii}.$$

2. Sejam $\mathbf{A} \in \mathbb{R}^{n \times n}$ e $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ os valores próprios de \mathbf{A} , então

$$\prod_{i=1}^n \lambda_i = |\mathbf{A}|.$$

3. Os valores próprios de uma matriz real simétrica são reais, isto é, se $\mathbf{A} \in \mathbb{R}^{n \times n}$ tal que $\mathbf{A}^T = \mathbf{A}$ então $\lambda_i \in \mathbb{R}$.
4. Os valores próprios de uma matriz real antissimétrica são complexos, isto é, se $\mathbf{A} \in \mathbb{R}^{n \times n}$ tal que $\mathbf{A}^T = -\mathbf{A}$ então $\lambda_i = \lambda'_i i$ onde $\lambda'_i \in \mathbb{R}$.
5. $\mathbf{A} + \alpha\mathbf{I}$ tem os mesmo vetores próprios de \mathbf{A} , com valores próprios $\lambda_i + \alpha$ onde λ_i são os valores próprios de \mathbf{A} .
6. \mathbf{A}^{-1} tem os mesmos vetores próprios de \mathbf{A} , com valores próprios iguais aos inversos dos valores próprios de \mathbf{A} , isto é, $\frac{1}{\lambda_i}$ onde λ_i são os valores próprios de \mathbf{A} . Pode ser visto ainda mais diretamente da seguinte forma:

$$\mathbf{A}\mathbf{s}_i = \lambda_i \mathbf{s}_i \Rightarrow \mathbf{s}_i = \lambda_i \mathbf{A}^{-1} \mathbf{s}_i \Rightarrow \mathbf{A}^{-1} \mathbf{s}_i = \frac{1}{\lambda_i} \mathbf{s}_i.$$

7. Sejam $\sigma \in \mathbb{R} \setminus \{0\}$ e $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ simétricas tais que $\mathbf{A} = \sigma^2 \mathbf{B}$. Os valores próprios da matriz \mathbf{A} são os mesmos da matriz \mathbf{B} multiplicando por σ^2 .

Definição 2.1.2. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$. O raio espectral de \mathbf{A} , denotado por $\rho(\mathbf{A})$, é definido pela igualdade $\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|$.*

Teorema 2.1.3. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica e $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}^n\}$, então*

$$\lambda_1 = \min \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \text{ e } \lambda_n = \max \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad (2.1)$$

onde os valores próprios são ordenados tais que $\lambda_1 \leq \dots \leq \lambda_n$.

A razão $\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$, $\mathbf{x} \neq \mathbf{0}^n$ é chamada de **quociente de Rayleigh**.

2.2 Norma Vetorial e Norma Matricial

Apresentam-se, de seguida, os conceitos de norma vetorial e norma matricial bem como as suas propriedades para matrizes simétricas definidas positivas e a sua relação com valores próprios dessas matrizes (Chong e Żak, 2004(3)).

Definição 2.2.1. (Norma Vetorial)

Uma norma, ou norma vetorial, em \mathbb{R}^n é uma aplicação $\|\cdot\|$ que associa a cada elemento $\mathbf{x} \in \mathbb{R}^n$ um número real $\|\mathbf{x}\|$, que possui as seguintes propriedades:

1. para todo $\mathbf{x} \in \mathbb{R}^n$, tem-se que $\|\mathbf{x}\| \geq 0$ com $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$ (positividade);
2. para todo $\mathbf{x} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ tem-se que $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ (homogeneidade);
3. para todo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ tem-se que $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (desigualdade triangular).

A norma de um vetor não difere, por definição, do módulo de um número real na medida em que representa o comprimento desse vetor.

Normas usuais para vetores no espaço vetorial real \mathbb{R}^n

- **Norma do Máximo:**

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

- **Norma-r:**

$$\|\mathbf{x}\|_r = \left(\sum_{i=1}^n |x_i|^r \right)^{\frac{1}{r}}.$$

- O caso particular em que $r=1$ corresponde à **Norma de Manhattan:**

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|.$$

- O caso particular em que $r=2$ corresponde à **Norma Euclidiana:**

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^n |x_i|^2}.$$

Definição 2.2.2. (Normas Equivalentes)

Considere o espaço vetorial real \mathbb{R}^n . Dizemos que as normas vetoriais $\|\cdot\|_\alpha$ e $\|\cdot\|_\beta$ em \mathbb{R}^n são equivalentes se existem constantes positivas m e M tais que:

$$m\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq M\|\mathbf{x}\|_\alpha, \text{ para todo } \mathbf{x} \in \mathbb{R}^n.$$

Em particular, obtém-se que $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$ e $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$ e, portanto,

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2 \leq n\|\mathbf{x}\|_\infty.$$

O conceito de norma vetorial pode também ser ampliado para matrizes, no sentido de se poder “medir” matrizes.

Uma norma matricial é uma aplicação $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ que satisfaz as mesmas propriedades que as normas vetoriais definidas em 2.2.1.

Definição 2.2.3. (Norma Matricial Induzida)

Uma norma matricial induzida é uma norma $\|\cdot\|_{ab} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ definida como

$$\|\mathbf{A}\|_{ab} = \max \left\{ \|\mathbf{A}\mathbf{x}\|_a : \mathbf{x} \in \mathbb{R}^n \text{ com } \|\mathbf{x}\|_b = 1 \right\},$$

onde $\|\cdot\|_a$ é uma norma vetorial em \mathbb{R}^m e $\|\cdot\|_b$ é uma norma vetorial em \mathbb{R}^n .

Seja $\mathbf{A} \in \mathbb{R}^{m \times n}$. Quando a mesma norma vetorial r é usada em ambos os espaços, escrevemos $\|\mathbf{A}\|_r$ para representar a norma matricial induzida por uma norma vetorial r . Aqui estão alguns exemplos dessas normas matriciais:

- **Norma Matricial induzida pela Norma de Manhattan**

$$\|\mathbf{A}\|_1 = \max_j \left(\sum_{i=1}^m |a_{ij}| \right) \quad (\text{maior soma absoluta das colunas}).$$

- **Norma Matricial induzida pela Norma do Máximo**

$$\|\mathbf{A}\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}| \right) \quad (\text{maior soma absoluta das linhas}).$$

- **Norma Matricial induzida pela Norma Euclidiana**

$$\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^T \mathbf{A})}.$$

Observe-se que nem todas as normas matriciais são normas induzidas. Um exemplo é a norma de Frobenius dada por:

- **Norma de Frobenius:**

$$\|\mathbf{A}\|_F = \sqrt{\text{traço}(\mathbf{A}^T \mathbf{A})} = \sqrt{\sum_{i,j=1}^n a_{ij}^2}.$$

Para qualquer norma induzida temos $\|\mathbf{I}\|_* = 1$, mas $\|\mathbf{I}\|_F = \sqrt{n}$.

Lema 2.2.4. *Toda a norma induzida é submultiplicativa, ou seja,*

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$$

Designa-se o **Número de Condição** de uma matriz \mathbf{A} invertível, $\text{Cond}(\mathbf{A})$, o produto da norma de \mathbf{A} pela norma da sua inversa. Em particular, para qualquer norma induzida tem-se,

$$1 = \|\mathbf{I}\|_* \leq \|\mathbf{AA}^{-1}\|_* \leq \|\mathbf{A}\|_* \|\mathbf{A}^{-1}\|_* = \text{Cond}(\mathbf{A}).$$

Além disso, é possível definir algumas normas matriciais como sendo uma função que depende de valores próprios.

Teorema 2.2.5. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica com valores próprios $\lambda_1, \dots, \lambda_n$. Então,*

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \lambda_i^2}.$$

Teorema 2.2.6. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica com valores próprios $\lambda_1, \dots, \lambda_n$. Então,*

$$\|\mathbf{A}\|_2 = \lambda_{max} = \max \{|\lambda_j|, j = 1, \dots, n\} = \rho(\mathbf{A}).$$

2.3 Matrizes Simétricas Definidas Positivas

Definição 2.3.1. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica ($\mathbf{A} = \mathbf{A}^T$). Dizemos que \mathbf{A} é **definida positiva** se $\forall \mathbf{z} \in \mathbb{R}^n \setminus \{\mathbf{0}^n\}$, $\mathbf{z}^T \mathbf{A} \mathbf{z} > 0$.*

Teorema 2.3.2. *Se $\mathbf{A} \in \mathbb{R}^{n \times n}$ é definida positiva, então \mathbf{A} é invertível e a sua inversa é também definida positiva.*

Teorema 2.3.3. *Se $\mathbf{A} \in \mathbb{R}^{n \times n}$ é definida positiva, então as submatrizes principais $\mathbf{A}_k, k = 1, \dots, n$ são também definidas positivas.*

Teorema 2.3.4. *Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica. As seguintes afirmações são equivalentes*

1. \mathbf{A} é definida positiva;
2. Os valores próprios de \mathbf{A} são estritamente positivos;
3. Os valores próprios de todas as submatrizes principais \mathbf{A}_k são positivos;
4. Existe uma matriz $\mathbf{R} \in M_{n \times n}(\mathbb{R})$ não singular tal que $\mathbf{A} = \mathbf{R}^T \mathbf{R}$.

Seja $\mathbf{A} \in \mathbb{R}^{n \times n}$. Um resultado importante para a compreensão de uma futura conclusão é, por meio do Teorema 2.3.4 (1. \leftrightarrow 2.) e pela propriedade 6,

$$\text{se } \lambda_1 \text{ é o menor valor próprio de } \mathbf{A} \text{ então } \lambda_1 = \frac{1}{\rho(\mathbf{A}^{-1})}.$$

Para além disso, através do Teorema 2.2.6, obtém-se

$$\lambda_1 = \frac{1}{\|\mathbf{A}^{-1}\|_2},$$

onde $\|\mathbf{A}^{-1}\|_2$ é a norma matricial induzida pela norma euclidiana.

Capítulo 3

Metodologia Estatística para Dados Longitudinais

Na tentativa de sistematizar ou formalizar alguns aspetos do mundo real, constroem-se **modelos estatísticos** que para além de uma parte determinística, incorporam uma componente de variação aleatória que pode ser intrínseca ao processo real ou uma consequência de medições imperfeitas. Podemos ainda aumentar a complexidade do modelo à medida que entendemos mais sobre o processo.

Apesar da grande evolução sentida nos últimos anos, nas áreas da Estatística e da Ciência de Dados, ainda não é possível obtermos modelos totalmente perfeitos, isto é, modelos que representam ou explicam 100 % da realidade e, portanto, o primeiro passo para a seleção de um modelo estatístico adequado deve ter como base conhecer o contexto e as questões de interesse do problema. Neste sentido, antes da formalização usual utilizada em modelos para análise de dados longitudinais, é apresentada uma breve introdução da definição de dados longitudinais, as principais características e os objetivos destes estudos.

Estamos perante um conjunto de dados proveniente de um estudo longitudinal quando estes são gerados por medidas repetidas de uma ou mais variáveis resposta ao longo do tempo para cada indivíduo, num ou mais grupos em estudo. Um exemplo bastante simplista é o de seleccionar vários indivíduos com a mesma idade, e medir altura (e/ou peso) numa sequência de tempos. Deste modo, para cada unidade amostral não se retira uma única medida Y_{ij} , mas uma sequência, Y_i , de medidas e, portanto, quando falamos de replicação referimo-nos ao número de indivíduos e não ao número de medições individuais.

O objetivo principal de um estudo longitudinal é caracterizar as mudanças da variável resposta com o tempo. Um dos principais focos desse tipo de estudo é identificar alterações dentro de um indivíduo, mas também é importante determinar se essas alterações estão relacionadas a outros fatores previamente escolhidos, chamados de covariáveis, que não o

tempo (Cabral e Gonçalves, 2011 (2)). Em outras palavras, um estudo longitudinal visa entender como uma variável resposta muda ao longo do tempo e se essas alterações estão relacionadas a outros fatores selecionados. Em alguns casos, pode acontecer de o valor das covariáveis se alterar ao longo do tempo.

Por um lado, em investigações envolvendo este tipo de dados, dispomos de um dado número de réplicas independentes do mesmo processo, ou seja, assumimos independência entre os indivíduos e, por outro, em cada indivíduo existe uma ordem implícita nos tempos das suas observações que sujeita o modelo a conter uma componente de correlação “serial” entre as observações. As medições repetidas de cada vetor resposta tendem a estar correlacionadas e assim, a modelação da estrutura de autocorrelação desempenha um papel fundamental na análise de dados longitudinais, ou seja, esta estrutura tem uma forte influência na estimação dos parâmetros do modelo a ajustar aos dados. Para além disso, estudos longitudinais têm a capacidade de distinguir a variação de \mathbf{Y} ao longo do tempo para um indivíduo da variação de \mathbf{Y} entre indivíduos, o que, futuramente, denominaremos como efeitos aleatórios.

A análise de dados longitudinais combina técnicas de análise multivariada e análise de séries temporais. Estes métodos, em termos de desenho do estudo, são considerados flexíveis no sentido que não obrigam a que todos os indivíduos tenham os dados recolhidos nos mesmos tempos e o mesmo número de observações.

Dizemos que o estudo foi desenhado para ser **balanceado** quando o esquema de recolha de dados determina que todas as unidades amostrais devem ser observadas nos mesmos tempos, os quais podem ou não ser igualmente espaçados. Se a planificação do desenho determinar que conjuntos diferentes de unidades amostrais sejam observados em conjuntos diferentes de tempos ou permitir que os dados sejam medidos irregularmente ao longo do tempo, ele é dito **não balanceado** (Singer et al., 2018 (29)).

O que acontece algumas vezes, na prática, é que, apesar de o estudo ser desenhado para ser balanceado, não se consegue que todos os indivíduos tenham o mesmo número de observações e de que estas tenham sido medidas nas mesmas ocasiões, o que leva à existência de valores omissos (*missing data*), pelo que a base de dados gerada transforma-se, assim, numa base de dados não balanceada. Além disso, estes casos originam implicações na matriz de desenho uma vez que os indivíduos vão ter um número de linhas diferente e, portanto, devem-se encontrar métodos de análise capazes de trabalhar com dados não equilibrados sem retirarmos da análise os indivíduos com dados omissos. Deste modo, é importante ter em consideração as diferentes razões que levaram à omissão e como podem estar associados ao processo de medidas repetidas.

Mecanismo de omissão de dados é o modelo que descreve a probabilidade com que a resposta é ou não observada num determinado momento (Little e Rubin, 1987 (23)). Na

literatura existem importantes distinções entre diferentes mecanismos de omissão de dados, os quais podem ser classificados em:

- Omissão completamente aleatória (**MCAR** - *missing completely at random*);
- Omissão aleatória (**MAR** - *missing at random*);
- Omissão não aleatória (**MNAR** - *missing not at random*).

(Schafer e Graham, 2002 (25)) fornecem uma revisão da literatura com uma descrição mais detalhada destes modelos levantando questões que ainda permaneciam sem solução.

(Diggle, Heagerty, Liang e Zeger, 2002 (6)), propõem métodos para testar se os dados ausentes em ensaios longitudinais são de omissão não aleatória (MCAR).

(Hedden et al., 2009 (13)) para além de sugerirem vários métodos de análise para estes dados com perda de seguimento, expõem um estudo de simulação que compara o poder de alguns métodos quando a quantidade de dados ausentes e o mecanismo variam. Passemos então à descrição mais detalhada dos mecanismos de dados omissos.

O primeiro cenário de dados ausentes é o mais ideal, pois ocorre quando a probabilidade da resposta estar omissa, não está correlacionada com quaisquer variáveis resposta ou quaisquer variáveis de interesse, ou seja, as razões não estão relacionadas com o tópico do estudo. Alguns exemplos deste tipo de dados incluem-se nos chamados **dados faltantes na sequência temporal** (*intermittent missing*), como, por exemplo, o instrumento de medição não estar a funcionar no momento e **perder indivíduos em determinado ponto do estudo** (*loss to follow-up*) como é o caso do paciente mudar de endereço e, por isso, deixar de fazer parte do estudo. Prova-se que num estudo longitudinal os dados MCAR não têm influência em termos de modelação e inferência, isto é, não há enviesamento dos resultados, o que não acontece nos próximos casos, que previamente sabemos que se não forem cuidadosamente tratados têm um grande impacto na análise.

Diz-se que se tem MAR quando a probabilidade da resposta estar omissa depende de valores de resposta ou de covariáveis observadas até ao momento da visita perdida, mas não daqueles que estão em falta (Schafer e Olsen, 1998 (26)). Para entender a suposição MAR, consideremos duas situações: 1) a ausência do valor resposta depende de uma covariável binária “depressão”, que diz respeito ao facto de um indivíduo já ter estado, ou não, numa fase de depressão, em que os indivíduos deprimidos são mais propensos a não completar o ensaio clínico do que os que nunca sofreram desta doença e, portanto, o valor ausente não é distribuído aleatoriamente por toda a amostra de indivíduos (Hedden et al., 2009 (13)); 2) Num ensaio clínico sobre perda de peso em que o indivíduo, em consulta, descobre que não conseguiu perder peso e acaba por abandonar o ensaio, este abandono está relacionado com

os valores observados da resposta (Cabral e Gonçalves, 2011 (2)). Na prática, este tipo de dados são muito comuns e, ao contrário do que acontece com os MCAR, se na distribuição conjunta da resposta \mathbf{Y}_i não for corretamente modelado o mecanismo de omissão, alguns métodos de análise de dados longitudinal não produzem estimativas válidas.

Por último, diz-se que se tem MNAR quando a probabilidade da resposta estar omissa depende de qualquer resultado não observado devido a visitas perdidas. Neste caso, os dados ausentes podem ser dependentes de qualquer um dos resultados não observados. Por exemplo, num estudo de prevenção de recaídas, em que um participante ao momento em que ocorreu a primeira visita perdida tenha exames negativos de drogas na urina. Se for conhecido que o motivo da visita perdida foi porque o participante sabia que a sua triagem de drogas na urina seria positiva, a visita perdida seria classificada como MNAR (Hedden et al., 2009 (13)).

(Sousa, 2011 (30)) fornece uma revisão de modelos longitudinais que lidam com dados omissos potencialmente informativos. Em particular, considera modelos onde o evento é o abandono do estudo.

Estudar a natureza da variável resposta é fundamental para a decisão do tipo de modelos a utilizar. Em geral, para dados longitudinais contínuos, os modelos paramétricos utilizados na sua análise estão baseados no pressuposto da distribuição gaussiana - Dados Longitudinais Gaussianos. Nos modelos paramétricos utilizados na análise de dados longitudinais discretos, o pressuposto da distribuição gaussiana já não é considerado - Dados Longitudinais Não Gaussianos (Cabral e Gonçalves, 2011 (2)).

3.1 Modelos para Dados Longitudinais

Na notação estabelecida em seguida, representamos parâmetros por letras gregas (por exemplo, $\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}$) e quantidades observadas por letras romanas (por exemplo, $\mathbf{y}, \mathbf{x}, \mathbf{t}$).

Seja y_{ij} o valor da resposta para o indivíduo $i = 1, \dots, m$ no ponto de tempo t_{ij} onde $j = 1, \dots, n_i$, realização da variável aleatória Y_{ij} . Designamos o vetor das variáveis resposta do indivíduo i por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$.

Os modelos para dados longitudinais incorporam duas componentes de modelação: a modelação para a média da população, usualmente de regressão, e a modelação da estrutura de variância-covariância presente nos dados. O valor esperado e a variância populacionais de \mathbf{Y}_i são, respetivamente,

$$E[\mathbf{Y}_i] = \boldsymbol{\mu}_i \quad (n_i \times 1) \quad \text{e} \quad Var[\mathbf{Y}_i] = \mathbf{V}_i \quad (n_i \times n_i),$$

onde o elemento (j, k) da matriz \mathbf{V}_i representa a covariância entre Y_{ij} e Y_{ik} para $j \neq k$ e a

variância de Y_{ij} , para $j = k$. O perfil do indivíduo i é a sequência de medições efetuadas $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$.

Denote-se por \mathbf{R}_i ($n_i \times n_i$) a matriz de correlação de \mathbf{Y}_i , cujos elementos da diagonal principal são 1 e para o elemento da posição (j, k) onde $j \neq k$ representa o parâmetro de correlação entre Y_{ij} e Y_{ik} .

Inclui-se também um conjunto de p variáveis explicativas dadas pelo vetor

$$\mathbf{x}_{ij} = (x_{ij}^1, \dots, x_{ij}^p)^T,$$

com dimensão p , que podem ser dependentes do tempo ou medidas apenas na linha de base, sendo

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}^T \\ \mathbf{x}_{i2}^T \\ \vdots \\ \mathbf{x}_{in_i}^T \end{bmatrix} = \begin{bmatrix} x_{i1}^1 & x_{i1}^2 & \dots & x_{i1}^p \\ x_{i2}^1 & x_{i2}^2 & \dots & x_{i2}^p \\ \vdots & \vdots & \ddots & \vdots \\ x_{in_i}^1 & x_{in_i}^2 & \dots & x_{in_i}^p \end{bmatrix} \quad (n_i \times p)$$

a matriz de covariáveis associada a \mathbf{Y}_i onde o elemento x_{jp} é o valor na p -ésima covariável explicativa do indivíduo i no tempo t_{ij} . Note-se que esta matriz corresponde a apenas um indivíduo pelo que, é possível ter-se um número de linhas diferente em cada i . Além disso, usualmente, para todo i e quaisquer j , $x_{ij}^1 = 1$.

A matriz que representa o valor das covariáveis de todos os indivíduos em todos os tempos é dada por

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \quad (N \times p),$$

onde $N = \sum_{i=1}^m n_i$ é o número de observações total.

O conjunto de todas as variáveis resposta é dado pelo vetor

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1^T \\ \mathbf{Y}_2^T \\ \vdots \\ \mathbf{Y}_m^T \end{bmatrix} \quad (N \times 1).$$

3.1.1 Modelo Linear Geral

A abordagem mais comum para medições repetidas é baseada em modelos de regressão, como é o caso do **Modelo Linear Geral com erros correlacionados**, em que admite os mecanismos de omissão de dados MCAR e MAR. Este modelo pressupõe independência entre os indivíduos, onde cada medição é a realização de uma variável aleatória gaussiana, ou seja,

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \epsilon_{ij}, \quad (3.1)$$

onde $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})$ é um vetor de dimensão p de coeficientes de regressão a serem estimados e ϵ_{ij} é uma variável aleatória gaussiana com média zero e representa o quanto distantes os valores da resposta estão da média do modelo, $\mathbf{x}_{ij}^T \boldsymbol{\beta}$.

Em estudos transversais ϵ_{ij} seriam variáveis aleatórias mutuamente independentes, o que não se espera em estudos longitudinais onde ϵ_{ij} de um mesmo indivíduo estão correlacionados, isto é,

- ϵ_{ij} de indivíduos diferentes vamos assumir como independentes:

$$\text{Cov}(\epsilon_{ij}, \epsilon_{lk}) = 0 \quad i \neq l,$$

- ϵ_{ij} de um mesmo indivíduo podem estar correlacionados:

$$\text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \theta_{jk} \quad j \neq k,$$

$$\text{Cov}(\epsilon_{ij}, \epsilon_{ij}) = \text{Var}[\epsilon_{ij}] = \sigma_j^2.$$

Para a análise destes modelos utilizamos notação matricial. Seja, então, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ o vetor aleatório gaussiano multivariado para o indivíduo i cuja média segue um modelo linear e os erros não são independentes,

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \text{MVN}(\mathbf{0}^{n_i}, \mathbf{V}_i(\boldsymbol{\theta})), \quad (3.2)$$

onde $E[\mathbf{Y}_i] = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = \mathbf{X}_i \boldsymbol{\beta}$ é o vetor das médias em todos os tempos do indivíduo i , $\text{Var}[\mathbf{Y}_i] = \mathbf{V}_i(\boldsymbol{\theta})$ é a matriz de de variância-covariância entre todas as observações do indivíduo i e $\boldsymbol{\theta}$ corresponde ao vetor de parâmetros que estão envolvidos nessa matriz. Note-se que $\boldsymbol{\theta}$ não está indexado a i .

Juntando os modelos especificados para todos os indivíduos descrito em 3.2,

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}^N, \mathbf{V}(\boldsymbol{\theta})), \quad (3.3)$$

onde

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \quad \text{e} \quad \text{Var}[\mathbf{Y}] = \mathbf{V}(\boldsymbol{\theta}).$$

Dada a independência entre indivíduos, $\mathbf{V}(\boldsymbol{\theta})$ é uma matriz diagonal de blocos ($N \times N$),

$$\mathbf{V}(\boldsymbol{\theta}) = \begin{bmatrix} \boxed{\mathbf{V}_1(\boldsymbol{\theta})} & 0 & 0 & 0 \\ 0 & \boxed{\mathbf{V}_2(\boldsymbol{\theta})} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boxed{\mathbf{V}_m(\boldsymbol{\theta})} \end{bmatrix}.$$

Deste modo, a distribuição da variável aleatória \mathbf{Y} é uma distribuição multivariada Gaussiana dada por

$$\mathbf{Y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

Consideremos que os erros são homocedásticos, com variância constante ao longo do tempo igual a σ^2 . Deste modo, podemos simplificar a matriz $\mathbf{V}(\boldsymbol{\theta}) = \sigma^2 \mathbf{R}(\boldsymbol{\phi})$ onde a matriz $\mathbf{R}(\boldsymbol{\phi})$ representa a matriz de correlação de \mathbf{Y} definida em cima e $\boldsymbol{\theta}^T = (\sigma^2, \boldsymbol{\phi}^T)$.

Estimadores dos Mínimos Quadrados Ponderados

Em estatística, o método dos mínimos quadrados ordinários (OLS), baseado em modelos de regressão linear, ou seja, os efeitos fixos são estimados através de uma função linear que depende do conjunto de variáveis explicativas, é um método utilizado para estimar os parâmetros desconhecidos de um modelo, pelo princípio de minimizar a soma das distâncias verticais quadradas entre as respostas observadas no conjunto de dados e as respostas previstas pela aproximação linear.

Consideremos, então, o problema de estimar o parâmetro de regressão $\boldsymbol{\beta}$. O estimador de mínimos quadrados ponderado (*weighted least squares*) de $\boldsymbol{\beta}$, usando uma matriz \mathbf{W} de pesos simétrica ($\mathbf{W}^T = \mathbf{W}$), é o vetor $\hat{\boldsymbol{\beta}}_W$ que minimiza a forma quadrática

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Ora,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) &= 0 \\ \Leftrightarrow \mathbf{X}^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= 0 \\ \Leftrightarrow \hat{\boldsymbol{\beta}}_W &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}. \end{aligned} \tag{3.4}$$

Note-se que para uma qualquer matriz de pesos \mathbf{W} , mesmo que não seja a correta, este estimador é não enviesado.

$$\begin{aligned} E[\hat{\beta}_W] &= E[(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} E[\mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{X}) \boldsymbol{\beta} = \boldsymbol{\beta}, \quad \forall \mathbf{W}. \end{aligned}$$

A eficiência relativa a $\hat{\beta}_W$ pode ser calculada a partir da respetiva matriz de variância dada por,

$$Var[\hat{\beta}_W] = \sigma^2 \left\{ (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \right\} \mathbf{R}(\phi) \left\{ \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \right\}.$$

Em particular, obtêm-se os seguintes casos especiais:

1. Se $\mathbf{W} = \mathbf{I}$ obtemos o estimador dos mínimos quadrados

$$\begin{aligned} \hat{\beta}_I &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ Var[\hat{\beta}_I] &= \sigma^2 \left\{ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right\} \mathbf{R}(\phi) \left\{ \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \right\}. \end{aligned} \quad (3.5)$$

Um uso ingênuo de OLS seria ignorar a estrutura de correlação dos dados e basear a estimativa do intervalo para $\hat{\beta}_I$ na fórmula da variância dada em 3.5 substituindo σ^2 pelo seu estimador usual, (Diggle et al.,2002(6))

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X} \hat{\beta}_I)^T (\mathbf{y} - \mathbf{X} \hat{\beta}_I)}{N - p}.$$

2. Se $\mathbf{W} = \mathbf{R}^{-1}(\phi)$ obtemos o estimador de máxima verosimilhança (este caso será analisado, posteriormente, com mais detalhe)

$$\begin{aligned} \hat{\beta}_R &= (\mathbf{X}^T \mathbf{R}^{-1}(\phi) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1}(\phi) \mathbf{y}, \\ Var[\hat{\beta}_R] &= (\mathbf{X}^T \mathbf{V}^{-1}(\rho) \mathbf{X})^{-1}. \end{aligned}$$

Este é o estimador de mínimos quadrados ponderado mais eficiente.

Note-se, a partir da equação 3.4, que os valores dos estimadores pontuais são influenciados pela matriz de pesos \mathbf{W} - mas não precisamos de conhecer σ^2 , pois $\hat{\beta}$ não se altera por mudanças proporcionais em todos os elementos de \mathbf{W} . Contudo, para identificar a matriz

ideal \mathbf{W} é necessário conhecer a estrutura de correlação completa que é imposta no modelo, $\mathbf{R}(\boldsymbol{\phi})$. É interessante questionarmos quanta perda de eficiência pode resultar do uso de uma estrutura de correlação errada relativamente à verdadeira estrutura de correlação dos dados. Diferentes estruturas de correlação resultam em estimativas de parâmetros diferentes e podem levar a diferentes conclusões acerca das relações entre as variáveis e, por isso, impor ao modelo uma estrutura de correlação adequada é fundamental. A escolha do tipo de correlação pode ser baseada na estrutura de correlação dos dados observados para a variável resposta. Para isto, é necessário fazer uma boa análise exploratória da estrutura de correlação dos dados, como mostraremos na secção 3.2.

Deste modo, quando não queremos impor nenhuma estrutura específica ao modelo, uma alternativa é a utilização da matriz empírica de correlação dos resíduos (após uma regressão linear assumindo independência), como estimador da verdadeira matriz \mathbf{R} . Desta forma, obtemos estimadores consistentes para os erros padrão dos estimadores pontuais. Ao estimador $\hat{\boldsymbol{\beta}}_{\mathbf{W}}$ referido na equação 3.4 onde \mathbf{W} é a matriz de correlações empírica dos resíduos chamamos **Estimadores Robustos**.

Estimadores de máxima verosimilhança - ML

Outra estratégia para a estimação destes parâmetros é considerar a estimativa simultânea dos parâmetros de interesse, nomeadamente, os parâmetros da média $\boldsymbol{\beta}$, o parâmetro da variância σ^2 e os parâmetros das correlações entre as observações dos tempos dados em $\boldsymbol{\phi}$, usufruindo da função de verosimilhança (*Likelihood function*) da amostra, representada por $L_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi} \mid \mathbf{Y})$, caracterizada através da função de densidade de probabilidade do vetor \mathbf{Y} , tomada como função dos parâmetros do modelo assumido.

O que se pretende é encontrar os valores dos parâmetros $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}$ que maximizam esta probabilidade de ocorrência. Às estatísticas $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{ML}(\mathbf{Y})$, $\hat{\sigma}_{ML}^2 = \hat{\sigma}_{ML}^2(\mathbf{Y})$ e $\hat{\boldsymbol{\rho}}_{ML} = \hat{\boldsymbol{\rho}}_{ML}(\mathbf{Y})$ chamamos **estimadores de máxima verosimilhança** (ou estimador ML como abreviatura de *maximum likelihood*). Este método intitula-se, naturalmente, por **método da Verosimilhança Máxima**.

A função de verosimilhança para dados gaussianos baseado no vetor de \mathbf{Y} , assumindo erros correlacionados, é dado por

$$L_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi} \mid \mathbf{Y}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{V}(\boldsymbol{\theta})|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}.$$

Uma vez que a função logarítmica é uma função crescente, maximizar a função de verosimilhança, $L_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})$, é equivalente a maximizar o logaritmo da função de verosimilhança (log-verosimilhança), $l_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) = \log(L_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi} \mid \mathbf{Y}))$. Na prática, maximizar

$$\begin{aligned}
l_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{V}(\boldsymbol{\theta})|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{1}{2} \left\{ N \log(2\pi\sigma^2) + \log(|\mathbf{R}(\boldsymbol{\phi})|) + \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{R}^{-1}(\boldsymbol{\phi}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}. \quad (3.6)
\end{aligned}$$

Se derivar a função log-verosimilhança em ordem a $\boldsymbol{\beta}$ obtém-se o estimador de máxima verosimilhança (*maximum likelihood estimator*) para $\boldsymbol{\beta}$ dado por

$$\begin{aligned}
\frac{\partial l_{ML}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})}{\partial \boldsymbol{\beta}} = 0 &\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{R}^{-1}(\boldsymbol{\phi}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) = 0 \\
&\Leftrightarrow \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \\
&\Leftrightarrow \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{y} = 0 \\
&\Leftrightarrow \hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\phi}) = (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{y}. \quad (3.7)
\end{aligned}$$

Como já foi observado, num caso particular dos estimadores dos mínimos quadrados, este é um estimador centrado com matriz de variância dada por,

$$\begin{aligned}
Var[\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\phi})] &= \sigma^2 \left\{ (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \right\} \mathbf{R}(\boldsymbol{\phi}) \left\{ \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X} (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1} \right\} \\
&= \sigma^2 (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1}. \quad (3.8)
\end{aligned}$$

Além disso, deve-se ressaltar o facto de que mesmo que não estejamos sob o caso homocedástico, $\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\phi}) = (\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{y}$ continua a ser um estimador centrado com variância dada por 3.8.

Voltando à função 3.6, $l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})$, e substituindo $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})$ obtém-se

$$l_{ML}^*(\sigma^2, \boldsymbol{\phi}) = -\frac{1}{2} \left\{ N \log(2\pi\sigma^2) + \log(|\mathbf{R}(\boldsymbol{\phi})|) + \frac{1}{\sigma^2} RSS(\boldsymbol{\phi}) \right\} \quad (3.9)$$

onde

$$RSS_{ML}(\boldsymbol{\phi}) = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\phi}))^T \mathbf{R}^{-1}(\boldsymbol{\phi}) (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\phi})).$$

Note-se que RSS apenas depende de parâmetros $\boldsymbol{\phi}$ e não depende de σ^2 . Assim, derivando a função 3.9 em ordem a σ^2 e fixando novamente a estrutura de $\mathbf{R}(\boldsymbol{\phi})$, obtemos o estimador de máxima verosimilhança para σ^2 dado por

$$\frac{\partial l_{ML}^*(\sigma^2, \boldsymbol{\phi})}{\partial \sigma^2} = 0 \Leftrightarrow \frac{\partial}{\partial \sigma^2} \left(-\frac{1}{2} \left\{ N \log(2\pi\sigma^2) + \frac{1}{\sigma^2} RSS(\boldsymbol{\phi}) \right\} \right) = 0$$

$$\Leftrightarrow \hat{\sigma}_{ML}^2(\boldsymbol{\phi}) = \frac{RSS(\boldsymbol{\phi})}{N}. \quad (3.10)$$

Substituindo σ^2 por $\hat{\sigma}_{ML}^2(\boldsymbol{\phi})$ na função $l_{ML}^*(\sigma^2, \boldsymbol{\phi})$ ficamos com a expressão reduzida da função log-verosimilhança que depende apenas do conjunto de parâmetros $\boldsymbol{\phi}$ assumindo uma estrutura específica em \mathbf{R} .

$$\begin{aligned} l_{ML}^{**}(\boldsymbol{\phi}) &= -\frac{1}{2} \left\{ N \log \left(\frac{2\pi}{N} RSS(\boldsymbol{\phi}) \right) + \log(|\mathbf{R}(\boldsymbol{\phi})|) + N \right\} \\ &\propto -\frac{1}{2} \left\{ N \log(RSS(\boldsymbol{\phi})) + \log(|\mathbf{R}(\boldsymbol{\phi})|) \right\}. \end{aligned} \quad (3.11)$$

Geralmente, assumindo uma estrutura de correlação para \mathbf{R} , o cálculo da maximização da função 3.11 em ordem a $\boldsymbol{\phi}$ é feito recorrendo a métodos numéricos de otimização. Desta forma, são obtidos os estimadores ML de $\boldsymbol{\phi}$ - $\hat{\boldsymbol{\phi}}_{ML}$.

Os estimadores de máxima verosimilhança para $\boldsymbol{\beta}$ e σ^2 são obtidos substituindo $\boldsymbol{\phi}$ em 3.7 e 3.10 pelo seu estimador ML $\hat{\boldsymbol{\phi}}_{ML}$, donde os estimadores de $\boldsymbol{\beta}$ e σ^2 são estimadores consistentes dados por

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{ML} &= (\mathbf{X}^T \hat{\mathbf{R}}_{ML}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{R}}_{ML}^{-1} \mathbf{y}, \\ \hat{\sigma}_{ML}^2 &= \frac{RSS(\hat{\boldsymbol{\phi}}_{ML})}{N}, \end{aligned}$$

onde $\hat{\mathbf{R}}_{ML} = \hat{\mathbf{R}}(\hat{\boldsymbol{\theta}}_{ML})$.

O estimador ML de $\boldsymbol{\beta}$ é assintoticamente não correlacionado com os estimadores ML de σ^2 e $\boldsymbol{\phi}$ e, portanto, para efeitos de inferência, obtém-se as seguintes distribuições assintóticas (2):

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{ML} &\sim MVN\left(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1}\right), \\ \begin{bmatrix} \hat{\boldsymbol{\phi}}_{ML} \\ \log(\hat{\sigma}_{ML}) \end{bmatrix} &\sim MVN\left(\begin{bmatrix} \boldsymbol{\phi} \\ \log(\sigma) \end{bmatrix}, \mathbf{I}(\sigma, \boldsymbol{\phi})\right), \end{aligned} \quad (3.12)$$

onde $\mathbf{I}(\sigma, \boldsymbol{\phi})$ é a matriz empírica de informação de Fisher dada por:

$$\mathbf{I}(\sigma, \boldsymbol{\phi}) = \begin{bmatrix} \frac{\partial^2 l^*(\sigma^2, \boldsymbol{\phi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^T} & \frac{\partial^2 l^*(\sigma^2, \boldsymbol{\phi})}{\partial \log(\sigma) \partial \boldsymbol{\phi}^T} \\ \frac{\partial^2 l^*(\sigma^2, \boldsymbol{\phi})}{\partial \boldsymbol{\phi} \partial \log(\sigma)} & \frac{\partial^2 l^*(\sigma^2, \boldsymbol{\phi})}{\partial^2 \log(\sigma)} \end{bmatrix}.$$

Usualmente utiliza-se a função logaritmo quando se pretende encontrar uma aproximação à distribuição gaussiana.

Note-se também que a matriz de desenho \mathbf{X} é necessária para a estimação dos parâmetros σ^2 e ϕ , o que significa que, se a matriz \mathbf{X} tiver a forma incorreta, não teremos estimadores consistentes de σ^2 e ϕ . Uma estratégia seria assumir um modelo saturado (isto é, com todas as variáveis explicativas em \mathbf{X}), e as suas associações possíveis, para a estimação de σ^2 e de ϕ e, posteriormente, utilizar esses estimadores para a inferência sobre β . No entanto, em estudos longitudinais, as estimativas ML de parâmetros de variância tendem a ser enviesadas em amostras finitas, uma vez que o procedimento de estimação não leva em conta a perda de graus de liberdade que resulta da estimativa de parâmetros de regressão. Este viés pode ser bastante pesado em situações em que o número de indivíduos, m , é pequeno em comparação com o número de parâmetros de regressão (20).

Deste modo, com o intuito de reduzir o viés na estimativa de ϕ , Diggle et al. (6) apresenta outra abordagem para este tipo de modelos - **o método de máxima verosimilhança restrita (REML)**, embora seja menos importante quando a dimensionalidade de β é pequena.

Estimadores de máxima verosimilhança restrita - REML

A estratégia subjacente ao **Método da Máxima Verosimilhança Restrita** consiste na transformação linear de \mathbf{Y} dada por $\mathbf{U}^{N \times 1} = \mathbf{A}^T \mathbf{Y}$ onde a matriz \mathbf{A} é ortogonal às colunas da matriz \mathbf{X} e de característica máxima ($N \times (N - p)$). À semelhança do que acontece no método anterior, pretende-se é encontrar os valores dos parâmetros β, σ^2, ϕ que maximizam a função de verosimilhança não de \mathbf{Y} , mas agora em relação a \mathbf{U} , que segue uma distribuição multivariada gaussiana dada por:

$$\mathbf{U} \sim MVN(\mathbf{0}^N, \mathbf{A}^T \mathbf{V}(\theta) \mathbf{A}).$$

A principal diferença entre a expressão do logaritmo de verosimilhança, dado em 3.6, e o logaritmo da verosimilhança restrita reside no termo $-\frac{1}{2} \log(|\mathbf{X}^T \mathbf{V}^{-1}(\phi) \mathbf{X}|)$ (Harville, 1974 (11)), isto é,

$$\begin{aligned} l_{REML}(\beta, \sigma^2, \phi) &= \text{constante} - \frac{1}{2} \left\{ N \log(\sigma^2) + \log(|\mathbf{R}(\phi)|) + \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{R}^{-1}(\phi) (\mathbf{y} - \mathbf{X}\beta) \right\} \\ &\quad - \frac{1}{2} \log(|\mathbf{X}^T \mathbf{V}^{-1}(\phi) \mathbf{X}|) \\ &= \text{constante}' + l_{ML}(\beta, \sigma^2, \phi) - \frac{1}{2} \log(|\mathbf{X}^T \mathbf{V}^{-1}(\phi) \mathbf{X}|). \end{aligned}$$

Note-se que o novo termo não depende de β e, portanto, fixando a estrutura da matriz

\mathbf{R} , o estimador de máxima verosimilhança restrita para $\boldsymbol{\beta}$ é dado por

$$\hat{\boldsymbol{\beta}}_{REML}(\boldsymbol{\phi}) = (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{y}. \quad (3.13)$$

É importante salientar que, apesar da sua forma matemática ser a mesma que foi obtida para o estimador ML, estes não são iguais.

O estimador de máxima verosimilhança restrita para σ^2 é

$$\hat{\sigma}_{REML}^2(\boldsymbol{\phi}) = \frac{RSS(\boldsymbol{\phi})}{N - p}, \quad (3.14)$$

onde

$$RSS_{REML}(\boldsymbol{\phi}) = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{REML}(\boldsymbol{\phi}))^T \mathbf{R}^{-1}(\boldsymbol{\phi}) (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{REML}(\boldsymbol{\phi})).$$

Finalmente, o estimador REML de $\boldsymbol{\phi}$ maximiza

$$l^{**}_{REML}(\boldsymbol{\phi}) = -\frac{1}{2} \left\{ N \log(RSS(\boldsymbol{\phi})) + \log(|\mathbf{R}(\boldsymbol{\phi})|) + \log(|\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X}|) \right\}. \quad (3.15)$$

A partir da estrutura bloco-diagonal explícita de $\mathbf{R}(\boldsymbol{\phi})$ denotada, anteriormente, por $\mathbf{R}_i(\boldsymbol{\phi})$ (matriz de correlação ($n_i \times n_i$) entre todas as observações do indivíduo i medidas no vetor de tempos \mathbf{t}_i) a equação $l^{**}_{REML}(\boldsymbol{\phi})$, dada por 3.15, pode ser escrita como

$$l^{**}_{REML}(\boldsymbol{\phi}) = -\frac{1}{2} \left\{ N \log(RSS(\boldsymbol{\phi})) + \sum_{i=1}^m \log(|\mathbf{R}_i(\boldsymbol{\phi})|) + \sum_{i=1}^m \log(|\mathbf{X}_i^T \mathbf{R}_i^{-1}(\boldsymbol{\phi}) \mathbf{X}_i|) \right\}.$$

Observe que nesta equação estão envolvidos o cálculo de m determinantes e inversas de matrizes de ordens $n_i \times n_i$. Contudo, quando o desenho do estudo é balanceado, é natural que o número de matrizes distintas $\mathbf{R}_i(\boldsymbol{\phi})$ seja muito mais pequeno que m , isto é, vários indivíduos são medidos o mesmo número de vezes e nos mesmos tempos. Além disso, a dimensionalidade de $\boldsymbol{\phi}$ não excede, usualmente, quatro, o que não acarreta algoritmos de otimização muito sofisticados.

Analogamente ao caso anterior, os estimadores de máxima verosimilhança restrita $\hat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2$ são obtidos substituindo $\boldsymbol{\phi}$ em 3.13 e 3.14 pelo seu estimador REML $\hat{\boldsymbol{\phi}}$.

Para efeitos de inferência sobre $\boldsymbol{\beta}$ obtém-se

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\phi}) \sim MVN\left(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{R}^{-1}(\boldsymbol{\phi}) \mathbf{X})^{-1}\right).$$

É possível constatar-se que continuamos a ter boas aproximações para este estimador se

substituímos os valores desconhecidos pelas suas estimativas REML, isto é,

$$\hat{\boldsymbol{\beta}} \sim MVN\left(\boldsymbol{\beta}, \hat{\sigma}^2(\mathbf{X}^T \mathbf{R}^{-1}(\hat{\boldsymbol{\phi}})\mathbf{X})^{-1}\right).$$

Tal como acontece com os estimadores ML, (Pinheiro,2000(21)) prova, também, que os estimadores REML são consistentes e assintoticamente gaussianos. As características das distribuições que os descrevem são idênticas às apresentadas em 3.12 substituindo os estimadores ML pelos seus respetivos estimadores REML e $l_{ML}^*(\sigma^2, \boldsymbol{\phi})$ por $l_{REML}^*(\sigma^2, \boldsymbol{\phi})$. Além disso, prova-se que os estimadores das componentes da variância são independentes de \mathbf{A} , ou seja, seja qual for a matriz \mathbf{A} obtemos os mesmos estimadores.

3.1.2 Modelo Linear Misto

Os modelos lineares mistos para dados longitudinais Gaussianos são uma extensão dos modelos lineares e a sua característica básica é a inclusão de efeitos aleatórios no modelo de regressão. Deste modo, estes modelos combinam efeitos fixos e efeitos aleatórios para representar, com mais precisão, a estrutura de variância-covariância dos dados. Os efeitos fixos são parâmetros associados a toda a população que está a ser estudada, enquanto os efeitos aleatórios estão associados aos indivíduos selecionados aleatoriamente da população. Ao incluir efeitos fixos e aleatórios, os modelos lineares mistos permitem a descrição das mudanças dentro de cada indivíduo ao longo do tempo e fornecem uma representação mais flexível da estrutura de variância-covariância dos dados. Isso torna os modelos lineares mistos particularmente úteis para analisar dados longitudinais ou dados coletados durante um período de tempo dos mesmos indivíduos (2).

À semelhança com o modelo linear geral, este modelo também admite os mecanismos de omissão de dados MCAR e MAR.

Seja, novamente, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ o vetor aleatório gaussiano multivariado para o indivíduo $i = 1, \dots, m$. O modelo linear geral misto de efeitos aleatórios é definido por

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{D}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \mathbf{b}_i \sim MVN(\mathbf{0}^q, \mathbf{G}), \quad \boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}^{n_i}, \boldsymbol{\Sigma}_i) \quad (3.16)$$

onde:

- $\boldsymbol{\beta}$ é o vetor dos efeitos fixos ($p \times 1$);
- \mathbf{X}_i matriz de covariáveis dos efeitos fixos ($n_i \times p$);
- \mathbf{b}_i é o vetor dos efeitos aleatórios ($q \times 1$);

- D_i matriz de covariáveis dos efeitos aleatórios ($n_i \times q$);
- ϵ_i é o vetor dos erros aleatórios dentro do grupo ($n_i \times 1$);
- b_i e ϵ_i são independentes para diferentes indivíduos;
- b_i e ϵ_i são independentes entre si.

As matrizes \mathbf{G} e Σ_i são matrizes definidas positivas com dimensões $q \times q$ e $n_i \times n_i$, respetivamente. A média segue um modelo linear dada por $\mathbf{X}_i\boldsymbol{\beta}$ e a matriz de variância-covariância decompõe-se em $\mathbf{V}_i = \mathbf{D}_i\mathbf{G}\mathbf{D}_i^T + \Sigma_i$.

Juntando os modelos especificados para todos os indivíduos descrito em 3.16, obtém-se

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\mathbf{b} + \boldsymbol{\epsilon}, \quad \mathbf{b} \sim MVN(\mathbf{0}^N, \mathbf{G}), \quad \boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}^{n_i}, \Sigma_i)$$

onde

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \quad \text{e} \quad Var[\mathbf{Y}] = \mathbf{D}\mathbf{G}\mathbf{D}^T + \Sigma = \mathbf{V}.$$

Dada a independência entre indivíduos, \mathbf{D} ($N \times mq$), \mathbf{G} ($mq \times mq$) e Σ ($N \times N$) são matrizes diagonais de blocos e, portanto, $\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_m)$ é também uma matriz diagonal de blocos ($N \times N$) decomposta da seguinte maneira,

$$\mathbf{D}\mathbf{G}\mathbf{D}^T = \begin{bmatrix} \boxed{\mathbf{D}_1} & 0 & 0 & 0 \\ 0 & \boxed{\mathbf{D}_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boxed{\mathbf{D}_m} \end{bmatrix} \begin{bmatrix} \boxed{\mathbf{G}} & 0 & 0 & 0 \\ 0 & \boxed{\mathbf{G}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boxed{\mathbf{G}} \end{bmatrix} \begin{bmatrix} \boxed{\mathbf{D}_1^T} & 0 & 0 & 0 \\ 0 & \boxed{\mathbf{D}_2^T} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boxed{\mathbf{D}_m^T} \end{bmatrix}$$

$$+ \begin{bmatrix} \boxed{\Sigma_1} & 0 & 0 & 0 \\ 0 & \boxed{\Sigma_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boxed{\Sigma_m} \end{bmatrix}.$$

Deste modo, a distribuição da variável aleatória \mathbf{Y} é uma distribuição multivariada Gaussiana dada por

$$\mathbf{Y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \mathbf{D}\mathbf{G}\mathbf{D}^T + \Sigma).$$

Modelo Linear Misto básico

Em particular, quando o modelo 3.16 assume que $\Sigma_i = \tau^2 \mathbf{I}_i$ designa-se por **modelo linear misto básico** e a matriz de variância-covariância de \mathbf{Y}_i assume a forma de

$$\mathbf{V}_i = \mathbf{D}_i \mathbf{G} \mathbf{D}_i^T + \tau \mathbf{I}_i.$$

3.1.3 Modelo Linear Misto Geral

Tipicamente, podemos incluir nos modelos para dados longitudinais pelo menos três fontes qualitativamente diferentes de variação aleatória.

- **Efeitos Aleatórios**

Esta é uma componente de variabilidade entre indivíduos, isto é, são características individuais que influenciam todas as medidas do mesmo paciente. Talvez o exemplo mais simples seja a tendência para medidas com valores altos ou baixos de alguns indivíduos. Esta variação surge quando as unidades de investigação formam realizações de uma variável estocástica com uma determinada variância, a qual fornece a heterogeneidade dessa população.

- **Correlação em Série** (estrutura de correlação no tempo)

Esta é a componente de variabilidade dentro do indivíduo ao longo do tempo, isto é, é o tipo de variação estocástica que resulta de uma correlação entre pares de medidas no mesmo indivíduo. Esta estrutura depende das distâncias temporais entre os pares de medidas. Tipicamente, espera-se que observações muito próximas sejam fortemente correlacionadas e a correlação torna-se mais fraca à medida que a distância temporal aumenta. Em escalas de tempo suficientemente pequenas, este tipo de estrutura é inevitável.

- **Erros de medição** (Ruído branco)

Esta é a componente de variabilidade não explicada. Quando a medição envolve máquinas e mecanismos, medidas duplicadas num mesmo indivíduo num mesmo momento, pode incluir substancial variabilidade.

Há várias maneiras de incorporar estas componentes de variação em modelos específicos. Se não quisermos ignorar totalmente a componente de efeitos aleatórios, uma estratégia seria assumir uma estrutura para os erros, tal como foi apresentada em 3.16, fazendo uma distinção entre o erro de medição puro da variabilidade dentro do indivíduo.

Aqui a ideia é muito similar, ou seja, assumimos a estrutura para os erros apresentada em 3.1, fazendo uma distinção entre o erro de medição puro da variabilidade dentro e entre os indivíduos (30). Isto pode ser expresso formalmente por

$$\boldsymbol{\epsilon}_{ij} = \mathbf{d}_{ij}^T \mathbf{b}_i + W_i(t_{ij}) + z_{ij}, \quad (3.17)$$

onde utilizando a terminologia descrita, \mathbf{b}_i é a componente dos efeitos aleatórios, $W_i(t_{ij})$ a componente de correlação em série, interpretados como a variabilidade entre e dentro dos indivíduos, respetivamente, e z_{ij} representa a componente de variabilidade não explicada.

O modelo linear geral de efeitos mistos é, então, definido da seguinte forma

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{d}_{ij}^T \mathbf{b}_i + W_i(t_{ij}) + z_{ij}, \quad \mathbf{b}_i \sim MVN(\mathbf{0}^q, \mathbf{G}), \quad z_{ij} \sim N(0, \tau^2) \quad (3.18)$$

onde:

- $\boldsymbol{\beta}$ é o vetor dos efeitos fixos ($p \times 1$);
- \mathbf{x}_{ij} vetor de covariáveis dos efeitos fixos ($p \times 1$);
- \mathbf{b}_i é o vetor dos efeitos aleatórios ($q \times 1$);
- \mathbf{d}_{ij} vetor de covariáveis dos efeitos aleatórios ($q \times 1$);
- $W_i(t_{ij})$ processo Gaussiano contínuo no tempo tal que
 1. $E[W_i(t_{ij})] = 0$;
 2. $Var[W_i(t_{ij})] = \sigma^2$;
 3. $Corr[W_i(t_{ij}), W_i(t_{ik})] = \rho(|t_{ij} - t_{ik}|) = h_{ijk}$.
- z_{ij} é o erro de medição puro;
- \mathbf{G} é uma matriz definida positiva ($q \times q$);
- \mathbf{b}_i, W_i são independentes para diferentes indivíduos;
- z_{ij} são independentes para diferentes indivíduos e diferentes tempos;
- \mathbf{b}_i, W_i e z_i são independentes entre si.

Juntando os modelos especificados para todos os indivíduos descrito em 3.18, obtém-se

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{D}_i \mathbf{b}_i + W_i(t) + \mathbf{z}_i$$

onde:

- $\mathbf{b}_i \sim MVN(\mathbf{0}^q, \mathbf{G})$;
- $\mathbf{z}_i \sim MVN(\mathbf{0}^{n_i}, \tau^2 \mathbf{I}_i)$;
- $W_i(t)$ processo Gaussiano contínuo no tempo tal que

1. $E[W_i(t)] = \mathbf{0}$;

2. $Var[W_i(t)] = \sigma^2 \mathbf{H}_i = \sigma^2 \begin{bmatrix} 1 & h_{i12} & h_{i13} & \dots & h_{i1n_i} \\ h_{i21} & 1 & h_{i23} & \dots & h_{i2n_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{in_i1} & h_{in_i2} & h_{in_i3} & \dots & 1 \end{bmatrix}$.

Deste modo,

$$Var[\boldsymbol{\epsilon}_i] = \mathbf{D}_i \mathbf{G} \mathbf{D}_i^T + \sigma^2 \mathbf{H}_i + \tau \mathbf{I}_i \quad (3.19)$$

e, uma vez mais, como todos os nossos modelos admitem a propriedade de que as medidas de indivíduos diferentes são independentes, eliminamos o índice i e escrevemos, simplesmente,

$$Var[\boldsymbol{\epsilon}] = \mathbf{D} \mathbf{G} \mathbf{D}^T + \sigma^2 \mathbf{H} + \tau \mathbf{I}.$$

Em particular, em algumas aplicações práticas, pretende-se mesmo ignorar a existência de efeitos aleatórios e de erros de medição e, assim, obtém-se **o modelo linear geral** dado em 3.3. Por outro lado, se quisermos incorporar no modelo os efeitos aleatórios e os erros de medição ignorando a correlação em série, **o modelo linear misto básico** apresentado em 3.1.2 é adequado.

3.2 Modelos para a Estrutura de Correlação

Cada modelo tem as suas especificidades e, por isso, ter em conta as características dos dados ou o tipo de estudo a realizar é imprescindível para a análise de uma estrutura de correlação. Nesta secção, apresentamos, para estudos balanceados e estudos observacionais, uma descrição das possíveis estruturas de correlação disponíveis na literatura estatística.

Para estudos balanceados ir-se-á apresentar as possíveis estruturas em forma matricial, devido à possibilidade de discretização dos tempos.

3.2.1 Estruturas de Correlação para dados balanceados

O modelo mais complexo para a estrutura de correlação designa-se por **Modelo de correlação saturado**. Este incorpora o número máximo possível de parâmetros $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2, \theta_{12}, \dots, \theta_{1n}, \theta_{23}, \dots, \theta_{2n}, \dots, \theta_{(n-1)n})$ para um modelo paramétrico

$$\mathbf{V}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & \dots & \theta_{1n_i} \\ \theta_{21} & \sigma_2^2 & \theta_{23} & \dots & \theta_{2n_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n_i1} & \theta_{n_i2} & \theta_{n_i3} & \dots & \sigma_n^2 \end{bmatrix}$$

onde se impõe a existência de um único parâmetro para cada variância em cada tempo e um único parâmetro para definir a correlação entre duas observações de tempos distintos.

Uma das vantagens de um estudo longitudinal é a possibilidade de medições incompletas, isto é, não é necessário ter-se medições em todos os tempos, de todos os indivíduos. Por conseguinte, se n for o número de tempos distintos existente na base de dados, o número de parâmetros a estimar é $\frac{n(n+1)}{2}$. Além disso, é expectável que as estimativas para os parâmetros da variância não variem muito uma vez que nos referimos ao mesmo processo de medição.

Uma perspetiva relevante, na medida em que fornece grande flexibilidade aos modelos, é que a matriz $\mathbf{V}(\boldsymbol{\theta})$ pode ser decomposta em duas componentes, uma de estrutura de variância e outra de estrutura de correlação, o que permite a modelação separada destas duas estruturas. Esta decomposição da matriz num produto de matrizes mais simples é dada da seguinte forma

$$\mathbf{V}(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\sigma})\mathbf{R}(\boldsymbol{\phi})\mathbf{C}(\boldsymbol{\sigma})$$

onde

$$\mathbf{C}(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & \mathbf{0} & & \\ & & \ddots & & \\ & & & \ddots & \\ \mathbf{0} & & & & \sigma_n \end{bmatrix} \quad \mathbf{R}(\boldsymbol{\phi}) = \begin{bmatrix} 1 & \phi_{12} & \dots & \phi_{1n} \\ \phi_{21} & 1 & \dots & \phi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n1} & \phi_{n2} & \dots & 1 \end{bmatrix}.$$

Uma simplificação bastante usual, a partir do modelo saturado, é assumir que estamos num caso homocedástico, isto é, assumir que a variância é constante ao longo do tempo, $Var[\epsilon_{ij}] = \sigma^2$,

$$\mathbf{C}(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma & & & & \\ & \sigma & \mathbf{0} & & \\ & & \ddots & & \\ & & & \ddots & \\ \mathbf{0} & & & & \sigma \end{bmatrix}.$$

Neste caso, o número de parâmetros a estimar é $\frac{n(n-1)+2}{2}$.

Examinemos, agora, casos particulares das estruturas de correlação.

Simetria Composta

Neste momento, em vez de assumirmos um parâmetro em cada posição de correlação, vamos tomar um único parâmetro de correlação ϕ , isto é, a correlação é igual entre quaisquer duas medidas de tempos diferentes, quer estejam a uma distância temporal grande ou pequena. A estrutura

$$\mathbf{R}(\phi) = \begin{bmatrix} 1 & \phi & \dots & \phi \\ \phi & 1 & \dots & \phi \\ \vdots & \vdots & \ddots & \vdots \\ \phi & \phi & \dots & 1 \end{bmatrix}$$

chamamos **Correlação Simetria Composta** (*Compound Symmetry*) e ao parâmetro ϕ designamos por **Coefficiente de Correlação Intraclasse**. Neste caso, facilmente se conclui que existem 2 parâmetros a ser estimados.

Em particular, a correlação entre medidas diferentes de um mesmo indivíduo pode ser tão pequena que ϕ não seja estatisticamente diferente de zero. Neste caso, obtemos exatamente a mesma estrutura de correlação para dados longitudinais não correlacionados dada por:

$$\mathbf{R}(\phi) = \begin{bmatrix} 1 & & & \\ & 1 & \mathbf{0} & \\ & & \ddots & \\ & \mathbf{0} & & 1 \end{bmatrix}.$$

Facilmente se conclui que, para esta situação, apenas existe um parâmetro a estimar na modelação da estrutura de correlação.

Estruturas de Correlação Serial

De seguida, são apresentadas famílias de estrutura de correlação que incluem diferentes classes de modelos lineares estacionários (2). O pressuposto destas estruturas está muito relacionado com a escala temporal em que os estudos são feitos, como por exemplo, quando fazemos medições numa distância temporal muito pequena seria expectável que essas medidas estivessem altamente correlacionadas.

Nestes modelos, vamos assumir que os intervalos de tempo entre as observações são inteiros e, ainda, que os exemplos que se seguem são válidos para medidas igualmente espaçadas no tempo. Sem perda de informação, consideremos o valor 1 como a distância temporal entre dois tempos consecutivos.

Modelos Autoregressivos - Ar(p)

Os modelos autoregressivos constituem um tipo muito importante de processos estacionários. Estes expressam-se através da imposição de uma dependência linear entre uma observação num determinado instante e os valores das observações anteriores, isto é,

$$\epsilon_{it} = \phi_1 \epsilon_{i(t-1)} + \dots + \phi_p \epsilon_{i(t-p)} + a_{it},$$

onde a_{it} é um ruído homocedástico com valor médio zero e é independente das observações anteriores.

Ao número de observações anteriores p incluídas no modelo chamamos de ordem do modelo autoregressivo. Nestes casos, existem p parâmetros a estimar.

Um exemplo bastante utilizado é o **autoregressivo de ordem 1 - Ar(1)**, que pode facilmente ser descrito pela matriz

$$\mathbf{R}(\phi) = \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \ddots & \vdots \\ \phi^2 & \phi & 1 & \ddots & \phi^2 \\ \vdots & \ddots & \ddots & \ddots & \phi \\ \phi^{n-1} & \dots & \phi^2 & \phi & 1 \end{bmatrix}, \quad \text{para } |\phi| < 1.$$

Modelos de Médias Móveis - MA(q)

Relativamente aos processos de médias móveis, estes são sempre estacionários e expressam-se por uma função linear de termos de ruído, isto é, da forma

$$\epsilon_{it} = \psi_1 a_{i(t-1)} + \dots + \psi_q a_{i(t-q)} + a_{it}.$$

Ao número de termos de ruído q incluídos no modelo chamamos de ordem do modelo de médias móveis. Nestes casos, existem q parâmetros a estimar.

Seja $k = |t_{ik} - t_{ij}|$. A fórmula geral para a função de correlação de um processo MA(q) pode ser obtida por:

$$\text{Corr}(\epsilon_{ij}, \epsilon_{ik}) = \begin{cases} 0, & \text{se } k \geq q + 1 \\ \frac{\psi_k + \psi_1 \psi_{k-1} + \dots + \psi_{k-q} \psi_k}{1 + \psi_1^2 + \dots + \psi_q^2}, & \text{se } k = 1, \dots, q \end{cases}.$$

Note-se que as observações que estão a uma distância temporal superior a q unidades não estão correlacionadas.

A título de exemplo, o **modelo de médias móveis de ordem 1 - MA(1)**, descreve-se pela seguinte matriz:

$$\mathbf{R}(\psi) = \begin{bmatrix} 1 & \frac{\psi}{1+\psi^2} & 0 & \dots & 0 \\ \frac{\psi}{1+\psi^2} & 1 & \frac{\psi}{1+\psi^2} & \ddots & \vdots \\ 0 & \frac{\psi}{1+\psi^2} & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \frac{\psi}{1+\psi^2} \\ 0 & \dots & 0 & \frac{\psi}{1+\psi^2} & 1 \end{bmatrix}.$$

Modelos Autoregressivos de Médias Móveis - ARMA(p,q)

Tal como o nome indica, estes modelos são uma combinação dos modelos autoregressivos e dos modelos de médias móveis.

Um proceso $ARMA(p,q)$ é definido pela seguinte expressão

$$\epsilon_{it} = \sum_{j=1}^p \phi_j \epsilon_{i(t-j)} + \sum_{j=1}^q \psi_j a_{i(t-j)} + a_{it}.$$

Deste modo, este modelo admite $p + q$ parâmetros de correlação a estimar.

Em consequência de $ARMA(p,0)=AR(p)$ e $ARMA(0,q)=MA(q)$, os modelos $AR(p)$ e $MA(q)$ são casos particulares destes.

O **modelo ARMA(1,1)** apresenta um decaimento exponencial para distâncias temporais superiores a 1 mas permite maior flexibilidade na primeira autocorrelação (2),

$$\mathbf{R}(\phi, \psi) = \begin{bmatrix} 1 & \psi & \psi\phi & \dots & \psi\phi^{n-2} \\ \psi & 1 & \psi & \ddots & \vdots \\ \psi\phi & \psi & 1 & \ddots & \psi\phi \\ \vdots & \ddots & \ddots & \ddots & \psi \\ \psi\phi^{n-2} & \dots & \psi\phi & \psi & 1 \end{bmatrix}.$$

O diagnóstico que pode ser uma grande ajuda para a especificação do modelo, é a matriz do gráfico de dispersão empírica dos dados. Para o modelo linear geral misto, (Zimmerman, 2000) (37) propõe, em conjunto com a realização desta matriz do gráfico de dispersão empírica dos dados, um diagnóstico da matriz de gráfico de dispersão parcial de regressão comum denominada PRISM, para identificar certos tipos de estruturas de correlação, em particular os que aqui apresentamos.

3.2.2 Estruturas de Correlação Serial para tempo contínuo

Até agora vimos casos em que temos dados balanceados onde podemos averiguar a estrutura da matriz de correlação empírica dos dados.

Para dados não balanceados, onde os indivíduos são medidos em tempos completamente distintos, não é possível fazer esta discretização da função de correlação. Neste sentido, é necessário estudar possíveis funções que traduzam essa correlação como sendo uma função contínua no tempo, que depende apenas da distância temporal e de um parâmetro ϕ a estimar, ou seja, $Corr(\epsilon_{ij}, \epsilon_{ik}) = \text{função}(\phi, t_{ij}, t_{ik})$.

Autoregressivo para tempo contínuo - CAR(p)

A título de exemplo, o modelo AR(1) pode ser generalizado para tempo contínuo, o qual designamos por CAR(1) (2) onde a função de correlação é dada por,

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = \phi^{|t_{ik} - t_{ij}|}.$$

Casos particulares das funções genéricas de Matérn, nomeadamente a **função de correlação exponencial** e a **função de correlação gaussiana**, são também exemplos bastante utilizados em modelos de dados longitudinais. Vejamos as suas características:

Função de correlação exponencial

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = \exp\left\{-\frac{1}{\phi} |t_{ik} - t_{ij}|\right\}, \quad \phi > 0.$$

Função de correlação gaussiana

$$Corr(\epsilon_{ij}, \epsilon_{ik}) = \exp\left\{-\frac{1}{\phi} (t_{ik} - t_{ij})^2\right\}, \quad \phi > 0.$$

Seja $u = t_{ik} - t_{ij}$. Uma das diferenças mais importantes entre estas duas funções de correlação, ilustrada na Figura 3.1, é o comportamento perto de $u = 0$. A função gaussiana tem um aspeto idêntico a um “S” invertido horizontalmente (efeito espelho), isto é, começa com valores de correlação próximos de 1 decrescendo, posteriormente, muito rapidamente, enquanto a exponencial vai decrescendo de uma forma muito menos acentuada à medida que a distância temporal entre medidas aumenta.

Para além disso, esta figura também sugere o impacto potencial do valor do parâmetro ϕ nestas duas funções cuja curvatura muda dependendo do valor de ϕ .

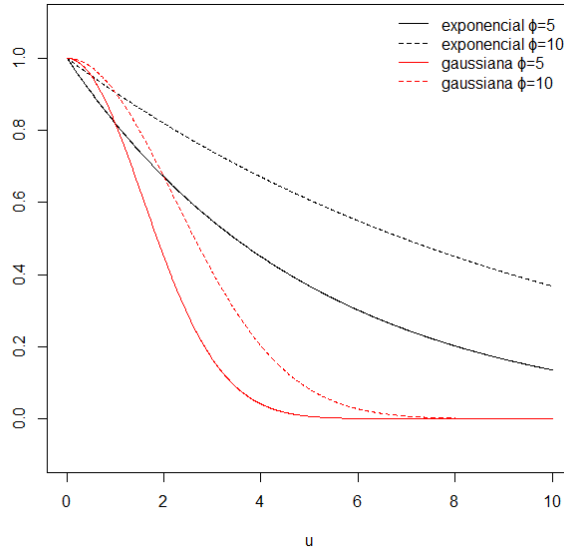


Figura 3.1: Função de correlação exponencial e gaussiana para $\phi = 5, 10$

Para percebermos melhor este impacto, observe-se que a estrutura gaussiana para $\phi = 5$ atinge correlações próximas de zero quando $u = 5$, enquanto se estivermos em $\phi = 10$ a função decresce para zero já posteriormente. Assim, este parâmetro é interpretado como sendo a distância a partir da qual a correlação deixa, de alguma forma, de ter impacto. Vejamos, por exemplo, que as diferenças das correlações de duas observações a uma distância temporal de ϕ e a uma distância de $\phi + 1$ irá ser muito pequena, o que não acontece se compararmos duas observações a uma distância de $\phi - 1$ e ϕ .

3.2.3 Função Variograma

Uma das grandes preocupações em estudos longitudinais é ajustar da melhor maneira possível a estrutura de correlação. Contudo, como já referido, definir qual estrutura de correlação deve ser ajustada aos dados pode não ser uma tarefa fácil. Assim, é imprescindível estudar várias estruturas da matriz de correlação para o modelo. Uma hipótese é começar por transformar as observações y_{ij} nos resíduos obtidos com ajuste dos mínimos quadrados ($r_{ij} = y_{ij} - x_{ij}\hat{\beta}_W$) e, conseqüentemente, modelar e analisar a estrutura de correlação dos resíduos da maneira mais saturada possível reconhecendo padrões característicos dentro dessa estrutura. A partir daí, iremos obter estruturas mais simplificadas, que dependem de um conjunto mais reduzido de parâmetros, e que conseguimos interpretar. Para um grande número de r_{ij} , este deve apresentar propriedades semelhantes a ϵ_{ij} .

Em 3.3 são apresentados vários métodos de seleção que nos permite avaliar e comparar modelos com estruturas de correlações distintas.

Para estudos observacionais, em que os indivíduos podem ser medidos em quaisquer tempos, não conseguimos fazer este tipo de análise exploratória. Assim, é necessário encontrar métodos que nos permitem explorar a estrutura de correlação de uma forma contínua. Uma estratégia bastante utilizada é usufruir de uma função **variograma** (dos resíduos).

O variograma é uma ferramenta importante que nos permite explorar, descrever e interpretar qual será a estrutura de correlação existente em qualquer tipo de dados empíricos (podem ser longitudinais, espaciais, etc.). Uma diferença clara entre a função de covariância e o variograma é que o primeiro é uma função direta da associação entre duas variáveis, enquanto que o segundo mede a dissociação. Os variogramas são mais gerais do que as funções de covariância, e muitas propriedades importantes foram inicialmente estabelecidas para as funções de covariância e, portanto, a estimação deste é mais simples que a estimação da função de covariância. Para além disso, admite vantagens de estimação em que não necessita de nenhuma estimativa do valor esperado e adapta-se facilmente a observações não estacionárias.

Para um processo estocástico à custa da distância temporal entre duas observações (u) o variograma é dado por

$$V(u) = \frac{1}{2} \text{Var}\{Y(t) - Y(t - u)\}. \quad (3.20)$$

Um conjunto de processos que tem um papel relevante para o estudo de modelos longitudinais são os **processos estacionários de 2^a ordem** ou **fracamente estacionários**. Num processo estacionário de 2^a ordem, tem-se que:

1. o valor médio não depende de t , isto é, $\mu(t) = \mu$;
2. a variância não depende de t , isto é, $\sigma^2(t) = \sigma^2$;
3. a covariância de $Y(t_1)$ e $Y(t_2)$ depende apenas do desfasamento $|t_2 - t_1|$, isto é,

$$\text{Cov}\{Y(t_1), Y(t_2)\} = \gamma(|t_2 - t_1|).$$

A partir de agora, trataremos o variograma para estes processos, os quais designamos simplesmente por **estacionários**.

Deste modo, pode substituir-se a expressão do variograma definido em 3.20 por

$$V(u) = \text{Var}[Y(t)] - \text{Cov}\{Y(t), Y(t - u)\}.$$

Atendendo nesta expressão, seguem-se três exemplos teóricos particulares dos apresentados em 3.17.

Variograma teórico assumindo correlação em série (modelo linear geral)

Este caso representa os erros que se traduzem apenas na componente de correlação em série, dada por

$$\epsilon_i = W_i(t).$$

Por 3.19 obtém-se

$$Var[\epsilon_i] = \sigma^2 \mathbf{H}_i.$$

Nesta situação, a variância para cada ϵ_i é σ^2 . Se os elementos de \mathbf{H}_i são especificados por uma função de autocorrelação $\rho(u)$ tal que $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, então a relação entre a função de covariância e a função de correlação é dada por $\gamma(u) = \sigma^2 \rho(u)$. Portanto, para processos estacionários, o variograma é bem definido e pode escrever-se da seguinte forma

$$V(u) = \sigma^2(1 - \rho(u)).$$

Note-se que $V(0) = 0$ e $V(u)$ tende para σ^2 à medida que a distância temporal aumenta. Tipicamente, $V(u)$ é uma função crescente em u porque a correlação, $\rho(u)$, diminui com o aumento de u .

Apesar de, numa perspetiva matemática, a relação do variograma ser inversa à função de correlação e, portanto, serem equivalentes, numa perspetiva estatística é mais fácil explorar a função variograma comparativamente com a função de correlação $\rho(u)$ para dados não balanceados.

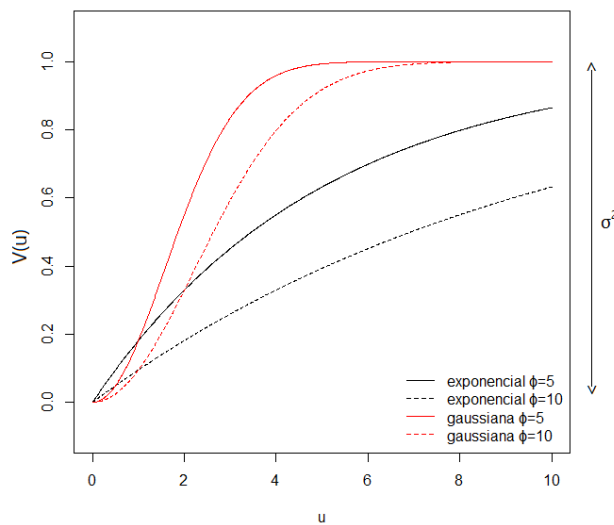


Figura 3.2: Variogramas teóricos assumindo $\sigma^2 = 1$ para o modelo de correlação exponencial e gaussiano, em $\phi = 5$ e 10.

Em muitas aplicações, o variograma empírico é semelhante ao modelo de correlação gaussiano, ilustrado na Figura 3.2, em que inicialmente há um aumento lento à medida que u aumenta, depois um aumento acentuado e, finalmente, um aumento mais lento à medida que se aproxima da sua assíntota.

Variograma teórico assumindo correlação em série e erros de medição

Este caso representa os erros que se traduzem na soma de duas componentes: a de correlação em série e a dos erros de medição

$$\epsilon_i = W_i(t) + z_i.$$

Por 3.19, obtém-se

$$Var[\epsilon_i] = \sigma^2 \mathbf{H}_i + \tau \mathbf{I}_i.$$

Neste caso, a variância para cada ϵ_i é dada por $\sigma^2 + \tau^2$. Se os elementos de \mathbf{H}_i são especificados por uma função de correlação $\rho(u)$ tal que $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$ então, uma vez mais, obtém-se a relação $\gamma(u) = \sigma^2 \rho(u)$ e, portanto, o variograma é definido da seguinte maneira

$$V(u) = \tau^2 + \sigma^2(1 - \rho(u)).$$

A figura 3.3 mostra uma das principais características presente em modelos que admitem erros de medição: à medida que a distância temporal diminui, $V(u)$ não tende para zero, mas para τ^2 .

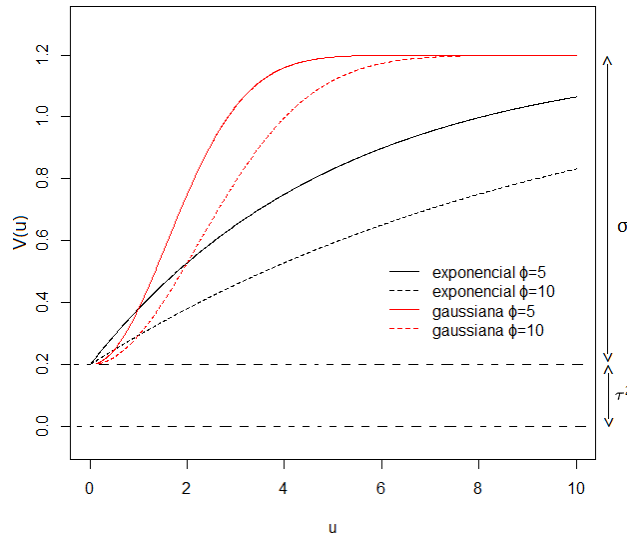


Figura 3.3: Variogramas teóricos assumindo $\tau^2 = 0.2$ e $\sigma^2 = 1$ para o modelo de correlação exponencial e gaussiano em $\phi = 5$ e 10.

Variograma teórico assumindo todas as fontes de variação (modelo linear misto geral)

Assumamos, agora, que num conjunto de dados longitudinais estão presentes as três fontes de variação aleatória, isto é,

$$\epsilon_i = \mathbf{D}_i \mathbf{b}_i + W_i(t) + \mathbf{z}_i.$$

Em particular, assumamos, também, que \mathbf{b}_i é uma variável aleatória gaussiana univariada com média zero e variância ν^2 . Tomemos, ainda, que \mathbf{D}_i é o vetor ($n_i \times 1$) em que todos os seus elementos são iguais a 1. Assim, obtém-se

$$\text{Var}[\epsilon_i] = \nu^2 \mathbf{J}_i + \sigma^2 \mathbf{H}_i + \tau \mathbf{I}_i$$

onde $\mathbf{J}_i = \mathbf{D}_i \mathbf{D}_i^T$ é uma matriz ($n_i \times n_i$) cujos elementos são iguais a 1.

Neste caso, a variância para cada ϵ_i é dada por $\sigma^2 + \tau^2 + \nu^2$. Se os elementos de \mathbf{H}_i são, uma vez mais, especificados por uma função de correlação $\rho(u)$ tal que $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, então a relação $\gamma(u) = \nu^2 + \sigma^2 \rho(u)$ retorna o variograma com a mesma forma do variograma obtida no caso anterior,

$$V(u) = \tau^2 + \sigma^2(1 - \rho(u)).$$

A figura 3.4 mostra os variogramas teóricos assumindo $\tau^2 = 0.2$, $\sigma^2 = 1$ e $\nu^2 = 0.2$ para o modelo de correlação exponencial e gaussiano em $\phi = 5$ e 10.

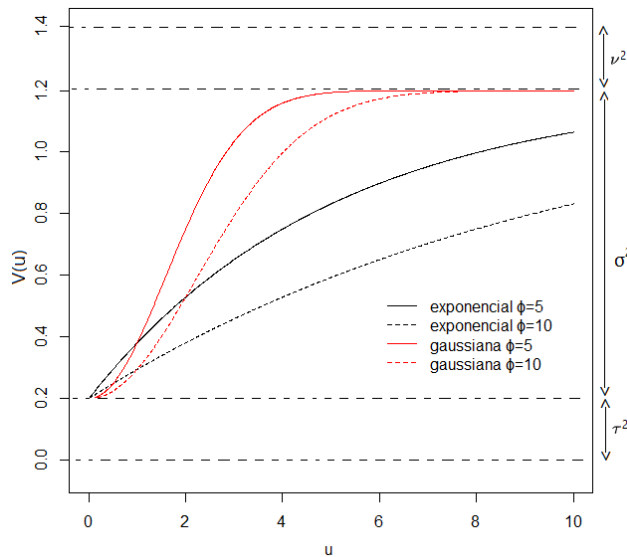


Figura 3.4: Variogramas teóricos assumindo $\tau^2 = 0.2$, $\sigma^2 = 1$ e $\nu^2 = 0.2$ para o modelo de correlação exponencial e gaussiano em $\phi = 5$ e 10.

Variograma experimental, \hat{V}

Como já analisado, o variograma consegue fornecer, para além da forma de $\rho(u)$, a representação gráfica da decomposição da variabilidade total dos dados em 3 componentes: variabilidade de diferença do comportamento entre indivíduos (heterogeneidade entre indivíduos), variabilidade ao longo do processo dentro do indivíduo e a variabilidade que não se consegue explicar. Mesmo quando as três componentes de variabilidade são, em princípio, necessárias, apenas uma delas, ou duas, são dominantes, na prática. Deste modo, vamos agora determinar a função variograma para os dados observados.

Ora, por 1, o processo $Y(t)$ tem valor esperado constante e, portanto, o variograma pode também ser expresso da seguinte forma

$$V(u) = \frac{1}{2}E\{(Y(t) - Y(t - u))^2\}.$$

À semelhança do que acontece com dados balanceados, se usarmos, em alternativa ao processo estocástico Y , r como os resíduos obtidos após ajustar uma regressão linear com todas as covariáveis presentes na base de dados, podemos determinar a função variograma para os dados observados através da função dada por

$$V(u) = \frac{1}{2}E\{(r(t) - r(t - u))^2\}.$$

Para encontrar o estimador para $V(u)$ designemos

$$v_{ijil} = \frac{1}{2}(r_{ij} - r_{il})^2,$$

onde r_{ij} e r_{il} são os resíduos preliminares para a média da resposta j e l , respetivamente, do mesmo indivíduo i , isto é, as diferenças destes resíduos devem ser feitas dentro do mesmo indivíduo para tempos distintos - e não entre tempos distintos de indivíduos distintos. Deste modo, $\hat{V}(u)$ é a média de todas as quantidades v_{ijil} tal que $|t_{ij} - t_{il}| \approx u$ e ainda, para os modelos em que incorporamos erros de medição, se considerarmos $V(0) = \tau^2$, então $\hat{\tau}^2$ é a média de todas as quantidades v_{ijkj} tal que $i \neq k$.

Para estimar a variância do processo, que denotamos por $\hat{\sigma}^2$, calculamos a média de todas as quantidades v_{ijkl} quer em medidas para o mesmo indivíduo em tempos distintos, quer em medidas de indivíduos diferentes.

De modo geral, o variograma experimental $\hat{V}(u)$ é dado por

$$\hat{V}(u) = \frac{1}{2|N(u)|} \sum_{N(u)} (r(t_{ij}) - r(t_{ik}))^2,$$

onde $N(u) = \{(t_{ij}, t_{ik}) : |t_{ij} - t_{ik}| \approx u, u \in \mathbb{R}\}$ e $|N(u)| =$ número de pares em $N(u)$.

Deste modo, comparando com os variogramas teóricos obtidos, o variograma empírico dos resíduos sugere informação bastante importante sobre que tipos de fontes de variação estão subjacentes no modelo que estamos a trabalhar, para assim as podermos incorporar e avaliar.

Na prática, é muito comum, a partir de um certo valor de u , o estimador do variograma empírico dos dados deixar de ser uma função crescente e passar a ser uma função decrescente, o que na interpretação associada a esta função é incoerente. Isto acontece porque para distâncias temporais pequenas possuímos mais informação, na medida em que temos mais pares de observações a essas distâncias do que para distâncias maiores e, portanto, o variograma vai ter uma menor precisão para valores de u grandes. Consequentemente, não se deve valorizar da mesma maneira o estimador do variograma para distâncias temporais grandes.

Além disso, técnicas paramétricas e não paramétricas de suavização do variograma empírico devem ser utilizadas para dados irregularmente espaçados, como, por exemplo, o agrupamento de diferenças temporais em bandas da mesma largura.

3.3 Ajuste e Seleção do Modelo

Após a análise exploratória dos dados, (Diggle et. al., 2002 (6)) assume que as linhas gerais para o processo de ajustar um modelo longitudinal são:

1. Formulação de um modelo provisório
 - Escolher uma estrutura para a média e uma estrutura de covariância através dos mínimos quadrados do modelo mais complicado e recorrendo ao variograma empírico.
2. Estimação de parâmetros
 - Usar REML ou ML para estimar parâmetros da estrutura de covariância e média.
3. Inferência de parâmetros
 - Calcular intervalos de confiança testando hipóteses sobre os parâmetros de interesse direto.
4. Métodos de diagnóstico
 - Verificar se o modelo se ajusta aos dados através de uma representação gráfica da média e do variograma teóricos e empíricos dos dados.

Relativamente à escolha do melhor modelo, podemos optar pelos **critérios de informação de Akaike** (AIC), **Akaike corrigido** (AICc) e **Bayesiano** (BIC).

O primeiro critério, proposto por Hirotugu Akaike em 1973, é uma fórmula matemática que depende do valor da função log-verosimilhança e do número de parâmetros que está a ser utilizado no modelo, dada por

$$AIC = -2L(\hat{\theta}) + 2p.$$

Prova-se que se estivermos a comparar dois modelos, o modelo a optar é o que tiver menor AIC, com uma diferença de pelo menos 5 unidades, ou seja, se a diferença entre os AIC's de dois modelos for menor que cinco podemos dizer que estes são significativamente iguais.

O critério de informação de Akaike corrigido foi desenvolvido em 1987 por Bozdogan e representa uma correção para o AIC, tendo em consideração que este pode apresentar um mau desempenho quando o número de parâmetros, p , excede em muito o tamanho da amostra, n (19).

$$AICc = -2L(\hat{\theta}) + 2p + \frac{p(p+1)}{n-p-1}.$$

O critério BIC (Schwarz, 1978) define-se por:

$$BIC = -2L(\hat{\theta}) + p \log(n).$$

Quando pretendemos comparar dois modelos encaixados que diferem apenas na estrutura de efeitos fixos, ou seja, quando as covariáveis de um dos modelos for um subconjunto do outro, também podemos optar pelo **teste de razão de verosimilhança** (*Likelihood-ratio test*). Este teste é válido apenas se o método utilizado para estimar os parâmetros fixos for o método de verosimilhança máxima (ML) (2).

A estatística de teste da razão de verosimilhança para testar a hipótese nula é dada por

$$2 \log \left(\frac{L_1}{L_0} \right) \sim \chi^2_{(k_1 - k_0)},$$

onde L_0 e L_1 representam a verosimilhança do modelo mais restrito (com menos parâmetros) e o modelo mais geral, respetivamente, e $k_1 - k_0$ a diferença entre o número de parâmetros dos dois modelos. O modelo mais adequado sob a hipótese nula é o modelo restrito.

A sobreposição do variograma empírico e a do variograma teórico é, também, uma técnica visual admissível que nos permite detetar qual poderá ser a estrutura de correlação mais adequada.

3.4 Modelos Conjuntos

Introdução à análise de sobrevivência

Estamos em contexto de análise de sobrevivência quando o nosso objetivo for estudar o tempo que decorre desde um momento muito bem especificado até um determinado evento de interesse. Para este tipo de análise é necessário definir muito bem algumas características tais como o **evento de interesse**, a **origem no tempo**, a **escala no tempo** e a **variável resposta**.

Se considerarmos dados médicos, a maior parte das aplicações referem-se ao tempo de vida de um indivíduo sendo a morte o nosso evento de interesse e a origem no tempo poderá ser, por exemplo, a data da cirurgia (4). Neste caso, a nossa variável resposta é o tempo desde a cirurgia até à morte. Outro exemplo para o evento de interesse é a recidiva, isto é, o reaparecimento dos sintomas de uma doença que já tinha sido curada no mesmo indivíduo ou recaída. A escala no tempo poderá ser em dias, meses, etc. Nos dados oncológicos, os indivíduos iniciam o estudo na data de cirurgia, ao contrário do que acontece nos ensaios clínicos em que iniciam o estudo no mesmo momento.

Para além das diferentes áreas da medicina, estes métodos podem também ser utilizados em áreas das ciências sociais (tal como o tempo de duração que um indivíduo executa uma determinada tarefa), da economia (por exemplo, o tempo que um indivíduo está desempregado), da engenharia (como o tempo de vida de uma componente eletrónica), etc.

As dificuldades específicas relacionadas com a análise de sobrevivência resultam, em grande parte, de uma característica peculiar, muitas vezes presente nos dados de sobrevivência conhecida por **censura**, que, em termos gerais, ocorre quando temos alguma informação sobre o tempo de sobrevivência individual (dentro de certos intervalos), mas não sabemos o tempo exatamente. Daí, podemos obter tempos de sobrevivência desconhecidos para um subconjunto do grupo de estudo.(16)

Podemos categorizar amplamente a censura em três grupos com base na razão pela qual a censura ocorreu (34):

- **Censura à direita:** Observa-se um determinado valor de tempo de sobrevivência inferior ou igual ao verdadeiro valor do tempo de sobrevivência. Na generalidade dos casos, este é o tipo de censura mais frequente e pode ocorrer por 3 motivos em especial: perdas de seguimento (causas que não tem a ver com o estudo em si, como mudança de residência), abandonos do estudo e, na maior parte dos casos, por término do estudo.
- **Censura à esquerda:** É o contrário de censura à direita, isto é, observa-se um determinado valor de sobrevivência maior ou igual ao real tempo de sobrevivência. Este

tipo de censura é menos frequente e é mais difícil de lidar em metodologias estatísticas usuais. Por exemplo, quando o estudo consiste no tempo de reaparecimento de um cancro, este é normalmente diagnosticado através de exames clínicos, biópsia, entre outros, mas nunca sabemos quando, efetivamente, foi o início do cancro e, portanto, assumimos a data do exame como a data do evento de interesse, o que será certamente um valor de tempo superior ao tempo exato.

- **Censura Intervalar:** É um tipo de censura bastante frequente (não tanto como censura à direita) e, felizmente, já conseguimos encontrar alguns métodos que permitem lidar com este tipo de dados. Em algumas situações, acontece não se saber exatamente o tempo até ao evento de interesse, mas sabemos que o evento de interesse ocorre entre 2 momentos. Um caso muito comum deste tipo de censura é, por exemplo, quando um indivíduo está a ser acompanhado por uma determinada doença e faz visitas periódicas ao seu médico a cada 3 meses. Quando, num determinado momento, este não observou o evento de interesse, mas, na secção consecutiva, já observou, o investigador não sabe, precisamente, qual é o tempo exato até ao evento de interesse, apenas que o evento ocorreu durante um dado intervalo de tempo, neste caso de 3 meses.

Apesar de tais tempos de sobrevivência censurados subestimarem o verdadeiro tempo de ocorrência, estas observações censuradas devem ser utilizadas na análise, pois fornecem informações importantes sobre o tempo de vida dos pacientes. Em termos práticos, poderíamos estar a remover as observações dos indivíduos com melhor prognóstico (aqueles que no final do estudo ainda continuam vivos) e, portanto, a omissão destas pode acarretar conclusões enviesadas. Dessa forma, uma das principais vantagens da análise de sobrevivência em comparação com metodologias clássicas, é poder incorporar a modelagem de dados censurados.

Além disso, embora seja um conceito totalmente diferente, um outro fenómeno geralmente confundido com a censura é a chamada **truncatura**. A truncatura consiste em excluir os indivíduos que não são relevantes para o estudo em questão. Por exemplo, para estudar a taxa de sobrevivência a determinada doença em determinada localidade, recorrendo a dados do recenseamento eleitoral da respetiva junta de freguesia, apenas considerar-se-ão os sujeitos maiores de idade. Dessa forma, existe uma truncatura à esquerda, uma vez que os menores de idade são excluídos do estudo. (28)

Se fixarmos o início (A) e o fim do estudo (B) podemos observar os tempos t_i e os tempos potenciais de censura de cada indivíduo C_i . Portanto, temos associadas duas variáveis:

- T_i : Tempo exato;
- C_i : Tempo potencial de censura.

Na prática, o que as bases de dados fornecem é, essencialmente,

$$(F_i, \delta_i), \quad F_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i) = \begin{cases} 1 & T_i \leq C_i, \\ 0 & T_i > C_i \end{cases}.$$

As variáveis **time** F_i e **status** δ_i são fundamentais e são elas que irão caracterizar a análise de sobrevivência. Time indica o tempo de sobrevivência ou tempo de censura observado e Status é uma variável categórica com valor 0 se for um tempo censurado ou incompleto e 1 se o evento de interesse foi observado.

Seja T uma variável aleatória contínua não negativa que representa o tempo até ao evento de interesse de um determinado indivíduo. De seguida são apresentadas algumas das funções importantes.

Função de densidade, $f(t)$

A função de densidade, $f(t)$, pode ser interpretada como a probabilidade de um indivíduo sofrer um evento num intervalo de tempo muito pequeno e pode ser representada por

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}.$$

Função de distribuição, $F(t)$

Função cumulativa da função densidade de probabilidade:

$$F(t) = P(T \leq t) = \int_0^t f(u) du,$$

$$F'(t) = f(t), \quad F(0) = 0, \quad \lim_{t \rightarrow +\infty} F(t) = 1.$$

Função de sobrevivência, $S(t)$

A Função de Sobrevivência indica-nos qual é a probabilidade de sobreviver para além de um determinado momento.

$$S(t) = P(T > t) = \int_t^{+\infty} f(u) du = 1 - \int_0^t f(u) du = 1 - F(t),$$

$$S'(t) = -f(t), \quad S(0) = 1, \quad \lim_{t \rightarrow +\infty} S(t) = 0.$$

A função de sobrevivência é uma função contínua e monótona não crescente. A partir dela, podemos resumir os principais objetivos da análise de sobrevivência (28):

- Estimar a função de sobrevivência a partir dos dados recolhidos, em particular, em função de fatores que sejam considerados de risco. O método de Kaplan-Meier fornece um estimador não paramétrico desta função;
- Comparar funções de sobrevivência entre grupos (a uma menor função de sobrevivência corresponde uma maior mortalidade). Por exemplo, podemos querer comparar se dois tratamentos conduzem às mesmas sobrevivências ou se um conduz a mais sobrevivências do que outro - no caso de doença coronária ou nos estudos oncológicos podemos comparar um tratamento contra o efeito placebo;
- Estudar a influência de covariáveis de interesse na função de sobrevivência (por exemplo, determinar se o aumento do valor de determinada covariável faz aumentar ou diminuir a função de sobrevivência no intervalo de estudo - para este caso, podemos aplicar um modelo de regressão).

Função de Hazard (Risco), $h(t)$

Risco instantâneo de observar o evento de interesse, sabendo que o indivíduo sobreviveu até aquele momento. Dito de outra forma, mede a taxa instantânea de observarmos o evento de interesse considerando que o indivíduo sobreviveu até ao evento t , ou seja, a probabilidade de sobreviver mais do que o tempo t .

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

A função de risco pode ser constante, crescente ou decrescente. Existe uma relação entre a função de risco e a função de sobrevivência dada por

$$\int_0^t h(u) du = -\log(S(t)).$$

À medida que diminui a função de risco, aumenta a função de sobrevivência e vice-versa.

Função de Hazard (Risco) Comulativa, $H(t)$

$$H(t) = \int_0^t h(u) du,$$

$$H(t) = -\log(S(t)).$$

Existem também alguns parâmetros de interesse que se pretendem estimar neste tipo de análise, tais como:

1. **Valor Médio** (valor esperado da variável aleatória T)

$$\mu = E[T] = \int_0^{+\infty} u f(u) du = \int_0^{+\infty} S(u) du.$$

2. **Tempo de Vida Residual Médio (MRL)**

De grosso modo, para entendermos melhor a interpretação desta quantidade, atentemos na seguinte questão: dado que o indivíduo sobreviveu para além do tempo t, qual é o tempo de vida residual médio, isto é, o tempo que lhe resta.

$$mrl(t) = E[T - t \mid T > t] = \frac{\int_t^{+\infty} S(u) du}{S(t)}.$$

3. **Variância**

$$Var[T] = 2 \int_0^{+\infty} u S(u) du - \left[\int_0^{+\infty} S(u) du \right]^2.$$

4. **Quantis**

$$S(t_p) = 1 - p, \text{ isto é, } t_p = \inf\{t : S(t) \leq 1 - p\}.$$

Para concluir, destacam-se os livros (Wang e Lee, 2003)(18) e (Klein e Kleinbaum, 2012)(17) que fornecem boas introduções elementares à análise de sobrevivência juntamente com a apresentação de vários exemplos e exercícios teóricos e práticos.

Modelação Conjunta de Dados Longitudinais e de Sobrevivência

Como observamos em vários exemplos anteriormente descritos, em estudos clínicos e médicos, dados longitudinais e de tempo até o evento de interesse são frequentemente considerados medidas de saúde importantes e, na maioria das vezes, surgem efetivamente em conjunto (14).

Questões de interesse em tais estudos requerem a análise separada dos resultados registados. No entanto, no caso do processo de resposta longitudinal e o do resultado do tempo até ao evento de interesse não serem independentes, adotar metodologias separadas deste tipo de dados podem levar a resultados ineficientes ou tendenciosos. Assim, aplicar abordagens em

que conseguimos modelar os dois processos em simultâneo é imprescindível para a obtenção de inferências válidas e eficientes(12).

Vários são os artigos que trabalham estas propriedades. Em 2004, Guo e Carlin desenvolveram uma abordagem conjunta totalmente bayesiana considerando que estes processos são independentes, dado um processo gaussiano bivariado latente de ligação. Esta abordagem providenciou melhorias significativas nas estimativas dos parâmetros de interesse como, por exemplo, os tempos médios de sobrevivência em comparação com os resultados obtidos usando métodos tradicionais bayesianos de probabilidade separada (10)).

Estudos mais recentes, como o realizado por Markos Erango em 2018 onde selecionou 550 amostras aleatórias de pacientes com HIV/SIDA que estavam em acompanhamento de terapia antirretroviral de janeiro de 2007 a outubro de 2017 no Arba Minch General Hospital, na Etiópia, também demonstrou que com a utilização de modelos conjuntos obtiveram-se resultados consistentes com alta precisão (8).

Neste sentido, o estudo destes modelos nos últimos anos tem sido uma área de pesquisa estatística que tem recebido muita atenção, motivada principalmente por ensaios clínicos de HIV/SIDA, em particular, pela modelagem conjunta de dados de sobrevivência e contagens longitudinais de CD4 (15).

Alguns dos primeiros trabalhos foram apresentados por (Self e Pawitan, 1992) (27), (DeGruttola e Tu, 1994) (5), (Faucett e Thomas, 1996) (9) e (Wulfsohn e Tsiatis, 1997) (36). Evidencia-se ainda um artigo fornecido por (Tsiatis e Davidian, 2004) (33) e, mais recentemente, outro concedido por (Sousa, 2011) (30), os quais fornecem excelentes revisões gerais das diferentes abordagens existentes de modelos conjuntos. Para aplicações de modelação conjunta em linguagem de programação R, o livro de (Dimitris Rizopoulos, 2012) (22) apresenta implementações destes modelos através do pacote disponível *JM* e por (Williamson et al., 2008) (35) através do pacote *joiner*.

Um artigo de grande destaque na formulação do problema deste trabalho foi apresentado por Diggle et.al, 2008)(7), em que são apresentadas três tipos de abordagens diferentes: modelagem paramétrica de efeitos aleatórios, modelagem semi-paramétrica e uma abordagem baseada no pressuposto de que o vetor conjunto de respostas das medidas repetidas juntamente com os resultados de sobrevivência seguem uma distribuição multivariada gaussiana.

Relativamente a abordagens de efeitos aleatórios ou parâmetros compartilhados para modelar conjuntamente medições repetidas e dados de tempo de evento, apesar de serem conceitualmente atraentes, flexíveis e essenciais em muitas configurações, a sua aplicação usual dispõe de algumas limitações devido a dificuldades computacionais. Métodos rápidos mas aproximados foram desenvolvidos para algumas metodologias de modelo conjunto, mas ainda assim a sua implementação permanece difícil, a menos que a componente de efeitos aleatórios

seja de baixa dimensão.(1)

Modelos semi-paramétricos, no sentido de Robins et.al (24), têm como principal foco as propriedades dos perfis médios de resposta do processo de medição, fornecendo estimativas consistentes sob suposições mínimas e rubusidade face ao comportamento não-Gaussiano. A sua desvantagem é que os métodos não são apropriados para responder a questões mais complexas sobre a natureza da dependência entre os processos de medição e tempo até o evento.(7)

A última abordagem é particularmente importante neste trabalho, na medida em que é sobre este modelo que fazemos a análise teórica e prática de possíveis estruturas de correlação e extraímos informação relevante dos dados. Este estudo encontra-se, mais especificamente, no capítulo quatro.

Na metodologia de modelação conjunta podem ser formulados vários objetivos, de acordo com o foco principal da análise (31): (i) analisar o resultado de tempo, levando em consideração o efeito de um resultado longitudinal como covariável dependente do tempo medido com erro (ii) analisar o resultado longitudinal na presença de tempo de abandono informativo (não aleatório) e (iii) analisar os efeitos das covariáveis de interesse em ambos os tipos de resultados (longitudinal e tempo até ao evento de interesse) simultaneamente.

Uma questão importante é se a ocorrência do evento no tempo F termina ou não, ou afeta as propriedades da sequência de medição Y . No que se segue, assumimos que o processo Y permanece bem definido após a ocorrência do evento, mas pode ou não ser observável. Deste ponto de vista, o nosso modelo é um modelo para uma sequência de tempo completa Y e tempo de evento F , mesmo quando a sequência Y é apenas parcialmente observada.

Uma vez que a distribuição conjunta para as duas variáveis aleatórias não é clara, admitem-se duas fatorizações diferentes da distribuição conjunta $[Y, F]$ utilizando a regra de Bayes:

Modelos de Mistura de Padrões

$$[Y, F] = [F][Y|F]$$

Modelos de Seleção

$$[Y, F] = [Y][F|Y] .$$

A estratégia do modelo depende principalmente da natureza do problema estatístico e das questões científicas a serem respondidas. Embora, matematicamente, os modelos descrevam exatamente a mesma distribuição conjunta, eles têm diferentes interpretações estatísticas e, conseqüentemente, aplicabilidades distintas a problemas individuais. Os modelos de seleção são usados principalmente quando a inferência está nos parâmetros do modelo de tempo até o evento, melhorando a inferência ao permitir a correlação nas medições longitudinais. Em oposição, quando o interesse principal está na trajetória longitudinal, que pode estar associada a um padrão de evento, os modelos de mistura de padrões são mais usados de forma comum. Portanto, as duas abordagens diferentes levam a diferentes entendimentos e inferên-

cias dos parâmetros do modelo, juntamente com diferentes visões sobre como armazenar os dados. (30)

Os modelos apresentados podem ser alargados para incorporar efeitos aleatórios, os chamados modelos aleatórios de mistura de padrões e modelos aleatórios de seleção. Os efeitos aleatórios individuais não observados nos modelos de seleção estão incluídos no modelo longitudinal marginal, enquanto que nos modelos de mistura de padrões estes vêm na distribuição marginal dos tempos do evento. Portanto, ao modelar conjuntamente medições repetidas Y , tempos de evento F e efeitos aleatórios U , as distribuições conjuntas são,

Modelos Aleatórios de Mistura de Padrões

$$[Y, F, U] = [U][F|U][Y|F]$$

Modelos Aleatórios de Seleção

$$[Y, F, U] = [U][Y|U][F|Y] .$$

Os **Modelos Conjuntos de Efeitos Aleatórios** assumem que tanto as medições repetidas quanto o tempo do evento dependem de um efeito aleatório não observado, estes especificados por meio de uma determinada distribuição bivariada. O modelo conjunto de efeitos aleatórios é descrito assumindo independência condicional entre Y e F dados os efeitos aleatórios $U = (U_1, U_2)$, como

$$[Y, F, U] = [U][Y|U_1][F|U_2].$$

Em modelos conjuntos de efeitos aleatórios, a associação entre medidas longitudinais e tempo até o evento é completamente determinada pela estrutura de correlação entre os dois efeitos aleatórios U_1 e U_2 . Além disso, sugere que ambos os processos são uma resposta conjunta a um processo individual específico não observado e assumidos como condicionalmente independentes, dada uma variável latente U .

Um modelo em que todas as associações entre os processos F , Y e U são possíveis chamamos de **Modelo Saturado**.

Capítulo 4

Modelo Gaussiano Transformado

Uma das abordagens para especificar a distribuição conjunta de um vetor de medidas repetidas $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$ e um resultado de tempo até à ocorrência de um evento de interesse F_i é o chamado **Modelo Gaussiano transformado**. Este pressupõe que o processo de medição é descrito por um modelo gaussiano linear e para F assume a distribuição log-Gaussiana (7). O modelo é de transformação simples, totalmente paramétrico e define-se da seguinte forma:

$$(\mathbf{Y}, \log F) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad i.i.d. \quad (4.1)$$

onde

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_Y, \mu_S) \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{V}(\boldsymbol{\theta}) & \mathbf{g}(\boldsymbol{\gamma}) \\ \mathbf{g}^T(\boldsymbol{\gamma}) & \eta^2 \end{bmatrix}.$$

Para o estudo desta abordagem consideramos m realizações independentes e identicamente distribuídas de $(\mathbf{Y}_i, \log F_i)$ e, portanto, observações de indivíduos diferentes são consideradas independentes e observações do mesmo indivíduo em tempos diferentes estão correlacionadas. O tempo do evento F_i para o indivíduo i pode ser considerado uma observação completa, censurada à direita ou censurada por intervalo.

A estrutura de variância-covariância do modelo contém três componentes de modelação: a da variância-covariância das medidas repetidas dada pela matriz $\mathbf{V}(\boldsymbol{\theta})$, a variância da variável de sobrevivência η^2 e a covariância entre \mathbf{Y} e $S = \log F$ dada pelo vetor $\mathbf{g}(\boldsymbol{\gamma})$.

Em contrapartida à escolha de um modelo paramétrico para a matriz de variância $\mathbf{V}(\boldsymbol{\theta})$, cujo problema encontra-se bem estudado (ver secção 3.2), a escolha de um modelo para $\mathbf{g}(\boldsymbol{\gamma})$ é menos direta, principalmente porque, para alguns indivíduos, os tempos de medição podem ser posteriores ao tempo do evento F e, dependendo da natureza do evento em questão, a ocorrência de F pode ou não encerrar o processo de medição (7).

4.1 Função Likelihood

Por um lado, é usual calcular-se a contribuição de verosimilhança para cada indivíduo por meio da fatorização de modelos de seleção, $[\mathbf{Y}, S] = [\mathbf{Y}] \times [S|\mathbf{Y}]$.

A verosimilhança da sequência de medição longitudinal do indivíduo i é representada pela distribuição gaussiana multivariada $\mathbf{Y}_i \sim MVN(\boldsymbol{\mu}_{\mathbf{Y}_i}, \mathbf{V}_i(\boldsymbol{\theta}))$. A verosimilhança do tempo do evento, F_i , é uma densidade gaussiana univariada, uma probabilidade de cauda ou a diferença entre duas probabilidades de cauda, dependentemente de F_i ser observado, censurado à direita ou censurado por intervalo.

Segue, então, que $S | \mathbf{Y} \sim N(\mu_{S|\mathbf{Y}}, \sigma_{S|\mathbf{Y}}^2)$, onde

$$\mu_{S|\mathbf{Y}} = \mu_S + \mathbf{g}^T(\boldsymbol{\gamma})\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}) \quad \text{e} \quad \sigma_{S|\mathbf{Y}}^2 = \eta^2 - \mathbf{g}^T(\boldsymbol{\gamma})\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{g}(\boldsymbol{\gamma}).$$

A contribuição de verosimilhança de um tempo do evento observado $F_i = f_i$ é, então, a densidade de $N(\mu_{S|\mathbf{Y}}, \sigma_{S|\mathbf{Y}}^2)$ avaliada em $s_i = \log f_i$. Para um tempo censurado à direita, $F_i > f_i$, a contribuição é $1 - \Phi\{(s_i - \mu_{S|\mathbf{Y}_i})/\sigma_{S|\mathbf{Y}}\}$. Para um censurado por intervalo $f_i < F_i < f'_i$ a contribuição é $[\Phi\{(s'_i - \mu_{S|\mathbf{Y}_i})/\sigma_{S|\mathbf{Y}}\} - \Phi\{(s_i - \mu_{S|\mathbf{Y}_i})/\sigma_{S|\mathbf{Y}}\}]$, onde $s'_i = \log f'_i$.

Estes cálculos produzem a probabilidade correta de ocorrência, ou não, do evento.

Por outro lado, à semelhança do que acontece nos modelos de dados longitudinais, a matriz $\boldsymbol{\Sigma}$ pode ser decomposta num produto de matrizes mais simples

$$\boldsymbol{\Sigma} = \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \sigma_n \\ & & & & \eta \end{bmatrix}}_W \underbrace{\begin{bmatrix} \mathbf{R}(\boldsymbol{\phi}) & \mathbf{h}(\boldsymbol{\rho}) \\ \mathbf{h}^T(\boldsymbol{\rho}) & 1 \end{bmatrix}}_C \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \sigma_n \\ & & & & \eta \end{bmatrix}}_{W^T}$$

onde W é uma matriz diagonal cujos elementos não nulos representam a estrutura de variância do modelo conjunto e C é a matriz de correlação, ou seja, uma matriz definida positiva com todos os elementos da diagonal iguais a um.

Esta decomposição em duas componentes fornece ao modelo uma maior flexibilidade na medida em que permite a modelação destas duas estruturas de correlação separadamente.

Seja $\mathbf{Z}_i = (\mathbf{Y}_i, \log F_i)$. A função de verosimilhança é dada por

$$L(\text{parâmetros}|\mathbf{Z}) = \prod_{i=1}^m f((y_{i1}, y_{i2}, \dots, y_{in_i}, s_i))$$

$$\begin{aligned}
&= \prod_{i=1}^m \frac{1}{(2\pi)^{\frac{n_i+1}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{Z}_i - \boldsymbol{\mu}_i)^T (\boldsymbol{\Sigma}_i)^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}_i)\right\} \\
&= \prod_{i=1}^m \frac{1}{(2\pi)^{\frac{n_i+1}{2}} |\mathbf{W}_i| |\mathbf{C}_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \left(\mathbf{W}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}_i)\right)^T \mathbf{C}_i^{-1} \left(\mathbf{W}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}_i)\right)\right\}.
\end{aligned}$$

Assumir-se-á, de agora em diante, que o processo de medição é estacionário, isto é, que a variância é constante ao longo do tempo - $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$. Além disso,

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{D}^{-1} & -\mathbf{D}^{-1} \mathbf{h}(\boldsymbol{\rho}) \\ -\mathbf{h}^T(\boldsymbol{\rho}) \mathbf{D}^{-1} & 1 + \mathbf{h}^T(\boldsymbol{\rho}) \mathbf{D}^{-1} \mathbf{h}(\boldsymbol{\rho}) \end{bmatrix}$$

onde $\mathbf{D} = \mathbf{R}(\boldsymbol{\phi}) - \mathbf{h}(\boldsymbol{\rho}) \mathbf{h}^T(\boldsymbol{\rho})$ apenas depende de parâmetros de correlação.

Deste modo, o logaritmo da função de verosimilhança (log-verosimilhança) é dado por

$$\begin{aligned}
&\sum_{i=1}^m -\frac{n_i+1}{2} \log(2\pi) - n_i \log(\sigma) - \log(\eta) - \frac{1}{2} \log(|\mathbf{D}_i|) - \frac{1}{2\sigma^2} (\mathbf{Y}_i - \boldsymbol{\mu}_{\mathbf{Y}_i})^T \mathbf{D}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_{\mathbf{Y}_i}) \\
&\quad + \frac{1}{\sigma\eta} (S_i - \mu_{S_i}) \mathbf{h}^T \mathbf{D}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_{\mathbf{Y}_i}) - \frac{1}{2\eta^2} (S_i - \mu_{S_i})^2 (1 + \mathbf{h}^T \mathbf{D}^{-1} \mathbf{h})
\end{aligned}$$

Note-se que a obtenção dos parâmetros ótimos, utilizando o método de inferência baseado em verosimilhança, requer técnicas numéricas sofisticadas.

4.2 Estudo de Estrutura de Correlação

Nesta secção são sugeridas e analisadas algumas estruturas de correlações possíveis para o modelo Gaussiano Transformado e ainda, estabelecidas relações entre a correlação do processo de medidas repetidas e a estrutura de correlação conjunta, assumindo que a variância é constante ao longo do tempo.

Neste enquadramento, a matriz de variância-covariância do modelo pode ser reescrita da seguinte forma

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 \mathbf{R}(\boldsymbol{\phi}) & \sigma\eta \mathbf{h}(\boldsymbol{\rho}) \\ \sigma\eta \mathbf{h}^T(\boldsymbol{\rho}) & \eta^2 \end{bmatrix} \quad (4.2)$$

onde $\mathbf{R}(\boldsymbol{\phi})$ é a matriz de correlação do processo de medição e $\mathbf{h}(\boldsymbol{\rho})$ é o vetor de correlações entre \mathbf{Y} e S .

Questão de interesse: Para que vetores de \mathbf{h} , matrizes \mathbf{R} e escalares σ^2 e η^2 a matriz Σ é definida positiva?

A condição necessária e suficiente é

$$\Leftrightarrow \|\mathbf{h}\|_2^2 < \frac{1}{\|\mathbf{R}^{-1}\|_2} \quad \text{ou} \quad \|\mathbf{R}^{-1}\|_2 < \frac{1}{\|\mathbf{h}\|_2^2}. \quad (4.3)$$

Demonstração.

Seja $\Sigma = \begin{bmatrix} \mathbf{V}(\boldsymbol{\theta}) & \mathbf{g}(\gamma) \\ \mathbf{g}^T(\gamma) & \eta^2 \end{bmatrix}$ uma matriz definida positiva. Por definição, $\forall \mathbf{x} \in \mathbb{R}^{n+1} \setminus \{0\}$, $\mathbf{x}^T \Sigma \mathbf{x} > 0$.

Note-se que, de Σ ser uma matriz definida positiva, em particular, \mathbf{V} também é uma matriz definida positiva (Teorema 2.3.3) e, portanto, $\forall \mathbf{z} \in \mathbb{R}^n \setminus \{0\}$, $\mathbf{z}^T \mathbf{V} \mathbf{z} > 0$.

Seja $\mathbf{x} = \begin{bmatrix} \mathbf{z} \\ \alpha \end{bmatrix}$ em que ($\mathbf{z} = \mathbf{0}^n$ e $\alpha \in \mathbb{R} \setminus \{0\}$) ou ($\mathbf{z} \in \mathbb{R}^n \setminus \{0\}$ e $\alpha \in \mathbb{R}$).

- Caso $\mathbf{z} = \mathbf{0}^n$ e $\alpha \in \mathbb{R} \setminus \{0\}$:

$$\mathbf{x}^T \Sigma \mathbf{x} = \begin{bmatrix} \mathbf{0}^n & \alpha \end{bmatrix} \begin{bmatrix} \mathbf{V}(\boldsymbol{\theta}) & \mathbf{g}(\gamma) \\ \mathbf{g}^T(\gamma) & \eta^2 \end{bmatrix} \begin{bmatrix} \mathbf{0}^n \\ \alpha \end{bmatrix} = \eta^2 \alpha^2 > 0.$$

Chegamos, assim, a uma condição universal.

- Caso $\mathbf{z} \in \mathbb{R}^n \setminus \{0\}$ e $\alpha \in \mathbb{R}$:

Então,

$$\mathbf{x}^T \Sigma \mathbf{x} = \begin{bmatrix} \mathbf{z} & \alpha \end{bmatrix} \begin{bmatrix} \mathbf{V}(\boldsymbol{\theta}) & \mathbf{g}(\gamma) \\ \mathbf{g}^T(\gamma) & \eta^2 \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \alpha \end{bmatrix} = \eta^2 \alpha^2 + (2\mathbf{g}^T \mathbf{z})\alpha + \mathbf{z}^T \mathbf{V} \mathbf{z} > 0.$$

Sem perda de generalidade, assumir-se-á que $\|\mathbf{z}\|_2 = 1$. E, portanto, $\mathbf{g}^T \mathbf{z} \leq \|\mathbf{g}\|_2$ para todo \mathbf{z} , isto é,

$$\eta^2 \alpha^2 + 2\|\mathbf{g}\|_2 \alpha + \mathbf{z}^T \mathbf{V} \mathbf{z} > 0. \quad (4.4)$$

A equação 4.4 de 2º grau, em α , é satisfeita, se e só se, não tem raízes reais e, isto acontece, se e só se, o binómio discriminante $4\|\mathbf{g}\|_2^2 - 4\eta^2(\mathbf{z}^T \mathbf{V} \mathbf{z})$ for negativo. Daqui resulta

$$\frac{\|\mathbf{g}\|_2^2}{\eta^2} < \mathbf{z}^T \mathbf{V} \mathbf{z}.$$

Uma vez que, pela equação 2.1,

$$\lambda_{\min}(\mathbf{V}) = \min_{\|\mathbf{z}\|_2=1} \mathbf{z}^T \mathbf{V} \mathbf{z},$$

obtem-se

$$\frac{\mathbf{g}^T \mathbf{g}}{\eta^2} < \lambda_{\min}(\mathbf{V}),$$

onde $\lambda_{\min}(\mathbf{V})$ é menor valor próprio de \mathbf{V} .

Deste modo, através da equação 4.2

$$\begin{aligned} & \frac{(\sigma\eta\mathbf{h}^T)(\sigma\eta\mathbf{h})}{\eta^2} < \lambda_{\min}(\mathbf{V}) \\ \Leftrightarrow & (\sigma\mathbf{h}^T)(\sigma\mathbf{h}) < \sigma^2\lambda_{\min}(\mathbf{R}) \quad (\text{Propriedade 7}) \\ \Leftrightarrow & \mathbf{h}^T \mathbf{h} < \lambda_{\min}(\mathbf{R}) = \frac{1}{\lambda_{\max}(\mathbf{R}^{-1})} = \frac{1}{\rho(\mathbf{R}^{-1})} = \frac{1}{\rho(\mathbf{R}^{-1})}, \end{aligned}$$

onde $\rho(\mathbf{R}^{-1})$ é o raio espectral de \mathbf{R}^{-1}

$$\Leftrightarrow \mathbf{h}^T \mathbf{h} < \frac{1}{\|\mathbf{R}^{-1}\|_2}.$$

Conclui-se, assim, que a estrutura de correlação $\mathbf{R}(\phi)$ restringe as formas admissíveis para $\mathbf{h}(\rho)$, que devem respeitar a condição de que

$$\Leftrightarrow \|\mathbf{h}\|_2^2 < \frac{1}{\|\mathbf{R}^{-1}\|_2} \quad \text{ou} \quad \|\mathbf{R}^{-1}\|_2 < \frac{1}{\|\mathbf{h}\|_2^2}$$

onde $\|\bullet\|$ é a norma matricial induzida pela norma euclidiana. \square

Este resultado indica que o modelo pressupõe que, perante correlações altas para as medidas longitudinais, não há “espaço” suficiente para correlacionar estas medidas com outra variável, nomeadamente, com o tempo de sobrevivência. E vice-versa.

Consequentemente, poderíamos utilizar um método diagnóstico onde se verificaria, para uma base de dados real, se esse pressuposto é satisfeito, isto é, se alta correlação entre o processo longitudinal implicaria correlações entre longitudinal e de sobrevivência baixas e vice-versa.

Em suma, apenas se deve aplicar o modelo Gaussiano Transformado se este pressuposto for satisfeito. Donde, sugere-se estudar, em primeiro lugar, a estrutura empírica dos dados para o processo longitudinal, matriz \mathbf{R} , e depois utilizar um vetor \mathbf{h} nas condições, agora, exigidas.

Na literatura estatística, existem grandes contribuições para estruturas de correlações longitudinais (ver secção 3.2), ao contrário do que acontece na componente conjunta. Neste sentido, são propostas e analisadas possíveis estruturas de correlação para a estrutura de correlação conjunta, cuja interpretação é bastante intuitiva. A ideia principal é caracterizá-la como sendo uma função de correlação que, além de depender de um conjunto finito de

parâmetros, necessita dos tempos em que as observações são medidas, tendo o último tempo um papel fundamental.

Chamemos às duas primeiras função de **correlação conjunta exponencial** e função de **correlação conjunta gaussiana**, respetivamente.

1. $Corr(Y_{ij}, S_i) = \text{função}(\rho, t_{ij}, t_{in_i}) = \exp\left\{-\frac{1}{\rho} |t_{ij} - t_{in_i}|\right\}$ (exponencial)
2. $Corr(Y_{ij}, S_i) = \text{função}(\rho, t_{ij}, t_{in_i}) = \exp\left\{-\frac{1}{\rho}(t_{ij} - t_{in_i})^2\right\}$ (gaussiana)

Para exemplificar, denote-se $u = t_{in_i} - t_{ij}$ a distância temporal entre o tempo t_{ij} e o último tempo de medição do indivíduo i . Suponha-se que esta distância, para qualquer indivíduo, é, no máximo, 8 e consideremos $\rho = 1, 3, 5$ e 7 .

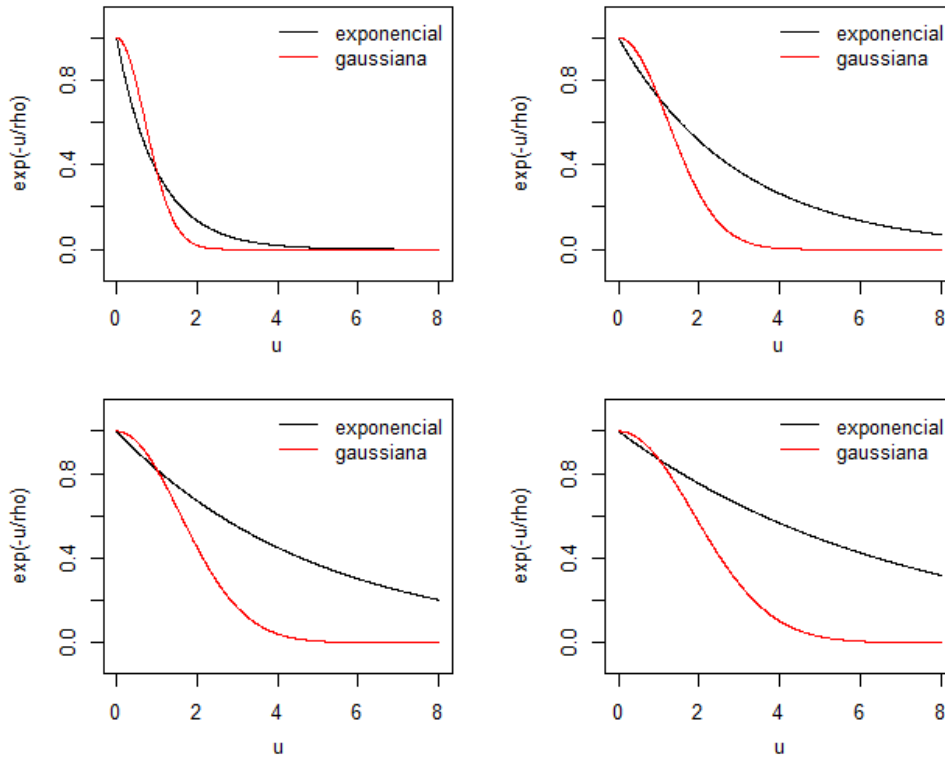


Figura 4.1: Função de correlação exponencial e gaussiana para $\rho = \frac{1}{5}, \frac{3}{7}$.

Como expectável, para tempos próximos temos correlações altas e à medida que a distância temporal entre t_{in_i} e t_{ij} aumenta, as correlações vão diminuindo.

Na correlação gaussiana, destaca-se a presença de uma distância a partir da qual as correlações deixam de ter impacto. Por exemplo, para $\rho = 5$, as correlações que estão a uma distância temporal de 1 e de 2 são significativamente diferentes, aproximadamente

0.9 e 0.4, respetivamente, o que não acontece para as distâncias temporais de 5 e de 6, ambas aproximadamente 0.1. Isto acontece quando a função começa a ter maior curvatura e, portanto, para distâncias temporais maiores, o impacto de u é menor, relativamente à correlação.

De realçar que, em ambas as correlações, $Corr(Y_{in_i}, S_i) = 1$.

As duas correlações seguintes são muito semelhantes às correlações 1. e 2. - funções de correlação conjunta exponencial e gaussiana não dos tempos mas, agora, dos logaritmos dos tempos.

$$\begin{aligned}
3. \quad Corr(Y_{ij}, S_i) &= \text{função } (\rho, \log(t_{ij}), \log(t_{in_i})) \\
&= \exp\left\{-\frac{1}{\rho} \left| \log(t_{ij}) - \log(t_{in_i}) \right| \right\} \quad (\text{exponencial dos log}) \\
&= \exp\left\{-\frac{1}{\rho} \left| \log\left(\frac{t_{ij}}{t_{in_i}}\right) \right| \right\}
\end{aligned}$$

$$\text{Como } 0 < t_{ij} < t_{in_i} \Leftrightarrow 0 < \frac{t_{ij}}{t_{in_i}} < 1$$

$$\begin{aligned}
&= \exp\left\{\frac{1}{\rho} \log\left(\frac{t_{ij}}{t_{in_i}}\right)\right\} \\
&= \exp\left\{\log\left(\frac{t_{ij}}{t_{in_i}}\right)^{\frac{1}{\rho}}\right\} \\
&= \left(\frac{t_{ij}}{t_{in_i}}\right)^{\frac{1}{\rho}}
\end{aligned}$$

$$\begin{aligned}
4. \quad Corr(Y_{ij}, S_i) &= \text{função } (\rho, \log(t_{ij}), \log(t_{in_i})) \\
&= \exp\left\{-\frac{1}{\rho} (\log(t_{ij}) - \log(t_{in_i}))^2\right\} \quad (\text{gaussiana dos log}) \\
&= \exp\left\{-\frac{1}{\rho} \left(\log\left(\frac{t_{ij}}{t_{in_i}}\right)\right)^2\right\}
\end{aligned}$$

Defina-se $v = \frac{t_{ij}}{t_{in_i}}$, que varia de 0 a 1 e, a título de exemplo, consideremos, novamente, $\rho = 1, 3, 5$ e 7 .

Para as funções de correlação 1. e 2. existe uma distância a partir da qual as correlações deixavam de ter impacto. Para estes casos, a figura 4.2 sugere que existe uma distância até à qual as correlações não são muito distintas, com correlações bastante altas. Novamente, em $\rho = 5$, é notável que quando $t_{ij} > 0.4t_{in_i}$ as correlações são bastante próximas atinguindo valores compreendidos entre 0.8 e 1.

De salientar que, para ambas as correlações, $Corr(Y_{in_i}, S_i) = 1$.

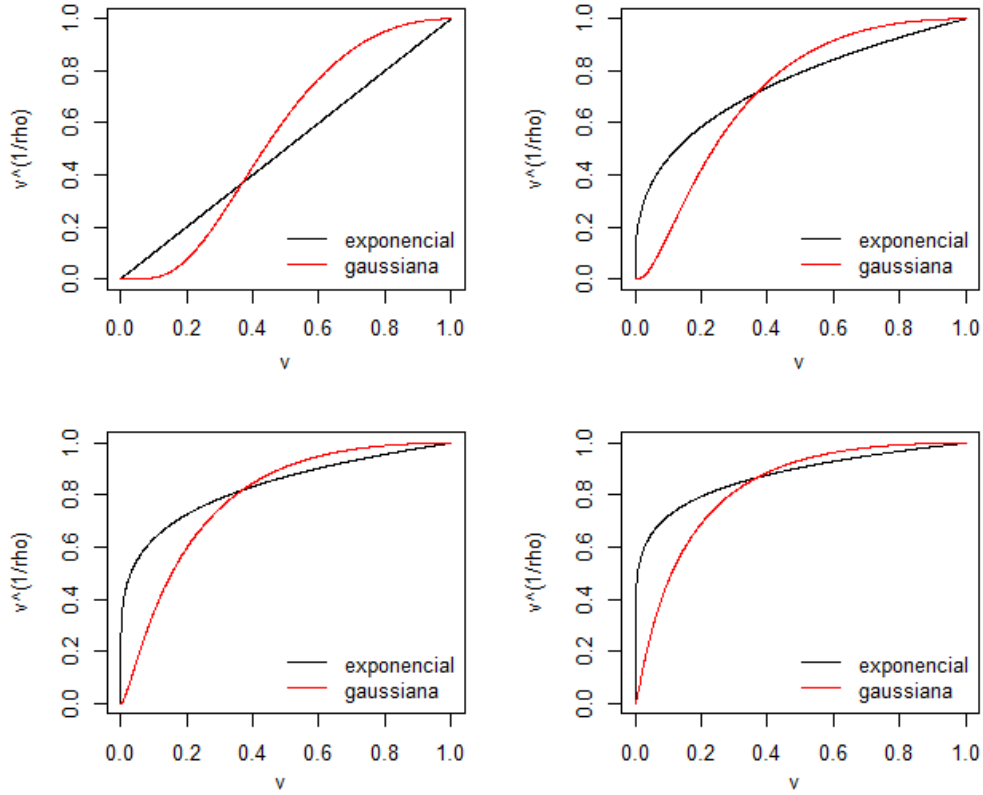


Figura 4.2: Função de correlação exponencial e gaussiana dos logaritmos dos tempos para $\rho = \begin{matrix} 1 & 3 \\ 5 & 7 \end{matrix}$.

5. $Corr(Y_{ij}, S_i) =$ função $(\rho, t_{ij}, t_{in_i}) = \rho^{|t_{in_i} - t_{ij}|}$ (autoregressivo)

6. $Corr(Y_{ij}, S_i) =$ função $(\rho, t_{ij}, t_{in_i}) = \rho$ (constante)

7. $Corr(Y_{ij}, S_i) =$ função $(\rho, t_{ij}, t_{in_i}) = Corr(Y_{ij}, Y_{in_i}) \times \rho$

Note-se que esta função depende da correlação das medidas repetidas. Uma interpretação aproximada disto é que a associação entre os processos de medição e tempo até ao evento não pode ser mais forte do que a associação dentro do processo de medição.

8. $Corr(Y_{ij}, S_i) =$ função $(\rho, \delta, t_{ij}, t_{in_i}) = \exp\left\{-\frac{1}{\rho} |t_{ij} - (t_{in_i} + \delta)|\right\}$

9. $Corr(Y_{ij}, S_i) =$ função $(\rho, t_{ij}, t_{in_i}) = \exp\left\{-\frac{1}{\rho}(t_{ij} - t_{in_i} + \delta)^2\right\}$

Estas função de correlação são muito semelhantes às funções 1. e 2.. A ideia aqui representada é, em vez de se utilizar o último tempo de cada indivíduo, é-lhe adicionado um parâmetro de tempo δ para, assim, existir a possibilidade de estimar, em média, o espaçamento da última observação e o tempo do evento de interesse.

Em termos práticos, para encontrar um modelo apropriado que se ajuste a uma base de dados é necessário definir que estrutura de correlação se deve utilizar em $\mathbf{h}(\boldsymbol{\rho})$. Explorar a estrutura de correlação empírica, geralmente, fornece boas indicações de qual poderá ser a verdadeira estrutura de correlação dos dados apenas quando estamos a trabalhar em dados que para todos os indivíduos ocorreu o evento de interesse.

Deste modo, surge a ideia de analisar o vetor empírico das correlações entre as observações de cada tempo com o tempo até ao evento de interesse quando estamos perante um estudo balanceado. Em estudos observacionais, em que os indivíduos podem ser medidos em tempos distintos, não conseguimos fazer esse tipo de análise exploratória.

Para estes casos, e apenas quando todas as observações não são censuradas, uma “espécie” de variograma definido por

$$\begin{aligned} V(u) &= \frac{1}{2} \text{var}\{S_i - Y(t_{in_i} - u)\} \\ &= \frac{1}{2} \{\eta^2 + \sigma^2 - 2g(u)\} \\ &= \frac{1}{2} \{\eta^2 + \sigma^2 - 2\eta\sigma h(u)\} \\ &= \frac{(\eta - \sigma)^2}{2} + \eta\sigma(1 - h(u)) \end{aligned}$$

poderá ser considerado.

4.3 Estudo de Simulação

Através de estudos de simulação, esta seção fornece um estudo de casos particulares de estruturas de correlação para o modelo Gaussiano Transformado, assumindo correlações específicas, descritas nos subcapítulos e capítulos anteriores, quer para estruturas de correlação longitudinal, quer para estrutura de correlação conjunta.

Correlação exponencial para $\mathbf{h}(\boldsymbol{\phi})$

Assumindo que estamos perante dados balanceados com $t_{ij} = t_j, \forall j = 1, \dots, n$ e todos os valores ausentes de \mathbf{Y}_i medidos nos tempos $t_j < S_i$ são considerados completamente aleatórios (*missing completely at random*), vamos agora descrever com algum detalhe a relação dos parâmetros utilizados para modelar a estrutura de correlação de Y , $\mathbf{R}(\boldsymbol{\phi})$, com $\boldsymbol{\rho}$ presente na

estrutura de correlação exponencial entre \mathbf{Y} e \mathbf{S} dada por

$$h_j = \text{Corr}(Y_{ij}, S_i) = \exp\left\{-\frac{1}{\rho} |t_j - t_n|\right\}, \forall j \in 1, \dots, n. \quad (4.5)$$

Primeiramente, é analisado o caso em que os tempos são igualmente espaçados, isto é, $t_j - t_{j-1} = a, \forall j \in 2, \dots, n$. Deste modo,

$$h_j = \text{Corr}(Y_{ij}, S_i) = \exp\left\{-\frac{(n-j)a}{\rho}\right\}, \forall j \in 1, \dots, n. \quad (4.6)$$

Ora,

$$\begin{aligned} \|\mathbf{h}\|_2^2 &= \sum_{j=1}^n h_j^2 = \sum_{j=1}^n \left(\epsilon^{-\frac{(n-j)a}{\rho}}\right)^2 = \sum_{j=1}^n h_j^2 = \sum_{j=1}^n \epsilon^{-\frac{2(n-j)a}{\rho}} \\ &= \epsilon^{-\frac{2na}{\rho}} \sum_{j=1}^n \epsilon^{\frac{2ja}{\rho}} = \epsilon^{-\frac{2na}{\rho}} \sum_{j=1}^n \left(\epsilon^{\frac{2a}{\rho}}\right)^j \\ &= \frac{\epsilon^{-\frac{2na}{\rho}} \epsilon^{\frac{2a}{\rho}} \left(\epsilon^{\frac{2na}{\rho}} - 1\right)}{\epsilon^{\frac{2a}{\rho}} - 1} = \frac{\epsilon^{-\frac{2na}{\rho}} \left(\epsilon^{\frac{2na}{\rho}} - 1\right)}{1 - \epsilon^{-\frac{2a}{\rho}}} = \frac{1 - \epsilon^{-\frac{2na}{\rho}}}{1 - \epsilon^{-\frac{2a}{\rho}}} \\ &= \frac{1 - \left(\epsilon^{-\frac{2a}{\rho}}\right)^n}{1 - \epsilon^{-\frac{2a}{\rho}}} = \frac{1 - \left(h_{n-1}^2\right)^n}{1 - h_{n-1}^2} = \left(h_{n-1}^2\right)^{n-1} + \left(h_{n-1}^2\right)^{n-2} + \dots + h_{n-1}^2 + 1 \end{aligned}$$

Logo,

$$\|\mathbf{R}^{-1}\|_2 < \frac{1}{\|\mathbf{h}\|_2^2} = \frac{1}{\left(h_{n-1}^2\right)^{n-1} + \left(h_{n-1}^2\right)^{n-2} + \dots + h_{n-1}^2 + 1} \leq 1.$$

Quanto maior for a penúltima correlação, menor serão as normas definidas em \mathbf{R} . Se considerarmos a correlação constante para $\mathbf{h}(\rho)$ dada por:

$$h_j = \text{Corr}(Y_{ij}, S_i) = \rho, \forall j \in 1, \dots, n. \quad (4.7)$$

Ora,

$$\|\mathbf{h}\|_2^2 = \sum_{j=1}^n h_j^2 = \sum_{j=1}^n \rho^2 = n\rho^2.$$

Logo,

$$\|\mathbf{R}^{-1}\|_2 < \frac{1}{\|\mathbf{h}\|_2^2} = \frac{1}{n\rho^2}.$$

Comece-se, agora, por estabelecer uma correlação longitudinal particular e retirar respectivas propriedades da estrutura conjunta.

- Correlação Independente

Este caso assume a estrutura de correlação independente para a matriz \mathbf{R} , isto é, uma matriz diagonal em que todos os elementos da diagonal principal são 1 e, portanto, $\|\mathbf{h}\|^2 < 1$.

$$\mathbf{R}_i(\phi) = \begin{bmatrix} 1 & & & \\ & 1 & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \ddots \\ & & & & 1 \end{bmatrix}$$

Ora, $\|\mathbf{R}^{-1}\|_2 = 1 \Rightarrow \|\mathbf{h}\|_2^2 < 1$. O que implica que nenhuma das correlações conjuntas pode ser igual a 1.

- Correlação Simetria composta (*Compound Symmetry*)

$$\mathbf{R}_i(\phi) = \begin{bmatrix} 1 & & & \\ & 1 & & \phi \\ & & \ddots & \\ & \phi & & \ddots \\ & & & & 1 \end{bmatrix}$$

Ora,

$$\mathbf{R}_i^{-1}(\phi) = a \begin{bmatrix} -(n-2)\phi - 1 & & & \\ & -(n-2)\phi - 1 & & \phi \\ & & \ddots & \\ & \phi & & \ddots \\ & & & & -(n-2)\phi - 1 \end{bmatrix}$$

onde $a = \frac{1}{(n-1)\phi^2 - (n-2)\phi - 1}$. Deste modo, $\|\mathbf{R}^{-1}\|_2 = \frac{1}{1-\phi}$ e, portanto, $\|\mathbf{h}\|_2^2 < 1 - \phi$.

Note-se que quanto maior for o valor de ϕ , menor poderá ser o comprimento do vetor h e, ainda, a correlação exponencial não poderá definir h .

- Correlação Autoregressiva para tempos igualmente espaçados

Se $\mathbf{R}_i(\phi) = \text{Corr}(Y_{ij}, Y_{ik}) = \phi^{|j-k|}$ então $\|\mathbf{R}^{-1}\|_2 > \frac{1}{1-\phi}$. Logo, conclui-se que $\|\mathbf{h}\|_2^2 < 1 - \phi$.

Capítulo 5

Conclusões e Trabalho Futuro

Os modelos para dados longitudinais são utilizados para medidas repetidas em indivíduos ao longo do tempo. Estes permitem explorar como variáveis mudam ao longo do tempo e como as variações no tempo são correlacionadas entre si. A escolha da estrutura de correlação apropriada é crucial para a precisão e a validade dos resultados. Uma seleção incorreta pode levar a estimativas tendenciosas e a interpretações erradas. Portanto, é importante selecionar cuidadosamente a estrutura de correlação apropriada para garantir análises precisas e confiáveis.

O Modelo Gaussiano Transformado é um caso particular de modelos conjuntos conhecido pela sua simplicidade. Esta simplicidade é estendida para inferência baseada em verosimilhança, ajuste de eventos censurados à direita e censurados por intervalo, e diagnósticos facilmente verificáveis.

Além disso, através da decomposição da matriz de correlação em três componentes de correlação, é possível admitir um novo pressuposto para este modelo, o qual é definido pela expressão $\|\mathbf{R}^{-1}\|_2 < \frac{1}{\|\mathbf{k}\|_2^2}$, ou dito de outra maneira, existe uma relação inversa entre os “comprimentos” das estruturas de correlação longitudinal e conjunta. Com isso, é possível descrever a associação entre observações de medidas repetidas e tempos de eventos para este modelo, desde que sejam considerados dados balanceados, no sentido em que os tempos de medição para qualquer indivíduo devem ser um subconjunto de um único conjunto finito.

No entanto, uma das principais desvantagens deste modelo é quando o tempo do evento acaba a sequência de medição, como em caso de o evento de interesse ser a morte.

Como resultado deste pressuposto adicional, uma futura proposta de trabalho é encontrar um método diagnóstico que verificar, para uma base de dados real, se este é satisfeito.

Referências Bibliográficas

- [1] J. Barrett, P. Diggle, R. Henderson, and D. Taylor-Robinson. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B*, 77(1):131–148, 2015.
- [2] M. S. Cabral and M. H. Gonçalves. *Análise de Dados Longitudinais*. SPE, 2011.
- [3] E. K. Chong and S. H. Zak. *An introduction to optimization*. John Wiley & Sons, 2004.
- [4] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.
- [5] T. X. De Gruttola V. Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, Dec 1994.
- [6] P. Diggle, P. Heagerty, K.-Y. Liang, S. Zeger, et al. *Analysis of longitudinal data*. Oxford university press, 2002.
- [7] P. J. Diggle, I. Sousa, and A. G. Chetwynd. Joint modelling of repeated measurements and time-to-event outcomes: The fourth armitage lecture. *Statistics in Medicine*, 27(16):2981–2998, Jul 20 2008.
- [8] M. A. Erango. Comparison analysis of separate and joint models in case of time-to-death event of hiv/aids patients under art follow-up. *Open Access Medical Statistics*, 2018.
- [9] C. L. Faucett and D. C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15 15:1663–85, 1996.
- [10] X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The american statistician*, 58(1):16–24, 2004.
- [11] D. Harville. Bayesian inference for variance components using only error contrasts. *Biometrika*, 61(2):383–385, 1974.
- [12] L. Hatfield, J. Hodges, and B. Carlin. Joint models: when are treatment estimates improved? *Statistics and its interface*, 7:439–453, 01 2014.
- [13] S. L. Hedden, R. F. Woolson, R. E. Carter, Y. Palesch, H. P. Upadhyaya, and R. J. Malcolm. The impact of loss to follow-up on hypothesis tests of the treatment effect for

- several statistical methods in substance abuse clinical trials. *Journal of Substance Abuse Treatment*, 37(1):54–63, Jul 2009.
- [14] R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.
- [15] J. G. Ibrahim, H. Chu, and L. M. Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801, Jun 2010.
- [16] J. P. Klein and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.
- [17] D. G. Kleinbaum, M. Klein, et al. *Survival analysis: a self-learning text*, volume 3. Springer, 2012.
- [18] E. T. Lee and J. Wang. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.
- [19] A. R. Moura. Critérios de seleção de modelos: Um estudo comparativo. Master’s thesis, Jul 2021.
- [20] G. Papageorgiou. Restricted maximum likelihood estimation of joint mean-covariance models. *Canadian Journal Of Statistics-Revue Canadienne de Statistique*, 40(2):225–242, Jun 2012.
- [21] J. C. Pinheiro and D. M. Bates. Springer Science Business Media, 2000.
- [22] D. Rizopoulos. *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press, 2012.
- [23] L. RJa and D. Rubin. Statistical analysis with missing data. *New York*, 1987.
- [24] J. M. Robins, A. Rotnitzky, and L. P. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association*, 90(429):106–121, 1995.
- [25] J. Schafer and J. Graham. Missing data: Our view of the state of the art. *Psychological Methods*, 7(2):147–177, Jun 2002.
- [26] J. Schafer and M. Olsen. Multiple imputation for multivariate missing-data problems: A data analyst’s perspective. *Multivariate Behavioral Research*, 33(4):545–571, 1998.

- [27] S. Self and Y. Pawitan. Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology*, pages 231–255. Springer, 1992.
- [28] P. Serranho. *Notas de análise de sobrevivência: uma introdução com R*. Repositório Aberto da Universidade Aberta, 2015.
- [29] J. M. Singer, J. S. Nobre, and F. M. M. Rocha. Universidade de São Paulo, Departamento de Estatística, Jun 2018.
- [30] I. Sousa. A review on joint modelling of longitudinal measurements and time-to-event. *Revstat-Statistical Journal*, 9(1):57, Mar 2011.
- [31] L. Teixeira, I. Sousa, A. Rodrigues, and D. Mendonça. Joint modelling of longitudinal and competing risks data in clinical research. 2019.
- [32] M. Thulasidas. *Linear Algebra for Computer Science*. Asian Books, Dez 2021.
- [33] A. Tsiatis and M. Davidian. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14(3):809–834, Jul 2004.
- [34] P. Wang, Y. Li, and C. K. Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 51(6), FEB 2019.
- [35] P. R. Williamson, R. Kolamunnage-Dona, P. Philipson, and A. G. Marson. Joint modelling of longitudinal and competing risks data. *Statistics in medicine*, 27(30):6426–6438, 2008.
- [36] M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53 1:330–9, 1997.
- [37] D. L. Zimmerman. Viewing the correlation structure of longitudinal data through a prism. *American Statistician*, 54(4):310–318, Nov 2000.