



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE GRADUAÇÃO EM BACHARELADO EM ESTATÍSTICA**

ELYDA CAMYLA DE ARAÚJO ALVES

**MODELOS LINEARES MISTOS APLICADOS A DADOS DE PLANEJAMENTO E
EXPERIMENTO AGRONÔMICO**

CAMPINA GRANDE - PB

2023

ELYDA CAMYLA DE ARAÚJO ALVES

**MODELOS LINEARES MISTOS APLICADOS A DADOS DE PLANEJAMENTO E
EXPERIMENTO AGRONÔMICO**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Ricardo Alves de Olinda

CAMPINA GRANDE - PB

2023

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

A474m Alves, Elyda Camyla de Araujo.
Modelos lineares mistos aplicados a dados de planejamento e experimento agrônômico [manuscrito] / Elyda Camyla de Araujo Alves. - 2023.
43 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.

"Orientação : Prof. Dr. Ricardo Alves de Olinda, Coordenação do Curso de Estatística - CCT. "

1. Efeitos fixos. 2. Efeitos aleatórios. 3. Modelo misto. I.

Título

21. ed. CDD 519.5

ELYDA CAMYLA DE ARAÚJO ALVES

MODELOS LINEARES MISTOS APLICADOS A DADOS DE PLANEJAMENTO E
EXPERIMENTO AGRONÔMICO

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Trabalho aprovado em : 24/11/2023.

BANCA EXAMINADORA



Prof. Dr. Ricardo Alves de Olinda (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Antonio Samuel Alves da Silva
Universidade Federal Rural de Pernambuco
Departamento de Estatística e Informática

Dedico aos meus pais, Maria Raquel e Sebastião.

AGRADECIMENTOS

Primeiramente agradeço a Deus, meu refúgio e fortaleza, por ter me guiado e dado forças para superar todos os obstáculos dessa trajetória, sem a presença Dele nada disso seria possível.

A minha mãe, Maria Raquel, por todo apoio e cuidado e ao meu pai, Sebastião Alves, pelo incentivo aos estudos, vocês foram essenciais para conclusão deste curso.

Ao meu orientador, Dr. Ricardo Alves de Olinda, por sua dedicação, paciência e orientações.

A todos os professores que contribuíram com seus conhecimentos e experiências, em especial aos professores Ana Patrícia Peixoto, Tiago Almeida de Oliveira, Gustavo Henrique Esteves, Diana Maia e Ricardo Alves de Olinda, por todas as oportunidades de aprofundamento do aprendizado.

Agradeço aos meus amigos e colegas de turma Hellen Silva, Marcela Araujo, Vitoria Souza, Gislania Cauanny, Davi Barbosa e Gabriel Graciano, pela amizade, companheirismo e ajuda. Os momentos de descontração e risos foram fundamentais para tornar o percurso mais leve e agradável.

“A sabedoria vale mais do que a prata, e dá mais lucro que o ouro.
Ela é mais valiosa do que as pérolas e não existe objeto preciso que se iguale a ela”
(Provérbios 3, 14-15)

RESUMO

Neste trabalho utilizou-se modelo linear misto aplicado aos dados obtido de um experimento agrônômico, onde investiga os rendimentos de catorze linhagens de feijão tepary (*Phaseolus acutifolius* Gray.) em três localidades semiáridas do Departamento de Progreso, na Guatemala, as quais são Jícaro, Rancho e Estância. Modelos convencionais não podem ser usados, pois eles não levam em consideração as diferentes variâncias dos níveis (municípios) do fator local, ao qual, conduz à falta de homogeneidade dos resíduos. Neste sentido, o modelo linear misto, por ser mais flexível em relação aos pressupostos sobre os resíduos, pois permite a modelagem de **G** e **R** que são, respectivamente, matrizes de variância e covariância dos efeitos aleatórios e residual, torna-se uma alternativa viável. Ademais, estimaram-se os parâmetros por meio do método da máxima verossimilhança restrita, tendo como efeitos fixos os fatores linhagens, local e interação entre eles, e os blocos aninhados ao local foram tratados como aleatórios. Ao nível de 0,05 de significância, tem-se que os fatores linhagem, local e a interação entre eles são estatisticamente significativos, (p-valor <0,0001). Pelo teste de comparação múltipla de agrupamento DGC (Rienzo, Guzman e Casanoves), pôde-se concluir para os locais onde foram montados os experimentos, que no município Jícaro as linhagens L-242-22 (629,44 a), L-242-43 (615,07 a) e L-242-25 (612,90 a) tiveram maiores destaque no rendimento médio na cultura da produção de feijão tepary, nas quais são estatisticamente similares. Em Rancho, todas as linhagens apresentaram produção média iguais, exceto L-38 (197,53 c) que difere de todas e teve rendimento médio baixo. E para Estância as linhagens L-242-9 (402,25 a) e L-37 (255,34 a) foram as que mostraram melhores rendimentos médios kg/ha. As linhagens L-242-22 (449,66 a), L-242-25 (447,35 a), L-242-43 (436,48 a) e L-242-9 (417,35 a) foram as que mostraram maiores rendimentos médios em relação às outras, além do mais, não há diferença entre elas. Em síntese, o modelo linear misto mostrou ser mais adequado aos dados, uma vez que as suposições sobre os erros foram alcançadas, por outro lado, o modelo linear clássico apresentou inferências afetadas pela heterocedasticidade inerente ao experimento.

Palavras-chave: efeitos fixos; efeitos aleatórios; modelo misto; modelo.

ABSTRACT

In this work, a mixed linear model was applied to the data obtained from an agronomic experiment, where it investigates the yields of fourteen lines of tepary bean (*Phaseolus acutifolius* Gray.) in three semi-arid localities of the Department of Progreso, Guatemala, which are Jícaro, Rancho and Estância. Conventional models cannot be used, because they do not take into account the different variances of the levels (municipalities) of the local factor, which leads to the lack of homogeneity of the residuals. In this sense, the mixed linear model, because it is more flexible in relation to the assumptions about the residuals, as it allows the modeling of **G** and **R**, which are, respectively, matrices of variance and covariance of random and residual effects, becomes a viable alternative. In addition, the parameters were estimated using the restricted maximum likelihood method, having as fixed effects the factors lineages, location and interaction between them, and the blocks nested to the site were treated as random. At the level of 0.05 of significance, the factors lineage, location and the interaction between them are statistically significant (p -value < 0.0001). By the multiple comparison test of DGC grouping (Rienzo, Guzman and Casanoves), it was possible to conclude that in the municipality of Jícaro the lines L-242-22 (629.44 a), L-242-43 (615.07 a) and L-242-25 (612.90 a) had greater prominence in the average yield in the tepary bean crop, in which they are statistically similar. In Rancho, all lines presented equal average production, except L-38 (197.53 c) which differs from all of them and had a low average yield. And for Estancia, the lines L-242-9 (402.25 a) and L-37 (255.34 a) were the ones that showed the best average yields kg/ha. The lines L-242-22 (449.66 a), L-242-25 (447.35 a), L-242-43 (436.48 a) and L-242-9 (417.35 a) showed the highest average yields in relation to the others, in addition, there is no difference between them. In summary, the mixed linear model proved to be more adequate to the data, since the assumptions about the errors were reached, on the other hand, the classical linear model presented inferences affected by the heteroscedasticity inherent to the experiment.

Keywords: fixed effects; random effects; mixed model; model.

LISTA DE ILUSTRAÇÕES

Figura 1 – Rendimento do feijão tepary em kg/ha de acordo com os três locais do experimento.	28
Figura 2 – Análise gráfica do rendimento das linhagens de feijão tepary em kg/ha conforme os três locais onde foram montados os experimentos em blocos completos casualizados.	29
Figura 3 – Análise gráfica do rendimento das catorze linhagens de feijão tepary em kg/ha.	30
Figura 4 – Interação entre os fatores local e linhagem.	32
Figura 5 – Análise dos resíduos para o modelo utilizado na tese de Hernández (1997). .	33
Figura 6 – Análise da variância dos resíduos condicionais studentizados do modelo A. .	36
Figura 7 – Análise da homogeneidade e normalidade dos resíduos condicionais para o modelo B, com heterogeneidade corrigida pela função de variância $\sigma_y^2 = \sigma^2 \delta_{Sij}^2$.	37

LISTA DE TABELAS

Tabela 1	– Análise descritiva do rendimento do feijão tepary em kg/ha nos três locais onde foram montados os experimentos em blocos completos casualizados.	27
Tabela 2	– Análise descritiva do rendimento das linhagens de feijão tepary em kg/ha conforme os locais onde foram montados os experimentos em blocos completos casualizados.	28
Tabela 3	– Rendimento das catorze linhagens de feijão tepary em kg/ha.	29
Tabela 4	– Análise da variância do modelo usado por Hernández (1997), para testar a igualdade das médias dos tratamentos em relação aos rendimentos de feijão tepary.	31
Tabela 5	– P-valores dos testes de Shapiro-Wilk e Bartlett conforme cada localidade onde foram montados os experimentos em blocos completos casualizados.	34
Tabela 6	– Estimativas dos critérios de ajuste, Akaike (AIC), Bayesiano de Schwarz (BIC), razão de verossimilhança com a probabilidade associada (P-valor), dos modelos A e B.	38
Tabela 7	– Testes de hipóteses sequenciais dos efeitos fixos e variância dos efeitos aleatórios ($\hat{\sigma}_B^2$) e residual ($\hat{\sigma}^2$) do modelo B.	38
Tabela 8	– Estimativas e comparações das médias em kg/ha das linhagens pelo teste de comparações múltiplas de agrupamentos DGC (Rienzo, Guzman e Casanoves). Médias acompanhadas da mesma letra não diferem pelo teste de comparação múltipla ao nível de 5% de significância.	39
Tabela 9	– Estimativas e comparações das médias em kg/ha das linhagens conforme os locais pelo teste de comparações múltiplas de agrupamento DGC (Rienzo, Guzman e Casanoves). Médias acompanhadas da mesma letra não diferem pelo teste de comparação múltipla ao nível de 5% de significância.	43

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i>
BLUP	<i>Best Linear Unbiased Predictor</i>
BLUE	<i>Best Linear Unbiased Estimator</i>
BIC	<i>Bayesian Information Criterion</i>
DGC	(Rienzo, Guzman e Casanoves)
MQ	Mínimos quadrados
MV	Máximo verossimilhança
MVR	Máximo verossimilhança restrita
MLM	Modelos lineares mistos

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Marco Histórico	14
2.2	Métodos Estatísticos	15
2.2.1	<i>Estruturas das matrizes G e R</i>	16
2.3	Estimação	18
2.4	Seleção de modelos	22
2.4.1	<i>Teste da razão de verossimilhança</i>	22
2.4.2	<i>Critérios de informação</i>	22
2.5	Análise de resíduos	23
2.6	Teste de comparação múltipla	24
3	RESULTADOS E DISCUSSÕES	26
3.1	Experimento	26
3.2	Estatística Descritiva	27
3.3	Modelos Estatísticos	30
3.3.1	<i>Modelos lineares mistos</i>	34
4	CONSIDERAÇÕES FINAIS	40
	REFERÊNCIAS	41
	APÊNDICE A – PRIMEIRO APÊNDICE	43

1 INTRODUÇÃO

Um experimento é considerado um trabalho antecipadamente planejado, onde são definidos os fatores, seus níveis, unidades experimentais e a variável resposta, com objetivo de tirar conclusões, isto é, comparar, medir ou prever os efeitos dos tratamentos (níveis) sob a variável resposta (BANZATTO; KRONKA, 2006). A experimentação é regida por três princípios básicos que garantem uma interpretação robusta e precisa, são elas: repetição, responsável por assegurar observações suficientes para estimar o erro experimental, e assim detectar indicativo de diferença entre tratamentos; casualização é a distribuição de forma aleatória dos tratamentos sob a unidade experimental, para garantir que todos os níveis tenham as mesmas possibilidades sem que nenhum seja favorecido ou prejudicado; controle local, este princípio só é utilizado quando a homogeneidade do experimento é incerta, e para contornar esse problema é constituído blocos homogêneos contendo todos os tratamentos com intuito de diminuir o erro experimental e aumentar a eficiência do delineamento (GARCIA; LEITE, 2006).

Os efeitos dos fatores sobre a variável resposta em um experimento podem ser considerados efeitos fixos quando os níveis do fator são determinados pelo pesquisador conforme o seu interesse, neste caso a inferência é particular aos níveis do fator em questão; e/ ou efeitos aleatórios quando o conjunto de níveis é uma amostra aleatória de uma população de possíveis tratamentos (os níveis são selecionados por meio de sorteios), isto é, o propósito é inferir sobre a população que os tratamentos foram extraídos, considerando que os efeitos é uma variável aleatória é viável usar a função de distribuição de probabilidade para caracterizar a distribuição dos efeitos relativos à amostra de tratamentos (BALZARINI; MACCHIAVELLI; CASANOVES, 2004).

Tendo em vista, a maneira que os tratamentos são alocados nas parcelas é definido o modelo de delineamento experimental, onde são empregadas as análises estatísticas para os dados observados de acordo com cada tipo de delineamento (GARCIA; LEITE, 2006). Os tipos mais usuais são o delineamento inteiramente casualizado (DIC), o delineamento em blocos casualizados (DBC) e o delineamento em quadrado latino (DQL). Os modelos estatísticos derivados de diferentes delineamentos podem ser caracterizados como modelo linear clássico, porém, os pressupostos de normalidade, independência e homocedasticidade sobre os erros tornar o modelo linear tradicional mais restritivo, onde na realidade muitas das vezes as pressuposições não são satisfeitas, quando isso ocorre uma opção é a transformação dos dados originais em uma tentativa que as suposições sejam atendidas, todavia, nem sempre isso é útil.

Em 1950 o estatístico estadunidense Charles Roy Henderson começou os estudos e desenvolvimento das equações de modelos mistos, onde inclui os efeitos fixos e os efeitos aleatórios (além do erro experimental) na sua estrutura, a qual promove as estimações e previsões não viesadas respectivamente dos efeitos. Os modelos mistos são considerados flexíveis, por permitirem a incorporação da correlação entre observações, a modelagem da heterocedasticidade da variância, e no caso das observações que não são oriundas de uma distribuição normal,

aplicam-se os modelos generalizados mistos, ainda mais, esse tipo de modelo lida com problemas intrínsecos a dados desbalanceados e incompletos.

Levando em conta, as vantagens que os modelos mistos têm em relação ao relaxamento de suposições sobre os erros, eles têm sido largamente empregados em diversas linhas de pesquisas, como, na agropecuária, onde Oliveira et al. (2004) usaram a metodologia para examinar a variabilidade genética de procedência e progênes de umbuzeiro; na zootecnia em que Garcia et al. (2012) empregaram os modelos lineares generalizados mistos na avaliação genética da prenhez precoce na raça Nelore; no aspecto socioeconômicos na qual Pietrafesa, Silva e Pietrafesa (2022) aplicaram os modelos como ferramenta para análise do índice de desempenho dos municípios do estado de Goiás. Os modelos mistos, por serem mais amplos, na maioria das vezes proporcionam melhor ajuste aos conjuntos de dados reais.

Portanto, o objetivo deste trabalho é utilizar modelos lineares mistos aplicados aos dados obtidos de um experimento agrônômico, onde investiga os rendimentos de catorze linhagens de feijão tepary (*Phaseolus acutifolius Gray.*) em três localidades semiáridas do Departamento de Progreso, na Guatemala. Os feijões tepary apresentam melhor produção em climas áridos e quentes, é um tipo de leguminosa rica em proteína e fibra solúvel, que são fontes de nutrientes usados no controle do colesterol e diabete (HAENSELBLATT, 2023). Os locais Jícaro, Estância e Rancho, onde foram montados os experimentos, sofrem com a má distribuição da chuva que dificulta o plantio de feijões nas regiões, então, o feijão tepary por sua característica pode ser uma alternativa de produtividade nas localidades (HERNÁNDEZ, 1997).

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo será revisado toda teoria estatística dos métodos que serão abordados no trabalho.

2.1 Marco Histórico

O estatístico e geneticista britânico Ronald Aylmer Fisher, considerado um dos fundadores da estatística moderna, foi responsável por desenvolver várias técnicas estatísticas aplicadas à biologia (ROBERTSON, 2003). Foi a partir de Fisher que começaram os estudos da análise de variância (ANOVA), que verifica as diferenças entre as médias de três ou mais grupos independentes, cujo foi o precursor dos modelos lineares e dos modelos lineares mistos. Em 1925, ele publicou o livro "*Statistical Methods for Research Workers*", que descreveu os princípios da ANOVA e sua aplicação em diversos campos da ciência. Muitas das suas contribuições mais relevantes foram desenvolvidas durante sua estadia na Estação Experimental Agrícola de Rothamsted (ROBERTSON, 2003).

Frank Yates estatístico inglês, foi nomeado estatístico assistente na Estação Experimental de Rothamsted em agosto de 1931, onde conheceu Ronald Aylmer Fisher e conjuntamente em 1934, eles publicaram um artigo que descreveu a aplicação da ANOVA em experimentos com parcelas subdivididas, que são um tipo de experimento com estrutura hierárquica (ROBERTSON, 2005).

O estatístico americano Charles Roy Henderson em 1950 desenvolveu as equações de modelo misto de Henderson no qual a abordagem de Fisher-Yates foram usados como base. Diante de sua tese de pós-graduação, ele teve que lidar com os problemas dos dados serem multi-ambientais, apresentarem muitos fatores genéticos aleatórios e valores faltantes, onde nessa época não havia métodos para estimar componentes de variância com tal complexidade (VLECK, 1998). Então, para superar esse obstáculo em seu trabalho, ele criou dois métodos, I e III, em que foi possível a estimativa imparcial de componentes de variância na presença de fatores de confusão com fatores ambientais fixos, em seguida, em 1951, o método II foi apresentado em uma conferência no Norte da Carolina, onde conheceu Fisher, que o incentivou a publicar os métodos. Em 1953, os métodos de estimação dos componentes das variâncias, I, II, III foram publicadas no artigo *Biometrics*, que se tornou referencial (VLECK, 1998).

Em 1955, Henderson conheceu o estatístico neozelandês Shayle Robert Searle, que o ajudou com as provas das propriedades das equações do modelo misto de Henderson através da álgebra matricial (VLECK, 1998).

No ano de 1967, Hartley e Rao propuseram o método de máxima verossimilhança para estimação dos componentes da variância para modelos lineares mistos, onde os estimadores de máxima verossimilhança apresentam a maioria das propriedades almejadas pela inferência estatística para estimativas adequadas, entretanto tal método é viesado por causa da imposição de restrição de não negatividade e perda de graus de liberdade devido à estimação dos efeitos fixos.

Patterson e Thompson em 1971 desenvolveu o método de máxima verossimilhança restrita, que é a modificação do MV, com a finalidade de eliminar o viés (RESENDE et al., 1996).

Os avanços computacionais permitiram a implementação de modelos lineares mistos em programas de *software* estatísticos que intensificaram o uso e estudo destes.

2.2 Métodos Estatísticos

Modelos lineares mistos são responsáveis de incorporam tanto efeitos fixos quanto efeitos aleatórios (além do erro experimental) em suas estruturas. Os efeitos fixos são aqueles que são constantes em todas as unidades experimentais, a qual o interesse é comparar a diferença entre médias ou medir o impacto desta sobre a variável resposta, já os efeitos aleatórios são aqueles que variam entre as unidades experimentais e são incluídos para modelar a heterogeneidade entre elas e a correlação entre observações dentro da mesma unidade, isto é, são empregados para modelar a variação que não é explicada pelos efeitos fixos.

A equação matricial do modelo linear misto é dado por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2.1)$$

Onde, \mathbf{Y} é um vetor de dimensão $(n \times 1)$ que contém os valores da variável resposta de todas unidades experimentais, \mathbf{X} é uma matriz de especificação dos efeitos fixos de ordem $(n \times p)$, $\boldsymbol{\beta}$ é um vetor de parâmetros desconhecidos relacionados aos efeitos fixos $(p \times 1)$, \mathbf{Z} assim como \mathbf{X} também representa uma matriz de especificação (ambas de posto completo), porém associada aos efeitos aleatórios, por isso não inclui uma coluna de 1's para o termo constante, sua dimensão é $(n \times q)$, \mathbf{u} é um vetor $(q \times 1)$ de elementos aleatórios que considera-se ter distribuição $N(\mathbf{0}, \mathbf{G})$ e $\boldsymbol{\varepsilon}$ é um vetor $(n \times 1)$ de erros aleatórios, atribuídos aos efeitos de outros fatores não controláveis, pressupõe que os erros sigam distribuição $N(\mathbf{0}, \mathbf{R})$ (BALZARINI; MACCHIAVELLI; CASANOVES, 2004).

Assume-se que o vetor de parâmetros aleatórios \mathbf{u} é independente de $\boldsymbol{\varepsilon}$, além disso $\mathbb{E}(\mathbf{u}) = \mathbf{0}$ e $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, então

$$Cov \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \quad (2.2)$$

Em que \mathbf{G} e \mathbf{R} são matrizes de variância e covariância, respectivamente de \mathbf{u} e $\boldsymbol{\varepsilon}$, que de acordo com Nobre (2004), elas são quadráticas, positivas definidas.

Para o modelo marginal, a matriz de covariância do vetor \mathbf{Y} é determinada por

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad (2.3)$$

E seu valor esperado é $\mathbf{X}\boldsymbol{\beta}$, deste modo $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. Nota-se que a esperança da variável resposta depende unicamente dos efeitos fixos, enquanto a matriz de variância e covariância é definida pelos efeitos aleatórios. Além disso, a variância dos efeitos aleatórios é responsável

por captura a variabilidade entre as unidades experimentais, ao passo que a variância residual descreve a variabilidade dentro de cada unidade experimental, então, a variação total é a soma da variabilidade entre e intra nível (OLIVEIRA; MARTINI; CATANHO, 2022).

No modelo condicional, a esperança de \mathbf{Y} é dado conforme um nível específico do efeito aleatório, $\mathbb{E}(\mathbf{Y}_i | \mathbf{u}_i)$. Tem-se que $\mathbf{Y}_i | \mathbf{u}_i \sim \mathbf{N}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i, \mathbf{R}_i)$ em que $i = 1, \dots, m$, sendo m o número total de unidade amostral. Todavia, as inferências são baseadas no modelo marginal, a não ser que uma abordagem Bayesiana seja aplicada à análise (BARBOSA, 2009).

2.2.1 Estruturas das matrizes \mathbf{G} e \mathbf{R}

A matriz de variância e covariância dos efeitos aleatórios é denotada por \mathbf{G} , que corresponde a variação entre níveis, na qual, sua diagonal principal consiste na variância de cada nível da variável classificada aleatório, já os demais elementos são as covariâncias entre os níveis, deste modo, quando os efeitos aleatórios não são assumidos correlacionados \mathbf{G} é uma matriz diagonal. Por exemplo, uma matriz \mathbf{G} com $q=3$ efeitos aleatórios, em que σ_a^2 representa o componente de variância entre níveis.

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_a^2 & 0 \\ 0 & 0 & \sigma_a^2 \end{bmatrix}$$

Por meio das matrizes de covariância \mathbf{G} e \mathbf{R} são especificadas as estruturas de correlação e a heterogeneidade da variância do experimento. Entretanto, em situações em que os componentes de variância dos efeitos aleatórios são pequenos ou não são interpretáveis, a correlação e a variância são modeladas diretamente pela matriz de covariância dos resíduos \mathbf{R} , quais os padrões são estabelecidas mediante submatrizes blocos \mathbf{R}_i (BALZARINI; MACCHIAVELLI; CASANOVES, 2004).

A matriz \mathbf{R}_i , que consiste na covariância dentro de cada nível, pode ser expressa como a combinação da variância dos erros juntamente com a correlação entre eles,

$$\mathbf{R}_i = \boldsymbol{\Sigma}_i \mathbf{C}_i \boldsymbol{\Sigma}_i$$

em que $\boldsymbol{\Sigma}_i$ corresponde a matriz da variância dos erros dentro de cada nível, no qual ela é diagonal com seus elementos positivos para garantir unicidade, já que $\boldsymbol{\Sigma}_i$ não é única. \mathbf{C}_i é uma matriz de correlação positiva definida, isto é, admite inversa clássica, e os seus elementos da diagonal principal são iguais a 1 (BARBOSA, 2009).

Para modelar a variância dos erros na presença de variáveis independentes Rienzo, Macchiavelli e Casanoves (2017) define a variância como,

$$\text{var}(\varepsilon_i) = \sigma^2 g^2(\mu_i, \mathbf{v}_i, \boldsymbol{\delta}), \quad i = 1, \dots, n_i$$

onde $g(\cdot)$ é uma função de variância, $\mu_i = \mathbb{E}(y_i)$, \mathbf{v}_i é um conjunto de covariáveis e $\boldsymbol{\delta}$ é o vetor de parâmetros da variância.

A seguir são apresentadas algumas das estruturas da matriz de covariância \mathbf{R}_i , para $n_i = 3$:

- Componente de variância

$$\mathbf{R}_i = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

- Simetria composta com variação idêntica

$$\mathbf{R}_i = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 \end{bmatrix}$$

- Diagonal (D)

$$\mathbf{R}_i = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

- Simetria composta com variação heterogênea

$$\mathbf{R}_i = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 \end{bmatrix}$$

- Modelo não estruturado

$$\mathbf{R}_i = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

- Estrutura autorregressivo de primeira ordem com homogeneidade da variância, AR(1)

$$\mathbf{R}_i = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{bmatrix}$$

- Estrutura autorregressiva de ordem ρ com média móvel de ordem γ , ARMA(ρ, γ)

$$\mathbf{R}_i = \begin{bmatrix} \sigma^2 & \gamma & \gamma\rho \\ \gamma & \sigma^2 & \gamma \\ \gamma\rho & \gamma & \sigma^2 \end{bmatrix}$$

À vista disso, os modelos lineares mistos se tornam menos rígidos, pois a modelagem das matrizes de variância e covariância \mathbf{G} e \mathbf{R} , possibilita a flexibilização dos pressupostos de independência e homocedasticidade das observações dos dados. Garantindo a representação da complexidade de fenômenos que seriam inadequados compreendê-los por meio de modelos lineares clássica, por consequência de quebra de suposições sobre os erros.

2.3 Estimação

Existem diversos métodos de estimação para os parâmetros dos modelos lineares mistos, em geral, os mais usados são os métodos mínimos quadrados (MQ), máxima verossimilhança (MV) e máxima verossimilhança restrita (MVR). O método mais indicado pela literatura para estimar os componentes das variâncias é a de máxima verossimilhança restrita, isso levando em consideração que \mathbf{R} e \mathbf{G} são desconhecidas, a vantagem da MVR sobre os demais é que o resultado gerado pelos estimadores dos parâmetros das matrizes de covariâncias apresentam viés menor, isso ocorre porque o método de estimação leva em conta os graus de liberdade usados para estimar os efeitos fixos (DEMÉTRIO; BRIEN, 2009).

Pela máxima verossimilhança, o vetor contendo todos os componentes de variância pode ser estimado maximizando a função de verossimilhança $L(\boldsymbol{\beta}, \mathbf{G}, \mathbf{R}) = L(\boldsymbol{\beta}, \boldsymbol{\alpha})$, onde $\boldsymbol{\alpha}$ é o vetor de todos os componentes da variância em \mathbf{R} e \mathbf{G} , conforme Balzarini, Macchiavelli e Casanoves (2004) a função de verossimilhança é baseada na distribuição normal multivariada com vetor de média e matriz de covariância \mathbf{V} , em população normal. A maximização consiste no logaritmo, $\log(L)$,

$$L = \frac{\exp\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}}{(2\pi)^{(1/2)n}|\mathbf{V}|^{(1/2)}} \quad (2.4)$$

onde o logaritmo, $\log(L)$ é dado por

$$\log(L) = K - \frac{1}{2}[\log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \quad (2.5)$$

sendo, $K = -\frac{1}{2}n\log(2\pi)$.

Ao maximizar a verossimilhança em relação a $\boldsymbol{\beta}$ e σ^2 , os estimadores destes são obtidos, em que $N = \sum_{i=1}^M n_i$ é o número total de dados do estudo (COSTA, 2010),

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N}$$

Este método tem a desvantagem de apresentar estimativas com viés negativa dos parâmetros da estrutura de covariância, já que os graus de liberdade usados para estimar os efeitos fixos não são considerados (DEMÉTRIO; BRIEN, 2009).

A máxima verossimilhança restrita é uma ampliação do método da máxima verossimilhança, criado por Patterson e Thompson (1971) para eliminar o viés presente no estimador da MV da variância ao estimar os efeitos fixos (BALZARINI; MACCHIAVELLI; CASANOVES, 2004). Neste método a restrição é imposta na função de verossimilhança para remover o efeito dos parâmetros fixos, ficando apenas com os componentes aleatórios para ser estimados. Isso é feito, substituindo o vetor das observações \mathbf{Y} pelos resíduos baseados nos efeitos fixos $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, onde $\hat{\boldsymbol{\beta}}$ é independente $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$.

A função de verossimilhança conjunta para os parâmetros de efeito fixo e para os componentes da variância é dado pelo produto das verossimilhanças, denotado por:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{y}) = L(\boldsymbol{\alpha}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})L(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}, \boldsymbol{\alpha})$$

Então, a função de verossimilhança de $L(\boldsymbol{\alpha}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ é expressa pela seguinte razão

$$L(\boldsymbol{\alpha}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{L(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{y})}{L(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}, \boldsymbol{\alpha})} = |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-(1/2)}|\mathbf{V}|^{-(1/2)}\exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\right\}$$

em que,

$$\log(L(\boldsymbol{\alpha}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})) = K - \frac{1}{2}[\log|\mathbf{V}| - \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1} + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] \quad (2.6)$$

onde, a estimativa de $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$.

Maximizando a verossimilhança de acordo com os parâmetros, tem-se os estimadores,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p}$$

Uma vez que os componentes da variância em \mathbf{R} e \mathbf{G} são estimados, é possível obter os melhores preditores lineares não viesados (BLUP-*Best Linear Unbiased Predictor*) dos efeitos aleatórios e as melhores estimações dos coeficientes dos efeitos fixos (BLUE-*Best Linear Unbiased Estimator*). Fazendo uso da maximização da densidade conjunta de \mathbf{Y} e \mathbf{u} quando segue distribuição conjunta normal multivariada, $\boldsymbol{\beta}$ e \mathbf{u} são estimados e preditos, respectivamente, solucionado as equações de Henderson, onde são obtidos o BLUP e BLUE.

Maximizando a densidade conjunta de \mathbf{Y} e \mathbf{u} ,

$$f(\mathbf{y}, \mathbf{u} | \boldsymbol{\beta}, \mathbf{G}, \mathbf{R}) \propto |\mathbf{R}|^{-1/2}|\mathbf{G}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) - \frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}\right\} \quad (2.7)$$

Tem-se que o logaritmo da função é expresso por

$$\begin{aligned}
l &\propto |\mathbf{R}| + |\mathbf{G}| + (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u} \\
&= |\mathbf{R}| + |\mathbf{G}| + \mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y} - 2\mathbf{Y}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - 2\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} \\
&\quad + 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u}
\end{aligned}$$

Derivando l em relação a $\boldsymbol{\beta}$ e \mathbf{u}

$$\begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \mathbf{u}} \end{bmatrix} \times \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} - \mathbf{G}^{-1}\mathbf{u} \end{bmatrix} \quad (2.8)$$

Igualando-os a zero, obtêm-se

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{G}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix} \quad (2.9)$$

Reescrevendo o sistema (2.9), são obtidas as equações de Henderson

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \times \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix} \quad (2.10)$$

Que multiplicando as matrizes tem-se

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix} \quad (2.11)$$

Igualando a segunda linha de ambas as matrizes de (2.11)

$$\begin{aligned}
\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \\
(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}
\end{aligned}$$

Colocando $\mathbf{Z}'\mathbf{R}^{-1}$ em evidência no segundo membro

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{X})$$

Multiplicando e pré-multiplicando ambos os membros por $(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}$

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{X})$$

Portanto, BLUP-*Best Linear Unbiased Predictor* é definido por

$$\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{X}) \quad (2.12)$$

Agora igualando a primeira linha de ambas as matrizes de (2.11)

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \quad (2.13)$$

substituindo $\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{X})$ em (2.13)

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}[(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{X})] = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$

aplicando a propriedade distributiva e colocando $\hat{\boldsymbol{\beta}}$ em evidência

$$\begin{aligned} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\hat{\boldsymbol{\beta}}\mathbf{X} &= \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ [\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}]\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{aligned}$$

Multiplicando e pré-multiplicando ambos os membros por

$$[\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}]^{-1}$$

então,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}]^{-1} \\ &\quad \times [\mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y}] \end{aligned}$$

colocando \mathbf{X}' , \mathbf{X} e \mathbf{Y} em evidência tem-se

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{X}]^{-1}\mathbf{X}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{Y}$$

Sabe-se que $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, logo $\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}$

Deste modo, $\hat{\boldsymbol{\beta}}$ pode ser reescrito por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (2.14)$$

Sendo assim, a expressão (2.14) determina o BLUE-*Best Linear Unbiased Estimator*, a qual $\hat{\boldsymbol{\beta}}$ apresenta matriz de variância e covariância aproximadamente $(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$.

Nobre (2004) cita algumas das características dos de $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{u}}$, que são funções lineares de \mathbf{Y} e, espera-se que $\mathbb{E}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{0}$ assim como $\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{0}$, isto é, o BLUP e BLUE são não viciados, respectivamente, para \mathbf{u} e $\boldsymbol{\beta}$, ainda mais, eles são nomeados os melhores preditor e estimador, por minimizar o erro quadrático médio (EQM) da predição e estimação.

A estatística F para os efeitos fixos de teste de hipóteses é dado por:

$$\frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\mathbf{K}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})\mathbf{K})^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}})}{r(\mathbf{K})} \approx F_{(\varphi_N, \varphi_D)}$$

Onde \mathbf{K} é uma matriz de contrastes que especifica quais parâmetros serão testados, $r(\mathbf{K}) = \text{rank}(\mathbf{K})$ é o número de linhas linearmente independentes da matriz \mathbf{K} , $F_{(\varphi_N, \varphi_D)}$ corresponde a distribuição F com graus de liberdade $\varphi_N = r(\mathbf{K})$ para o numerador, e os graus de liberdade do denominador φ_D são estimados a partir dos dados por vários métodos, como Satterthwaite e Kenward-Roger, que segundo Balzarini, Macchiavelli e Casanoves (2004) são recomendados para situações onde há efeitos aleatórios e modelagem de covariância residual.

2.4 Seleção de modelos

Quando diferentes modelos são estimados aos dados analisados é preciso adotar critérios de seleção de ajuste, para identificar o modelo que melhor se ajusta aos dados disponíveis e fornece uma representação adequada do fenômeno em estudo (BALZARINI; MACCHIAVELLI; CASANOVES, 2004).

2.4.1 Teste da razão de verossimilhança

É utilizado para comparar dois modelos com base nas suas funções de verossimilhança, no qual um modelo complexo (com mais parâmetros) é confrontado a um modelo restrito (com o número de parâmetros reduzidos). Sendo assim, o teste tem como hipótese nula que o modelo restrito é mais adequado, em contra partida, a hipótese alternativa é que o modelo complexo representar melhor os dados. A estatística do teste é obtida através da fórmula

$$\xi_{RV} = -2[\log(l(\tilde{\theta})) - \log(l(\hat{\theta}))] \quad (2.15)$$

em que, $\theta = (\beta', \alpha')'$ é o vetor de todos os parâmetros do modelo, $\log(l(\tilde{\theta}))$ é o logaritmo da verossimilhança do modelo reduzido e $\log(l(\hat{\theta}))$ do modelo com mais parâmetros. A razão de verossimilhança segue uma distribuição qui-quadrado assintoticamente, com graus de liberdade igual à diferença do número de parâmetros estimados entre os dois modelos χ_r^2 . Quando não se rejeita a hipótese nula isso significa que o modelo mais simples explica tão bem quanto um modelo mais completo, então, pelo princípio da parcimônia escolhe o modelo menos complexo.

Entretanto, esse teste é empregado apenas quando um dos dois modelos comparados é um caso especial do outro, além disso, conforme Nobre (2004) ele não é propício para avaliar os efeitos fixos quando os parâmetros dos modelos são estimados pelo MVR, isso porque o método exclui tais efeitos.

2.4.2 Critérios de informação

O AIC (*Akaike Information Criterion*) e o BIC (*Bayesian Information Criterion*) são dois critérios comumente utilizados para seleção de modelos, onde ambos penalizam a verossimilhança com a finalidade de escolher o modelo mais parcimonioso, ou seja, que apresentam o menor número de parâmetros possível e ao mesmo tempo explique bem o comportamento da variável dependente. Desta maneira, evita a inclusão de parâmetros irrelevante e complexidades desnecessárias.

O AIC é calculado usando a seguinte fórmula:

$$AIC = -2L(\hat{\theta}) + 2d \quad (2.16)$$

em que, $L(\hat{\theta})$ é o logaritmo da função de máxima verossimilhança (completa ou restrita) e $d = p + q$ é o número de parâmetros do modelo avaliado que leva em conta o número de

parâmetros estimados para o efeito fixo p e aleatório q . Quanto menor o valor do AIC, melhor o modelo é considerado em termos de ajuste aos dados.

O BIC também penaliza a complexidade do modelo, mas de forma mais rigorosa do que o AIC, isso porque ao adicionar $\ln(n)$ no qual n é o tamanho da amostra juntamente com d aumenta a penalização à medida que o tamanho da amostra aumenta, favorecendo modelos mais simples.

O BIC é calculado usando a seguinte fórmula:

$$BIC = -2L(\hat{\theta}) + d\ln(n) \quad (2.17)$$

Segundo Pinheiro e Bates (2000) quando os parâmetros dos modelos são estimados pelo MVR os resultados do AIC, BIC e log-verossimilhança, apenas podem ser comparados modelos que expressam a mesma estrutura de efeito fixos. Além disso, caso os critérios de informação indique seleção de modelos diferentes no contexto de MVR, é sugerido por Gurka (2006) o uso do BIC para a escolha do modelo.

2.5 Análise de resíduos

Os resíduos são estimações dos erros, a análise destes é fundamental para avaliar se o modelo pode representar com precisão ou não os dados. Tal análise é feita com intuito de verificar se os resíduos segue as pressuposições de normalidade, homocedasticidade, independência dos erros e verificar se existe valores discrepantes. Como os modelos mistos lineares apresentam duas fonte de variações (os efeitos aleatórios \mathbf{u} e os erros marginais ξ), dessa forma, há mais de um tipo de resíduo, em que cada um é conveniente para verificar algum tipo de suposição (NOBRE, 2004). Três tipos de vetores de erros podem ser definidos:

- **Erros condicionais:** $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbb{E}[\mathbf{Y}|\mathbf{u}] = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}$;

Sendo assim, os resíduos são obtidos pela diferença entre os valores observados e os valores ajustados (previstos) pelo modelo, $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$. Eles são denominados de condicionais, porque refletem a componente de efeitos fixos e efeitos aleatórios do modelo. São úteis para verificar se os efeitos aleatórios estão modelando a estrutura de dependência dos dados e se o modelo está capturando as variações entre os grupos. Pinheiro e Bates (2000) aconselha o uso de $\hat{\boldsymbol{\varepsilon}}$ contra os valores de $\hat{\mathbf{Y}}$ para verificar os pressupostos de normalidade, homocedasticidade e pontos discrepantes. Porém Nobre (2004) sugere a padronização deste resíduo, já que este pode apresentar diferentes variâncias por ser um resíduo confundido, pois depende de outros tipos de erros (HILDEN-MINTON, 1995).

- **Erros marginais:** $\boldsymbol{\xi} = \mathbf{Y} - \mathbb{E}[\mathbf{Y}] = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$

Nos quais os resíduos são resultados de $\hat{\boldsymbol{\xi}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Z}\hat{\mathbf{u}} + \boldsymbol{\varepsilon}$, a principal diferença entre os resíduos condicionais e os resíduos marginais é que estes últimos são calculados

ignorando os efeitos aleatórios do modelo. Esse tipo de resíduo é útil para avaliar a qualidade dos efeitos fixos do modelo, ou seja, quão bem as covariáveis estão explicando as variações nos dados. Para avaliar a pressuposição de linearidade entre $\mathbb{E}[\mathbf{Y}]$ e as variáveis independentes \mathbf{X} , Hilden-Minton (1995) recomenda criar um gráfico para dos resíduos $\hat{\xi}$ versus os valores das covariáveis.

- **Efeitos aleatórios:** $\mathbf{Zu} = \mathbb{E}[\mathbf{Y}|\mathbf{u}] - \mathbb{E}[\mathbf{Y}]$, é o desvio entre o valor predito e o valor médio populacional estimado, são úteis para analisar se existe unidades experimentais discrepantes (NOBRE, 2004).

2.6 Teste de comparação múltipla

Os testes de comparação múltipla são utilizados para comparar diferentes médias, no caso paramétrico, e determinar se há diferenças estatisticamente significativas entre elas. O teste empregando neste trabalho foi proposto por Rienzo, Guzman e Casanoves (1999). DGC é um método baseado em *cluster* (agrupamento) para identificar grupos de médias não homogêneas.

Como descrito por Rienzo, Guzman e Casanoves (1999), considera-se x_1, \dots, x_k um conjunto de médias não correlacionadas calculadas a partir de amostras aleatórias de tamanho n de distribuições normais $N(\mu_i, \sigma^2)$, $i = 1, \dots, k$, onde é definida uma matriz $\mathbf{D} = \{d_{ij}\}$ em que $d_{ij} = |x_i - x_j| / \sqrt{(s^2/n)}$ sendo s^2 a variância comum dentro dos grupos. De forma, que é aplicado o método de agrupamento na matriz \mathbf{D} , na qual são construídos grupos de médias homogêneas, ou seja, que são semelhantes, cujos grupos formados são heterogêneo entre eles. Um dendograma é gerado, em que os nós terminais são as k médias de tratamento, onde os agrupamentos são arranjados conforme a distância entre as médias, isto é, distâncias menores significa que as médias são mais parecidas. A distância euclidiana entre elas é calculada como

$$q(S_k, S_L) = \frac{1}{\#(S_k)\#(S_L)} \sum_{x_i \in S_k; x_j \in S_L} d_{ij}$$

Sendo S_K e S_L *clusters* distintos, senão $q(S_k, S_L) = 0$. O critério de corte de nível α para o dendograma é obtido por $c = EPM \times Q_{K,n;1-\alpha}$, em que EPM é uma estimativa do erro padrão de média, a distribuição Q é explicada por Rienzo, Guzman e Casanoves (1999), o dendograma termina no nó raiz à distância de q_{K-1} , se Q for a distância do nó raiz da variável aleatória, o quantil $1 - \alpha$ da distribuição de Q sob a hipótese $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ poderia ser usado para construir um teste de tamanho α , deste modo, qualquer valor $Q \geq Q_{1-\alpha}$ levaria à rejeição desta hipótese. Como Q é uma média dessas distâncias (sob ligação média), a distribuição nula de Q depende de k e n , os 95 e 99% pontos percentuais da distribuição nula de Q , foi obtida por simulação de Monte Carlo (Tabelas I e II do Apêndice do artigo de Rienzo, Guzman e Casanoves (1999)), para um número selecionado de tratamentos (k) e repetições (n). Esse tipo de teste é indicado quando muitas médias de tratamento são comparadas, pois os testes de comparações tradicionais sofrem com a sobreposição.

Este teste de comparação múltipla se encontra implementado no *software* estatístico *InfoStat*, no qual é utilizado por Rienzo, Macchiavelli e Casanoves (2017) na comparação de tratamentos para modelos lineares mistos.

3 RESULTADOS E DISCUSSÕES

Os dados usados no estudo estão apoiados sob a tese Hernández (1997), que consiste em um experimento agrônômico, onde investiga os rendimentos de catorze linhagens de feijão tepary (*Phaseolus acutifolius Gray*), que passou por melhoramento genético, em três localidades semiáridas do Departamento de Progreso, na Guatemala.

Esta espécie de leguminosa da família do feijão Fabaceae, é nativa do sudoeste dos Estados Unidos e do norte do México, onde é cultivada há milhares de anos pelos povos indígenas. Os feijões tepary são adaptáveis aos climas áridos e quentes, o que os torna resistentes a condições difíceis de cultivo.

Tendo em vista isso, o objetivo do autor era saber o rendimento em kg/ha dos cultivares do feijão tepary nas localidades de Jícaro; Estância, San Cristóbal A.C e Rancho, San Agustín Acasaguastlán, pois as linhagens que apresentassem os melhores rendimentos seriam uma alternativa para produção de feijão nos municípios. Essas regiões sofrem com a má distribuição das chuvas, isso porque os locais são cercados de montanhas e os ventos oriundos do Mar do Caribe direciona as nuvens para as partes mais altas onde ocorre a precipitação.

3.1 Experimento

Para a coleta dos dados da análise foi montado um experimento em blocos completos casualizados, com 4 repetições (blocos) e 14 tratamentos (linhagens) nas três localidades, em que era averiguado os rendimentos de cada linhagem.

Conforme descrito por Hernández (1997) a unidade experimental corresponde a um tratamento onde se semeou três grãos de feijão por postura experimental a cada 0,30m entre plantas, com espaçamento entre sulcos de 0,40m . A parcela experimental tinha uma área total de 451,20m², com 48m de comprimento e 9,40m de largura, já a parcela útil com efeito de 0,5m de borda e 1m cabeça a área era 399m². Os tratamentos foram distribuídos aleatoriamente nos blocos, e os terrenos foram preparados com enxadas.

A semeadura aconteceu em agosto/setembro, nas quais as linhagens do feijão utilizada para avaliação foram:

1. L-18 semente preta, coletada no mercado de Hermosillo, Sonora, México;
2. L-242-43 semente preta,coletada no mercado de Hermosillo, Sonora, México;
3. L-242-7 semente preta,coletada no mercado de Hermosillo, México;
4. L-39 semente preta,coletada no mercado de Hermosillo, México;
5. L-246-19 colheita, comprada no mercado de Hermosillo Sonora, México, setembro de 1976;

6. L-242-22 semente separada de L-242, coletada no mercado de Hermosillo, Sonora, México;
7. L-242-9 semente separada de L-241, coletada no mercado de Hermosillo, Sonora, México;
8. L-246-9 Tepary cultivado, coletado no mercado de Hermosillo, Sonora, México, setembro de 1976;
9. L-242-25 semente separada de L-242, coletado no mercado de Hermosillo, Sonora, México;
10. L-242-38 semente separada de L-242, coletado no mercado de Hermosillo, Sonora, México;
11. L-38 semente branca, coletada no mercado de Hermosillo, Sonora, México;
12. L-30 semente de cor marrom, coletada no mercado de Hermosillo, Sonora, México;
13. L-35 semente de cor marrom, coletada no mercado de Hermosillo, Sonora, México;
14. L-37 semente de cor marrom, coletada no mercado de Hermosillo, Sonora, México.

A colheita ocorreu quando a plantação alcançou a maturidade fisiológica e principalmente quando 90% das vargens estavam secas. O rendimento de grãos da parcela foi determinada para cada tipo de tratamento e expressa em kg/ha (HERNÁNDEZ, 1997).

3.2 Estatística Descritiva

Para resumir os dados foram calculados a média de rendimento, cuja é a medida onde a maioria das observações está próxima, e para examinar a variabilidade dos valores em torno da média utiliza-se o desvio padrão, a qual é a raiz quadrada positiva da variância, como retratado por Bussab e Morettin (2017) tal medida de dispersão pode ser interpretada como desvio cometido ao tentar substituir cada observação pela média, também foi usado a medida-resumo mediana que corresponde o valor central que divide a amostra em duas partes iguais.

Tabela 1 – Análise descritiva do rendimento do feijão tepary em kg/ha nos três locais onde foram montados os experimentos em blocos completos casualizados.

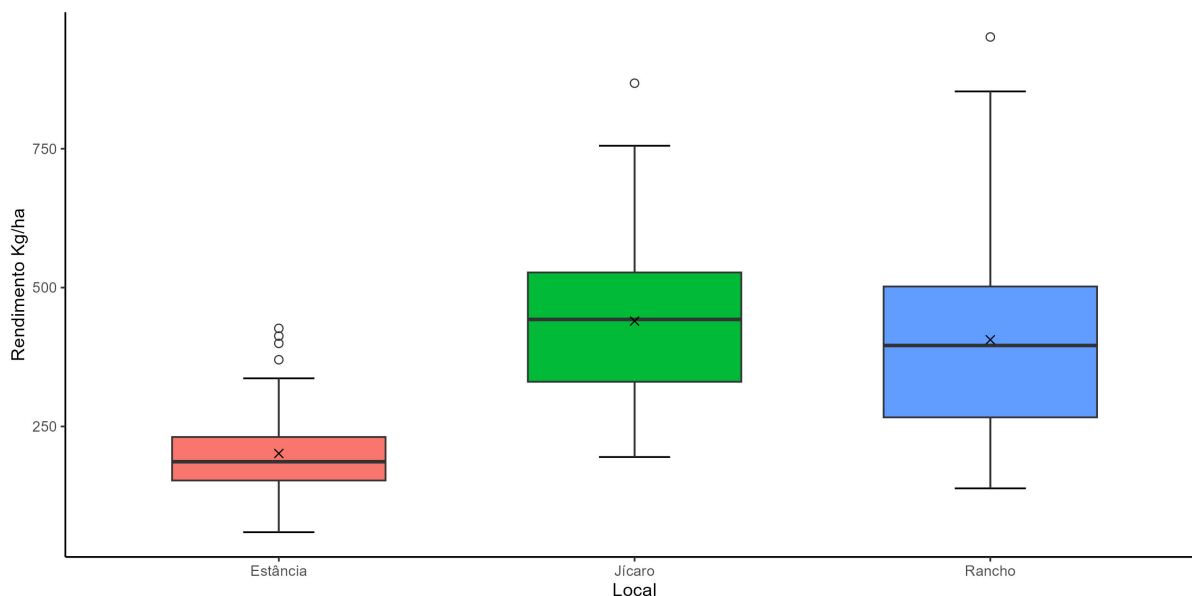
Local	Média	Desvio padrão	Mediana
Estância	201,3	77,99	186,5
Jícaro	439,7	155,79	442,7
Rancho	406,11	187,24	395,76

Fonte: Elaborada pela autora, 2023.

Mediante a Tabela 1, em média o município que apresentou maior rendimento foi Jícaro com 439,7 kg/ha e desvio padrão de 155,79 kg/ha, seguida por Rancho com média igual 406,11

kg/ha com desvio para mais ou para menos de 187,24 kg/ha e o menor rendimento médio comparado aos demais locais foi de Estância com 201,3 kg/ha e desvio padrão de 77,99 kg/ha.

Figura 1 – Rendimento do feijão tepary em kg/ha de acordo com os três locais do experimento.



Fonte: Elaborada pela autora, 2023.

As estatísticas anteriores podem ser representadas graficamente por meio dos *box plots*. Pela a Figura 1, é possível notar que a variabilidade do rendimento entre os locais se destoam, o que sugere que a variância entre eles não são homogêneas.

Tabela 2 – Análise descritiva do rendimento das linhagens de feijão tepary em kg/ha conforme os locais onde foram montados os experimentos em blocos completos casualizados.

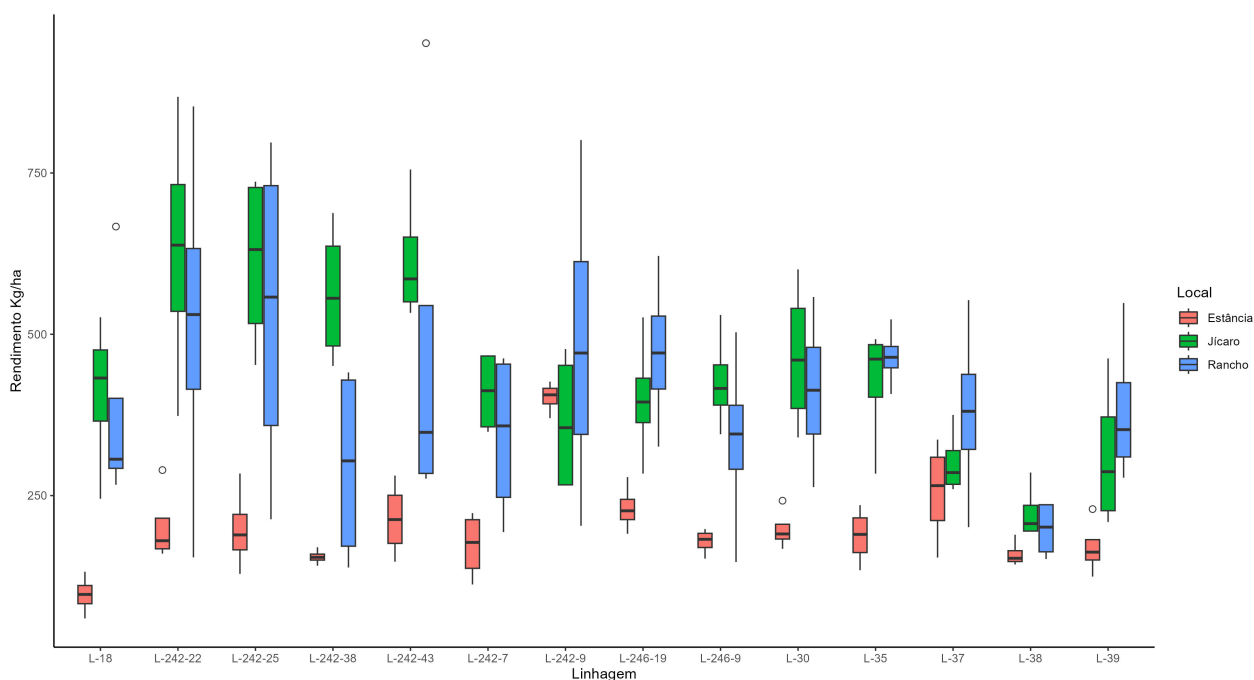
Linhagem	Estância		Jícara		Rancho	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
L-18	96,28	30,04	408,96	119,9	386,6	187,86
L-242-22	202,39	59,41	629,44	205,98	517,13	286,75
L-242-25	197,72	64,88	612,9	140,31	531,42	269,89
L-242-38	154,94	11,75	562,63	110,07	296,75	158,44
L-242-43	213,51	58,93	615,07	99,77	480,86	319,28
L-242-7	172,45	52,38	410,32	65,16	343,02	134,58
L-242-9	402,25	24,11	363,29	112,82	486,52	253,06
L-246-19	230,55	36,55	400,05	99,16	472,33	122,51
L-246-9	178,71	19,99	426,79	77	335,26	145,93
L-30	197,64	31,62	465,18	117,18	411,86	124,77
L-35	187,29	44,11	424,87	96,36	464,8	47,26
L-37	255,34	80,7	301,54	52,04	378,82	144,49
L-38	159,55	20,59	223,4	42,89	197,53	44,77
L-39	169,54	43,61	311,42	116,09	382,64	118,8

Fonte: Elaborada pela autora, 2023.

A Tabela 2 traz, que na localidade Estância a linhagem que exibiu maior rendimento médio foi L-242-9 com 402,25 kg/ha, por outro lado, L-18 teve o menor rendimento médio

na região com 96,28 kg/ha. No município de Jícaro o maior valor médio foi de L-242-22 com 629,44 kg/ha e a menor média foi da linhagem L-38 com 223,4 kg/ha. O experimento no local de Rancho, a maior média de rendimento foi de L-242-25 com 531,42 kg/ha, já L-38 apresentou a menor média com 197,53 kg/ha. A Figura 2 transmite a análise gráfica da dispersão.

Figura 2 – Análise gráfica do rendimento das linhagens de feijão tepary em kg/ha conforme os três locais onde foram montados os experimentos em blocos completos casualizados.



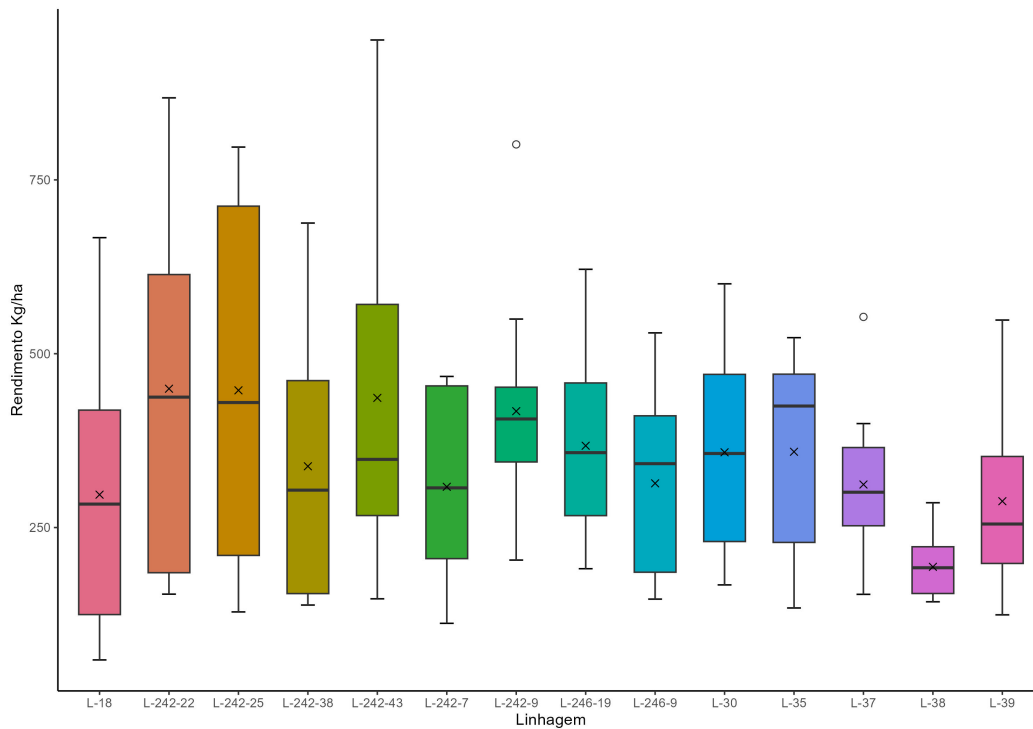
Fonte: Elaborada pela autora, 2023.

Tabela 3 – Rendimento das catorze linhagens de feijão tepary em kg/ha.

Linhagem	Média	Desvio padrão	Mediana
L-18	297,28	189,53	283,73
L-242-22	449,66	265,71	437,56
L-242-25	447,35	248,15	429,79
L-242-38	338,11	203,32	303,8
L-242-43	436,48	248,71	348
L-242-7	308,6	133,34	307,05
L-242-9	417,35	154,85	406,15
L-246-19	367,64	135,43	357,7
L-246-9	313,59	137,76	341,86
L-30	358,23	151,15	356,48
L-35	358,99	141,57	424,74
L-37	311,9	105,07	300,8
L-38	193,49	43,75	192,17
L-39	287,87	128,85	255,06

Fonte: Elaborada pela autora, 2023.

Figura 3 – Análise gráfica do rendimento das catorze linhagens de feijão tepary em kg/ha.



Fonte: Elaborada pela autora, 2023.

Analisando a Tabela 3 e graficamente a Figura 3, a linhagem que apresenta menor rendimento médio (a média é representada pelo símbolo ×) foi L-38 com um valor entorno de 193,49 kg/ha. Já as que retrataram maiores produtividades foram L-242-22, L-242-25, L-242-43 e L-242-9 com rendimento médio respectivamente de 449,66 kg/ha; 447,35 kg/ha, 436,48 kg/ha e 417,35. Entretanto, precisa-se investigar se essas diferenças são estatisticamente significativas.

3.3 Modelos Estatísticos

Os modelos estatísticos são usados para entender a associação entre as variáveis, isto é, mede impacto de uma variável sobre outra, além disso, ajudam a determinar se as evidências observadas são estatisticamente significativas ou apenas resultado do acaso. Desta maneira, o intuito é identificar das 14 linhagens do feijão tepary as que mostraram melhores rendimentos e quais delas são mais produtivas para cada local.

O seguinte modelo foi empregado por Hernández (1997), para explicar a complexidade dados:

$$Y_{ijk} = \mu + L_i + B_{ij} + \tau_k + (L\tau)_{ik} + E_{ijk} \quad \begin{cases} i = 1, 2, 3 & \text{localidade,} \\ j = 1, 2, 3, 4 & \text{bloco,} \\ k = 1, 2, \dots, 14 & \text{tratamento.} \end{cases}$$

em que, Y_{ijk} é a variável resposta, isto é, o rendimento na localidade i , no bloco j , do tratamento k ; μ a média geral; L_i efeito da localidade i ; B_{ij} efeito do bloco j aninhado a localidade i ; τ_k efeito do tratamento k ; $(L\tau)_{ik}$ efeito da interação entre tratamento e localidade; E_{ijk} erro da unidade experimental.

As hipóteses de interesse a serem testados são:

- Para o fator linhagem

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_{14}$$

$$H_1 : \tau_1 \neq \tau_2 \neq \dots \neq \tau_{14} \quad \text{para pelo menos um par com } k \neq k'$$

- Para o fator Local

$$H_0 : L_1 = L_2 = L_3$$

$$H_1 : L_1 \neq L_2 \neq L_3 \quad \text{para pelo menos um par com } i \neq i'$$

- Para a interação entre linhagem e localidade

$$H_0 : (L\tau)_{ik} = 0 \quad \forall i, k$$

$$H_1 : (L\tau)_{ik} \neq 0 \quad \text{para pelo menos um par com } ik$$

Para esse experimento os efeitos dos blocos aninhados aos locais não é relevante para a prática. Agora, tomando o nível de significância $\alpha = 0,05$; rejeita-se a hipótese nula quando P-valor, probabilidade de se adquirir um valor da estatística de teste superior ou igual ao observado, é menor do que o nível de significância.

O modelo convencional utilizado por Hernández (1997) teve como resultado da análise variância as seguintes estatísticas presente na Tabela 4:

Tabela 4 – Análise da variância do modelo usado por Hernández (1997), para testar a igualdade das médias dos tratamentos em relação aos rendimentos de feijão tepary.

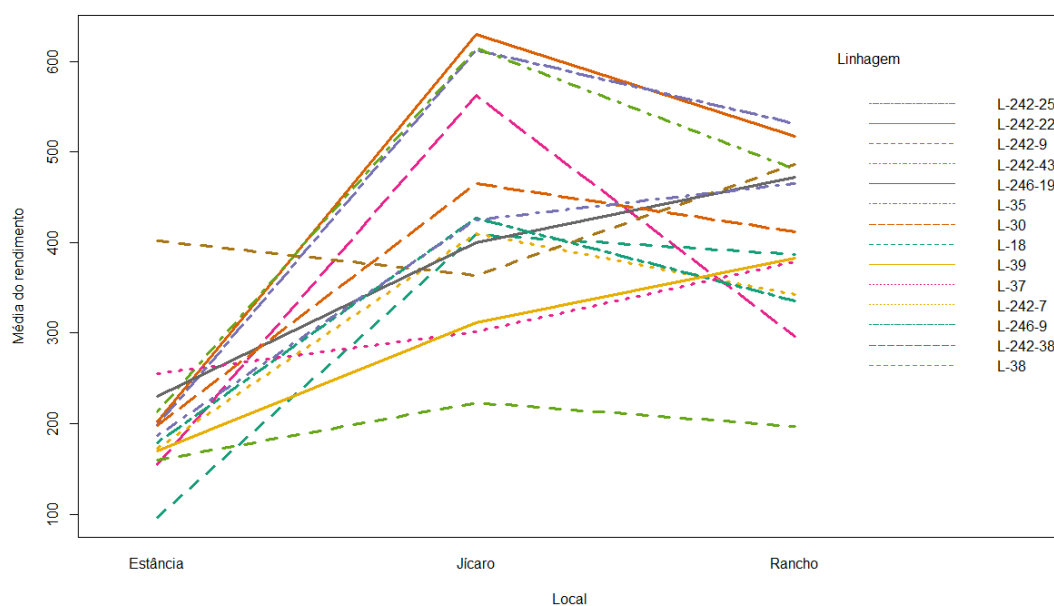
Fonte de Variação	G.L	S.Q.	Q. M.	F	P-valor
Linhagem	13	811652	62435	3,604	<0,001
Local	2	1865057	932529	53,829	<0,001
Linhagem x local	26	709704	27296	1,576	0,0535
Blocos(local)	9	49329	5481	0,316	0,9683
Resíduo	117	2026910	17324		
Total	167	5462652			

Fonte: Elaborada pela autora, 2023.

Deste forma, ao nível de 0,05 de significância existem evidências estatísticas que as linhagens e os locais apresentam efeitos diferentes sobre o rendimento médio do feijão tepary em

kg/ha, já que o p-valor de ambos os fatores são inferior a $\alpha = 0,05$. Por outro lado, a interação entre eles não é significativa, pois $p\text{-valor}=0,0535 > 0,05$; o que indica que os efeitos das linhagens sobre o rendimento médio não depende da localidade. Porém, através da análise do gráfico da interação (Figura 4) entre os fatores é possível notar que há interação complexa entre eles, isto é, as linhagens respondem de maneira diferente aos locais, já que existe intersecção das linhas que representam as linhagens em cada localidade, caso não houve interação as linhas seriam paralelas.

Figura 4 – Interação entre os fatores local e linhagem.

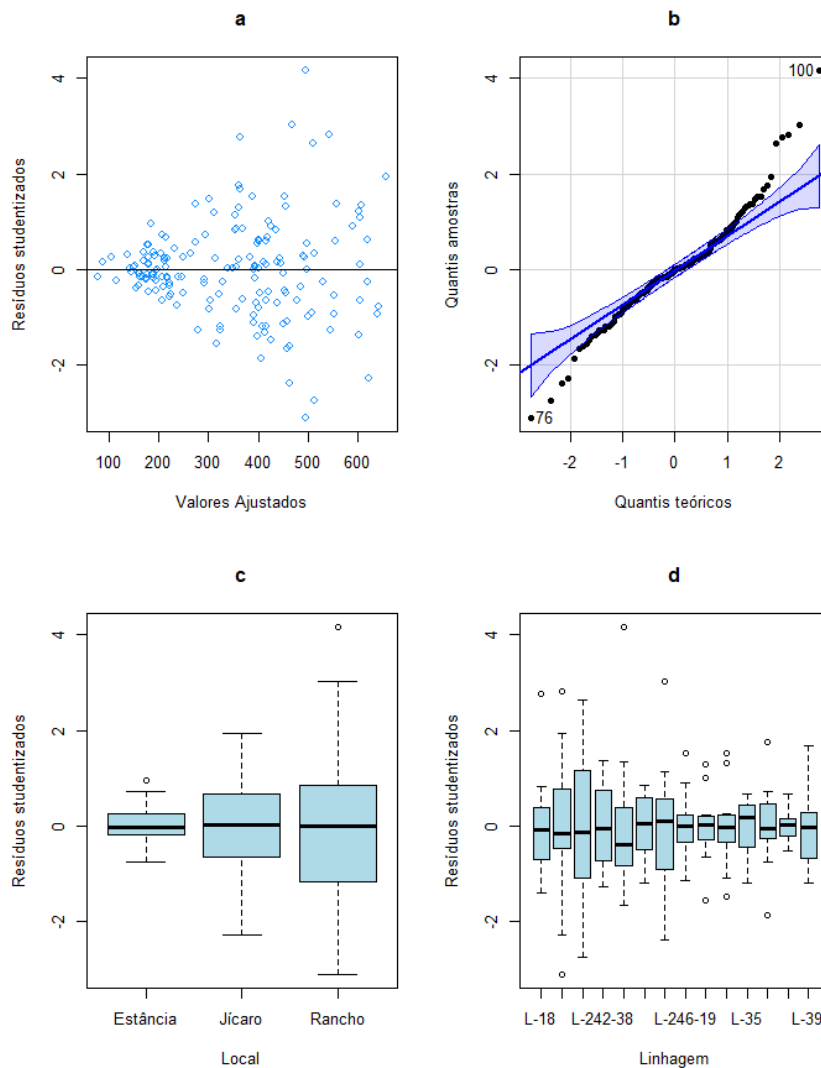


Fonte: Elaborada pela autora, 2023.

Portanto, é preciso investigar o motivo que está provocando essa discordância entre os resultados da análise da variância e o gráfico de interação, então, levando em consideração isso, para saber se o modelo é apropriado para os dados, e assim garantir que as inferências feitas sejam válidas e confiáveis, é necessário que tal modelo atenda as suposições de normalidade, homocedasticidade e independência dos resíduos.

A normalidade dos resíduos é importante, pois testes de hipóteses e intervalos de confiança assumem distribuição normal para validação da análise. A homogeneidade da variância é crucial, ela é considerada uma das mais graves quebras de suposição (ALBUQUERQUE, 2016), se as variâncias dos resíduos forem diferentes entre os grupos, os resultados sobre os efeitos dos fatores podem ser afetados. Já a independência dos resíduos assegura que não tenha influência entre as observações dentro de cada bloco, pois caso houver correlação isso pode acarretar a inferências enviesadas e imprecisas. Esses importantes pressupostos podem ser verificados por meio de gráficos.

Figura 5 – Análise dos resíduos para o modelo utilizado na tese de Hernández (1997).



Fonte: Elaborada pela autora, 2023.

O gráfico (a) da Figura 5, analisa a homocedasticidade (resíduos studentizados versus valores ajustados), para ser atendido o critério os pontos (resíduos) devem estar dispersos aleatoriamente em torno de zero. Nesta situação, como a dispersão dos pontos apresenta um padrão, formato “corneta” fechada à esquerda e aberta à direita, isso sugere que a homogeneidade das variâncias não está sendo atendida.

O gráfico (b) da Figura 5 testa a normalidade dos resíduos, se a distribuição deles é normal, os pontos devem estar sobrepostos a linha azul e dentro do envelope. Portanto, os resíduos do modelo em questão aparentemente não seguem distribuição normal, a falta de normalidade dos resíduos pode ser consequência da quebra da suposição de homogeneidade da variância;

Já os gráficos (c) e (d) da Figura 5, são referentes a análise da homocedasticidade conforme os fatores local e linhagem, respectivamente. A heterocedasticidade pode ter como maior causa as distintas variâncias para os municípios Estância, Jícara e Rancho.

Tendo em conta a análise dos resíduos do modelo usado por Hernández (1997), observa-se que as conclusões e inferências sobre o estudo não são válidas, em virtude que os pressupostos não foram atendidos, que como consequência a interação entre linhagem e local é afetada, logo o modelo não é adequado para explicar a complexidade do experimento. Por esse motivo, é importante a análise dos resíduos do modelo proposto, para verificar o ajuste deste aos dados, antes das interpretações dos resultados.

Com o objetivo de investigar a causa da falta de homocedasticidade, uma análise individual para cada local foi feita, onde o modelo estatístico considerado foi:

$$Y_{ij} = \mu + \tau_i + B_j + E_{ij} \quad \begin{cases} i = 1, 2, \dots, 14 & \text{tratamento,} \\ j = 1, 2, 3, 4 & \text{bloco.} \end{cases}$$

onde, Y_{ij} é a variável resposta; μ a média geral; τ_i efeito do i -ésimo tratamento; B_j efeito j -ésimo bloco; E_{ij} erro da unidade experimental.

Por conseguinte, foi aplicado o teste de Shapiro-Wilk para verificar a normalidade dos resíduos, as hipóteses a serem testadas são, H_0 : Os resíduos seguem uma distribuição normal; H_1 : Os resíduos não seguem uma distribuição normal. Já para avaliar a homogeneidade foi usado o teste de Bartlett, que tem como hipóteses, H_0 : Os resíduos são homocedásticos; H_1 : Os resíduos não são homocedásticos, é recomendado o uso desse teste apenas se os resíduos seguir normalidade. Ambos os teste foi avaliado ao nível de 5% de significância.

Tabela 5 – P-valores dos testes de Shapiro-Wilk e Bartlett conforme cada localidade onde foram montados os experimentos em blocos completos casualizados.

Teste	Estância	Jícaro	Rancho
Shapiro-Wilk	0,926	0,640	0,344
Bartlett	0,174	0,824	0,197

Fonte: Elaborada pela autora, 2023.

Diante da Tabela 5, há evidências que os pressupostos referente a normalidade e homogeneidade da variância dos resíduos foram atendidos, pois as hipóteses nulas dos testes nos três locais não são rejeitadas, já que os p-valores são superiores a 0,05, o que leva a concluir que a falta de homogeneidade pode ser provocada pelos diferentes níveis (municípios) do fator localidade.

3.3.1 Modelos lineares mistos

Desta maneira, pensando no resultado da análise dos resíduos do modelo estatístico usado por Hernández (1997), nos quais não são homogêneos e como consequência disso eles também não segue distribuição normal, os modelos lineares mistos são uma alternativa, por serem mais flexíveis em relação às suposições sobre os resíduos, por exemplo, a heterocedasticidade pode ser trabalhada por meio da modelagem das matrizes de variância e covariância \mathbf{G} e \mathbf{R} .

O primeiro passo para construir o modelo linear misto é definir o que serão efeito fixo e aleatório, e se preciso escolher uma estrutura de covariância. Segundo Hajduk e Gallois (2022), os efeitos aleatórios são fatores categóricos de agrupamento, a qual o interesse não é medir o impacto deste na variável resposta, porém sabe-se que ele pode ter influência sobre a variabilidade inerente aos dados. Haja em vista isso, será tratado como efeito aleatório os blocos aninhados aos locais, uma vez que o objetivo não é comparar os rendimentos médios de blocos aninhados a cada local, mas saber o quanto a variação é atribuível a isso.

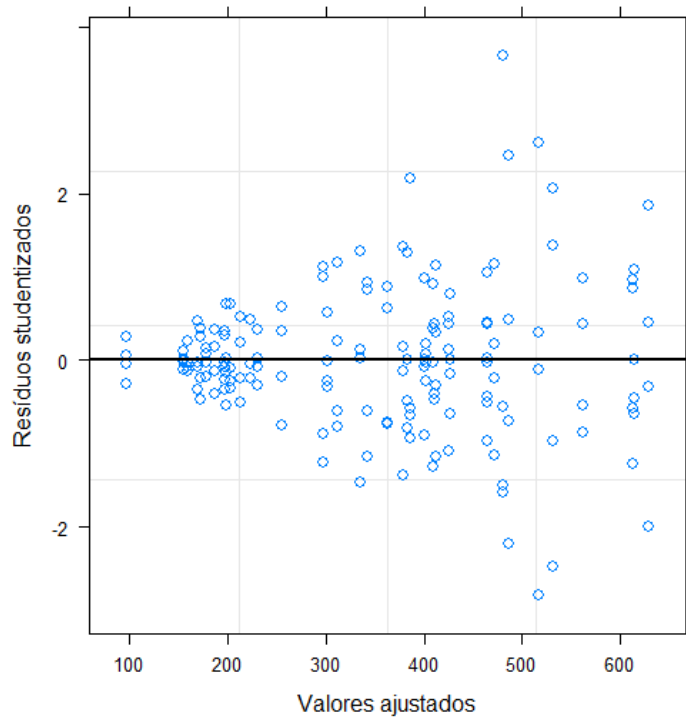
Já os efeitos fixos, os níveis dos fatores são escolhidos pelo pesquisador, devido que a finalidade é fazer inferência sobre as médias específicas de cada nível, pois se acredita que tal fator tenha efeito sob a variável resposta. Desta forma, como o propósito é saber rendimento médio das linhagens, dos locais e da interação entre esses dois fatores, então eles são retratados como efeitos fixos.

Sendo assim, o primeiro modelo linear misto ajustado (nomeado por modelo A), pelo método da máxima verossimilhança restrita, tem como efeitos fixos os fatores linhagens, local e interação entre eles, e os blocos aninhados aos locais são tratados como aleatórios, já os erros foram considerados independentes e variância constante entre as localizações.

$$\begin{aligned}
 \text{(modelo A): } Y_{ijk} = \mu + L_i + B_{ij} + \tau_k + (L\tau)_{ik} + \varepsilon_{ijk} & \quad \left\{ \begin{array}{l} i = 1, 2, 3 \quad \text{localidade,} \\ j = 1, 2, 3, 4 \quad \text{bloco,} \\ k = 1, 2, \dots, 14 \quad \text{linhagem.} \end{array} \right.
 \end{aligned}$$

onde, Y_{ijk} é a variável resposta, isto é, o rendimento na localidade i , no bloco j , da linhagem k ; μ a média geral; B_{ij} o efeito do bloco j aninhado a localidade i ; τ_k efeito da linhagem k ; $(L\tau)_{ik}$ efeito da interação entre linhagem e localidade; ε_{ijk} erro da unidade experimental. Em que, $B_{ij} \sim N(0, \sigma_B^2)$ e $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

Figura 6 – Análise da variância dos resíduos condicionais studentizados do modelo A.



Fonte: Elaborada pela autora, 2023.

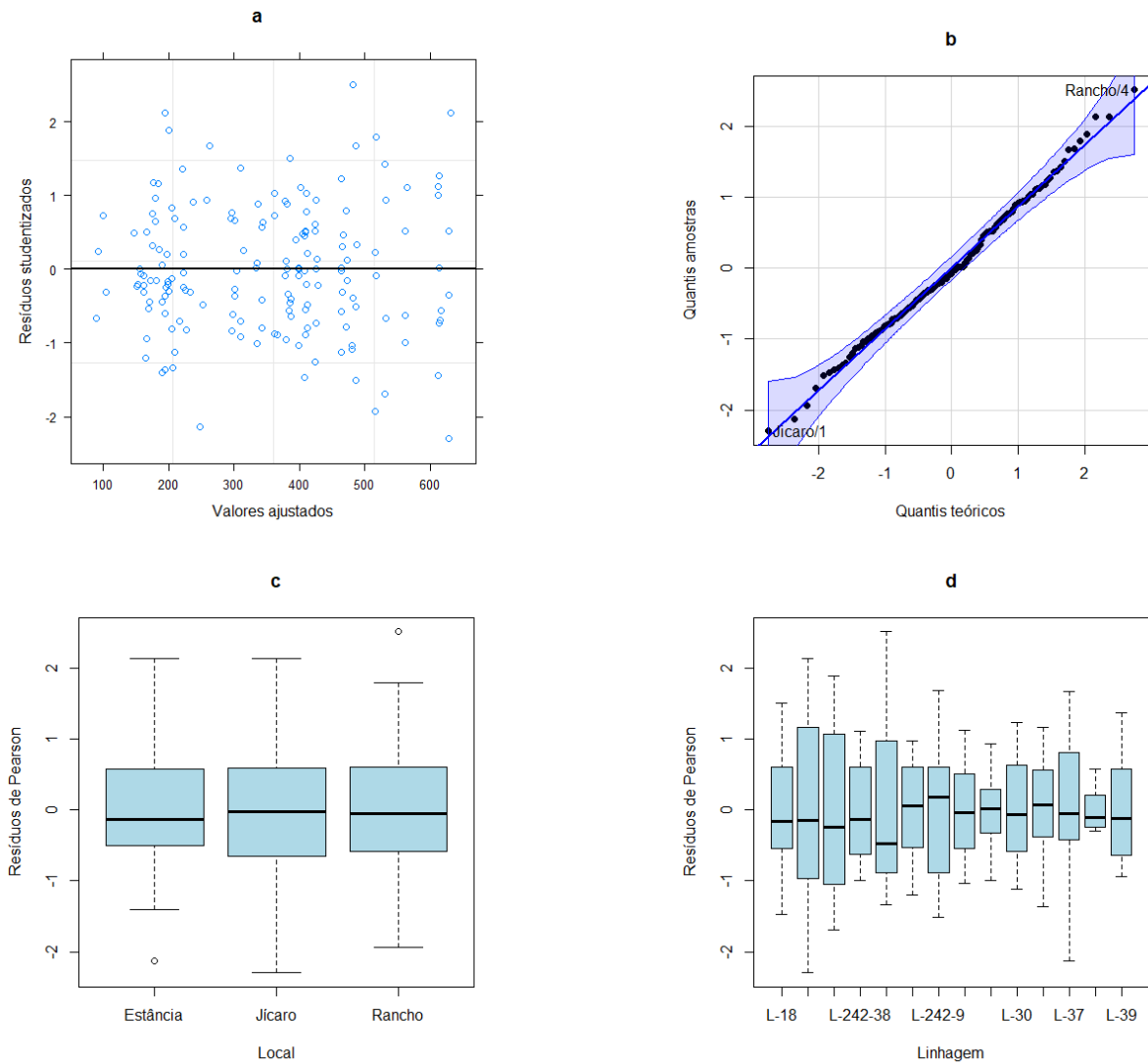
No entanto, apenas o efeito aleatório não é suficiente para explicar a heterocedasticidade inerente ao problema, isso é nítido no gráfico da Figura 6, onde os resíduos do modelo A apresenta um padrão formato "corneta".

Então, um (modelo B) com os mesmos efeitos fixos e aleatório do modelo A, estimado também pela MVR foi criado, todavia, a sua estrutura de variância foi modelada pensando na heterogeneidade provocada pelo fator local, isso foi feito usando uma função de variância, $\sigma_y^2 = \sigma^2 \delta_{Sij}^2$, que considera diferentes variâncias por níveis de um fator, neste caso, diferentes variações para Estância, Jícara e Rancho, os erros foram considerados não correlacionados.

$$\text{(modelo B): } Y_{ijk} = \mu + L_i + B_{ij} + \tau_k + (L\tau)_{ik} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, 3 & \text{localidade,} \\ j = 1, 2, 3, 4 & \text{bloco,} \\ k = 1, 2, \dots, 14 & \text{linhagem.} \end{cases}$$

onde, Y_{ijk} é a variável resposta, isto é, o rendimento na localidade i , no bloco j , da linhagem k ; μ a média geral; B_{ij} o efeito do bloco j aninhado a localidade i ; τ_k efeito da linhagem k ; $(L\tau)_{ik}$ efeito da interação entre linhagem e localidade; ε_{ijk} erro da unidade experimental. Em que, $B_{ij} \sim N(0, \sigma_B^2)$ e $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, e diferentes variâncias são levados em conta para Estância, Jícara e Rancho, $\sigma_y^2 = \sigma^2 \delta_{Sij}^2$.

Figura 7 – Análise da homogeneidade e normalidade dos resíduos condicionais para o modelo B, com heterogeneidade corrigida pela função de variância $\sigma_y^2 = \sigma^2 \delta_{Sij}^2$.



Fonte: Elaborada pela autora, 2023.

Por meio do gráfico (a) da Figura 7, é possível notar que o pressuposto da homogeneidade dos resíduos foram alcançados com a modelagem da variância para o fator local, já que os pontos estão dispersos de forma aleatória entorno de zero (linha horizontal). Além disso, os resíduos atendem a suposição de normalidade, pois, agora a maioria dos pontos estão sobrepostos na linha azul e dentro do envelope, como pode ser visto no gráfico (b) da Figura 7. Portanto, a heterocedasticidade é realmente dada pelas diferentes variâncias de locais, e modelos estatísticos convencionais seriam incapazes de capturar tal complexidade.

A fim de comparar o ajuste dos modelos A e B aos dados, foram empregados critérios estatísticos de seleção de modelo, representados na Tabela 6.

Tabela 6 – Estimativas dos critérios de ajuste, Akaike (AIC), Bayesiano de Schwarz (BIC), razão de verossimilhança com a probabilidade associada (P-valor), dos modelos A e B.

Modelos	Estimação	G.L.	AIC	BIC	log veross	-2log veross	P-valor
Modelo A	MVR	44	1727,230	1852,026	-819,6150		
Modelo B	MVR	46	1663,197	1793,666	-785,5986	68,03283	<0,0001

Fonte: Elaborada pela autora, 2023.

Os critérios AIC e BIC do modelo B são menores do que do modelo A, ou seja, ele apresenta melhor adequação aos dados, também foi realizada a comparação pelo teste de razão de verossimilhança, ao nível de 0,05 de significância, onde a hipótese nula é o modelo A contra a hipótese alternativa que é o modelo B, e como o p-valor é inferior 0,05, existe forte evidência de que o modelo B é estatisticamente melhor que o modelo A. Contudo, é importante salientar que os critérios mais indicado neste contexto seriam apenas AIC e BIC, pois ambos os modelos foram estimados pelo método da máxima verossimilhança restrita e é aconselhável por Nobre (2004) que o teste de razão de verossimilhança não é convincente neste caso.

Então, tendo em mente a análise dos resíduos e os critérios de seleção, o modelo B foi o que melhor se ajustou aos dados, atendendo as suposições exigidas, logo, ele será o modelo final, usado para obter as inferências desejadas de maneira confiável.

Tabela 7 – Testes de hipóteses sequenciais dos efeitos fixos e variância dos efeitos aleatórios ($\hat{\sigma}_B^2$) e residual ($\hat{\sigma}^2$) do modelo B.

Fonte de Variação	G.L. Numerador	G.L. Denominador	F	P-valor
Intercepto	1	117	1533,3386	<0,0001
Local	2	9	113,0581	<0,0001
Linhagem	13	117	9,0157	<0,0001
Linhagem x local	26	117	3,4015	<0,0001
		$\hat{\sigma}^2 = 1956,789$	$\hat{\sigma}_B^2 = 89,127$	

Fonte: Elaborada pela autora, 2023.

Ao nível de 0,05 de significância, tem-se que os fatores linhagem, local e a interação entre eles são estatisticamente significativos, p-valores menores que 0,05, isto é, pelo menos uma linhagem apresenta comportamento diferente sobre o rendimento médio kg/ha, assim também acontece para o fator local, ainda mais, há evidências que os efeitos das linhagens sobre o rendimento médio depende da localidade (mutuamente).

Mediante a análise da Tabela 7, as médias dos níveis dos fatores não são iguais, contudo, para identificar quais níveis se diferem entre si é necessário aplicar um teste de comparação múltipla, neste caso, será utilizado o teste de agrupamento DGC (Rienzo, Guzman e Casanoves). Médias acompanhadas da mesma letra não diferem pelo teste de comparação múltipla ao nível de 5% de significância.

Assim sendo, Estância apresentou o menor rendimento médio em kg/ha (201,30 b) de feijão tepary comparado aos demais, por outro lado, as localidades Jícaro (439,70 a) e Rancho (406,11 a) tiveram produções médias estatisticamente iguais.

Tabela 8 – Estimativas e comparações das médias em kg/ha das linhagens pelo teste de comparações múltiplas de agrupamentos DGC (Rienzo, Guzman e Casanoves). Médias acompanhadas da mesma letra não diferem pelo teste de comparação múltipla ao nível de 5% de significância.

Local	Média
L-242-22	449,66 a
L-242-25	447,35 a
L-242-43	436,48 a
L-242-9	417,35 a
L-246-19	367,65 b
L-35	358,99 b
L-30	358,23 b
L-242-38	338,11 b
L-246-9	313,58 b
L-37	311,90 b
L-242-7	308,60 b
L-18	297,28 b
L-39	287,87 b
L-38	193,49 c

Fonte: Elaborada pela autora, 2023.

A Tabela 8, apresenta que as linhagens L-242-22, L-242-25, L-242-43 e L-242-9 foram as que mostraram maiores rendimentos médios em relação às outras, além do mais, não há diferença entre elas pelo teste de comparação múltipla. Todavia, L-38 exibiu um rendimento muito baixo levando em conta as outras linhagens.

No município Jícaro as linhagens L-242-22 (629,44 a), L-242-43 (615,07 a) e L-242-25 (612,90 a) tiveram maiores destaque no rendimento médio na cultura da produção de feijão tepary, nas quais são estatisticamente similares. Em Rancho, todas as linhagens apresentou produção média iguais, exceto L-38 (197,53 c) que difere de todas e teve rendimento médio baixo. E para Estância as linhagens L-242-9 (402,25 b) e L-37 (255,34 b) foram a que mostraram melhores rendimentos médios kg/ha (a Tabela 9, referente a está análise se encontra no Apêndice A).

As estimativas dos componentes da variância em **G** e **R** pode ser representados por $\hat{\sigma}_B = 9,44$ que é o desvio padrão da variação entre os aninhamentos de blocos e locais e $\hat{\sigma} = 44,24$ que refere ao desvio padrão da variação residual que não são explicadas pelos efeitos fixos e aleatórios, os parâmetros estimados da função de variância para cada local são $\hat{\delta}_E = 1$; $\hat{\delta}_J = 2,51$; $\hat{\delta}_R = 4,23$.

4 CONSIDERAÇÕES FINAIS

Diante do exposto, o modelo linear misto mostrou ser mais adequado aos dados obtidos do experimento em blocos completos casualizados da tese de Hernández (1997), uma vez que retrata a heterocedasticidade inerente ao fator local, o que não é possível em modelos estatísticos convencional por ser mais restritivo em relação aos pressupostos dos resíduos.

Sendo assim, foi possível concluir, ano nível de 5% de significância, que os fatores linhagem, local e a interação são estatisticamente significativas (p -valor $<0,0001$). Portanto, foi aplicado o teste de comparações múltiplas de agrupamento DGC (Rienzo, Guzman e Casanoves), em que foi constatado, que o município Estância apresentou o menor rendimento médio de feijão tepary comparado aos demais, por outro lado, as localidades Jícaro e Rancho tiveram produções médias estatisticamente iguais.

As linhagens L-242-22, L-242-25, L-242-43 e L-242-9 foram as que mostraram maiores rendimentos médios em relação às outras, além disso, elas não se diferem entre elas, todavia, L-38 exibiu um rendimento muito baixo levando em conta as outras linhagens.

No município Jícaro as linhagens L-242-22, L-242-43 e L-242-25 tiveram maiores destaque no rendimento médio, nas quais são estatisticamente similares. Em Rancho, todas as linhagens apresentaram produção média iguais, exceto L-38 que difere de todas e teve rendimento médio baixo. E para Estância as linhagens L-242-9 e L-37 foram as que mostraram melhores rendimentos médios kg/ha.

Como sugestão futura de aprimoramento do trabalho, pretende-se construir um modelo linear misto tendo os locais Jícaro e Rancho agrupados como uma única localidade, para verificar a homoscedasticidade dos resíduos nesta circunstância.

REFERÊNCIAS

- ALBUQUERQUE, A. M. M. de O. *Análise estatística de um experimento aninhado com uma aplicação em dados de algodão*. Trabalho de conclusão de curso, 2016. Citado na página 32.
- BALZARINI, M.; MACCHIAVELLI, R.; CASANOVES, F. *Aplicaciones de Modelos Mixtos en Agricultura y Forestería*. [S.l.: s.n.], 2004. Citado 7 vezes nas páginas 12, 15, 16, 18, 19, 21 e 22.
- BANZATTO, D. A.; KRONKA, S. do N. *Experimentação agrícola*. [S.l.: s.n.], 2006. Citado na página 12.
- BARBOSA, M. *Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos*. Dissertação (mestrado), 2009. Citado na página 16.
- BUSSAB, W. de O.; MORETTIN, P. A. *Estatística Básica*. [S.l.: s.n.], 2017. Citado na página 27.
- COSTA, T. R. da. *Modelos Lineares Mistos: Uma aplicação na produção de leite de vacas da raça sindi*. Mestrado, 2010. Citado na página 18.
- DEMÉTRIO, C.; BRIEN, C. J. Formulating mixed models for experiments, including longitudinal experiments. 2009. Citado na página 18.
- FISHER, R. A. Statistical methods for research workers. In: _____. *Breakthroughs in Statistics: Methodology and Distribution*. New York, NY: Springer New York, 1992. p. 66–70. ISBN 978-1-4612-4380-9. Disponível em: <https://doi.org/10.1007/978-1-4612-4380-9_6>. Nenhuma citação no texto.
- GARCIA, D. et al. Modelos lineares generalizados mistos na avaliação genética da prenhez precoce na raça nelore. 2012. Citado na página 13.
- GARCIA, S. L. R.; LEITE, H. G. *Curso de estatística experimental*. [S.l.: s.n.], 2006. Citado na página 12.
- GURKA, M. J. Selecting the best linear mixed model under reml,"the american statistician. *American Statistician*, v. 60, p. 20–26, Fev 2006. Citado na página 23.
- HAENSELBLATT. *O que são as informações sobre o feijão tepary no cultivo do feijão tepary*. 2023. Disponível em: <<https://pt.haenselblatt.com/articles/edible-gardens/what-are-tepary-beans-information-on-tepary-bean-cultivation.html>>. Citado na página 13.
- HAJDUK, G. K.; GALLOIS, E. *INTRODUCTION TO LINEAR MIXED MODELS*. 2022. Disponível em: <<https://ourcodingclub.github.io/tutorials/mixed-models/?s=08>>. Citado na página 35.
- HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. 1967. Citado na página 14.
- HERNÁNDEZ, A. P. *Evaluación de catorce líneas de frijol tepari (Phaseolus acutifolius Gray) en tres localidades de EL Progreso*. Tese, outubro 1997. Citado 2 vezes nas páginas 13 e 27.
- HILDEN-MINTON, J. *Multilevel diagnostics for mixed and hierarchical linear models*. Thesis, 1995. Citado 2 vezes nas páginas 23 e 24.

NOBRE, J. S. *Métodos de diagnóstico para modelos lineares mistos*. Artigo, 2004. Citado 6 vezes nas páginas 15, 21, 22, 23, 24 e 38.

OLIVEIRA, A. A. de; MARTINI, A. M. Z.; CATANHO, C. de T. pós-graduação, *Princípios de Planejamento e Análise de Dados em Ecologia*. 2022. Citado na página 16.

OLIVEIRA, V. R. de et al. Variabilidade genética de procedências e progênies de umbuzeiro via metodologia de modelos lineares mistos (reml/blup). 2004. Disponível em: <<https://www.scielo.br/j/rbf/a/Xpmw9wXnHj6srVLwPRHTPZc/?lang=pt>>. Citado na página 13.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. 1971. Citado 2 vezes nas páginas 15 e 19.

PIETRAFESA, P. A.; SILVA, T. F. do Nascimento Melo da; PIETRAFESA, J. P. Modelos lineares mistos como ferramenta estatística para análise do índice de desempenho dos municípios do estado de Goiás. 2022. Citado na página 13.

PINHEIRO, J. C.; BATES, D. M. *Mixed-Effects Models in S and S-PLUS*. [S.l.: s.n.], 2000. Citado na página 23.

RESENDE, M. D. V. de et al. Estimaco de componentes de variaco e predico de valores genéticos pelo método da máxima verossimilhança restrita (reml) e melhor predico linear não viciada (blup) em pinus. 1996. Citado na página 15.

RIENZO, J. A. D.; GUZMAN, A. W.; CASANOVES, F. A multiple-comparisons method based on the distribution of the root node distance of a binary tree. v. 7, 1999. Disponível em: <<https://link.springer.com/article/10.1198/10857110260141193>>. Citado 8 vezes nas páginas 6, 9, 10, 24, 38, 39, 40 e 43.

RIENZO, J. A. D.; MACCHIAVELLI, R.; CASANOVES, F. *Modelos Lineales Mixtos Aplicaciones en InfoStat*. 2017. Disponível em: <<https://www.infostat.com.ar/index.php?mod=page&id=34>>. Citado 2 vezes nas páginas 16 e 25.

ROBERTSON, J. O. e E. *Ronald Aylmer Fisher*. 2003. Disponível em: <<https://mathshistory.st-andrews.ac.uk/Biographies/Fisher/>>. Citado na página 14.

ROBERTSON, J. O. e E. *Frank Yates*. 2005. Disponível em: <<https://mathshistory.st-andrews.ac.uk/Biographies/Yates/>>. Citado na página 14.

VLECK, L. D. V. *Charles Roy Henderson*. [S.l.: s.n.], 1998. Citado na página 14.

APÊNDICE A – PRIMEIRO APÊNDICE

Tabela 9 – Estimativas e comparações das médias em kg/ha das linhagens conforme os locais pelo teste de comparações múltiplas de agrupamento DGC (Rienzo, Guzman e Casanoves). Médias acompanhadas da mesma letra não diferem pelo teste de comparação múltipla ao nível de 5% de significância.

Local	Linhagem	Média
Jícaro	L-242-22	629,44 a
Jícaro	L-242-43	615,07 a
Jícaro	L-242-25	612,90 a
Jícaro	L-242-38	562,63 b
Rancho	L-242-25	531,42 b
Rancho	L-242-22	517,13 b
Rancho	L-242-9	486,52 b
Rancho	L-242-43	480,86 b
Rancho	L-246-19	472,33 b
Jícaro	L-30	465,17 b
Rancho	L-35	464,80 b
Jícaro	L-246-9	426,79 b
Jícaro	L-35	424,87 b
Rancho	L-30	411,86 b
Jícaro	L-242-7	410,32 b
Jícaro	L-18	408,95 b
Estância	L-242-9	402,25 b
Jícaro	L-246-19	400,05 b
Rancho	L-18	386,61 b
Rancho	L-39	382,65 b
Rancho	L-37	378,82 b
Jícaro	L-242-9	363,29 b
Rancho	L-242-7	343,02 b
Rancho	L-246-9	335,26 b
Jícaro	L-39	311,42 b
Jícaro	L-37	301,54 b
Rancho	L-242-38	296,75 b
Estância	L-37	255,34 b
Estância	L-246-19	230,55 c
Jícaro	L-38	223,40 c
Estância	L-242-43	213,51 c
Estância	L-242-22	202,40 c
Estância	L-242-25	197,73 c
Estância	L-30	197,65 c
Rancho	L-38	197,53 c
Estância	L-35	187,29 c
Estância	L-246-9	178,71 c
Estância	L-242-7	172,45 c
Estância	L-39	169,54 c
Estância	L-38	159,55 c
Estância	L-242-38	154,95 c
Estância	L-18	96,28 d

Fonte: Elaborada pela autora, 2023.