UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE ESTATÍSTICA

KELLY PEREIRA DE LIMA

MODELOS MULTINÍVEIS

FORTALEZA
2014

KELLY PEREIRA DE LIMA

MODELOS MULTINÍVEIS

Monografia apresentada ao curso de Estatística do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Ronald Targino Nojosa

FORTALEZA

2014

KELLY PEREIRA DE LIMA

MODELOS MULTINÍVEIS

Monografia apresentada ao curso de Estatística do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Estatística.

Aprovada em __/__/____.

BANCA EXAMINADORA

____

Prof. Dr. Ronald Targino Nojosa (Orientador)
Universidade Federal do Ceará (UFC)

____

Prof. Dr. Juvêncio Santos Nobre
Universidade Federal do Ceará (UFC)

____

Prof[a]. Dr[a]. Maria Jacqueline Batista
Universidade Federal do Ceará (UFC)

Dedico esse trabalho

À Deus acima de tudo.
Aos meus pais Maria Lúcia Avelino e Ozarias Marques, pelo apoio e a confiança que sempre depositaram em mim. Esta monografia dedico a vocês minha fonte de inspiração.

"O que vale na vida não é o ponto de partida e sim a caminhada. Caminhando e semeando, no fim terás o que colher."

Cora Coralina

"Tudo o que é bom dura o tempo necessário para ser inesquecível."

Fernando Pessoa

AGRADECIMENTOS

Foram anos de muita dedicação, perseverança, superação e satisfação pessoal ao longo do curso de Estatística. Ao concluir mais uma etapa da minha vida, agradeço a todas as pessoas que estiveram me apoiando.

Agradeço primeiramente a Deus que me concedeu e permitiu o gozo da saúde e da capacidade de aprender.

Aos meus pais pelo amor incondicional, pela formação do meu caráter e pelo exemplo que são de homem e mulher. Não tenho palavras para agradece todos os momentos que vocês me proporcionaram. Essa vitória não é minha é de vocês. Amo vocês.

Aos meus irmãos, sobrinhos e tios pelo apoio e palavras de incentivos durante toda minha graduação.

Uma pessoa muito especial, Adaílton pelo seu infinito conselho espiritual, suas sábias palavras e por acreditar muito em mim. Muito obrigada por sua ajuda.

Agradeço ao meu orientador Professor Ronald Targino, por todos os ensinamentos estatísticos e pela sua orientação. Além disso, agradeço por me aceitar como sua orientanda. Foi e será um privilégio poder contar com a sua ajuda.

Um agradecimento muito especial, Raquel Araújo (a minha SMA), por permanecemos sempre junta nesta longa caminhada, pela palavras de incentivos e por cada momento vivenciado durante essa jornada. Obrigado por ser essa pessoa tão especial que encanta a todos que convivem com você.

A Tia Nukácia e Tia Régia, por me considerar um membro da sua família e ensinar que a mulher tem uma força extraordinária. Além disso, muitíssimo obrigado por cada conselho, ensinamento e hospedagem. Vocês são à força da mulher brasileira.

Não posso deixar de agradecer três pessoas muito especiais que encontrei durante a graduação, são eles: Janaína Marques, Yuri Alves e Bruno Pinheiro (o Galo), quantos estudos, discussões, troca de ideias, telefonemas e reuniões na casa do Yuri (agradecendo a dona Elienice por cada lanche preparado), com isso conseguimos prosseguir no curso.

Um muito obrigado a elas: Laura Mlr, Janaina Mlr e Raquel Mlr. Vocês me ensinaram o verdadeiro significado da palavra companheirismo e sempre pensar no próximo.

A todos os professores do DEMA quero deixar o meu muito obrigado, pois cada um contribuiu para o meu crescimento dentro do curso. Quero deixar um agradecimento ao prof. Juvêncio pelos momentos de aprendizagem durante as disciplinas ministradas que foram de suma importância. Além disso, quero deixar um agradecimento especial aos professores Sílvia e Maurício, pela sua ajuda durante o estudo para o mestrado.

Agradeço a Márgeri, Luiza e Erione pelo ajudar no momento de aperreio e por sempre estarem dispostas ajudar.

Ao Júnior, funcionário da biblioteca da Matemática, por está sempre dispostos a ajudar e por quebrar "uns galhos".

Agradeço cada pessoa que contribuiu de forma direta e indireta na construção desta monografia. Sempre terão o meu muito obrigado.

RESUMO

Modelar um conjunto de dados é buscar uma representação matemática fidedigna para o fenômeno em estudo; o caso ideal seria especificar um modelo parcimonioso, ou seja, aquele que envolva um número mínimo de parâmetros a serem estimados sem comprometer a explicação do modelo escolhido. Os modelos clássicos de regressão linear são frequentemente aplicados nas mais diversas áreas do conhecimento, tais como: agronomia, ciências sociais, administração, engenharias, biologia, educação, etc. Entretanto, em situações em que os dados investigados possuem uma estrutura de agrupamento ou hierarquia, ou seja, que são caracterizados pela presença de unidades experimentais agregadas em outras unidades maiores, estes dados necessitam de um tratamento diferenciado. Com isso, as técnicas estatísticas que tem sido mais utilizadas são as dos modelos multiníveis que foram desenvolvidos para análise de dados que possuem uma estrutura hierárquica, ou seja, estrutura de grupo, pois levam em consideração a dependência existente dentro de cada nível hierárquico e entre os níveis. O presente trabalho tem por objetivo apresentar a estrutura dos modelos multiníveis, além de compará-los com os modelos clássicos de regressão linear. Será utilizada uma aplicação a partir dos microdados do SAEB com edição de 2011. Todos os procedimentos e implementações foram feitos usando o *software* de código livre `R`.

Palavras-chave: **Estrutura hierárquica. Modelos lineares. *Software* `R`.**

ABSTRACT

Model a dataset is to find a reliable mathematical representation for the phenomenon under study, the ideal case would be to specify a parsimonious model, ie, one that involves a minimum number of parameters to be estimated without compromising the explanation of the chosen model. The classical linear regression models are often applied in various areas of knowledge, such as agronomy, social sciences, management, engineering, biology, education, and so on. However, in situations where the data have investigated a structure hierarchy or grouping, or which are characterized by the presence of aggregated units in other experimental larger units, these data need special treatment. Thus, a statistical technique that has been most used are those of the multilevel models that have been developed for the analysis of data that have a hierarchical structure, or group structure, because they take into account the dependency within each hierarchical level and between levels. This paper aims to present the structure of multilevel models, and compare them with the classical linear regression models. An application from microdata Saeb with the 2011 edition will be used. All procedures and implementations were done using the *software* free `R` code.

Keywords: **Hierarchical structure. Linear models. Software `R`.**

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O modelo multinível (MM) ou modelo linear hierárquico (MLH) leva em consideração uma estrutura aninhada para os dados, combinando informações de variáveis de diferentes níveis e reduzindo a perda de informação (GOLDSTEIN, 1995; BRYK; RAUDENBUSH, 1992).

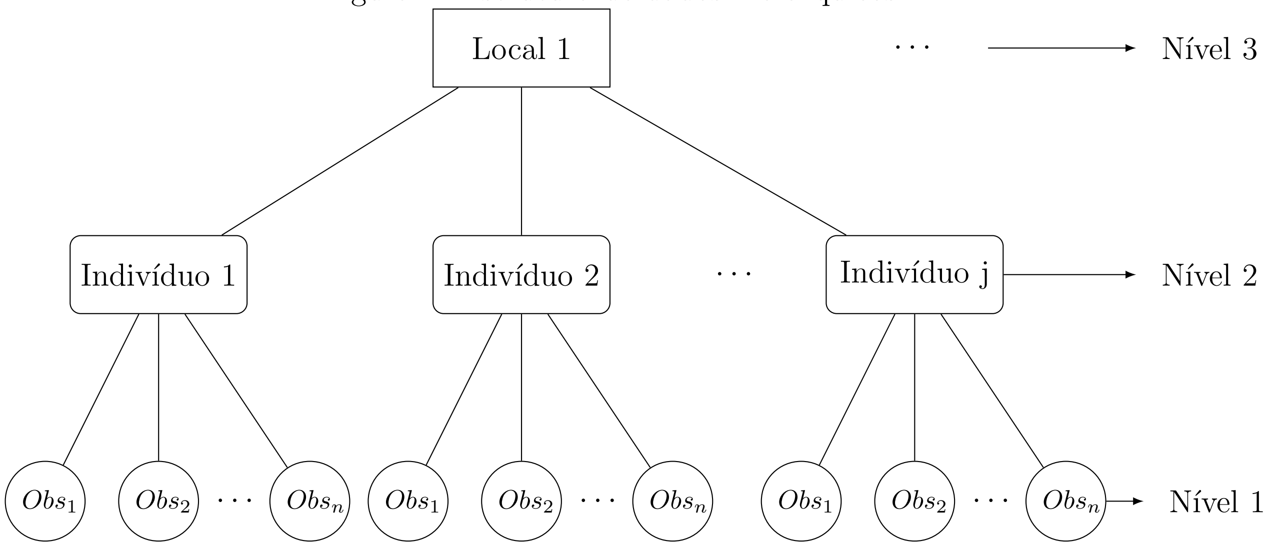
Os modelos clássicos de regressão são muitos usados nas mais diversas áreas de pesquisa, um exemplo que pode-se citar é na área educacional quando se quer explicar a proficiência alcançada pelos alunos a partir de variáveis como sexo, condições socioeconômica e experiência do professor.

Vale ressaltar que os modelos clássicos empregam quatro suposições fundamentais para as características dos dados: linearidade, normalidade, independência entre as observações e homocedasticidade. Porém, para dados com uma estrutura de hierarquia ou multinível a independência para as observações, já não se verifica.

No modelo de regressão tradicional assume-se a independência das observações, caso essa suposição seja violada (isso acontece quando têm-se dados hierárquicos) resultados equivocados podem ser obtidos na análise. Posteriormente será abordado um exemplo em que inicialmente não se considera a estrutura hierárquica e em seguida considera-se, assim evidenciando que os resultados são diferentes (ver próxima seção). De acordo Cruz (2010), outra situação possível é quando os valores dos estimadores dos erros padrões são bem menores em relação ao modelo de regressão tradicional, tornando assim os resultados falsamente significativos. Além do mais uma consequência da dependência entre as observações é a subestimação dos erros padrões dos coeficientes da regressão.

Um exemplo de estrutura hierárquica pode ser visto na Figura 1. Para este caso, as observações representam o nível 1, os indivíduos representam o nível 2 e as localidades representam o nível 3.

Figura 1: Estrutura de dados hierárquicos



"Estruturas hierárquicas de dados são caracterizadas pela presença de unidades experimentais agrupadas em outras unidades maiores, em que as fontes de variabilidade se encontram aninhadas" (Bogutchi, 2010, p.34). Para analisar esse tipo de estrutura citada por Bogutchi existem os MM's que passaram a ser usado a partir do final da década de 80, pois os modelos clássicos não levam em consideração a estrutura multinível dos dados.

Modelos para esse tipo de dados apresentam diversas nomenclaturas na literatura tais como: "modelo de coeficiente aleatório" Leeuw & Kreft (1986, *apud* Valente 2007), "modelo de componente de variância" Longford (1987, *apud* Valente 2007) e "modelo linear hierárquico" Raudenbush & Bryk (1992, *apud* Valente 2007).

Nos MM's existem uma especificação para cada um dos níveis de hierarquia e a incorporação de efeitos aleatórios associado a cada um desses níveis. O presente trabalho tem por objetivo apresentar a estrutura dos MM's, além de compará-los com os modelos clássicos de regressão linear.

1.1 Motivação

Os MM's são empregados para a análise de estudos que agrupam indivíduos dentro de seus grupos ou contextos sociais, analisando os efeitos combinados tantos das variáveis individuais como das variáveis de grupos, possibilitando a investigação da interação das variáveis nos mais diferente níveis de agrupamento. As estruturas hierárquicas podem ser encontradas nas mais diversas áreas tais como: demografia, sociologia, saúde, educação, esporte e epidemiologia. A seguir serão ilustrados alguns exemplos.

Quando os especialistas desejam investigar de que forma as diferenças no desenvolvimento da economia nacional podem interferir na relação entre grau educacional dos adultos e a taxa de fertilidade. Geralmente, essa investigação combina indicadores econômicos (como IDH, índice de Gini e outros) recolhidos a nível nacional com as informações relacionada a nível domiciliar (como a educação e a fertilidade). Desta maneira, se tem dois níveis que são países e domicílios. Nessa área, Demografia, pode-se citar Lobo (2007) que faz uma aplicação dos modelos multiníveis com enfoque bayesiano para analisar a evolução da mortalidade por doenças isquêmicas do coração, no estado do Rio de Janeiro.

Na Sociologia, Soares e Alves (2007) analisaram os impactos dos processos escolares sobre os resultados dos alunos com o intuito de diminuir a diferença de resultados escolares entre grupos sociais, a partir de dados longitudinais.

Na saúde, Cruz (2008) verificou a existência de estrutura hierárquica na epidemiologia, pois ao fazer uma investigação examinou uma relação de dependência entre a saúde dos indivíduos e as zonas geográficas onde habitam. Além do que pode-se ter a estrutura de dependência como os tratamentos recebidos pelos pacientes e o serviço de saúde. Sicheiri e Moura (2009) utilizaram a regressão multinível para verificar as variações no índice de massa corporal (IMC) entre adultos segundo os fatores individuais e a característica do ambiente da cidade.

Na educação, grande parte dos estudos relacionam a proficiência dos alunos com algumas características (tais como idade, gêneros, estrutura da escola, experiência do professor e etc). Além do mais, na literatura são considerados para esse tipo de estudo, três níveis: alunos (1º nível), turmas (2º nível) e escolas (3º nível). Segundo Cruz (2010), esse tipo de estudo pode acontecer quando o pesquisador tem interesse em investigar como as variáveis professor, disciplina, habitação (local onde vive) influenciam na classificação escolar do aluno. Nesse mesmo contexto, cita-se Albernaz, Ferreira e Franco (2002) que desenvolveram uma pesquisa para avaliar, por meio dos dados SAEB (Sistema Nacional de Educação Brasileira), o efeito de variáveis como a escolaridade do professor e a qualidade da infraestrutura física no desempenho dos estudantes. O português Valente (2007) utilizou o MLH para estudar a relevância do apoio das escolas nas perspectivas profissionais dos alunos do 10º ano de escolaridade, sendo adotado um modelo com dois níveis: alunos e escolas. Ainda na área da educação, Murillo (2008) faz referência ao MM como uma ferramenta fundamental, pois o mesmo relaciona os fatores escolares que podem afetar ou estão associados ao rendimento dos estudantes.

No esporte, Maia *et al* (2003) demonstraram a necessidade de considerar a estrutura hierárquica. Considerando isso, Lopes *et al* (2010) modelaram o desempenho da coordenação motora e valores de aptidão e atividade física, associados à saúde de crianças entre 6 a 10 anos de idade da Região Autónoma dos Açores (um arquipélago da República Portuguesa).

A partir dos exemplos citados, observa-se que existem diversas razões para a utilização dos modelos multiníveis. Quando usados de forma correta, estes modelos possibilitam obter melhores estimativas para os coeficientes de regressão do que com os modelos tradicionais, além da sua grande flexibilidade no que se diz respeito à modelagem da estrutura da variância dos dados em função das variáveis explicativas, permitindo analisar dados com a variância não homogênea. Além disso, detalham melhor o comportamento da variância, a existência de variáveis correlacionadas, característica dentro de cada nível, dissimilhanças, agrupamento ou cruzamento entre níveis.

Com o intuito de ilustrar a importância do estudo dos modelos multiníveis, será apresentado a seguir um exemplo fictício retirado de Chaves (2012), onde a autora se baseada na ideia apresentada em Demidenko (2004).

Considera-se a relação entre descontos (x) e vendas (y). Seja (x_i, y_i), em que $i = 1, \ldots, N$, uma amostra obtida a partir dos valores observados dos descontos concedidos (%) e do número de vendas (em milhares de unidades). Na Figura 2 (a) têm-se o gráfico de dispersão com a reta ajustada através dos mínimos quadrados ordinários, enquanto que na Figura 2 (b) têm-se o mesmo gráfico e as retas ajustadas para cada tipo de mercadoria, evidenciando um paradoxo. Esse paradoxo ocorre por que se desconsiderar a estrutura dos dados conclui-se que as vendas tendem a diminuir com valores menores dos descontos e se tal estrutura for considerada, as vendas tendem a aumentar com a diminuição dos descontos, o que nos parece ser uma medida plausível.

Considerando o modelo de regressão clássico, temos o seguinte modelo:

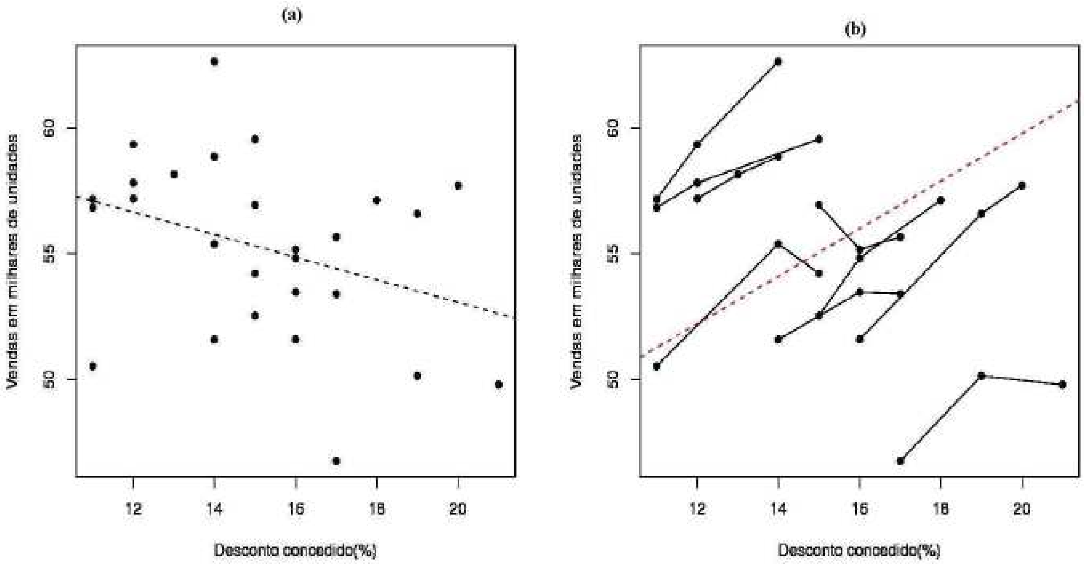
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \ldots N, \tag{1.1}$$

em que ε_i representa o erro aleatório com variância (σ^2) constante, ou seja, assume-se homocedasticidade para os erros e, consequentemente, para as mercadorias (variáveis dependentes). Em outras palavras, assume-se que os dados são coletados com "condições" similares, homogêneas. A partir da Figura 2 (b), verifica-se que os diferentes tipos de mercadoria variam muito em termos de desconto e venda, caracterizando uma condição não homogênea. Um modelo para representar tal situação é dado abaixo:

$$Y_{ij} \quad = \quad \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, n_j; j = 1, \ldots, J, \tag{1.2}$$

em que Y_{ij} representam a venda e X_{ij} o desconto concedido, da i - ésima observação da j-ésima mercadoria, β_{0j} é o intercepto da j-ésima mercadoria e $\beta_1 j$ é coeficiente de inclinação associados ao desconto concedido da i-ésima observação da j-ésima mercadoria, ε_{ij} é a fonte de variação associada à i-ésima observação da j-ésima mercadoria , n_j representa o número de observações da j-ésima mercadoria e $N = \sum_{j=1}^{J} n_j$.

Figura 2: Relação entre vendas e descontos, (a) a reta ajustada através dos mínimos quadrados ordinários, (b) as retas ajustadas para cada tipo de mercadoria



Fonte: Retirado de Chaves 2012

Assim sendo, ao desconsiderar a estrutura dos dados pode-se levar a conclusões erradas.

1.2 Custos e benefícios

Antes de começar a discutir sobre os modelos multiníveis, apresentamos alguns pontos relevantes do emprego da regressão clássica:

- Previsão para variáveis contínuas e discretas;
- Inclusão de preditores categóricos, utilizado variáveis indicadores;

- Modelagem de interações entre variáveis.

De acordo com Gelman e Hill (2007), com o uso dos modelos multiníveis é possível:

- Incluir variáveis ao nível dos indivíduos e dos grupos para estimação dos coeficientes de regressão;

- Modelar a variação entre o nível ao grupo e ao nível individual. A modelagem multinível é conveniente quando queremos modelar a variação desses coeficientes em grupos, ou em conta a variação de nível de grupo em a incerteza para os coeficientes de nível individual;

- Estimar os coeficientes de regressão para grupos específicos. Por exemplo, o problema de estimar os níveis de radão de medições em diversos municípios em Minnesota. Com um modelo multinível, podemos obter razoável estimativas, mesmo para municípios com tamanhos de amostras pequenos que seria difícil por meio de regressão clássica.

Um modelo multinível exige pressupostos adicionais além daqueles de regressão clássica, cada nível do modelo corresponde à sua própria regressão com seu próprio conjunto de pressupostos como aditividade, linearidade, independência, igualdade de variância e normalidade.

Uma alternativa bem simples para o uso da modelagem multinível é a regressão clássica quando ignora-se a variação com relação ao nível do grupo (estimação de diferentes coeficientes) ou combina-se regressões (uma única reta de regressão com indivíduos e grupos).

Existe casos limitantes em que as abordagens clássica e multinível coincidem. Isso acontece quando existe pouca variação do nível de grupo. O modelo multinível reduz-se a regressão clássica sem indicadores dos grupos e, de forma inversa, quando os coeficientes do nível grupo variam significativamente (levando em consideração quando comparado com os erros padrões estimados), a modelagem multinível reduz-se a regressão tradicional com indicadores dos grupos.

Quando se tem um número de grupo pequeno e não tem-se informação suficiente para estimar com precisão as variações com relação ao nível do grupo opta-se pelo uso dos modelos lineares clássicos.

1.3 Plataforma computacional

Essa trabalho foi desenvolvido utilizando o sistema de tipografia livre LaTeX, pois facilita o desenvolvimento da escrita e edição do texto principalmente com notações matemáticas. Para maiores detalhes sobre o sistema de tipografia usado nessa monografia basta visitar o site `http://www.latex-project.org/`.

Os resultados numéricos apresentados nessa monografia foram obtidos usando o *software* livre R, um ambiente de programação e análise de dados que pode ser obtido através do site `http://www.R-project.org`. Esse *software* dispõe de pacotes com funções especifica. Os pacotes **lme4**, **nlme** e **multilevel** foram usados para a manipulação, análise e estimação dos modelos multiníveis.

2 MODELO MULTINÍVEL

2.1 Introdução

O modelo linear hierárquico, ver por exemplo em Bryk e Raudenbush (1992), leva em consideração a estruturação de agrupamento dos dados e certamente essa estruturação reflete na especificação do modelo. Posteriormente, esse modelo aparece como modelo multinível (Goldstein, 1995). Como se sabe no modelo de regressão clássico, o intercepto e o coeficiente de inclinação são parâmetros fixos (efeitos fixos), porém no modelo multinível estes podem ser parâmetros aleatórios (efeitos aleatórios) que sofrem influência do nível hierárquico mais alto.

De acordo com McCulloch *et al.*(2008), no contexto dos modelos multiníveis, os conceitos de efeitos aleatórios e fixos são de suma importância. Segundo McCulloch *et al.*(2008) os efeitos fixos são considerados fixos quando se querer inferir somente para os níveis que foram inseridos no experimento ou considerados na avaliação e são aleatórios quando os níveis usado em um experimento ou avaliação são oriundos de uma amostra aleatória de uma população de níveis e se deseja inferir para a população inteira.

Com o objetivo de ilustrar as diferenças entre a regressão tradicional e a multinível usaremos o exemplo retirado de Valente (2007) que considera uma amostra hipotética de J escolas de uma população onde existem n_j estudantes em cada escola j. Toma-se o índice j para as escolas ($j = 1, 2, \cdots J$) e o índice i para os alunos individualmente ($i = 1, 2, \cdots, n_j$). Começa-se, então, com o modelo de mínimos quadrado ordinários, e também, por levar em consideração uma única escola.

Inicialmente considere o modelo de mínimos quadrados ordinários e que não haja distinção entre as escolas, isto é, trataremos com um único grupo de tamanho n de alunos. A abordagem é a de Regressão Linear Simples (RLS).

Seja Y_i uma variável resposta que representa a resposta para o aluno i pertencente a uma qualquer escola j, X_i uma variável explicativa associada a este aluno e ε_i um erro que provoca distorção na variável resposta. O modelo RLS é dado por:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad\qquad i = 1, 2, \cdots, n, \tag{2.1}$$

em que,

- β_0 é intercepto;
- β_1 é a alteração no valor esperado de Y_i quando X_i é alterado em uma unidade;
- $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$. [1]

Segundo Valente (2007) este é um modelo individual em que a variável resposta só depende das características individuais dada por X_i, ou seja, é ajustado um única reta para todos os grupos (escolas) e sendo suprimida qualquer diferença contextual que possa ser importante. Além disso, não será possível pelo identificar a escola a qual o aluno pertence.

Conforme Aguerre (2003) existem limitações para o uso do método dos mínimos quadrados, pois esse método ignora algumas propriedades como a correlação e a heterogeneidade. Por exemplo, quando se trabalha como os alunos, ignora-se o agrupamento das escolas devido ao suposição de ausência de autocorrelação. Já ao trabalhar ao nível das escolas, ignora-se a heterogeneidade que pode existir nos alunos de uma escola.

Suponha agora se queria observar duas escolas. Desta forma, o modelo fica:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}, \tag{2.2}$$

em que,

- Y_{ij} é a variável resposta do i-ésimo aluno para a j-ésima escola;
- X_{ij} é a variável explicativa medida no i - ésimo aluno agregrado na j-ésima escola;
- β_{0j} é o intercepto da j-ésima escola;
- β_{1j} é o coeficiente de inclinação associada à variável explicativa X_{ij} do i-ésimo aluno da j-ésima escola;

[1] i.d.d = independentes e identicamente distribuídas

- ε_{ij} é a fonte de variação associada ao i-ésimo aluno agregrado na j-ésima escola, com $\varepsilon_{ij} \sim \mathbb{N}(0, \sigma^2)$.

Diferente do modelo (2.1) agora o intercepto e o coeficiente de inclinação possuem o índice j representado a variação de um grupo para o outro (de escola para escola), ou seja, cada escola pode ter um intercepto β_{0j} e um coeficiente de inclinação β_{1j}, que serão (ou poderão ser) diferentes para cada escola.

Considerando o intercepto (β_{0j}) e o coeficiente de inclinação (β_{1j}) desconhecidos e aleatórios é necessário desenvolver um modelo para predize-los. Considerando $\beta_{0j} \sim \mathbb{N}(\gamma_{00}, \tau_{00})$, no qual τ_{00} é a variância populacional do intercepto e $\beta_{1j} \sim \mathbb{N}(\gamma_{10}, \tau_{11})$, no qual τ_{11} é a variância populacional do coeficiente de inclinação, decompõem o modelo ao nível do grupo, como pode ser visto no modelo abaixo. Os coeficientes atuam como variáveis resposta ao nível 2:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \tag{2.3}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \tag{2.4}$$

em que,

- γ_{00} é o intercepto médio para todas as escolas;
- γ_{10} é o coeficiente de inclinação médio para todas as escolas;
- u_{0j} é o erro aleatório para cada escola associado ao intercepto médio e tendo que $u_{0j} \sim N(0, \tau_{00})$;
- u_{1j} é o erro aleatório para cada escola associado ao coeficiente de inclinação médio e tendo que $u_{1j} \sim N(0, \tau_{11})$.

Segundo Valente (2007) pode-se estimar um parâmetro suplementar a $\text{Cov}(\beta_{0j}, \beta_{1j})$. Como a $\text{Cov}(u_{0j}, u_{1j}) = \tau_{01}$, então similarmente, $\text{Cov}(\beta_{0j}, \beta_{1j}) = \tau_{01}$, em que τ_{01} é a covariância populacional entre β_{0j} e β_{1j} ao nível da escola. Para os erros aleatórios u_{0j} e u_{1j} assume-se a independência com relação a ε_{ij}, ou seja, $Cov(\tau_{0j}, \epsilon_{ij}) = Cov(\tau_{1j}, \epsilon_{ij}) = 0$.

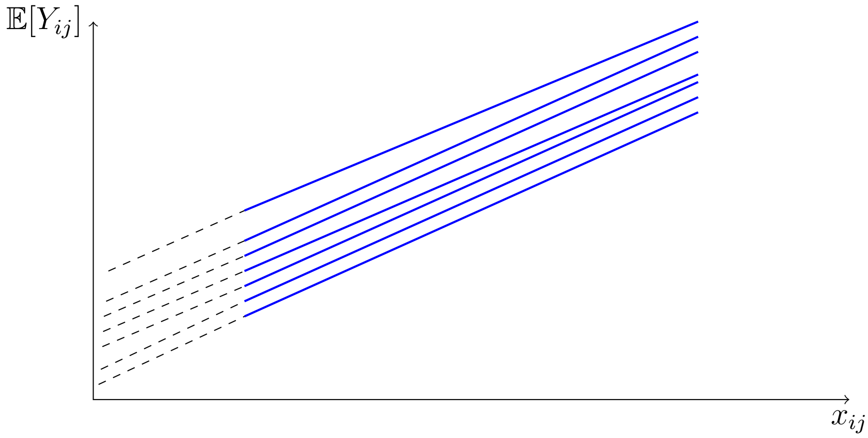
De acordo com Bergamo (2002), verifica-se, pelos modelos (2.3) e (2.4), que os coeficientes de regressão não possuem o mesmo intercepto e nem o mesmo coeficiente de inclinação, essa possível diferença pode ser esclarecida através dos efeitos aleatórios u_{0j} e u_{1j}.

Em termos práticos, pode-se ter três situações diferentes para o modelo acima: as retas de regressão para cada escola são diferentes somente no intercepto (Figura 3); somente no coeficiente de inclinação (Figura 4); diferem nos dois parâmetros, o que leva a pensar em construir um modelo que considere os distintos níveis de hierarquia (Figura 5). Em seguida estuda-se modelos em que se considera as três situações citada acima.

2.1.1 *Modelo de intercepto aleatório*

O modelo de intercepto aleatório possui retas com diferentes médias e para uma visualização melhor do comportamento desse modelo verificar na Figura 3.

Figura 3: Variação do Intercepto



Para obtenção do modelo abaixo bastar substituir o modelo (2.3) que representa o intercepto para o nível 2 , no modelo (2.2) que representa o modelo para o nível 1, onde o coeficiente constante de regressão β_{1j} será agora denotado por γ_{10}, para indicar que é um parâmetro fixo em todo o modelo. Então, se obtêm o seguinte modelo:

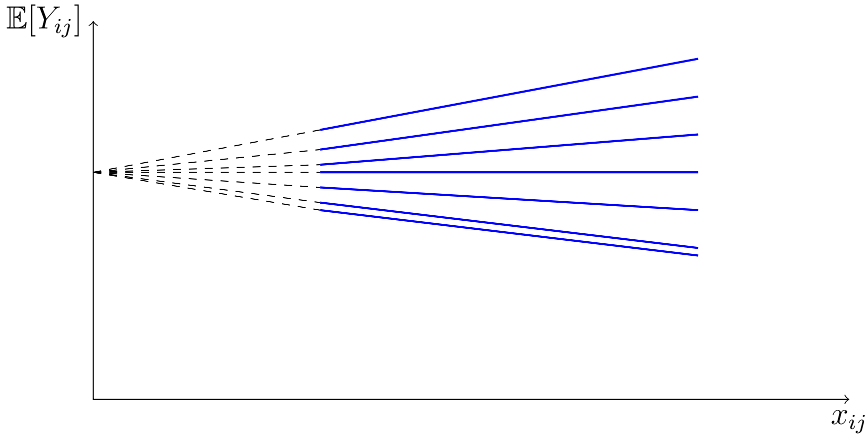
$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \underbrace{u_{0j} + \varepsilon_{ij}}_{\text{parte aleatória}}.$$

Existe quatro parâmetros para estimação nesse modelo, são eles: γ_{00} e γ_{10}, que são como os coeficientes de regressão múltipla clássica e τ_{00} e σ^2 são os parâmetros aleatórios. A parte aleatória associada ao intercepto tem variância τ_{00}, representado a variabilidade do intercepto entre escolas. O erro de nível 1 (ε_{ij}) possui variância σ^2 e representa a variabilidade intra-escola.

2.1.2 *Modelo de inclinação aleatória*

Quando as retas de regressão variam somente no coeficiente de inclinação, como pode ser observado na Figura 4, o modelo para tal situação é chamado de coeficientes de inclinações aleatórios.

Figura 4: Variando a inclinação



Essa variação no coeficiente de inclinação acontece por causa do efeito de X_{ij} que pode variar de acordo com as escolas. Portanto, as escolas que apresentarem um coeficiente de inclinação ou coeficiente de regressão diferentes terão um coeficiente β_{1j} variando aleatoriamente com relação ao nível da escola. E para obter o modelo basta substituir (2.4) em (2.2), ver modelo abaixo:

$$Y_{ij} = \beta_{0j} + \gamma_{01}X_{ij} + \underbrace{u_{1j}X_{ij} + \varepsilon_{ij}}_{\text{parte aleatória}},$$

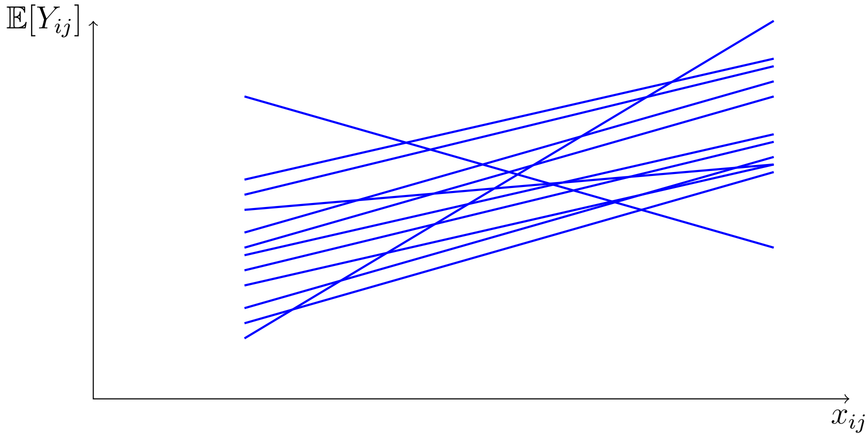
em que, o intercepto constante β_{0j} será agora denotado por γ_{00}, para indicar que é um parâmetro fixo em todo o modelo. Então, o modelo é dado por:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \underbrace{u_{1j}X_{ij} + \varepsilon_{ij}}_{\text{parte aleatória}}.$$

2.1.3 *Modelo de intercepto e inclinação aleatórios*

Quando as retas de regressão variam no intercepto e no coeficiente de inclinação, como pode ser observado na Figura 5, o modelo é chamado de modelo com intercepto e inclinação aleatórios.

Figura 5: Variando intercepto e coeficientes de inclinação



O modelo é obtido ao substituir (2.3) e (2.4) em (2.2), resultando:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \underbrace{u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}}_{\text{parte aleatória}} . \tag{2.5}$$

Este modelo possui variáveis que contém efeitos fixos e aleatórios, sendo designado por modelo misto. O modelo (2.5) comporta a principal diferença entre os modelos multiníveis e os modelos de regressão clássicos. Como pode-se observar que este modelo apresenta três variáveis aleatórias:

- ε_{ij}, uma variável aleatório no nível do aluno;
- u_{0j}, uma variável aleatória no nível da escola;
- $u_{1j}X_{ij}$ designado como uma interação aleatória entre os grupos e a variável explicativa X_{ij}.

Além da variação do intercepto e inclinação, pode-se verificar que as unidades dentro de cada grupo - no caso em estudo os alunos - podem possuir similaridade, ou seja, o efeito do grupo (turma ou escola) sobre as unidades individuais podem ser de grande importância. Assim sendo, as observações dentro de cada grupo não serão independentes e ao aplicar a regressão clássica viola-se a suposição de independência.

Como pode-se observar os alunos não estão unicamente relacionados com a sua escola, mas também possuem diferenças entre si. Logo, o objetivo não é estudar uma única variável explicativa, mas tudo que está interligado com a causa da variação. Com isso, procura-se um modelo que leve em conta toda a variabilidade existente entre as variáveis (escola) e com ela possa adicionar outras características diversas de cada uma delas e dos alunos, ou seja, um modelo que considere essa estrutura de dados de forma hierárquica. Com isso, pode-se considerar que os modelos multiníveis são uma generalização da regressão linear, em que intercepto e, possivelmente, inclinação podem variar de acordo com grupo.

2.2 Modelos multiníveis

Os modelos multiníveis são extensões dos modelos de regressões lineares clássica, no qual se elaboram várias regressões para todos os níveis em estudo. Com isso, o modelo entre os níveis estão inter-relacionados, pois os coeficientes de inclinação do nível 1 estão adicionados num nível 2 com variáveis explicativas. Antes de discutir sobre os modelos multiníveis, deve-se ter noções de três conceitos fundamentais como: correlação intraclasse, coeficiente fixo e aleatório.

A correlação intraclasse é o grau de dependência dos indivíduos, caso esse valor seja muito próximo de zero, os indivíduos dentro do mesmo grupo são distintos entre si como os que pertencem a outro grupo. Sendo assim, torna-se desnecessário o agrupamento dos dados, ou seja, as observações são independentes.

Nos modelos de regressão tradicionais, os parâmetros estimados são coeficientes fixos, isto é, comum a todos os indivíduos. Por outro lado, nos modelos multiníveis têm-se que os coeficientes aleatórios são variáveis que distribuem segundo uma função distribuição de probabilidade. Esses modelos são formados por duas partes: uma geral, comum em todos os grupos, que é conhecida como parte fixa e outra que representa a caracterização de cada grupo, que varia e sua estimação é feita através da variância nos diferentes níveis.

Segundo Ferrão (2003, *apud* Valente 2007) os modelos multiníveis acomodam de forma natural e parcimonioso a estrutura hierárquica da população em estudo, tratando o intercepto e coeficiente de inclinação como variáveis aleatórias, permitido os estudos das variações no grupo e intra grupo.

Nessa monografia será considerado modelos com dois níveis, nos quais as observações são classificadas como unidades de nível 1 e unidades de nível 2. Além disso, os modelos possuem uma variável resposta (Y) que é medida no nível individual, uma variável explicativa (X) no nível 1 e outra variável explicativa (W) no nível 2. Os modelos podem ser considerados como um sistema hierárquico de equações de regressão.

Apresentaremos o modelo multinível com dois níveis começando pela sua forma geral e a partir dessa forma geral obteremos alguns modelos particulares.

2.2.1 *Forma geral para modelo multinível com 2 níveis*

Considera-se que o modelo tenha Q variáveis explicativas X no nível 1, indicada pelo índice $q = 1, 2, \ldots, Q$ e P variáveis explicativas W no nível 2, sendo indicada por $p = 1, \ldots, P$. O modelo para o nível 1 é dada por:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \ldots + \beta_{Qj}X_{Qij} + \varepsilon_{ij}. \tag{2.6}$$

E para o nível 2 é dada por:

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_{1j} + \ldots + \gamma_{qP}W_{Pj} + u_{qj} = \gamma_{q0} + \sum_{p=1}^{P} \gamma_{qp}W_{pj} + u_{qj}, \tag{2.7}$$

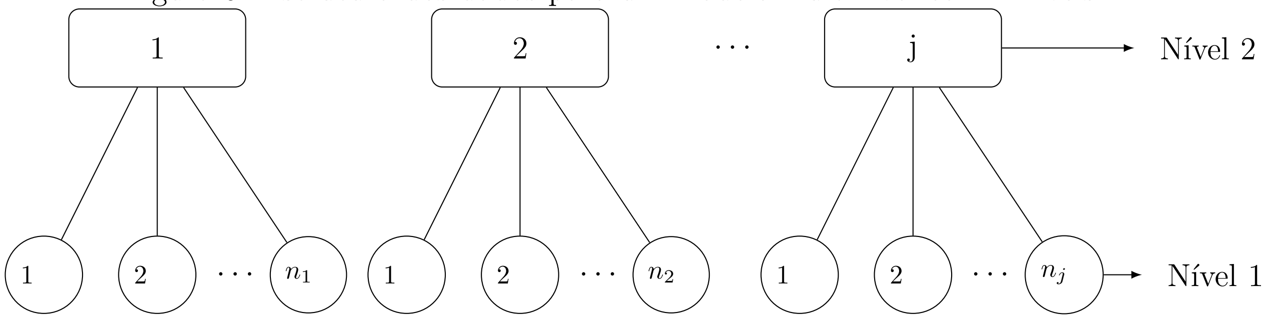
em que, $i = 1, \ldots, n_j$, $j = 1, \ldots, J$, $q = 0, 1, \ldots, Q$ e $p = 1, \ldots, P$.

Vale ressaltar que a inclusão de variáveis explicativas no nível 2, com exceção daquela que representa β_{0j}, resulta no surgimento de interações entre as variáveis dos dois níveis do modelo.

2.2.2 *Modelo multinível com 2 níveis*

No modelo multinível com dois níveis têm-se que o índice i representa o i-ésima unidade do nível 1 (dos indivíduos)($i = 1, \ldots, n_j$) da j-ésima unidade do nível 2 (do grupo) ($j = 1, \ldots, J$), ou seja, ocorrem n_j unidades do nível 1 para cada unidade j do nível 2. Na Figura 6 pode-se observar uma estrutura de dado para um modelo multinível com dois níveis.

Figura 6: Estrutura dos dados para um modelo multinível com 2 níveis



Conforme Bergamo (2002), considerando o caso de uma variável resposta Y e uma única variável explicativa do nível 1, a partir do modelo (2.6). O modelo para o nível 1 é dada por:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}, \tag{2.8}$$

com $i = 1, 2, \ldots, n_j$ e $j = 1, 2, \ldots, J$, em que

- Y_{ij} é a variável resposta do i-ésimo indivíduo do nível 1 para o j-ésimo nível 2;
- X_{ij} é a variável explicativa na sua medida original;
- β_{0j} é o intercepto para a j-ésima unidade do nível 2;
- $\beta_{0j} \sim N(\gamma_{00}, \tau_{00})$;
- τ_{00} é a variância populacional dos interceptos;
- β_{1j} é a inclinação associada à variável explicativa X_{ij} da i-ésima unidade do nível 1 para a j-ésima unidade do nível 2;
- $\beta_{1j} \sim N(\gamma_{10}, \tau_{11})$
- τ_{11} é a variância populacional das inclinações;
- ε_{ij} é o erro aleatório associado à i-ésima unidade do nível 1 associado a j-ésima unidade do nível 2;
- $\varepsilon_{ij} \sim N(0, \sigma^2)$ e $\varepsilon_{ij}'s$ são independentes.

Para o modelo do nível 2 sem variável explicativa para esse nível (P=0), considera que $q = 0, 1$ no modelo (2.7), e o modelo para esse nível fica:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j}, \quad (2.9) \\ \beta_{1j} &= \gamma_{10} + u_{1j}, \end{aligned}$$

em que,

- γ_{00} é o valor esperado dos interceptos nas unidades do nível 2;
- γ_{10} é o valor esperado dos interceptos das inclinações na população de unidades do nível 2;
- u_{0j} é o erro aleatório da j-ésima unidade do nível 2 em β_{0j};
- u_{1j} é o erro aleatório da j-ésima unidade do nível 2 em β_{1j};
- τ_{10} é a covariância entre β_{0j} e β_{1j}.

Com as seguinte suposições:

- $u_{0j} \sim N(0, \tau_{00})$ e $u'_{0j}s$ independentes;
- $u_{1j} \sim N(0, \tau_{11})$ e $u'_{1j}s$ independentes;
- $\text{cov}(u_{0j}, u_{1j}) = \tau_{01}$;
- $u'_{0j}s$ e $u'_{1j}s$ são independentes dos $\varepsilon'_{ij}s$.

Considerando $q = 0, 1$ e $P = 1$ no modelo (2.7), então tem-se a inclusão da variável explicativa W_j no modelo associado ao nível 2 para explicar a variabilidade entre as unidades desse nível. O modelo do nível 2 com essa inclusão é dada por:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} W_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} W_j + u_{0j}, \end{aligned} \tag{2.10}$$

em que,

- γ_{00} é o valor esperado do intercepto para W_j;
- γ_{10} é o valor esperado da inclinação para W_j;
- γ_{01} é o coeficiente de regressão associado à variável explicativa W_j do nível 2 relativo ao intercepto;
- γ_{11} é o coeficiente de regressão associado à variável explicativa W_j do nível 2 relativo à inclinação;

- u_{0j} é o erro aleatório da j-ésima unidade do nível 2 em β_{0j} para W_j;
- u_{1j} é o erro aleatório da j-ésima unidade do nível 2 em β_{1j} para W_j;
- τ_{00} é a variância populacional dos interceptos corrigida por W_j;
- τ_{11} é a variância populacional das inclinações corrigida por W_j.

2.2.2.1 *Centralização das variáveis em estudo*

Um aspecto muito importante é a centralização da variável explicativa, com relação ao nível 1 e 2, pois a centralização torna adequada a interpretação do intercepto da regressão. De acordo com Natis (2000) ao utilizar a variável explicativa X_{ij} em sua escala original pode-se ter um problema na interpretação do intercepto β_{0j}, pois se X_{ij} fosse altura, não faz sentindo interpretar β_{0j} como o valor esperado para um indivíduo do grupo j com altura igual a zero.

A centralização de variáveis é importante, pois a interpretação dos parâmetros principalmente de β_{0j} do modelo depende como as variáveis explicativas do nível 1 são consideradas. Conforme Nezlek (2001) a centralização nos modelos MM's pode ser tratada com três formas: não haver centralização, centralização na grande média e centralização na média do grupo. Como pode ser vista na Tabela 1:

Tabela 1: Centralização da variável em estudo

Condição de Centralização	Mudança em β_{0j}	Interpretação
$\bar{\mathrm{X}}_{ij}$ na escala original	$\beta_{0j} = \mathbb{E}(Y_{ij} \mid X_{ij} = 0)$	β_{0j} é o valor esperado da variável resposta Y_{ij} quando X_{ij} for igual a zero.
($\bar{\mathrm{X}}_{ij}$ - $\bar{\mathrm{X}}_{..}$) centralizada na sua média geral	$\beta_{0j} = \mu_{Y_j} - \beta_{1j}$ (X_{ij} - $\bar{\mathrm{X}}_{..}$)	β_{0j} é interpretado com a média do j-ésimo grupo do nível 2 ajustado para a variável X.
(X_{ij} - $\bar{\mathrm{X}}_{.j}$) centralizada na média do grupo	$\beta_{0j} = \mu_{Y_j}$	β_{0j} é interpretado como a média não ajustado da variável resposta Y_{ij}.

Fonte: Elaborada pela autora

De acordo com Bryk e Raudenbush (2002) não existe nenhuma regra simples para representar todos os casos, portanto a decisão para a centralização é baseada na estrutura dos dados e hipóteses de interesse.

2.2.3 *Submodelos multiníveis*

Os submodelos podem ser obtidos a partir do modelo multinível com 2 níveis representados pelos modelos (2.9) e (2.10) com a anulação de alguns termos. Nesta seção será apresentados alguns submodelos, entre eles o Anova [2] com 1 fator e efeito aleatório, Regressão de médias como respostas, Análise de Covariância (ANCOVA) com 1 Fator e Efeitos Aleatórios, Regressão com Coeficientes Aleatórios, Interceptos e inclinações como respostas e Modelo com Inclinações Variando Não Aleatoriamente.

2.2.4 *Anova com 1 fator e efeito aleatório*

De acordo com Bryk e Raudenbush (2002), esse modelo é o mais simples, pois não possui variáveis explicativas em nenhum dos dois níveis, além disso, a inclinação β_{1j} no modelo (2.8) do nível 1 é nula, para todo j. Com isso, os modelos do nível 1 e nível 2 são dados, respectivamente, por:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}. \tag{2.11}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}, \tag{2.12}$$

em que,

- γ_{00} é a média da variável resposta para a população;
- u_{0j} é o efeito aleatório associado ao j-ésimo grupo e com $u_{0j} \overset{i.i.d}{\sim} N(0, \tau_{00})$, em que $\text{Cov}(u_{0j}, \varepsilon_{ij}) = 0$.

Para se obter o modelo combinado basta substituir (2.12) em (2.11):

$$Y_{ij} = \underbrace{\gamma_{00}}_{\text{parte fixa}} + \underbrace{u_{0j} + \varepsilon_{ij}.}_{\text{parte aleatória}} \tag{2.13}$$

A variância de Y_{ij} é dada por:

$$\text{Var}(Y_{ij}) = \tau_{00} + \sigma^2. \tag{2.14}$$

[2] Análise de Variância

Nota-se que a variância de Y_{ij} decompõe-se em duas partes independentes: a variância dos erros do nível 1 (dos indivíduos) denotada por ε_{ij} e a variância dos erros do nível 2 (do grupo) representada por τ_{00}, definidos por u_{0j}.

O coeficiente de correlação é um parâmetro de grande importância, pois representa a proporção da variância da resposta explicada pela variabilidade entre as unidades do nível 2. Esse coeficiente é dado por:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \ . \tag{2.15}$$

Os valores para esse coeficiente variam no intervalo [0, 1], $\rho = 0$ significa que toda a variância em Y deve-se ao nível 1, não sofrendo qualquer influência do nível 2. Para o caso em que $\rho = 1$, toda a variabilidade em Y é devida as diferenças no nível 2, não sofrendo qualquer influência do nível 1.

2.2.5 *Regressão de médias como respostas*

Segundo Ramos (2009), nestes modelos são adicionados variáveis explicativas no nível 2, buscando explicar a variabilidade dos coeficiente β_{0j} entre as unidades do nível 2. O modelo de média em relação ao nível 1 é igual ao caso para Anova com 1 fator e efeito aleatórios, ou seja, o modelo para o nível 1 está definido em (2.11) e nível 2 é dado por:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}, \tag{2.16}$$

em que, β_{0j} é o valor esperado da variável resposta de um modelo de regressão linear, no qual as variáveis explicativas correspondem as características do grupo j. No nível 2, tem-se a variável explicativa W.

O modelo combinado é obtido ao substituir (2.16) em (2.11):

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{01} W_j}_{\text{parte fixa}} + \underbrace{\varepsilon_{ij}\ .}_{\text{parte aleatória}} \tag{2.17}$$

No qual,

- γ_{00} é o intercepto médio dos grupos para quando W_j for igual a zero;
- γ_{01} é a diferença média entre os J grupos;
- u_{0j} é o efeito aleatório do j-ésimo grupo sobre o intercepto para $W_j = 0$

O coeficiente intra-classe continua sendo calculado da mesma forma e com a mesma interpretação do modelo Anova com 1 fator e efeito aleatórios, porém agora é chamado de coeficiente intra-classe condicionado, pois está corrigido pela variável W_j.

2.2.6 *ANCOVA com 1 fator e efeitos aleatórios*

Esse é um modelo no qual considera-se que as inclinações não variam e não são afetadas pelas características do grupo. Com as seguintes restrições:

- $\gamma_{01} = 0$ e $\gamma_{11} = 0$ para o coeficiente do nível 2;
- $u_{1j} = 0$.

Então, para esse modelo, nos níveis 1 e 2, respectivamente, temos:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}. \tag{2.18}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{2.19}$$

$$\beta_{1j} = \gamma_{10}. \tag{2.20}$$

O modelo combinado é dado por:

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10} X_{ij}}_{\text{parte fixa}} + \underbrace{\varepsilon_{ij} + u_{0j}}_{\text{parte aleatória}}. \tag{2.21}$$

O modelo apresenta somente uma variável explicativa que está associada ao coeficiente γ_{10} que é considerado constante e sem nenhum efeito aleatório, para cada unidade de j do nível 2, ou seja, esse coeficiente é igual para o grupo.

2.2.7 *Regressão com coeficientes aleatórios*

Neste modelo consideram β_{0j} e β_{1j} como coeficientes aleatórios, os interceptos e inclinações variando aleatoriamente. O modelo para o nível 1 e para nível 2, respectivamente, é dado por:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}. \tag{2.22}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}, \tag{2.23}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \tag{2.24}$$

em que,

- β_{0j} é o intercepto para j-ésima unidade do nível 2 e representa o valor esperado da variável resposta Y_{ij} quando $X_{ij} = 0$;
- β_{1j} é o coeficiente de inclinação associado a variável explicativa X_{ij} da i-ésima unidade do primeiro nível para a j-ésima unidade do segundo nível;
- γ_{00} é o valor esperado dos interceptos dos J grupos;
- γ_{10} é o valor esperado das inclinações dos J grupos;
- u_{0j} é o efeito aleatório da j-ésima unidade do segundo nível em β_{0j};
- u_{1j} é o efeito aleatório da j-ésima unidade do segundo nível em β_{0j};
- $u_{0j} \overset{iid}{\sim} N(0, \tau_{00})$ e $u_{1j} \overset{iid}{\sim} N(0, \tau_{11})$;
- $\text{Cov}(u_{0j}, \varepsilon_{ij}) = 0$ e $\text{Cov}(u_{1j}, \varepsilon_{ij}) = 0$.

Os efeitos u_{0j} e u_{1j} possuem uma distribuição normal multivariada com uma matriz de variância e covariância dados por:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_p(0, \Omega_u) \tag{2.25}$$

$$\Omega_u = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}, \tag{2.26}$$

em que,

- $\text{Var}(u_{0j}) = \tau_{00}$;
- $\text{Var}(u_{1j}) = \tau_{11}$;
- $\text{Cov}(u_{0j}, u_{1j}) = \tau_{01}$.

Os coeficientes τ_{00}, τ_{11} e τ_{01} são variâncias e covariâncias não condicionais, pois o modelo não possui um preditor no segundo nível.

Para obtenção do modelo combinado basta substituir (2.23) e (2.24) em (2.22) :

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}X_{ij}}_{\text{parte fixa}} + \underbrace{u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}}_{\text{parte aleatória}} . \tag{2.27}$$

O modelo acima é formado por uma parte fixa $\gamma_{00} + \gamma_{10}X_{ij}$ como função de regressão e outra parte aleatória $u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}$, em que o componente u_{0j} é o efeito com relação ao j-ésimo grupo sobre a média, $u_{1j}X_{ij}$ é o efeito aleatório do j-ésimo grupo sobre β_{1j} e ε_{ij} é o erro aleatório do nível 1.

2.2.8 *Interceptos e inclinações como respostas*

Neste modelo é adicionada a variável (W_j) no modelo do nível 2, pois esse variável ajuda a explicar a variabilidade tanto dos interceptos com das inclinações. Para a obtenção desse modelo, basta adicionar a variável W_j nos modelos (2.23) e (2.24):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \tag{2.28}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}, \tag{2.29}$$

em que,

- γ_{00} é o valor esperado do intercepto quando $W_j = 0$;
- γ_{01} é o coeficiente de regressão associado a variável explicativa W_j do nível 2 com relação ao intercepto;
- γ_{11} é o coeficiente de regressão associado a variável W_j do nível 2 relacionado à inclinação;
- u_{0j} é o efeito aleatório da j-ésima unidade do nível 2 sobre o intercepto quando $W_j = 0$;
- u_{1j} é o efeito aleatório da j-ésima unidade do nível 2 sobre a inclinação para $W_j = 0$;

Com isso, pode-se escrever os efeitos aleatório da seguinte forma:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_p(0, \Omega_u), \tag{2.30}$$

$$\Omega_u = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}, \tag{2.31}$$

em que,

- $\text{Var}(u_{0j}) = \tau_{00}$ é a variância populacional dos interceptos corrigida por W_j;
- $\text{Var}(u_{1j}) = \tau_{11}$ é a variância populacional das inclinações corrigida por W_j;
- $\text{Cov}(u_{0j}, u_{1j}) = \tau_{01} = \text{Cov}(\beta_{0j}, \beta_{1j})$;
- $\text{Cov}(u_{0j}, \varepsilon_{ij}) = 0$ e $\text{Cov}(u_{1j}, \varepsilon_{ij}) = 0$.

Para este modelo τ_{00}, τ_{11} e τ_{01} são componentes de variância e covariância residuais em β_{0j} e β_{1j} quando W_j é excluída.

O modelo combinado é dado por:

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}X_{ij}W_j}_{\text{parte fixa}} + \underbrace{u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}\ .}_{\text{parte aleatória}} \tag{2.32}$$

Verifica-se que o modelo envolve tanto a variável X_{ij} do nível 1 com a W_j do nível 2, sendo uma parte fixa e outra aleatória, além disso $X_{ij}W_j$ é chamado interação de nível cruzado com a função de permite que a inclinação média de X_{ij} varie com W_j. De acordo com Ramos (2009), as variáveis explicativas X_{ij} e W_j dos níveis 1 e 2, respectivamente, podem ser consideradas centralizadas na média amostral global. Essa centralização das variáveis explicativas na média amostral global ocorre quando os interceptos β_{0j} não podem receber o valor zero.

2.2.9 *Modelo com inclinações variando não aleatoriamente*

"O modelo com inclinações variando não aleatoriamente é obtido quando a variância residual (u_{1j}) da inclinação β_{1j} é bem próxima de zero. Em outras palavras, após o controle de W_j, pouca ou quase nenhuma variância da inclinação deverá restar para explicar o modelo"(Bryk e Raudenbush, 1992). Com isso, pode-se considerar que $\tau_{11} = 0$. Assim, o modelo para os níveis 1 e 2, respectivamente, é dado por:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}. \tag{2.33}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{10}W_j + u_{0j}. \tag{2.34}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j. \tag{2.35}$$

Substituindo o modelo (2.34) e (2.35) em (2.33), encontra-se o modelo combinado:

$$Y_{ij} = \underbrace{\gamma_{00} + \gamma_{10} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij}}_{\text{parte fixa}} + \underbrace{u_{0j} + \varepsilon_{ij}}_{\text{parte aleatória}}. \tag{2.36}$$

Verifica-se neste modelo que as inclinações sofrem uma variação de grupo para grupo, mas não de forma aleatória. Isto pode ser visualizada pelo modelo (2.33), onde β_{1j} varia somente em função de W_j.

3 INFERÊNCIA

3.1 Notação matricial

A notação matricial será considerada nessa seção, pois a mesma facilita a manipulação para estimação dos parâmetros. De acordo com Natis (2000) e Ramos(2009), o modelo geral do nível 1 com Q variáveis é dado por:

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \qquad j = 1, \ldots J, \tag{3.1}$$

em que, $\boldsymbol{Y}_j$ é o vetor resposta do nível 1, $\boldsymbol{X}_j$ é a matriz das variáveis explicativas do nível 1, $\boldsymbol{\beta}_j$ é o vetor dos coeficientes do nível 1 para o grupo j e $\boldsymbol{\varepsilon}_j$ é o vetor de erro aleatório. Assim, sua notação matricial é representada por:

$$\boldsymbol{Y}_j = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{n_j j} \end{bmatrix}, \boldsymbol{X}_j = \begin{bmatrix} 1 & X_{11j} & \ldots & X_{Q1j} \\ 1 & X_{12j} & \ldots & X_{Q2j} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & X_{1n_j j} & \ldots & X_{Qn_j j} \end{bmatrix}, \boldsymbol{\beta}_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{Qj} \end{bmatrix}, \boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j j} \end{bmatrix}. \tag{3.2}$$

Com relação ao nível 2, a sua representação matricial é dada por:

$$\boldsymbol{\beta}_j = \boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{u}_j, \tag{3.3}$$

em que, $\boldsymbol{\gamma}$ representa o vetor de parâmetros fixos com $\boldsymbol{\gamma'}_q = [\gamma_{q0} \;\; \gamma_{q1} \;\; \ldots \;\; \gamma_{qP}]$, em que $q = 0, 1, \ldots Q$, $\boldsymbol{W}_j$ é a matriz de variáveis explicativas do nível 2 com $\boldsymbol{W}_{qj} = [1 \;\; W_{1j} \;\; W_{2j} \;\; \ldots \;\; W_{Pj}]$ e $\boldsymbol{u}_j$ é o vetor de parâmetros aleatórios. Sendo assim, a sua representação matricial é dada por:

$$\boldsymbol{W}_j = \begin{bmatrix} \boldsymbol{W}_{0j} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{2j} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{W}_{Qj} \end{bmatrix} \boldsymbol{\gamma} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_Q \end{bmatrix} \boldsymbol{u}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{Qj} \end{bmatrix}. \tag{3.4}$$

Dessa forma, o modelo combinado na forma matricial é dada por:

$$\boldsymbol{Y}_j = \boldsymbol{X}_j \boldsymbol{W} j \boldsymbol{\gamma} + \boldsymbol{X}_j \boldsymbol{u}_j + \boldsymbol{\varepsilon}_j. \tag{3.5}$$

Considerando agora $\boldsymbol{A}_j = \boldsymbol{X}_j \boldsymbol{W} j$ e substituindo no modelo (3.5) têm-se a fórmula abaixo:

$$\boldsymbol{Y}_j = \boldsymbol{A}_j \boldsymbol{\gamma} + \boldsymbol{X}_j \boldsymbol{u}_j + \boldsymbol{\varepsilon}_j. \tag{3.6}$$

Agora considerando todas as J unidades do nível 2 para reescrever o modelo (3.6) temos:

$$\boldsymbol{Y} = \boldsymbol{A} \boldsymbol{\gamma} + \boldsymbol{X} \boldsymbol{u} + \boldsymbol{\varepsilon}. \tag{3.7}$$

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_J \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{A}_J \end{bmatrix}, \boldsymbol{\gamma} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_J \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{X}_J \end{bmatrix},$$

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_0 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_J \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_J \end{bmatrix}.$$

Para o modelo (3.6) assume-se normalidade, pois há necessidade para os testes estatísticos.

$$\boldsymbol{\varepsilon}_j \sim N(0, \boldsymbol{R}) \qquad \boldsymbol{R} = \sigma^2 \boldsymbol{I}_{n_j}, \tag{3.8}$$

$$\boldsymbol{u}_j \sim N(0, \boldsymbol{G}), \tag{3.9}$$

em que, $\boldsymbol{I}_{n_j}$ é uma matriz identidade com dimensão n_j e $\boldsymbol{G}$ é a matriz de variância e covariância.

3.2 Método de estimação

De acordo com Bryk e Raudenbush (2002) existem três parâmetros a serem estimados em um modelo multinível com 2 níveis, são eles: efeito fixo, coeficiente aleatório do nível 1 e componentes de variância e covariância. Diversas técnicas de estimação são utilizadas no modelo multinível, pois o mesmo possui diversos tipos de parâmetros como citado acima. Os coeficientes β_j no nível 1 pode ser fixos ou aleatórios, os coeficientes γ são fixos no nível 2 e as variâncias e covariância entre os níveis 1 e 2 são chamados de componentes de variância e covariância. Em geral, são necessários processos interativos para a estimação desses parâmetros.

Atualmente, algoritmos têm sido implementados em *softwares* para a estimação e o ajuste dos modelos multiníveis. Os Mínimos Quadrados Iterativos Generalizados (MQIG) e os Mínimos Quadrados Interativos Generalizados Restritos (MQIGR) são exemplos desses algoritmos.

Na seção a seguir, serão abordados tópicos referentes à estimação nos modelos multiníveis. Para maiores detalhes teóricos sugerimos Littell et al (2006), Demidenko (2004) e McCulloch (2008). Para a estimação dos parâmetros será considerado, neste trabalho, o modelo (3.7) e os métodos dos Mínimos Quadrados Ponderados (MQP) ou Mínimos Quadrados Generalizados (MQG).

3.2.1 *Estimação dos efeitos fixos* ($\boldsymbol{\gamma}$)

Os mínimos quadrados ponderados (MQP) ou MQG são utilizados para estimar $\boldsymbol{\gamma}$, como é visto abaixo:

$$\hat{\boldsymbol{\gamma}} = (\boldsymbol{A}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{Y}, \tag{3.10}$$

em que,

$$\boldsymbol{V} = \text{var}(\boldsymbol{Y}) = \boldsymbol{XGX}' + \boldsymbol{R}. \tag{3.11}$$

Com **A** representando uma matriz com dimensão $N \times J$ com $N = \sum_{j=1}^{J} n_j$, $\hat{\boldsymbol{V}}$ é a estimativa para $\boldsymbol{V}$ com $\boldsymbol{G}$ e $\boldsymbol{R}$, sendo $\boldsymbol{R}$ uma matriz com dimensão N x N, onde $\boldsymbol{G}$ e $\boldsymbol{R}$ serão substituído pelos seus estimadores de máxima verossimilhança (MV). Os elementos das matrizes $\boldsymbol{G}$ e $\boldsymbol{R}$ são chamados de componente de variância e são estimados pela máxima verossimilhança restrita (MVR)[1] ou pela máxima verossimilhança (MV), como será mostrado na seção seguinte. A variância de estimada $\hat{\boldsymbol{\gamma}}$ é dada por:

$$\widehat{var(\hat{\boldsymbol{\gamma}})} = (\boldsymbol{A}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{A})^{-1}. \tag{3.12}$$

3.2.2 *Estimação dos componentes de variâncias (R e G)*

Caso o tamanho amostral n_j seja iguais para todos os níveis, existe uma fórmula de forma fechada para estimar os parâmetros de variância e covariância. Quando o tamanho amostral n_j é diferente usa-se métodos numéricos iterativos para obter as estimativas. Geralmente, esses métodos são baseados em técnicas de estimação por MV. Os estimadores de MV de $\boldsymbol{R}$ e $\boldsymbol{G}$ são encontrados maximizando a função de log-verossimilhança dada por:

$$l_{\text{MV}}(\boldsymbol{GR}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{N}{2}\log(\boldsymbol{r}'\boldsymbol{V}^{-1}\boldsymbol{r}) - \frac{N}{2}\left[1 + \log\frac{2\pi}{N}\right], \tag{3.13}$$

em que,

$$\boldsymbol{r} = \boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A})'\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{Y}. \tag{3.14}$$

Se o número de unidades do nível 2 (J) é grande, então, os estimadores gerados pela MV são aproximadamente iguais aos estimadores gerados pela MVR. Os estimadores MVR dos componentes de variância e covariância são baseados nos resíduos, que são obtidos após a estimação dos efeitos fixos pelo MQP ou pelo MQG. Os estimadores de MVR levam em consideração o número de graus de liberdade usado nas estimativas dos efeitos fixos quando se estima os componentes de variância e covariância. Os estimadores

[1]: Patterson e Thompson (1971) apresentaram uma modificação ao método de MV que passou a ser denominado MVR por Harville(1997). O método MVR mantém as demais propriedades do MV. O MVR é equivalente a ML para um conjunto de dados que tenham sido padronizados. Além disso, o MVR é iterativo e, portanto, demanda grandes recursos computacionais.

MVR de $\boldsymbol{G}$ e $\boldsymbol{R}$ são encontrados maximizando a função de log-verossimilhança seguinte:

$$l_{\mathrm{MV}}(\boldsymbol{GR}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}\log(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A}) - \frac{(N-p)}{2}\log(\boldsymbol{r}'\boldsymbol{V}^{-1}\boldsymbol{r}) - \frac{(N-p)}{2}\left[1 + \log\frac{2\pi}{(N-p)}\right], \tag{3.15}$$

em que $\boldsymbol{r} = \boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A})'\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{Y}$ e $p = rank(\boldsymbol{A})$.

3.2.3 *Predição dos efeitos aleatórios* ($\boldsymbol{u}$)

Os estimadores dos efeitos aleatórios podem ser obtidos quando se substitui (3.10) na função obtida quando deriva-se l_{MV}, com relação a $\boldsymbol{\gamma}$ e $\boldsymbol{u}_j$. Deste modo, tem-se:

$$\hat{\boldsymbol{u}_j} = \hat{\boldsymbol{G}}\boldsymbol{X'}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{A}\hat{\boldsymbol{\gamma}}). \tag{3.16}$$

3.3 Teste de hipóteses

Os teste de hipóteses, uniparamétricos e multiparamétricos, para os efeitos fixos e os componentes de variância e covariância de um modelo multinível serão apresentados nessa seção.

3.3.1 *Teste de hipótese para efeito fixo*

Neste teste, o interesse é investigar se os efeitos fixos estimados são significativamente diferentes de zero. A hipótese de interesse para testar um único efeito fixo é dada por:

$$H_0 : \gamma_{qp} = 0.$$

A estatística de teste é calculada tomando a razão entre o estimador MV ou MRV e o seu erro-padrão estimado e é dada por:

$$t = \frac{\hat{\gamma_{pq}}}{\sqrt{\widehat{Var}(\hat{\gamma_{pq}})}},$$

em que $\hat{\gamma_{pq}}$ é o estimador do efeito fixo (γ_{pq})e $\sqrt{\widehat{Var}(\hat{\gamma_{pq}})}$ é a variância estimada de $\hat{\gamma_{pq}}$.

De acordo com Ramos(2009), a estatística de teste possui uma distribuição t-Student com $J - P - 1$ graus de liberdade para dados balanceados e para algumas situações de dados não balanceados.

No caso que se tenha mais de um efeito fixo simultaneamente, se faz necessário o uso de contraste e o teste é realizado por meio da estatística de Wald. A notação usada nessa seção será baseada em Ramos (2009).

Suponha que o vetor de efeitos fixos é dada por:

$$\boldsymbol{\gamma} = (\gamma_{11} \quad \gamma_{12} \quad \gamma_{21} \quad \gamma_{22})'.$$

Então, por exemplo, deseja-se testar a seguinte hipótese:

$$H_0 : \gamma_{11} = \gamma_{22} = 0.$$

Reescrevendo a hipótese na forma matricial têm-se que:

$$H_0 : \boldsymbol{C}'\boldsymbol{\gamma} = 0,$$

em que $\boldsymbol{C}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$

Sendo $\hat{\boldsymbol{\gamma}}$ o estimador do vetor de efeitos fixos $\boldsymbol{\gamma}$ e $\hat{\boldsymbol{V}}_{\boldsymbol{\gamma}}$ a variância estimada de $\hat{\boldsymbol{\gamma}}$. Assim sendo, a variância é dada por:

$$\widehat{Var}(\boldsymbol{C}'\hat{\boldsymbol{\gamma}}) = \boldsymbol{C}'\hat{\boldsymbol{V}}\boldsymbol{C} = \hat{\boldsymbol{V}}_{\boldsymbol{C}},$$

e sob H_0 a estatística calculada para o teste de Wald é dado por:

$$H = \hat{\boldsymbol{\gamma}}'\boldsymbol{C}\hat{\boldsymbol{V}}^{-1}\boldsymbol{C}'\hat{\boldsymbol{\gamma}},$$

em que, a distribuição assintótica para essa estatística é uma distribuição χ^2 (qui-quadrado) com número de graus de liberdade igual ao número de linha de $\boldsymbol{C}$.

3.3.2 *Teste de hipótese para componente de variância e covariância*

Segundo Ramos(2009) seja τ_{qq} a variância dos coeficientes u_{qj} e para verificar se esse coeficiente é fixo ou aleatório, pode-se testar a hipótese nula da seguinte maneira:

$$H_0 : \tau_{qq} = 0,$$

em que τ_{qq} é um elemento de $\boldsymbol{G}$. Para o caso da hipótese nula for rejeitada conclui-se que há variação aleatória nos β_{qj}. A estatística para testar a hipótese acima é dada por:

$$z = \frac{\widehat{\tau_{qq}}}{\sqrt{\widehat{Var}(\widehat{\tau_{qq}})}}. \tag{3.17}$$

Para grande amostras têm-se que a estatística (3.17) possui uma distribuição aproximadamente normal e a estimativa de $\widehat{\tau_{qq}}$ é obtida através do inverso da matriz de informação de Fisher.

Uma outra alternativa para testar a hipótese é usando as estimativas obtidas através do método dos Mínimos Quadrados Ordinários(MQO), ou seja, o mesmo método usado para a estimação dos modelos lineares clássicos. Porém, esse teste só é possível de ser usado caso todos ou pelo menos a maioria dos grupos possuírem dados suficientes para calcular as estimativas de MQO. Dessa maneira, considera-se o seguinte modelo:

$$\beta_{qj} = \gamma_{q0} + \sum_{p=1}^{P} \gamma_{qp} W_{pj}. \tag{3.18}$$

Considerando o modelo em (3.18) a sua estatística de teste é dada por:

$$E = \sum_{j=1}^{J} \frac{(\widehat{\beta_{qj}} - \widehat{\gamma_{q0}} - \sum_{p=1}^{P} \widehat{\gamma_{q0}} W_{pj})^2}{\hat{V}_{qqj}}. \tag{3.19}$$

Essa estatística tem aproximadamente uma distribuição χ^2 com $J - P - 1$ graus de liberdade. Para o caso que se queria testar mais de um componente de variância e covariância simultaneamente, pode-se utilizar o teste de razão de verossimilhanças. Com isso, na hipótese nula têm-se que todos esses componentes são nulos.

$$H : D_0 - D_1, \tag{3.20}$$

em que D_0 e D_1 são as "Deviance" relacionadas a cada um dos modelos testados. O D_0 e D_1 são calculados da seguinte forma, respectivamente:

$$D_0 = -2log(L_0),$$

$$D_1 = -2log(L_1),$$

em que L_0 é o valor da verossimilhança associada com a estimativa de MV do modelo completo e L_1 é o valor da verossimilhança associada com a estimativa de MV do modelo alternativo.

A estatística de teste (3.20), sob H_0, possui distribuição assintótica χ^2 com m graus de liberdade, em que m é a diferença entre os graus de liberdade dos dois modelos.

3.4 Seleção de modelo

A escolha do modelo nem sempre é fácil, pois leva-se em consideração que o modelo deve se adaptar da melhor maneira a descrição dos dados em estudo. Com isso, existem critérios para fazer tal seleção, entre esses critérios têm-se o critério de informação de Akaike (AIC), o *deviance* e o coeficiente de determinação (R^2).

3.4.1 AIC

Esse critério é usado para comparar os modelos com os mesmos efeitos fixos, mas com estruturas de covariância diferentes. O AIC para modelo multinível é dada pela seguinte fórmula:

$$AIC = D + 2q, \tag{3.21}$$

em que D é o deviance e q é o número de parâmetros estimados.

Assim, o modelo que apresentar o menor valor do AIC pode ser considerado o mais apropriado para representar o fenômeno em estudo. Além disso, AIC é um índice de qualidade de ajustamento que compara os modelos que estão sendo ajustado para um mesmo conjunto de dados, utilizando o mesmo método de estimação.

3.4.2 *Deviance*

Na análise de dados deve-se considerar que quanto maior for o número de variáveis explicativas em estudo incluída no modelo, menor será o valor do *deviance*, ou seja, melhor será o modelo. Porém, o modelo deve ser parcimonioso em relação ao número de variáveis explicativas. Desta forma, faz-se necessário saber se as variáveis ou os parâmetros que são acrescentada ao modelo contribuem de forma significativa para o melhoramento do ajuste ou não.

3.4.3 *Coeficiente de determinação*

O coeficiente de determinação, denotado por R^2, é a variabilidade total da variável resposta que é explicada pelas variáveis explicativas no modelo. Nos modelos de regressão clássicos, ele mede a capacidade explicativa do modelo e sua estatística é dada por:

$$R^2 = \frac{\text{var}(Y) - \text{var}\left(Y - \sum_{j=1}^{n} \beta_j X_j\right)}{\text{var}(Y)}, \tag{3.22}$$

em que, $\text{var}\left(Y - \sum_{j=1}^{n} \beta_j X_j\right)$ representa a variância residual do modelo com variáveis explicativas e $\text{var}(Y)$, a variância do modelo nulo, ou seja, modelo sem variáveis explicativas.

Segundo (Ferrão, 2003, p.56 apud Valente, 2007, p. 79), nos modelos multiníveis, nos quais apenas o intercepto tem componente aleatório, a variância total pode ser decomposta em duas parcelas: em variância de nível 1 (σ_ε^2) e variância de nível 2 (σ_{u0}^2). Com isso, calcula-se o coeficiente de determinação, que é representado por R_1^2 e R_2^2. Assim, têm-se a fórmula para calcular R_1^2 e R_2^2 é dada por, respectivamente:

$$R_1^2 = \left(\frac{\sigma_{\varepsilon|b}^2 - \sigma_{\varepsilon|m}^2}{\sigma_{\varepsilon|b}^2}\right), \tag{3.23}$$

$$R_2^2 = \left(\frac{\sigma_{u0|b}^2 - \sigma_{u0|m}^2}{\sigma_{u0|b}^2}\right), \tag{3.24}$$

em que, $\sigma_{\varepsilon|b}^2$ é a variância residual do nível 1 para o modelo nulo e $\sigma_{\varepsilon|m}^2$ é a variância residual do nível 1 para modelo de comparação.

em que, $\sigma_{u0|b}^2$ é a variância residual do nível 2 para o modelo nulo e $\sigma_{u0|m}^2$ é a variância residual do nível 2 para modelo de comparação.

Como pode-se verificar, o R^2 é adaptado para modelos multiníveis, pois o mesmo representa um indicador aproximado do percentual de variância explicada em cada nível do modelo. Geralmente, o R^2 é usado com uma estatística reserva para a seleção do modelo ou tomada de decisão do modelo final.

4 APLICAÇÃO

Neste capítulo são apresentadas as análises dos microdados do SAEB (Sistema Nacional de Educação Brasileira). O banco de dados está disponível no site do Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP). Vale ressaltar que o SAEB foi criado em 1988 pelo Ministério da Educação e Cultura (MEC) com a finalidade de avaliar o sistema educacional do Brasil. Encontra-se no site do INEP as edições do SAEB para os anos de 90, 93, 95, 97, 99, 01, 03, 05 e 11.

Para melhor exemplificar a técnica de MM será utilizado apenas os microdados da edição de 2011 do SAEB. Além disso, foram considerados somente os alunos que estudaram em escolas da capital cearense. Além do mais, os selecionados preencheram o questionário socioeconômico (que também está disponível no site do INEP) e obtiveram nota na prova de proficiência em matemática. Vale enfatizar que apenas algumas variáveis foram utilizadas para a análise.

4.1 Banco de dados

O conjunto de dados contém uma amostra de 20491 alunos distribuídos entre 267 escolas da capital cearense. Para a utilização do modelo multinível considerou-se dois níveis, sendo o nível 1 será associado aos alunos e o nível 2 para as escolas. Além do mais, a proficiência em matemática foi considerada a variável resposta e como variáveis explicativas: sexo, idade, raça, tipo de rede de ensino e participação em atividade continuada feita pelo diretor da escola. Vale ressaltar que as variáveis em estudo fazem parte do questionário socioeconômico dos alunos e diretores. Na Tabela 2 encontram-se a codificação e a descrição com relação ao banco de dados de cada variável que será usada no modelo.

Tabela 2: Descrição e codificação das variáveis em estudo

Variável	Descrição	Codificação
Sexo	Sexo do Aluno	0-Masculino 1 -Feminino
Raca	Como o aluno se considera	0 -Branco 1 -Para as demais
Idade	Idade do aluno	8 a 15 anos (Idade centralizada)
Rede	Tipo de Rede de Ensino	0 -Pública 1 -Privada
AtivPartic	Participou atividade de formação continuada	0 - Não 1 -Sim
Profic	Proficiência em Matemática	0 a 500

4.2 *Softwares*

Atualmente, existe um conjunto de ferramentas computacionais que podem ser usadas para explorar, avaliar e compreender a ampla técnica dos modelos multiníveis. Para a utilização dos MM's alguns algoritmos foram implementados em diversos *softwares*, entre eles HLM, MLwiN, SAS e R. No SAS, um programa computacional pago, pode-se destacar o pacote `PROC MIXED` (ver em Littell et al (2006)) sendo um pacote adequado para o ajuste e a estimação dos modelos multiníveis.

Para ajustar os modelos propostos será utilizado o *software* R, um programa computacional livre e gratuito. Esse *software* disponibiliza diversos pacotes, para o ajuste e estimação dos MM's, tais como *nlem*, *lme4* e o *multilevel*. Os pacotes *lme4* e *nlem* serão usados para a análise, e as funções para especificar o modelo são *lmer*() e *lme*(), respectivamente.

4.3 Modelo de regressão linear clássico - M0

Considerando os dados do SAEB, uma opção inicial de modelagem é através de um modelo de regressão linear clássico. Para a selecionar o melhor modelo foi utilizado a função *step*(). Vale ressaltar que o comando seleciona o melhor modelo por meio do AIC. Desta forma têm-se o modelo final resultante é:

$$profic_i \;=\; \beta_0 + \beta_1 Sexo + \beta_2 Idade + \beta_3 Rede + \varepsilon_i, \qquad i = 1, 2, \ldots 20491.$$

Na Tabela 3 são apresentadas as estimativas para os parâmetros do modelo de regressão clássico. Verifica-se que todos os parâmetros são significativos, ao nível de significância de 5%.

Tabela 3: Estimativa dos parâmetros para o modelo final

Parâmetro	Estimativa	Erro padrão	t valor	Nível descritivo
Intercepto	204,2791	0,3847	531,03	$< 0,001$
Idade	-5,5130	0,2094	-26,33	$< 0,001$
Sexo	-11,7480	0,5472	-21,47	$< 0,001$
Rede	39,1969	1,8539	21,14	$< 0,001$

O coeficiente de determinação do modelo foi 0,0735, sendo próximo de 0. Essas estimativas serão comparadas com as fornecidas pelo MM, considerando já que existe indicação de dependência nas observações, pois há um estrutura hierárquica no sistema educacional. Além do mais, uma das suposições do modelo linear clássico é que as observações não podem ter dependências. Com isso, uma solução plausível seria considerar a estrutura dos dados na análise, assim sendo o MM pode ser um boa alternativa para esse problema.

4.4 Modelos multiníveis

Nessa seção serão expostos alguns MM's, considerando dois níveis. O índice i representa e identifica o aluno e o índice j indica a que escola ele pertence. Com isso, serão definidas as variáveis de cada nível em estudo.

Variáveis explanatórias para o nível 1:

- Sexo_{ij}: o sexo do i-ésimo aluno da j-ésima escola;
- Raca_{ij}: a raça do i-ésimo aluno da j-ésima escola;
- Idade_{ij}: a idade do i-ésimo aluno da j-ésima escola.

No nível 2, escola, as variáveis explanatórias são as seguintes:

- Rede_{j}: tipo de rede de ensino da j-ésima escola;
- AtivPartic_{j}: participação de atividade do diretor da j-ésima escola.

A variável resposta considerada é $Profic_{ij}$: a proficiência em matemática do i-ésimo aluno da j-ésima escola.

4.4.1 *Modelo Anova com 1 fator e efeito aleatório - M1*

É considerado o modelo mais simples, pois não existem variáveis explicativas em nenhum dos dois níveis. A partir dele calcula-se o valor do coeficiente intraclasse. O modelo M1 é apresentado abaixo:

$$\begin{aligned} Profic_{ij} &= \beta_{0j} + \varepsilon_{ij} & \varepsilon_{ij} &\sim N(0, \sigma^2) \\ \beta_{0j} &= \gamma_{00} + u_{0j} & u_{0j} &\sim N(0, \tau_{00}) \end{aligned}$$

Na Tabela 4 são apresentadas as estimativas e o erro padrão para os estimadores.

Tabela 4: Estimativas dos parâmetros do modelo nulo (M1)

Efeito Fixo	Estimativas	Erro-padrão	t valor	Nível descritivo
γ_{00}	199,757	0,927	215,469	$< 0,001$
Efeito Aleatório	Estimativa			
τ_{00}	204,6848			
σ^2	1455,6294			

De acordo com a Tabela 4a estimativa para o intercepto (efeito fixo) mostra que a proficiência média dos alunos é 199,757 pontos, sendo significativamente diferente de zero, uma vez que o nível descritivo associado é $< 0,001$. Pode-se verificar que da variância total de 1660,3142; 204,6848 é devido a variabilidade entre escolas e 1455,6294 é devido a variabilidade intra escolas. Outra informação obtida a partir da tabela acima é o coeficiente intraclasse que é dado por:

$$\rho = \frac{\tau_{00}}{\sigma^2 + \tau_{00}} = \frac{204,6848}{1455,6294 + 204,6848} = 0,1232$$

O coeficiente intraclasse de 0,1232 significa que aproximadamente 12% da variância da proficiência em matemática pode ser atribuída a escola. Esse valor indica que a suposição de independência de regressão linear possivelmente não é adequada para esses dados, assim justificando o uso do MM. Porém, existe a necessidade de verificar se τ_{00} é significativo, ou seja, se existe variação no intercepto, o que é feito comparando os modelos com intercepto aleatório e sem intercepto aleatório. Essa comparação pode ser vista na Tabela 5.

Tabela 5: Teste de razão de verossimilhança

Modelo	gl	AIC	BIC	logverossimilhança	T.R.V	valor-p
com intercepto aleatório	3	208024,5	208048,3	-104009,3		
sem intercepto aleatório	2	209803,1	209819,0	-104899,6	1780,6	$< 0,001$

Com o resultado do valor-p $< 0,0001$ rejeita-se a hipótese de que o componente de variâncias é nulo. Desta forma, têm-se que o intercepto aleatório é significativo, ou seja, existe uma variação significativa entre as escolas.

4.4.2 *Modelo com o intercepto aleatório e variáveis explicativas no nível 1- M2*

Para este modelo, adicionam-se as variáveis explicativas *Sexo*, *Raca* e *Idade* associadas aos alunos no nível 1, além disso considera-se o intercepto aleatório. Assim sendo, o modelo M2 é dado por:

$$\begin{aligned}
Profic_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Raca_{ij} + \beta_{3j} Idade_{ij} + \varepsilon_{ij} \\
\beta_{0j} &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30}
\end{aligned}$$

Na Tabela 6 temos os valores das estimativas para o M2:

Tabela 6: Estimativas para os parâmetros do M2

Efeito Fixo	Estimativas	Erro-padrão	t valor	Nível descritivos
γ_{00}	204,8506	1,0405	196,8632	$< 0,001$
γ_{10}	-11,4110	0,5275	-21,6293	$< 0,001$
γ_{20}	0,6253	0,6466	0,9669	0,3336
γ_{30}	-5,1758	0,2055	-25,1851	$< 0,001$
Efeito Aleatório	**Estimativas**			
τ_{00}	180,2205			
σ^2	1389,3284			

Através da Tabela 6 pode-se verificar que apenas a variável raça não foi significativa para o modelo. Embora as variáveis explicativas sejam associadas aos alunos, pode-se verificar que a variância intra escola, ou variância entre alunos, sofreu redução.

4.4.3 *Modelo com intercepto e inclinações aleatórios - M3*

Considerando todos os coeficientes variando aleatoriamente, defini-se o M3 que pode ser especificado da seguinte forma:

$$\begin{aligned} Profic_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Idade_{ij} + \varepsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \end{aligned}$$

A Tabela 7 contém as estimativas dos parâmetros do modelo e verifica-se ao nível de significância de 5% que os coeficientes de efeitos fixos são significativos.

Tabela 7: Estimativa dos parâmetros do M3

Efeito Fixo	Estimativas	Erro-padrão	t valor	Nível Descritivo
γ_{00}	205,1944	0,9506	215,86	$< 0,001$
γ_{10}	-11,4545	0,5474	-20,93	$< 0,001$
γ_{20}	-5,2850	0,2189	-24,14	$< 0,001$
Efeito Aleatório	**Estimativas**			
τ_{00}	798,277			
τ_{11}	5,495			
τ_{22}	1,616			
σ^2	1387,786			

4.4.4 *Modelo com variáveis explicativas no nível 2 - M4*

Nesse modelo adiciona-se as variáveis explicativas *Rede* e *Atividade prática* no nível 2, com a finalidade de explicar a variabilidade das escolas. A estrutura do M4 é dada abaixo:

$$\begin{aligned} Profic_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Idade_{ij} + \varepsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01} Rede_j + \gamma_{02} AtivPartic_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \end{aligned}$$

Na Tabela 8 são apresentadas as estimativas dos parâmetros do modelo e pela mesma verifica-se que o nível descritivo da variável *ativpartic* não foi significativo, ou seja, a atividade continuada do diretor não afeta a proficiência do aluno.

Tabela 8: Estimativa dos parâmetros do M4

Efeito Fixo	Estimativas	Erro-padrão	t valor	Nível descritivo
γ_{00}	202,1834	2,0965	96,4344	$< 0,001$
γ_{10}	-11,4018	0,5272	-21,6232	$< 0,001$
γ_{20}	-5,0996	0,2054	-24,8274	$< 0,001$
γ_{01}	32,7330	3,5742	9,1579	$< 0,001$
γ_{02}	1,7951	2,2175	0,8095	0,4182
Efeito Aleatório	**Estimativa**			
τ_{00}	132,0127			
σ^2	1389,2710			

4.4.5 *Modelo com variável explicativa no nível 2 - M5*

Após a retirada dessa variável obtêm-se as estimativas apresentadas na Tabela 9:

Tabela 9: Estimativa dos parâmetros do M5

Efeito Fixo	Estimativas	Erro-padrão	t valor	Nível descritivo
γ_{00}	203,7437	0,8251	246,9185	$< 0,001$
γ_{10}	-11,4002	0,5272	-21,6208	$< 0,001$
γ_{20}	-5,0998	0,2053	-24,8291	$< 0,001$
γ_{01}	32,6558	3,5774	9,1282	$< 0,001$
Efeito Aleatório	**Estimativa**			
τ_{00}	132,4664			
σ^2	1389,2624			

4.4.6 *Escolha do modelo*

A escolha do modelo final será feita através do AIC, que é uma medida de ajuste do modelo, ou seja, por meio dele pode-se concluir qual é o melhor modelo para representar a natureza do banco de dados. Na Tabela 9 verifica-se o grau de liberdade e o valor do AIC. Por esse critério, o modelo a ser escolhido é o que apresentar o menor valor para o AIC.

Tabela 10: AIC para os modelos ajustados

Modelo	Grau de Liberdade	AIC
M0	5	208242,1
M1	3	208026,2
M2	6	207057,9
M3	10	207034,4
M4	7	206986,9
M5	6	206985,5

Pelo critério do AIC o modelo indicado é o M5 - que possui variável rede no nível 2 - pois possui um AIC igual a 206985,5. O modelo selecionado para explicar a proficiência em matemática do aluno i da escola j é:

$$Profic_{ij} \quad = \quad \gamma_{00} + \gamma_{01} Rede_j + \gamma_{10} Sexo_{ij} + \gamma_{20} Idade_{ij} + u_{0j} + \varepsilon_{ij}$$

Ao comparar as estimativas do modelo M5 com o M0 verifica-se um redução nessas estimativas

5 CONSIDERAÇÕES FINAIS

Os modelos multiníveis estão sendo muito usados nas mais diversas áreas de estudo, como por exemplo em Educação, pois vêm cada vez mais crescendo suas aplicações, fundamentações teóricas e ferramentas computacionais. Nesse trabalho, a teoria e aplicação dos modelos multiníveis foram apresentadas considerando os modelos lineares e que possuem distribuição normal.

Conforme os objetivos deste estudo, pode-se concluir que quando os dados estão organizados segundo uma estrutura hierárquica, é de suma importância levar em consideração essa estrutura, uma vez que não fazê-la pode implicar em uma superestimação dos coeficientes do modelo em estudo. Além disso, vários MM's com 2 níveis foram apresentados, desde os mais simples, em que não existe nenhuma variável explicativa, até os mais complexos, em que existe variáveis nos 2 níveis do modelo.

Na modelagem com os microdados do SAEB pode-se verificar que ao desconsiderar a estrutura hierárquica utilizado a regressão linear clássica obteve-se estimativas um pouco mais elevadas quando comparadas com as obtidas usado os modelos multiníveis. Além disso, ao utiliza o modelo linear clássico houve uma possível subestimação das estimativas dos erros padrões, pois o mesmo não considera as duas fontes de variação encontradas nos modelos multiníveis.

O *software* R mostrou-se muito útil no ajuste dos modelos multiníveis e na estimação dos parâmetros envolvidos nos modelos, facilitando, assim, o desenvolvimento desse trabalho.

Como continuidade para deste estudo seria interessante:

- Adicionar mais variáveis ao modelo na tentativa de melhorar a explicação da proficiência dos alunos; uma variável interessante a ser acrescentada no modelo seria o nível socioeconômico dos alunos;
- Fazer uma análise de diagnóstico para testar as suposições do modelo, assim como identificar pontos influentes. Para verificar mais detalhes ver em Nobre (2004),

Hilden-Minton (1995) , Simonoff (2013) e Pires (2009);

- Usar uma modelagem com estimação bayesiana. Para maiores detalhes verificar em Natis (2000).

REFERÊNCIAS

AGUERRE, T. F. Métodos Estadístico de Estimación de los Efectos de la Aplicación al Estudio de las Escuelas Eficaces. **REICE- Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, 2003, Vol. I, nº 2.

ALBERNAZ, A., FERREIRA, F. H., FRANCO, C. **Qualidade e eqüidade da educação fundamental brasileira**. Rio de Janeiro: PUC-Rio, 2002.

BERGAMO, Genevile Carife. **Aplicação de Modelos Multiníveis na Análise de Dados de Medidas Repetidas no Tempo**. 2002. 70 f. Dissertação (Mestrado) - Curso de Agronomia, Departamento de Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, São Paulo, 2002. Cap. 5.

BOGUTCHI, Tânia Fernandes. **MODELOS LINEARES HIERÁRQUICOS APLICADOS À GEOGRAFIA:** um estudo da avaliação do Ensino Fundamental em Minas. 2010. 139f. Tese (Doutorando em Geografia) -Tratamento da Informação Espacial, Pontifícia Universidade Católica de Minas Gerais, 2010. Gerais

BRYK, A.; RAUDENBUSH, S. **Hierarchical Linear Models:** applications and data analysis methods. London: Sage Publications, 1992.

BRYK, A.; RAUDENBUSH, S. **Hierarchical Linear Models:** applications and data analysis methods. London: Sage Publications, 2002, 2ed.

CHAVES, Nathalia Lima. **Modelos para dados pré-teste/pós-teste no contexto clássico e bayesiano**. 2012. 87 f. Monografia (Especialização) - Curso de Estatística, Departamento de Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, 2012.

CRUZ, F., **Modelos Multinível**, Revista per. Epidemiol., Vol. 12, nº 3, 2008.

CRUZ, Cláudia C. M. S. da. **Modelos Multi-nível:** Fundamentos e Aplicações. 2010. 200f. Dissertação (Mestrando em Matemática, Estatística e Computação), Universidade Aberta, Lisboa, 2010.

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

GELMAN, Andrew; HILL, Jennifer. **Data Analysis Using Regression and Multilevel/ Hiearchical Models**. New York: Cambridge University Press, 2007.

GOLDSTEIN, Harvey. **Multilevel Statistical Models**. London: Institute Of Education Multilevel Models, Project, 1995.

HARVILLE, D.A. **Maximum likelihood approaches to variance component estimation and to related problems**. Journal of the American Statistics Association, Washington, v.72, p.320-328, 1977.

HILDEN-MINTON, James Andrew.**Multilevel Diagnostics for Mixed and Hierarchical Linear Models**. 1995. 120 f. Dissertação (Mestrado) - Curso de Philosophy In Mathematics, University Of California, Los Angeles, 1995.

LITTELL, Ramon C. et al. **SAS for Mixed Models**. 2. ed. Nc: Sas Institute Inc., 2006.

LOBO, C. **Modelagem Bayesiana da Evolução da Mortalidade por Doenças Isquêmicas do Coração, no Estado do Rio de Janeiro, no Período de 1980 a 2002**. Dissertação de Mestrado, Departamento de Métodos Estatístico, IM-UFRJ, 2007.

LOPES, V. P.; RODRIGUES, L. P.; MAIA, J. A. R.; MALINA, R. M. **Motor coordination as predictor of physical activity in childhood**. Scandinavian Journal of Medicine & Science in Sports, 2010.

MAIA, J.A.; LOPES, V.P.; SILVA, R.G.; SEABRA, A.; FERREIRA, J.; CARDOSO, M. 2003. **Modelação hierárquica ou multi-nível. Uma metodologia estatística e um instrumento útil de pensamento na investigação em Ciências do Desporto**. Revista Portuguesa de Ciências do Desporto, vol.3, nº 1, 92-107 p.

MCCULLOCH, Charles E. *et al.* **Generalizad, Linear, and Mixed Models**. 2. ed. New Jersey: John Wiley & Sons, 2008.

MONETTE, Georges; KWAN, Qing Shao And Ernest. **A First Look at Multilevel Models**. New York: Institute For Social Research Statistical Consulting Service, 2002. 96 p. (York University)

MIGON, Hélio S.; SOUZA, Aparecida D.P.; SCHMIDT, Alexandra M.. **Modelos Hierárquicos e Aplicações**. São Pedro: ABE, 2008. p. 1 - 151.

MURILLO, F. J. Torrecilla. **Los modelos multinivel como herramienta para la investigación educativa**. Magis, Revista Internacional de Investigación en Educación,1, 45-62, 2008.

NATIS, Lílian. **Modelos Lineares Hierárquico**. 2000. 73 f. Dissertação (Mestrado) - Curso de Estatística, Universidade de São Paulo, São Paulo, 2000.

NEZLEK, J. B. **Multilevel Random Coefficient Analyses of Event and Interval Contingent Data in Social and Personality Psychology Research**. Personality and Social Psychology Bulletin, Vol. 27, nº 7, Julho 2001, 771-785.

NOBRE, Juvêncio Santos. **Métodos de Diagnóstico para Modelos Lineares Mistos**. 2004. 99 f. Dissertação (Mestrado) - Curso de Estatística, Universidade de São Paulo, São Paulo, 2004.

PATTERSON, H.D.; THOMPSON, R.**Recovery of inter-block information when block sizes are unequal**. Biometrika, London, v.58, p.545-554, 1971.

PIRES, Juliana Freitas. **INFLUÊNCIA LOCAL ATRAVÉS DA CURVATURA NORMAL EM MODELOS MULTINÍVEIS**. 2009. 88 f. Dissertação (Mestrado) - Curso de Estatística, Universidade Federal de Pernambuco, Recife, 2009.

R Core Team (2013). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http:www.R-project.org.

RAMOS, Manoel Wallace Alves. **A modelagem de um índice de produção científica através de modelos lineares generalizados hierárquicos**. 2009. 66 f. Dissertação (Mestrado) - Curso de Estatística, Departamento de Estatística, Universidade Federal de Pernambuco, Recife, 2009.

SICHIERI, R.; MOURA, E.C.. **Análise multinível das variações no índice de massa corporal entre adultos**. Brasil, 2009, Ver. Saúde Pública, vol.43, Supl. 2, São Paulo.

SIMONOFF, Jeffrey S. Regression tree-based diagnostics for linear multilevel models. **Statistical Modelling: An International Journal**. EUA, p. 459-480. out. 2013.

SOARES, J.F.; ALVES, M.T.G. **As pesquisas sobre o efeito das escolas:**contribuições metodológicas para a sociologia da educação. 2007. Brasília.

SULLIVAN, Lisa M. et al. **AN INTRODUCTION TO HIERARCHICAL LINEAR MODELLING.** Boston: University School Of Public Health, 1999.

VALENTE, V. **Estudo da ≪Relevância do apoio da Escola nas perspectivas profissionais dos alunos do 10ºAno de escolaridade≫ com aplicação dos Modelos Lineares Hierárquicos**,Portugal, Universidade Aberta (Tese), 2007, 294p.

ANEXO - COMANDOS NO R

```
# Pacote para a aplicação do modelo multinivel

require(lme4)
require(nlme)
require(multilevel)

# Leitura dos dados

dados<- read.table(choose.files())

# Fixando o banco de dados

attach(dados)

# Visualizando os primeiros valores

head(dados)

# Nome das variaveis

names(dados)

# recodificando a variável idade

idade<-idade-mean(idade)

sexo<-factor(sexo)

rede<-factor(rede)
```

```
raca<-factor(raca)

ativpart<-factor(ativpart)

# Modelo de regressão linear clássico

modre<-lm(proficiencia~idade+ sexo+ raca +rede+ativpart)

step(modre,direction = c("backward"))

# Modelo de regressão linear clássico escolhido

modref<-lm(proficiencia ~  sexo + idade + rede)

summary(modref)

# Analise de variância

anova(modref)

# Modelo Nulo- Não contém nenhuma variável - Anova com 1 fator aleatórios -M1

mod1<-lme(proficiencia ~ 1, random = ~ 1 | escola, data=dados,method="ML")

# ML metodo de maximo verossimilhança

summary(mod1)

# Estimativas as estimativas do componente de variancia

vcomp<- as.numeric(VarCorr(mod1)[, 1]); vcomp

# Coeficiene intral classe

tau.sq <- as.numeric(VarCorr(mod1)[1,1])
```

```
sigma.sq <- as.numeric(VarCorr(mod1)[2,1])

icc<-tau.sq/(tau.sq+sigma.sq); icc

# Verificando se existe intercepto aleatorio

mod.nulo.c<-lme(proficiencia~1, random= ~1|escola, data=dados)

mod.nulo<-gls(proficiencia~1, data=dados)

anova(mod.nulo.c,mod.nulo)

# Modelo com as variaveis explicativas com coeficiente aleatório- M2

mod2<-lme(proficiencia ~ 1 + sexo + raca + idade,

random = ~ 1 | escola, data=dados,method="ML")
summary(mod2)

# Estimativas as estimativas de componte de varincia

vcomp1<- as.numeric(VarCorr(mod2)[, 1]); vcomp1

# Coeficiene intral classe

tau.sq <- as.numeric(VarCorr(mod2)[1,1])

sigma.sq <- as.numeric(VarCorr(mod2)[2,1])

icc2<-tau.sq/(tau.sq+sigma.sq); icc2

# Modelo com intercepto e inclinação aleatorio

mod3<-lmer(proficiencia ~ 1 + sexo + idade +(sexo+idade|escola),
data=dados,REML=F)
```

```
summary(mod3)

# Estimativa do componente de variância

vcomp2<- as.numeric(VarCorr(mod3)[,3]); vcomp2

# Modelo com variáveis no segundo nível

mod4<-lme(proficiencia ~ 1 + sexo + idade +rede+ ativpart,
random = ~ 1 | escola, data=dados,method="ML")
summary(mod4)

# Estimativa do componente de variância

vcomp3<- as.numeric(VarCorr(mod4)[,1]); vcomp3

# Modelo com variáveis no segundo nível sem ativpart

mod5<-lme(proficiencia ~ 1 + sexo + idade + rede,
random = ~ 1 | escola, data=dados,method="ML")

summary(mod5)

# Estimativa do componente de variância

vcomp4<- as.numeric(VarCorr(mod5)[,1]); vcomp4

#Escolha do modelo

AIC(modref,mod1,mod2,mod3,mod4,mod5)
```