



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Bacharelado em Estatística

**PREDIÇÃO DO CUSTO DA
PRODUTIVIDADE DE ARROZ POR
MEIO DE MODELO MULTINIVEL**

Leticia Gabriela Ferreira Dias

Uberlândia-MG

2022

Leticia Gabriela Ferreira Dias

**PREDIÇÃO DO CUSTO DA
PRODUTIVIDADE DE ARROZ POR
MEIO DE MODELO MULTINIVEL**

Trabalho de conclusão de curso apresentado à Co-
ordenação do Curso de Bacharelado em Estatística
como requisito parcial para obtenção do grau de
Bacharel em Estatística.

Orientador: Prof. Dr. Janser Moura Pereira

Uberlândia-MG

2022



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, _____ de _____ de 20_____

BANCA EXAMINADORA

Prof. Dr. Janser Moura Pereira

Profa. Dra. Nádia Giareta Biase

Prof. Dr. Marcelo Tavares

**Uberlândia-MG
2022**

AGRADECIMENTOS

À Universidade Federal de Uberlândia (UFU) pela possibilidade da formação e realização de meus estudos.

À Deus por essa conquista, pela força e energia para concluir o curso.

À minha mãe Mônica por todo apoio necessário no decorrer do curso e não deixar-me desistir nunca.

Aos meus amigos, dentre os quais a Ivana, o Carlos, a Geovanna e a Brenda por toda ajuda e incentivo.

Aos meus professores, em especial ao meu orientador Janser Moura Pereira, pelo apoio, incentivo, ensinamento e paciência.

E a todos, que direta ou indiretamente fazem parte da minha formação.

RESUMO

Tendo em vista a importância da produção do arroz no país, sabendo de sua enorme significância para a exportação, o cálculo dos custos de produtividade se torna necessário para a maximização dos lucros. Assim, o presente trabalho tem como objetivo utilizar modelagem multinível para prever o custo total da Produtividade do arroz de alguns municípios brasileiros. Ou seja, a análise multinível permite conhecer as particularidades dos municípios acerca da Produtividade do arroz, considerando efeitos que podem impactar significativamente no custo da Produtividade, de forma a auxiliar os gestores na tomada de decisão. A pesquisa foi realizada com base nos dados levantados da CONAB (Companhia Nacional de Abastecimento) dos custos de Produtividade do arroz no período de 2016 a 2022. A base de dados é composta por 10 (dez) variáveis, na qual apenas 4 (quatro) foram selecionadas a compor o modelo final, sendo elas “Ano”, “Custo Atividade”, “Custo Variável” e “Tipo de Agricultura” escolhidas partir dos critérios de análise adotados. Em relação ao nível tempo, o modelo indica que há uma variação significativa do custo total, ou seja, o custo total segue uma tendência temporal linear. Com base no modelo final ajustado, observa-se que as variáveis que caracterizam os municípios auxiliam na explicação da variabilidade do custo total ao longo dos anos. Os resultados indicam que 46,9% da proporção da variação do custo total pode ser atribuído às características dos municípios.

Palavras-chave: HML2, HLM3, Step-up strategic, Estrutura hierárquica, Modelos Mistos.

ABSTRACT

Considering the rice's productivity importance of rice's in Brazil, knowing the huge significance to exportation, the producing cost estimate made profit maximization necessary. Thus, the present paperwork has as purpose use multilevel modeling to predict the total cost of productivity rice in some Brazilian municipalities. In other words, the multilevel analyze enables knowing specific features of producing rice, considering effects that significantly impacts the producing cost, in a way to help managers to take decisions. The search was performed based in CONAB (National Supply Company) dataset of all the producing rice costs over 2016-2021. This dataset is composed by 10 (ten) variables, where only 4 (four) were selected to the final model, they are: "Year", "Activity Cost", "Variable Cost" and "Agriculture Type" chosen by the adopted analyzes criteria. Relating to the time level, the model indicates the existence of a significant total cost variation, it means the total cost has linear temporal trend. Based in the final model, it is observed that the variables that characterize the municipalities assist in the explanation of the total cost variability over the years. The results indicates 46,9% of the total cost variation proportion as municipalities characteristics.

Keywords: HML2, HLM3, Step-up strategic, Hierarchical structure, Mixed Models.

SUMÁRIO

Lista de FigurasI

Lista de TabelasII

1	Introdução	1
2	Materiais E Métodos	3
2.1	Contextualização do Modelo Hierárquico Linear de três níveis – HLM3.4
2.2	Contextualização do Modelo Hierárquico Linear de dois níveis – HLM2.9
2.3	Step-up strategic – passos para construção do modelo final.10
2.4	Métodos de estimação dos coeficientes de um modelo multinível..12
2.5	Escolha do melhor modelo.13
2.6	Multicolinearidade.14
2.7	Coefficiente de correlação intraclasse – ICC.14
2.8	Autocorrelação dos resíduos.15
3	Resultados E Discussões	17
3.1	Análise dos dados.17
3.2	Análise de Regressão Multinível.27
3.2.1	Etapa 1.27
3.2.2	Etapa 2.28
3.2.3	Etapa 3.30
3.2.4	Etapa 4.36
3.2.5	Etapa 5.38
3.2.6	Etapa 6.41
4	Conclusões	44
	Referências Bibliográficas	45

LISTA DE FIGURAS

2.1	Anagrama da hierarquia de um modelo multinível.	4
3.1	Gráfico de Produtividade x custo total por município.	22
3.2	Gráfico de custo da atividade x custo total por município.	22
3.3	Gráfico de custo variável x custo total por município.	23
3.4	Gráfico de custo fixo x custo total por município.	23
3.5	Gráfico de renda fatores x custo total por município.	24
3.6	Gráfico de Produtividade x custo total por ano.	24
3.7	Gráfico de custo da atividades x custo total por ano.	25
3.8	Gráfico de custo variável x custo total por ano.	25
3.9	Gráfico de custo fixo x custo total por ano.	26
3.10	Gráfico de renda fatores x custo total por ano.	26
3.11	Gráfico de tendência ano x custo total por município.	27

LISTA DE TABELAS

2.1	Resumo da notação utilizada na modelagem com três níveis..	8
2.2	Descrição dos níveis e variáveis explicativas consideradas na análise multinível.	9
3.1	Resultados das medidas descritivas das variáveis quantitativas..	17
3.2	Resultados das médias, desvios padrões e coeficientes de variação das variáveis ao longo do período analisado..	17
3.3	Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao tipo de agricultura..	18
3.4	Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao tipo de agricultura e ano..	19
3.5	Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao município..	20
3.6	Matriz de correlação das variáveis que serão utilizadas na análise multinível.	21
3.7	Resultados de ajuste dos modelos nulos por meio de máxima verossimilhança restrita..	28
3.8	Comparação entre os modelos M3 e M4 ambos com estrutura de erros autorregressivos AR(1)..	29
3.9	Resultados do ajuste do modelo com variável ano e intercepto aleatório	29
3.10	Comparação entre os modelos HLM2 Nulo e M3 ambos com estrutura de erros autorregressivos AR(1)..	30
3.11	Tabela de correlação linear entre variável Custo Total e variáveis independentes	31
3.12	Comparação entre os modelos M4 e M5 ambos com estrutura de erros autorregressivos AR(1)..	31
3.13	Resultados do ajuste do modelo que considera a interação entre "Ano" e "Custo Atividade" com intercepto aleatório (Modelo M4), com estrutura de erros autorregressivos AR(1)..	32
3.14	Comparação entre os modelos M6 e M7 ambos com estrutura de erros autorregressivos AR(1)..	33
3.15	Resultados do ajuste do modelo sem a interação entre "Ano" e "Custo Atividade" com intercepto aleatório (Modelo M6) e com estrutura de erros autorregressivos AR(1)..	33
3.16	Comparação entre os modelos M3 e M6 ambos com estrutura de erros autorregressivos AR(1)..	34
3.17	Comparação entre os modelos M6 e demais modelos da etapa 3 todos com estrutura de erros autorregressivos AR(1)..	35
3.18	Comparação entre os modelos M6 e demais modelos da etapa 4 todos estrutura de erros autorregressivos AR(1)..	36
3.19	Resultados do ajuste do modelo sem a interação entre "Ano", "Custo Atividade", "Custo Variável" com intercepto aleatório (Modelo M12), com estrutura de erros autorregressivoscm AR(1)..	38

3.20	Comparação entre os modelos M12 e demais modelos da etapa 5 todos com estrutura de erros autorregressivos AR(1)..39
3.21	Resultados do ajuste do modelo sem a interação entre “Ano”, “Custo Atividade”, “Custo Variável”, “Tipo de Agricultura” com intercepto aleatório (Modelo M20), com estrutura de erros autorregressivos AR(1)..40
3.22	Comparação entre os modelos M20 e demais modelos da etapa 6 com estrutura de erros autorregressivos AR(1)..41
3.23	Estimativas dos coeficientes referentes ao ajuste do modelo final, contextualizado por municípios..43

1. INTRODUÇÃO

A necessidade de controlar custos, fez com que a sua apuração se tornasse de grande importância desde o início do capitalismo, pois era devido à contabilidade de custos que o comerciante tinha resposta se estava tendo lucro ou não. Neste contexto, a contabilidade de custos era usada como um instrumento seguro para controlar as variações de custos e de vendas e também para avaliar o crescimento ou o retrocesso do negócio [20].

De acordo com a ABIARROZ (Associação Brasileira da Indústria do Arroz), o Brasil representa 75% do arroz produzido pelo Mercado Comum do Sul (Mercosul), tornando-se assim o país fora da Ásia com maior produtividade de arroz do mundo, além de ter uma exportação média de 1,2 milhões de toneladas por ano, evidenciando a importância do cultivo do mesmo [19]. Devido à complexidade nas atividades que envolvem a Produtividade primária desse setor e as mudanças significativas dessas atividades durante os anos, já que de acordo com o diretor da ANA (Agência Nacional de Águas e Saneamento Básico), Marcelo Cruz, "No Brasil, 40% do volume de água captado para irrigação é destinado ao arroz. Com 90% de sua produtividade total sob irrigação, a cultura é chave tanto para as discussões de segurança hídrica quanto de segurança alimentar, de desenvolvimento regional e de recursos hídricos. O mapeamento dessas lavouras é fundamental para tomadas de decisão nas políticas agrícolas, de desenvolvimento regional e de recursos hídricos" [2], o interesse em ganhar competitividade mercadológica promoveu aos agentes privados o interesse pelo cálculo do custo de produtividade que, mesmo difícil de apurar devido a divergência dos sistemas de produtividade praticados, é necessário para a gestão profissional.

Portanto, fica evidente que tanto para a abertura quanto para tomada de decisões de um empreendimento, o custo de produtividade é um fator extremamente importante, devido a competitividade de mercado, para estudos de fabricação relacionados a viabilidade econômica e a análise de sobrevivência de uma companhia, com a finalidade de maximizar lucros e até mesmo impactar na diminuição do custo do produto para o cliente [17].

Destaca-se que os modelos multiníveis também são conhecidos como modelos hierárquicos lineares ou modelos mistos [8]. Os modelos hierárquicos lineares (HLM) ocupam-se de analisar efeitos aleatórios de indivíduos agrupados em níveis que causam variância significativa no resultado final. Um anagrama da hierarquia de um modelo multinível (HLM3 – modelo hierárquico linear com 3 níveis) pode-se perceber a presença de três níveis: municípios (nível 2) agrupados em unidades federativas (nível 3) além de levar em consideração medidas repetidas ao longo do tempo, o ano (nível 1) que agrupará os demais níveis apontados com a finalidade de aumentar

significativamente a precisão do custo previsto.

De acordo com Ferrão [6], o ajuste de modelos hierárquicos lineares de três níveis (HLM3) permitem:

- obtenção de estimativas eficientes dos coeficientes de regressão e de suas variâncias;
- tornam-se corretos erros-padrão, intervalos de confiança e testes de hipóteses (levando-se em conta informações sobre o agrupamento/contexto dos dados);
- variáveis explicativas mensuradas em cada nível da hierarquia (exploração mais detalhada do impacto e da contribuição de cada nível para a variabilidade da variável resposta);
- estabelecer listas comparativas das unidades de segundo nível.

Sendo assim, o trabalho possibilita avaliar as seguintes hipóteses:

- Hipótese 1: Há variabilidade no custo total do arroz ao longo dos seis anos analisados;
- Hipótese 2: Há variabilidade no custo total, ao longo dos seis anos analisados entre os municípios da mesma unidade federativa;
- Hipótese 3: Há variabilidade no custo total, ao longo dos seis anos analisados, entre os municípios provenientes de unidades federativas diferentes.

Dessa forma, o trabalho tem como objetivo identificar fatores que podem influenciar nos valores dos custos de produtividade do arroz no Brasil. Nesse contexto, a modelagem multinível é uma solução viável, pois permite identificar os efeitos aleatórios que contribuem para a mudança significativa dos custos, ou seja, permite avaliar se o custo da Produtividade se comporta de forma diferenciada entre os municípios e unidades federativas ao longo dos anos. Permitindo auxiliar os produtores da área a tomar decisões e melhorar a Produtividade, além de medir a capacidade de pagamento e definir a viabilidade econômica de uma tecnologia alternativa a partir do custo que poderá ser previsto com maior precisão.

2. MATERIAIS E MÉTODOS

Os dados que serão utilizados referem-se à Produtividade (kg/ha) e ao custo de Produtividade de arroz (R\$/ha), disponibilizados pela Companhia Nacional de Abastecimento [4]. A base contém informações sobre Produtividade (de agricultura familiar e empresarial) e os custos de Produtividade do arroz de 11 municípios distribuídos em 6 estados do Brasil, no período de 2016 a 2021. Nesta base de dados é possível obter informações sobre:

Período ou ano: 2016, 2017, 2018, 2019, 2020 e 2021;

Estados ou unidades federativas: Maranhão, Mato Grosso, Mato Grosso do Sul, Rio Grande do Sul, Roraima e Santa Catarina;

Municípios: Balsas - MA, Boa Vista - RR, Cachoeira do Sul - RS, Dourados - MS, Massaranduba - SC, Meleiro - SC, Pelotas - RS, Restinga Seca - RS, Santo Antônio da Patrulha - RS, Sorriso - MT e Uruguaiana - RS;

Tipo de agricultura utilizada na Produtividade: agricultura familiar ou agricultura empresarial;

X1: Produtividade do arroz em quilogramas por município em cada ano;

X2: custo da atividade (despesa de custeio da lavoura);

X3: custo variável (despesas com transporte externo, armazenagem, seguros, assistência técnica, impostos etc.);

X4: custo fixo (despesas de depreciações de máquinas e implementos, encargos sociais, manutenções etc.);

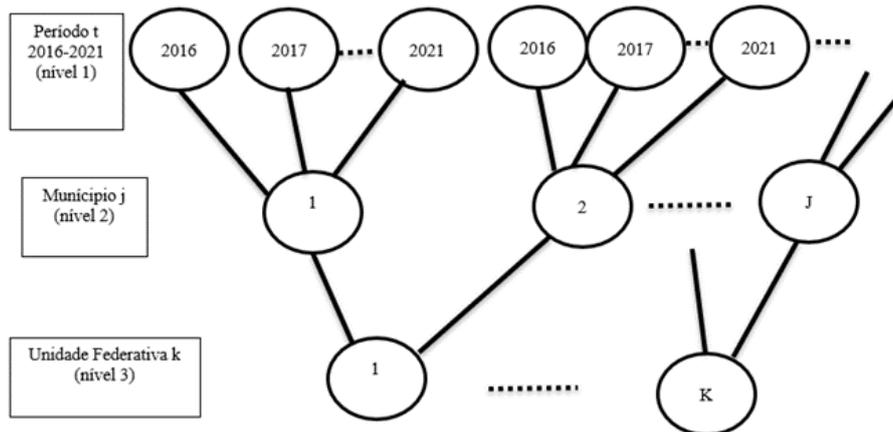
X5: renda de fatores (despesas remuneração esperada, terra e arrendamento);

Y: custo total.

Portanto, pretende-se nesse trabalho ajustar modelos multiníveis que estabeleça uma relação linear entre o custo total da Produtividade de arroz (variável dependente) em função de variáveis independentes como a Produtividade, o tipo de agricultura, o custeio da lavoura, entre outras despesas das cidades produtoras: Balsas - MA, Boa Vista - RR, Cachoeira do Sul - RS, Dourados - MS, Massaranduba - SC, Meleiro - SC, Pelotas - RS, Restinga Seca - RS, Santo Antônio da Patrulha - RS Sorriso - MT e Uruguaiana - RS. Ou seja, ajustar modelos hierárquicos lineares com a finalidade de encontrar, descrever e predizer o comportamento desses custos de forma que possa auxiliar gestores ligados ao setor no planejamento de investimentos. Cabe ressaltar que todas as análises estatísticas foram realizadas no freeware R [24].

Na Figura 2.1, é apresentado um anagrama da hierarquia de um modelo hierárquico de três níveis (HLM3), isto é, modelo multinível com três níveis. O modelo multinível determina cons-

tructos de pesquisa que consideram a existência de estruturas aninhadas de dados, em que determinadas variáveis apresentam variação entre unidades distintas que representam grupos, porém não entre observações pertencentes a um mesmo grupo [7]. Desse modo, a princípio, pode-se propor uma estrutura de dados em 3 níveis com medidas repetidas no tempo/ano (nível 1), aninhados em municípios (níveis 2), e estes em unidades federativas (nível 3), conforme Figura 1. No entanto, na próxima seção, será apresentado a contextualização do modelo multinível e conseqüentemente, justificar-se-á porque da escolha de um modelo hierárquico de dois níveis ao invés de três níveis.



Fonte: Elaborada pela autora.

Figura 2.1: Anagrama da hierarquia de um modelo multinível

2.1 CONTEXTUALIZAÇÃO DO MODELO HIERÁRQUICO LINEAR DE TRÊS NÍVEIS – HLM3

De acordo com [7] um modelo hierárquico de três níveis é composto por um submodelo de cada nível.

Logo, para o submodelo do nível 1, que apresenta "p" as variáveis explicativas (Z_1, \dots, Z_p) referentes às unidades i ($i = 1, \dots, n$), tem-se a expressão [9]:

$$Y_{ijk} = \pi_{0jk} + \sum_{(p=1)}^P \pi_{pjk} \cdot Z_{pjk} + e_{ijk} \quad (1)$$

Em que [9]:

Y_{ijk} : variável resposta para a i -ésima observação de nível 1 aninhada na unidade j do nível 2 e na unidade k do nível 3 ;

Z_{pjk} : p -ésima variável explicativa de nível 1 para a i -ésima observação aninhada na unidade j do nível 2 e na unidade k do nível 3 ;

π_{pjk} : coeficientes de nível 1, onde $p = 0, 1, \dots, P$;

e_{ijk} : termos de erro com distribuição normal, com média igual a zero e variância σ^2 .

Para o submodelo do nível 2, que apresenta "q" variáveis explicativas (X_1, \dots, X_q) referentes às unidades j ($j = 1, \dots, J$), tem-se a expressão [9]:

$$\pi_{pjk} = b_{p0k} + \sum_{(s=1)}^{Q_p} b_{pqk} \cdot X_{qjk} + r_{pjk} \quad (2)$$

Em que [9]:

π_{pjk} : coeficientes de nível 1, onde $p = 0, 1, \dots, P$;

X_{qjk} : q-ésima variável explicativa aninhada na unidade j do nível 2 e na unidade k do nível 3;

b_{pqk} : coeficientes de nível 2, onde $q = 0, 1, \dots, Q_p$;

r_{pjk} : efeito aleatório do nível 2, onde para cada j tem-se o vetor $(r_{0jk}, r_{1jk}, \dots, r_{pjk})'$ de distribuição normal multivariada com cada elemento possuindo média zero e variância $\tau_{\pi rpp}^2$.

Por fim, para o submodelo do nível 3, que apresenta "s" variáveis explicativas (W_1, \dots, W_s) referentes às unidades k ($k = 1, \dots, K$), tem-se a expressão [9]:

$$b_{pqk} = \gamma_{pq0} + \sum_{(s=1)}^{S_{pq}} \gamma_{pqs} \cdot W_{sk} + u_{pqk} \quad (3)$$

Em que [9]:

b_{pqk} : coeficientes de nível 2, onde $q = 0, 1, \dots, Q_p$;

W_{sk} : s-ésima variável explicativa de nível 3 aninhada na unidade k do nível 3;

γ_{pqs} : coeficientes de nível 3, onde $s = 0, 1, \dots, S_{pq}$;

u_{pqk} : efeito aleatório do nível 3, onde para cada k tem-se o vetor $(u_{0qk}, u_{1qk}, \dots, u_{pqk})'$ de distribuição normal multivariada com cada elemento possuindo média zero e variância $\tau_{\pi rpp}^2$.

De forma bastante simplista, a seguir será ilustrado a contextualização de um modelo hierárquico de três níveis com medidas repetidas no tempo, com intercepto e inclinação aleatórios, ou seja, o primeiro nível apresenta apenas a variável "período (Ano)" referentes às unidades t ($t = 1, \dots, T_j$) de nível 1; o segundo nível, uma única variável explicativa "X" referentes às unidades j ($j = 1, \dots, J$) de nível 2; e o terceiro nível, uma única variável explicativa "W" referentes às unidades k ($k = 1, \dots, K$) de nível 3 [9]:

- Nível 1: incluindo a variável período (Ano) do nível 1.

$$Y_{tjk} = \pi_{0jk} + \pi_{1jk} \cdot \text{período} + e_{tjk}; \text{ com } e_{tjk} \sim NID(0, \sigma^2) \quad (4)$$

Em que [9]:

Y_{tjk} : variável resposta no tempo t (tempo ou período) aninhada na unidade j do nível 2 e na unidade k do nível 3;

Período: variável explicativa de nível 1 para a observação t (tempo t) aninhada na unidade j do nível 2 e na unidade k do nível 3;

π_{pjk} : coeficientes de nível 1, onde $p = 0, 1$;

e_{tjk} : termos de erro com distribuição normal, com média igual a zero e variância σ^2 .

Nível 2: incluindo a variável X do nível 2.

$$\pi_{0jk} = b_{00k} + b_{01k} \cdot X_{jk} + r_{0jk}; \text{ com } r_{0jk} \sim NID(0, \tau_{\pi 0}^2) \quad (5)$$

$$\pi_{1jk} = b_{10k} + b_{11k} \cdot X_{jk} + r_{1jk}; \text{ com } r_{1jk} \sim NID(0, \tau_{\pi 1}^2) \quad (6)$$

Em que [9]:

π_{pjk} : coeficientes de nível 1, onde $p = 0, 1$;

X_{jk} : variável explicativa do nível 2 aninhada na unidade j do nível 2 e na unidade k do nível 3;

b_{pqk} : coeficientes de nível 2, onde $p = 0, 1$ e $q = 0, 1$;

r_{pjk} : efeito aleatório do nível 2, onde para cada j tem-se o vetor $(r_{0jk}, r_{1jk})'$ de distribuição normal bivariada : $r_{0jk} \sim NID(0, \tau_{\pi 0}^2)$ e $r_{1jk} \sim NID(0, \tau_{\pi 1}^2)$

Nível 3: incluindo a variável W do nível 3.

$$b_{00k} = \gamma_{000} + \gamma_{001} \cdot W_k + u_{00k}; \text{ com } u_{00k} \sim NID(0, \tau_b^2_{00}) \quad (7)$$

$$b_{01k} = \gamma_{010} + \gamma_{011} \cdot W_k + u_{01k}; \text{ com } u_{01k} \sim NID(0, \tau_b^2_{01}) \quad (8)$$

$$b_{10k} = \gamma_{100} + \gamma_{101} \cdot W_k + u_{10k}; \text{ com } u_{10k} \sim NID(0, \tau_b^2_{10}) \quad (9)$$

$$b_{11k} = \gamma_{110} + \gamma_{111} \cdot W_k + u_{11k}; \text{ com } u_{11k} \sim NID(0, \tau_b^2_{11}) \quad (10)$$

Em que [9]:

b_{pqk} : coeficientes de nível 2, onde $p = 0,1$ e $q = 0, 1$;

W_k : variável explicativa do nível 3 aninhada na unidade k do nível 3;

Y_{pqs} : coeficientes de nível 3, onde $p = 0,1$, $q = 0, 1$ e $s = 0,1$;

u_{pqk} : efeito aleatório do nível 3, onde para cada k tem-se o vetor $(u_{00k}, u_{01k}, u_{10k}, u_{11k})'$ de distribuição normal multivariada: $u_{00k} \sim NID(0, \tau_b^2_{00})$, $u_{01k} \sim NID(0, \tau_b^2_{01})$, $u_{10k} \sim NID(0, \tau_b^2_{10})$ e $u_{11k} \sim NID(0, \tau_b^2_{11})$.

Para um modelo final é feita a combinação das expressões (4) a (10) resultando em [9]:

$$Y_{tjk} = (\gamma_{000} + \gamma_{001} \cdot W_k + \gamma_{010} \cdot X_{jk} + \gamma_{011} \cdot W_k \cdot X_{jk} + u_{00k} + u_{01k} \cdot X_{jk} + r_{0jk})$$

intercepto com efeitos aleatórios

$$+ (\gamma_{100} + \gamma_{101} \cdot W_k + \gamma_{110} \cdot X_{jk} + \gamma_{111} \cdot W_k \cdot X_{jk} + u_{10k} + u_{11k} \cdot X_{jk} + r_{1jk}) \cdot \text{período}_{jk}$$

inclinação com efeitos aleatórios

$$+ e_{tjk}(11)$$

Em que [9]:

período_{jk} : período t aninhado na unidade j do nível 2 e na unidade k do nível 3;

X_{jk} : variável explicativa do nível 2 aninhada na unidade j do nível 2 e na unidade k do nível 3;

W_k : variável explicativa do nível 3 aninhada na unidade k do nível 3;

Y_{tjk} : variável resposta no período t aninhada na unidade j do nível 2 e na unidade k do nível 3;

γ_{000} : representa o valor esperado da variável dependente no instante inicial e quando $X = W = 0$ (intercepto geral);

γ_{001} : : representa o incremento no valor esperado da variável dependente no instante inicial (alteração no intercepto) para determinada unidade j de nível 2 pertencente a uma unidade k de nível 3 quando houver alteração unitária na característica W de k, mantendo as demais variáveis constantes.

γ_{010} : representa o incremento no valor esperado da variável dependente no instante inicial para determinada unidade jk quando houver alteração unitária na característica X de j, mantendo as demais variáveis constantes;

γ_{011} : representa o incremento no valor esperado da variável dependente no instante inicial para determinada unidade jk quando houver alteração unitária no produto W.X, mantendo as demais variáveis constantes;

u_{00k} e u_{01k} : termos de erro que indicam a existência de aleatoriedade nos interceptos sendo que o último incide sobre alterações na variável X;

γ_{100} : representa a alteração no valor esperado da variável dependente quando houver alteração unitária no período de análise (mudança na inclinação em razão da evolução temporal unitária), mantendo as demais variáveis constantes;

γ_{101} : representa a alteração no valor esperado da variável dependente em razão da evolução temporal unitária para determinada unidade jk quando houver alteração unitária na característica W, mantendo as demais variáveis constantes;

γ_{110} : representa a alteração no valor esperado da variável dependente em razão da evolução temporal unitária para determinada unidade jk quando houver alteração unitária na característica X, mantendo as demais variáveis constantes;

γ_{111} : representa a alteração no valor esperado da variável dependente em razão da evolução temporal unitária para determinada unidade jk quando houver alteração unitária na característica W.X, mantendo as demais variáveis constantes;

u_{10k} e u_{11k} : termos de erro que indicam a existência de aleatoriedade nas inclinações sendo que o último também incide sobre alterações na variável X.

e_{tjk} : termos de erro com distribuição normal, com média igual a zero e variância σ^2 , ou seja, $e_{tjk} \sim NID(0, \sigma^2)$.

r_{0jk} e r_{1jk} : são os efeitos aleatórios do nível 2, assumindo-se, para cada unidade j, que o vetor $(r_{0j}, r_{1j})'$ apresenta distribuição normal bivariada: $r_{0jk} \sim NID(0, \tau_{\pi 0}^2)$ e $r_{1jk} \sim NID(0, \tau_{\pi 1}^2)$.

$u_{00k}, u_{01k}, u_{10k}$ e u_{11k} : são os efeitos aleatórios do nível 3, assumindo-se, para cada unidade k, que o vetor $(u_{00k}, u_{01k}, u_{10k}, u_{11k})'$ apresenta distribuição normal multivariada: $u_{00k} \sim NID(0, \tau_b^2_{00}); u_{01k} \sim NID(0, \tau_b^2_{01}); u_{10k} \sim NID(0, \tau_b^2_{10})$ e $u_{11k} \sim NID(0, \tau_b^2_{11})$.

Conforme Fávero e Belfiore [9], a expressão (11) facilita a visualização da influência que o intercepto e a inclinação podem sofrer em função de termos aleatórios decorrentes da existência de comportamentos distintos da variável dependente ao longo do tempo para cada uma das unidades do nível 2 (distintas séries de tempo), e esse fenômeno pode ser decorrente das características dessas unidades, bem como das características dos grupos a que pertencem tais unidades.

Na Tabela 2.1 é apresentado um resumo sobre a notação que foi adotada para os coeficientes na modelagem multinível com três níveis.

Tabela 2.1: Resumo da notação utilizada na modelagem com três níveis.

Nível	Termos	Notação
1 (Ano)	Coeficiente	π
	Erro	e
	Variâncias (erros)	σ^2
2 (Município)	Coeficientes	b
	Erro	r
	Variâncias (erros)	τ_{π}^2
3 (Unidade Federativa)	Coeficientes	γ
	Erro	u
	Variâncias (erros)	τ_b^2

Fonte: Elaborada pela autora.

Além disso, cabe destacar que para se ajustar um modelo hierárquico de três níveis (HLM3), é necessário que cada nível possua no mínimo uma variável. Na Tabela 2.2 a seguir destacam-se as variáveis (explicativas) que foram consideradas na análise multinível e seus respectivos níveis. Cujas variáveis dependentes são o custo total que foi observado ao longo do período de 2016 a 2021.

Tabela 2.2: Descrição dos níveis e variáveis explicativas consideradas na análise multinível.

Nível	Variável	Fonte
Nível 1: Evolução temporal	Ano: 2016, 2017, 2018, 2019, 2020 3 2021	CONAB
Nível 2: Município	X1:Produtividade do arroz em quilogramas	
	X2: Custo atividade: despesa de custeio da lavoura	
	X3: Custo variável: despesas com transporte externo, armazenagem, seguros, assistência técnica, impostos etc.	
	X4: Custo fixo : despesas de depreciações de máquinas e implementos, encargos sociais, manutenções etc.	
	X5: Renda fatores: despesas remuneração esperada, terra e arrendamento.	
	Tipo de agricultura utilizada na Produtividade: agricultura familiar ou agricultura empresarial.	
Nível 3: Unidade Federativa	Não teve nenhuma variável aninhada nesse nível, pois tomou-se como referência apenas a base de dados da CONAB. No entanto, nada impediria buscar variáveis a partir de outras fontes que poderiam ser aninhadas ao nível 3. Além disso, a variância associada ao nível 3 é não significativa, o que inviabiliza a consideração de um modelos HML3.	

Fonte: Elaborada pela autora.

Portanto, percebe-se que embora consiga aninhar a base de dados em três níveis (mesmo não possuindo variáveis para esse nível), não será possível ajustar um modelo hierárquico de três níveis, pois a variância associada ao nível 3 é não significativa. Sendo assim, as hipóteses a serem testadas passam a ser:

H1: Há variabilidade no custo total do arroz ao longo dos seis anos analisados;

H2: Há variabilidade no custo total, ao longo dos seis anos analisados entre os municípios.

Logo, no presente trabalho será ajustado um modelo hierárquico de dois níveis (HLM2).

2.2 CONTEXTUALIZAÇÃO DO MODELO HIERÁRQUICO LINEAR DE DOIS NÍVEIS – HLM2

A seguir será ilustrado a contextualização de um modelo hierárquico de dois níveis com medidas repetidas no tempo (HLM2), com intercepto e inclinação aleatório, com o primeiro nível contendo apenas a variável “Ano” referentes às unidades t ($t = 1, \dots, T_j$) de nível 1; e o segundo nível, uma única variável explicativa X referentes às unidades j ($j = 1, \dots, J$) de nível 2 [9]:

- Nível 1: incluindo a variável período (Ano) do nível 1.

$$Y_{tj} = \pi_{0j} + \pi_{1j} \cdot \text{período} + e_{tj}; \text{ com } e_{tj} \sim NID(0, \sigma^2) \quad (12)$$

Em que [9]:

Y_{tj} : variável resposta aninhada no período t (nível 1) e na unidade j do nível 2;

π_{pj} : coeficientes de nível 1, onde $p = 0, 1$;

e_{tj} : termos de erro com distribuição normal, com média igual a zero e variância σ^2 .

- Nível 2: incluindo a variável X do nível 2.

$$\pi_{0j} = b_{00} + b_{01}.X_j + r_{0j}; \text{ com } r_{0j} \sim NID(0, \tau_{\pi 0}^2) \quad (13)$$

$$\pi_{1j} = b_{10} + b_{11}.X_j + r_{1j}; \text{ com } r_{1j} \sim NID(0, \tau_{\pi 1}^2) \quad (14)$$

Em que [9]:

π_{pj} : coeficientes de nível 1, onde $p = 0, 1$;

X_j : variável explicativa de nível 2 aninhada na unidade j do nível 2;

b_{pq} : coeficientes de nível 2, onde $p = 0, 1$ e $q = 0, 1$;

r_{0j} e r_{1j} : são os efeitos aleatórios do nível 2, assumindo-se, para cada unidade j , que o vetor $(r_{0j}, r_{1j})'$ apresenta distribuição normal bivariada: $r_{0j} \sim NID(0, \tau_{\pi 0}^2)$ e $r_{1j} \sim NID(0, \tau_{\pi 1}^2)$.

Para um modelo final é feita a combinação das expressões (12) a (14) resultando em [9]:

$$Y_{tj} = (b_{00} + b_{01}.X_j + r_{0j}) + (b_{10} + b_{11}.X_j + r_{1j}).periodo_j + e_{tj}. \quad (15)$$

$$Y_{tj} = b_{00} + b_{01}.X_j + b_{10}.periodo_j + b_{11}.X_j.periodo_j + r_{0j} + r_{1j}.periodo_j + e_{tj}. \quad (16)$$

Em que [9]:

Y_{tj} : variável resposta aninhada no período t (nível 1) e na unidade j do nível 2;

X_j : variável explicativa do nível 2 aninhada na unidade j do nível 2;

b_{pq} : coeficientes de nível 2, onde $p = 0, 1$ e $q = 0, 1$;

r_{0j} e r_{1j} : são os efeitos aleatórios do nível 2, assumindo-se, para cada unidade j , que o vetor $(r_{0j}, r_{1j})'$ apresenta distribuição normal bivariada: $r_{0j} \sim NID(0, \tau_{\pi 0}^2)$ e $r_{1j} \sim NID(0, \tau_{\pi 1}^2)$.

e_{tj} : termos de erro com distribuição normal, com média igual a zero e variância σ^2 .

2.3 STEP-UP STRATEGIC – PASSOS PARA CONSTRUÇÃO DO MODELO FINAL

A seguir será apresentado o passo a passo para a construção do modelo final, também denominado de modelo completo.

Fávero e Confortini [8] e Fávero e Belfiore [7] explicam que o modelo completo é representado em etapas, por equações que consideram os efeitos aleatórios, onde, primeiramente (passo 1),

testam-se os modelos sem as variáveis preditoras para estimar a partição da variância entre os dois níveis, denominados de modelos nulos, ou seja, é analisada a significância das mudanças no Custo total por meio do teste de qui-quadrado χ^2 . O teste consiste em avaliar se as variâncias são estatisticamente diferentes de zero. Se constatada significância das variâncias, a premissa multinível é atendida. Logo, pode-se prosseguir com o ajuste de modelos multiníveis. Na seção de resultados será apresentado os resultados acerca das significâncias das variâncias, por meio do teste de qui-quadrado.

O próximo passo (passo 2) consiste em ajustar modelos multiníveis, com a incorporação das variáveis preditoras de nível 1, no caso, a variável “Período (Ano)”. De acordo com Fávero e Confortini [8], ajustam-se dois modelos que incluem um componente de tendência no nível 1. O primeiro modelo sem efeitos aleatórios, testa apenas se o “Custo total” segue uma tendência temporal linear. Já o segundo modelo com efeitos aleatórios testa se há variância significativa da tendência do “Custo total” entre municípios ao longo do tempo.

Para verificar qual dos modelos apresenta melhor ajuste, foi adotado o Critério de Informação de Akaike (AIC), o Critério de Informação Bayesiano (BIC) e o Teste de Razão de Verossimilhanças (TRV). A regra de decisão baseada nos critérios AIC e BIC consiste em quanto menor for o valor do critério de informação, melhor é o modelo. Já o TRV é utilizado para comparação entre modelos aninhados diferindo somente na estrutura de efeitos fixos. A hipótese nula do teste TRV é que o modelo mais restrito (modelo aninhado) é adequado. Maiores detalhes sobre AIC, BIC e TRV serão apresentados nas próximas seções.

De acordo com Fávero e Confortini [8], a significância das mudanças individuais é testada de duas formas:

(i) Aplica-se o teste de χ^2 para comparar as variâncias entre o modelo com o efeito do ano (modelo de tendência linear) e o modelo nulo;

(ii) Aplica-se o teste t para os efeitos fixos e χ^2 para os componentes de variância. Se for observado significância do efeito fixo para “Ano”, sugere que o efeito do tempo é constante para todos os municípios. Já a inclusão de efeitos aleatórios possibilita a existência de variabilidade significativa no “Custo total”, ao longo do tempo, entre municípios.

No caso do presente estudo, observou-se por meio do AIC e do BIC que o modelo com efeitos aleatórios (intercepto aleatório) se apresentou com melhor ajuste, sendo este utilizado para o estudo.

Após essa verificação, a próxima etapa (passo 3) consiste na inclusão de variáveis preditoras do nível 2, e o critério de seleção das variáveis para compor o modelo é ponderado levando em conta a significância do parâmetro e também a questão da multicolinearidade. Uma vez definido o modelo dessa etapa, a próxima etapa (passo 4) consiste em ajustar/atualizar o modelo definido anteriormente considerando/incluindo a interação das variáveis selecionadas anteriormente com a variável “Ano”. Com a inclusão das interações existe grandes chances de termos problemas de multicolinearidade. Portanto, é importante avaliar de forma criteriosa a questão da multicolinearidade. Guardado todas as recomendações finalmente tem-se o modelo completo (modelo final), que também deve atender a premissa dos modelos multiníveis, ou seja,

deve-se verificar a significância das variâncias.

2.4 MÉTODOS DE ESTIMAÇÃO DOS COEFICIENTES DE UM MODELO MULTINÍVEL.

Como já mencionado anteriormente, os modelos multiníveis também são conhecidos como modelos hierárquicos lineares ou modelos mistos [8]. A nomenclatura de modelo misto se deve ao fato de que o modelo contém parâmetros de efeitos fixos em β e de efeitos aleatórios em u . A lei geral de formação de um modelo misto é especificada conforme a expressão a seguir [18]:

$$Y = X\beta + Zu + \varepsilon, \quad (17)$$

Em que:

Y caracteriza o vetor de observações da variável dependente Y ;

X caracteriza a matriz conhecida dos valores X_{ij} ;

β refere-se ao vetor de parâmetros de efeitos fixos desconhecidos;

ε representa o vetor não-observado de erros aleatórios gaussianos independente e identicamente distribuídos, ou seja, $\varepsilon \sim NID(0, I\sigma^2)$;

Z caracteriza a matriz de planejamento conhecida, podendo ser constituída de variáveis contínuas ou classificatórias (“dummy”);

u representa o vetor de parâmetros de efeitos aleatórios desconhecidos.

Além disso, tem-se que u e ε são normalmente distribuídos com média e variância, respectivamente [18]:

$$E \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ e } \text{Var} \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}. \quad (18)$$

Logo, percebe-se que os modelos mistos (modelos multiníveis e modelo hierárquico linear) permitem correlações e variâncias heterogêneas para os erros [18]. Portanto, fica evidente que a estimação no modelo misto é mais complexa do que no modelo linear clássico. Consequentemente, entende-se também que o método de mínimos quadrado ordinários não é o método mais apropriado. Uma alternativa seria o método de mínimos quadrados generalizados, pois possibilita a ponderação das correlações e heterogeneidades de variâncias. De acordo com a literatura, supondo que u e ε são normalmente distribuídos, a melhor aproximação baseia-se na verossimilhança [16]. Existem vários métodos de estimação pressupondo normalidade, no entanto, abordaremos apenas os métodos de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML).

No método de máxima verossimilhança (ML), as estimativas obtidas são sempre não negativas para os componentes da variância. O método ML não considera a perda da estimação dos efeitos fixos do modelo, e consequentemente, produz estimativas viesadas principalmente em

pequenas amostras ($n < 30$) [22]. Já no método de máxima verossimilhança restrita (REML), cada observação é dividida em duas partes independentes, uma refere-se aos efeitos fixos e outra aos efeitos aleatórios, de modo que a função densidade de probabilidade das observações seja dada pela soma das funções densidades de cada parte [3]. Para amostras maiores, a diferença entre as estimativas via ML e as estimativas REML não são significativas [22].

Contudo, destaca-se que o método de estimação adotado para estimação dos parâmetros no presente trabalho foi o de máxima verossimilhança restrita (REML).

2.5 ESCOLHA DO MELHOR MODELO

Conforme seção anterior, os parâmetros foram estimados por meio do método de máxima verossimilhança restrita. A comparação entre os modelos realizou-se por intermédio do Critério de Informação de Akaike (AIC), do Critério de Informação Bayesiano (BIC) e do Teste de Razão de Verossimilhanças (TRV), quando possível. A seguir é apresentado informações sobre cada critério.

O Critério de Informação de Akaike (AIC), proposto por Akaike [1], é dado por:

$$\text{AIC} = -2\log(L) + 2p, \quad (19)$$

em que L é a verossimilhança do modelo e p é o número de parâmetros do modelo.

O Critério de Informação Bayesiano (BIC), proposto por Schwarz [21], é dado por:

$$\text{BIC} = -2\log(L) + 2p \cdot \log(n), \quad (20)$$

em que L é a verossimilhança do modelo, p é o número de parâmetros do modelo e n é o número total de observações. A regra de decisão baseado nos critérios AIC e BIC consiste em quanto menor for o valor do critério de informação, melhor é o modelo. Os dois critérios são muito semelhantes, sendo o critério BIC mais sensível ao número de parâmetros incluídos no modelo, penalizando o que tem mais parâmetros [15].

O Teste de Razão de Verossimilhanças (Likelihood Ratio Test - TRV) é utilizado para comparação entre modelos aninhados diferindo somente na estrutura de efeitos fixos. A estatística do teste é:

$$\text{TRV} = 2\log(L_1/L_0) = 2[\log(L_1) - \log(L_0)]_a \sim \chi^2_{p_1 - p_0} \quad (21)$$

A estatística TRV segue distribuição assintótica de qui-quadrado com $(p_1 - p_0)$ graus de liberdade;

L_1 é a verossimilhança do modelo com mais parâmetros;

L_0 é a verossimilhança do modelo aninhado e $(p_1 - p_0)$ é diferença entre o número de parâmetros dos dois modelos.

A hipótese nula do teste TRV é que o modelo mais restrito (modelo aninhado) é adequado. A terminologia efeitos fixos refere-se as estimativas dos coeficientes das variáveis explicativas (independentes), enquanto que, o termo efeitos aleatórios (variáveis) refere-se as estimativas de variâncias e covariâncias do modelo ajustado [5].

2.6 MULTICOLINEARIDADE

Foi realizado o estudo de multicolinearidade, a partir da inclusão das variáveis explicativas no modelo multinível. Uma medida muito utilizada no estudo de multicolinearidade é o Fator de Inflação da Variância (VIF). O fator de inflação da variância é uma medida para analisar a magnitude da multicolinearidade dos termos do modelo. Um VIF menor que 5 indica uma correlação baixa desse preditor com outros preditores. Um valor entre 5 e 10 indica uma correlação moderada, enquanto valores VIF maiores que 10 são um sinal de correlação alta e não tolerável de preditores de modelo [10]. No entanto, deve ser ter muita atenção e não ficar preso aos valores de VIF referendados, pois se a multicolinearidade for um problema, o modelo demonstra que os preditores em questão não parecem estar associados de forma confiável com o resultado (estimativas baixas, altos erros padrão), mesmo que os parâmetros sejam significativos [14]. O problema de multicolinearidade torna a estimativa dos parâmetros imprecisa, por conta de um alto valor do erro padrão, o que não é conveniente estatisticamente [13][23]. Em síntese, o critério adotado para a eliminação/retirada da variável leva em consideração:

- (i) se a variável explicativa apresenta uma correlação significativa com outra(s) variável(is) explicativa(s);
- (ii) se os sinais das estimativas dos coeficientes (inclinações) do modelo em questão divergem dos sinais das correlações;
- (iii) se as estimativas são baixas e os erros padrões são altos, mesmo que os parâmetros sejam significativos;
- (iv) Valor de VIF acima de 5 ($VIF > 5$). Se isso ocorre, a variável deve ser retirada do modelo.

2.7 COEFICIENTE DE CORRELAÇÃO INTRACLASSE – ICC

Goldstein [11] afirma que o ICC também pode ser interpretado como a parcela da variância total devida aos grupos que definem cada nível, ou seja, o ICC representa a homogeneidade entre os indivíduos associados a um mesmo grupo e simultaneamente, a heterogeneidade entre grupos distintos. A seguir tem-se que:

1. Correlação intraclasse de nível 1 (tempo):

$$\rho_{(Ano)} = \frac{\sigma^2}{\tau_b^2 + \tau_\pi^2 + \sigma^2} \quad (22)$$

2. Correlação intraclasse de nível 2 (Município):

$$\rho_{(Mun.)} = \frac{\tau_{\pi}^2}{\tau_b^2 + \tau_{\pi}^2 + \sigma^2} \quad (23)$$

3. Correlação intraclasse de nível 3 (Unidade Federatiava):

$$\rho_{(UF)} = \frac{\tau_b^2}{\tau_b^2 + \tau_{\pi}^2 + \sigma^2} \quad (24)$$

Em que:

$\rho_{(Ano)}$ é o ICC do nível 1; $\rho_{(Mun.)}$ é o ICC do nível 2; $\rho_{(UF)}$ é o ICC do nível 3; τ_b^2 é a variância dos resíduos do nível macro (nível 3 - UF); τ_{π}^2 é a variância do nível intermediário (nível 2 - Município); σ^2 é a variância dos resíduos e_{tjk} do nível micro (nível 1 - Ano).

Se a variância no nível macro é não significativa e a variância do nível 2 é significativa, ajusta-se um modelo multinível com dois níveis. Logo, para o modelo multinível com dois níveis, pode-se definir [11]:

1. Correlação intraclasse de nível 1 (tempo):

$$\rho_{(Ano)} = \frac{\sigma^2}{\tau_{\pi}^2 + \sigma^2} \quad (25)$$

2. Correlação intraclasse de nível 2 (Município):

$$\rho_{(Mun.)} = \frac{\tau_{\pi}^2}{\tau_{\pi}^2 + \sigma^2} \quad (26)$$

Em que:

$\rho_{(Ano)}$ é o ICC do nível 1; $\rho_{(Mun.)}$ é o ICC do nível 2; τ_{π}^2 é a variância do nível intermediário (nível 2 - Município); σ^2 é a variância dos resíduos e_{tj} do nível micro (nível 1 - Ano).

2.8 AUTOCORRELAÇÃO DOS RESÍDUOS

Sabe-se que os dados foram coletados ao longo do tempo. Portanto, durante o ajuste dos modelos multiníveis com dois níveis, considerou-se uma estrutura de autocorrelação de primeira ordem para os resíduos, ou seja, admitiu-se que os erros são autocorrelacionados na forma de um processo autorregressivo estacionário de primeira ordem AR (1), e_{tj} , da seguinte forma [12]:

$$e_{tj} = \phi u_{tj-1} + \varepsilon_{tj} \quad (27)$$

em que ϕ é o parâmetro de autocorrelação e ε_t , o ruído branco. Assim,

$$E(\varepsilon_{tj}) = 0, \quad E(\varepsilon_{tj}^2) = \sigma_\varepsilon^2 \quad \text{e} \quad E(\varepsilon_{tj}\varepsilon_{(tj-h)}) = 0 \quad \text{se} \quad h \neq 0$$

e

$$-1 \leq \phi \leq 1.$$

Logo, o erro e_{tj} presente na equação (16) será conforme a expressão (27).

3. RESULTADOS E DISCUSSÕES

3.1 ANÁLISE DOS DADOS

No primeiro momento será apresentado algumas estatísticas descritivas acerca das variáveis presentes no banco de dados como propósito preliminar da análise exploratória dos dados. Na Tabela 3.1 são apresentados estatísticas descritivas das variáveis quantitativas presentes na base de dados.

Tabela 3.1: Resultados das medidas descritivas das variáveis quantitativas.

Variável	Mínimo	Máximo	Média	Mediana	Desvio Padrão	CV
Produtividade (kg/ha)	2.400	8.000	6.502	7.225	1.796,30	27,63%
Custo Atividade (R\$/ha)	1.213	5.466	3.613	3.871	1.113,41	30,82%
Custo Variável (R\$/ha)	200,4	1.473,90	837,6	860,5	374,15	44,67%
Custo Fixo (R\$/ha)	39,44	3.471,69	915,29	559,54	895,95	97,89%
Renda Fatores (R\$/ha)	61,01	1.561,09	544,57	504,11	351,37	64,52%
Custo Total (R\$/ha)	2.170	10.108	5.910	6.183	2.036,91	34,47%

Fonte: Elaborada pela autora.

Apartir da Tabela 3.1, as variáveis que apresentaram maior dispersão reativa foram respectivamente: “Custo Fixo”, “Renda Fatores” e “Custo Variável”.

Na Tabela 3.2, são apresentadas as médias, desvios padrões e coeficientes de variação das variáveis ao longo do período analisado.

Tabela 3.2: Resultados das médias, desvios padrões e coeficientes de variação das variáveis ao longo do período analisado.

Ano	Estatística	Custo Total (R\$/ha)	Produtividade (R\$/ha)	Custo Atividade (R\$/ha)	Custo Variável (R\$/ha)	Custo Fixo (R\$/ha)	Renda Fatores (R\$/ha)
2016	Média	5.050,41	6.382	3.287,31	752,26	394,67	616,17
	Desvio Padrão	1.618,29	1.837	1.043,29	334,91	222,40	395,35
	CV	32,04%	28,78%	31,74%	44,52%	56,35%	64,16%
2017	Média	5.063,56	6.400	3.307,47	797,03	369,87	589,20
	Desvio Padrão	1.618,59	1.842	1.050,53	360,29	191,41	269,14
	CV	31,97%	28,78%	31,76%	45,20%	51,75%	45,68%
2018	Média	5.130,61	6.400	3.323,99	760,35	384,05	662,22
	Desvio Padrão	1.558,77	1.842	1.062,75	361,98	201,51	244,60
	CV	30,38%	28,78%	31,97%	47,61%	52,47%	36,94%

2019	Média	5.965,59	6.609	3.706,50	849,50	611,36	798,23
	Desvio Padrão	2.020,13	1.892	1.206,99	409,34	384,50	418,91
	CV	33,86%	28,63%	32,56%	48,19%	62,89%	52,48%
2020	Média	6.437,98	6.609	3.644,17	877,33	1.487,58	428,91
	Desvio Padrão	1.959,31	1.892	1.047,14	390,45	723,80	250,58
	CV	30,43%	28,63%	28,74%	44,50%	48,66%	58,42%
2021	Média	7.813,36	6.609	4.407,42	988,99	2.244,23	172,72
	Desvio Padrão	2.194,52	1.892	1.080,23	416,46	1.049,33	122,41
	CV	28,09%	28,63%	24,51%	42,11%	46,76%	70,87%

Fonte: Elaborada pela autora.

A partir da Tabela 3.2 pode-se observar que há uma tendência de aumento na média dos custos totais anualmente mesmo sem tanto aumento na média da “Produtividade”, como esperado em condições normais, sendo esse aumento maior no período de 2019 a 2021, sendo possível questionar se essa alavancagem advém da influência do Covid 19 ou até mesmo por conta da inflação no período mencionado. Também fica evidente maiores valores de desvio padrão nos anos de 2019, 2020, sendo possível observar que os custos que mais impactaram no aumento são os custos com despesa de custeio da lavoura, despesas de depreciações de máquinas e implementos, encargos sociais, e manutenções, ou seja, o “Custo de atividade” e o “Custo Fixo”.

Na Tabela 3.3, são apresentadas as médias, desvios padrões e coeficientes de variação das variáveis quantitativas a partir do tipo de agricultura utilizada na Produtividade (Empresarial ou Familiar).

Tabela 3.3: Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao tipo de agricultura.

Variável	Agricultura Empresarial			Agricultura Familiar		
	Média	Desvio Padrão	CV	Média	Desvio Padrão	CV
Produtividade (kg/ha)	6.265,39	2.255,64	36%	6.842,59	652,76	9,54%
Custo Atividade (R\$/ha)	3.581,26	1.279,57	35,73%	3.658,38	838,19	22,91%
Custo Variável (R\$/ha)	888,42	406,55	45,76%	764,13	314,54	41,16%
Custo Fixo (R\$/ha)	877,70	796,08	90,7%	969,59	1.037,02	106,95%
Renda Fatores (R\$/ha)	470,01	284,85	60,09%	646,50	414,48	64,11%
Custo Total (R\$/ha)	5.821,39	2.139,06	36,75%	6.038,61	1.912,10	31,67%

Fonte: Elaborada pela autora.

Na Tabela 3.3 percebe-se que a agricultura familiar apresenta maior média de Produtividade (Kg/ha), quando comparado com a agricultura empresarial. Em contrapartida, a agricultura familiar apresenta maiores valores médios de despesas, com exceção aos valores médios destinados às despesas com transporte externo, armazenagem, seguros, assistência técnica, e impostos (custo variável). Também é possível observar maiores valores de desvios padrões concentrados

na agricultura empresarial, o que implica em uma maior variabilidade do volume de Produtividade e do Custo Total no cultivo empresarial do que no cultivo familiar.

Na Tabela 3.4, são apresentadas as médias, desvios padrões e coeficientes de variação das variáveis quantitativas a partir do tipo de agricultura utilizada na Produtividade (Empresarial ou Familiar) ao longo do tempo.

Tabela 3.4: Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao tipo de agricultura e ano.

Tipo de Agricultura	Variáveis	Estatísticas	Anos						
			2016	2017	2018	2019	2020	2021	
Empresarial	Produtividade (kg/ha)	Média	6.278,57	6.278,57	6.278,57	6.250	6.250	6.250	
		Desvio Padrão	2.303,96	2.303,96	2.303,96	2.553,23	2.553,23	2.553,23	
		CV	36,7%	36,7%	36,7%	40,85%	40,85%	40,85%	
	Custo Atividade (R\$/ha)	Média	3.393,18	3.419,18	3.442,30	3.646,30	3.426,50	4.241,61	
		Desvio Padrão	1.276,21	1.308,28	1.340,18	1.520,34	1.256,16	1.311,46	
	Custo Variável (R\$/ha)	Média	833,78	866,86	852	896,46	902,33	997,86	
		Desvio Padrão	376,33	403,28	396,12	469,06	454,94	504,35	
	Custo Fixo (R\$/ha)	Média	487,35	468,62	488,58	578,89	1.404,12	2.036,75	
		Desvio Padrão	118,09	99,92	109,79	109,29	757,80	1.214,77	
	Renda Fatores (R\$/ha)	Média	571,32	633,91	580,14	582,75	288,18	127,18	
		Desvio Padrão	220,27	242,24	225,87	379,61	297,17	41,33	
	Custo Total (R\$/ha)	Média	5.285,64	5.388,58	5.363,02	5.704,39	6.021,12	7.403,39	
		Desvio Padrão	1.792,00	1.910,29	1.926,28	2.307,63	2.316,49	2.715,14	
			CV	33,9%	35,45%	35,92%	40,45%	38,47%	36,67%
	Familiar	Produtividade (kg/ha)	Média	6.562,50	6.612,50	6.612,50	7.040	7.040	7.040
			Desvio Padrão	747,64	775	775	610,74	610,74	610,74
			CV	11,39%	11,72%	11,72%	8,68%	8,86%	8,86%
		Custo Atividade (R\$/ha)	Média	3.102,03	3.111,96	3.116,95	3.778,74	3.905,38	4.606,39
Desvio Padrão			546,65	418,18	287,69	860,73	782,61	822,54	
Custo Variável (R\$/ha)		Média	609,59	674,81	599,95	793,15	847,32	978,36	
		Desvio Padrão	219,09	275,90	262,77	369,58	346,93	339,68	
Custo Fixo (R\$/ha)		Média	232,47	197,06	201,13	650,33	1.587,73	2.493,21	
		Desvio Padrão	286,11	198,95	202,80	543,84	754,26	875,28	
Renda Fatores (R\$/ha)		Média	694,66	510,94	805,85	1.056,81	597,78	227,36	
		Desvio Padrão	641,14	333,57	232,75	324,56	267,17	168,76	
Custo Total (R\$/ha)		Média	4.638,78	4.494,78	4.723,88	6.279,03	6.938,21	8.305,32	
		Desvio Padrão	1.397,12	869,76	575,75	1.822,28	1.521,89	1.506,72	
			CV	30,12%	19,35%	12,19%	29,02%	21,93%	18,14%

Fonte: Elaborada pela autora.

Já na Tabela 3.4 é possível manter a percepção, agora mais clara ao agrupar por anos, de maior média de Produtividade (kg/ha) na agricultura familiar ao se comparar com a agricultura empresarial. Porém, diferentemente da Tabela 3 consegue-se observar uma menor média de custos na agricultura familiar, sendo a média geral impactada pelo aumento que ocorreu no período de 2019 a 2021, ao contrário da agricultura empresarial que não sofreu um aumento tão drástico na média de custos neste mesmo período. Também é possível observar a maior dispersão relativa dos custos de “Renda Fatores” na agricultura familiar no último ano e “Custo Fixo” na agricultura familiar nos primeiros anos indicando a heterogeneidade dos dados dessas variáveis.

Na Tabela 3.5, são apresentadas as médias, desvios padrões e coeficientes de variação das variáveis quantitativas por Município.

Tabela 3.5: Resultados das médias, desvios padrões e coeficientes de variação das variáveis quantitativas em relação ao município.

Município	Estatísticas	Custo Total (R\$/ha)	Produtividade (kg/ha)	Custo Atividade (R\$/ha)	Custo Variável (R\$/ha)	Custo Fixo (R\$/ha)	Renda Fatores (R\$/ha)
Balsas	Média	2.493,92	2.400	1.499,80	226,19	518,21	249,72
	Desvio Padrão	492,76	0	408,99	36,43	150,43	10649
	CV	19,76%	0,00%	27,27%	16,11%	29,03%	42,64%
Boa Vista	Média	5.531,20	8.000	4.064,65	827,12	409,43	230
	Desvio Padrão	495,07	0	494,27	28,25	51,72	79,22
	CV	8,95%	0,00%	12,16%	3,42%	12,63%	34,44%
Cachoeira do Sul	Média	8.165,49	7.250	4.687,81	1.150,56	1.588,23	738,90
	Desvio Padrão	1.595,39	54,77	480,37	119,18	1.232,30	359,12
	CV	19,54%	0,76%	10,25%	10,36%	77,59%	48,60%
Dourados	Média	5.048,73	5.750	3.404,44	633,75	662,17	348,38
	Desvio Padrão	1.885,60	273,86	1.013,89	204,76	746,29	250,92
	CV	37,35%	4,76%	29,78%	32,31%	112,70%	72,02%
Masaranduba	Média	5.423,48	7.100	3.316,62	347,24	971,97	787,66
	Desvio Padrão	1.532,34	438,18	471,29	74,60	1.391,80	286,88
	CV	28,25%	6,17%	14,21%	21,48%	143,19%	36,42%
Meleiro	Média	6.173,35	7.325	3.876,53	937,10	687,59	672,13
	Desvio Padrão	1.846,02	82,16	813,74	164,49	848,50	508,51
	CV	29,90%	1,12%	20,09%	17,55%	123,40%	75,66%
Pelotas	Média	7.605,50	7.925	4.668,29	1.270,19	1.145,29	521,74
	Desvio Padrão	1.210,35	82,16	446,75	60,73	1.098,70	250,58
	CV	15,91%	1,04%	9,57%	4,78%	95,93%	48,03%
Restinga Seca	Média	5.732,08	6.967	3.304,59	892,30	758,52	776,67
	Desvio Padrão	522,18	81,65	171,55	97,95	465,93	468,52
	CV	9,11%	1,17%	5,19%	10,98%	61,43%	60,32%
Santo Antônio da Patrulha	Média	7.599,24	7.200	4.334,17	1.227,68	1.370,17	667,21
	Desvio Padrão	1.516,50	328,63	515,31	159,53	1.189,27	331,80
	CV	19,96%	4,56%	11,89%	12,99%	86,80%	49,73%

	Média	3.875,55	3.600	2.116,90	471,37	921,83	365,46
Sorriso	Desvio Padrão	843,45	0	380,94	50,21	591,68	172,15
	CV	21,76%	0,00%	18,00%	10,65%	64,19%	47,11%
	Média	7.364,21	8.000	4.467,12	1.299,82	1.034,83	632,44
Uruguaiana	Desvio Padrão	943,06	0	233,97	122,10	973,56	310,47
	CV	12,81%	0,00%	5,24%	9,93%	94,08%	49,09%

Fonte: Elaborada pela autora.

A partir da Tabela 3.5, é possível observar que há uma variação expressiva na média de custo total entre os municípios, sendo que Balsas foi a cidade que apresentou menor média e desvio R\$2.493,92 e R\$492,76, respectivamente. Enquanto as maiores médias de custos totais se concentram em Pelotas, Santo Antônio da Patrulha e Uruguaiana: R\$7.605,50, R\$7.599,24 e R\$7.364,21, respectivamente. Outro ponto que merece atenção são os altos valores dos coeficientes de variação para as variáveis “Custo Fixo” e “Renda Fatores”, apenas o município Boa Vista apresentou menor dispersão relativa para “Custo Fixo”.

No Tabela 3.6 a seguir apresenta-se a matriz de correlação e os valores-p das correlações entre as variáveis que serão utilizadas na análise multinível. Acima da diagonal principal tem-se as estimativas dos coeficientes de correlação linear de Pearson e abaixo da diagonal principal tem-se os valores-p do teste de correlação linear de Pearson.

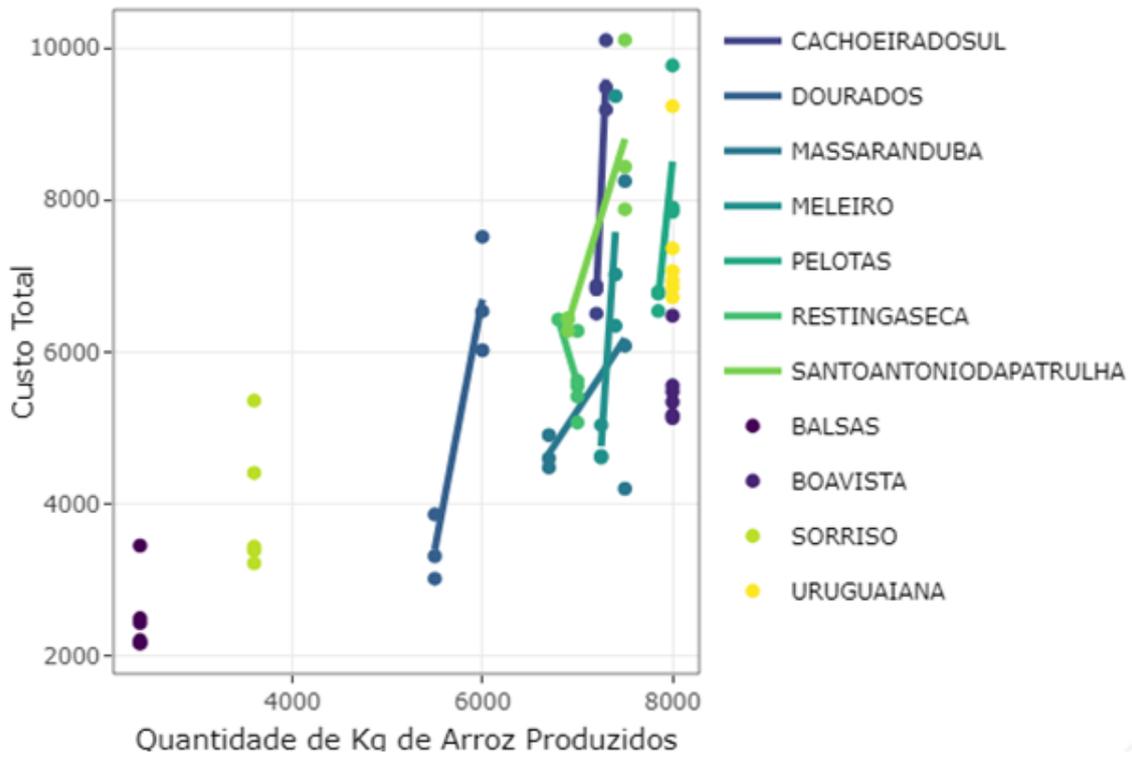
Tabela 3.6: Matriz de correlação das variáveis que serão utilizadas na análise multinível

	Ano	Produtividade	Custo Atividade	Custo Variável	Custo Fixo	Renda Fatores	Tipo Agricultura	Custo Total
Ano	1	0,0540	0,3088	0,1989	0,7040	-0,3585	0,0812	0,4532
Produtividade	0,6667	1	0,8424	0,7315	0,1734	0,2918	0,1592	0,7214
Custo Atividade	0,01164	<0,000	1	0,8677	0,4515	0,1514	0,0343	0,9307
Custo Variável	0,1093	<0,000	<0,000	1	0,4013	0,1305	-0,1646	0,8570
Custo Fixo	<0,000	0,1638	0,0001	0,0008	1	-0,3217	0,0508	0,7049
Renda Fatores	0,0031	0,0175	0,225	0,2964	0,0084	1	0,2432	0,1377
Tipo agricultura	0,5169	0,2017	0,7844	0,1867	0,6853	0,0491	1	0,0528
Custo Total	0,0002	<0,000	<0,000	<0,000	<0,000	0,2701	0,6735	1

Fonte: Elaborada pela autora.

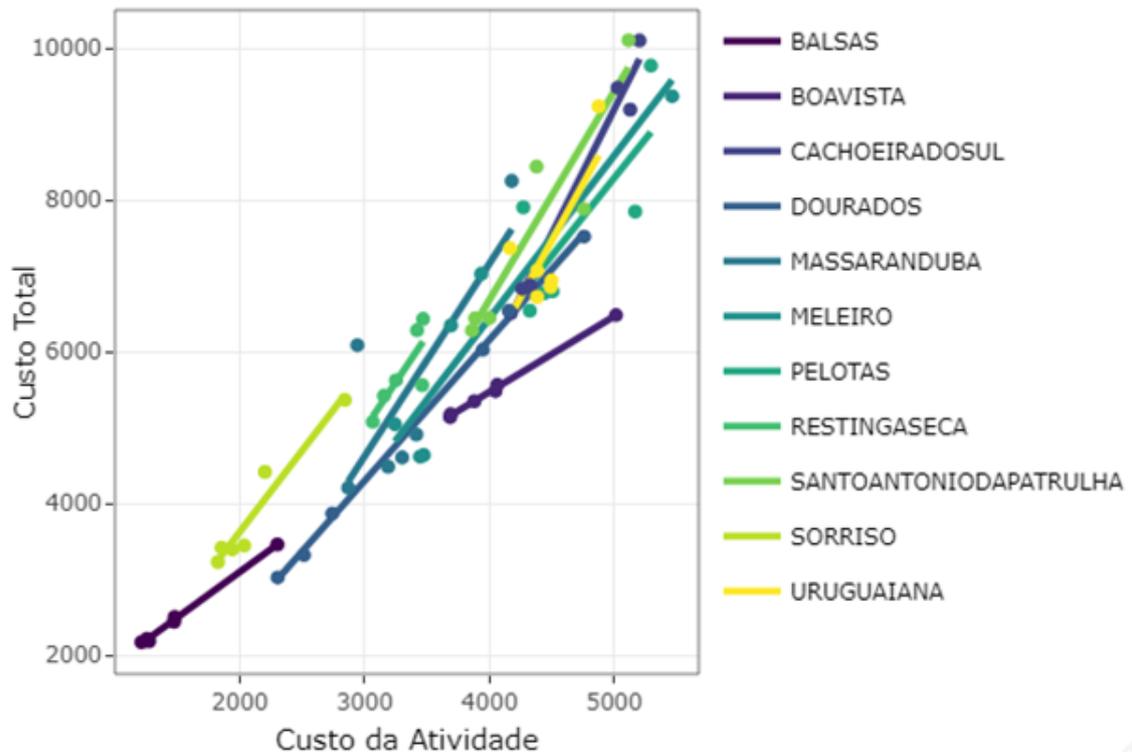
Pode-se observar no Tabela 3.6 que há uma correlação de mais de 80% entre as variáveis “Produtividade” e “Custo Atividade”, “Custo Atividade” e as variáveis “Custo Variável” e “Custo Total” e “Custo Variável” e “Custo Total”, enquanto as menores correlações estão entre as variáveis “Ano” e as variáveis “Produtividade” e “Tipo de Agricultura”, “Custo Atividade” e “Tipo de Agricultura” e entre “Custo Fixo” e “Tipo de Agricultura”.

Nas Figuras 3.1, 3.2, 3.3, 3.4 e 3.5, são apresentados os gráficos das variáveis independentes versus custo total por município com o objetivo de averiguar comportamento heterocedástico do nível 2.



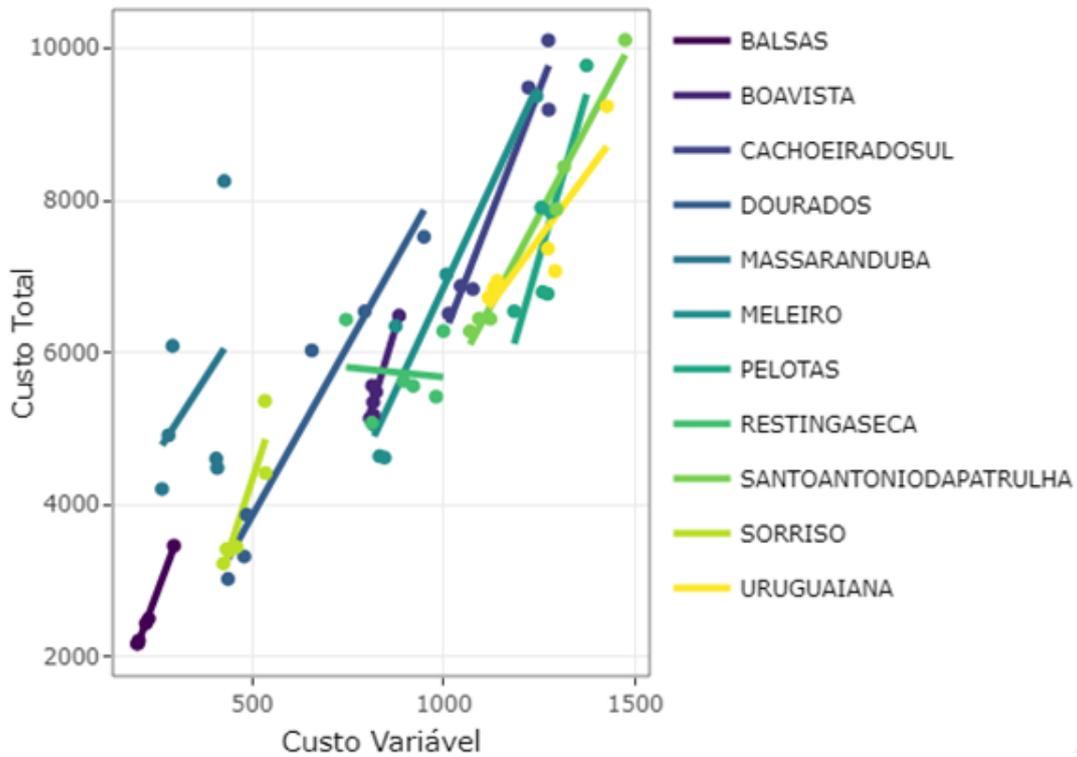
Fonte: Elaborada pela autora.

Figura 3.1: Gráfico de Produtividade x custo total por município



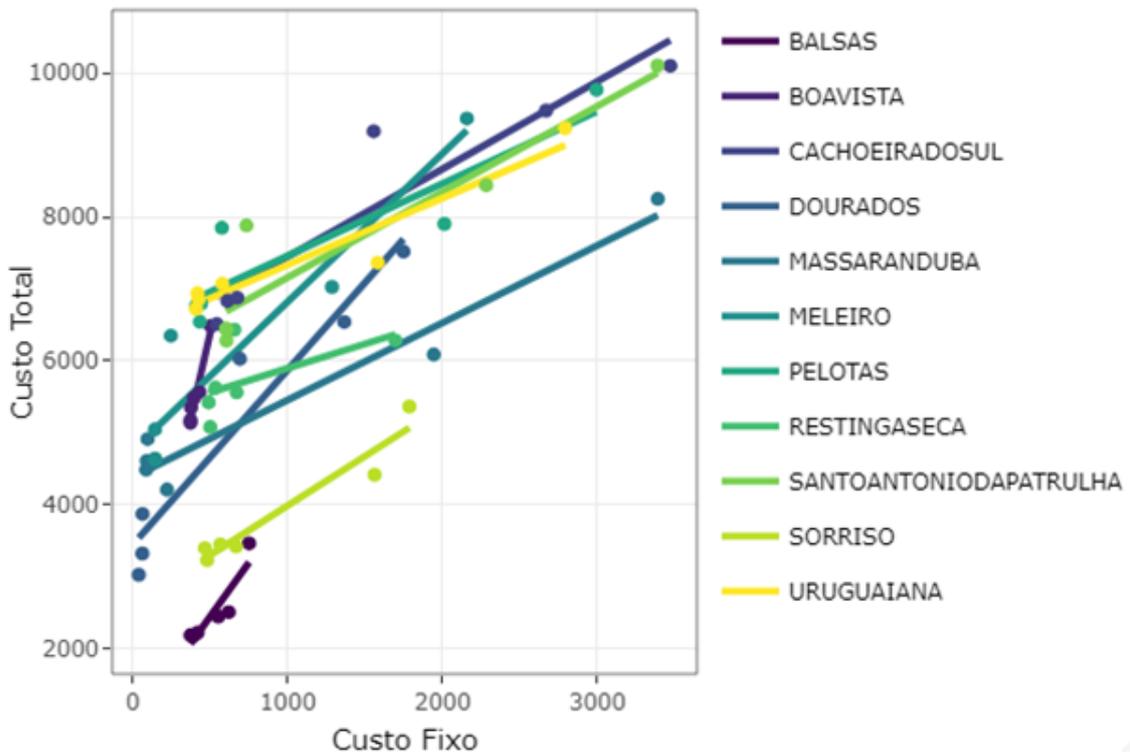
Fonte: Elaborada pela autora.

Figura 3.2: Gráfico de custo da atividade x custo total por município



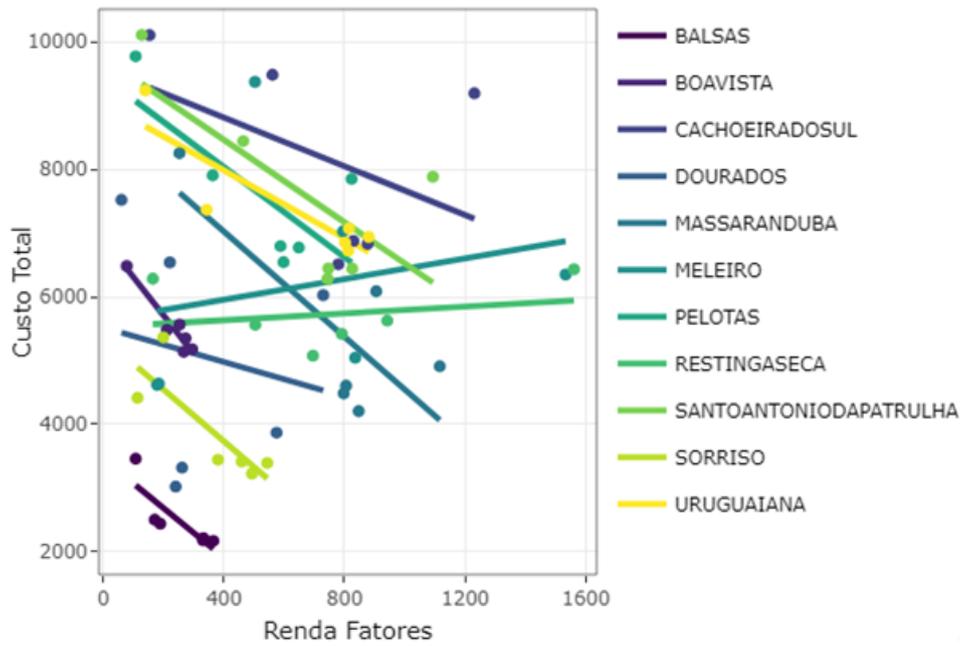
Fonte: Elaborada pela autora.

Figura 3.3: Gráfico de custo variável x custo total por município



Fonte: Elaborada pela autora.

Figura 3.4: Gráfico de custo fixo x custo total por município

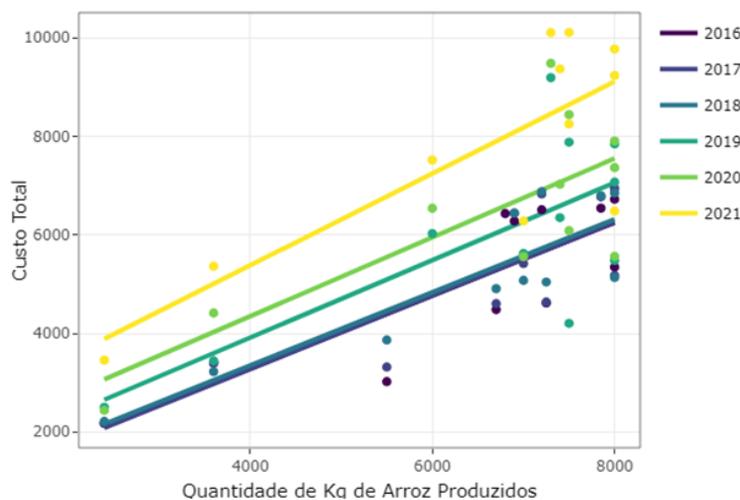


Fonte: Elaborada pela autora.

Figura 3.5: Gráfico de renda fatores x custo total por município

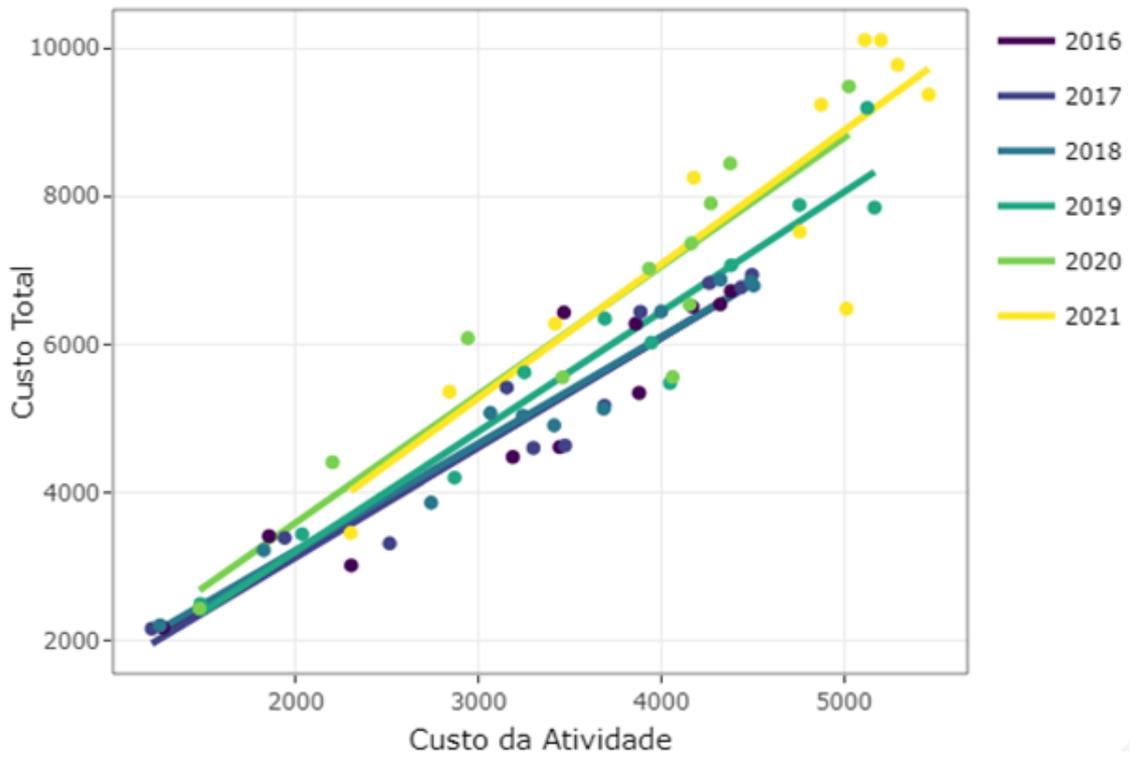
As Figuras 3.2, 3.3, 3.4 e 3.5 evidenciam, no contexto município, que as variáveis custo atividade, custo variável, custo fixo e renda fatores, aparentemente apresentam a mesma inclinação variando apenas o intercepto. Quanto a variável Produtividade deve-se ter cautela pois, a variação é pequena ou não existe variação dependendo do município, Figura 3.1.

Agora, nas Figuras 3.6, 3.7, 3.8, 3.9 e 3.10, são apresentados os gráficos das variáveis independentes x custo total por ano com o objetivo de averiguar comportamento heterocedástico do nível 1.



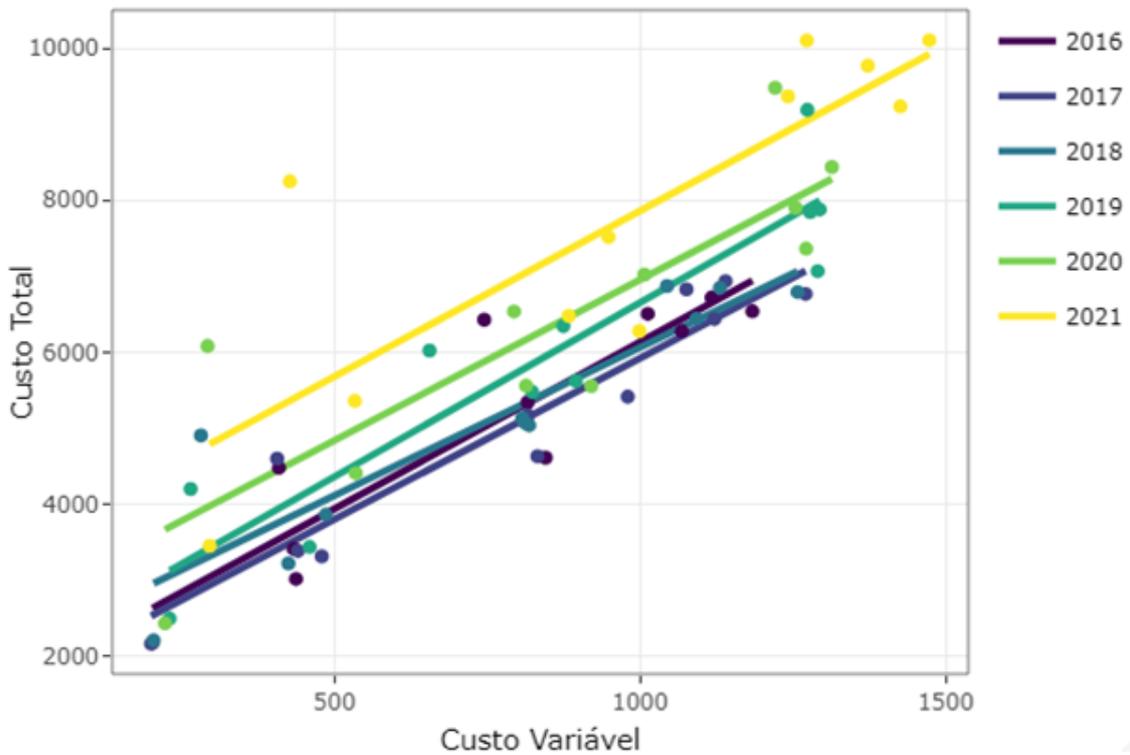
Fonte: Elaborada pela autora.

Figura 3.6: Gráfico de Produtividade x custo total por ano



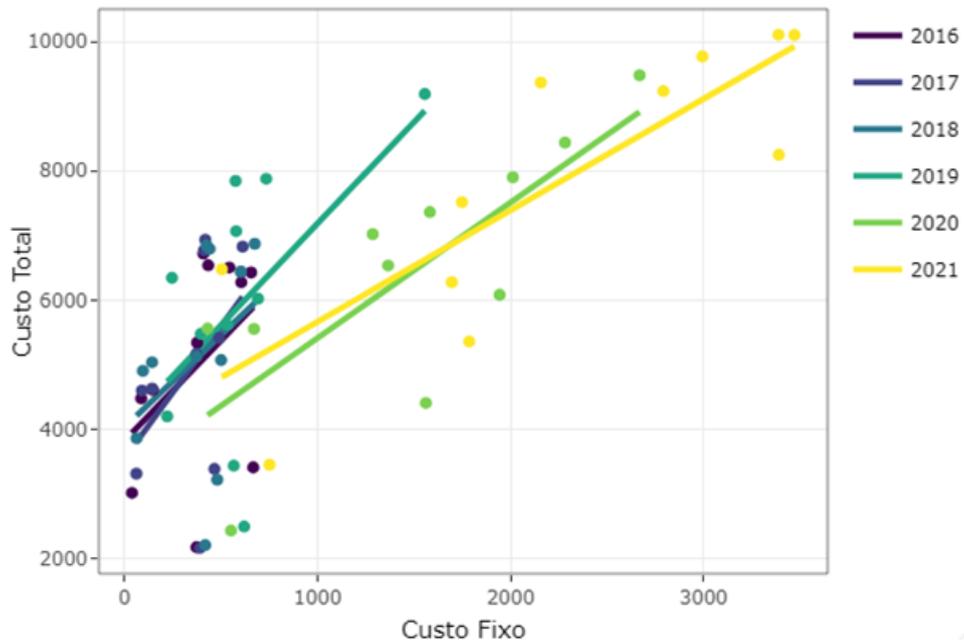
Fonte: Elaborada pela autora.

Figura 3.7: Gráfico de custo da atividades x custo total por ano



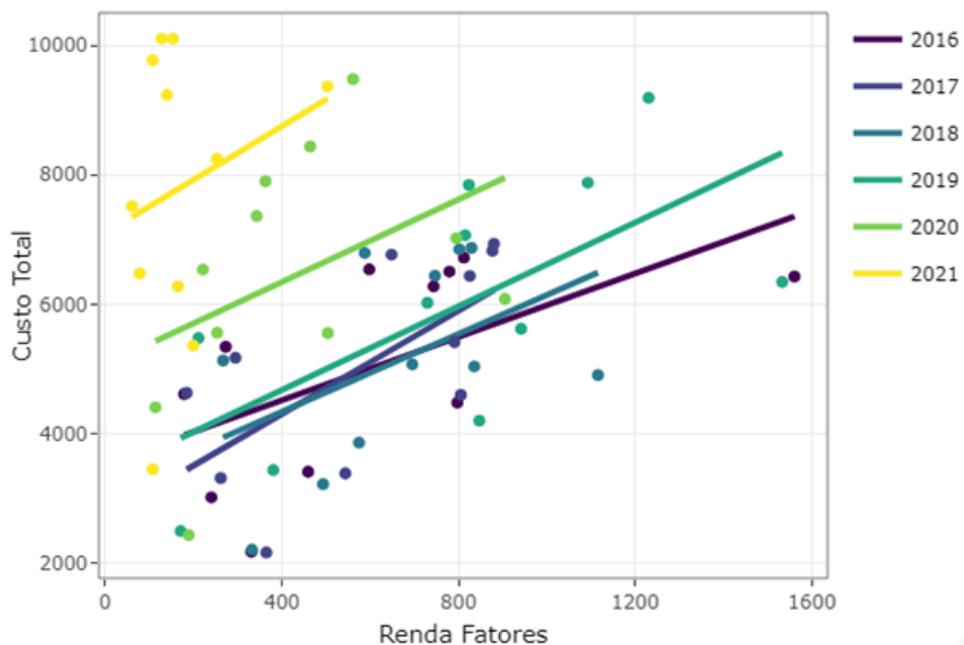
Fonte: Elaborada pela autora.

Figura 3.8: Gráfico de custo variável x custo total por ano



Fonte: Elaborada pela autora.

Figura 3.9: Gráfico de custo fixo x custo total por ano



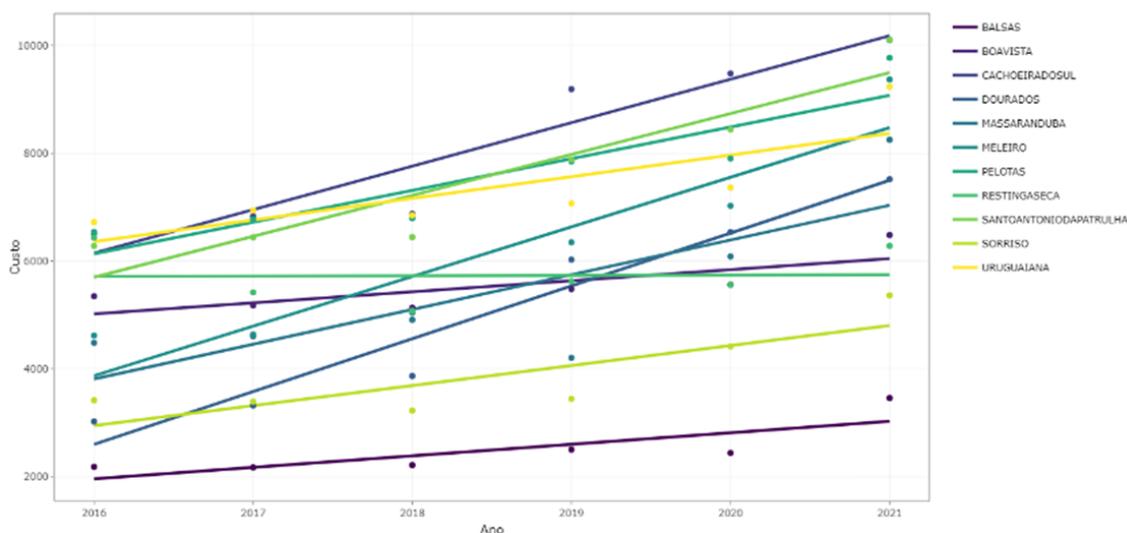
Fonte: Elaborada pela autora.

Figura 3.10: Gráfico de renda fatores x custo total por ano

Nas Figuras 3.6, 3.7, 3.8, 3.9 e 3.10, no contexto ano, as variáveis Produtividade, custo variável e renda aparentemente apresentam a mesma inclinação variando apenas o intercepto. Já as variáveis custo atividade e custo fixo aparentemente apresentam a mesma inclinação e o mesmo intercepto. Portanto, a partir das figuras 3.1 a 3.10 tem-se uma direção quanto à característica do intercepto e da inclinação no modelo multinível, isto é, foi possível perceber de

forma geral e informal que aparentemente apenas o intercepto será aleatório.

Finalmente, na Figura 3.11 é apresentado o gráfico do custo total por município ao longo dos anos, com o objetivo de averiguar comportamento do custo total ao longo dos anos.



Fonte: Elaborada pela autora.

Figura 3.11: Gráfico de tendência ano x custo total por município

De forma geral na Figura 4.11 pode-se perceber para a maioria dos municípios, que os custos totais apresentam inclinações (tendências) iguais e interceptos distintos ao longo dos anos.

3.2 ANÁLISE DE REGRESSÃO MULTINÍVEL

3.2.1 ETAPA 1

Para enfatizar a importância da análise multinível, realiza-se inicialmente o ajuste do modelo vazio (modelo nulo), somente com intercepto aleatório, isto é, sem qualquer variável explicativa. Uma vez proposto o ajuste do modelo nulo com intercepto aleatório deve-se verificar se a premissa da análise multinível é atendida. Nesse procedimento de análise do modelo multinível estima-se a variância dos erros aleatórios em cada nível de hierarquia. Se constatada significância das variâncias, a premissa multinível é atendida. No entanto, também será proposto o ajuste de um modelo nulo clássico (sem efeito aleatório) para confrontar qual é o melhor modelo.

Outro ponto que merece destaque é que como as medidas são tomadas ao longo do tempo, então, será considerado uma estrutura de erros autocorrelacionados de primeira ordem (AR(1)).

Na Tabela 3.7, tem-se os resultados obtidos a respeito da estimação dos dois modelos nulos (com e sem intercepto aleatório).

Tabela 3.7: Resultados de ajuste dos modelos nulos por meio de máxima verossimilhança restrita.

Modelo nulo com intercepto aleatório com AR(1) - (HLM2 Nulo)				
Efeito fixo				
Coeficiente	Estimativa	Erro Padrão	Estatística t	
Intercepto	5.950,38	522,33	11,39	
Efeito variáveis				
Níveis	Variância	Desvio Padrão	Valor-p	ICC
Intercepto (Nível Município)	2.664.922	1.632,46	0,047	0,65
Resíduo	1.457.040	1.207,08	< 0,000	–
Modelo nulo via Mínimos Quadrado - (OLS Nulo)				
Coeficiente	Estimativa	Erro Padrão	Estatística t	
Intercepto	5.910,30	250,7	23,57	
Teste da razão de verossimilhança – TRV				
OLS Nulo versus HLM2 Nulo			Estatística χ^2	Valor-p
			61,58	< 0,000

Fonte: Elaborada pela autora.

Tem-se na Tabela 3.7, ao nível de significância de 5%, por intermédio do teste da razão de verossimilhança que o modelo nulo com intercepto aleatório (HLM2 Nulo) é melhor que o modelo nulo sem efeito aleatório (OLS Nulo). Além disso, ao nível de significância de 5%, verifica-se que as variâncias do modelo HLM2 Nulo são estatisticamente diferentes de zero. Portanto, a premissa do modelo multinível é atendida. Outro ponto que merece destaque é o coeficiente de correlação intraclasse (ICC). O ICC indica que aproximadamente 65% da proporção da variação do custo total pode ser atribuída às características dos municípios.

3.2.2 ETAPA 2

Prosseguindo com a análise multinível, será incluído uma variável por vez do nível 1. Nesse caso, por se tratar de medidas repetidas no tempo, será incluída apenas a variável Ano no nível 1. Portanto, tem-se duas situações possíveis (dois modelos possíveis): (i) modelo com variável Ano e intercepto aleatório (M3); (ii) modelo com variável Ano com intercepto aleatório e inclinação aleatória (M4). A comparação dos modelos será por meio do AIC, BIC e do teste da razão de verossimilhança.

Primeiramente, serão apresentados na Tabela 3.8, os resultados de comparação dos dois

modelos conforme AIC, BIC e teste da razão de verossimilhança e depois os resultados do ajuste do melhor modelo será apresentado na Tabela 3.9.

Tabela 3.8: Comparação entre os modelos M3 e M4 ambos com estrutura de erros autorregressivos AR(1).

Modelo	Avaliadores do ajuste	
	AIC	BIC
Modelo com Ano e intercepto aleatório (M3)	1.079,13	1.087,77
Modelo com Ano, intercepto e inclinação aleatórios (M4)	1.083,09	1.096,05
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p
M3 vs M4	0,04	0,97

Fonte: Elaborada pela autora.

De acordo com a Tabela 3.8, ao nível de significância de 5%, não rejeita-se hipótese nula do teste da razão de verossimilhança, isto é, o modelo mais restrito é adequado. Logo, o modelo M3 é a escolha mais apropriada. Já a regra de decisão baseada nos critérios AIC e BIC consiste na escolha do modelo que apresentar menores valores de critério de informação. Portanto, quanto aos critérios de informação, o modelo M3 apresenta melhor ajuste quando comparado com o modelo M4.

A seguir será apresentado na Tabela 3.9, os resultados do ajuste do modelo M3, isto é, Modelo com a variável Ano e intercepto aleatório com estrutura de erros autoregressivos AR(1).

Tabela 3.9: Resultados do ajuste do modelo com variável ano e intercepto aleatório .

Modelo com a variável Ano e intercepto aleatório (M3) com estrutura de erros AR(1)				
Efeito fixo				
Coeficiente	Estimativa	Erro Padrão	Estatística t	Valor-p
Intercepto	-1.083.631,40	123.192,26	-8,79	< 0,000
Ano	539,8	61,03	8,84	< 0,000
Efeito variáveis				
Níveis	Variância	Desvio Padrão	Valor-p	ICC
Intercepto (Nível Município)	2.861.303	1.691,53	< 0,000	0,83
Resíduo	606.054	778,49	< 0,000	-
Avaliadores do ajuste	AIC	BIC		
	1.079,13	1.087,77		

Fonte: Elaborada pela autora.

A partir da Tabela 3.9, ao nível de significância de 5%, verifica-se que as variâncias do modelo M3 são estatisticamente diferentes de zero. Portanto, a premissa do modelo multinível

continua sendo atendida. Já o ICC indica que aproximadamente 83% da proporção da variação do custo total pode ser atribuída às características dos municípios. Em relação a variável Ano, o modelo indica que há uma variação significativa do custo total, ou seja, o custo total segue uma tendência temporal linear positiva. Esse resultado converge ao resultado encontrado na matriz de correlação. Agora, o próximo passo será confrontar o melhor modelo da etapa anterior com o melhor modelo dessa etapa, ou seja, agora deve-se comparar os modelos HLM2 Nulo como o modelo M3. Os dois modelos serão avaliados por meio do AIC, BIC e do teste da razão de verossimilhança.

Na Tabela 3.10, são apresentados os resultados de comparação dos modelos HLM2 Nulo e M3 por meio do AIC, do BIC e do teste da razão de verossimilhança.

Tabela 3.10: Comparação entre os modelos HLM2 Nulo e M3 ambos com estrutura de erros autorregressivos AR(1).

Modelo	Avaliadores do ajuste	
	AIC	BIC
Modelo Nulo com intercepto aleatório (HLM2 Nulo)	1.136,44	1.142,96
Modelo com Ano e intercepto aleatório (M3)	1.079,13	1.087,77
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p
HLM2 Nulo vs M3	59,3	< 0,000

Fonte: Elaborada pela autora.

Baseado na Tabela 3.10, ao nível de significância de 5%, rejeita-se hipótese nula do teste da razão de verossimilhança, isto é, o modelo mais restrito não é adequado. Logo, o modelo M3 é a escolha mais apropriada. Já a regra de decisão baseada nos critérios AIC e BIC consiste na escolha do modelo que apresentar menores valores de critério de informação. Portanto, quanto aos critérios de informação, o modelo M3 apresenta melhor ajuste quando comparado com o modelo HLM2 Nulo.

Portanto, encerra-se esta etapa e conclui-se que o modelo M3 é o melhor modelo ajustado, até o momento.

3.2.3 ETAPA 3

Agora nessa etapa, será adicionado uma variável por vez do nível 2 ao modelo M3, ou seja, será determinado qual é a primeira variável do nível 2 que será incorporada ao modelo M3. Essa escolha será ponderada a partir da correlação linear de Pearson entre as variáveis preditoras e a variável dependente “custo total” para analisarmos por onde começarmos o processo. Na Tabela 3.11 são apresentadas as estimativas das correlações lineares entre Custo Total e as variáveis preditoras.

Tabela 3.11: Tabela de correlação linear entre variável Custo Total e variáveis independentes

Variáveis	Correlação
Custo Total x Custo Atividade	0,93
Custo Total x Custo Variável	0,86
Custo Total x Produtividade	0,72
Custo Total x Custo Fixo	0,71
Custo Total x Renda Fatores	0,14
Custo Total x Fixo	0,05

Fonte: Elaborada pela autora.

A partir da Tabela 3.11, tem-se em ordem crescente das correlações lineares entre Custo Total e as variáveis preditoras, sendo assim a ordem com a qual serão propostos os ajustes dos modelos mais adiante.

Vale ressaltar que nessa etapa continua-se admitindo que os erros estão correlacionados conforme a estrutura de autocorrelação de primeira ordem (AR(1)). O primeiro modelo a ser ajustado nessa etapa irá conter a variável “Ano” (nível 1), a variável “Custo Atividade” (nível 2) e também a interação entre elas. Assim como nos ajustes dos modelos apresentados anteriormente, a comparação de todos os modelos será por meio do teste da razão de verossimilhança, do AIC e BIC. Também é importante destacar/adiantar que a interação da variável “Ano” com todas as variáveis explicativas do nível 2 não permaneceram em nenhum modelo, seja por conta da não significância do parâmetro associada a interação ou por conta de “problemas” de multicolinearidade. Além disso, tem-se a questão de poder considerar em todos os modelos as duas possibilidades: (i) modelo com intercepto aleatório; e (ii) modelo com intercepto aleatório e inclinação aleatória. Essa etapa será a última etapa do processo de ajuste de modelos multi-níveis e demanda um trabalho muito árduo. Portanto, para elucidar o processo nesse primeiro momento será destacado de forma detalhada todos os ajustes de modelos necessários que envolvem as variáveis “Ano” e “Custo Atividade”. Depois os demais resultados serão apresentados de forma sintetizada.

Portanto, será apresentado no primeiro instante, os resultados do teste de razão de verossimilhança dos modelos com a variável “Ano”, “Custo Atividade” e “Ano:Custo Atividade”, considerando as duas possibilidades: (i) modelo com intercepto aleatório (M4); e (ii) modelo com intercepto aleatório e inclinação aleatória (M5), conforme Tabela 3.12.

Tabela 3.12: Comparação entre os modelos M4 e M5 ambos com estrutura de erros autorregressivos AR(1).

Modelo com a variável Ano, Custo Atividade e Ano: Custo Atividade	Avaliadores do ajuste	
	AIC	BIC
Com intercepto aleatório(M4)	1.006,63	1.019,39
Com intercepto aleatório e inclinação aleatório (M5)	1.010,60	1.027,61
Teste da Razão de Verossimilhança M4 vs M5	Estatística χ^2 0,03	Valor-p 0,98

Fonte: Elaborada pela autora.

Pode-se concluir pelo teste da razão da razão de verossimilhança e/ou por meio do AIC e BIC que o modelo mais “adequado” é o modelo M4 quando comparado com o modelo M5. No entanto, para esse modelo deve-se verificar a questão da premissa multinível e se há problemas de multicolinearidade. Portanto, na Tabela 3.13, tem-se os resultados a respeito da estimação do modelo com interação entre Ano e Custo Atividade com intercepto aleatório (modelo M4).

Tabela 3.13: Resultados do ajuste do modelo que considera a interação entre “Ano” e “Custo Atividade” com intercepto aleatório (Modelo M4), com estrutura de erros autorregressivos AR(1).

Modelo com a variável Ano , Custo Atividade, Ano:Custo e intercepto aleatório (M4)					
Efeito fixo					
Coefficiente	Estimativa	Erro Padrão	Estatística t	Valor-p	VIF
Intercepto	265.696,24	259.062,17	1,02	0,3	–
Ano	-131,4	128,35	-1,02	0,31	13,1
Custo Atividade	-194,54	67,68	-2,87	0,01	515.130,87
Ano:Custo Atividade	0,1	0,03	2,89	0,01	517.815,25
Efeitos variáveis					
Níveis	Variância	Desvio Padrão	Valor-p	ICC	
Intercepto (Município)	243.222,50	493.176	0,073	0,54	
Resíduo	204.603,40	452.331	< 0,000		
Avaliadores do ajuste		AIC	BIC		
		1.006,63	1.019,39		

Fonte: Elaborada pela autora.

Com os resultados apresentados na Tabela 3.13, percebe-se que o modelo M4 têm sérios problemas de multicolinearidade, pois apresentou altos valores de VIF. Além disso, os sinais das estimativas dos coeficientes das variáveis “Ano” e “Custo Atividade” são negativos.

No entanto, esses resultados divergem das correlações apresentadas na Tabela 3.11. Quanto a premissa multinível, tem-se que as variâncias são significativamente diferentes de zero, ao nível de 10%. Diante desses resultados, o próximo passo é realizar ajuste de mais dois modelos descartando/excluindo a interação “Ano:Custo Atividade”. A seguir na Tabela 3.14 serão apresentados os resultados do teste de razão de verossimilhança dos modelos com a variável “Ano” e “Custo Atividade”, considerando as duas possibilidades: (i) modelo com intercepto aleatório (M6); e (ii) modelo com intercepto aleatório e inclinação aleatória (M7).

Tabela 3.14: Comparação entre os modelos M6 e M7 ambos com estrutura de erros autorregressivos AR(1).

Modelo com a variável Ano e Custo Atividade	Avaliadores do ajuste	
	AIC	BIC
Com intercepto aleatório (M6)	1.007,41	1.018,13
Com intercepto aleatório e inclinação aleatória (M7)	1.011,38	1.026,38
Teste da Razão de Verossimilhança M6 vs M7	Estatística χ^2	Valor-p
	0,029	0,98

Fonte: Elaborada pela autora.

Conforme Tabela 3.14, pode-se concluir pelo teste da razão da razão de verossimilhança e/ou por meio do AIC e BIC que o modelo mais adequado é o modelo M6 quando comparado com o modelo M7. No entanto, para esse modelo deve-se verificar a questão da premissa multinível e se há problemas de multicolinearidade.

Portanto, na Tabela 3.15, tem-se os resultados a respeito da estimação do modelo com as variáveis “Ano”, “Custo Atividade” e intercepto aleatório (modelo M6).

Tabela 3.15: Resultados do ajuste do modelo sem a interação entre “Ano” e “Custo Atividade” com intercepto aleatório (Modelo M6) e com estrutura de erros autorregressivos AR(1).

Modelo com a variável Ano , Custo Atividade e intercepto aleatório (M6)					
Efeito fixo					
Coeficiente	Estimativa	Erro Padrão	Estatística t	Valor-p	VIF
Intercepto	-440.944,70	89.956,08	-4,9	< 0,000	–
Ano	218,6	44,67	4,89	< 0,000	1,31
Custo Atividade	1,6	0,11	13,81	< 0,000	1,31
Efeitos variáveis					
Níveis	Variância	Desvio Padrão	Valor-p	ICC	
Intercepto (Município)	205.162,70	452,94	0,084	0,46	
Resíduo	236.699,10	486,51	< 0,000		
Avaliadores do ajuste	AIC	BIC			
	1.007,41	1.018,13			

Fonte: Elaborada pela autora.

De acordo com os resultados da Tabela 3.15, percebe-se que o modelo M6 não apresenta problemas de multicolinearidade, pois apresentou baixos valores de VIF. Além disso, os sinais

das estimativas dos coeficientes das variáveis “Ano” e “Custo Atividade” são positivos, ou seja, esses resultados convergem com as correlações apresentadas na Tabela 3.11. Quanto a premissa multinível, tem-se que as variâncias são significativamente diferentes de zero, ao nível de 10%.

No entanto, ainda é necessário comparar o modelo M6 com o modelo M3 (melhor modelo da etapa anterior). A seguir na Tabela 3.16 serão apresentados os resultados do teste de razão de verossimilhança dos modelos M3 e M6.

Tabela 3.16: Comparação entre os modelos M3 e M6 ambos com estrutura de erros autorregressivos AR(1).

Modelos	Avaliadores do ajuste	
	AIC	BIC
Ano e intercepto aleatório (M3)	1.079,13	1.087,77
Ano e Custo Variável (M7)	1.007,41	1.018,13
Teste da Razão de Verossimilhança M3 vs M6	Estatística χ^2	Valor-p
	73,72	<0,000

Fonte: Elaborada pela autora.

Na Tabela 3.16, pode-se concluir pelo teste da razão da razão de verossimilhança e/ou por meio do AIC e BIC que o modelo mais adequado é o modelo M6 quando comparado com o modelo M3.

Diante dos resultados apresentados até o momento fica evidente o quanto a análise multinível exige trabalho e atenção. Esse procedimento, deve ser repetido para as variáveis “Custo Variável”, “Produtividade”, “Custo Fixo”, “Renda Fatores” e “Tipo de Agricultura”. Isso significa que devem ser ajustados modelos com e sem interação entre variável “Ano” e cada uma das variáveis citadas anteriormente, considerando as possibilidades de intercepto aleatório e/ou intercepto aleatório e inclinação aleatória. Portanto, a partir desse instante, os resultados serão apresentados de forma sintetizada para definir qual é o modelo mais adequado nessa etapa 3.

A seguir será apresentado na Tabela 3.17, os resultados do teste de razão de verossimilhança, do AIC e BIC dos modelos da etapa 3 que devem ser comparados com o modelo M6. É importante destacar/adiantar que a interação da variável “Ano” com todas as demais variáveis explicativas do nível 2 não permaneceram em nenhum modelo, seja por conta da não significância do parâmetro associada a interação ou por conta de “problemas” de multicolinearidade. Além disso, os modelos com intercepto aleatório apresentaram melhores ajustes quando comparados com os modelos com intercepto aleatório e inclinação aleatória, visto que foi observado significância de variância somente para o intercepto. Portanto, a partir desse ponto, os resultados que serão apresentados consideram modelos com intercepto aleatório, com estrutura de erros AR(1) e sem interação.

Tabela 3.17: Comparação entre os modelos M6 e demais modelos da etapa 3 todos com estrutura de erros autorregressivos AR(1).

Modelos com apenas intercepto aleatório	Avaliadores do ajuste		Melhor Modelo
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M6
Ano e Custo Variável (M7)	1.036,17	1.046,89	
Teste da Razão de Verossimilhança M7 vs M6	Estatística χ^2	Valor-p	
	28,764	<0,000	
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M6
Ano e Produtividade (M8)	1.067,32	1.078,03	
Teste da Razão de Verossimilhança M8 vs M6	Estatística χ^2	Valor-p	
	59,90	<0,000	
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M6
Ano e Custo Fixo (M9)	1.042,12	1.052,83	
Teste da Razão de Verossimilhança M9 vs M6	Estatística χ^2	Valor-p	
	34,707	<0,000	
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M6
Ano e Renda Fatores (M10)	1.080,90	1.091,62	
Teste da Razão de Verossimilhança M10 vs M6	Estatística χ^2	Valor-p	
	73,493	<0,000	
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M6
Ano e Tipo de Agricultura (M11)	1.063,70	1.074,42	
Teste da Razão de Verossimilhança M11 vs M6	Estatística χ^2	Valor-p	
	56,295	<0,000	

Fonte: Elaborada pela autora.

A partir dos resultados apresentados na Tabela 3.17, tem-se pelo teste da razão da razão de verossimilhança, do AIC e do BIC que o modelo M6 é o modelo mais adequado na etapa

3. Portanto, o modelo da etapa 3 inclui a variável “Ano” e “Custo Atividade”, com intercepto aleatório, com estrutura de erros AR(1) e sem interação.

3.2.4 ETAPA 4

Nessa etapa o modelo M6 será atualizado, ou seja, será incluído ao modelo M6 uma variável por vez. A seguir destacam-se a ordem em que as variáveis serão incluídas: “Custo Variável”, “Produtividade”, “Custo Fixo”, “Renda Fatores” e “Tipo de Agricultura”, respectivamente. Novamente, ressaltar que devem ser ajustados modelos com e sem interação entre variável “Ano” e cada uma das variáveis citadas anteriormente, considerando as possibilidades de intercepto aleatório e/ou intercepto aleatório e inclinação aleatória. No entanto, como mencionado anteriormente, os resultados serão apresentados de forma sintetizada para definir qual é o modelo mais adequado nessa etapa 4.

A seguir será apresentado na Tabela 3.18, os resultados do teste de razão de verossimilhança, do AIC e BIC dos modelos da etapa 4 que devem ser comparados com o modelo M6. Novamente destaca-se que a interação da variável “Ano” com as variáveis explicativas “Custo Variável”, “Produtividade”, “Custo Fixo”, “Renda Fatores” e “Tipo de Agricultura” não permaneceram em nenhum modelo, seja por conta da não significância do parâmetro associada a interação ou por conta de “problemas” de multicolinearidade. Além disso, os modelos com intercepto aleatório apresentaram melhores ajustes quando comparados com os modelos com intercepto aleatório e inclinação aleatória, visto que foi observado significância de variância somente para o intercepto. Portanto, a partir desse ponto, os resultados que serão apresentados consideram modelos com intercepto aleatório, com estrutura de erros AR(1) e sem interação.

Tabela 3.18: Comparação entre os modelos M6 e demais modelos da etapa 4 todos estrutura de erros autorregressivos AR(1).

Modelos com apenas intercepto aleatório	Avaliadores do ajuste		Melhor Modelo (Ocorrências)
	AIC	BIC	
Ano e Custo Atividade (M6)	1.007,41	1.018,13	M12
Ano, Custo Atividade e Custo Variável (M12)	1.002,47	1.015,23	
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p	
M6 vs M12	6,93	0,0084	
	AIC	BIC	
Ano, Custo Atividade e Custo Variável (M12)	1.002,47	1.015,23	M12
Ano, Custo Atividade e Produtividade (M13)	1.011,79	1.024,55	(M13 -
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p	Produtividade
M13 vs M12	9,31	<0,000	é não significativa)

	AIC	BIC	
Ano, Custo Atividade e Custo Variável (M12)	1.002,47	1.015,23	"M14" (Ano é não significativa e o sinal do coeficiente é negativo - Multicolinearidade). Portanto, melhor modelo é M12.
Ano, Custo Atividade e Produtividade (M14)	928,58	914,34	
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p	
M12 vs M14	73,89	<0,000	
	AIC	BIC	
Ano, Custo Atividade e Custo Variável (M12)	1.002,47	1.015,23	M12
Ano, Custo Atividade e Tipo Agricultura (M15)	1.010,44	1.023,21	(Renda
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p	Fatores é não
M15 vs M12	7,97	<0,000	significativa)
	AIC	BIC	
Ano, Custo Atividade e Custo Variável (M12)	1.002,47	1.015,23	M16
Ano, Custo Atividade e Tipo de Agricultura (M16)	995,01	1.007,77	(Tipo de
Teste da Razão de Verossimilhança	Estatística χ^2	Valor-p	Agricultura é
M12 vs M16	7,46	<0,000	não
			significativa). Portanto melhor modelo é M12.

Fonte: Elaborada pela autora.

A partir dos resultados apresentados na Tabela 3.18, destaca-se: (i) que no modelo M13 a variável “Produtividade” é não significativa; (ii) que o modelo M14 apresenta problemas de multicolinearidade; (iii) que no modelo M16 a variável “Tipo de Agricultura” é não significativa. Portanto, tem-se pelo teste da razão da razão de verossimilhança, do AIC e do BIC que o modelo M12 é o modelo mais adequado na etapa 4.

Na Tabela 3.19, apresentam-se os resultados da estimação do modelo com as variáveis “Ano”, “Custo Atividade”, “Custo Variável”, com intercepto aleatório, com estrutura de erros AR(1) e sem interação (modelo M12).

Tabela 3.19: Resultados do ajuste do modelo sem a interação entre “Ano”, “Custo Atividade”, “Custo Variável” com intercepto aleatório (Modelo M12), com estrutura de erros autorregressivoscm AR(1).

Modelo com a variável Ano , Custo Atividade, Custo Variável e intercepto aleatório (M12)					
Efeito fixo					
Coeficiente	Estimativa	Erro Padrão	Estatística t	Valor-p	VIF
Intercepto	-452.843,50	86.487,38	-5,23	< 0,000	–
Ano	224,5	42,94	5,22	< 0,000	1,35
Custo Atividade	1,2	0,17	7,17	< 0,000	3,49
Custo Variável	1,4	0,57	2,55	0,0135	3,05
Efeitos variáveis					
Níveis	Variância	Desvio Padrão	IC95% Desvio Padrão	Valor-p	ICC
Intercepto (Município)	162.435,20	403,03	[217,21; 747,84]	0,113	0,42
Resíduo	222.156,60	471,33	[389,64; 570,16]	< 0,000	
Avaliadores do ajuste	AIC	BIC			
	1.002,47	1.015,23			

Fonte: Elaborada pela autora.

3.2.5 ETAPA 5

De forma análoga a etapa anterior, o modelo M12 será atualizado, ou seja, será incluído ao modelo M12 uma variável por vez. A seguir destacam-se a ordem em que as variáveis serão incluídas: “Produtividade”, “Custo Fixo”, “Renda Fatores” e “Tipo de Agricultura”, respectivamente. Novamente, ressalta-se que devem ser ajustados modelos com e sem interação entre variável “Ano” e cada uma das variáveis citadas anteriormente, considerando as possibilidades de intercepto aleatório e/ou intercepto aleatório e inclinação aleatória. No entanto, como mencionado anteriormente, os resultados serão apresentados de forma sintetizada para definir qual é o modelo mais adequado nessa etapa 5.

A seguir será apresentado na Tabela 3.20, os resultados do teste de razão de verossimilhança, do AIC e BIC dos modelos da etapa 5 que devem ser comparados com o modelo M12. Novamente destaca-se que a interação da variável “Ano” com as variáveis explicativas “Produtividade”, “Custo Fixo”, “Renda Fatores” e “Tipo de Agricultura” não permaneceram em nenhum modelo, seja por conta da não significância do parâmetro associada a interação ou por conta de “problemas” de multicolinearidade. Além disso, os modelos com intercepto aleatório apresentaram melhores ajustes quando comparados com os modelos com intercepto aleatório e inclinação aleatória, visto que foi observado significância de variância somente para o intercepto. Portanto, a partir desse ponto, os resultados que serão apresentados consideram modelos com intercepto aleatório, com estrutura de erros AR(1) e sem interação.

Tabela 3.20: Comparação entre os modelos M12 e demais modelos da etapa 5 todos com estrutura de erros autorregressivos AR(1).

Modelos com apenas intercepto aleatório	Avaliadores do ajuste		Melhor Modelo (ocorrências)
	AIC	BIC	
Ano, Custo Atividade e Custo Variável (M12)	1.002,48	1.015,24	M12
Ano, Custo Atividade, Custo Variável e Produtividade (M17)	1.006,67	1.021,45	
Teste da Razão de Verossimilhança M12 vs M17	Estatística χ^2	Valor-p	
	2,19	0,14	
Ano, Custo Atividade e Custo Variável (M12)	AIC	BIC	"M18" (Intercepto e Ano são não significativos; sinal negativo do coeficiente associado a Ano - Multicolinearidade). Portanto, melhor modelo é M12
	1.002,48	1.015,24	
Ano, Custo Atividade, Custo Variável e Custo Fixo (M18)	925,89	940,67	
Teste da Razão de Verossimilhança M12 vs M18	Estatística χ^2	Valor-p	
	78,5854	<0,0001	
Ano, Custo Atividade e Custo Variável (M12)	AIC	BIC	M12
	1.002,48	1.015,24	
Ano, Custo Atividade, Custo Variável e Renda Fatores (M19)	1.005,02	1.019,80	
Teste da Razão de Verossimilhança M12 vs M19	Estatística χ^2	Valor-p	
	0,55	0,46	
Ano, Custo Atividade e Custo Variável (M12)	AIC	BIC	M20
	1.002,48	1.015,24	
Ano, Custo Atividade, Custo Variável e Tipo de Agricultura (M20)	988,63	1.003,40	
Teste da Razão de Verossimilhança M12 vs M20	Estatística χ^2	Valor-p	
	15,84	<0,001	

Fonte: Elaborada pela autora.

A partir dos resultados apresentados na Tabela 3.20, destaca-se que no modelo M18 os parâmetros associados ao intercepto e a variável “Ano” são não significativos, e o sinal do

coeficiente associado à variável “Ano” no modelo M18 é negativo, isto é, diverge do resultado encontrado a partir da correlação linear entre “Custo Total” e “Ano”. Portanto, tem-se pelo teste da razão da razão de verossimilhança, do AIC e do BIC que o modelo M20 é o modelo mais adequado na etapa 5.

Na Tabela 3.21, apresentam-se os resultados da estimação do modelo com as variáveis “Ano”, “Custo Atividade”, “Custo Variável”, “Tipo de Agricultura”, com intercepto aleatório, com estrutura de erros AR(1) e sem interação (modelo M20).

Tabela 3.21: Resultados do ajuste do modelo sem a interação entre “Ano”, “Custo Atividade”, “Custo Variável”, “Tipo de Agricultura” com intercepto aleatório (Modelo M20), com estrutura de erros autorregressivos AR(1).

Modelo com a variável Ano , Custo Atividade, Custo Variável, Tipo de Agricultura e intercepto aleatório (M20)					
Efeito fixo					
Coeficiente	Estimativa	Erro Padrão	Estatística t	Valor-p	VIF
Intercepto	-435.328,10	85.493,14	-5,09	< 0,000	–
Ano	215,72	42,46	5,08	< 0,000	1,39
Custo Atividade	1,17	0,17	6,76	< 0,000	3,58
Custo Variável	1,67	0,59	2,86	0,0061	3,13
Tipo Ag. Familiar	445,48	255,9	1,74	0,0877	1,07
Efeitos variáveis					
Níveis	Variância	Desvio Padrão	IC95% Desvio Padrão	Valor-p	ICC
Intercepto (Município)	186.251,90	431,57	[229,70; 810,84]	0,12	0,47
Resíduo	210.884,50	459,22	[378,94; 556,51]	<0,000	
Avaliadores do ajuste	AIC	BIC			
	988,628	1003,404			
Pressupostos dos resíduos					
Testes	Normalidade	Independência	Homogeneidade de Variância		
	Shapiro-Wilk	Durbin-Watson	Breusch-Pagan		
Valor-p	0,29	0,30	0,99		

Fonte: Elaborada pela autora.

Com base nos dados apresentados na Tabela 3.21, tem-se que todas as variáveis são significativas e que o modelo não apresenta problemas de multicolinearidade, pois os valores de VIF são baixos. Quanto a premissa multinível pode-se afirmar que foi atendida pois, os intervalos de confiança de 95% para o desvio padrão não contém o valor zero em sua extensão. Logo, as variâncias são significativamente diferentes de zero.

3.2.6 ETAPA 6

De forma análoga a etapa anterior, o modelo M20 será atualizado, isto é, será incluído ao modelo M20 uma variável por vez. A seguir destacam-se a ordem em que as variáveis serão incluídas: “Produtividade”, “Custo Fixo” e “Renda Fatores”, respectivamente. Novamente, ressalto que devem ser ajustados modelos com e sem interação entre variável “Ano” e cada uma das variáveis citadas anteriormente, considerando as possibilidades de intercepto aleatório e/ou intercepto aleatório e inclinação aleatória. No entanto, como mencionado anteriormente, os resultados serão apresentados de forma sintetizada para definir qual é o modelo mais adequado nessa etapa 6.

A seguir será apresentado na Tabela 3.22, os resultados do teste de razão de verossimilhança, do AIC e BIC dos modelos da etapa 6 que devem ser comparados com o modelo M20. Novamente destaca-se que a interação da variável “Ano” com as variáveis explicativas “Produtividade”, “Custo Fixo” e “Renda Fatores” não permaneceram em nenhum modelo, seja por conta da não significância do parâmetro associada a interação ou por conta de “problemas” de multicolinearidade.

Além disso, os modelos com intercepto aleatório apresentaram melhores ajustes quando comparados com os modelos com intercepto aleatório e inclinação aleatória, visto que foi observado significância de variância somente para o intercepto. Portanto, a partir desse ponto, serão apresentados resultados que consideram modelos com intercepto aleatório, com estrutura de erros AR(1) e sem interação.

Tabela 3.22: Comparação entre os modelos M20 e demais modelos da etapa 6 com estrutura de erros autorregressivos AR(1).

Modelos com apenas intercepto aleatório	Avaliadores do ajuste		Melhor Modelo (Ocorrência)
	AIC	BIC	
Ano, Custo Atividade, Custo Variável e Tipo de Agricultura (M20)	988,63	1.003,40	"M21"
Ano, Custo Atividade, Custo Variável, Tipo de Agricultura e Produtividade (M21)	992,25	1.009,00	(Produtividade é não significativa; sinal do coeficiente negativo - Multicolinearidade).
Teste da Razão de Verossimilhança M20 vs M21	Estatística χ^2	Valor-p	Portanto, o melhor modelo é M20
	14,23	<0,00	
	AIC	BIC	
Ano, Custo Atividade, Custo Variável e Tipo de Agricultura (M20)	988,63	1.003,40	"M22"
Ano, Custo Atividade, Custo Variável, Tipo de Agricultura e Custo Fixo (M22)	913,97	930,78	(Intercepto, Ano e Tipo de Agricultura são não significativos; sinal do coeficiente Ano - negativo Multicolinearidade).
Teste da Razão de Verossimilhança M20 vs M22	Estatística χ^2	Valor-p	Portanto, melhor modelo é M20
	92,50	<0,00	

	AIC	BIC	
Ano, Custo Atividade, Custo Variável e Tipo de Agricultura (M20)	988,63	1.003,40	"M23"
Ano, Custo Atividade, Custo Variável e Renda Fatores (M23)	991,55	1.008,30	(Renda Fatores é não significativo). Portanto, o melhor
Teste da Razão de Verossimilhança M20 vs M23	Estatística χ^2	Valor-p	modelo é M20.
	14,93	<0,000	

Fonte: Elaborada pela autora.

A partir dos resultados apresentados na Tabela 3.22, destaca-se que no modelo M21 o parâmetro associado a variável “Produtividade” é não significativo e o sinal do coeficiente associado à variável “Produtividade” no modelo M21 é negativo, isto é, diverge do resultado encontrado a partir da correlação linear entre “Custo Total” e “Produtividade”. Já no modelo M22, os parâmetros associados ao intercepto, à variável “Ano” e variável “Tipo de Agricultura” são não significativos. Além disso o sinal do coeficiente associado a variável “Ano” é negativo, isto é, diverge do resultado encontrado a partir da correlação linear entre “Custo Total” e “Ano”.

Quanto ao modelo M23, o parâmetro associado à variável “Renda Fatores” é não significativo. Dadas as circunstâncias pode-se dizer que o modelo M20 é o modelo mais adequado na etapa 6. Portanto, nenhuma variável foi incluída nessa etapa, então o processo de seleção de variáveis se encerra na etapa 6 e o modelo final é modelo M20.

Os resultados do ajuste do modelo M20 já foram apresentados anteriormente na Tabela 3.21. Percebe-se a partir da Tabela 3.21 que as premissas acerca dos resíduos foram todas atendidas, conforme os testes de normalidade (Shapiro-Wilk), independência (Durbin-Watson) e homogeneidade de variâncias (Breusch-Pagan) dos resíduos. Além disso, destaca-se na Tabela 3.21, a partir do intervalo de confiança do desvio padrão que a premissa para o ajuste do modelo multinível também foi atendida, isto é, a variância do nível 2 (186.251,9) é significativamente diferente de zero. Ainda a partir da Tabela 3.21 pode-se afirmar que as variáveis “Custo Atividade”, “Custo Variável” e “Tipo de Agricultura” referentes aos municípios são representativas para a diferenciação do custo total médio e das taxas de crescimento deste custo entre os municípios ao longo do tempo.

Em relação às variáveis que são associadas ao município (nível 2) verifica-se que “Custo Atividade” e “Custo Variável”, apresentaram resultados positivos e altamente significativos (valores-p < 0,01). Quanto à variável “Tipo de Agricultura”, significativa a 10%, observa-se que o custo total médio para Produtividade de arroz é maior para a agricultura familiar quando comparado com a agricultura empresarial.

Já a variável “Ano”, o modelo indica que há uma variação significante do custo total, ou seja, o custo total segue uma tendência temporal linear positiva. Além disso, cerca de 47% da proporção da variação do custo total pode ser atribuído às características dos municípios. Finalmente, será apresentado na Tabela 3.23 as estimativas finais dos coeficientes do modelo final (intercepto aleatório e estrutura de erros AR(1)).

Tabela 3.23: Estimativas dos coeficientes referentes ao ajuste do modelo final, contextualizado por municípios.

Municípios	Intercepto	Ano	Custo Atividade	Custo Variável	Tipo de Agricultura
Boa Vista	-435.908,27	215,72	1,17	1,68	445,48
Dourados	-435.768,77	215,72	1,17	1,68	445,48
Meleiro	-435.719,59	215,72	1,17	1,68	445,48
Restinga Seca	-435.445,91	215,72	1,17	1,68	445,48
Pelotas	-435.384,44	215,72	1,17	1,68	445,48
Uruguaiana	-435.336,18	215,72	1,17	1,68	445,48
Balsas	-435.130,80	215,72	1,17	1,68	445,48
Santo Antônio da Patrulha	-435.024,25	215,72	1,17	1,68	445,48
Cachoeira do Sul	-434.989,57	215,72	1,17	1,68	445,48
Massaranduba	-434.983,68	215,72	1,17	1,68	445,48
Sorriso	-434.918,10	215,72	1,17	1,68	445,48

Fonte: Elaborada pela autora.

Com base na Tabela 3.23 percebe-se claramente que todas as variáveis são de efeito fixo, pois o único coeficiente que varia é o intercepto. Tendo assim os valores 215,72; 1,17; 1,68; e 445,48 para as variáveis “Ano”, “Custo Atividade”, “Custo Variável”, “Custo Agricultura” respectivamente, independente do município que será usado no cálculo do custo total, variando somente o valor do intercepto no modelo.

4. CONCLUSÕES

Através deste trabalho, com o uso da modelagem multinível, foi possível avaliar as duas hipóteses levantadas: H1 - há variabilidade no custo total do arroz ao longo dos seis anos analisados; e H2 - há variabilidade no custo total, ao longo dos seis anos analisados, entre os municípios. Os resultados mostram que as duas hipóteses não podem ser refutadas, isto é: (i) pode-se afirmar que em relação ao nível tempo, o modelo indica que há uma variação significativa do custo total, ou seja, o custo total segue uma tendência temporal linear, (ii) o custo total, ao longo dos seis anos analisados varia entre os municípios. Conforme modelo final, observa-se que as variáveis “Ano”, “Custo Atividade”, “Custo Variável” e “Tipo de Agricultura” que caracterizam os municípios auxiliam na explicação da variabilidade do custo total ao longo dos anos. Os resultados indicam que 46,9% da proporção da variação do custo total pode ser atribuído às características dos municípios.

Portanto, fica evidente que a modelagem multinível permite que sejam identificadas e analisadas as heterogeneidades individuais (municípios) e entre grupos a que pertencem esses indivíduos. Logo, a partir dos modelos multiníveis (modelos mistos), torna-se possível a inclusão de componentes aleatórios em cada nível da análise. É justamente esse o ponto que diferencia a modelagem multinível dos tradicionais modelos de regressão estimados via Mínimos Quadrados Ordinários (MQO). Além disso, os modelos tradicionais estimados por MQO não conseguem levar em consideração a estrutura aninhada dos dados e, conseqüentemente, geram estimadores viesados dos parâmetros [9].

Para estudos futuros pretende-se considerar mais municípios de um mesmo estado. Acredita-se que assim será possível identificar uma possível variabilidade também na média de custos entre unidades federativas distintas e melhorar ainda mais a capacidade de predição do modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Akaike, H.: *A new look at the statistical model identification*. IEEE transactions on automatic control, 19(6):716–723, 1974.
- [2] ASCOM, A.: *Mapeamento da CONAB e da ANA identifica 1,3 milhão de hectares de arroz irrigado.*, 2020. <https://www.conab.gov.br/ultimas-noticias/3569-mapeamento-da-conab-e-daana-identifica-1-3-milhao-de-hectares-de-arroz-irrigado-no-brasil>, acessado em 10/03/2022.
- [3] Battisti, I. D. E.: *Análise de dados epidemiológicos incorporando planos amostrais complexos. 2008. 198 p.* Tese de Doutorado, Tese (Doutorado em Epidemiologia)-Faculdade de Medicina, Universidade . . . , 2008.
- [4] CONAB: *Arroz.* <https://www.conab.gov.br/info-agro/custos-de-producao/planilhas-de-custo-de-producao/itemlist/category/791-arroz>, acessado em 21/01/2022.
- [5] Diaz, M. D. M.: *Efetividade no ensino superior brasileiro: aplicação de modelos multinível à análise de resultados do exame nacional de cursos*. Revista Economia, 8(1):93–120, 2007.
- [6] Ferrão, M. E.: *Introdução aos Modelos de Regressão Multinível em Educação*. Komed, 2003.
- [7] FÁVERO, L. P.; CONFORTINI, D.: *Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®*. pp. 855–938, 2020.
- [8] FÁVERO, LUIZ PAULO; BELFIORE, P.: *Modelos multinível de coeficientes aleatórios e os efeitos firma, setor e tempo no mercado acionário Brasileiro, 2010*, 2010. <https://www.scielo.br/j/pope/a/5FqmXZfgcfMwpyJZ8HvCVLq/?lang=pt>, acessado em 20/12/2021.
- [9] FÁVERO, LUIZ PAULO; BELFIORE, P.: *Manual de análise de dados*. 1ª ed., 2017.
- [10] Gareth, J., Daniela, W., Trevor, H. e Robert, T.: *An introduction to statistical learning: with applications in R*. Springer, 2013.
- [11] Goldstein, H.: *Multilevel Statistical Models Edward Arnold*.
- [12] HOFFMANN, R.; VIEIRA, S.: *Análise de regressão: uma introdução à econometria*. HUCITEC, 3ª ed., 1998.

- [13]KUTNER, M. H.: *Applied linear models*. New York: McGraw-Hill Irwin, 5ª ed., 2005.
- [14]McElreath, R.: *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2ª ed., 2020.
- [15]Nunes, A. M. R.: *O modelo linear misto multinível na análise do efeito do desbaste de pinheiros na recuperação ecológica de uma pedreira calcária*. Mestrado em Bioestatística, 2010.
- [16]Patterson, H. D. e Thompson, R.: *Recovery of inter-block information when block sizes are unequal*. *Biometrika*, 58:545–554, 1971.
- [17]REIS, T.: *Custo de Produtividade: a contabilidade fundamental de uma empresa*. <https://www.suno.com.br/artigos/custo-producao/>, acessado em 26/02/2022.
- [18]RIBOLDI, J.; PRETTO, K. F. M. B. G. L. A.: *Metodologia de Modelos Mistos: Aplicações à Epidemiologia*. 2005.
- [19]RICE, B.: *Perfil da Produtividade*, 2022. <http://brazilianrice.com.br/br/sobre-o-brasil/>, acessado em 10/02/2022.
- [20]SANTOS, J. J.: *Contabilidade e Análise de custos: Modelo contábil, Método de depreciação, ABC – Custeio Baseado em Atividades, análise atualizada de encargos sociais sobre salários*. 5ª ed., 2005.
- [21]SCHWARZ, G. E.: *Estimating the dimension of a model*. *Annals of Statistics* 6 (2) : 461—464. 1978.
- [22]SNIJDERS, T.; BOSKER, R.: *An introduction to basic and advanced multilevel modeling*, 1999.
- [23]TAMHANE, A. C. . D. D. D.: *Statistics and Data Analysis: from elementary to intermediate*. 2000.
- [24]TEAM, R. C.: *Software R*. 2021. <https://www.r-project.org/>, acessado em 10/12/2022.