

Olympio Teixeira Neto

*Modelos Assimétricos com Efeitos Aleatórios  
para Dados Pré-Teste/Pós-Teste*

São Carlos

Abril de 2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Olympio Teixeira Neto

*Modelos Assimétricos com Efeitos Aleatórios  
para Dados Pré-Teste/Pós-Teste*

Orientador:  
Prof. Dr. Josemar Rodrigues

DEPARTAMENTO DE ESTATÍSTICA - UFSCAR

São Carlos  
Abril de 2007

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

T266ma

Teixeira Neto, Olympio.

Modelos assimétricos com efeitos aleatórios para dados pré-teste/pós-teste / Olympio Teixeira Neto. -- São Carlos : UFSCar, 2007.

117 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2007.

1. Modelos lineares (Estatística). 2. Técnicas de inferência. 3. Distribuição (Probabilidades). 4. Inferência Bayesiana. I. Título.

CDD: 511.8 (20<sup>a</sup>)

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Olympio Teixeira Neto e aprovada pela comissão julgadora como parte dos requisitos para obtenção do título de Mestre em Estatística.

**Banca Examinadora:**

Prof. Dr. Josemar Rodrigues  
Orientador

Profa. Dra. Maria Aparecida de Paiva  
Franco  
Des/UFSCar

Profa. Dra. Márcia D'Élia Branco  
IME/USP

*Aos meus pais.*

*"Sabemos incrivelmente pouco, e é espantoso que apesar disso saibamos tanto, e mais assombroso ainda que tão pouco conhecimento nos dê tanto poder."*

Bertrand Russell (1872-1970), filósofo e matemático inglês.

# *Agradecimentos*

- A Deus, pelo dom da vida.
- Aos meus pais, Olympio e Therezinha pelas oportunidades, motivação e apoio que sempre me ofereceram durante toda minha vida, jamais poupando esforços para que eu pudesse realizar tudo o que fiz até hoje. Dedico este trabalho a vocês.
- Às minhas irmãs, Maria Teresa, Cecília, Sílvia e Fabiana e aos meus cunhados, por sempre acreditarem e confiarem em mim, ouvindo minhas reclamações, meus pedidos, minhas “explicações”. Enfim, agradeço a oportunidade de convivermos nesta família.
- Aos meus sobrinhos mais velhos, Carol, Neto e Guta por sempre torcerem por mim, pela amizade e força que prestaram durante toda nossa vida juntos. Agradeço aos meus sobrinhos mais novos, Luiz Pedro, Vitória e Gabriel pela alegria e “perturbação” em toda esta trajetória.
- Ao meu orientador e amigo, Prof. Josemar, pela atenção, ensinamentos, compreensão, entusiasmo e, principalmente, paciência durante o período que convivemos juntos para estudar e elaborarmos este texto. Agradeço, acima de tudo, a perseverança em encontrar algo onde não se via nada.
- Aos meus professores da graduação, principalmente, Profas. Cida e Estela. Muito obrigado pela confiança e acima de tudo, amizade. Serei sempre muito grato a vocês.
- Aos meus amigos Júnior, Camila, Adriana, Carol, Fábio, Leandro, Cortez, e todos que aqui não cito, por compreenderem que, às vezes, estive ausente para poder realizar este trabalho.
- Aos professores da banca examinadora, Prof. Dra. Maria Aparecida e Prof. Dra. Márcia Branco, pelos comentários e contribuições para a elaboração desta dissertação.

- À FAPESP pelo importante apoio financeiro que possibilitou, em grande parte, desenvolvermos este trabalho.

# *Resumo*

Os modelos para dados pré e pós-testes são de grande utilidade nas mais diversas áreas da ciência. Eles são caracterizados como estudos longitudinais cujas observações são tomadas antes e após a aplicação de tratamentos ou covariáveis para estudo de seu efeito no tempo, assim, os dados são, usualmente, correlacionados, de modo que esta dependência pode ser modelada pela introdução de efeitos aleatórios, caracterizando um modelo misto.

Em especial, propusemos o estudo do modelo multiplicativo de Sef & Singer (SINGER; ANDRADE, 1997; SEF, 1999) para a análise de um conjunto de dados contendo índices de placa bacteriana em crianças submetidas ao efeito de algumas covariáveis. O modelo de Sef & Singer contém diversas restrições que podem contaminar o erro, tornando-o assimétrico. É neste cenário que introduzimos a distribuição normal assimétrica sob a representação de Sahu & Dey (SAHU et al., 2003), de modo a modelar o efeito aleatório como uma distribuição seminormal e o erro normal, tal que a assimetria é controlada por um parâmetro modulador que indica quão assimétrico será o erro final.

A estimação dos parâmetros foi feita com implementação do algoritmo E-M na inferência clássica. A inovação está na análise bayesiana, cujo modelo, sob uma reparametrização adequada, define uma distribuição à priori conjunta. A marginal desta distribuição à priori conjunta leva à distribuição T-Gama, para o parâmetro modulador de assimetria da representação de Sahu & Dey, generalizando a priori t-Student de Rodriguez (2005) **Palavras-chave:** *Algoritmo E-M, Análise de Covariância, Distribuição de*

*Kummer, Inferência Bayesiana, Distribuição Normal Assimétrica Multivariada, Modelos Mistos, Modelos Multiplicativos, Modelos para Dados Pré-Teste/ Pós-Teste, Regressão Linear Múltipla.*

# *Abstract*

The Pretest/Posttests models are special cases of repeated measures studies. The designs aim basically to compare pre and post treatment in experimental units during a defined period of time, thus, indicating a longitudinal study with two repetitions.

Furthermore, there is a strong dependence between measures, resulting in random effects included in the model. Special attention will be given to a multiplicative model (SINGER; ANDRADE, 1997; SEF, 1999), which presents as advantage, richly informative parameter characteristics. A data set establishing the index plaque of children of both sexes, pre and post supervised brushing employing two different kinds of dental brushes will be applied. The normal distribution might be not realistic due to several restrictions present in the model and related to the nature of the phenomenon under study.

The purpose of this study is to enable a flexible distribution to the multiplicative model through introduction of the skew-normal as defined by (SAHU et al., 2003), applying the EM algorithm and afterwards, Bayesian procedures in order to compare the estimates results obtained under these models.

From the Bayesian point of view, we have introduced, using a suitable reparametrization, the t-gamma prior distribution. One of its marginal is a new prior for a skewness parameter of Sahu & Dey representation, generalizing Rodríguez's t-Student prior (RODRÍGUEZ, 2005).

**Keywords:** *Bayesian Inference, Covariance Analysis, E-M Algorithm, Kummer's Distribution, Multiple Linear Regression, Multiplicative Models, Multivariate Skew Normal, Pretest/Posttest Models, Random Effects Models.*

## *Lista de Figuras*

|    |  |       |
|----|--|-------|
| 1  | Funções densidade de normais assimétricas univariadas. . . . .   | p. 18 |
| 2  | Escova dental do tipo Huger. . . . .   | p. 39 |
| 3  | Histograma da diferença entre as medidas pós-teste/pré-teste. . . . .  | p. 41 |
| 4  | Gráficos de dispersão entre as medidas pós e pré-teste em função das classes. . . . .  | p. 42 |
| 5  | Curvas individuais para os modelos estimados em função das classes . .   | p. 44 |
| 6  | Gráfico de dispersão de resíduos por valores preditos . . . . .  | p. 45 |
| 7  | Comparação entre a aproximação de Chaibub Neto & Branco com relação à distribuição normal . . . . .                              | p. 55 |
| 8  | Comparação entre a priori de Jeffreys e a aproximação pela t-student. .  | p. 56 |
| 9  | Comparação entre as prioris de Jeffreys e o resultado da indução sobre a distribuição uniforme (-1,1). . . . .                   | p. 57 |
| 10 | Distribuições das prioris do parâmetro de modulação da assimetria. . .   | p. 58 |
| 11 | Gráficos com algumas curvas definidas pela distribuição de Kummer, variando-se os parâmetros. . . . .                            | p. 65 |
| 12 | Distribuição conjunta . . . . .  | p. 67 |
| 13 | Formas da distribuição do parâmetro de modulação de assimetria em função dos hiperparâmetros da priori gama. . . . .             | p. 70 |
| 14 | Distribuições T-gama original e induzida. . . . .  | p. 73 |
| 15 | Escovas dentais do tipo monobloco. . . . .   | p. 74 |
| 16 | Diagramas de dispersão entre as medidas pré-teste/pós-teste para os dados da aplicação 6.3. . . . .                              | p. 74 |
| 17 | Histograma das diferenças entre as medidas pós e pré-testes em função da combinação de tratamentos para a aplicação 6.3. . . . . | p. 75 |

|    |  |       |
|----|--|-------|
| 18 | Curvas individuais para os valores pré e pós-testes para a aplicação 6.3.                                  | p. 76 |
| 19 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 1 - Parte 1. . . . . | p. 81 |
| 20 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 1 - Parte 2. . . . . | p. 82 |
| 21 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 1 - Parte 3. . . . . | p. 83 |
| 22 | Gráficos diagnósticos do modelo 1. . . . .   | p. 84 |
| 23 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 2 - Parte 1. . . . . | p. 86 |
| 24 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 2 - Parte 2. . . . . | p. 87 |
| 25 | Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros - Modelo 2 - Parte 3. . . . . | p. 88 |
| 26 | Gráficos diagnósticos do modelo 2. . . . .   | p. 89 |
| 27 | Gráficos diagnósticos do modelo 3. . . . .   | p. 90 |

# *Sumário*

## **I**

### **Clássica**

|          |  |       |
|----------|--|-------|
| <b>1</b> | <b>Introdução</b>  | p. 14 |
| <b>2</b> | <b>Distribuição Normal Assimétrica Univariada</b>  | p. 17 |
| 2.1      | Definições . . . . .   | p. 17 |
| 2.2      | Propriedades das Distribuições Normais Assimétricas . . . . .                            | p. 18 |
| <b>3</b> | <b>Modelos para Dados Pré-Teste/Pós-Teste</b>  | p. 22 |
| 3.1      | Modelo Multiplicativo de Sef & Singer . . . . .  | p. 23 |
|          | <b>Modelo</b> . . . . .  | p. 23 |
| 3.1.1    | Modelos para Dados Pré-Teste/Pós-Teste Normal Assimétrico .                              | p. 26 |
| 3.1.2    | Modelo Normal Assimétrico . . . . .  | p. 27 |
| 3.1.3    | Modelo para Dados Pré-Teste/Pós-Teste utilizando a Representação de Sahu & Dey . . . . . | p. 28 |
| <b>4</b> | <b>Algoritmo E-M</b>   | p. 30 |
| 4.1      | As etapas E e M . . . . .  | p. 31 |
| 4.2      | Erro Padrão no Algoritmo E-M . . . . .   | p. 33 |
| 4.3      | Algoritmo E-M para dados Pré e Pós-Testes . . . . .                                      | p. 33 |
| 4.3.1    | Aplicação . . . . .  | p. 38 |
| 4.3.1.1  | Resultados . . . . .   | p. 41 |
| 4.3.1.2  | Dificuldades Técnicas . . . . .  | p. 47 |

|  |           |
|--|-----------|
| <b>II Inferência Bayesiana</b>   | <b>49</b> |
| <b>5 Introdução</b>  | p. 50     |
| 5.1 Priori de Jeffreys . . . . .   | p. 51     |
| 5.2 Priori para Parâmetro de Assimetria . . . . .                          | p. 53     |
| 5.3 Definição das prioris para o modelo de Sef & Singer. . . . .           | p. 56     |
| <b>6 Formulação do modelo bayesiano</b>                                    | p. 60     |
| 6.1 A distribuição de Kummer . . . . .                                     | p. 63     |
| 6.1.1 Momentos da distribuição de Kummer . . . . .                         | p. 64     |
| 6.1.2 Ilustração . . . . .   | p. 64     |
| 6.2 Reparametrização do modelo . . . . .                                   | p. 64     |
| 6.2.1 Caracterizações da Priori T-Gama . . . . .                           | p. 71     |
| 6.3 Aplicação . . . . .  | p. 73     |
| 6.3.1 Resultados . . . . .   | p. 79     |
| 6.3.1.1 Modelo 1 - Prioris usuais ou canônicas . . . . .                   | p. 80     |
| 6.3.1.2 Modelo 2 - Priori T-Gama não-informativa . . . . .                 | p. 83     |
| 6.3.1.3 Modelo 3 - Análise de covariância simples com erro normal. . . . . | p. 86     |
| <b>7 Considerações finais e propostas futuras</b>                          | p. 91     |
| 7.1 Modelos para dados pré-teste/pós-teste . . . . .                       | p. 91     |
| 7.2 Representação de Sahu & Dey . . . . .                                  | p. 92     |
| 7.2.1 Estimação pelo algoritmo E-M . . . . .                               | p. 94     |
| 7.3 Inferência Bayesiana . . . . .   | p. 94     |
| <b>Apoio Financeiro</b>  | p. 98     |
| <b>Referências</b>   | p. 99     |
| <b>Apêndice A – Justificativas do algoritmo E-M</b>                        | p. 102    |

|  |  |        |
|--|--|--------|
| A.1  | Convergência . . . . .                       | p. 102 |
| A.1.1  | Propriedades do Algoritmo E-M . . . . .      | p. 103 |
| A.1.1.1  | Vantagens . . . . .                          | p. 103 |
| A.1.1.2  | Desvantagens . . . . .                       | p. 104 |
| <b>Apêndice B – Conjunto de Dados</b>                                    |  | p. 105 |
| B.1  | Dados da aplicação 4.3.1 . . . . .           | p. 105 |
| B.2  | Dados da Aplicação 6.3 . . . . .             | p. 106 |
| <b>Apêndice C – Modelos Multivariados para Dados Pré-Teste/Pós-Teste</b> |  | p. 107 |
| C.1  | Testes T - Pareados . . . . .                | p. 107 |
| C.2  | Planejamento com Medidas Repetidas . . . . . | p. 108 |
| C.2.1  | Procedimento de Análise . . . . .            | p. 109 |
| <b>Apêndice D – Programas em R</b>                                       |  | p. 112 |
| <b>Apêndice E – Programas em WinBUGS</b>                                 |  | p. 115 |

*Parte I*

*Inferência Clássica*

# 1 *Introdução*

A busca por modelos estatísticos gerais sob os quais pesam poucas imposições vem sendo uma constante na literatura científica desta área. Muitos autores apresentaram suas propostas, mas em alguns casos, as soluções são muito complicadas para a vida cotidiana, e estas são somente utilizadas no mundo acadêmico. Outras vezes, idéias ótimas não podem ser aproveitadas porque os problemas computacionais são de tal ordem que impossibilitam o seu uso com economicidade. Na tentativa de encontrar um caminho intermediário entre estas dificuldades é que propusemos o estudo aqui apresentado.

Em poucas palavras, pode-se definir um experimento com dados de pré-teste/pós-teste como sendo uma abordagem experimental em que uma medida é obtida antes da aplicação de um tratamento (medida pré-teste) e após aplicação do tratamento, a segunda medida é tomada (medida pós-teste). Claramente *trata-se de um estudo longitudinal* com duas unidades de tempo distintas, ou também como um experimento com medidas repetidas. São diversas as situações em que experimentos com dados pré-teste/pós-teste são realizados e modelos específicos para este tipo de abordagem já são muito consagrados, dentre eles, pode-se citar : Modelos Multivariados com medidas repetidas, Análise de Variância e Covariância (SINGER; ANDRADE, 1997; SEF, 1999).

Neste trabalho propomos o estudo do modelo multiplicativo (SINGER; ANDRADE, 1997; SEF, 1999) para dados pré-teste/pós-teste (que agora será chamado pelo nome dos autores, *Sef & Singer*), utilizado com um conjunto de dados muito específico da área de Odontologia. O modelo em questão é totalmente interpretável com relação ao problema e como característica principal, impõe-se como um modelo misto que é facilmente linearizado por meio de uma transformação logarítmica.

Os autores utilizaram distribuições normais tanto para os efeitos aleatórios quanto para os erros, mas dadas as características do fenômeno que estudaram, nas quais, muitas restrições são impostas ao modelo, as distribuições subjacentes, por eles utilizadas, podem não ser realistas.

Desta forma, para a flexibilização das distribuições envolvidas nesta modelagem, lançamos mão da classe de distribuições normais assimétricas. Esta classe é capaz de captar as assimetrias que podem estar envolvidas em um conjunto de dados, quando existem fatores externos que tornam as amostras seletivas, ou seja, quando há perturbações naturais em populações teoricamente normais. Essas perturbações podem levar às assimetrias aqui comentadas. É interessante ressaltar que as distribuições normais assimétricas carregam consigo propriedades inerentes à distribuição normal, como por exemplo, a forma quadrática de uma distribuição normal assimétrica também segue uma distribuição  $\chi^2$ , sob condições específicas (AZZALINI; VALLE, 1996).

Há várias maneiras de se representar uma distribuição normal assimétrica, mas tendo em mente que o modelo é multiplicativo e, quando linearizado, torna-se aditivo, então, é possível e natural ao problema a inclusão de efeitos aleatórios para se captar a correlação entre medidas tomadas da mesma unidade experimental ou indivíduo. Assim, tem-se, ao final, um modelo misto. É nesta abordagem que recorreremos à representação de uma distribuição normal assimétrica segundo a definição de Sahu et al. (2003). Durante a apresentação e definição do problema poderemos observar que antes, um modelo longitudinal (que é o caso geral quando se trata de medidas pré-teste/pós-teste), cujos parâmetros são difíceis de estimar computacionalmente, pode ser interpretado como uma análise de covariância. Ainda, como se verificará, o próprio modelo misto poderá ser testado com relação ao modelo de regressão linear múltipla, bastando para tanto, uma configuração específica de um parâmetro da distribuição proposta por Sahu et al. (2003), mesmo que na prática, o modelo resultante não seja o mais adequado.

Este trabalho será dividido da seguinte maneira:

- Parte I - Inferência Clássica
  - Introdução das Normais Assimétricas e suas propriedades;
  - Modelo para dados Pré-Teste/Pós-Teste;
  - Definição do modelo em estudo e definição para sua utilização utilizando a abordagem de Sahu et al. (2003);
  - Algoritmo E-M;
  - Aplicação utilizando-se o algoritmo E-M.
  
- Parte II - Inferência Bayesiana
  - Introdução à Priori de Jeffreys;

- Priori de Jeffreys para o coeficiente de assimetria utilizando o resultado de Rodríguez (2005);
- Formulação do modelo para dados Pré-Teste/Pós-Teste;
- Introdução à distribuição de Kummer e a Priori de Kummer;
- Aplicações.

## 2 *Distribuição Normal Assimétrica Univariada*

### 2.1 Definições

Vários autores em diversos trabalhos, tais como: Azzalini & Valle (1996), Sahu et al. (2003), Lachos (2004), Liu & Dey (2005), definiram muitas distribuições normais assimétricas. Elas surgem de acordo com as necessidades de maior ou menor generalização. Nesta dissertação exploraremos o caso uniparamétrico com os seus três parâmetros: locação, escala e forma.

A maior parte destas demonstrações está disponível em Rodríguez (2005), e caso contrário, a fonte será citada expressamente.

**Definição 2.1.1** *Uma variável aleatória  $Y \in \mathbb{R}$  segue uma distribuição normal assimétrica com vetor de locação  $\boldsymbol{\mu} \in \mathbb{R}$  e de escala  $\sigma^2 > 0$  se sua função densidade de probabilidade é dada por*

$$f_Y(y) = \frac{2}{\sigma} \phi_1 \left( \frac{y - \mu}{\sigma} \right) \Phi_1 \left[ \left( \frac{y - \mu}{\sigma} \right) \lambda \right], \text{ com } y \text{ e } \lambda \in \mathbb{R}, \quad (2.1)$$

sendo representada por  $Y \sim SN(\mu, \sigma^2, \lambda)$ . Quando em (2.1),  $\mu = 0$  e  $\sigma^2 = 1$ , obtém-se  $Y \sim SN(\lambda)$ , a distribuição normal assimétrica padrão.

É fácil verificar que se  $Z \sim SN(\lambda)$ , a transformação  $Y = \mu + \sigma^2 Z$  leva a  $Y \sim SN(\mu, \sigma^2, \lambda)$ .

A figura 1 representa as curvas de densidade para algumas distribuições normais assimétricas padrão com diferentes parâmetros de assimetria. O modelo normal é um caso particular dos modelos assimétricos, pois se obtém aquela distribuição tomando  $\lambda = 0$ .

Pode-se mostrar que a densidade apresentadas em (2.1) é densidade de probabilidade.

Existem muitas outras formas de se derivar a distribuição normal assimétrica, mas a

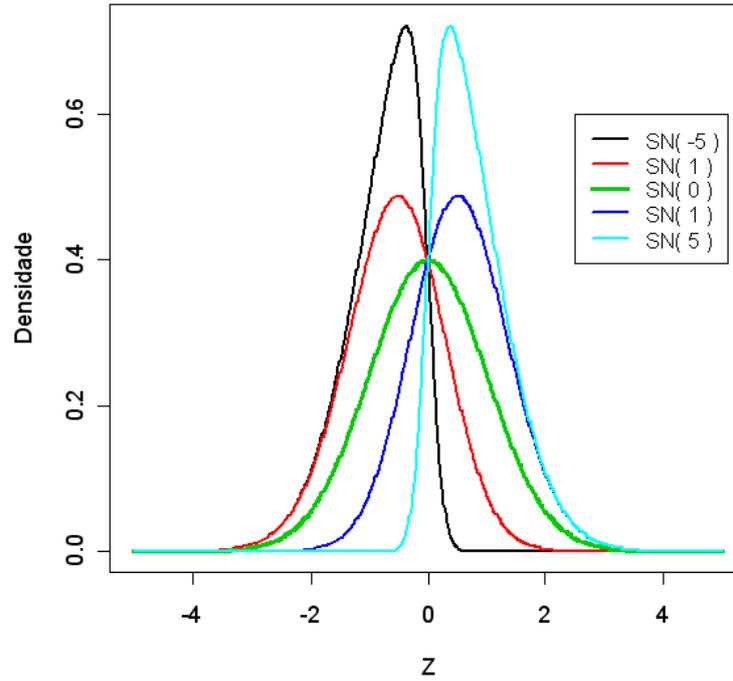


Figura 1: Funções densidade de normais assimétricas univariadas com diferentes parâmetros de assimetria  $\lambda$ .

representação de Sahu & Dey será a mais útil para o desenvolvimento desta dissertação, como veremos posteriormente.

## 2.2 Propriedades das Distribuições Normais Assimétricas

Muitas propriedades das distribuições normais assimétricas podem ser derivadas de sua função geradora de momentos (f.g.m.). No entanto, para se demonstrar a forma da f.g.m., utilizaremos um importante resultado apresentado por Lachos (2004).

**Lema 2.2.1** *Seja  $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Então, para algum vetor fixo  $\mathbf{a}$  de dimensão  $k$  e alguma matriz fixa  $\mathbf{B}$  de dimensão  $k \times n$ , tem-se*

$$E[\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \Phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}') \quad (2.2)$$

$$E[\phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'). \quad (2.3)$$

**Proof.** A demonstração está disponível em Lachos (2004, p. 22). ■

Com a importante afirmação do lema 2.2.1, a f.g.m. da distribuição normal assimétrica pode ser facilmente encontrada. Observe a seguir.

**Proposição 2.2.1** *Seja  $Z \sim SN(\lambda)$ . Então, a função geradora de momentos é dada por*

$$M_Z(s) = 2 \exp\left(\frac{s^2}{2}\right) \Phi(\delta s), \text{ com } s \in \mathbb{R}, \delta = \frac{\lambda}{\sqrt{(1 + \lambda^2)}}. \quad (2.4)$$

**Proof.** Utilizando a definição de uma função geradora de momentos, temos:

$$\begin{aligned} M_Z(s) &= E(e^{sz}) = \int_{-\infty}^{\infty} 2e^{sz} \phi(z) \Phi(\lambda z) dz \\ &= 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2sz + s^2)} e^{\frac{s^2}{2}} \Phi(\lambda z) dz \\ &= 2e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - s^2)} e^{\frac{s^2}{2}} \Phi(\lambda z) dz \end{aligned}$$

Fazendo a transformação de variável  $x = z - s$ ,

$$M_Z(s) = 2e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \phi(x) \Phi(\lambda x + \lambda s) dx.$$

Utilizando-se o lema 2.2.1, obtemos:

$$M_Z(s) = 2 \exp\left(\frac{s^2}{2}\right) \Phi(\delta s).$$

■

A extensão para a família de normais assimétricas com parâmetros de locação-escala é facilmente obtida, utilizando-se com a proposição a seguir, o que formaliza a expressão em (2.1).

Podemos encontrar a f.g.m. de uma classe mais geral, resultados estes derivados da proposição 2.2.1. Seguem abaixo os corolários.

**Corolário 2.2.1** *Seja  $Y \sim SN_n(\mu, \sigma^2, \lambda)$ . A f.g.m. de  $Y$  é dada por:*

$$M_Y(s) = 2e^{s\mu + \frac{s^2 \sigma^2}{2}} \Phi(\delta s), \text{ com } \delta = \frac{\lambda}{\sqrt{(1 + \lambda^2)}}.$$

**Proof.** A demonstração do corolário 2.2.1 está disponível em Rodríguez (2005, p. 16).

■

Do corolário em 2.2.1 seguem estes interessantes resultados anunciados no seguinte corolário:

**Corolário 2.2.2** *Se  $Z \sim SN(\lambda)$ , então*

1.  $-Z \sim SN(-\lambda)$ ;
2.  $aZ \sim SN(a\lambda)$ , com  $a \in \mathbb{R}$ ;

A representação formal das distribuições normais assimétricas não é muito interessante do ponto de vista computacional. Desse modo, Henze (1986) definiu uma maneira estocástica de se reescrever esta distribuição de probabilidade.

**Proposição 2.2.2** *Fazendo  $Z \sim SN(\lambda)$ , temos,*

$$Z = \delta |X_0| + \sqrt{(1 - \delta^2)} X_1, \text{ sendo } \delta = \frac{\lambda}{\sqrt{(1 + \lambda^2)}}, \quad (2.5)$$

com  $|X_0| \sim HN(0,1)$ ,  $X_1 \sim N(\mathbf{0},1)$  e  $|X_0|$  independente de  $\mathbf{X}_1$ . *HN representa a distribuição “half normal” ou seminormal.*

**Proof.** A demonstração dessa proposição pode ser encontrada originalmente em Henze (1986), mas também pode ser vista em Rodríguez (2005, p. 13). ■

O resultado da proposição 2.2.2 é muito importante na formulação de modelos assimétricos e fornece uma maneira prática de se gerar amostras desta família de distribuições, parametrizar os modelos e desenvolver algoritmos para a estimação, como veremos posteriormente, tanto para a inferência clássica como para a bayesiana.

Agora, apresentamos a caracterização da média e variância de uma variável aleatória normal assimétrica.

**Proposição 2.2.3** *Seja  $Z \sim SN_n(\lambda)$ . Então,*

$$E(Z) = \sqrt{\frac{2}{\pi}} \delta \quad (2.6)$$

$$Var(Z) = 1 - \frac{2}{\pi} \delta^2. \quad (2.7)$$

**Proof.** *Considerando a proposição 2.2.2 e calculando-se a esperança de uma variável  $|X_0| \sim HN(0,1)$ , temos*

$$E(|X_0|) = \int_0^\infty (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{u^2}{2}\right] du = \sqrt{\frac{2}{\pi}}.$$

Como  $E(X_1) = 0$ , então, aplicando-se em (2.5),

$$\begin{aligned} E[Z] &= E\left[\delta|X_0| + \sqrt{(1-\delta^2)}X_1\right] = \delta E(|X_0|) + \sqrt{(1-\delta^2)}E(X_1) \\ &= \delta\sqrt{\frac{2}{\pi}}. \end{aligned}$$

Agora, para a variância de  $Z$ ,  $\text{var}(|X_0|) = 1 - \frac{2}{\pi}$  e  $\text{var}(X_1) = 1$ . Ainda,  $|X_0|$  é independente de  $X_1$ .

$$\begin{aligned} \text{Var}(Z) &= \text{Var}\left[\delta|X_0| + \sqrt{(1-\delta^2)}X_1\right] = \delta^2\left(1 - \frac{2}{\pi}\right) + (1-\delta^2) \\ &= 1 - \frac{2}{\pi}\delta^2. \end{aligned}$$

■

Para se obter as esperanças do modelo com parâmetros de locação-escala, basta utilizar a relação linear  $Y = \mu + \sigma^2 Z$ , com  $Z \sim SN(\lambda)$  e aplicá-la juntamente com as propriedades básicas da esperança e da variância de uma variável aleatória.

### *3 Modelos para Dados Pré-Teste/Pós-Teste*

Em muitas situações práticas existe a necessidade de se analisar dados que são obtidos antes e após a aplicação de algum tipo de tratamento. É neste contexto que surgem os **Modelos para Dados Pré-Teste/Pós-Teste**. Esta classe de modelos, casos particulares dos planejamentos com medidas repetidas e especialmente, dos modelos para dados longitudinais quando o número de repetições da medida ao longo do tempo é de tamanho 2, formam uma rica e elegante ferramenta quando se deseja saber o ganho que se obtém de algum tratamento no intervalo de tempo de estudo, sob um planejamento de medidas repetidas. Esta ferramenta apresenta numerosas aplicações em vários campos da ciência, inclusive com muitas formas para se abordar o mesmo problema.

Muitos autores estudaram esta forma de planejar e modelar experimentos, propondo modificações significativas no aspecto e forma, dentre eles, citamos: Brogan & Kutner (1980), Laird & Ware (1982), STANEK III (1998). Mais recentemente, temos: Singer & Andrade (1997) e Sef (1999).

Genericamente, a forma inicial de abordagem para os dados pré-teste/pós-teste foi realizada com a utilização de metodologias **ANOVA** e **ANCOVA**, utilizando-se da medida pré-teste como covariável. Há também a alternativa do modelo linear multivariado, utilizando-se para isso de **modelos com medidas repetidas em blocos** (SEF, 1999) e **testes t multivariados** (RENCHEK, 2002). A natureza dos dados leva rapidamente a estes tipos de abordagens, no entanto, quando alguma sofisticação é dada ao modelo, como uma maior flexibilização quanto à forma da matriz de variâncias-covariâncias ou a adição de efeitos aleatórios, eles podem ser intratáveis sob estas perspectivas. A inferência bayesiana também parece ter um bom ferramental para a solução destes tipos de problemas, desde que se obtenham distribuições a priori adequadas para os parâmetros, como será visto na segunda parte desta dissertação.

Nas próximas seções apresentaremos as principais técnicas para se abordar este pro-

blema e formularemos uma nova proposta que leva em conta as características dos conjuntos de dados a serem utilizados. No **apêndice C** algumas noções sobre os modelos multivariados são mostradas e mais informações podem ser obtidas em Rencher (2002).

### 3.1 Modelo Multiplicativo de Sef & Singer

Os modelos que podem ser vistos no apêndice C são gerais e podem, perfeitamente, ser utilizados para análise de dados pré-teste/pós-teste. Muitos outros também existem e que não foram tratados aqui, como os planejamentos em parcelas divididas, por exemplo.

Nesta dissertação trataremos de um modelo específico definido por Singer & Andrade (1997) e Sef (1999), agora denominado por *Sef & Singer* para trabalhar com dois conjuntos de dados provenientes de um estudo sobre o índice de placa bacteriana dentária antes e após a utilização dois tratamentos distintos. A apresentação do problema se dá na seção 4.3.1. Outro exemplo, exposto em 6.3, será dado com dois tratamentos: Escovas e uso de dentifrícios.

Tendo em vista o conjunto de dados referente ao índice de placa bacteriana antes e depois da aplicação de algum tratamento, Singer & Andrade (1997) e Sef (1999) propuseram, em um artigo e dissertação de mestrado, respectivamente, a utilização de um modelo multiplicativo em que os parâmetros estimados têm interpretação prática sobre o problema.

É notório que um índice de placa bacteriana é uma medida que, em especial, deve ter sempre um valor positivo e este tipo de restrição torna-se, às vezes, inadequada a utilização da distribuição normal pelas assimetrias que possam estar envolvidas. Sob este aspecto, proporemos, a seguir, a utilização da distribuição normal assimétrica de modo a flexibilizar este modelo. Outras considerações serão feitas posteriormente, com uma inovação na forma de se abordar este tipo de modelagem, onde uma regressão múltipla simples substitui ao final a necessidade de modelos multivariados para este fenômeno.

**Modelo** Para satisfazer as condições do experimento pré-teste/pós-teste envolvendo os índices de placa bacteriana dental, são exigidas algumas restrições (SEF, 1999) que são próprias da natureza dos dados.

1. Os índices pré-teste nulos implicam em índices pós-teste também nulos, ou seja, o modelo parte da origem no plano cartesiano;

2. Os índices são sempre positivos;
3. Pode haver heteroscedasticidade;
4. O relacionamento entre índices pré-teste/pós-teste não é necessariamente linear;
5. Existe correlação entre as medidas tomadas do mesmo indivíduo;
6. A distribuição das observações não é necessariamente normal.

A restrição *número 6* é uma das inovações desta dissertação, uma vez que os autores estudaram este tipo de modelo sob a condição de normalidade para o erro e efeitos aleatórios (SINGER; ANDRADE, 1997; SEF, 1999). Ainda, os autores justificam a utilização do modelo multiplicativo por causa da restrição 1, ou seja, a necessidade de que o modelo parta da origem, implicando que medidas pré-testes nulas impliquem em medidas pós-testes nulas.

Assim, define-se

$$y_{ijk} = \beta_{ij} x_{ijk}^{\gamma_{ij}} \varepsilon_{ijk} \quad (3.1)$$

tal que:

$y_{ijk}$  é a medida tomada pós-teste para o  $k$ -ésimo indivíduo, ( $k = 1, \dots, n$ ) sob os tratamentos  $i$  e  $j$ ,  $i = 1,2$  e  $j = 1,2$ . No caso desta dissertação, tratamento pode ser interpretados ora como sexo e ora como tipo de escova, ou uso ou não de dentifrício, por isso adotou-se o termo genérico “tratamento” para estas covariáveis ;

$x_{ijk}$  é a medida tomada pré-teste para  $k$ -ésimo indivíduo, ( $k = 1, \dots, n$ ) sob os tratamentos  $i$  e  $j$ ,  $i = 1,2$  e  $j = 1,2$ , considerada fixa neste caso;

$\beta_{ij}$  representa o coeficiente que pode ser interpretado como placa bacteriana residual, para os tratamentos  $i$  e  $j$ ,  $i = 1,2$  e  $j = 1,2$ ;

$\gamma_{ij}$  representa o coeficiente de uniformidade que revela a velocidade com que a placa bacteriana é removida com a utilização dos tratamentos  $i$  e  $j$ ,  $i = 1,2$  e  $j = 1,2$ ;

$\varepsilon_{ijk}$  erro aleatório para o  $k$ -ésimo indivíduo, ( $k = 1, \dots, n$ ) e tratamentos  $i$  e  $j$ ,  $i = 1,2$  e  $j = 1,2$ .

Este modelo tem uma característica interessante como observaram os autores:  $E(y_{ijk}) = \beta_{ij} x_{ijk}^{\gamma_{ij}} E(\varepsilon_{ijk})$  e  $\text{Var}(y_{ijk}) = (\beta_{ij} x_{ijk}^{\gamma_{ij}})^2 \text{Var}(\varepsilon_{ijk})$ . Se  $\gamma_{ij} \neq 1$ , então o modelo é não-linear. Ainda, a razão entre esperança da medida pós-teste  $y_{ijk}$ , e o valor considerado fixo da

medida pré-teste,  $x_{ijk}$ , é denotada por  $\frac{E(y_{ijk})}{x_{ijk}} = \frac{\beta_{ij} x_{ijk}^{\gamma_{ij}} E(\varepsilon_{ijk})}{x_{ijk}} = \beta_{ij} x_{ijk}^{\gamma_{ij}-1} E(\varepsilon_{ijk})$ , o que implica nas seguintes situações para o índice pós-teste:

$$\left\{ \begin{array}{l} \gamma_{ij} > 1: \text{variação no índice de placa é crescente com relação ao índice pré-teste;} \\ \gamma_{ij} = 1: \text{variação no índice de placa é constante com relação ao índice pré-teste;} \\ \gamma_{ij} < 1: \text{variação no índice de placa é decrescente com relação ao índice pré-teste.} \end{array} \right.$$

Claramente o modelo multiplicativo é não-linear, mas pode ser linearizado por uma transformação logarítmica, o que resulta em

$$\ln(y_{ijk}) = \ln(\beta_{ij}) + \gamma_{ij} \ln(x_{ijk}) + \ln(\varepsilon_{ijk}). \quad (3.2)$$

O resultado da transformação em (3.2) é linear em  $\ln(x_{ijk})$  e ainda é possível renomear os novos parâmetros do modelo transformado de forma a torná-lo mais adequado à situação em estudo.

Na teoria inicial sobre o modelo multiplicativo (SINGER; ANDRADE, 1997; SEIF, 1999), adota-se que  $\ln(\varepsilon_{ijk})$  segue uma distribuição gaussiana, o que leva, naturalmente, a que  $\varepsilon_{ijk}$  tenha distribuição lognormal. Ademais, é de conhecimento já consolidado em Estatística que medidas tomadas do mesmo indivíduo em tempos diferentes podem trazer consigo correlação. Uma maneira de abordar estes problemas é por meio da inclusão de efeitos aleatórios, o que leva ao desmembramento do erro. Isto será feito no modelo transformado.

O modelo (3.2) pode ser reparametrizado como:

$$z_{ijk} = \alpha_{ij} + \gamma_{ij} w_{ijk} + \zeta_{ijk}, \quad (3.3)$$

em

$$z_{ijk} = \ln(y_{ijk}), \alpha_{ij} = \ln(\beta_{ij}), w_{ijk} = \ln(x_{ijk}) \text{ e } \zeta_{ijk} = \ln(\varepsilon_{ijk}).$$

Se considerarmos um modelo de efeitos aleatórios de forma a incluir as correlações intra-indivíduos, teremos a decomposição de  $\zeta_{ijk}$  como apresentado a seguir:

$$\zeta_{ijk} = \eta_k + \nu_{ijk} \quad (3.4)$$

sendo que

$$\begin{aligned}\eta_k &\sim H \text{ (H é a distribuição dos efeitos aleatórios) e,} \\ \nu_{ijk} &\sim G \text{ (G é a distribuição do erro residual).}\end{aligned}$$

Os autores consideraram tanto H como G distribuições normais (SINGER; ANDRADE, 1997; SEF, 1999), mas podemos imaginar que esta suposição pode não ser adequada, haja vista as restrições inerentes ao fenômeno em estudo, dentre elas, a possível heteroscedasticidade (que não afeta a normalidade diretamente, mas sim, as suposições para modelos de ANCOVA e ANOVA), a positividade das medidas pré e pós-testes obtidas, que por si só, já geram processos de seletividade na amostra, o que leva as observações a se comportarem assimetricamente, que, às vezes, não propiciam o uso da distribuição normal nessas situações.

Esta dissertação se propõe a flexibilizar o modelo justamente nas distribuições subjacentes ao efeito aleatório e ao erro, considerando-as na classe das distribuições normais assimétricas. Assim, o modelo ficaria:

$$z_{ijk} = \alpha_{ij} + \gamma_{ij}w_{ijk} + \underbrace{\eta_k + \nu_{ijk}}_{\zeta_{ijk}}. \quad (3.5)$$

### 3.1.1 Modelos para Dados Pré-Teste/Pós-Teste Normal Assimétrico

O modelo em estudo nessa dissertação pode ser abordado de várias maneiras. Poder-se-ia realizar um estudo longitudinal com efeitos aleatórios (LACHOS, 2004) considerando os dois tempos (pré e pós-teste), utilização de um modelo ANOVA (STANEK III, 1998; SINGER; ANDRADE, 1997) de forma a se ajustar a diferença obtida entre os valores pós-teste e pré-teste (medida esta chamada de *ganho*), modelos de regressão multivariada entre muitos outros.

O que propomos aqui é a utilização de um modelo de regressão linear múltipla que engloba os efeitos aleatórios e os erros em uma única distribuição assimétrica, de tal modo que, dependendo de um parâmetro específico, a assimetria será modulada. Para tanto, é necessária a suposição de que o efeito aleatório é distribuído por uma seminormal. A seguir mostraremos a estrutura deste modelo, bem como suas características.

### 3.1.2 Modelo Normal Assimétrico

Para a obtenção dos modelos de regressão normais assimétricos, é necessário especificar distribuições normais assimétricas tanto para o efeito aleatório quanto para o erro, quando houver evidências de assimetrias nestes dois componentes do modelo. A vantagem desta modelagem é que a distribuição dos efeitos aleatórios e erros são assimétricas e como já se viu, a distribuição normal simétrica é um caso especial das distribuições normais assimétricas, de modo que dependendo do valor do parâmetro de assimetria, estes modelos serão encaixados.

Assim, para o modelo como em (3.5), adotamos

$$\begin{aligned}\eta_k &\sim SN(0, \sigma_\eta^2, \lambda_\eta), \\ \nu_{ijk} &\sim SN(0, \sigma_\nu^2, \lambda_\nu), \\ \eta_k &\perp \nu_{ijk},\end{aligned}\tag{3.6}$$

em que o símbolo  $\perp$  indica independência entre estas variáveis.

Isto implica em situações tais que:

$$\left\{ \begin{array}{l} \lambda_\eta = 0: \text{ o modelo tem efeitos aleatórios com distribuição normal;} \\ \lambda_\nu = 0: \text{ o modelo tem erros com distribuição normal;} \\ \lambda_\eta = 0 \text{ e } \lambda_\nu = 0: \text{ o modelo tem erros e efeitos aleatórios com distribuição normal.} \end{array} \right.$$

Obviamente, a presença de parâmetros de assimetria não-nulos em ambas as variáveis aleatórias implicam no modelo normal assimétrico para efeitos aleatórios e erros.

A principal dificuldade com esta especificação do modelo é na estimação dos parâmetros de assimetria. Rodríguez (2005) verificou que a estimação por máxima verossimilhança é problemática, principalmente nos casos de multimodalidade da distribuição assimétrica subjacente, pois os coeficientes de assimetria têm alta probabilidade de serem estimados como infinitos. Aqui, a recomendação é a utilização do algoritmo E-M, método que é também uma estimação por máxima verossimilhança, mas que minimiza os problemas quanto à estimação infinita destes parâmetros e é relativamente fácil de ser implementado, mas requer uso de computação intensiva. O caso de modelos de efeitos aleatórios e erros distribuídos por normais assimétricas não será objeto de estudo desta dissertação. Mais detalhes, consultar o completo trabalho de Lachos (2004).

### 3.1.3 Modelo para Dados Pré-Teste/Pós-Teste utilizando a Representação de Sahu & Dey

A abordagem que estamos propondo para modelos como em (3.5) é uma nova maneira de se explicar o erro e o efeito aleatório. Sahu et al. (2003) e Liu & Dey (2005) propuseram que uma distribuição normal assimétrica pode ser gerada de maneira semelhante à proposição 2.2.2, que a partir de agora será denominada por *representação de Sahu & Dey*.

**Proposição 3.1.1** *Uma distribuição normal assimétrica (2.1) pode ser definida, segundo a representação de Sahu & Dey, pela seguinte combinação linear: sejam  $\eta \sim N(0, \sigma_\eta^2)$ ,  $\nu \sim N(0, \sigma_\nu^2)$  com  $\rho \in \mathbb{R}$ . Assim,*

$$\zeta = \rho|\eta| + \nu \sim SN \left( 0, \sigma_\nu^2 + \rho^2 \sigma_\eta^2, \frac{\rho \sigma_\eta}{\sigma_\nu} \right). \quad (3.7)$$

**Proof.** *Inicialmente já é conhecida a forma da função geradora de momentos de uma distribuição normal assimétrica que é dada pelo corolário 2.2.1. Assim, usando as propriedades da f.g.m. e algumas manipulações algébricas,*

$$\begin{aligned} M_\zeta(t) &= M_{\rho|\eta|+\nu}(t) = E \left[ e^{t(\rho|\eta|+\nu)} \right] \\ &= E \left( e^{t\rho|\eta|} \right) E \left( e^{t\nu} \right) \\ &= \exp \left( \frac{1}{2} t^2 \sigma_\nu^2 \right) E \left( e^{t\rho|\eta|} \right) \\ &= \exp \left( \frac{1}{2} t^2 \sigma_\nu^2 \right) \int_0^\infty e^{t\rho\eta} \phi(\eta|0, \sigma_\eta^2) d\eta \\ &= 2 \exp \left[ \frac{1}{2} t^2 (\sigma_\nu^2 + \rho^2 \sigma_\eta^2) \right] \Phi(t\rho\sigma_\eta). \end{aligned} \quad (3.8)$$

A expressão (3.8) é igual a função geradora de momentos de uma distribuição normal assimétrica unidimensional e se rearranjarmos os termos, teremos

$$\begin{aligned} M_\zeta(t) &= 2 \exp \left[ \frac{1}{2} t^2 (\sigma_\nu^2 + \rho^2 \sigma_\eta^2) \right] \Phi(t\rho\sigma_\eta) \\ &= 2 \exp \left[ \frac{1}{2} t^2 (\sigma_\nu^2 + \rho^2 \sigma_\eta^2) \right] \Phi \left( \frac{\rho \sigma_\eta}{\sqrt{\sigma_\nu^2 + \rho^2 \sigma_\eta^2}} \sqrt{\sigma_\nu^2 + \rho^2 \sigma_\eta^2} t \right); \end{aligned}$$

logo,  $\zeta \sim SN \left( 0, \sigma_\nu^2 + \rho^2 \sigma_\eta^2, \delta = \frac{\rho \sigma_\eta}{\sqrt{\sigma_\nu^2 + \rho^2 \sigma_\eta^2}} \right)$ . Como se vê, o coeficiente de assimetria depende de  $\rho$ . Para que  $\rho$  fique em função do parâmetro de assimetria  $\lambda$ , então basta

fazer a transformação  $\lambda = \frac{\delta}{\sqrt{1 - \delta^2}} = \frac{\rho\sigma_\eta}{\sigma_\nu}$ , o que conclui a prova. ■

A representação (3.7) é muito semelhante à decomposição do erro descrito em (3.5), de modo que estão conexas e **a inclusão de um parâmetro  $\rho$  à frente do efeito aleatório fará com que este novo integrante do modelo module a assimetria final**. Assim, poderíamos adotar esta estrutura para o modelo de forma que, dependendo do valor coeficiente  $\rho$  estimado, duas situações poderiam ser encontradas. Se  $\hat{\rho} \approx 0$ , então o modelo seria interpretado como uma análise de covariância univariada com distribuição normal.

**Definição 3.1.1** *O modelo para dados pré-teste/pós-teste proposto por Sef & Singer (3.1) e linearizado por uma transformação logarítmica (3.5), com a inclusão de efeitos aleatórios para o intercepto pela decomposição do erro, segundo a distribuição normal assimétrica sob a representação Sahu & Dey, é definido como*

$$\begin{aligned} z_{ijk} &= \alpha_{ij} + \gamma_{ij}w_{ijk} + \overbrace{\sigma_\eta\rho|\eta_k| + \sigma_\nu\nu_{ijk}}^{\zeta_{ijk}}, \\ \eta_k &\sim N(0,1), \\ \nu_{ijk} &\sim N(0,1), \eta_k \perp \nu_{ijk} \\ \zeta_{ijk} &\sim SN\left(0, \sigma_\nu^2 + \rho^2\sigma_\eta^2, \frac{\rho\sigma_\eta}{\sigma_\nu}\right), \end{aligned} \quad (3.9)$$

onde  $i = 1,2; j = 1,2; k = 1, \dots, n$ , onde  $\alpha_{ij}$  e  $\gamma_{ij}$  são parâmetros do modelo e  $\rho \in \mathbb{R}$ .

Utilizando-se inferência bayesiana, é possível, graças aos métodos MCMC (Monte Carlo com cadeias de Markov) (GAMERMAN; LOPES, 2006), estimar os parâmetros desse modelo, como também construir intervalos de predição. Na inferência clássica, os intervalos de confiança também podem ser obtidos por meio de métodos computacionais intensivos (reamostragem), como o bootstrap (GENTLE et al., 2004).

Na parte II desta dissertação estudaremos este modelo sob a óptica da inferência bayesiana, inclusive encontrando distribuições a priori para os parâmetros, mas por hora, será apresentada a teoria básica para a implementação do algoritmo E-M, ferramenta que será muito útil para a resolução dos problemas pela inferência clássica.

## 4 Algoritmo E-M

O algoritmo EM, “Expectation Maximization” (Esperança - Maximização) é um método de computação iterativa amplamente utilizado para o cálculo de estimativas de máxima verossimilhança em dados incompletos. Assim, como um dos alicerces da estatística é a obtenção destas estimativas, a sua utilidade é incontestável.

No contexto geral, pode-se considerar como *dados incompletos* aqueles elementos que não estejam disponíveis em uma amostra, (dados faltantes, ou “missings”), ou que sejam variáveis aleatórias não observáveis diretamente, de forma que é possível *prever* seus valores de modo a completar a amostra. Também, o algoritmo E-M é muito interessante na tentativa de tornar verossimilhanças complexas em formas mais simples (GENTLE et al., 2004; TANNER, 2005). Essa situação é comum, mas não uma regra.

Para iniciar o estudo do algoritmo, considere o vetor com os dados observados  $\mathbf{y}$ , que têm função densidade de probabilidade  $g(\mathbf{y}; \Psi)$ , tal que  $\Psi$  seja um vetor de dimensão  $n$  contendo os parâmetros desconhecidos da distribuição de  $\mathbf{Y}$ . O objetivo geral é maximizar a verossimilhança

$$L(\Psi) = g(\mathbf{y}; \Psi), \quad (4.1)$$

como uma função de  $\Psi$  sobre o espaço de parâmetros  $\Omega$ .

Se a função em (4.1) se encontra sob as condições de regularidade (ver Cramér, 1946), então as raízes de sua derivada primeira quando igualadas a  $\mathbf{0}$  (ou a mesma situação para a função de log-verossimilhança  $\ln L(\Psi)$ ), serão os estimadores de máxima verossimilhança,  $\hat{\Psi}$ , que têm, em condições especiais, as propriedades mais adequadas para um estimador.

Nas aplicações estatísticas quase sempre se impõe a necessidade de estimação de parâmetros, mesmo quando a função de verossimilhança é extremamente complicada, o que dificulta, fatalmente, a maximização dessas funções, dificuldade esta que pode ser analítica, computacional ou ambas. Exemplos destas dificuldades são: as estimações de misturas de distribuições, modelos com dados censurados e principalmente, de interesse nessa dissertação, a estimação de parâmetros de distribuições normais assimétricas. Mui-

tas dessas dificuldades podem ser resolvidas com a introdução de variáveis aleatórias não observáveis diretamente ou fictícias (latentes) associadas ao problema, de forma a *ampliar* o conjunto de dados originais para que se possa trabalhar com o método da máxima verossimilhança de maneira adequada. A este *conjunto de dados ampliados* dá-se o nome de *dados completos*. A função de verossimilhança recebe o nome de *verossimilhança com dados incompletos e completos*, quando ela é estruturada utilizando-se o conjunto de originais e aumentados, respectivamente.

Feita estas considerações iniciais, o algoritmo E-M fica caracterizado como um método genérico de estimação de parâmetros por máxima verossimilhança para um conjunto de dados incompletos pela formulação de um problema associado com dados completos, sempre buscando uma maneira mais simples de se obter os estimadores de máxima verossimilhança quando esta, originalmente, é complicada, ou ainda, quando o parâmetro de interesse não é diretamente observável somente com a amostra disponível (GENTLE et al., 2004; TANNER, 2005). Não se pode negligenciar, também, a utilidade do método, quando existem muitos dados faltantes ou “missings” no conjunto de dados originais, e por algum motivo, o método requer que estes conjuntos estejam completos. Nestas situações o algoritmo, em uma etapa específica, entra para “completar” este conjunto de dados e assim, permitir aplicação do método.

Rapidamente, pode-se dizer que o algoritmo tem um caráter iterativo, que pode claramente ser dividido em duas etapas. Uma etapa de cálculo da *esperança* (etapa-E) e a etapa da *maximização* (etapa-M) de uma função de verossimilhança para cada uma de suas iterações.

## 4.1 As etapas E e M

Para formalizar o entendimento do algoritmo, suponhamos que  $\mathbf{x}$  seja um vetor contendo os dados completos ou ampliados e que  $\mathbf{z}$  denote o vetor de dados ausentes (“missings”), também chamado de vetor de *variáveis latentes*. Assim, teremos que  $\mathbf{X}' = (\mathbf{Y}, \mathbf{Z})'$ . Vamos definir por  $g_c(\mathbf{x}, \Psi)$  a f.d.p. de um vetor aleatório  $\mathbf{X}$  correspondente ao vetor de dados completos  $\mathbf{x}$ . Se  $\mathbf{X}$  for completo, podemos escrever a função de log-verossimilhança como

$$\ln L_c(\Psi) = \ln g_c(\mathbf{x}, \Psi). \quad (4.2)$$

O algoritmo E-M tenta solucionar o problema da verossimilhança dos dados incompletos (4.1) indiretamente, atualizando aquela função por meio de (4.2), que contém todos os

dados completos. Como (4.1) tem valores não observáveis, quando comparada com (4.2), esta, é, então, substituída por sua esperança condicional dado o vetor de variáveis  $\mathbf{y}$  e os valores atuais das estimativas de  $\Psi$ . Logo, podemos definir claramente o algoritmo E-M.

**Algoritmo 4.1.1** *Expectation-Maximization (E-M)*

1. Defina os valores iniciais de  $\Psi$ ;
2. Etapa - E: Calcule  $Q(\Psi, \Psi^{(k)})$ , onde

$$Q(\Psi, \Psi^{(k)}) = E_{\Psi^{(k)}} [\ln L_c(\Psi) | \mathbf{y}]; \quad (4.3)$$

3. Etapa - M: Faça  $\Psi^{(k+1)}$  ser qualquer valor de  $\Psi \in \Omega$  que maximize  $Q(\Psi, \Psi^{(k)})$ :

$$Q(\Psi, \Psi^{(k+1)}) \geq Q(\Psi, \Psi^{(k)}) \quad \forall \Psi \in \Omega; \quad (4.4)$$

4. Repita os passos 2 e 3 até que a convergência  $\varepsilon > 0$  desejada seja alcançada.

O operador  $E_{\Psi^{(k)}}(*)$  representa a esperança de uma variável aleatória utilizando o vetor de parâmetros  $\Psi$  na k-ésima iteração. Dessa forma, temos que o resultado imediatamente anterior de uma iteração do algoritmo é utilizado para se “calibrar”, ou “completar” a verossimilhança dos dados completos, pela substituição das variáveis latentes, pelas esperanças condicionais do vetor de parâmetro, dado os dados observados do problema em estudo. Esta “atualização” da verossimilhança é que permite o cálculo das estimativas de máxima verossimilhança de maneira mais simples, na maioria dos casos. Esta simplificação decorre do fato de que podemos escrever (4.2) como

$$\ln L_c(\Psi) = \ln g_c[(\mathbf{y}, \mathbf{z}), \Psi]$$

e o interesse é remover, matematicamente, a variável  $\mathbf{z}$  da função de verossimilhança e o operador esperança realiza função.

$$\begin{aligned} Q(\Psi, \Psi^{(k)}) &= E_{\Psi^{(k)}} [\ln L_c(\Psi) | \mathbf{y}] \\ &= E_{\Psi^{(k)}} [\ln g_c[(\mathbf{y}, \mathbf{z}), \Psi] | \mathbf{y}] \\ &= \int [\ln g_c[(\mathbf{y}, \mathbf{z}), \Psi] | \mathbf{y}] f(\mathbf{z} | \Psi^{(k)}, \mathbf{Y}) d\mathbf{z}. \end{aligned} \quad (4.5)$$

Em (4.5) não existe a presença das variáveis latentes  $\mathbf{z}$ , que foram removidas pela integração, de forma que houve uma *calibração* da log-verossimilhança (4.2). Então, encontrar a distribuição de probabilidade  $f(\mathbf{z}|\Psi^{(k)}, \mathbf{Y})$  é fundamental na construção do algoritmo e essa busca é uma parte muito importante no método, pois, a escolha adequada leva à log-verossimilhanças calibradas mais simples (FIGUEIREDO, 2004).

Uma justificativa para (4.4) será mostrada no apêndice A.

## 4.2 Erro Padrão no Algoritmo E-M

A finalidade de uma estimação pelo método E-M é encontrar os valores de  $\hat{\Psi}$  que maximizam a função (4.2). Uma aproximação dos erros padrão de  $\hat{\Psi}$  pode ser feita utilizando-se a *matriz hessiana* de  $Q$  resultante da iteração final do algoritmo e posteriormente, teoria normal assintótica para a definição de intervalos de confiança. Este procedimento pode ser muito complicado e várias abordagens para este problema são discutidas na literatura. Ver Tanner (2005) .

Como será visto na seção 4.3, o modelo multiplicativo de *Sef & Singer* normal assimétrico, como propomos nesta dissertação, não tem uma verossimilhança muito simples para se poder calcular a sua derivada segunda objetivando a obtenção do erro padrão, com o intuito de se construir intervalos de confiança utilizando teoria assintótica. Assim, optamos por simulações bootstrap para a obtenção do erro padrão das estimativas dos parâmetros de interesse, e a partir disso, os intervalos de confiança não-paramétricos (EFRON; TIBHIRANI, 1994) foram construídos.

## 4.3 Algoritmo E-M para o Modelo Multiplicativo de Sef & Singer

Considerando o modelo multiplicativo em (3.3), é possível obter, de maneira relativamente simples, o algoritmo E-M para a estimação dos parâmetros de regressão e da distribuição normal assimétrica subjacente ao modelo misto. Colocando o modelo em

forma matricial, temos:

$$\mathbf{Z}_k = \begin{bmatrix} z_{11k} \\ z_{12k} \\ z_{21k} \\ z_{22k} \end{bmatrix} = \begin{bmatrix} \alpha_{11} + \gamma_{11}w_{11k} \\ \alpha_{12} + \gamma_{12}w_{12k} \\ \alpha_{21} + \gamma_{21}w_{21k} \\ \alpha_{22} + \gamma_{22}w_{22k} \end{bmatrix}. \quad (4.6)$$

A matriz em (4.6) pode ainda ser empilhada, definindo:

$$\begin{bmatrix} \overbrace{1 & w_{11k} & 0 & 0 & 0 & 0 & 0 & 0}^{\mathbf{W}_k^*} \\ 0 & 0 & 1 & w_{12k} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & w_{21k} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & w_{22k} \end{bmatrix} \begin{pmatrix} \overbrace{\alpha_{11} \\ \gamma_{11} \\ \alpha_{12} \\ \gamma_{12} \\ \alpha_{21} \\ \gamma_{21} \\ \alpha_{22} \\ \gamma_{22}}^{\boldsymbol{\gamma}^*} \end{pmatrix} \quad (4.7)$$

tal que  $\mathbf{W}_k^*$  é a matriz de planejamento de dimensão  $4 \times 8$  para o  $k$ -ésimo indivíduo e  $\boldsymbol{\gamma}^*$  é o vetor com os parâmetros  $\alpha_{ij}$  e  $\gamma_{ij}, i = 1, 2; j = 1, 2$ , de dimensão 8. Ainda, temos que

$$\boldsymbol{\zeta}_k = (\rho\eta_k + \nu_{ijk}) \mathbf{I}_4, \quad (4.8)$$

onde  $\mathbf{I}_4$  é uma matriz identidade de dimensão 4.

Para finalidades práticas de notação, adotaremos também, a seguinte matriz notação para representar as linhas de  $\mathbf{W}_k$ :

$$\begin{aligned} \mathbf{R}_1 &= \mathbf{W}_{1k}^{*'} = \begin{pmatrix} 1 & w_{11k} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ &\quad \vdots \\ \mathbf{R}_4 &= \mathbf{W}_{4k}^{*'} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & w_{22k} \end{pmatrix}, \end{aligned} \quad (4.9)$$

onde  $\mathbf{R}_l = \mathbf{W}_{lk}^{*'$ , com dimensão  $1 \times 8$  representa a  $l$ -ésima ( $l = 1, \dots, 4$ ) da matriz  $\mathbf{W}_k$  e,

$$\left\{ \begin{array}{l} z_{1k}^* = z_{11k} \\ z_{2k}^* = z_{12k} \\ z_{3k}^* = z_{21k} \\ z_{4k}^* = z_{22k} \end{array} \right. \text{ e } \left\{ \begin{array}{l} \nu_{1k}^* = \nu_{11k} \\ \nu_{2k}^* = \nu_{12k} \\ \nu_{3k}^* = \nu_{21k} \\ \nu_{4k}^* = \nu_{22k} \end{array} \right. \quad (4.10)$$

tal que  $z_{lk}^*$  representa cada uma das observações pós-teste em função dos tratamentos a que ele é submetido, e  $\nu_{lk}^*$ , o resíduo do modelo, com  $l = 1, \dots, 4$ .

Assim, tomando-se (3.3) e utilizando a forma de Sahu & Dey (3.7) sob a representação matricial (4.6-4.10), temos, lembrando a definição 3.1.1:

$$\begin{aligned} \mathbf{Z}_k &= \mathbf{W}_k^* \boldsymbol{\gamma}_k^* + (\rho \sigma_\eta |\eta_k| + \sigma_\nu \nu_{lk}^*) \mathbf{I}_4 \\ \eta_k &\sim N(0,1) \\ \boldsymbol{\nu}_k^* &\sim N(0,1) \mathbf{I}_4, \quad k = 1, \dots, n \end{aligned} \quad (4.11)$$

Logo,  $(\rho \sigma_\eta |\eta_k| + \sigma_\nu \nu_{lk}^*) \sim SN\left(0, \sigma_\nu^2 + \rho^2 \sigma_\eta^2, \frac{\rho \sigma_\eta}{\sigma_\nu}\right)$ ,  $k = 1, \dots, n$  e  $l = 1, \dots, 4$ . Observe que nesta representação, assumindo variâncias unitárias para  $\eta_k$  e  $\nu_{lk}^*$ , a representação em (4.11) é idêntica à da normal assimétrica padrão de Azzalini (2006),  $SN(0, 1 + \rho^2, \rho)$ , com o  $\rho$  cumprindo o papel de parâmetro de assimetria. Com a expansão das formas matriciais, teremos:

$$z_{lk}^* = \mathbf{R}_l \boldsymbol{\gamma}^* + \rho \sigma_\eta \eta_k + \sigma_\nu \nu_{lk}^*, \quad l = 1, \dots, 4; k = 1, \dots, n. \quad (4.12)$$

O modelo em (4.12) pode ser escrito hierarquicamente, de forma a facilitar a implementação do algoritmo e, posteriormente, a sua especificação utilizando a inferência bayesiana. Assim,

$$\begin{aligned} z_{lk}^* | \eta_k &\sim N(\mathbf{R}_l \boldsymbol{\gamma}^* + \rho \sigma_\eta \eta_k, \sigma_\nu^2) \\ \eta_k &\sim N(0,1) I_{(0,\infty)}(\eta_k) \equiv HN(0,1), \quad l = 1, \dots, 4; k = 1, \dots, n, \end{aligned} \quad (4.13)$$

o que leva à função densidade de probabilidade de  $f(z_{lk}^*; \eta_k) = f(z_{lk}^* | \eta_k) \times f(\eta_k)$ . Mas,

também, é necessário conhecer-se a distribuição de  $f(\eta_k|z_{lk}^*)$  para se obter a esperança de  $E(\eta_k|Z_{lk}^*)$ , cujo objetivo é estimar, no passo E, o valor de cada uma das variáveis aleatórias  $\eta_k$ . Uma maneira de se encontrar esta distribuição é por meio da definição de probabilidade condicional. Utilizando-se a definição de probabilidade condicional, temos que

$$f(\eta_k|z_{lk}^*) = \frac{f(z_{lk}^*; \eta_k)}{\int_{\mathbb{R}} f(z_{lk}^*; \eta_k) d\eta_{ijk}}. \quad (4.14)$$

A expressão (4.14) pode ser encontrada analiticamente, mas como a integral do denominador é constante em relação a  $z_{lk}^*$ , então

$$f(\eta_k|z_{lk}^*) \propto f(z_{lk}^*; \eta_k) = f(z_{lk}^*|\eta_k) \times f(\eta_k) \quad (4.15)$$

Desse modo, a distribuição condicional da variável latente, dadas as variáveis observadas, resulta na seguinte função densidade de probabilidade:

$$\begin{aligned} f(\eta_k|z_{lk}^*) &\propto \frac{1}{\sigma_\nu} \exp\left[-\frac{1}{2\sigma_\nu^2} (z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^* + \rho\sigma_\eta\eta_k)^2\right] \exp\left(-\frac{1}{2}\eta_k\right) I_{(0,\infty)}(\eta_k) \\ &= \frac{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}{\sigma_\nu} \exp\left\{-\frac{1}{2}\left[\frac{\sigma_\nu^2 + \rho^2\sigma_\eta^2}{\sigma_\nu^2} \left(\eta_k - \frac{\rho\sigma_\eta(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sigma_\nu^2 + \rho^2\sigma_\eta^2}\right)^2\right]\right\} I_{(0,\infty)}(\eta_k) \\ &\sim N\left(\frac{\rho\sigma_\eta(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sigma_\nu^2 + \rho^2\sigma_\eta^2}, \frac{\sigma_\nu^2}{\sigma_\nu^2 + \rho^2\sigma_\eta^2}\right) I_{(0,\infty)}(\eta_k). \end{aligned} \quad (4.16)$$

Para o cálculo de  $E(\eta_k|Z_{lk}^*)$ , será necessário o seguinte resultado, como apresentado por Lachos (2004):

**Teorema 4.3.1** *Seja  $X \sim N(\mu, \sigma^2)$ . Então, para uma constante real  $\alpha$ , segue que*

$$\begin{aligned} E(X|X > \alpha) &= \mu + \frac{\phi_1\left(\frac{\alpha-\mu}{\sigma}\right)}{1 - \Phi_1\left(\frac{\alpha-\mu}{\sigma}\right)}\sigma \\ E(X^2|X > \alpha) &= \mu^2 + \sigma^2 + \frac{\phi_1\left(\frac{\alpha-\mu}{\sigma}\right)}{1 - \Phi_1\left(\frac{\alpha-\mu}{\sigma}\right)}(\mu + \alpha)\sigma. \end{aligned}$$

Em particular, quando  $\alpha = 0$ , então

$$E(X|X > 0) = \mu + \frac{\phi_1\left(\frac{\mu}{\sigma}\right)}{\Phi_1\left(\frac{\mu}{\sigma}\right)}\sigma \text{ e}$$

$$E(X^2|X > 0) = \mu^2 + \sigma^2 + \frac{\phi_1\left(\frac{\mu}{\sigma}\right)}{\Phi_1\left(\frac{\mu}{\sigma}\right)}\mu\sigma.$$

**Proof.** A prova deste resultado pode ser encontrada em Johnson (1970). ■

Utilizando-se o resultado do teorema 4.3.1 e reescrevendo a distribuição (4.16) como

$$f[(\eta_k|z_{lk}^*)|\eta_k > 0],$$

podemos facilmente calcular as esperanças, como segue:

$$E(\eta_k|Z_{lk}^*) = \frac{\rho\sigma_\eta(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sigma_\nu^2 + \rho^2\sigma_\eta^2} + \frac{\phi_1\left[\frac{\lambda(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}\right]}{\Phi_1\left[\frac{\lambda(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}\right]} \frac{\sigma_\nu}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}} \text{ e} \quad (4.17)$$

$$E(\eta_k|Z_{lk}^*) = \left[\frac{\rho\sigma_\eta(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sigma_\nu^2 + \rho^2\sigma_\eta^2}\right]^2 + \frac{\sigma_\nu^2}{\sigma_\nu^2 + \rho^2\sigma_\eta^2} +$$

$$+ \frac{\phi_1\left[\frac{\lambda(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}\right]}{\Phi_1\left[\frac{\lambda(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}\right]} \frac{\rho\sigma_\eta\sigma_\nu(z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^*)}{(\sigma_\nu^2 + \rho^2\sigma_\eta^2)^{\frac{3}{2}}}, \quad (4.18)$$

para  $k = 1, \dots, n; l = 1, \dots, 4$  e  $\lambda = \frac{\rho\sigma_\eta}{\sigma_\nu}$ .

A construção do algoritmo E-M está praticamente resolvida, bastando calcular-se a função de verossimilhança. Como temos uma amostra aleatória, então a distribuição de  $Z_{lk}^*|\eta_k$  é dada hierarquicamente por (4.13). Assim, a função log-verossimilhança, que será utilizada na etapa M do algoritmo, é dada como segue:

$$l(Z_{lk}^*|\eta_k, \boldsymbol{\theta}') = -\frac{n}{2} \ln(\sigma_\nu) - \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^4 \left[ \left( \frac{z_{lk}^* - \mathbf{R}_l\boldsymbol{\gamma}^* + \rho\sigma_\eta\eta_k}{\sigma_\nu} \right)^2 + \eta_k \right]. \quad (4.19)$$

Finalmente, podemos escrever o algoritmo.

#### Algoritmo 4.3.1 Algoritmo E-M

1 - Iniciar com  $\boldsymbol{\theta}'^{(0)} = (\boldsymbol{\gamma}^{*(0)}, \rho^{(0)}, \sigma_\eta^{2(0)}, \sigma_\nu^{2(0)})'$  arbitrários;

Para um  $i$  qualquer,

2 - (**Passo E**) Calcular  $\tilde{\eta}_k^{(i)} = E\left(\eta_k | Z_{lk}^*, \boldsymbol{\theta}'^{(i)}\right)$  e  $\tilde{\eta}_k^{2(i)} = E\left(\eta_k | Z_{lk}^*, \boldsymbol{\theta}'^{(i)}\right)$  a partir das expressões (4.17) e (4.18), respectivamente;

3 - (**Passo M**) Maximizar a log-verossimilhança (4.19) tal que

$$\boldsymbol{\theta}'^{(i)} = \arg \max_{\boldsymbol{\theta}'} \left\{ l\left(Z_{lk}^* | \tilde{\eta}_k^{(i-1)}, \boldsymbol{\theta}'^{(i-1)}\right) \right\};$$

4 - Repetir os **passos 2 e 3** até que a tolerância  $\varepsilon > 0$  seja alcançada.

A maximização, utilizando-se o algoritmo (4.3.1), pode ser realizada em programas estatístico. Nesta dissertação utilizamos o R (R Development Core Team, 2006).

### 4.3.1 Aplicação

Este estudo foi realizado na Faculdade de Odontologia da Universidade de São Paulo, cujos dados foram retirados do trabalho de Singer & Andrade (1997). Os dados apresentados na tabela 11, do apêndice B.1, correspondem às medidas do índice de placa bacteriana obtidas de 14 crianças em período pré-escolar do sexo feminino e, 12 do sexo masculino, antes e depois do uso de escovas convencionais e outra, alternativa do tipo “Hugger” (figura 2). O objetivo da pesquisa era comparar as duas escovas com respeito à eficácia em remover a placa, uma vez que ela é um coadjuvante importantíssimo na prevenção da cariogênese. Alguns cuidados foram tomados para se evitar efeitos residuais das escovas, como um intervalo de 14 dias entre os usos das mesmas. Esta experiência poderia utilizar-se de um planejamento “crossover”, mas todas as medidas do índice de placa com escovas do tipo “Hugger” foram obtidas antes que os das escovas convencionais, prejudicando, assim, planejamentos desse tipo.

Singer & Andrade (1997) e Sef (1999) utilizaram, então, o modelo multiplicativo para explicar os efeitos do tipo de escova e do sexo do indivíduo na redução do índice de placa bacteriana pré e pós-teste. A seguir definiremos o modelo em estudo e suas características. É importante ressaltar que, para fins de comparação com os modelos estudados Singer &



Figura 2: Escova dental do tipo Huger.

Andrade (1997) e Sef (1999), e por simplicidade, esta nova abordagem será definida como um modelo com interceptos aleatórios.

Considerando o modelo geral e suas propriedades (3.1), e acrescentando os índices para melhor defini-lo, temos:

$$y_{ijk} = \beta_{ij} x_{ijk}^{\gamma_{ij}} \varepsilon_{ijk}, \quad (4.20)$$

tal que  $y_{ijk}(x_{ijk})$  representa a medida pós (pré) - teste do  $k$ -ésimo indivíduo do  $i$ -ésimo sexo ( $i = 1,2$ ), utilizando o  $j$ -ésimo tipo de escova ( $j = 1,2$ ). O coeficiente de uniformidade para o sexo  $i$  e escova  $j$  é representado por  $\gamma_{ij}$ , e  $\beta_{ij}$  representa a placa bacteriana residual.

Aplicando-se a transformação logarítmica, tem-se

$$z_{ijk} = \alpha_{ij} + \gamma_{ij} w_{ijk} + \zeta_{ijk}, \quad (4.21)$$

com

$$z_{ijk} = \ln(y_{ijk}), \alpha_{ij} = \ln(\beta_{ij}), w_{ijk} = \ln(x_{ijk}) \text{ e } \zeta_{ijk} = \ln(\varepsilon_{ijk}).$$

Sob o modelo linear (4.21) é que as estimações serão feitas. O erro  $\zeta_{ijk}$  deve ser desmembrado para que a representação da distribuição normal assimétrica de Sahu &

Dey possa ser escrita, representando, então, o modelo de efeitos aleatórios. Assim,

$$\begin{aligned}
 z_{ijk} &= \alpha_{ij} + \gamma_{ij} w_{ijk} \overbrace{\sigma_{\eta} \rho |\eta_k| + \sigma_{\nu} \nu_{ijk}}^{\zeta_{ijk}}, \quad k = 1, \dots, n \\
 \eta_k &\sim N(0, 1) \\
 \nu_{ijk} &\sim N(0, 1) \\
 \eta_k &\perp \nu_{ijk} \\
 \zeta_{ijk} &\sim SN\left(0, \sigma_{\nu}^2 + \sigma_{\eta}^2 \rho^2, \frac{\rho \sigma_{\eta}}{\sigma_{\nu}}\right), \\
 k &= 1, \dots, n; \quad j = 1, 2; \quad i = 1, 2.
 \end{aligned} \tag{4.22}$$

Ao modelo multiplicativo de Sef & Singer, foi acrescentado nesta fase, o parâmetro  $\rho$ , que indicará, por meio do conjunto de dados, qual a distribuição de  $\zeta_{ijk}$ , que pode ser normal assimétrica ou normal simétrica, quando  $\rho = 0$ . Além disso, é imediato o teste para o modelo de interceptos aleatórios (modelo misto) ou não, dependendo, somente, deste parâmetro  $\rho$ . Se  $\hat{\rho} = 0$ , então, tem-se um modelo *ANCOVA* simples normal, fato este que pode tornar a análise insuficiente, uma vez que os dados pré e pós-testes devem incluir um efeito de correlação, como já discutido anteriormente. Se  $\rho \neq 0$ , estaremos com um modelo misto normal assimétrico, que pode facilmente ser tratado como um modelo de regressão linear normal assimétrico com a parametrização apresentada em (4.22).

Diversos trabalhos e autores, entre eles, Rodríguez (2005) e Freitas (2005), já estudaram as propriedades dos modelos de regressão assimétricos e portanto, a implementação é relativamente simples, apesar das dificuldades computacionais mencionadas naqueles trabalhos.

Os resultados obtidos por Singer & Andrade (1997) e Sef (1999) podem não ser realistas, no sentido de que as várias restrições não são totalmente contempladas quando se impõe que o efeito aleatório e os resíduos sejam distribuídos normalmente. A figura 3 mostra que a diferença entre as medidas pós-teste/pré-teste (ganho) são sempre negativas, ou seja, há uma forte tendência à assimetria à esquerda.

A figura 4 mostra que existe uma relação que parece ser linear entre os índices pós e

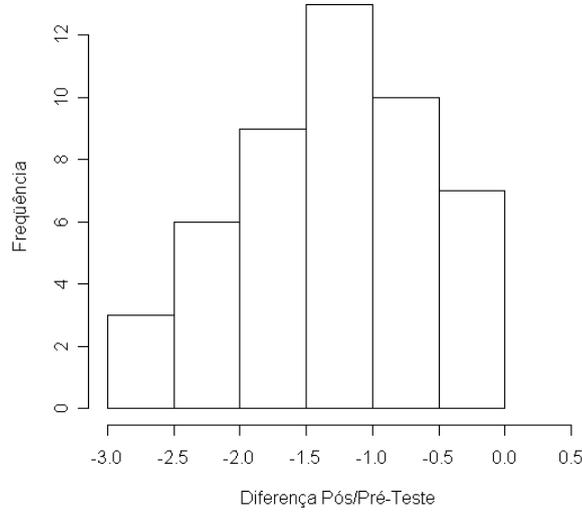


Figura 3: Histograma da diferença entre as medidas pós-teste/pré-teste.

pré-teste. Dessa maneira, utilizaremos o algoritmo E-M apresentado em 4.3 para estimar o modelo definido em (4.22) e comparar os resultados com aqueles utilizando o modelo normal.

Os interceptos, além da correção que sofrerão por conta do efeito aleatório, devem ser corrigidos ainda para que se obtenha a mesma interpretação que se tem em uma regressão linear, onde o erro tem média zero; esta correção será dada por  $\alpha_{ij} + \sqrt{\frac{2}{\pi}}\delta\sigma$ , tal que  $\sigma = \sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}$  e  $\delta = \frac{\rho\sigma_\eta}{\sqrt{\sigma_\nu^2 + \rho^2\sigma_\eta^2}}$ .

#### 4.3.1.1 Resultados

Para estimação dos parâmetros do modelo,

$$\boldsymbol{\theta}' = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \rho, \sigma_\eta^2, \sigma_\nu^2)',$$

utilizou-se o aplicativo R e o código encontra-se no apêndice D. Como o modelo multiplicativo não pode conter valores de medidas iguais a 0, aqueles indivíduos que apresentaram esta característica foram removidos do conjunto de dados, de forma que, ao final, a amostra continha as medidas de 24 indivíduos. Os valores iniciais foram escolhidos por meio da estimação via modelos mistos normal, e atribuindo valor 0 para  $\rho$ . A tabela 1 apresenta

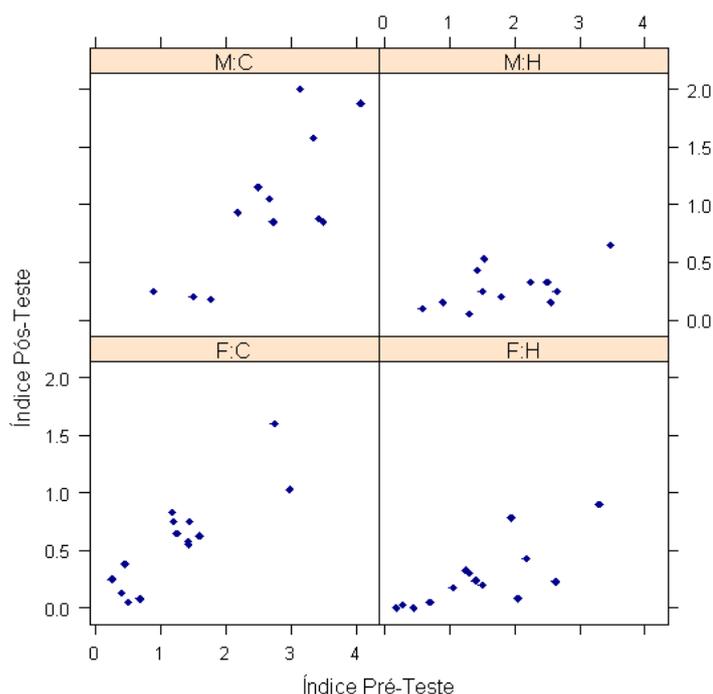


Figura 4: Gráficos de dispersão entre as medidas pós e pré-teste em função das classes : M: Sexo masculino, F: Sexo feminino, H: Escova do tipo "Hugger" e C: Escova convencional.

estes resultados e usando o modelo normal, como propõem Singer & Andrade (1997) e Sef (1999).

O resultado obtido aqui para o modelo normal difere daquele obtido por Sef (1999) pelo fato de que houve a exclusão dos indivíduos com algumas das medidas pré ou pós-teste faltantes, a fim de evitar problemas computacionais, como matrizes mal-condicionadas, por exemplo. A análise inicial indica que há certa tendência dos resíduos a serem dispersos assimetricamente.

Optou-se pelo critério AIC (Akaike Information Criteria) como uma maneira de se selecionar o “melhor” modelo, tal qual fizeram Sef (1999), Lachos (2004), Freitas (2005).

Notadamente, existe uma diferença significativa quanto aos parâmetros estimados entre os modelos (tabela 1). O modelo normal apresenta uma baixa estimativa da variância do efeito aleatório, um forte indicativo de que o modelo misto é desnecessário para este conjunto de dados, apesar de não muito adequado, haja vista o efeito de correlação. No entanto, o modelo SN não indica a mesma situação, uma vez que obtivemos um  $\rho$  pontual

maior que 0, que é significativo. Sem conhecer o intervalo de confiança para esta estatística, não é possível verificar, de fato, se a hipótese de que  $\rho \neq 0$  é verdadeira para este conjunto de dados. Esta é uma dificuldade da técnica utilizando-se o algoritmo E-M, a necessidade de utilização de métodos bootstrap, ou outros métodos, para a obtenção de intervalos de confiança.

| Fatores |        |               | Modelos              |                    |
|---------|--------|---------------|----------------------|--------------------|
| Sexo    | Escova | Coefficientes | Normal               | SN                 |
| Fem     | Conv   | $\alpha_{11}$ | -0.851 (0.187)       | -1.071 (0.153)     |
| Mas     | Conv   | $\alpha_{21}$ | -1.719 (0.448)       | -1.965 (0.415)     |
| Fem     | Hug    | $\alpha_{12}$ | -1.965 (0.210)       | -2.102 (0.156)     |
| Mas     | Hug    | $\alpha_{22}$ | -1.880 (0.275)       | -1.591 (0.336)     |
| Fem     | Conv   | $\gamma_{11}$ | 1.012 (0.266)        | 0.774 (0.460)      |
| Mas     | Conv   | $\gamma_{21}$ | 1.606 (0.452)        | 1.659 (0.414)      |
| Fem     | Hug    | $\gamma_{12}$ | 1.214 (0.296)        | 1.639 (0.356)      |
| Mas     | Hug    | $\gamma_{22}$ | 0.791 (0.386)        | 0.463 (0.477)      |
|         |        |               | $\rho$               | -6.530 (0.500)     |
|         |        |               | $\sigma_{\eta}^{2*}$ | 0.021              |
|         |        |               | $\sigma_{\nu}^{2*}$  | 0.0003             |
|         |        |               | $\lambda$            | -427.808 (234.074) |
| AIC     |        |               | 113.504              | -275.64            |

Tabela 1: Estimativas pelos modelos normal e normal assimétrico. Valores entre parênteses indicam o desvio padrão estimado para o modelo normal, obtido de maneira usual e o desvio padrão bootstrap para o modelo SN

Os coeficientes estimados, dados pelo vetor  $\hat{\gamma}^*$ , seguem o mesmo sinal, tanto no modelo normal, quanto no modelo SN, mas em grandezas diferentes. Esta diferença fica nítida, graficamente, ao se analisar a figura 5, em que a modelagem dos valores pós-teste para o modelo normal assimétrico é mais conservadora, pois indica uma menor redução nos índices de placa.

A forma dos gráficos de resíduos por valores preditos (figura 6) para o modelo linearizado (logarítmico) são semelhantes, mas a densidade estimada nos histogramas de resíduos, para o modelo SN é mais suave, o que já é um fator favorável, pois diminui a tendência à bimodalidade existente nos resíduos do modelo normal. Ainda, um fator favorável ao modelo SN, para este tipo de análise, é o critério AIC. Enquanto o valor

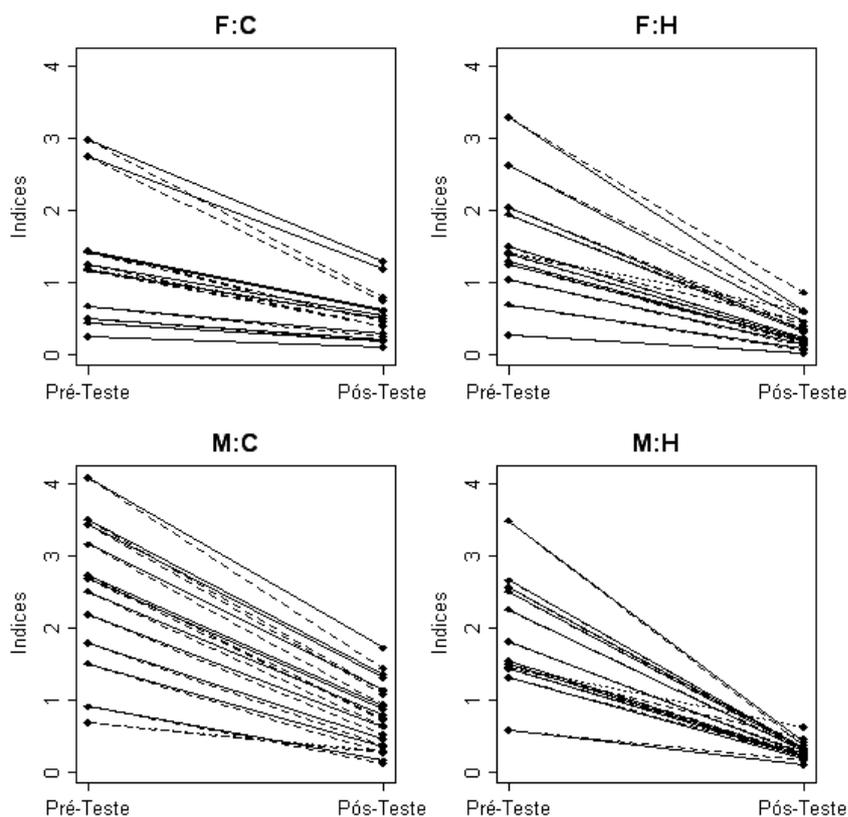


Figura 5: Curvas individuais para os modelos estimados em função das classes. A linha cheia indica o modelo normal e a linha tracejada o modelo normal assimétrico.

do AIC para o modelo normal foi positivo (113.5), o cálculo deste resultou em um valor negativo (-275.6) para o outro, e quanto menor o valor AIC, mais adequado é o modelo, ou seja, por este critério, escolher-se-ia o modelo normal assimétrico.

Como já fora dito, uma desvantagem da estimação via algoritmo E-M na inferência clássica é a dificuldade em se encontrar intervalos de confiança. A técnica, por si, não fornece, de imediato, um meio de se encontrar quantidades pivotais para se calcular estes intervalos, baseados em uma distribuição de probabilidade conhecida. Desta maneira, lançou-se mão da simulação bootstrap (EFRON; TIBHIRANI, 1994), em que os indivíduos sob estudo (todos com ambas as medidas pré/pós-teste, conforme o procedimento utilizado nesta aplicação), são selecionados aleatoriamente com reposição e um novo conjunto de dados é formado, e utilizando-se os mesmos valores iniciais, um novo modelo é estimado. Isto foi repetido por 500 vezes e um intervalo de confiança não-paramétrico limitado pelos quantis 2,5% e 97,5% pôde ser obtido. No caso do modelo normal, como ele já foi

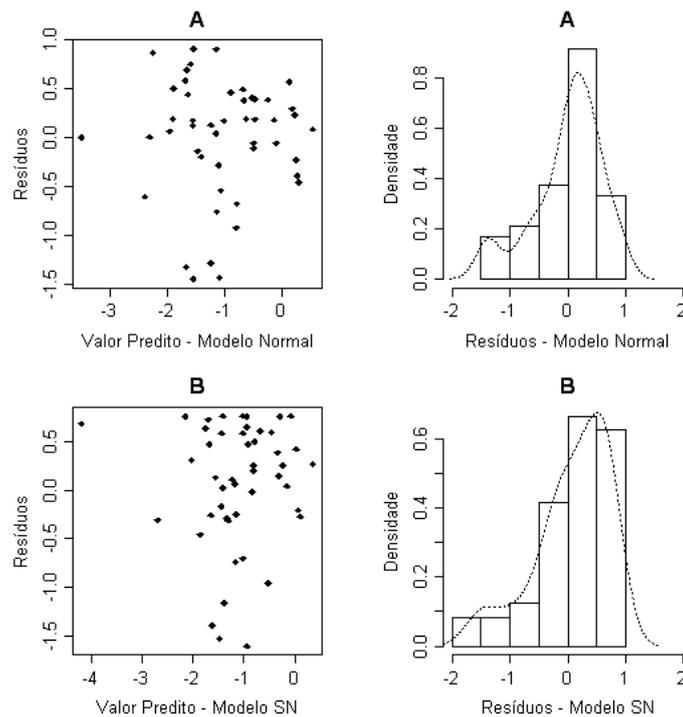


Figura 6: Gráfico de dispersão de resíduos por valores preditos. A: Modelo Normal (Sef & Singer); B: Modelo SN (Sahu & Dey).

exaustivamente estudado e com as técnicas assintóticas disponíveis, a variabilidade dos estimadores é obtida facilmente em programas estatísticos, como no pacote 'nlme' em R (VENABLES; RIPLEY, 2002; R Development Core Team, 2006) ou 'Proc Mixer' em *SAS*<sup>®</sup>.

A funcionalidade do modelo multiplicativo para dados pré-teste/pós-teste, principalmente quando aplicado a este conjunto de dados, é a interpretabilidade dos parâmetros. Assim, as tabelas a seguir apresentam os intervalos de confiança (IC) de 95%, assintóticos para o modelo normal e bootstrap para o modelo SN, de forma que comentários serão feitos posteriormente.

A análise da tabela 2, que contém os intervalos de confiança assintóticos de 95% com os coeficientes já em escala natural (por isso utilizando a notação  $\beta$ , ao invés de  $\alpha$ ), revelam que os coeficientes são significativos a 5%. O que chama a atenção neste caso, é que não podemos rejeitar a hipótese de que  $\gamma_{ij} = 1$  para qualquer uma das combinações entre sexo e escova. Assim, segundo este modelo, a redução do índice de placa é constante. Ainda, a placa residual tende a ser maior em crianças do sexo feminino e utilizando escovas convencionais (o maior coeficiente é  $\beta_{11}$  e esta conclusão pode ser observada na figura 5,

| IC 95%        |                 |            |                 |
|---------------|-----------------|------------|-----------------|
| Parâmetro     | Limite Inferior | Estimativa | Limite Superior |
| $\beta_{11}$  | 0.296           | 0.427      | 0.615           |
| $\beta_{21}$  | 0.075           | 0.179      | 0.431           |
| $\beta_{12}$  | 0.093           | 0.140      | 0.212           |
| $\beta_{22}$  | 0.089           | 0.153      | 0.261           |
| $\gamma_{11}$ | 0.491           | 1.012      | 1.378           |
| $\gamma_{21}$ | 0.721           | 1.606      | 2.483           |
| $\gamma_{12}$ | 0.634           | 1.214      | 1.627           |
| $\gamma_{22}$ | 0.034           | 0.791      | 1.330           |

Tabela 2: Intervalos de confiança assintóticos de 95% para os coeficientes do modelo multiplicativo normal.

$F:C$ ). Para as outras combinações, estes coeficientes são relativamente semelhantes.

| IC 95% Bootstrap |                 |            |                 |
|------------------|-----------------|------------|-----------------|
| Parâmetro        | Limite Inferior | Estimativa | Limite Superior |
| $\beta_{11}$     | 0.247           | 0.371      | 0.503           |
| $\beta_{21}$     | 0.051           | 0.058      | 0.246           |
| $\beta_{12}$     | 0.068           | 0.077      | 0.185           |
| $\beta_{22}$     | 0.093           | 0.099      | 0.370           |
| $\gamma_{11}$    | 0.572           | 0.774      | 2.567           |
| $\gamma_{21}$    | 1.133           | 1.659      | 2.678           |
| $\gamma_{12}$    | 0.574           | 1.639      | 2.135           |
| $\gamma_{22}$    | -0.814          | 0.463      | 1.227           |
| $\rho$           | -8.572          | -6.530     | -2.621          |
| $\lambda$        | -745.329        | -427.808   | -2.526          |

Tabela 3: Intervalos de confiança bootstrap de 95% para os coeficientes do modelo multiplicativo SN com os interceptos já corrigidos.

A tabela 3 representa os intervalos de confiança bootstrap de 95% para os parâmetros estimados do modelo SN para dados pré/pós-teste. Como estes intervalos de confiança são não-paramétricos, eles não concordam com os erros padrão apresentados na tabela 1. Uma rápida análise indica que o coeficiente estimado para aqueles indivíduos do sexo masculino, que utilizaram escovas do tipo “Hugger”, não seja significativo.

Neste modelo com distribuição SN, não se rejeita, também, a hipótese de que  $\gamma_{ij} = 1$ , o que implica em dizer que a taxa de remoção da placa bacteriana, independentemente da escova ou do sexo é constante. Houve uma diminuição significativa, a um nível significância de 5%, nos coeficientes que indicam a placa bacteriana residual, quando estes são comparados com os do modelo normal. Os valores estimados para a medida pós-teste

do modelo que utiliza a distribuição normal assimétrica com a representação de Sahu & Dey tendem a ser menores quando comparados com o modelo utilizando a distribuição normal.

Especificamente, tem-se que a hipótese de que um modelo normal é rejeitada, pois há evidências de que  $\rho \neq 0$ . Claramente os dados são muito assimétricos ( $\lambda$  muito negativo), e isso é importante, uma vez que é proposta uma modificação do modelo que incorpore esta possibilidade de captação das assimetrias. Além disso, o modelo normal associado (Sef & Singer) indica que o modelo de regressão linear múltipla pode ser adequado, haja vista que a variância do efeito aleatório foi muito baixa, e no modelo estimado com as distribuições normais assimétricas, o modelo é misto, uma vez que o IC para  $\rho$  não inclui o valor nulo.

No que se propõe com este modelo, obteve-se o esperado. Ele é capaz de captar as assimetrias e ainda ser um modelo misto com todas as características inerentes a essa ferramenta estatística.

#### 4.3.1.2 Dificuldades Técnicas

Apesar de haver tido êxito na obtenção de estimativas do modelo multiplicativo linearizado por transformação logarítmica, muitos problemas são encontrados quanto aos aspectos computacionais. Rodríguez (2005) relata problemas com a estimação do parâmetro de forma de uma distribuição normal assimétrica, principalmente quando se utiliza o método da máxima verossimilhança tradicional, pois podem não existir estes estimadores ou podem haver multimodalidades. Com a possibilidade de utilização do algoritmo E-M para a obtenção destas estimativas de máxima verossimilhança, este problema é minimizado, mas ainda, falta robustez ao método, pois diferentes valores iniciais levam a diferentes valores estimados para o parâmetro de forma.

Ao utilizar a representação de Sahu & Dey, este problema fica exposto, pois há uma hiperparametrização do parâmetro de assimetria, pois ele fica em função de  $\rho$ ,  $\sigma_\nu$  e  $\sigma_\eta$ . Assim, há grande dificuldade em estimar com robustez e eficiência este importante cons-

tituinte do modelo e da distribuição de probabilidade subjacente.

Ao escolher os valores iniciais, deve-se priorizar aqueles obtidos com o modelo misto normal, e definir  $\rho^{(0)} = 0$  como valor inicial. É interessante observar que os coeficientes lineares são robustos aos valores iniciais, talvez porque o método de estimação, quando se utiliza o algoritmo E-M, fica muito próximo ao método de mínimos quadrados.

A construção de intervalos de confiança utilizando-se quantidades pivotais fica, também, prejudicada quando se lança mão do algoritmo E-M para obtenção de estimativas. Assim, adicionalmente, métodos bootstrap são recomendados para se encontrar tais estatísticas, e para tanto, computadores com grande capacidade de processamento e algum tempo são necessários para obtê-las com eficiência.

Quanto ao modelo em si, a dificuldade encontrada é relacionada aos valores nulos, que não são permitidos para as medidas pré ou pós-teste, pelo fato de que a estimação e a inferência é feita sob um modelo linearizado pela transformação logarítmica.

## *Parte II*

### *Inferência Bayesiana*

## 5 *Introdução*

Com o avanço dos recursos computacionais e pela implementação, nestes sistemas, de algoritmos como o Monte Carlo Cadeias de Markov (MCMC), a inferência bayesiana deu um grande salto, tornando-se, de fato, uma ferramenta estatística com muitos potenciais.

A abordagem bayesiana implica em utilizar probabilidades como forma de se obter medir o grau de credibilidade das informações ou parâmetros (PAULINO et al., 2003) e isso faz com que o conhecimento prévio sobre o fenômeno em estudo faça parte da computação dos parâmetros de interesse do pesquisador, que ao se basear no conhecimento apriorístico e um conjunto de dados, tem a capacidade de fazer inferências a cerca de parâmetros desconhecidos, dado a amostra que ele dispõe.

Esta caracterização torna a inferência bayesiana versátil, pois os parâmetros estão fixos no espaço paramétrico, como na inferência clássica, mas agora, podem receber distribuições de probabilidades, uma vez que são desconhecidos, de modo que estas distribuições indiquem o grau de incerteza sobre eles. Enfim, a inferência bayesiana trata dos problemas probabilisticamente, levando em conta o conhecimento prévio sobre os fenômenos para qual esta ferramenta é utilizada.

Ainda, como parte da versatilidade desta abordagem estatística, existe a possibilidade de não se inferir nada ou muito pouco à priori sobre o espaço paramétrico de interesse, o que leva à definição das *prioris não-informativas*.

A arte da inferência bayesiana, está, então, diretamente relacionada à busca por uma priori, informativa ou não, que possa representar bem os conhecimentos prévios sobre o espaço paramétrico relativos aos fenômenos em estudo, e quando não se quiser inse-

rir informações, por falta de conhecimento ou qualquer outro motivo, a distribuição a priori deve levar consigo o mínimo ou nenhum conhecimento prévio sobre aquele espaço paramétrico.

Dessa maneira, várias técnicas foram desenvolvidas para se definir uma distribuição que represente a credibilidade prévia, ou a ausência dela, e nesta dissertação trataremos somente das *prioris não-informativas de Jeffreys* para um modelo com distribuição normal assimétrica.

Rodríguez (2005), em sua dissertação de mestrado, encontra uma aproximação para a priori de Jeffreys para o parâmetro de assimetria das normais assimétricas baseada na distribuição t-Student. Mostraremos, posteriormente, que existe uma extensão dessa priori para a representação de Sahu & Dey, baseada na distribuição de Kummer, e denominada Priori de Kummer, de forma que obtivemos uma forma fechada para a priori do parâmetro de modulação da assimetria  $\rho$ , com a vantagem de se mostrar versátil, pois, pode ser informativa ou não, dependendo dos hiperparâmetros escolhidos. Estas considerações serão apresentadas na seção 6.2.

Começaremos, então, pela introdução da priori de Jeffreys, a seguir por apresentar o resultado de Rodríguez (2005) e o estudo da priori T-Gama.

## 5.1 Priori de Jeffreys

Existem muitas situações na prática estatística em que o pesquisador sabe muito pouco sobre fenômeno em estudo, de forma que o conhecimento é vago com relação à informação amostral (PAULINO et al., 2003). Então, é praxe utilizar distribuições a priori que informem muito pouco ou nada sobre o espaço paramétrico em questão, de modo que a informação prévia seja difusa. Nestes casos, se diz que existe uma ignorância sobre o espaço paramétrico à priori. Assim, buscou-se um método que padronizasse e fosse objetivo para esta procura por *prioris* vagas e que não se restringisse somente à utilização da distribuição uniforme, pois esta não é invariável sobre transformações injetivas, como cita Paulino et al. (2003). Obteve-se sucesso nesta busca, quando foi baseada na medida

de informação de Fisher sobre  $\theta \in \mathfrak{R}$ . Devemos lembrar que a informação de Fisher pode ser definida, para o caso univariado como:

$$I(\theta) = -E \left[ \frac{\partial^2 \ln f(x|\theta)}{\partial \theta^2} \mid \theta \right]. \quad (5.1)$$

Em condições de regularidade, temos que  $I(\theta)$  pode ser simplificada pela seguinte expressão:

$$I(\theta) = E \left[ \left( \frac{\partial \ln f(x|\theta)}{\partial \theta} \right)^2 \mid \theta \right]. \quad (5.2)$$

O resultado da expressão (5.2) é derivado do conhecimento que se tem de que a esperança do escore  $\ln f(x|\theta)$  é nula e portanto, a derivada segunda daquela função é equivalente à variância do escore, de modo que a informação de Fisher pode ser dada, em condições de regularidade, pela esperança do quadrado de sua derivada primeira.

**Definição 5.1.1 (Priori de Jeffreys)** *Dada uma função de probabilidade  $f(x|\theta)$ , com  $\theta \in \mathfrak{R}$ , então a priori de Jeffreys para  $h(\theta)$  é dada por:*

$$J(\theta) = [I(\theta)]^{\frac{1}{2}}, \quad (5.3)$$

onde  $I(\theta)$  é dada pela expressão (5.1).

A principal vantagem da priori de Jeffreys é a sua invariância sob transformações injetivas (PAULINO et al., 2003), mas, em muitos casos pode resultar em funções densidades impróprias, o que não garante que a posteriori seja própria.

O interesse principal desta dissertação é utilizar as prioris vagas para o parâmetro de assimetria, uma vez que a forma de Sahu et al. (2003) envolve este parâmetro e ele é importante para todo o problema.

Rodríguez (2005) estudou uma distribuição a priori vaga de Jeffreys para o parâmetro de assimetria  $\lambda$  (2.1), e esta distribuição pôde ser aproximada por uma distribuição própria, de forma que utilizaremos estes resultados para construir um modelo bayesiano

para o tratamento do modelo multiplicativo de Sef & Singer sob a representação de Sahu & Dey.

## 5.2 Priori não-informativa de Jeffreys para o parâmetro de assimetria

Recordando a forma básica de uma normal assimétrica padrão definida por Azzalini (1985), temos que  $Z \sim SN(\lambda)$  com densidade dada pela expressão (2.1) tal que  $\mu = 0$  e  $\sigma^2 = 1$ . A informação de Fisher  $I(\lambda)$  (5.1) para a distribuição precisa ser calculada, no entanto, temos que encontrar a derivada primeira da log-verossimilhança para a distribuição, com uma amostra inicialmente. Assim,

$$\frac{\partial \ln f(z|\lambda)}{\partial \lambda} = \frac{z\phi(\lambda z)}{\Phi(\lambda z)}. \quad (5.4)$$

Agora, a informação de Fisher por ser calculada, utilizando-se a expressão (5.4) e a função densidade de probabilidade da normal assimétrica padrão.

$$\begin{aligned} I(\lambda) &= E \left[ \left( \frac{\partial \ln f(z|\lambda)}{\partial \lambda} \right)^2 \right] \\ &= \int_{-\infty}^{\infty} \frac{z^2 \phi^2(\lambda z)}{\Phi^2(\lambda z)} 2\phi(z) \Phi(\lambda z) dz \\ &= \int_{-\infty}^{\infty} 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\ &= \int_0^{\infty} 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + \int_{-\infty}^0 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz \\ &= \int_0^{\infty} 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z) [1 - \Phi(\lambda z)]} dz. \end{aligned} \quad (5.5)$$

A integral resultante (5.5) não é tratável analiticamente. No entanto, existe um resultado apresentado por Rodríguez (apud CHAIBUB NETO; BRANCO, 2003) que faz a aproximação de um conjunto de funções do integrando simplificando-o.

**Proposição 5.2.1** *Rodríguez (apud CHAIBUB NETO; BRANCO, 2003).*

$$\frac{\phi(z)}{\pi\Phi(z)[1-\Phi(z)]} \approx \underbrace{\frac{1}{\sqrt{2\pi}\left(\frac{\pi}{2}\right)} \exp\left[-\frac{z^2}{2\left(\frac{\pi^2}{4}\right)}\right]}_{N\left(0, \frac{\pi^2}{4}\right)}. \quad (5.6)$$

**Proof.** A proposição 5.2.1 é apresentada em Rodríguez (2005) e referências sobre a demonstração podem ser encontradas naquele trabalho. ■

Se observarmos o integrando de (5.5), notaremos que a aproximação poderá ser utilizada, bastando apenas, ajustá-la quanto ao argumento das funções. A figura 7 indica que a aproximação é excelente. Assim, utilizando (5.6) em (5.5), temos:

$$\begin{aligned} I(\lambda) &= \int_0^\infty 2z^2\phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)[1-\Phi(\lambda z)]} dz. \\ &\approx \int_0^\infty 2z^2\phi(z) \frac{\pi}{2\left(\frac{\pi^2}{4}\right)} \exp\left[-\frac{\lambda^2 z^2}{\left(\frac{\pi^2}{4}\right)}\right] dz \\ &= \frac{2}{\pi} \left[1 + \frac{2\lambda^2}{\left(\frac{\pi^2}{4}\right)}\right]^{-\frac{1}{2}} \underbrace{\int_0^\infty z^2\phi\left(z\left|0, \left[1 + \frac{2\lambda^2}{\left(\frac{\pi^2}{4}\right)}\right]^{-1}\right.\right) dz}_{E(Z^2|Z>0)}. \end{aligned} \quad (5.7)$$

O integrando na expressão (5.7) é a esperança de  $Z^2$  dado que  $Z > 0$ . Como esta variável é normal, então podemos utilizar o teorema 4.3.1 e calcular esta esperança analiticamente. Então, utilizando o resultado temos que ao final,

$$I(\lambda) \propto \frac{2}{\pi} \left[1 + \frac{2\lambda^2}{\left(\frac{\pi^2}{4}\right)}\right]^{-\frac{3}{2}}, \quad (5.8)$$

e, portanto, uma aproximação da priori de Jeffreys para  $\lambda$  é dada por

$$J(\lambda) \propto \sqrt{\frac{2}{\pi}} \left[1 + \frac{2\lambda^2}{\left(\frac{\pi^2}{4}\right)}\right]^{-\frac{3}{4}}, \quad (5.9)$$

que é o núcleo de uma distribuição

$$t\left(0, \frac{\pi^2}{4}, \frac{1}{2}\right). \quad (5.10)$$

Em outras palavras,  $J(\lambda) \propto t\left(0, \frac{\pi^2}{4}, \frac{1}{2}\right)$ , onde o símbolo  $\propto$  significa “**proporcional a**”.

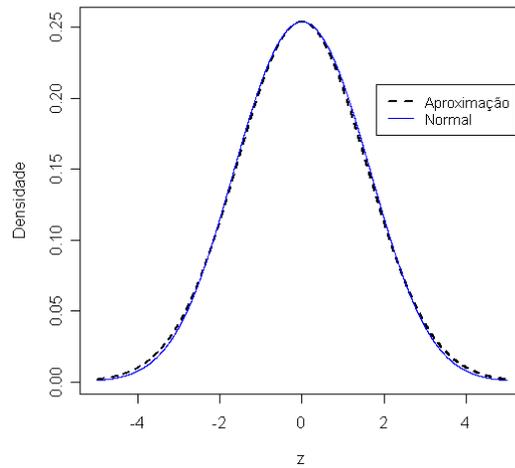


Figura 7: Comparação entre a aproximação de Chaibub Neto & Branco com relação à distribuição normal

Ter uma distribuição a priori própria implica em sempre obter uma posteriori própria (GELMAN et al., 2000; PAULINO et al., 2003), e isso é uma grande vantagem na análise bayesiana da posteriori, principalmente quando existe a intenção de utilizar algoritmos MCMC para a estimação dos parâmetros de interesse.

A figura 7 mostra que a aproximação da priori de Jeffreys pela distribuição t-Student é muito boa, com a vantagem de não necessitar de métodos de integração para a obtenção dos valores de  $J(\lambda)$ .

Outra consideração deve ser feita quando se trabalha com prioris de normais assimétricas: É comum formular a priori para o coeficiente de assimetria  $\delta$ , tal que  $\delta \sim U(-1,1)$ .

No entanto, define-se o coeficiente de assimetria como  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ , e portanto, a transformação desta variável aleatória induz  $\lambda \propto t\left(0, \frac{1}{2}, 2\right)$ , o que leva muito mais informação para a priori do que se se utilizasse, somente, a priori aproximada para  $\lambda$ . A figura 9 indica este relacionamento. Note que há uma massa de probabilidade muito grande em torno de 0, o que acontece em menor grau com a priori de Jeffreys.

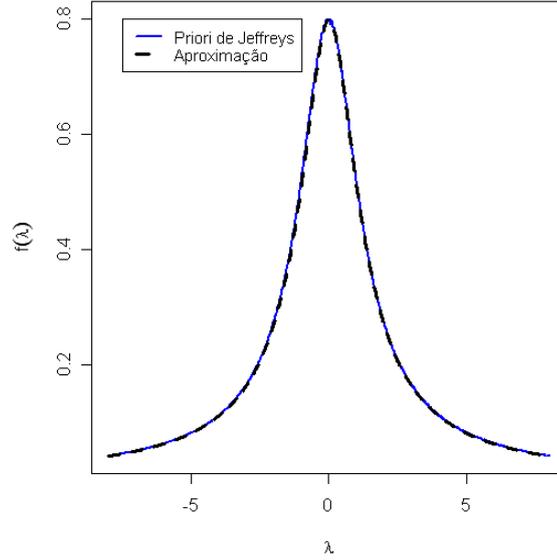


Figura 8: Comparação entre  $J(\lambda)$  e sua aproximação pela  $t\left(0, \frac{\pi^2}{4}, \frac{1}{2}\right)$ .

### 5.3 Definição das prioris para o modelo de Sef & Singer.

Relembrando a definição do modelo multiplicativo para dados pré-teste/pós-teste de Sef & Singer (SINGER; ANDRADE, 1997; SEF, 1999), com erro normal assimétrico como definido por Sahu et al. (2003) e Liu & Dey (2005), podemos, agora, definir prioris para os parâmetros envolvidos no modelo multiplicativo (4.22). Naquela expressão, temos que normal assimétrica é dada pela seguinte relação

$$\begin{aligned}\zeta_{ijk} &= \sigma_\eta \rho |\eta_k| + \sigma_\nu \nu_{ijk}, \\ \eta_k &\sim N(0,1) \text{ e } \nu_{ijk} \sim N(0,1), \\ \eta_k &\perp \nu_{ijk}, \\ \zeta_{ijk} &\sim SN\left(0, \sigma_\nu^2 + \rho^2 \sigma_\eta^2, \lambda = \frac{\rho \sigma_\eta}{\sigma_\nu}\right).\end{aligned}$$

O parâmetro de assimetria  $\lambda$  depende do parâmetro que modula a distribuição semi-normal  $\rho \in \mathfrak{R}$ , que é o contaminante para tornar a distribuição normal clássica, assimétrica. Ainda, devido a um resultado de Rodríguez (2005), aproximação da priori para o parâmetro de assimetria é  $\lambda \propto t\left(0, \frac{\pi^2}{4}, \frac{1}{2}\right)$ , logo, se fizermos uma transformação de variáveis aleatórias, considerando dadas (fixas) as variâncias  $\sigma_\eta^2$  e  $\sigma_\nu^2$ , poderemos saber qual é a

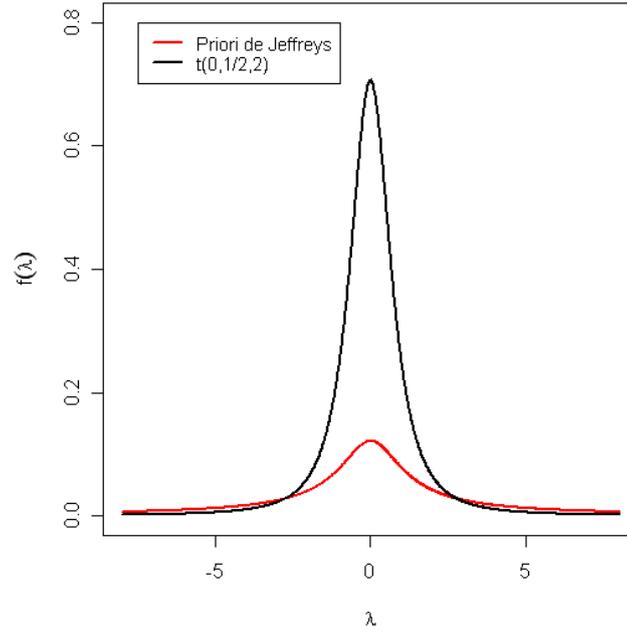


Figura 9: Comparação entre as prioris de Jeffreys e o resultado da indução sobre  $\lambda$  de  $\delta \sim U(-1,1)$ .

mistura de distribuições que são induzidas a  $\rho|\sigma_\eta^2, \sigma_\nu^2$ , pela própria distribuição t-Student aproximada para  $\lambda$ , segundo a aproximação de Rodríguez (2005) e pelas distribuições gamas invertidas que serão atribuídas às variâncias  $\sigma_\nu^2$  e  $\sigma_\eta^2$ . Assim, lembrando que a distribuição t-Student é linear quanto às transformações de variáveis aleatórias,

$$\rho = \frac{\lambda\sigma_\nu}{\sigma_\eta} \text{ com } \lambda \sim t\left(0, \frac{\pi^2}{4}, \frac{1}{2}\right) \implies \rho|\sigma_\eta^2, \sigma_\nu^2 \sim t\left[0, \frac{(\pi\sigma_\nu)^2}{4\sigma_\eta^2}, \frac{1}{2}\right]. \quad (5.11)$$

Observe que se  $\sigma_\eta^2 = \sigma_\nu^2$ , ou seja, a variância do erro e efeitos aleatórios são iguais, a priori em (5.11) é a mesma que definida por Rodríguez (2005) (expressão 5.10). A figura 10 mostra algumas formas da mistura de prioris induzidas para  $\rho$ .

Nesta figura se nota, considerando  $\sigma_\nu^2$  fixo, que a medida que se aumentar o parâmetro de escala da variável seminormal  $\sigma_\eta^2$ , menos informativa fica a distribuição a priori para  $\rho$ . Já o aumento da variância da distribuição normal no erro,  $\sigma_\nu^2$ , mais vaga se torna a distribuição. Isto acontece porque o parâmetro de escala resultante da distribuição assimétrico do modelo em estudo é dada por  $\sigma_\nu^2 + \rho^2\sigma_\eta^2$ , e quando  $\sigma_\eta^2 \rightarrow 0$  e  $\sigma_\nu^2 \rightarrow \infty$ , então  $\rho \rightarrow \infty$ , que faz o parâmetro de escala final tender ao infinito, o que caracteriza a

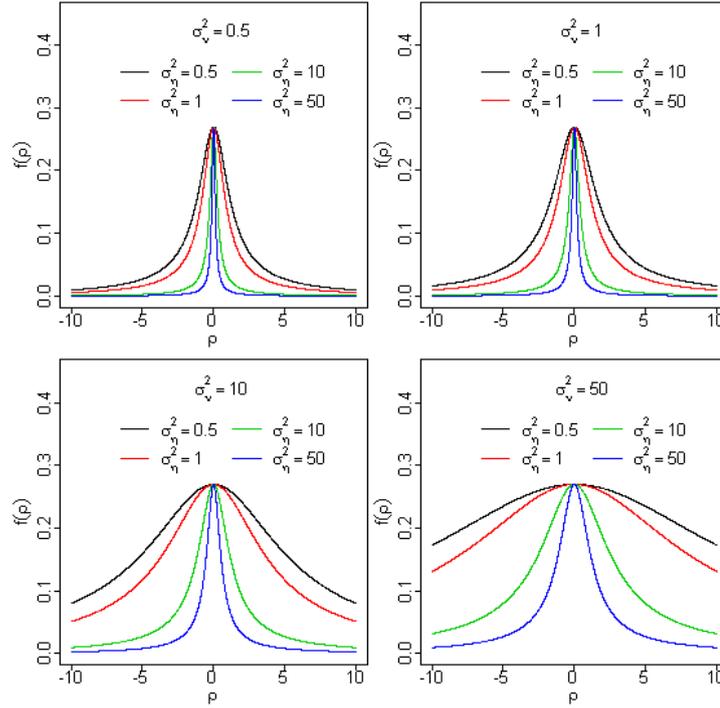


Figura 10: Distribuições das priors de  $\rho$  com vários valores para  $\sigma_\eta^2$  e  $\sigma_\nu^2$ .

vagueza de uma informação. Quando as variâncias são iguais, chega-se à priori proposta por Rodríguez (2005), como já foi citado anteriormente.

Para fins computacionais, poder-se-ia utilizar, também, uma distribuição normal centrada em 0 com variância relativamente grande para representar a vagueza, pois os valores de  $\rho$  estão definidos para todos os reais no espaço paramétrico. No entanto, a distribuição t-Student tem a vantagem de ter caudas mais pesadas, dando, portanto, mais credibilidade a valores afastados de 0 com relação à distribuição normal.

Os últimos parâmetros para os quais devem ser definidas distribuições a priori são os coeficientes do modelo. Lembrando-se da notação adotada para representar o modelo de Sef & Singer, (expressão 4.22), é natural imaginar que os coeficientes denotados por  $\alpha_{ij}$  e  $\gamma_{ij}, i = 1, 2, j = 1, 2$ , mais genericamente podem ser agrupados no vetor  $\gamma^*(4.7)$ , estão sob o espaço paramétrico dos números reais, de forma que se aplica adotar uma distribuição normal centrada em 0 com variância relativamente grande para representar a vagueza para cada um destes parâmetros. Computacionalmente esta é a melhor forma de se atacar o problema. As variâncias podem ter, a priori, distribuições gamas invertidas.

Com *prioris* definidas, podemos, a seguir, formalizar o modelo multiplicativo de Sef & Singer para dados pré-teste/pós-teste com erro subjacente definido pela representação da normal assimétrica segundo Sahu & Dey.

## 6 *Formulação do modelo bayesiano*

Para iniciar a formulação do modelo multiplicativo para dados pré-teste/pós-teste de Singer & Andrade com erros normais assimétricos, devemos lembrar sua forma básica dada pelas expressões em (4.11), sua forma matricial, cuja definição é apresentada pelas expressões 4.6-4.10 e reescrevê-lo sob a forma estocástica, representada em (4.12). Aquelas expressões são transcritas aqui para que se definam os parâmetros do modelo.

$$\begin{aligned} \mathbf{Z}_k &= \mathbf{W}_k^* \boldsymbol{\gamma}_k^* + (\sigma_\eta \rho |\eta_k| + \sigma_\nu \nu_{lk}^*) \mathbf{I}_4 \\ \eta_k &\sim N(0,1) \\ \boldsymbol{\nu}_k^* &\sim N(0,1) \mathbf{I}_4, \quad k = 1, \dots, n. \end{aligned} \tag{6.1}$$

A verossimilhança de  $\mathbf{Z}_k$  pode ser escrita de forma hierárquica de modo possamos definir a distribuição a posteriori. Inicialmente este processo será tratado para uma amostra de tamanho um, de modo que os índices serão omitidos. Ainda, a matriz  $\mathbf{W}_k^*$  representa a matriz de planejamento do experimento e  $\boldsymbol{\gamma}^*$ , o vetor de coeficientes de dimensão  $8 \times 1$ .

Representado  $\mathbf{Z}_k$  de forma hierárquica, com a representação matricial dada em (4.11) e formulações que seguem a esta expressão, temos:

$$Z_{lk}^* \mid (|\eta_k|, \boldsymbol{\gamma}^*, \rho, \sigma_\eta^2, \sigma_\nu^2) \sim N(\mathbf{R}_l \boldsymbol{\gamma}^* + \rho \sigma_\eta |\eta_k|, \sigma_\nu^2); \tag{6.2}$$

$$|\eta_k| \sim HN(0,1), \tag{6.3}$$

para  $l = 1, \dots, 4; k = 1, \dots, n$ .

A verossimilhança de  $\mathbf{Z}_k$  é definida em dois estágios, mas como se nota, muitos parâmetros precisam de priors, e nesse caso, todas não-informativas, e quando possível,

prioris conjugadas, uma vez que temos distribuições normais nas formas condicionais do modelo. As distribuições a priori, são, portanto:

$$\begin{aligned} \rho | (\sigma_\eta^2, \sigma_\nu^2) &\sim t \left[ 0, \frac{(\pi\sigma_\nu)^2}{4\sigma_\eta^2}, \frac{1}{2} \right], \\ \frac{1}{\sigma_\nu^2} &\sim \text{Gama} \left( \frac{a}{2}, \frac{b}{2} \right), \\ \frac{1}{\sigma_\eta^2} &\sim \text{Gama} \left( \frac{a}{2}, \frac{b}{2} \right) \text{ e,} \\ \boldsymbol{\gamma}^* &\sim N_8(0, \boldsymbol{\Sigma}), \end{aligned} \tag{6.4}$$

onde o  $\lim(a,b) \rightarrow 0$  e  $\boldsymbol{\Sigma}$  é a matriz diagonal, de dimensão  $8 \times 8$ , com cada um dos elementos da diagonal tendendo ao infinito. Note que as marginais de  $f(\boldsymbol{\gamma}^*)$  são normais com média 0 e variância, digamos,  $\psi^2$ , e que essas distribuições normais são independentes.

Os parâmetros do modelo são interpretáveis, como já foi exposto nos capítulos anteriores, de sorte que não repetiremos aqui estas interpretações. A priori mais importante é a do parâmetro de modulação da assimetria,  $\rho$ , uma vez que a distribuição t-Student atribuída a ele é resultado de uma mistura de distribuições induzida pela aproximação da priori de Jeffreys para o parâmetro de assimetria  $\lambda$ , mostrado por Rodríguez (2005), ponderada pelos valores de  $\sigma_\nu^2$  e  $\sigma_\eta^2$ , que são distribuições gamas invertidas vagas (“flat”), para assegurar que, pouca, ou nenhuma informação à priori seja fornecida àquelas variâncias. Ao inverso destes parâmetros de escala podem ser atribuídas distribuições gama.

É de se imaginar que a posteriori não tenha uma forma analítica simples, mas com os recursos computacionais existentes hoje, a implementação de algoritmos MCMC poupados de cálculos integrais complexos (GAMERMAN; LOPES, 2006), como o que aqui propomos.

Enfim, a posteriori do modelo multiplicativo de Sef & Singer para dados pré-teste/pós-teste com distribuição normal assimétrica segundo a forma de Sahu & Dey é dada por

$$\begin{aligned} f(\boldsymbol{\theta} | Z_{lk}^*) &\propto \overbrace{f(z_{lk}^* | |\eta_k|, \boldsymbol{\gamma}^*, \rho, \sigma_\eta^2, \sigma_\nu^2)}^{\text{Verossimilhança}} \times f(|\eta_k|) \times f(\boldsymbol{\gamma}^*) \\ &\times f(\rho | (\sigma_\eta^2, \sigma_\nu^2)) \times f(\sigma_\nu^2) \times f(\sigma_\eta^2), \end{aligned} \tag{6.5}$$

onde  $\boldsymbol{\theta}' = (\boldsymbol{\gamma}^*, \rho, \sigma_\eta^2, \sigma_\nu^2)'$  e  $f(\cdot)$  são as funções densidade de probabilidade. A expressão (6.5) pode ser escrita de maneira analítica, para se mostrar o grau de complexidade das integrais envolvidas.

$$\begin{aligned}
 f(\boldsymbol{\theta} | Z_{lk}^*) &\propto \overbrace{\prod_{k=1}^n \prod_{l=1}^4 \left\{ \frac{1}{\sqrt{\sigma_\nu^2}} \exp \left[ -\frac{1}{2} \left( \frac{z_{lk}^* - \mathbf{R}_l \boldsymbol{\gamma}^* - \rho \sigma_\eta |\eta_k|}{\sigma_\nu} \right)^2 \right] \times \exp \left( -\frac{1}{2} \eta_k \right) I_{\{0, \infty\}}^{(\eta_k)} \right\}}^{\text{Verossimilhança}} \\
 &\times \boldsymbol{\Sigma}^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \boldsymbol{\gamma}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}^{*'} \right) \times \frac{\frac{1}{2}^{\frac{1}{4}}}{\sqrt{\frac{(\pi \sigma_\nu)^2}{4 \sigma_\eta^2}}} \left[ 1 + \frac{\rho^2}{\frac{(\pi \sigma_\nu)^2}{4 \sigma_\eta^2}} \right]^{-\frac{3}{4}} \\
 &\times \left( \frac{a}{2} \right)^{\frac{b}{2}} \left( \frac{1}{\sigma_\nu^2} \right)^{\frac{a}{2}-1} \exp \left( -\frac{b}{2 \sigma_\nu^2} \right) \times \left( \frac{a}{2} \right)^{\frac{b}{2}} \left( \frac{1}{\sigma_\eta^2} \right)^{\frac{a}{2}-1} \exp \left( -\frac{b}{2 \sigma_\eta^2} \right).
 \end{aligned} \tag{6.6}$$

A distribuição t-Student pode ser definida como uma mistura de distribuições normais com uma distribuição gama para o inverso de seu parâmetro de escala (GAMERMAN; LOPES, 2006). A proposição 6.0.1 mostra a forma da mistura. Esta definição é importante quando se pretende implementar esta distribuição no WinBUGS e não há interesse em utilizar a técnica Metropolis em Gibbs. Mais sobre o OpenBUGS pode ser visto em Thomas (2006) e Thomas et al. (2006).

**Proposição 6.0.1** *Seja  $X | \mu, \sigma^2, V = v \sim N \left( \mu, \frac{\sigma^2}{v} \right)$  e  $V \sim \text{Gama} \left( \frac{a}{2}, \frac{a}{2} \right)$ , com a conhecido. Então,*

$$X | \mu, \sigma^2 \sim t_a(\mu, \sigma^2),$$

com  $\mu$  sendo o parâmetro de locação,  $a$ , o número de graus de liberdade e  $\sigma^2$ , o parâmetro de escala desta distribuição t-Student.

Dessa maneira, a distribuição  $t_{\frac{1}{2}} \left( 0, \frac{(\pi \sigma_\nu)^2}{4 \sigma_\eta^2} \right)$ , que é a priori induzida para o parâmetro de modulação da normal assimétrica escrita na forma de Sahu & Dey pode ser escrita como

$$\rho | \sigma_\nu^2, \sigma_\eta^2 \sim t_{\frac{1}{2}} \left( 0, \frac{(\pi \sigma_\nu)^2}{4 \sigma_\eta^2} \right) \stackrel{\text{mistura}}{\equiv} \begin{cases} \rho | Q \sim N \left( 0, \frac{(\pi \sigma_\nu)^2}{4 q \sigma_\eta^2} \right) \\ Q \sim \text{Gama} \left( \frac{1}{4}, \frac{1}{4} \right) \end{cases}. \tag{6.7}$$

Seria interessante conhecer a distribuição marginal de  $\rho$  ponderada pelas variâncias do modelo. Na próxima seção, será introduzida uma forma fechada para a distribuição

deste parâmetro, levando em conta uma reparametrização envolvendo as variâncias.

## 6.1 A distribuição de Kummer

Com o objetivo de se conseguir uma forma fechada para a distribuição à priori marginal de  $\rho$ , será introduzida agora a distribuição de Kummer (CHITTA, 1995).

**Definição 6.1.1 (Distribuição de Kummer)** *Uma variável aleatória  $X$  tem uma distribuição de Kummer ( $X \sim K[\alpha, \beta, \gamma, \delta]$ ) com parâmetros  $\alpha, \beta, \gamma$  e  $\delta$  ( $\alpha; \beta; \gamma e \delta > 0$ ) se é contínua e se a função densidade de probabilidade é dada por:*

$$f(x|\alpha, \beta, \gamma, \delta) = C \frac{x^{\alpha-1} \exp(-\beta x)}{(1+\delta x)^\gamma} I_{(0, \infty)}^{(x)}, \quad (6.8)$$

onde  $C$ , a constante de padronização, é dada da forma:

$$C^{-1} = \frac{\Gamma(\alpha)}{\delta^\alpha} U\left(\alpha, \alpha + 1 - \gamma, \frac{\beta}{\delta}\right) \quad (6.9)$$

e  $U(a, b, z)$  é a função de Kummer com representação integral dada por

$$\Gamma(a) U(a, b, z) = \int_0^\infty \exp(-zt) t^{a-1} (1+t)^{b-a-t} dt. \quad (6.10)$$

A função de Kummer pertence à classe das funções hipergeométricas confluentes (ABRAMOWITZ; STEGUN, 1972, Cap.13) e está implementada em diversos softwares matemáticos como, por exemplo, no pacote fOptions do R (R Development Core Team, 2006) ou Maple<sup>®</sup> (MONAGAN et al., 2005).

Esta função não é definida para  $z = 0$ , mas, Chitta (apud ARMERO; BAYARRI, 1993) define que:

$$U(a, b, 0) = \int_0^\infty t^{a-1} (1+t)^{b-a-1} dt \quad (6.11)$$

tal que 6.11 converge para  $b < 1$ , e resultando em

$$U(a, b, 0) = \frac{\Gamma(1-b)}{\Gamma(a+1-b)}, \text{ para } b < 1. \quad (6.12)$$

Ainda, Chitta (1995) faz um paralelo entre a distribuição de Kummer com a distribuição gama, uma vez que ambas são derivadas de funções matemáticas básicas. Desta maneira, a distribuição gama é um caso especial da distribuição de Kummer quando  $\gamma = 0$  e quando  $\gamma > \alpha$  e  $X \sim K(\alpha, \beta = 0, \gamma, \delta)$ ,  $\frac{(\gamma - \alpha)\delta}{\alpha} X \sim F(2\alpha, 2\gamma - 2\alpha)$ . Esta relação é interessante computacionalmente para efeitos de geração de número aleatórios da distribuição de Kummer a partir de uma distribuição F, que está implementada em quase todos os softwares estatísticos. Pode ser visto mais sobre a função de Kummer em Abramowitz & Stegun (1972, Cap. 13).

### 6.1.1 Momentos da distribuição de Kummer

Os momentos da distribuição  $X \sim K(\alpha, \beta, \gamma, \delta)$  são obtidos em função de  $U$ , tal que

$$E(X^k) = \frac{\Gamma(\alpha + k) U(\alpha + k, \alpha + k + 1 - \gamma, \frac{\beta}{\delta})}{\delta^k \gamma \alpha U(\alpha, \alpha + 1 - \gamma, \frac{\beta}{\delta})}. \quad (6.13)$$

Assim, por exemplo,

$$E(X) = \frac{\Gamma(\alpha + 1) U(\alpha + 1, \alpha + 2 - \gamma, \frac{\beta}{\delta})}{\delta \gamma \alpha U(\alpha, \alpha + 1 - \gamma, \frac{\beta}{\delta})}.$$

### 6.1.2 Ilustração

Aqui apresentaremos alguns gráficos da distribuição de Kummer para diferentes valores dos parâmetros  $\alpha, \beta, \gamma$  e  $\delta$ . Os gráficos da figura 11 representam algumas dessas curvas, variando-se os parâmetros.

## 6.2 Reparametrização do modelo

*A reparametrização é utilizada usualmente para simplificar as distribuições de probabilidade subjacentes aos modelos, de modo a facilitar a implementação das técnicas de estimação MCMC, melhorando desta forma a convergência* (PAULINO et al., 2003).

Nesta dissertação a reparametrização será utilizada a fim de se encontrar a distribuição à priori do parâmetro  $\rho$  que modula a assimetria das normais assimétricas, estendendo o

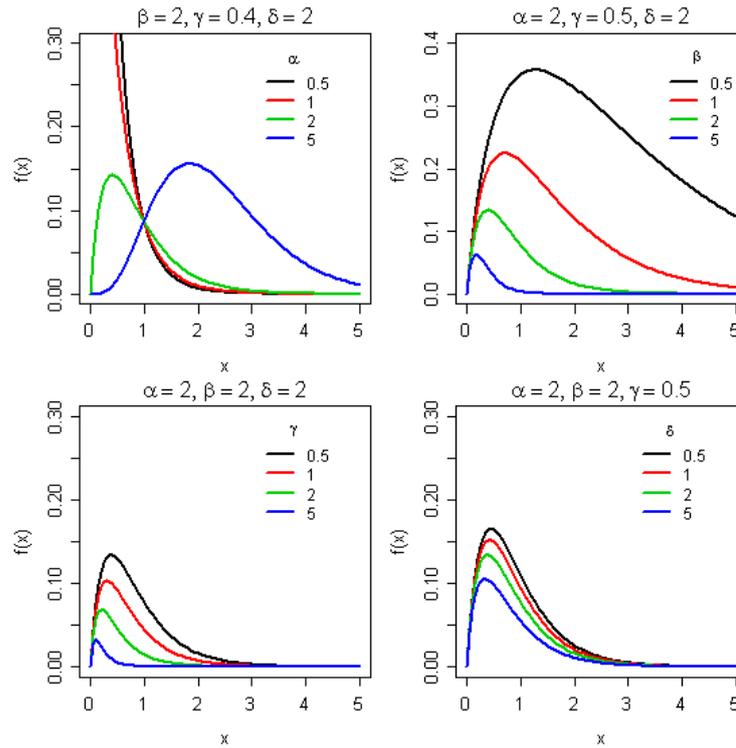


Figura 11: Gráficos com algumas curvas definidas pela distribuição de Kummer, variando-se os parâmetros.

trabalho de Rodríguez (2005) para a distribuição assimétrica sob a forma de Sahu & Dey . Sem a reparametrização, a busca pela distribuição de  $\rho$  é muito difícil, resultando em uma integral que não pode ser resolvida analiticamente de maneira trivial, impossibilitando assim, encontrar uma forma fechada para a mesma.

Assim, a partir da expressão (5.11), temos que

$$\rho|\sigma_{\eta}^2, \sigma_{\nu}^2 \sim t \left[ 0, \frac{\pi^2 \sigma_{\nu}^2}{4\sigma_{\eta}^2}, \frac{1}{2} \right], \quad (6.14)$$

onde  $\frac{\pi^2 \sigma_{\nu}^2}{4\sigma_{\eta}^2}$  representa o parâmetro de escala da distribuição t-Student. Assim, definamos

$$\tau = \frac{4\sigma_{\eta}^2}{\pi^2 \sigma_{\nu}^2} \quad (6.15)$$

como a “precisão” da distribuição t-Student, em analogia com a precisão da distribuição normal onde o parâmetro de escala é também a variância, e seu inverso, a precisão propriamente dita.

É usual atribuir-se uma priori gama para a precisão de uma distribuição, de modo que definiremos, por analogia,

$$\tau \sim \text{gama} \left( \frac{a}{2}, \frac{b}{2} \right). \quad (6.16)$$

Agora, a distribuição de  $\rho|\tau \sim t \left( 0, \frac{1}{\tau}, \frac{1}{2} \right)$  é definida como:

$$f(\rho|\tau) = \frac{\overbrace{\sqrt{2}}^{c_1}}{B\left(\frac{1}{4}, \frac{1}{2}\right)} \sqrt{\tau} \left[ 1 + \left(\frac{1}{2}\right)^{-1} \tau \rho^2 \right]^{-\frac{3}{4}} I_{(-\infty, \infty)}^{(\rho)}, \quad (6.17)$$

onde  $B(a,b)$  é a função beta, definida como

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, a > 0, b > 0. \quad (6.18)$$

A priori de  $\tau$  (6.16) é dada por:

$$f(\tau) = \frac{\overbrace{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{c_2}}{\Gamma\left(\frac{a}{2}\right)} \tau^{\frac{a}{2}-1} \exp\left(-\frac{b}{2}\tau\right) I_{(0, \infty)}^{(\tau)}. \quad (6.19)$$

O teorema seguinte trata da distribuição conjunta do vetor aleatório  $(\rho, \tau)$ .

**Definição 6.2.1 (Distribuição conjunta)** *Considere o vetor aleatório  $(\rho, \tau)$ , tal que*

$$\rho|\tau \sim t\left(0, \frac{1}{\tau}, \frac{1}{2}\right), e \quad (6.20)$$

$$\tau \sim \text{gama}\left(\frac{a}{2}, \frac{b}{2}\right). \quad (6.21)$$

Então, a distribuição conjunta  $f(\rho, \tau)$ , dada pelo produto da densidade condicional de  $\rho|\tau$  pela densidade gama de  $\tau$  é definida por:

$$f(\rho, \tau) = c_1 c_2 \tau^{\frac{a+1}{2}-1} \exp\left(-\frac{b}{2}\tau\right) \left[ 1 + \left(\frac{1}{2}\right)^{-1} \tau \rho^2 \right]^{-\frac{3}{4}} I_{(-\infty, \infty)}^{(\rho)} I_{(0, \infty)}^{(\tau)}, \quad (6.22)$$

onde  $c_1 = \frac{\sqrt{2}}{B\left(\frac{1}{4}, \frac{1}{2}\right)}$ ,  $c_2 = \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)}$ ,  $B(*)$  é a função beta, definida em (6.18),  $a < \frac{1}{2}$  e  $b \in \mathfrak{R}^+$ .

A f.d.p. dada pela definição 6.2.1 serve como ponto de partida para se encontrar a

distribuição condicional de  $\tau$  dado  $\rho$ , como apresentado no teorema 6.2.2.

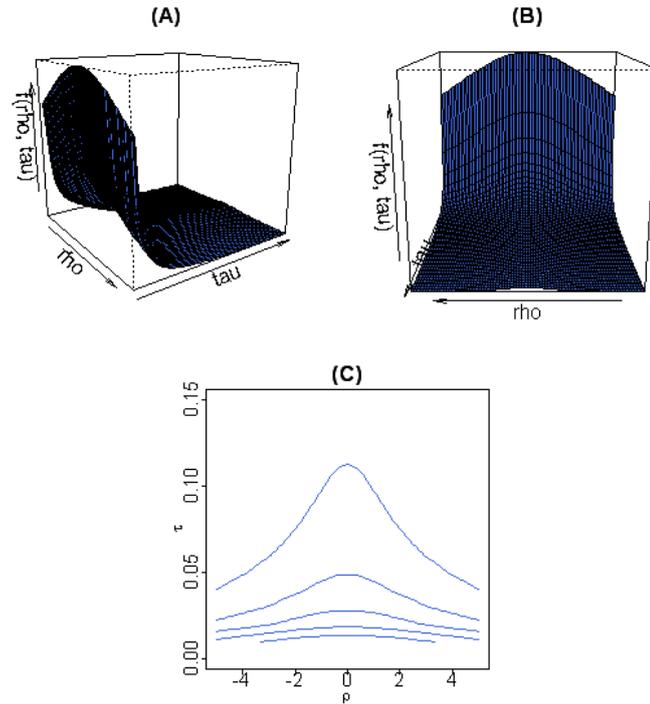


Figura 12: Distribuição conjunta de  $f\left(\frac{2}{5}, \frac{1}{8}, 3\right)$ . (A) e (B): Superfície em ângulos distintos; (C): Curvas de nível.

**Teorema 6.2.2** *Se o vetor aleatório  $(\rho, \tau)$  apresenta distribuição conjunta dada por (6.22) com parâmetros  $a$  e  $b$ , então a distribuição condicional de  $\tau | \rho$  é a distribuição de Kummer (definição 6.1.1), com os parâmetros  $\frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}$  e  $2\rho^2$ .*

$$\tau | \rho \sim K\left(\frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}, 2\rho^2\right), \quad (6.23)$$

com  $\tau, a, b > 0$  e  $\rho \neq 0$ .

**Proof.** A demonstração do teorema 6.2.2 é imediata, bastando encontrar a condicional

completa de  $f(\rho, \tau)$ .

$$\begin{aligned}
f(\tau, \rho) &= c_1 c_2 \tau^{\frac{a}{2} + \frac{1}{2} - 1} \exp\left(-\frac{b}{2}\tau\right) \left[1 + \left(\frac{1}{2}\right)^{-1} \tau \rho^2\right]^{-\frac{3}{4}} \\
&= \frac{c_1 c_2 \tau^{\frac{a}{2} + \frac{1}{2} - 1} \exp\left(-\frac{b}{2}\tau\right)}{(1 + 2\tau \rho^2)^{\frac{3}{4}}} \\
&= \frac{c_1 c_2}{C} \underbrace{\tau^{\frac{a}{2} + \frac{1}{2} - 1} \exp\left(-\frac{b}{2}\tau\right) I_{(0, \infty)}^{(\tau)}}_{K\left(\tau, \frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}, 2\rho^2\right)} \\
&\sim K\left(\frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}, 2\rho^2\right). \tag{6.24}
\end{aligned}$$

onde  $C^{-1}$  é definido como em (6.9),  $c_1$  e  $c_2$  apresentados na definição 6.2.1. ■

É interessante o resultado do teorema 6.2.2, pois a distribuição da “precisão” de uma t-Student que é ponderada por uma distribuição gama resulta na distribuição de Kummer com os parâmetros  $\alpha = \frac{a+1}{2}, \beta = \frac{b}{2}, \gamma = \frac{3}{4}$  e  $\delta = 2\rho^2$ . Assim, a implementação do algoritmo de Gibbs pode ser feita, desde que o software onde ela for executada contenha a função de Kummer também implementada, e ainda, é formular prioris diretamente para o parâmetro  $\tau$ , sendo elas informativas ou não. Vale notar também que o valor atribuído ao hiperparâmetro  $\frac{a}{2}$  da distribuição à priori gama é restrito, pois será somente neste intervalo que a função de Kummer será limitada. O teorema 6.2.3 apresenta uma inédita distribuição a priori para  $\rho$  que é conectada com os resultados obtidos por Rodríguez (2005) para prioris para o parâmetro de assimetria de uma normal assimétrica, mas agora estendidos para a representação daquela distribuição segundo Sahu & Dey.

**Teorema 6.2.3 (Priori T-Gama)** *Se o vetor aleatório  $(\rho, \tau)$  apresenta distribuição conjunta (expressão 6.22) com parâmetros  $a$  e  $b$ , então a distribuição marginal de  $\rho$  [Priori T-Gama, representada por  $TG(a, b)$ ] é dada por*

$$TG(a, b) = \begin{cases} \frac{\sqrt{2\pi}}{B\left(\frac{1}{4}, \frac{1}{2}\right)B\left(\frac{1}{2}, \frac{a}{2}\right)} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{U\left(\frac{a+1}{2}, \frac{a}{2} + \frac{3}{4}, \frac{b}{4\rho^2}\right)}{(2\rho^2)^{\frac{a+1}{2}}} & \rho \neq 0 \\ \frac{\sqrt{\pi}}{\sqrt{\frac{b}{4}}B\left(\frac{1}{4}, \frac{1}{2}\right)B\left(\frac{a}{2}, \frac{1}{2}\right)} & \rho = 0 \end{cases}, \tag{6.25}$$

onde  $C^{-1}$  é dado por (6.9),  $c_1$  e  $c_2$  são as constantes de normalização das distribuições

em (6.17) e (6.19), respectivamente,  $U(*)$  é a função de Kummer (6.10) e  $B(*)$  é a função beta (6.18),  $a < \frac{1}{2}$  e  $b > 0$ .

**Proof.** Para se encontrar a f.d.p de  $TG(a,b)$ , para  $\rho \neq 0$  basta resolver a integral abaixo, com  $a < \frac{1}{2}$  e  $b > 0$ .

$$\begin{aligned}
TG(a,b) &= \int_0^\infty f(\rho,\tau) d\tau \\
&= \int_0^\infty c_1 c_2 \tau^{\frac{a+1}{2}-1} \exp\left(-\frac{b}{2}\tau\right) \left[1 + \left(\frac{1}{2}\right)^{-1} \tau \rho^2\right]^{-\frac{3}{4}} d\tau \\
&= \frac{c_1 c_2}{C} \int_0^\infty \overbrace{C \frac{\tau^{\frac{a+1}{2}-1} \exp\left(-\frac{b}{2}\tau\right)}{[1 + 2\tau\rho^2]^{\frac{3}{4}}}}^{K\left(\tau, \frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}, 2\rho^2\right)} d\tau \\
&= \frac{c_1 c_2}{C} \times 1 \\
&= \frac{\Gamma\left(\frac{1}{2}\right) \sqrt{2}}{B\left(\frac{1}{4}, \frac{1}{2}\right)} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{\Gamma\left(\frac{a+1}{2}\right) U\left(\frac{a+1}{2}, \frac{a}{2} + \frac{3}{4}, \frac{b}{4\rho^2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{a}{2}\right) (2\rho^2)^{\frac{a+1}{2}}} \\
&= \frac{\sqrt{2\pi}}{B\left(\frac{1}{4}, \frac{1}{2}\right) B\left(\frac{1}{2}, \frac{a}{2}\right)} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{U\left(\frac{a+1}{2}, \frac{a}{2} + \frac{3}{4}, \frac{b}{4\rho^2}\right)}{(2\rho^2)^{\frac{a+1}{2}}} I_{(-\infty, 0)(0, \infty)}^{(\rho)}, \quad (6.26)
\end{aligned}$$

onde  $C^{-1}$ ,  $c_1$  e  $c_2$ ,  $U(*)$  e  $B(*)$  estão definidos no enunciado do teorema 6.2.3, com  $0 < a < \frac{1}{2}$ ,  $b > 0$  e  $\rho \neq 0$ .

Para  $\rho = 0$ , a demonstração é trivial, simplesmente fazendo  $\rho = 0$  e integrando em relação a  $\tau$ , mantendo-se as restrições definidas para o caso  $\rho \neq 0$ . A expressão a seguir ilustra este processo.

$$\int_0^\infty f(\tau, \rho = 0) d\tau = \frac{\sqrt{\pi}}{\sqrt{\frac{b}{4}} B\left(\frac{1}{4}, \frac{1}{2}\right) B\left(\frac{a}{2}, \frac{1}{2}\right)}. \quad (6.27)$$

■

A priori T-Gama (6.26) pode ser interpretada como a média de uma distribuição t-Student com a “precisão” ponderada pela distribuição gama, e ela é flexível, sendo importante esta característica no sentido de que a busca por prioris para este parâmetro fica facilitada, além de ser uma **inovação obter uma priori fechada**, mesmo que de-

pendendo de uma função matemática não muito bem comportada computacionalmente. A figura 13 apresenta gráficos com variações nos parâmetros da priori de  $\tau$ , de modo a verificar o comportamento da distribuição de  $\rho$ . Um problema computacional surge, pois os softwares aproximam o valor daquela função, tornando os pontos na vizinhança de 0 instáveis e por isso, apresentando comportamentos anômalos nos gráficos, mas sabemos que tendem assintoticamente para o valor da expressão em (6.27).

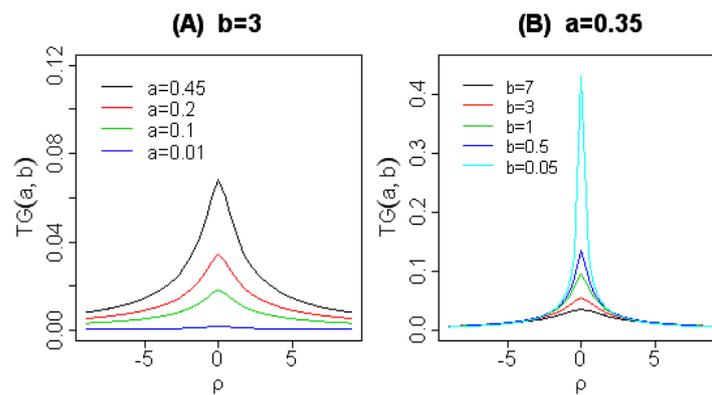


Figura 13: Formas da distribuição de  $\rho$  em função dos hiperparâmetros da priori gama para  $\tau$ . (A): variando-se  $a$  com  $b$  fixo; (B): variando-se  $b$  com  $a$  fixo.

Há uma restrição para o valor do hiperparâmetro  $a$  da priori para  $\tau$  ( $0 < a < \frac{1}{2}$ ), de modo que a escolha desse valor não pode ser arbitrária em toda reta real positiva. As curvas definidas pela priori T-Gama têm o aspecto de sino, característico das distribuições normais e t. Quando se aumenta o valor do hiperparâmetro  $a$ , no máximo a  $\frac{1}{2}$ , e fixando-se  $b$ , a priori se torna mais informativa (figura 13, (A)). No caso de termos  $a$  fixado e variando-se  $b$ , o contrário ocorre, quanto maior o valor desse hiperparâmetro, mais informativa fica a distribuição à priori (figura 13, (B)). Em situações onde não se tem conhecimento prévio algum sobre  $\rho$  e se deseja expressar vagueza pela ausência de informação à priori, é natural “espalhar” a massa de probabilidade em todo espaço paramétrico de  $\rho$  e “deixar que os dados” indiquem a correta localização do parâmetro de modulação da assimetria. Assim, a priori T-Gama parece acomodar tanto os casos em que haja informação à priori quanto aqueles casos em que há incerteza nesta informação prévia.

### 6.2.1 Caracterizações da Priori T-Gama

A distribuição à priori T-Gama é o resultado da marginalização da função distribuição conjunta t-gama, como já fora mostrado na definição 6.2.1. Assim, ela é um resultado indireto da mistura de uma distribuição t com graus de liberdade fracionário  $\left(\frac{1}{2}\right)$  e de uma distribuição gama.

Um importante teorema da Teoria das probabilidades (ROSS, 1976), mostra que, considerando  $X$  e  $Y$  variáveis aleatórias, então

$$E [E (X | Y)] = E (X). \quad (6.28)$$

Assim, sabemos que  $\rho | \tau \sim t_{\frac{1}{2}} \left(0, \frac{1}{\tau}\right)$  e também que somente existe esperança matemática para a função t-Student quando o número de graus de liberdade for maior que um. Feitas estas considerações, temos

$$E [E (\rho | \tau)] = E (\rho),$$

mas  $E (\rho | \tau)$  é a esperança de uma t-Student com  $\frac{1}{2}$  grau de liberdade, que não tem, portanto, momentos definidos; logo, não existe média nem variância para a distribuição à priori T-Gama, por esta parametrização.

Por outro lado, se utilizássemos a indução da t-Student pela distribuição uniforme definida entre  $-1$  e  $1$ , (RODRÍGUEZ, 2005), a distribuição à priori para o parâmetro de assimetria  $\lambda \sim t_2 \left(0, \frac{1}{2}\right)$ , de forma que esta distribuição tem média e variância, o que implicaria em encontrar, também a média e a variância induzida para  $\rho$ , da mesma maneira que foi utilizada para se chegar à priori T-Gama. Para evitar repetições, apresentamos a seguir somente o resultados, no seguinte teorema.

**Teorema 6.2.4** *Se a distribuição à priori para o parâmetro de assimetria de uma distribuição normal assimétrica for induzida por uma distribuição t-Student com parâmetro de locação 0, parâmetro de escala  $\frac{1}{2}$  e 2 graus de liberdade, então:*

1. A distribuição  $\rho | \sigma_{\eta}^2, \sigma_{\nu}^2 \sim t_2 \left(0, \frac{\sigma_{\nu}^2}{2\sigma_{\eta}^2}\right)$ , tal que podemos reparametrizar o parâmetro

de escala desta distribuição *t*-Student como um parâmetro de “precisão” como

$$\tau = \frac{2\sigma_\eta^2}{\sigma_\nu^2}, \quad (6.29)$$

de forma que  $\rho|\tau \sim t_2(0, \frac{1}{\tau})$  e a priori para  $\tau \sim \text{gama}(\frac{a}{2}, \frac{b}{2})$ ;

2. A distribuição conjunta é dada, sob esta reparametrização, como:

$$f(\tau, \rho) = \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{a}{2})\sqrt{2}} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{\tau^{\frac{a+1}{2}-1} \exp(-\frac{b}{2}\tau)}{\left[1 + \frac{\rho^2}{2}\tau\right]^{\frac{3}{2}}} I_{(0,\infty)}^{(\tau)} I_{(-\infty,0)(0,\infty)}^{(\rho)} \quad (6.30)$$

com  $a < 2, b > 0$ ;

3. A distribuição condicional  $\tau|\rho \propto K\left(\frac{a+1}{2}, \frac{b}{2}, \frac{3}{2}, \frac{\rho^2}{2}\right)$  e por fim,

4. A priori T-Gama, portanto, é dada por:

$$TG(a, b) = \sqrt{\frac{\pi}{2}} \frac{1}{B(1, \frac{1}{2}) B(\frac{a}{2}, \frac{1}{2})} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{U\left(\frac{a+1}{2}, \frac{a}{2}, \frac{b}{\rho^2}\right)}{\left(\frac{\rho^2}{2}\right)^{\frac{a+1}{2}}} I_{(-\infty,0)(0,\infty)}^{(\rho)} \quad (6.31)$$

com  $a < 2, b > 0$ .

**Proof.** As demonstrações serão omitidas, mas são semelhantes às exibidas na seção 6.2.

■

Agora, é possível encontrar a média da distribuição T-Gama dada em (6.31). Conhecendo a esperança da esperança condicional de uma variável aleatória (expressão 6.28), temos:

$$E[E(\rho|\tau)] = 0,$$

uma vez que  $\rho|\tau \sim t_2(0, \frac{1}{\tau})$ . Outra relação entre distribuições condicionais é dada pela relação (ROSS, 1976):

$$Var(\rho) = E[Var(\rho|\tau)] + Var[E(\rho|\tau)]. \quad (6.32)$$

A  $Var(\rho|\tau) = \frac{2}{2-2} = \infty$ , uma vez que a variância de uma distribuição t-student é dada por  $\frac{gl}{gl-2}$ , onde  $gl$  é o número de graus de liberdade, que em nosso caso é 2. Logo, toda a variância de  $\rho$  calculada pela expressão (6.32) vai a infinito, de forma que  $Var(\rho) = \infty$ .

Para ilustrar as formas de ambas prioris, fixou-se que  $a = \frac{1}{5}$  e  $b = \frac{1}{3}$  e o gráfico de ambas as distribuições foram construídos juntas e estão na figura 14.

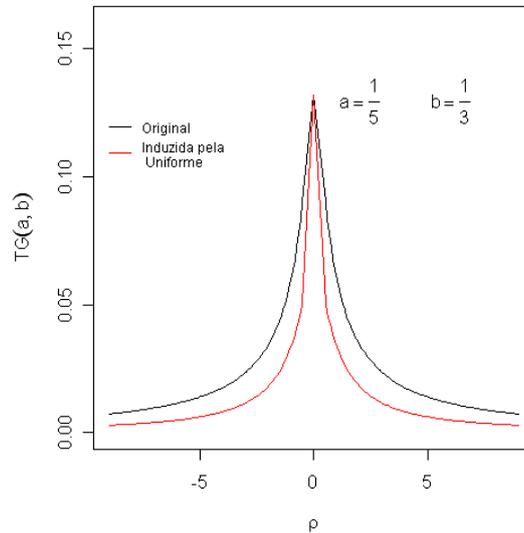


Figura 14: Distribuições à T-Gama com as parametrizações originais e segundo a indução pela uniforme.

### 6.3 Aplicação

O conjunto de dados analisados é referente à tabela apresentada no apêndice D. Os dados<sup>1</sup> foram obtidos de Sef (1999) e são referentes às medidas de placa bacteriana antes e após a utilização de uma escova dental convencional e outra, de baixo custo, denominada de monobloco (figura 15), associada ao uso de dentifrício. A amostra em análise é composta de 32 crianças entre 4 e 6 anos. Os indivíduos que participavam do estudo foram divididos em 4 grupos com 8 indivíduos em cada, de forma que todos utilizassem os dois tipos de escova com ou sem dentifrício. Foi dado, também, o tempo de uma semana entre a aplicação dos tratamentos nos grupos para se evitar quaisquer efeitos residuais.

A análise descritiva inicial dos dados indica que existe uma relação aparentemente linear entre a medida pré-teste e pós-teste, como evidencia a figura 16.

<sup>1</sup>O conjunto de dados está disponível em <http://www.ime.usp.br/~jmsinger/Dados/Sef1999.xls>.



Figura 15: Escovas dentais do tipo monobloco.

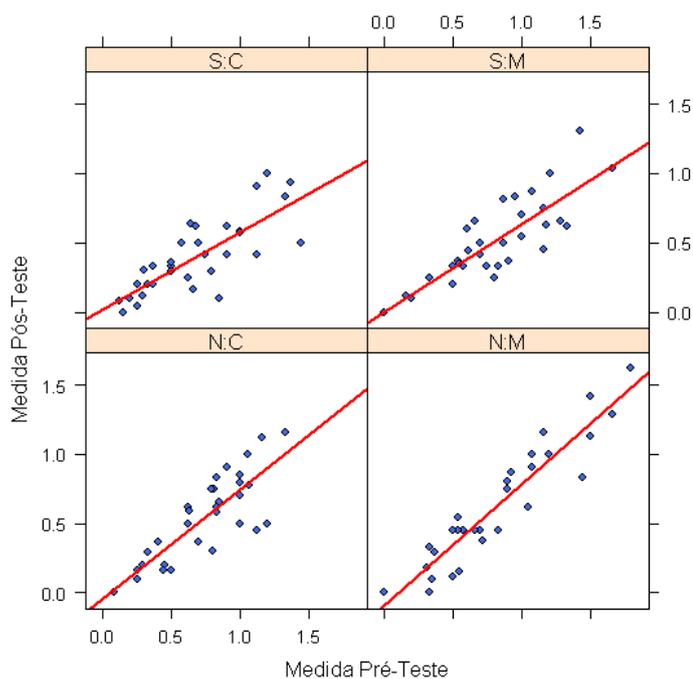


Figura 16: Diagramas de dispersão entre as medidas pré-teste/pós-teste. **S:C**: Dentifricio e escova convencional; **N:C**: Sem dentifricio e escova convencional; **S:M**: Dentifricio e escova monobloco; **N:M**: Sem dentifricio e escova monobloco.

Quando se verifica os histogramas das diferenças entre as medidas pós e pré-testes, é clara a evidência de que há assimetrias em todas as combinações de tratamentos, como mostra a figura 17.

É interessante observar que uma medida geral de assimetria resulta em  $-0,973$  para a diferença entre as medidas pós e pré-testes. Na tabela 4 foram calculadas as assimetrias das diferenças em função da combinação dos tratamentos.

Nesta tabela 4 se observa que existe assimetria para quaisquer combinações de tratamentos, confirmando o que a figura 17 indica. As curvas individuais por tratamentos é

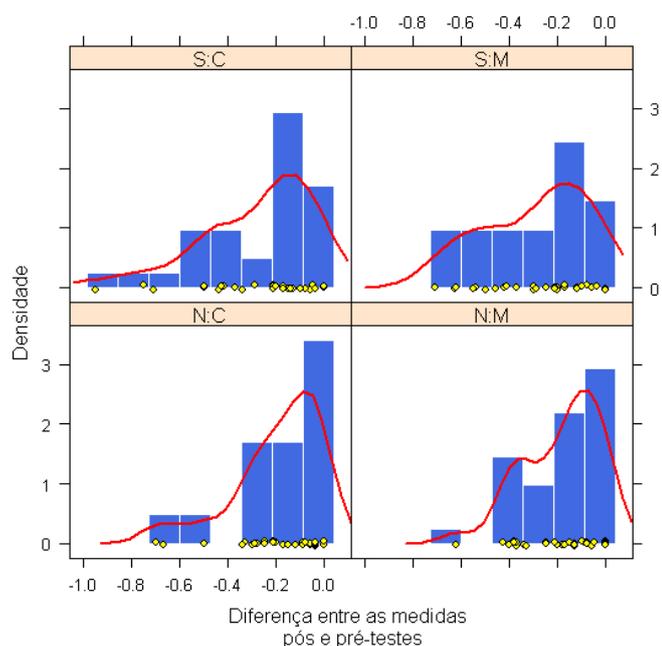


Figura 17: Histograma das diferenças entre as medidas pós e pré-testes em função da combinação de tratamentos. **S:C**: Dentifrício e escova convencional; **N:C**: Sem dentifrício e escova convencional; **S:M**: Dentifrício e escova monobloco; **N:M**: Sem dentifrício e escova monobloco.

Tabela 4: Momentos de assimetria para a diferença entre as medidas pós e pré-testes em função da combinação dos tratamentos

| Momentos de Assimetria |       |        |        |
|------------------------|-------|--------|--------|
| N:C                    | S:C   | N:M    | S:M    |
| -1,198                 | 0,987 | -0,718 | -0.406 |

indicada pela figura 18. Observe que há redução maior quando se utiliza dentifrício (S:C e S:M), aparentemente independente do tipo de escova utilizada.

Sef (1999), utilizando este mesmo conjunto de dados para a formulação de um modelo multiplicativo, chamado aqui nesta dissertação de Sef & Singer, faz inúmeras suposições (já citadas na seção 3.1), inclusive a de normalidade dos resíduos. No entanto, a evidência inicial é de que esta condição não seja plausível, mesmo após a transformação logarítmica sugerida. Assim, utilizaremos os conceitos já estabelecidos para o modelo de Sef & Singer sob a óptica bayesiana definida neste capítulo para estimar os parâmetros, com a suposição de que os erros e efeitos aleatórios podem ser expressos, de maneira única, por uma distribuição normal assimétrica utilizando a parametrização de Sahu & Dey.

Considerando a notação expressa para o modelo multiplicativo de Sef & Singer na

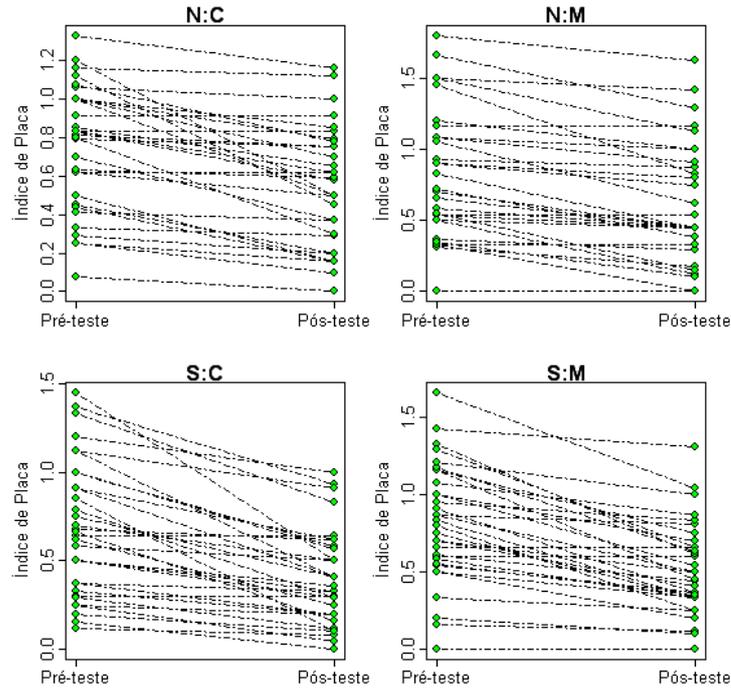


Figura 18: Curvas individuais para os valores pré e pós-testes. **S:C**: Dentifrício e escova convencional; **N:C**: Sem dentifrício e escova convencional; **S:M**: Dentifrício e escova monobloco; **N:M**: Sem dentifrício e escova monobloco.

definição (3.1.1) e as distribuições atribuídas para o modelo bayesiano em (6.1), será, então, construído o modelo para o conjunto de dados utilizados neste estudo.

Considere o modelo multiplicativo

$$y_{ijk} = \beta_{ij} x_{ijk}^{\gamma_{ij}} \varepsilon_{ijk}, \quad i = 1, 2, \quad j = 1, 2 \quad \text{e} \quad k = 1, \dots, n, \quad (6.33)$$

tal que  $y_{ijk}(x_{ijk})$  representa a medida pós (pré) - teste do  $k$ -ésimo indivíduo utilizando ou não dentifrício ( $i = 1$  para a não utilização e  $i = 2$ , caso contrário) e utilizando-se do  $j$ -ésimo tipo de escova ( $j = 1$ , escova convencional e  $j = 2$ , escova monobloco). O coeficiente de uniformidade para o uso de dentifrício  $i$  e escova do tipo  $j$  é representado por  $\gamma_{ij}$ , e  $\beta_{ij}$  representa a placa bacteriana residual.

Aplicando-se a transformação logarítmica, obteremos resultados semelhantes às expressões em (4.21) (haverá necessidade de remover aqueles elementos da amostra que possuem valores nulas em medidas pré ou pós-testes, por causa do tipo de transformação utilizada), associando-as com as distribuições em (6.1), teremos, de maneira formal,

lembrando que  $\mathbf{W}_k^*$  dada por (4.7):

$$\begin{aligned}\mathbf{Z}_k &= \mathbf{W}_k^* \boldsymbol{\gamma}^* + (\sigma_\eta^2 \rho |\eta_k| + \sigma_\nu^2 \nu_{lk}^*) \mathbf{I}_4 \\ \eta_k &\sim N(0,1) \\ \nu_{lk}^* &\sim N(0,1) \\ \nu_{lk}^* &\perp \eta_k,\end{aligned}\tag{6.34}$$

para ,  $k = 1, \dots, n$  e  $l = 1, \dots, 4$ .

A incerteza quanto aos parâmetros do modelo, fixos no espaço paramétrico, pode ser representada pelas distribuições a priori. Aqui escolhemos prioris não-informativas para modelos 1 e 2; para o modelo 3 supôs-se que  $\rho = 0$ , ou seja, os dados foram tratados, como análises de covariâncias com erro normal. Assim, a precisão (inverso dos parâmetros de escala) são distribuídas a priori como distribuições gama com hiperparâmetros tendendo a zero para os modelos 1 e 2 e os coeficientes do modelo, dados pelo vetor  $\boldsymbol{\gamma}^*$ , por distribuições normais vagas. As principais diferenças ficam por conta das prioris para o parâmetro de modulação da assimetria ( $\rho$ ):

- No **modelo 1**, por uma  $t_{\frac{1}{2}}\left(0, \frac{(\pi\sigma_\nu)^2}{4\sigma_\eta^2}\right)$ , segundo a indução pelo resultado de Rodríguez (2005) para a priori do parâmetro de assimetria  $\lambda$ ; este pode ser chamado de modelo com prioris usuais ou canônicas. Assim:

$$\begin{aligned}\rho | (\sigma_\eta^2, \sigma_\nu^2) &\sim t\left[0, \frac{(\pi\sigma_\nu)^2}{4\sigma_\eta^2}, \frac{1}{2}\right], \\ \frac{1}{\sigma_\nu^2} &\sim \text{Gama}\left(\frac{a}{2}, \frac{b}{2}\right), \\ \frac{1}{\sigma_\eta^2} &\sim \text{Gama}\left(\frac{a}{2}, \frac{b}{2}\right), \\ \boldsymbol{\gamma}^* &\sim N_8(0, \Sigma).\end{aligned}$$

- no **modelo 2** utilizou a reparametrização da “precisão” da distribuição t, de modo que  $\rho$  utiliza a distribuição priori T-Gama não-informativa (hiperparâmetros da distribuição gama atribuída à priori para  $\tau$  com valores pequenos). As distribuições

à priori, foram, portanto:

$$\begin{aligned} \tau &= \frac{4\sigma_\eta^2}{\pi^2\sigma_\nu^2} \\ TG(a,b) &\Rightarrow \begin{cases} \rho | \tau \sim t \left[ 0, \frac{1}{\tau}, \frac{1}{2} \right], \\ \tau \sim Gama \left( \frac{a}{2}, \frac{b}{2} \right), \end{cases} \\ \frac{1}{\sigma_\eta^2} &\sim Gama \left( \frac{a}{2}, \frac{b}{2} \right), \\ \gamma^* &\sim N_8(0, \Sigma). \end{aligned}$$

- no **modelo 3** não foi necessária priori para  $\rho$ , uma vez que se fixou seu valor em 0 e, finalmente. Logo,

$$\begin{aligned} \rho &= 0, \\ \frac{1}{\sigma^2} &\sim Gama \left( \frac{a}{2}, \frac{b}{2} \right), \\ \gamma^* &\sim N_8(0, \Sigma). \end{aligned}$$

O método MCMC utilizado via WinBUGS pode ser problemático quando os parâmetros do modelo não são ortogonais (ver manual do programa) e além disso, não está definido um gerador de variáveis aleatórias para a distribuição t-Student com graus de liberdade fracionários, de modo que aquela distribuição pode ser gerada de duas formas: como uma mistura de normal com gama (6.7) ou via algoritmo Metropolis-Hastings, que é baseado na implementação do núcleo da distribuição t-Student com os graus de liberdade fracionários. Optamos pela segunda forma, pois a convergência se dá de maneira mais estável.

A reparametrização da distribuição condicional de  $\rho$  dados  $\sigma_\eta^2$  e  $\sigma_\nu^2$  (seção 6.2), utilizada nos modelos 2 e 4 está baseada em tornar a “precisão” da distribuição t-Student ponderada pela gama, o que resulta na distribuição conjunta t-gama, cuja marginal de  $\rho$  é a priori T-Gama (6.2.3). Assim, ela será construída da seguinte maneira:

$$\begin{aligned}
\tau &= \frac{4\sigma_\eta^2}{\pi^2\sigma_\nu^2} \\
\rho|\tau &\sim t\left[0, \frac{1}{\tau}, \frac{1}{2}\right], \\
\tau &\sim \text{Gama}\left(\frac{a}{2}, \frac{b}{2}\right), \\
\frac{1}{\sigma_\nu^2} &\sim \text{Gama}\left(\frac{a}{2}, \frac{b}{2}\right), \\
\gamma^* &\sim N_8(0, \Sigma).
\end{aligned} \tag{6.35}$$

O que se espera com esta reparametrização é facilitar a otimização dos parâmetros quando a implementação do algoritmo MCMC for feita, de modo a obter uma relativa estabilidade nas iterações e a convergência mais rápida, o que implica em utilização de cadeias menores e assim, economia de recursos computacionais.

A programação utilizada está disponível no apêndice E. Mais informações sobre simulação MCMC, consultar Gamerman & Lopes (2006).

### 6.3.1 Resultados

Todos os modelos foram construídos considerando a geração de duas cadeias longas com valores iniciais distintos (cerca de 100.000 valores gerados para cada uma delas) e um período de “burn-in” também alto, (cerca de 25.000). Estas cadeias longas são necessárias, pois a correlação entre os valores gerados são altos e para minimizá-los, foram tomados os valores espaçados numa distância de 40 elementos entre cada geração, formando uma amostra efetiva em torno de 3200 elementos em cada cadeia, resultando, ao total, após a convergência e diminuição do efeito de correlação, 6400 amostras efetivas das mais de 200000 geradas em todo o processo para cada modelo.

Tabela 5: Resultados da estimação dos parâmetros para o **Modelo 1**.

| <b>Modelo 1</b>   |   |   |              |          |         |            |        |        |
|-------------------|---|---|--------------|----------|---------|------------|--------|--------|
| Parâmetros        | E | P | Estatísticas |          |         | Int. Pred. |        |        |
|                   |   |   | Média        | D.P.     | Erro MC | 2,5%       | 50%    | 97,5%  |
| $\alpha_{11}$     | C | N | -0,037       | 0,345    | 0,007   | -0,950     | -0,308 | 0,095  |
| $\alpha_{12}$     | M | N | -0,337       | 0,346    | 0,007   | -0,896     | -0,269 | 0,124  |
| $\alpha_{21}$     | C | S | -0,657       | 0,354    | 0,007   | -1,235     | -0,564 | -0,173 |
| $\alpha_{22}$     | M | S | -0,515       | 0,348    | 0,007   | -1,073     | -0,438 | -0,033 |
| $\gamma_{11}$     | C | N | 1,084        | 0,143    | 0,003   | 0,810      | 1,079  | 1,376  |
| $\gamma_{12}$     | M | N | 1,084        | 0,138    | 0,003   | 0,828      | 1,080  | 1,366  |
| $\gamma_{21}$     | C | S | 0,932        | 0,106    | 0,002   | 0,727      | 0,930  | 1,147  |
| $\gamma_{22}$     | M | S | 0,983        | 0,127    | 0,003   | 0,734      | 0,984  | 1,234  |
| $\alpha_{11}^*$   | C | N | -0,384       | 0,098    | 0,002   | -0,577     | -0,385 | -0,192 |
| $\alpha_{12}^*$   | M | N | -0,346       | 0,089    | 0,002   | -0,519     | -0,349 | -0,170 |
| $\alpha_{21}^*$   | C | S | -0,666       | 0,095    | 0,002   | -0,849     | -0,668 | -0,482 |
| $\alpha_{22}^*$   | M | S | -0,524       | 0,085    | 0,002   | -0,689     | -0,523 | -0,365 |
| $\sigma_{\eta}^2$ |   |   | 0,008        | 0,057    | 0,001   | < 0,000    | 0,001  | 0,041  |
| $\sigma_{\nu}^2$  |   |   | 0,100        | 0,022    | < 0,00  | 0,067      | 0,096  | 0,151  |
| $\rho$            |   |   | -0,707       | 13,888   | 0,283   | -25,543    | -0,299 | 24,790 |
| $\lambda$         |   |   | -0,113       | 1,472    | 0,030   | -2,080     | -0,278 | 2,089  |
| $\sigma^2$        |   |   | 0,302        | 0,098    | 0,002   | 0,173      | 0,284  | 0,537  |
| DIC               |   |   | -624,9       | Deviance | 102,4   |            |        |        |

E: Escova (C: Convencional, M: Monobloco); P: Dentífrício (S: Sim, N: Não); D.P.: Desvio padrão,  $\sigma^2$ : parâmetro de escala da normal assimétrica.

### 6.3.1.1 Modelo 1 - Prioris usuais ou canônicas

Os resultados da tabela 5 que são provenientes do **modelo 1**. Note que os coeficientes indicados por um \* foram corrigidos por uma constante dada por

$$\alpha_{ij}^* = \alpha_{ij} + \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} \sigma, \quad (6.36)$$

para que a regressão seja centrada em 0, conforme indica Rodríguez (2005) e Freitas (2005). É importante observar que a distribuição dos resíduos corrigidos ( $\tilde{\zeta}_{ijk}$ ), com este ajuste segue uma distribuição  $SN\left(-\sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} \sigma, \sigma^2, \lambda\right)$ . As figuras 19, 20 e 21 indicam o histórico das simulações MCMC e as densidades empíricas estimadas.

Há problemas quanto às convergências dos parâmetros do **modelo 1** representando os coeficientes do modelo de regressão, principalmente antes da centralização, talvez causada pela presença de bimodalidade ou vice-versa. É clara a dificuldade de convergência do

parâmetro  $\rho$  e chama a atenção seu grande intervalo de predição que inclui o 0, indicado a ausência de assimetria, pois não se poderia rejeitar a hipótese nula de que  $\rho = 0$ .

Ainda, quanto à análise dos parâmetros temos que os coeficientes ajustados ( $\alpha_{ij}^*$ ) e os coeficientes de uniformidade ( $\gamma_{ij}$ ) são significativamente diferentes de 0, resultado este auferido pelos intervalos de predição de 95%. Vale ressaltar que a distribuição à posteriori para esses parâmetros é unimodal.

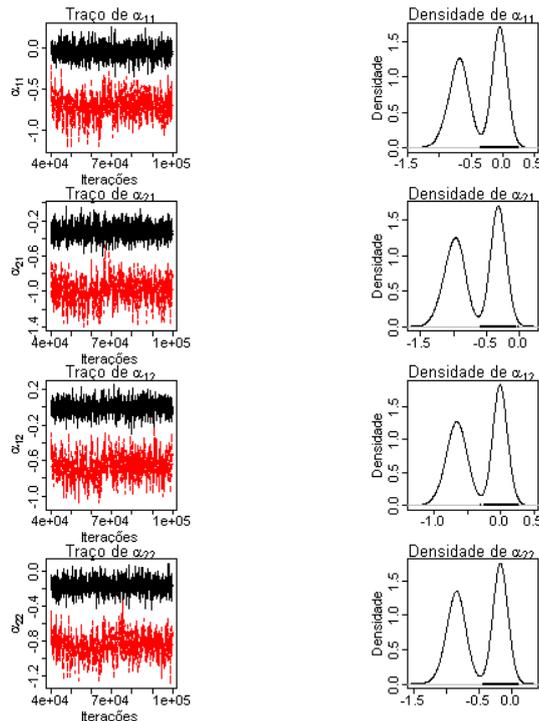


Figura 19: Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros  $\alpha_{11}, \alpha_{12}, \alpha_{21}$  e  $\alpha_{22}$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

Os coeficientes  $\gamma_{ij}, i = 1, 2; j = 1, 2$ , representam a taxa de remoção de placa para cada combinação dos tratamentos e os intervalos de predição para estes quatro parâmetros contém o valor 1, de modo que a utilização das escovas e dentifrício não a afetam diretamente, sendo iguais. Os coeficientes expressos por  $\alpha_{ij}^*, i = 1, 2; j = 1, 2$ , estão em escala logarítmica e a sua transformação em escala natural é apresentada na tabela 6.

Os coeficientes ajustados  $\beta_{ij}$  indicam o que já era esperado. Todas as combinações de dentifrícios e escovas diminuem a quantidade de placa pós-teste, e quanto menor este coeficiente, menor será o índice pós-teste. No caso, a utilização de escova convencional

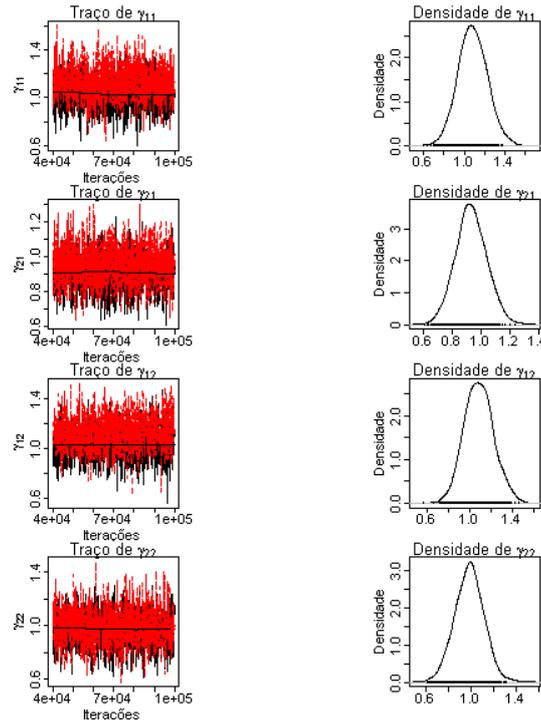


Figura 20: Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros  $\gamma_{11}, \gamma_{12}, \gamma_{21}$  e  $\gamma_{22}$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

| Parâmetro    | Níveis | Estimativa             | Interv. Pred. |       |
|--------------|--------|------------------------|---------------|-------|
|              |        |                        | 2,5%          | 97,5% |
| $\beta_{11}$ | C N    | $\exp(-0,384) = 0,681$ | 0,562         | 0,825 |
| $\beta_{12}$ | M N    | $\exp(-0,346) = 0,708$ | 0,595         | 0,844 |
| $\beta_{21}$ | C S    | $\exp(-0,666) = 0,514$ | 0,428         | 0,618 |
| $\beta_{22}$ | M S    | $\exp(-0,524) = 0,592$ | 0,502         | 0,694 |

Tabela 6: Transformação dos coeficientes  $\alpha_{ij}$  para a escala natural,  $\beta_{ij} = \exp(\alpha_{ij})$

associada ao uso do dentifrício ( $\beta_{21}$ ) leva à maior redução de placa. No entanto, parece não haver diferenças significativas entre as combinações de tratamentos.

Sef (1999), utilizando o mesmo conjunto de dados, mas com a metodologia usual, chega às mesmas conclusões, mas com coeficientes estimados relativamente diferentes. O importante na análise é verificar o poder preditivo deste modelo, dada as inúmeras aplicações importantes que ele pode vir a desempenhar em Odontologia. A figura 22 contém alguns gráficos diagnósticos dos resíduos. Nestes gráficos (figura 22) há fortes evidências de que os resíduos sejam realmente distribuídos assimetricamente (gráfico B) e o gráfico Q-Q para a distribuição normal assimétrica (gráfico C) parece não indicar um

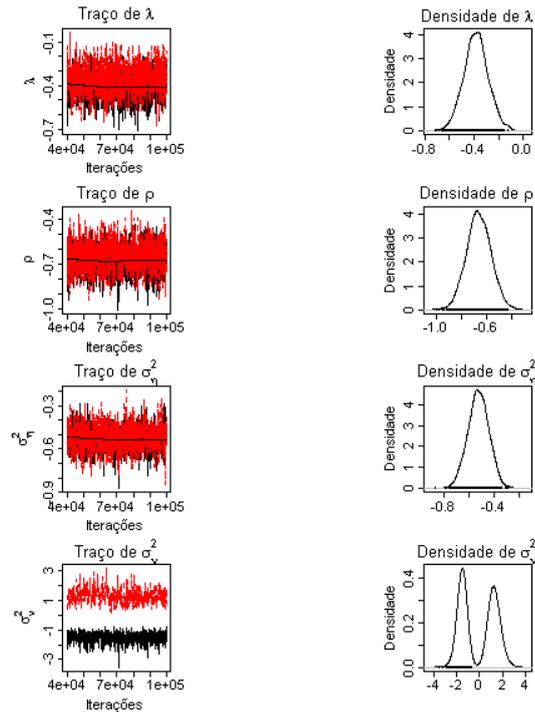


Figura 21: Resultados do histórico de atualizações do algoritmo MCMC para os parâmetros  $\lambda, \rho, \sigma_\eta^2$  e  $\sigma_\nu^2$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

bom ajuste entre os resíduos e aquela distribuição, de modo que, a princípio, o modelo parece sub-ajustado, apesar da aparente homogeneidade da variância (gráfico A). Alguns pontos discrepantes podem ser observados neste gráfico.

### 6.3.1.2 Modelo 2 - Priori T-Gama não-informativa

Os resultados relativos ao **modelo 2** estão disponíveis na tabela 7.

É fácil verificar que os coeficientes do **modelo 2**, são muito semelhantes ao **modelo 1** (tabela 5), mas o coeficiente modulador da assimetria ( $\rho$ ), os parâmetros de escala e os relacionados à normal assimétrica são diferentes. Ainda é notório que o valor do parâmetro de assimetria ( $\lambda$ ) e o próprio  $\rho$  são significativamente diferentes de 0 para esta abordagem, diferente do que aconteceu no modelo 1. A estimativa da razão  $\tau = \frac{4\sigma_\eta^2}{\pi^2\sigma_\nu^2} \approx 5,80$ . Ainda, como pode ser visto nas figuras 23, 24 e 25, existe convergência para  $\rho$  e  $\tau$ , mas a faixa de variação das amostras geradas são pequenas, talvez influenciada pelo algoritmo utilizado, Metropolis-Hastings, que é sabidamente menos precisa que a amostragem por

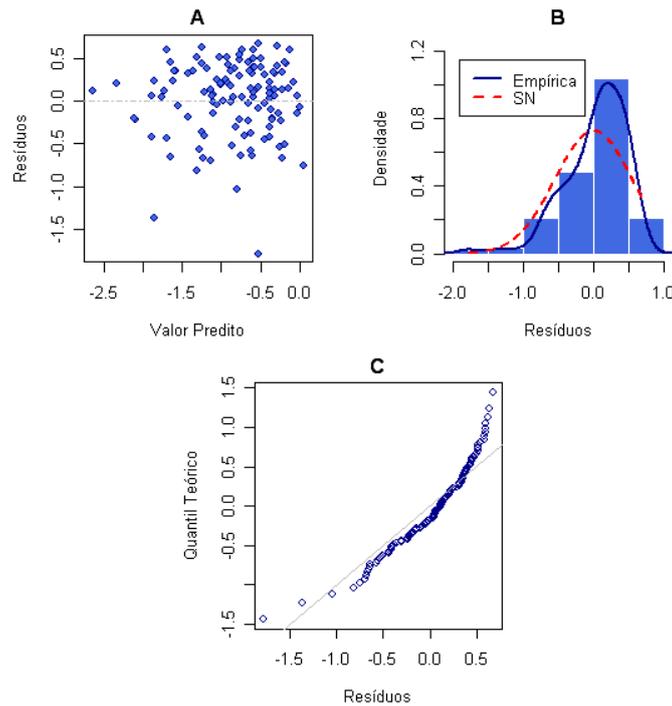


Figura 22: Gráficos diagnósticos do **modelo 1** em escala logarítmica. **A**: Gráfico de resíduos por valores preditos; **B**: Histograma com as curvas estimadas empiricamente e por uma  $SN(0,049; 0,302; \lambda = -0,113)$ ; **C**: Gráfico Q-Q para a normal assimétrica e os resíduos.

Gibbs (GELMAN et al., 2000), mas para nosso conjunto de dados, este algoritmo se mostrou mais estável que a construção da t-Student por meio da mistura de distribuições normais-gama (6.7). Observe também que há uma grande variabilidade para o valor de  $\tau$ , que leva a um intervalo de predição de 95% muito grande e a presença de bimodalidade.

Quando a aplicação 4.3.1 foi executada, o critério para escolha do modelo foi dada pelo AIC (Akaike Information Criteria), mas em modelos bayesianos, este critério não é conveniente, por isso adotou-se o critério DIC (Deviance Information Criteria) (PAULINO et al., 2003, p. 352). Analisando-se os resultados os modelos 1 e 2, este último mais adequado, pois apresenta menor DIC (no entanto a causa deste “excelente” ajuste deve ser questionada e pesquisada com mais cuidado, uma vez que pelo critério utilizado, o modelo 2 é muito superior ao modelo 1). A análise de resíduos do ajuste é apresentada na figura 26. Os resíduos corrigidos ( $\widetilde{\zeta}_{ijk}$ ) deverão seguir uma distribuição teórica dada por  $SN(0,307; 0,181; -2,122)$ , que não é muito diferente, graficamente, do modelo 1. (figuras 22 e 26). A qualidade do ajuste melhora com relação ao modelo 1, pois, ao menos, o

Tabela 7: Resultados da estimação dos parâmetros para o **Modelo 2**.

| <b>Modelo 2</b> |   |   |              |       |          |            |        |        |
|-----------------|---|---|--------------|-------|----------|------------|--------|--------|
| Parâmetros      | E | P | Estatísticas |       |          | Int. Pred. |        |        |
|                 |   |   | Média        | D.P.  | Erro MC  | 2,5%       | 50%    | 97,5%  |
| $\alpha_{11}$   | C | N | -0,059       | 0,087 | 0,001    | -0,228     | -0,063 | 0,114  |
| $\alpha_{12}$   | M | N | -0,015       | 0,083 | 0,001    | -0,172     | -0,016 | 0,153  |
| $\alpha_{21}$   | C | S | -0,329       | 0,091 | 0,001    | -0,502     | -0,331 | -0,147 |
| $\alpha_{22}$   | M | S | -0,179       | 0,083 | 0,001    | -0,341     | -0,181 | -0,012 |
| $\gamma_{11}$   | C | N | 1,032        | 0,124 | 0,002    | 0,788      | 1,033  | 1,272  |
| $\gamma_{12}$   | M | N | 1,030        | 0,120 | 0,002    | 0,793      | 1,031  | 1,273  |
| $\gamma_{21}$   | C | S | 0,911        | 0,101 | 0,001    | 0,711      | 0,909  | 1,109  |
| $\gamma_{22}$   | M | S | 0,974        | 0,118 | 0,002    | 0,744      | 0,973  | 1,206  |
| $\alpha_{11}^*$ | C | N | -0,405       | 0,086 | 0,001    | -0,577     | -0,403 | -0,241 |
| $\alpha_{12}^*$ | M | N | -0,361       | 0,082 | 0,001    | -0,524     | -0,361 | -0,202 |
| $\alpha_{21}^*$ | C | S | -0,674       | 0,090 | 0,001    | -0,847     | -0,673 | -0,497 |
| $\alpha_{22}^*$ | M | S | -0,524       | 0,080 | 0,001    | -0,687     | -0,523 | -0,374 |
| $\tau$          |   |   | 5,795        | 5,893 | 0,086    | 0,003      | 3,883  | 14,690 |
| $\sigma_v^2$    |   |   | 0,090        | 0,013 | < 0,000  | 0,068      | 0,089  | 0,122  |
| $\sigma_\eta^2$ |   |   | 1,291        | 1,284 | 0,019    | 0,001      | 1,295  | 2,611  |
| $\rho$          |   |   | -3,663       | 4,607 | 0,067    | -16,690    | -1,430 | -0,203 |
| $\lambda$       |   |   | -1,456       | 0,296 | 0,004    | -2,122     | -1,425 | -0,975 |
| $\sigma^2$      |   |   | 0,284        | 0,076 | 0,001    | 0,181      | 0,268  | 0,495  |
| DIC             |   |   | -418530,0    |       | Deviance | 92,19      |        |        |

E: Escova (C: Convencional, M: Monobloco); P: Dentifrício(S: Sim, N:Não); D.P.: Desvio padrão,  $\sigma^2$ : parâmetro de escala da normal assimétrica.

sub-ajuste diminui muito ( figura 26, B).

Dessa maneira, podemos teorizar que a reparametrização é boa em vários aspectos, principalmente pelo fato de ser possível encontrar uma distribuição fechada para  $\rho$ , a Priori T-Gama.

| Parâmetro    | Níveis | Estimativa             | Interv. Pred. |       |
|--------------|--------|------------------------|---------------|-------|
|              |        |                        | 2,5%          | 97,5% |
| $\beta_{11}$ | C N    | $\exp(-0,405) = 0,667$ | 0,562         | 0,786 |
| $\beta_{12}$ | M N    | $\exp(-0,361) = 0,697$ | 0,592         | 0,817 |
| $\beta_{21}$ | C S    | $\exp(-0,674) = 0,510$ | 0,429         | 0,608 |
| $\beta_{22}$ | M S    | $\exp(-0,524) = 0,592$ | 0,503         | 0,688 |

Tabela 8: Transformação dos coeficientes  $\alpha_{ij}$  para a escala natural,  $\beta_{ij} = \exp(\alpha_{ij})$  para o **modelo 2**

A tabela 8 mostra a transformação dos coeficientes  $\alpha_{ij}, i = 1,2; j = 1,2$ , em  $\beta_{ij}$ , sua escala natural por meio da transformação exponencial. O que ocorre era esperado, haja

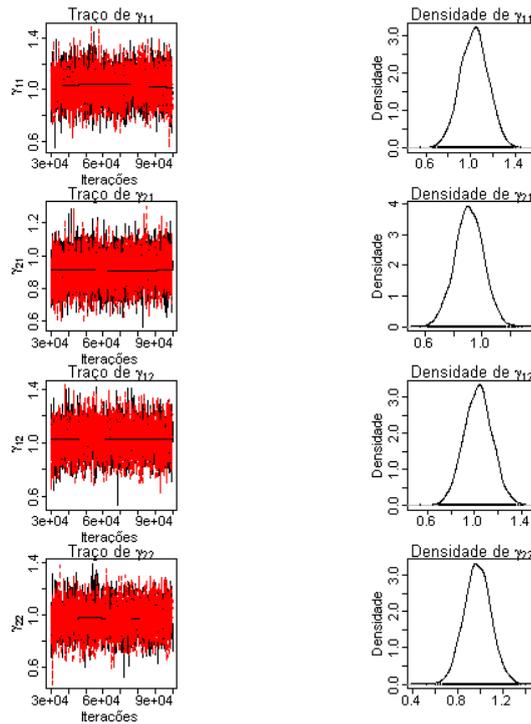


Figura 23: Resultados do histórico de atualizações do algoritmo MCMC do **modelo 2** para os parâmetros  $\alpha_{11}, \alpha_{12}, \alpha_{21}$  e  $\alpha_{22}$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

vista que os valores dos coeficientes estimados e centralizados para o **modelo 2** foram muito próximos do **modelo 1**, e neste já havia indícios de que não havia diferenças significativas entre os índices pré e pós-testes utilizando-se ou não dentifrício e com escova dental convencional ou monobloco. O resultado concorda com aqueles obtidos por Sef (1999), mas com metodologia diferente da abordada aqui. Assim, a taxa de remoção de placa parece constante ( $\gamma_{ij}$  são estimados muito próximos de 1) e os parâmetros indicativos da quantidade residual de placa bacteriana são também próximos em ambos os modelos (tabelas 5 e 6 para o **modelo 1** e 7 e 8 para o **modelo 2**). O menor valor obtido para o coeficiente de placa bacteriana residual ( que implica em menores valores pós-teste) foi para aqueles indivíduos que utilizam escova convencional associada ao dentifrício.

### 6.3.1.3 Modelo 3 - Análise de covariância simples com erro normal.

Para fins de comparação, um modelo com a restrição de  $\rho = 0$ , que corresponderia ao modelo normal, foi simulado, com os valores iniciais para os parâmetros iguais aos estimados para o **modelo 1**. Espera-se que os coeficientes fiquem muito próximo daqueles

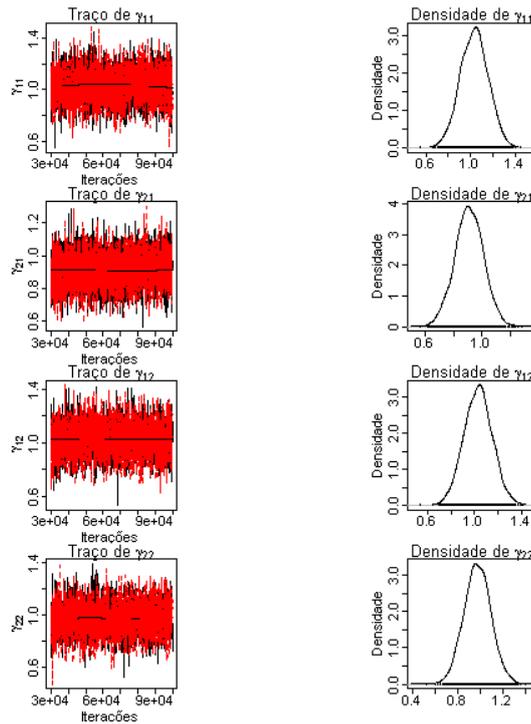


Figura 24: Resultados do histórico de atualizações do algoritmo MCMC do **modelo 2** para os parâmetros  $\gamma_{11}, \gamma_{12}, \gamma_{21}$  e  $\gamma_{22}$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

encontrados em Sef (1999), uma vez que estarão sob prioris não-informativas, o que seria equivalente à obtenção dos estimadores dos parâmetros por máxima verossimilhança na inferência clássica. Este modelo não é muito plausível pois exclui a possibilidade de captação de efeitos aleatórios, o que se espera em modelos longitudinais. Ele foi feito somente para comparação. A tabela 9 apresenta os resultados utilizando-se o **modelo 3** com a restrição de que  $\rho = 0$ .

Com os resultados da tabela 9, temos que todos os coeficientes estimados são diferentes de 0, segundo a análise dos intervalo de predição de 95%.

O interesse, conforme visto anteriormente, é obter os coeficientes em escala original, uma vez que foi estimado em escala logarítmica. A tabela 10 apresenta os resultados em escala original para o **modelo 3**.

Os resultados da tabela 10 indicam que a escova convencional com a utilização de dentifrício tem o maior poder de remoção residual de placa, uma vez que o coeficiente é menor. No entanto, os intervalos de predição de 95% não indicam diferenças significativas

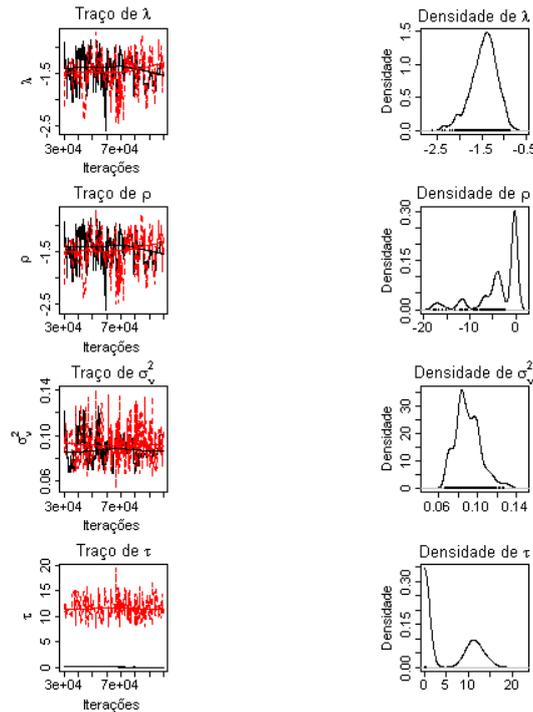


Figura 25: Resultados do histórico de atualizações do algoritmo MCMC do **modelo 2** para os parâmetros  $\lambda, \rho, \tau$  e  $\sigma_v^2$  (gráficos à esquerda) e densidades empíricas estimadas da posteriori destes parâmetros (direita).

Tabela 9: Resultados da estimação dos parâmetros para o **Modelo 3**.

| Modelo 3      |       |   |              |       |                       |            |        |        |
|---------------|-------|---|--------------|-------|-----------------------|------------|--------|--------|
| Parâmetros    | E     | P | Estatísticas |       |                       | Int. Pred. |        |        |
|               |       |   | Média        | D.P.  | Erro MC               | 2,5%       | 50%    | 97,5%  |
| $\alpha_{11}$ | C     | N | -0,329       | 0,092 | $1,10 \times 10^{-3}$ | -0,511     | -0,330 | -0,148 |
| $\alpha_{12}$ | M     | N | -0,289       | 0,085 | $1,01 \times 10^{-3}$ | -0,457     | -0,290 | -0,124 |
| $\alpha_{21}$ | C     | S | -0,626       | 0,094 | $1,12 \times 10^{-3}$ | -0,815     | -0,626 | -0,444 |
| $\alpha_{22}$ | M     | S | -0,500       | 0,081 | $9,66 \times 10^{-4}$ | -0,656     | -0,499 | -0,343 |
| $\gamma_{11}$ | C     | N | 1,151        | 0,140 | $1,75 \times 10^{-3}$ | 0,863      | 1,148  | 1,443  |
| $\gamma_{12}$ | M     | N | 1,179        | 0,115 | $1,68 \times 10^{-3}$ | 0,903      | 1,178  | 1,453  |
| $\gamma_{21}$ | C     | S | 0,964        | 0,088 | $1,37 \times 10^{-3}$ | 0,732      | 0,965  | 1,192  |
| $\gamma_{22}$ | M     | S | 0,960        | 0,137 | $1,63 \times 10^{-3}$ | 0,694      | 0,958  | 1,234  |
| $\rho$        |       |   | 0,000        | —     | —                     | —          | —      | —      |
| $\lambda$     |       |   | 0,000        | —     | —                     | —          | —      | —      |
| $\sigma^2$    |       |   | 0,159        | 0,021 | $2,52 \times 10^{-4}$ | 0,123      | 0,157  | 0,206  |
| DIC           | 301,3 |   | Deviance     |       | 292,2                 |            |        |        |

E: Escova (C: Convencional, M: Monobloco); P: Dentifrício (S: Sim, N: Não); D.P.: Desvio padrão,  $\sigma^2$ : parâmetro de escala da normal assimétrica.

para os índices de placa antes e após os tratamentos, resultado este que também concorda com aqueles obtidos por Sef (1999).

Pelo critério DIC, o modelo 2 seria, ainda, o de escolha, uma vez que apresenta o menor

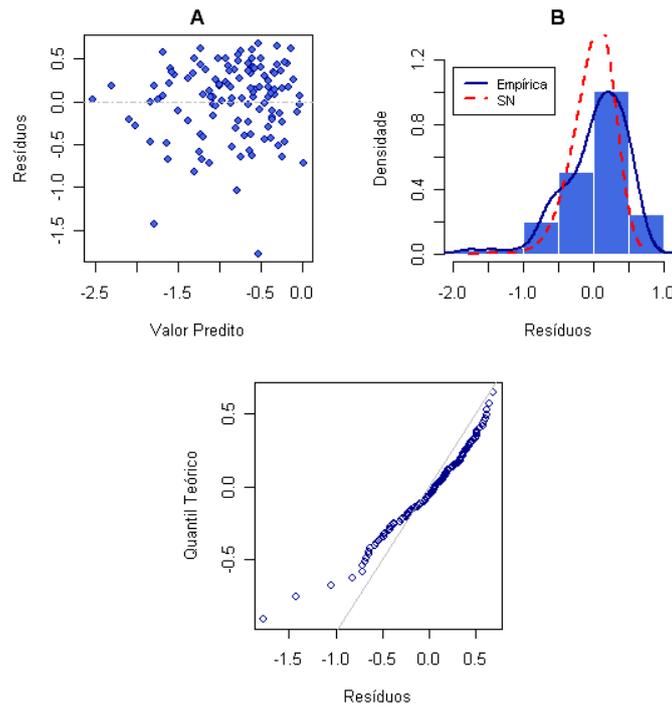


Figura 26: Gráficos diagnósticos do **modelo 2** em escala logarítmica. **A:** Gráfico de resíduos por valores preditos; **B:** Histograma com as curvas estimadas empiricamente e por uma  $SN(0,487; 0,380; -6,996)$ ; **C:** Gráfico Q-Q para a normal assimétrica e os resíduos.

| Parâmetro    | Níveis | Estimativa             | Interv. Pred. |       |
|--------------|--------|------------------------|---------------|-------|
|              |        |                        | 2,5%          | 97,5% |
| $\beta_{11}$ | C N    | $\exp(-0,329) = 0,720$ | 0,600         | 0,862 |
| $\beta_{12}$ | M N    | $\exp(-0,289) = 0,749$ | 0,663         | 0,883 |
| $\beta_{21}$ | C S    | $\exp(-0,626) = 0,535$ | 0,443         | 0,641 |
| $\beta_{22}$ | M S    | $\exp(-0,500) = 0,607$ | 0,519         | 0,710 |

Tabela 10: Transformação dos coeficientes  $\alpha_{ij}$  para a escala natural,  $\beta_{ij} = \exp(\alpha_{ij})$  para o **modelo 3**

valor desta estatística. O modelo normal (modelo 3) apresentou um DIC positivo, o que o torna o menos adequado, segundo o critério adotado. Com finalidade de ilustração, foram apresentados alguns gráficos diagnósticos do modelo 3 (figura 27) para simples comparação com o ajuste do **modelo 1** e **2** (figura 22 e 26, respectivamente).

Nos gráficos da figura 27 é possível notar que a distribuição normal não é adequada para o modelo (gráfico C). Observa-se, no gráfico B desta figura, que a curva estimada fica relativamente distante da massa de dados expressa pelo histograma e que o gráfico de quantil, comparando os resíduos com uma distribuição normal teórica estimada a partir dos dados fica muito fora da diagonal, o que indica ausência de normalidade, pelo menos,

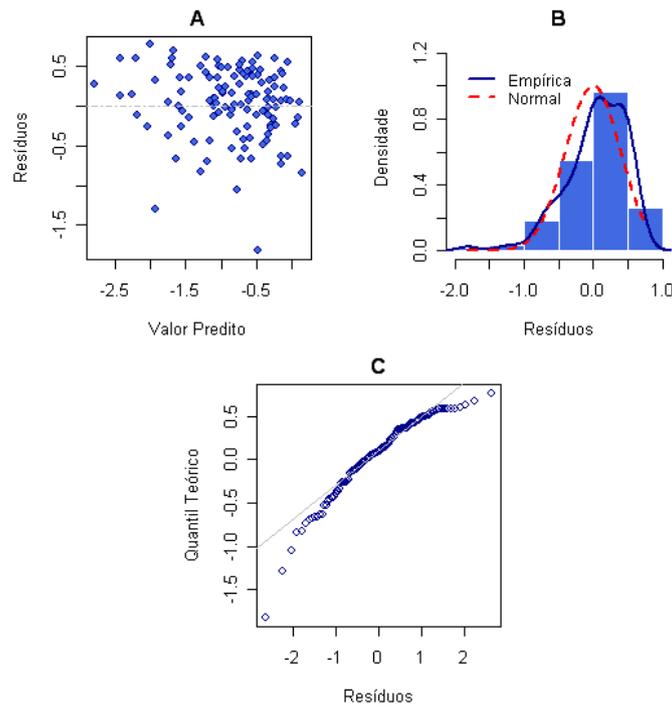


Figura 27: Gráficos diagnósticos do **modelo 3** em escala logarítmica. **A:** Gráfico de resíduos por valores preditos; **B:** Histograma com as curvas estimadas empiricamente e por uma  $N(0; 0,159)$ ; **C:** Gráfico Q-Q para a distribuição normal e os resíduos.

no aspecto gráfico. O teste formal de Shapiro-Wilks, cuja hipótese nula é definida com os resíduos distribuídos por uma normal, é rejeitada em quaisquer níveis de significância usuais ( $p\text{-valor} = 9,367 \times 10^{-6}$ ).

Dessa maneira, o modelo de Sef & Singer utilizando distribuições normais assimétricas na forma de Sahu & Dey sob a reparametrização proposta para a “precisão” da distribuição de t de Student (seção 6.2) apresentou melhores resultados, indicando um ajuste interessante e um razoável caráter preditivo.

Assim, os nossos resultados concordam com aqueles obtidos por Singer & Andrade (1997) e Sef (1999), mas com a vantagem de podermos, neste trabalho, quantificarmos a assimetria existente nos dados pelas restrições impostas ao modelo de Sef & Singer.

## 7 *Considerações finais e propostas futuras*

Neste capítulo faremos uma revisão do que apresentamos nesta dissertação, dividindo-o pelos assuntos abordados; é evidente que todos eles são conexos, no entanto, a análise individual dos pontos é importante porque muitos detalhes são deixados em aberto, pois imaginamos que um pouco mais de reflexão sobre estes assuntos pode resultar em melhorias e extensões daquilo que aqui já desenvolvemos.

Inicialmente, vamos tratar sobre a proposta inicial, os modelos para dados pré-teste/pós-teste.

### 7.1 Modelos para dados pré-teste/pós-teste

A proposição inicial do trabalho foi estudar e compreender melhor a estrutura de um modelo para dados pré-teste/pós-teste, com o auxílio de um conjunto de dados reais sobre o índice de placa bacteriana em crianças antes e depois de estarem submetidas às covariáveis do estudo. A análise de dados pré-teste/pós-teste é amplamente empregada em várias áreas da ciência e este fato chama a atenção, pois as abordagens são todas baseadas em modelos lineares, obviamente com efeitos aleatórios, uma vez que os dados são correlacionados. Vários autores já estudaram estes modelos, e algum deles foram citados nesta dissertação, mas nos chamou a atenção um modelo multiplicativo proposto por Singer & Andrade (1997) e posteriormente detalhado por Sef (1999), em sua dissertação de mestrado. O modelo, que aqui chamamos de *Sef & Singer*, é especial, pois existem várias restrições associadas a ele, principalmente a necessidade de que a reta estimada

passa pelo intercepto e outras mais que possivelmente afetariam a característica dos erros serem simétricos, como foi proposto à época. Além disso, ele foi utilizado para analisar o conjunto de dados aqui mencionado e seus parâmetros tinham interessantes interpretações sobre o problema da medida de placa bacteriana. Desta forma, imaginamos adicionar ao modelo erros distribuídos por normais assimétricas, buscando verificar uma melhora do ajuste, haja vista que a diferença entre a medida pós-teste e pré-teste, denominada *ganho* (STANEK III, 1998) era muito assimétrica. Assim, utilizamos a idéia do modelo de Sef & Singer com a adição de erros normais assimétricos que norteou todo o trabalho após esta definição.

O modelo em si, frisamos, é extremamente interessante, pois, dá aos seus parâmetros interpretações que são de fácil assimilação e “visualização” pelo pesquisador. Alguns estudos (AOKI et al., 2001) e (AOKI et al., 2003) já abordaram os modelos para dados pré-teste e pós-teste com introdução de erros nas variáveis, uma vez que o índice de placa bacteriana é uma medida, e como tal, sujeita à imprecisão. Dessa forma, deixamos em aberto a questão da *introdução de erro nas variáveis sob a abordagem de Sef & Singer*, mas espera-se que haja uma grande contribuição à exatidão das estatísticas obtidas a partir desta modelagem, uma vez que outra fonte de variação, o erro de medida poderá ser estimado e controlado.

Como foi dito, o modelo estudado aqui consistia de adicionar assimetria aos erros, de forma que trataremos na próxima seção de uma nova abordagem que inclui, ao mesmo tempo, as desejadas capacidades de detecção de assimetrias e o próprio efeito aleatório, a representação de Sahu & Dey para as normais assimétricas.

## 7.2 Representação de Sahu & Dey

A representação de Sahu & Dey (SAHU et al., 2003; LIU; DEY, 2005), definida na proposição 3.1.1, é uma maneira eficiente de se representar a normal assimétrica, principalmente em termos computacionais. Ela é muito semelhante à forma estocástica de Henze, e portanto, pode facilmente ser simulada em computadores onde haja algum gerador de

números pseudo-aleatórios normais.

Basicamente a representação de Sahu & Dey é dada por

$$\zeta = \rho |\eta| + \nu,$$

onde  $|\eta| \sim HN(0, \sigma_\eta^2)$  e  $\nu \sim N(0, \sigma_\nu^2)$ . Os valores da seminormal (HN) contaminam os valores da distribuição normal por meio de um parâmetro  $\rho$ , que aqui denominamos de *parâmetro modulador de assimetria*, mas que, ao final, faz o papel do parâmetro de assimetria  $\lambda$  da normal assimétrica padrão de Sahu et al. (2003). Este fato fica evidente quando o parâmetro de escala da seminormal ( $\sigma_\eta^2$ ) e a variância da normal ( $\sigma_\nu^2$ ) são, ambos, idênticos a 1. Se observarmos a forma de Sahu & Dey, veremos que ela é muito similar à forma da decomposição do erro no modelo de Sef & Singer (definição 3.1.1). Logo, podemos representar a decomposição do erro do modelo como a soma de efeito aleatório e erro propriamente dito, de modo que, se considerarmos, no modelo de Sef & Singer, os efeitos aleatórios distribuídos por uma seminormal e os erros como normais, acrescentando-se o parâmetro modulador de assimetria de forma multiplicativa à frente do efeito aleatório, teremos um modelo misto com a capacidade de captar as assimetrias, uma grande inovação desta dissertação. Pela própria configuração do parâmetro  $\rho$ , podemos, agora, quantificar quão assimétrico são os dados.

Para trabalhos futuros, *pode ser estendida a representação de Sahu & Dey para o caso multivariado*, com aplicação em modelos longitudinais onde não somente os interceptos são aleatórios. Imaginamos, que, de alguma forma, uma matriz contendo os parâmetros moduladores de assimetria deve ser construída e esta questão é deixada em aberto pela exigüidade do tempo disponível para desenvolver esta dissertação.

Utilizamos duas maneiras distintas de obter estimativas para o modelo de Sef & Singer sob a representação de Sahu e Dey. Agora trataremos da estimação pelo algoritmo E-M, utilizado na inferência clássica. A estimação pela inferência bayesiana terá um capítulo à parte, uma vez que novas contribuições foram dadas.

### 7.2.1 Estimação pelo algoritmo E-M

No final da parte I foram relacionadas diversas dificuldades técnicas em se obter estimativas dos parâmetros envolvidos em nossa modelagem pelo algoritmo E-M. A teoria para a aplicação do algoritmo E-M foi mostrada com detalhes e a elaboração de um programa para estimá-los, também apresentada no apêndice D, não é muito trabalhosa, no entanto, o que observamos foi uma grande sensibilidade do algoritmo aos valores iniciais. Este problema foi resolvido procurando-se uma região factível no espaço paramétrico para buscar a solução utilizando-se o mesmo conjunto de dados em um modelo misto com distribuição normal, tanto para o efeito aleatório como para o erro propriamente dito, e fixando-se o parâmetro modulador da assimetria em 0.

Após exaustivas tentativas obtivemos um resultado que concordava com o trabalho de Singer & Andrade (1997), mas outro problema foi a obtenção do erro padrão para cada um dos parâmetros estimados. A solução foi utilizar a técnica bootstrap, que é recomendada no importante livro de estatística computacional (GENTLE et al., 2004). Assim, *seria necessário desenvolver algumas melhorias para aplicação do algoritmo E-M* para nossa modelagem, tornando-o mais rápido e eficiente, ao mesmo tempo em que melhorasse a sua convergência, economizando tempo e recursos computacionais.

## 7.3 Inferência Bayesiana

Foi no estudo dos modelos pré-teste/pós-teste sob a abordagem assimétrica sob a representação de Sahu & Dey que esta dissertação apresenta seus resultados mais interessantes e úteis, além de serem inovadores. A abordagem bayesiana da distribuição normal assimétrica já foi assunto explorado com qualidade e resultados importantes por Rodríguez (2005) em sua dissertação de mestrado. Naquele texto há o desenvolvimento de uma aproximação da priori não-informativa de Jeffreys para o parâmetro de assimetria da distribuição normal assimétrica por meio de uma distribuição t-Student representada por  $t_{\frac{1}{2}}\left(0, \frac{\pi^2}{4}\right)$ . Este resultado é a base para todo o desenvolvimento posterior que apresentamos aqui. O parâmetro modulador da assimetria de Sahu & Dey pode ser definido como

uma função dos parâmetros de escala da distribuição seminormal e normal que definem a representação, especificamente,  $\rho = \frac{\lambda\sigma_\nu}{\sigma_\eta}$ , onde  $\lambda$  é o parâmetro de assimetria da normal assimétrica canônica (Azzalini). Quando  $\sigma_\eta^2 = \sigma_\nu^2$ , então, estamos sob a t-Student definida por Rodríguez (2005). Assim, se é definida uma aproximação da priori de Jeffreys para  $\lambda$ , então, por meio de uma transformação de variáveis aleatórias podemos chegar a uma priori não informativa para  $\rho \mid \sigma_\eta^2, \sigma_\nu^2 \sim t_{\frac{1}{2}} \left( 0, \frac{\pi^2 \sigma_\nu^2}{4\sigma_\eta^2} \right)$ . Sob esta parametrização, fica muito complicado encontrar uma distribuição fechada para a marginal de  $\rho$ , de modo que se definiu

$$\tau = \frac{4\sigma_\eta^2}{\pi^2\sigma_\nu^2}, \tau \sim \text{gama} \left( \frac{a}{2}, \frac{b}{2} \right),$$

com  $\tau$  fazendo papel de “precisão” (em analogia à distribuição normal) da distribuição t-Student. Escolheu-se  $\tau$  desta forma, pois é usual utilizar prioris gamas para a precisão e dessa maneira, o produto de  $f(\rho \mid \tau) \times f(\tau)$  apresenta um resultado que é aparentemente novo: surge a definição da **distribuição conjunta** dada pela expressão (6.22). Ela tem dois parâmetros,  $a$  e  $b$ , provenientes da distribuição gama, tal que  $0 < a < \frac{1}{2}$  e  $b > 0$ . A restrição ao parâmetro  $a$  é devida à função de Kummer, uma função matemática resultante de soluções de uma equação diferencial especial e pertence à classe das funções hipergeométricas congruentes (ABRAMOWITZ; STEGUN, 1972, para detalhes). Esta função se mostrou valiosa em nosso trabalho, pois, define a f.d.p. de Kummer [definição 6.1.1, simbolicamente,  $K(\alpha, \beta, \gamma, \delta)$ ], com suporte em  $\mathfrak{R}^+$  e possui quatro parâmetros. Ela foi importante por duas razões:

- possibilitou-nos encontrar a distribuição condicional de  $\tau \mid \rho$  e;
- permitiu-nos encontrar uma priori fechada para a distribuição marginal de  $\rho$ .

Dessa feita, apresentamos algumas de suas propriedades e chegamos a um importante resultado: a distribuição condicionada de  $\tau \mid \rho \sim K \left( \frac{a+1}{2}, \frac{b}{2}, \frac{3}{4}, 2\rho^2 \right)$  que é apresentada pelo teorema 6.2.2. Com este teorema, é possível, então, formular algoritmos MCMC para estimação dos parâmetros, pois se conhece agora a condicional completa.

O resultado mais importante e final desta dissertação, e que pode vir a ser, realmente de grande serventia para a inferência bayesiana foi encontrar a **distribuição priori T-Gama**, dada pelo teorema 6.2.3, cuja função densidade é apresentada a seguir.

$$TG(a,b) = \begin{cases} \frac{\sqrt{2\pi}}{B(\frac{1}{4}, \frac{1}{2})B(\frac{1}{2}, \frac{a}{2})} \left(\frac{b}{2}\right)^{\frac{a}{2}} \frac{U\left(\frac{a+1}{2}, \frac{a}{2} + \frac{3}{4}, \frac{b}{4\rho^2}\right)}{(2\rho^2)^{\frac{a+1}{2}}} & \rho \neq 0 \\ \frac{\sqrt{\pi}}{\sqrt{\frac{b}{4}}B(\frac{1}{4}, \frac{1}{2})B(\frac{a}{2}, \frac{1}{2})} & \rho = 0 \end{cases}.$$

Observe que a densidade depende diretamente da função de Kummer. A restrição nestas novas distribuições (a conjunta e a priori T-Gama), é que o parâmetro  $a < \frac{1}{2}$ . Este fato decorre porque a função de Kummer só é definida no ponto 0 se o parâmetro  $\frac{a}{2} + \frac{3}{4} < 1$ . Assim, só existe f.d.p. se esta restrição for satisfeita. Observe, na figura ?? que se liberarmos os parâmetros, as áreas lá definidas não são unitárias, e portanto, não podem ser funções densidade de probabilidade. Uma interessante interpretação para a distribuição T-Gama é que quanto mais se acredite na relação  $\sigma_\eta^2 = \sigma_\nu^2$ , mais aquela distribuição se aproximaria da T-Student definida por Rodríguez (2005), mas da maneira como este problema foi parametrizado, a prova formal ou gráfica precisaria de mais estudos, e por isso, poderá ser explorada em artigos futuros.

Aplicamos estes conceito em três estimações de modelos para ilustrar, e os resultados foram muito semelhantes, indicando que existe uma robustez do modelo quanto à distribuição à priori adotada. O critério DIC foi utilizado para a escolha do modelo, mas em um dos casos ele se apresentou extremamente negativo, indicando um excelente ajuste, o que é passível de atenção, pois graficamente aquele ajuste (modelo 2 - Priori de Kummer não-informativa) não é muito diferente dos resultados encontrados no modelo 1 (prioris usuais). O modelo 3 somente foi utilizado para exemplificar um caso com parâmetro  $\rho = 0$ . Não encontramos nenhuma diferença quanto à interpretação resultados obtidos por Sef (1999), mas acreditamos que a nossa estimação seja mais coerente com o problema, pois captamos as assimetrias envolvidas. *Uma nova série de análises pode ser realizada utilizando-se a parametrização dada no teorema 6.2.4, com o objetivo de se verificar o desempenho do modelo sob aquelas distribuições à priori.*

Enfim, podemos dizer que a priori T-Gama é uma boa ferramenta à disposição da inferência bayesiana, mas as restrições que se impõe para sua existência pode ser problema em outras situações, pois ela não é definida para o ponto 0. Ainda, computacionalmente, os algoritmos que calculam as função de Kummer, essencial para a priori T-Gama, apresentam instabilidade numérica na vizinhança de 0, o que dificulta a utilização nessas regiões da reta real. Uma alternativa que pode ser para discussões futuras *seria encontrar um relacionamento entre a função de Kummer e funções mais conhecidas, como, por exemplo, a função de Bessel, ou outras mais estáveis, e tentar, por meio dela, obter uma distribuição de probabilidade que permita unificar a aproximação t-Student com a priori T-Gama* ou ainda, generalizar a função T-Gama de forma que no limite e em condições especiais, haja a aproximação desta distribuição pela aproximação da priori de Jeffreys definida por Rodríguez (2005).

Outra interessante matéria de estudo seria a utilização das *prioris de referência* construídas de maneira objetiva, como propõem Bernardo & Smith (1994). A vantagem é que, em algum ponto do método, a construção da priori de referência depende da Informação de Fisher, e já definimos ao longo do trabalho uma aproximação baseada na distribuição t-Student; ainda, ao contrário da Priori T-Gama, aquela teria média e variância finitas.

## *Apoio Financeiro*

Esta dissertação de mestrado contou com o apoio financeiro da **FAPESP**.

(**Processo 05/03377-1**)

## *Referências*

- ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of mathematical functions*. USA: National Bureau of Standards, 1972. Disponível em: <<http://www.math.sfu.ca/~cbm/aands>>.
- AOKI, R. et al. Bayesian analysis of null intercept error-in-variables regression for pretest/post-test data. *Journal of Applied Statistics*, v. 30, n. 1, p. 3–12, 2003.
- AOKI, R.; BOLFARINE, H.; SINGER, J. M. Null intercept measurement error regression models. *Sociedad de Estadística e Investigación Operativa*, v. 10, n. 2, p. 441–457, 2001.
- AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, v. 12, n. 22, p. 171–178, 1985.
- AZZALINI, A. *R package sn: The skew-normal and skew-t distributions (version 0.4-1)*. Università di Padova, Italia, 2006. Disponível em: <URL <http://azzalini.stat.unipd.it/SN>>.
- AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. *Biometrika*, v. 83, n. 4, p. 715–726, 1996.
- BERNARDO, J. M.; SMITH, A. F. M. *Bayesian theory*. [S.l.]: Wiley, 1994.
- BROGAN, D. R.; KUTNER, M. K. Comparative analysis of pretest-posttest research designs. *The American Statistician*, v. 34, n. 4, p. 229–232, 1980.
- CHITTA, S. M. P. *Aplicação dos métodos bayesianos nos sistemas de filas*. Dissertação (Mestrado) — ICMC/USP, São Carlos, 1995.
- EFRON, B.; TIBHIRANI, R. J. *An introduction to the bootstrap*. 1. ed. [S.l.]: CRC Press, 1994. 436 p.
- FIGUEIREDO, M. A. T. *Lecture notes on the E-M algorithm*. Portugal, 2004.
- FREITAS, L. A. *Modelo de regressão com erros normais assimétricos: uma abordagem bayesiana*. Dissertação (Mestrado) — Des/UFSCar, São Carlos, 2005.
- GAMERMAN, D.; LOPES, H. F. *Markov chain monte carlo: Stochastic simulation for bayesian inference*. 2. ed. [S.l.]: Chapman & Hall-CRC, 2006. 323 p.
- GELMAN, A. et al. *Bayesian data analysis*. 1. ed. [S.l.]: Chapman & Hall-CRC, 2000.
- GENTLE, J.; HARDLE, W.; MORI, Y. *Handbook of computational statistics: concepts and methods*. 1. ed. [S.l.]: Springer, 2004. 1070 p.

- HENZE, N. A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, v. 13, p. 271–275, 1986.
- JOHNSON, N. L. Distributional theory of spherical distributions and location-scale parameter. *Sankhya*, v. 32, p. 419–430, 1970.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 5. ed. [S.l.]: Prentice Hall, 2002. 767 p.
- LACHOS, V. H. *Modelos lineares mistos assimétricos*. Tese (Doutorado) — IME/USP, São Paulo, 2004.
- LAIRD, N. M.; WARE, J. H. Random effects models for longitudinal data. *Biometrics*, v. 38, p. 963–974, 1982.
- LIU, J.; DEY, D. K. Modeling random effects for multilevel binomial regression models. Sob revisão. 2005. Disponível em: <<http://merlot.stat.uconn.edu/~junfeng/publication.html>>.
- MONAGAN, M. B. et al. *Maple 10 Programming Guide*. Waterloo ON, Canada: Maplesoft, 2005.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística bayesiana*. 1. ed. Lisboa, Portugal: Fundação Calouste Gulbenkian, 2003. 446 p.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2006. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.
- SUYAMA, E. (Ed.). *Modelos de efeitos aleatórios para dados longitudinais*. Ribeirão Preto, SP: RBRAS, 1995.
- RENCHER, A. C. *Methods of Multivariate Analysis*. 2. ed. [S.l.]: Wiley Intescience, 2002. 738 p.
- RODRÍGUEZ, C. L. B. *Inferência bayesiana no modelo normal assimétrico*. Dissertação (Mestrado) — IME/USP, São Paulo, 2005.
- ROSS, S. *A first course in probability*. 1. ed. New York, USA: Macmillan Publishing, 1976.
- SAHU, S.; DEY, D. K.; BRANCO, M. D. A new class of multivariate skew distributions with applications to bayesian regression models. *The Canadian Journal of Statistics*, v. 31, n. 2, p. 129–150, 2003.
- SEF, H. C. *Modelos multiplicativos de regressão para dados pré-teste/pós-teste em blocos*. Dissertação (Mestrado) — IME/USP, São Paulo, 1999.
- SINGER, J. M.; ANDRADE, D. F. Models for analysis of pretest/posttest data. *Biometrics*, v. 53, n. 2, p. 729–735, 1997.
- STANEK III, E. J. Choosing a pretest-posttest analysis. *The American Statistician*, v. 42, p. 129–150, 1998.

- 
- TANNER, M. A. *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*. 3. ed. [S.l.]: Springer, 2005. 220 p.
- THOMAS, A. The BUGS language. *R News*, v. 6, n. 1, p. 17–21, March 2006. Disponível em: <<http://CRAN.R-project.org/doc/Rnews/>>.
- THOMAS, A. et al. Making BUGS open. *R News*, v. 6, n. 1, p. 12–17, March 2006. Disponível em: <<http://CRAN.R-project.org/doc/Rnews/>>.
- VENABLES, W. N.; RIPLEY, B. D. *Modern applied statistics with s*. 4. ed. [S.l.]: Springer-Verlag, 2002. 495 p.

## *APÊNDICE A – Justificativas do algoritmo E-M*

### A.1 Convergência

Recordando-se da notação utilizada no capítulo 4, seja  $k(\mathbf{x}|\mathbf{y}; \Psi) = \frac{g_c(\mathbf{x}; \Psi)}{g(\mathbf{y}; \Psi)}$  a probabilidade condicional de  $\mathbf{X}$  dado  $\mathbf{Y} = \mathbf{y}$ . A verossimilhança completa pode ser expressa como

$$\begin{aligned} \ln L_C(\Psi) &= \ln g_C(\mathbf{x}; \Psi) \\ &= \ln L(\Psi) + \ln k(\mathbf{x}|\mathbf{y}; \Psi). \end{aligned} \tag{A.1}$$

Tomando-se a esperança em ambos os lados da igualdade em (A.1) com respeito à distribuição de  $\mathbf{X}|\mathbf{Y}$ , utilizando-se a estimativa obtida a partir da  $k$ -ésima iteração do algoritmo para  $\Psi, \Psi^{(k)}$ , temos:

$$Q(\Psi; \Psi^{(k)}) = \ln L(\Psi) + H(\Psi; \Psi^{(k)}), \tag{A.2}$$

tal que  $H(\Psi, \Psi^{(k)}) = E_{\Psi^{(k)}}[\ln k(\mathbf{x}|\mathbf{y}; \Psi) | \mathbf{Y}]$ . Segue de (A.2) que

$$\begin{aligned} \ln L(\Psi^{(k+1)}) - \ln L(\Psi^{(k)}) &= \left[ Q(\Psi^{(k+1)}; \Psi^{(k)}) - Q(\Psi^{(k)}; \Psi^{(k)}) \right] \\ &\quad - \left[ H(\Psi^{(k+1)}; \Psi^{(k)}) - H(\Psi^{(k)}; \Psi^{(k)}) \right]. \end{aligned} \tag{A.3}$$

Pela desigualdade de Jensen,

$$H(\Psi^{(k+1)}; \Psi^{(k)}) \leq H(\Psi^{(k)}; \Psi^{(k)}), \tag{A.4}$$

e como  $Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi^{(k)}; \Psi^{(k)})$ , então a desigualdade (A.4) é não-negativa, logo a função de verossimilhança é não-decrescente após qualquer uma das iterações do algoritmo E-M.

Em outras palavras,

$$L(\Psi^{(k+1)}) \geq L(\Psi^{(k)}) \text{ para } k = 0, 1, 2, \dots \quad (\text{A.5})$$

e assim, existirá a convergência do processo iterativo, mas não há garantias de que seja obtido um ponto de máximo absoluto, podendo, em alguns casos, resultar em convergência em um ponto de sela (GENTLE et al., 2004; TANNER, 2005). Esta última afirmação é válida sob as seguintes condições:

a)  $\left. \frac{\partial Q(\Psi; \Psi^{(k)})}{\partial \Psi} \right|_{\Psi = \Psi^{(k+1)}} = \mathbf{0}$  e,

b) A seqüência  $\{\Psi^{(k)}\}$  converge para algum ponto  $\Psi^*$  e  $\ln k(\mathbf{x}|\mathbf{y}; \Psi)$  é suficientemente lisa (ou melhor, bem-comportada).

Se as condições a) e b) forem satisfeitas, implica que  $\left. \frac{\partial L(\Psi)}{\partial \Psi} \right|_{\Psi = \Psi^*} = \mathbf{0}$ , ou seja,  $\Psi^*$  representa um valor crítico da função de log-verossimilhança.

## A.1.1 Propriedades do Algoritmo E-M

### A.1.1.1 Vantagens

- 1 - É estável, com cada iteração aumentando a log-verossimilhança;
- 2 - Sob algumas condições, alcança a convergência global;
- 3 - É de fácil implementação computacional em muitos casos;
- 4 - Pode ser utilizado para prover dados ausentes ou faltantes.

**A.1.1.2 Desvantagens**

- 1 - Não fornece, automaticamente uma matriz de variâncias e covariâncias, para assim, obterem-se os erros padrões;
- 2 - Pode ter uma lenta convergência;
- 3 - Em alguns problemas, seu uso pode ser dificultado pela própria forma da verossimilhança ou pela dificuldade de se obter as esperanças condicionadas das variáveis latentes.

## *APÊNDICE B - Conjunto de Dados*

### B.1 Dados da aplicação 4.3.1

| Indivíduo | Sexo | Tipo de Escova     |                     |                    |                     |
|-----------|------|--------------------|---------------------|--------------------|---------------------|
|           |      | Hugger             |                     | Convencional       |                     |
|           |      | Antes<br>Escovação | Depois<br>Escovação | Antes<br>Escovação | Depois<br>Escovação |
| 1         | F    | 2.18               | 0.43                | 1.20               | 0.75                |
| 2         | F    | 2.05               | 0.08                | 1.43               | 0.55                |
| 3         | F    | 1.05               | 0.18                | 0.68               | 0.08                |
| 4         | F    | 1.95               | 0.78                | 1.45               | 0.75                |
| 5         | F    | 0.28               | 0.03                | 0.5                | 0.05                |
| 6         | F    | 2.63               | 0.23                | 2.75               | 1.60                |
| 7         | F    | 1.50               | 0.20                | 1.25               | 0.65                |
| 8         | F    | 0.45               | 0.00                | 0.40               | 0.13                |
| 8         | F    | 0.70               | 0.05                | 1.18               | 0.83                |
| 10        | F    | 1.30               | 0.30                | 1.43               | 0.58                |
| 11        | F    | 1.25               | 0.33                | 0.45               | 0.38                |
| 12        | F    | 0.18               | 0.00                | 1.60               | 0.63                |
| 13        | F    | 3.30               | 0.90                | 0.25               | 0.25                |
| 14        | F    | 1.40               | 0.24                | 2.98               | 1.03                |
| 15        | M    | 0.90               | 0.15                | 3.35               | 1.58                |
| 16        | M    | 0.58               | 0.10                | 1.50               | 0.20                |
| 17        | M    | 2.50               | 0.33                | 4.08               | 1.88                |
| 18        | M    | 2.25               | 0.33                | 3.15               | 2.00                |
| 19        | M    | 1.53               | 0.53                | 0.90               | 0.25                |
| 20        | M    | 1.43               | 0.43                | 1.78               | 0.18                |
| 21        | M    | 3.48               | 0.65                | 3.50               | 0.85                |
| 22        | M    | 1.80               | 0.20                | 2.50               | 1.15                |
| 23        | M    | 1.50               | 0.25                | 2.18               | 0.93                |
| 24        | M    | 2.55               | 0.15                | 2.68               | 1.05                |
| 25        | M    | 1.30               | 0.05                | 2.73               | 0.85                |
| 26        | M    | 2.65               | 0.25                | 3.43               | 0.88                |

Tabela 11: Resultado do estudo conduzido pela Faculdade de Odontologia - USP-SP

## B.2 Dados da Aplicação 6.3

|    | Com dentifício |        |           |        | Sem dentifício |        |           |        |
|----|----------------|--------|-----------|--------|----------------|--------|-----------|--------|
|    | Convencional   |        | Monobloco |        | Convencional   |        | Monobloco |        |
|    | Antes          | Depois | Antes     | Depois | Antes          | Depois | Antes     | Depois |
| 1  | 0.68           | 0.62   | 0.87      | 0.81   | 0.81           | 0.75   | 0.93      | 0.87   |
| 2  | 0.64           | 0.64   | 0.60      | 0.60   | 0.85           | 0.65   | 0.90      | 0.75   |
| 3  | 0.15           | 0.00   | 0.50      | 0.20   | 0.45           | 0.20   | 0.31      | 0.18   |
| 4  | 0.62           | 0.25   | 0.75      | 0.33   | 0.62           | 0.50   | 0.58      | 0.45   |
| 5  | 0.25           | 0.04   | 0.91      | 0.37   | 0.83           | 0.58   | 0.70      | 0.45   |
| 6  | 0.37           | 0.33   | 0.50      | 0.33   | 0.25           | 0.16   | 0.33      | 0.33   |
| 7  | 0.25           | 0.20   | 0.95      | 0.83   | 0.91           | 0.91   | 0.54      | 0.54   |
| 8  | 1.37           | 0.93   | 1.43      | 1.31   | 1.06           | 1.00   | 1.80      | 1.62   |
| 9  | 0.75           | 0.41   | 1.33      | 0.62   | 0.83           | 0.83   | 1.20      | 1.00   |
| 10 | 0.50           | 0.33   | 0.61      | 0.44   | 0.44           | 0.16   | 0.72      | 0.38   |
| 11 | 1.33           | 0.83   | 1.29      | 0.66   | 1.00           | 0.70   | 1.08      | 0.91   |
| 12 | 1.45           | 0.50   | 1.66      | 1.04   | 1.00           | 0.50   | 1.66      | 1.29   |
| 13 | 1.12           | 0.41   | 1.16      | 0.75   | 1.12           | 0.45   | 1.16      | 1.16   |
| 14 | 1.12           | 0.91   | 0.87      | 0.50   | 1.16           | 1.12   | 0.83      | 0.45   |
| 15 | 0.33           | 0.20   | 0.58      | 0.33   | 0.33           | 0.29   | 0.58      | 0.45   |
| 16 | 0.50           | 0.29   | 0.54      | 0.37   | 0.29           | 0.20   | 0.54      | 0.45   |
| 17 | 0.58           | 0.50   | 0.66      | 0.66   | 0.41           | 0.37   | 0.50      | 0.45   |
| 18 | 0.91           | 0.62   | 1.00      | 0.70   | 0.62           | 0.62   | 1.45      | 0.83   |
| 19 | 0.30           | 0.30   | 0.70      | 0.50   | 1.20           | 0.50   | 0.90      | 0.80   |
| 20 | 1.00           | 0.58   | 1.08      | 0.87   | 0.70           | 0.37   | 1.08      | 1.00   |
| 21 | 0.91           | 0.41   | 1.16      | 0.45   | 1.33           | 1.16   | 0.50      | 0.45   |
| 22 | 0.50           | 0.36   | 1.18      | 0.63   | 0.63           | 0.59   | 1.50      | 1.13   |
| 23 | 0.70           | 0.50   | 0.70      | 0.41   | 0.79           | 0.75   | 0.37      | 0.29   |
| 24 | 0.12           | 0.08   | 0.16      | 0.12   | 0.29           | 0.20   | 0.00      | 0.00   |
| 25 | 0.37           | 0.20   | 0.33      | 0.25   | 0.83           | 0.62   | 0.33      | 0.33   |
| 26 | 0.29           | 0.12   | 0.00      | 0.00   | 0.08           | 0.00   | 0.50      | 0.12   |
| 27 | 0.79           | 0.29   | 1.00      | 0.54   | 1.00           | 0.79   | 0.66      | 0.45   |
| 28 | 1.20           | 1.00   | 0.20      | 0.10   | 1.00           | 0.85   | 1.05      | 0.62   |
| 29 | 0.66           | 0.16   | 0.83      | 0.33   | 0.50           | 0.16   | 0.33      | 0.00   |
| 30 | 0.20           | 0.10   | 0.55      | 0.35   | 0.80           | 0.30   | 0.55      | 0.15   |
| 31 | 0.85           | 0.10   | 0.80      | 0.25   | 0.25           | 0.10   | 0.35      | 0.10   |
| 32 | 1.00           | 0.57   | 1.21      | 1.00   | 1.07           | 0.78   | 1.50      | 1.42   |

Tabela 12: Tabela com os índices de placa bacteriana antes e depois da aplicação dos tratamentos para os dentes anteriores

## *APÊNDICE C – Modelos Multivariados para Dados Pré-Teste/Pós-Teste*

### C.1 Testes T - Pareados

Os Testes T são utilizados quando existem duas amostras tomadas de um mesmo indivíduo e que são naturalmente correlacionadas entre si de tal forma que a observação  $x_i$  da 1ª amostra é pareada com a observação  $y_i$  da segunda amostra (RENCHEER, 2002). O interesse é estudar a diferença  $d_i = x_i - y_i$ . No caso multivariado, os tratamentos são os mesmos, mas se apresentam mais variáveis simultaneamente.

A estrutura de dados para este planejamento é apresentada na tabela 13.

Suponha que  $\mathbf{x}$  e  $\mathbf{y}$  têm distribuição normal multivariada de forma que a matriz de variância-covariância apresente algum elemento que indique o relacionamento entre estas variáveis. No caso de um planejamento com medidas pré-teste/pós-teste, podemos considerar estes elementos da matriz de variância como sendo as correlações entre as medidas obtidas, tal que a medida  $x_i$ , representa a observação pós-teste e por conseqüência,  $y_i$ , a pré-teste. Não está no escopo desta dissertação a abordagem detalhada desta técnica, de forma que maiores referências estão disponíveis em livros de Análise Multivariada, como em Rencher (2002) e Johnson & Wichern (2002).

A desvantagem deste tipo de planejamento é que só se permite a inserção de uma única covariável com 2 níveis, que o torna muito restritivo.

Tabela 13: Estrutura do Teste T - Pareado Multivariado

| Indivíduo | Amostra 1      | Amostra 2      | $\mathbf{d}_i = \mathbf{x}_i - \mathbf{y}_i$ |
|-----------|----------------|----------------|--|
| 1         | $\mathbf{x}_1$ | $\mathbf{y}_1$ | $\mathbf{d}_1$                               |
| 2         | $\mathbf{x}_2$ | $\mathbf{y}_2$ | $\mathbf{d}_2$                               |
| $\vdots$  | $\vdots$       | $\vdots$       | $\vdots$                                     |
| n         | $\mathbf{x}_n$ | $\mathbf{y}_n$ | $\mathbf{d}_n$                               |

## C.2 Planejamento com Medidas Repetidas

Os planejamentos com medidas repetidas são realizados quando um indivíduo tem alguma característica medida em vários tratamentos diferentes ou em pontos diferentes do tempo, como define Suyama (1995). Os tratamentos podem ser resultados de testes, uso de medicamentos, estímulos e muitos outros. Se o indivíduo é medido em pontos sucessivos do tempo, então se pode estar interessado em determinar um polinômio de grau  $n$  de tal forma que uma curva possa ser ajustada para representar a forma de crescimento do indivíduo ou o comportamento da unidade experimental. Este tipo de técnica é chamado de **Curvas de Crescimento**.

Quando está se comparando médias dos tratamentos aplicados a cada indivíduo, está-se analisando, portanto, os fatores *intra-indivíduos*. Quando se dá a análise das médias dos grupos, dá-se o nome de análise *entre - indivíduos*. Nos planejamentos Pré-Teste/Pós-Teste é este tipo de abordagem em que se tem interesse: analisar o comportamento destas unidades experimentais entre os tratamentos e dentro do próprio tratamento.

Normalmente este planejamento é analisado como um modelo univariado chamado de **Split-Plot**, com as unidades experimentais aninhadas em um fator, digamos, B que faz o cruzamento com as repetições das medidas, podendo ser definido como um fator A. Este modelo, descrito desta maneira é representado como:

$$Y_{ijr} = \mu + B_i + S_{(i)j} + A_r + BA_{ir} + \epsilon_{ijr} \quad (\text{C.1})$$

Onde A, B e BA denotam os fatores envolvidos e o subscrito (i)j em S indica que os sujeitos estão aninhados ao fator B. A tabela 14 mostra a forma esquemática de apresentação deste modelo.

Os fatores A e B são considerados fixos e os sujeitos, aleatórios. Uma possível crítica a esta forma de abordar o problema, à primeira vista, é que um dos pressupostos da ANOVA é de que os elementos das amostras são independentes entre si. É muito difícil, que ao se obter a repetição da medida em uma mesma unidade experimental, este fato acima descrito aconteça, pois naturalmente existirá uma relacionamento entre as medidas tomadas consecutivamente. No entanto, pode-se mostrar (Wilks, 1946) que se a matriz da distribuição geradora dos dados tiver uma forma específica, a ANOVA, e por conseqüência, os testes F envolvidos serão válidos. A esta matriz dá-se o nome de *Matriz Simétrica Composta* (*Compound Symmetry*) e a sua forma é apresentada abaixo:

$$\text{Cov}(\mathbf{y}_{ij}) = \Sigma = \sigma^2 [(1 - \rho)\mathbf{I} + \rho\mathbf{J}] \quad (\text{C.2})$$

Para esta matriz, temos que  $\mathbf{I}$  é a matriz identidade de ordem  $n$  e  $\mathbf{J}$  é uma matriz quadrada

Tabela 14: Esquema de k-grupos com Medidas Repetidas

| Fator B        |                 | Fator A - Med. Rep. |                  |     |                    |                   |
|----------------|-----------------|---------------------|------------------|-----|--------------------|-------------------|
| (Grupo)        | Indivíduos      | A <sub>1</sub>      | A <sub>2</sub>   | ... | A <sub>p</sub>     |                   |
| B <sub>1</sub> | S <sub>11</sub> | (y <sub>111</sub>   | y <sub>112</sub> | ... | y <sub>11p</sub> ) | =y' <sub>11</sub> |
|                | S <sub>12</sub> | (y <sub>121</sub>   | y <sub>122</sub> | ... | y <sub>12p</sub> ) | =y' <sub>12</sub> |
|                | S <sub>13</sub> | (y <sub>131</sub>   | y <sub>132</sub> | ... | y <sub>13p</sub> ) | =y' <sub>13</sub> |
|                | ⋮               | ⋮                   | ⋮                | ⋮   | ⋮                  | ⋮                 |
| B <sub>2</sub> | S <sub>1n</sub> | (y <sub>1n1</sub>   | y <sub>1n2</sub> | ... | y <sub>1np</sub> ) | =y' <sub>1n</sub> |
|                | S <sub>21</sub> | (y <sub>211</sub>   | y <sub>212</sub> | ... | y <sub>21p</sub> ) | =y' <sub>21</sub> |
|                | S <sub>22</sub> | (y <sub>221</sub>   | y <sub>222</sub> | ... | y <sub>22p</sub> ) | =y' <sub>22</sub> |
|                | S <sub>23</sub> | (y <sub>231</sub>   | y <sub>232</sub> | ... | y <sub>23p</sub> ) | =y' <sub>23</sub> |
| ⋮              | ⋮               | ⋮                   | ⋮                | ⋮   | ⋮                  | ⋮                 |
|                | S <sub>2n</sub> | (y <sub>2n1</sub>   | y <sub>2n2</sub> | ... | y <sub>2np</sub> ) | =y' <sub>2n</sub> |
|                | ⋮               | ⋮                   | ⋮                | ⋮   | ⋮                  | ⋮                 |
|                | ⋮               | ⋮                   | ⋮                | ⋮   | ⋮                  | ⋮                 |
| B <sub>k</sub> | S <sub>k1</sub> | (y <sub>k11</sub>   | y <sub>k12</sub> | ... | y <sub>k1p</sub> ) | =y' <sub>k1</sub> |
|                | S <sub>k2</sub> | (y <sub>k21</sub>   | y <sub>k22</sub> | ... | y <sub>k2p</sub> ) | =y' <sub>k2</sub> |
|                | S <sub>k3</sub> | (y <sub>k31</sub>   | y <sub>k32</sub> | ... | y <sub>k3p</sub> ) | =y' <sub>k3</sub> |
|                | ⋮               | ⋮                   | ⋮                | ⋮   | ⋮                  | ⋮                 |
|                | S <sub>kn</sub> | (y <sub>kn1</sub>   | y <sub>kn2</sub> | ... | y <sub>knp</sub> ) | =y' <sub>kn</sub> |

preenchida por  $\mathbf{1}$ 's.

Como visto em (C.2), os efeitos aleatórios para cada indivíduo é constante e a variância entre indivíduos também é constante. A desvantagem, portanto, é que normalmente em dados Pré-Teste/Pós-Teste, a variabilidade entre - indivíduos não é constante, o que pode diminuir a flexibilidade do modelo e levar a inferências errôneas. Feita estas considerações, nosso modelo não mais ANOVA, mas sim, MANOVA e agora pode ser definido como:

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_{ij} \quad (\text{C.3})$$

Para o modelo em (C.3) temos que  $\mathbf{y}_{ij}$  é o vetor das medidas obtidas do sujeito  $S_{ij}$ ,  $\boldsymbol{\beta}_i$  é um vetor com os p efeitos principais para o fator B e  $\boldsymbol{\epsilon}_{ij}$  é o vetor aleatório dos erros, onde  $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ .

### C.2.1 Procedimento de Análise

Para analisar o modelo de medidas repetidas em (C.3), inicialmente testa-se o fator A, comparando-se as médias das variáveis  $y_i, i = 1, \dots, p$  contra as médias dos níveis de B, lembrando-se que as p variáveis correspondem aos níveis de A. O vetor de médias

$\mu_i, i = 1, \dots, k$  correspondem aos níveis do fator B e são estimados por  $\bar{y}_i, i = 1, \dots, k$ . Definimos também  $\bar{\mu} = \sum_{i=1}^k \mu_i / k$ . O estimador de  $\bar{\mu} = \bar{y}_{..} = \sum_{i=1}^k \bar{y}_i / k$ .

A estatística de teste é:

$$T^2 = N(\mathbf{C}\bar{\mathbf{y}}_{..})'(\mathbf{C}\mathbf{S}_{\mathbf{pl}}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}}_{..}) \quad (\text{C.4})$$

Onde a estatística em (C.4) é distribuída por uma  $T^2 \sim \frac{(N-1)(p-1)}{(N-p-1)} F_{p-1, N-p-1}$  (p medidas repetidas em  $N = \sum n_i$  indivíduos).  $\mathbf{S}_{\mathbf{pl}}$  é a matriz de variância estimada do modelo,  $\mathbf{S}_i = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (x_j - \bar{x})(x_j - \bar{x})'$  e  $\mathbf{S}_{\mathbf{pl}} = \frac{\sum_{i=1}^k (n_i-1)\mathbf{S}_i}{\sum_{i=1}^k n_i - k}$ .

A matriz  $\mathbf{C}$  forma o contraste entre os níveis dos fatores. No caso do modelo em (C.1), deve existir um  $\mathbf{C}_{\mathbf{A}}$ , contraste para testar os efeitos de  $\mathbf{A}$ ,  $\mathbf{C}_{\mathbf{B}}$ , para os efeitos de  $\mathbf{B}$  e  $\mathbf{C}_{\mathbf{BA}}$  para os efeitos de  $\mathbf{BA}$ . Estas matrizes devem ser ortogonais e dessa maneira, a estatística  $T^2$  fica independente de  $\mathbf{C}$ . A matriz  $\mathbf{C}$  em geral deve ter posto p-1.

### Exemplo

Suponha que estamos tratando de um modelo semelhante a (C.1) somente onde  $\mathbf{A}$  tem 2 níveis e  $\mathbf{B}$  também tem 2 níveis (p=2 repetições). Então, deseja-se testar:

•Fator A

$$\begin{cases} H_0 : \mu_1 + \mu_2 = \mu_3 + \mu_4 \\ H_1 : \mu_1 + \mu_2 \neq \mu_3 + \mu_4 \end{cases}$$

A matriz correspondente é dada por:

$$\mathbf{C}_{\mathbf{A}} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}$$

•Fator B

$$\begin{cases} H_0 : \mu_1 + \mu_4 = \mu_2 + \mu_3 \\ H_1 : \mu_1 + \mu_4 \neq \mu_2 + \mu_3 \end{cases}$$

A matriz correspondente é dada por:

$$\mathbf{C}_{\mathbf{B}} = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}$$

•Fator BA

A interação entre A e B é obtida pela multiplicação de cada linha de  $\mathbf{C}_A$  por cada linha de  $\mathbf{C}_B$ . Então, o resultado é:

$$\mathbf{C}_{BA} = \begin{bmatrix} 1 & -1 & 1 & -1 \end{bmatrix}$$

Para finalizar os testes, basta verificar se as estatísticas  $T^2$  obtidas são significantes em determinado nível de confiança de  $100(1 - \alpha)\%$ .

## *APÊNDICE D – Programas em R*

```
#####Algoritmo E-M para o modelo de SAHU
# Parâmetros
# beta, s2u,s2e,rho
#
#Função Lambda
#Calcula o lambda em função dos dos parâmetros
lambda=function(rho,s2u,s2e)
{
  lamb=rho*s2u/s2e
  return(lamb)
}
#####
#Função esp.U
# Calcula a Esperança do vetor
# variáveis latentes no 1° e 2° momentos.
esp.U=function(beta,rho,s2u,s2e,X,Y)
  {
mu=(Y-as.vector(X*%beta))
var.e=rho^2*s2u+s2e
lamb=lambda(rho,s2u,s2e)
#Esperança do primeiro momento
espu=(rho*sqrt(s2u)*mu)/var.e+
(dnorm(lamb*mu/sqrt(var.e),0,1)/
pnorm(lamb*mu/sqrt(var.e),0,1))*sqrt(s2e/var.e)
#Esperança do segundo momento
espu2= s2u*(rho*mu/var.e)^2+s2e/var.e+
(dnorm(lamb*mu/sqrt(var.e),0,1)/pnorm(lamb*mu/sqrt(var.e),0,1))*
rho*sqrt(s2u*s2e)*mu/var.e^(3/2)
res=list(U=espu,U2=espu2)
return(res)
}
####Cálculo da Log-verossimilhança
logver=function(beta,rho,s2u,s2e,e.u,e.u2,X,Y)
{
mu=as.vector(X*%beta)
n=length(X)
lv=-n/2*log(s2e)-(1/(2*s2e))*(sum((Y-mu)^2)-
```

```

2*sum((Y-mu)*e.u)*rho*sqrt(s2u)**
rho^2*s2u*sum(e.u2))-1/2*sum(e.u2)
return(lv)
}
#####
#Algoritmo propriamente dito
#0 programa a seguir realiza os cálculos
#necessários para a estimação dos parâmetros
alg.EM=function(beta,rho,s2u,s2e,X,Y,tol=1.0E-5,iterm=1000)
{
t0=c(rho,s2u,s2e)
t1=numeric(0)
fat=1
j=0
while(tol<fat & iterm>j)
{
#Passo E
e.u=esp.U(beta,t0[1],t0[2],t0[3],X,Y)
#Passo M
#Atualização de beta
beta.i=as.vector(solve(t(X)%*%X)%*%t(X)%*(Y-t0[1]*sqrt(t0[2])*e.u$U))
mu=as.vector(Y-X%*%beta.i)
#Atualização de rho
t1[1]=sum(mu*e.u$U)/(sqrt(t0[2])*sum(e.u$U2))
#Atualização de s2u
t1[2]=(sum(mu*e.u$U)^2)/(t0[1]*sum(e.u$U2))^2
#Atualização de s2e
#
t1[3]=(sum(mu)^2-2*t1[1]*sqrt(t1[2])*sum(mu*e.u$U)+
t1[1]^2*t1[2]*sum(e.u$U2))/nrow(X)
#Cálculo da logver
e.u1=esp.U(beta.i,t1[1],t1[2],t1[3],X,Y)
lv0=logver(beta,t0[1],t0[2],t0[3],e.u$U,e.u$U2,X,Y)
lv1=logver(beta.i,t1[1],t1[2],t1[3],e.u1$U,e.u1$U2,X,Y)
fat=abs((lv1-lv0)/lv1)
beta=beta.i
t0=t1;j=j+1
}
return(list(beta=beta,rho=t0[1],s2u=t0[2],s2e=t0[3],logv=lv1,iter=j))
}

#### Bootstrap
indiv=unique(usr$ind)
usr=usr
N=500
param=list(beta=matrix(0,nc=8,nr=N),rho=numeric(N),
s2u=numeric(N),s2e=numeric(N))
for(j in 1:N)

```

```
{
  samp=sample(indiv,24,replace=T)
  for( i in 1:24)
  {
    usr[c(2*i-1,2*i),]=usar[usar$ind==samp[i],]
  }
  X=model.matrix(~sexo:esc+sexo:esc:lant-1,usr)
  Y=usr$ldep
  rho=-10
  s2e=1
  s2u=0.1
  beta0=rep(0,8)
  eps=1
  while(eps>1.0E-6)
  {
    a=esp.U(beta0,rho,s2u,s2e,X,Y)
    e.u=a$U
    e.u2=a$U2
    rm(a)
    l0=logver(beta0,c(rho,s2u,s2e),e.u,e.u2,X,Y)
    est=optim(par=c(rho,s2u,s2e),fn=logver,X=X,Y=Y,
    e.u=e.u,e.u2=e.u2,beta=beta0)
    rho=est$par[1]
    s2e=est$par[3]
    s2u=est$par[2]
    beta0=as.vector(solve(t(X)%*%X)%*%t(X)%*%(Y-rho*sqrt(s2u)*e.u))
    l1=logver(beta0,c(rho,s2u,s2e),e.u,e.u2,X,Y)
    eps=abs((l1-l0)/l0)
  }
  param$beta[j,]=beta0
  param$rho[j]=rho
  param$s2u[j]=s2u
  param$s2e[j]=s2e
}
```

## *APÊNDICE E - Programas em WinBUGS*

```

#### Modelo Bayesiano ####
#Priori t-student para Rho e tau=precisão da t-student#

Modelo
{
for(i in 1:N)
{
for(j in 1:nc)
{
mu.i[i,j]<-b[j]*(X[i,j])
}
}
#Efeitos Aleatórios
for(i in 1:32)
{
eta[i]~dnorm(0,1)I(0,)
}
for(i in 1:N)
{
for(j in 1:32)
{
ea.i[i,j]<-eta[j]*(ea[i,j])
}
}

#Construção da verossimilhança
for(i in 1:N)
{
mu[i]<-sum(mu.i[i,])+rho*sqrt(s2eta)*sum(ea.i[i,])
y[i]~dnorm(mu[i],tau2ni)
}
#Prioris não-informativas
#prioris dos coeficientes
for(i in 1:nc)
{
b[i]~dnorm(0,1.0E-6)
}

```

```
correc<-0.7978846*lambda/sqrt(1+pow(lambda,2))*sqrt(s2)
for(j in 1:4)
{
bc[j]<-b[j]+correc
}
#Modelo 1
#Prioris precisão
#tau2ni~dgamma(1.0E-3,1.0E-3)
#tau2eta~dgamma(1.0E-3,1.0E-3)
#s2ni<-1/tau2ni
#s2eta<-1/tau2eta

#Modelo 2
tau2ni~dgamma(1.0E-3,1.0E-3)
s2ni<-1/tau2ni
#Priori para o modelo 2 - Priori de Kummer não informativa
tau2r~dgamma(1.0E-2,1.0E-2)
s2eta<-2.467401*tau2r*s2ni

#lambda e S2
lambda<-rho*sqrt(s2eta/s2ni)
s2<-s2ni+pow(rho,2)*s2eta

#locacao da t
#loc<-2.467*s2ni/s2eta

#Geração da t com 1/2 graus de liberdade (Metropolis-Hastings)
zero<-0
rho~dflat()
phi<--0.5*log(1/tau2r)+(3/4)*log(1+2*pow(rho,2)*tau2r)
zero~dpois(phi)
}
```

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)