

# Modelos de Regressão Linear Mistos para dados discretos: Uma abordagem utilizando MCMC através do Stan integrado ao R.

Felipe Vieira - RA 160424

Guilherme Artoni - RA 160318

The Stan project develops a probabilistic programming language that implements full Bayesian statistical inference via Markov Chain Monte Carlo, rough Bayesian inference via ‘variational’ approximation, and (optionally penalized) maximum likelihood estimation via optimization. In all three cases, automatic differentiation is used to quickly and accurately evaluate gradients without burdening the user with the need to derive the partial derivatives.

## 1 INTRODUÇÃO

A degeneração macular relacionada à idade (DMRI) é uma doença atualmente sem cura que ocorre em uma parte da retina chamada mácula e que leva a perda progressiva da visão central. A DMRI é uma alteração muito comum em pessoas com mais de 55 anos, sendo a causa mais frequente de baixa acuidade visual nessa faixa etária. Com o intuito de avaliar se um novo medicamento para DMRI tem poder competitivo com o principal existente atualmente, iremos comparar por meio de modelos de regressão linear a qualidade da acuidade visual de pacientes com DMRI sob ambos tratamentos durante aproximadamente dois meses.

### MOTIVAÇÃO SOBRE O MODELO:

- Modelo com efeitos fixos e aleatórios.
- Considerar variabilidade de perfis individuais.
- Estudar e estimar componentes da variância descompactando e adicionando complexidade aos erros.
- Resolver problemas de interdependência dos dados.

### Importante em diversas disciplinas

- ANOVA com efeitos mistos (Estatística, Econometria)
- Modelos lineares hierárquicos (Educação)
- Modelos de efeitos contextuais (Sociologia)

### MOTIVAÇÃO SOBRE OS DADOS/EXPERIMENTO

- Dados disponibilizados pelo Grupo de Estudo de Terapias Farmacológicas para Degeneração Macular (GETFDM) em 1997.
- Trata-se de informações sobre ensaios clínicos aleatorizados realizados em diferentes centros de estudos.
- O objetivo era comparar um tratamento experimental chamado *interferon- $\alpha$*  e o placebo para pacientes diagnosticados com Degeneração Macular Relacionada a Idade (DMRI).
- Os dados mostrados são em relação ao placebo e a maior dose administrada do *interferon- $\alpha$* .

### MOTIVAÇÃO SOBRE A DOENÇA

- Pacientes com DMRI gradativamente perdem a visão.
- Durante os ensaios, a qualidade da visão de cada um dos 240 pacientes foi medida no início e após 4, 12, 24 e 52 semanas.
- A qualidade da visão foi medida através da quantidade de letras que os pacientes foram capazes de ler em gráficos de visão padronizados.
- Segue que temos dados longitudinais para cada paciente em forma de medidas da qualidade de sua visão.

No que segue apresentaremos a metodologia utilizada para investigar os desempenhos de ambos tratamentos, ou seja a estrutura do modelo uma análise descritiva dos dados, o modelo proposto, assim como os resultados obtidos com o mesmo.

## 2 METODOLOGIA

### 2.1 MODELO MISTO

Conforme discutido em Júlio M. Singer (2018), em geral podemos analisar dados provenientes de estudos com medidas repetidas por meio de modelos mistos da forma:

$$\mathbf{Y}_{j(k_j \times 1)} = \mathbf{X}_{j(k_j \times p)}\boldsymbol{\beta}_{(p \times 1)} + \mathbf{Z}_{j(k_j \times q)}\mathbf{b}_{j(q \times 1)} + \boldsymbol{\xi}_{j(k_j \times 1)}$$

Onde  $j = 1, 2, \dots, n$  é o indivíduo

- $\mathbf{Y}_j = (y_{j1}, \dots, y_{jk_j})$ , onde  $k_j$ : número de avaliações realizadas no indivíduo  $j$ .
- $\mathbf{X}_j$ : matriz de planejamento associada aos efeitos fixos para o indivíduo  $j$ .
- $\boldsymbol{\beta}$ : vetor de efeitos fixos
- $\mathbf{Z}_j$ : matriz de planejamento associada aos efeitos aleatórios para o indivíduo  $j$ .
- $\mathbf{b}_j$ : vetor de efeitos aleatórios associado ao indivíduo  $j$ .
- $\boldsymbol{\xi}_j$ : vetor de erros associado ao indivíduo  $j$ .

Em muitos casos é razoável supor que

$$\mathbf{b}_j \sim \mathcal{N}_k(\mathbf{0}, \mathcal{D}) \text{ e } \boldsymbol{\xi}_j \sim \mathcal{N}_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_j)$$

em que  $\mathcal{D}$  e  $\boldsymbol{\Sigma}_j$  são matrizes simétricas definidas positivas e, além disso, que  $\mathbf{b}_j$  e  $\boldsymbol{\xi}_j$  são variáveis aleatórias independentes.

Sob esse modelo, o vetor de respostas associado à  $i$ -ésima unidade amostral tem distribuição normal multivariada com vetor de médias e matriz de covariâncias dados, respectivamente por

$$E(\mathbf{Y}_j) = \mathbf{X}_j\boldsymbol{\beta}$$

e

$$V(\mathbf{Y}_j) = \mathbf{Z}_j\mathcal{D}\mathbf{Z}_j' + \boldsymbol{\Sigma}_j$$

Em que primeira componente  $\mathbf{Z}_j\mathcal{D}\mathbf{Z}_j'$  modela a dispersão dos perfis individuais de resposta e a segunda componente  $\boldsymbol{\Sigma}_j$  está relacionada com a dispersão da resposta em torno dos perfis individuais.

### 2.1.1 MODELO PARA A ESTRUTURA DE COVARIÂNCIA

Grande parte do esforço empregado na modelagem de dados com medidas repetidas se concentram na estrutura de covariância. Em geral, o modelo para a matriz de covariâncias  $\mathbf{Z}_j \mathcal{D} \mathbf{Z}_j' + \Sigma_j$  deve depender da maneira pela qual as observações foram obtidas e do conhecimento sobre o mecanismo gerador das observações. Quando  $\Sigma_j = \sigma^2 \mathbf{I}_{n_j}$  o modelo é chamado de modelo de independência condicional homocedástico, indicando que as  $n_j$  observações da  $j$ -ésima unidade amostral são condicionalmente independentes dado  $b_j$ .

Das possíveis estruturas de covariância disponíveis na literatura estatística, adotaremos a Estrutura Uniforme, tomando  $n_j = 4$  temos

$$\begin{aligned} \mathbf{V}_j &\equiv \mathbf{Z}_j \mathcal{D} \mathbf{Z}_j' + \sigma^2 \mathbf{I}_{n_j} \\ &= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \tau \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau \end{pmatrix} \end{aligned}$$

Pela estrutura da matriz de variâncias e covariâncias temos um coeficiente de correlação comum para todos os pares de variáveis  $\rho = \tau / (\sigma^2 + \tau)$ . Pelo fato do  $\tau$  ser não negativo, implica que  $\rho$  é também não negativo.

## 2.2 INFERÊNCIA BAYESIANA

Segundo Manly (1997) a ideia básica por trás da Inferência Bayesiana é mudar as probabilidades para os parâmetros tomando valores numéricos específicos para novas probabilidades como um resultado da coleta de mais dados, com essa mudança sendo alcançada através do Teorema de Bayes. Como um exemplo da abordagem Bayesiana, suponha que temos interesse no valor de um parâmetro  $\theta$  de uma determinada população, e que antes de qualquer informação ser observada é de alguma forma possível afirmar que  $\theta$  deve assumir um dos valores entre  $\theta_1, \theta_1, \dots, \theta_n$  e que a probabilidade de o valor ser  $\theta_i$  é  $\pi(\theta_i)$ . Suponha também que alguns dados novos são coletados e a probabilidade de observar estes dados é  $\pi(dados|\theta_i)$  se de fato  $\theta = \theta_i$ . Então o Teorema de Bayes afirma que a probabilidade de  $\theta$  ser igual a  $\theta_i$ , dado novas observações, é

$$\pi(\theta_i|dados) = \frac{\pi(dados|\theta_i)\pi(\theta_i)}{\sum_{j=1}^n \pi(dados|\theta_j)\pi(\theta_j)}, \quad (??)$$

onde  $\pi(\theta_i|dados)$  é a distribuição a posteriori de  $\theta$ . Contudo frequentemente lidamos com situações em que vários parâmetros estão envolvidos, de tal forma que no geral

$$\pi(\theta_1, \theta_2, \dots, \theta_p|dados) \propto \pi(dados|\theta_1, \theta_2, \dots, \theta_p)\pi(\theta_1, \theta_2, \dots, \theta_p), \quad (??)$$

ou seja, a distribuição a posteriori de vários parâmetros dado um conjunto de dados é proporcional a probabilidade dos dados quando conhecidos os parâmetros multiplicada pela probabilidade a priori dos parâmetros.

## 2.3 MONTE CARLO MARKOV CHAIN

### 2.3.1 INTRODUÇÃO ÀS CADEIAS DE MARKOV

Considere uma sequência de variáveis aleatórias discretas  $\{X_0, X_1, X_2, \dots\}$  e espaço de estados denotado por  $S = \{s_1, s_2, \dots, s_k\}$ . A sequência de variáveis aleatórias  $\{X_0, X_1, X_2, \dots\}$  é uma Cadeia de Markov (CM), se

$$p(X_t | X_{t-1}, X_{t-2}, \dots, X_0) = p(X_t | X_{t-1}),$$

para  $t = 1, 2, \dots$ , ou seja, dada  $X_{t-1}$ , a distribuição de  $X_t$  independe de suas predecessoras,  $X_{t-2}, X_{t-3}, \dots$ . Em um instante  $t$  qualquer, a probabilidade que o processo mude de um estado  $X_t = s_i$  para um estado  $X_{t+1} = s_j$  é dada pela matriz de transição,  $P = \{p_{ij}\}$ . A restrição natural sobre a matriz de transição é que a soma das linhas seja 1,  $\sum_j p_{ij} = 1$ , para todo  $i$ .

Uma CM admite uma distribuição de equilíbrio  $\pi$  se existir  $\pi$  tal que  $\pi = \pi P$ .

#### IRREDUTIBILIDADE

Uma CM, cujas as variáveis aleatórias são discretas, é dita irredutível se for uma cadeia em que partindo-se de um estado qualquer, pode-se atingir qualquer estado, inclusive o inicial, em um número finito de transições.

#### REVERSIBILIDADE

Uma CM é dita reversível se a probabilidade de estar em um estado  $s_i$  e mover-se para  $s_j$  é igual à probabilidade de estar no estado  $s_j$  e mover-se para o estado  $s_i$ , ou seja,

$$\pi(i)p(i, j) = \pi(j)p(j, i)$$

A denominação reversível se deve ao fato que, atendida a condição acima, a mesma lei de transição se aplica à cadeia tanto no sentido natural de evolução do tempo,  $t, t+1, t+2, \dots$ , quanto no sentido contrário  $t, t-1, t-2, \dots$ . A reversibilidade de uma CM está associada ao fato da mesma admitir uma distribuição de equilíbrio.

### 2.3.2 AMOSTRADOR DE GIBBS

O Amostrador de Gibbs (AG) é um método para aproximar uma distribuição multivariada tomando somente amostras de distribuições univariadas. O benefício deste método com Inferência Bayesiana é que torna relativamente fácil amostrar de uma distribuição a posteriori multivariada até mesmo quando o número de parâmetros envolvidos é muito grande.

Suponha que a distribuição a posteriori tenha função de densidade  $\pi(\theta_1, \theta_2, \dots, \theta_p)$  para os  $p$  parâmetros  $\theta_1, \theta_2, \dots, \theta_p$  e seja  $\pi(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$  a função densidade condicional para  $\theta_i$  dado os valores dos outros parâmetros. O problema então é gerar um grande número de amostras aleatórias da distribuição a posteriori com o objetivo de aproximar a própria distribuição e distribuições de várias funções dos parâmetros. Isto é feito tomando arbitrariamente valores iniciais  $\{\theta_1(0), \theta_2(0), \dots, \theta_p(0)\}$  para os  $p$  parâmetros e em seguida mudá-los um a um selecionando novos valores como segue:

$$\begin{aligned}\theta_1(1) & \text{ é escolhido de } \pi(\theta_1 | \theta_2(0), \theta_3(0), \dots, \theta_p(0)) \\ \theta_2(1) & \text{ é escolhido de } \pi(\theta_2 | \theta_1(1), \theta_3(0), \dots, \theta_p(0)) \\ \theta_3(1) & \text{ é escolhido de } \pi(\theta_3 | \theta_1(1), \theta_2(1), \theta_4(0), \dots, \theta_p(0)) \\ & \vdots \\ \theta_p(1) & \text{ é escolhido de } \pi(\theta_p | \theta_1(1), \theta_2(1), \dots, \theta_{p-1}(1))\end{aligned}$$

Nesse ponto todos os valores iniciais foram substituídos, o que representa um ciclo completo do algoritmo. O processo então é repetido muitas vezes produzindo a sequência  $\{\theta_1(1), \theta_2(1), \dots, \theta_p(1)\}, \{\theta_1(2), \theta_2(2), \dots, \theta_p(2)\}, \dots, \{\theta_1(N), \theta_2(N), \dots, \theta_p(N)\}$  a qual caracteriza uma CM, pois em cada etapa do algoritmo a mudança é feita dependendo apenas do valor atual de  $\theta$ .

### 2.3.3 METROPOLIS-HASTINGS

Considere uma distribuição  $\pi$  da qual se deseja gerar uma amostra através de uma CM. De acordo com Gamerman (1996) deve-se construir um núcleo de transição  $p(\theta, \phi)$  de forma que  $\pi$  seja a distribuição de equilíbrio da cadeia. Uma forma simples de fazer isso é através de cadeias onde o núcleo  $p$  satisfaça

$$\pi(\theta)p(\theta, \phi) = \pi(\phi)p(\phi, \theta), \quad \forall (\theta, \phi) \quad (??)$$

essa equação é também conhecida como equação de equilíbrio detalhado. Embora não seja necessária, ela é suficiente para que  $\pi$  seja a distribuição de equilíbrio da cadeia. Como vimos na seção ??? essa é a condição de reversibilidade da cadeia.

O núcleo  $p(\theta, \phi)$  é constituído de 2 elementos: um núcleo de transição arbitrário  $q(\theta, \phi)$  e uma probabilidade  $\alpha(\theta, \phi)$  de forma que,

$$p(\theta, \phi) = q(\theta, \phi)\alpha(\theta, \phi), \quad \text{se } \theta \neq \phi$$

Portanto, o núcleo de transição define uma densidade  $p(\theta, \cdot)$  para todos os valores diferentes de  $\theta$ . Consequentemente, resta uma probabilidade positiva da cadeia ficar em  $\theta$  dada por

$$p(\theta, \theta) = 1 - \int q(\theta, \phi)\alpha(\theta, \phi)d\phi$$

Logo, o núcleo de transição define uma distribuição mista para o novo estado  $\phi$  da cadeia. Para  $\phi \neq \theta$  essa distribuição tem densidade e para  $\phi = \theta$ , essa distribuição atribui uma probabilidade positiva.

A expressão mais comum para a probabilidade de aceitação é

$$\alpha(\theta, \phi) = \min\left\{1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}\right\}$$

onde a razão que aparece na expressão é chamada de razão de teste.

Em termos práticos isso significa que a simulação de uma amostra de  $\pi$  usando a cadeia de Markov pode ser esquematizada da seguinte forma:

- i) inicialize o contador de iterações da cadeia  $j = 1$  e tome um valor arbitrário para  $\theta^{(0)}$ ;
- ii) mova a cadeia para um novo valor  $\phi$  gerado da densidade  $q(\theta^{(j-1)}, \cdot)$ ;
- iii) calcule a probabilidade de aceitação do movimento  $\alpha(\theta^{(j-1)}, \phi)$ . Se o movimento for aceito,  $\theta^{(j)} = \phi$ , caso contrário  $\theta^{(j)} = \theta^{(j-1)}$  e a cadeia não se move;
- iv) mude o contador de  $j$  para  $j + 1$  e retorne ao item ii) até a convergência.

A etapa iii) é realizada após a geração de uma quantidade uniforme  $u \sim U(0, 1)$  independente de todas as outras variáveis. Se  $u \leq \alpha$ , o movimento é aceito e se  $u > \alpha$  o movimento não é permitido. O núcleo de transição  $q$  define apenas uma proposta de movimento que pode ou não ser confirmado por  $\alpha$ . Por esse motivo,  $q$  é normalmente chamado de proposta e, quando considerado como uma densidade ou distribuição condicional  $q(\theta, \cdot)$  é chamado de densidade ou distribuição condicional proposta.

### 2.3.4 CONEXÃO ENTRE AMOSTRADOR DE GIBBS E METROPOLIS-HASTINGS

O AG pode ser visto como um caso particular do Metropolis-Hastings (MH) onde as distribuições geradoras de candidatos são as condicionais da distribuição de interesse e o candidato é aceito com probabilidade 1. No AG a cadeia está restrita a se mover paralelamente aos eixos enquanto em MH este movimento é livre quando a geração é feita em blocos. Isto se explica em função de no AG atualizarmos um elemento do vetor aleatório por vez, enquanto que no MH isto se faz em bloco.

Considere a situação em que para um vetor aleatório com  $K$  elementos, sendo que alguns deles tem densidades condicionais bem definidas e fáceis de gerar, enquanto que os demais não tem esta facilidade. Neste caso podemos fazer uso combinado do AG para as variáveis fáceis de gerar e do MH para as demais. Esta combinação recebe o nome de Metropolis-within-Gibbs.

O procedimento Metropolis-within-Gibbs, combina aspectos interessantes de MH e AG, gerando direto da distribuição condicional quando for conveniente, caracterizando um passo do AG, e aplicando um passo MH quando a geração direta da condicional for difícil.

### 2.3.5 VANTAGENS E DESVANTAGENS DO MÉTODO

Dois fatores fazem este algoritmo útil. Primeiro que pode ser mostrado que  $\{\theta_1(i), \theta_2(i), \dots, \theta_p(i)\}$  segue a distribuição com densidade  $\pi(\theta_1, \theta_2, \dots, \theta_p)$  para valores grandes de  $i$ . Segundo que amostrar observações da distribuição condicional é frequentemente relativamente mais fácil, tornando o método de fácil implementação.

Complicações aparecem por que os conjuntos sucessivos de valores amostrais geradores podem ser correlacionados, porém pode ser resolvido tomando somente valores a partir da  $r$ -ésima etapa da sequência, com  $r$  grande o suficiente para garantir que os valores tenham correlações negligenciáveis. Paralelamente podem ser geradas várias sequências diferentes com valores iniciais escolhidos aleatoriamente e somente os conjuntos de valores finais  $\{\theta_1(N), \theta_2(N), \dots, \theta_p(N)\}$  serem mantidos e comparados.

## 3 APLICAÇÃO

### 3.1 ANÁLISE DESCRITIVA

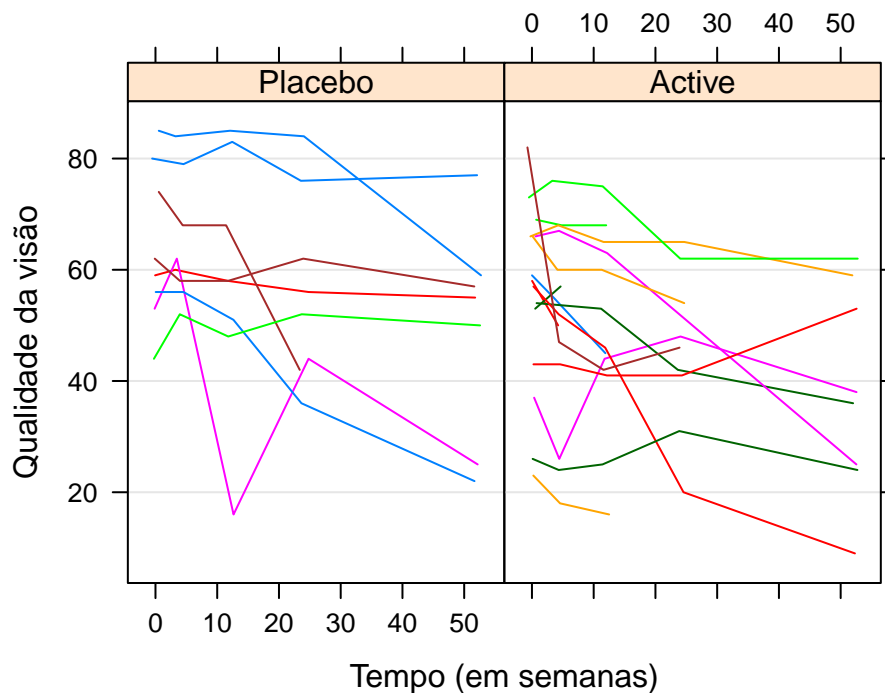


Figura 1: Gráfico de perfis de alguns indivíduos selecionados aleatoriamente

O experimento iniciou-se com 240 pacientes sendo que 119 receberam um placebo e 121 receberam a droga. Com o passar das semanas a quantidade de pessoas no estudo foram diminuindo, isso ocorreu por conta de efeitos colaterais sentidos em ambos os tratamentos. Por conta disso, ao fim do experimento obtivemos alguns dados faltantes conforme mostra a Tabela 1.

Médias e medianas amostrais das medidas de qualidade da visão						
Tempo	Placebo			Active		
	Núm. de Indiv.	Média	Mediana	Núm. de Indiv.	Média	Mediana
Início	119	55,34	56,0	121	54,58	57,0
4º semana	117	53,97	54,0	114	50,91	52,0
12º semana	117	52,87	53,0	110	48,67	49,5
24º semana	112	49,33	50,5	102	45,46	45,0
52º semana	105	44,44	44,0	90	39,10	37,0

Tabela 1: Número de indivíduos, médias e medianas amostrais das medidas de qualidade da visão de cada tempo observado.

Podemos observar um decrescimento ao longo do tempo das médias e medianas da medida de qualidade da visão conforme mostra a Figura 2. Nota-se também um aumento da variabilidade dos dados coletados nas últimas semanas, o aumento no número de dados faltantes pode ser uma possível causa para o crescimento dessa variabilidade. Há também um forte indício de simetria nas distribuições de ambos tratamentos por conta de que médias e medianas apresentaram valores próximos conforme indicado na Tabela 1.

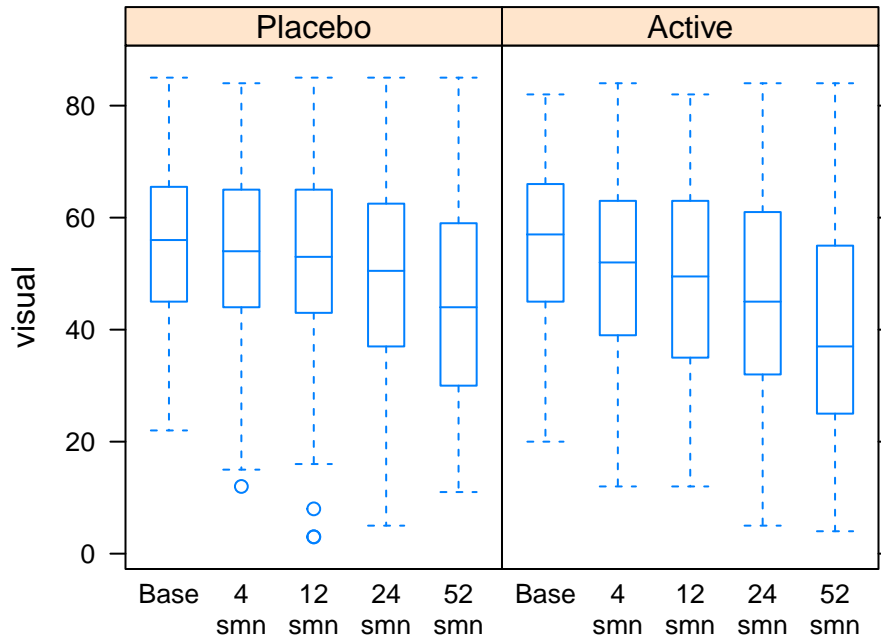


Figura 2: Boxplots das medidas de qualidade da visão de cada tempo observado.

A partir da matriz de variância e covariância  $\Sigma$  observa-se que há um aumento da variabilidade dos dados coletados nas últimas semanas, em concordância com as informações contidas nos boxplots. Considerando as correlações  $D$ , estas sugerem uma forte ou moderada correlação entre os tratamentos e uma diminuição entre as últimas medidas tomadas, possivelmente consequência dos dados faltantes.

• **Matriz de variancias e covariancias e matriz de correlações amostrais:**

$$\Sigma = \begin{pmatrix} 220.31 & 206.71 & 196.24 & 193.31 & 152.71 \\ 206.71 & 246.22 & 224.79 & 221.27 & 179.23 \\ 196.24 & 224.79 & 286.21 & 257.77 & 222.68 \\ 193.31 & 221.27 & 257.77 & 334.45 & 285.23 \\ 152.71 & 179.23 & 222.68 & 285.23 & 347.43 \end{pmatrix} \quad D = \begin{pmatrix} 1.00 & 0.89 & 0.78 & 0.71 & 0.55 \\ 0.89 & 1.00 & 0.85 & 0.77 & 0.61 \\ 0.78 & 0.85 & 1.00 & 0.83 & 0.71 \\ 0.71 & 0.77 & 0.83 & 1.00 & 0.84 \\ 0.55 & 0.61 & 0.71 & 0.84 & 1.00 \end{pmatrix}$$

### 3.2 MODELAGEM

Vamos considerar agora o seguinte modelo:

$$Y_{it} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2it} + \beta_3 x_{3i} + \beta_4 x_{2it} x_{3i} + b_{0i} + \xi_{it}, \quad (4)$$

- $Y_{it}$  é a qualidade da visão do paciente  $i$  ( $i = 1, \dots, 240$ ) no tempo  $t$  ( $t = 1, 2, 3, 4$ , correspondendo aos valores 4º, 12º, 24º e 52º semana, respectivamente).
- $x_{1i}$  é o valor inicial da qualidade da visão.
- $x_{2it}$  é o tempo  $t$  de medição no paciente  $i$ .
- $x_{3i}$  é o indicador do tratamento, 0 se placebo e 1 caso contrário.
- $x_{2it} x_{3i}$  é a interação entre as duas covariáveis.
- $\beta_0$  é o intercepto geral.
- $\beta_1$  é o incremento positivo ou negativo no valor esperado de  $Y_{it}$  quando variado em uma unidade o valor inicial da qualidade da visão.
- $\beta_2$  é o incremento positivo ou negativo na valor esperado de  $Y_{it}$ , quando acrescido o tempo em uma semana entre as que foram observadas.
- $\beta_3$  é o efeito geral positivo ou negativo no valor esperado de  $Y_{it}$  causado pelo tratamento.
- $\beta_4$  é o incremento positivo ou negativo sobre o valor esperado de  $Y_{it}$ , gerado pela variação do tempo em uma semana entre as que foram observadas sobre o paciente  $i$  que estava sob tratamento.
- $b_{0i}$  é o efeito aleatório específico para cada paciente. Tal que  $b_{0i} \sim \mathcal{N}(0, \tau) \forall i$ .
- $\xi_{it}$  é o erro aleatório. Tal que  $\xi_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2) \forall i$  e  $t$ .
- $b_{0i}$  representa uma variação específica do  $\beta_0$  para cada paciente.

Em notação matricial, o modelo para o sujeito  $i$  com o conjunto completo das quatro medidas da qualidade da visão é expresso por:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{pmatrix} = \begin{pmatrix} 1 & x_{1i} & 4 & x_{3i} & 4x_{3i} \\ 1 & x_{1i} & 12 & x_{3i} & 12x_{3i} \\ 1 & x_{1i} & 24 & x_{3i} & 24x_{3i} \\ 1 & x_{1i} & 52 & x_{3i} & 52x_{3i} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} b_{0i} + \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \xi_{i3} \\ \xi_{i4} \end{pmatrix}$$

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\xi}_i$$

Com  $\mathcal{D} \equiv \tau$  e  $\mathcal{R}_i \equiv \sigma^2 \mathbf{I}_4$ , no qual  $\mathbf{I}_4$  é a matrix identidade 4 x 4.



	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	9.43	0.07	2.67	4.13	7.61	9.43	11.28	14.66	1356.27	1
beta[2]	0.82	0.00	0.04	0.74	0.79	0.82	0.85	0.91	1328.35	1
beta[3]	-0.21	0.00	0.02	-0.26	-0.23	-0.21	-0.20	-0.17	5446.26	1
beta[4]	-2.46	0.03	1.49	-5.40	-3.45	-2.47	-1.42	0.38	1993.87	1
beta[5]	-0.05	0.00	0.03	-0.12	-0.07	-0.05	-0.03	0.01	5267.90	1
sigma_e	8.65	0.00	0.24	8.18	8.48	8.64	8.81	9.14	5342.71	1
sigma_p	9.02	0.01	0.54	7.99	8.65	9.00	9.38	10.09	4211.24	1

A partir dos resultados presentes na Tabela ?? obtidos através do ajuste do modelo temos que o valor esperado da qualidade da visão dos pacientes é de 9.43. Enquanto que a cada unidade acrescida ao valor inicial da qualidade da visão aumentamos em 0.82 unidades na resposta. Já o incremento no valor esperado de  $Y_{it}$  quando acrescido o tempo de uma semana para a próxima observada quando o paciente não está sob o tratamento  $interferon-\alpha$  é de -0.21 unidades. Ao passo que quando o paciente está sob o tratamento  $interferon-\alpha$  esse incremento apresenta o valor de -2.51. Com isso, temos indícios que até o final do experimento as pessoas enxergam quase uma letra a menos do que no início apenas pelo efeito do tempo. Percebemos também que os valores estimados de  $\beta_3$  e  $\beta_4$  são negativos, como eles representam o efeito geral causado pelo tratamento e o efeito gerado pela interação entre tratamento e tempo respectivamente, isso mostra que o tratamento em questão não é eficaz, tendo resultados piores do que o tratamento já existente.

**Estimativa da Matriz de Variâncias e Covariâncias Condicionais, Variâncias e Covariâncias Marginais e Correlações, respectivamente:**

$$\begin{pmatrix} 74.82 & 0 & 0 & 0 \\ 0 & 74.82 & 0 & 0 \\ 0 & 0 & 74.82 & 0 \\ 0 & 0 & 0 & 74.82 \end{pmatrix} \begin{pmatrix} 156.18 & 81.36 & 81.36 & 81.36 \\ 81.36 & 156.18 & 81.36 & 81.36 \\ 81.36 & 81.36 & 156.18 & 81.36 \\ 81.36 & 81.36 & 81.36 & 156.18 \end{pmatrix} \begin{pmatrix} 1 & 0.52 & 0.52 & 0.52 \\ 0.52 & 1 & 0.52 & 0.52 \\ 0.52 & 0.52 & 1 & 0.52 \\ 0.52 & 0.52 & 0.52 & 1 \end{pmatrix}$$

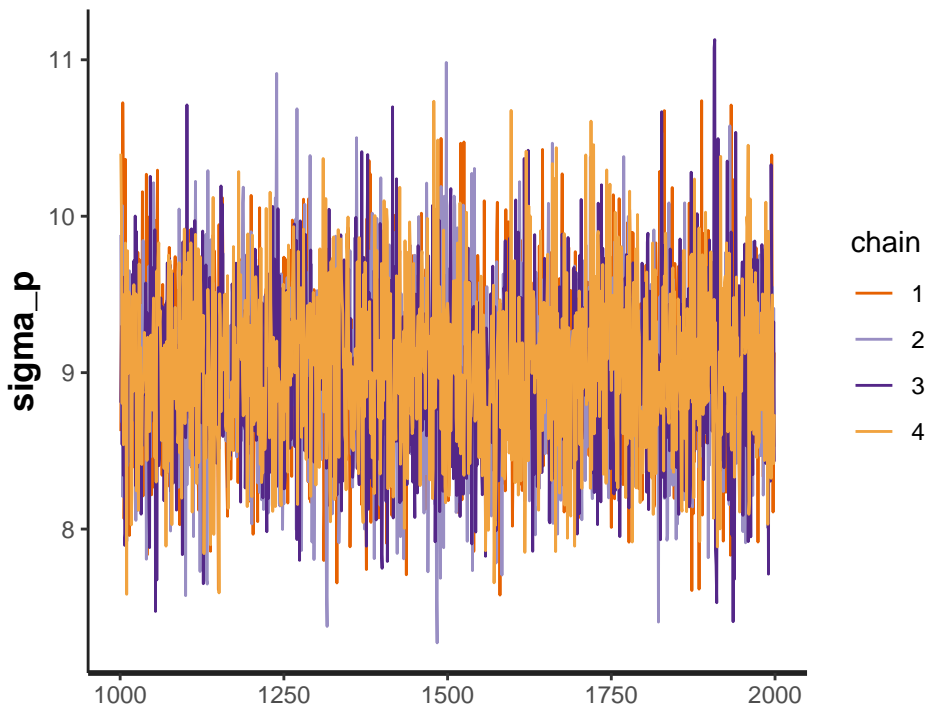


Figura 3: Gráfico mostrando o resultado sobre convergência da CM, após descartado o warm-up.

Como descrito na seção de metodologia para diminuir a influência dos valores iniciais, geralmente são descartados a primeira metade de cada sequência. Dessa forma nossas inferências são baseadas na suposição de que a distribuição dos valores simulados para valores grandes de  $i$  são próximos da distribuição alvo. Segue na Figura 3, a segunda metade dos valores de 4 CM simuladas para estimação do valor de  $\tau$  representadas em cores diferentes. O fato de estarem todas sobrepostas é um ótimo sinal de convergência das sequências.

## 4 CONSIDERAÇÕES FINAIS

### Referências

- Gamerman, Dani. 1996. *Simulação Estocástica via Cadeias de Markov*. Associação Brasileira de Estatística.
- Júlio M. Singer, Francisco Marcelo M. Rocha, Juvêncio S. Nobre. 2018. “Análise de Dados Longitudinais.” Versão parcial preliminar.
- Manly, Bryan F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall.