



Universidade de Brasília
IE - Instituto de Exatas
Departamento de Estatística

Modelos Lineares Mistos: Uma Abordagem Bayesiana

Alex Luiz Martins Matheus da Rocha

Dissertação de Mestrado

Orientador: Professor Eduardo Yoshio Nakano

Brasília
2017

Sumário

Lista de Figuras	iv
Lista de Tabelas	vi
Resumo	vii
Abstract	viii
1 Introdução	1
2 Referencial Teórico	3
2.1 Modelos Lineares Generalizados - MLG	3
2.1.1 Regressão Linear	4
2.2 Modelos Mistos	5
2.2.1 Modelo de Coeficientes Aleatórios	6
2.2.2 Modelos Lineares Generalizados Mistos - Ligação Canônica	10
2.3 Predição dos efeitos aleatórios (BLUP)	15
2.4 Inferência Bayesiana	17
3 Metodologia	20
3.1 Verossimilhança, priori e posteriori	21
3.2 Estimativa dos parâmetros	22
3.3 BLUP para o intercepto	23

3.4	Construção da reta predita	24
4	Aplicação em conjunto de dados	25
4.1	Aplicação - 2 níveis hierárquicos.	25
4.1.1	Ajuste do Modelo Aleatório - dois níveis	28
4.1.2	Ajuste do Modelo de Regressão Hierárquica - dois níveis	35
4.2	Aplicação - três níveis hierárquicos	40
4.2.1	Ajuste do Modelo Aleatório - três níveis	43
5	Conclusão	48
	Referências Bibliográficas	50

Lista de Figuras

2.1	Resultados invertidos pela Abordagem Usual e com Efeitos Mistos .	6
4.1	<i>Boxplot</i> das variáveis Extroversão, Popularidade e Popularidade dentro de turmas	27
4.2	Dispersão das variáveis Popularidade e Extroversão	28
4.3	MCMC - Traço e densidade empírica dos parâmetros do Modelo Aleatório	30
4.4	MCMC - Traço e densidade empírica do ICC (ρ)	31
4.5	Intervalo de Credibilidade dos BLUP - Efeito Predito das Turmas .	33
4.6	Intervalo de Credibilidade dos BLUP - Intercepto Médio das Turmas	34
4.7	MCMC - Traço e Densidade Empírica dos Parâmetros da Regressão Hierárquica	36
4.8	Reta Predita pelo Modelo	37
4.9	Intervalo de Credibilidade dos BLUP - Efeito Predito da Turma - Modelo com Regressor	39
4.10	Retas Ajustadas com BLUP	40
4.11	<i>Boxplot</i> da variável Estresse, da média do estresse em cada enfermaria e Estresse dentro de Hospitais	42
4.12	MCMC - Traço e densidade empírica dos parâmetros do Modelo Aleatório 3 níveis	44
4.13	Ajuste do ICC (2.8)	45
4.14	Ajuste do ICC (2.9)	46

4.15 Ajuste do ICC (2.10)	47
-------------------------------------	----

Lista de Tabelas

4.1	Descrição das Turmas - Fonte [J.J.Hox, 2010]	26
4.2	Estimativas Clássicas e Bayesianas dos Parâmetros do Modelo Ale- atorio	29
4.3	BLUP para o intercepto	32
4.4	Estimativas Clássicas e Bayesianas dos Parâmetros da Regressão Hierárquica	35
4.5	BLUP para o intercepto - Modelo com Regressor	38
4.6	Descrição dos Hospitais. Fonte [J.J.Hox, 2010]	41
4.7	Estimativas Clássicas e Bayesianas dos Parâmetros do Modelo Ale- atorio 3 níveis	43

Resumo

Estudos em que a população de interesse possui estrutura hierárquica ou multinível são cada vez mais frequentes nas áreas de educação e saúde, onde, por exemplo, tem-se o desejo de avaliar determinada característica de alunos dentro de escolas ou pacientes dentro de hospitais. Nessa situação, modelos hierárquicos são mais adequados do que modelos que não levam em consideração a hierarquia.

Esses modelos incorporam a estrutura de dependência dos dados, tornando as estimativas mais realistas e não viesadas. Esses modelos fazem parte da classe de modelos mistos, que possui efeitos fixos e mais de um efeito aleatório em sua composição. Este trabalho apresenta aplicações do modelo de regressão linear hierárquica, utilizando a abordagem bayesiana para estimação dos parâmetros.

Concluiu-se que esses modelos apresentam ganhos expressivos nas estimativas intervalares dos parâmetros, sem desrespeitar os pressupostos teóricos de um modelo mais simples, proporcionando estimativas não viesadas.

Palavras-chave: Regressão Multinível, Modelos Mistos, Modelos Hierárquicos, Inferência Bayesiana.

Abstract

Studies in which the population of interest has a multilevel or hierarchical structure are increasingly frequent in the areas of education and health, when one has the desire to evaluate a certain characteristic of students clustered within schools or patients clustered within hospitals. In this situation, hierarchical models are more appropriate than models that do not take hierarchy into account.

These models incorporate data dependency structure, making estimates more realistic and unbiased. These models are also called mixed models, containing both fixed effects and more than one random effect in its composition. This work presents applications of the linear hierarchical regression model, using bayesian methods of estimation.

It was concluded that these models present expressive gains in the interval estimates of the parameters, without disrespecting the assumptions of a simpler model, providing unbiased estimates.

Keywords: Multilevel Regression, Mixed Models, Hierarchical Models, Bayesian Inference.

Capítulo 1

Introdução

Os modelos estatísticos permitem uma representação aproximada da realidade sobre um certo conjunto de variáveis, a partir de uma amostra. Dentre esses modelos, a regressão é utilizada quando o foco do estudo é a relação entre uma variável resposta e uma ou mais variáveis explicativas. Para fazer inferências sobre os parâmetros do modelo, é necessário um conjunto de pressupostos teóricos que descrevem a distribuição de probabilidade que melhor representa os dados, bem como outras suposições como a independência entre as observações e homocedasticidade.

No entanto, é possível que nem todos os pressupostos sejam satisfeitos, o que pode levar à conclusões incorretas, estimativas viesadas, erro padrão sub ou superestimado, entre outros. A suposição de independência, presente nos modelos de regressão linear, por exemplo, pode não representar a realidade em casos que a população tem estrutura hierárquica ou nos casos em que a amostra for obtida por meio de uma amostragem complexa, por conglomerados ou estratificada. Nesses casos, indivíduos ou variáveis dentro de um mesmo grupo podem apresentar características semelhantes, indicando que não há independência entre eles.

Por esse motivo, modelos que incorporam a estrutura hierárquica da população e as dependências entre as observações são importantes para a inferência estatística. Os modelos hierárquicos, ou aninhados, fazem parte de uma classe de

modelos chamada de modelos com efeitos mistos que são mais flexíveis quanto às suas suposições. Nesse caso, a amostragem não necessariamente precisa ser simples e as observações não precisam ser independentes.

Dentro dos modelos mistos, temos os modelos lineares generalizados mistos (MLGM), que são MLG com mais de um efeito aleatório. Um desses modelos é a regressão hierárquica, cuja construção é similar à um modelo bayesiano. Muitos softwares estatísticos, como SPSS, SAS e MlwiN possuem procedimentos para a estimação clássica dos parâmetros desses modelos. No entanto, procedimentos para a abordagem bayesiana são mais difíceis de se encontrar, sendo que esta abordagem traz vantagens na inferência dos parâmetros do modelo, que serão discutidas nesse trabalho, em troca de procedimentos computacionalmente intensivos, por meio de simulações.

Nesse contexto, o objetivo desse trabalho é fazer a estimação bayesiana dos parâmetros do MLGM e comparar com os resultados da estimação clássica, bem como verificar significância de parâmetros e diagnóstico. Será considerado o caso da regressão linear hierárquica e o uso software R para obtenção dos resultados.

A metodologia apresentada nesse trabalho foi ilustrada por meio de dois conjuntos de dados da literatura [[J.J.Hox, 2010](#)], na qual o autor apresentou a análise dos dados apenas pela abordagem clássica. O primeiro conjunto possui uma estrutura hierárquica de dois níveis e se refere à alunos dentro de escolas. O segundo, se refere à uma estrutura hierárquica de três níveis, na qual temos enfermeiras dentro de enfermarias dentro de hospitais.

Capítulo 2

Referencial Teórico

2.1 Modelos Lineares Generalizados - MLG

Uma das formas de analisar o padrão de associação e interação entre uma variável de interesse, denominada variável resposta, e outras variáveis, denominadas variáveis explicativas, é por meio de modelos estatísticos. Os parâmetros determinam a intensidade e a importância dos efeitos e inferências podem ser feitas sobre esses parâmetros para avaliar quais variáveis realmente afetam a resposta. Os valores preditos pelo modelo melhoram a estimativa da média da variável resposta nos possíveis valores das variáveis explicativas.

Uma classe de modelos muito importante na modelagem estatística é a de Modelos Lineares Generalizados (*MLG*), que envolve uma variável resposta univariada, variáveis explicativas e n observações independentes, levando em consideração uma distribuição de probabilidade que pertença à família exponencial. Esta classe é caracterizada por 3 componentes:

i) Componente Aleatório

É a especificação da distribuição de probabilidade da variável resposta Y , que deve pertencer à família exponencial de distribuições.

ii) Componente Sistemático

Especifica as variáveis explicativas (x_1, \dots, x_k) , por meio de uma combinação linear, denominada preditor linear, que é dada por:

$$\alpha + \beta_1 x_1 + \dots + \beta_k x_k.$$

Interações entre variáveis explicativas, por exemplo $x_k = x_1 x_2$, são permitidas.

iii) Função de Ligação

Especifica a função $g(\cdot)$, monótona e diferenciável, que relaciona $\mu = E(Y)$ com o preditor linear, isto é, conecta os componentes aleatório e sistemático:

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k.$$

Os MLG permitem explicar a resposta Y em função de variáveis explicativas e são muito flexíveis devido à possibilidade da utilização de diferentes distribuições para a variável resposta e diferentes formas para a função de ligação. Dessa maneira, pode-se selecionar um modelo que representa melhor os dados e com maior precisão.

2.1.1 Regressão Linear

O caso mais simples de MLG é a regressão linear, utilizada para uma variável resposta numérica, que considera a ligação canônica, ou identidade, $g(\mu) = \mu = E(Y)$ e Y com distribuição normal. Sua forma, seguindo a notação de MLG, é dada por:

$$\mu(x_1, \dots, x_k) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k. \quad (2.1)$$

Em (2.1), o parâmetro β_k indica que para um acréscimo de 1 unidade em x_k , a média de Y aumenta em β_k unidades e α é o intercepto, isto é, se todos os valores de x forem 0, Y é, em média, α .

A Equação (2.2) mostra a notação usual desse modelo, na qual $\epsilon \sim N(0, \sigma^2)$. Observe que aplicando a esperança em ambos os lados de (2.2), obtém-se (2.1).

$$Y = \alpha + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon. \quad (2.2)$$

2.2 Modelos Mistos

Um modelo estatístico que possui mais de um termo ou efeito aleatório, além dos efeitos fixos, é chamado de modelo com efeitos mistos ou simplesmente modelo misto. Os casos discutidos na Seção (2.1) possuíam apenas um efeito aleatório. Na Equação (2.2), o único efeito aleatório é o erro entre indivíduos ϵ que tem distribuição normal. Os demais parâmetros α e β_k são efeitos fixos.

A decisão sobre considerar um efeito como fixo ou aleatório não é trivial nem única e depende muito do estudo que está sendo feito, do interesse do pesquisador ou do contexto dos dados. Por exemplo, o delineamento em blocos casualizados é usualmente tratado com um modelo de apenas um efeito aleatório, que é o erro entre as observações. No entanto, podemos considerar o efeito dos blocos como aleatório e proceder com a análise inferencial usando a teoria de modelos mistos.

Nas próximas seções serão apresentados alguns casos de modelos com mais de um efeito aleatório, com foco em modelos de regressão hierárquica, que são um caso particular dos modelos lineares generalizados mistos (MLGM).

Um caso onde essa abordagem produz resultados completamente contrários ao modelo usual pode ser vista em [Demidenko, 2004]. É possível que uma regressão linear usual tenha inclinação negativa, enquanto que uma regressão hierárquica apresenta uma inclinação positiva, conforme figura abaixo. Isso acontece pois o modelo hierárquico incorpora a estrutura de variância e covariância dos dados, com suposições mais flexíveis e realistas.

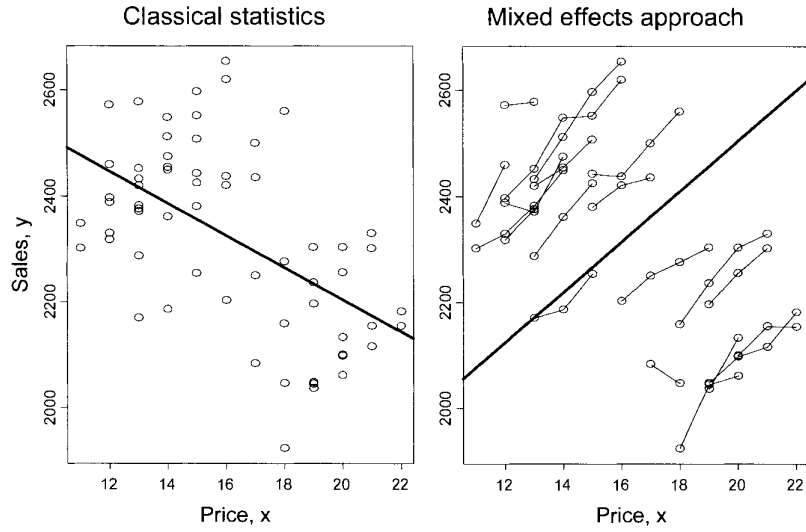


Figura 2.1: Resultados invertidos pela Abordagem Usual e com Efeitos Mistos

2.2.1 Modelo de Coeficientes Aleatórios

Também conhecido como modelo nulo, modelo de componentes da variância, ou simplesmente modelo aleatório, o modelo de coeficientes aleatórios tem grande importância em modelos hierárquicos. Nesse caso, possui apenas o intercepto como efeito fixo e dois efeitos aleatórios: um que representa a diferença entre indivíduos e outro que representa a diferença entre grupos. Sua utilidade é estimar as variâncias para calcular o coeficiente de correlação intraclasse.

Considere o caso de uma estrutura hierárquica com dois níveis, por exemplo alunos dentro de turmas. A variável resposta, denotada por y , deve estar no menor nível, que nesse caso é o nível aluno. Seja $i = 1, \dots, I$, o número de grupos ou *clusters*, e $j = 1, \dots, n_i$ é o número de observações em cada grupo. O modelo é dado por:

$$y_{ij} = \mu + G_i + e_{ij}. \quad (2.3)$$

Em (2.3), y_{ij} é a j -ésima observação do grupo i ; μ é o intercepto que é um efeito

fixo; $e_{ij} \sim N(0, \sigma_E^2)$ é o erro no menor nível, que indica diferença de indivíduos, dentro de grupos; e $G_i \sim N(0, \sigma_G^2)$ é o erro no maior nível, que indica diferença entre grupos. Este efeito pode ser visto como uma reparametrização considerando o intercepto como aleatório, isto é $\mu_i = \mu + G_i \sim N(\mu, \sigma_G^2)$ e, portanto, $\mu_i - \mu = G_i \sim N(0, \sigma_G^2)$. Os efeitos aleatórios são independentes entre si.

Tem-se que (2.3) não explica a variabilidade de y , apenas a decompõe em dois termos independentes: e_{ij} com variância σ_E^2 e G_i com variância σ_G^2 , que são chamados de componentes da variância.

Pela independência entre os efeitos aleatórios, tem-se que $Cov(y_{ij}, y_{kl}) = 0$, ou seja, indivíduos em grupos diferentes ($i \neq k$) possuem covariância zero. Para indivíduos dentro do mesmo grupo, ($j \neq l$), $Cov(y_{ij}, y_{il}) = \sigma_G^2$. O coeficiente de correlação intraclasses ρ é dado por:

$$\rho = \frac{Cov(y_{ij}, y_{il})}{\sqrt{Var(y_{ij})Var(y_{il})}} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}. \quad (2.4)$$

Assim, ρ é a proporção da variância explicada pela estrutura de agrupamento na população. Também pode ser interpretado como a correlação esperada entre dois indivíduos escolhidos aleatoriamente dentro do mesmo grupo. Quanto maior for esse coeficiente, maior a evidência de que a estrutura hierárquica deve ser levada em conta no modelo com covariáveis.

A notação matricial facilita o entendimento do modelo. Esta é dada por:

$$\mathbf{Y} = \mathbf{1}\mu + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi} \quad (2.5)$$

- $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}\sigma_G^2)$; $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{I}\sigma_E^2)$; independentes entre si.
- $Var(\mathbf{Y}) = \mathbf{Z}\mathbf{Z}'\sigma_G^2 + \mathbf{I}\sigma_E^2 = \mathbf{V}$
- $\mathbf{Y} \sim N(\mathbf{1}\mu, \mathbf{V})$

Por exemplo, considerando $I = 2$ grupos com n_i observações cada, 2 para o grupo 1 e 3 para o grupo 2, $i = 1, 2$ e $j = 1, \dots, n_i$, temos:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{23} \end{bmatrix}.$$

Com a matriz de variância e covariância de \mathbf{Y} dada por:

$$\mathbf{V} = \begin{bmatrix} \sigma_G^2 + \sigma_E^2 & \sigma_G^2 & 0 & 0 & 0 \\ \sigma_G^2 & \sigma_G^2 + \sigma_E^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_G^2 + \sigma_E^2 & \sigma_G^2 & \sigma_G^2 \\ 0 & 0 & \sigma_G^2 & \sigma_G^2 + \sigma_E^2 & \sigma_G^2 \\ 0 & 0 & \sigma_G^2 & \sigma_G^2 & \sigma_G^2 + \sigma_E^2 \end{bmatrix}.$$

Com a suposição de normalidade para \mathbf{Y} , a função de verossimilhança é dada por:

$$L(\mu, \mathbf{V} | \mathbf{Y}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{Y}-\mathbf{1}\mu)' \mathbf{V}^{-1}(\mathbf{Y}-\mathbf{1}\mu)\right)}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}}. \quad (2.6)$$

E a log-verossimilhança é:

$$l(\mu, \mathbf{V} | \mathbf{Y}) = l = -\frac{1}{2}(\mathbf{Y}-\mathbf{1}\mu)' \mathbf{V}^{-1}(\mathbf{Y}-\mathbf{1}\mu) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{V}|).$$

A extensão para três ou mais níveis é análoga, mas bastante complicada, especialmente pelo fato de termos muitos parâmetros. Nesse caso, coeficientes de regressão das variáveis explicativas no menor nível podem variar em ambos segundo e terceiro níveis, além de que os coeficientes de regressão do segundo nível podem variar no terceiro nível.

Para evitar um modelo super parametrizado, não se deve incluir interações de alta ordem, a não ser que isso seja importante no estudo.

Pode-se também definir o coeficiente de correlação intra-classe. Considerando

o modelo com 3 níveis e sem variáveis explicativas, sendo que agora i indica o número de grupos no terceiro nível, j indica o número de grupos no segundo nível e k o número de observações.

$$y_{ijk} = \mu + G_{1i} + G_{2j} + e_{ijk}. \quad (2.7)$$

Agora G_{1i} representa o erro do terceiro nível com variância $\sigma_{G_1}^2$ e G_{2i} representa o erro do segundo nível com variância $\sigma_{G_2}^2$. Em notação matricial, temos $\mathbf{Y} \sim N(\mathbf{1}\mu, \mathbf{V})$. Mas aqui, \mathbf{V} é tal que $var(y_{ijk}) = \sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2$, $cov(y_{ijk}, y_{ijk'}) = \sigma_{G_1}^2 + \sigma_{G_2}^2$, $cov(y_{ijk}, y_{ij'k'}) = \sigma_{G_1}^2$ e $cov(y_{ijk}, y_{i'j'k'}) = 0$. Assim, temos dois métodos para calcular ρ (Algina, 2000):

i) Método 1 (Davis & Scott, 1995)

Para o nível 2:

$$\rho_2 = \frac{\sigma_{G_2}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2}. \quad (2.8)$$

Para o nível 3:

$$\rho_3 = \frac{\sigma_{G_1}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2}. \quad (2.9)$$

ii) Método 2 (Siddiqui, Hedeker, Flay & Hu, 1996)

Para o nível 2:

$$\rho_2 = \frac{\sigma_{G_1}^2 + \sigma_{G_2}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2}. \quad (2.10)$$

Para o nível 3:

$$\rho_3 = \frac{\sigma_{G_1}^2}{\sigma_{G_1}^2 + \sigma_{G_2}^2 + \sigma_E^2}. \quad (2.11)$$

O método 1 identifica a proporção da variância explicada no nível 2 e nível 3. O segundo método representa uma estimativa para a correlação esperada entre dois elementos escolhidos aleatoriamente dentro do mesmo grupo. O método a ser utilizado depende da interpretação desejada.

2.2.2 Modelos Lineares Generalizados Mistos - Ligação Canônica

Este modelo também é conhecido como regressão hierárquica ou regressão multinível e permite examinar relações entre variáveis medidas em diferentes níveis de dados com essa estrutura. Para utilizá-lo, é necessário ter, além de uma estrutura hierárquica na população, uma variável resposta quantitativa medida no menor nível. Pode-se ter variáveis explicativas em quaisquer um dos níveis.

Considerando o caso de uma estrutura hierárquica com 2 níveis, assumindo que temos P variáveis explicativas x no menor nível, indexadas por $p = 1, \dots, P$. Similarmente, temos Q variáveis explicativas w no maior nível, indexadas por $q = 1, \dots, Q$. A variável resposta, que está no menor nível, é denotada por y . Além disso, $i = 1, \dots, I$, é o número de grupos ou *clusters*, e $j = 1, \dots, n_i$ é o número de observações em cada grupo e $n = \sum_{i=1}^I n_i$ é o total de observações. A equação do modelo completo é dada por:

$$y_{ij} = \mu + \sum_{p=1}^P \beta_p x_{pij} + \sum_{q=1}^Q \gamma_q w_{qi} + \sum_{p=1}^P \sum_{q=1}^Q \theta_{pq} x_{pij} w_{qj} + \sum_{p=1}^P \tau_{pi} x_{pij} + G_i + e_{ij}. \quad (2.12)$$

Claramente 2.12 possui muitos parâmetros. Em regressão hierárquica é ainda mais importante que sejam incluídas apenas as variáveis e interações que forem de fato importantes para o estudo, pois esse tipo de modelo pode ficar facilmente super parametrizado. Em (2.12), temos:

i) Efeitos fixos:

y_{ij} é a resposta do j -ésimo indivíduo do grupo i .

x_{ij} é o valor da variável explicativa, de nível indivíduo, do j -ésimo indivíduo do grupo i .

w_{ij} é o valor da variável explicativa, de nível grupo, do grupo i -ésimo grupo.

μ é o intercepto.

β_p é o coeficiente de regressão da p -ésima variável explicativa no menor nível.

γ_q é o coeficiente de regressão da q -ésima variável explicativa no maior nível.

θ_{pq} é o coeficiente de regressão da interação entre níveis (*Cross-level interaction*).

ii) Efeitos aleatórios:

τ_{pi} é o termo aleatório que indica se o coeficiente de regressão para o preditor x_p varia entre grupos.

G_i é o erro no maior nível, que indica diferença entre grupos.

e_{ij} é o erro no menor nível, que indica diferença dentro de grupos.

iii) Suposições do modelo:

Suposição inicial é de relação linear entre a variável resposta e as explicativas.

$e_{ij} \sim N(0, \sigma_E^2)$ com variância constante (Homocedasticidade).

τ_{pi} e G_i são independentes de e_{ij} e tem distribuição normal multivariada com média 0. $\text{Var}(G_i) = \sigma_G^2$ é a variância do erro entre grupos. $\text{Var}(\tau_{pi}) = \sigma_{\tau_p}^2$ é a variância dos coeficientes de regressão entre grupos. Em geral, as covariâncias entre G_i e τ_{pi} não são assumidas 0.

As interpretações dos parâmetros do modelo (2.12) são análogas à regressão linear usual, onde, por exemplo, para cada unidade da variável x_p temos um aumento (no caso do coeficiente β_p ser positivo e não termos interação) de y , em média, de β_p unidades, mantidas as demais variáveis constantes. Na presença de interação, a interpretação depende de outra variável explicativa.

A notação usual de um modelo misto é:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi}.$$

$\mathbf{X}\boldsymbol{\beta}$ é a matriz de efeitos fixos, $\mathbf{Z}\boldsymbol{\alpha}$ é a matriz de efeitos aleatórios e $\boldsymbol{\xi}$ é o vetor dos erros, também aleatórios. Tem-se que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, onde $\mathbf{V} = \mathbf{Z}Var(\boldsymbol{\alpha})\mathbf{Z}' + \mathbf{I}\sigma_E^2$. Observe que no caso em que apenas os grupos são aleatórios, $\mathbf{V} = \mathbf{Z}\mathbf{Z}'\sigma_G^2 + \mathbf{I}\sigma_E^2$, conforme visto em 2.5.

O modelo de regressão linear multinível se encaixa nessa notação. Primeiramente considerando o modelo com todos os efeitos fixos, mas sem coeficientes de regressão aleatórios, dado por:

$$y_{ij} = \mu + \sum_{p=1}^P \beta_p x_{pij} + \sum_{q=1}^Q \gamma_q w_{qi} + \sum_{p=1}^P \sum_{q=1}^Q \theta_{pq} x_{pij} w_{qj} + G_i + e_{ij}. \quad (2.13)$$

A equação (2.13) segue direto de (2.12). Considerando os vetores de efeitos fixos $\mathbf{1}$, \mathbf{x} representando x_{pij} , \mathbf{w} representando w_{qi} e \mathbf{k} para as interações $x_{pij}w_{qj}$, a notação matricial é dada por:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1} & \mathbf{x} & \mathbf{w} & \mathbf{k} \end{bmatrix} \begin{bmatrix} \mu \\ \boldsymbol{\beta}_p \\ \boldsymbol{\gamma}_q \\ \boldsymbol{\theta}_{pq} \end{bmatrix} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi}. \quad (2.14)$$

Assim, em (2.14) temos a mesma notação matricial do modelo de componentes da variância para os efeitos aleatórios. Para os efeitos fixos, todos são juntados na mesma matriz $\mathbf{X}\boldsymbol{\beta}$.

Novamente considerando o exemplo em que se tem $I = 2$ grupos com n_i observações cada, 2 para o grupo 1 e 3 para o grupo 2, $i = 1, 2$ e $j = 1, \dots, n_i$, com duas variáveis explicativas x_1 e x_2 no menor nível e uma variável w_1 no maior nível, e considerando interação entre níveis com as variáveis x_1 e w_1 , temos que $p = q = 1$ e a notação matricial baseada em (2.13) e (2.14) é:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & x_{111} & x_{211} & w_{11} & x_{111}w_{11} \\ 1 & x_{112} & x_{212} & w_{11} & x_{112}w_{11} \\ 1 & x_{121} & x_{221} & w_{12} & x_{121}w_{12} \\ 1 & x_{122} & x_{222} & w_{12} & x_{122}w_{12} \\ 1 & x_{123} & x_{223} & w_{12} & x_{123}w_{12} \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \gamma_1 \\ \theta_{11} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{23} \end{bmatrix}$$

A inclusão de um regressor β aleatório complica a matriz de variância e covariância \mathbf{V} . Suponha agora que temos um efeito β aleatório. Cada grupo possui um regressor β_i e β é o regressor médio de todos os grupos. Temos que $\beta_i = (\beta + \tau_i)x_{ij} \sim N(\beta, \sigma_\tau^2)$ e portanto $\tau \sim N(0, \sigma_\tau^2)$ e conforme (2.12), o modelo se torna:

$$y_{ij} = \mu\beta x_{ij} + \tau_i x_{ij} + G_i + e_{ij} \quad (2.15)$$

Naturalmente, esperamos que τ_i e G_i não sejam independentes, pois quando ajustamos uma reta aos dados, uma modificação no intercepto pode impactar a inclinação. Por exemplo, se temos dados em que tanto a inclinação quanto o intercepto são positivos, aumentar um implica em reduzir o outro, caso contrário, a reta não se ajusta aos dados. Portanto, tem-se que $Cov(G_i, \tau_i) = \sigma_{\tau G}$. Observe que ainda temos uma estrutura bloco diagonal de matriz de variância e covariância, pois $Cov(G_i, \tau_{i'}) = 0$, para $i \neq i'$.

Para ilustrar esse caso, considere novamente $I = 2$ grupos com n_i observações cada, $n_1 = 2$ para o grupo 1 e $n_2 = 3$ para o grupo 2, $i = 1, 2$ e $j = 1, \dots, n_i$. Suponha que temos apenas uma variável explicativa x no menor nível. A notação matricial resultante, baseada em (2.14) é dada por:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \end{bmatrix} \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \begin{bmatrix} 1 & 0 & x_{11} & 0 \\ 1 & 0 & x_{12} & 0 \\ 0 & 1 & 0 & x_{21} \\ 0 & 1 & 0 & x_{22} \\ 0 & 1 & 0 & x_{23} \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{23} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\xi}$$

Nesse caso, $\mathbf{V} = \text{Var}(\mathbf{Z}\boldsymbol{\alpha}) + \mathbf{I}\sigma_E^2$. Observe que:

$$\mathbf{Z}\boldsymbol{\alpha} = \begin{bmatrix} G_1 + \tau_1 x_{11} \\ G_1 + \tau_1 x_{12} \\ G_2 + \tau_2 x_{21} \\ G_2 + \tau_2 x_{22} \\ G_2 + \tau_2 x_{23} \end{bmatrix}$$

Logo, $\text{Cov}(G_i + \tau_i x_{ij}, G_i + \tau_i x_{ij}) = \text{Var}(G_i) + \text{Var}(\tau_i x_{ij}) + 2\text{Cov}(G_i, \tau_i x_{ij})$ para a diagonal. Para as demais entradas dentro da matriz bloco diagonal V_i do grupo i , $\text{Cov}(G_i + \tau_i x_{ij}, G_i + \tau_i x_{ij'}) = \text{Var}(G_i) + \text{Cov}(\tau_i x_{ij}, \tau_i x_{ij'}) + \text{Cov}(G_i, \tau_i x_{ij}) + \text{Cov}(G_i, \tau_i x_{ij'})$. As demais entradas são iguais à 0, pois vimos que $\text{Cov}(G_i, \tau_{i'}) = 0$, para $i \neq i'$. Portanto, \mathbf{V} é dada por:

$$\mathbf{V} = \begin{bmatrix} V_1 & \mathbf{0} \\ \mathbf{0} & V_2 \end{bmatrix} + \mathbf{I}\sigma_E^2 ; \quad V_1 = \begin{bmatrix} \sigma_G^2 + \sigma_\tau^2 x_{11}^2 + 2x_{11}\sigma_{\tau G} & \sigma_G^2 + \sigma_\tau^2 x_{11}x_{12} + (x_{11} + x_{12})\sigma_{\tau G} \\ \sigma_G^2 + \sigma_\tau^2 x_{12}^2 + 2x_{12}\sigma_{\tau G} \end{bmatrix}$$

$$V_2 = \begin{bmatrix} \sigma_G^2 + \sigma_\tau^2 x_{21}^2 + 2x_{21}\sigma_{\tau G} & \sigma_G^2 + \sigma_\tau^2 x_{21}x_{22} + (x_{21} + x_{22})\sigma_{\tau G} & \sigma_G^2 + \sigma_\tau^2 x_{21}x_{23} + (x_{21} + x_{23})\sigma_{\tau G} \\ \sigma_G^2 + \sigma_\tau^2 x_{22}^2 + 2x_{22}\sigma_{\tau G} & \sigma_G^2 + \sigma_\tau^2 x_{22}x_{23} + (x_{22} + x_{23})\sigma_{\tau G} \\ \sigma_G^2 + \sigma_\tau^2 x_{23}^2 + 2x_{23}\sigma_{\tau G} \end{bmatrix}$$

Observe agora que a matriz \mathbf{V} é mais complexa, pois inclui duas variâncias, uma covariância e as observações x_{ij} . Além disso, $\sigma_{\tau G}$ pode ser negativo, o que

pode dificultar algoritmos de otimização, caso não seja incluída uma restrição. Observe que a correlação entre dois indivíduos dentro do mesmo grupo é dada por:

$$Cor(x_{ij}, x_{ij'}) = \frac{\sigma_G^2 + \sigma_\tau^2 x_{ij} x_{ij'} + (x_{ij} + x_{ij'}) \sigma_{\tau G}}{\sqrt{\sigma_G^2 + \sigma_\tau^2 x_{ij}^2 + 2x_{ij} \sigma_{\tau G} + \sigma_E^2} \sqrt{\sigma_G^2 + \sigma_\tau^2 x_{ij'}^2 + 2x_{ij'} \sigma_{\tau G} + \sigma_E^2}}. \quad (2.16)$$

A Expressão 2.16 implica que há uma restrição para a métrica entre os parâmetros σ_G^2 , σ_τ^2 , $\sigma_{\tau G}$ e σ_E^2 . De fato, essa restrição é dada por:

1. $\sigma_G^2 + \sigma_\tau^2 x_{ij}^2 + 2x_{ij} \sigma_{\tau G} + \sigma_E^2 > 0$
2. $\sigma_G^2 + \sigma_\tau^2 x_{ij'}^2 + 2x_{ij'} \sigma_{\tau G} + \sigma_E^2 > 0$
3. $|Cor(x_{ij}, x_{ij'})| \leq 1, \forall i, j$

2.3 Predição dos efeitos aleatórios (BLUP)

Uma propriedade relevante nos modelos mistos é a possibilidade de fazer predições individuais mesmo que a quantidade de parâmetros estimados seja maior que o número de grupos. Considere o modelo:

$$y_{ij} = \mu + \beta x_{ij} + G_i + e_{ij}, \quad (2.17)$$

o qual representa uma regressão hierárquica com apenas uma variável explicativa no menor nível. Como discutido na Seção 2.2.1, G_i pode ser visto como uma reparametrização $\mu_i = \mu + G_i \sim N(\mu, \sigma_G^2)$ e, portanto, $\mu_i - \mu = G_i \sim N(0, \sigma_G^2)$. Logo, cada grupo possui um intercepto μ_i e o efeito fixo μ representa o intercepto médio entre todos os grupos. O modelo misto permite a predição desse efeito aleatório μ_i , ou seja, é possível obter um valor para o intercepto de cada grupo. Isso

significa que se tivermos 10 grupos, podemos 'estimar' um efeito para cada um dos 10 grupos, além dos outros parâmetros do modelo.

O estimador do efeito aleatório, na realidade, é chamado de preditor, pois não é exatamente um estimador. Esse preditor é conhecido como melhor preditor linear não viesado (*Best linear unbiased predictor - BLUP*).

Para se obter os BLUP do efeito aleatório de grupos G_i , é necessário calcular o valor esperado de μ_i dado y_i , ou seja, ajusta-se o modelo (2.17) e com as estimativas dos efeitos fixos e os resíduos, calcula-se:

$$E(\boldsymbol{\mu}_i|Y_i) = Cov(\boldsymbol{\mu}_i, Y_i)Cov^{-1}(Y_i)(Y_i - X_i\boldsymbol{\beta}) = \sigma_G^2 Z_i(\sigma_G^2 + I_i\sigma_E^2)^{-1}(Y_i - X_i\boldsymbol{\beta}). \quad (2.18)$$

Em (2.18), $\boldsymbol{\mu}_i$ é o efeito aleatório do grupo i , Y_i é a resposta y no grupo i e X_i é a variável explicativa no grupo i . Assim, o BLUP é dado pela esperança condicional do efeito aleatório, dado os dados e o modelo ajustado. O BLUP também pode ser calculado para qualquer outro efeito aleatório. Assim, pode-se fazer um gráfico que inclui as diferentes retas ajustadas para cada grupo, bem como prever o valor esperado de um novo indivíduo de qualquer grupo.

Observa-se que a Equação (2.18) resulta no efeito individual que cada grupo tem sobre o intercepto geral, que é uma média dos grupos. Dessa maneira, obtemos o intercepto individual de cada grupo somando o efeito daquele grupo com o intercepto médio dos grupos $\hat{\mu}$, conforme equação abaixo.

$$\hat{\mu}_i = E(\boldsymbol{\mu}_i|Y_i) + \hat{\mu}. \quad (2.19)$$

O modelo (2.17) supõe que os interceptos variam entre grupos, mas os regressores não variam. Isso significa que as retas ajustadas com o BLUP são paralelas. Um modelo como esse pode ser aplicado no caso que os grupos apresentam incrementos iguais, isto é, o regressor é o mesmo para todos os grupos, mas o intercepto varia.

Nota-se que, se a variância $\hat{\sigma}_G^2$ do efeito aleatório G_i não for significativa, os

efeitos $\hat{\mu}_i$ serão muito próximos. Logo, é possível fazer um diagnóstico do modelo após o cálculo dos BLUP, pois se $\hat{\mu}_i$ não apresentar diferença entre grupos, temos um indício de que não há variabilidade entre esses grupos. O mesmo vale para qualquer regressor que for considerado como aleatório dentro do modelo.

É importante observar que o modelo de regressão hierárquica (2.20) com coeficientes aleatórios

$$y_{ij} = \mu_i + \beta x_{ij} + G_i + e_{ij}, \quad (2.20)$$

não é equivalente à (2.17). De fato, (2.20) considera que o efeito de grupo é aleatório, mas, no entanto, também considera os interceptos como se fossem efeitos fixos. Seguindo os mesmos passos de estimação do modelo (2.17), obtemos que σ_G^2 e suas variâncias são 0, o que gera um erro na estimação. Empiricamente, pode-se observar que o resultado para as estimativas pontuais desse modelo é muito próximo de se ajustar uma regressão linear para separadamente para cada grupo.

Caso tenhamos medidas repetidas para cada indivíduo, (2.20) pode ser ajustado ([Demidenko, 2004]).

2.4 Inferência Bayesiana

Na teoria de probabilidade, o teorema de Bayes permite obter a probabilidade de um evento a partir de um conhecimento ou informação a priori que esteja relacionado com esse evento. Além disso, também é possível utilizá-lo para se obter distribuições de probabilidade de uma certa variável. Dessa maneira, dizemos que as probabilidades são atualizadas conforme mais informação são obtidas. Sua forma matemática para dois eventos A e B é descrita por:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.21)$$

A abordagem bayesiana para inferência estatística segue diretamente de (2.21). Na inferência clássica, assumimos que o parâmetro θ é fixo e desconhecido, mas no

paradigma bayesiano assumimos que o parâmetro que se deseja estimar é aleatório e atribuímos à ele uma distribuição a priori, que reflete a informação conhecida sobre esse parâmetro. Obtida uma amostra que contenha informações sobre esse parâmetro, podemos atualizar sua distribuição usando (2.21). A distribuição a posteriori, para um parâmetro θ dado a observação da variável aleatório Y , é dada por:

$$h(\theta|Y) = \frac{h(Y, \theta)}{h(Y)} = \frac{h(Y|\theta)h(\theta)}{\int h(Y, \theta)d\theta}. \quad (2.22)$$

Em (2.22), o denominador $\int h(Y, \theta)d\theta$ não depende de θ , portanto é apenas uma constante normalizadora, que torna $h(\theta|Y)$ uma distribuição de probabilidade. Além disso, $h(Y|\theta) = L(\theta; Y)$ é a função de verossimilhança. Logo podemos dizer que:

$$h(\theta|Y) \propto L(\theta; Y)h(\theta). \quad (2.23)$$

Em (2.23), temos que a posteriori $h(\theta|Y)$ é proporcional ao produto da verossimilhança $L(\theta; Y)$ com a distribuição a priori $h(\theta)$. Observe que o suporte da distribuição a priori precisa ser o mesmo do parâmetro θ .

A inferência bayesiana satisfaz o princípio da verossimilhança, que diz que para uma constante c que não depende de θ , duas verossimilhanças proporcionais $L(\theta; Y) = cL(\theta; Y)$ devem proporcionar a mesma inferência sobre θ .

As estimativas de interesse seguem diretamente da distribuição a posteriori. O estimador de bayes é o valor que minimiza o risco, isto é, o valor esperado a posteriori de uma função de perda $l(\hat{\theta}, \theta)$,

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} E(l(\hat{\theta}, \theta)|Y) = \int_{\Theta} l(\hat{\theta}, \theta)h(\theta|Y)d\theta. \quad (2.24)$$

Tem-se que (2.24) depende da função perda escolhida. Nos casos mais utilizados, o estimador de bayes é:

- i) Para a perda quadrática: $l(\hat{\theta}, \theta) = [\hat{\theta} - \theta]^2$, $\hat{\theta} = E(\theta|Y)$ a média a posteriori.
- ii) Para a perda absoluta $l(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$, $\hat{\theta} = Med(\theta|Y)$ a mediana a posteriori.
- iii) Para a perda 0-1:

$$l(\hat{\theta}, \theta) = \begin{cases} 1 & \text{se } |\hat{\theta} - \theta| > \epsilon; \\ 0 & \text{Caso contrário.} \end{cases}$$

Tem-se que $\hat{\theta} = Moda(\theta|Y)$ a moda a posteriori.

A estimativa pontual não dá uma ideia do quanto o parâmetro varia em torno daquele valor. Uma analogia aos intervalos de confiança clássicos é dada pelas regiões R de credibilidade, onde para $0 < \gamma < 1$:

$$P(\theta \in R|Y) = \int_R h(\theta|Y) d\theta \geq \gamma. \quad (2.25)$$

Em (2.25), dado um γ , a região de credibilidade R não é única. Algumas propriedades são interessantes para esse intervalo, como por exemplo comprimento mínimo. Alguns métodos utilizados para se chegar em um intervalo adequado são:

- i) O intervalo onde os pontos tem maior densidade de probabilidade (HPD, *highest posterior density*). Isso significa que, dentro desse intervalo, qualquer ponto tem densidade maior ou igual à outros pontos fora desse intervalo. Dessa maneira, este é o intervalo com menor comprimento.
- ii) O intervalo simétrico é tal que a probabilidade abaixo é tão provável quanto acima, isto é, para (c_1, c_2) , $P(\theta < c_1|Y) = \frac{\gamma}{2}$ e $P(\theta > c_2|Y) = \frac{\gamma}{2}$.
- iii) Se a média existir, é possível criar um intervalo cuja média está no centro.

Como consideramos que o parâmetro é aleatório, dado um intervalo de credibilidade para θ , tal que (2.25) seja satisfeito, temos que a probabilidade do parâmetro estar no intervalo é de $100\gamma\%$.

Capítulo 3

Metodologia

Duas aplicações foram consideradas neste trabalho, ambas utilizando conjuntos de dados presentes em [J.J.Hox, 2010], com estrutura hierárquica e variável resposta quantitativa, para aplicar os modelos discutidos no Capítulo 2. O primeiro possui uma estrutura hierárquica de dois níveis e se refere à alunos dentro de escolas. O segundo, se refere à uma estrutura hierárquica de três níveis, na qual temos enfermeiras dentro de enfermarias dentro de hospitais.

Antes do ajuste do modelo, foi realizada uma análise descritiva dos dados, para apresentar as principais características das variáveis de interesse e como essas se comportam dentro da estrutura hierárquica.

Na modelagem, primeiramente ajustou-se o modelo aleatório para calcular o coeficiente de correlação intraclasse ρ (2.4), que indica a necessidade ou não da utilização do modelo hierárquico. Verificada essa necessidade, ajustou-se o modelo com variáveis explicativas. Para os dados das enfermeiras, apenas o modelo aleatório foi ajustado.

A estimação clássica foi feita pelo método de máxima verossimilhança. As estimativas bayesianas foram obtidas conforme descrito a seguir.

Para o resultado final do modelo, considerou-se as estimativas bayesianas.

3.1 Verossimilhança, priori e posteriori

Na Seção 2.4 foi dito que é necessário uma distribuição a priori para se fazer inferência bayesiana sobre os parâmetros desejados. Para todos os parâmetros, uma priori difusa foi utilizada, isto é, uma distribuição com alta variância, caracterizada por pouca precisão a respeito de quais valores o parâmetro pode assumir. Amostras da distribuição a posteriori foram obtidas por meio do comando `MCMCmetrop1R` do pacote 'MCMCPack' do R (R core team, 2017), que utiliza um algoritmo MCMC (*Markov Chain Monte Carlo*).

As amostras obtidas no MCMC foram de tamanho 15000 para os dados com dois níveis, sendo as 5000 primeiras descartadas (*burnin*) e de 20000 para três níveis, com *burnin* maior de 10000, pois a cadeia não se movimentava entre as iterações 5000 à 8000.

i) Modelo Aleatório - dois níveis

Conforme exposto na Seção 2.2.1, o modelo aleatório dado por (2.5) é tal que $\mathbf{Y} \sim N(\mathbf{1}\mu, \mathbf{V})$, uma normal multivariada, e função de verossimilhança igual a (2.6). Para μ utilizou-se uma normal com média 0 e variância 1000. Para os componentes da variância, utilizou-se a distribuição gama com parâmetros iguais a 10^{-3} .

Dessa maneira, para o vetor de parâmetros $\theta' = (\mu, \sigma_G^2, \sigma_E^2)$, a posteriori, é dada por

$$h(\theta|Y) \propto h(\mu)h(\sigma_G^2)h(\sigma_E^2)L(\theta; Y),$$

sendo $L(\theta; Y) = L(\mu, \mathbf{V}|\mathbf{Y})$ em (2.6). O método de estimação utilizado apresentou boa convergência para os parâmetros μ e σ_E^2 mas σ_G^2 não convergiu bem para 10 ou menos grupos, mas os grupos em si podiam ter tamanhos pequenos. Foi necessário calibrar a taxa de aceitação do algoritmo, de 80% inicialmente, para 40% posteriormente. Os intervalos de credibilidade utilizados foram os HPD. O método delta foi utilizado para se obter um intervalo

de confiança para ρ , de modo a comparar com o resultado bayesiano.

ii) MLGM - Ligação Canônica - dois níveis

Para os parâmetros β do modelo (2.12) também foi usada uma normal com média 0 e variância 1000. Tem-se que $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V})$, logo a posteriori é dada da mesma maneira que o modelo aleatório, para $\theta' = (\mu, \beta_1, \dots, \beta_k, \sigma_G^2, \sigma_E^2)$

$$h(\theta|Y) \propto h(\mu)h(\beta_1)\dots h(\beta_k)h(\sigma_G^2)h(\sigma_E^2)L(\theta; Y),$$

na qual $L(\theta; Y) = L(\mu, \mathbf{V}|\mathbf{Y})$ em 2.6 com $1\mu=\mathbf{X}\beta$. O algoritmo MCMC convergiu bem para os parâmetros de interesse, sem apresentar problemas para β_k .

iii) Modelo Aleatório - três níveis

Para o modelo de três níveis, somente o modelo aleatorio foi considerado, de forma que $\mathbf{Y} \sim N(\mathbf{1}\mu, \mathbf{V})$, também normal multivariada, com matriz \mathbf{V} conforme modelo (2.7). Foi utilizado uma priori difusa gama com parâmetros iguais a 10^{-3} , para as três variâncias $\sigma_{G_1}^2$, $\sigma_{G_2}^2$ e σ_E^2 e para o intercepto μ , uma normal com média 0 e variância 1000. Portanto, para $\theta' = (\mu, \sigma_{G_1}^2, \sigma_{G_2}^2, \sigma_E^2)$, a posteriori é dada por:

$$h(\theta|Y) \propto h(\mu)h(\sigma_{G_1}^2)h(\sigma_{G_2}^2)h(\sigma_E^2)L(\theta; Y)$$

3.2 Estimativa dos parâmetros

Para a abordagem clássica, utilizou-se o método de máxima verossimilhança em ambos modelos e o método delta para o intervalo de confiança do coeficiente de correlação intraclasse (icc) ρ , no modelo aleatório.

Para a abordagem bayesiana, utilizou-se a média a posteriori como estimativa pontual e o intervalo de credibilidade HPD, conforme descrito na Seção 2.4. A

posteriori de ρ foi obtida a partir das posteriores de σ_G^2 e σ_E^2 , calculando-se (2.4) para cada vetor de resultados do MCMC.

Todos os intervalos de confiança e credibilidade utilizados utilizaram um $\alpha = 95\%$.

Com a posteriori, também verificou-se a significância dos parâmetros, em particular para ρ , pois este indica se a diferença entre grupos justifica o uso de um modelo hierárquico. Nesse trabalho, consideramos que, em uma estrutura hierárquica de dois níveis, a estimativa pontual $\rho \geq 20\%$ já é uma indicação de que o modelo multinível é adequado, contanto que este seja significativamente diferente de zero.

3.3 BLUP para o intercepto

Utilizou-se a fórmula descrita na Seção 2.18 para ambas abordagens, tanto no modelo aleatório quanto no modelo com regressores. Para a abordagem clássica, a fórmula é aplicada diretamente após o ajuste do modelo. Já para a abordagem bayesiana, é necessário calcular o BLUP para cada resultado da cadeia gerada pelo MCMC. Feito isso, temos a posteriori do efeito individual de cada grupo. Com a posteriori do efeito médio dos grupos, pela fórmula (2.19), obtemos a posteriori do intercepto individual de cada grupo. Finalmente, o estimador de bayes e os intervalos de credibilidade são calculados.

Observa-se que, para se obter os BLUP pela abordagem bayesiana, é necessário calcular os resíduos e a inversa da matriz de variância e covariância, para cada vetor de resultados do MCMC, o que pode ser bastante extensivo, principalmente quando a amostra e o número de grupos é grande.

3.4 Construção da reta predita

Após a estimação dos parâmetros do modelo, a reta predita foi construída com base no intervalo da amplitude da variável explicativa. Suponha que X seja uma variável explicativa e sua amplitude no intervalo $[a,b]$. Podemos dividir esse intervalo em 1000 ou 10000 partes equidistantes e para cada parte, calcula-se a reta $\mu + X\beta$ para cada um dos valores da cadeia MCMC.

Feito isso, para cada ponto do intervalo temos uma posteriori da reta predita. O valor da reta para esse ponto é a média a posteriori. Da mesma maneira, calcula-se o HPD para cada ponto e traça-se a reta com limites superiores e inferiores.

Capítulo 4

Aplicação em conjunto de dados

Como mencionado no Capítulo 3, dois conjuntos de dados de [J.J.Hox, 2010] foram considerados. O primeiro possui uma estrutura hierárquica de dois níveis e se refere à alunos dentro de escolas. A variável resposta numérica considerada foi a popularidade do aluno e tomou-se como variável explicativa a sua extroversão, também numérica. O segundo conjunto de dados se refere à uma estrutura hierárquica de três níveis, na qual temos enfermeiras dentro de enfermarias dentro de hospitais. Nesse caso, a variável resposta numérica é o estresse da enfermeira.

4.1 Aplicação - 2 níveis hierárquicos.

Os dados consistem de 2000 alunos dentro de 100 turmas. Por praticidade, considerou-se uma amostra de $I = 25$ turmas, resultando em $n = 516$ alunos. A Tabela 4.1 descreve o comportamento médio das variáveis Y popularidade (resposta) e X extroversão (explicativa) em cada turma.

Tabela 4.1: Descrição das Turmas - Fonte [J.J.Hox, 2010]

Variável	n_i	\bar{y}_i	\bar{x}_i	$Cor_i(y,x)$	
Turmas	1	20	5,075	4,850	0,515
	2	23	5,434	4,782	0,519
	3	17	4,188	4,294	0,566
	4	20	6,140	5,100	0,465
	5	17	4,888	5,882	0,320
	6	26	5,134	4,807	0,524
	7	20	5,505	4,650	0,501
	8	20	6,345	5,050	0,538
	9	22	5,259	5,227	0,545
	10	19	4,552	5,736	0,470
	11	20	4,840	5,050	0,140
	12	21	4,061	4,809	0,620
	13	17	4,876	5,176	0,343
	14	20	4,770	5,400	0,580
	15	21	4,385	6,476	0,765
	16	19	5,778	4,421	0,584
	17	20	5,710	4,800	0,163
	18	25	5,316	5,000	0,472
	19	21	5,838	5,047	0,262
	20	21	4,304	5,523	0,645
	21	18	6,133	5,222	0,344
	22	22	3,963	6,000	0,668
	23	23	4,556	4,869	0,472
	24	21	3,480	4,761	0,865
	25	23	4,852	4,739	0,599
Geral	25	516	5,011	5,103	0,324

Na Tabela 4.1, temos que a média geral de Y é 5,011 e de X é 5,103. Pode-se verificar que Y apresenta médias diferentes nas turmas, o que nos dá indícios de que o modelo multinível pode ser adequado, ou seja, há diferença entre grupos. Além disso, a correlação entre X e Y é de 0,324. Mas, olhando individualmente a correlação dentro de cada turma, podemos observar que esta é maior que a correlação geral em 21 dos 25 grupos, o que pode não só indicar que há diferença do efeito de X em cada turma, mas também que a inclusão dos grupos na estrutura

de variância e covariância de Y pode influenciar o efeito de X .

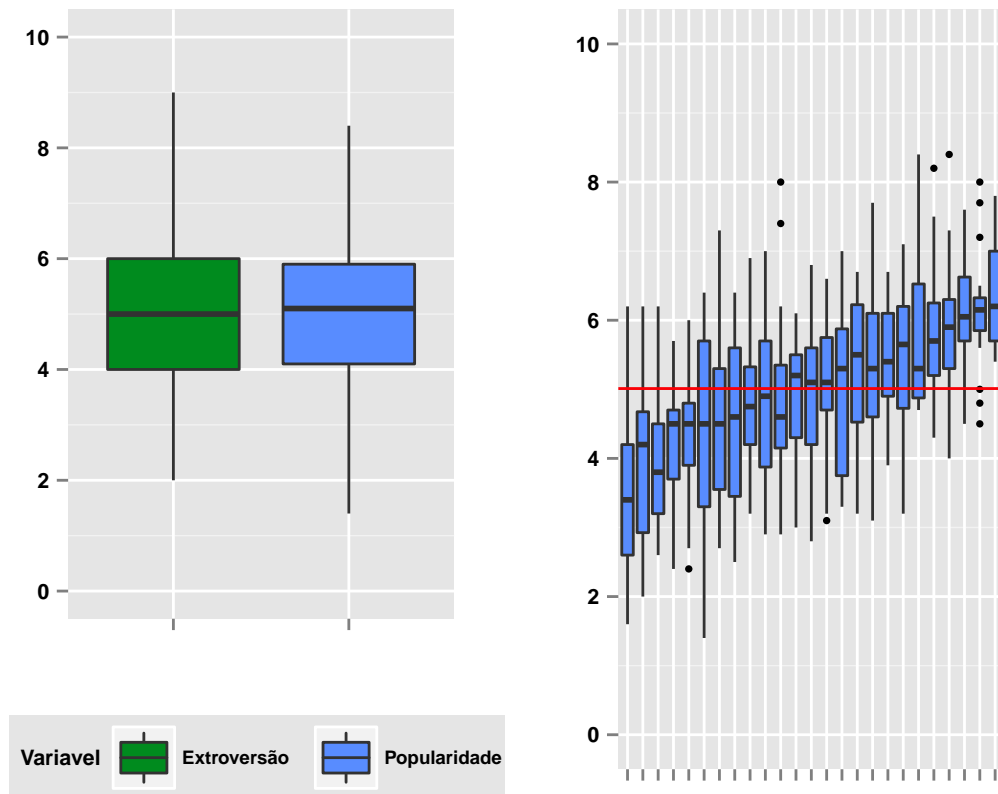


Figura 4.1: *Boxplot* das variáveis Extroversão, Popularidade e Popularidade dentro de turmas

A Figura 4.1 apresenta os *Boxplot* das variáveis consideradas. Pode-se observar que 75% das observações, tanto de X quanto de Y , estão em torno de 4 à 6. À direita da mesma figura, tem-se os *Boxplot* de Y , para cada turma, ordenados da menor média de Y dentro daquele grupo até a maior média, sendo a reta horizontal vermelha a média geral. Pode-se observar que Y está distribuído de uma maneira razoavelmente diferente em cada grupo, o que também indica que o modelo multinível pode ser adequado.

Antes de ajustar o modelo de regressão hierárquica, convém ajustar a regressão

linear usual como uma análise tanto preliminar quanto comparativa, para se ter uma ideia da relação entre X e Y e das diferenças entre o multinível e o usual.

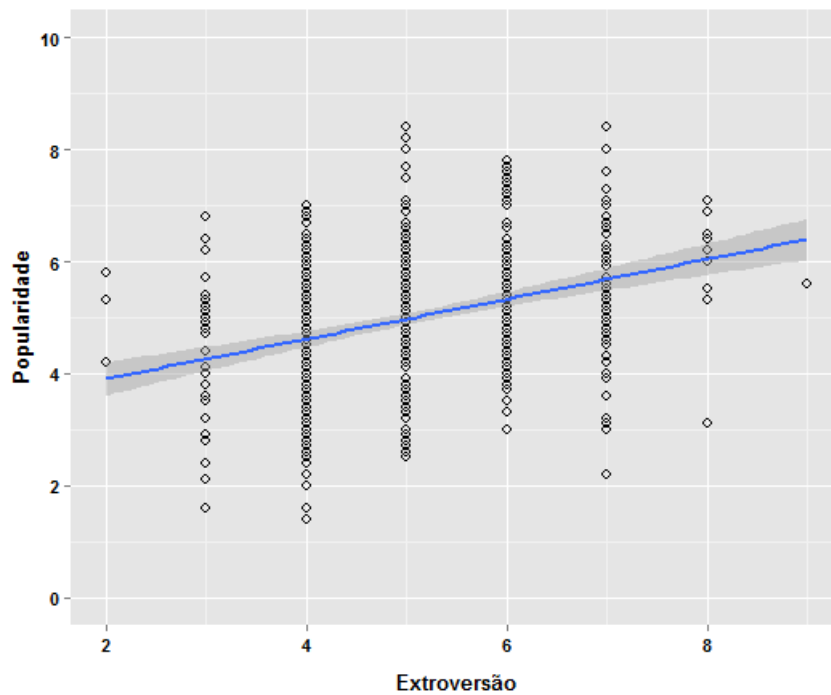


Figura 4.2: Dispersão das variáveis Popularidade e Extroversão

Na Figura 4.2 temos o ajuste da reta de regressão linear, com $\mu = 3,193$ e $\beta = 0,356$, indicando que quanto maior X , maior será a resposta Y .

4.1.1 Ajuste do Modelo Aleatório - dois níveis

A Tabela 4.2 apresenta as estimativas dos parâmetros do modelo (2.3) para ambas abordagens clássica e bayesiana.

Tabela 4.2: Estimativas Clássicas e Bayesianas dos Parâmetros do Modelo Aleatório

Abordagem	μ (IC 95%)	σ_G^2 (IC 95%)	σ_E^2 (IC 95%)	ρ (IC 95%)
Clássica (EMV)	5,015 (4,731 ; 5,298)	0,466 (0,175 ; 0,757)	1,174 (1,027 ; 1,321)	0,284 (0,154 ; 0,414)
Bayesiana (Média Post.)	5,014 (4,710 ; 5,295)	0,523 (0,226 ; 0,854)	1,175 (1,028 ; 1,320)	0,302 (0,173 ; 0,432)

Verifica-se que não houve muita diferença entre as estimativas clássica e bayesiana para μ e σ_E^2 . Para σ_G^2 , observou-se uma leve diferença entre as abordagens: 0,466 para o clássico e 0,523 para o bayesiano, refletindo também nos intervalos de confiança e credibilidade. No entanto, essa diferença não impactou fortemente a correlação intraclasses ρ com intervalo de confiança de (0,154; 0,414) e de credibilidade (0,173; 0,432).

Utilizando as estimativas bayesianas como resultado, tem-se que a proporção da variabilidade de Y explicada pelo agrupamento da população é de $\rho = 30\%$, indicando que nessa situação um modelo multinível é mais adequado para explicar a variabilidade de Y . Conforme discutido na Seção 2.2.1, a importância desse modelo para a regressão hierárquica é justamente verificar essa necessidade.

Na estimação bayesiana, a densidade a posteriori dos parâmetros é obtida numericamente via MCMC, desta forma, é necessário verificar a convergência da cadeia. A Figura 4.3 apresenta o traço e a densidade empírica dos parâmetros do modelo aleatório.

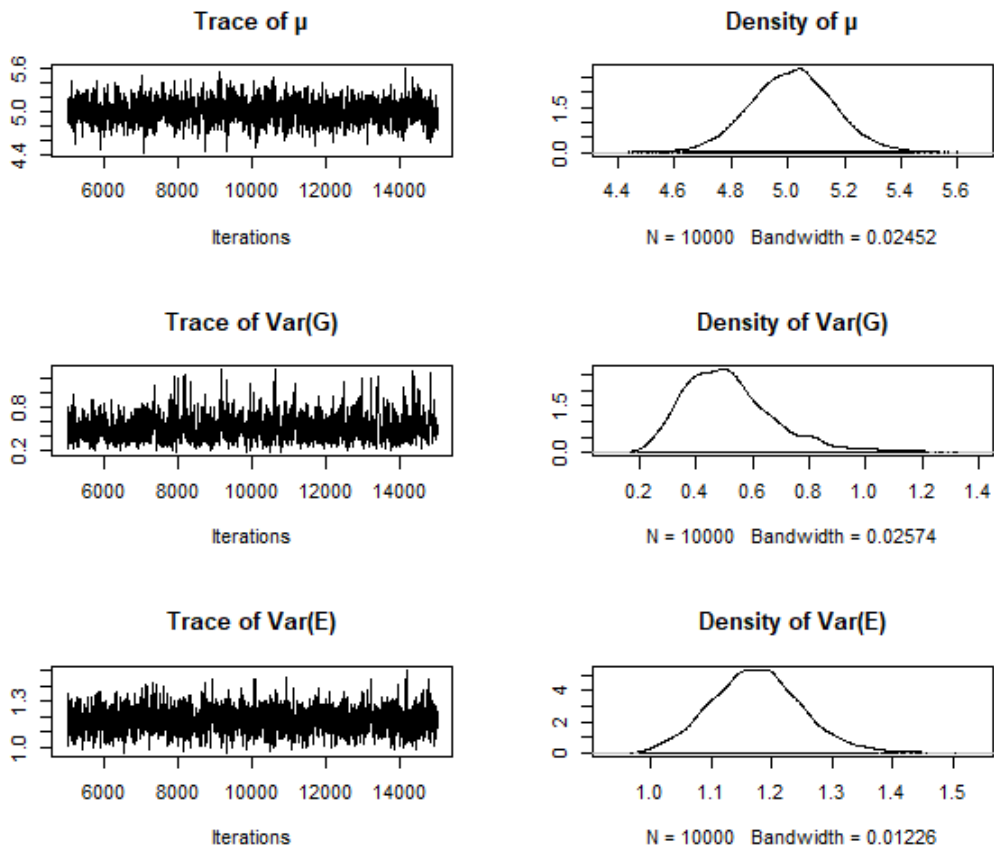


Figura 4.3: MCMC - Traço e densidade empírica dos parâmetros do Modelo Aleatório

Na Figura 4.3, mostrando na ordem μ , σ_G^2 e σ_E^2 , o traço indica uma forma próxima de um ruído branco, indicando que o algoritmo convergiu bem para esses parâmetros. Dessa maneira, a amostra obtida pelo MCMC pôde ser usada como uma amostra da distribuição a posteriori dos parâmetros de interesse. O mesmo verificou-se para o coeficiente de correlação intraclasse ρ , conforme Figura 4.4.

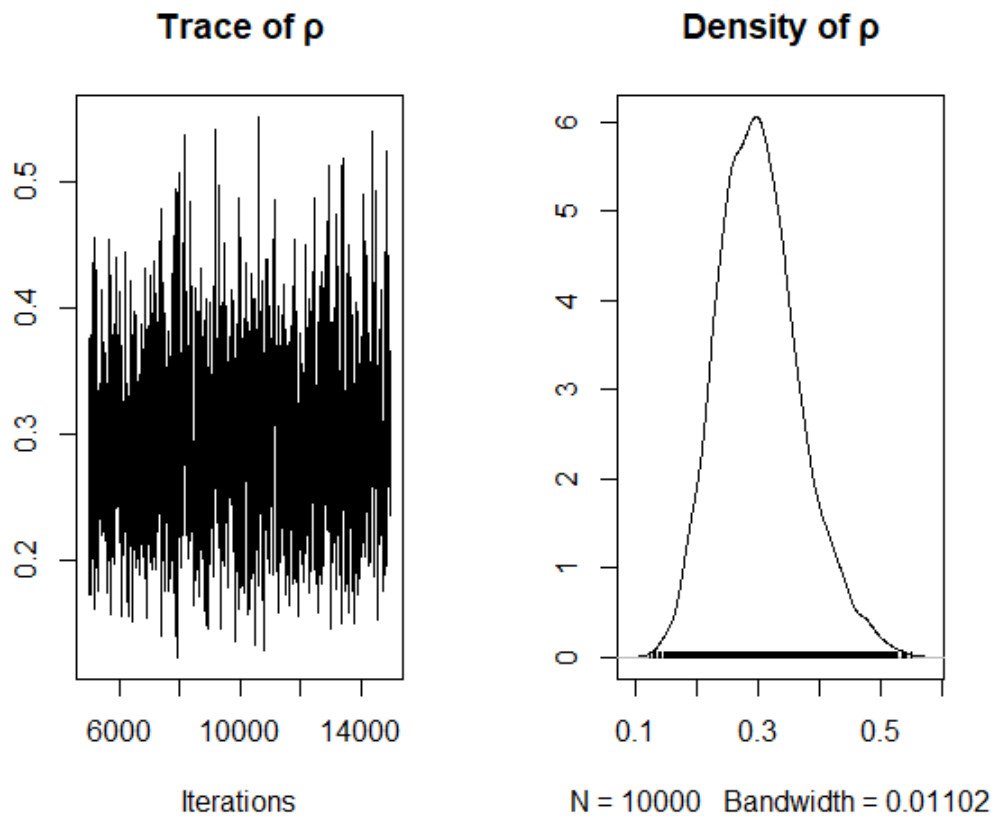


Figura 4.4: MCMC - Traço e densidade empírica do ICC (ρ)

Antes de ajustar o modelo com variáveis explicativas, calculou-se os BLUP para o intercepto, para avaliar a estimação em ambas abordagens.

Tabela 4.3: BLUP para o intercepto

Turma	Clássico	Bayesiano
1	0,053	0,054
2	0,378	0,379
3	-0,720	-0,722
4	0,999	1,002
5	-0,110	-0,110
6	0,108	0,109
7	0,435	0,437
8	1,181	1,185
9	0,218	0,220
10	-0,408	-0,408
11	-0,155	-0,155
12	-0,851	-0,852
13	-0,120	-0,120
14	-0,217	-0,217
15	-0,562	-0,562
16	0,674	0,677
17	0,617	0,619
18	0,273	0,274
19	0,734	0,737
20	-0,634	-0,635
21	0,981	0,985
22	-0,943	-0,945
23	-0,413	-0,413
24	-1,370	-1,372
25	-0,146	-0,146

Observa-se, de acordo com a Tabela 4.3, que ambas abordagens produziram resultados próximos, diferindo na terceira casa decimal apenas em alguns casos. Conforme discutido na Seção 2.3, os BLUP acima se referem à (2.18), então os valores acima se referem ao efeito individual de cada grupo. Para se obter o intercepto específico de cada turma, deve-se somar à este efeito o intercepto médio, conforme Equação (2.19).

Uma vantagem imediata na estimação bayesiana é que podemos facilmente calcular intervalos de credibilidade para os BLUP. Esses intervalos estão representados

na Figura 4.5.

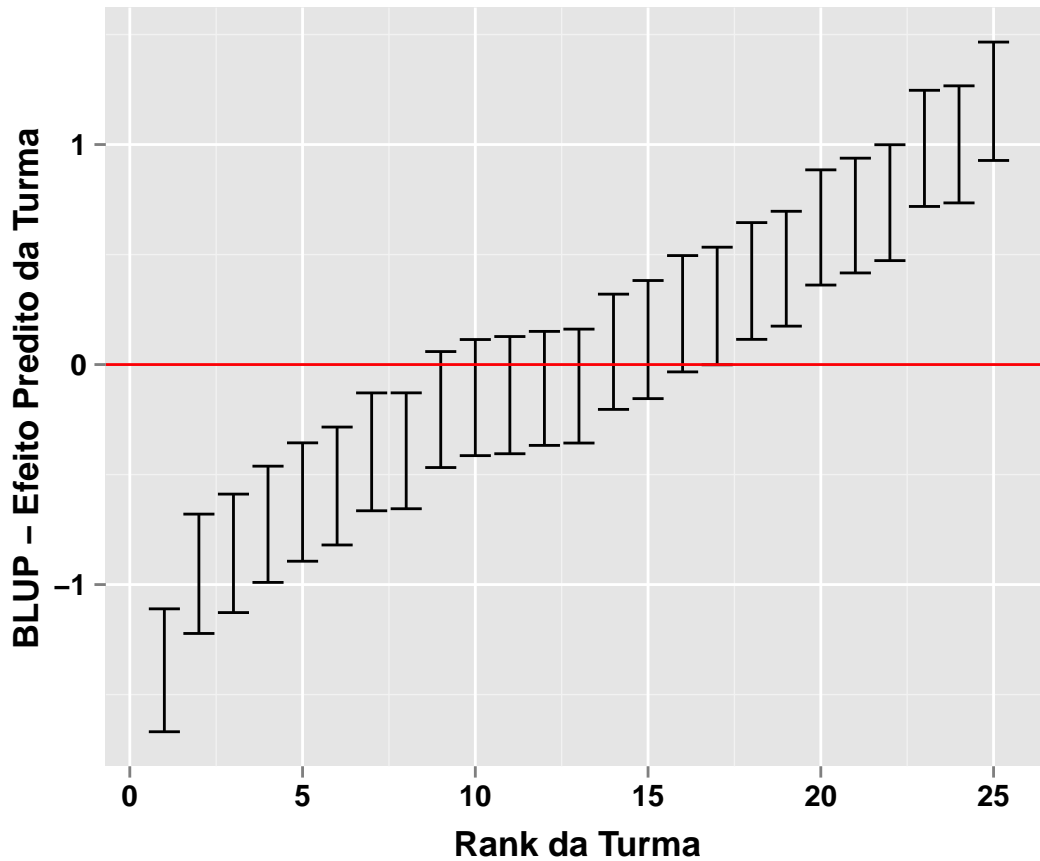


Figura 4.5: Intervalo de Credibilidade dos BLUP - Efeito Predito das Turmas

A Figura 4.5 mostra os BLUP de cada turma com seus limites superior e inferior do intervalo de credibilidade de 95%, ordenados do grupo com menor valor do BLUP até o maior. Se não houvesse diferença entre turmas, então esperaríamos que todos os intervalos passassem por 0. Dessa maneira, esse gráfico serve como uma espécie de diagnóstico para verificar se realmente há ou não diferença entre grupos que justifique o modelo multinível.

Assim, temos que as 8 turmas com menor BLUP estão com seus intervalos de credibilidade abaixo de 0, enquanto que as 8 com maiores médias estão acima. Isso é uma indicação de que realmente há efeito de grupos na variável resposta Y .

O mesmo gráfico pode ser feito para o intercepto específico de cada turma, conforme Equação (2.19) (Figura 4.6).

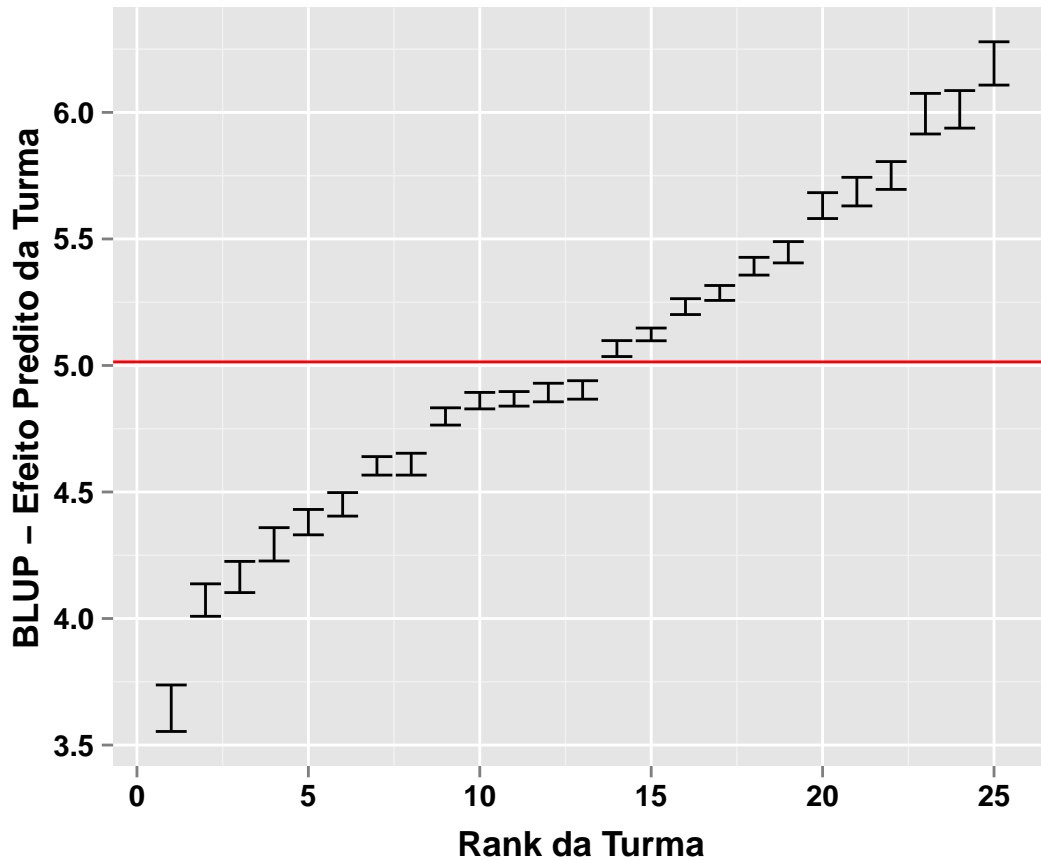


Figura 4.6: Intervalo de Credibilidade dos BLUP - Intercepto Médio das Turmas

Um resultado interessante na Figura 4.6, é que os intervalos HPD, construídos a partir da posteriori de (2.19), foram bem menores que os anteriores. Esse gráfico ainda apresenta a mesma conclusão do gráfico anterior, de que temos diferença entre grupos. No entanto, dessa vez olhamos para aqueles grupos que estão acima ou abaixo do intercepto geral, representado pela reta horizontal vermelha.

4.1.2 Ajuste do Modelo de Regressão Hierárquica - dois níveis

A Tabela 4.4 apresenta as estimativas dos parâmetros do modelo de regressão hierárquica (2.12), com apenas um regressor no menor nível, fixo.

Tabela 4.4: Estimativas Clássicas e Bayesianas dos Parâmetros da Regressão Hierárquica

Abordagem	μ (IC 95%)	β (IC 95%)	σ_G^2 (IC 95%)	σ_E^2 (IC 95%)
Clássica (EMV)	2,470 (1,967 ; 2,973)	0,498 (0,422 ; 0,574)	0,625 (0,253 ; 0,996)	0,869 (0,760 ; 0,978)
Bayesiana (Média Post.)	2,485 (2,025 ; 2,989)	0,498 (0,421 ; 0,568)	0,693 (0,343 ; 1,115)	0,877 (0,778 ; 1,001)

Nesse caso, tem-se que ambas abordagens foram próximas nas estimativas de todos os parâmetros μ , β , σ_G^2 , σ_E^2 . Tem-se que o intervalo de credibilidade dos parâmetros de interesse nos indica que os parâmetros são significativamente diferente de zero.

Novamente, é necessário verificar a convergência das cadeias geradas.

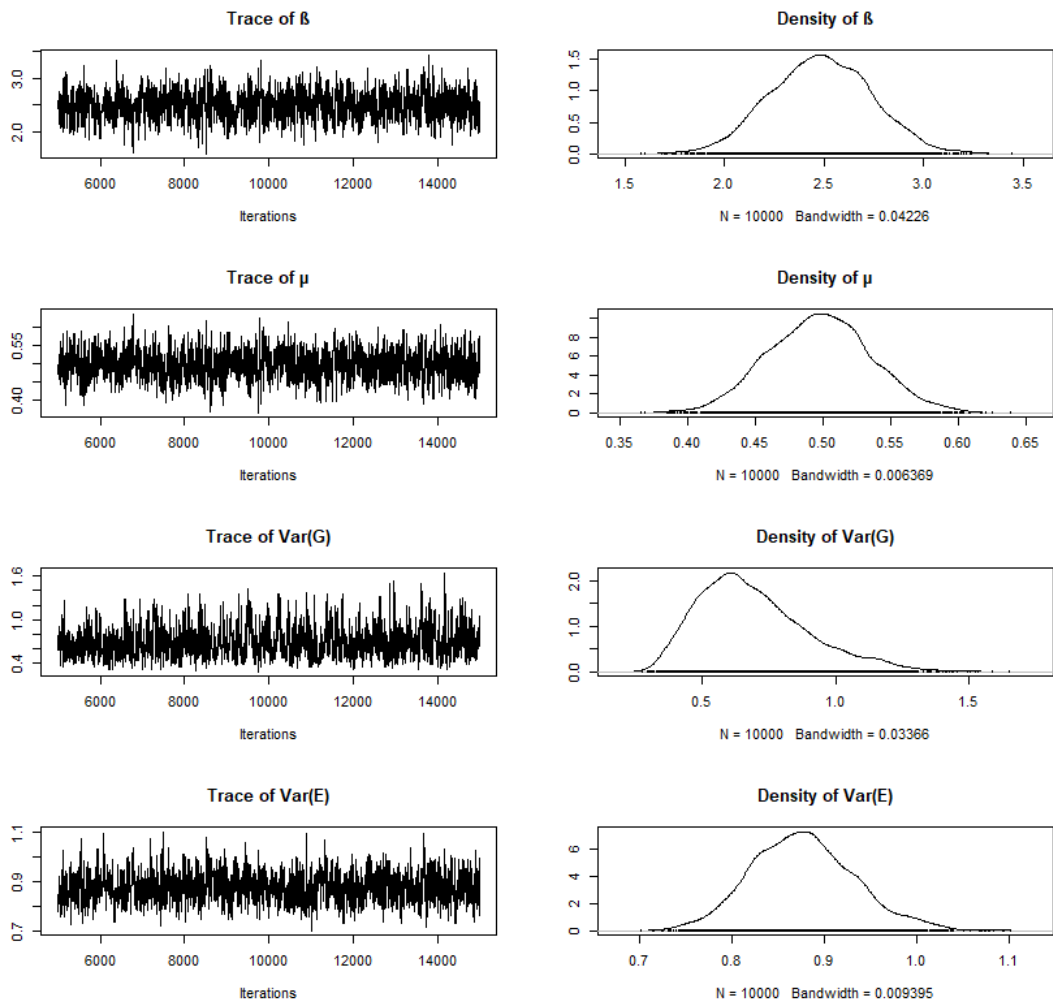


Figura 4.7: MCMC - Traço e Densidade Empírica dos Parâmetros da Regressão Hierárquica

De acordo com a Figura 4.7, apresentando na ordem μ , β , σ_G^2 e σ_E^2 , tem-se boa convergência do MCMC para todos os parâmetros.

Assim, temos que $\beta = 0,498$ indica que um aumento de 1 unidade na extroversão do aluno resulta em um aumento de 0,498 unidades na popularidade desse aluno, em média. O valor $\mu = 2,485$ indica a popularidade esperada de um aluno com 0 de extroversão. A Figura 4.8 apresenta a reta predita pelo modelo, segundo a abordagem bayesiana.

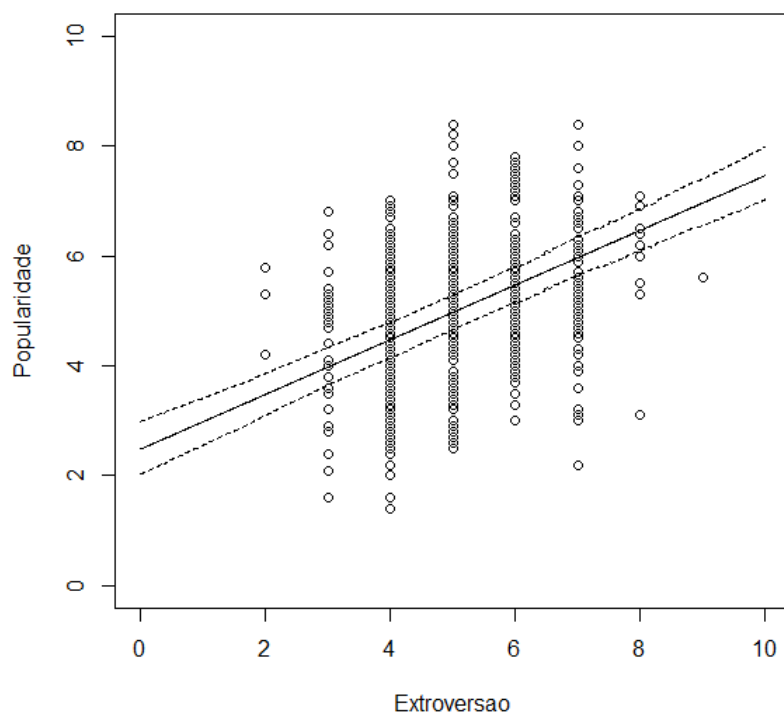


Figura 4.8: Reta Predita pelo Modelo

A Figura 4.8 mostra a dispersão da variável explicativa extroversão e a variável resposta popularidade, com a reta predita pelo modelo e limites inferior e superior do intervalo HPD de 95% de credibilidade.

É importante ressaltar que a reta ajustada na Figura 4.8 leva em conta o intercepto médio dos grupos. Com o BLUP, podemos calcular uma reta predita para cada turma.

Tabela 4.5: BLUP para o intercepto - Modelo com Regressor

Turma	Clássico	Bayesiano
1	0,181	0.163
2	0,563	0.535
3	-0,404	-0.402
4	1,088	1.043
5	-0,492	-0.487
6	0,261	0.242
7	0,692	0.658
8	1,310	1.258
9	0,177	0.160
10	-0,748	-0.736
11	-0,141	-0.149
12	-0,778	-0.768
13	-0,166	-0.172
14	-0,377	-0.378
15	-1,268	-1.243
16	1,065	1.018
17	0,818	0.780
18	0,344	0.323
19	0,824	0.787
20	-0,887	-0.874
21	1,019	0.973
22	-1,449	-1.421
23	-0,330	-0.333
24	-1,317	-1.291
25	0,019	0.006

Para esse modelo, observamos uma maior diferença entre os BLUP bayesiano e clássico, quando comparada ao modelo anterior, mesmo que esta ainda seja pequena. Novamente convém fazer um gráfico com os intervalos de credibilidade dos BLUP para verificar a diferença entre grupos.

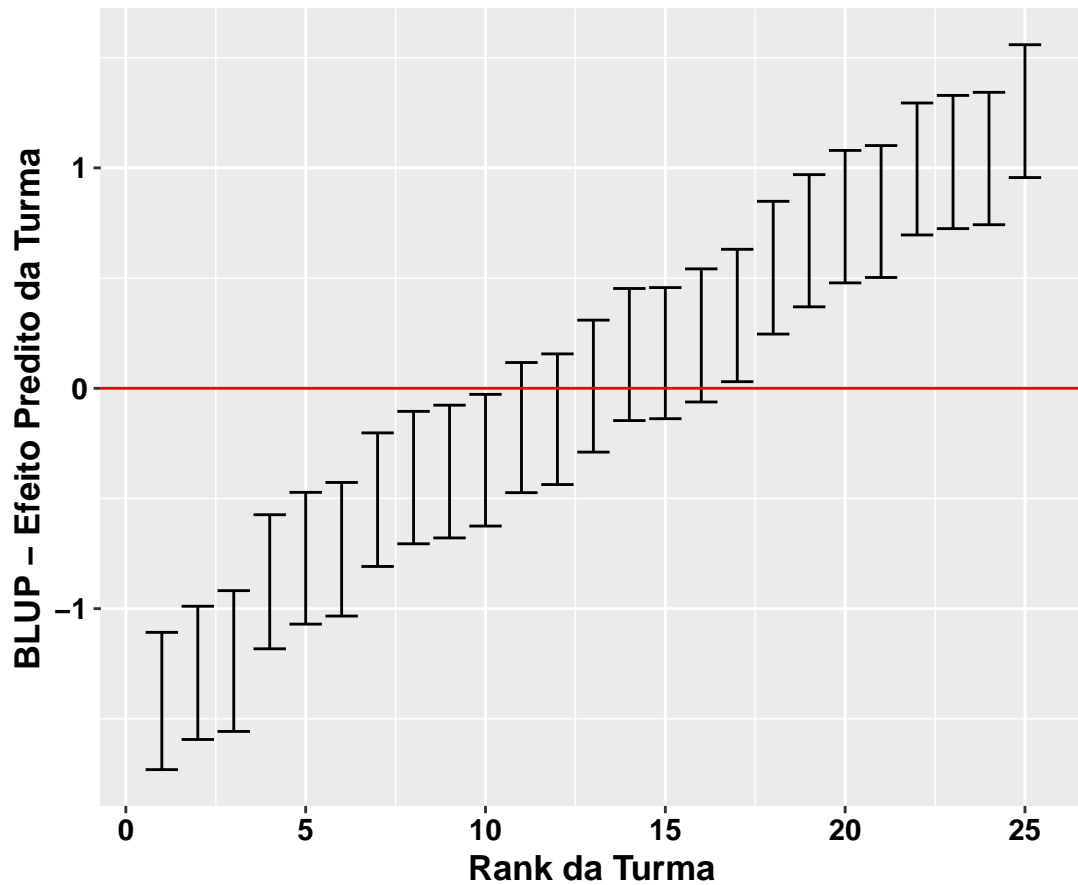


Figura 4.9: Intervalo de Credibilidade dos BLUP - Efeito Predito da Turma - Modelo com Regressor

Analogamente ao modelo anterior, a Figura 4.9 mostra os BLUP ordenados do menor para o maior BLUP de cada turma na Tabela 4.5, com seus limites superior e inferior de credibilidade. Observa-se que há diferença entre grupos, pelo fato dos 10 grupos com menor BLUP estarem com seus intervalos abaixo de 0 e os 9 grupos de maior BLUP estarem acima.

De posse das posteriores dos parâmetros do modelo e dos BLUP, podemos ajustar uma reta predita para cada grupo, conforme descrito na Seção 3.4 e apresentado na Figura 4.10.

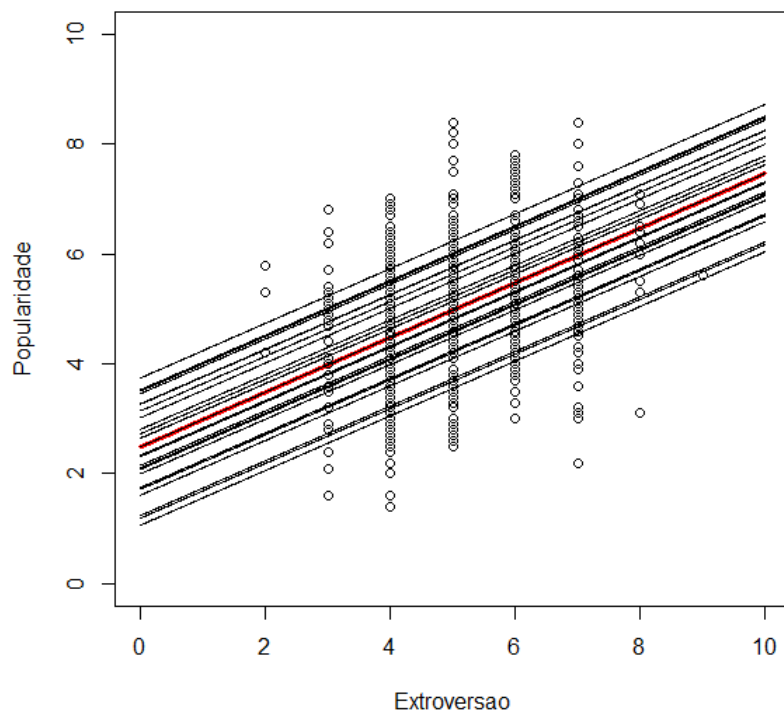


Figura 4.10: Retas Ajustadas com BLUP

Como esperado pelo modelo com apenas interceptos aleatórios, as retas de cada turma são paralelas, conforme Figura 4.10. A reta vermelha representa a reta predita pelo intercepto médio, enquanto que as demais retas representam as retas de cada grupo. Os intervalos de credibilidade podem facilmente ser calculados para cada reta, pois temos todas as posteriores.

4.2 Aplicação - três níveis hierárquicos

Os dados consistem de 1000 enfermeiras dentro de 25 hospitais. Cada hospital possui 4 enfermarias, com 9 à 13 enfermeiras cada, totalizando 100 enfermarias. A Tabela 4.6 descreve o comportamento da variável Y estresse (resposta) em cada

hospital. Nesse exemplo, não considerou-se covariáveis.

Tabela 4.6: Descrição dos Hospitais. Fonte [J.J.Hox, 2010]

Variável		n_i	m_i	\bar{y}_i
Hospitais	1	4	36	5,361
	2	4	36	4,805
	3	4	37	4,918
	4	4	38	5,289
	5	4	36	5,500
	6	4	36	4,583
	7	4	40	4,700
	8	4	40	5,275
	9	4	40	3,600
	10	4	40	5,550
	11	4	43	4,697
	12	4	51	4,941
	13	4	52	4,423
	14	4	50	4,480
	15	4	46	5,195
	16	4	40	5,125
	17	4	36	4,916
	18	4	39	4,820
	19	4	40	5,725
	20	4	40	5,250
	21	4	40	3,775
	22	4	36	5,388
	23	4	36	5,777
	24	4	36	5,166
	25	4	36	5,694
Geral	25	100	1000	4,977

Na Tabela 4.6, n_i é o número de enfermarias do hospital i e m_i é a quantidade de enfermeiras no hospital i , tem-se que a média geral de Y é 4,977. Observa-se que a média dessa variável em cada grupo não aparenta ser tão diferente da geral, mas nos hospitais 9 e 21 a média foi de 3,6 e 3,7, respectivamente, o que pode justificar o modelo multinível.

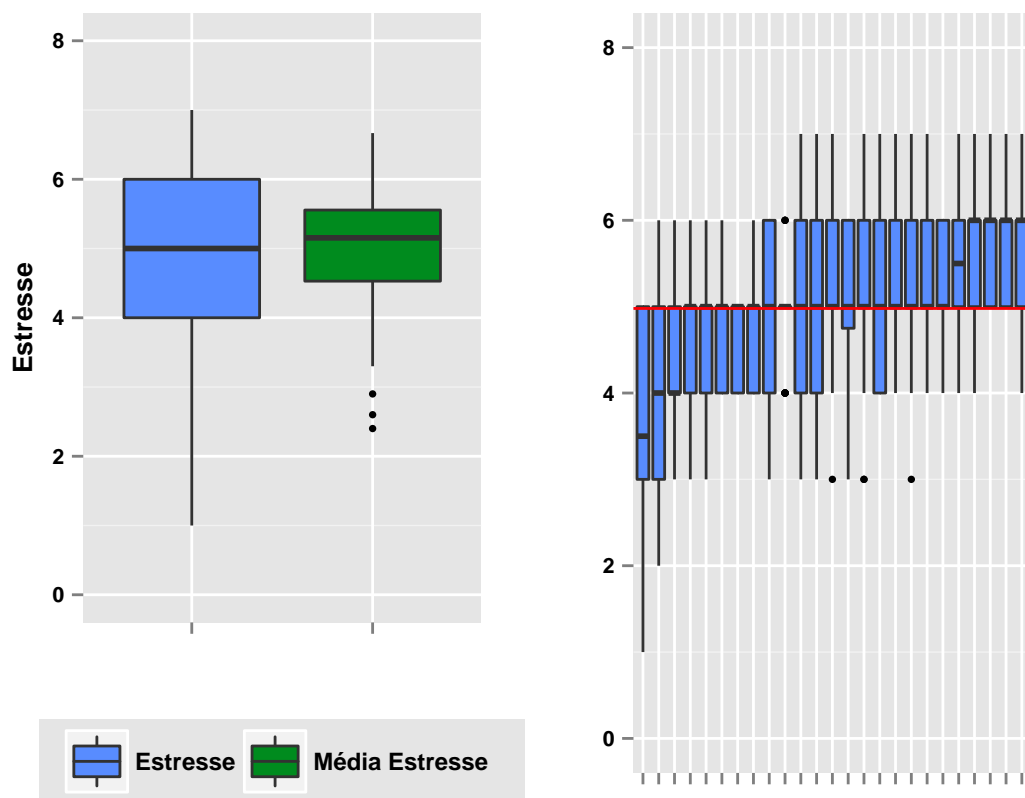


Figura 4.11: *Boxplot* da variável Estresse, da média do estresse em cada enfermaria e Estresse dentro de Hospitais

A Figura 4.11 à esquerda apresenta os *boxplot* da variável Y e da média de Y em cada enfermaria, por uma facilidade de visualização, pois temos 100 enfermarias. Se não houver diferença entre grupos, espera-se que a média do estresse em cada enfermaria não seja muito diferente uma das outras. Pode-se observar que três enfermarias aparentam ter média bastante diferente das demais. Novamente, à direita tem-se os *Boxplot* de Y , para cada hospital, ordenados da menor média de Y dentro daquele grupo até a maior média, sendo a reta horizontal vermelha a média geral.

Observa-se que os hospitais parecem estar divididos em dois grupos com comportamentos de estresse parecidos: os que tem 75% de suas observações acima da

média e os que tem 75% abaixo. No entanto, todos os boxplots passam pela média de Y , o que pode indicar que não há diferença entre grupos.

4.2.1 Ajuste do Modelo Aleatório - três níveis

A Tabela 4.7 apresenta as estimativas dos parâmetros do modelo aleatório com 3 níveis, de acordo com (2.7).

Tabela 4.7: Estimativas Clássicas e Bayesianas dos Parâmetros do Modelo Aleatório 3 níveis

Abordagem	μ (IC 95%)	$\sigma_{G_1}^2$ (IC 95%)	$\sigma_{G_2}^2$ (IC 95%)	σ_E^2 (IC 95%)
Clássica (EMV)	5,000 (4,789 ; 5,212)	0,162 (0,000 ; 0,329)	0,4888 (0,322 ; 0,654)	0,301 (0,273 ; 0,329)
Bayesiana (Média Post.)	5,004 (4,782 ; 5,202)	0,151 (0,000 ; 0,341)	0,528 (0,348 ; 0,729)	0,302 (0,276 ; 0,331)

Ambas as abordagens foram próximas nas estimativas e intervalos de confiança e credibilidade para μ , $\sigma_{G_1}^2$ e σ_E^2 . Uma diferença maior, porem aparentemente pouco expressiva, foi para o parâmetro $\sigma_{G_2}^2$, que é a variância do nível 2, enfermarias. Pelo intervalo de credibilidade, podemos verificar a significância dos parâmetros, exceto para a variabilidade do maior nível (hospitais) $\sigma_{G_1}^2$, no qual suspeitamos que não seja significativamente diferente de zero, pois seu intervalo contém o zero.

A Figura 4.12 apresenta o traço e a densidade empírica dos parâmetros do modelo aleatório, na ordem μ , $\sigma_{G_1}^2$, $\sigma_{G_2}^2$ e σ_E^2 .

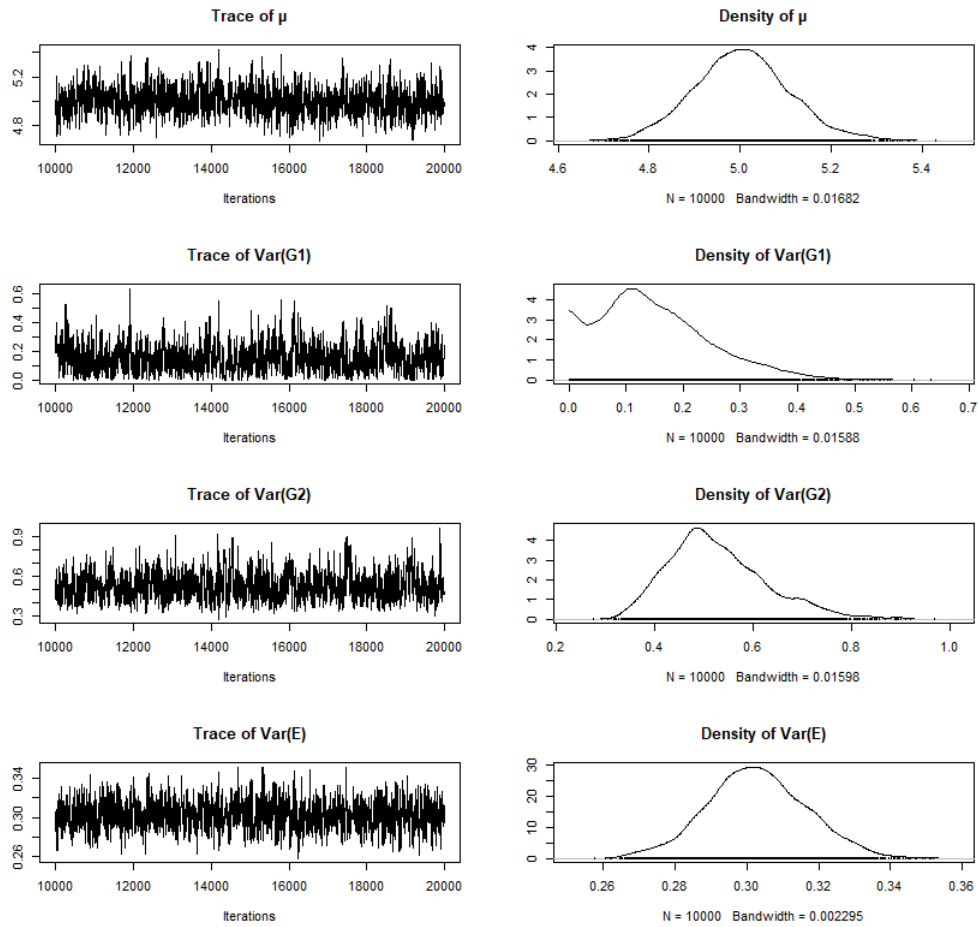


Figura 4.12: MCMC - Traço e densidade empírica dos parâmetros do Modelo Aleatório 3 níveis

Tem-se boa convergência para μ , $\sigma_{G_2}^2$ e σ_E^2 . Para $\sigma_{G_1}^2$, o algoritmo não convergiu muito bem, provavelmente por conta da alta incidência de valores próximos de zero, o que leva à suspeita de que o terceiro nível pode não explicar muito da variabilidade de Y . Ou seja, os hospitais podem não explicar a variabilidade do estresse da enfermeira.

Na Seção 2.2.1 vários coeficientes de correlação intraclasses foram apresentados. O primeiro método, dado pelas equações (2.8) e (2.9) identificam a proporção da variância explicada no nível 2 e nível 3, respectivamente. As Figuras 4.13 e 4.14 apresentam o traço e densidade empírica desses icc 's.

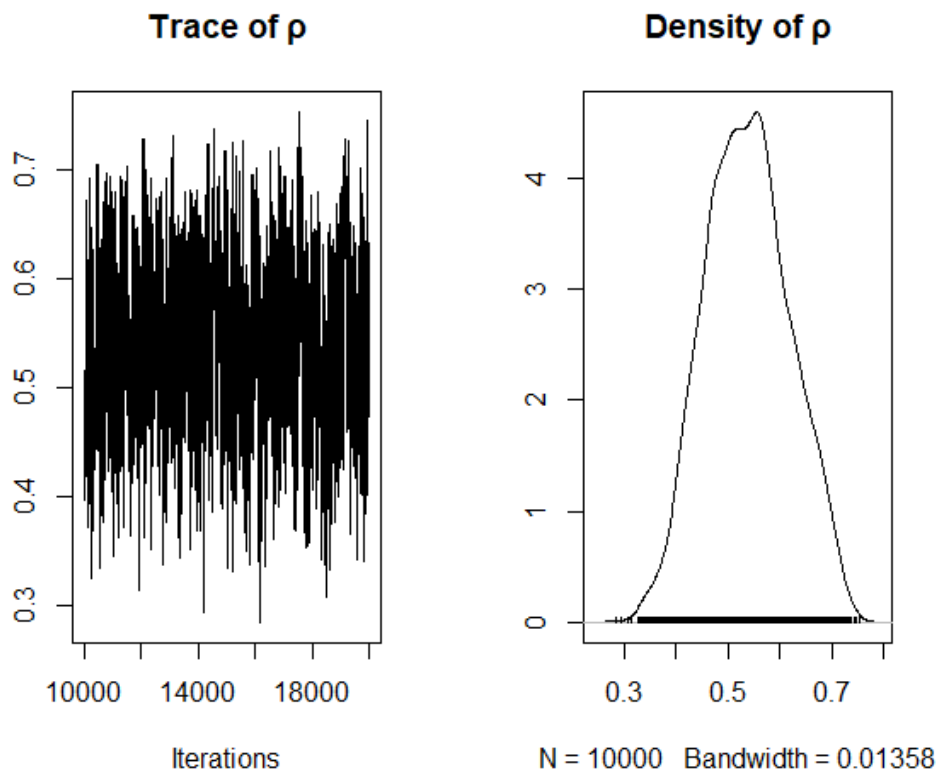


Figura 4.13: Ajuste do ICC (2.8)

A Figura 4.13 apresenta o icc conforme método 1 Equação (2.8). No nível dois, o valor obtido pelo estimador de Bayes foi 0.537, com intervalo de 95% credibilidade de (0,401;0,706), indicando que a enfermaria responde por mais da metade da variabilidade do estresse da enfermeira.

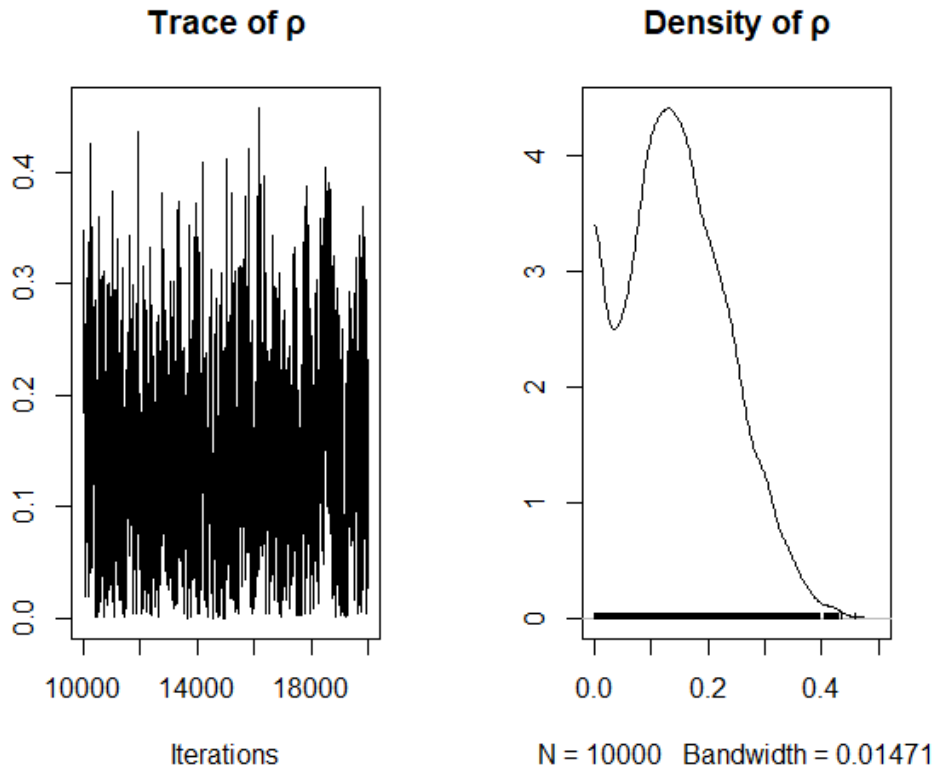


Figura 4.14: Ajuste do ICC (2.9)

No nível três, a proporção é de 0,150, com intervalo de 95% credibilidade de (0,000;0,304), um valor baixo que somado ao fato da grande quantidade de valores próximos zeros obtidas na amostra posteriori do parâmetro $\sigma_{G_1}^2$ indica que o hospital não ajuda muito a explicar o estresse da enfermeira. Além disso, o intervalo de credibilidade inclui zero, dando maiores indícios que esse parâmetro não é significativo. Isso não é surpreendente, pois verificamos na Figura 4.11 que todos os hospitais poderiam apresentar diferença no estresse.

O método 2, dado pelas Equações (2.10) e (2.11), representa uma estimativa para a correlação esperada entre dois elementos escolhidos aleatoriamente dentro do mesmo grupo. A Figura 4.15, apresenta o traço e densidade empírica do icc dado pela Fórmula (2.10).

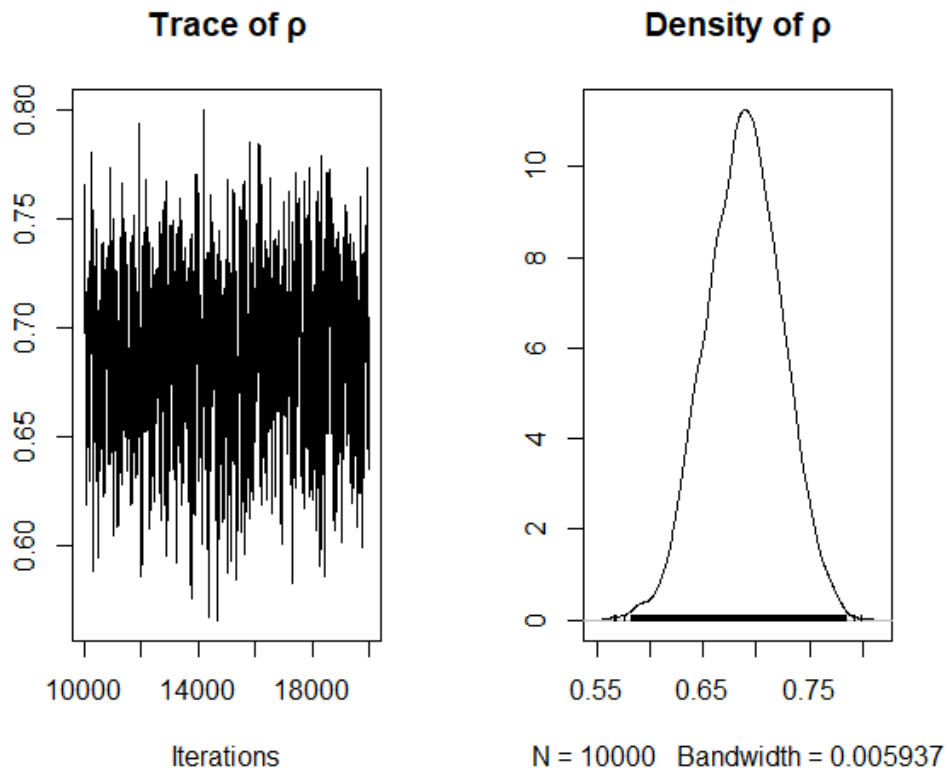


Figura 4.15: Ajuste do ICC (2.10)

A estimativa de ρ obtida pelo estimador de Bayes foi de 0,688, com intervalo de 95% credibilidade de (0,618; 0,753), indicando que mais de 60% da variância de Y é explicada pela estrutura hierárquica dos hospitais e enfermarias. Assim, embora os hospitais não tenham representação muito forte na variabilidade de Y , as enfermarias têm. Logo um modelo com estrutura hierárquica é adequado para esses dados.

Capítulo 5

Conclusão

A partir das análises feitas, conclui-se que modelos hierárquicos, embora mais complexos, apresentam ganho expressivo nas estimativas intervalares dos parâmetros, sem desrespeitar os pressupostos teóricos de um modelo mais simples, proporcionando estimativas não viesadas.

A inferência bayesiana nesse tipo modelo apresenta ganhos expressivos na construção dos intervalos de credibilidade dos parâmetros de interesse. Por não depender da teoria assintótica, a construção desses intervalos para funções dos parâmetros, como o ICC e o BLUP, é feita de maneira imediata, de posse da posteriori. Em compensação, cálculos extensivos são necessários para se chegar à esses resultados.

Os intervalos de credibilidade para os BLUP tornaram mais simples o processo de verificação da necessidade do uso de um modelo hierárquico. Além disso, a inferência bayesiana também permite construir facilmente a reta preditiva para cada grupo, incluindo seus respectivos intervalos de credibilidade.

Um ponto importante na estimação bayesiana é verificar a convergência do MCMC. Para esse tipo de modelo, notou-se que há uma maior dificuldade do algoritmo convergir para a variância do maior nível, quando a quantidade de grupos é menor que 10.

O modelo foi aplicado em dois exemplos da literatura e a partir das análises

obtidas foi possível comprovar a necessidade do uso de uma regressão hierárquica ao invés de regressão usual.

Neste trabalho, foi considerado *prioris* difusas em todos os modelos. No entanto, caso tenhamos informações *a priori* a respeito dos parâmetros, podemos incluí-las na análise por meio da distribuição *a priori*.

O próximo passo natural desse modelo é ajustá-lo com regressores aleatórios, aplicando o mesmo procedimento para se obter os BLUP e a reta predita. Além disso, esse trabalho apresentou apenas um caso dos modelos lineares generalizados mistos, que é o que possui ligação canônica. No entanto, modelos hierárquicos com outras funções de ligação podem ser ajustados, como por exemplo a regressão logística hierárquica.

Referências Bibliográficas

CORDEIRO, Gauss; DEMÉTRIO, Clarice (2009). **Modelos Lineares Generalizados e Extensões**. Piracicaba, São Paulo.

DEMIDENKO, Eugene. **Mixed Models: Theory and Applications**. Hoboken, New Jersey: John Wiley & Sons, 2004.

EHLERS, R. S. (2011). **Inferência Bayesiana**. USP, São Paulo.

HOX, J. J.. **Multilevel analysis: techniques and applications**. Second Edition. Great Britain: Routledge, 2010.

ROCHA, ALEX L.M.M. (2014). **Regressão Logística Multinível: Uma aplicação de Modelos Lineares Generalizados Mistos**. UnB, Brasília.

SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance Components**. Hoboken, New Jersey: John Wiley & Sons, 2006.