

USO DE MODELOS MULTINÍVEIS NA ANÁLISE DE DADOS DE MEDIDAS REPETIDAS NO TEMPO: UMA APLICAÇÃO A DADOS DO PROGRAMA DE ATENÇÃO AO IDOSO

Genevile Carife BERGAMO¹
José Eduardo CORRENTE²

- **RESUMO:** Em muitos trabalhos científicos, é comum encontrar os dados estruturados de forma hierárquica, ou seja, os indivíduos em estudo estão agrupados em unidades de nível mais baixo, que por sua vez pertencem a unidades de um nível mais alto e assim sucessivamente. Na análise desse tipo de dados é importante levar em conta a estrutura hierárquica uma vez que não fazê-la pode implicar na superestimação dos coeficientes do modelo em estudo. Assim, para facilitar a análise de dados seguindo uma estrutura hierárquica, foram desenvolvidos os modelos multiníveis. Tais modelos consideram toda a variabilidade existente para os dados tanto num mesmo nível como nos diferentes níveis da hierarquia. No caso da análise de dados de medidas repetidas no tempo, uma estrutura hierárquica em dois níveis pode ser considerada, organizando as ocasiões de medidas, no primeiro nível, para cada indivíduo no segundo nível. Neste trabalho é feita uma abordagem dos modelos multiníveis para vários níveis da hierarquia, bem como os métodos de estimação e teste dos parâmetros envolvidos no modelo. Essa estrutura foi utilizada para analisar dados provenientes do Programa de Atenção ao Idoso (PAI), desenvolvido no ambulatório municipal Dr. Plínio do Prado Coutinho em Alfenas, MG, em que foram observadas as variáveis Índice de Massa Corpórea (IMC) e Pressão Arterial dos idosos durante 22 meses. As análises foram feitas com o uso do *software* SAS, v.8.0.
- **PALAVRAS-CHAVE:** Modelos multiníveis; modelos hierárquicos; medidas repetidas; pressão arterial; índice de massa corpórea.

1 Introdução

Em estudos biológicos, bem como em outras áreas do conhecimento, é comum encontrar os dados estruturados de forma hierárquica, ou seja, os indivíduos em

¹Universidade de Alfenas - UNIFENAS, CEP 37130-000, Alfenas, MG, Brasil.

²Departamento de Bioestatística, Instituto de Biociências, Universidade Estadual Paulista - UNESP, CEP 18618-000, Botucatu, SP, Brasil. E-mail: jecorren@ibb.unesp.br

estudo estão agrupados em unidades de nível mais baixo, que por sua vez pertencem a unidades de um nível mais alto e assim sucessivamente. Reconhecer e levar em conta a estrutura hierárquica dos dados é importante, uma vez, que não fazê-la pode implicar na superestimação dos coeficientes do modelo em estudo.

Em modelos de regressão tradicionais, existe a suposição de independência entre os indivíduos. No entanto, quando os dados são estruturados em hierarquias, as unidades de um mesmo nível pertencentes a uma unidade de nível mais alto, raramente são independentes. Isso acontece porque essas unidades compartilham de um mesmo ambiente ou apresentam características semelhantes. Assim, a suposição de independência foi violada e passa a existir correlação entre essas unidades. É preciso que o modelo a ser ajustado leve em conta tais correlações. Para isso, foram desenvolvidos modelos multiníveis, os quais contemplam uma estrutura hierárquica dos dados, considerando todas as correlações existentes entre as observações nos diferentes níveis de hierarquia.

A utilização de modelos multiníveis está inicialmente relacionada a um estudo com crianças do ensino fundamental conduzido nos anos 70 na Inglaterra. Bennett (1976) alegou que as crianças expostas à uma maneira formal de ensinar a ler, exibiam um aprendizado maior do que as não expostas. Os dados foram analisados utilizando técnicas de regressão múltipla tradicionais, considerando as crianças individualmente como as unidades de análise e ignorando os agrupamentos para professores e classes. Os resultados obtidos foram estatisticamente significantes. Subseqüentemente, Aitkin et al. (1981) demonstraram que, ao considerar a análise com as crianças agrupadas em classes, as diferenças significantes desapareciam, e as que se submeteram ao ensino formal não apresentaram qualquer diferença em relação às demais (Goldstein, 2003). Como exemplo, a Figura 1 mostra uma estrutura hierárquica, considerando um estudo em que o interesse está na avaliação das escolas por meio do desempenho de seus alunos e professores. Assim, os alunos estão agrupados por professores e esses agrupados por escolas.

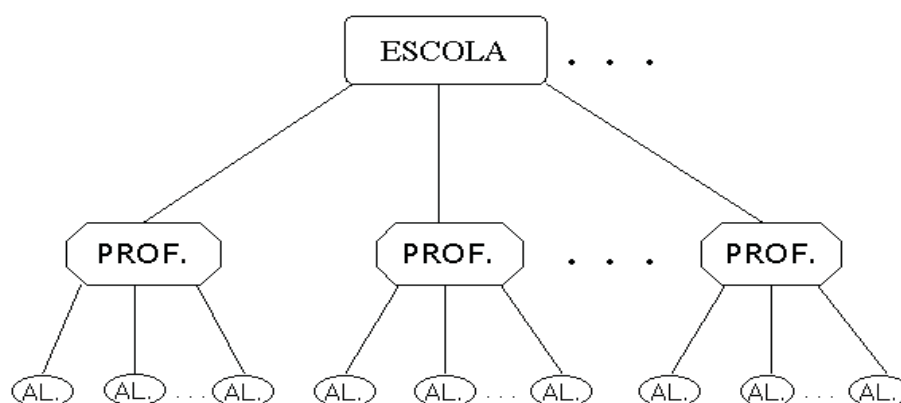


FIGURA 1 - Estrutura dos dados para um modelo hierárquico em três níveis.

Nos modelos multiníveis, as unidades de cada nível são vistas como tendo um efeito aleatório, ou seja, são amostras aleatórias de uma população dessas unidades. Esses efeitos aleatórios são responsáveis no modelo por coeficientes aleatórios que levam em conta a variabilidade entre essas unidades, seja de forma simples, por meio da variabilidade nos interceptos, ou de forma mais complexa, pela variabilidade nas inclinações. Além disso, os modelos multiníveis com efeitos aleatórios possuem a vantagem, em relação aos tradicionais, de acomodar a estrutura hierárquica aos dados.

Aitkin e Longford (1986) iniciaram, assim, uma série de procedimentos, resultando nas idéias centrais dos modelos multiníveis. Bryk e Raudenbush (1992) discutiram modelos lineares multiníveis com 2 e 3 níveis para aplicações específicas a dados educacionais e experimentos com medidas repetidas. A Figura 2 mostra uma estrutura hierárquica com medidas repetidas no tempo, em que as ocasiões de medida estão no primeiro nível da estrutura hierárquica. Aitkin e Longford (1986) iniciaram, assim, uma série de procedimentos, resultando nas idéias centrais dos modelos multiníveis. Bryk e Raudenbush (1992) discutiram modelos lineares multiníveis com 2 e 3 níveis para aplicações específicas a dados educacionais e experimentos com medidas repetidas. A Figura 2 mostra uma estrutura hierárquica com medidas repetidas no tempo, em que as ocasiões de medida estão no primeiro nível da estrutura hierárquica.

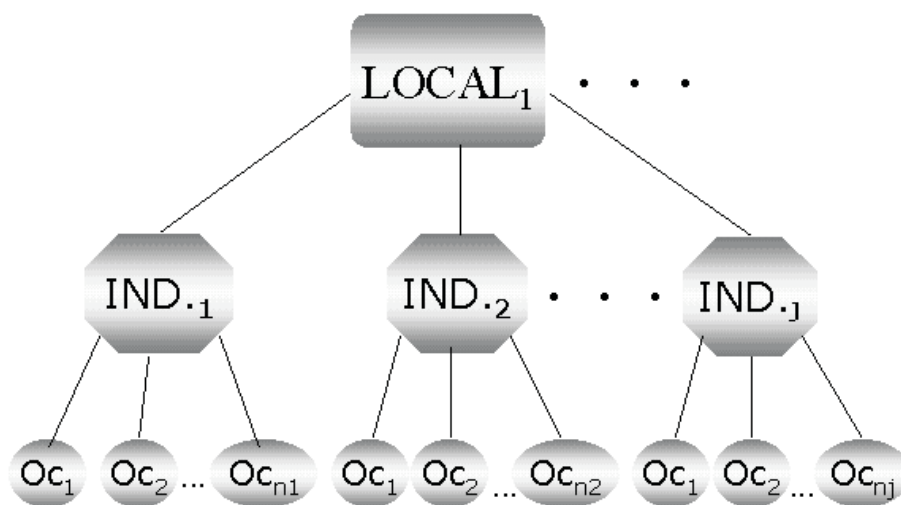


FIGURA 2 - Estrutura dos dados para um modelo hierárquico em três níveis, com medidas repetidas no tempo.

Longford (1993) fornece uma orientação mais teórica dos modelos multiníveis incluindo a discussão de um modelo multinível para análise fatorial, modelos com respostas categorizadas e outros modelos multivariados. Atualmente, o

desenvolvimento desses modelos está voltado para as áreas de dados com respostas discretas, modelos de séries temporais, classificações cruzadas, dados perdidos e modelos não-lineares, com aplicação na educação, epidemiologia, geografia, crescimento animal e vegetal, pesquisas domiciliares, entre outras (Goldstein, 2003).

1.1 Estrutura geral dos modelos multiníveis

De acordo com Neter et al. (1990), no modelo de regressão linear simples, a relação entre a variável resposta e a variável explanatória geralmente é escrita como:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{com } i = 1, \dots, n,$$

em que

Y_i representa a resposta do i -ésimo indivíduo;

β_0 o valor esperado da variável resposta Y_i para X_i igual a zero;

β_1 a mudança esperada em Y_i quando X_i aumenta de uma unidade;

X_i é variável explanatória do i -ésimo indivíduo;

ϵ_i é o erro associado ao i -ésimo indivíduo, com as suposições:

$$\epsilon_i \sim N(0, \sigma^2) \text{ e } \epsilon_i \text{'s independentes.}$$

Ao se considerar várias retas de regressão, que podem ser originadas de experimentos realizados sob as mesmas condições, pode-se ter um conjunto de retas para explicar a variabilidade dos experimentos. Assim, comparando-se dois experimentos, o que possui maior intercepto é o que proporciona um melhor resultado na origem. O que possui menor inclinação mostra uma relação com a variável explanatória mais fraca, o que o torna menos preditivo em relação ao outro experimento. Desse modo, se o objetivo é estudar não apenas um determinado evento, mas estudar tudo o que os envolve, é possível ajustar um modelo que leve em conta toda a variabilidade entre os experimentos e incorpore os diferentes aspectos de cada um deles. Uma maneira de se levar em conta tal tipo de análise é considerar os dados de acordo com uma estrutura hierárquica.

Em modelos hierárquicos ou modelos multiníveis com dois níveis, as observações são classificadas como unidades de nível 1 e unidades de nível 2, ou seja, ocorrem n_j unidades do nível 1 para cada unidade j ($j = 1, 2, \dots, J$) do nível 2. Neste caso, os dados não precisam ser balanceados, isto é, não é necessário que $n_j = n_k$ para $j \neq k$.

Assim, os modelos para o nível 1 são desenvolvidos separadamente em cada unidade j do nível 2, levando-se em consideração uma possível variação de interceptos e inclinações. Considerando o caso de uma variável resposta Y e uma única variável explanatória X do nível 1, o modelo no nível 1 é da forma:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij}^* + \epsilon_{ij}, \tag{1}$$

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

- Y_{ij} é a variável resposta do i -ésimo indivíduo do nível 1 para o j -ésimo nível 2,
- X_{ij}^* a variável explanatória em sua medida original centrada na média geral $\bar{X}_{..}$ ou centrada na média de uma unidade do nível 2 $\bar{X}_{.j}$, medidas na i -ésima unidade do nível 1 agrupadas para a j -ésima unidade do nível 2;
- β_{0j} é o intercepto para a j -ésima unidade do nível 2;
- β_{1j} é a inclinação associada à variável explanatória X_{ij}^* da i -ésima unidade do nível 1 para a j -ésima unidade do nível 2;
- ϵ_{ij} é o erro aleatório associado à i -ésima unidade do nível 1 agrupado para a j -ésima unidade do nível 2, com as suposições:

$$\epsilon_{ij} \sim N(0, \sigma^2) \text{ e } \epsilon_{ij}\text{'s independentes.}$$

A interpretação dos parâmetros do modelo (particularmente os interceptos β_{0j} para todo j) depende do modo como as variáveis explanatórias do nível 1 são consideradas. Pode-se dizer que:

- se a variável explanatória do nível 1, X_{ij}^* , é considerada em sua medida original, isto é, como X_{ij} , então o intercepto β_{0j} é o valor esperado da variável resposta Y_{ij} quando X_{ij} for igual a zero;
- agora, se a variável explanatória do nível 1, X_{ij}^* , está centrada na sua média geral, isto é, centrada em $\bar{X}_{..}$, então, o intercepto β_{0j} , é interpretado como a média da j -ésima unidade do nível 2 ajustada para a variável X ;
- mas, se a variável explanatória X_{ij}^* , do nível 1, é centrada em $\bar{X}_{.j}$, ou seja, na respectiva média da unidade do nível 2, então o intercepto β_{0j} é interpretado como a média não ajustada da variável resposta Y_{ij} ;
- finalmente, se a variável explanatória do nível 1, X_{ij}^* é centrada em qualquer outro valor médio, então, o intercepto β_{0j} é interpretado como o valor esperado da variável resposta Y_{ij} para cada valor da variável explanatória.

No modelo para o nível 2, considerando as variáveis respostas Y_{ij} e a variável explanatória X_{ij}^* , têm-se J modelos iguais aos mostrados em (1) em que, para cada modelo há interceptos β_{0j} 's e inclinações β_{1j} 's diferentes, para $j = 1, \dots, J$. Para os modelos no nível 2, estes coeficientes de regressão são considerados como variáveis resposta, isto é:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \nu_{0j} \\ \beta_{1j} &= \gamma_{10} + \nu_{1j}, \end{aligned} \tag{2}$$

em que

γ_{00} e γ_{10} são os valores esperados dos interceptos e das inclinações nas unidades do nível 2, respectivamente;

ν_{0j} e ν_{1j} são os efeitos aleatórios da j -ésima unidade do nível 2 no intercepto β_{0j} e na inclinação β_{1j} , respectivamente ;

τ_{00} e τ_{11} são as variâncias populacionais dos interceptos e das inclinações, respectivamente;

τ_{01} é a covariância entre β_{0j} e β_{1j} , com as suposições:

$\nu_{0j} \sim N(0, \tau_{00})$ e ν_{0j} 's independentes;

$\nu_{1j} \sim N(0, \tau_{11})$ e ν_{1j} 's independentes;

$Cov(\nu_{0j}, \nu_{1j}) = \tau_{01}$;

ν_{0j} 's e ν_{1j} 's independentes dos ϵ_i 's.

Em resumo,

$$\beta_{0j} \sim N(\gamma_{00}, \tau_{00}), \quad \beta_{1j} \sim N(\gamma_{10}, \tau_{11}) \quad \text{e} \quad Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}.$$

Nota-se que, com as equações (2), as unidades do nível 2 não apresentam o mesmo intercepto e a mesma inclinação e os seus efeitos aleatórios, ν_{0j} e ν_{1j} , ajudam a esclarecer estas possíveis diferenças.

Além disso, é possível incluir variáveis explanatórias relacionadas ao nível 2 para explicar a variabilidade entre as unidades do nível 2. Assim, considere as variáveis W_j para o nível 2. Os modelos para este nível serão agora da forma:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + \nu_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_j + \nu_{1j}, \end{aligned} \tag{3}$$

em que

β_{0j} e β_{1j} são o intercepto e a inclinação para a j -ésima unidade do nível 2, respectivamente;

γ_{00} e γ_{10} são os valores esperados do intercepto e da inclinação para W_j ;

γ_{01} e γ_{11} são os coeficientes de regressão associados à variável explanatória W_j do nível 2 relativo ao intercepto e inclinação, respectivamente;

ν_{0j} e ν_{1j} são os efeitos aleatórios da j -ésima unidade do nível 2 sobre o intercepto e inclinação para W_j , respectivamente;

τ_{00} e τ_{11} são as variâncias populacionais dos interceptos e das inclinações corrigida pela variável W_j , respectivamente;

τ_{01} é a covariância entre β_{0j} e β_{1j} , com as suposições:

$\nu_{0j} \sim N(0, \tau_{00})$ e ν_{0j} 's independentes;

$\nu_{1j} \sim N(0, \tau_{11})$ e ν_{1j} 's independentes;

$Cov(\nu_{0j}, \nu_{1j}) = \tau_{01}$;

ν_{0j} 's e ν_{1j} 's independentes dos ϵ_i 's.

Note que, da mesma forma que foi vista anteriormente para a variável explanatória no nível 1 para o modelo dado em (1), pode-se considerar W_j em sua escala original ou uma variável W_j^* centrada na média geral \bar{W} . Incluindo a variável W_j , τ_{00} , τ_{11} e τ_{01} são agora componentes de variância e covariância condicionais ou residuais.

Substituindo (3) na equação (1), tem-se:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j^* + \gamma_{10}W_j^*X_{ij}^* + \gamma_{11}W_j^*X_{ij}^* + \nu_{0j} + \nu_{1j}X_{ij}^* + \epsilon_{ij}. \quad (4)$$

O modelo combinado (4) envolve as variáveis explanatórias X_{ij}^* do nível 1 e W_j^* do nível 2, um termo entre os níveis $W_j^*X_{ij}^*$ e um termo complexo $\nu_{0j} + \nu_{1j}X_{ij}^* + \epsilon_{ij}$, considerado como um erro complexo. Ainda, tem-se que os erros não são independentes nas unidades de nível 1, ou seja, há uma dependência entre as unidades do nível 1 agrupadas dentro de cada unidade do nível 2 em termos de ν_{0j} e ν_{1j} . Além disso, as variâncias dos erros podem não ser homogêneas se ν_{0j} e ν_{1j} assumirem diferentes valores dentro de cada unidade do nível 2. Esses modelos também podem ser escritos em notação matricial.

As variáveis explanatórias X e W para os níveis 1 e 2, respectivamente, também podem ser consideradas em suas medidas originais ou centradas. A partir dessa estrutura, Sullivan et al. (1999) apresentam uma tabela com os casos especiais para os modelos descritos.

O modelo (4) não é uma forma na qual o método dos mínimos quadrados ordinários possa ser usado para estimar os parâmetros. Isto porque, nesse tipo de estrutura, diferentes parâmetros são usados. Especificamente, os coeficientes β_j no nível 1 podem ser fixos ou aleatórios; no nível 2, os coeficientes γ são considerados fixos, as variâncias e co-variâncias entre os níveis 1 e 2 são chamados componentes de variância. Assim, para estimar os parâmetros envolvidos no modelo, são necessários processos iterativos. Há várias sugestões de métodos de estimação para esse caso, mas neste trabalho foi utilizado o método REML. Maiores detalhes sobre os processos de estimação dos parâmetros envolvidos no modelo podem ser encontrados em Littell et al. (1996) e em Searle et al. (1992).

Para verificar a significância dos parâmetros tanto no nível 1 como no nível 2, uma hipótese de interesse é feita levando-se em conta se o efeito é fixo ou aleatório. No caso de efeito fixo, o teste é baseado na estatística t-Student e no do efeito ser aleatório é usada uma estatística de teste com distribuição aproximada. Em geral, os *softwares* já realizam esses testes para a verificação da existência de tais efeitos.

Singer (1998) comenta que o PROC MIXED do SAS é um programa flexível para o ajuste de modelos multiníveis. Isso por esse programa ter sido desenvolvido para analisar modelos mistos, em que são levados em conta efeitos fixos e aleatórios, permitindo considerar interceptos e inclinações como efeitos aleatórios e medir a variabilidade existente nos níveis e entre níveis. Um problema também citado pela autora, é o fato de que o usuário pode não estar familiarizado com a linguagem SAS. Assim, uma descrição detalhada da aplicação do PROC MIXED para modelos multiníveis pode ser encontrada no referido artigo.

Diante deste fato, o objetivo deste trabalho consiste em avaliar a aplicação dos modelos multiníveis na análise de dados de medidas repetidas no tempo por meio da análise dos dados referente ao “Programa de Atenção ao Idoso” (PAI) desenvolvido no Ambulatório Dr. Plínio do Prado Coutinho em Alfenas, MG. Esses dados foram coletados durante 22 semanas de pacientes com mais de 60 anos e foram observadas as pressões arteriais sistólica e diastólica dos pacientes em pé e sentado. Também foram levadas em conta as variáveis Índice de Massa Corpórea (IMC), Sexo, Raça e Estado Civil dos idosos.

2 Metodologia

2.1 Análise dos dados do Programa de Atenção ao Idoso - “PAI”

2.1.1 Descrição do experimento

Os dados foram coletados de um Programa de Atenção ao Idoso (PAI), criado pelo Ministério da Saúde e com implementação facultativa pelos municípios.

A Secretaria Municipal de Saúde de Alfenas, MG, implementou o PAI no Ambulatório Dr. Plínio do Prado Coutinho em 1999, tendo já passado por algumas reformulações e atualmente consta de equipe formada por médico geriatra, enfermeira, assistente social, fisioterapeuta e técnico de enfermagem. Com esta equipe prestam assistência na qualidade de participantes consultores fonoaudiólogo, odontólogo, psicólogo, terapeuta ocupacional e nutricionista.

As atividades do programa são desenvolvidas no período da manhã, de segunda à sexta-feira, atendendo a todos os clientes referenciados das Unidades Básicas de Saúde, com idade igual ou superior à 60 anos.

Como protocolo é realizada avaliação multidimensional desses clientes, ou seja, os aspectos físicos, estrutura de suporte social, estado psicológico e habilidade funcional. Todos os clientes passam obrigatoriamente pela consulta médica e de enfermagem, e pelo assistente social, sendo encaminhados aos demais profissionais quando necessário. Além desse atendimento são realizadas visitas domiciliares a todo cliente que inicia o Programa, que podem ser repetidos sempre que houver necessidade.

Os dados obtidos pelo “PAI” referem-se ao sexo, raça e estado civil dos idosos e as medidas de pressão arterial, peso e altura tomados a cada consulta, no período entre março de 2000 e dezembro de 2001, num total de 22 meses (ocasiões). Como o retorno à cada consulta não ocorre todos os meses, isso leva a uma ausência de informação mensal de cada idoso; dos 166 pesquisados apenas 18 encontravam-se no programa em março de 2000, e outros possivelmente podem ter abandonado o programa, gerando um número grande de observações faltantes.

2.1.2 Modelos para a análise dos dados do "PAI"

Para a utilização de um modelo multinível, considerou-se as ocasiões da consulta como variável de nível 1 e os idosos de nível 2. Como variáveis-resposta, considerou-se a pressão arterial sistólica e diastólica tomadas sentado e em pé e, como variáveis explanatórias, o índice de massa corpórea ($IMC = \text{kg/m}^2$), sexo, raça e estado civil. A variável IMC foi centrada no seu valor médio, sendo utilizada a notação IMCC. Assim, o modelo foi estruturado como segue:

Variáveis do nível 1 (ocasião da consulta)

- $pass_{ij}$: valor da pressão arterial sistólica sentado na i -ésima ocasião para o idoso j ;
- $pads_{ij}$: valor da pressão arterial diastólica sentado na i -ésima ocasião para o idoso j ;
- $pasp_{ij}$: valor da pressão arterial sistólica em pé na i -ésima ocasião para o idoso j ;
- $padp_{ij}$: valor da pressão arterial diastólica em pé na i -ésima ocasião para o idoso j ;
- $imcc_{ij}$: valor do índice de massa corpórea centrado na sua média geral para i -ésima ocasião no j -ésimo idoso.

Variáveis do nível 2 (idoso)

- $sexo_j = \begin{cases} 1 & \text{se o idoso é do sexo feminino} \\ 0 & \text{se o idoso é do sexo masculino} \end{cases}$
- $raça_j = \begin{cases} 1 & \text{se o idoso é da raça branca} \\ 0 & \text{em caso contrário} \end{cases}$
- $estado\ civil_j = \begin{cases} 1 & \text{se o idoso é casado} \\ 0 & \text{em caso contrário} \end{cases}$

Foram conduzidas análises para todas as variáveis resposta obtidas do programa. Assim, como uma notação geral, usou-se pa^{**} , para denotar a pressão arterial. Para cada caso em separado, será utilizada a notação: PASS - pressão arterial sistólica sentado, PADS - pressão arterial diastólica sentado, PASP - pressão arterial sistólica em pé, PADP - pressão arterial diastólica em pé. Para os testes de hipóteses e obtenção dos p-valores, foram utilizadas o teste t-Student para os efeitos fixos, e a aproximação normal para os componentes de variância, dados por (Sullivan et al. 1999).

Inicialmente, foi considerado um modelo apenas com os interceptos dado por:

$$\begin{aligned} pa_{ij}^{**} &= \beta_{0j} + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \nu_{0j}, \end{aligned} \quad (5)$$

em que

$\epsilon_{ij} \sim N(0, \sigma^2)$, ϵ_{ij} 's independentes;
 $\nu_{0j} \sim N(0, \tau_{00})$ e ν_{0j} 's independentes;
 $Cov(\nu_{0j}, \nu_{1j}) = \tau_{01}$ e
 ν_{0j} 's independentes dos ϵ_i 's.

Assim, no nível 1, a variável-resposta da ocasião, pa^{**} é a soma de um intercepto para os idosos de uma ocasião (β_{0j}) com um erro aleatório (ϵ_{ij}) associado à i -ésima ocasião no j -ésimo idoso. No nível 2, o intercepto do nível dos idosos é a soma da média global (γ_{00}) com uma série de desvios aleatórios em relação a essa média (ν_{0j}).

Adicionando-se a seguir, a variável IMCC ao modelo como uma variável explanatória, tem-se:

$$\begin{aligned} pa_{ij}^{**} &= \beta_{0j} + \beta_{1j}(imc_{ij} - M)_{ij} + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \nu_{0j} \\ \beta_{1j} &= \gamma_{10}, \end{aligned} \quad (6)$$

em que

M é a média geral da variável imc_{ij} ;
 $\epsilon_{ij} \sim N(0, \sigma^2)$, ϵ_{ij} 's independentes;
 $\nu_{0j} \sim N(0, \tau_{00})$ e ν_{0j} 's independentes;
 ν_{0j} 's independentes dos ϵ_i 's.

Com a variável explanatória do nível 1 centrada na média geral $imcc = (imc_{ij} - M)$, tem-se a mudança esperada na pa^{**} do j -ésimo idoso quando o IMC aumenta de uma unidade em relação à sua média geral M . Assim, o intercepto β_{0j} é a pa^{**} esperada no j -ésimo idoso quando ela atinge um IMC igual a M .

Considerando-se, a seguir, o IMCC com efeito aleatório para o nível dos idosos, ou seja, não apenas a pressão em uma determinada ocasião está relacionada com um IMCC dessa ocasião, mas também que a relação entre os IMCC pode variar entre dos idosos, tem-se:

$$\begin{aligned} pa_{ij}^{**} &= \beta_{0j} + \beta_{1j}(imc_{ij} - M)_{ij} + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \nu_{0j} \\ \beta_{1j} &= \gamma_{10} + \nu_{1j}, \end{aligned} \quad (7)$$

em que

M é a média geral da variável imc_{ij} ;
 $\epsilon_{ij} \sim N(0, \sigma^2)$, ϵ_{ij} 's independentes;

$\nu_{0j} \sim N(0, \tau_{00})$ e ν_{0j} 's independentes;
 $\nu_{1j} \sim N(0, \tau_{11})$ e ν_{1j} 's independentes;
 $Cov(\nu_{0j}, \nu_{1j}) = \tau_{01}$ e
 ν_{0j} 's e ν_{1j} 's independentes dos ϵ_i 's.

O modelo (7) permite ao intercepto e a inclinação variarem nos idosos, aumentando a matriz de variâncias e co-variâncias para representar os efeitos aleatórios nos idosos, isto é, os componentes de variância do intercepto e da inclinação são adicionados de um componente de co-variância entre o intercepto e a inclinação (τ_{01}). A matriz de co-variância pode ser representada de uma forma simples, como:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix}.$$

Adicionando as variáveis explanatórias do segundo nível: sexo, raça e estado civil ao modelo (7), tem-se:

$$\begin{aligned} pa_{ij}^{**} &= \beta_{0j} + \beta_{1j}imcc + \beta_2masculino_j + \beta_3n\tilde{a}o\ brancos_j + \beta_4n\tilde{a}o\ casados_j + \epsilon_{ij} \\ \beta_{0j} &= \gamma_{00} + \nu_{0j} \\ \beta_{1j} &= \gamma_{10} + \nu_{1j}, \end{aligned} \tag{8}$$

em que as suposições dos efeitos aleatórios são análogas às do modelo (7).

3 Resultados e discussão

Considerando a variável pressão arterial sistólica sentado, $pass_{ij}$, e seguindo o modelo dado em (5), as estimativas dos parâmetros considerados são mostradas na Tabela 1.

Tabela 1 - Estimativa dos parâmetros do modelo (5) considerando apenas os interceptos para os dados de pressão arterial sistólica sentado - "PAI"

Parâmetros	Estimativas	Erro Padrão	p-valor
γ_{00}	142,577	1,333	< 0,0001
τ_{00}	237,840	32,203	< 0,0001
σ^2	315,303	14,178	< 0,0001
Estatística	-2*log-veros.	10201,950	-

Neste modelo foi considerado o intercepto como um efeito fixo. A estimativa de 142,577 mmHg dada na Tabela 1, representa a pressão média arterial sistólica sentado dos idosos. Analisando os resíduos considerados como efeitos aleatórios, observa-se uma maior variação entre ocasiões entre os idosos ($\sigma^2 = 315,303$) do que

apenas entre os idosos ($\tau_{00} = 237,840$). Isso indica a existência de uma influência das ocasiões na pressão arterial sistólica sentado para cada idoso.

Uma outra forma de verificar a variação na pressão arterial sistólica sentado seria estimar a correlação intragrupo (neste caso, intra-idosos), indicada por ρ , a qual estima a parte da variação total que ocorre entre as unidades do nível 2 (idosos). Este valor é:

$$\hat{\rho} = \frac{237,840}{237,840 + 315,303} = 0,4300,$$

e mede até que ponto a pressão arterial sistólica sentado medida nas ocasiões em um mesmo idoso assemelha-se quando comparada com as pressões nos diferentes idosos. Tem-se que, neste caso, esse valor indica apenas 43% de variabilidade intra-idosos. Ocorre que o modelo ajustado é bastante simples, podendo incluir outras variáveis medidas no programa obtendo-se um possível aumento nesse valor.

Ainda no modelo considerado, o valor da estatística -2 vezes o logaritmo da função de verossimilhança foi de 10.201,950, obtida ao considerar 1.154 valores do total de 3.652 observações. Essa estatística será utilizada na comparação desse modelo simples com modelos mais complexos, por meio de um teste χ^2 para a razão das verossimilhanças.

Continuando com a variável pressão arterial sistólica sentado, $pass_{ij}$ e seguindo agora o modelo dado em (6), as estimativas dos parâmetros considerados são mostradas na Tabela 2.

Tabela 2 - Estimativa dos parâmetros do modelo (6) considerando IMCC como uma variável explanatória para os dados de pressão arterial sistólica sentado - "PAI"

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	141,899	1,348	< 0,0001
γ_{01}	1,094	0,260	< 0,0001
τ_{00}	234,810	32,137	< 0,0001
σ^2	309,694	14,012	< 0,0001
Estatística	-2*log-veros.	10039,010	-

Tem-se na Tabela 2, um intercepto de 141,899 mmHg representando a pressão arterial sistólica sentado esperada no j -ésimo idoso quando o IMCC for igual a 25,2482 (valor médio M para a variável). A estimativa da variável IMCC foi positiva (1,094), indicando um aumento da pressão arterial sistólica sentado com o aumento do IMCC, ou seja a pressão arterial sistólica sentado aumenta de 1,094 mmHg, quando o IMCC aumenta de uma unidade em relação à média geral 25,2482. Para os resíduos, observa-se pouca alteração em relação ao modelo considerando apenas os interceptos, ou seja, entre ocasiões dentro dos idosos ($\sigma^2 = 309,694$) e entre os idosos ($\tau_{00} = 234,810$). Isso indica uma influência das ocasiões na pressão arterial sistólica sentado para cada idoso, mesmo após a inclusão da variável IMCC.

Nesse caso a correlação intra-idosos é dada por:

$$\hat{\rho} = \frac{234,810}{234,810 + 309,694} = 0,4312,$$

o que praticamente não muda em relação ao modelo anterior.

A estatística -2 vezes o logaritmo da função de verossimilhança para esse modelo é igual a 10.039,010, considerando 1.138 valores do total de 3.652 observações. Isso mostra uma redução do modelo com a variável IMCC em relação ao modelo só com os interceptos (10.201,950 - 10.039,010 = 162,94). Assim, o valor da diferença obtida entre os dois modelos se comparado com o quantil de uma distribuição χ^2 com 1 grau de liberdade, tem-se que uma redução significativa nessa estatística ($p < 0,0001$). Verifica-se, desse modo, que há uma variação da pressão arterial de idoso para idoso com diferentes IMCC.

Fazendo o ajuste para a variável $pass_{ij}$ no modelo (7), ou seja, fazendo a variável IMCC agora como um efeito aleatório para o nível dos idosos tem-se os resultados mostrados na Tabela 3.

Tabela 3 - Estimativa dos parâmetros do modelo (7) considerado o IMCC com efeito aleatório para o nível dos idosos para os dados de pressão arterial sistólica sentado - "PAI"

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	142,019	1,329	< 0,0001
γ_{10}	1,118	0,266	< 0,0001
τ_{00}	227,804	37,072	< 0,0001
τ_{11}	0,147	0,918	0,8728
τ_{01}	3,103	3,374	0,3577
σ^2	309,506	14,016	< 0,0001
Estatística	-2*log-veros.	10038,000	-

De acordo com a Tabela 3, a pressão arterial sistólica sentado dos idosos controlada pelo IMCC das ocasiões é de 142,019 e o valor 1,118 é a inclinação média representando a relação entre essa pressão e o IMCC das ocasiões.

A matriz de covariância é dada por:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 227,804 & 3,103 \\ 3,103 & 0,147 \end{pmatrix}.$$

O valor 227,804 representa a variação nos interceptos, enquanto o valor 0,147 representa a variação nas inclinações e 3,103 representa a co-variância entre os interceptos e as inclinações. Nesse caso, nota-se que existe uma grande variabilidade nos interceptos. Em outras palavras, as pressões arteriais sistólicas sentado dos idosos diferem, em média, mesmo após a inclusão do efeito IMCC. Já as inclinações

não têm o mesmo significado, uma vez que apresentaram um erro padrão alto (0,918). Nota-se também que a co-variância entre os interceptos e as inclinações é muito pequena em relação ao seu erro padrão (3,374). Assim, não há evidência de que os efeitos dos IMCC das ocasiões sobre as diferentes valores de pressão arterial sistólica sentado difiram, dependendo da pressão arterial sistólica sentado média no idoso.

Para o ajuste do modelo (7), o valor da estatística -2 vezes o logaritmo da função de verossimilhança foi de 10038,000 considerando 1.138 valores de um total de 3.652 observações. Isso indica uma pequena redução do modelo, considerando a variável IMCC como efeito aleatório para o nível dos idosos em relação ao modelo considerando a variável IMCC apenas como explanatória fixa (10.039,010 - 10.038,000 = 1,010). Comparando o valor da diferença das estatísticas com o quantil de uma distribuição χ^2 com 2 graus de liberdade, tem-se que o p-valor é de 0,6035, comprovando a não-significância dos modelos ajustados. Desse modo, quando se considera as inclinações como efeitos aleatórios, não existe uma melhora no modelo, indicando que não existe uma variação da pressão arterial sistólica sentado de idoso para idoso através das ocasiões.

Adicionando, agora, as variáveis explanatórias do segundo nível, ou seja, o sexo, a raça e o estado civil, tem-se os resultados para a variável $pass_{ij}$ com o modelo (8) mostrados na Tabela 4.

Tabela 4 - Estimativa dos parâmetros do modelo (8) considerando o IMCC com efeito aleatório para o nível dos idosos mais as variáveis explanatórias desse nível para os dados de pressão arterial sistólica sentado - "PAI"

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	142,568	2,164	< 0,0001
γ_{10}	1,146	0,275	< 0,0001
β_2	0,633	2,988	0,8325
β_3	-4,301	3,088	0,1656
β_4	0,933	2,753	0,7351
τ_{00}	227,635	37,270	< 0,0001
τ_{11}	0,272	0,954	0,7756
τ_{01}	3,306	3,459	0,3392
σ^2	309,210	14,011	< 0,0001
Estatística	-2*log-veros.	10036,030	-

Analisando a Tabela 4, nota-se que pouca alteração ocorreu nas estimativas dos parâmetros, não alterando as conclusões obtidas para a variável resposta $pass_{ij}$ considerando o modelo (7), mesmo após a inclusão das variáveis explanatórias do nível dos idosos. Isso pode ser observado pelo valor da estatística -2 vezes o logaritmo da função de verossimilhança, que apresenta uma pequena redução nos

valores ($10038,000 - 10036,030 = 1,970$). Novamente, se comparar esse valor obtido com o quantil de uma distribuição χ^2 com 3 graus de liberdade, tem-se um p-valor de 0,5787, indicando a não-significância para o modelo adicionado das variáveis sexo, raça e estado civil para o nível dos idosos.

Portanto, o modelo final a ser considerado para a variável $pass_{ij}$ leva em consideração apenas a variável explanatória IMCC fixa, de acordo com o modelo (6).

A Figura 3 representa graficamente o valor estimado da variável $pass_{ij}$, de acordo com o modelo (6), e observa-se diferenças nos interceptos e igualdade nas inclinações.

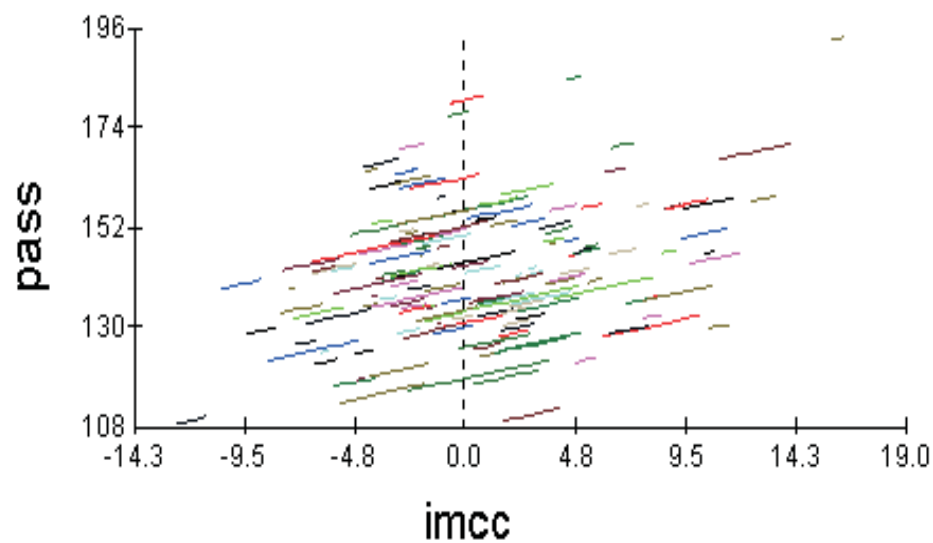


FIGURA 3 - Gráfico do modelo final para a variável $pass_{ij}$.

Considerando agora a variável pressão arterial diastólica sentado, $pads_{ij}$, e seguindo a seqüência de modelos dada por (5), (6), (7) e (8), tem-se os resultados apresentados na Tabela 5.

Analisando a Tabela 5, verifica-se que as estimativas dos parâmetros referente às variáveis explanatórias sexo, raça e estado civil são não significativas, mostrando apenas uma variação nos interceptos, referente aos parâmetros τ_{00} e τ_{10} . Com isso, o modelo final a ser considerado para a variável $pads_{ij}$ deve levar em consideração apenas a variável explanatória IMCC.

Tabela 5 - Estimativa dos parâmetros do modelo (8) considerando o IMCC com efeito aleatório para o nível dos idosos mais as variáveis explanatórias desse nível para os dados de pressão arterial diastólica sentado - "PAI"

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	83,412	1,003	< 0,0001
γ_{10}	0,582	0,130	< 0,0001
β_2	1,131	1,380	0,4137
β_3	1,102	1,435	0,4437
β_4	-0,440	1,276	0,7307
τ_{00}	44,668	8,087	< 0,0001
τ_{11}	0,037	0,198	0,8518
τ_{01}	1,175	0,725	0,0419
σ^2	100,378	4,539	< 0,0001
Estatística	-2*log-veros.	8691,148	-

A Figura 4 representa graficamente o valor estimado da variável $pads_{ij}$, de acordo com o modelo (6), na qual observa-se também diferenças nos interceptos e igualdade nas inclinações.

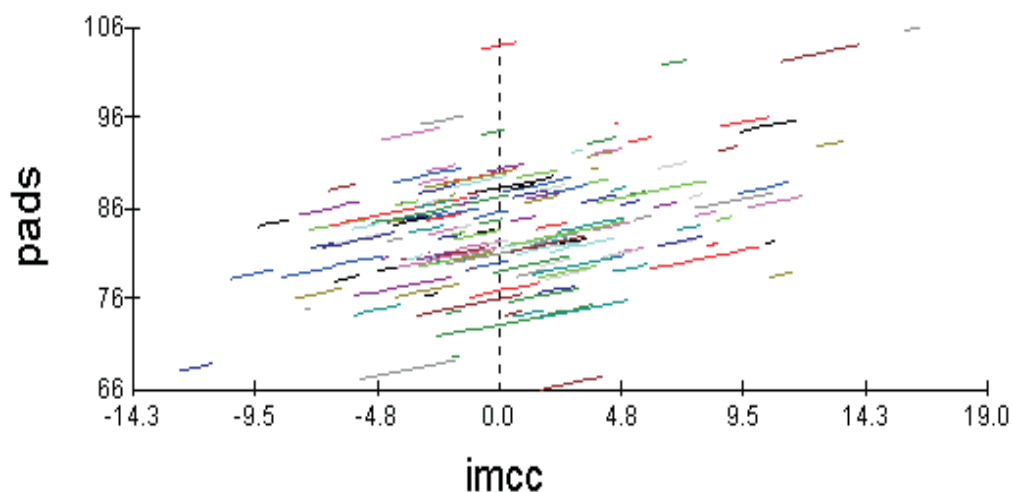


FIGURA 4 - Gráfico do modelo final para a variável $pads_{ij}$.

Para a variável pressão arterial sistólica em pé, $pasp_{ij}$, seguindo a mesma seqüência de modelos visto anteriormente, tem-se que os resultados são apresentados na Tabela 6.

Tabela 6 - Estimativa dos parâmetros do modelo (8) considerando o IMCC com efeito aleatório para o nível dos idosos mais as variáveis explanatórias desse nível para os dados de pressão arterial sistólica em pé - "PAI"

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	150,509	4,813	< 0,0001
γ_{10}	0,388	0,884	0,6625
β_3	0,404	9,344	0,9657
β_4	3,378	7,909	0,6714
τ_{00}	357,061	144,950	0,0138
τ_{11}	4,435	5,135	0,3878
τ_{01}	9,449	17,541	0,5901
σ^2	364,263	68,536	< 0,0001
Estatística	-2*log-veros.	955,905	-

Da mesma forma que o ajuste anterior, a Tabela 6 mostra que não houve diferença significativa nas estimativas dos parâmetros referente às variáveis explanatórias sexo, raça e estado civil, revelando apenas uma diferença nos interceptos. Assim sendo, o modelo final a ser considerado para a variável $pasp_{ij}$, leva em consideração apenas a variável explanatória IMCC.

O resultado apresentado graficamente (Figura 5) representa o valor estimado da variável $pasp_{ij}$, de acordo com o modelo (6), observando-se diferenças nos interceptos e igualdade nas inclinações.

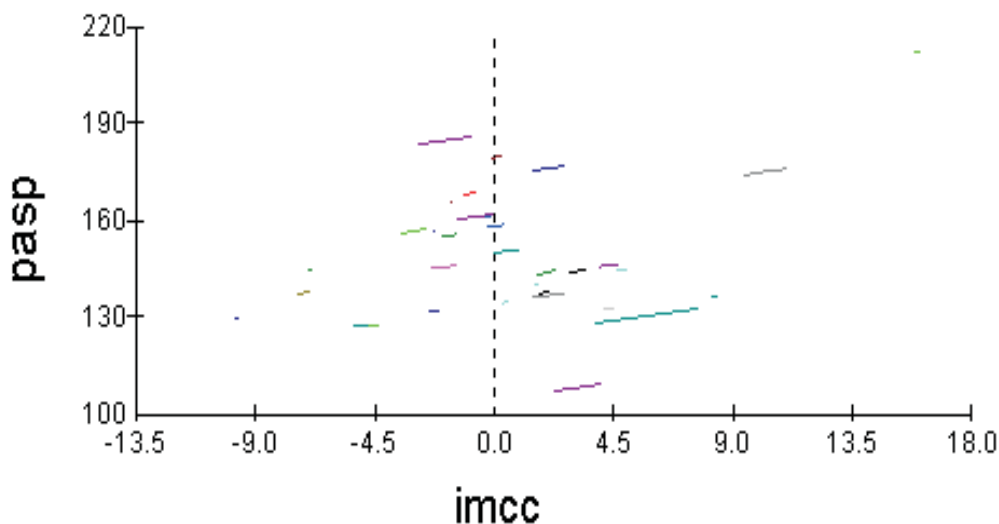


FIGURA 5 - Gráfico do modelo final para a variável $pasp_{ij}$.

Para a variável pressão arterial diastólica em pé, $padp_{ij}$, seguindo a mesma seqüência de ajuste, tem-se os resultados apresentados na Tabela 7.

Tabela 7 - Estimativa dos parâmetros do modelo (7) considerando o IMCC como efeito aleatório para o nível dos idosos para os dados de pressão arterial diastólica em pé - “PAI”

Parâmetros	Estimativas	Erro padrão	p-valor
γ_{00}	89,541	1,839	< 0,0001
γ_{10}	0,666	0,345	0,0588
τ_{00}	99,460	32,557	0,0023
τ_{11}	0,000	0,000	-
τ_{01}	0,000	0,000	-
σ^2	101,446	19,158	< 0,0001
Estatística	-2*log-veros.	802,234	-

Neste caso, a matriz de co-variâncias, não é positiva definida, assim sendo, os valores de τ_{11} e τ_{01} não são estimados e não há como adicionar as variáveis explanatórias sexo, raça e estado civil do segundo nível. Novamente o modelo final a ser considerado para a variável $padp_{ij}$ leva em consideração apenas a variável explanatória IMCC, ou seja, o modelo (6).

O resultado apresentado graficamente (Figura 6) representa o valor estimado da variável $padp_{ij}$, de acordo com o modelo (6), observando-se diferenças nos interceptos e igualdade nas inclinações.

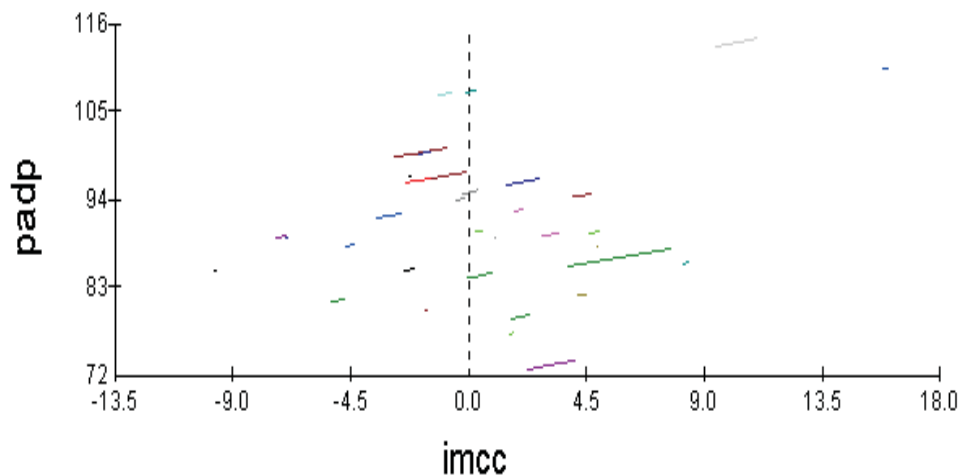


FIGURA 6 - Gráfico do modelo final para a variável $padp_{ij}$.

Conclusões

De acordo com os objetivos deste estudo, pode-se concluir que:

- quando os dados estão organizados segundo uma estrutura hierárquica, é importante levar em conta essa estrutura, uma vez que não fazê-lo pode implicar na superestimação dos coeficientes do modelo em estudo;
- quando se tem dados de medidas repetidas no tempo é interessante considerar uma estrutura hierárquica para esses dados, já que é possível verificar as possíveis correlações existentes em cada nível da estrutura com as possíveis variações dos tempos em cada indivíduo;
- o *software* SAS através do PROC MIXED mostrou-se muito eficiente na estimação dos parâmetros envolvidos no modelo, não apresentando quaisquer problemas com relação à presença de observações faltantes nas diferentes ocasiões de medida; seu uso também é interessante por fornecer os níveis de significância para os testes dos parâmetros estimados;
- para os dados do “PAI”, levar em conta a estrutura hierárquica dos dados revelou que as diferentes medidas de pressão arterial estão relacionadas positivamente com o IMC ao longo do tempo e, revelou também que a variação na pressão arterial não depende de sexo, raça e estado civil;

Agradecimentos

Os autores agradecem aos pareceristas pelas sugestões e contribuições para a melhor apresentação deste trabalho.

BERGAMO, G. C; CORRENTE, J. E. Multilevel models applied in the analysis of repeated measure data: an application to Elderly Care Program (ECP). *Rev. Mat. Estat.*, São Paulo, v.22, n.3, p.81-101, 2004.

- **ABSTRACT:** *It is common to find hierarchically structured data in several scientific studies, that is, the studied subjects are nested in the lowest level units belonging to the highest level units, and so on. To analyze this sort of data, it is important to take the hierarchical structure into account; otherwise, coefficients can be overestimated in the studied model. Then, in order to make data analysis according to the hierarchical structure easier, multilevel models were developed. Such models take into account all the existing variability for the data at the same level as well as for those in different levels of the hierarchy. In the case of repeated measure data, a two-level hierarchical structure can be considered, organizing the occasions at the first level for each subject at the second level. In this study, the multilevel models for several levels as well as the estimation methods and tests for the involved parameters in the model are approached. As an application, data from the Elderly Care Program (ECP), developed at outpatient clinic Dr. Plinio do Prado Coutinho in Alfenas, MG, where the Body Mass Indexes and Blood Pressure Rates of 22 elderly patients were observed during 22 months, were used.*

- **KEYWORDS:** *Multilevel models; hierarchical models; repeated measures; blood pressure; body mass index.*

Referências

- AITKIN, M.; LONGFORD, N. Statistical modelling in school effectiveness studies. *J. R. Stat. Soc.*, Series A, Cambridge, v.149, p.1-43, 1986.
- AITKIN, M.; ANDERSON, D.; HINDE, J. Statistical modelling of data on teaching styles. *J. R. Stat. Soc.*, Series A, Cambridge, v.144, p.148-161, 1981.
- BENNETT, N. *Teaching styles and pupil progress*. Cambridge: Harvard University Press, 1976. 201p.
- BRYK, A. S.; RAUDENBUSH, S. W. *Hierarchical linear models*. Newbury Park: Sage, 1992. 202p.
- GOLDSTEIN, H. Multilevel mixed linear model analysis using iterative generalized least square. *Biometrika*, London, v.73, n.1, p.43-56, 1986.
- GOLDSTEIN, H. *Multilevel statistical models*. 3rd. ed. London: Edward Arnold, 2003. 253p.
- LITTELL, R. C. et al. *SAS system for mixed models*. Cary: SAS Institute, 1996. 633p.
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. *Applied linear statistical models: regression, analysis of variance and experimental designs*. 3rd. ed. Burr Ridge: R.D. Irwin, 1990. 1181p.
- SAS INSTITUTE *SAS/STAT user's guide, version 6*. 4th. ed. Cary. 1990. v.2, 796p.
- SEARLE, S. R.; CASELLA, G.; McCULLOCH, C. E. *Variance components*. New York: John Wiley, 1992. 501p.
- SINGER, J. D. Using SAS proc mixed to fit multilevel models, hierarchical models, and individual growth models. *J. Educ. Behav. Stat.*, Washington, v.24, n.4, p.323-355, 1998.
- SULLIVAN, L. M.; DUKES, K.; LOSINA, E. Tutorial in biostatistics an introduction to hierarchical linear modelling. *Stat. Med.*, Chichester, v.18, p.855-888, 1999.

Apêndice

Programas SAS para a análise dos dados do "PAI"

Programa para o modelo (5), considerando apenas os interceptos

```
proc mixed nclprint covtest; class idoso; model pass= /solution;  
random intercept/subject=idoso;
```

Programa para o modelo (6), considerando IMCC como uma variável explanatória

```
proc mixed nclprint covtest; class idoso; model pass = imcc  
/solution ddfm=bw; random intercept /subject=idoso ;
```

Programa para o modelo (7), considerando IMCC com efeito aleatório para o nível dos idosos

```
proc mixed nclprint covtest; class idoso; model pass=  
imcc/solution ddfm=bw; random intercept imcc/ subject=idoso  
type=un;
```

Programa para o modelo (8), considerando IMCC com efeito aleatório para o nível dos idosos adicionando as variáveis explanatórias desse nível (sexo, raça, estado civil)

```
proc mixed nclprint covtest; class idoso; model pass= imcc sexo  
cor estciv/solution ddfm=bw; random intercept imcc/ subject=idoso  
type=un;
```

Recebido em 14.09.2004.

Aprovado após revisão em 08.12.2004.