

Lista 02 - Planejamento e Análise de Experimentos (MAE0316)

Caio M. de Almeida - 15444560 Eduardo Yukio G. Ishihara - 15449012
Gustavo S. Garone - 15458155 Ian B. Loures - 15459667
João Victor G. de Sousa - 15463912

11 de novembro de 2025

Nesta lista, usaremos “.” como separador decimal e quatro dígitos decimais.

Exercício 3

Análise Exploratória de Dados (AED)

Consta-se três variáveis qualitativas: a variável “group”, dividido entre “C” para suplementados por cálcio e “P” para placebo; a variável “visit”, que indica o número da visita (1, 2, 3, 4 ou 5), espaçadas aproximadamente um semestre entre si, indicando dados longitudinais; e a variável “person”, indicando o número da participante no estudo. Como variáveis quantitativas, temos a variável “age”, medida de forma contínua em anos, a variável resposta “bmd”, que indica a densidade mineral óssea (DMO), medida em g/cm² e a variável “ctime”, uma medida contínua do tempo das observações, transformadas das datas das visitas.

Iniciaremos por uma análise numérica para então partirmos para a análise gráfica.

Observamos na Tabela 1 que, em ambas as condições, a média da densidade mineral óssea aumenta com o passar do tempo. Contudo, o grupo suplementado parece apresentar crescimento maior, especialmente entre a segunda e terceira visita, quando comparado com o grupo placebo. Por outro lado, a mediana diferença de crescimento maior entre os grupos da primeira para segunda visita.

Partiremos para a análise gráfica para melhor visualizarmos essas diferenças. Foi orientado atenção ao conteúdo longitudinal dos dados. Temos três variáveis temporais disponíveis: a variável “visit”, que indica o número da visita (1 a 5); a variável “ctime”, que indica o tempo contínuo em que as observações foram realizadas; e a variável “age”, que indica a idade das participantes no momento da medição. Por conta da colinearidade, precisaremos escolher uma dessas variáveis para representar o fator temporal na modelagem, em especial se a idade inicial das participantes for similar.

Tabela 1: Análise descritiva numérica da DMO

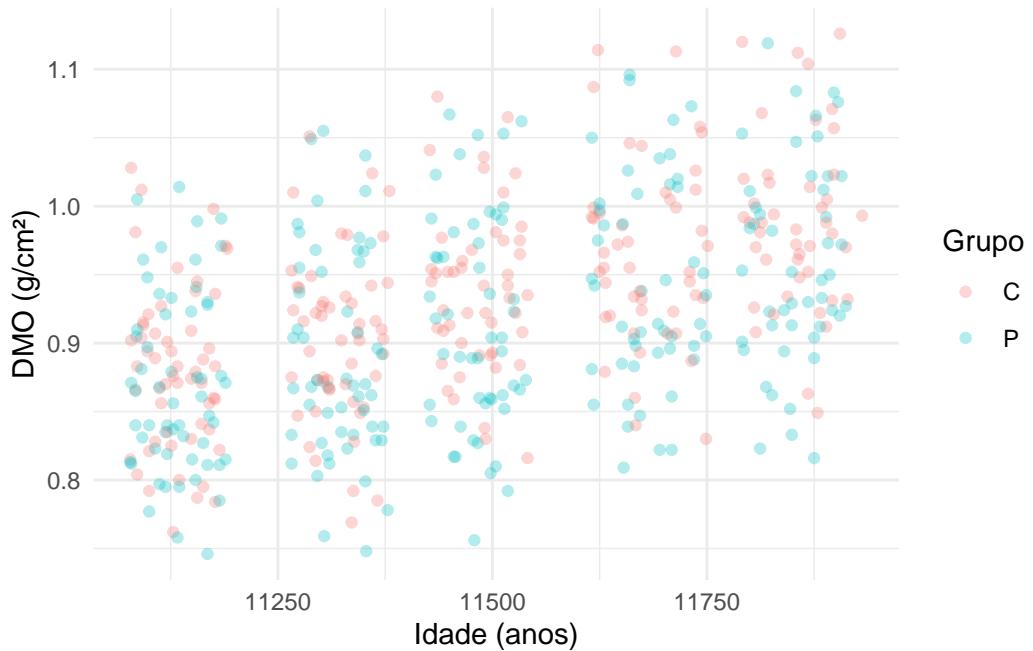
(a) Placebo

visit	n	media	sd	min	mediana	max
1	57	0.870	0.066	0.746	0.861	1.014
2	53	0.890	0.075	0.748	0.869	1.055
3	51	0.914	0.078	0.756	0.904	1.067
4	48	0.942	0.075	0.809	0.925	1.096
5	47	0.958	0.074	0.816	0.950	1.119

(b) Suplementado

visit	n	media	sd	min	mediana	max
1	55	0.880	0.060	0.762	0.879	1.028
2	52	0.903	0.059	0.769	0.905	1.051
3	48	0.938	0.059	0.816	0.939	1.080
4	46	0.965	0.065	0.830	0.955	1.114
5	44	0.988	0.063	0.849	0.986	1.126

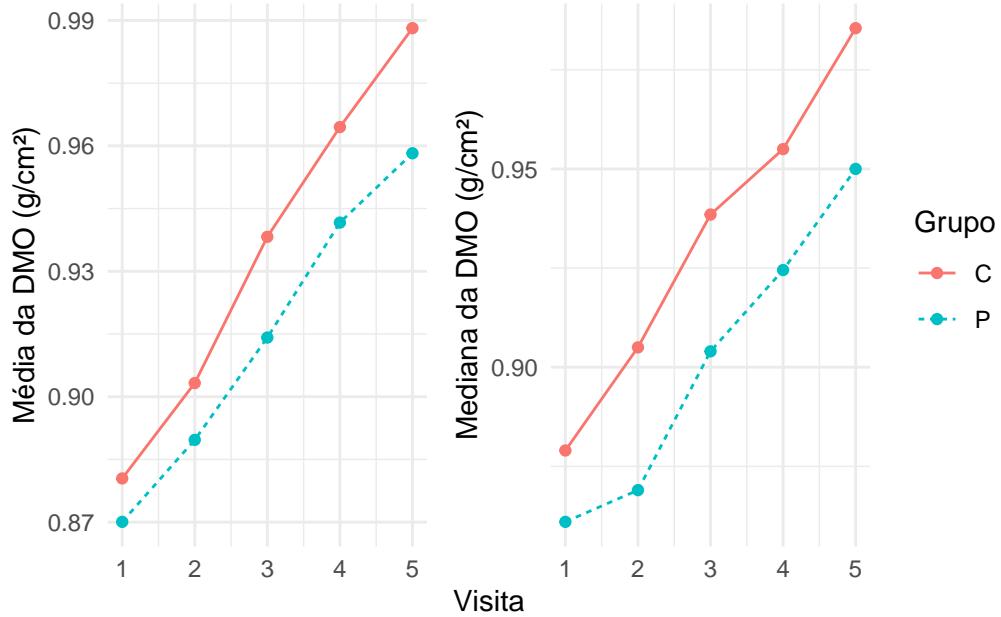
Figura 1: Evolução da densidade mineral óssea ao longo do tempo contínuo



Os gráficos na Figura 1 mostram a evolução da densidade mineral óssea ao longo do tempo contínuo para ambos os grupos. Observa-se que o grupo suplementado com cálcio apresenta um aumento mais acentuado na densidade mineral óssea *média* em comparação com o grupo placebo, especialmente entre a segunda e terceira visita, enquanto a diferença entre crescimento das *medianas* é mais acentuada entre a primeira e segunda visita. Ambos resultados sugerem que pode existir um efeito de interação entre a idade e a suplementação, com esta sendo mais efetiva em crianças mais novas.

Levaremos isso em consideração na modelagem, atentando-nos a possível existência de interação. Verificaremos se essa tendência é a mesma ao analisarmos em função das visitas, como esperamos.

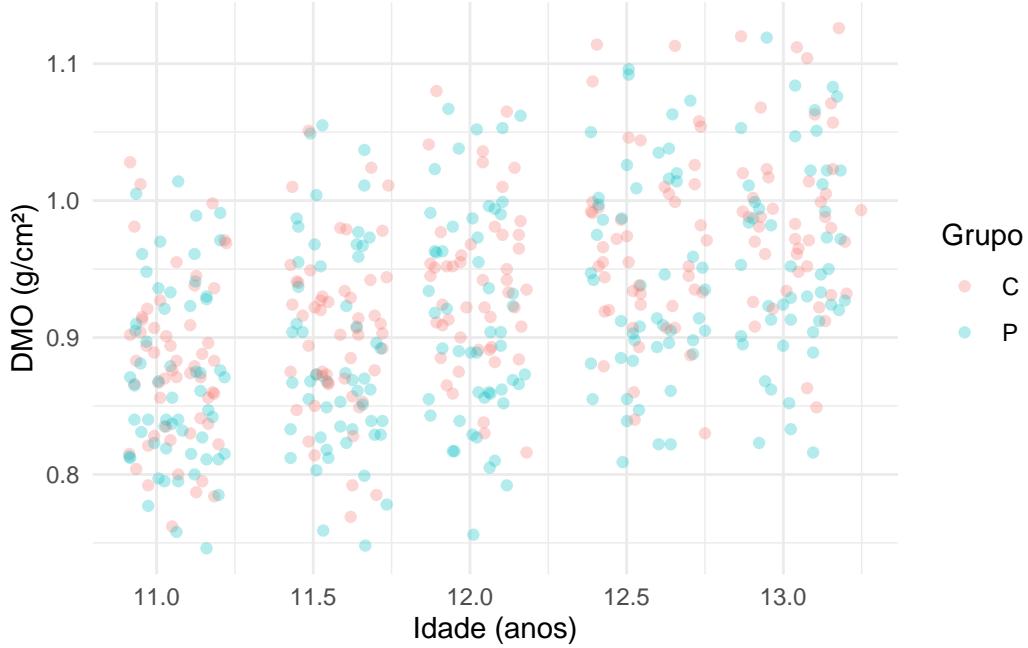
Figura 2: Evolução da densidade mineral óssea ao longo das visitas



Os gráficos na Figura 2 mostram a evolução da densidade mineral óssea ao longo das visitas para ambos os grupos. Observa-se que os resultados são similares aos observados na Figura 1, confirmando que ambas variáveis retratam o fator temporal similarmente.

Analisaremos agora a evolução da densidade mineral óssea em função da idade das participantes. Esperamos que, caso as idades iniciais das participantes sejam similares, o gráfico seja similar ao da Figura 1.

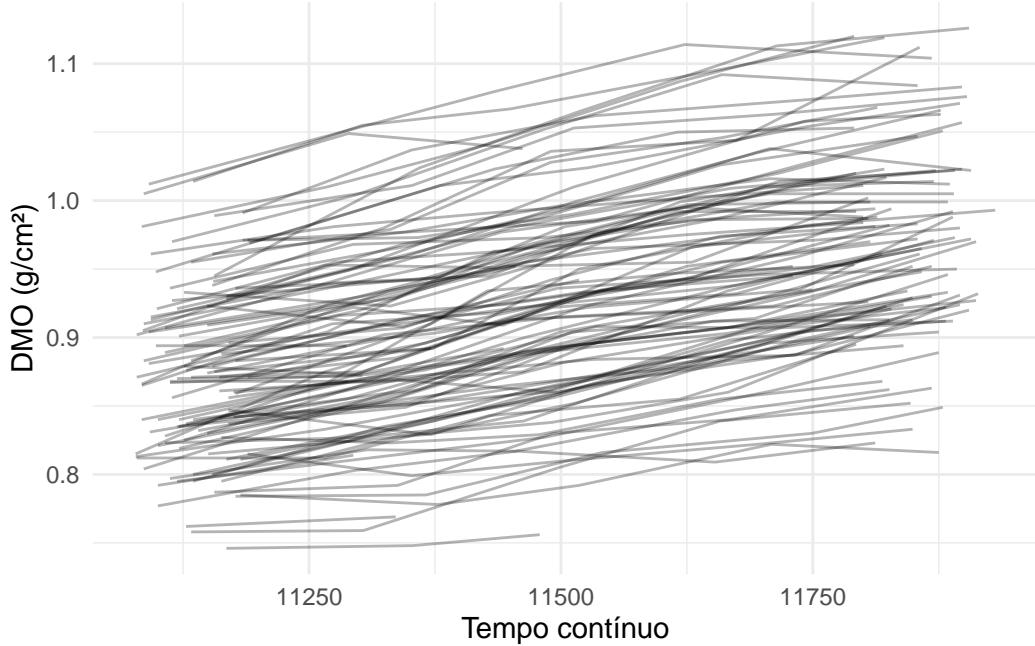
Figura 3: Evolução da DMO ao longo do tempo contínuo, separada por idade



O gráfico da Figura 3 mostra a dispersão da densidade mineral óssea em função da idade das participantes. Observa-se que este gráfico é quase idêntico ao gráfico da Figura 1, sugerindo que ambas variáveis retratam o fator temporal similarmente, uma vez que, como indicam, as participantes iniciaram sua participação no estudo com idades próximas. Disso, concluímos que tanto a variável “ctime” quanto a variável “age” podem ser utilizadas para representar o fator temporal quantitativamente na modelagem, enquanto a variável “visit” pode ser utilizada como variável categórica alternativa.

Finalmente, resta-nos uma análise dos perfis individuais das participantes para identificação de efeito aleatório.

Figura 4: Perfis da DMO das participantes ao longo do tempo contínuo



Pelo gráfico da Figura 4, observa-se que os perfis individuais das participantes apresentam variação considerável em seus níveis iniciais de densidade mineral óssea, sugerindo a presença de um efeito aleatório de intercepto. Por outro lado, a baixa variação na inclinação dos perfis sugere que pode não haver efeito aleatório de inclinação. Ademais, não está claro se, caso de fato exista efeito aleatório de inclinação, este seria correlacionado com o efeito aleatório de intercepto. Levaremos essas observações em consideração na modelagem.

Formulação dos modelos

Para este modelo, utilizaremos a variável “Age” para representação do aspecto temporal por sua fácil interpretação biológica. A variável “visit” será incluída indiretamente como as observações de um mesmo indivíduo ao longo do tempo (ver modelo adiante). A variável “ctime” não será utilizada por redundância e colinearidade.

Somada à variável “Age”, temos a variável “group”, que indica se a participante foi suplementada com cálcio ou recebeu placebo, e a interação entre “group” e “age”, que pode capturar efeitos diferenciados da suplementação em diferentes idades, como sugere a Análise Exploratória de Dados.

Finalmente, precisaremos incluir efeitos aleatórios para capturar a variação entre as participantes, conforme aponta o gráfico da Figura 4, e tratar da dependência entre observações de uma mesma participante. Durante nossa análise, faremos testes para determinar a necessidade de incluir efeitos aleatórios de intercepto e inclinação, bem como a possível correlação entre estes. Pode ser que precisaremos também tratar da heterocedasticidade esperada nos dados longitudinais, uma vez que a variância dos erros pode variar com a idade (visitas) das participantes.

Para modelarmos essas variáveis em um modelo de efeitos mistos lineares com dados longitudinais, iniciaremos por um modelo completo da forma:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i5} \end{bmatrix} = \begin{bmatrix} A_{i1} & G_i & A_{i1}G_i \\ A_{i2} & G_i & A_{i2}G_i \\ \vdots & \vdots & \vdots \\ A_{i5} & G_i & A_{i5}G_i \end{bmatrix} \begin{bmatrix} \tau \\ \beta \\ (\tau, \beta) \end{bmatrix} + \begin{bmatrix} 1 & A_{i1} \\ 1 & A_{i2} \\ \vdots & \vdots \\ 1 & A_{i5} \end{bmatrix} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i5} \end{bmatrix}$$

em que A_{ij} é a idade da participante i na visita j , G_i é um indicador binário (dummy) que vale 1 se a participante i foi suplementada com cálcio e 0 caso contrário, $A_{ij}G_i$ é a interação entre idade e grupo, b_{0i} é o efeito aleatório de intercepto da participante i , b_{1i} é o efeito aleatório de inclinação da participante i e ϵ_{ij} é o erro aleatório da observação ij . Ademais,

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b0}^2 & \sigma_{b0b1} \\ \sigma_{b0b1} & \sigma_{b1}^2 \end{bmatrix} \right)$$

e

$$\boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i5} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 |A_{i1}|^\delta & 0 & \dots & 0 \\ 0 & \sigma^2 |A_{i2}|^\delta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 |A_{i5}|^\delta \end{bmatrix} \right)$$

Dessa forma, podemos modear o efeito aleatório de intercepto e inclinação, bem como a correlação entre estes. O modelo trata a heterocedasticidade esperada nos dados longitudinais, modelando a variância dos erros como uma função da idade das participantes, com parâmetro δ a ser estimado.

Partiremos deste modelo completo e o reduziremos conforme nossas análises progredirem.

Implementação e comparação dos modelos

Como estratégia, criaremos todos os modelos possíveis de efeitos aleatórios e ajuste para heterocedasticidade pela `varPower`, totalizando oito modelos. Faremos a comparação entre estes modelos utilizando os critérios AIC e BIC, selecionando o melhor modelo dentre estes. Então, partiremos para análise diagnóstica desse modelo. Caso não satisfaça os pressupostos, partiremos para o próximo modelo melhor classificado pelos critérios AIC e BIC, repetindo o processo até encontrarmos um modelo satisfatório.

Finalmente, com melhor modelo de efeitos aleatórios selecionado e melhor ajuste para heterocedasticidade, faremos a seleção do modelo de efeitos fixos, com interesse especial na inclusão ou não da interação entre idade e grupo. Finalmente, com o modelo finalizado, faremos a análise dos resíduos para verificação dos pressupostos do modelo e, se bem ajustado, decidiremos se há efeito significativo ou não da suplementação com cálcio na densidade mineral óssea dessa população.

Seleção do modelo de efeitos aleatórios e ajuste de heterocedasticidade

Tabela 2: Comparação dos modelos de efeitos aleatórios e ajuste de heterocedasticidade

Modelo	AIC	BIC
Completo com correlação	-2334.11	-2300.45
Completo com correlação e ajuste	-2265.52	-2223.44
Completo sem correlação	-2262.92	-2233.46
Completo sem correlação com ajuste	-2261.54	-2227.87
Sem inclinação	-2239.75	-2214.49
Sem inclinação com ajuste	-2239.23	-2209.77
Sem intercepto	-2251.85	-2226.60
Sem intercepto com ajuste	-2250.13	-2220.67
Sem efeito aleatório	-1269.87	-1248.78
Sem efeito aleatório com ajuste	-1233.83	-1208.58

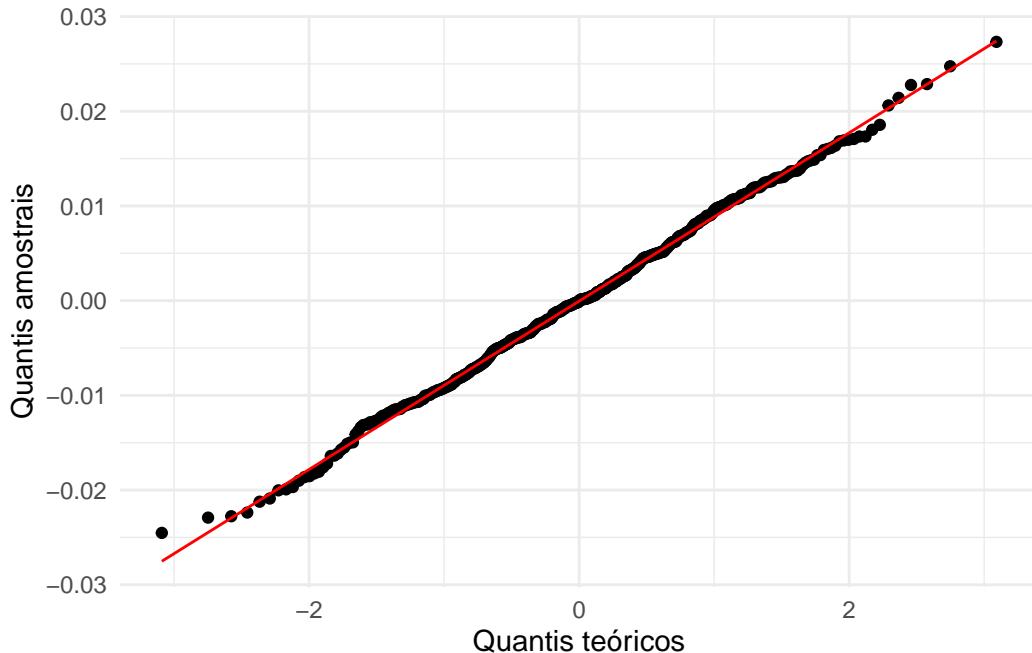
O modelo com correção ajustado não convergiu com o método `varPower`, por isso, utilizamos `varConstPower`.

Da Tabela 2, observa-se que o modelo completo com correlação e sem ajuste apresenta os menores valores de AIC e BIC, caracterizando-o como melhor candidato dentre os testados. Partiremos para a análise diagnóstica deste modelo para verificar se satisfaz os pressupostos.

Normalidade

Começamos pela análise gráfica via QQ-Plot:

Figura 5: Comparação dos quantis



O gráfico da Figura 5 aponta bom ajuste, vamos confirmar com teste de Shapiro-Wilk, com resultados em Tabela 3.

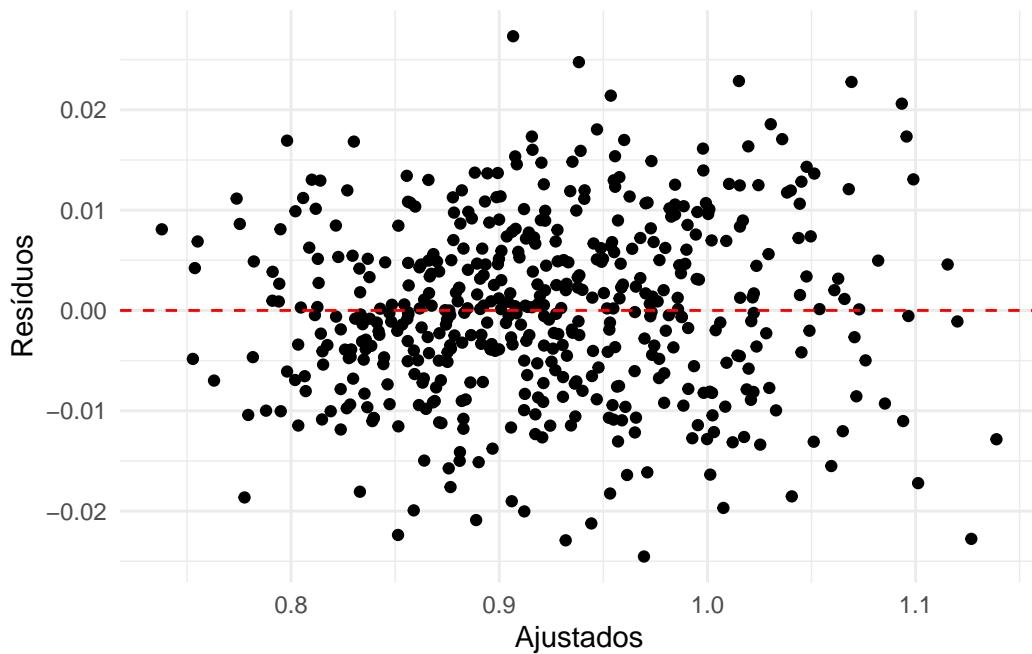
Tabela 3: Saída do teste de Shapiro-Wilk sobre os resíduos

Estatística	Valor-p
0.9978	0.7776

Com normalidade não rejeitada, partimos para teste da homocedasticidade.

Homocedasticidade

Figura 6: Resíduos x Ajustados do Modelo com Correlação



O gráfico da Figura 6 reafirma média dos resíduos próxima de zero (como visto no QQ-plot da Figura 5), e dá indícios de homocedasticidade, com resíduos distribuídos aleatoriamente em torno de zero, o que também reforça a independência obtida pela inclusão dos efeitos aleatórios.

Confirmaremos a homocedasticidade com o teste de Bartlett, com resultado na Tabela 4.

Tabela 4: Saída do teste de Bartlett para homocedasticidade entre os grupos (P x C)

Estatística	Valor-p
0.3235	0.5695

Podemos também realizar o teste de Levene, cujos resultados são apresentados na Tabela 5.

Tabela 5: Saída do teste de Levene para homocedasticidade entre os grupos (P x C)

Estatística	Valor-p
0.8616	0.3537
NA	NA

Concluímos destas estatísticas que não há evidências suficientes para rejeitarmos a homocedasticidade dos resíduos do modelo completo com correlação. Dessa forma, não será necessário considerar o modelo com ajuste para heterocedasticidade, e podemos partir para seleção dos efeitos fixos.

Seleção dos efeitos fixos

Da análise anterior, o modelo completo com correlação e sem ajuste foi selecionado como melhor candidato dentre os modelos de efeitos aleatórios testados, satisfazendo os pressupostos de normalidade e homocedasticidade dos resíduos, assim como a independência entre observações após tratamento com efeito aleatório. Partiremos para a seleção do modelo de efeitos fixos, testando a inclusão ou não da interação entre idade e grupo, e se existe efeito da suplementação com cálcio na densidade mineral óssea.

Primeiramente, conferimos um sumário de nosso modelo na Tabela 6.

Tabela 6: Sumário do modelo completo com correlação

Termo	Estimativa	EP	GL	t	Valor-p
Intecepto	0.2818	0.0255	387	11.05	0
Idade	0.0542	0.0022	387	24.64	0
Grupo(P)	0.0867	0.0357	110	2.43	0.0167
Idade:Grupo(P)	-0.0089	0.0031	387	-2.89	0.0041

Da Tabela 6, observa-se que o termo de interação entre idade e grupo é significativo ao nível de 5%, indicando que o efeito da suplementação com cálcio na densidade mineral óssea varia conforme a idade das participantes. De fato, por apresentar uma estimativa negativa, ocorre o que observamos na Análise Exploratória de Dados: a suplementação com cálcio é mais efetiva em crianças mais novas, com o efeito diminuindo conforme a idade aumenta.

Como não há parâmetros a serem removidos do modelo, este é nosso modelo final, mantendo a forma completa com correlação, mas sem ajuste para heterocedasticidade. Dessa forma, nosso modelo final é:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i5} \end{bmatrix} = \begin{bmatrix} A_{i1} & G_i & A_{i1}G_i \\ A_{i2} & G_i & A_{i2}G_i \\ \vdots & \vdots & \vdots \\ 1 & A_{i5} & A_{i5}G_i \end{bmatrix} \begin{bmatrix} \tau \\ \beta \\ (\tau, \beta) \end{bmatrix} + \begin{bmatrix} 1 & A_{i1} \\ 1 & A_{i2} \\ \vdots & \vdots \\ 1 & A_{i5} \end{bmatrix} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i5} \end{bmatrix}$$

em que A_{ij} é a idade da participante i na visita j , G_i é um indicador binário (dummy) que vale 1 se a participante i foi suplementada com cálcio e 0 caso contrário, $A_{ij}G_i$ é a interação entre idade

e grupo, b_{0i} é o efeito aleatório de intercepto da participante i , b_{1i} é o efeito aleatório de inclinação (slope, coeficiente angular) da participante i e ϵ_{ij} é o erro aleatório da observação ij . Ademais,

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b0}^2 & \sigma_{b0b1} \\ \sigma_{b0b1} & \sigma_{b1}^2 \end{bmatrix} \right)$$

e

$$\boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i5} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \right)$$

Com esse modelo, podemos explicar a variação da densidade mineral óssea em função da idade, da suplementação com cálcio e da interação entre estes, possibilitando a predição de valores em cada visita, além de capturar a variação entre as participantes via efeitos aleatórios de intercepto e inclinação.

Podemos dizer que há evidências para eficácia da suplementação com cálcio na densidade mineral óssea dessa população, com efeito mais acentuado em crianças mais novas, conforme indicado pela interação significativa e negativa entre idade e grupo.