

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Análise de modelos de regressão multiníveis simétricos

Marina Mitie Gishifu Osio

Orientadora: Profa. Dra. Cibele Maria Russo Noveli

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA.*

USP – São Carlos
Junho de 2013

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

O82a Osio, Marina Mitie Gishifu
Análise de modelos de regressão multiníveis
simétricos / Marina Mitie Gishifu Osio; orientadora
Cibele Maria Russo Noveli. -- São Carlos, 2013.
83 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2013.

1. Regressão. 2. Modelos multiníveis simétricos.
3. Dados educacionais. I. Noveli, Cibele Maria
Russo, orient. II. Título.

*À minha querida família
e à todos que me
apoiaram sempre que
necessário.*

Agradecimentos

Agradeço a Deus por me dar saúde, força e por colocar pessoas que são verdadeiras luzes que iluminam a minha vida.

À minha querida família, incluindo vó Diva C. Trofino, tia Lourdes A. T. Bastos e ao querido Erick F. Osio (in memoriam), por tudo que sou.

À Professora Cibele, minha querida orientadora, pessoa admirável, responsável e dedicada, pelas orientações e contribuições na elaboração deste trabalho e também pela paciência, apoio, segurança, amizade e incentivo transmitidos durante todo o mestrado.

À admirável e querida amiga Alessandra Viveiro, que é um exemplo de doçura e meiguice misturadas com inteligência e determinação, por me animar com suas doces palavras de incentivo, mesmo longe. Como foram importantes as sms que chegavam nos momentos mais precisos.

Ao querido amigo Bruno Feres de Souza, que tanto admiro, pela agradável companhia, pelas boas lembranças e quantas, pela sua conduta que aumentou minha fé em Deus, pelos muito “sem choro!”, pela paciência, por sempre tirar minhas incontáveis dúvidas (embora tenha limitado, por pouco tempo, em três perguntas diárias sobre LaTeX, me diverti com esse fato), quanto riso, quanto auxílio, e mais ainda, pelas sugestões e contribuições que enriqueceram esse trabalho, por conseguir me animar quando tudo parecia ficar impossível e, principalmente, pela enorme ajuda, que ajuda!

Ao admirável e querido amigo Flavio L. Carvalho, pela grande amizade e respeito, pela parceria nas “infinitas” horas de estudo, pela ajuda, incentivo e força dada em momentos difíceis, por me manter atualizada, por passar os entretenimentos, pela ajuda dada nos estudos, pela agradável companhia e por ser um exemplo de que dedicação e determinação traz, com certeza, o sucesso. Ah, agradeço também por conseguir quebrar minha resistência ao uso do cinto de segurança. Que exemplo!

À querida amiga Rossy M. Pinheiro, que considero uma irmã, pela companhia, apoio e força quando me senti sozinha, sua companhia foi imprescindível.

Ao querido amigo Vinícius Siqueira, que tenho grande admiração e respeito, por sempre tirar minhas dúvidas, pela disposição em ajudar de alguma forma, por ouvir minhas lamentações, pelo carinho, pelo ânimo que contagia e também pela agradável companhia. Seria perfeito se torcesse para outro time, de preferência o meu.

Aos amigos Alina, Catia, Cecília, Dani, Daniel, Danilo, Juliana, Letícia, Luciane, Marcia, Sandra Souza, Rayani, Rose, Rose Pazz, Sarah, Tonton, Willian Oliveira, à família Vaccare, à turma da estatística, aos amigos da escola e a tantos outros que não mencionei aqui mas que me ajudaram de alguma forma, me incentivaram e acreditaram na possibilidade de ter esse sonho realizado.

Às professoras que participaram da banca do exame de qualificação e/ou da defesa de mestrado: Reiko Aoki (ICMC-USP), Roseli Aparecida Leandro (ESALQ-USP) e Viviana Giampaoli (IME-USP), que muito contribuíram com as correções e sugestões me fazendo sentir muito mais orgulho desse trabalho.

A todos os professores que contribuíram na minha formação e aos funcionários do ICMC, em especial aos da segurança, pois me sentia segura e assim estudava tranquilamente nos laboratórios de estudo, inclusive nos finais de semana, até tarde da noite e foram tantas a vezes que saí de madrugada.

Agradeço aos amigos Alina Marcondes, André Rossi, Danilo Alvares, Murillo Carneiro, Noemi Veiga, Rodrigo Calvo e Valéria Carvalho, pessoas que foram assistir e me dar aquela força no dia da defesa de mestrado. Que força!!!

Enfim, agradeço à todas as pessoas, amigos de longe e de perto, que pensam que não tiveram participação nesse trabalho, mas saibam com toda certeza que contribuíram indiretamente na realização desse trabalho, um grande sonho! “Aqueles que passam por nós, não vão sós, não nos deixam sós, deixam um pouco de si, levam um pouco de nós (Antoine de Saint-Exupéry)”. Amigos são anjos que Deus envia pra nos ajudar. Sinto-me abençoada.

Finalmente, impossível esquecer-los, e não poderia deixar de dedicar esse trabalho também aos meus amores, Gunther e Lothar, fiéis companheiros de todas as horas e que me fizeram forte, além de me dar amor incondicional e alegrias. Saudades imensas!!!

Resumo

O uso de modelos multiníveis é uma alternativa interessante para analisar dados que estão estruturados de forma hierárquica, pois permite a obtenção de diferentes estimativas de parâmetros relativos a grupos distintos e, ao mesmo tempo, leva em consideração a dependência entre as observações em um mesmo grupo. Neste trabalho, desenvolvemos e aplicamos modelos de regressão multiníveis simétricos, a fim de fornecer alternativas ao modelo usual, sob normalidade. Além disso, apresentamos uma breve análise de diagnóstico e estudo de simulação. Como motivação, consideramos dados educacionais, a fim de avaliar se o número de reprovações no histórico escolar do aluno e a infraestrutura da escola são variáveis relevantes que afetam o baixo desempenho dos alunos do ensino básico na disciplina de Matemática.

Abstract

The use of multilevel models is an interesting alternative to analyze data that is structured in a hierarchical manner, since it allows the obtention of different parameters estimates for distinct groups and, at the same time, it takes into account the dependence of observations in the same group. In this dissertation, we develop and apply symmetrical multilevel regression models, for the purpose of providing alternatives to the usual model, under normality. Furthermore we present a brief diagnostics analysis and a simulation study. As motivation, we consider educational data in order to assess whether the number of failures in school history of students and the school infrastructure are important variables that affect the low performance of elementary school students in Mathematics.

Sumário

1	Introdução	1
1.1	Justificativa	1
1.2	Revisão Bibliográfica	4
2	Motivação	9
2.1	Conjunto de dados	9
2.2	Análise descritiva e exploratória	10
3	Modelos de regressão multiníveis normais	13
3.1	Modelo de regressão com um nível	13
3.2	Modelos de regressão com dois níveis	14
3.3	Modelo de regressão multinível com t covariáveis	18
3.4	Estimação dos parâmetros	19
3.4.1	Funções Escore	21
3.4.2	Matriz de Informação de Fisher	23
3.4.3	Processo de estimação dos parâmetros	23
4	Modelos de regressão multiníveis simétricos	27
4.1	Modelos simétricos	27
4.2	Modelos multiníveis simétricos	32
4.2.1	Processo de estimação dos parâmetros	32
4.3	Seleção de modelos	35
5	Análise de diagnóstico	37
5.1	Diagnóstico de influência	37
5.1.1	Influência local proposta por Cook (1986)	38
5.1.2	Abordagem de influência local para modelos mistos	40

5.2	Análise de resíduos	41
5.2.1	Resíduos condicionais	42
5.2.2	Resíduo marginal	43
6	Aplicações	45
7	Estudo de simulação	57
8	Discussão e trabalhos futuros	63
A	Acesso aos dados	65
B	Comandos em R	73
	Referências Bibliográficas	77

Lista de Figuras

1.1	Estrutura hierárquica de três níveis.	3
2.1	Histograma, gráfico de quantis e boxplot referentes a amostra de notas dos alunos na prova de Matemática.	11
2.2	Densidade alisada das notas de matemática.	12
3.1	Comparação do rendimento escolar médio de duas escolas A e B, para diferentes números de reprovações.	15
4.1	Gráficos da função de densidade de probabilidades de algumas distribuições simétricas.	28
6.1	Gráficos para avaliar o ajuste do modelo normal sem variáveis explicativas.	47
6.2	Gráficos para avaliar o ajuste do modelo sob normalidade considerando o número de reprovações para explicar o rendimento.	49
6.3	Gráficos das funções densidades de probabilidades de algumas distribuições elípticas e o histograma das notas normalizadas.	50
6.4	Gráfico do AIC para escolher o melhor parâmetro para o ajuste do modelo.	52
6.5	Gráficos para avaliar o ajuste do modelo normal com duas covariáveis.	52
6.6	Gráficos para avaliar o ajuste do modelo t-Student com duas covariáveis.	53
6.7	Gráficos para avaliar o ajuste do modelo exponencial potência.	54
6.8	Autovalores relativos às direções de influência.	54
6.9	Gráficos de autovetores absolutos vs. índices (escolas).	55

Lista de Tabelas

2.1	Medidas de posição referentes as notas de matemática da amostra de dados de São Carlos.	10
2.2	Medidas de posição referentes a notas das avaliações da amostra de São Carlos na escala do SAEB.	10
2.3	Medidas de dispersão das notas de Matemática dos alunos de São Carlos.	12
4.1	Função geradora de densidades de algumas distribuições simétricas. . .	30
4.2	Expressões de $w_g(u)$ de algumas distribuições simétricas	31
4.3	Expressões de $w'_g(u)$ de algumas distribuições simétricas	31
4.4	Expressões de d_{gi} e f_{gi} de algumas distribuições elípticas.	36
5.1	Propósito dos gráficos de resíduos condicionais.	43
5.2	Propósito dos gráficos de resíduos marginais.	44
6.1	Estimativas dos parâmetros do modelo sem variáveis explicativas. . . .	46
6.2	Estimativas dos parâmetros do modelo considerando a variável explicativa número de reprovações.	48
6.3	Estimativas dos parâmetros com erros padrão (E. P.) sob os modelos normal, t-Student e exponencial potência.	51
6.4	Estimativas dos parâmetros, mudança relativa sob os modelos normal, t-Student e exponencial potência com todas as observações, sem 12 e 19 e excluindo apenas a 19 e AIC dos modelos ajustados.	56
7.1	Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição normal.	59

- 7.2 Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição t-Student. 60
- 7.3 Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição exponencial potência. 61

Introdução

Os modelos multiníveis representam uma importante ferramenta para a modelagem de dados organizados em uma estrutura hierárquica. É bastante discutido na literatura que o uso desses modelos pode fornecer análises mais adequadas do que os modelos de regressão usuais, pois leva em consideração a possível correlação existente entre dados originalmente pertencentes a um mesmo grupo. Alguns trabalhos já mostraram que ignorar a estrutura de agrupamentos pode levar a conclusões incorretas a respeito de um determinado problema.

Neste capítulo, apresentamos uma justificativa para desenvolvimento e análises de modelos multiníveis simétricos, que representam alternativas aos modelos de regressão usuais, pois possibilitam obter estimativas mais robustas para os parâmetros, melhorando assim as inferências acerca de um fenômeno. Com isso, pode-se intervir de forma mais eficiente na resolução do problema relacionado ao baixo desempenho dos alunos nas avaliações de Matemática. Além disso, apresentamos a revisão bibliográfica com os trabalhos mais relevantes relacionados ao tema da dissertação de mestrado.

1.1 Justificativa

A educação de qualidade é um grande desafio de países em desenvolvimento. O baixo desempenho de estudantes brasileiros nas avaliações tem causado grande preocupação

para órgãos responsáveis pela educação. Como exemplo pode-se citar o resultado apresentado pelo relatório do Programa Internacional de Avaliação de Alunos (PISA) de 2009 em que os estudantes brasileiros obtiveram uma pontuação de 386 em matemática, quando a média sugerida como adequada seria de 496. A presença sistemática de instrumentos de avaliação do índice de rendimento escolar, como o Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP), o Sistema Nacional de Avaliação da Educação Básica (SAEB) e o Programa Internacional de Avaliação de Alunos (PISA), possibilitam a realização de análises mais precisas do aproveitamento dos alunos e de características das escolas que podem impactar no aproveitamento escolar. Assim, periodicamente, os estudantes da rede pública são submetidos às avaliações externas promovidas pela Secretaria do Estado da Educação com o objetivo de avaliar seu desempenho em diversos componentes curriculares, possibilitando, desse modo, o diagnóstico do sistema de ensino e o monitoramento das políticas públicas de educação. Essas avaliações são feitas por meio de provas específicas de Matemática, Língua Portuguesa, Ciências Humanas, entre outras. Adicionalmente, os estudantes respondem a questionários socioeconômicos, onde são coletadas informações de características pessoais e familiares, tais como nível educacional dos responsáveis, renda familiar, idade de ingresso no sistema escolar, entre outras. Constam também nesse questionário informações associadas à escola, como infra-estrutura e práticas pedagógicas. Com esses dados, pode-se investigar o relacionamento entre o rendimento escolar e características socioeconômicas, culturais e ambientais, por exemplos, grau de escolaridade dos professores e ou dos responsáveis pelos alunos, se a escola tem ou não biblioteca e sala de informática, se o espaço físico da escola é adequado, limpo, organizado e seguro, ou seja, muitas são as características que podem impactar no nível de aproveitamento dos estudantes.

Em muitas pesquisas, é comum os dados serem organizados de forma hierárquica, ou seja, as informações coletadas dos indivíduos podem estar separadas em diferentes níveis de agrupamento. A organização dos dados em estrutura hierárquica pode ser importante para a compreensão do problema estudado. Por exemplo, para investigar como as características do local de trabalho influenciam na produtividade dos funcionários, considere uma grande empresa, que tem unidades espalhadas em várias regiões, os funcionários estão agrupados em diferentes setores, de acordo com suas funções, tais como, atendimento, limpeza, segurança, etc. Logo, pode-se observar que existe uma hierarquia de agrupamento. Inicialmente, os funcionários estão separados em equipes de trabalhos, em seguida em setores de atividade, inseridas numa unidade que representa uma das redes de uma grande empresa. Portanto, percebe-se, que uma das principais características de uma estrutura hierárquica, é que os agrupamentos estão inseridos

em outros grupos, ou seja, unidades experimentais estão agrupadas em outras maiores. Essa estrutura de organização dos dados, ou esse tipo de agrupamento é importante tanto para a determinação de quais características do local de trabalho podem influenciar na produtividade dos mesmos, quanto para a compreensão de outros fenômenos. Além disso, deve-se considerar uma possível correlação existente entre unidades experimentais pertencentes a um mesmo grupo.

Dados estruturados de forma hierárquica podem ser observados também em muitas outras áreas como saúde, educação, estudos que envolvam aspectos ambientais, em pesquisas sociais, etc. A estrutura hierárquica é visivelmente observada nos dados relacionados à educação, uma vez que os alunos estão agrupados por turmas (nível 1), as turmas estão separadas em períodos (nível 2), os períodos estão agrupadas em escolas (nível 3) e as escolas são de cidades ou regiões (nível 4). Assim, devido às características de organização desses dados, é razoável que uma estrutura seja proposta para a análise dos fatores determinantes do rendimento escolar em matemática dos estudantes brasileiros. Apresentamos na Figura 1.1 a ilustração de uma estrutura hierárquica com três níveis, onde características relacionadas aos alunos, turmas e escolas estão, respectivamente, nos níveis 1, 2 e 3.

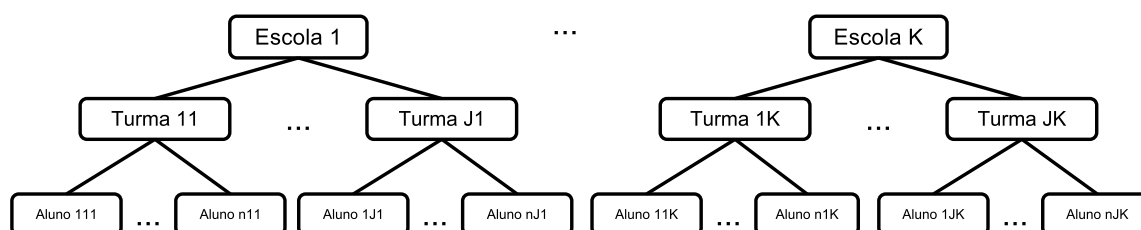


Figura 1.1: Estrutura hierárquica de três níveis.

Tendo em vista que os dados relacionados à educação apresentam, em geral, estrutura hierárquica, de acordo com Goldstein (2011a), para estudar a eficácia educacional, uma possibilidade é utilizar modelos multiníveis, pois permitem que cada um dos níveis seja especificado separadamente e que posteriormente seja reunidos em um único modelo. Além disso, permitem a incorporação de efeitos aleatórios associados a cada um dos níveis de hierarquia (Natis, 2000). Assim, a análise de modelos multiníveis é uma ferramenta estatística muito apropriada para analisar dados dessa natureza.

Um problema bastante comum discutido na literatura é o fato de possíveis observações atípicas afetarem os modelos em desenvolvimento. Nossa proposta é evitar essa influência

considerando distribuições simétricas, que incluem distribuições com caudas mais pesadas ou mais leves que as da distribuição normal (Cysneiros et al., 2005).

Diante do exposto, o presente projeto tem como objetivo estudar, desenvolver e analisar alternativas adequadas para empregar os modelos de regressão multiníveis simétricos em um conjunto de dados da área educacional, que é indicado e utilizado em conjuntos de dados com características hierárquicas. No modelo linear multinível, as unidades estão agrupadas em níveis, esses níveis poderão ser reagrupados em um nível superior. Por exemplo, considerando a hierarquia do conjunto de dados relacionado aos alunos de uma determinada escola, espera-se que a infra-estrutura dessa escola tenha a mesma influência no rendimento de diferentes turmas nessa mesma escola. Entretanto, esse rendimento pode se diferenciar entre as turmas, pois outra variável, por exemplo a prática pedagógica do professor, pode ser um fator de influência no desempenho dessas turmas. Nos modelos de regressão tradicionais, uma das suposições mais comuns é a independência entre os indivíduos, o que não ocorre nos problemas aqui discutidos. Dessa forma, o modelo de regressão multinível pode ser uma alternativa aos modelos tradicionais, pois as observações dos indivíduos estão relacionadas entre os níveis e não considerar esse aspecto um possível resultado causaria a superestimação da variância. Assim, utilizar modelos de regressão multinível para dados estruturados na forma hierárquica pode fornecer estimativas mais precisas para os parâmetros. Como consequência desse estudo, obtivemos resultados que podem contribuir para a compreensão de elementos relacionados ao baixo rendimento observado nas avaliações de matemática e além disso nortear ações pedagógicas e administrativas para aumentar o desempenho dos estudantes das escolas públicas.

Na próxima seção apresentamos uma revisão bibliográfica detalhada relacionada ao modelo em estudo, cujo objetivo é descrever os principais trabalhos de investigação científica na área.

1.2 Revisão Bibliográfica

A literatura dos modelos multiníveis é relativamente recente. É bastante comum encontrar, em trabalhos dessa área, comparações de modelos multiníveis com modelos usuais, que assumem independência mesmo para observações em um grupo comum. Em diversos estudos, observou-se uma discordância nas conclusões desses diferentes modelos, o que pode enfatizar a importância de se considerar a possível estrutura de dependência entre observações em um mesmo grupo e as consequências indesejadas de se ignorar a estrutura hierárquica.

Alguns dos trabalhos mais importantes que tratam de modelos de regressão linear multiníveis podem ser encontrados em Raudenbush and Bryk (2002), Gelman and Hill (2007), Goldstein (2011b) e Hox (2010). Entretanto, não foi encontrado trabalhos na literatura considerando modelos multiníveis assumindo distribuições simétricas, o que pode proporcionar maior flexibilidade para a distribuição dos erros e ao mesmo tempo produzir estimativas mais robustas a pontos atípicos no conjunto de dados.

Resultados interessantes utilizando modelos de regressão multiníveis foram apresentados em Zanini et al. (2011), cujo interesse era verificar quais variáveis estão relacionados com a mortalidade neonatal. Esse estudo mostrou que no modelo tradicional, o baixo peso ao nascer é a variável que está mais relacionada com a variável resposta, enquanto que o uso do modelo multinível apresentou a existência de pequenos efeitos de determinantes contextuais relacionados com a mortalidade neonatal.

Na área educacional, Natis (2001) descreve uma aplicação do modelo de regressão hierárquico onde utiliza o referido modelo para estudar a habilidade em Língua Portuguesa dos alunos da 4^a série das escolas públicas. Para a autora, os modelos lineares hierárquicos constituem uma nova formulação para os modelos de efeitos aleatórios, que permitem uma descrição e análise mais apropriada das diferentes fontes de variação.

Outro trabalho, também na área educacional e de grande relevância foi descrito por Ferrão (2003). Em seu livro, a autora justifica porque o modelo de regressão linear clássico mostra-se inadequado e descreve as vantagens da utilização do modelo de regressão multinível para análise de dados com estrutura de agrupamento. Ou seja, os modelos de regressão multiníveis podem ser mais adequados para compreender quais características dos agentes envolvidos (alunos, professores, diretores) e das escolas, que justificam a variabilidade entre os resultados referentes à eficácia escolar. Um dos resultados, ilustrado por meio de gráfico e utilizando o modelo de regressão tradicional, mostrou que o impacto do nível socioeconômico dos estudantes no desempenho escolar existe e é igual em todas as escolas, no entanto, o resultado difere se for utilizado o modelo multinível, uma vez que um dos resultados das análises sugeriu que as escolas da amostra onde os alunos atingem notas mais elevadas também são aquelas em que o desempenho escolar é impactado pela sua origem social.

Em 2004, para medir a eficácia do sistema educacional brasileiro, Jesus and Laros (2004), realizaram um estudo que utilizou modelos lineares hierárquicos como ferramenta estatística uma vez que esses modelos incorporam naturalmente a estrutura organizada em hierarquia. Segundo os autores, 79 % da variabilidade do desempenho entre as escolas pode ser atribuída às variáveis relacionadas à infra-estrutura da escola e ao perfil sócio econômico dos alunos.

Como existem diferenças entre alunos, entre turmas e entre escolas, Soares (2005), em seu estudo, utilizou os modelos de regressão multiníveis, considerando três níveis de hierarquia, o que permitiu avaliar a proporção da variabilidade das proficiências dos alunos da 4ª série na avaliação de Língua Portuguesa no estado de Minas Gerais.

Por não considerar a estrutura hierárquica presente nos dados das pesquisas educacionais, Bennett (1976), concluiu que o método de ensino tradicional na Inglaterra era melhor do que o método progressivo. Posteriormente, Aitkin et al. (1981), refez as análises considerando a hierarquia do conjunto de dados e concluiu que não havia evidência de que os métodos fossem diferentes.

Na área de saúde, ao analisar dados de medidas repetidas ao longo do tempo, em que as ocasiões de medidas estão no 1º nível da estrutura hierárquica, Bergamo (2002) utilizou o modelo multinível em dois conjunto de dados. O primeiro foi usado para relacionar a pressão arterial com as variáveis explicativas: índice de massa corpórea, sexo, raça e estado civil de um idoso (nível 2). O segundo conjunto de dados foi utilizado para relacionar o teor de proteínas no leite de 79 vacas australianas, coletados durante 19 semanas e submetidas a três diferentes dietas. Em seu estudo, as conclusões foram que não levar em conta a hierarquia em um conjunto de dados, pode implicar na superestimação dos coeficientes do modelo em estudo e, além disso, quando se tem dados de medidas repetidas no tempo é interessante considerar a estrutura hierárquica.

A utilização dos modelos lineares multiníveis, na área de saúde, tem crescido bastante devido ao fato do interesse em determinar as variáveis que influenciam no desenvolvimento de determinada doença, (Pires, 2009). Segundo a autora, pode haver interesse em avaliar, em uma determinada empresa, se a idade (nível 1) e o setor de trabalho (nível 2) são variáveis que influenciam na variável resposta. Para a autora, a aplicação dos modelos lineares multiníveis pode ser mais atraente na área financeira onde o interesse seria estudar o volume de aplicações na caderneta de renda fixa em função das características dos clientes de agências e explicar as diferenças entre agências bancárias.

Recentemente, Castro (2011) utilizou um modelo de regressão multinível de variância com estrutura complexa para estimar a razão de variâncias da ingestão de energia e nutrientes e também para calcular o número de medidas dietéticas em crianças separadas segundo faixa etária. Segundo a autora a variância intrapessoal tem estrutura complexa pois não é isenta de erros e é de natureza variável. Em seu trabalho essa variância foi obtida para cada criança e modelada como uma função da idade. Nos modelos foram considerados quatro níveis, todos com distribuições normais e com média zero. A ferramenta estatística utilizada para estimar a variância intrapessoal e

interpessoal foi o modelo de regressão multinível de variância com estrutura complexa cujos resultados apontaram para mudanças na variância intrapessoal com a idade.

Uma alternativa para a técnica de regressão múltipla, que leva em consideração a estrutura hierárquica, é a análise multinível, que é a opção metodológica correta para estabelecer as relações entre as variáveis (Laros and Marciano, 2008). Para mostrar, de forma prática, como funciona uma análise multinível, os autores utilizaram o modelo multinível de regressão com dois níveis, utilizando a equação e o método composto em cinco passos descrito em Hox (2010) e usando a base de dados NELS:88 (*National Educational Longitudinal Study*) cujo propósito foi identificar o efeito do fator socioeconômico, também sobre o desempenho educacional.

O trabalho mais recente que trata de modelos multiníveis foi desenvolvido por Manghi (2012) que propõe uma generalização dos modelos normais multiníveis denominada modelos elípticos multiníveis e desenvolve aspectos desses modelos, por exemplos testes de hipóteses para os efeitos fixos e parâmetros de variância e covariância e análise de resíduos. Manghi (2012) mostra que o modelo elíptico multinível pode ser utilizado com eficiência na modelagem de dados com estrutura hierárquica e comprova que o modelo multinível com distribuição t-Student acomoda melhor observações aberrantes do que o modelo com distribuição normal multinível.

No próximo capítulo apresentaremos a motivação para o nosso trabalho, incluindo análises descritivas a respeito do conjunto de dados que motivou esse estudo.

Motivação

Como motivação para o desenvolvimento de modelos multiníveis simétricos, consideramos um conjunto de dados educacionais, cujo objetivo é verificar a influência de fatores externos no aproveitamento escolar em estudantes do ensino fundamental. Mais detalhes são apresentados a seguir.

2.1 Conjunto de dados

Avaliações externas são promovidas periodicamente pela Secretaria da Educação para fazer um diagnóstico do sistema de ensino e ainda, monitorar as políticas públicas da educação. Além das provas específicas, os alunos respondem um questionário de onde são coletadas informações para o estudo dos fatores associados à aprendizagem, a fim de identificar quais são aqueles que contribuem para a qualidade do ensino. Muitas são as variáveis que podem estar relacionadas com o desempenho dos estudantes, inclusive os novos hábitos da sociedade, o que justifica a necessidade de se fazer pesquisas para avaliar a qualidade da educação.

Recentemente, uma das políticas públicas adotadas para melhorar o aproveitamento dos estudantes foi a implementação, em algumas escolas, do sistema de tempo integral, em que o aluno permanece na escola por dois períodos para se dedicar às atividades propostas. Dessa forma, um possível estudo, por exemplo, poderia verificar se o número

de aulas semanais de matemática é uma variável que está relacionada com o rendimento dos alunos nas avaliações anuais por meio de um modelo de regressão linear.

Em uma primeira análise, investigaremos a relação do desempenho com alguns dos fatores, disponíveis no conjunto de dados por exemplo, número de reprovações no histórico escolar. Os dados utilizados nesse estudo estão disponível no site do Instituto Nacional de Estudos e Pesquisas Educacionais e referem-se à avaliação do SAEB do ano de 2007 aplicado na oitava série, denominado atualmente o nono ano do ensino fundamental. Mais detalhes sobre o acesso aos dados e a tabela de dados da amostra utilizada se encontram no apêndice A.

2.2 Análise descritiva e exploratória

As análises preliminares e descritivas foram realizadas utilizando o software R (R Core Team, 2012) e também o pacote computacional *SAS*, especialmente, para a manipulação inicial do conjunto de dados. Esse conjunto de dados contém informações a respeito de alunos, professores e também das características das escolas de todo o Brasil. Para este trabalho, vamos considerar, uma amostra de dados da cidade de São Carlos - SP.

A Tabela 2.1 apresenta algumas medidas de posição do conjunto de dados de uma amostra de alunos de oitava série das escolas da cidade de São Carlos. Os dados se referem a proficiência do aluno em Matemática cujo desempenho varia de 0 a 425.

Tabela 2.1: Medidas de posição referentes as notas de matemática da amostra de dados de São Carlos.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
157,939	219,299	240,752	247,278	280,684	375,107

A Tabela 2.2 informa as mesmas medidas a respeito da proficiência em Matemática, calculada na escala do Sistema de Avaliação do Ensino Básico - SAEB.

Tabela 2.2: Medidas de posição referentes a notas das avaliações da amostra de São Carlos na escala do SAEB.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
-1,646	-0,548	-0,164	-0,048	0,549	2,239

Utilizando apenas os dados da amostra da cidade de São Carlos, foram obtidos os seguintes gráficos que mostram características dos dados de proficiência em Matemática dos alunos nessa amostra.

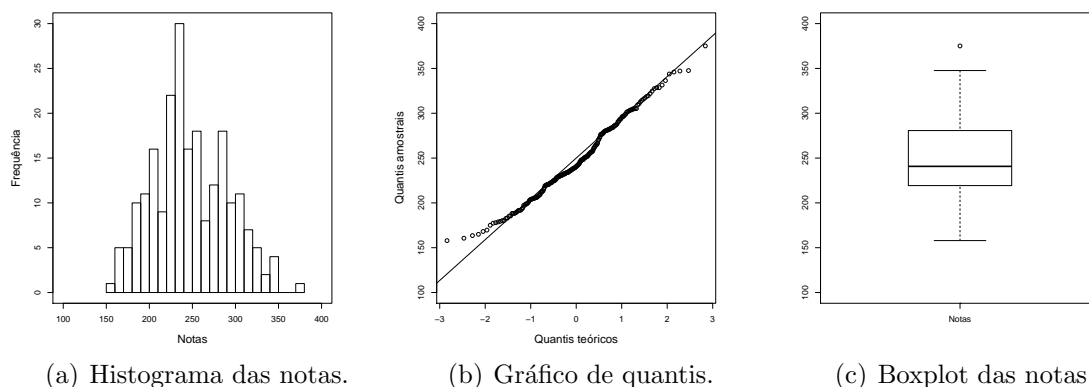


Figura 2.1: Histograma, gráfico de quantis e boxplot referentes a amostra de notas dos alunos na prova de Matemática.

A Figura 2.1(a) mostra a distribuição das notas da prova de Matemática que sugere que os dados apresenta uma distribuição simétrica. O gráfico dos quantis na Figura 2.1(b) também indica que a distribuição de frequência dos dados é simétrica sugerindo uma distribuição com cauda mais leve do que a normal. Já o boxplot da Figura 2.1(c) mostra que os dados da amostra apresentam valores extremos e informa as estatísticas sumárias, por exemplo, nota mínima, mediana e máxima da amostra considerada.

Uma medida de posição útil para descrever a nota obtida pelos alunos de São Carlos na avaliação de matemática é a moda. A moda é uma estatística relacionada ao ponto máximo da curva de densidade no caso contínuo. A moda amostral das notas de matemática de São Carlos está em torno de 229,0.

O cálculo da assimetria, medida que indica o grau de distorção da distribuição em relação a uma distribuição simétrica, indicou o valor de 0,225 para os dados de São Carlos. Para isso foi utilizada a função *skewness* do pacote *fBasics* do software R, ou seja a distribuição tem uma cauda pouco mais pesada do lado direito (assimetria positiva). A Figura 2.2 mostra que a distribuição dos dados é aproximadamente simétrica. Vale observar que existe nesse gráfico uma aparente bimodalidade, que pode ocorrer pelo fato da possível existência de grupos, o que reforça a necessidade do modelo multinível. Outra possibilidade, que pode ser considerada em um trabalho futuro, é a suposição de mistura de distribuições.

A curtose é outra medida que indica a formato da curva em relação ao achatamento, ou seja, indica a intensidade das frequências na vizinhança dos valores centrais. As distribuições cuja curtose assume o valor 3 são usualmente classificadas como mesocúrticas, característica da distribuição normal, as que tem valores menores que 3 são classificadas como platicúrticas e nos demais casos são chamadas de leptocúrticas.

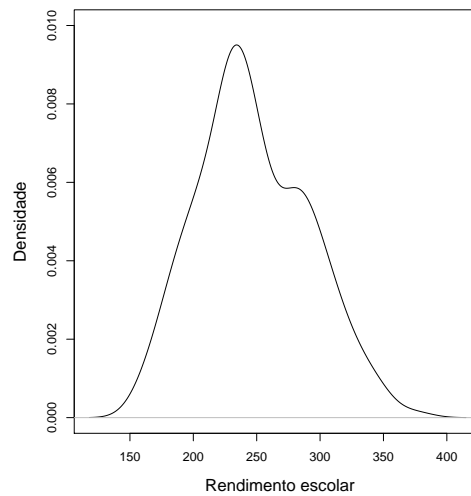


Figura 2.2: Densidade alisada das notas de matemática.

O conjunto de dados de São Carlos apresenta uma distribuição levemente platicúrtica pois o valor da curtose é de 2,717. Esse valor foi obtido utilizando a função *kurtosis* também do pacote *fBasics* do software R, que calcula uma estimativa da medida de curtose de Pearson. Vale ressaltar que como a curtose observada não é exatamente 3, pode fazer sentido propor modelos simétricos, alternativas ao modelo gaussiano, cuja curtose é 3.

As medidas de dispersão informam sobre o grau da variação dos valores observados, ou seja, servem para avaliar o quanto os dados são ou não homogêneos. Na Tabela 2.3 apresentamos algumas medidas de dispersão que foram obtidas utilizando a função *proc univariate* do pacote estatístico SAS.

Tabela 2.3: Medidas de dispersão das notas de Matemática dos alunos de São Carlos.

Desvio padrão	Variância	Amplitude	Amplitude interquartil
43,609	1901,816	217,168	61,385

No próximo capítulo propomos o uso de modelos de regressão multiníveis com a suposição de normalidade para analisar dados estruturados de forma hierárquica.

Modelos de regressão multiníveis normais

Neste capítulo introduzimos alguns modelos de regressão que serão ajustados aos dados da amostra de notas de Matemática de São Carlos, iniciando por um modelo de regressão multinível mais simples e apresentamos o processo para o ajuste do modelo aos dados, ou seja, a estimação dos parâmetros do modelo. Neste capítulo, supomos a normalidade para a distribuição dos erros.

3.1 Modelo de regressão com um nível

Para investigar a relação entre o rendimento escolar (Y) de Matemática dos n alunos de uma determinada escola, e a variável explicativa número de reprovações no histórico escolar (X), um possível modelo seria:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \text{ com } i = 1, 2, \dots, n. \quad (3.1)$$

As suposições usuais são:

- $E(\varepsilon_i) = 0$,

- $Var(\varepsilon_i) = \sigma^2$,
- $Cov(\varepsilon_i, \varepsilon_j) = 0$, para $i \neq j$.

De forma geral, para explicar o rendimento escolar, considere:

Y_i : variável resposta observada no indivíduo i , representa o rendimento escolar;

X_i : variável explicativa observada no indivíduo i , representa uma característica relacionada ao aluno;

β_0 : intercepto, é o valor esperado do rendimento para um aluno com $X_i = 0$;

β_1 : mudança esperada no rendimento médio quando X_i aumenta em uma unidade ou muda de categoria;

ε_i : erro aleatório associado ao i -ésimo aluno, referente a outras variáveis ou fontes de variação não observáveis.

Com esse modelo, pode-se estudar a relação entre uma característica relacionada ao aluno e seu rendimento do mesmo grupo. Entretanto, o modelo de regressão usual, com um nível, não é apropriado quando se tem dados organizados em hierarquia, pois características relacionadas a escolas podem ter um impacto razoável na variável resposta. Além disso, conforme discutido no Capítulo 1, vários trabalhos na literatura mostraram que, neste caso, a modelagem multinível é mais adequada, pois pode produzir estimativas mais precisas quando se tem dados agrupados de forma hierárquica.

Devido a sua importância para este estudo, apresentamos na próxima seção, uma introdução ao modelo de regressão com dois níveis, com as expressões matemáticas associadas e suas interpretações.

3.2 Modelos de regressão com dois níveis

Suponha que o modelo de regressão simples, com um nível, seja ajustado para dois conjuntos de dados distintos de duas escolas. Em ambos, a variável dependente Y_i é o desempenho escolar do aluno i da oitava série em matemática e a variável independente X_i é o número de reprovações que o aluno i , com $i = 1, \dots, n$, (número de alunos na escola), apresenta em seu histórico escolar, onde $X_i = 0$ indica que o aluno nunca foi reprovado em matemática, $X_i = 1$ representa que o aluno teve apenas uma reprovação em seu histórico e $X_i = 2$ indica que o aluno foi reprovado mais de uma vez em sua vida acadêmica. Utilizando dados reais de duas escolas da cidade de São Carlos, e para essas duas escolas consideradas, percebe-se que as variáveis, desempenho e número de reprovações são, geralmente, inversamente proporcionais, ou

seja, o aluno apresenta maior rendimento quando, em seu histórico, não há nenhuma reprovação. Neste caso, ao ajustar modelos de regressão linear usual para cada uma dessas escolas e supondo que esses alunos estejam em condições similares, é possível que as estimativas dos parâmetros (coeficientes angulares e lineares) de cada um dos modelos sejam diferentes, como se observa na Figura 3.1. Essa expectativa ocorre pois existe a possibilidade de que aspectos relacionados à escola possam influenciar o rendimento médio em matemática no final do ensino fundamental.

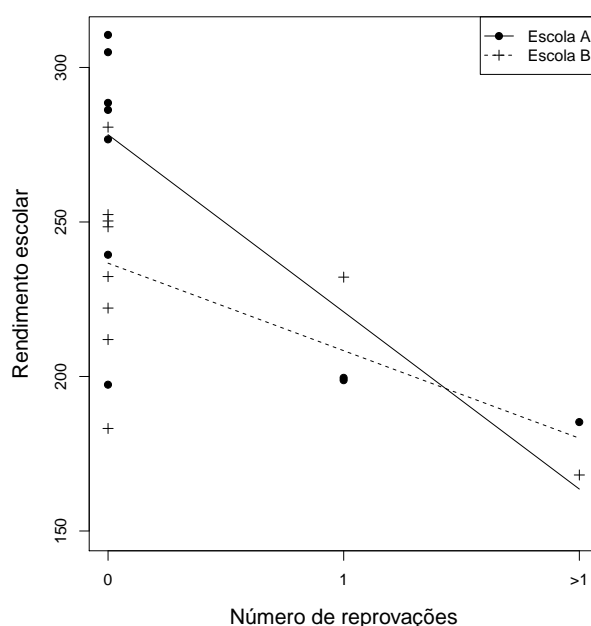


Figura 3.1: Comparação do rendimento escolar médio de duas escolas A e B, para diferentes números de reprovações.

O coeficiente linear (intercepto) de um modelo ajustado indica o desempenho médio dos alunos de cada uma das escolas quando $X_i = 0$, ou seja, representa o rendimento médio dos alunos que nunca foram reprovados de cada uma das escolas. A escola cujo modelo apresentar um coeficiente linear maior, é aquela na qual o desempenho médio dos alunos que nunca foram reprovados é maior. O coeficiente angular, por sua vez, apresenta a variação no desempenho escolar para cada aumento de uma unidade no valor da variável explicativa, ou seja, apresenta o impacto no rendimento escolar quando o número de reprovações aumenta em uma unidade. Assim, maior coeficiente linear e angular podem ser indicativos de que a escola apresenta, respectivamente, maior desempenho médio “inicial” (quando $X_i = 0$) e maior aumento de rendimento com a variação do número de reprovação no histórico escolar.

Entretanto, se o coeficiente angular do modelo ajustado para a escola que apresentar maior intercepto for menor, é possível que o rendimento escolar médio dos alunos das duas escolas se iguale em média em algum momento. Além disso, a escola cujo modelo apresentar o menor coeficiente angular será a escola em que o número de reprovações do aluno exercerá menor impacto em seu desempenho, ou seja, o rendimento desses alunos será menos influenciado pela variação da covariável número de “reprovações”. Outro aspecto relevante é a possibilidade de que haja outros fatores ligados à escola que possam influenciar no aumento ou na redução no rendimento desses alunos.

Na comparação de duas escolas, a covariável se diz menos preditiva para a escola cujo modelo tem menor coeficiente angular. Logo, nessa escola, a relação rendimento e número de reprovações no histórico escolar é mais fraca. Portanto, com o mesmo objetivo que é estudar as variáveis que estão relacionadas com o rendimento do aluno, e considerando o modelo de dois níveis descrito por Natis (2001), que considera que os n alunos tenham sido selecionados de J escolas com n_j alunos cada uma, cujas escolas selecionadas serão incorporadas no modelo que leva em conta a possível variação de interceptos e inclinações entre as escolas, temos:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}, \quad (3.2)$$

em que $i = 1, 2, \dots, n_j$; $j = 1, 2, \dots, J$ e $n = \sum_{j=1}^J n_j$;

Y_{ij} : variável resposta observada no indivíduo i da j -ésima escola;

X_{ij} : variável explicativa observada no indivíduo i da j -ésima escola;

β_{0j} : intercepto, valor esperado do rendimento para o aluno da j -ésima escola no nível 2;

β_{1j} : é a inclinação associada à variável explicativa X_{ij} do i -ésimo aluno no nível 1 para a j -ésima escola no nível 2;

ε_{ij} : erro aleatório associado ao indivíduo i da j -ésima escola ;

As suposições usuais são:

- $E(\varepsilon_{ij}) = 0$,
- $Var(\varepsilon_{ij}) = \sigma_j^2$,
- $Cov(\varepsilon_{ij}, \varepsilon_{kj}) = 0$, para $i \neq k$; $i, k = 1, \dots, n_j$ e $j = 1, \dots, J$.

As suposições mais comuns na literatura são a independência dos elementos dentro do mesmo grupo, nos diferentes níveis. No modelo de regressão tradicional, os coeficientes do ajuste do modelo, β_{0j} e β_{1j} são considerados iguais a β_0 e β_1 , respectivamente, para

todas as escolas. No modelo de regressão multinível, assume-se que os modelos de diferentes escolas podem ter diferentes interceptos e diferentes coeficientes de inclinação e também que os erros são independentes e $\varepsilon_j \sim N_j(0, \sigma^2 I_j)$. Neste caso, uma possível alternativa é utilizar o modelo de regressão multinível em que os coeficientes são variáveis aleatórias e que podem ser modelados da seguinte forma:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}, \quad (3.3)$$

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_j + b_{qj}, \quad (3.4)$$

em que

$q = 0, 1$ e $j = 1, 2, \dots, J$,

γ_{00} e γ_{10} são os valores esperados, respectivamente, do intercepto β_{0j} e da inclinação β_{1j} para a covariável $W_j = 0$ para o nível 2;

γ_{01} e γ_{11} são os coeficientes da regressão associados a covariável explicativa W_j relacionado, respectivamente, ao intercepto β_{0j} e inclinação β_{1j} ;

b_{qj} é o efeito aleatório associado a β_{qj} .

Para este modelo, as suposições para os erros b_{qj} são:

- $E(b_{qj}) = 0$,
- $Var(b_{qj}) = \tau_{qq}$,
- $Cov(b_{qj}, b_{q'j}) = \tau_{qq'}$, com $q \neq q'$ em que,

τ_{00} : variância populacional dos interceptos;

τ_{11} : variância populacional das inclinações;

τ_{01} : covariância entre o elementos de $\mathbf{b}_j = (b_{0j}, b_{1j})'$;

$\mathbf{b}_j \sim N_2(\mathbf{0}, D)$, com D uma matriz de variâncias e covariâncias positiva definida.

A equação (3.4) prevê o rendimento médio para alunos sem reprovação de uma escola quando $q = 0$ e a variação média no rendimento escolar quando $q = 1$. Substituindo (3.4) na equação (3.3) obtemos a expressão

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_{1j}^{(0)} + \gamma_{10}X_{ij} + \gamma_{11}W_{1j}^{(1)}X_{ij} + b_{0j} + b_{1j}X_{ij} + \varepsilon_{ij}; \quad (3.5)$$

Na equação (3.5), $W_{1j}^{(0)}$ e $W_{1j}^{(1)}$ são variáveis preditoras de β_{0j} e β_{1j} , respectivamente.

Utilizando a notação matricial, esse modelo assume, para uma escola j , a forma que segue:

$$\mathbf{Y}_j = X_j W_j \boldsymbol{\gamma} + X_j \mathbf{b}_j + \boldsymbol{\varepsilon}_j. \quad (3.6)$$

Em que:

$$\mathbf{Y}_j = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{n_jj} \end{bmatrix}; \quad X_j = \begin{bmatrix} 1 & X_{(1)1j} \\ 1 & X_{(1)2j} \\ \vdots & \vdots \\ 1 & X_{(1)n_jj} \end{bmatrix}; \quad \boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_jj} \end{bmatrix}; \quad \boldsymbol{\gamma}_j = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix};$$

$$W_j = \begin{bmatrix} 1 & W_{1j}^{(0)} & 0 & 0 \\ 0 & 0 & 1 & W_{1j}^{(1)} \end{bmatrix}; \quad \mathbf{b}_j = \begin{bmatrix} b_{0j} \\ b_{1j} \end{bmatrix} \text{ e } D = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix},$$

Os efeitos aleatórios \mathbf{b}_j permitem aos modelos levar em conta a possível correlação existente entre grupos, possibilitando assim o ajuste de diferentes modelos para dados correlacionados, mas ao mesmo tempo representam um deslocamento em relação ao modelo geral ajustado.

No Capítulo 6, apresentaremos aplicações de alguns desses modelos.

3.3 Modelo de regressão multinível com t covariáveis

Sejam $X_{(m)ij}$, com $m = 1, \dots, p$, $i = 1, \dots, n_j$ e $j = 1, \dots, J$, e $W_{kj}^{(q)}$, com $k = 1, \dots, s$, e $q = 0, 1, \dots, p$, covariáveis dos primeiro e segundo níveis, respectivamente. Temos que o modelo de dois níveis com essas $t = p + s$ covariáveis é dado por:

$$\text{(nível 1)} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{(1)ij} + \beta_{2j}X_{(2)ij} + \dots + \beta_{pj}X_{(p)ij} + \varepsilon_{ij},$$

$$\text{(nível 2)} \quad \beta_{qj} = \gamma_{q0} + \sum_{k=1}^s \gamma_{qk}W_{kj}^{(q)} + b_{qj},$$

em que $i = 1, 2, \dots, n_j$, $j = 1, 2, \dots, J$ e $q = 0, 1, \dots, p$.

Neste modelo, é usual supor que $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, $b_{qj} \sim N(0, \sigma_b^2)$, b_{qj}' s são independentes entre si e b_{qj}' s são independentes dos erros ε_{ij}' s. Usando a notação matricial, o modelo pode ser reescrito como:

$$\text{(nível 1)} \quad \mathbf{Y}_j = X_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j,$$

$$\text{(nível 2)} \quad \boldsymbol{\beta}_j = W_j \boldsymbol{\gamma} + \mathbf{b}_j,$$

ou, realizando a devida substituição:

$$\mathbf{Y}_j = X_j W_j \boldsymbol{\gamma} + X_j \mathbf{b}_j + \boldsymbol{\varepsilon}_j \quad (3.7)$$

onde, para a j -ésima escola, tem-se que:

$\mathbf{Y}_j = (Y_{1j}, \dots, Y_{n_jj})'$ é o vetor ($n_j \times 1$) de variáveis resposta dos alunos da escola j ,

X_j é a matriz $n_j \times (p + 1)$ de variável explicativa no nível 1,

$\beta_j = (\beta_{0j}, \dots, \beta_{pj})'$ é um vetor $((p+1) \times 1)$ de parâmetros desconhecidos,

$\gamma = (\gamma^{(0)}, \dots, \gamma^{(p)})'$ é o vetor $((s+1)(p+1) \times 1)$ de efeitos fixos,

$\varepsilon_j \sim N(\mathbf{0}, \sigma_j^2 I_{n_j})$ é o vetor $(n_j \times 1)$ dos erros aleatórios,

$b_j \sim N(\mathbf{0}, \sigma_b^2)$ é o vetor $((p+1) \times 1)$ dos efeitos aleatórios associado ao nível 2,

W_j é a matriz de dimensão $(p+1) \times (s+1)(p+1)$, em que seus elementos são nulos ou da forma $W_j^{(k)}$, que é um vetor de preditores de β_{qj} de dimensão $1 \times (s+1)$ e $W_{sj}^{(k)}$ é a s -ésima covariável do segundo nível. Desta forma, diz-se que:

$$\begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{n_j j} \end{bmatrix} = \begin{bmatrix} 1 & X_{(1)1j} & \cdots & X_{(p)1j} \\ 1 & X_{(1)2j} & \cdots & X_{(p)2j} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{(1)n_j j} & \cdots & X_{(p)n_j j} \end{bmatrix} \begin{bmatrix} W_j^{(0)} & 0 & \cdots & 0 \\ 0 & W_j^{(1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_j^{(p)} \end{bmatrix} \begin{bmatrix} \gamma^{(0)} \\ \gamma^{(1)} \\ \vdots \\ \gamma^{(p)} \end{bmatrix} + \begin{bmatrix} 1 & X_{(1)1j} & \cdots & X_{(p)1j} \\ 1 & X_{(1)2j} & \cdots & X_{(p)2j} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{(1)n_j j} & \cdots & X_{(p)n_j j} \end{bmatrix} \begin{bmatrix} b_{0j} \\ b_{1j} \\ \vdots \\ b_{pj} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j j} \end{bmatrix};$$

$$W_j^{(k)} = \begin{bmatrix} 1 & W_{1j}^{(k)} & \cdots & W_{sj}^{(k)} \end{bmatrix}, k = 0, 1, \dots, p$$

$$\gamma^{(k)} = \begin{bmatrix} \gamma_{k0} & \gamma_{k1} & \cdots & \gamma_{ks} \end{bmatrix}'.$$

3.4 Estimação dos parâmetros

Considere o modelo com p variáveis independentes no primeiro nível:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \dots + \beta_{pj}X_{pij} + \varepsilon_{ij},$$

para $i = 1, 2, \dots, n_j$, $j = 1, 2, \dots, J$. O processo para encontrar as estimativas $\widehat{\beta}_{kj}$, $k = 0, \dots, p$, $j = 1, 2, \dots, J$, dos parâmetros desconhecidos β_{kj} , é comumente chamado de ajuste do modelo aos dados.

O método de máxima verossimilhança é um dos mais utilizados para se obter as estimativas dos coeficientes nos modelos multiníveis, pois tem a vantagem de produzir estimativas que são assintoticamente eficientes e consistentes (Hox, 2010), quando as condições de regularidades estão satisfeitas, por exemplo, o parâmetro β é identificável, $f(y|\beta)$ é diferenciável em β , o espaço paramétrico contém um conjunto aberto ω do

qual o verdadeiro valor do parâmetro está no interior de ω , a densidade $f(y|\boldsymbol{\beta})$ é três vezes diferenciável em relação a $\boldsymbol{\beta}$. Mais detalhes e informações a respeito das condições de regularidades podem ser vistos em Casella and Berger (2001).

No método de máxima verossimilhança, podem ser usadas duas funções, uma é chamada de função de verossimilhança completa, onde os coeficientes da regressão e os componentes da variância são incluídos na função de verossimilhança. A outra função é chamada de função de verossimilhança restrita, no qual os componentes da variância são incluídos na função de verossimilhança e os coeficientes da regressão são estimados na segunda etapa.

Para cada escola j , o modelo utilizado pode ser representado segundo a equação (3.7). Consequentemente, temos que $\mathbf{Y}_j|\mathbf{b}_j \sim N(X_j W_j \boldsymbol{\gamma}, \sigma^2 I_{n_j n_j})$, onde $\boldsymbol{\gamma}$ é o vetor dos efeitos fixos, \mathbf{b}_j é o vetor dos efeitos aleatórios, X_j é a matriz de variáveis explicativas no nível 1, W_j é a matriz descrita na Seção 3.3 e $\mathbf{b}_j \sim N_q(\mathbf{0}, D)$ com

$$D = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1q} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2q} \\ \vdots & \cdots & \ddots & \vdots \\ \tau_{q0} & \tau_{q1} & \cdots & \tau_{qq} \end{bmatrix}$$

$\boldsymbol{\varepsilon}_j$ é o vetor de erros aleatórios e $\boldsymbol{\varepsilon}_j \sim N_{n_j}(\mathbf{0}, \Psi_j)$ onde, $\Psi_j = \sigma^2 I_{n_j n_j}$ e I é a matriz identidade.

Para considerar dados das J escolas, podemos escrever um único modelo na forma:

$$\mathbf{Y} = XW\boldsymbol{\gamma} + X\mathbf{b} + \boldsymbol{\varepsilon} = X(W\boldsymbol{\gamma} + \mathbf{b}) + \boldsymbol{\varepsilon},$$

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ com } \boldsymbol{\beta} = W\boldsymbol{\gamma} + \mathbf{b},$$

onde \mathbf{Y} é um vetor de dimensão n , W é uma matriz do tipo $(p+1)J \times (p+1)(s+1)$, $\boldsymbol{\gamma}$ e \mathbf{b} são vetores de tamanhos $(s+1)(p+1)$ e $(p+1)J$, respectivamente e $\boldsymbol{\varepsilon}$ vetor de tamanho n .

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix}, X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_J \end{bmatrix}, W = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_J \end{bmatrix}, \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}^{(0)} \\ \boldsymbol{\gamma}^{(1)} \\ \vdots \\ \boldsymbol{\gamma}^{(p)} \end{bmatrix},$$

$$\boldsymbol{\gamma}^{(k)} = \begin{bmatrix} \gamma_{k0} \\ \gamma_{k1} \\ \vdots \\ \gamma_{ks} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_J \end{bmatrix}, \mathbf{b}_j = \begin{bmatrix} b_{0j} \\ b_{1j} \\ \vdots \\ b_{pj} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{bmatrix} \text{ e } \boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j} \end{bmatrix}$$

Assim, temos que, $\mathbf{Y} \sim N_{n_j}(XW\boldsymbol{\gamma}, \Sigma)$ e $\Sigma = XDX' + \Psi$, com $\Psi = \sigma^2 I$.

O procedimento para estimar os parâmetros de um modelo multinível é um tanto complexo, portanto, será descrito o processo para um modelo hierárquico com 2 níveis, tendo em vista que a de nível superior pode ser feita de maneira similar.

Seja \mathbf{Y}_j o vetor de variáveis respostas do j -ésimo grupo e \mathbf{b}_j o vetor dos efeitos aleatórios. A distribuição conjunta de \mathbf{Y}_j e \mathbf{b}_j é normal multivariada, ou seja:

$$\begin{bmatrix} \mathbf{Y}_j \\ \mathbf{b}_j \end{bmatrix} \sim N_{n_j+q+1} \left\{ \begin{bmatrix} X_j W_j \boldsymbol{\gamma} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} X_j D X_j' + \Psi_j & X_j D \\ D X_j' & D \end{bmatrix} \right\} \quad \text{em que,}$$

$$\text{Cov}(\mathbf{Y}_j, \mathbf{b}_j) = \text{Cov} [(X_j W_j \boldsymbol{\gamma} + X_j \mathbf{b}_j + \boldsymbol{\varepsilon}_j) \mathbf{b}_j] = X_j \text{Cov}(\mathbf{b}_j, \mathbf{b}_j) + \text{Cov}(\boldsymbol{\varepsilon}_j, \mathbf{b}_j) = X_j D.$$

Uma possibilidade para a estimação dos parâmetros é considerar o modelo multinível como modelo misto e utilizar os métodos usuais de estimação. Por exemplo, pode-se considerar o modelo marginal de \mathbf{Y} , o que possibilita o ajuste do modelo pelo método escore de Fisher. Assim, no modelo marginal, a distribuição de \mathbf{Y}_j é dada por:

$$\mathbf{Y}_j \sim N_{n_j}(X_j W_j \boldsymbol{\gamma}, \Sigma_j = X_j D X_j' + \Psi_j).$$

Seja $f(\mathbf{y}_j, \boldsymbol{\theta}) = (2\pi)^{-\frac{n_j}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma}) \right\}$, a função densidade de probabilidade e $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\tau})$.

Note que os elementos de Σ_j são funções dos parâmetros de $\boldsymbol{\tau} = (\tau_0, \tau_{rs})$, em que $\tau_0 = \sigma^2$ e τ_{rs} é o rs -ésimo elemento de D , para $r, s = 1, 2, \dots, q$.

Seja $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança para $\boldsymbol{\theta}$, então:

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^J L_j(\boldsymbol{\theta}) \propto -\frac{1}{2} \sum_{j=1}^J \{ \log |\Sigma_j| + u_j \}, \quad \text{onde } u_j = (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma}).$$

As estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ podem ser obtidas por meio das soluções das equações:

$$\mathbf{U}_\gamma = \mathbf{0},$$

$\mathbf{U}_\tau = \mathbf{0}$, onde \mathbf{U}_γ e \mathbf{U}_τ são as funções escore, cujas expressões serão dadas na próxima subseção.

3.4.1 Funções Escore

Como é complicado obter estimadores explícitos a partir das equações dadas, vamos considerar o método iterativo escore de Fisher, que consiste em determinar as soluções do sistema

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_\gamma \\ \mathbf{U}_\tau \end{bmatrix} = \mathbf{0}.$$

A diferença desse método para o também bastante utilizado Newton-Rapshon é que o método iterativo de Newton-Rapshon utiliza a matriz de informação observada enquanto que o método escore de Fisher, utiliza matriz de informação de Fisher.

As funções escore para $\boldsymbol{\theta}$ são determinadas derivando o logaritmo da função de verossimilhança em relação a cada elemento de $\boldsymbol{\theta}$. Deste modo, obtemos a expressão da função escore para $\boldsymbol{\gamma}$ da seguinte maneira:

$$\mathbf{U}_\gamma = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = \sum_{j=1}^J \frac{-1}{2} \frac{\partial u_j}{\partial \boldsymbol{\gamma}},$$

$$\mathbf{U}_\gamma = \sum_{j=1}^J -\frac{1}{2} \{ \Sigma_j^{-1} [(\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' (-X_j W_j) + (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}) (-W_j' X_j')] \},$$

$$\mathbf{U}_\gamma = \sum_{j=1}^J -\frac{1}{2} [-2 \Sigma_j^{-1} W_j' X_j' (\mathbf{Y}_j - W_j X_j \boldsymbol{\gamma})] = \sum_{j=1}^J W_j' X_j' \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}),$$

Para simplificar, considere $\mathbf{r}_j = \mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}$ para obter:

$$\mathbf{U}_\gamma = \sum_{j=1}^J W_j' X_j' \Sigma_j^{-1} \mathbf{r}_j. \quad (3.8)$$

A função escore para $\boldsymbol{\tau}$ é dada por:

$$\mathbf{U}_\tau = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}}, \text{ onde } \mathbf{U}_\tau = (U_{\tau_0}, U_{\tau_1}, \dots, U_{\tau_q}), \text{ então:}$$

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \frac{\partial \log |\Sigma_j|}{\partial \tau_k} + \sum_{j=1}^J \left(-\frac{1}{2} \frac{\partial u_j}{\partial \tau_k} \right),$$

Como $\frac{\partial \log |\Sigma_j|}{\partial \tau_k} = \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \right]$, obtemos:

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \right] + \sum_{j=1}^J -\frac{1}{2} \left\{ \frac{\partial [(\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})]}{\partial \tau_k} \right\},$$

Considerando $\mathbf{r}_j = (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})$ obtemos a expressão:

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \right] + \frac{1}{2} \sum_{j=1}^J [\mathbf{r}_j' \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1} \mathbf{r}_j],$$

E finalmente, como $\frac{\partial \Sigma_j^{-1}}{\partial \tau_k} = -\frac{\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1}}{\partial \tau_k}$, escrevemos a função escore para $\boldsymbol{\tau}$ da seguinte maneira:

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \left\{ \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \right] - \mathbf{r}_j' \frac{\partial \Sigma_j^{-1}}{\partial \tau_k} \mathbf{r}_j \right\}.$$

Maiores detalhes podem ser visto em Osorio (2006) e Savalli (2005).

3.4.2 Matriz de Informação de Fisher

A matriz de informação de Fisher é obtida por meio da expressão:

$$K_{\theta\theta} = -E \left(\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \right)$$

e assume a forma bloco diagonal $K_{\theta\theta} = D(K_{\gamma\gamma}, K_{\tau\tau})$ ou seja, é da forma:

$$K_{\theta\theta} = \begin{bmatrix} K_{\gamma\gamma} & \mathbf{0} \\ \mathbf{0} & K_{\tau\tau} \end{bmatrix},$$

em que

$$K_{\gamma\gamma} = \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i, \quad (3.9)$$

$K_{\tau\tau} = [K_{\tau_r\tau_s}]$ e o rs -ésimo elemento de $K_{\tau\tau}$ é dado pela expressão:

$$K_{\tau_r\tau_s} = \sum_{i=1}^J \frac{1}{2} \text{tr} \left[\Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_r} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_s} \right]. \quad (3.10)$$

em que $\frac{\partial \Sigma_i}{\partial \tau_k}$ representa uma matriz de derivadas de cada elemento de Σ_i em relação ao parâmetro τ_k .

3.4.3 Processo de estimação dos parâmetros

Nesta seção, apresentamos o processo para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ considerando que os dados têm distribuição normal. O processo consiste, basicamente, em duas etapas e maiores detalhes podem ser vistos em Osorio (2006).

O método iterativo que utiliza a matriz de informação de Fisher, consiste em determinar as soluções do sistema $\begin{bmatrix} U_{\boldsymbol{\gamma}} \\ U_{\boldsymbol{\tau}} \end{bmatrix} = \mathbf{0}$, em que:

$$U_{\boldsymbol{\gamma}} = \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} (\mathbf{Y}_i - X_i W_i \boldsymbol{\gamma}) \quad \text{e} \quad U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \left\{ \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \right] - \mathbf{r}_j' \frac{\partial \Sigma_j^{-1}}{\partial \tau_k} \mathbf{r}_j \right\}.$$

Assim, para obter as estimativas dos parâmetros referentes aos efeitos fixos e dos componentes de variância, podemos utilizar o algoritmo score de Fisher que se resume em duas etapas:

Etapa 1.

Atualizar $\boldsymbol{\gamma}^{(r+1)}$ como:

$$\hat{\boldsymbol{\gamma}}^{(r+1)} = \hat{\boldsymbol{\gamma}}^{(r)} + K_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1} U_{\hat{\boldsymbol{\gamma}}}. \quad (3.11)$$

Substituindo as equações (3.9) e (3.8) na equação acima, obtemos:

$$\begin{aligned} \hat{\boldsymbol{\gamma}}^{(r+1)} &= \hat{\boldsymbol{\gamma}}^{(r)} + \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{r}_i \right) \\ &= \hat{\boldsymbol{\gamma}}^{(r)} + \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} (\mathbf{Y}_i - X_i W_i \hat{\boldsymbol{\gamma}}^{(r)}) \right) \\ &= \hat{\boldsymbol{\gamma}}^{(r)} + \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{Y}_i - \sum_{i=1}^J X_i' W_i' \Sigma_i^{-1} X_i W_i \hat{\boldsymbol{\gamma}}^{(r)} \right) \\ &= \hat{\boldsymbol{\gamma}}^{(r)} + \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{Y}_i - \\ &\quad \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right) \hat{\boldsymbol{\gamma}}^{(r)} \\ &= \hat{\boldsymbol{\gamma}}^{(r)} + \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{y}_i - \hat{\boldsymbol{\gamma}}^{(r)} \\ &= \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{y}_i. \end{aligned}$$

Das operações realizadas, verificamos que a equação (3.11) é equivalente a seguinte expressão:

$$\hat{\boldsymbol{\gamma}}^{(r+1)} = \left(\sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} X_i W_i \right)^{-1} \sum_{i=1}^J W_i' X_i' \Sigma_i^{-1} \mathbf{Y}_i, \quad (3.12)$$

em que $\Sigma_i = X_i D X_i' + \Psi_j$ e o índice $r + 1$, indica o termo da $(r + 1)$ -ésima iteração.

Etapa 2. Atualizar as estimativas de $\boldsymbol{\tau}^{(r+1)}$ como:

$$\hat{\boldsymbol{\tau}}^{(r+1)} = \hat{\boldsymbol{\tau}}^{(r)} + (K_{\boldsymbol{\tau}\boldsymbol{\tau}})^{-1} U_{\hat{\boldsymbol{\tau}}}, \quad (3.13)$$

em que, $K_{\boldsymbol{\tau}\boldsymbol{\tau}} = [K_{\tau_r \tau_s}]$, onde $K_{\tau_r \tau_s} = \sum_{i=1}^J \frac{1}{2} \text{tr} \left[\Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_r} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_s} \right]$, e

$$U_{\tau_k} = -\frac{1}{2} \sum_{i=1}^J \left\{ \text{tr} \left[\Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \right] - \mathbf{r}_i' \frac{\partial \Sigma_i^{-1}}{\partial \tau_k} \mathbf{r}_i \right\}.$$

Deve-se repetir as etapas 1 e 2 até atingir convergência.

Uma possibilidade para os valores iniciais de $\boldsymbol{\gamma}^{(0)}$ e $\boldsymbol{\tau}^{(0)}$ no modelo gaussiano é utilizar as estimativas de máxima verossimilhança considerando o modelo de regressão linear e para os modelos não normais, os valores iniciais podem ser as estimativas obtidas do modelo sob normalidade.

Modelos de regressão multiníveis simétricos

No presente capítulo propomos o uso de modelos de regressão multiníveis simétricos, apresentamos algumas de suas propriedades, incluindo a definição da variável aleatória com distribuição simétrica e apresentamos algumas das distribuições pertencentes a classe simétrica e também a forma do modelo multinível simétrico.

4.1 Modelos simétricos

A suposição de normalidade na distribuição dos erros em alguns modelos nem sempre é adequada, pois em alguns casos os dados podem ter distribuições com caudas mais leves ou mais pesadas que as da distribuição normal. Além disso, observações atípicas podem influenciar de forma desproporcional um ajuste sob normalidade. Durante algum tempo, uma das alternativas mais comuns era utilizar algum artifício, por exemplo, uma transformação na variável resposta para obter pelo menos a simetria. Entretanto, estudos mostraram que, mesmo que fossem ajustados modelos com erros normalmente distribuídos, as estimativas obtidas para os coeficientes do modelo eram sensíveis a observações extremas, que são conhecidas na literatura, como aberrantes ou *outliers*. Tais estudos motivaram o desenvolvimento de metodologias robustas

contra essas observações extremas que não eram explicados pelo modelo. Dentre as metodologias adotadas, destacam-se os que assumem distribuições simétricas para os componentes aleatórios do modelo, que minimizam a influência de *outliers*, ou seja, assume, para os erros, distribuições com caudas mais pesadas ou mais leve que as da distribuição normal. Grande parte desses resultados podem ser encontrados em Fang and Anderson (1990).

A classe das distribuições simétricas inclui as distribuições t-Student, exponencial potência, logística, normal contaminada, normal, entre outras.

A Figura 4.1 apresenta os gráficos das densidades das distribuições simétricas t-Student com $\nu = 1$ e $\nu = 5$ e da exponencial potência com $\lambda = 0,8$ e $\lambda = 1,4$, comparadas com o gráfico da distribuição normal, que podem minimizar a influência das observações aberrantes. Em particular, no caso da t-Student, um parâmetro adicional, ν , o número de graus de liberdade, controla a curtose da distribuição e no caso exponencial potência, o parâmetro de forma que controla o peso das caudas. Esses parâmetros devem ser estimados ou fixados.

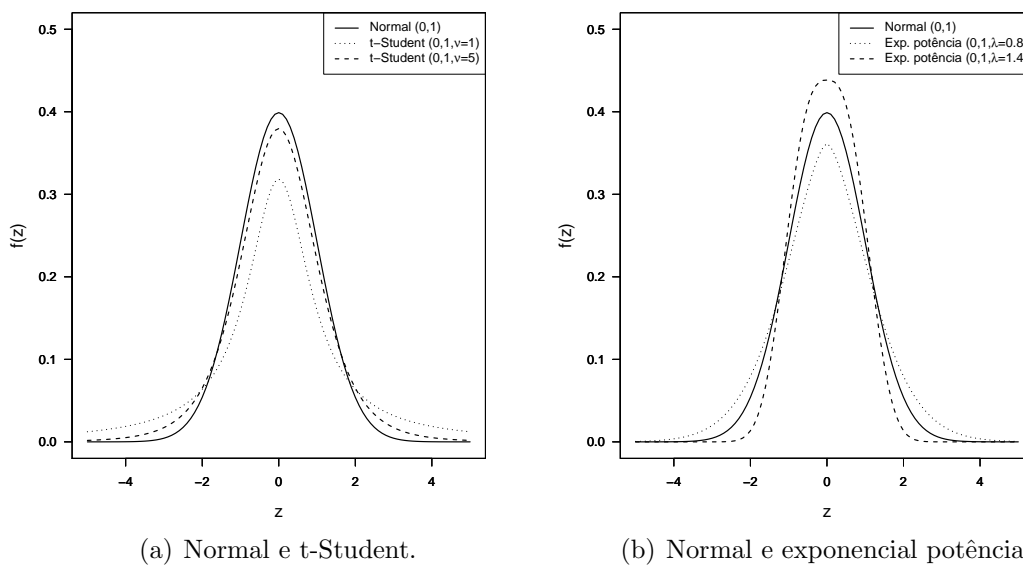


Figura 4.1: Gráficos da função de densidade de probabilidades de algumas distribuições simétricas.

Em Russo (2010), a autora faz uma discussão sobre as distribuições com contornos elípticos e introduz os modelos elípticos não lineares com efeitos mistos. O modelo simétrico é um caso particular e portanto, maiores detalhes podem ser vistos no referido trabalho.

Cysneiros et al. (2005) trabalham com a variável aleatória contínua simétrica e discutem as principais propriedades dessas distribuições, dentre elas a distribuição normal, que é a mais conhecida e amplamente utilizada nos estudos de modelagem estatística aplicada em diversas áreas.

A seguir, apresentamos a definição da variável aleatória que pertence à classe das distribuições simétricas e a definição de uma variável aleatória com distribuição elíptica com as notações mais comuns utilizadas na literatura..

Definição 4.1.1. *Seja Y uma variável aleatória com suporte em \mathbb{R} , com parâmetros de locação $\mu \in \mathbb{R}$ e de escala $\sigma^2 > 0$ com função de densidade de probabilidade dada por*

$$f(y, \mu, \sigma^2) = \frac{1}{\sqrt{\sigma^2}} g(u), \quad y \in \mathbb{R} \quad (4.1)$$

para alguma função $g(\cdot)$ denominada função geradora de densidade, em que $u = \frac{(y-\mu)^2}{\sigma^2}$, com $g(u) > 0$, para $u > 0$, e $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$. Nessas condições, dizemos que Y é uma variável aleatória simétrica e denotamos por $Y \sim S(\mu, \sigma^2)$.

Definição 4.1.2. (*distribuição elíptica*) *Dizemos que um vetor m -dimensional Y_m tem distribuição elíptica multivariada $\mathbf{Y} \sim El_m(\boldsymbol{\mu}, \Sigma, g)$, com parâmetro de posição $\boldsymbol{\mu} \in \mathbb{R}^m$ e matriz de escala Σ (positiva definida), se sua função densidade é dada por:*

$$f(\mathbf{y}) = |\Sigma|^{-\frac{1}{2}} g[(\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})] \quad (4.2)$$

em que $g : \mathbb{R} \rightarrow [0, \infty)$ é tal que $\int_0^\infty u^{\frac{m}{2}-1} g(u) du < \infty$ e g é conhecida como a função geradora de densidades.

A distribuição simétrica mais conhecida e utilizada é a Normal cuja função geradora de densidades é da forma

$$g(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u}{2}\right), \quad u > 0$$

As funções geradoras de densidades para alguns modelos simétricos são apresentadas na Tabela 4.1, onde para cada distribuição, os valores de $c_i, i = 1, \dots, k$, são determinadas considerando as respectivas distribuições e cujas expressões podem ser encontradas em Cysneiros et al. (2005).

Tabela 4.1: Função geradora de densidades de algumas distribuições simétricas.

Distribuição	Função geradora de densidades $g(\mathbf{u})$
t-Student (μ, σ^2, ν)	$c_1 \left(1 + \frac{u}{\nu}\right)^{-(\nu+1)/2}$
t-Student generalizada (μ, σ^2, s, r)	$c_2 \left(1 + \frac{u}{\gamma}\right)^{-(\nu+1)/2}$
Logística I (μ, σ^2)	$c_3 \frac{\exp(-u)}{(1 + \exp(-u))^2}$
Logística II (μ, σ^2)	$c_4 \frac{\exp(-u^{1/2})}{(1 + \exp(-u^{1/2}))^2}$
Logística generalizada (μ, σ^2, m)	$c_5 \left\{ \frac{\exp(-\alpha\sqrt{u})}{(1 + \exp(-\alpha\sqrt{u}))^2} \right\}^m$
Exponencial potência (μ, σ^2, k)	$c_k \exp(-u^{1/(1+k)}/2)$

Os modelos simétricos de regressão são aqueles em que os erros $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são variáveis aleatórias independentes cuja função densidade é da forma dada na Definição 4.1.1.

É usual trabalhar com as quantidades w_g e w'_g definidas como:

$$w_g(u_j) = \frac{\partial \log g(u_j)}{\partial u_j} \quad \text{e} \quad w'_g(u_j) = \frac{\partial w_g(u_j)}{\partial u_j}.$$

Essas quantidades são utilizadas para determinar as estimativas dos parâmetros dos modelos que pressupõem erros com distribuições simétricas. Em particular, as quantidades w_g representam pesos e podem controlar a influência de informações atípicas nas estimativas dos parâmetros, por exemplo.

Nas Tabelas 4.2 e 4.3 apresentamos, respectivamente, alguns exemplos das distribuições pertencentes a classe das distribuições simétricas com as respectivas expressões $w_g(u)$ e $w'_g(u)$. As quantidades c_1, \dots, c_k são constantes.

Tabela 4.2: Expressões de $w_g(u)$ de algumas distribuições simétricas

Distribuição	$w_g(u)$
Normal(μ, Σ)	$-\frac{1}{2}$
t-Student(μ, Σ, ν)	$-\frac{\nu + 1}{2(\nu + u)}$
t-Student generalizada(μ, Σ, r, s)	$-\frac{r + 1}{2(s + u)}$
Logística I(μ, Σ)	$-\tanh\left(\frac{u}{2}\right)$
Logística II(μ, Σ)	$-\frac{\exp(-\sqrt{u}) - 1}{(-2\sqrt{u})[1 + \exp(-\sqrt{u})]}$
Logística generalizada(μ, Σ, m)	$\frac{-\alpha m[\exp(-\alpha\sqrt{u}) - 1]}{(-2\sqrt{u})[1 + \exp(-\alpha\sqrt{u})]}$
Exponencial potência(μ, Σ, k)	$-\frac{1}{2(1 + k)u^{\frac{k}{k+1}}}$

Tabela 4.3: Expressões de $w'_g(u)$ de algumas distribuições simétricas

Distribuição	$w'_g(u)$
Normal(μ, Σ)	0
t-Student(μ, Σ, ν)	$\frac{(\nu + 1)}{2(\nu + u)^2}$
t-Student generalizada(μ, Σ, r, s)	$\frac{(r + 1)}{2(s + u)^2}$
Logística I(μ, Σ)	$-\operatorname{sech}\left(\frac{u}{2}\right)/2$
Logística II(μ, Σ)	$\frac{2 \exp(-\sqrt{u})\sqrt{u} + \exp(-2\sqrt{u}) - 1}{-4u^{\frac{3}{2}}[1 + \exp(-\sqrt{u})]^2}$
Logística generalizada(μ, Σ, m)	$\frac{-\alpha m}{4} \frac{2\alpha \exp(-\alpha\sqrt{u})\sqrt{u} + \exp(-2\alpha\sqrt{u}) - 1}{u^{\frac{3}{2}}[1 + \exp(-\alpha\sqrt{u})]^2}$
Exponencial potência(μ, Σ, k)	$\frac{k}{(1 + k)^2 2u^{\frac{(2k+1)}{(1+k)}}$

Algumas distribuições pertencentes a classe simétrica podem ser vistas como extensões da distribuição Normal, por exemplo, a distribuição t-Student com número de graus de liberdade $\nu \rightarrow \infty$ ou a exponencial potência com parâmetro de forma $\lambda = 1$. Podemos encontrar trabalhos que discutem as propriedades dessas distribuições, por exemplo, em Cysneiros (2004).

4.2 Modelos multiníveis simétricos

Sejam $X_{(m)}$, $m = 1, \dots, p$ e $W_{(k)}$, $k = 1, \dots, s$ covariáveis dos primeiro e segundo níveis, respectivamente. Temos que o modelo de dois níveis com essas $t = p + s$ covariáveis é dado por:

$$(\text{nível 1}) \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{(1)ij} + \beta_{2j}X_{(2)ij} + \dots + \beta_{pj}X_{(p)ij} + \varepsilon_{ij}, \quad (4.3)$$

$$(\text{nível 2}) \quad \beta_{qj} = \gamma_{q0} + \sum_{k=1}^s \gamma_{qk}W_{kj}^{(q)} + b_{qj}, \quad (4.4)$$

em que $i = 1, 2, \dots, n_j$, $j = 1, 2, \dots, J$, $q = 0, 1, \dots, p$,

$\varepsilon_{ij} \sim S(0, \sigma_e^2)$, $b_{qj} \sim S(0, \sigma_b^2)$,

γ_{q0} : valor esperado de β_{qj} quando $\sum_{k=1}^s \gamma_{qk}W_{kj}^{(q)} = 0$;

γ_{qk} : é o coeficiente de regressão associado à variável explicativa $W_{kj}^{(q)}$;

b_{qj} : é o efeito aleatório da j -ésima unidade sobre os parâmetros β_{qj} da regressão.

4.2.1 Processo de estimação dos parâmetros

Devido a complexidade do procedimento para estimar os parâmetros do modelo multinível simétrico para o caso geral, descreveremos o processo para um modelo com dois níveis. O modelo simétrico com dois níveis pode ser visto como um modelo com efeito mistos expresso na forma $\mathbf{Y}_j = X_j W_j \boldsymbol{\gamma} + X_j \mathbf{b}_j + \boldsymbol{\epsilon}_j$, com $j = 1, 2, \dots, J$, como já mencionado na seção 3.4. Aqui, assumimos que \mathbf{b}_j e $\boldsymbol{\epsilon}_j$ são variáveis aleatórias com distribuição simétrica multivariada (elíptica). Assim, similarmente aos modelos elípticos mistos que podem ser vistos em Cysneiros et al. (2005) e Russo (2010), podemos supor que a distribuição conjunta de \mathbf{Y}_j e \mathbf{b}_j é elíptica.

$$\begin{bmatrix} \mathbf{Y}_j \\ \mathbf{b}_j \end{bmatrix} \sim El_{n_j+p+1} \left\{ \begin{bmatrix} X_j W_j \boldsymbol{\gamma} \\ 0 \end{bmatrix}; \begin{bmatrix} \sigma^2 I + X_j D X_j' & X_j D \\ D X_j' & D \end{bmatrix} \right\}, \quad (4.5)$$

em que $i = 1, 2, \dots, J$, X_j é a matriz de covariáveis, $\boldsymbol{\gamma}$ é o vetor dos parâmetros fixos, W_j é a matriz de variáveis que especifica os efeitos aleatórios, \mathbf{b}_j é o vetor dos efeitos aleatórios e D é a matriz de variâncias e covariâncias.

A partir da estrutura hierárquica, podemos trabalhar com o modelo marginal, que tem distribuição simétrica multivariada ou elíptica denotado por $\mathbf{Y}_j \sim El(X_j W_j \boldsymbol{\gamma}, \Sigma_j = X_j D X_j' + \sigma^2 I)$, com função densidade dada pela expressão:

$$f(\mathbf{y}_j, \boldsymbol{\theta}) = |\Sigma_j|^{-\frac{1}{2}} g((\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})),$$

Note que os elementos de Σ_j são funções dos componentes de $\boldsymbol{\theta}$ em que $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\tau})$ onde $\tau_0 = \sigma^2$ e τ_r é o r -ésimo elemento de D , para $r = 1, 2, \dots, q$.

Seja $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança para $\boldsymbol{\theta}$, então:

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^J L_j(\boldsymbol{\theta}) = \sum_{j=1}^J \left[-\frac{1}{2} \log |\Sigma_j| + \log g(u_j) \right],$$

em que g é uma função contínua e diferenciável,

$$u_j = (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma}),$$

$$w_g(u_j) = \frac{\partial \log g(u_j)}{\partial u_j} = \frac{g'(u_j)}{g(u_j)},$$

$$w'_g(u_j) = \frac{\partial w_g(u_j)}{\partial u_j}.$$

A estimativa de máxima verossimilhança dos componentes $\boldsymbol{\theta}$ pode ser obtida através das soluções das equações $\mathbf{U}_\gamma = \mathbf{0}$ e $\mathbf{U}_\tau = \mathbf{0}$, onde \mathbf{U}_γ e \mathbf{U}_τ são chamadas de funções escore. Pela complexidade de obter estimadores explícitos a partir das equações dadas, vamos considerar o método iterativo escore de Fisher, que consiste em determinar as soluções do sistema $\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_\gamma \\ \mathbf{U}_\tau \end{bmatrix} = \mathbf{0}$, utilizando no procedimento a matriz de informação de Fisher.

As funções escore para $\boldsymbol{\theta}$ são determinadas derivando logaritmo da função de verossimilhança em relação a cada elemento de $\boldsymbol{\theta}$. Deste modo, obtemos a expressão da função escore para $\boldsymbol{\gamma}$ da seguinte maneira:

$$\mathbf{U}_\gamma = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = \sum_{j=1}^J \frac{g'(u_j)}{g(u_j)} \frac{\partial u_j}{\partial \boldsymbol{\gamma}}$$

$$\mathbf{U}_\gamma = \sum_{j=1}^J \frac{g'(u_j)}{g(u_j)} \{ \Sigma_j^{-1} [(\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' (-X_j W_j) + (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}) (-W_j' X_j')] \}$$

$$\mathbf{U}_\gamma = \sum_{j=1}^J \frac{g'(u_j)}{g(u_j)} \{ -2 \Sigma_j^{-1} W_j' X_j' (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}) \} = \sum_{j=1}^J W_g \{ -2 W_j' X_j' \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}) \}$$

Para simplificar, considere $q_j(u_j) = -2w_g(u_j)$ e $\mathbf{r}_j = \mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}$ para obter:

$$\mathbf{U}_\gamma = \sum_{j=1}^J q_j(u_j) W_j' X_j' \Sigma_j^{-1} \mathbf{r}_j$$

$$\hat{\boldsymbol{\gamma}} = q_j(u_j) W_j' X_j' \Sigma_j^{-1} \mathbf{Y}_j [q_j(u_j) W_j' X_j' \Sigma_j^{-1} X_j W_j]^{-1} \quad (4.6)$$

A função escore para $\boldsymbol{\tau}$ é dada por:

$$\mathbf{U}_\tau = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \text{ onde } \mathbf{U}_\tau = (U_{\tau_0}, U_{\tau_1}, \dots, U_{\tau_q}), \text{ então:}$$

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \frac{\partial \log |\Sigma_j|}{\partial \tau_k} + \sum_{j=1}^J \frac{g'(u_j)}{g(u_j)} \frac{\partial u_j}{\partial \tau_k}$$

Como $\frac{\partial \log |\Sigma_j|}{\partial \tau_k} = \text{tr} [\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k}]$, obtemos

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^{n_j} \text{tr} [\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k}] + \sum_{j=1}^J W_g(u_j) \frac{\partial [(\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})]}{\partial \tau_k}.$$

Para simplificar, foi considerado $q_j(u_j) = -2w_g(u_j)$, então, $w_g(u_j) = -\frac{1}{2}q_j(\boldsymbol{\theta})$ e portanto,

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \text{tr} [\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k}] - \frac{1}{2} \sum_{i=1}^J q_j(u_j) \frac{[(\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' \partial \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})]}{\partial \tau_k}$$

Considerando $\mathbf{r}_j = (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})$ obtemos a expressão

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \text{tr} [\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k}] + \frac{1}{2} \sum_{j=1}^J q_j(u_j) [\mathbf{r}_j' \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1} \mathbf{r}_j]$$

E finalmente, como $\frac{\partial \Sigma_j^{-1}}{\partial \tau_k} = -\frac{\Sigma_j^{-1} \partial \Sigma_j \Sigma_j^{-1}}{\partial \tau_k}$, escrevemos a função escore para $\boldsymbol{\tau}$ da seguinte maneira:

$$U_{\tau_k} = -\frac{1}{2} \sum_{j=1}^J \{ \text{tr} [\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_k}] - q_j(u_j) \mathbf{r}_j' \frac{\partial \Sigma_j^{-1}}{\partial \tau_k} \mathbf{r}_j \}.$$

E para estimar os parâmetros referentes aos efeitos fixos e os componentes de variância, pode-se usar o processo iterativo que utiliza a matriz de informação de Fisher, dado por:

$$\hat{\boldsymbol{\gamma}}^{(r+1)} = \left[\sum_{j=1}^n q_j^{(r)} W_j' X_j' \Sigma_j^{-(r)} X_j W_j \right]^{-1} \left[\sum_{j=1}^n q_j^{(r)} W_j' X_j' \Sigma_j^{-(r)} \mathbf{Y}_j \right],$$

e

$$\hat{\boldsymbol{\tau}}^{(r+1)} = \hat{\boldsymbol{\tau}}^{(r)} + (K_{\tau\tau})^{-1} \mathbf{U}_{\hat{\boldsymbol{\tau}}},$$

para $r = 0, 1, 2, \dots$; $\boldsymbol{\tau} = (\sigma^2, \boldsymbol{\tau}')'$; $u_j = (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})' \Sigma_j^{-1} (\mathbf{Y}_j - X_j W_j \boldsymbol{\gamma})$; $q_j = -2 \frac{g'(u_j)}{g(u_j)} = -2w_g(u_j)$ e $l(\boldsymbol{\theta})$ é o logaritmo da função de verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\tau}')'$. Os parâmetros $\boldsymbol{\gamma}$ e $\boldsymbol{\tau}$ são ortogonais e portanto, a matriz de informação de Fisher para o vetor $\boldsymbol{\theta}$ é da forma bloco diagonal $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = D(K_{\boldsymbol{\gamma}\boldsymbol{\gamma}}, K_{\boldsymbol{\tau}\boldsymbol{\tau}})$ ou seja, é da forma:

$$K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} K_{\boldsymbol{\gamma}\boldsymbol{\gamma}} & 0 \\ 0 & K_{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix},$$

em que

$$K_{\boldsymbol{\gamma}\boldsymbol{\gamma}} = \sum_{j=1}^J \frac{4d_{gi}}{m_j} W_j' X_j' \Sigma_j^{-1} X_j W_j, \quad (4.7)$$

$$K_{\boldsymbol{\tau}\boldsymbol{\tau}} = [K_{\tau_r \tau_s}],$$

$d_{gi} = E[w_g^2(u_j) \mathbf{U}_j]$, $\mathbf{U}_j = \|Z_j\|^2$, $Z_j = \Sigma_j^{-\frac{1}{2}} (\mathbf{y}_j - X_j W_j \boldsymbol{\gamma})$ e o rs -ésimo elemento de $K_{\boldsymbol{\tau}\boldsymbol{\tau}}$ é dado pela próxima expressão

$$K_{\tau_r \tau_s} = \sum_{j=1}^J \left\{ \frac{a_{rsi}}{4} \left(\frac{4f_{gi}}{m_j(m_j+2)} - 1 \right) + \frac{2f_{gi}}{m_j(m_j+2)} \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_r} \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_s} \right] \right\}, \quad (4.8)$$

onde

$$f_{gi} = E\{w_g^2(u_i) \mathbf{U}_i^2\}$$

$$a_{rsi} = \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_r} \right] \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_s} \right] \text{ com } r, s = 1, \dots, q+2.$$

As formas d_{gi} e f_{gi} das distribuições normal, t-Student e exponencial potência encontram-se na tabela 4.4 e mais detalhes sobre as expressões dos elementos da matriz de informação de Fisher podem ser vistas em Russo et al. (2011).

4.3 Seleção de modelos

Após o ajuste do modelo, é preciso selecionar o melhor modelo de regressão linear multinível entre os ajustados para explicar o desempenho dos alunos nas avaliações de Matemática. Existem muitos critérios para a seleção de modelos e utilizando tais critérios podemos encontrar um modelo mais parcimonioso, ou seja, um modelo que envolva o menor número de parâmetros, mas que seja mais próximo possível da realidade.

Tabela 4.4: Expressões de d_{gi} e f_{gi} de algumas distribuições elípticas.

Distribuição	d_{gi}	f_{gi}
Normal $N_m(\mu, \Sigma)$	$\frac{m_i}{4}$	$\frac{m_i(m_i + 2)}{4}$
t-Student $St_m(\mu, \Sigma, \nu)$	$\frac{m_i(\nu + m_i)}{4(\nu + m_i + 2)}$	$\frac{m_i(m_i + 2)(\nu + m_i)}{4(\nu + m_i + 2)}$
Exponencial potência $EP_m(\mu, \Sigma, \lambda)$	$\frac{\lambda^2}{2^{1/\lambda}} \frac{\Gamma(\frac{m_i-2}{2\lambda} + 2)}{\Gamma(\frac{m_i}{2\lambda})}$	$\frac{m_i(m_i + 2\lambda)}{4}$

O critério de informação Akaique (AIC) e o critério de informação Bayesiano (BIC) são utilizados com muita frequência para encontrar o melhor modelo para o fenômeno em estudo (Burnahm and Anderson, 2002). Ambos utilizam o logaritmo da função de verossimilhança para medir o bom ajuste, mas definem diferentes penalizações. As expressões comumente utilizadas para a seleção de modelos são dadas a seguir.

$$AIC = -2 \log l(\hat{\theta}) + 2p;$$

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

Em que p é o número de parâmetros e n é o número de observações na amostra. Na comparação de modelos, o modelo com o menor valor de AIC ou BIC é assumido como o melhor modelo entre todos os ajustados.

Análise de diagnóstico

Muitos problemas podem ser modelados por meio de modelos mistos, em particular o modelo multinível com dois níveis. Neste capítulo, trataremos sobre o diagnóstico de influência local em modelos lineares mistos e também sobre a análise de resíduos para verificar a adequabilidade das suposições do modelo proposto aos dados de interesse.

5.1 Diagnóstico de influência

Nos estudos de modelagem estatística, a análise de diagnóstico é uma etapa muito importante e utiliza conjunto de ferramentas para avaliar a qualidade do ajuste do modelo proposto aos dados e ainda para verificar a coerência das suposições iniciais. Métodos de estimação utilizando a função de verossimilhança podem ser sensíveis a observações aberrantes, especialmente no modelo normal, e o diagnóstico de influência inclui técnicas que permitem identificar observações que podem influenciar desproporcionalmente as estimativas dos parâmetros. O diagnóstico de influência é usado para investigar vários aspectos do modelo ajustado, permitindo a validação das suposições do modelo proposto. Esse diagnóstico inclui, basicamente, dois métodos de análise de influência: local e global. Segundo Souza (2006), uma observação é influente se uma pequena perturbação nos dados ou no modelo causar alguma mudança inferencial significativa. A técnica de influência local, em particular, possivelmente identifica observações influentes

por meio de perturbações nos dados ou no modelo, enquanto a influência global utiliza alguma medida como DFBETA, DFFITS e D-Cook (Cook, 1977; Belsey et al., 1980) para analisar as mudanças nos modelos ajustados quando é induzida a exclusão de um subconjunto de observações.

Para realizar uma análise de diagnóstico, a técnica de influência local tem se constituído de uma ferramenta muito importante com ampla utilização. Por exemplo, recentemente, foi aplicada por Souza (2006) no modelo de regressão logística. Em outro trabalho, um amplo estudo de sensibilidade para modelos não lineares elípticos com efeitos mistos foi desenvolvido por Russo et al. (2011). Já em modelos lineares com efeitos mistos, sua aplicação pode ser vista em (Osorio, 2006; Nobre, 2004; Lesaffre and Verbeke, 1998).

De forma geral, a influência local consiste em analisar, por meio de uma medida adequada de influência, a robustez das estimativas dos parâmetros ajustados quando pequenas perturbações são introduzidas no modelo ou nos dados (Paula, 2004). Neste contexto, a ponderação de casos é um esquema de perturbação que foi considerada no trabalho de Lobos (2010) para avaliar o afastamento das suposições do modelos log-Birnbaum-Saunders mistos. Russo (2006), além de fazer um estudo sobre influência local proposta por Cook (1986), aborda a técnica de exclusão de casos, em que uma observação é excluída de cada vez e o modelo de regressão com erros nas variáveis com intercepto nulo é ajustado novamente para verificar as mudanças ocorridas. Uma aplicação do modelo misto para estudar dados longitudinais pode ser visto em Osorio et al. (2007) e um estudo de influência local através da curvatura normal em modelos multiníveis pode ser visto em Pires (2009).

5.1.1 Influência local proposta por Cook (1986)

A popularidade da influência local se dá pelo fato de poder ser aplicada a qualquer problema em que se conheça a função de verossimilhança. Nessa técnica, a ideia principal consiste em efetuar pequenas perturbações nos dados ou no modelo e verificar se os resultados são alterados de forma significativa. Esse método foi proposto por Cook (1986) e é descrito e utilizado em muitos trabalhos envolvendo modelos estatísticos, por exemplo, em Russo (2010); Souza (2006); Nobre (2004); Paula (2004); Lesaffre and Verbeke (1998).

Sejam:

θ o vetor de parâmetros;

$L(\theta)$ a função de verossimilhança;

$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{j=1}^J L_j(\boldsymbol{\theta})$, em que $L_j(\boldsymbol{\theta}) = -\frac{1}{2} \log |\Sigma_j| + \log g(u_j)$;

$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{j=1}^J w_j L_j(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança perturbada;

$\boldsymbol{\omega} = (w_1, w_2, \dots, w_J)'$ o vetor de perturbações;

$\boldsymbol{\omega}_o = (1, 1, \dots, 1)'$ o vetor de não perturbação tal que $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_o) = \ell(\boldsymbol{\theta})$;

Para verificar a influência das perturbações nas estimativas de $\boldsymbol{\theta}$, a medida mais utilizada é o afastamento da verossimilhança, dada pela relação a seguir:

$$LD(\boldsymbol{\omega}) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\},$$

onde $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ são os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ em $L(\boldsymbol{\theta})$ e $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ respectivamente e $LD(\boldsymbol{\omega})$ contém informações sobre a influência da perturbação.

Como a análise de $LD(\boldsymbol{\omega})$ para todos os elementos de $\boldsymbol{\omega}$ é inviável, Cook (1986) propôs o estudo do comportamento local em torno de $LD(\boldsymbol{\omega})$ considerando uma superfície formada pelos elementos do vetor $\boldsymbol{\alpha}(\boldsymbol{\omega})$, denotado por gráfico de influência, onde

$$\boldsymbol{\alpha}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega}' \\ LD(\boldsymbol{\omega}) \end{pmatrix}$$

e a ideia básica foi analisar como $\boldsymbol{\alpha}(\boldsymbol{\omega})$ desvia-se do plano tangente em $\boldsymbol{\omega}_o$ e também como a função se comporta em torno de $\boldsymbol{\omega}_o$. Maiores detalhes podem ser vistos em Cook (1986); Paula (2004); Russo (2006) e Osorio (2006).

O método busca analisar o gráfico de $LD(\boldsymbol{\omega}_o + a\mathbf{d})$ após ter selecionado uma direção unitária \mathbf{d} e $a \in \mathbb{R}$. $LD(\boldsymbol{\omega}_o + a\mathbf{d})$ apresenta um mínimo local (em $a = 0$) e neste caso o gráfico, cuja curvatura chamaremos $C_{\mathbf{d}}$, pode ser visto como o círculo de melhor ajuste em $\boldsymbol{\omega}_o$. A maior curvatura contém as observações que mais influenciam em $LD(\boldsymbol{\omega})$. A expressão para a curvatura $C_{\mathbf{d}}$, proposta por (Cook, 1986) tem a forma

$$C_{\mathbf{d}} = 2|\mathbf{d}'\Delta'\ddot{L}^{-1}\Delta\mathbf{d}|, \quad (5.1)$$

onde $-\ddot{L}$ é a matriz de informação observada, em que

$$\ddot{L} = \left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad (5.2)$$

$$\Delta = \left. \frac{\partial^2 L(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}'} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{(\boldsymbol{\omega})}}. \quad (5.3)$$

Podemos notar possíveis observações que tem uma influência desproporcional sob pequenas perturbações em $LD(\boldsymbol{\omega})$ no gráfico de índices de \mathbf{d}_{max} . Por exemplo, sob o esquema de ponderação de casos, tratado mais adiante, uma possibilidade é considerar a matriz $\Delta' \ddot{L}^{-1} \Delta$ e determinar o autovetor \mathbf{d}_{max} correspondente ao maior autovalor $C \mathbf{d}_{max}$, onde o gráfico de \mathbf{d}_{max} com respectivos índices podem revelar observações influentes.

5.1.2 Abordagem de influência local para modelos mistos

A classe elíptica possui distribuições com caudas mais pesadas do que as da normal, por exemplo t-Student, que pode acomodar melhor as observações aberrantes e isso justifica o interesse de estudar diagnóstico de influência em modelos elípticos com efeitos mistos, além de ter poucos trabalhos que investigam a obtenção de medidas de influência local nesses modelos (Osorio, 2006).

O modelo multinível com dois níveis pode ser visto como um modelo misto, que é dado pela expressão

$$\mathbf{Y}_j = X_j W_j \boldsymbol{\gamma} + X_j \mathbf{b}_j + \boldsymbol{\epsilon}_j; \quad (5.4)$$

A distribuição conjunta de \mathbf{Y}_j e \mathbf{b}_j é da forma dada em (4.5), em que

\mathbf{Y}_j tem distribuição marginal com média $X_j W_j \boldsymbol{\gamma}$ e variância $\text{Var}(\mathbf{Y}_j) = \Sigma_j$ em que $\Sigma_j = X_j D X_j' + \sigma^2 I_{n_j}$, ou seja $\mathbf{Y}_j \sim El_{n_j}(X_j W_j \boldsymbol{\gamma}, \Sigma_j)$.

Sejam $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}')'$ o vetor dos parâmetros a serem estimados onde $\boldsymbol{\beta}' = (\gamma_{q0}, \gamma_{q1}, \dots, \gamma_{qp})$, $\tau_0 = \sigma^2, \tau_{rs}$ é o rs -ésimo elemento de D , para $r, s = 1, 2, \dots, q$; $L(\boldsymbol{\theta}) = \sum_{j=1}^J L_j(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança em que,

$$L_j(\boldsymbol{\theta}) = -\frac{1}{2} \{ \log |\Sigma_j| + \mathbf{r}_j' \Sigma_j^{-1} \mathbf{r}_j + n_j \log(2\pi) \}; \quad (5.5)$$

$\mathbf{r}_j = \mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}$ para $j = 1, \dots, J$.

Usando resultado de diferenciação de matrizes que podem ser visto em Osorio (2006) obtemos a matriz de informação observada

$$-\ddot{L}_j(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 L_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} -\ddot{L}_{11,j} & -\ddot{L}_{12,j} \\ -\ddot{L}'_{12,j} & -\ddot{L}_{22,j} \end{bmatrix}$$

em que

$$\begin{aligned} \ddot{L}_{11,j} &= 2W_j' X_j' \hat{\Sigma}_j^{-1} \{ w_g(\hat{u}_j) \hat{\Sigma}_j + 2w_g'(\hat{u}_j) \hat{\mathbf{r}}_j \hat{\mathbf{r}}_j' \} \hat{\Sigma}_j^{-1} X_j W_j, \\ \ddot{L}_{12,j} &= 2X_j W_j \hat{\Sigma}_j^{-1} \{ w_g(\hat{u}_j) \hat{\Sigma}_j + w_g'(\hat{u}_j) \hat{\mathbf{r}}_j \hat{\mathbf{r}}_j' \} \hat{\Sigma}_j^{-1} \frac{\partial \Sigma_j}{\partial \sigma^2} \hat{\Sigma}_j^{-1} \hat{\mathbf{r}}_j, \end{aligned}$$

$\ddot{L}_{22,j} = \frac{\partial^2 L_j(\boldsymbol{\theta})}{\partial \tau \partial \tau'}$ e cada elemento é da forma:

$$\begin{aligned} \frac{\partial^2 L_j(\boldsymbol{\theta})}{\partial \tau_k \partial \tau_l} &= \frac{1}{2} tr \left\{ \Sigma^{-1} \left[\frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1} \frac{\partial \Sigma}{\partial \tau_l} - \frac{\partial^2 \Sigma_j}{\partial \tau_k \partial \tau_l} \right] \right\} \\ &+ \mathbf{r}'_j \Sigma^{-1} \left\{ w'_g(u_j) \frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1} \mathbf{r}_j \mathbf{r}'_j \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_l} - w_g(u_j) \frac{\partial^2 \Sigma_j}{\partial \tau_k \partial \tau_l} \right. \\ &+ \left. w_g(u_j) \frac{\partial \Sigma_j}{\partial \tau_k} \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_l} + w_g(u_j) \frac{\partial \Sigma_j}{\partial \tau_l} \frac{\partial \Sigma_j}{\partial \tau_k} \right\} \Sigma_j^{-1} \mathbf{r}_j; \end{aligned}$$

$k, l = 0, 1, \dots, q$

Para identificar possíveis pontos influentes no modelo proposto utilizaremos o esquema de perturbação denominado de ponderação de casos que pode identificar observações que exercem grande influência no processo de estimação. Um estudo de diagnóstico de influência local sob o esquema de ponderação de casos em modelos lineares mistos normais e em modelos elípticos não lineares com efeito misto podem ser visto, respectivamente, em Lesaffre and Verbeke (1998) e Russo (2010).

O esquema de ponderação de casos consiste em atribuir uma ponderação para a função de verossimilhança dada em (5.5), ou seja, $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{j=1}^J \omega_j L_j(\boldsymbol{\theta})$, em que $\boldsymbol{\omega}_{J \times 1}$ é o vetor de pesos que pertence a um conjunto aberto de \mathbb{R}^J , com J igual à dimensão do vetor de perturbação. Neste caso, as expressões dos elementos da matriz $\Delta = [\Delta_1 \ \Delta_2]'$ de dimensão número de elementos de $\boldsymbol{\theta}$ dado na equação (5.3) por J são da forma:

$$\begin{aligned} \Delta_1 &= \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial w_j} \Big|_{\hat{\boldsymbol{\theta}}, w_0} = \hat{q}_j W'_j X'_j \hat{\Sigma}_j^{-1} (\mathbf{Y}_j - X_j W_j \hat{\boldsymbol{\beta}}), \\ \Delta_2 &= \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \tau_{rs} \partial w_j} \Big|_{\hat{\boldsymbol{\theta}}, w_0} = -\frac{1}{2} \left\{ tr \hat{\Sigma}_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_{rs}} - \hat{q}_j \hat{\mathbf{r}}'_j \hat{\Sigma}_j^{-1} \frac{\partial \Sigma_j}{\partial \tau_r} \hat{\Sigma}_j^{-1} \hat{\mathbf{r}}_j \right\}, \end{aligned}$$

para $r = 0, 1, s = 0, 1, \dots, q$, $\hat{q}_j = -2W_g(\hat{u}_j)$ e $\hat{\mathbf{r}}_j = \mathbf{Y}_j - X_j W_j \hat{\boldsymbol{\beta}}$.

5.2 Análise de resíduos

O resíduo é a diferença entre a quantidade observada e o valor estimado por um modelo de regressão. O objetivo da análise de resíduos é avaliar essa distância utilizando um conjunto de técnicas que além de serem utilizadas para verificar as suposições assumidas para o modelo, servem para analisar a qualidade do ajuste e auxiliam por meios de gráficos, na busca de pontos discrepantes e ainda, podem indicar se o modelo proposto é adequado. A análise de resíduos é um dos aspectos desenvolvidos por Manghi

(2012) para verificar características relacionadas ao ajuste e suposições estabelecidas para modelos elípticos multiníveis.

Considerando os resíduos para análise de diagnóstico do modelo misto dado em (5.4), que possui mais de uma fonte de variação, uma vez que os erros variam entre indivíduos dentro do grupo e entre grupos, Nobre (2004) define diferentes tipos de vetores de erros por exemplo, erros condicionais e erros marginais cujas expressões são da forma $\boldsymbol{\epsilon}_j = \mathbf{Y}_j - X_j W_j \boldsymbol{\gamma} - X_j \mathbf{b}$ e $\mathbf{e}_j = \mathbf{Y}_j - X_j W_j \boldsymbol{\gamma}$, respectivamente. No modelo misto, cada tipo de resíduo é útil para avaliar algum tipo de suposição inicial e no modelo sem efeito aleatório, os dois tipos de resíduos coincidem. A seguir vamos estudar a utilidade dos resíduos condicionais e marginais onde $\hat{\boldsymbol{\gamma}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\gamma}$ e $\hat{\mathbf{b}}$ é o melhor preditor linear de \mathbf{b} .

5.2.1 Resíduos condicionais

O resíduo condicional $\hat{\boldsymbol{\epsilon}}_i$ do modelo (5.4) é obtido usando a expressão

$$\hat{\boldsymbol{\epsilon}}_j = \mathbf{Y}_j - X_j W_j \hat{\boldsymbol{\gamma}} - X_j \hat{\mathbf{b}}, \quad (5.6)$$

Na expressão acima, $\hat{\boldsymbol{\gamma}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\gamma}$ dada na equação (4.6) e $\hat{\mathbf{b}}$ é o preditor de \mathbf{b} dado pela expressão $\hat{\mathbf{b}} = DX'\Sigma^{-1}(\mathbf{y} - XW\hat{\boldsymbol{\gamma}})$.

O resíduo condicional pode ser útil para identificar observações discrepantes, além de verificar a suposição de normalidade do erro condicional por meio do gráfico dos valores dos resíduos contra os valores ajustados. Pelo fato dos elementos de $\hat{\boldsymbol{\epsilon}}$ apresentarem diferentes variâncias uma alternativa para a expressão em (5.6) para identificar *outliers* seria utilizar resíduos condicionais padronizados. Nobre and Singer (2007) também propõem a padronização dos resíduos condicionais e descreve em detalhes sobre o uso dessa proposta para verificar algumas das suposições assumidas para o modelo misto, além disso, consideram de grande utilidade definir uma versão estudentizada para análises de resíduos que não depende da escala e permite uma comparação entre eles. Manghi (2012) considera essa versão e denomina de resíduos condicionais studentizados. Maiores detalhes sobre padronização dos resíduos condicionais em modelos linear misto também podem ser vistos em Savalli (2005) e Pinho et al. (2012) que usa a seguinte expressão:

$$\hat{\boldsymbol{\epsilon}}_j^* = \frac{\hat{\boldsymbol{\epsilon}}_j}{\sqrt{\sigma^2 q_{jj}}}, \quad (5.7)$$

em que q_{ii} é o i -ésimo elemento da diagonal da matriz Q obtida da seguinte maneira: $Q = \sigma^2(\Sigma^{-1} - \Sigma^{-1}XW(W'X'\Sigma^{-1}XW)^{-1}W'X'\Sigma^{-1})$.

Na Tabela 5.1 apresentamos a utilidade do resíduo condicional e o respectivo gráfico que podem ser construído para análise.

Tabela 5.1: Propósito dos gráficos de resíduos condicionais.

Utilidade	Gráfico
Presença de <i>outliers</i>	$\hat{\epsilon}_k^*$ vs. índices das observações
Homoscedasticidade dos erros condicionais	$\hat{\epsilon}_k^*$ vs. valores ajustados
Normalidade dos erros condicionais	quantis teóricos

5.2.2 Resíduo marginal

O resíduo marginal do modelo misto proposto em (5.4) é dado pela expressão

$$\hat{e}_j = \mathbf{Y}_j - X_j W_j \hat{\gamma}, \quad (5.8)$$

e analisando o gráfico, similarmente ao modelo linear normal, podemos investigar a adequabilidade do modelo avaliando se as suposições de linearidade estão satisfeitas, ou seja, espera-se um comportamento aleatório em torno de zero quando a relação é linear. Outra utilidade dos resíduos marginais é para avaliar a estrutura de covariância onde para cada modelo a matriz escala corresponde à matriz de variância e covariância e assume uma forma específica, por exemplo, $\text{Var}(\mathbf{Y}_j) = \Sigma_j$ para o modelo normal e Lindsey and Lindsey (2006) mostram que para a distribuição t-Student tem se que $\text{Var}(\mathbf{Y}_j) = \frac{\nu}{\nu - 2} \Sigma_j$, ($\nu > 2$).

Uma alternativa para a expressão em (5.8) é utilizar a forma padronizada denominada de resíduo marginal studentizado (Manghi, 2012) que é a estimativa do erro padrão do resíduo \hat{e}_j .

A análise do gráfico dos resíduos marginais versus respectivos índices deve ser feita considerando os respectivos pesos estimados para cada grupo devido a influência que esses pesos podem afetar na estimação dos parâmetros, ou seja, dependendo do peso atribuído, observações aberrantes podem não ser influentes nas estimativas dos parâmetros. Na Tabela 5.2 apresentamos a utilidade do resíduo marginal e o respectivo gráfico que podem ser construídos para análise.

Outra técnica para a análise de diagnósticos é o envelope. Ela consiste na representação gráfica que apresenta bandas de confiança para os resíduos e foi sugerida por Atkinson

Tabela 5.2: Propósito dos gráficos de resíduos marginais.

Utilidade	Gráfico
Linearidade dos efeitos fixos. Identificar observações influentes	Resíduos vs. variáveis explicativas $\ \mathbf{I}_{n_j} - R_j R_j'\ $ vs. índices das escolas, em que $R_j = \widehat{\Sigma}_j^{-\frac{1}{2}} \hat{\mathbf{e}}_j$

(1985) para detectar afastamentos sérios da normalidade além de informar a existência de pontos aberrantes. O procedimento para a construção do envelope está descrito em Paula (2004), nas páginas 65 e 66.

Aplicações

Neste capítulo aplicamos os modelos lineares multiníveis discutidos nos Capítulos 3 e 4 ao conjunto de dados educacionais disponibilizado pelo INEP, cuja forma de acesso está descrita no apêndice A. Ajustamos os modelos multiníveis aos dados de uma amostra aleatória com 221 alunos da cidade de São Carlos, tal que cada escola é representada por, na média, 10 alunos. Para obter as estimativas dos parâmetros do modelo, foi utilizado o método iterativo *score* de Fisher, que utiliza as funções *score*, determinadas derivando-se o logaritmo da função de verossimilhança em relação a cada elemento do vetor de parâmetros desenvolvido na Seção 3.4.1.

Denotamos por Y_{ij} o rendimento escolar em Matemática do aluno i que estuda na j -ésima escola (variável resposta).

O modelo inicial, que Ferrão (2003) denomina de modelo nulo, sem variáveis explicativas, composto apenas pelo intercepto, é especificado pelas seguintes equações:

$$\text{(nível 1)} \quad Y_{ij} = \beta_{0j} + \varepsilon_{ij},$$

$$\text{(nível 2)} \quad \beta_{0j} = \gamma_{00} + b_{0j},$$

em que $i = 1, 2, \dots, n_j$, com n_j igual ao número de alunos em cada escola, $j = 1, 2, \dots, 21$ e $\sum_{j=1}^{21} n_j = 221$.

No modelo supracitado, temos que:

β_{0j} é o parâmetro desconhecido e representa o valor esperado do rendimento dos alunos que estudam na escola j ;

γ_{00} é a média global do rendimento;

$\varepsilon_{ij} \sim N(0, \sigma^2)$, ε_{ij} s independentes;

$b_{0j} \sim N(0, \tau_{00})$, b_{0j} s independentes;

ε_{ij} é o efeito aleatório associado ao nível 1 e;

b_{0j} é o efeito aleatório associado ao nível 2 e representa o afastamento do rendimento médio da escola j em relação à média global.

b_{0j} e ε_{ij} são independentes.

Os resultados do ajuste do modelo inicial constam na Tabela 6.1. Se a estimativa da variância entre escolas, τ_{00} , pudesse ser considerada nula, a variância do rendimento, seria atribuída à variabilidade das notas dos alunos, ou seja, o efeito da escola não se verificaria no rendimento escolar atingido pelos estudantes. Portanto, com esse modelo, pode-se estimar o efeito das escolas nos resultados das avaliações.

Tabela 6.1: Estimativas dos parâmetros do modelo sem variáveis explicativas.

Parâmetro	Estimativa	Erro Padrão
γ_{00}	247,303	4,929
τ_{00}	365,152	158,142
σ^2	1525,318	152,530

Como τ_{00} é estatisticamente diferente de zero, verifica-se que alguma das características relacionada a escola tem relativa influência nos resultados das avaliações obtidas pelos alunos. Uma quantidade que pode nos informar a magnitude da influência da escola é o coeficiente de correlação intra-escolar (ρ) que é obtida pela razão em que o antecedente é a variância entre as escolas (τ_{00}) e o conseqüente é a variância total da variável resposta ($\sigma^2 + \tau_{00}$). Quando $\rho \cong 0$, temos um indicativo de que as características relacionadas a escola em nada influencia nos resultados escolares, ou seja, que as escolas são homogêneas. Por outro lado, se $\rho \cong 1$, significa que as características individuais do aluno não tem peso significativo no desempenho escolar, ou seja, toda a variabilidade da variável resposta se deve à diferença entre as escolas. Com as estimativas da variância que constam na Tabela 6.1, o valor obtido para o coeficiente de correlação foi de, aproximadamente, 0,19 e esse valor indica que 19% da variância do rendimento escolar está associado a diferenças entre as escolas.

A análise dos resíduos é uma ferramenta básica para verificar se os pressupostos do modelo foram violados. Os gráficos da Figura 6.1 permitem verificar desvios da

suposição de normalidade dos erros condicionais ou presença de outliers para o primeiro modelo considerado.

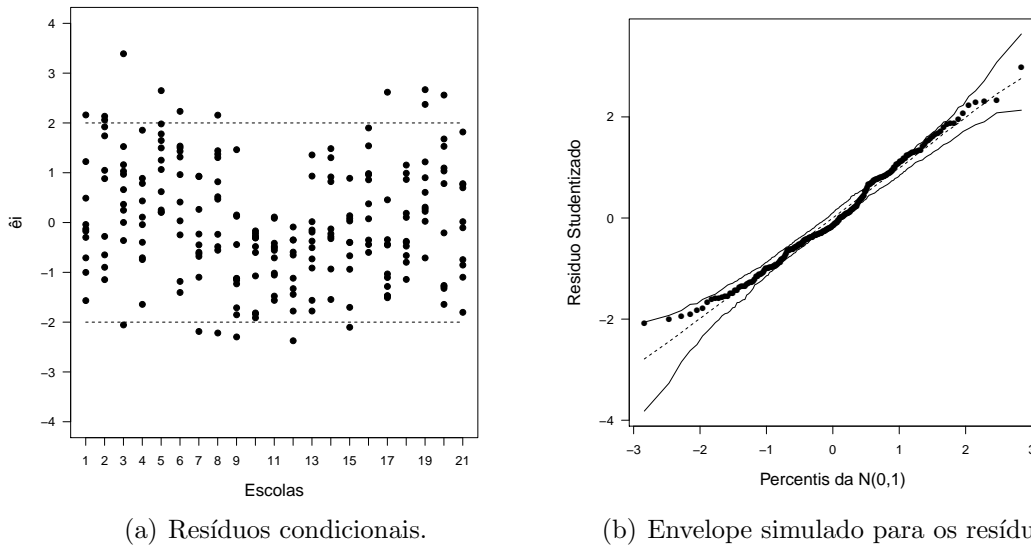


Figura 6.1: Gráficos para avaliar o ajuste do modelo normal sem variáveis explicativas.

Notamos que não há indícios para rejeitar a hipótese de que os resíduos seguem uma distribuição normal, ou seja, os pressupostos estão satisfeitos e podemos perceber que o modelo pode ser apropriado para o ajuste dos dados já que não há observações fora das bandas de confiança dos envelopes simulados, embora o primeiro gráfico apresente pontos atípicos, indicando presença de observações discrepantes e indicando que um modelo com caudas mais leves ou mais pesadas do que as da normal seja mais adequado.

O próximo modelo considerando uma variável explicativa, por exemplo, o número de reprovações na vida acadêmica que denotamos por $(Nurep)$, é dado pelas equações definidas no Capítulo 3, fazendo $X_{ij} = Nurep_{ij}$ na equação (3.3) e $W_j = 0$ na equação (3.4). Neste exemplo, a escola em que o aluno i estuda é identificada pelo índice j .

$$\text{(nível 1)} \quad Y_{ij} = \beta_{0j} + \beta_{1j}Nurep_{ij} + \varepsilon_{ij}$$

$$\text{(nível 2)} \quad \beta_{0j} = \gamma_{00} + b_{0j}$$

$$\beta_{1j} = \gamma_{10} + b_{1j}$$

em que:

$$i = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, 21, \quad \sum_{j=1}^{21} n_j = 221;$$

$\varepsilon_{ij} \sim N(0, \sigma^2)$. Temos que

$$\begin{bmatrix} b_{0j} \\ b_{1j} \end{bmatrix} \sim N_{n_j} \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right\}$$

Na Tabela 6.2, apresentamos os resultados do ajuste com a suposição de normalidade e podemos perceber que o rendimento sofre uma influência negativa se for considerada a covariável número de reprovações, ou seja, para cada unidade adicional no número de reprovações, mudança na categoria, o rendimento do aluno, diminui em média de 21,228 unidades. A correlação entre b_{0j} e b_{1j} é alto, indica que talvez seja suficiente considerar a covariável número de reprovações na parte fixa. E $\tau_{11} = 32,833$ com erro padrão alto, é um indicativo de que as inclinações da reta para cada escola pode ser consideradas a mesmas para todas as escolas.

Tabela 6.2: Estimativas dos parâmetros do modelo considerando a variável explicativa número de reprovações.

Parâmetro	Estimativa	Erro Padrão
γ_{00}	254,765	5,464
γ_{10}	-21,228	4,005
τ_{00}	462,237	194,176
τ_{10}	-169,198	118,141
τ_{11}	32,833	104,567
σ^2	1315,692	137,769

A variância entre as escolas, τ_{11} não é nula e isso pode significar que a influência da covariável número de reprovações do aluno no desempenho escolar varia entre as escolas. A covariância, τ_{10} , sendo negativa é um indicativo que há uma certa tendência para as escolas com rendimento médio alto estar associado ao número de reprovações mais baixo, ou seja, escolas em que os alunos atingem notas mais altas, são também frequentadas por alunos que não apresentam reprovações no histórico escolar.

Analisando os gráficos da Figura 6.2 há indícios para não rejeitar a hipótese de que os resíduos condicionais studentizados seguem uma distribuição normal, ou seja, há evidências para aceitar a suposição de normalidade para os erros do modelo em estudo, podemos perceber que os dados apresentam outliers e como não há observações fora das bandas de confiança do envelope simulado para os resíduos, o modelo ajustado parece ser adequado.

A infraestrutura da escola é uma possível variável que pode ter influência na variável resposta pois cada escola é classificada em uma das categorias, por exemplo, excelente, boa, razoável ou ruim. Portanto, as variáveis consideradas, neste outro exemplo, são

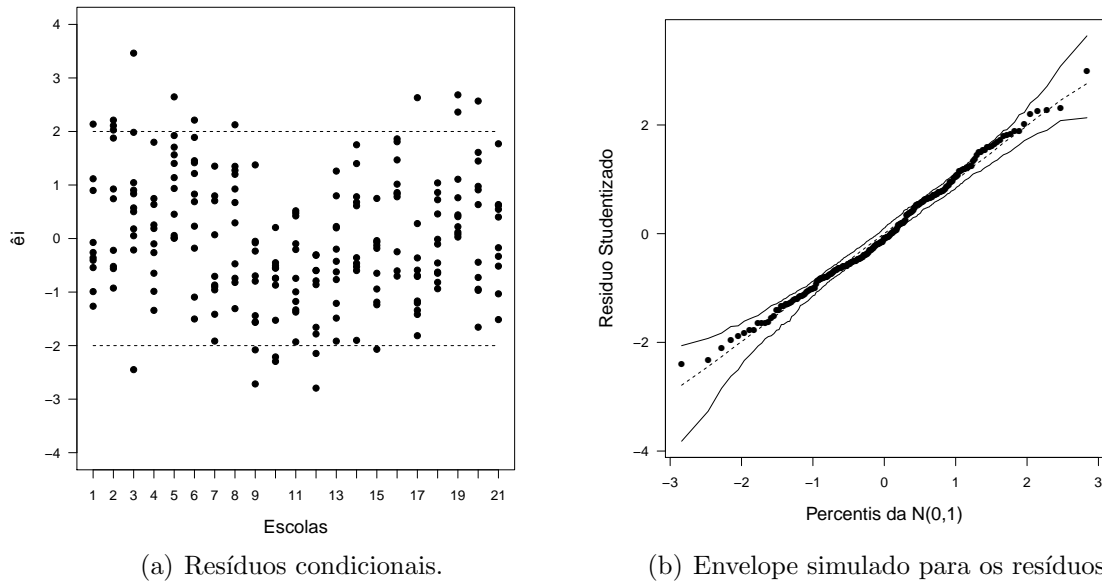


Figura 6.2: Gráficos para avaliar o ajuste do modelo sob normalidade considerando o número de reprovações para explicar o rendimento.

nomeadas por $Nurep_{ij}$ e $Infest_j$, que representam o número de reprovações que aluno i apresenta em seu histórico escolar e que estuda na escola j e, infraestrutura da escola j , respectivamente. Assim o modelo é especificado pelas equações:

$$\text{(nível 1)} \quad Y_{ij} = \beta_{0j} + \beta_{1j}Nurep_{ij} + \varepsilon_{ij}$$

$$\text{(nível 2)} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}Infest_j + b_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Infest_j + b_{1j}$$

em que:

$$i = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, 21, \quad \sum_{j=1}^{21} n_j = 221;$$

$$\varepsilon_{ij} \sim S(0, \sigma_e^2), \quad \varepsilon_{ij}'s \text{ independentes};$$

$$b_{qj} \sim S(0, \tau_{qq}) \quad q = 0, 1 \text{ e } b_{qj}'s \text{ independentes e};$$

$$\varepsilon_{ij} \text{ e } b_{qj} \text{ independentes entre si.}$$

Com a finalidade de encontrar o melhor modelo para explicar o rendimento escolar, utilizando uma variável no primeiro nível e outra no segundo nível, consideramos além da distribuição normal, as distribuições simétricas t-Student e exponencial potência, que pertencem à classe elíptica e, por atribuir pesos para as observações aberrantes, podem evitar a influência desproporcional das mesmas no cálculo das estimativas dos parâmetros. Assim, para este modelo, vamos comparar os resultados obtidos considerando essas três distribuições uma vez que a distribuição normal, é a mais

utilizada na literatura, a t-Student, por ter cauda mais pesada do que a normal, e exponencial potência, que dependendo do parâmetro de forma, λ , pode apresentar caudas mais leves ($\lambda > 1$) ou mais pesadas ($\lambda < 1$) do que as da normal.

A Figura 6.3 apresenta o histograma das notas padronizadas utilizados nesse trabalho e os gráficos de funções densidade de probabilidades das distribuições t-Student e exponencial potência, comparadas com o gráfico da função densidade da distribuição normal. Na Figura 6.4(a), percebemos que a exponencial potência com cauda leve ($\lambda = 1,4$) parece acomodar melhor os dados do que a com cauda pesada ($\lambda = 0,8$), e também melhor do que com a da normal. Com relação à distribuição t-Student, a Figura 6.4(b) nos sugere que independente do grau de liberdade adotado (1 ou 5), o uso da normal parecer ser mais adequado.

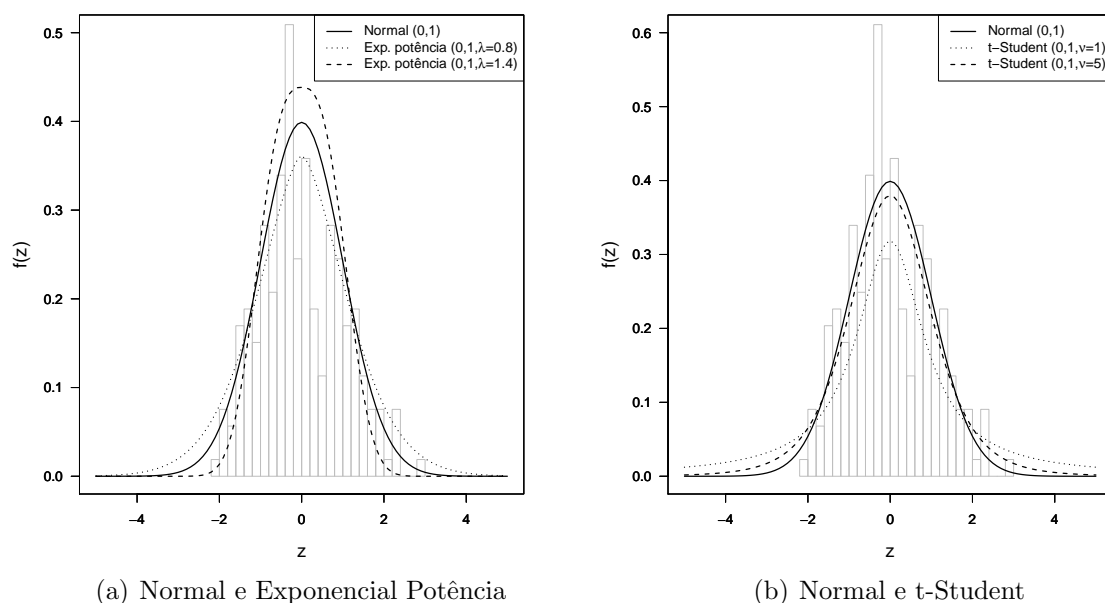


Figura 6.3: Gráficos das funções densidades de probabilidades de algumas distribuições elípticas e o histograma das notas normalizadas.

Na Tabela 6.4, apresentamos os resultados do ajuste desse modelo e uma das inferências sobre o valor estimado para a variância do coeficiente de inclinação, τ_{11} é que a influência no rendimento escolar das covariáveis consideradas varia de escola para escola. O valor negativo para γ_{10} indica que o número de reprovações na vida acadêmica afeta o desempenho e outro indicativo dos resultados, γ_{01} positivo, é que as condições das escolas não afetam o desempenho escolar porém o número de reprovações alto de um aluno associado as péssimas condições da escola faz com que o rendimento diminua consideravelmente.

Na Tabela 6.3, apresentamos os resultados do ajuste dos modelos, considerando as distribuições normal, t-Student com 5 graus de liberdade e exponencial potência com $\lambda = 1, 4$, escolhidos empiricamente utilizando o AIC (ver Figura 6.4(a)). Podemos perceber que as estimativas de γ para o modelo sob normalidade estão sempre entre as estimativas do modelo t-Student e exponencial potência, evidenciando que um dos modelos considerados apresentará melhor ajuste e que as inferências sobre as análises dos três modelos são similares. Como a matriz escala nos modelos elípticos representa a matriz de variância e covariância a menos de uma constante multiplicativa para cada modelo, não podemos comparar os valores de σ^2 e de τ (Russo, 2010).

Tabela 6.3: Estimativas dos parâmetros com erros padrão (E. P.) sob os modelos normal, t-Student e exponencial potência.

	Normal		t-Student		Exponencial Potência	
	Estimativa	(E. P.)	Estimativa	(E. P.)	Estimativa	(E. P.)
γ_{00}	245,154	(8,874)	239,071	(9,134)	247,847	(8,642)
γ_{01}	11,794	(8,792)	15,890	(9,051)	9,980	(8,550)
γ_{10}	-15,985	(6,704)	-13,180	(7,000)	-17,597	(6,541)
γ_{11}	-6,523	(7,132)	-8,655	(7,406)	-5,123	(6,969)
σ^2	1318,391	(138,024)	1196,812	(220,353)	3381,884	(307,309)
τ_{00}	407,479	(177,424)	388,131	(186,009)	998,108	(424,782)
τ_{01}	-145,562	(110,169)	-163,570	(116,363)	-329,728	(262,691)
τ_{11}	19,994	(100,664)	38,526	(103,718)	25,239	(242,534)

O Critério de Informação Akaike (AIC) foi utilizado para selecionar o melhor modelo considerando as três distribuições. Por ele, os dados escolares empregados podem ser melhor modelados segundo a exponencial potência pois esta apresentou valores mais baixos de AIC, seguido da normal e da t-Student. Esse resultado é coerente com as considerações feitas sobre a Figura 6.3.

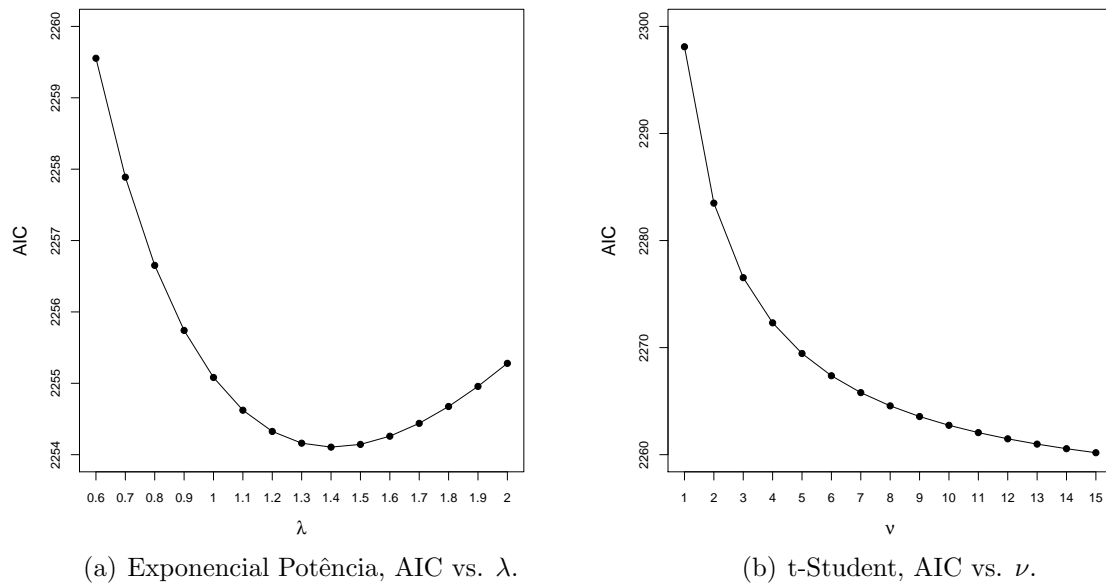


Figura 6.4: Gráfico do AIC para escolher o melhor parâmetro para o ajuste do modelo.

A seguir apresentamos os gráficos para análise de resíduos dos três modelos ajustados. A Figura 6.5 indica que os pressupostos dos erros estão satisfeitos e que considerar um modelo com cauda mais leve ou mais pesada do que o modelo normal pode diminuir a influência das observações consideradas aberrantes.

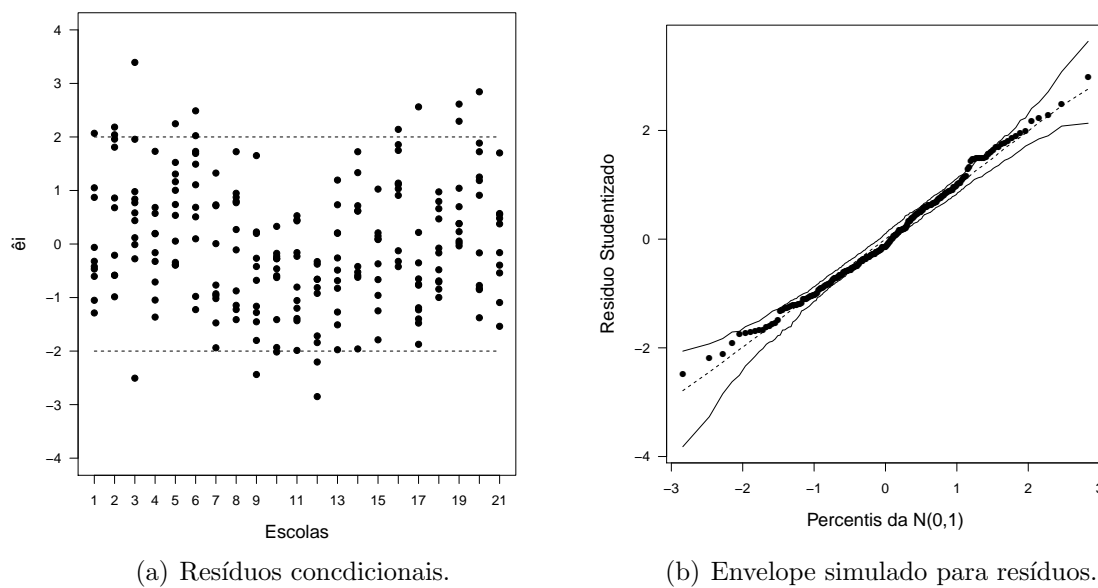


Figura 6.5: Gráficos para avaliar o ajuste do modelo normal com duas covariáveis.

Analisando os gráficos da Figura 6.6, podemos perceber que na Figura 6.6(b) não há observações fora da banda de confiança do envelope simulado para os resíduos indicando que os pressupostos do modelo estão satisfeitos, porém a existência de muitos pontos abaixo do intervalo considerado ideal, indica que o modelo sob normalidade visto anteriormente é o mais adequado.

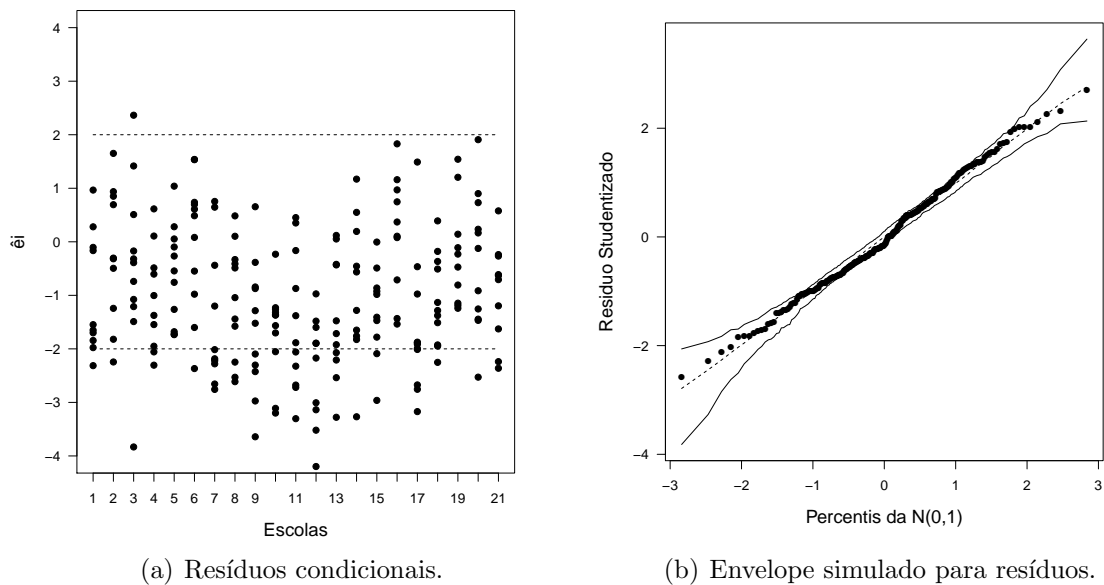


Figura 6.6: Gráficos para avaliar o ajuste do modelo t-Student com duas covariáveis.

Os gráficos da Figura 6.7, indicam claramente que o modelo exponencial potência com cauda mais leve, $\lambda = 1,4$, do que a distribuição normal é o mais apropriado para ajustar os dados educacionais aqui apresentados, que tem estrutura hierárquica. Podemos perceber que esse modelo acomoda melhor as observações discrepantes comparadas com os outros dois modelos e que também por não apresentar pontos fora da banda de confiança do envelope indica que as suposições iniciais estão satisfeitas e portanto, concluímos que é o modelo mais apropriado.

A análise de influência local sob o esquema de ponderação de casos foi aplicado para verificar observações influentes. Os autovalores relativos às direções de influência sob esse esquema de perturbação, podem ser visto na Figura 6.8. Como existe apenas um autovalor com magnitude alta, consideraremos o enfoque de Cook (1986), que considera a influência apenas da direção do autovetor referente ao maior autovalor.

Para encontrar observações influentes no conjunto de dados discutidos no Capítulo 2, construímos gráficos de índices de d_{max_i} sob o esquema de perturbação de ponderação

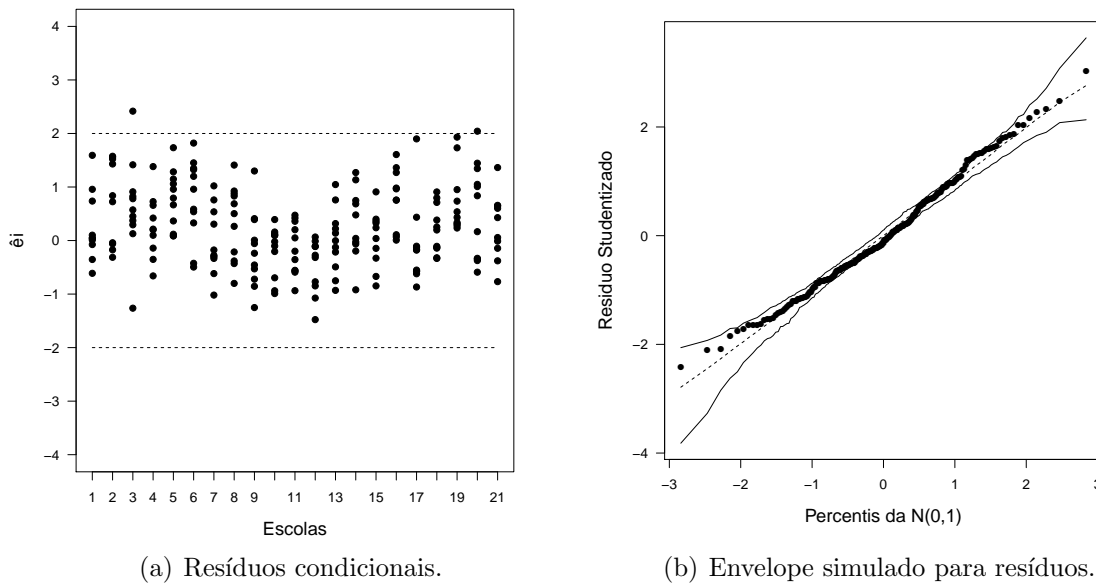


Figura 6.7: Gráficos para avaliar o ajuste do modelo exponencial potência.

de casos considerando as distribuições simétricas t de Student e exponencial potência, além da distribuição normal.

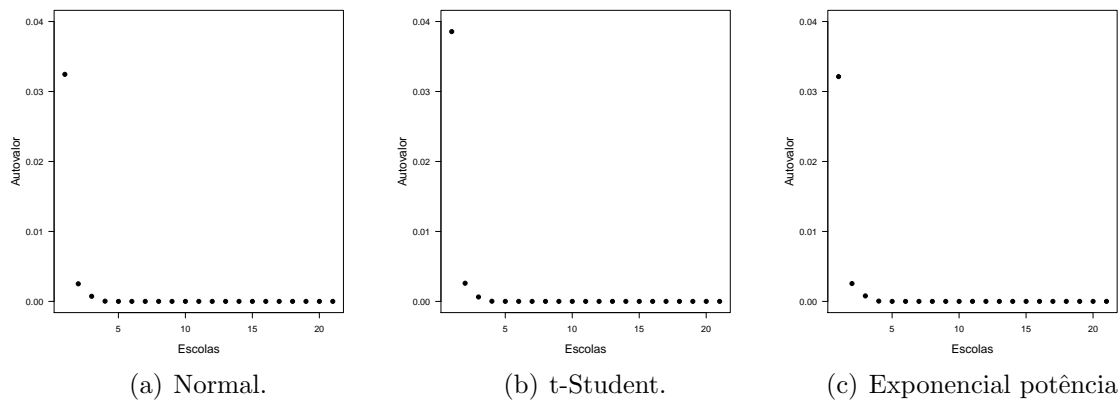


Figura 6.8: Autovalores relativos às direções de influência.

Na Figura 6.9, podemos perceber que algumas escolas parecem ter grande influência na estimação dos parâmetros nos três modelos considerados. Para verificar se, realmente, uma escola influencia de maneira desproporcional no modelo ajustado, consideramos a influência na direção $|d_{max}|$, a observação 16 que parece ser a mais influente no modelo normal mas que também se destoa nos outros dois modelos. Além dessa, analisamos a influência da observação 11 no modelo t-Student e da observação 3 no modelo exponencial potência.

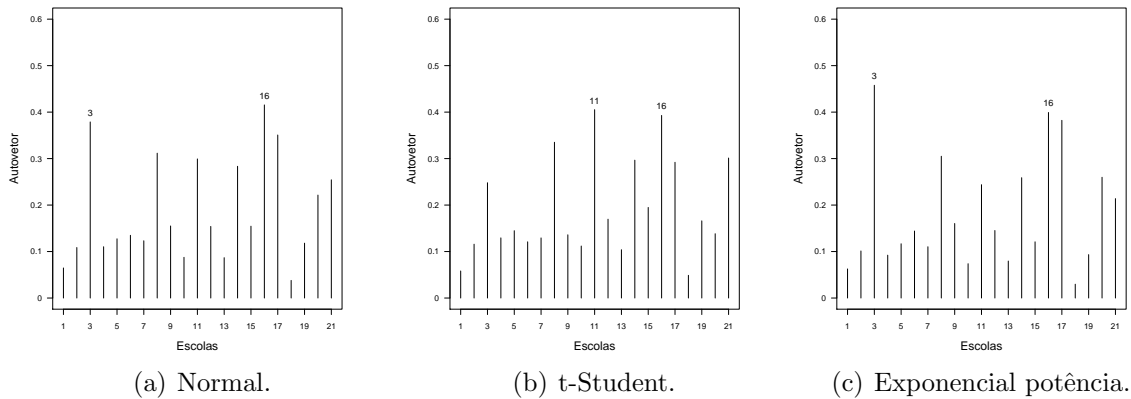


Figura 6.9: Gráficos de autovetores absolutos vs. índices (escolas).

Eliminando as observações que se destacam e usando o mesmo procedimento para a estimação dos parâmetros, obtivemos os valores que estão na Tabela 6.4 que contém também as estimativas de máxima verossimilhança do ajuste excluindo as duas observações mais influentes, a mudança relativa comparadas com as estimativas originais (com todas as observações) e os respectivos valores do AIC dos modelos ajustados cujos valores diminuíram para a escola mais influente. A mudança relativa da estimativa de cada parâmetro foi obtida calculando o valor absoluto da razão em que o antecedente é a diferença da estimativa excluindo as observações mais influente (ou a mais influente) e estimativa original, e o conseqüente é a estimativa original.

Concluimos que as estimativas mais afetadas foram nos elementos da matriz D , que as escolas influenciam nas estimativas uma vez que o AIC diminuiu e além disso, que distribuições com caudas mais pesadas do que a da distribuição normal são mais sensíveis para observações que aparecem como possivelmente influentes.

Tabela 6.4: Estimativas dos parâmetros, mudança relativa sob os modelos normal, t-Student e exponencial potência com todas as observações, sem 12 e 19 e excluindo apenas a 19 e AIC dos modelos ajustados.

Normal					
	Estimativa (com todas)	Estimativa (sem 3)	Mudança relativa	Estimativa (sem 16)	Mudança relativa
γ_{00}	245,154	244,675	0,002	241,112	0,016
γ_{01}	11,794	11,32	0,040	14,981	0,270
γ_{10}	-15,985	-15,828	0,010	-20,168	0,262
γ_{11}	-6,523	-7,502	0,150	-2,247	0,656
σ^2	1318,391	1240,544	0,059	1341,642	0,018
τ_{00}	407,479	433,733	0,064	398,647	0,022
τ_{01}	-145,562	-180,833	0,242	-217,718	0,496
τ_{11}	19,994	53,205	1,661	28,396	0,420
AIC	2255,082	2133,042		2149,520	
t-Student($\nu = 5$)					
	Estimativa (com todas)	Estimativa (sem 11)	Mudança relativa	Estimativa (sem 16)	Mudança relativa
γ_{00}	239,071	240,880	0,008	234,778	0,018
γ_{01}	15,890	17,230	0,084	19,158	0,206
γ_{10}	-13,180	-15,350	0,165	-18,245	0,384
γ_{11}	-8,655	-9,971	0,152	-4,137	0,522
σ^2	1196,812	1251,543	0,046	1156,008	0,034
τ_{00}	388,131	308,476	0,205	363,212	0,064
τ_{01}	-163,570	-34,796	0,787	-183,304	0,121
τ_{11}	38,526	-77,377	3,008	92,571	1,403
AIC	2269,451	2160,859		2163,910	
Exponencial		Potência ($\lambda = 1, 4$)			
	Estimativa (com todas)	Estimativa (sem 3)	Mudança relativa	Estimativa (sem 16)	Mudança relativa
γ_{00}	247,847	247,325	0,002	244,516	0,013
γ_{01}	9,980	9,384	0,060	12,643	0,267
γ_{10}	-17,597	-17,462	0,008	-21,790	0,238
γ_{11}	-5,123	-6,428	0,255	-1,017	0,802
σ^2	3381,884	3152,985	0,068	3448,358	0,020
τ_{00}	998,108	1085,850	0,088	983,762	0,014
τ_{01}	-329,728	-436,246	0,323	-483,899	0,468
τ_{11}	25,239	124,376	3,928	14,355	0,431
AIC	2254,106	2130,909		2149,595	

Estudo de simulação

Algumas questões importantes a serem respondidas quando se propõe uma nova metodologia em relação a uma metodologia usualmente utilizada são o ganho obtido com a nova proposta e as dificuldades existentes na sua implementação. Muitas vezes, o custo de se propor um novo modelo é bastante alto frente aos benefícios que ele traz. Quando se trata de modelos simétricos em comparação com os modelos usuais (assumindo distribuição normal), atribui-se diferentes pesos às observações e é comum que outros parâmetros sejam adicionados ao modelo, por exemplo o número de graus de liberdade no caso do modelo t-Student. Em um estudo recente, Russo et al. (2011) mostraram que as consequências de se assumir normalidade na existência de observações atípicas podem estar relacionadas com a sensibilidade do modelo em relação a estatísticas do teste, por exemplo.

O objetivo deste capítulo é desenvolver um breve estudo de simulação para avaliar as consequências da especificação incorreta da distribuição suposta como a das componentes aleatórias envolvidas no modelo e verificar se os modelos considerados nesse trabalho são sensíveis a essas distribuições. Assim, pretendemos considerar alguns cenários para gerar observações e para estimar os parâmetros do modelo ajustados àqueles dados. Por exemplo, poderemos gerar observações vindas de um modelo t-Student com poucos graus de liberdade e estimar os parâmetros de um modelo que assume distribuição normal. Em outra ocasião, podemos trabalhar com dados simulados vindos de um

modelo normal e ajustar um modelo exponencial potência com caudas mais leves ou mais pesadas do que as da distribuição normal. Como formas de avaliar os modelos ajustados, podemos considerar quantidades como o viés, variância ou erro quadrático médio das estimativas obtidas e também pelo critério de informação Akaike (*AIC*), entre outros.

O viés de um estimador $\hat{\theta}$ para estimar θ é definido como

$$b(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta,$$

assim o viés empírico é dado pela expressão

$$\hat{b}_i = \frac{\sum_{i=1}^M \hat{\theta}_{ij}}{M} - \theta_i,$$

onde \hat{b}_i é o viés empírico da estimativa $\hat{\theta}_{ij}$ para estimar θ_i , com M o número de amostras geradas. Aqui $\hat{\theta}_{ij}$ representa a estimativa de θ_i na j -ésima amostra gerada.

Como resultados, esperamos verificar se existem consequências preocupantes de se considerar uma distribuição incorreta no modelo multinível, ou em que ponto pode ser considerado aceitável supor normalidade quando na verdade uma distribuição com caudas mais pesadas seria adequada.

No estudo aqui realizado, o procedimento experimental descrito a seguir foi adotado para a geração das notas simuladas dos alunos. Inicialmente, assumimos que um vetor de parâmetros $\boldsymbol{\theta} = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma_e^2, \tau_{00}, \tau_{01}, \tau_{11})$ é conhecido. A partir dele, as quantidades seguintes são aleatoriamente geradas:

$$b_{0j} \sim El(0, \tau_{00}),$$

$$b_{1j} \sim El(0, \tau_{11}),$$

$$\epsilon_{ij} \sim El(0, \sigma_e^2),$$

em que $j = 1, \dots, J$ e $i = 1, \dots, n_j$ e por simplicidade, assumimos que D é diagonal. Com b_{0j} e b_{1j} independentes entre si e independentes dos ϵ_{ij} , assim caracterizados, calculam-se β_{0j} e β_{1j} para todas as J escolas utilizando-se a equação (3.3) do Capítulo 4. Aplicando-se esses valores e ϵ_{ij} à equação (3.4) daquele capítulo, temos um conjunto de notas simuladas para todos os n alunos considerados. Neste esquema de geração de notas, a distribuição elíptica da qual b_{0j} , b_{1j} e ϵ_{ij} são amostrados definem possíveis cenários de simulação. Aqui, consideram-se distribuições normal, t-Student com $\nu = 5$

e exponencial potência com $\lambda = 1.4$, com base nos modelos ajustados no Capítulo 6. Na discussão subsequente, analisamos os resultados da simulação para essas três situações.

No primeiro cenário, consideramos notas simuladas utilizando a distribuição normal. Para tanto, empregamos a função *rmvnorm* do pacote *mvtnorm* do R Core Team (2012) para gerar $M = 2000$ amostras, onde cada amostra tem 221 valores. Nesse caso, o θ teórico é fixado em valores próximos dos obtidos no ajuste do modelo normal aos dados educacionais originais (como realizado no Capítulo 6). Assim, observando-se a Tabela 6.3, temos os seguintes valores teóricos de parâmetros: $\gamma_{00} = 245,154$, $\gamma_{01} = 11,794$, $\gamma_{10} = -15,985$, $\gamma_{11} = -6,523$, $\sigma^2 = 1318,391$, $\tau_{00} = 407,479$, $\tau_{01} = -145,562$, $\tau_{11} = 19,994$. Para cada conjunto de dados simulado e utilizando o modelo descrito na Seção 4.2 foi feito o ajuste aos dados considerando as três distribuições supracitadas.

Na Tabela 7.1 estão as médias das estimativas dos ajustes das 2000 amostras geradas sob normalidade, com os respectivos vies, para os três modelos considerados. Percebe-se que em todos os casos, o vieses relacionados a γ_{00} , γ_{10} , γ_{01} e γ_{11} são pequenos, enquanto que para σ^2 , τ_{00} , τ_{01} e τ_{11} os vieses são maiores. A fim de avaliar os modelos, considerou-se também o AIC, cujos valores médios dos ajustes dos três modelos é mostrado na tabela. O valor do AIC para o modelo normal apresentou vantagem em relação ao dos demais modelos. Este resultado é coerente, dado que notas geradas nesse cenário apresentam b_{0j} , b_{1j} e ϵ_{ij} com distribuição normal.

Tabela 7.1: Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição normal.

	Normal		t-Student		Exponencial Potência	
	Estimativa	Viés	Estimativa	Viés	Estimativa	Viés
γ_{00}	245,148	-0,006	245,157	0,004	245,136	-0,017
γ_{01}	11,727	-0,068	11,688	-0,107	11,754	-0,040
γ_{10}	-16,170	-0,184	-16,149	-0,164	-16,196	-0,210
γ_{11}	-6,258	0,265	-6,296	0,227	-6,224	0,299
σ^2	1324,841	6,450	1164,555	-153,836	3420,061	2101,671
τ_{00}	347,123	-60,356	307,706	-99,773	881,183	473,704
τ_{01}	8,252	153,814	6,847	152,409	24,245	169,807
τ_{11}	-4,382	-24,375	-3,467	-23,460	-21,490	-41,483
AIC	2254,599		2265,448		2255,133	

No segundo cenário, consideramos notas simuladas utilizando a distribuição t-Student. Nele, a função *rmvnorm* do pacote *mvtnorm* do R Core Team (2012) foi empregada

para gerar $M = 2000$ amostras, onde cada amostra tem 221 valores. Nesse caso, o θ teórico é fixado do ajuste do modelo t-Student, conforme realizado no Capítulo 6. Assim, observando-se a Tabela 6.3, temos os seguintes valores de parâmetros: $\gamma_{00} = 239,071$, $\gamma_{01} = 15,890$, $\gamma_{10} = -13,180$, $\gamma_{11} = -8,655$, $\sigma^2 = 1196,812$, $\tau_{00} = 388,131$, $\tau_{01} = -163,570$, $\tau_{11} = 38,526$. Para cada conjunto de dados simulado e utilizando o modelo descrito na Seção 4.2 foi feito o ajuste aos dados considerando as três distribuições dadas na Tabela 4.4.

Na Tabela 7.2 estão as médias das estimativas dos ajustes das 2000 amostras geradas com distribuição t-Student, com os respectivos viés, para os três modelos considerados. Percebe-se, novamente, que em todos os casos, o vieses relacionados a γ_{00} , γ_{10} , γ_{01} e γ_{11} são pequenos, enquanto que para σ^2 , τ_{00} , τ_{01} e τ_{11} eles são maiores. Quanto aos valores de AIC, o modelo t-Student foi melhor quando comparado aos outros modelos. Tal resultado é esperado, pois as notas geradas nesse cenário apresentam b_{0j} , b_{1j} e ϵ_{ij} com distribuição t-Student.

Tabela 7.2: Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição t-Student.

	Normal		t-Student		Exponencial Potência	
	Estimativa	Viés	Estimativa	Viés	Estimativa	Viés
γ_{00}	238,734	-0,337	238,881	-0,190	238,675	-0,396
γ_{01}	16,330	0,441	16,246	0,356	16,377	0,488
γ_{10}	-13,105	0,076	-13,203	-0,022	-13,062	0,118
γ_{11}	-8,880	-0,225	-8,696	-0,041	-8,995	-0,340
σ^2	1906,559	709,747	1233,007	36,194	5452,935	4256,122
τ_{00}	526,213	138,082	445,067	56,936	1263,417	875,286
τ_{01}	6,653	170,223	2,186	165,756	35,146	198,716
τ_{11}	49,324	10,798	43,137	4,612	81,410	42,884
AIC	2332,158		2310,851		2348,856	

No terceiro cenário, consideramos notas simuladas utilizando a distribuição exponencial potência. Nele, geraram-se $M = 2000$ amostras, onde cada amostra tem 221 valores. Nesse caso, o vetor θ teórico é fixado do ajuste do modelo exponencial potência, tal qual realizado no Capítulo 6. Assim, observando-se a Tabela 6.3, temos os seguintes valores de parâmetros: $\gamma_{00} = 247,847$, $\gamma_{01} = 9,980$, $\gamma_{10} = -17,597$, $\gamma_{11} = -5,123$, $\sigma^2 = 3381,884$, $\tau_{00} = 998,108$, $\tau_{01} = -329,728$, $\tau_{11} = 25,239$. Para cada conjunto de dados simulado e utilizando o modelo descrito na Seção 4.2 foi feito o ajuste aos dados considerando as três distribuições dadas na Tabela 4.4.

Na Tabela 7.3 estão as médias das estimativas dos ajustes das 2000 amostras geradas com distribuição exponencial potência, com os respectivos vieses, para os três modelos considerados. Percebe-se, outra vez, que em todos os casos, o vieses relacionados a γ_{00} , γ_{10} , γ_{01} e γ_{11} são pequenos, enquanto que para σ^2 , τ_{00} , τ_{01} e τ_{11} eles são maiores. Quanto ao critério AIC, o modelo exponencial potência apresentou valor menor que o dos demais modelos. Tal resultado é razoável, uma vez que as notas geradas nesse cenário apresentam b_{0j} , b_{1j} e ϵ_{ij} com distribuição exponencial potência.

Tabela 7.3: Média das estimativas dos parâmetros do ajuste dos modelos normal, t-Student e exponencial potência, respectivos vieses e AIC médio, com dados simulados de uma distribuição exponencial potência.

	Normal		t-Student		Exponencial Potência	
	Estimativa	Viés	Estimativa	Viés	Estimativa	Viés
γ_{00}	247,572	-0,275	247,609	-0,238	247,553	-0,294
γ_{01}	10,061	0,081	10,023	0,043	10,079	0,099
γ_{10}	-17,581	0,016	-17,531	0,067	-17,598	-0,001
γ_{11}	-5,055	0,068	-5,096	0,028	-5,072	0,052
σ^2	1108,688	-2273,196	987,353	-2394,531	2848,135	-533,749
τ_{00}	173,786	-824,322	148,425	-849,683	448,527	-549,581
τ_{01}	9,752	339,480	8,061	337,788	27,308	357,036
τ_{11}	-13,657	-38,896	-10,876	-36,115	-41,813	-67,052
AIC	2206,759		2219,012		2206,750	

Discussão e trabalhos futuros

Neste trabalho, ajustamos e analisamos modelos multiníveis com distribuições simétricas e aplicamos aos dados com uma estrutura hierárquica de grupos. Para tanto, utilizamos, além da distribuição normal, modelos multiníveis com distribuição t-Student e exponencial potência, que podem produzir estimativas mais robustas a observações aberrantes ou que são mais adequados aos dados por terem caudas mais pesadas ou mais leves do que a da distribuição normal. Além disso, com os modelos multiníveis simétricos, avaliamos os dados educacionais a fim de identificar se o número de reprovações na vida acadêmica até o nono ano do ensino fundamental e a infraestrutura da escola são variáveis que influenciam o rendimento escolar. Na análise de diagnóstico, foi empregada a técnica de influência local para identificar a presença de observações aberrantes ou influentes nos modelos ajustados. Um estudo de simulação foi realizado para analisar o viés dos modelos ajustados, onde os valores teóricos considerados foram as estimativas obtidas dos respectivos ajustes dos modelos.

Ao aplicar os modelos multiníveis simétricos aos dados educacionais utilizados nesse trabalho, que tem estrutura hierárquica, concluímos que considerar uma distribuição com cauda mais leve do que a da distribuição normal produz estimativas melhores. Com isso, pudemos concluir que o alto número de reprovações afeta o desempenho dos alunos, que as más condições das escolas não influencia aproveitamento e que a associação dessas duas variáveis implica em um rendimento escolar mais baixo.

A eliminação das observações que se destacaram no conjunto de dados impactou nas estimativas dos parâmetros e no valor do AIC o que sugere que é importante considerar outros modelos que leva em consideração observações possivelmente influentes.

As análises dos resultados obtidos dos ajustes considerando dados simulados sugerem que existe diferença relativa no viés e que é muito importante considerar diferentes modelos para dados educacionais.

Uma das contribuições desse trabalho é nortear ações pedagógicas e administrativas, ou seja, auxiliar os gestores da educação na tomadas de decisões. Assim, pelas análises aqui apresentadas, pode-se indicar uma metodologia a ser seguida ou mostrar onde há uma grande necessidade de apoio pedagógico. Como conclusão, tem-se que o número de reprovações na vida acadêmica do aluno parece influenciar negativamente seu desempenho na avaliação de matemática. Além disso, as más condições das escolas parece não afetar o desempenho escolar por si só. Porém, o número de reprovações alto de um aluno associado as péssimas condições da escola faz com que o rendimento diminua consideravelmente.

Para pesquisas futuras, uma sugestão seria realizar um estudo para investigar o desempenho acadêmico assumindo distribuições assimétricas, uma vez que, utilizando os dados da cidade de São Carlos, o valor da assimetria indicou um pequeno grau de distorção da distribuição em relação a uma distribuição simétrica. Além disso, pode-se analisar mais variáveis explicativas e também aumentar o tamanho da amostra no estudo de simulação.

Acesso aos dados

A base de dados utilizada nesse trabalho foi disponibilizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) que pode ser acessada por meio de CD-ROM ou via download no seguinte endereço eletrônico:

<http://portal.inep.gov.br/web/guest/basica-levantamentos-acessar>

O software utilizado para a leitura e organização dos dados da avaliação foi o SAS e os programas para os inputs de leitura dos arquivos estão disponíveis no manual do usuário disponibilizada, também para download no site do INEP. Além dos comandos para leitura dos arquivos, também consta no manual do usuário, os dicionários das variáveis e os questionários do aluno, professor, diretor e também da escola.

Os modelos multiníveis desta dissertação foram aplicados aos dados de uma amostra aleatória com 221 alunos da cidade de São Carlos, tal que cada escola é representada em torno de 10 alunos por escola, selecionados de forma estratificada em relação ao número de reprovações dos mesmos. Na Tabela A.1, apresentamos o conjunto de dados da amostra, que contém nas primeira, segunda, terceira, quarta e quinta colunas, os códigos que identificam, respectivamente, aluno, escola, nota obtida pelo aluno na prova, a resposta dada pelo aluno para caracterizar a covariável número de reprovações e a resposta para caracterizar a covariável infraestrutura relacionada a escola. Na coluna “Reprovações”, as letras A, B e C representam, respectivamente, que o aluno não teve reprovações no histórico escolar, que teve apenas uma reprovação e que teve duas

ou mais reprovações ($X = 0, 1, 2$) descritas no Capítulo 3. Na coluna "Infraestrutura", as letras C, D e E representa, respectivamente, escolas razoáveis, boas e excelentes ($W = 0, 1, 2$).

Tabela A.1: Amostra de dados utilizada.

Aluno	Escola	Nota	Reprovações	Infraestrutura
5618883	35024430	293,301	A	D
5618927	35024430	328,541	A	D
5618892	35024430	220,568	A	D
5618947	35024430	245,774	A	D
5618953	35024430	236,043	A	D
5618864	35024430	242,397	A	D
5618898	35024430	240,752	A	D
5618962	35024430	265,682	B	D
5618925	35024430	188,281	B	D
5618949	35024430	209,591	C	D
5619054	35024442	327,569	A	D
5619020	35024442	236,796	A	D
5619029	35024442	324,737	A	D
5619023	35024442	319,545	A	D
5618996	35024442	286,704	A	D
5619046	35024442	280,432	A	D
5619003	35024442	222,829	A	D
5618999	35024442	312,741	B	D
5619006	35024442	213,450	B	D
5619055	35024442	204,127	C	D
5619404	35024485	232,335	A	D
5619380	35024485	263,579	A	D
5619394	35024485	220,6318	A	D
5619312	35024485	251,3449	A	D
5619358	35024485	317,0217	A	D
5619389	35024485	280,6438	A	D
5619329	35024485	245,6801	A	D
5619338	35024485	276,7954	A	D
5619315	35024485	185,4358	B	D
5619325	35024485	219,2996	C	D
5619473	35024512	255,2127	A	E
5619537	35024512	294,5164	A	E
5619466	35024512	270,6344	A	E
5619442	35024512	287,4361	A	E
5619440	35024512	347,1723	A	E

Continua na próxima página

Tabela A.1: Amostra de dados utilizada (continuação).

Aluno	Escola	Nota	Reprovações	Infraestrutura
5619450	35024512	321,9569	A	E
5619476	35024512	314,3811	A	E
5619453	35024512	254,9419	A	E
5619515	35024512	303,7672	A	E
5619427	35024512	256,7688	A	E
5619426	35024512	309,3824	A	E
5619685	35024557	331,4681	A	C
5619561	35024557	283,5492	A	C
5619639	35024557	296,9025	A	C
5619578	35024557	262,7724	A	C
5619733	35024557	202,6666	A	C
5619672	35024557	305,1761	A	C
5619697	35024557	303,854	A	C
5619604	35024557	248,5316	A	C
5619729	35024557	301,3079	B	C
5619610	35024557	194,2898	B	C
5619724	35024557	237,9147	C	C
5619775	35024612	205,9505	A	D
5619815	35024612	230,4102	A	D
5619791	35024612	282,2862	A	D
5619813	35024612	224,7917	A	D
5619812	35024612	221,6821	A	D
5619811	35024612	257,1986	A	D
5619789	35024612	223,653	A	D
5619838	35024612	281,9575	B	D
5619835	35024612	164,9665	B	D
5619832	35024612	238,5435	C	D
5619861	35024624	296,4538	A	E
5619966	35024624	238,3458	A	E
5619909	35024624	228,991	A	E
5619944	35024624	328,6069	A	E
5619996	35024624	301,651	A	E
5619937	35024624	299,0283	A	E
5619884	35024624	278,1465	A	E
5619913	35024624	264,9525	A	E
5620005	35024624	226,2631	A	E
5620012	35024624	266,787	B	E
5620088	35024624	163,5835	C	E
5620105	35036432	252,0005	A	C
5620127	35036432	230,5976	A	C
5620150	35036432	182,6767	A	C

Continua na próxima página

Tabela A.1: Amostra de dados utilizada (continuação).

Aluno	Escola	Nota	Reprovações	Infraestrutura
5620145	35036432	302,453	A	C
5620144	35036432	204,8172	A	C
5620109	35036432	200,7935	A	C
5620106	35036432	160,6034	A	C
5620132	35036432	252,9813	A	C
5620102	35036432	204,9852	B	C
5620119	35036432	177,3809	B	C
5620099	35036432	203,5924	C	C
5620164	35049992	175,0821	A	C
5620179	35049992	235,6267	A	C
5620161	35049992	229,0767	A	C
5620168	35049992	237,1859	A	C
5620182	35049992	239,1039	A	C
5620169	35049992	224,557	A	C
5620165	35049992	177,9994	A	C
5620175	35049992	206,8209	B	C
5620178	35049992	178,7653	B	C
5620172	35049992	240,9311	B	C
5620179	35049992	235,6267	A	C
5620250	35127887	188,2987	A	D
5620219	35127887	251,4261	A	D
5620213	35127887	207,4654	A	D
5620201	35127887	208,9063	A	D
5620244	35127887	229,171	A	D
5620205	35127887	220,49	A	D
5620200	35127887	250,4943	B	D
5620251	35127887	226,2235	B	D
5620241	35127887	191,533	B	D
5620225	35127887	228,0031	C	D
5620243	35127887	231,5885	C	D
5620303	35191528	234,1456	A	D
5620288	35191528	243,8899	A	D
5620279	35191528	224,919	A	D
5620289	35191528	157,9396	A	D
5620275	35191528	197,3129	A	D
5620274	35191528	192,9861	A	D
5620272	35191528	180,3754	A	D
5620300	35191528	233,9907	A	D
5620299	35191528	205,1984	B	D
5620270	35191528	222,6735	B	D
5620310	35900795	247,9028	A	D

Continua na próxima página

Tabela A.1: Amostra de dados utilizada (continuação).

Aluno	Escola	Nota	Reprovações	Infraestrutura
5620315	35900795	282,4355	A	D
5620346	35900795	233,1759	A	D
5620340	35900795	188,3326	A	D
5620357	35900795	228,2199	A	D
5620341	35900795	212,7906	A	D
5620368	35900795	240,0191	A	D
5620331	35900795	298,4582	A	D
5620355	35900795	180,2224	B	D
5620363	35900795	241,6094	B	D
5620365	35900795	219,7098	C	D
5620493	35905835	236,6332	A	D
5620419	35905835	238,783	A	D
5620400	35905835	281,721	A	D
5620411	35905835	242,4168	A	D
5620374	35905835	278,1763	A	D
5620388	35905835	303,1003	A	D
5620465	35905835	189,1288	A	D
5620430	35905835	296,2527	B	D
5620495	35905835	212,1348	B	D
5620394	35905835	235,082	C	D
5620529	35906773	250,3245	A	C
5620573	35906773	232,3388	A	C
5620581	35906773	248,4375	A	C
5620550	35906773	252,3922	A	C
5620539	35906773	211,9526	A	C
5620528	35906773	183,1818	A	C
5620518	35906773	280,6847	A	C
5620538	35906773	222,1397	A	C
5620567	35906773	232,1479	B	C
5620590	35906773	168,1264	C	C
5620631	35907923	283,4008	A	C
5620622	35907923	305,2151	A	C
5620632	35907923	284,2417	A	C
5620626	35907923	234,0317	A	C
5620634	35907923	318,7209	A	C
5620654	35907923	230,6482	A	C
5620645	35907923	224,7035	B	C
5620651	35907923	261,4008	B	C
5620653	35907923	279,5547	C	C
5620623	35907923	250,0905	C	C
5620735	35914988	230,2286	A	D

Continua na próxima página

Tabela A.1: Amostra de dados utilizada (continuação).

Aluno	Escola	Nota	Reprovações	Infraestrutura
5620714	35914988	234,2938	A	D
5620766	35914988	208,3169	A	D
5620694	35914988	264,4674	A	D
5620809	35914988	345,9877	A	D
5620781	35914988	230,8189	A	D
5620768	35914988	205,6196	A	D
5620761	35914988	191,8424	A	D
5620780	35914988	191,6865	B	D
5620824	35914988	190,3106	B	D
5620841	35914988	198,8341	C	D
5621011	35914997	290,8007	A	D
5621081	35914997	279,8999	A	D
5620940	35914997	254,287	A	D
5621038	35914997	251,1711	A	D
5620931	35914997	233,025	A	D
5621010	35914997	232,2939	A	D
5620981	35914997	284,5986	A	D
5620930	35914997	222,2602	A	D
5621053	35914997	204,1826	B	D
5621118	35914997	217,1764	B	D
5621113	35914997	229,4048	C	D
7923915	35077461	270,0694	A	D
7923919	35077461	281,1364	A	D
7923982	35077461	293,096	A	D
7923947	35077461	347,6448	A	D
7923975	35077461	258,8698	A	D
7923931	35077461	257,4785	A	D
7923926	35077461	255,7244	A	D
7923948	35077461	336,5202	A	D
7923892	35077461	248,2054	B	D
7923902	35077461	220,5219	C	D
7924007	35077495	343,8082	A	C
7924069	35077495	239,4177	A	C
7924036	35077495	288,5687	A	C
7924030	35077495	286,2915	A	C
7924021	35077495	276,755	A	C
7924057	35077495	310,5612	A	C
7924015	35077495	197,369	A	C
7924011	35077495	304,9639	A	C
7923999	35077495	198,8175	B	C
7924068	35077495	199,5869	B	C

Continua na próxima página

Tabela A.1: Amostra de dados utilizada (continuação).

Aluno	Escola	Nota	Reprovações	Infraestrutura
7924006	35077495	185,2711	C	C
7924089	35224169	315,6807	A	D
7924088	35224169	243,3263	A	D
7924118	35224169	276,019	A	D
7924078	35224169	273,5336	A	D
7924142	35224169	276,6121	A	D
7924092	35224169	219,2271	A	D
7924107	35224169	215,12	B	D
7924136	35224169	247,9018	B	D
7924143	35224169	179,4253	B	D
7924114	35224169	205,9411	C	D

Comandos em R

O seguinte código em R é utilizado para realizar o ajuste dos parâmetros do modelo multinível simétrico apresentado no Capítulo 4 para os dados escolares fornecidos no Apêndice A. Como pré-condição, ele supõe que os dados da Tabela A.1 estejam armazenados na variável `dados`. Sua pós-condição são as variáveis `beta` e `xk`, contendo, respectivamente, as estimativas de γ e τ , previamente definidos.

```
# O programa para estimar os parâmetros da regressão começa aqui.

# Carrega os dados
dados <- read.table('dados_tabela_apêndice_A.txt')

nr_escolas <- length(unique(dados[,2])) # Indica o nr de escolas (21, no caso)
s<-function(i) { ma[i]:(ma[i+1]-1) } # A função s(i) retorna os alunos da escola i
Imi <- function(mi){diag(rep(1,mi))} # Matriz identidade com dimensão mixmi

t <- cbind(table(dados[,2]))
o <- t[as.character(unique(dados[,2])),] # Indica a quantidade de alunos por escola

# As 6 linhas seguintes constroem o ma, ou seja, onde começa e onde termina os alunos em cada escola.
ma <- 1
for(i in 1:nr_escolas)
{
ma <- c(ma,ma[length(ma)]+o[i])
}
ma <- as.numeric(ma)

# Aqui, inicializam-se as variáveis se acordo com as nomenclaturas da regressão
```

```

y <- dados[,3]
x <- dados[,4]
w <- dados[,5]
n <- length(y) # Indica a quantidade de alunos

In <- function(n){diag(rep(1,n))} # Constrói a matriz identidade de tamanho nxn

p<-4 # Quantidade de parâmetros a estimar: gamma00, gamma01, gamma10, gamma11
u<-rep(0,nr_escolas)
mi <- as.numeric(o) # Quantidade de alunos nas escolas

# Estimativas iniciais: usa-se lm pro modelo multinível linear
data_lm <- as.data.frame(cbind(dados[,3],dados[,4],dados[,5])) # Matriz com (y,x,w)
colnames(data_lm) <- c('y','x','w')
l <- lm(y ~ x*w,data=data_lm) # Realiza a regressão linear

# As 4 linhas abaixo inicializam os gammas que se quer estimar
gamma00 <- as.numeric(l$coefficients[1])
gamma01 <- as.numeric(l$coefficients[3])
gamma10 <- as.numeric(l$coefficients[2])
gamma11 <- as.numeric(l$coefficients[4])

beta<-c(gamma00,gamma01,gamma10,gamma11) #0 4 gammas que se quer estimar sao agrupados no vetor "beta"

sigma2= (summary(l)$sigma)^2 # Estimativa inicial de sigma2
d11<-1; d12<--1; d22<-1 # Estimativas iniciais das variâncias
xk<-c(sigma2,d11,d12,d22) # Esse eh vetor corresponde ao vetor Tau na dissertação
xkmais <- xk # Esse xkmais é o xk na iteração seguinte
D<-matrix(c(d11,d12,d12,d22),nrow=2,byrow=T) # Matriz das variâncias

z <- cbind(rep(1,length(x)),x)

# As 3 linhas abaixo fazem a estimativa da variância (por escola), e depois o inverso de Sigma
Sigma<-matrix(0,nrow=n,ncol=n)
for (i in 1:nr_escolas) Sigma[s(i),s(i)] <- z[s(i),] %*% D %*% t(z[s(i),])+ sigma2 * In(mi[i])
SigmaInv<-svd.inverse(Sigma) #qr.solve(Sigma,tol=1e-100000)

vetorPar = c(c(beta),sigma2,d11,d12,d22) # Todos os parâmetros que se quer estimar

J <- cbind(rep(1,n),w,x,x*w)

# Variáveis utilizadas considerando o modelo normal
Wgi = rep(-.5,nr_escolas)
WgLinha = rep(0,nr_escolas)
dgi = mi/4
fgi = mi*(mi+2)/4
vi = -2*Wgi

# O método iterativo para o cálculo dos parâmetros começa aqui.

controleWhile <- 10 # Realiza o controle se deve ou não sair do laço
contador <- 0 # informa a quantidade de vezes que o laço iterou.

while(abs(controleWhile)>1e-3 && contador <50)
{

```

```

contador <- contador+1
vetorPar0 <- vetorPar

# As 8 linhas abaixo fazem a atualização dos valores do vetor beta
M1 <- matrix(0,ncol=ncol(J),nrow=ncol(J))
M2 <- matrix(0,ncol=1,nrow=ncol(J))
for(i in 1:nr_escolas)
{
  M1 <- M1 + (4*dgi[i]/mi[i]*t(J[s(i),]) %% SigmaInv[s(i),s(i)] %% J[s(i),])
  M2 <- M2 + vi[i]*t(J[s(i),])%% SigmaInv[s(i),s(i)] %% (y[s(i)] - (J[s(i),] %% beta) )
}
beta <- beta + qr.solve(M1,tol=1e-100000) %% M2

r <- y - (J %% beta)

# Calcula a função Escore de Fisher
Escore = function()
{
  c11<-c12<-c13<-c14<-rep(0,nr_escolas)

  for (i in 1:nr_escolas)
  {
    c11[i] <- sum(diag(SigmaInv[s(i),s(i)])) -vi[i] * t((r)[s(i)]) %%
      SigmaInv[s(i),s(i)] %% SigmaInv[s(i),s(i)] %% (r)[s(i)]
    c12[i] <- sum(diag(SigmaInv[s(i),s(i)] %% (z[s(i),1] %% t(z[s(i),1])))) -
      vi[i] * t((r)[s(i)]) %% SigmaInv[s(i),s(i)] %% (z[s(i),1] %%
      t(z[s(i),1])) %% SigmaInv[s(i),s(i)] %% (r)[s(i)]
    c13[i] <- sum(diag(SigmaInv[s(i),s(i)] %% ((z[s(i),1] %% t(z[s(i),2])) +
      (z[s(i),2] %% t(z[s(i),1])))) -vi[i] * t((r)[s(i)]) %%
      SigmaInv[s(i),s(i)] %% ((z[s(i),1] %% t(z[s(i),2]))+(z[s(i),2] %%
      t(z[s(i),1])))) %% SigmaInv[s(i),s(i)] %% (r)[s(i)]
    c14[i] <- sum(diag(SigmaInv[s(i),s(i)] %% (z[s(i),2] %% t(z[s(i),2])))) -
      vi[i] * t((r)[s(i)]) %% SigmaInv[s(i),s(i)] %% (z[s(i),2] %%
      t(z[s(i),2])) %% SigmaInv[s(i),s(i)] %% (r)[s(i)]
  }
  c(-1/2*sum(c11), -1/2*sum(c12), -1/2*sum(c13), -1/2*sum(c14))
}

# Calcula a matriz de informação de Fisher
MatrizInformacao<-function()
{
  brsi11<-brsi12<-brsi13<-brsi14<-brsi22<-brsi23<-brsi24<-brsi33<-brsi34<-brsi44<-rep(0,nr_escolas)
  for (i in 1:nr_escolas)
  {
    brsi11[i] <- sum(diag(SigmaInv[s(i),s(i)]))^2
    brsi12[i] <- sum(diag(SigmaInv[s(i),s(i)])) *sum(diag(SigmaInv[s(i),s(i)] %%
      (z[s(i),1] %% t(z[s(i),1]))))
    brsi13[i] <- sum(diag(SigmaInv[s(i),s(i)])) *sum(diag(SigmaInv[s(i),s(i)] %%
      (z[s(i),1] %% t(z[s(i),2]) + (z[s(i),2] %% t(z[s(i),1]))))
    brsi14[i] <- sum(diag(SigmaInv[s(i),s(i)])) *sum(diag(SigmaInv[s(i),s(i)] %%
      (z[s(i),2] %% t(z[s(i),2]))))
    brsi22[i] <- sum(diag(SigmaInv[s(i),s(i)] %% (z[s(i),1] %% t(z[s(i),1]))))^2
    brsi23[i] <- sum(diag(SigmaInv[s(i),s(i)] %% (z[s(i),1] %% t(z[s(i),1])))) *
      sum(diag(SigmaInv[s(i),s(i)] %% ((z[s(i),1] %% t(z[s(i),2])) +
      ((z[s(i),2] %% t(z[s(i),1]))))

```

```

brsi24[i] <- sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*% t(z[s(i),1]))) *
  sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])))))
brsi33[i] <- sum(diag(SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*% t(z[s(i),2])) +
  ((z[s(i),2] %*% t(z[s(i),1]))))))^2
brsi34[i] <- sum(diag(SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*% t(z[s(i),2])) +
  ((z[s(i),2] %*% t(z[s(i),1])))))) * sum(diag(SigmaInv[s(i),s(i)] %*%
  (z[s(i),2] %*% t(z[s(i),2])))))
brsi44[i] <- sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])))))^2
}
K11<-K12<-K13<-K14<-K22<-K23<-K24<-K33<-K34<-K44<-U1<-U2<-U3<-U4<-rep(0,nr_escolas)
for (i in 1:nr_escolas)
{
  K11[i] <- brsi11[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% SigmaInv[s(i),s(i)]))
  K12[i] <- brsi12[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*%
    t(z[s(i),1])))))
  K13[i] <- brsi13[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*%
    t(z[s(i),2]) + (z[s(i),2] %*% t(z[s(i),1]))))))
  K14[i] <- brsi14[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*%
    t(z[s(i),2])))))
  K22[i] <- brsi22[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*% t(z[s(i),1])) %*%
    SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*% t(z[s(i),1])))))
  K23[i] <- brsi23[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*% t(z[s(i),1])) %*%
    SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*% t(z[s(i),2]))+(z[s(i),2] %*%
    t(z[s(i),1]))))))
  K24[i] <- brsi24[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),1] %*% t(z[s(i),1])) %*%
    SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])))))
  K33[i] <- brsi33[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*% t(z[s(i),2]))+
    ((z[s(i),2] %*% t(z[s(i),1])))) %*% SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*%
    t(z[s(i),2]))+(z[s(i),2] %*% t(z[s(i),1]))))))
  K34[i] <- brsi34[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])) %*%
    SigmaInv[s(i),s(i)] %*% ((z[s(i),1] %*% t(z[s(i),2]))+(z[s(i),2] %*%
    t(z[s(i),1]))))))
  K44[i] <- brsi44[i]/4*(4*fgi[i]/(mi[i]*(mi[i]+2))-1)+2*fgi[i]/(mi[i]*(mi[i]+2))*
    sum(diag(SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])) %*%
    SigmaInv[s(i),s(i)] %*% (z[s(i),2] %*% t(z[s(i),2])))))
}
matrix(c(sum(K11), sum(K12), sum(K13), sum(K14), sum(K12), sum(K22), sum(K23), sum(K24),
  sum(K13), sum(K23), sum(K33), sum(K34), sum(K14), sum(K24), sum(K34), sum(K44)),nrow=4,ncol=4,byrow=T)
}

S <- Escore()
matriz_informacao <- MatrizInformacao()

xkmais <- xk + svd.inverse(matriz_informacao) %*% S
xk <- xkmais

```

```
sigma2 <- xk[1]
d11 <- xk[2]
d12 <- xk[3]
d22 <- xk[4]

D <- matrix(c(d11,d12,d12,d22),nrow=ncol(z),byrow=T)

Sigma <- matrix(0,nrow=n,ncol=n)
for (i in 1:nr_escolas) Sigma[s(i),s(i)] <- z[s(i),] %*% D %*% t(z[s(i),])+ sigma2 * Imi(mi[i])
SigmaInv <- svd.inverse(Sigma)

# Variáveis utilizadas considerando o modelo normal
Wgi = rep(-.5,nr_escolas)
WgLinha = rep(0,nr_escolas)
dgi = mi/4
fgi = mi*(mi+2)/4
vi = -2*Wgi

# Vetor de parâmetros
vetorPar = c(c(beta),sigma2,d11,d12,d22)

# Atualiza a variável da condição de parada
controleWhile = max(abs(1-vetorPar[which(vetorPar0!=0)]/vetorPar0[which(vetorPar0!=0)]))

print(controleWhile)
}

# Realiza a impressão dos valores estimados para os parâmetros
print(beta)
print(round(xk,8))
```

Referências Bibliográficas

- Aitkin, M., Bennett, S. N., and Hesketh, J. (1981). Teaching styles and pupil progress: a re-analysis. *British Journal of Educational Psychology*, 51(2):170–186.
- Atkinson, A. C. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford.
- Belsey, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley and Sons, New York.
- Bennett, N. (1976). *Teaching styles and pupil progress*. Times Educational Supplement (London).
- Bergamo, G. C. (2002). Aplicação de modelos multiníveis na análise de dados de medidas repetidas no tempo. Dissertação de mestrado, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, Brasil.
- Burnahm, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Verlag, New York, 2^a edition.
- Casella, G. and Berger, R. L. (2001). *Statistical inference*. Duxbury Thompson Learning, California, USA.

- Castro, M. (2011). Variabilidade intrapessoal e interpessoal da ingestão de nutrientes de crianças brasileiras. Dissertação de mestrado, Departamento de Nutrição, Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, Brasil.
- Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):pp. 133–169.
- Cysneiros, F., Paula, G., and Galea, M. (2005). Modelos simétricos aplicados. Apostila da IX Escola de Modelos de Regressão, Águas de São Pedro, Brasil.
- Cysneiros, F. J. A. (2004). *Métodos Restritos e Validação de Modelos Simétricos de Regressão*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Fang, K.-T. and Anderson, T. W. (1990). *Statistical inference in elliptically contoured and related distributions*. Allerton Press, New York, USA.
- Ferrão, M. (2003). *Introdução aos modelos de regressão multinível em educação*. Editora Komedi.
- Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*, volume Analytical methods for social research. Cambridge University Press, New York, USA.
- Goldstein, H. (2011a). *Multilevel Statistical Models*. Wiley Series in Probability and Statistics. John Wiley & Sons.
- Goldstein, H. (2011b). *Multilevel Statistical Models*. Wiley Series in Probability and Statistics. John Wiley & Sons.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Taylor & Francis.
- Jesus, G. and Laros, J. (2004). Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, 3(2):93–106.

- Laros, J. and Marciano, J. (2008). Análise multinível aplicada a dados do nels: 88. *Estudos em Avaliação Educacional*, 19(40):263–278.
- Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, 54(2):570–582.
- Lindsey, J. and Lindsey, P. (2006). Multivariate distributions with correlation matrices for nonlinear repeated measurements. *Computational Statistics & Data Analysis*, 50(3):720–732.
- Lobos, C. M. V. (2010). *Modelos log-Birnbaum-Saunders mistos*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Manghi, R. (2012). Modelos elípticos multiníveis. Dissertação de mestrado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Natis, L. (2000). Modelos lineares hierárquicos. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade São Paulo, São Paulo.
- Natis, L. (2001). Modelos lineares hierárquicos. *Estudos em Avaliação Educacional*, 23:3–30.
- Nobre, J. S. (2004). Métodos de diagnóstico para modelos lineares mistos. Dissertação de mestrado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Nobre, J. S. and Singer, J. d. M. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, 49(6):863–875.
- Osorio, F. (2006). *Diagnóstico de influência em modelos elípticos com efeitos mistos*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Osorio, F., Paula, G., and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational statistics & data analysis*, 51(9):4354–4368.
- Paula, G. (2004). *Modelos de regressão: com apoio computacional*. IME-USP Sao Paulo:.

- Pinho, L. G. B., Nobre, J. S., and de Freitas, S. M. (2012). On linear mixed models and their influence diagnostics applied to an actuarial problem. *Chilean Journal of Statistics*, 3(1):57–73.
- Pires, J. F. (2009). Influência local baseada na curvatura normal em modelos multiníveis. Dissertação de mestrado, Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife, Brasil.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Raudenbush, S. and Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage Publications, Inc.
- Russo, C. M. (2006). Análise de um modelo de regressão com erros nas variáveis multivariado com intercepto nulo. Dissertação de mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- Russo, C. M. (2010). *Modelos não lineares elípticos para dados correlacionados*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Russo, C. M., Aoki, R., and Paula, G. A. (2011). Assessment of variance components in nonlinear mixed-effects elliptical models. *TEST*, 21(3):519–545.
- Savalli, C. (2005). *Teste do tipo score para componentes de variância em modelos elípticos lineares mistos*. Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil.
- Soares, T. (2005). Modelo de três níveis hierárquicos para a proficiência dos alunos de 4a série avaliados no teste de língua portuguesa do simave/proeb-2002. *Revista Brasileira de Educação*, 29:73–87.
- Souza, E. C. (2006). Análise de influência local no modelo de regressão logística. Dissertação de mestrado, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, Brasil.

- Zanini, R., Moraes, A., Giugliani, E., and Riboldi, J. (2011). Determinantes contextuais da mortalidade neonatal no rio grande do sul por dois modelos de análise. *Revista de Saude Pública*, 45(1):79–89.