

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos lineares mistos e generalizados mistos em estudos
de adaptação local e plasticidade fenotípica de *Euterpe edulis*

Ezequiel Abraham López Bautista

Tese apresentada para obtenção do título de
Doutor em Ciências. Área de concentração: Es-
tatística e Experimentação Agronômica

Piracicaba
2014

Ezequiel Abraham López Bautista
Engenheiro Agrônomo

**Modelos lineares mistos e generalizados mistos aplicados em estudos
de adaptação local e plasticidade fenotípica de *Euterpe edulis***

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

Tese apresentada para obtenção do título de
Doutor em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2014**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

López Bautista, Ezequiel Abraham

Modelos lineares mistos e generalizados mistos em estudos de adaptação local e plasticidade fenotípica de *Euterpe edulis*. / Ezequiel Abraham López Bautista - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.
123 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Adaptação local 2. Plasticidade fenotípica 3. Análise de grupos de experimentos
4. Modelos lineares mistos 5. Modelos lineares generalizados mistos I. Título

CDD 634.6
L864m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte -O autor"

DEDICATÓRIA

Aos meus pais,
Justiniano Rafael López Velásquez e
Basilía Bautista

(in memoriam) pelo carinho e bons
exemplos que me propiciaram.

Com amor, DEDICO.

AGRADECIMENTOS

A Deus por sempre estar presente em cada momento de minha vida, me dando forças.

À minha família, em especial a minha avó Concepción Vidalina Echeverría De León (*in memorian*), ao meu irmão Carlos Enrique (*in memorian*), às minhas irmãs Cristina, Celeste e Clara Luz, aos meus sobrinhos e às minhas sobrinhas, obrigado pelo carinho e força.

Ao meu orientador, Prof. Dr. Silvio Sandoval Zocchi, pelos ensinamentos e pela disposição em me ajudar.

Aos professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP, que me propiciaram condições para a realização deste trabalho.

Aos professores Pedro Henrique Santin Brancalion e Giancarlo Conde Xavier Oliveira, que cederam os dados utilizados neste trabalho.

Ao Programa Estudante Convênio de Pós-Graduação (PEC-PG) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), o suporte financeiro concedido para a realização dos meus estudos de doutorado.

À Universidade de São Carlos de Guatemala (USAC) e Facultad de Agronomía (FAUSAC), pelo afastamento e apoio econômico concedido para a realização dos meus estudos de doutorado.

Ao meu professor de português, Ary Dos Santos, pelo ensinamento e pela amizade.

Aos colegas docentes da FAUSAC, em especial a Marino Barrientos García, Byron González Ramírez e Víctor Álvarez Cajas (*in memorian*) pela amizade.

Aos amigos Paulo César Zacarias, Isaac Ponciano, Cristiano Nunes, Rose Moreira, Denise Viola, Dirceu Arraes, Tiago Santana, Alexandre Lavorenti, Elton Pereira, Pedro Javier Mansilla, Marcello Neiva, Vinicius Franzini, Heisler Gómez e Antônio Moita.

Aos colegas e amigos do doutorado, Edilan de Sant'Ana Quaresma, Cássio Dessotti, Guilherme Biz, Maurício Lordêlo, Ricardo Álves, Rodrigo Pescim, Simone Werner, Marina Mestre, Lucas Santana, Pedro Ramos Cerqueira, Ana Julia Righetto, Simone Sartorio, Luiz Ricardo Nakamura, José Nilton Cruz, Djair Durand, Alessandra Santos, Joseane Silva, Thiago Gentil, Thiago Oliveira, Natalia Martins, Maria Cristina Martins, Ricardo Klein, pela amizade e pela troca de conhecimentos.

“Para realizar grandes conquistas, devemos não apenas agir, mas também sonhar, não apenas planejar, mas também acreditar.”

(Anatole France).

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Plasticidade fenotípica	23
2.2 Adaptação local	25
2.3 Análise de grupos de experimentos	27
2.4 Modelos lineares	29
2.5 Modelos lineares mistos	30
2.5.1 Estruturas de covariância	32
2.5.2 Estimação dos componentes de variância	35
2.5.2.1 Método ANOVA	35
2.5.2.2 Método da Máxima Verossimilhança (ML)	36
2.5.2.3 Método da Máxima Verossimilhança Restrita (REML)	37
2.5.3 Estimação dos termos fixos e predição dos termos aleatórios simultaneamente	40
2.5.4 Inferência para parâmetros de efeitos fixos	43
2.5.5 Inferência para parâmetros de efeitos aleatórios	45
2.5.6 Análise de resíduos e diagnóstico em modelos lineares mistos	47
2.6 Modelos lineares generalizados	48
2.6.1 Estimação do vetor de parâmetros β por máxima verossimilhança	51
2.6.2 Função <i>Deviance</i> e estatística de Pearson X^2 generalizada	54
2.6.3 Estimação do parâmetro ϕ	56
2.6.4 Testes de hipóteses relativas aos parâmetros $\beta's$	57
2.6.5 Análise de resíduos e diagnósticos para modelos lineares generalizados	58
2.6.6 Superdispersão	59
2.7 Modelos lineares generalizados mistos	61
2.7.1 Inferência em modelos lineares generalizados mistos	63
2.7.2 Estimação por máxima verossimilhança	64
2.7.3 Estimação e inferência baseada em métodos de linearização	67

3 MATERIAL E MÉTODOS	73
3.1 Material	73
3.1.1 Ensaio de estabelecimento (semeadura) em transplantes recíprocos para avaliação da adaptação local e plasticidade fenotípica	74
3.1.2 Ensaio de crescimento de mudas (juvenis) em transplantes recíprocos para avaliação da plasticidade e adaptação local.	77
3.2 Métodos	78
3.2.1 Modelos lineares mistos para avaliar o teor da massa de matéria seca das plantas	78
3.2.1.1 Estudo da plasticidade fenotípica	82
3.2.1.2 Estudo da adaptação local	83
3.2.2 Modelos lineares generalizados mistos para avaliar a proporção de sementes germinadas	85
4 RESULTADOS E DISCUSSÃO	87
4.1 Avaliação da plasticidade fenotípica e da adaptação local com base nos teores de massa de matéria seca das plantas em ensaios de estabelecimento e de crescimento	87
4.1.1 Plasticidade fenotípica	93
4.1.2 Adaptação local	94
4.2 Avaliação da plasticidade fenotípica e adaptação local com base na proporção de sementes germinadas em ensaios de estabelecimento	101
4.2.1 Plasticidade fenotípica	103
4.2.2 Adaptação local	104
5 CONSIDERAÇÕES FINAIS	109
REFERÊNCIAS	111
ANEXOS	119

RESUMO

Modelos lineares mistos e generalizados mistos em estudos adaptação local e plasticidade fenotípica de *Euterpe edulis*

Este trabalho objetivou a avaliação da presença de plasticidade fenotípica e de adaptação local de três procedências de palmitreiro: Ombrófila Densa, Estacional Semidecidual e Restinga, em três locais no Estado de São Paulo: Parque Estadual da Ilha do Cardoso, Parque Estadual de Carlos Botelho e Estação Ecológica dos Caetetus, em ensaios de adaptação no estabelecimento (ou de semeadura) e de adaptação em juvenis (ou de crescimento). Os conjuntos de dados foram analisados utilizando estruturas de grupos de experimentos, com efeitos cruzados e aninhados. As variáveis relacionadas com a massa de matéria seca das plantas, nos dois ensaios, foram analisadas usando a abordagem de modelos lineares de efeitos mistos, por meio da incorporação de fatores de efeito aleatório, e fazendo uso do método da máxima verossimilhança restrita (REML) para estimação dos componentes de variância associados a tais fatores com um menor viés. Por outro lado, para a proporção de sementes germinadas, no ensaio de adaptação no estabelecimento, a análise estatística foi realizada a partir da abordagem dos modelos lineares generalizados mistos, sob a pressuposição de que a variável segue uma distribuição binomial, com função de ligação logito. O método da pseudo-verossimilhança foi empregado para obtenção da solução das equações de verossimilhança. Os resultados mostraram que as plantas originadas de sementes dos três biomas avaliados apresentaram um comportamento plástico, para todos os caracteres avaliados no ensaio de adaptação no estabelecimento. Com relação ao ensaio de adaptação em juvenis, a característica de plasticidade foi verificada somente para a massa de matéria seca da folha em plantas provenientes do bioma Estacional Semidecidual. A característica de adaptação local, apresentou-se de forma evidente no ensaio de adaptação no estabelecimento. Estes resultados evidenciaram que em cada local avaliado, as plantas originadas das sementes de diferentes procedências apresentaram um comportamento diferenciado nos caracteres relacionados à massa de matéria seca, podendo em alguns casos, tratar-se de adaptação local. Concluiu-se que os locais Carlos Botelho e Ilha do Cardoso são os mais favoráveis para a germinação das sementes de sua mesma procedência.

Palavras-chave: Adaptação local; Plasticidade fenotípica; Análise de grupos de experimentos; Modelos lineares mistos; Modelos lineares generalizados mistos

ABSTRACT

Linear mixed models and generalized mixed models applied in studies of local adaptation and phenotypic plasticity of *Euterpe edulis*

The aim of this work was to evaluate the presence of phenotypic plasticity and local adaptation of three provenances of the palm specie *Euterpe edulis*: Atlantic Rainforest, Seasonally Dry Forest and Restinga Forest, in permanent parcels inserted in three forest types of the São Paulo State (Brazil): Parque Estadual da Ilha do Cardoso, Parque Estadual de Carlos Botelho e Estação Ecológica dos Caetetus, in experiments of seedling establishment and juveniles plants growth. The data sets were analyzed using structures of groups of experiments, with crossed and nested effects. The variables related to dry matter content of plants in both assays were analyzed using linear mixed models (LMM) approach, through the incorporation of random effect factors, and using the restricted maximum likelihood method (REML) for estimation of variance components associated with these factors with a minor bias. On the other hand, germination proportion of the seeds at seedling establishment assay was analyzed using the generalized linear mixed models (GLMM) approach, under the assumption that the variable follows a binomial distribution, with logit link function. The pseudo-likelihood (PL) method was used to obtain the numerical solution of the likelihood equations. The results showed that, plants from seeds of the three biomes evaluated presented a plastic behavior for all characters assessed in the seedling establishment assay. In respect to juveniles adaptation assay, the phenotypic plasticity characteristic was observed only to the leaf dry matter content of plants from Seasonally Dry Forest biome. The local adaptation characteristic was clearly observed in the seedling establishment assay. These results showed that at each site evaluated, plants originating from seeds of different provenances exhibited different behavior on characters related to the dry matter content and may in some cases be local adaptation. It was concluded that locations Carlos Botelho and Ilha do Cardoso are the most favorable for seed germination of its same provenance.

Keywords: Local adaptation; Phenotypic plasticity; Joint analysis from agronomical essays; Linear mixed models; Generalized linear mixed models

LISTA DE FIGURAS

- Figura 1 - Locais de estudo de *Euterpe edulis* Martius, A - Parque Estadual da Ilha do Cardoso, B - Estação Ecológica dos Caetetus e C - Parque Estadual de Carlos Botelho. Reproduzido de BRANCALION, 2009 75
- Figura 2 - Croqui do experimento de estabelecimento (semeadura), para um local determinado 76
- Figura 3 - Croqui de: (a) um bloco qualquer mostrando a casualização das procedências das sementes: R (Restinga), E (Estacional Semidecidual) e O (Ombrófila Densa), as parcelas sombreadas pertencem a outro projeto; e (b) uma parcela com três sementes de cada planta-matriz do mesmo ecossistema 76
- Figura 4 - Diagrama de Hasse para obtenção dos números de graus de liberdades e esperanças dos quadrados médios para os ensaios de estabelecimento (semeadura) e crescimento de mudas (juvenis) 79
- Figura 5 - Gráficos dos resíduos condicionais estudentizados em função dos valores preditos, para o experimento de adaptação no estabelecimento, para as variáveis: (a) massa seca da folha, (b) massa seca do caule e (c) massa seca da parte aérea 91
- Figura 6 - Gráficos dos resíduos condicionais estudentizados em função dos valores preditos, para o experimento de adaptação em juvenis, para as variáveis: (a) massa seca da folha, (b) massa seca do caule, (c) massa seca da raiz e (d) massa seca total 91
- Figura 7 - Gráficos de quantil-quantil para os resíduos condicionais estudentizados no experimento de adaptação no estabelecimento, para as variáveis: (a) massa seca da folha, (b) massa seca do caule e (c) massa seca da parte aérea 92
- Figura 8 - Gráficos de quantil-quantil para os resíduos condicionais estudentizados para o experimento de adaptação em juvenis, para as variáveis: (a) massa seca da folha, (b) massa seca do caule, (c) massa seca da raiz e (d) massa seca total 92

- Figura 9 - Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule e (c) da parte aérea, por local dentro de cada procedência, no ensaio de adaptação no estabelecimento. As barras representam as médias originais e as linhas verticais o erro padrão da média. 95
- Figura 10 - Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule, (c) da raiz e (d) total, por local dentro de cada procedência, no ensaio de adaptação em juvenis. As barras representam as médias originais e as linhas verticais o erro padrão da média. 96
- Figura 11 - Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule e (c) da parte aérea, por procedência dentro de cada local, no ensaio de adaptação no estabelecimento. As barras representam as médias originais e as linhas verticais o erro padrão da média. 99
- Figura 12 - Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule, (c) da raiz e (d) total, por procedência dentro de cada local, no ensaio de adaptação em juvenis. As barras representam as médias originais e as linhas verticais o erro padrão da média. 100
- Figura 13 - Gráficos dos resíduos condicionais estudentizados, (a) em função dos valores preditos, (b) histograma, (c) quantil-quantil e (d) *box plot*, para o experimento de adaptação no estabelecimento 103
- Figura 14 - Comportamento da proporção média estimada de sementes germinadas (em escala logito) segundo o local para cada procedência 105
- Figura 15 - Comportamento da proporção esperada de sementes germinadas (em escala logito) segundo a procedência em cada local de estudo 106

LISTA DE TABELAS

Tabela 1 - Decomposição dos números de graus de liberdade e esperanças dos quadrados médios para os ensaios de estabelecimento (semeadura) e crescimento de mudas (juvenis)	80
Tabela 2 - Médias e desvios padrões (d.p.) da massa de matéria seca (mg) da folha, do caule e parte aérea das plantas das procedências Ombrófila (Omb.), Semidecidual (Sem.) e Restinga (Res.), nos locais Carlos Botelho (CB), Caetetus (CT) e Ilha do Cardoso (IC), no ensaio de adaptação no estabelecimento	87
Tabela 3 - Médias e desvios padrões (d.p.) da massa de matéria seca (mg) da folha, do caule, da raiz e total das plantas das procedências Ombrófila (Omb.), Semidecidual (Sem.) e Restinga (Res.), nos locais Carlos Botelho (CB), Caetetus (CT) e Ilha do Cardoso (IC), no ensaio de adaptação em juvenis	88
Tabela 4 - Estatística F e significância para as fontes de variação fixas consideradas nos ensaios de adaptação no estabelecimento e em juvenis	89
Tabela 5 - Estimativas dos componentes de variância e valores da estatística Wald- Z para testar sua significância	89
Tabela 6 - Porcentagem da variância total atribuída a cada componente de variância, por experimento e variável analisada	90
Tabela 7 - Estatística F e significância para o desdobramento do número de graus de liberdade de local \sharp procedência para as variáveis relacionadas com a massa de matéria seca, analisando os dados de locais dentro de cada nível de procedência	93
Tabela 8 - Estimativas (Est.) dos contrastes entre locais (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) da mesma procedência contra outros locais, estatística t e significância	94
Tabela 9 - Estatística F e significância para o desdobramento do número de graus de liberdade de local \sharp procedência para as variáveis relacionadas com a massa de matéria seca, analisando procedências dentro de cada nível de local	97

Tabela 10 -Estimativas (Est.) dos contrastes entre procedências do mesmo local (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) contra outras procedências, estatística t e significância	98
Tabela 11 -Proporções médias observadas de sementes germinadas de <i>E. edulis</i> segundo a procedência e local de instalação do experimento.	101
Tabela 12 -Estimativas e erros padrões dos componentes de variância, estatística χ^2 e níveis descritivos para o teste da razão de verossimilhanças	102
Tabela 13 -Estatística F e respectivos níveis descritivos (valores de p) para as fontes de variação fixas	102
Tabela 14 -Estatística F e níveis descritivos (valores de p) para o desdobramento do número de graus de liberdade de local \nmid procedência para a variável proporção de sementes germinadas, analisando os dados de locais dentro de cada nível de procedência	103
Tabela 15 -Estimativas, em escala logito e entre parênteses na escala original (proporção de sementes germinadas), dos contrastes entre locais (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) da mesma procedência contra outros locais, estatística t e níveis descritivos (valor de p)	104
Tabela 16 -Estatística F e valores de probabilidade (valores- p) para o desdobramento do número de graus de liberdade de local \nmid procedência para a variável proporção de sementes germinadas, analisando os dados de procedências dentro de cada nível de local	105
Tabela 17 -Estimativas, em escala logito e entre parênteses na escala original (proporção de sementes germinadas), dos contrastes entre procedências do mesmo local (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) contra outras procedências, estatística t e níveis descritivos (valor de p)	105

1 INTRODUÇÃO

A palmeira *Euterpe edulis* pertence à família Arecaceae, está distribuída naturalmente na Mata Atlântica, tendo distribuição preferencial ao longo do litoral brasileiro, desde o sul da Bahia (latitude 15°S) até o norte do Rio Grande do Sul (latitude 30°S). Esta palmeira possui estipe simples que cresce até 20 metros de altura e forma o palmito na base das folhas. Popularmente esta espécie é conhecida como palmitreiro juçara, jiçara ou ripa (LORENZI; MELLO FILHO, 2001).

O palmitreiro caracteriza-se por produzir palmito de excelente qualidade, com valor econômico elevado e amplamente consumido na alimentação humana, porém é uma planta que não rebrota na base e o corte implica em sua morte (CARVALHO, 2003). Além disso, é de extrema importância ecológica na cadeia alimentar do ecossistema florestal, pois apresenta altos níveis de interação com os animais e desempenha significativo papel na nutrição da fauna da Mata Atlântica, uma vez que seu fruto serve de alimento para aves e mamíferos, como roedores, marsupiais, primatas e morcegos (REIS; KAGEYAMA, 2000).

No entanto, as populações de palmiteiros foram drasticamente reduzidas devido à ação extrativista e ao corte indiscriminado, sendo atualmente encontrado apenas em áreas protegidas da Mata Atlântica, principalmente em locais de difícil acesso (FANTINI, 1997). Essas ações extrativistas têm inviabilizado a regeneração natural do palmitreiro necessária para recompor a população original, estando ameaçado de extinção. Perante esta situação, estudos sobre sua restauração ecológica são indispensáveis (BRANCALION, 2009). Uma parte desses estudos consiste na avaliação da adaptação local e da plasticidade fenotípica de espécies de plantas com ampla distribuição geográfica e ecológica, em diferentes locais.

A plasticidade fenotípica é a capacidade que mostram alguns genótipos de alterar de forma significativa sua expressão (variações morfológicas e/ou fisiológicas) em resposta a determinados fatores ambientais, sem que mudanças genéticas sejam necessárias, produzindo uma série de fenótipos diferentes (BRADSHAW, 1965; SCHLICHTING, 1986).

O maior interesse deste fenômeno é a possibilidade da plasticidade incluir processos ativos de adaptação ao estresse ambiental, aspecto fundamental para garantir a estabilidade futura das massas florestais, perante as alterações climáticas globais.

Por outro lado, a adaptação local é o processo evolutivo pelo qual um organismo se torna mais capaz de viver em seu (ou seus) habitat(s). As espécies que apresentam grande amplitude de distribuição e longos períodos de vida estão, geralmente, submetidas a condições bióticas (competição, parasitismo, cooperação etc) e abióticas (luz, pluviosidade, temperatura, umidade atmosférica, ventos, textura do solo etc) bem variadas. Esta heterogeneidade ambiental, no espaço e no tempo, conduz a longo prazo, ao desenvolvimento de adaptações locais entre as populações (SAVOLAINEN; PYHAJARVI; KNURR, 2007).

Num trabalho conduzido por Brancalion (2009), o objetivo foi avaliar a presença de plasticidade fenotípica e adaptação local de palmiteiros originários de três procedências, ou formações florestais (Ombrófila Densa, Estacional Semidecidual e Restiga) em três locais (unidades de conservação) do Estado de São Paulo (Parque Estadual de Carlos Botelho, Estação Ecológica de Caetetus e Parque Estadual da Ilha do Cardoso). Para isso, conduziu ensaios de estabelecimento e de crescimento de mudas.

Neste trabalho propõe-se, para análise dos resultados, a utilização de modelos lineares mistos, para as variáveis relacionadas com os valores de massa de matéria seca folha, do caule, da raiz e total. Por outro lado, para a análise da proporção de sementes germinadas propõe-se a utilização de modelos lineares generalizados mistos (MLGM).

O modelo linear misto (MLM), tal como utilizado neste trabalho, tem como pressuposições a normalidade dos resíduos do vetor de parâmetros de efeito aleatório. Para a estimação dos parâmetros e predição dos efeitos aleatórios, o método utilizado foi o da máxima verossimilhança restrita (REML), que fornece estimativas com menor viés.

As pressuposições consideradas para os MLGM, por outro lado, foram que a variável segue uma distribuição binomial, com estrutura do preditor linear correspondente a um modelo de grupos de experimentos, com fatores cruzados e aninhados, com efeitos fixos: local, procedência e a interação local e procedência, e efeitos aleatórios: blocos dentro de locais, plantas matrizes dentro de procedências e a interação entre local e plantas matrizes dentro de procedências.

As estimativas dos parâmetros dos efeitos fixos e a predição dos efeitos aleatórios foram obtidas a partir da maximização do logaritmo da função de pseudo-verossimilhança (PL), utilizando-se o algoritmo de otimização quase-Newton.

Na seção a seguir é realizada a revisão da literatura a respeito dos assuntos descritos nesta introdução, na Seção 3 são descritos os conjuntos de dados utilizados no trabalho. Nesta seção é também apresentada a metodologia adotada para analisar os dados. Na Seção 4 são apresentados os resultados e discussão para os dois experimentos e na Seção 5 são apresentadas as considerações finais sobre o trabalho.

2 REVISÃO BIBLIOGRÁFICA

Nesta seção são apresentados alguns conceitos relacionados com a plasticidade fenotípica e a adaptação local (definição e importância em estudos de restauração ecológica, formas de expressão, causas e principais metodologias para o estudo destes processos), notações e principais conceitos relacionados à análise de grupos de experimentos, modelos lineares mistos, modelos lineares generalizados (MLG) e MLG mistos (inferência para parâmetros de efeito fixo e aleatório e diagnóstico de resíduos).

2.1 Plasticidade fenotípica

A plasticidade fenotípica, segundo Bradshaw (1965) e Schlichting (1986), é a capacidade que mostram alguns genótipos de alterar de forma significativa sua expressão (variações morfológicas e/ou fisiológicas) em resposta a determinados fatores ambientais, sem que mudanças genéticas sejam necessárias, produzindo uma série de fenótipos diferentes.

Em plantas, segundo Fuzeto e Lomônaco (2000) a plasticidade fenotípica pode ser expressa no crescimento em altura, na anatomia e morfologia das estruturas vegetativas (partes encarregadas de todas as funções vitais, exceto a reprodução) e reprodutivas, na alocação absoluta e relativa de biomassa, e na taxa fotossintética e fenologia.

De acordo com Thompson (1991), genótipos que expressam grande variabilidade fenotípica em diferentes ambientes são considerados como plásticos, e os que mostram pequena variabilidade são denominados robustos. Apesar da existência dessa classificação, Bradshaw (1965) e Scheiner (1993) ressaltam que uma determinada característica pode ser plástica em resposta a um fator ambiental, mas não a outro, e que analogamente, caracteres distintos podem representar diferentes graus de plasticidade, concluindo que a plasticidade não é propriedade geral do genótipo, mas sim específica de um caráter genético ou de um conjunto de caracteres genéticos.

A plasticidade fenotípica, segundo Scheiner (1993), retrata ainda, a habilidade de um organismo de alterar sua fisiologia e/ou morfologia em decorrência de sua interação com o meio ambiente. Espera-se, deste modo, que uma população, ocupando um ambiente heterogêneo, apresente grande potencial plástico em suas características fisiológicas e/ou morfológicas (FUZETO; LOMÔNACO, 2000).

No entanto, deve-se notar que uma parte da variação fenotípica observada pode ser causada por variações aleatórias durante o desenvolvimento dos organismos, que não têm conexão com alguma influência ambiental (NOVOPLANSKY, 2002; BRADSHAW, 1965).

Chambel et al. (2005) citam um fator importante que atua sobre a plasticidade fenotípica: a temperatura que, atualmente, tem aumentado devido a mudanças climáticas. Adicionalmente, López (2009) cita os seguintes fatores que afetam a plasticidade fenotípica dos organismos vegetais: a competição interespecífica (competição entre duas espécies diferentes que disputam um mesmo nicho ecológico no mesmo local), a altura das ondas do mar, a quantidade de luz recebida, a resistência à salinidade, a quantidade de nutrientes ou o ambiente, em geral, em que vive o indivíduo.

Todos esses fatores podem, em determinadas situações, induzir a alterações genotípicas nos organismos, que podem fazer com que uma população seja beneficiada e obtenha um aumento na tolerância ambiental.

Para estudar a plasticidade fenotípica é frequente a instalação de ensaios em diferentes ambientes (locais), denominados ensaios de proveniências. Nestes, os materiais genéticos provenientes de diferentes locais são plantados em todos os locais, que possuem diferentes condições ambientais. Alternativamente podem ser realizados ensaios em ambientes com condições controladas (casas de vegetação, estufas, laboratórios) com gradientes em algum fator determinado (luminosidade, temperatura, disponibilidade hídrica, níveis de fertilidade de solos etc).

Scheiner (1993) salienta que, para quantificar a plasticidade, é necessária a reprodução dos genótipos em ambientes contrastados mediante propagação clonal ou partindo de famílias de meios irmãos ou irmãos completos.

Por sua vez, Martíns (2007) descreve os principais métodos de avaliação da plasticidade fenotípica, e cita que a análise de variância (ANOVA) para dois fatores é o método frequentemente utilizado para comparar a plasticidade da resposta de diferentes genótipos em uma série de ambientes. Dessa forma, esta análise permite identificar a variação devida aos genótipos, aos ambientes e à interação de ambos os fatores. Nos delineamentos experimentais que consideram o genótipo, um fator ambiental (local) e a sua interação, a variância ambiental se deve a respostas plásticas dos genótipos em geral, e a existência de interação se deve a uma variação genética nessa resposta plástica.

Em estudos sobre avaliação fenotípica nos quais a variável resposta medida é do tipo proporção ou contagem, têm sido utilizados para sua análise estatística, os MLGs ou os MLGMs.

Como exemplo, López (2009) avaliou a diferenciação adaptativa entre 21 populações de *Pinus canariensis* Chr. Sm. ex DC., em dois locais (Ilhas Canárias, Espanha e na região setentrional de Israel), entre 1999 e 2006, utilizando o delineamento de blocos incompletos. Uma das variáveis analisadas neste estudo, foi a proporção de árvores sobreviventes aos 5 e 8 anos depois da plantação. Análises individuais e conjuntas (séries de experimentos) foram realizadas, utilizando modelos logísticos, pressupondo uma distribuição binomial para a variável resposta, função de ligação logito e considerando o efeito de blocos como aleatório.

Ramírez (2010), por sua vez, avaliou a plasticidade fenotípica e diferenciação genética inter e intra-populacional de 13 populações de *Quercus suber* L., por meio de caracteres fisiológicos e marcadores moleculares, em dois locais da Espanha (Parque Nacional de Monfragüe e Parque Nacional de la Sierra de Andajúr), utilizando o delineamento casualizado em blocos. Análises individuais para a variável proporção de plantas sobreviventes aos 4 anos após iniciado os ensaios, foram realizadas, utilizando modelos logísticos, pressupondo uma distribuição binomial para a variável resposta, função de ligação logito e considerando o efeito de blocos como fixo.

Por outro lado, para analisar variáveis de natureza contínua e quando são considerados alguns fatores de efeitos aleatórios (por exemplo, blocos) ou aninhados (por exemplo, plantas dentro de procedências ou blocos dentro de locais), nos ensaios sobre plasticidade fenotípica têm sido utilizados os modelos lineares mistos, como descrito nos trabalhos de Eckhart, Geber e McGuire (2004), Angert e Schmeske (2005) e Cavallero et al. (2011), dentre outros.

2.2 Adaptação local

Combes (1946) define a adaptação como o conjunto de modificações provocadas na constituição de um organismo pela ação contínua de um meio diferente daquele onde, inicialmente, este se desenvolveu ou seus ascendentes. Savolainen, Pyhajarvi e Knurr (2007), por sua vez, definem a adaptação como o processo evolutivo pelo qual um organismo se torna mais capaz de viver em seu (ou seus) habitat(s). As espécies que

apresentam grande amplitude de distribuição e longos períodos de vida estão, geralmente, submetidas a condições bióticas e abióticas bem variadas.

Segundo os autores, os fatores bióticos são aqueles relacionados às relações ecológicas interespecíficas (aquelas que ocorrem entre seres de espécies diferentes, por exemplo, predação, competição, parasitismo, cooperação, mutualismo e comensalismo) e as intraespecíficas, que são as relações que se estabelecem entre seres vivos da mesma espécie (cooperação, competição e canibalismo). Entretanto, os fatores abióticos são os fatores climáticos (luz, pluviosidade, temperatura, umidade atmosférica, ventos etc.), fatores edáficos (ligados ao solo, como a textura, estrutura, composição química, pH, umidade, permeabilidade etc.) e fatores físico-químicos da água (temperatura, pH, salinidade, turbidez etc.).

Esta heterogeneidade ambiental, no espaço e no tempo, origina pressões seletivas diferentes que conduzem, a longo prazo, ao desenvolvimento de adaptações locais e à divergência ecotípica entre as populações.

A pressão seletiva é o termo designado para relacionar o papel do meio ambiente na seleção dos genes de uma população. Assim, dependendo do ambiente onde os organismos estão interagindo, por meio da seleção natural, alguns genes terão maior chances de sobreviver e serem passados a geração seguinte, do que outros. A pressão seletiva representa, portanto, um conjunto particular de características do ambiente que filtra determinados genes “direcionando” a evolução de determinadas características para a adaptação a este ambiente.

A instalação de ensaios de procedências é muito utilizada, especialmente nas espécies florestais, para estudar processos de adaptação entre populações. Esses ensaios consistem na instalação de populações de uma espécie, caracterizadas por diferentes condições ecológicas no seu lugar de origem, sob um ambiente comum. As possíveis diferenças populacionais observadas nos caracteres analisados podem ser atribuídas às diferenças genéticas (COLAUTTI; ECKERT; BARRETT, 2010).

A seguir será apresentada uma descrição da análise de grupos de experimentos, que são de uso frequente em estudos sobre plasticidade fenotípica e adaptação local de espécies vegetais.

2.3 Análise de grupos de experimentos

Análise de grupos de experimentos é uma técnica de uso frequente na experimentação agrônômica. Em inúmeras situações, ocorre a instalação de um grupo de experimentos por toda a região para a qual se deseja fazer inferência. Segundo Barbin (2013), todos os experimentos devem, preferivelmente, ter o mesmo número de tratamentos e repetições, mesmo delineamento experimental e tratos culturais, e repetidos por vários anos, porém instalados em locais (ambientes) distintos, com o objetivo de obter uma abrangência nas conclusões. No caso de se dispor de ensaios com número diferente de blocos ou repetições, Vencovsky e Barriga (1992), recomendam, para contornar tal problema, utilizar os procedimentos sugeridos por Cochran e Cox (1957).

De acordo com Martínez (1988), o primeiro estudo teórico do problema que surge ao combinar os resultados de uma série de experimentos similares, é atribuído a Cochran (1937). Posteriormente, Yates e Cochran (1938) sugerem um exame cuidadoso dos resultados experimentais antes de iniciar qualquer análise combinada, e atestam que o procedimento de análise de variância usual, apropriado para analisar os resultados dos experimentos, pode requerer modificação, devido à heterocedasticidade presente entre os diferentes experimentos. Assim, sugerem a partição das somas de quadrados associadas aos tratamentos e à interação de tratamentos \times locais em componentes ortogonais apropriados.

Uma análise de variância para grupos de experimentos em estudos de melhoramento de plantas foi desenvolvido por Rojas (1951), focalizando sua atenção na estimação dos componentes de variância. Kempthorne (1952), por sua vez, considera que na análise conjunta, existem duas possíveis dificuldades para a interpretação dos resultados na análise de variância. A primeira dificuldade é que a variância residual (σ^2) não é constante nos experimentos, e a segunda, é que o componente de variância referente à interação de tratamentos \times locais depende da combinação de tratamentos e locais. Posteriormente Rojas (1958) estudou a situação geral da análise de um grupo de experimentos distribuídos no tempo e no espaço, e desenvolveu uma teoria para solucionar o problema da heterogeneidade de variâncias.

Outras contribuições importantes para o estudo da análise conjunta de experimentos são apresentadas por Cochran (1954), Cochran e Cox (1957), Pimentel-Gomes e Guimarães (1958), Pavate (1961) e Pimentel-Gomes (1970), dentre outros.

Com relação à metodologia para a análise estatística dos grupos de experimentos, Barbin (2013) recomenda que inicialmente se devem fazer todas as análises individuais, isto é, uma análise para cada local e em seguida, examinar as grandezas dos quadrados médios residuais. Para que as análises individuais possam ser reunidas em uma análise conjunta, é preciso que haja homogeneidade de variâncias entre os experimentos, isto é, que os quadrados médios residuais (QMRes) não difiram muito entre si. Nesse aspecto, Pimentel-Gomes (2009) considera que se o quociente entre o maior e o menor QMRes for menor que 7, geralmente a análise conjunta poderá ser realizada sem grandes problemas.

Para testar a homogeneidade, por outro lado, pode-se aplicar qualquer um dos testes para verificação da homogeneidade de variâncias, por exemplo, o teste de Bartlett, descrito em Anderson e Bancroft (1952) ou em Steel e Torrie (1960), ou o teste F máximo de Hartley (1950). Nos casos em que a homogeneidade de variâncias não for detectada deve-se proceder da seguinte maneira, de acordo com Cochran (1954):

- (i) Considerar separadamente subgrupos de experimentos com quadrados médios residuais homogêneos. Análises conjuntas serão feitas para cada subgrupo.
- (ii) Alternativamente, podem-se reunir todos esses locais numa só análise conjunta, e no momento de tomar a decisão de rejeitar ou não a hipótese H_0 , por meio do teste F , aplicar o método proposto por Cochran (1954), que consiste em ajustar o número de graus de liberdade do resíduo médio, representado por v , e o número de graus de liberdade da interação Locais \times Tratamentos, representado por v' , obtidos da seguinte maneira:

$$v = \frac{\left[\sum_{k=1}^K \text{QM Resíduo (Local } k) \right]^2}{\sum_{k=1}^K \frac{[\text{QM Resíduo (Local } k)]^2}{\text{gl Resíduo (Local } k)}},$$

que corresponde a aplicar a equação de Satterthwaite (1946), com $\text{gl Resíduo (Local } k) < v < \sum_{k=1}^K \text{gl Resíduo (Local } k)$, e

$$v' = \frac{(I-1)(K-1)^2 V_1^2}{(K-2)V_2 + V_1^2},$$

em que I é o número de tratamentos e K é o número de locais (ou experimentos), e

$$V_1 = \frac{\sum_{k=1}^K \text{QM Resíduo (Local } k)}{K}, \quad V_2 = \frac{\sum_{k=1}^K [\text{QM Resíduo (Local } k)]^2}{K},$$

sendo que QM Resíduo (Local k) é o quadrado médio residual referente ao k -ésimo local ou experimento.

Mais recentemente, Littell et al. (2006) citam que, variâncias heterogêneas podem ser incorporadas na análise estatística dos dados, no contexto dos modelos lineares de efeito misto, especificando diferentes estruturas de variâncias e covariâncias residuais para os níveis de um fator (ou uma combinação dos níveis dos fatores avaliados), por exemplo, especificando variâncias diferentes entre locais com grande variabilidade, dando menor peso às observações com variância grande.

2.4 Modelos lineares

Considere-se, para fins de ilustração, um estudo sobre avaliação da plasticidade fenotípica e da adaptação local de I tratamentos (procedências) de efeito t_i ($i=1, \dots, I$) de uma determinada espécie vegetal, realizado em K locais (l_1, l_2, \dots, l_K), utilizando o delineamento casualizado em blocos em cada local, com J blocos (completos ou incompletos) de efeitos b_j ($j=1, \dots, J$), em que foi medida uma variável Y , e considerando ainda, os fatores procedências, locais e blocos como fixos, um modelo linear clássico utilizado para a análise desse conjunto de dados, é definido na forma matricial por Searle (1971), como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1)$$

em que: \mathbf{y} representa o vetor de dimensão n , de dados observados; \mathbf{X} é a matriz do delineamento, de dimensões $n \times p$; $\boldsymbol{\beta}$ é o vetor de parâmetros desconhecidos de efeitos fixos, de dimensão p ; e $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios, de dimensão n .

O objetivo do modelo linear clássico é modelar a média de \mathbf{y} , usando-se o vetor de parâmetros de efeitos fixos $\boldsymbol{\beta}$. Os componentes do vetor $\boldsymbol{\epsilon}$ são variáveis aleatórias independentes e identicamente distribuídas com média 0 e variância σ^2 . Pressupondo que $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, tem-se que $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$, com função de verossimilhança dada por:

$$L = L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{\exp \left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \left(\frac{\mathbf{I}}{\sigma^2} \right) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]}{(2\pi\sigma^2)^{\frac{n}{2}}}. \quad (2)$$

Considerando-se então, um vetor de observações \mathbf{y} , os estimadores de máxima verossimilhança dos parâmetros são obtidos maximizando-se a função L , dada em (2), ou o seu logaritmo, em relação a $\boldsymbol{\beta}$. Isto leva ao sistema de equações normais: $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, cuja solução é $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, desde que $\mathbf{X}'\mathbf{X}$ seja não singular. Tem-se, ainda, que $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ e $\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$.

Caso $(\mathbf{X}'\mathbf{X})^{-1}$ não exista, utiliza-se uma inversa generalizada $(\mathbf{X}'\mathbf{X})^{-}$ e o estimador de $\boldsymbol{\beta}$ é dado por: $\hat{\boldsymbol{\beta}}^0 = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$, com $E(\hat{\boldsymbol{\beta}}^0) = \boldsymbol{\beta}$ e $\text{Cov}(\hat{\boldsymbol{\beta}}^0) = (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-}\sigma^2$.

2.5 Modelos lineares mistos

De acordo com Balzarini (2002), o modelo linear clássico, utiliza como procedimento de estimação o método dos mínimos quadrados ordinários (ou *Ordinary Least Squares* -OLS), que é muito restrito devido às pressuposições, como independência dos resíduos e homogeneidade de variâncias, frequentemente não atendidas.

Considere o exemplo citado na seção 2.4, relacionado com a avaliação da plasticidade fenotípica e da adaptação local de I procedências em K locais, no caso em que os blocos sejam definidos de efeito aleatório e aninhados em cada local ($b_{j(k)}$). Além disso, seja incluído o efeito aninhado de M plantas matrizes, selecionadas aleatoriamente dentro de cada procedência i ($\gamma_{m(i)}$), e a ocorrência de desbalanceamento não planejado, decorrente da perda de parcelas. Nesta situação o uso do modelo linear clássico não é adequado, sendo necessário analisar os dados por meio de um modelo linear misto (MLM). Este é definido por West, Welch e Galecki (2007), como um modelo linear paramétrico para dados agrupados, longitudinais ou dados provenientes de experimentos com medidas repetidas, que inclui parâmetros de efeito fixo, além da média geral, e efeitos aleatórios associados com um ou mais fatores aleatórios, além do erro experimental.

Assim, esses modelos envolvem duas partes: uma parte descrevendo os efeitos aleatórios e a outra descrevendo os efeitos fixos. Os efeitos de um fator são considerados como fixos (constantes), quando os níveis em estudo, geralmente poucos, forem escolhidos pelo pesquisador, de modo que a inferência é restrita a esses níveis.

Por outro lado, Pinheiro e Bates (2000) citam que, os efeitos de um fator são considerados como aleatórios, quando os níveis em estudo correspondem a uma amostra aleatória de uma população de referência. Neste caso os níveis provêm de uma distribuição de probabilidade e a inferência é extrapolada para a população de referência.

Nos modelos mistos, a análise da parte aleatória consiste na predição dos efeitos aleatórios, na presença de efeitos fixos, e na estimação dos componentes de variância (variâncias associadas aos efeitos aleatórios de um modelo matemático).

A análise da parte fixa, por sua vez, consiste na estimação e testes de hipóteses sobre funções estimáveis dos efeitos fixos. Em geral, tanto a predição dos efeitos aleatórios quanto a estimação dos efeitos fixos dependem da estimação dos componentes de variância. Em termos matriciais o MLM pode ser expresso da seguinte forma, como descrito por Searle, Casella e McCulloch (1992):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (3)$$

em que: \mathbf{y} é o vetor de observações, de dimensão n ; \mathbf{X} é a matriz de delineamento dos efeitos fixos, conhecida, de dimensões $n \times (p+1)$; $\boldsymbol{\beta}$ é o vetor de parâmetros de efeitos fixos desconhecidos, de dimensão $p+1$; $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_b]$, sendo \mathbf{Z}_i de dimensões $n \times q_i$ a matriz de delineamento para o i -ésimo efeito aleatório; $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_b^T]^T$, o vetor de parâmetros de efeitos aleatórios, desconhecido, de dimensão q , em que \mathbf{u}_i possui dimensão q_i , sendo $q = \sum_{i=1}^b q_i$; e $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios desconhecidos, de dimensão n . Admite-se, ainda, que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$ e que, além disso, são variáveis não correlacionadas.

Em termos de modelos condicionais (ou hierárquicos), o modelo linear misto geral pode ser reescrito de tal modo que:

$$E(\mathbf{Y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u},$$

$$\mathbf{Y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}),$$

$$Cov(\mathbf{Y}|\mathbf{u}) = \mathbf{R},$$

e

$$\text{Cov}(\mathbf{u}) = \mathbf{G}.$$

Assim, marginalmente, tem-se que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, com $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ e,

$$\text{Var}(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}, \quad (4)$$

sendo \mathbf{G} de dimensões $q \times q$, a matriz de variâncias e covariâncias dos efeitos aleatórios \mathbf{u} , e \mathbf{R} de dimensões $n \times n$, a matriz de variâncias e covariâncias residual, que representa a variação intra-grupos. Assim, pode-se modelar a variância dos dados, da eq. (4), especificando a estrutura (ou forma) de \mathbf{G} e \mathbf{R} . Quando $\mathbf{R} = \sigma^2 \mathbf{I}_{(n)}$ e $\mathbf{Z} = \mathbf{0}$, o modelo misto se reduz ao modelo linear padrão: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

2.5.1 Estruturas de covariância

A modelagem da parte aleatória se realiza por meio da inclusão de uma matriz de variâncias e covariâncias (LITTELL et al., 2006), sendo que a necessidade de se incluírem parâmetros de variâncias e covariâncias pode surgir por várias razões, dentre elas:

- (i) as unidades experimentais sobre as quais são medidos, podem ser colocadas em grupos e os dados de um grupo comum são correlacionados. Isso pode ocorrer com dados de famílias, ninhadas, colônias e pessoas que habitam a mesma casa e
- (ii) medidas repetidas são tomadas sobre a mesma unidade experimental e são correlacionadas. A natureza dessas medidas pode ser multivariada. Exemplos comuns são os dados observados ao longo do tempo, chamados dados longitudinais.

Várias estruturas podem ser adotadas para as matrizes de variâncias e covariâncias \mathbf{G} e \mathbf{R} , dependendo do conhecimento que se tenha do fenômeno que se gera os dados. Algumas estruturas de covariância possíveis são apresentadas a seguir, admitindo-se uma situação simples com quatro ocasiões:

- (i) Não estruturada: todas as variâncias e covariâncias podem ser desiguais. Especifica uma matriz completamente geral, parametrizada em termos de variâncias e covariâncias. As variâncias são restritas a valores não negativos e as covariâncias não têm restrições.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

- (ii) Componentes de variância: matriz, geralmente utilizada como estrutura para \mathbf{G} e \mathbf{R} nos modelos ANOVA (todos os fatores são tratados como qualitativos), contendo apenas um parâmetro, σ^2 . Essa matriz supõe independência e homogeneidade de variâncias entre os componentes.

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

- (iii) Simetria composta: esta matriz contém dois parâmetros e admite homogeneidade de variâncias e covariâncias constantes. Exemplos de utilização desta matriz podem ser vistos em experimentos casualizados em blocos, para os quais o efeito de blocos é aleatório. Na matriz \mathbf{V} observam-se tantas matrizes de simetria composta quantos forem os blocos. Sendo assim, a matriz \mathbf{V} resultante será dita bloco diagonal.

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

- (iv) Autorregressiva de 1ª ordem - AR(1): esta matriz contém dois parâmetros, admite homogeneidade de variâncias e covariâncias decrescentes em função das distâncias entre as observações, ou seja, a correlação decresce à medida que se aumenta a distância entre as observações repetidas, pois ρ o parâmetro autoregressivo, para um processo estacionário, assume-se entre $-1 < |\rho| < 1$. Essa estrutura pode ser utilizada desde que as observações sejam igualmente espaçadas (no tempo ou no espaço).

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

(v) espacial (ou potência): nesta matriz todas as variâncias são iguais e a correlação entre observações separadas por uma distância de d unidades é ρ^d , sendo ρ a correlação entre observações. Seu uso é aconselhado para os casos em que as observações repetidas não são equidistantes,

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}$$

Para a escolha adequada da estrutura é necessário utilizar algum critério de seleção. Gbur et al. (2012) citam que o Critério de Informação de Akaike (AIC), o AIC corrigido (AICC) e o Critério de Informação Bayesiano (BIC), podem ser utilizados para comparar as estruturas de covariância. Estes critérios são baseados no logaritmo da função de verossimilhança (ML ou REML) e dependem do número de observações e de parâmetros do modelo. Suas expressões são dadas por:

$$AIC = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\kappa}}) + 2d,$$

$$AICC = AIC + \frac{2d(d+1)}{(n-d-1)} = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\kappa}}) + 2d \left(1 + \frac{d+1}{n-d-1} \right) = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\kappa}}) + 2d \frac{n}{n-d-1},$$

e

$$BIC = -2\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\kappa}}) + d \log(n),$$

sendo ℓ o logaritmo da função de verossimilhança, d o número de parâmetros de efeitos fixos e aleatórios estimados no modelo, $\hat{\boldsymbol{\kappa}}$ o vetor de estimativas dos componentes de variância e n o número de observações utilizadas na estimação desses parâmetros. A estrutura de covariâncias com menor valor do critério escolhido é considerada mais adequada.

2.5.2 Estimação dos componentes de variância

Barbin (1993) define os componentes de variância como as variâncias associadas aos efeitos aleatórios de um modelo estatístico-matemático, sendo que o seu conhecimento é de grande importância em genética e melhoramento genético animal e vegetal, pois a população e o método de melhoramento a serem utilizados dependem de algumas informações que podem ser obtidas a partir desses componentes.

No caso de modelos mistos, a solução das Equações de Modelos Mistos (EMM) de Henderson (1953), depende do conhecimento da matriz de variâncias e covariâncias \mathbf{V} , cuja estrutura é conhecida, porém, seus componentes não o são. Desse modo, torna-se necessário substituí-los por suas estimativas.

Diversos métodos têm sido propostos para estimar os componentes de variância, destacando-se os seguintes: método dos momentos ou método da análise de variância (ANOVA); método da máxima verossimilhança - ML(HARTLEY; RAO, 1967); método da estimação quadrática não-viesada de variância mínima - MIVQUE (RAO, 1971) e o método da máxima verossimilhança restrita - REML (PATTERSON; THOMPSON, 1971; HARVILLE, 1977). Existem ainda os métodos I, II e III de Henderson (1953) e aqueles derivados da estimação de funções quadráticas (MARCELINO; IEMMA, 2000).

Segundo Resende (2007), para obter estimativas tanto de ML quanto de REML, vários algoritmos têm sido desenvolvidos, dentre eles, o algoritmo esperança-maximização (EM), Escore de Fisher, Newton-Raphson e de Informação média (AI).

2.5.2.1 Método ANOVA

Este método consiste em igualar as esperanças dos quadrados médios de cada fonte de variação presente na análise da variância, aos seus respectivos quadrados médios e resolver as equações resultantes para cada componente de variância. Este método em geral é adequado para modelos simples, que envolvem dados balanceados. Os estimadores ANOVA apresentam muitas propriedades, por exemplo, são não-viesados e têm variância mínima, são funções de estatísticas suficientes, para as quais podem ser obtidas estimativas dos erros padrões associados, e uma aproximação dos números de graus de liberdade, por métodos como os propostos por Satterthwaite (1946), Fai e Cornelius (1996) e Kenward e Roger (1997). Além disso, nenhuma suposição da distribuição dos dados, além das suposições básicas sobre as variâncias e covariâncias já mencionadas é exigida.

Entretanto, quando os dados são não balanceados, não existe um único modo de se obter a tabela da análise da variância, levando a diferentes estimativas para um mesmo componente. Como uma desvantagem pode-se citar o fato de que esse método não exclui a ocorrência de estimativas negativas para os componentes de variância, fato que torna a propriedade de estimador não viesado pouco interessante.

Uma sugestão para contornar tal problema é utilizar a restrição do espaço paramétrico, ou seja, igualar as estimativas negativas a zero. Porém essa solução sacrifica a propriedade do estimador ser não viesado pelo método ANOVA (SEARLE; CASSELLA; McCULLOCH, 1992).

2.5.2.2 Método da Máxima Verossimilhança (ML)

O método da máxima verossimilhança (ML) foi desenvolvido por Fisher (1922), mas Hartley e Rao (1967) apresentaram a especificação matricial de um modelo misto e a derivação de equações ML para várias classes de modelos. Por outra parte, os trabalhos de Henderson (1953) tiveram grande impacto no desenvolvimento dos métodos de estimação de componentes de variância a partir de dados desbalanceados.

Em situações de dados desbalanceados, Shaw (1987) cita que os estimadores ML apresentam as seguintes propriedades desejáveis: suficiência (tal que o preditor condense o máximo possível a informação contida na amostra e não seja função dependente do parâmetro), consistência (refere-se ao aumento da precisão da estimativa com o aumento do tamanho da amostra), eficiência (o preditor apresenta variância mínima) e translação invariante (não afetados por mudanças nos efeitos fixos). Outra vantagem do ML é a geração de estimativas não negativas dos componentes de variância.

Para a estimação ML de componentes de variância os efeitos fixos devem ser conhecidos, caso contrário, são substituídos por suas estimativas obtidas por ML. Porém, na estimação dos componentes de variância, o método ML não considera a perda de número de graus de liberdade devido à estimação desses efeitos fixos, causando então o vício, que conduz a subestimativas dos parâmetros de variância e, portanto podem conduzir a inferências incorretas (RESENDE, 2007).

2.5.2.3 Método da Máxima Verossimilhança Restrita (REML)

Segundo Resende (2007), um método alternativo, baseado na verossimilhança, para inferir sobre os componentes de variância nos modelos mistos é o método da máxima verossimilhança restrita (REML), desenvolvido por Patterson e Thompson (1971).

Os estimadores dos componentes de variância, pelo método REML, têm sido amplamente adotados, porque eliminam o primeiro dos problemas encontrados no método ML, ou seja, leva em consideração os graus de liberdade envolvidos na estimação dos parâmetros fixos do modelo. Sendo assim, estimativas REML dos componentes de variância tendem a ser menos viesadas que as estimativas de ML, e o método permite também a imposição de restrições de não negatividade. Dessa forma, o método REML é o procedimento ideal de estimação de componentes de variância com dados desbalanceados. Além disso, para todos os casos de dados balanceados, as soluções fornecidas pelo método REML a partir dos modelos mistos coincidem com as soluções fornecidas pelo método ANOVA (McCULLOCH; SEARLE; NEUHAUS, 2001).

De acordo com Resende (2007), o método REML é uma ferramenta flexível para a estimação de componentes de variância e efeitos fixos, e predição de efeitos aleatórios. Além disso, apresenta as seguintes vantagens:

- (i) Pode ser aplicado a dados desbalanceados.
- (ii) Permite lidar com estruturas complexas de dados (ensaios com observações realizadas em diferentes anos e locais, por exemplo).
- (iii) Permite ajustar modelos que não podem ser acomodados pela ANOVA.
- (iv) Permite o ajuste de vários modelos alternativos, podendo-se escolher o que se ajusta melhor aos dados e que ao mesmo tempo, seja parcimonioso (apresenta menor número de parâmetros).
- (v) Permite a correção simultânea para os efeitos ambientais, estimação de componentes de variância e predição de valores genéticos.
- (vi) Permite maior flexibilidade na modelagem, contemplando plenamente a análise de dados correlacionados devido ao parentesco, distribuição temporal e espacial.

O procedimento REML requer que o vetor de dados \mathbf{Y} tenha distribuição normal multivariada. Entretanto, vários autores relatam que os estimadores REML são também apropriados quando não se verifica normalidade dos dados (HARVILLE, 1977; MEYER, 1989).

Ao contrário do método ML, que maximiza a função de verossimilhança de todos os contrastes, o método REML maximiza a função de verossimilhança conjunta de todos os contrastes de erros ou resíduos, $\mathbf{Y}^* = \mathbf{L}^T \mathbf{Y}$, em que \mathbf{L} é uma matriz com $[n - \text{posto}(\mathbf{X})]$ colunas, de posto completo, com colunas ortogonais às colunas da matriz \mathbf{X} , isto é, $\mathbf{L}^T \mathbf{X} = \mathbf{0}$. Em outras palavras, o método REML maximiza a parte da função de verossimilhança que é invariante aos efeitos fixos.

Dessa forma, considere a matriz $\mathbf{L} = [\mathbf{L}_1 \mathbf{L}_2]$, não singular, com \mathbf{L}_1 e \mathbf{L}_2 de dimensões $(n \times (p + 1))$ e $(n \times (n - p - 1))$, respectivamente e satisfazem:

$$\mathbf{L}_1^T \mathbf{X} = \mathbf{I}_{p+1} \text{ e } \mathbf{L}_2^T \mathbf{X} = \mathbf{0}.$$

Alcarde (2012) cita que, os dados podem ser então transformados de \mathbf{y} para $\mathbf{L}^T \mathbf{y}$, ou seja,

$$\mathbf{L}^T \mathbf{y} = [\mathbf{L}_1 \mathbf{L}_2]^T \mathbf{y} = \begin{bmatrix} \mathbf{L}_1^T \mathbf{y} \\ \mathbf{L}_2^T \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p+1} \boldsymbol{\beta} + \mathbf{L}_1^T \mathbf{Z} \mathbf{u} + \mathbf{L}_1^T \boldsymbol{\epsilon} \\ \mathbf{L}_2^T \mathbf{Z} \mathbf{u} + \mathbf{L}_2^T \boldsymbol{\epsilon} \end{bmatrix}.$$

O vetor aleatório $\mathbf{y}^T = [\mathbf{y}_1^* \mathbf{y}_2^*]^T$ tem distribuição com esperança e variância, respectivamente, dadas por:

$$E(\mathbf{Y}^*) = E \left[\begin{pmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{pmatrix} \right] = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}$$

e

$$Var(\mathbf{Y}^*) = Var \left[\begin{pmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{pmatrix} \right] = \begin{pmatrix} \mathbf{L}_1^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V} \mathbf{L}_2 \end{pmatrix}.$$

Logo,

$$\begin{pmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{pmatrix} \sim N \left[\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{L}_1^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V} \mathbf{L}_2 \end{pmatrix} \right].$$

A distribuição completa de $\mathbf{L}^T \mathbf{y}$ pode ser subdividida em uma distribuição condicional, $\mathbf{Y}_1^* | \mathbf{Y}_2^*$ para a estimação de $\boldsymbol{\beta}$, e uma distribuição marginal, baseada em

\mathbf{Y}_2^* , para a estimação dos componentes de variância (Alcarde, 2012). Esta distribuição marginal é a base da máxima verossimilhança restrita.

Por outro lado, seja $\boldsymbol{\kappa}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\alpha}^T)$ o vetor de componentes de variância, tal que $\boldsymbol{\gamma}$ contém os componentes de variância associados a \mathbf{u} e $\boldsymbol{\alpha}$ os componentes de variância associados a $\boldsymbol{\epsilon}$, sua estimação é baseada no logaritmo da função de verossimilhança restrita:

$$\ell_R = -\frac{1}{2} \left[\log \det (\mathbf{L}_2^T \mathbf{V}^{-1} \mathbf{L}_2) + \mathbf{y}^{T*} (\mathbf{L}_2^T \mathbf{V} \mathbf{L}_2)^{-1} \mathbf{y} \right],$$

que pode ser expressa como:

$$\ell_R = -\frac{1}{2} \log \det (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}) + \log \det \mathbf{V} + \mathbf{y}^T \mathbf{P} \mathbf{y}, \quad (5)$$

em que,

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$$

e ainda,

$$\mathbf{y}^T \mathbf{P} \mathbf{y} = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$

Segundo Alcarde (2012), as estimativas REML de κ_l , tal que $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_L)$, são obtidas calculando a função escore, dada por:

$$\mathbf{U}(\kappa_l) = \frac{\partial \ell_R}{\partial \kappa_l} = \frac{1}{2} \left[\text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \right) - \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \mathbf{y} \right],$$

e igualando a zero.

Note que os elementos da matriz informação observada são:

$$\begin{aligned} -\frac{\partial^2 \ell_R}{\partial \kappa_l \partial \kappa_k} = & -\frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \right) - \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right) + \\ & + \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \mathbf{P} \mathbf{y} - \frac{1}{2} \mathbf{y}^T \mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \mathbf{P} \mathbf{y}, \end{aligned}$$

e, os elementos da matriz informação esperada são:

$$E \left(-\frac{\partial^2 \ell_R}{\partial \kappa_l \partial \kappa_k} \right) = \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right).$$

No entanto, a solução para $\mathbf{U}(\kappa_l) = 0$ exige um algoritmo iterativo que utiliza uma estimativa inicial $(\kappa^{(0)})$ para κ e uma atualização $\kappa^{(1)}$. Um desses algoritmos é o “Escore de Fisher”, em que:

$$\kappa^{(1)} = \kappa^{(0)} + \mathbf{I}(\kappa^{(0)}, \kappa^{(0)})^{-1} \mathbf{U}(\kappa^{(0)}),$$

em que $\mathbf{U}(\kappa^{(0)})$ é o vetor escore e $\mathbf{I}(\kappa^{(0)}, \kappa^{(0)})$ representa a matriz de informação esperada de κ , avaliada em $\kappa^{(0)}$ (ALCARDE, 2012).

Na medida em que as dimensões da matriz invertida aumentam, a utilização do algoritmo Escore de Fisher pode levar a dificuldades computacionais, as quais são discutidas em Gilmour, Thompson e Cullis (1995). Uma alternativa é a utilização do algoritmo AI que apresenta propriedades de convergência semelhantes ao algoritmo Escore de Fisher, evitando a sobrecarga computacional (GILMOUR; THOMPSON; CULLIS, 1995).

2.5.3 Estimação dos termos fixos e predição dos termos aleatórios simultaneamente

As estimativas de β e \mathbf{u} são obtidas por meio do uso de funções baseadas na função de verossimilhança dos dados. Assim, se a função de densidade de probabilidade de \mathbf{y} é dada por:

$$f(\mathbf{y}) = \frac{1}{2\pi^{\frac{n}{2}} |\mathbf{ZGZ}^T + \mathbf{R}|^{\frac{1}{2}}} \exp \left\{ \frac{1}{2} [(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{ZGZ}^T + \mathbf{R})^{-1} (\mathbf{y} - \mathbf{X}\beta)] \right\},$$

a função densidade de probabilidade conjunta de \mathbf{y} e \mathbf{u} , por sua vez, pode ser escrita como o produto entre a função densidade de probabilidade condicional de \mathbf{y} , dado \mathbf{u} , e a função densidade de probabilidade de \mathbf{u} , ou seja,

$$f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u})f(\mathbf{u}),$$

ou ainda,

$$f(\mathbf{y}, \mathbf{u}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \times \exp \left\{ \frac{1}{2} [(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})] \right\} \times$$

$$\times \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{G}|^{\frac{1}{2}}} \exp \left\{ \frac{1}{2} [(\mathbf{u} - \mathbf{0})^T \mathbf{G}^{-1} (\mathbf{u} - \mathbf{0})] \right\},$$

cujo logaritmo denotado por ℓ é dado por:

$$\begin{aligned} \ell = & -n \log(2\pi) - \frac{1}{2} (\log |\mathbf{R}| + \log |\mathbf{G}|) + \frac{1}{2} (\mathbf{y}^T \mathbf{R}^{-1} \mathbf{y} - 2\mathbf{y}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} - 2\mathbf{y}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \\ & + 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{u}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{u}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u}). \end{aligned}$$

Derivando-se ℓ em relação a $\boldsymbol{\beta}$ e \mathbf{u} e igualando-se as expressões resultantes a 0, obtêm-se as Equações de Modelos Mistos (EMM) propostas por Henderson (1984):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}$$

Se \mathbf{G} e \mathbf{R} são conhecidas, então, o estimador de mínimos quadrados generalizados (*generalized least squares*-GLS) ou melhor estimador linear não viesado (*best linear unbiased estimator*-BLUE) de $\boldsymbol{\beta}$, é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y},$$

em que \mathbf{V} é dada pela eq. (4).

De forma análoga, o melhor preditor linear não viesado (*best linear unbiased predictor* -BLUP) de \mathbf{u} é dado por:

$$\hat{\mathbf{u}} = \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$

Esses estimadores são denominados “melhores” por minimizarem a variância amostral, “lineares”, pois são funções lineares de \mathbf{y} e “não viesados”, porque $E((BLUE(\boldsymbol{\beta}))) = \boldsymbol{\beta}$ e $E(BLUP(\mathbf{u})) = \mathbf{u}$. Além disso, a matriz de variâncias e co-variâncias de $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\mathbf{u}} - \mathbf{u})$ é dada por:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-}, \quad (6)$$

em que o símbolo $^{-}$ no expoente da matriz \mathbf{C} denota a inversa generalizada.

Entretanto, na maioria das vezes, $\text{Var}(\mathbf{y}) = \mathbf{V}$ é desconhecida, e consequentemente, as matrizes \mathbf{G} e \mathbf{R} são desconhecidas. Neste caso, assumindo-se uma certa perda de eficiência, os parâmetros em \mathbf{G} e \mathbf{R} podem ser substituídos pelas estimativas correspondentes, constituindo, respectivamente, as matrizes denotadas por $\hat{\mathbf{G}}$ e $\hat{\mathbf{R}}$. Logo, substituindo-se \mathbf{G} por $\hat{\mathbf{G}}$ e \mathbf{R} por $\hat{\mathbf{R}}$ em (6), tem-se a matriz $\hat{\mathbf{C}}$ dada por:

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-1},$$

como uma aproximação da matriz de variâncias e covariâncias de $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\mathbf{u}} - \mathbf{u})$. Neste caso, os termos BLUE e BLUP não se aplicam, sendo apropriado substituí-los por EBLUE (melhor estimador linear não viesado empírico ou *empirical best linear unbiased estimator*) e EBLUP (melhor preditor linear não viesado empírico ou *empirical best linear unbiased predictor*), respectivamente, de acordo com Littell et al. (2006). O termo empírico é adicionado, portanto, para indicar esse tipo de aproximação.

Segundo McLean e Sander (1988), a matriz $\hat{\mathbf{C}}$ pode ser reescrita como:

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{21} \\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22} \end{bmatrix},$$

em que,

$$\hat{\mathbf{C}}_{11} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1},$$

$$\hat{\mathbf{C}}_{21} = -\hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{X} \hat{\mathbf{C}}_{11},$$

e

$$\hat{\mathbf{C}}_{22} = \left(\mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1} \right)^{-1} - \hat{\mathbf{C}}_{21} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Z} \hat{\mathbf{G}}.$$

Nota-se que $\hat{\mathbf{C}}_{11}$ é a equação para obter estimativas de mínimos quadrados generalizados dos elementos da matriz de variâncias e covariâncias de $\hat{\boldsymbol{\beta}}$.

Logo, utilizando a matriz $\hat{\mathbf{C}}$, as estimativas dos parâmetros de $\boldsymbol{\beta}$ e as predições dos parâmetros de \mathbf{u} são obtidas como segue:

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

e

$$\hat{\mathbf{u}} = \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

em que $\hat{\mathbf{V}}^{-1}$ é obtida substituindo-se os parâmetros em \mathbf{V} , pelas estimativas correspondentes.

Por outro lado, objetivando a construção de intervalos de $(1 - \alpha) \times 100\%$ de confiança (IC) para as estimativas dos parâmetros de β_i , calculam-se os erros padrões $\sqrt{\widehat{Var}(\hat{\beta}_i)}$, a fim de que os ICs nos forneçam o campo de variação de cada uma das estimativas destes parâmetros, como segue:

$$\text{IC } (\hat{\beta}_i) = \hat{\beta}_i \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_i)},$$

sendo $i = 1, \dots, p$, α o nível de significância, $z_{\alpha/2}$ o valor tal que $P(|Z| < z_{\alpha/2}) = 1 - \alpha$, sendo Z uma variável com distribuição normal padronizada, $\widehat{Var}(\hat{\beta}_i)$, a variância associada ao parâmetro de efeito fixo β_i .

E ainda, os intervalos de $(1 - \alpha) \times 100\%$ de confiança (IC) relacionados aos componentes de variância ($\hat{\sigma}_i^2$), podem ser estimados usando o método delta (GBUR et al., 2012) em que estas estimativas são linearizadas por meio da função logaritmica (log) e possuem distribuição normal assintótica, e o IC na escala original da variável resposta é dado por:

$$\text{IC } (\hat{\sigma}_i^2) = \exp \left(\log(\hat{\sigma}_i^2) \pm z_{1-\alpha/2} \hat{\sigma}_i^{-2} \sqrt{\widehat{Var}(\hat{\sigma}_i^2)} \right)$$

2.5.4 Inferência para parâmetros de efeitos fixos

Molenberghs e Verbeke (2005) e West, Welch e Gálecki (2007), dentre outros, apresentam os testes de Wald, t e Wald- F , que podem ser utilizados para testar a significância dos termos fixos, ou de uma combinação linear dos mesmos, em modelos mistos.

O teste de Wald aproximado pode ser realizado para cada β_i em $\boldsymbol{\beta}$, com $i=1, \dots, p$, assim como a obtenção do respectivo intervalo de confiança. Seja \mathbf{L} uma matriz de constantes conhecidas e de posto completo c ($c \leq p$), de dimensões $c \times p$, e considerando-se as hipóteses:

$$H_0 : \mathbf{L}\boldsymbol{\beta} = 0 \text{ contra } H_1 : \mathbf{L}\boldsymbol{\beta} \neq 0 \quad (7)$$

Tem-se que a estatística de Wald (W) é dada por:

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T \left[\mathbf{L} \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\hat{\boldsymbol{\kappa}}) \mathbf{X}_i \right)^{-1} \mathbf{L}^T \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

em que $\hat{\boldsymbol{\kappa}}$ é o vetor das estimativas dos parâmetros das matrizes de variância e covariância \mathbf{G} e \mathbf{R} , denominados componentes de variância. Sob H_0 a estatística W tem distribuição assintótica χ^2 com r graus de liberdade.

Por outro lado, devido ao fato de que no teste de Wald não se considerar a variabilidade introduzida pela estimação dos componentes de variância e poder subestimar a variação dos efeitos fixos, Molenberghs e Verbeke (2005) sugerem a utilização dos testes t e Wald- F para testar hipóteses sobre os parâmetros de efeito fixo.

O teste t é frequentemente utilizado para testar hipóteses do tipo:

$$H_0 : \beta_i = 0 \text{ contra } H_1 : \beta_i \neq 0,$$

em que a estatística t é calculada da seguinte maneira:

$$t = \frac{\hat{\beta}_i}{\sqrt{\text{Var}(\hat{\beta})}},$$

que, sob a hipótese nula, segue uma distribuição t de Student com ν graus de liberdade, que dependem exclusivamente dos dados e são calculados utilizando métodos como o de Satterhwaite (1946), Fai e Cornelius (1996) e Kenward e Roger (1997).

Por sua vez, a estatística F do teste Wald- F é definida por:

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T \left[\mathbf{L} \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\hat{\boldsymbol{\kappa}}) \mathbf{X}_i \right)^{-1} \mathbf{L}^T \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\text{posto}(\mathbf{L})},$$

que segue uma distribuição F aproximada, com número de graus de liberdade do numerador dado pelo posto da matriz \mathbf{L} e número de graus de liberdade do denominador utilizando métodos como o de Satterhwaite (1946), Fai e Cornelius (1996) e Kenward e Roger (1997).

Uma alternativa é a aplicação do teste de χ^2 associado ao teste da razão de verossimilhanças (LRT) cuja estatística,

$$2 \log \left(\frac{L_2}{L_1} \right) = 2[\log(L_2) - \log(L_1)],$$

em que L_1 representa a função de verossimilhança do modelo aninhado ou restrito (o modelo mais simples, referente à hipótese nula) e L_2 , a função de verossimilhança do modelo de referência (o modelo mais geral ou completo). Neste caso, os modelos aninhado e de referência devem conter os mesmos componentes de variância e mesmas estruturas para as matrizes \mathbf{G} e \mathbf{R} , porém diferentes conjuntos de parâmetros fixos. Essa prática permite verificar a importância dos termos fixos do modelo, uma vez que a diferença entre tais modelos encontra-se apenas com relação a esses termos.

Para este caso, a estatística para o teste da razão de verossimilhança segue, assintoticamente, a distribuição χ^2 com número de graus de liberdade igual à diferença entre os números de parâmetros de efeito fixo dos modelos em questão. Entretanto, para os casos em que os parâmetros encontram-se na fronteira do espaço paramétrico, a estatística do teste da razão de verossimilhança segue uma mistura de distribuições χ^2 (SELF; LIANG, 1987).

2.5.5 Inferência para parâmetros de efeitos aleatórios

De acordo com Resende (2007), o uso da análise de variância para a construção de testes F para os efeitos aleatórios em modelos desbalanceados é muito difícil. Isto porque é necessária a obtenção dos quadrados médios a partir dos componentes de variância e seus multiplicadores, que são muito difíceis de ser computados sob desbalanceamento. Há, no entanto, uma maneira mais formal para testar os efeitos aleatórios, ou seja, para verificar se determinado efeito aleatório necessita permanecer no modelo. Essa abordagem formal baseia-se em estatísticas fundamentadas na verossimilhança.

Segundo Pinheiro e Bates (2000), os modelos de referência e aninhado, devem ser estimados utilizando o mesmo procedimento. West, Welch e Gajewski (2007), por sua vez, sugerem o uso do método REML para a estimação dos componentes de variância, já que proporciona estimativas menos viesadas, comparadas com o método ML. Além disso, quando o teste é realizado para componentes aleatórios, a especificação da parte fixa deve ser a mesma para os dois modelos.

Uma das estatísticas utilizadas para testar as hipóteses $H_0 : \sigma_i^2 = 0$ e $H_1 : \sigma_i^2 > 0$ (existe variabilidade entre os níveis do fator aleatório i) é a Z de Wald, que é calculada dividindo-se a estimativa do parâmetro aleatório por seu erro padrão assintótico. Os erros padrões assintóticos são obtidos a partir da inversa da matriz de derivada segunda da verossimilhança, em relação a cada um dos parâmetros de efeito aleatório.

A estatística Z de Wald é válida para grandes amostras, mas ela pode ser incerta para pequenos conjuntos de dados e para parâmetros tais como componentes de variância, que apresentam uma distribuição assimétrica ou distribuição amostral limite.

Segundo Resende (2007), uma melhor alternativa é utilizar o teste da razão de verossimilhança (*Likelihood Ratio Test* -LRT), e recomenda calcular previamente a relação $\frac{\hat{\sigma}^2}{s(\hat{\sigma}^2)}$, em que $\hat{\sigma}^2$ é a estimativa de um componente de variância de um determinado efeito aleatório e $s(\hat{\sigma}^2)$ seu respectivo desvio padrão; e aplicar o LRT apenas quando $1 < \frac{\hat{\sigma}^2}{s(\hat{\sigma}^2)} < 2$.

Quanto à distribuição da estatística do teste da razão de verossimilhança restrita sob a hipótese de nulidade, West, Welch e Galecki (2007) discriminam dois casos que dependem se os valores dos componentes de variância envolvidos na hipótese estão, ou não, na fronteira do espaço paramétrico. Os dois casos são:

- (i) Os parâmetros de covariância referentes à hipótese de nulidade não estão na fronteira do espaço paramétrico, sendo que, o interesse está na verificação da homogeneidade de variâncias, ou ainda, se a covariância entre dois efeitos aleatórios é igual a zero. Nesses casos, a estatística segue assintoticamente a distribuição χ^2 com número de graus de liberdade igual à diferença entre o número de parâmetros nos modelos de referência e aninhado.
- (ii) Os parâmetros de covariância estão na fronteira do espaço paramétrico: são os casos em que se deseja verificar se um efeito aleatório deve, ou não, permanecer no modelo. Neste caso, Stram e Lee (1994) demonstraram que a estatística para o teste da razão de verossimilhanças para um único parâmetro de variância, que se encontra na fronteira do espaço paramétrico, segue uma mistura de distribuições χ^2 , $0,5\chi_0^2 + 0,5\chi_1^2$. Para os casos em que k parâmetros se encontram na fronteira do espaço paramétrico, a estatística segue também uma mistura de distribuições χ^2 , porém, nesse caso a mistura é dada por $0,5\chi_0^2 + 0,5\chi_k^2$.

Self e Liang (1987) apresentaram, adicionalmente, as distribuições para outros casos, como teste simultâneo para parâmetros de variância para os quais um se encontra na fronteira do espaço paramétrico e outro não.

2.5.6 Análise de resíduos e diagnóstico em modelos lineares mistos

Os diagnósticos devem ser parte do processo de construção de modelos e análise de conjunto de dados. Os resíduos são utilizados para examinar as suposições do modelo estatístico-matemático e detectar a presença de *outliers* e possíveis observações influentes.

Lembrando que as médias marginal e condicional nos modelos mistos são dadas por $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ e $E(\mathbf{Y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, respectivamente, dois tipos de resíduos são apresentados a seguir:

- (i) Resíduos marginais, que consistem da diferença entre o valor observado e a média marginal estimada. Neste caso, o vetor \mathbf{r}_m de resíduos marginais é definido como:

$$\mathbf{r}_m = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

- (ii) Resíduos condicionais, que consistem da diferença entre o valor observado e o valor predito da observação. Neste caso, o vetor \mathbf{r}_c de resíduos condicionais é definido como:

$$\mathbf{r}_c = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}} = \mathbf{r}_m - \mathbf{Z}\hat{\mathbf{u}}.$$

Por outro lado, segundo Gregoire, Schabenberger e Barrett (1995), dadas as matrizes $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T$ e $\mathbf{K} = \mathbf{I} - \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}^T \hat{\mathbf{V}}^{-1}$, tem-se que

$$\widehat{Var}(\mathbf{r}_m) = \hat{\mathbf{V}} - \mathbf{Q}$$

e

$$\widehat{Var}(\mathbf{r}_c) = \mathbf{K}(\hat{\mathbf{V}} - \mathbf{Q})\mathbf{K}^T.$$

Alcarde (2012), entretanto considera que, os resíduos \mathbf{r}_m e \mathbf{r}_c não são ade-

quados para diagnósticos, pois podem apresentar correlações, mesmo para dados não correlacionados, e possuem difícil interpretação quando são incorporadas variâncias distintas ao modelo. Um modo de minimizar tais problemas consiste em trabalhar com os resíduos padronizados ou os resíduos de Pearson, descritos a seguir:

(i) Resíduo marginal estudentizado

$$r_{mi}^{estudentizado} = \frac{r_{mi}}{\sqrt{\widehat{Var}(r_{mi})}};$$

(ii) Resíduo condicional estudentizado

$$r_{ci}^{estudentizado} = \frac{r_{ci}}{\sqrt{\widehat{Var}(r_{ci})}};$$

(iii) Resíduo marginal de Pearson

$$r_{mi}^{Pearson} = \frac{r_{mi}}{\sqrt{\widehat{Var}(y_i)}};$$

(iv) Resíduo condicional de Pearson

$$r_{ci}^{Pearson} = \frac{r_{ci}}{\sqrt{\widehat{Var}(y_i|\mathbf{u})}}.$$

A autora recomenda, ainda, considerar o melhor preditor linear não viesado (BLUP) de \mathbf{u} , para diagnosticar os efeitos aleatórios. West, Welch e Galecki (2007) sugerem adicionalmente, a utilização de gráficos de diagnósticos padrões, ou seja, histogramas, gráficos de quantil-quantil e gráficos de dispersão para a verificação da normalidade dos resíduos, e nesse caso, dos resíduos condicionais estudentizados.

2.6 Modelos lineares generalizados

Metodologias para modelagem de dados na forma de proporções e de contagens foram propostas antes da década de 1970 (FINNEY, 1947; GRIZZLE; STARMER; KOCH, 1969). No entanto, a unificação do procedimento de modelagem foi proposta por Nelder e Wedderburn (1972), que desenvolveram a teoria dos modelos lineares generalizados (MLG), que são uma extensão dos modelos lineares clássicos.

Neste caso, não são necessárias as pressuposições de normalidade e homocedasticidade para a variável resposta, cuja distribuição deve pertencer à família exponencial de distribuições, e trabalha-se com os dados em sua forma original, ou seja, sem a necessidade de fazer transformações (McCULLAGH; NELDER, 1989).

De acordo com Cordeiro e Demétrio (2007), os modelos lineares generalizados consistem de três componentes:

- (i) Componente aleatório do modelo: a variável resposta é representada por um conjunto de variáveis aleatórias independentes Y_1, \dots, Y_n com distribuição que pertence à família exponencial na forma canônica com médias μ_1, \dots, μ_n , ou seja, $E(Y_i) = \mu_i, i = 1, 2, \dots, n$; e parâmetro de dispersão constante $\phi > 0$. A família exponencial engloba as distribuições, normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens;
- (ii) Componente sistemático: é constituído pelas variáveis explicativas que entram na forma de uma soma linear de seus efeitos, ou seja,

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j = \mathbf{x}_i^T \boldsymbol{\beta}, \text{ ou } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

em que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ é a matriz do modelo com elementos $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros e $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ é o preditor linear;

- (iii) Função de ligação: relaciona o componente aleatório ao componente sistemático, isto é,

$$\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta},$$

sendo $g(\cdot)$ uma função monótona e diferenciável, que determina a escala em que a linearidade é suposta. Os parâmetros β_1, \dots, β_p não são sujeitos a restrições, e, portanto, $g(\mu_i)$ pode assumir qualquer valor em $(-\infty, \infty)$, e assim, a forma da função de ligação apropriada é determinada em alguma escala pelo domínio de variação de $E(Y_i) = \mu_i$. Para observações estritamente positivas, como no caso de contagens, Hinkley, Reide e Snell (1991) citam que, a média deve ficar restrita a $(0, \infty)$. Utiliza-se:

$$g(\mu_i) = \eta_i = \sum_{j=1}^p x_{ij}\beta_j,$$

para descrever a relação entre os componentes do modelo linear generalizado, sendo $g(\cdot)$ a função de ligação e η_i o preditor linear associado a Y_i . A escolha da função de ligação depende da distribuição definida em (i). Por exemplo, a função de ligação canônica logística pode ser utilizada para a distribuição binomial.

Seja a variável aleatória Y_i , cuja função densidade de probabilidade para o caso contínuo, ou função de probabilidade para o caso discreto, que pode ser escrita na forma:

$$f(y_i; \theta; \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\}, \quad (8)$$

em que $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas, θ_i é um parâmetro canônico ou natural. Em geral, usa-se $a_i(\phi) = \frac{\phi}{w_i}$ sendo ϕ denominado parâmetro de escala ou dispersão (e seu inverso ϕ^{-1} , uma medida de precisão), e w_i os pesos a priori. De modo mais geral, tem-se a classe de modelos de dispersão exponencial (JØRGENSEN, 1997) que inclui situações em que ϕ é desconhecido.

O parâmetro canônico, θ_i , é uma função da média e, portanto, pode ser relacionado ao preditor linear. Uma escolha natural da função de ligação é

$$\theta_i = \eta_i = g(\mu_i),$$

sendo conhecida como função de ligação canônica.

O valor esperado e a variância de Y_i com distribuição pertencente à família (8) são dados por:

$$E(Y_i) = b'(\theta_i) = \mu_i$$

e

$$Var(Y_i) = a(\phi) b''(\theta_i) = a(\phi) V(\mu_i),$$

sendo $b'(\theta_i)$ e $b''(\theta_i)$ as derivadas de 1ª e 2ª ordens de $b(\theta_i)$ em relação a θ_i , respectivamente. A função $b''(\theta_i)$ pode ser expressa por $V(\mu_i)$, pois depende apenas de μ_i , sendo chamada de função de variância.

Como exemplo de distribuição pertencente à família exponencial tem-se a distribuição binomial. Considere-se a variável aleatória Y_i que representa o número de

sucessos em amostras de tamanhos m_i com distribuição binomial $B(m_i, \pi_i)$, com probabilidade de sucesso π_i ($0 < \pi_i < 1$), então, sua função de probabilidade é expressa por:

$$P(Y_i = y_i) = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}, y_i = 0, 1, 2, \dots, m_i. \quad (9)$$

Desenvolvendo-se a eq. (9), obtém-se:

$$\begin{aligned} f(y_i; \pi_i) &= \exp \left[\log \binom{m_i}{y_i} + y_i \log \pi_i + (m_i - y_i) \log(1 - \pi_i) \right], \\ &= \exp \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \log(1 - \pi_i) + \log \binom{m_i}{y_i} \right], \end{aligned}$$

e comparando-se com (8), tem-se: $\phi=1$, $\theta_i = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \log \left(\frac{\mu_i}{m_i - \mu_i} \right) \Rightarrow \mu_i = \frac{m_i e^{\theta_i}}{(1 + e^{\theta_i})}$, $b(\theta_i) = -m_i \log(1 - \pi_i) = m_i \log(1 + e^{\theta_i})$, e $c(y_i; \phi) = \log \binom{m_i}{y_i}$.

Portanto, a distribuição binomial pertence à família exponencial de distribuições, com $E(Y) = b'(\theta) = \frac{m_i e^{\theta_i}}{1 + e^{\theta_i}} = \mu_i$, $Var(Y_i) = b''(\theta_i) = \frac{m_i e^{\theta_i}}{(1 + e^{\theta_i})^2} = \frac{\mu_i}{m_i} (1 - \mu_i) = V(\mu_i)$.

2.6.1 Estimação do vetor de parâmetros β por máxima verossimilhança

Para a estimação dos parâmetros lineares β_1, \dots, β_p do modelo linear generalizado, o método comumente utilizado é o da máxima verossimilhança. Considerando uma variável aleatória Y_i com observações $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$ de uma distribuição pertencente à família exponencial, a função de verossimilhança é dada por

$$L(\boldsymbol{\theta}, \phi, \mathbf{y}) = \prod_{i=1}^n f(y_i; \theta_i, \phi) = \exp \left[\sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\} \right],$$

cujo logaritmo é definido por (CORDEIRO; DEMÉTRIO, 2007):

$$\ell = \ell(\boldsymbol{\theta}, \phi, \mathbf{y}) = \sum_{i=1}^n \ell(\theta; \phi; y) = \sum_{i=1}^n \left\{ \frac{1}{a(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\}, \quad (10)$$

sendo que $E(Y_i) = (\mu_i)$, $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ e $\theta_i = q(\mu_i)$.

De acordo com Demétrio (2002), uma propriedade da família exponencial de distribuições é que seus elementos satisfazem as condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança (10) é dado unicamente pela solução do sistema de equações $\mathbf{U}_{\boldsymbol{\theta}} = \frac{\partial \ell}{\partial \boldsymbol{\theta}} = \mathbf{0}$, ou equivalentemente, $\mathbf{U}_{\boldsymbol{\beta}} = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{0}$. Tem-se então, que o vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança, em que:

$$\begin{aligned} U_j &= \sum_{i=1}^n \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{1}{a_i(\phi)} [y_i - b'(\theta_i)] \frac{1}{\frac{\partial \mu_i}{\partial \eta_i}} \frac{\partial \mu_i}{\partial \theta_i} x_{ij}, \end{aligned}$$

pois $E(Y_i) = \mu_i = b'(\theta_i)$ e $\frac{\partial \mu_i}{\partial \theta_i} = V(\mu_i)$.

Logo,

$$U_j = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}. \quad (11)$$

Em geral, as equações $U_j = 0, j = 1, 2, \dots, p$ não são lineares, de forma que para a sua solução são necessários procedimentos iterativos, sendo Newton-Raphson e escore de Fisher os mais utilizados.

Considerando-se que se deseja obter a solução do sistema de equações $\mathbf{U}_{\boldsymbol{\beta}} = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{0}$ e usando-se a versão multivariada do método de Newton-Raphson, tem-se:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left[\mathbf{I}_o^{(m)} \right]^{-1} \mathbf{U}^{(m)},$$

sendo $\boldsymbol{\beta}^{(m)}$ e $\boldsymbol{\beta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $(m+1)$, $\mathbf{U}^{(m)}$ o vetor escore, com elementos $\frac{\partial \ell}{\partial \beta_j}$ avaliado no passo m e $\mathbf{I}_o^{(m)}$ a matriz de informação observada com elementos $-\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_{j'}}$, $j, j'=1, \dots, p$, avaliados em $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}$. Trocando-se a matriz de informação observada \mathbf{I}_o pela matriz de informação esperada de Fisher \mathfrak{F} , tem-se a solução pelo método escore de Fisher, isto é,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left[\mathfrak{F}^{(m)} \right]^{-1} \mathbf{U}^{(m)}, \quad (12)$$

sendo que \mathfrak{F} tem elementos dados por $\mathfrak{F}_{jj'} = E \left(-\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_{j'}} \right) = E \left(\frac{\partial \ell}{\partial \beta_j} \frac{\partial \ell}{\partial \beta_{j'}} \right)$, que é a

matriz de covariâncias dos U_j 's. Multiplicando-se ambos os lados da eq.(12) por \mathfrak{S} tem-se:

$$\mathfrak{S}^{(m)} \boldsymbol{\beta}^{(m+1)} = \mathfrak{S}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{U}^{(m)}. \quad (13)$$

Assim, usando-se a eq.(11), obtém-se:

$$\begin{aligned} \mathfrak{S}_{jj'} &= E(U_j U_{j'}) = \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} (y_i - \mu_i)^2 \frac{1}{[V(\mu_i)]^2} \left[\frac{\partial \mu_i}{\partial \eta_i} \right]^2 x_{ij} x_{ij'} \\ &= \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} a_i(\phi) V(\mu_i) \frac{1}{[V(\mu_i)]^2} \left[\frac{\partial \mu_i}{\partial \eta_i} \right]^2 x_{ij} x_{ij'} \\ &= \sum_{i=1}^n \frac{1}{[a_i(\phi)]^2} \frac{1}{V(\mu_i)} \left[\frac{\partial \mu_i}{\partial \eta_i} \right]^2 x_{ij} x_{ij'}, \end{aligned}$$

e fazendo-se $a_i(\phi) = \frac{\phi}{w_i}$, como $\phi > 0$, constante, w_i peso a priori e $W_i = \frac{w_i}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$, tem-se:

$$\mathfrak{S} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \mathbf{X},$$

com elementos $\mathfrak{S}_{jj'} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i x_{ij'}$, \mathbf{X} , a matriz do modelo e $\mathbf{W} = \text{diag}(W_1, W_2, \dots, W_n)$.

No caso das funções de ligação canônicas $W_i = w_i V(\mu_i)$, pois $\frac{\partial \theta_i}{\partial \mu_i} = \frac{\partial \eta_i}{\partial \mu_i} = V^{-1}(\mu_i)$.

Além disso, rearranjando-se os termos de U_j tem-se:

$$U_j = \sum_{i=1}^n \frac{w_i}{\phi} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \sum_{i=1}^n \frac{1}{\phi} x_{ij} W_i \frac{\partial \eta_i}{\partial \mu_i} (y_i - \mu_i),$$

e portanto, o vetor escore \mathbf{U} fica:

$$\mathbf{U} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W} \Delta (\mathbf{y} - \boldsymbol{\mu}),$$

com $\Delta = \text{diag} \left\{ \frac{\partial \eta_1}{\partial \mu_1}, \frac{\partial \eta_2}{\partial \mu_2}, \dots, \frac{\partial \eta_n}{\partial \mu_n} \right\} = \text{diag} \{g'(\mu_1), g'(\mu_2), \dots, g'(\mu_n)\}$.

Logo, substituindo \mathfrak{S} e \mathbf{U} na eq. (13) tem-se:

$$\frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} + \frac{1}{\phi} \mathbf{X}^T \mathbf{W}^{(m)} \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)},$$

ou, ainda,

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \left[\mathbf{X} \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} \right],$$

e fazendo-se $\mathbf{z}^{(m)} = \mathbf{X}\boldsymbol{\beta}^{(m)} + \boldsymbol{\Delta}^{(m)}(\mathbf{y} - \boldsymbol{\mu})^{(m)} = \boldsymbol{\eta}^{(m)} + \boldsymbol{\Delta}^{(m)}(\mathbf{y} - \boldsymbol{\mu})^{(m)}$, denominada como variável dependente ajustada, tem-se:

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (14)$$

ou ainda,

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (15)$$

que tem a forma da solução das equações normais, para o modelo linear obtida pelo método dos quadrados mínimos ponderados, exceto que nesse caso a solução $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$ é obtida por processo numérico iterativo. Observando-se que a expressão (15) independe de ϕ .

2.6.2 Função *Deviance* e estatística de Pearson X^2 generalizada

O ajuste de um modelo, com valores estimados $\hat{\boldsymbol{\mu}}$, a um conjunto de dados observados, \mathbf{y} , pode ser verificado por meio da estatística *scaled deviance* (traduzida como desvio por CORDEIRO, 1986) proposta por Nelder e Wedderburn (1972), com expressão dada por:

$$S_p(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y}) = -2\{\ell(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y}) - \ell(\bar{\mathbf{y}}, \phi, \mathbf{y})\},$$

sendo p o número de parâmetros do modelo, $\ell(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y})$ o logaritmo da função de verossimilhança para o modelo sob pesquisa e $\ell(\bar{\mathbf{y}}, \phi, \mathbf{y})$ o logaritmo da função de verossimilhança para o modelo saturado. Para o caso particular em que $\phi = 1$, esta estatística reduz-se à deviance, isto é,

$$D(\hat{\boldsymbol{\mu}}; \mathbf{y}) = -2\{\ell(\hat{\boldsymbol{\mu}}; \mathbf{y}) - \ell(\bar{\mathbf{y}}; \mathbf{y})\}.$$

Outra medida da discrepância de ajuste de um modelo a um conjunto de dados é a estatística de Pearson X^2 generalizada, que em muitos casos, é preferida em relação à *deviance*, por facilidade de interpretação. A estatística de Pearson é dada por:

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{V}(\hat{\mu}_i)},$$

sendo $\hat{V}(\hat{\mu}_i)$ a função de variância estimada sob o modelo que está sendo ajustado aos dados. A estatística de Pearson generalizada tem uma forma equivalente expressa em termos da variável independente ajustada do algoritmo (15):

$$X^2 = (\mathbf{z} - \hat{\boldsymbol{\eta}})^T \hat{\mathbf{W}} (\mathbf{z} - \hat{\boldsymbol{\eta}}).$$

Para respostas com distribuição normal, X^2 é igual à soma de quadrados do resíduo e

$$\frac{X^2}{\sigma^2} \sim \chi_{n-p}^2 \text{ exata.}$$

Para dados provenientes das distribuições binomial e de Poisson, em que $\phi = 1$, X^2 é a estatística original de Pearson, comumente usada na análise dos modelos logístico e log-linear para tabelas multidimensionais e que pode ser escrita na forma

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i},$$

sendo O_i a frequência observada e E_i a frequência esperada.

Pode-se mostrar que $S_p(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y})$ e $\frac{X^2}{\phi}$ têm distribuição χ_{n-p}^2 exata, quando a distribuição é normal e distribuição χ_{n-p}^2 assintótica, para as distribuições binomial e de Poisson, sob determinadas condições de regularidade.

Como objetivo geral da análise, deseja-se selecionar um modelo que seja parcimonioso e explique bem os dados. Portanto, para um modelo bem ajustado, espera-se que a *deviance* residual esteja próxima do número de graus de liberdade do resíduo do modelo. Uma maneira de obter uma deviance próxima do número de graus de liberdade é aumentar o número de parâmetros no modelo, mas com isso, aumenta-se também a complexidade na interpretação. O ideal é encontrar modelos mais simples com deviance moderada.

Outros critérios comumente utilizados para a seleção de modelos são os critérios de informação de Akaike (AIC) proposto por Akaike (1974), o critério de informação de Bayes (BIC) proposto por Schwarz (1978) e o critério AIC corrigido para pequenas amostras (AICC) proposto por Hurvich e Tsai (1989), e dados por:

$$AIC = -2\ell(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y}) + 2p,$$

$$BIC = -2\ell(\hat{\boldsymbol{\mu}}, \phi, \mathbf{y}) + p \log(n),$$

e

$$AICC = AIC + \frac{2p(p+1)}{(n-p-1)},$$

em que p é o número de parâmetros estimados no modelo quando as observações são independentes. O melhor modelo é aquele que apresenta menor valor para o critério escolhido. Convém ressaltar, ainda, que os modelos a serem comparados devem usar as mesmas observações, as mesmas variáveis explicativas e a mesma variável de resposta.

2.6.3 Estimação do parâmetro ϕ

Para as distribuições binomial e Poisson tem-se que $\phi = 1$. Quando desconhecido (distribuições normal e normal inversa $\phi = \sigma^2$ e gama $\phi = \nu^{-1}$, admite-se que seja o mesmo para todas as observações, isto é, constante. Segundo Demétrio (2002), a estimação de ϕ é necessária para a obtenção dos erros padrões dos $\hat{\beta}'s$, intervalos de confiança e testes de hipóteses para os $\beta's$ etc. Nestes casos, os métodos mais utilizados para a estimação de ϕ são: método da máxima verossimilhança, método dos momentos e perfil de verossimilhança.

O uso do método da máxima verossimilhança é sempre possível em teoria, mas pode se tornar intratável computacionalmente quando não existe solução explícita. Se ϕ é o mesmo para todas as observações, a estimativa de máxima verossimilhança de $\boldsymbol{\beta}$ independe de ϕ , mas já a matriz de variâncias e covariâncias dos $\beta's$ envolve esse parâmetro (DEMÉTRIO, 2002). Interpretando o logaritmo da função de verossimilhança $\ell(\boldsymbol{\beta}, \phi; \mathbf{y})$ como função de $\boldsymbol{\beta}$ e de ϕ , dado \mathbf{y} , a estimativa de máxima verossimilhança para ϕ é obtida pela solução da equação:

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = 0,$$

sendo que para as distribuições normal e normal inversa $\hat{\phi} = \frac{1}{n} D_p$.

Por outro lado, o método dos momentos fornece uma outra estimativa, também não consistente, para ϕ . Esse método baseia-se no fato, nem sempre verdadeiro, de que $S_p \sim \chi^2_{n-p}$. Neste caso,

$$E(S_p) = \frac{1}{\phi} E(D_p) \cong n - p,$$

e, portanto,

$$\tilde{\phi} = \frac{D_p}{n - p},$$

sendo D_p a *deviance* sob o modelo corrente (sob pesquisa).

Uma estimativa considerada melhor que a anterior é dada por:

$$\tilde{\phi} = \frac{D_m}{n - m},$$

sendo D_m a *deviance* sob o modelo maximal. Espera-se que para um modelo maximal bem ajustado aos dados S_m tenha um valor mais próximo da esperança da qui-quadrado de referência, isto é, $E(S_m) = \frac{1}{\tilde{\phi}} E(D_m) \cong n - m$. Para o modelo normal tem-se que:

$$\tilde{\phi} = \frac{D_m}{n - m} = \frac{SQRes}{n - m},$$

é a estimativa usual de σ^2 e é não viesada, mas para os outros modelos isso não acontece, em geral.

Uma outra maneira de se estimar ϕ é baseada na estatística de Pearson X^2 de Pearson e é dada por:

$$\phi^* = \frac{X^2}{n - m},$$

que nem sempre é imparcial, porém, é consistente (DEMÉTRIO, 2002).

2.6.4 Testes de hipóteses relativas aos parâmetros $\beta's$

Os métodos de inferência nos modelos lineares generalizados, baseiam-se, fundamentalmente, na teoria da máxima verossimilhança. De acordo com esta teoria, Demétrio (2002) apresenta três estatísticas para testar hipóteses relativas aos parâmetros $\beta's$, que são deduzidas de distribuições assintóticas de funções adequadas das estimativas dos $\beta's$, sendo elas: i) razão de verossimilhanças, ii) Wald e iii) escore; assintoticamente equivalentes que sob H_0 e para ϕ conhecido, convergem para uma variável com distribuição χ^2_p . Ressalta, entretanto que a razão de verossimilhanças leva a um teste uniformemente mais poderoso.

No caso em que há interesse no teste de hipótese do vetor β com um todo, isto é, no teste das hipóteses:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ contra } H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

tem-se as estatísticas:

(i) razão de verossimilhanças: $\Lambda = -2 \left[\ell(\hat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}_0) \right],$

(ii) estatística de Wald: $W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \hat{\mathfrak{S}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$ e

(iii) estatística escore: $S_R = \mathbf{U}(\boldsymbol{\beta}_0)^T \hat{\mathfrak{S}}_0^{-1} \mathbf{U}(\boldsymbol{\beta}_0),$

em que $\ell(\hat{\boldsymbol{\beta}})$ e $\ell(\boldsymbol{\beta}_0)$ são os valores do logaritmo da função de verossimilhança (10) em $\hat{\boldsymbol{\beta}}$ e $\boldsymbol{\beta}_0$, respectivamente, $\mathbf{U}(\boldsymbol{\beta}_0)$ e \mathfrak{S}_0 são o vetor escore e a matriz de informação avaliadas em $\boldsymbol{\beta}_0$, e $\hat{\mathfrak{S}}$ a matriz de informação de Fisher avaliada na estimativa de máxima verossimilhança (EMV) de $\boldsymbol{\beta}$. De acordo com Demétrio (2002), na estatística de Wald, $\hat{\mathfrak{S}}$ pode ser substituída por \mathfrak{S}_0 para definir uma estatística de Wald modificada que é assintoticamente equivalente.

2.6.5 Análise de resíduos e diagnósticos para modelos lineares generalizados

As técnicas para análise de resíduos e diagnósticos são utilizadas para verificar se um determinado modelo é apropriado para os dados. Essas técnicas podem ser utilizadas para detectar uma falha sistemática do modelo ou falhas particulares como, por exemplo, presença de *outliers*. Uma análise fácil para verificar a qualidade do ajuste é observar se o valor da *deviance* residual está próximo do número de graus de liberdade do resíduo e quando isso ocorre, significa que o modelo está bem ajustado aos dados.

De acordo com Demétrio (2002), as técnicas usadas para análise de resíduos e diagnósticos para modelos lineares generalizados são semelhantes às usadas para modelos lineares clássicos, com algumas adaptações. Assim, por exemplo, na verificação da pressuposição de linearidade para o modelo linear clássico usam-se os vetores \mathbf{y} e $\hat{\boldsymbol{\mu}}$ enquanto que para o modelo linear generalizado devem ser utilizados \mathbf{z} , a variável dependente ajustada, e $\hat{\boldsymbol{\beta}}$, o preditor linear.

Os tipos de resíduos mais utilizados para os modelos lineares generalizados são:

(i) Resíduos ordinários: $r_i = y_i - \hat{\mu}_i = y_i - g^{-1}(\eta_i);$

- (ii) Resíduos de Pearson generalizados: $r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)}}$, sendo $\hat{\phi}$ uma estimativa consistente do parâmetro ϕ e w_i um peso a priori (na maior parte dos casos igual a 1);

- (iii) Resíduos de Pearson generalizados estudentizados internamente

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)(1 - h_i)}},$$

Sendo h_i um elemento da diagonal da matriz de projeção $\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{\frac{1}{2}}$.

- (iv) Componentes de deviance

$$r_i^D = \text{sinal}(y_i - \hat{\mu}_i) \sqrt{\frac{2w_i}{\phi} [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\tilde{\theta}_i) - b(\hat{\theta}_i)]};$$

sendo $\tilde{\theta}_i$ e $\hat{\theta}_i$ as estimativas do parâmetro canônico sob os modelos saturado e corrente, respectivamente.

- (v) Componentes de deviance estudentizado internamente

$$r_i^{D'} = \frac{r_i^D}{\sqrt{\hat{\phi}(1 - h_i)}}.$$

2.6.6 Superdispersão

A utilização de MLG na análise de dados tem se tornado de uso frequente, principalmente com o avanço dos recursos computacionais disponíveis. De acordo com Gbur et al. (2012), um dos cuidados que se deve tomar na análise de dados, principalmente no caso de variáveis discretas, é com a superdispersão que pode ocorrer.

Um conjunto de dados, às vezes, apresenta variabilidade maior do que a esperada pelos modelos probabilísticos padrões, ou seja, uma deviance residual maior do que a esperada. Esse fenômeno é conhecido como superdispersão e Hinde e Demétrio (1998) apresentam algumas das possíveis causas deste fenômeno dadas a seguir:

- (i) Variabilidade do material experimental, que pode ser devida à variabilidade individual, gerando um componente aleatório adicional que não é levado em consideração na análise do modelo básico;
- (ii) Correlação entre respostas individuais, que pode ocorrer entre indivíduos do mesmo grupo, por exemplo, no estudo de doenças de plantas, pode haver uma correlação entre plantas da mesma unidade experimental;
- (iii) Amostragem por conglomerados (*cluster*) ;
- (iv) Omissão de variáveis não observadas;
- (v) Contagem com excessos de zeros.

O fato de não considerar a superdispersão na análise dos dados pode levar à estimação incorreta dos erros padrões, sendo os mesmos super ou subestimados, consequentemente uma avaliação incorreta da significância dos parâmetros do modelo. Portanto, é essencial verificar a presença de superdispersão no modelo linear generalizado ou no modelo linear generalizado misto para garantir que as inferências derivadas do modelo ajustado sejam precisas.

A distribuição padrão para análise de dados de proporção é a distribuição binomial, enquanto que para contagens é a Poisson. Essas distribuições têm como pressuposições:

- (i) Independência entre as observações e
- (ii) A mesma probabilidade de sucesso no caso de proporções, ou a mesma média no caso de contagens, para todos os indivíduos.

Se uma destas suposições não é satisfeita, a variação residual pode ser maior do que aquela predita pelo modelo, ou seja,

- (i) Dados de proporção com $Var(Y_i) > m_i \pi_i (1 - \pi_i)$ e
- (ii) Dados de contagem com $Var(Y_i) > \mu_i$,

sendo Y_i uma variável resposta independente.

Nestes casos tem-se que $\phi > 1$, fato conhecido por superdispersão. Pode ocorrer, também, a subdispersão, situação em que $\phi < 1$.

Diferentes modelos e métodos de estimação têm sido propostos na literatura para resolver o problema da superdispersão, como os descritos por: Collett (1991), Williams (1982), McCullagh e Nelder (1989) e Lindsey (1997), entre outros. Hinde e Demétrio (1998) apresentam modelos que incorporam a superdispersão e discutem os métodos de estimação dos parâmetros e técnicas para a verificação do ajuste. Essas técnicas podem ser estudadas, inicialmente, por procedimentos que se enquadrem, de uma forma geral, em dois grupos:

- (i) Assumir uma forma mais geral para a função de variância, possivelmente incluindo parâmetros adicionais que podem ser estimados por quase-verossimilhança (modelos de média-variância), pseudo-verossimilhança e método dos momentos.
- (ii) Assumir um modelo de dois estágios para a resposta, isto é, assumir que o parâmetro do modelo para a resposta tem alguma distribuição, levando a modelos de probabilidade compostos, como o binomial negativo, beta-binomial, Poisson-normal, logístico-normal e probit-normal podendo-se utilizar para estimação dos parâmetros o método da máxima verossimilhança ou métodos aproximados como em (i).

2.7 Modelos lineares generalizados mistos

A teoria de modelos lineares generalizados considera apenas o estudo de fatores de efeitos fixos. Uma extensão natural são modelos que se ajustam a dados obtidos a partir de experimentos em que os níveis de um fator foram selecionados de uma população de níveis, isto é, são aleatórios.

Desta forma, como uma extensão dos modelos lineares generalizados (MLG), McCulloch (2003) cita os modelos lineares generalizados mistos (MLGM), que incorporam efeitos aleatórios no preditor linear de um modelo pertencente à família exponencial. De forma análoga aos modelos mistos, os MLGM contêm pelo menos um efeito fixo e pelo menos um efeito aleatório.

Considere a distribuição condicional de \mathbf{Y} dado \mathbf{u} , sendo \mathbf{Y} o vetor de valores da variável resposta consistente de elementos condicionalmente independentes (não necessariamente) com densidade pertencente à família exponencial e \mathbf{u} o vetor de efeitos

aleatórios, formalmente tem-se, de acordo com Molenberghs e Verbeke (2005) o modelo a seguir:

$$Y_i|\mathbf{u}_i \sim f_{Y_i|\mathbf{u}_i}(y_i|\mathbf{u}_i)$$

com

$$f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}_i) = \exp \left\{ \frac{w_i}{\phi} [y_i\theta_i - b(\theta_i)] + c(y_i; \phi) \right\}. \quad (16)$$

visto que $\mu_i = b'(\theta_i)$ decorre:

$$E(Y_i|\mathbf{u}_i) = \mu_i,$$

que é a média condicional da variável aleatória Y_i dado \mathbf{u}_i , que é funcionalmente ligada à parte sistemática por uma função de ligação:

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}_i, \quad (17)$$

e

$$Var(Y_i|\mathbf{u}_i) = \phi V(\mu_i),$$

em que $g(\cdot)$ é a função de ligação; \mathbf{x}_i é a i -ésima linha da matriz de delineamento associada aos efeitos fixos; $\boldsymbol{\beta}$ é o vetor de parâmetros dos efeitos fixos; \mathbf{z}_i é a i -ésima linha da matriz do modelo associada aos efeitos aleatórios e \mathbf{u} é o vetor dos parâmetros dos efeitos aleatórios.

A parte sistemática do modelo (17) inclui tanto os efeitos fixos como os aleatórios. Assume-se que:

$$\mathbf{u}_i \sim f_U(\mathbf{u}). \quad (18)$$

De acordo com Molenberghs e Verbeke (2005), um fato comum é pressupor que \mathbf{u} segue uma distribuição conhecida, sendo usual atribuir distribuição normal multivariada aos efeitos aleatórios, ou seja, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$. Uma discussão mais aprofundada sobre os efeitos aleatórios não-normais é dada por Lee e Nelder (1996). Considerando o modelo condicional (16), tem-se que:

$$E(Y_i) = E[E(Y_i|\mathbf{u}_i)] = E[\mu_i] = E[g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}_i)],$$

que, em geral, não pode ser simplificado devido à presença de funções não-lineares em $g^{-1}(\cdot)$. A variância marginal de Y_i , por sua vez, é dada por:

$$\begin{aligned} Var(Y_i) &= Var[E(Y_i|\mathbf{u}_i)] + E[Var(Y_i|\mathbf{u}_i)] \\ &= V[\mu_i] + E[a(\phi)V(\mu_i)] \\ &= Var[g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}_i)] + E\{a_i(\phi)Var[g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}_i)]\} \end{aligned}$$

em que $a_i(\phi) = \frac{w_i}{\phi}$, não sendo possível simplificações sem fazer suposições específicas sobre a forma de $g(\cdot)$ e/ou a distribuição condicional de \mathbf{Y} .

O uso de efeitos aleatórios também introduz uma correlação entre observações que tenham algum efeito em comum. Assumindo-se independência condicional dos elementos Y_i , tem-se:

$$\begin{aligned} Cov(Y_i, Y_j) &= Cov[E(Y_i|\mathbf{u}_i), E(Y_j|\mathbf{u}_i)] + E[Cov(Y_i, Y_j|\mathbf{u}_i)] \\ &= Cov(\mu_i, \mu_j) + E(0) \\ &= Cov[g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}_i), g^{-1}(\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{z}_j^T\mathbf{u}_i)]. \end{aligned}$$

Segundo Costa (2003), os estimadores resultantes dependem da função geradora de momentos da variável aleatória.

2.7.1 Inferência em modelos lineares generalizados mistos

De acordo com Gbur et al.(2012) a inferência em modelos lineares generalizados mistos (MLGM) envolve a estimação dos coeficientes dos efeitos fixos, $\boldsymbol{\beta}$, predição dos parâmetros de f_U e, possivelmente estimar o parâmetro de escala, ϕ . Além disso, o teste de hipóteses a respeito dos parâmetros desconhecidos em \mathbf{R} e \mathbf{G} . Da mesma forma como nos modelos lineares generalizados, o processo de inferência é realizado na escala da função de ligação e não na escala original da variável resposta.

A estimação em MLGM está baseada no método da máxima verossimilhança e Costa (2003) apresenta várias abordagens para estimar os parâmetros do modelo (16). Dentre elas, Schall (1991) sugeriu estimação de máxima verossimilhança similar ao que é utilizado para modelos mistos; Breslow e Clayton (1993) estudaram um tipo de estimador de máxima verossimilhança marginal; McGilchrist (1994) recomendou o melhor preditor linear não-viesado, enquanto que Lee e Nelder (1996) introduziram um método geral chamado estimação de máxima verossimilhança hierárquica.

Gbur et al. (2012) reportam duas abordagens computacionais básicas presentes na literatura, para obter as soluções das equações de verossimilhança:

- (i) Integração numérica para o cálculo e a maximização numérica da verossimilhança (para aproximar a função objetivo). Várias técnicas são utilizadas para calcular a aproximação: métodos de Laplace, métodos de quadratura de Gauss-Hermite, algoritmo de Monte Carlo, cadeias de Markov etc. como principal vantagem destes métodos, Pinheiro e Bates (2000) destacam a eficiência computacional para aproximar a função de verossimilhança e a maior eficiência dos testes de razão de verossimilhança. No entanto, o seu uso torna-se mais complicado à medida que se aumenta o número de efeitos aleatórios.
- (ii) Métodos de linearização usando expansões de séries de Taylor para aproximar o modelo. Neles se empregam expansões para aproximar o modelo utilizando como base pseudodados com poucos componentes não lineares, sendo os procedimentos mais comuns, os de Quasi-verossimilhança (QL) propostos por Breslow e Clayton (1993) e os de Pseudo-verossimilhança (PL) de Wolfinger e O'Connell (1993).

Berdugo (2010) cita como vantagens, que os modelos linearizados possuem uma forma relativamente simples, que tipicamente pode ser ajustada baseada unicamente na média e na variância da forma linearizada, por outra parte, os modelos com distribuição conjunta complicada são facilmente tratados e, ainda, é possível abordar modelos com um grande número de efeitos fixos e aleatórios, assim como com estruturas aninhadas e cruzadas. Porém, são criticados principalmente por fornecer estimadores viesados e inconsistentes.

2.7.2 Estimação por máxima verossimilhança

De (16), (17) e (18) pode-se escrever a função de verossimilhança .

$$L = \prod_{i=1}^n \int f(\mathbf{y}_i | \mathbf{u}) f_U(\mathbf{u}) d\mathbf{u} = f_Y(\mathbf{y}), \quad (19)$$

sendo que a integração é sobre a distribuição de \mathbf{u} , de dimensões q . Nos casos mais simples, a integração numérica para o cálculo da verossimilhança é direta e, conseqüentemente, a maximização numérica da função de verossimilhança, não é difícil, já que o logaritmo da

função de verossimilhança é a soma das contribuições independentes de cada agrupamento, que envolve apenas uma integral de dimensão única, que pode ser calculada usando-se técnicas de quadratura (Costa, 2003).

A seguir, serão apresentadas as equações de verossimilhança para parâmetros fixos e aleatórios, conforme descrito por Costa (2003)

Equações de verossimilhança para parâmetros fixos

Embora uma solução para as equações de verossimilhança seja numericamente difícil, pode-se obtê-las de uma forma mais simples. De (19), tem-se:

$$\ell = \log f_Y(\mathbf{y}). \quad (20)$$

Assim,

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[\log \int f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u} \right] \\ &= \frac{1}{\int f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u}} \left[\frac{\partial}{\partial \boldsymbol{\beta}} \int f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u} \right] \\ &= \frac{1}{f_Y(\mathbf{y})} \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \right] f_U(\mathbf{u}) d\mathbf{u}, \end{aligned} \quad (21)$$

pois $f_U(\mathbf{u})$ não envolve $\boldsymbol{\beta}$. Mas,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) &= \frac{1}{f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u})} \left[\frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \right] f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \\ &= \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \right] f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \end{aligned}$$

Então,

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \frac{1}{f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u})} \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) f_{Y|U}(\mathbf{y}|\mathbf{u}) \right] f_U(\mathbf{u}) d\mathbf{u} \\
&= \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \right] \frac{f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) f_U(\mathbf{u}) d\mathbf{u}}{f_Y(\mathbf{y})} \\
&= \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{Y|\mathbf{u}}(\mathbf{y}|\mathbf{u}) \right] f_{U|Y}(\mathbf{u}|\mathbf{y}) d\mathbf{u}.
\end{aligned} \tag{22}$$

Além disso, $\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \mathbf{W} \boldsymbol{\Delta}(\mathbf{y} - \boldsymbol{\mu})$,

sendo

$$\mathbf{W} = \text{diag} \{W_i\} = \text{diag} \left[a_i(\phi) \text{Var}(\mu_i) \frac{\partial \eta_i}{\partial \mu_i} \right]^2$$

e

$$\boldsymbol{\Delta} = \text{diag} \left\{ \frac{\partial \eta_i}{\partial \mu_i} \right\}.$$

Portanto,

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \int \mathbf{X}^T \mathbf{W}^* (\mathbf{y} - \boldsymbol{\mu}) f_{U|Y}(\mathbf{u}|\mathbf{y}) d\mathbf{u},$$

em que $\mathbf{W}^* = \text{diag} \left[a(\phi) \text{V}(\mu_i) \frac{\partial \eta_i}{\partial \mu_i} \right]^{-1}$. Logo,

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y} E[\mathbf{W}^*|\mathbf{y}] - \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu}|\mathbf{y}],$$

e consequentemente,

$$\mathbf{X}^T \mathbf{y} E[\mathbf{W}^*|\mathbf{y}] = \mathbf{X}^T E[\mathbf{W}^* \boldsymbol{\mu}|\mathbf{y}].$$

Equações de verossimilhança para parâmetros aleatórios

Um resultado similar ao obtido na equação (22) pode ser encontrado para equações de máxima verossimilhança para os parâmetros na distribuição $f_U(\mathbf{u})$. Denotando $\boldsymbol{\gamma}$ os parâmetros dos efeitos aleatórios, de (20) tem-se:

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\gamma}} &= \frac{1}{f(\mathbf{y})} \frac{\partial f(\mathbf{y})}{\partial \boldsymbol{\gamma}} \\
&= \frac{1}{f(\mathbf{y})} \frac{\partial}{\partial \boldsymbol{\gamma}} \left[\int f(\mathbf{y}|\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \right] \\
&= \int \left[\frac{\partial}{\partial \boldsymbol{\gamma}} f(\mathbf{y}|\mathbf{u}) \right] \frac{f(\mathbf{u})}{f(\mathbf{y})} d\mathbf{u} + \int \frac{f(\mathbf{y}|\mathbf{u})}{f(\mathbf{y})} \left[\frac{\partial}{\partial \boldsymbol{\gamma}} f(\mathbf{u}) \right] d\mathbf{u} \\
&= \int \frac{f(\mathbf{y}, \mathbf{u})}{f(\mathbf{y})} \frac{1}{f(\mathbf{u})} \left[\frac{\partial}{\partial \boldsymbol{\gamma}} f(\mathbf{u}) \right] d\mathbf{u} \\
&= \int f_{U|\mathbf{y}}(\mathbf{u}, \mathbf{y}) \left[\frac{\partial}{\partial \boldsymbol{\gamma}} \log f_U(\mathbf{u}) \right] d\mathbf{u} \\
&= E \left[\frac{\partial}{\partial \boldsymbol{\gamma}} \log f_U(\mathbf{u}) | \mathbf{y} \right],
\end{aligned}$$

que não pode ser simplificada sem que se especifique uma forma para a distribuição dos efeitos aleatórios.

Para a obtenção dos estimadores de máxima verossimilhança, Vieira (2008) apresenta um resumo dos principais algoritmos atualmente implementados nos *softwares* estatísticos. Destacando dentre os algoritmos para otimização de funções, os métodos de Newton-Raphson, da secante, *scoring* de Fisher e Quase-Newton; e dentre os algoritmos para integração numérica, os métodos da Quadratura de Newton-Côtes, de Laplace e Quadratura de Gauss-Hermite.

2.7.3 Estimação e inferência baseada em métodos de linearização

No ajuste dos MLG como proposto por Nelder e Wedderburn (1972), as equações de estimação tem a seguinte forma matricial (eq. 15):

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{W} \mathbf{z},$$

em que \mathbf{X} , de dimensões $n \times p$, é a matriz de delineamento do modelo, \mathbf{W} , de dimensões $p \times p$, é uma matriz diagonal de pesos dada por $\frac{1}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$, e \mathbf{z} , uma variável dependente ajustada definida em termos da variável original Y por:

$$\mathbf{z} = \boldsymbol{\eta} + (\mathbf{y} - \boldsymbol{\mu}) \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}} \right).$$

A estimação é implementada mediante um processo iterativo no qual se atualiza a matriz de pesos, o vetor de parâmetros e a pseudo-variável. Esta ideia é estendida por Wolfinger e O'Connell (1993) a um MLGM aproximando a expressão $g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$ por meio de uma série de Taylor de primeira ordem centrada nos valores de $\tilde{\boldsymbol{\beta}}$ e $\tilde{\mathbf{u}}$, ou seja:

$$\boldsymbol{\mu} \approx g^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\boldsymbol{\Delta}} \left[\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \mathbf{Z}(\mathbf{u} - \tilde{\mathbf{u}}) \right], \quad (23)$$

em que $\tilde{\boldsymbol{\Delta}} = \left(\frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)_{\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{u}}}$ e $\tilde{\boldsymbol{\eta}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}$, sendo $\tilde{\boldsymbol{\beta}}$ e $\tilde{\mathbf{u}}$ estimativas e predições para os vetores de efeitos fixos e aleatórios, respectivamente.

Reorganizando a eq.(23), obtém-se:

$$E(\mathbf{P}|\mathbf{u}) = \tilde{\boldsymbol{\Delta}}^{-1} [\boldsymbol{\mu} - g^{-1}(\tilde{\boldsymbol{\eta}})] + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \quad (24)$$

em que $\mathbf{P} = \tilde{\boldsymbol{\Delta}}^{-1} [\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})] + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}$ é uma pseudo-variável com esperança e variância condicionais dadas por $E(\mathbf{P}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ e $Var(\mathbf{P}|\mathbf{u}) = \mathbf{S} = \tilde{\boldsymbol{\Delta}}^{-1} \mathbf{G} \tilde{\boldsymbol{\Delta}}^{-1}$.

Com base nessas aproximações, Wolfinger e O'Connell (1993) consideram que a variável \mathbf{P} segue o modelo linear misto:

$$\mathbf{P} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \text{ com } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{S}). \quad (25)$$

As equações (24) e (25) implicam que a pseudo-resposta \mathbf{P} tem uma distribuição aproximadamente normal com média $\mathbf{X}\boldsymbol{\beta}$ e variância $Var(\mathbf{P}|\mathbf{u}) = \mathbf{S} = \tilde{\boldsymbol{\Delta}}^{-1} \mathbf{G} \tilde{\boldsymbol{\Delta}}^{-1}$, em que \mathbf{G} , de dimensão q , contém todos os parâmetros das matriz de variâncias e covariâncias dos efeitos aleatórios.

Estas especificações levam ao logaritmo da função de pseudo-verossimilhança $\ell(\boldsymbol{\Psi}; \mathbf{p})$ e ao logaritmo da função de verossimilhança restrita $\ell_R(\boldsymbol{\Psi}; \mathbf{p})$, dadas, respectivamente, por:

$$\ell(\Psi; \mathbf{p}) = -\frac{1}{2} \log |\mathbf{V}(\Psi)| - \frac{1}{2} \mathbf{r}^T \mathbf{V}(\Psi)^{-1} \mathbf{r} - \frac{f}{2} \log\{2\pi\} \quad \text{e} \quad (26)$$

$$\begin{aligned} \ell_R(\Psi; \mathbf{p}) = & -\frac{1}{2} \log |\mathbf{V}(\Psi)| - \frac{1}{2} \mathbf{r}^T \mathbf{V}(\Psi)^{-1} \mathbf{r} + \\ & -\frac{1}{2} \log |\mathbf{X}^T \mathbf{V}(\Psi)^{-1} \mathbf{X}| - \frac{f-k}{2} \log\{2\pi\} \end{aligned} \quad (27)$$

em que $\mathbf{r} = \mathbf{p} - \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{p}$, f é a soma de todas as frequências utilizadas na análise e r o posto da matriz \mathbf{X} .

Neste caso, os parâmetros de Ψ são estimados usando técnicas de otimização como: Newton-Raphson, quase-Newton, Dual quasi-Newton etc. O processo de estimação inicia tomando valores iniciais de $\tilde{\beta}$ e $\tilde{\mathbf{u}}$ para formar \mathbf{r} e substituí-lo em (26) e (27). Logo se maximiza alguma destas expressões com relação a Ψ e ao obter um estimador $\hat{\Psi}$ se atualizam $\tilde{\beta}$ e $\tilde{\mathbf{u}}$ usando:

$$\tilde{\beta} = [\mathbf{X}^T \mathbf{V}(\hat{\Psi})^{-1} \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{V}(\hat{\Psi})^{-1} \mathbf{p}$$

$$\tilde{\mathbf{u}} = \hat{\mathbf{D}} \mathbf{Z}^T \mathbf{V}(\hat{\Psi})^{-1} \hat{\mathbf{r}}$$

O processo de otimização-atualização se repete até que algum critério de convergência é atingido, geralmente a mudança relativa é menor ou igual que algum valor predefinido (McCULLOCH; SEARLE, 2001).

No caso em que a distribuição condicional contém um parâmetro de dispersão ϕ , sua estimação é realizada somente no procedimento PQL, por meio da expressão:

$$\hat{\phi} = \frac{1}{m} \hat{\mathbf{r}}^T \mathbf{V}(\hat{\Psi})^{-1} \hat{\mathbf{r}}.$$

A escolha de m depende do método de estimação, de modo que para o método da máxima verossimilhança (ML), $m = f$ e no método da máxima verossimilhança restrita (REML), $m = f - k$.

Inferência para parâmetros de efeitos fixos

De acordo com Littell et al. (2006), de maneira análoga aos MLMs, nos MLGMS baseados na linearização, o procedimento de inferência para os parâmetros de efeito fixo é realizado a partir da função de predição $\mathbf{K}\boldsymbol{\beta} + \mathbf{M}\mathbf{u}$, sendo o teste de Wald um dos mais utilizados, cuja estatística W é dada por:

$$W = (\mathbf{L}^T \hat{\boldsymbol{\Psi}})^T (\mathbf{L}^T \mathbf{C} \mathbf{L}) (\mathbf{L}^T \hat{\boldsymbol{\Psi}}),$$

em que:

$$\mathbf{L}^T = [\mathbf{K} \ \mathbf{M}], \hat{\boldsymbol{\Psi}} = [\tilde{\boldsymbol{\beta}}^T, \tilde{\mathbf{u}}^T]^T_e \quad \mathbf{C} = \begin{bmatrix} \mathbf{X}^T \mathbf{S}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{S}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{S}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{S}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}. \quad (28)$$

Quando em (28) \mathbf{S} independe de algum parâmetro de escala ϕ , a estatística de Wald se distribui assintoticamente como uma variável χ^2 com número de graus de liberdade $\nu_1 = \text{posto}(\mathbf{L})$ (LITTELL et al., 2006). Nos casos em que \mathbf{S} depende de um parâmetro de escala desconhecido ϕ , prefere-se utilizar a estatística de Wald dividida por ν_1 , obtendo-se a estatística F_W :

$$F_W = W/\nu_1$$

. Esta estatística está assintoticamente distribuída como uma variável $F(\nu_1, \nu_2)$, em que $\nu_1 = \text{posto}(\mathbf{L})$ e ν_2 são os graus de liberdade utilizados para estimar $\mathbf{L}^T \mathbf{C} \mathbf{L}$, que devem ser aproximados usando métodos como os propostos por Satterthwaite, Kenward Roger, entre outros.

Inferência sobre os componentes de variância

As hipóteses relativas aos componentes de variância da matriz \mathbf{G} podem ser avaliadas por meio do teste da razão de verossimilhança (LRT), cujos detalhes são apresentados por Casella e Berger (2002). O método se fundamenta em comparar as verossimilhanças de dois modelos, nos quais os parâmetros são obtidos de dois espaços Ω_o de dimensão s_o e Ω de dimensão s com parâmetros $\Omega_o \subset \Omega$ e $s_o < s$. Se $\hat{\boldsymbol{\theta}}_o$ e $\hat{\boldsymbol{\theta}}$ são os estimadores de máxima verossimilhança em cada modelo, a estatística (Λ) é definida como:

$$\Lambda = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_o)],$$

em que ℓ é a função logaritmo de verossimilhança, que é avaliada nas estimativas de cada parâmetro dentro dos espaços previamente definidos. Quando nenhum dos parâmetros especificados no modelo nulo pertence à fronteira de Ω , a estatística Λ possui uma distribuição assintótica χ^2 com $s - s_o$ graus de liberdade. Em caso contrário, quando pelo menos um dos parâmetros no modelo nulo pertence à fronteira de Ω , Pinheiro e Bates (2000) afirmam que a distribuição da estatística Λ é uma mistura de distribuições χ^2 .

3 MATERIAL E MÉTODOS

3.1 Material

O palmitreiro (*Euterpe edulis* Martius), segundo Reis (2000), é uma planta nativa do domínio floresta tropical Atlântica do Brasil, que tem um estipe (caule lenhoso) único de 5 a 15 m de altura e é incapaz de produzir perfilhos, o que acarreta na morte da planta após corte do palmito. Brancalion (2009) salienta que esta espécie tem sido considerada importante para os frugívoros (animais que se alimentam de frutas) no bioma da Mata Atlântica, uma vez que produz frutos que são consumidos pelo menos por 30 tipos de aves e 15 espécies de mamíferos, principalmente em períodos de escassez de alimentos.

Assim, de modo a ilustrar a metodologia apresentada, foram utilizados dados relativos a dois grupos de experimentos com transplantes recíprocos, sendo o primeiro, um ensaio de estabelecimento de plântulas ou de semeadura e o segundo, um ensaio de crescimento de juvenis. Estes ensaios foram conduzidos por Brancalion (2009), com o objetivo de avaliar a adaptação local e a plasticidade fenotípica de três procedências de palmitreiro em parcelas permanentes inseridas em três formações florestais do Estado de São Paulo.

Nesses ensaios, o autor coletou sementes das populações de *E. edulis* presentes em três tipos de florestas do bioma da Mata Atlântica. Nestes mesmos locais foi instalado cada um dos experimentos, envolvendo indivíduos das três procedências. A seguir são descritas algumas características desses tipos de florestas:

- (i) Restinga (*Restinga Forest*), no Parque Estadual da Ilha do Cardoso (IC). Este parque está localizado no litoral sul do Estado de São Paulo, na divisa com o Estado do Paraná, abrangendo uma área aproximada de 151 km², situando-se entre as coordenadas geográficas: 48°05'42" e 48 53'48" longitude oeste, e 25°03'05" e 25°18'18" latitude sul. A altitude média é de 7 m, temperatura média anual de 22,4°C, precipitação anual de 2261 mm, sem estação seca, clima equatorial com alta influência do oceano.
- (ii) Ombrófila Densa (*Atlantic Rainforest*). Situa-se no Parque Estadual de Carlos Botelho (CB) na Serra de Paranapiacaba, região sudeste do estado de São Paulo, ocupando uma área de 376,44 km², entre as coordenadas geográficas 47°56'57" longitude

oeste e 24°07'53" latitude sul, com um relevo acidentado, indo de 50 a 975 m de altitude, a temperatura média anual é de 21,8°C, precipitação anual de 1582 mm, sem estação seca, clima equatorial.

- (iii) Estacional Semidecidual (*Seasonally Dry Forest*) na Estação Ecológica dos Caetetus (CT), localizada entre as coordenadas geográficas 49°42'05" longitude oeste e 22°24'11" latitude sul, tem área de 21,79 km², e altitudes variando de 500 a 690 m, temperatura média anual: 21,4°C, precipitação anual de 1303 mm, com estação seca no inverno (déficit de água= 10 mm), clima sub-tropical úmido.

Essas populações foram encontradas em crescimento no entorno de 10 ha de parcelas permanentes de cada tipo de floresta, que são protegidas em reservas naturais do Estado de São Paulo. Na Figura 1 mostram-se os locais onde foram coletadas as sementes de *E. edulis* e instalados os experimentos. Os quadrados negros dentro de cada área representam as parcelas permanentes de 10 ha cujo entorno foi utilizado para a coleta de sementes.

A coleta das sementes foi realizada no início da época de frutificação em 10 plantas matrizes (plantas mães, das quais são coletadas sementes ou propágulos para a produção de mudas) em duas trilhas de cada reserva natural, em maio de 2008 e em maio de 2009. As plantas matrizes escolhidas estavam separadas por uma distância mínima de 100 m, a fim de reduzir as chances de fluxo gênico entre elas.

A polpa foi retirada dos frutos do palmitreiro colocando-os em uma tela de arame e logo enxaguados abundantemente em água corrente. Os frutos foram processados separadamente para cada grupo, evitando misturas de sementes entre plantas matrizes e procedências diferentes. A massa seca e teor de umidade foram avaliados de acordo com o método do forno (105°C \pm 3°C por 24 horas), utilizando 30 sementes por planta matriz.

3.1.1 Ensaio de estabelecimento (semeadura) em transplantes recíprocos para avaliação da adaptação local e plasticidade fenotípica

Brancaion (2009) relata que o ensaio de estabelecimento foi iniciado em maio de 2008, sete dias após a coleta de sementes para evitar a perda da viabilidade de sementes recalcitrantes de *E. edulis*. Essas sementes caracterizam-se por não sofrerem dessecação natural na planta matriz ao longo do processo de maturação, sendo dispersas

com elevados teores de água que, se reduzidos a um nível considerado crítico, levarão à rápida perda da viabilidade e até à morte, e mesmo quando a umidade for mantida em nível adequado durante o armazenamento, sua longevidade é curta (NEVES, 1994).

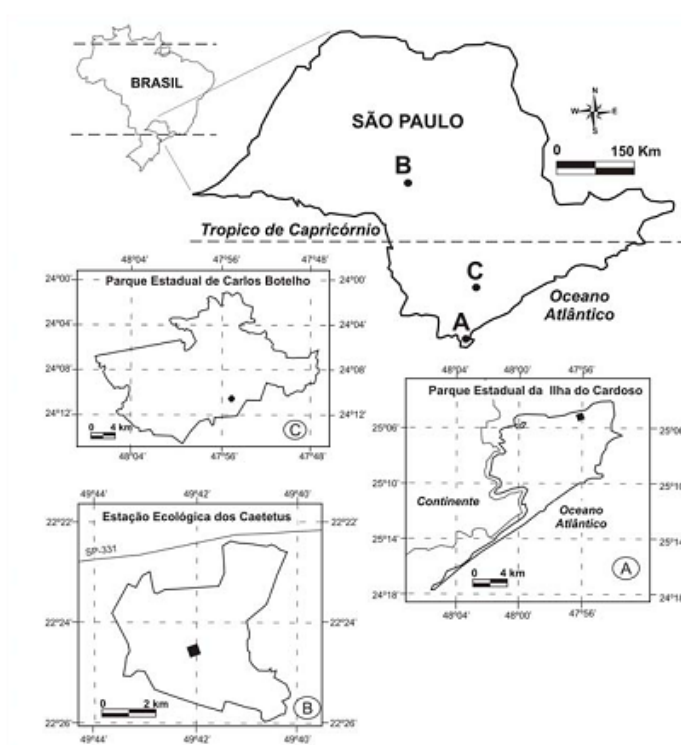


Figura 1 – Locais de estudo de *Euterpe edulis* Martius, A - Parque Estadual da Ilha do Cardoso, B - Estação Ecológica dos Caetetus e C - Parque Estadual de Carlos Botelho. Reproduzido de BRANCALION, 2009

Ao longo de dois transectos traçados em forma aleatória em cada tipo de floresta foram alocados dez blocos experimentais, e situados a 100 m um do outro e 30 m à direita ou à esquerda do transecto, alternadamente, como ilustra a Figura 2.

Cada bloco experimental teve três parcelas utilizadas neste ensaio com as três procedências atribuídas aleatoriamente. Em cada parcela, foram semeadas 30 sementes de cada procedência, ou seja, três sementes para cada planta matriz e dez plantas matrizes por tipo de floresta (Figura 3).

Dado que cada planta matriz forneceu sementes para todos os experimentos, Brancalion (2009) considera que pelo menos metade do seu genoma foi exatamente repetido nos três tipos de floresta. Sementes das mesmas plantas matrizes de cada procedência foram utilizadas em todos os experimentos.

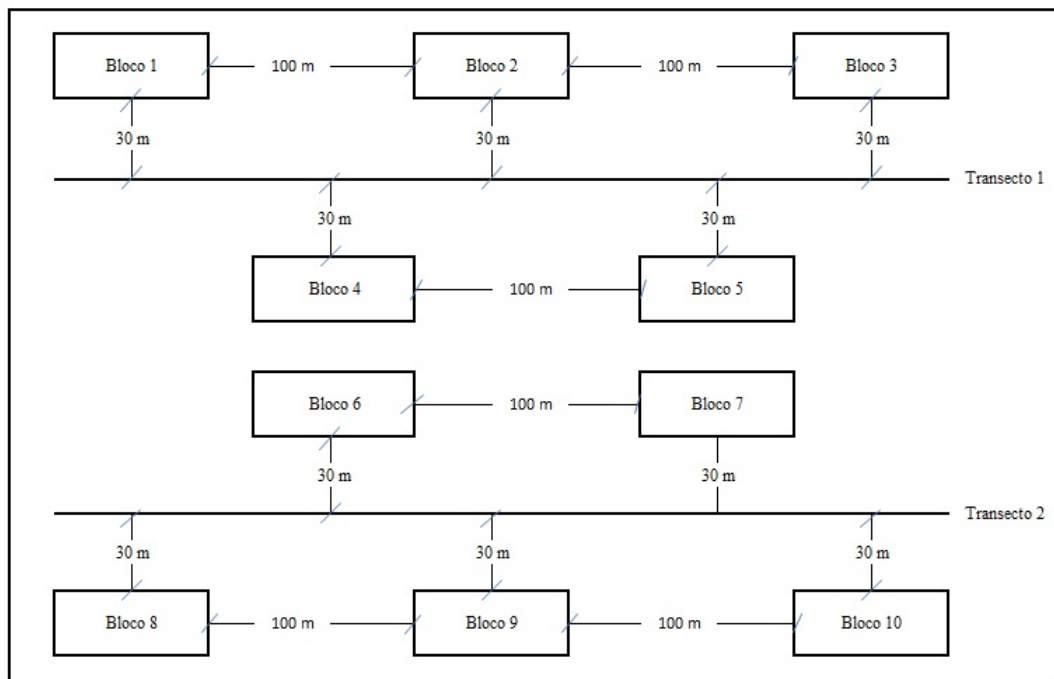


Figura 2 – Croqui do experimento de estabelecimento (semeadura), para um local determinado

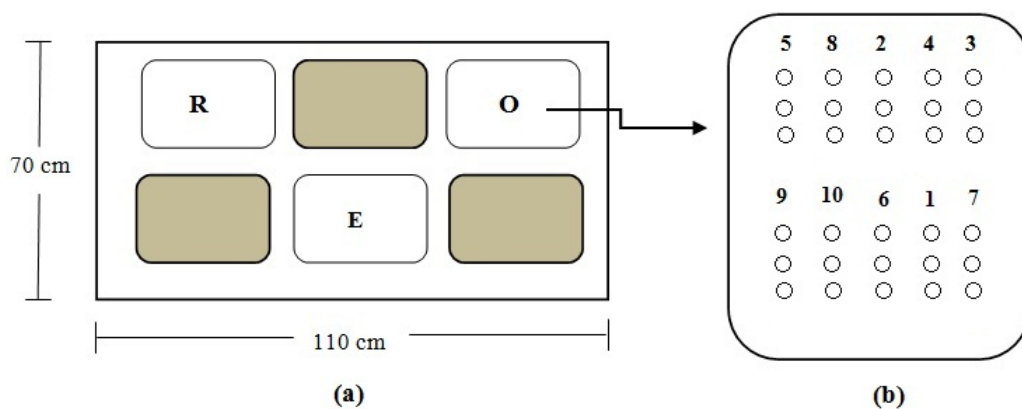


Figura 3 – Croqui de: (a) um bloco qualquer mostrando a casualização das procedências das sementes: R (Restinga), E (Estacional Semidecidual) e O (Ombrófila Densa), as parcelas sombreadas pertencem a outro projeto; e (b) uma parcela com três sementes de cada planta-matriz do mesmo ecossistema

A localização exata das sementes de cada planta matriz e proveniência foi marcada com etiquetas de plástico. Os blocos experimentais foram protegidos por gaiolas de exclusão para evitar a remoção de sementes por mamíferos e a deposição de sementes por agentes de dispersão. As gaiolas foram feitas com barras de ferro (110 cm de comprimento \times 70 cm largura \times 15 cm de altura) e cobertas por uma malha de arame, com aberturas circulares de 1 cm. As sementes foram enterradas no solo a uma profundidade equivalente à metade do seu diâmetro, e a outra metade se deixou exposta. Assim é a prática comum dos viveiristas que produzem essas espécies comercialmente.

Neste experimento avaliou-se o número de plântulas presentes nas parcelas aos 90, 180 e 270 dias após a semeadura. As plântulas, por sua vez, foram marcadas individualmente com etiqueta de plástico, a fim de avaliar a emergência e a mortalidade durante o período. A população final de plântulas foi avaliada na última contagem (270 dias), quando a parte aérea de todas as plântulas foi removida das parcelas e inserida individualmente em sacos de papel. A parte aérea das plântulas foi seca em estufa a 72°C durante 48 horas, para obter a massa de matéria seca da folha e do caule.

As variáveis respostas medidas neste ensaio foram: proporção de sementes germinadas por planta matriz de cada procedência, valores médios de massa de matéria seca (em miligramas) da folha, do caule e da parte aérea (folha e caule), por planta matriz de cada procedência.

3.1.2 Ensaio de crescimento de mudas (juvenis) em transplantes recíprocos para avaliação da plasticidade e adaptação local.

Grupos de sementes de cada planta matriz das procedências avaliadas foram semeadas separadamente, em caixas plásticas contendo areia. Mudas de 90 dias de idade foram transplantadas em tubos de plásticos de 56 cm³ preenchidos com uma mistura livre de nutrientes orgânicos, onde as mudas dependiam exclusivamente de suas reservas do endosperma para o desenvolvimento, o que é desejável porque um substrato adubado poderia mascarar os efeitos das condições naturais do solo sobre o crescimento em cada área de estudo.

As mudas foram mantidas em viveiro florestal e, posteriormente, utilizadas em transplantes recíprocos realizados em novembro de 2008, no início da estação chuvosa. Os dez blocos experimentais foram construídos 2 m longe dos blocos utilizados

na avaliação descrita anteriormente no ensaio de caracteres adaptativos em mudas de estabelecimento. Cada bloco experimental teve três parcelas com as três procedências distribuídas aleatoriamente. Cada parcela esteve constituída por dez plantas, uma de cada planta matriz.

As plantas produzidas foram removidas das parcelas após 12 meses, e secas no forno. O desempenho durante o crescimento das mudas foi avaliado medindo a massa de matéria seca (mg) da raiz, da folha, do caule e total, por planta matriz de cada procedência.

3.2 Métodos

3.2.1 Modelos lineares mistos para avaliar o teor da massa de matéria seca das plantas

Para que os conjuntos de dados pudessem ser analisados de acordo com a proposta dos modelos lineares mistos, inicialmente foram identificados os fatores presentes em cada experimento, discriminando-os como fator aleatorizado ou fator não aleatorizado (BRIEN, 2007), sendo identificadas também suas camadas (BRIEN, 1983) e as respectivas fórmulas estruturais, indicando assim a relação entre os mesmos.

Embora os ensaios contenham dados faltantes, para uma melhor compreensão da sua estrutura, considerou-se uma série de experimentos, balanceados, realizados com I procedências em K locais, e utilizando um delineamento casualizado em blocos, com J repetições em cada local, e ainda, M plantas matrizes associadas a cada procedência. Construiu-se a tabela de decomposição dos números de graus de liberdade e esperanças dos quadrados médios (Tabela 1), fazendo uso do diagrama de Hasse (TAYLOR; HILTON, 1981), que é apresentado na Figura 4.

Segundo Machado et al. (2005) o diagrama de Hasse é uma poderosa ferramenta visual utilizada na representação da estrutura dos fatores de um delineamento experimental e fornece uma valiosa perspectiva complementar para a análise de variância e as técnicas de análise, por meio de uma conexão entre a descrição verbal do experimento e o correspondente modelo linear estatístico.

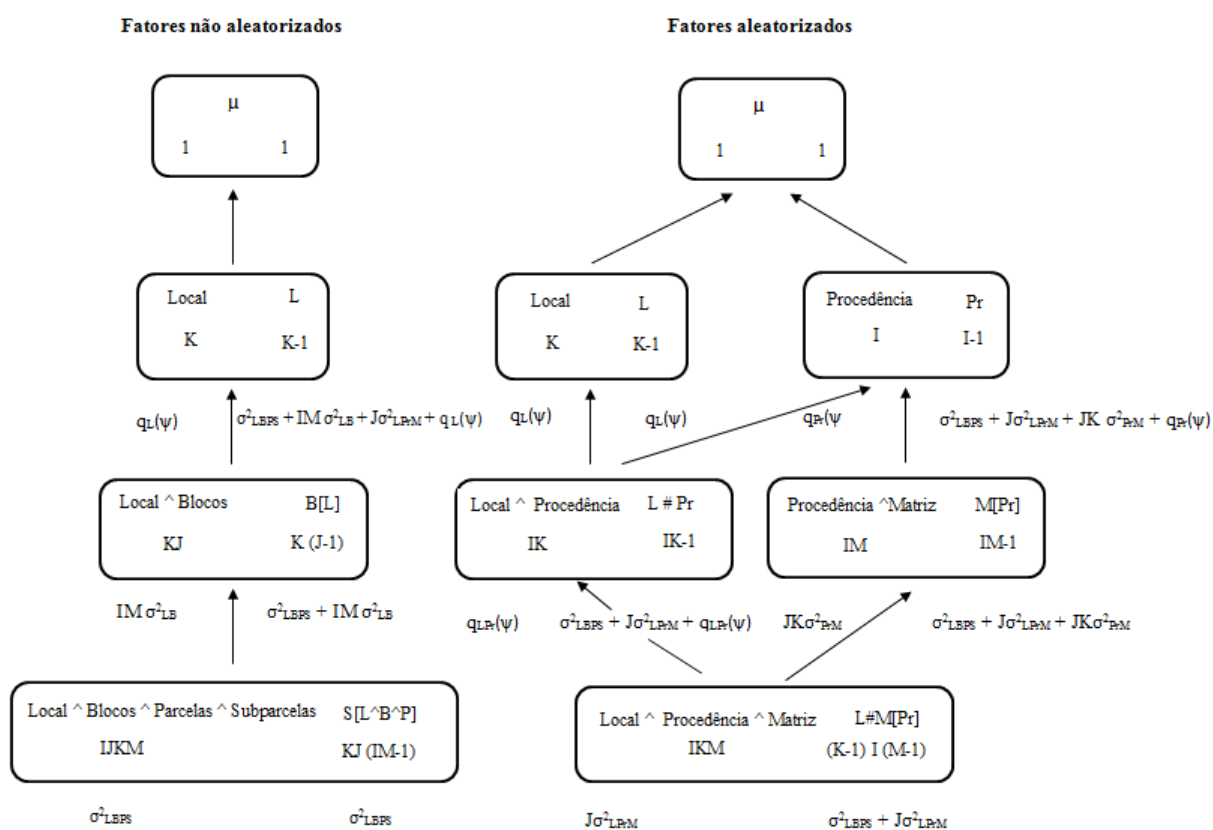


Figura 4 – Diagrama de Hasse para obtenção dos números de graus de liberdades e esperanças dos quadrados médios para os ensaios de estabelecimento (semeadura) e crescimento de mudas (juvenis)

Logo, com a estrutura definida, os dados foram analisados de acordo com o modelo estatístico matemático, a seguir:

$$y_{ijkm} = \mu + \tau_i + b_{j(k)} + \ell_k + \tau\ell_{ik} + \delta_{m(i)} + \ell\delta_{m(ik)} + \varepsilon_{ijkm}, \quad (21)$$

em que: $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$, $m = 1, \dots, M$, y_{ijkm} é a resposta observada referente à m -ésima planta matriz da i -ésima procedência, no j -ésimo bloco do k -ésimo local; μ é a constante associada à todas as observações; τ_i é o efeito da i -ésima procedência (efeito fixo), $b_{j(k)}$ é o efeito aleatório do j -ésimo bloco no k -ésimo local, $b_{j(k)} \sim N(0, \sigma_b^2)$, ℓ_k é o efeito fixo do k -ésimo local; $\tau\ell_{ik}$ representa o efeito da interação entre a i -ésima procedência e o k -ésimo local, $\delta_{m(i)}$ é o efeito aleatório da m -ésima planta-matriz dentro da i -ésima procedência, $\delta_{m(i)} \sim N(0, \sigma_\delta^2)$, $\ell\delta_{m(ik)}$ é o efeito aleatório da m -ésima planta-matriz dentro da ik -ésima interação local \times procedência, $\ell\delta_{m(ik)} \sim N(0, \sigma_{\ell\delta}^2)$ e ε_{ijkm} é o erro experimental aleatório associado às observações y_{ijkm} , $\varepsilon_{ijkm} \sim N(0, \sigma_\varepsilon^2)$.

Considerando-se $b_{j(k)}$, $\delta_{m(i)}$, $\ell\delta_{m(ik)}$ e ε_{ijkl} como independentes, a variância de uma observação é dada por: $\sigma_b^2 + \sigma_\delta^2 + \sigma_{\ell\delta}^2 + \sigma_\varepsilon^2$.

Tabela 1 – Decomposição dos números de graus de liberdade e esperanças dos quadrados médios para os ensaios de estabelecimento (semeadura) e crescimento de mudas (juvenis)

Fontes de variação	gl	E[QM]
Locais	$K - 1$	$\sigma_\varepsilon^2 + IM\sigma_{LB}^2 + J\sigma_{LP r M}^2 + q_L(\psi)$
Blocos [Locais]	$K(J - 1)$	$\sigma_\varepsilon^2 + IM\sigma_{LB}^2$
Subparcelas [L \wedge B \wedge P]	$KJ(IM - 1)$	
Procedências	$I - 1$	$\sigma_\varepsilon^2 + J\sigma_{LP r M}^2 + KJ\sigma_{P r M}^2 + q_{P r}(\psi)$
Locais#Procedências	$(I - 1)(K - 1)$	$\sigma_\varepsilon^2 + J\sigma_{LP r M}^2 + q_{LP r}(\psi)$
Matriz(Procedências)	$I(M - 1)$	$\sigma_\varepsilon^2 + KJ\sigma_{P r M}^2$
L#M(Procedências)	$(K - 1)(I(M - 1))$	$\sigma_\varepsilon^2 + J\sigma_{LP r M}^2$
Resíduo	$K(J - 1)(IM - 1)$	σ_ε^2
Total	$IJKM - 1$	

Optou-se pelo método da máxima verossimilhança restrita -REML (PATTERSON; THOMPSON, 1971), para a estimação dos componentes de variância, e pelo método dos mínimos quadrados generalizados, para a estimação dos parâmetros do vetor de efeitos fixos, utilizando um nível de 5% de significância. Além disso foi utilizada a estrutura de componentes de variância para modelar a matriz \mathbf{G} de efeitos aleatórios e para a matriz residual, $\mathbf{R} = \sigma^2 \mathbf{I}$.

Para avaliar a significância dos fatores de efeito fixo e sua interação, foi utilizado o teste Wald- F , como descrito na seção 2.5.4, cujas estatísticas são dadas por:

(i) Local (L):

$$F_L = \frac{QM_{Local}}{(QM_{B[L]} + QM_{L\sharp M(Pr)} - QM_{Residuo})} \sim F_{(K-1), \nu, \alpha},$$

com ν graus de liberdade no denominador, dados pela aproximação de Satterthwaite (1946), ou seja,

$$\nu = \frac{(QM_{B[L]} + QM_{L\sharp M(Pr)} - QM_{Residuo})^2}{\frac{(QM_{B[L]})^2}{gl_{B[L]}} + \frac{(QM_{L\sharp M(Pr)})^2}{gl_{L\sharp M(Pr)}} - \frac{(QM_{Residuo})^2}{gl_{Residuo}}}.$$

(ii) Procedência (Pr):

$$F_{Pr} = \frac{QM_{Pr}}{QM_{M[P]}} \sim F_{(I-1), I(M-1), \alpha},$$

(iii) Interação Local \times Procedência ($L\sharp Pr$):

$$F_{L\sharp Pr} = \frac{QM_{L\sharp Pr}}{QM_{L\sharp M[P]}} \sim F_{[(I-1)(K-1)], [(K-1)I(M-1)], \alpha}.$$

Por outro lado, para testar as hipóteses referentes aos componentes de variância, $H_0 : \sigma_i^2 = 0$ e $H_1 : \sigma_i^2 > 0$ (existe variabilidade entre os níveis do fator aleatório i), foi utilizado o teste Z de Wald, descrito na seção 2.5.5.

Os diagnósticos foram realizados observando-se os gráficos de dispersão para os resíduos condicionais estudentizados, conforme descrito na seção 2.5.6. A verificação da normalidade tanto para o vetor de parâmetros de efeito aleatório quanto dos resíduos, foi feita utilizando os gráficos de quantil-quantil, e a homocedasticidade por meio do gráfico dos resíduos condicionais estudentizados contra os valores preditos.

A seguir será descrita a metodologia utilizada para estudar a plasticidade fenotípica e a adaptação local do palmitreiro, com base nas variáveis relacionadas com o teor de massa de matéria seca das plantas.

3.2.1.1 Estudo da plasticidade fenotípica

Dado o modelo (21), considere $\mu_{ik} = \mu + \tau_i + \ell_k + \tau\ell_{ik}$, sendo μ_{ik} a média da interação entre a i -ésima procedência e o k -ésimo local. Com a finalidade de testar a hipótese de nulidade de que não há plasticidade fenotípica detectável entre as populações nos K ambientes testados contra a hipótese alternativa de que há plasticidade fenotípica, ou seja, os caracteres medidos na mesma população submetida a ambientes diferentes são significativamente diferentes foi realizado o desdobramento do número de graus de liberdade do fator local dentro dos níveis de procedência, sendo testadas as hipóteses seguintes:

$$\begin{cases} H_0 : \mu_{11} = \mu_{12} = \dots = \mu_{1K} \\ H_1 : \text{pelo menos uma } \mu_{1k} \neq \mu_{1k'} \text{ para } k \neq k' \end{cases},$$

$$\begin{cases} H_0 : \mu_{21} = \mu_{22} = \dots = \mu_{2K} \\ H_1 : \text{pelo menos uma } \mu_{2k} \neq \mu_{2k'} \text{ para } k \neq k' \end{cases},$$

$$\vdots$$

$$\begin{cases} H_0 : \mu_{I1} = \mu_{I2} = \dots = \mu_{IK} \\ H_1 : \text{pelo menos uma } \mu_{Ik} \neq \mu_{Ik'} \text{ para } k \neq k' \end{cases},$$

em que μ_{ik} , é a média de uma variável aleatória em particular, observada em cada i procedência nos k locais, sendo $i = 1, 2, \dots, I$ e $k = 1, 2, \dots, K$.

Adicionalmente, para testar se o próprio local é o mais favorável para o desenvolvimento de plantas da sua mesma procedência, foi construído um conjunto de contrastes, avaliando as hipóteses:

$$\begin{cases} H_0 : \mu_{11} - \frac{\sum_{k=1}^K \mu_{1k}}{K-1} = 0; \ k \neq 1 \\ H_1 : \mu_{11} - \frac{\sum_{k=1}^K \mu_{1k}}{K-1} \neq 0; \ k \neq 1 \end{cases},$$

$$\begin{cases}
H_0 : \mu_{22} - \frac{\sum_{k=1}^K \mu_{2k}}{K-1} = 0; \quad k \neq 2 \\
H_1 : \mu_{22} - \frac{\sum_{k=1}^K \mu_{2k}}{K-1} \neq 0; \quad k \neq 2 \\
\vdots \\
H_0 : \mu_{IK} - \frac{\sum_{k=1}^K \mu_{Ik}}{K-1} = 0; \quad k \neq K \\
H_1 : \mu_{IK} - \frac{\sum_{k=1}^K \mu_{Ik}}{K-1} \neq 0; \quad k \neq K
\end{cases},$$

em que μ_{ik} , são as médias de uma variável aleatória em particular, observadas na procedência i no local k , com $i=k$, comparadas com as médias da mesma procedência nos $K-1$ locais restantes.

3.2.1.2 Estudo da adaptação local

Por outro lado, com a finalidade de testar a hipótese nula de que nenhuma adaptação foi necessária para a ocupação de mais de um bioma por parte de cada espécie contra a hipótese alternativa de que as populações locais são de fato ecótipos e apresentam adaptações específicas, foi realizado o desdobramento do número de graus de liberdade do fator procedências dentro de locais. Neste caso foram testadas as hipóteses:

$$\begin{cases}
H_0 : \mu_{11} = \mu_{21} = \dots = \mu_{I1} \\
H_1 : \text{pelo menos uma } \mu_{i1} \neq \mu_{i'1} \text{ para } i \neq i'
\end{cases},$$

$$\begin{cases}
H_0 : \mu_{12} = \mu_{22} = \dots = \mu_{I2} \\
H_1 : \text{pelo menos uma } \mu_{i2} \neq \mu_{i'2} \text{ para } i \neq i'
\end{cases},$$

$$\vdots$$

$$\begin{cases}
H_0 : \mu_{1K} = \mu_{2K} = \dots = \mu_{IK} \\
H_1 : \text{pelo menos uma } \mu_{iK} \neq \mu_{i'K} \text{ para } i \neq i'
\end{cases},$$

em que μ_{ik} , são as médias de uma variável aleatória em particular, das procedências $i = 1, 2, \dots, I$ num determinado k .

Adicionalmente, para testar se cada procedência está mais adaptada ao seu próprio ambiente em comparação com os outros ambientes onde foi plantada (adaptação local), foi construído um conjunto de contrastes, avaliando as hipóteses:

$$\left\{ \begin{array}{l} H_0 : \mu_{11} - \frac{\sum_{i=1}^I \mu_{i1}}{I-1} = 0; \ i \neq 1 \\ \\ H_1 : \mu_{11} - \frac{\sum_{i=1}^I \mu_{i1}}{I-1} \neq 0; \ i \neq 1 \end{array} \right. ,$$

$$\left\{ \begin{array}{l} H_0 : \mu_{22} - \frac{\sum_{i=1}^I \mu_{i2}}{I-1} = 0; \ i \neq 2 \\ \\ H_1 : \mu_{22} - \frac{\sum_{i=1}^I \mu_{i2}}{I-1} \neq 0; \ i \neq 2 \end{array} \right. ,$$

$$\vdots$$

$$\left\{ \begin{array}{l} H_0 : \mu_{IK} - \frac{\sum_{i=1}^I \mu_{iK}}{I-1} = 0; \ i \neq I \\ \\ H_1 : \mu_{IK} - \frac{\sum_{i=1}^I \mu_{iK}}{I-1} \neq 0; \ i \neq I \end{array} \right. ,$$

em que μ_{ik} , são as médias de uma variável aleatória em particular, observadas nas i procedências no local k , com $i=k$, comparadas com as médias do mesmo local nas $I-1$ procedências restantes.

As análises foram feitas usando-se o procedimento *mixed* implementado no software estatístico SAS ® (SAS INSTITUTE, 2011) , sendo que os programas são apresentados no Anexo A. Para todos os testes empregou-se o nível de 5% de significância.

3.2.2 Modelos lineares generalizados mistos para avaliar a proporção de sementes germinadas

A proporção de sementes germinadas (y_{ijkm}) foi registrada aos 270 dias após a semeadura e analisada usando um MLGM sob a pressuposição de distribuição binomial da variável em estudo. Usando a função de ligação logito, a média condicional da proporção de plantas germinadas, $E[y_{ijkm}|b_{j(k)}, \delta_{m(i)}, \ell\delta_{m(ik)}]$, pode ser expressa em termos do preditor linear η_{ijkm} , com a inclusão dos efeitos aleatórios (para modelar a variação extra-binomial), como:

$$\eta_{ijkm} = \log \left(\frac{\pi_{ijkm}}{1 - \pi_{ijkm}} \right) = \mu + \tau_i + b_{j(k)} + \ell_k + \tau\ell_{ik} + \delta_{m(i)} + \ell\delta_{m(ik)},$$

em que $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$, $m = 1, \dots, M$,

π_{ijkm} é a probabilidade de uma semente da planta mãe m e procedência i ter germinado na repetição j no local k , μ é a média geral, τ_i é o efeito da i -ésima procedência (efeito fixo), $b_{j(k)}$ é o efeito aleatório do j -ésimo bloco no k -ésimo local, $b_{j(k)} \sim N(0, \sigma_b^2)$, ℓ_k é o efeito fixo do k -ésimo local; $\tau\ell_{ik}$ representa o efeito da interação entre a i -ésima procedência e o k -ésimo local, $\delta_{m(i)}$ é o efeito aleatório da m -ésima planta-matriz dentro da i -ésima procedência, $\delta_{m(i)} \sim N(0, \sigma_\delta^2)$, $\ell\delta_{m(ik)}$ é o efeito aleatório da m -ésima planta-matriz dentro da ik -ésima interação local \times procedência, $\ell\delta_{m(ik)} \sim N(0, \sigma_{\ell\delta}^2)$. Considerou-se $b_{j(k)}, \delta_{m(i)}, \ell\delta_{m(ik)}$ como independentes.

As estimativas dos parâmetros de efeito fixo e a predição dos parâmetros de efeito aleatório foram obtidas a partir da maximização do logaritmo da função de pseudo-verossimilhança (baseado no método REML em um modelo linearizado), utilizando a técnica de otimização Quase-Newton duplamente iterativo.

O valor ajustado da proporção de sementes germinadas é dado por: $\hat{\pi}_i = \frac{\exp(\hat{\eta})}{1 + \exp(\hat{\eta})}$.

Como medida do ajuste do modelo foi utilizada a estatística X^2 generalizada de Pearson, que tem a forma: $X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{V}(\hat{\mu}_i)}$, sendo y_i os valores observados, $\hat{\mu}_i$ os valores estimados pelo modelo e $\hat{V}(\hat{\mu}_i)$ a função de variância estimada para a distribuição em estudo. No caso da distribuição binomial, $B(n, \pi)$ tem-se que $V(\mu) = \mu \left(1 - \frac{\mu}{n}\right)$.

Para estimar o parâmetro de escala ou dispersão ϕ , foi utilizada a expressão: $\phi^* = \frac{X^2}{n - m}$, em que n é o tamanho da amostra e m é o número de parâmetros estima-

dos. Para dados provenientes das distribuições binomial e Poisson, tem-se $\phi=1$, portanto valores distantes de um, são um indicativo de falta de ajuste do modelo.

Para avaliar a significância dos fatores de efeito fixo e suas interações, foi utilizado o teste Wald- F , como descrito na seção 2.7.3. Por outro lado, para testar as hipóteses referentes aos componentes de variância da matriz \mathbf{G} foi utilizado o teste da razão de verossimilhanças.

Adicionalmente, foram calculados resíduos condicionais estudentizados (r_e) na escala logito, definidos pela expressão seguinte: $r_e = \frac{p - \hat{\eta}}{\widehat{Var}[p|u]}$, em que p denota os valores dos "pseudo-dados" que surgem a partir do algoritmo de linearização usado para obter o estimador de máxima verossimilhança, $\widehat{Var}[p|\mathbf{u}]$ é a variância estimada de p , condicional aos efeitos aleatórios contidos no vetor \mathbf{u} (GBUR et al., 2012), e construídos gráficos (histograma, *box plot*, quantil-quantil e de r_e contra o preditor linear), para verificar o ajuste do MLGM.

Posteriormente, com o objetivo de testar a hipótese de nulidade de que não há plasticidade fenotípica detectável entre as populações nos ambientes testados contra a hipótese alternativa de que há plasticidade fenotípica, ou seja, as proporções de sementes germinadas da mesma população submetidas a ambientes diferentes são significativamente diferentes foi realizado o desdobramento do número de graus de liberdade do fator locais dentro de procedências.

Por outro lado, com a finalidade de testar a hipótese nula de que nenhuma adaptação foi necessária para a ocupação de mais de um bioma por parte de cada procedência contra a hipótese alternativa de que as populações locais são de fato ecótipos e apresentam adaptações específicas, foi realizado o desdobramento do número de graus de liberdade do fator procedências dentro de locais.

As análises foram feitas usando-se o procedimento *glimmix* implementado no software estatístico SAS (SAS INSTITUTE, 2011), sendo que os programas são apresentados no Anexo B. Para todos os testes empregou-se o nível de 5% de significância.

4 RESULTADOS E DISCUSSÃO

Na seção 4.1 serão apresentados os resultados e discussões das análises estatísticas referentes à avaliação da plasticidade fenotípica e da adaptação local dos palmiteiros, para os dois grupos de experimentos, considerando como variáveis de resposta, a massa de matéria seca de diferentes partes das plantas e total, ajustando e avaliando a adequação de um modelo linear de efeitos mistos. Posteriormente, na seção 4.2, serão exibidos os resultados e discussões das análises estatísticas, considerando desta vez, as proporções de sementes germinadas apenas no experimento de adaptação no estabelecimento, ajustando e avaliando a adequação de um modelo linear generalizado de efeitos mistos.

4.1 Avaliação da plasticidade fenotípica e da adaptação local com base nos teores de massa de matéria seca das plantas em ensaios de estabelecimento e de crescimento

Os valores médios e os respectivos desvios padrões para as variáveis consideradas nos diferentes ensaios, por local e procedência dos palmiteiros, estão apresentados nas Tabelas 2 e 3.

Tabela 2 – Médias e desvios padrões (d.p.) da massa de matéria seca (mg) da folha, do caule e parte aérea das plantas das procedências Ombrófila (Omb.), Semidecidual (Sem.) e Restinga (Res.), nos locais Carlos Botelho (CB), Caetetus (CT) e Ilha do Cardoso (IC), no ensaio de adaptação no estabelecimento

Local	Proc.	Folha		Caule		Parte aérea	
		Média	d.p.	Média	d.p.	Média	d.p.
CB	Omb.	0,1027	0,0256	0,0768	0,0300	0,1794	0,0527
	Sem.	0,0682	0,0267	0,0544	0,0224	0,1226	0,0455
	Res.	0,0715	0,0295	0,0513	0,0228	0,1227	0,0495
CT	Omb.	0,0735	0,0328	0,0571	0,0206	0,1306	0,0508
	Sem.	0,0588	0,0294	0,0449	0,0229	0,1036	0,0502
	Res.	0,0578	0,0240	0,0411	0,0170	0,0988	0,0381
IC	Omb.	0,1307	0,0356	0,1215	0,0352	0,2521	0,0637
	Sem.	0,0949	0,0396	0,0877	0,0349	0,1826	0,0689
	Res.	0,0943	0,0359	0,0835	0,0350	0,1777	0,0669

A Tabela 2 permitiu verificar que no ensaio de adaptação no estabelecimento, a procedência Ombrófila apresentou os maiores valores médios da massa de matéria seca das partes das plantas, em todos os locais avaliados. Com relação aos locais, no parque Ilha do Cardoso apresentaram-se os melhores resultados. Além disso, as plantas das diferentes procedências apresentaram um comportamento diferenciado ao serem semeadas nos distintos locais, o que dá indicio da existência de plasticidade fenotípica.

Tabela 3 – Médias e desvios padrões (d.p.) da massa de matéria seca (mg) da folha, do caule, da raiz e total das plantas das procedências Ombrófila (Omb.), Semi-decidual (Sem.) e Restinga (Res.), nos locais Carlos Botelho (CB), Caetetus (CT) e Ilha do Cardoso (IC), no ensaio de adaptação em juvenis

Local	Proc.	Folha		Caule		Raiz		Total	
		Média	d.p.	Média	d.p.	Média	d.p.	Média	d.p.
CB	Omb.	0,1870	0,1062	0,1912	0,1055	0,1455	0,0633	0,5236	0,2402
	Sem.	0,1595	0,1023	0,1847	0,1044	0,1679	0,0802	0,5121	0,2567
	Res.	0,2288	0,1306	0,2189	0,1233	0,1654	0,0828	0,6131	0,3077
CT	Omb.	0,2468	0,1048	0,2216	0,1016	0,1621	0,0797	0,6306	0,2519
	Sem.	0,2329	0,1133	0,2077	0,1083	0,1787	0,0847	0,6194	0,2898
	Res.	0,2861	0,1106	0,2645	0,1160	0,1934	0,0704	0,7440	0,2635
IC	Omb.	0,1895	0,1090	0,2091	0,0883	0,1346	0,0510	0,5332	0,2176
	Sem.	0,1711	0,1250	0,2169	0,1264	0,1608	0,0828	0,5488	0,3161
	Res.	0,2801	0,1463	0,3093	0,1823	0,1870	0,1113	0,7764	0,4097

Com relação ao ensaio de adaptação em juvenis, a Tabela 3 permitiu verificar que não existe um padrão claro do comportamento da massa da matéria seca das plantas. Nenhum dos locais se mostrou evidentemente superior. Além disso nenhuma procedência se destacou entre as demais.

Em seguida, foram verificadas as significâncias dos parâmetros de efeito fixo, e as estatísticas para o teste Wald- F são apresentadas na Tabela 4, salientando que o nível de significância utilizado foi de 5%. A referida Tabela inclui os resultados obtidos nos ensaios de adaptação no estabelecimento e em juvenis, para a massa de matéria seca de diferentes partes das plantas de palmitreiro.

Observa-se na Tabela 4 diferenças significativas para os efeitos principais e a interação local#procedência, no ensaio de adaptação no estabelecimento. Os palmitreiros apresentaram maiores diferenças neste ensaio porque foi conduzido desde a semeadura, sendo afetado pelas condições ambientais dos locais. Por outro lado, no ensaio de adaptação em juvenis o único efeito significativo foi a procedência, nas variáveis massa de matéria seca da folha, do caule, da parte aérea e total.

Tabela 4 – Estatística F e significância para as fontes de variação fixas consideradas nos ensaios de adaptação no estabelecimento e em juvenis

Massa de matéria seca (MS)	F para fontes de variação fixas		
	Local	Procedência	Local \sharp Procedência
Ensaio de adaptação no estabelecimento			
Folha	29,46***	12,40**	2,24*
Caule	61,04***	15,95***	2,85*
Parte aérea	48,37**	14,50**	2,88*
Ensaio de adaptação em juvenis			
Folha	3,16	8,20**	0,86
Caule	1,80	4,36*	1,04
Raiz	0,75	1,29	0,76
Total	1,87	4,14*	0,97

(*) valor- $p < 0,05$; (**) valor- $p < 0,01$; (***) valor- $p < 0,001$

Com relação aos componentes de variância, na Tabela 5 são apresentadas as suas estimativas e os valores da estatística Wald- Z para testar suas significâncias. O componente de variância relacionado com o fator aleatório produzido pela m -ésima planta matriz dentro da ik -ésima interação local \sharp procedência, $(\ell\delta_{m(ik)})$, apresentou uma contribuição significativa apenas no caso da análise da massa de matéria seca do caule e do total, no ensaio de adaptação em juvenis. Os demais componentes de variâncias proporcionaram uma contribuição significativa em todos os casos.

Tabela 5 – Estimativas dos componentes de variância e valores da estatística Wald- Z para testar sua significância

Matéria seca	σ_b^2	Z	σ_δ^2	Z	$\sigma_{\ell\delta}^2$	Z	σ_ϵ^2	Z
Ensaio de adaptação no estabelecimento								
Folha	0,00012	2,59**	0,00017	2,80**	0,00001	0,51	0,00074	14,63***
Caule	0,00007	2,24**	0,00009	2,48**	0,00001	0,29	0,00066	14,63***
Parte aérea	0,00032	2,47**	0,00051	2,82**	0,00003	0,32	0,00235	14,63***
Ensaio de adaptação em juvenis								
Folha	0,00271	2,66**	0,00110	1,74*	0,00078	1,27	0,00976	12,34***
Caule	0,00338	2,78**	0,00170	2,07*	0,00115	1,88*	0,00903	12,43***
Raiz	0,00177	2,97**	0,00140	2,93**	0,00015	0,79	0,00342	12,43***
Total	0,02198	2,89**	0,01250	2,39**	0,00584	1,76*	0,04768	12,39***

(*) valor de $p < 0,05$; (**) valor de $p < 0,01$; (***) valor de $p < 0,001$.

As porcentagens da variância total atribuídas a cada componente de variância, por ensaio e por variável analisada são apresentadas na Tabela 6. Observa-se que a variância residual (σ_ϵ^2) representa a maior proporção da variância total dos dados. Por outro lado a variância devida ao componente ($\sigma_{\ell\delta}^2$) corresponde à menor contribuição para a variância total dos dados. Vale destacar que o componente de variância referente a plantas matrizes dentro de procedência (σ_δ^2) representa maior porcentagem da variabilidade total explicada em relação ao componentes de variância de blocos dentro de locais (σ_b^2) para o ensaio de estabelecimento, caso contrário foi verificado no ensaio de adaptação em juvenis.

Nas Figuras 5 e 6 são apresentados os gráficos dos valores preditos contra os resíduos condicionais estudentizados do modelo linear misto, para cada variável analisada, nos ensaios de adaptação no estabelecimento (semeadura) e de adaptação em juvenis (crescimento), respectivamente. Nota-se que a maioria dos resíduos condicionais estudentizados encontram-se no intervalo (-3,3), com distribuição aleatória em torno da média zero, sem apresentar qualquer tendência, satisfazendo a condição de independência, confirmando que o modelo utilizado é adequado para descrever o comportamento dos dados. Além disso, nota-se a presença de alguns pontos atípicos.

Tabela 6 – Porcentagem da variância total atribuída a cada componente de variância, por experimento e variável analisada

Massa de matéria seca	σ_b^2	σ_δ^2	$\sigma_{\ell\delta}^2$	σ_ϵ^2
Ensaio de adaptação no estabelecimento				
Folha	11,61	15,93	1,34	71,11
Caule	7,88	11,16	0,79	80,17
Parte aérea	9,96	15,96	0,81	73,27
Ensaio de adaptação em juvenis				
Folha	18,86	7,66	5,44	68,04
Caule	22,14	11,12	7,55	59,19
Raiz	26,25	20,83	2,15	50,77
Total	24,98	14,21	6,63	54,18

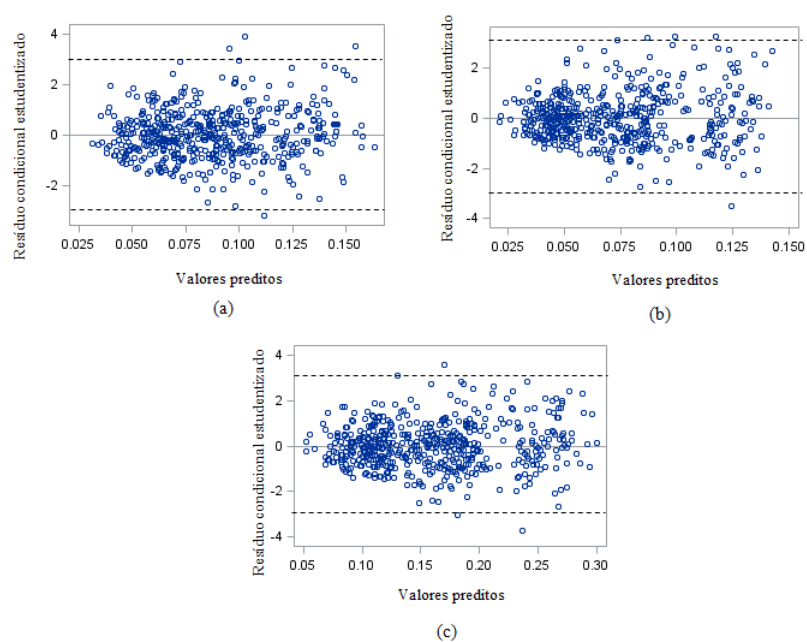


Figura 5 – Gráficos dos resíduos condicionais estudentizados em função dos valores preditos, para o experimento de adaptação no estabelecimento, para as variáveis: (a) massa seca da folha, (b) massa seca do caule e (c) massa seca da parte aérea

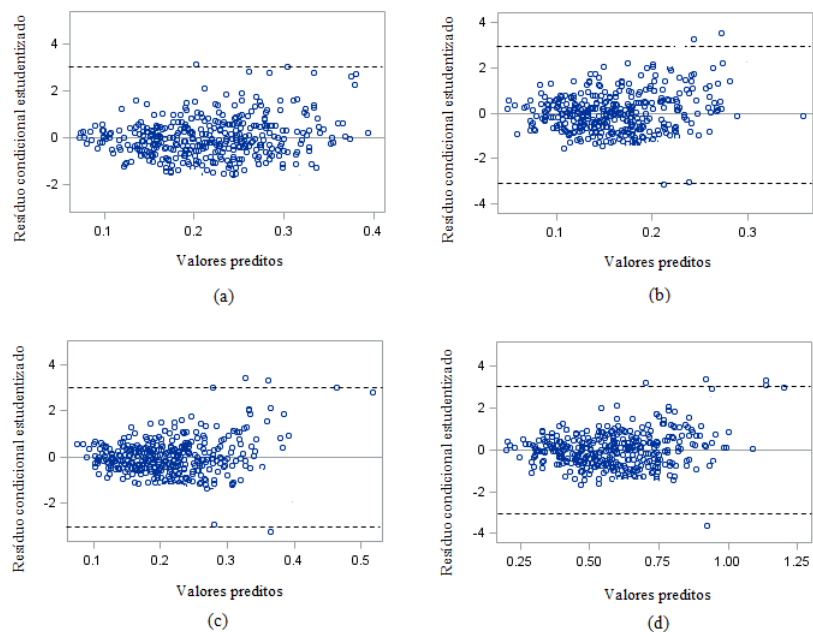


Figura 6 – Gráficos dos resíduos condicionais estudentizados em função dos valores preditos, para o experimento de adaptação em juvenis, para as variáveis: (a) massa seca da folha, (b) massa seca do caule, (c) massa seca da raiz e (d) massa seca total

Nas Figuras 7 e 8 são apresentados os gráficos quantil-quantil dos resíduos condicionais estudentizados contra os quantis teóricos da distribuição normal, relativos aos ensaios de adaptação no estabelecimento e de adaptação em juvenis, para cada uma das variáveis. Pode-se notar que em todos os casos é satisfeita a pressuposição de normalidade, porém, observam-se alguns pontos atípicos.

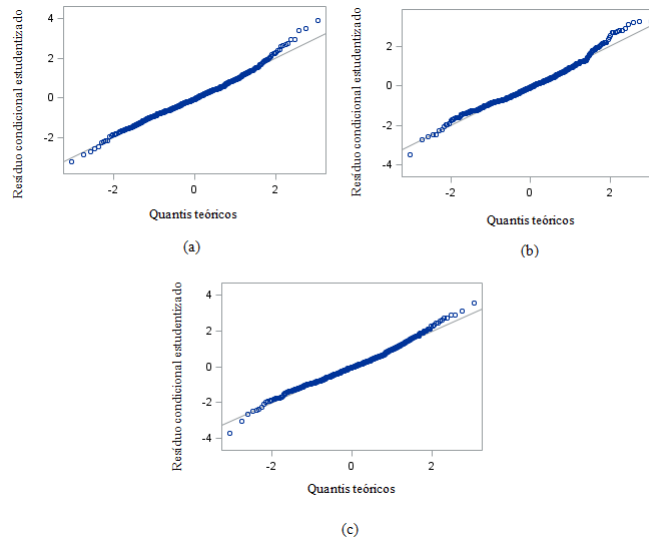


Figura 7 – Gráficos de quantil-quantil para os resíduos condicionais estudentizados no experimento de adaptação no estabelecimento, para as variáveis: (a) massa seca da folha, (b) massa seca do caule e (c) massa seca da parte aérea

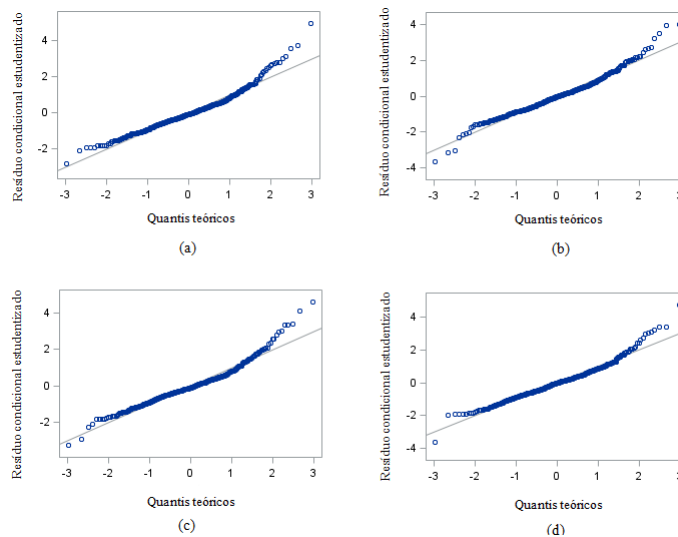


Figura 8 – Gráficos de quantil-quantil para os resíduos condicionais estudentizados para o experimento de adaptação em juvenis, para as variáveis: (a) massa seca da folha, (b) massa seca do caule, (c) massa seca da raiz e (d) massa seca total

4.1.1 Plasticidade fenotípica

Para estudar a presença de plasticidade fenotípica nos palmiteiros foi realizado o desdobramento do número de graus de liberdade de locais dentro dos níveis do fator procedência. Na Tabela 7, é apresentado o resumo dos resultados deste desdobramento.

Tabela 7 – Estatística F e significância para o desdobramento do número de graus de liberdade de local \sharp procedência para as variáveis relacionadas com a massa de matéria seca, analisando os dados de locais dentro de cada nível de procedência

Massa de matéria seca (MS)	Procedências		
	Ombrófila	Semidecidual	Restinga
Ensaio de adaptação no estabelecimento			
Folha	29,91***	14,18**	14,06***
Caule	57,95***	29,79***	22,52***
Parte área	48,22***	24,21***	20,86***
Ensaio de adaptação em juvenis			
Folha	1,93	3,32*	2,10
Caule	0,62	0,94	3,20
Raiz	0,95	0,42	0,69
Total	1,20	1,42	2,17

(*) valor de $p < 0,05$; (**) valor de $p < 0,01$; (***) valor de $p < 0,001$

Valores significativos da estatística F evidenciam a presença de plasticidade fenotípica, ou seja, plantas procedentes de um bioma determinado apresentam diferenças quanto ao comportamento de uma determinada variável medida, dependendo do local onde foram semeadas ou transplantadas. No ensaio de adaptação no estabelecimento (semeadura) apresentou-se a característica de plasticidade em todas as variáveis analisadas. Por outro lado, no ensaio de adaptação em juvenis, as plantas têm um comportamento não plástico (ou robusto), exceto para a variável massa de matéria seca da folha.

O resumo da análise de contrastes é apresentado na Tabela 8, para verificar se um local determinado é mais favorável para a produção de massa de matéria seca (dos caracteres medidos) de plantas originadas de sementes da sua mesma procedência, em cada experimento realizado.

Observa-se na Tabela 8, que plantas da proveniência Restinga tiveram melhores produções médias de massa de matéria seca da folha, do caule e da parte aérea (no ensaio de adaptação no estabelecimento), e do caule (no ensaio de adaptação em juvenis), quando foram semeadas ou transplantadas no seu próprio lugar de origem (Ilha do Cardoso).

Tabela 8 – Estimativas (Est.) dos contrastes entre locais (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) da mesma procedência contra outros locais, estatística t e significância

Massa de matéria seca (MS)	Ombrófila		Semidecidual		Restinga	
	CB vs outros		CT vs outros		IC vs outros	
	Est.	t	Est.	t	Est.	t
Ensaio de adaptação no estabelecimento						
Folha	-0,0016	-0,25	-0,0241	-3,87*	0,0329	4,68***
Caule	-0,0117	-2,12*	-0,0263	-5,07***	0,0379	6,31**
Parte aérea	-0,0101	-0,90	-0,0503	-4,79*	0,0707	5,90**
Ensaio de adaptação em juvenis						
Folha	-0,0418	-1,39	0,0796	2,50*	0,0228	0,74
Caule	-0,0345	-1,06	0,0210	0,61	0,0669	2,03*
Raiz	-0,0128	-0,61	0,0218	0,98	0,0145	0,68
Total	-0,0928	-1,17	0,1261	1,51	0,1041	1,29

(*) valor de $p < 0,05$; (**) valor de $p < 0,01$; (***) valor de $p < 0,001$

Com relação as plantas da proveniência Semidecidual, tiveram melhores produções médias de massa de matéria seca da folha quando transplantadas no seu próprio local (Caetetus), no ensaio de adaptação em juvenis. Situação contrária aconteceu, para as produções médias de massa de matéria seca obtidas no ensaio de adaptação no estabelecimento, onde as plantas da proveniência Semidecidual tiveram o pior desempenho no seu próprio local de origem.

Por outro lado, as plantas da proveniência Ombrófila tiveram as piores produções médias de massa de matéria de seca (principalmente no caso da MS do caule) quando foram semeadas ou transplantadas no seu próprio local (Carlos Botelho).

Nas Figuras 9 e 10 é apresentado o comportamento das médias de massa de matéria seca (mg) dos caracteres medidos, por local dentro de cada procedência, nos ensaios de adaptação no estabelecimento e adaptação em juvenis, respectivamente. De forma geral, nota-se que a característica de plasticidade fenotípica é evidente no ensaio de adaptação no estabelecimento (semeadura) para todos os caracteres avaliados.

4.1.2 Adaptação local

O resumo do teste F para o desdobramento do número de graus de liberdade de procedências dentro de locais, para estudar a presença de adaptação local é apresentado na Tabela 9.

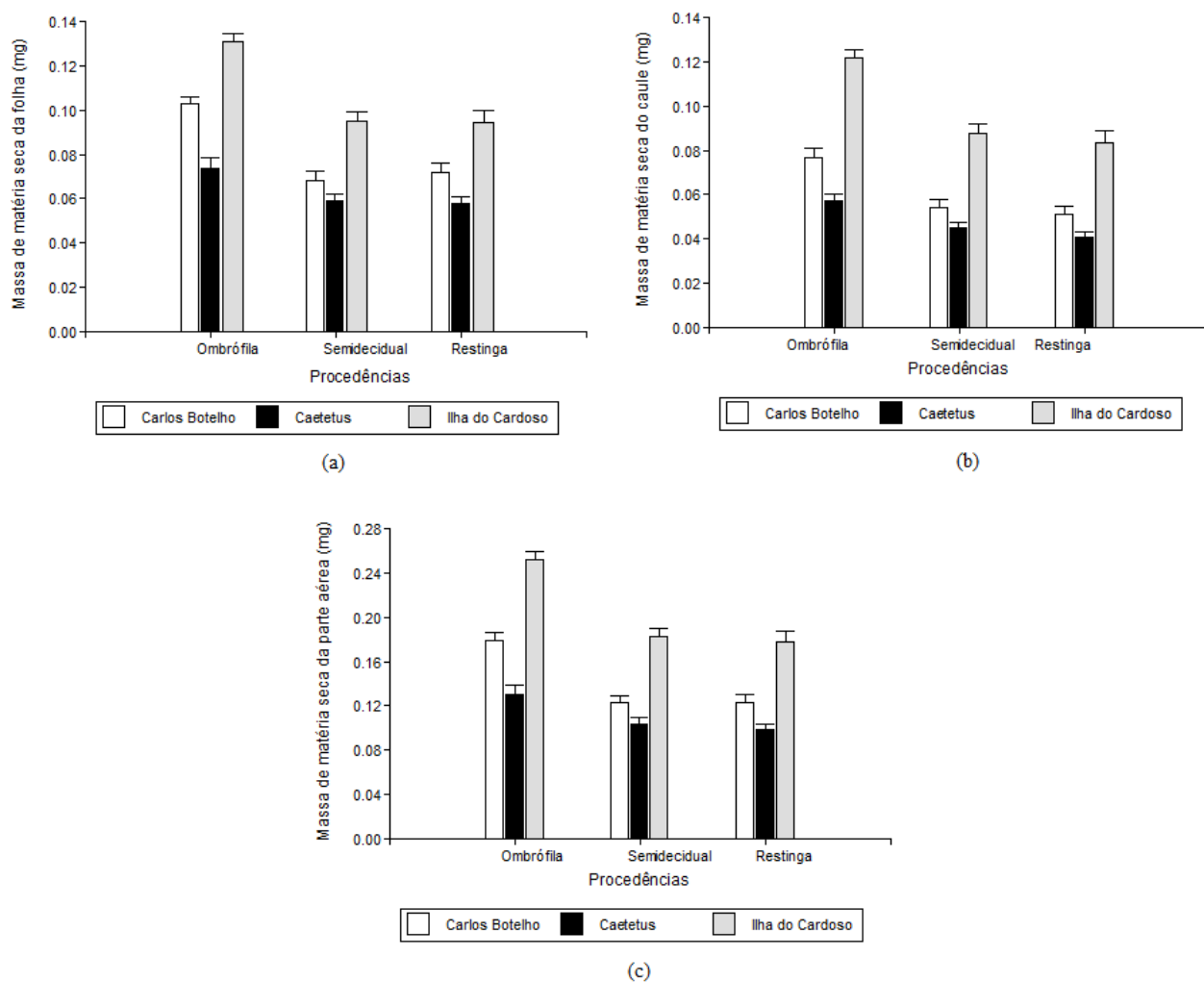


Figura 9 – Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule e (c) da parte aérea, por local dentro de cada procedência, no ensaio de adaptação no estabelecimento. As barras representam as médias originais e as linhas verticais o erro padrão da média.

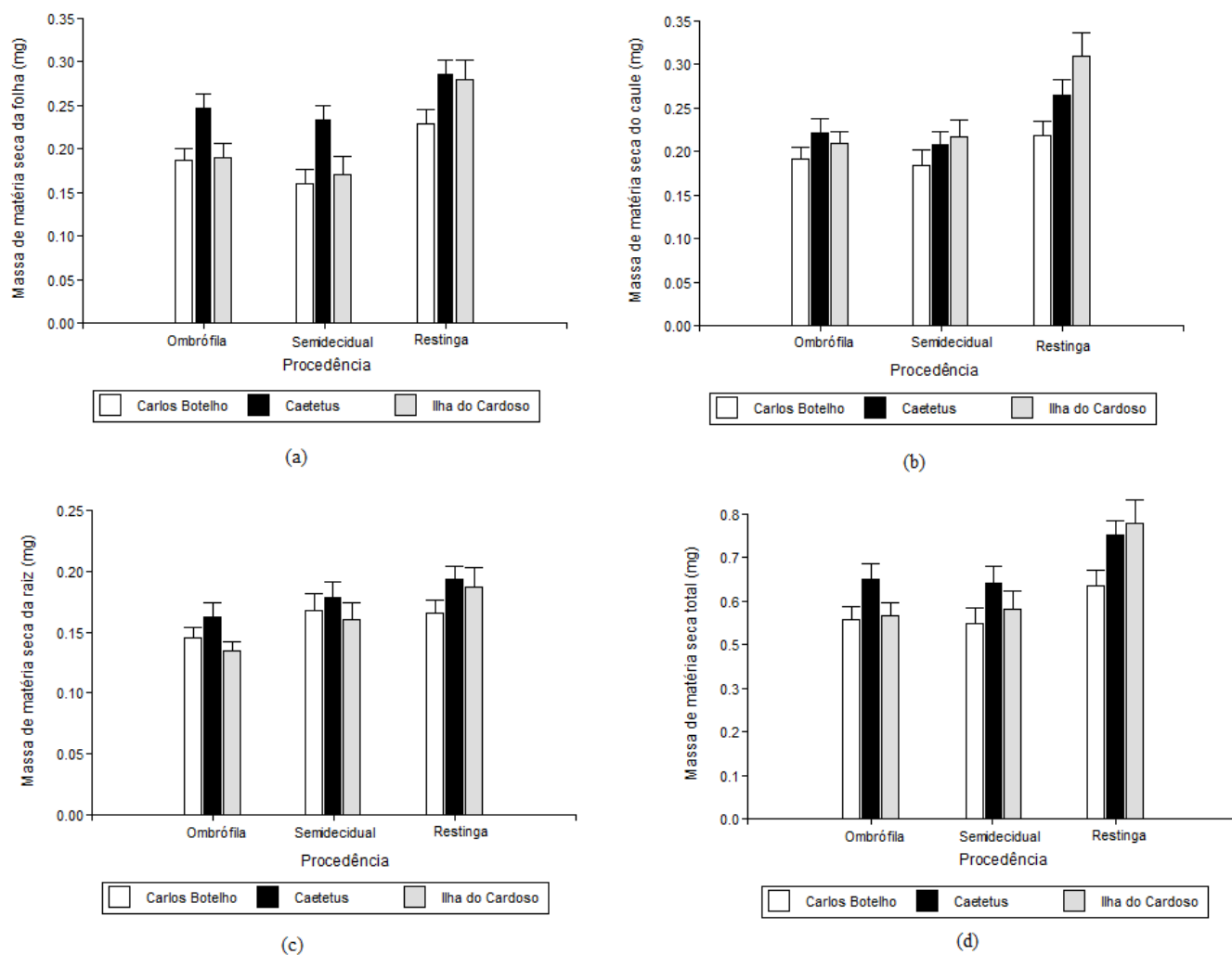


Figura 10 – Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule, (c) da raiz e (d) total, por local dentro de cada procedência, no ensaio de adaptação em juvenis. As barras representam as médias originais e as linhas verticais o erro padrão da média.

Tabela 9 – Estatística F e significância para o desdobramento do número de graus de liberdade de local \sharp procedência para as variáveis relacionadas com a massa de matéria seca, analisando procedências dentro de cada nível de local

Massa de matéria seca (MS)	Locais		
	Carlos Botelho	Caetetus	Ilha do Cardoso
Ensaio de adaptação no estabelecimento			
Folha	10,80***	2,97*	14,90***
Caule	8,60***	3,68*	22,37***
Parte aérea	10,76***	3,66*	19,55***
Ensaio de adaptação em juvenis			
Folha	3,81	1,89	7,93*
Caule	1,16	1,67	5,80*
Raiz	0,50	0,28	2,16
Total	1,61	1,37	5,70**

(*) valor de $p < 0,05$; (**) valor de $p < 0,01$; (***) valor de $p < 0,001$

Pode-se observar que no ensaio de adaptação no estabelecimento se apresenta de forma evidente a característica de adaptação, ou seja, que em cada local avaliado, as plantas das diferentes procedências apresentam um comportamento diferenciado em todos os caracteres avaliados, podendo em alguns casos tratar-se de adaptação local.

Por outro lado, no ensaio de adaptação em juvenis, apenas as plantas transplantadas no bioma Restinga (Parque Estadual Ilha do Cardoso) apresentam evidência de adaptação para as variáveis massa de matéria seca da folha, do caule e total. Brancalion (2009), salienta que o lento crescimento das plantas de *E. edulis*, principalmente em condições naturais de interior de floresta, somado ao período restritivo de duração do experimento (10 meses), podem ter limitado a possibilidade de identificação de adaptações locais para essa espécie nos outros biomas.

O resumo da análise de contrastes para verificar se plantas originadas de sementes de uma determinada procedência estão mais adaptadas ao seu próprio local, em comparação com plantas procedentes de sementes de outros locais, com relação à produção de massa de matéria seca em cada experimento realizado é apresentado na Tabela 10.

Neste caso conclui-se que, plantas originadas de sementes da formação florestal Restinga estão significativamente mais adaptadas ao seu próprio local (Ilha do Cardoso), em comparação com plantas procedentes de sementes de outros locais, ou seja, apresentam maiores valores médios de massa de matéria seca da folha, do caule e total, no ensaio de adaptação em juvenis, sugerindo, de acordo com Brancalion (2009) a existência de um ecótipo para essa formação florestal.

Tabela 10 – Estimativas (Est.) dos contrastes entre procedências do mesmo local (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) contra outras procedências, estatística t e significância

Massa de matéria seca (MS)	Local CB		Local CT		Local IC	
	Omb. vs outras		Sem. vs outras		Res. vs outras	
	Est.	t	Est.	t	Est.	t
Ensaio de adaptação no estabelecimento						
Folha	0,0325	4,64***	-0,0057	-0,84	-0,0163	-2,29*
Caule	0,0243	4,14***	-0,0040	-0,71	-0,0204	-3,40**
Parte aérea	0,0569	4,64***	-0,0098	-0,82	-0,0367	-2,94**
Ensaio de adaptação em juvenis						
Folha	-0,0011	-0,05	-0,0362	-1,46	0,0957	3,84***
Caule	-0,0022	-0,08	-0,0371	-1,37	0,0928	3,41**
Raiz	-0,0152	-0,83	-0,0017	-0,09	0,0371	1,98
Total	-0,0161	-0,25	-0,0720	-1,09	0,2255	3,39***

(*) valor de $p < 0,05$; (**) valor de $p < 0,01$; (***) valor de $p < 0,001$

Caso contrário acontece no ensaio de adaptação no estabelecimento, onde parecem estar menos adaptadas ao seu próprio local. Sobre esta situação, Brancalion (2009) comenta que a ausência de adaptações locais pode estar relacionada ao fato das condições para o estabelecimento de plantas nessa formação florestal terem sido as mais favoráveis dentre as áreas de estudo, com valores altos de germinação e de massa de matéria seca de plantas. Tal fato permitiu que essa formação florestal tivesse elevada densidade de plantas, de forma que a menor pressão de seleção para o estabelecimento pode não ter estimulado o surgimento de adaptações locais específicas para essa fase do ciclo de vida vegetal.

As plantas originadas de sementes da formação florestal Ombrófila Densa estão mais adaptadas ao seu próprio local (Carlos Botelho), apresentando maiores valores médios de massa de matéria seca da folha, do caule e da parte aérea, no caso do ensaio de adaptação no estabelecimento (semeadura).

Nas Figuras 11 e 12 é apresentado o comportamento da massa de matéria seca (mg) dos caracteres medidos, por procedência dentro de cada local, nos ensaios de adaptação no estabelecimento e adaptação em juvenis, respectivamente. De forma geral, nota-se que a característica de adaptação é evidente no ensaio de adaptação no estabelecimento (semeadura) para todos os caracteres avaliados.

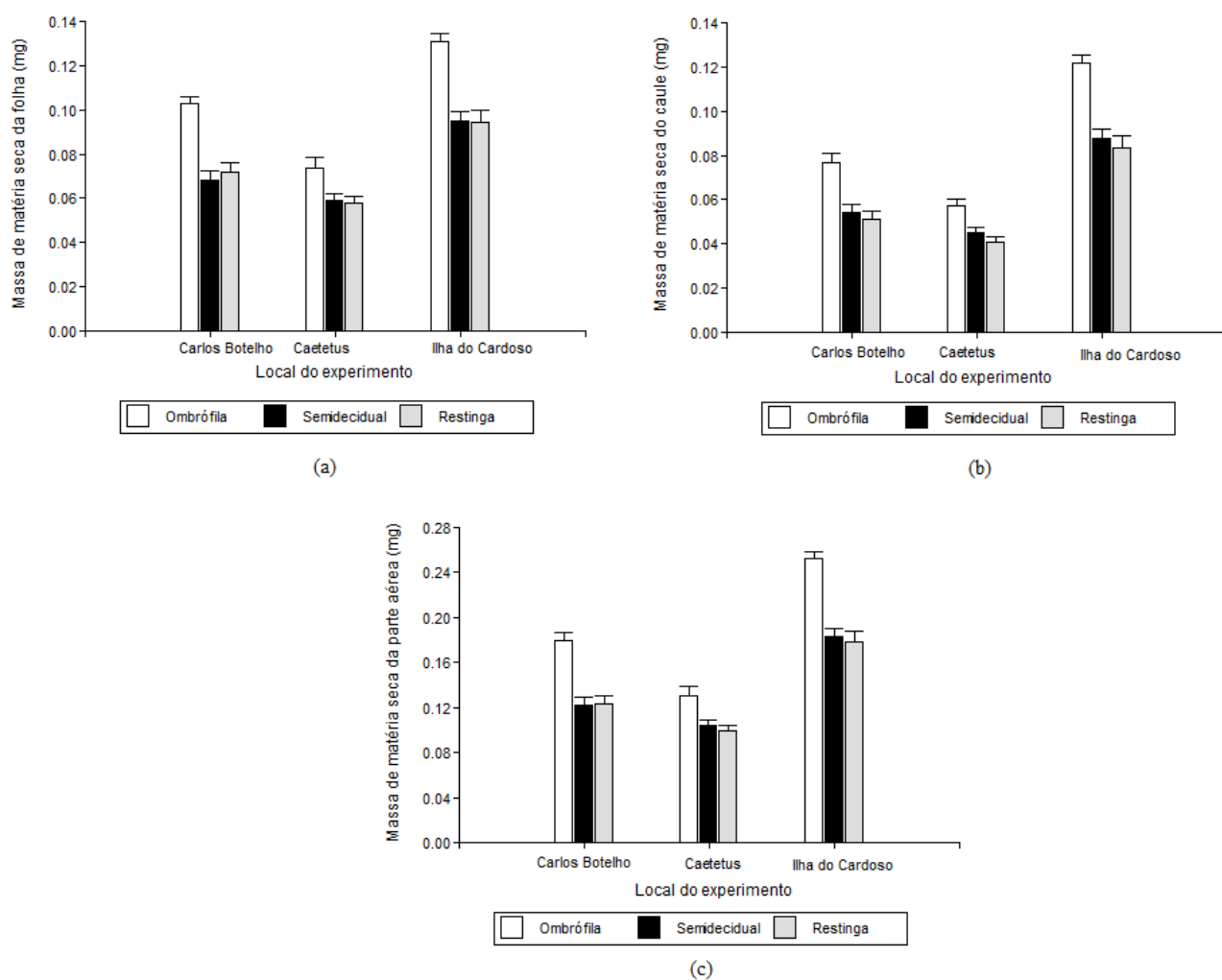


Figura 11 – Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule e (c) da parte aérea, por procedência dentro de cada local, no ensaio de adaptação no estabelecimento. As barras representam as médias originais e as linhas verticais o erro padrão da média.

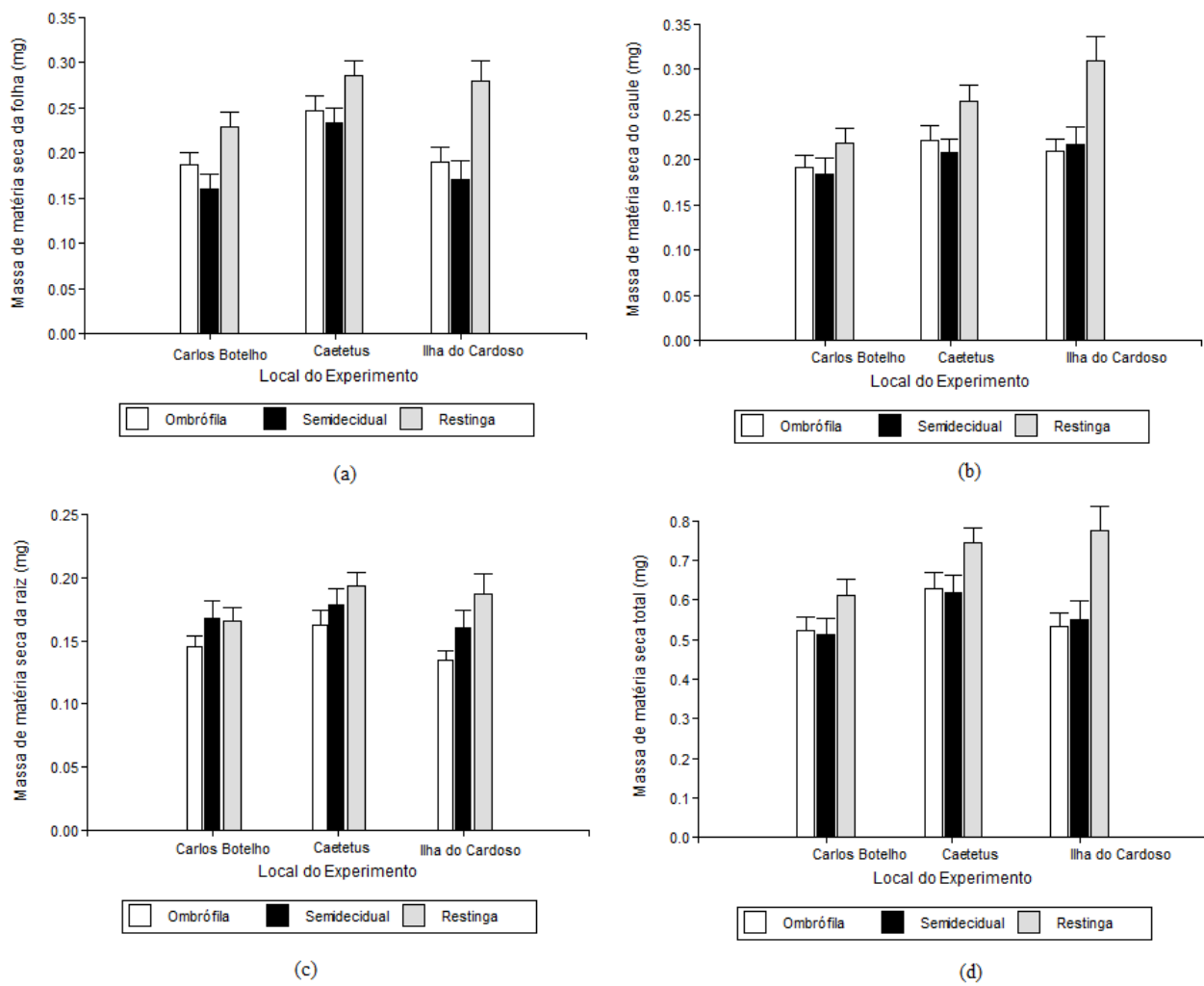


Figura 12 – Comportamento do teor médio de massa de matéria seca: (a) da folha, (b) do caule, (c) da raiz e (d) total, por procedência dentro de cada local, no ensaio de adaptação em juvenis. As barras representam as médias originais e as linhas verticais o erro padrão da média.

4.2 Avaliação da plasticidade fenotípica e adaptação local com base na proporção de sementes germinadas em ensaios de estabelecimento

A instalação e manutenção de plantações sustentáveis de *E. edulis*, em condições naturais, exigem estudos relacionados aos aspectos silviculturais do palmitreiro, como os relacionados à germinação de sementes, porque o vigor germinativo é qualitativamente essencial na produção de mudas. Frequentemente, determinadas procedências podem apresentar maiores taxas de germinação de sementes ao serem semeadas em diferentes locais. Provavelmente, isso propiciaria maior possibilidade de ocupação e aproveitamento mais rápido do substrato em condições naturais, favorecendo o estabelecimento dessas procedências, permitindo sua conservação e manejo sustentável.

Desse modo, um dos objetivos deste estudo foi avaliar o desempenho germinativo de sementes de três procedências de palmitreiro em três locais. As proporções médias observadas de sementes germinadas por local e procedência estão apresentadas na Tabela 11.

Tabela 11 – Proporções médias observadas de sementes germinadas de *E. edulis* segundo a procedência e local de instalação do experimento.

Local	Procedências			Média
	Ombrófila	Semidecidual	Restinga	
Carlos Botelho	0,7833	0,7875	0,7083	0,7597
Caetetus	0,2148	0,5037	0,3333	0,3506
Ilha do Cardoso	0,8567	0,7433	0,7267	0,7756
Média	0,6210	0,6765	0,5901	0,6292

Observa-se na Tabela 11, que o comportamento de *E. edulis* quanto à capacidade germinativa variou principalmente de acordo com o local onde foi realizado o experimento, sendo os locais Carlos Botelho e Ilha do Cardoso, onde se registraram os maiores valores de proporção média de sementes germinadas.

Na Tabela 12 são apresentadas as estimativas dos componentes de variância e seus respectivos erros padrões, os valores obtidos da estatística χ^2 e respectivos níveis descritivos (valores de p) para testar a hipótese nula de que cada componente de variância é igual a zero.

Tabela 12 – Estimativas e erros padrões dos componentes de variância, estatística χ^2 e níveis descritivos para o teste da razão de verossimilhanças

Componente de variância	Estimativa	Erro padrão	χ^2	Valor de p (*)
Bloco (Local)	0,3610	0,1249	82,11	< 0,0001
Matriz (Procedência)	0,0998	0,0487	82,11	< 0,0001
Local \sharp Matriz (Procedência)	0,0129	0,0424	7,23	0,0036

(*) Valor de p baseado em uma mistura de distribuições χ^2

Nota-se ainda na Tabela 12 que os três componentes de variância são significativos. Estes resultados mostram que estes componentes de variância no modelo são necessários para descrever adequadamente a variação dos dados. Por outro lado, aproximadamente 76% da variância total (em escala logito) pode ser atribuída ao efeito de bloco dentro de local, 21% ao efeito de planta matriz dentro de procedência e somente 3% ao efeito da interação local \sharp planta matriz dentro de procedência.

Os valores da estatística F e respectivos valores descritivos (valores de p) para o teste dos efeitos fixos na análise da proporção de sementes germinadas, são apresentados na Tabela 13.

Tabela 13 – Estatística F e respectivos níveis descritivos (valores de p) para as fontes de variação fixas

Efeito	F	Valor de p
Local	28,36	< 0,0001
Procedência	2,85	0,0754
Local \sharp Procedência	13,67	< 0,0001

Os resultados do teste para os efeitos fixos apresentados na Tabela 13, mostram que a probabilidade de germinação das sementes difere entre locais e entre a interação local \sharp procedência (em escala logito), considerando-se o nível de 5% de significância.

O conjunto de gráficos dos resíduos condicionais estudentizados é apresentado na Figura 13. Estes gráficos evidenciam que o modelo se ajusta bem aos dados. Por outro lado, o valor estimado do parâmetro de dispersão (ϕ) foi igual a 1,05, mostrando a ausência de super ou de subdispersão.

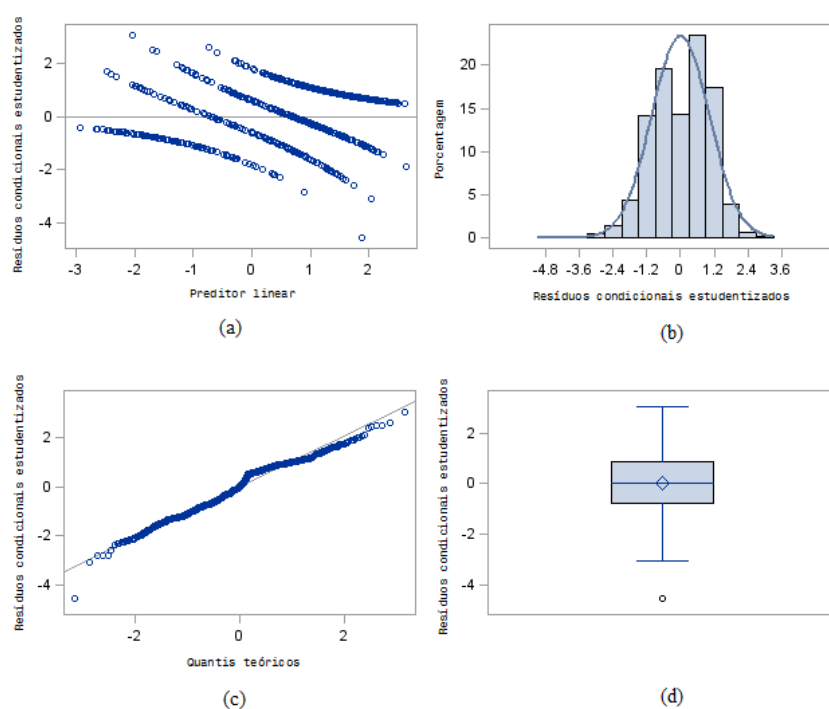


Figura 13 – Gráficos dos resíduos condicionais estudentizados, (a) em função dos valores preditos, (b) histograma, (c) quantil-quantil e (d) *box plot*, para o experimento de adaptação no estabelecimento

4.2.1 Plasticidade fenotípica

Para estudar a presença de plasticidade fenotípica, por sua vez, foi realizado o desdobramento do número de graus de liberdade de locais dentro de procedências, cujos resultados são apresentados na Tabela 14.

Tabela 14 – Estatística F e níveis descritivos (valores de p) para o desdobramento do número de graus de liberdade de local \times procedência para a variável proporção de sementes germinadas, analisando os dados de locais dentro de cada nível de procedência

Procedência	F	Valor de p
Ombrófila	46,94	< 0,0001
Semiresidual	8,43	0,0006
Restinga	17,24	< 0,0001

Considerando-se o nível de 5% de significância, os valores significativos da estatística F evidenciam a presença de plasticidade fenotípica, ou seja, sementes procedentes de uma formação florestal determinada apresentam variação na capacidade germinativa, dependendo das condições edafo-climáticas do local onde foram semeadas.

O resumo da análise do efeito do fator local dentro do fator procedências, por meio de contrastes, realizado com o objetivo de verificar se um local determinado é o mais favorável para a germinação de plantas originadas de sementes da sua mesma procedência, é apresentado na Tabela 15.

Tabela 15 – Estimativas, em escala logito e entre parênteses na escala original (proporção de sementes germinadas), dos contrastes entre locais (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) da mesma procedência contra outros locais, estatística t e níveis descritivos (valor de p)

Procedência	Contraste	Estimativa		t	Valor de p
Ombrófila	Local CB vs outros	1,1536	(0,2675)	3,55	0,0008
Semidecidual	Local CT vs outros	- 1,2265	(-0,2707)	-4,08	<0,0001
Restinga	Local IC vs outros	0,9356	(0,2173)	3,17	0,0025

A partir dos resultados apresentados nesta Tabela, considerando-se o nível de 5% de significância, observa-se que os locais Carlos Botelho e Ilha do Cardoso são os mais favoráveis para a germinação de sementes da sua mesma procedência, ao serem significativos os contrastes e apresentarem valores de proporção de germinação média superiores comparados com as médias das outras procedências avaliadas. No caso do local Caetetus (CT), apresentou as piores condições para a germinação das sementes, tanto da sua mesma procedência, quanto das outras avaliadas.

O comportamento da proporção estimada de sementes germinadas (em escala logito) segundo o local em que foi semeada cada procedência é apresentado na Figura 14. Observa-se que no local Caetetus apresentaram-se os menores valores médios da referida proporção (em escala logito).

4.2.2 Adaptação local

O resumo do teste F para o desdobramento do número de graus de liberdade de procedências dentro de locais, para estudar a adaptação local das procedências é apresentado na Tabela 16.

Na Tabela 16, considerando-se o nível de 5% de significância, observa-se que há diferenças significativas entre as proporções de sementes germinadas das diferentes formações florestais, quando são semeadas nos locais Caetetus e Ilha do Cardoso. Este fato revela a existência de adaptação local.

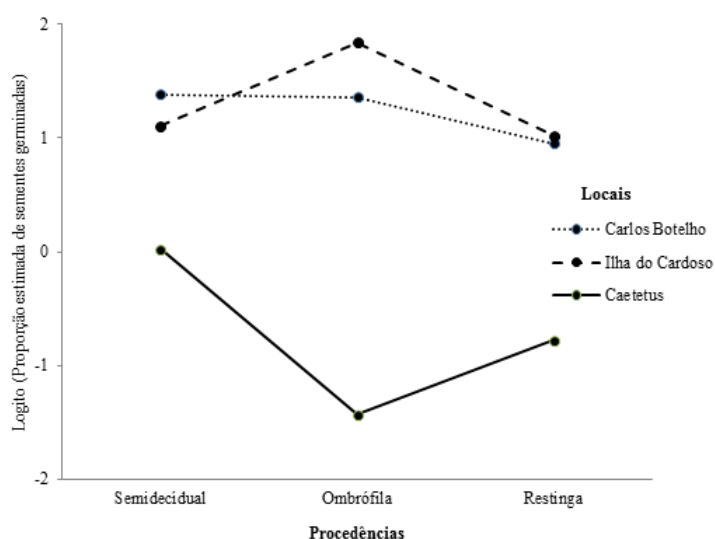


Figura 14 – Comportamento da proporção média estimada de sementes germinadas (em escala logito) segundo o local para cada procedência

Tabela 16 – Estatística F e valores de probabilidade (valores- p) para o desdobramento do número de graus de liberdade de local \times procedência para a variável proporção de sementes germinadas, analisando os dados de procedências dentro de cada nível de local

Local	F	Valor de p
Carlos Botelho	1,72	0,1885
Caetetus	16,79	< 0,0001
Ilha do Cardoso	5,69	0,0057

Finalmente, o resumo da análise do efeito do fator procedências dentro do fator local, por meio de contrastes, é apresentado na Tabela 17. Os testes apresentados foram realizados com o objetivo de verificar se plantas originadas de sementes de uma determinada procedência estão mais adaptadas ao seu próprio local, em comparação com plantas procedentes de sementes dos outros locais.

Tabela 17 – Estimativas, em escala logito e entre parênteses na escala original (proporção de sementes germinadas), dos contrastes entre procedências do mesmo local (CB=Carlos Botelho, CT=Caetetus e IC=Ilha do Cardoso) contra outras procedências, estatística t e níveis descritivos (valor de p)

Local	Contraste	Estimativa		t	Valor de p
Carlos Botelho	Proc. Ombrófila vs outras	0,1867	(0,0343)	0,80	0,4263
Caetetus	Proc. Semidecidual vs outras	1,1271	(0,2516)	5,33	<0,0001
Ilha do Cardoso	Proc. Restinga vs outras	-0,4512	(-0,0718)	-2,11	0,0399

Neste caso conclui-se que, considerando o nível de 5% de significância, plantas originadas de sementes da formação florestal Estacional Semidecidual estão mais adaptadas ao seu próprio local (Caetetus), em comparação com plantas procedentes de sementes dos outros locais, ou seja, apresentam maior germinação. No caso da formação florestal Restinga, as sementes procedentes deste bioma parecem estar menos adaptadas ao seu próprio local (Ilha do Cardoso), no caso da variável proporção de sementes germinadas.

Observa-se na Figura 15 que as menores proporções esperadas de sementes germinadas (em escala logito) foram obtidas no local Caetetus. A esse respeito, Brancalion (2009) cita que, a germinação das plantas foi prejudicada neste local provavelmente pela baixa disponibilidade hídrica nos primeiros meses após a semeadura, já que nesse local se apresenta uma estação seca bem definida que começa junto com o período de frutificação de palmitreiro.

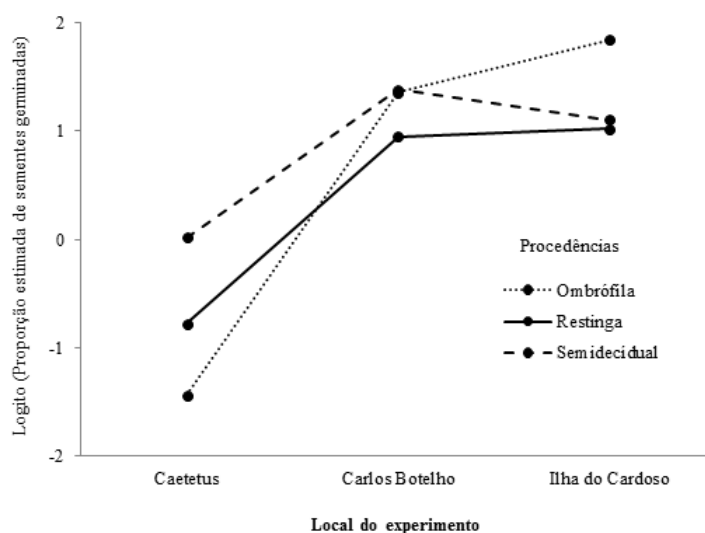


Figura 15 – Comportamento da proporção esperada de sementes germinadas (em escala logito) segundo a procedência em cada local de estudo

De acordo com Martins, Nakagawa e Bovi (1999), a intolerância à dessecação das sementes dessa espécie pode contribuir para aumentar a sensibilidade das mesmas ao estresse hídrico durante a germinação em condições de baixa disponibilidade hídrica, reduzindo o potencial fisiológico das sementes e consequentemente a emergência de plântulas em campo. Por outra parte, segundo Reis e Kageyama (2000), por se tratar de sementes recalcitrantes, uma redução no teor de umidade nas sementes (abaixo de 28%) ocasiona perda de viabilidade e, consequentemente, diminuição nas taxas de germinação.

Contudo, Brancalion (2009) cita que, mesmo nessa condição desfavorável para a sobrevivência e germinação de sementes recalcitrantes, sementes provenientes do mesmo local (Caetetus) apresentaram maiores proporções de germinação quando comparadas com as outras fontes de sementes. O mesmo autor também salienta, que a tolerância diferenciada das sementes à dessecação pode ser um indício de adaptação local para a ocorrência de *E. edulis* em florestas tropicais sujeitas ao déficit hídrico.

5 CONSIDERAÇÕES FINAIS

O emprego de técnicas de formulação de modelos lineares de efeitos mistos foi fundamental para o estudo da adaptação local e da plasticidade fenotípica de *Euterpe edulis*, quando avaliadas as variáveis relacionadas com a massa de matéria seca das plantas. As plantas originadas das sementes dos três biomas avaliados apresentaram um comportamento plástico em todos os caracteres avaliados, no ensaio de adaptação no estabelecimento (ou de semeadura). Por outro lado, no ensaio de adaptação em juvenis (ou de crescimento), a característica de plasticidade foi verificada unicamente na massa de matéria seca da folha, para as plantas provenientes do bioma Estacional Semidecidual.

A característica de adaptação local foi observada de forma evidente no ensaio de adaptação no estabelecimento (ou de semeadura), verificando que em cada local avaliado, as plantas originadas das sementes das diferentes procedências, apresentaram um comportamento diferenciado para todos os caracteres avaliados, podendo em alguns casos, tratar-se de adaptação local. Por outro lado, no ensaio de adaptação em juvenis, apenas as plantas transplantadas no bioma Restinga (Parque Estadual Ilha do Cardoso) apresentaram evidência de adaptação para a massa de matéria seca da folha, do caule e do total. As plantas originadas das sementes das formações florestais Ombrófila Densa e Restinga foram as únicas que apresentaram adaptação ao seu próprio local.

A respeito da variável proporção de sementes germinadas, a aplicação do modelo linear generalizado misto permitiu identificar diferenças significativas nos efeitos de local e da interação local \times procedência, levando em consideração os efeitos aleatórios relacionados aos blocos e às plantas matrizes, determinando a presença de plasticidade fenotípica nas plantas originadas das sementes das três procedências avaliadas. Além disso, determinou-se que os locais Carlos Botelho e Ilha do Cardoso são os mais favoráveis para a germinação das sementes da sua mesma procedência. Adicionalmente, a característica de adaptação mostrou-se presente nas plantas semeadas nos locais Caetetus e Ilha do Cardoso. Concluiu-se que as plantas originadas das sementes da formação florestal Estacional Semidecidual apresentaram melhor adaptação ao seu próprio local (Caetetus).

REFERÊNCIAS

- AKAIKE, H. A New Look at the Statistical Model Identification. **IEEE Transactions on automatic control**, New York, v. 19, n. 6, p. 716-723, Dec. 1974.
- ALCARDE, R. **Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R**. 2012. 156 p. Tese (Doutorado em Ciências) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2012.
- ANDERSON, R.L.; BANCROFT, T.A. **Statistical theory in research**. New York: McGraw-Hill, 1952. 399 p.
- ANGERT, A.; SCHEMSKE, D. The evolution of species' distributions: Reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. **Evolution**, Lancaster, v. 59, n. 8, p. 1671-1684, Aug. 2005.
- BALZARINI, M. Applications of mixed models in planta breeding. In: KANG, M.A. (Ed.) **Quantitative genetics, genomics and plant breeding**. New York: CABI Publishing, 2002. p. 353-363.
- BARBIN, D. **Componentes de variância: teoria e aplicações**. Piracicaba: FEALQ, 1993. 120 p.
- BARBIN, D. **Planejamento e análise estatística de experimentos agrônômicos**. Londrina: Mecnas, 2013. 214 p.
- BERDUGO, E. **Estudio comparativo entre los modelos Poisson-Normal y Binomial Negativo log-lineal mixto, para datos de conteo longitudinales**. 2010. 87 p. Trabajo de Grado (Magister en Ciencias Estadística) -Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, 2010.
- BRADSHAW, A.D. Evolutionary significance of phenotypic plasticity in plants. **Advances in Genetics**, New York, v. 13, p. 115-155, 1965.
- BRANCALION, P.H.S. **Contribuições de adaptações locais e da plasticidade em sementes e plântulas para a ocorrência de *Euterpe edulis* e *Syagrus romanzoffiana* em três formações florestais do Estado de São Paulo**. 2009. 153 p. Tese (Doutorado em Ciências), -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2009.
- BRESLOW, E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, Alexandria, v. 88, n. 421, p. 9-25, Mar. 1993.
- BRIEN, C.J. Analysis of variance tables based on experimental structure. **Biometrics**, Arlington, v. 39, n. 1, p. 53-59, Mar. 1983.
- BRIEN, C.J. Determining the analysis of variance table. In: **Statistical Modelling**. Disponível em: <<http://chris.brien.name/ee2/course/SM06.pdf>>. Acesso em: 16 mar. 2013>.

CARVALHO, P. E. R. **Espécies arbóreas brasileiras**. Colombo: Embrapa Florestas, 2003. 1039 p.

CASELLA, G.; BERGER, G. **Statistical Inference**. Belmont: Thomson Learning, 2002. 660 p.

CAVALLERO, L.; GALETTI, L.; LÓPEZ, D.; McCARGO, J.; MARTÍN, I. Morphological variation of the leaves of *Aechmea distichantha* Lem. plants from contrasting habitats of a Chaco forest: a trade-off between leaf area and mechanical support. **Revista Brasileira de Biociências**, Porto Alegre, v. 9, n. 4, p. 455-464, Out-Dez. 2011.

CHAMBEL, M.R.; CLIMENT, J.; GARCIA-ARRANZ, R.; ALÍA, R. Plasticidade fenotípica do crescimento e sobrevivência de proveniências de *Pinus pinaster*, *Pinus halapensis* e *Pinus sylvestris*. In: CONGRESSO NACIONAL FLORESTAL, 5., 2005, Viseu: SPCF, 2005. Comunicação oral.

COCHRAN, W.G. Problems arising in the analysis of a serie of similar experiments. **Journal of the Royal Statistical Society**. Supplement, London, v. 4, p. 102-118, 1937.

COCHRAN, W.G. The combination of estimates from different experiments. **Biometrics**, Arlington, v. 10, p. 101-129, 1954.

COCHRAN, W.G.; COX, G.M. **Experimental Desings**. New York: John Wiley, 1957. 611 p.

COLAUTTI, R.I.; ECKERT, C.G.; BARRETT, S.C.H. Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. **Proceedings of the Royal Society**. Series B, London, v. 277, n. 1689, p. 1799-1806, Feb. 2010.

COLLETT, D. **Modelling binary data**. London: Chapman & Hall, 1991. 369 p.

COMBES, R. **La forme des végétaux et le milieu**. Paris: Armand Colin, 1946. 222 p.

CORDEIRO, G.M. **Modelos Lineares Generalizados**. Campinas: Unicamp, 1986. 286 p.

CORDEIRO, G.M.; DEMÉTRIO, C.G.B. Modelos lineares generalizados. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12. **Minicurso**. Santa Maria: UFSM, 2007. 161 p.

COSTA, S.C. da **Modelos lineares generalizados mistos para dados longitudinais**. 2003. 101 p. Tese (Doutorado em Agronomia) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.

DOBSON, A.J. **An introduction to generalized linear models**. 2. ed. London: Chapman & Hall, 2001. 225 p.

DEMÉTRIO, C.G.B. **Modelos lineares generalizados em experimentação agronômica**. Piracicaba: ESALQ, Departamento de Ciências Exatas, 2002. 113 p.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete observations. **Journal of the Royal Statistical Society**. London, v. 39, p. 1-38, 1977.

ECKHART, V.; GEBER, M.; MCGUIRE, M. Experimental studies of adaptation in *Clarkia xantiana* I. sources of trait variation across a subspecies border. **Evolution**, Lancaster, v. 58, n. 1, p. 59-70, Jan. 2004.

FAI, A.H.T; CORNELIUS, P.I. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. **Journal of Statistical Computation and Simulation**, London, v. 54, n. 4, p. 363-378, 1996.

FANTINI, A. C. Estimativa da produção de palmito em plantas de palmitreiro (*Euterpe edulis* Martius) a partir de características fenotípicas. **Árvore**, Viçosa, v. 21, n. 1, p. 49-57, 1997.

FINNEY, D.J. The Estimation from Individual Records of the Relationship between Dose and Quantal Response. **Biometrika**, Oxford, v. 34, p. 320-334, Dez. 1947.

FISHER, R.A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society of London**. Series A, London, v. 222, p. 309-368, 1922.

FUZETO, A.P.; LOMÔNACO, C. Potencial plástico de *Cabralea canjara* subsp. *Polytricha* Penn. (Meliaceae) e seu papel na formação de ecótipos em áreas de cerrado e vereda, Uberlândia, MG. **Revista brasileira de Botânica**, São Paulo, v. 23, p. 169-176, 2000.

GBUR, E.E.; STROUP, W.W.; McCARTER, K.S.; DURHAM, S.; YOUNG, L.J.; CHRISTMAN, M.W.; KRAMER, M. **Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences**. 1. ed. Madison: American Society of Agronomy, Soil Science Society of America, Crop Science Society of America, 2012. 283 p.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B.R. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, n. 4, p. 1140-1450, Dec. 1995.

GREGOIRE, T.G.; SCHABENBERGER, O.; BARRETT, J.P. Linear Modelling of Irregularly Spaced, Unbalanced, Longitudinal Data from Permanent Plot Measurements. **Canadian Journal of Forest Research**, Ottawa, v. 25, n. 1, p. 137-156, Jan. 1995.

GRIZZLE, J.E.; STARMER, C.F.; KOCH, G.G. Analysis of categorical data by linear models. **Biometrics**, Arlington, v. 25, n. 3, Sept. p. 489-504, 1969.

HARTLEY, H.O. The Maximum F-Ratio as Short-Cut Test for Heterogeneity of Variances. **Biometrika**, Oxford, v. 37, p. 308-312, 1950.

HARTLEY, H.O.; RAO, J.N.K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, Oxford, v. 54, p. 93-108, June 1967.

HARVILLE, D.A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. **Journal of the American Statistical Association**, Alexandria, v. 72, n. 2, p. 320-338, June 1977.

- HENDERSON, C.R. Estimation of Variance and Covariance Components. **Biometrics**, Arlington, v. 9, n. 2, p. 226-252, June 1953.
- HENDERSON, C.R. **Applications of Linear Models in Animal Breeding**. Guelph: University of Guelph, 1984. 462 p.
- HINDE, J.P.; DEMÉTRIO, C.G.B. Overdispersion: models and estimation. **Computational Statistics and Data Analysis**, New York, v. 27, p. 151-170, 1998.
- HINKLEY, D.V.; REIDE, N.; SNELL, E.J. **Statistical Theory and Modelling**. London: Chapman and Hall, 1991. 349 p.
- HURVICH, C.M.; TSAI, C.-L. Regression and time series model selection in small samples. **Biometrika**, Oxford, v. 76, n. 2, p. 297-307, June 1989.
- JØRGENSEN, B. Exponential dispersion models (with discussion). **Journal of the Royal Statistical Society. Series B**, London, v. 49, p. 127-162, 1987.
- KEMPTHORNE, O. **The Design and Analysis of Experiments**. New York: Wiley, 1952. 631 p.
- KENWARD, M.G.; ROGER, J.H. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. **Biometrics**, Arlington, v. 53, n. 3, p. 983-997, Sept. 1997.
- LEE, Y.; NELDER, J.A. Hierarchical Generalized Linear Models. **Journal of the Royal Statistical Society. Series B**, London, v. 58, p. 619-678, 1996.
- LINDSEY, J. K. **Applying Generalized Linear Models**. New York: Springer-Verlag, 1997. 256 p.
- LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D.; SCHABENBERGER, O. **SAS ® System for Mixed Models**. 2. ed. Cary: SAS Institute Inc., 2006. 834 p.
- LÓPEZ, R.A. **Diferenciación adaptativa entre poblaciones de *Pinus canariensis* Chr. Sm. Ex DC**. 2009. 135 p. Tesis (Doctorado en Ingeniería de Montes) -Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros de Montes, Madrid, 2009.
- LORENZI, H.; MELLO FILHO, L. E. **As plantas tropicais de R. Burle Max**. São Paulo: Instituto Plantarum de Estudos da Flora, 2001. 504 p.
- MACHADO, A. de A.; SILVA, J.G.C., da; DEMÉTRIO, C.G.B.; FERREIRA, D.F. Estatística experimental: Uma abordagem baseada no planejamento e no uso de recursos computacionais. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., 2005, Londrina. **Minicurso**. Londrina: UEL, 2005. 300 p.
- MARCELINO, S.D.R.; IEMMA, A.F. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. **Scientia Agricola**, Piracicaba, v. 56, n. 4, p. 643-652, Out./Dez. 2000.

MARTÍNEZ GARZA, A. **Diseños Experimentales: Métodos y elementos de teoría**. México, D.F.: Trillas, 1988. 756 p.

MARTINS, C.C.; NAKAGAWA, J. ; BOVI, M.L.A. Tolerância à dessecação de sementes de palmito-vermelho (*Euterpe espirotosantensis* Fernandes). **Revista Brasileira de Botânica**, São Paulo, v. 22, p. 391-396. 1999.

MARTÍNS, P. **Plasticidade fenotípica de *Pinus pinaster* frente a la disponibilidad de nutrientes**. 2007. Trabajo de Investigación Tutelado (Maestría en Ingeniería de Montes) -Universidad de Santiago de Compostela, Escuela Politécnica Superior de Lugo, Departamento de Ingeniería Agroforestal, Lugo, 2007. 63 p.

McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2. ed. London: Chapman & Hall, 1989. 511 p.

McCULLOCH, C.E.; SEARLE, S.R.; NEUHAUS, J.M. **Generalized, linear, and mixed models**. 2. ed. New York: J. Wiley, 2001. 384 p.

McCULLOCH, C. E., Generalized Linear Mixed Models. NSF-CBMS Regional Conference Series in Probability and Statistics, v. 7. Beachwood, OH: Institute of Mathematical Statistics, 2003.

McGILCHRIST, C.A. Estimation in generalized mixed models. **Journal of the Royal Statistical Society**. Series B, London, v. 56, p. 61-69, 1994.

McLEAN, R.A.; SANDERS, W.L. Approximating degrees of freedom for standard errors in mixed linear models. **Proceedings of the Statistical Computing Section, American Statistical Association**, New Orleans, p. 50-59, 1988.

MEYER, K. Estimation of genetic parameters. In: HILLS, W.G.; MACKAY, T.F.C. **Evolution and animal breeding**. Wallingford: CAB International, 1989. p. 161-167.

MOLENBERGHS, G.; VERBEKE, G. **Linear Mixed Models for Discrete Longitudinal Data**. New York: Springer-Verlag, 2005. 706 p.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society**. Series A, London, v. 135, p. 370-384, 1972.

NEVES, C.S.V.J. Sementes recalcitrantes: revisão de literatura. **Pesquisa Agropecuária Brasileira**, Brasília, v. 29, n. 9, p. 1459-1467, 1994.

NOVOPLANSKY, A. Phenotypic plasticity in plants: Implications of non-cognitive behavior. **Evolutionary Ecology**, Houten, v. 16, n. 3, p. 177-188, 2002.

PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when blocks sizes are unequal. **Biometrika**, Oxford, v. 58, n. 3, p. 545-554, Dec. 1971.

PAVATE, M.V. Combined analysis if balanced incomplete block designs with some common treatmets. **Biometrics**, Arlington, v. 17, p. 111-119, 1961.

PIMENTEL-GOMES, F.; GUIMARÃES, R.F. Joint analysis of experiments in complete randomized blocks with some common treatments. **Biometrics**, Arlington, v. 14, p. 521-526, 1958.

PIMENTEL-GOMES, F. An extension of the method of joint analysis of experiments in complete randomized blocks **Biometrics**, Arlington, v. 26, p. 331-336, 1970.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15 ed. Piracicaba: FEALQ, 2009. 451 p.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528 p.

RAMÍREZ, J.A. **Plasticidad fenotípica y diferenciación genética inter- e intrapoblacional en *Quercus suber* L.: Evaluación mediante caracteres fisiológicos y marcadores moleculares**. 2010. 274 p. Tesis (Doctorado en Ingeniería de Montes) -Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros de Montes, Madrid, 2010.

RAO, C.R. Minimum variance quadratic unbiased estimation of variance components. **Journal of Multivariate Analysis**, Amsterdam, v. 1, p. 445-456, 1971.

REIS, A.; KAGEYAMA, P. Y. Dispersão de sementes de *Euterpe edulis* Martius Palmae. **Sellowia**, Itajaí, v. 49-52, p. 60-92, 2000.

RESENDE, M. D.V. de **Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes**. Colombo: Embrapa Florestas, 2007. 561 p.

ROJAS, B.A. da **Analysis of a group of experiments in combining ability in corn**. 1951. 120 p. Dissertation (M.S.) Iowa State College of Agriculture and Mechanic Arts, Ames, 1951.

ROJAS, B.A. da **The analysis of groups of similar experiments**. 1958. 161 p. Thesis (Ph.D.) Iowa State College of Agriculture and Mechanic Arts, Ames, 1958.

SAS INSTITUTE. **SAS STAT 9.3 user's guide**. Cary, 2011. Disponível em: <<http://support.sas.com/documentation/onlinedoc/stat/indexproc.html#stat131>>. Acesso em: 16 abr. 2012.

SATTERTHWAITE, F.E. Na Approximate Distribution of Estimates of Variance Components. **Biometrics Bulletin**, Washington, v. 2, n. 6, p. 110-114, Dec. 1946.

SAVOLAINEN, O.; PYHAJARVI, T.; KNURR, T. Gene flow and local adaptation in trees. **Annual Review of Ecology, Evolution & Systematics**, Palo Alto, v. 38 p. 595-619, 2007.

SCHALL, G. Estimation in generalized linear models with random effects. **Biometrika**, v. 78, n. 4, p. 719-727, Dec. 1991.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Jerusalem, v. 6, n. 2, p. 461-464, 1978.

SCHEINER, S.M. Genetics and evolution of phenotypic plasticity. **Annual Review of Ecology, Evolution & Systematics**, Palo Alto, v. 24, p. 35-68, 1993.

SCHLICHTING, C.D. The evolution of phenotypic plasticity in plants. **Annual Review of Ecology & Systematics**, Palo Alto, v. 17, p. 667-693, 1986.

SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance components**. New York: John Wiley, 1992. 536 p.

SELF, S.G.; LIANG, K-Y. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Test Under Nonstandard Conditions. **Journal of the American Statistical Association**, Alexandria, v. 82, n. 398, p. 605-610, 1987.

SHAW, R.G. Maximum-likelihood approaches to quantitative genetics of natural populations. **Evolution**, Lancaster, v. 41, p. 812-826, 1987.

STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics**. New York: McGraw-Hill, 1960. 481 p.

TAYLOR Jr., W.H.; HILTON, H.G. A Structure Diagram Symbolization for Analysis of Variance. **The American Statistician**, New York, v. 35, n. 2, p. 85-93, May. 1981.

THOMPSON, J.D. Phenotypic plasticity as a component of evolutionary chance. **Trends in Ecology & Evolution**, Amsterdam, v. 6, p. 246-249, 1991.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: SBG, 1992. 496 p. Publicado na Revista Brasileira de Genética, Ribeirão Preto, 1992.

VIEIRA, A.M.C. **Modelagem simultânea de média e dispersão e aplicações na pesquisa agrônômica**. 2008. 117 p. Tese (Doutorado em Ciências) -Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.

WEST, B.T.; WELCH, K.B.;GALECKI, A.T. **Linear mixed models: A practical guide using statistical software**. New York: Chapman & Hall, 2007. 376 p.

WILLIAMS, D. A. Extra-binomial variation in logistic linear models. **Journal of the Royal Statistical Society**. Series C, London, v. 31, n. 2, p. 144-148, 1982.

WOLFINGER, R., O'CONNELL, M. Generalized linear mixed models: a pseudo-likelihood approach. **Journal of Statistical Computation Simulation**, Blacksburg , v. 48, n. 3-4, p. 233-243, 1993.

YATES, F.; COCHRAN, W.G. The analysis of groups of experiments. **Journal of Agricultural Science**, Harpenden, v. 28, p. 556-580, 1938.

ANEXOS

ANEXO A - Códigos no SAS para o ajuste do modelo linear misto para dados de massa de matéria seca de *E. edulis*, nos ensaio de estabelecimento (semeadura) e de crescimento de mudas (juvenis).

```
ods graphics on;
PROC mixed data=sem covtest cl;
CLASS Local Bloco Proc Mat;
MODEL mfol= Local Proc Local*Proc/ddfm=satterthwaite residual;
Random Bloco(Local) Mat(Proc) Mat(Local*Proc);
lsmeans Proc*Local/slice=Local; /* Testes de adaptação */
contrast 'Local CB: Proc CB vs outras'
    Proc 2 -1 -1
    Proc*Local 2 -1 -1 0 0 0 0 0 0;
contrast 'Local CT: Proc CT vs outras'
    Proc -1 2 -1
    Proc*Local 0 0 0 -1 2 -1 0 0 0 ;
contrast 'Local IC: Proc IC vs outras'
    Proc -1 -1 2
    Proc*Local 0 0 0 0 0 0 -1 -1 2 ;
estimate 'Local CB: Proc CB vs outras'
    Proc 2 -1 -1
    Proc*Local 2 -1 -1 0 0 0 0 0 0 /divisor=2;
estimate 'Local CT: Proc CT vs outras'
    Proc -1 2 -1
    Proc*Local 0 0 0 -1 2 -1 0 0 0 /divisor=2;
estimate 'Local IC: Proc IC vs outras'
    Proc -1 -1 2
    Proc*Local 0 0 0 0 0 0 -1 -1 2 /divisor=2;

lsmeans Proc*Local/slice=Proc; /* Testes de plasticidade */

contrast 'Proc CB: Local CB vs outros'
    Local 2 -1 -1
    Proc*Local 2 0 0 -1 0 0 -1 0 0 ;
contrast 'Proc CT: Local CT vs outros'
    Local -1 2 -1
    Proc*Local 0 -1 0 0 2 0 0 -1 0 ;
contrast 'Proc IC: Local IC vs outros'
    Local -1 -1 2
    Proc*Local 0 0 -1 0 0 -1 0 0 2 ;
estimate 'Proc CB: Local CB vs outros'
    Local 2 -1 -1
    Proc*Local 2 0 0 -1 0 0 -1 0 0 /divisor=2;
estimate 'Proc CT: Local CT vs outros'
    Local -1 2 -1
    Proc*Local 0 -1 0 0 2 0 0 -1 0 /divisor=2;
estimate 'Proc IC: Local IC vs outros'
    Local -1 -1 2
    Proc*Local 0 0 -1 0 0 -1 0 0 2 /divisor=2;

RUN;
ods graphics off;
```

ANEXO B - Códigos no SAS para o ajuste do modelo linear generalizado misto para dados de proporção de sementes germinadas de *E. edulis*, no ensaio de estabelecimento (semeadura).

```

PROC glimmix DATA=germ plots=(residualpanel pearsonpanel studentpanel);
CLASS local proc rep mat;
MODEL y/n = local proc local*proc;
random rep(local);
random mat(proc);
random mat(local*proc);

covtest "mat(proc)" 0.;
covtest "rep(local)" .0.;
covtest "mat(local*proc)" ..0;

output out=new pred(ilink)= predi stderr(ilink)= sepredi pred= pred
stderr=sepred resid=resid student=student;

lsmeans local*proc/ilink plot=meanplot(sliceby=proc join);
lsmeans local*proc/ilink plot=meanplot(sliceby=proc join ilink);
lsmeans local*proc/ilink slice=Local; /* Teste de plasticidade */
lsmeans local*proc/ilink slice=Proc; /* Teste de adaptação local */

lsmeans proc/ilink;
lsmeans local/ilink;
lsmeans proc/ilink;
lsmeans local/ilink;

contrast 'Local CB: Proc CB vs outras'
    Proc 2 -1 -1
    Proc*Local 2 -1 -1 0 0 0 0 0 0;
contrast 'Local CT: Proc CT vs outras'
    Proc -1 2 -1
    Proc*Local 0 0 0 -1 2 -1 0 0 0 ;
contrast 'Local IC: Proc IC vs outras'
    Proc -1 -1 2
    Proc*Local 0 0 0 0 0 0 -1 -1 2 ;

estimate 'Local CB: Proc CB vs outras'
    Proc 2 -1 -1
    Proc*Local 2 -1 -1 0 0 0 0 0 0 /divisor=2;
estimate 'Local CT: Proc CT vs outras'
    Proc -1 2 -1
    Proc*Local 0 0 0 -1 2 -1 0 0 0 /divisor=2;
estimate 'Local IC: Proc IC vs outras'
    Proc -1 -1 2
    Proc*Local 0 0 0 0 0 0 -1 -1 2 /divisor=2;

```

```

contrast 'Proc CB: Local CB vs outros'
    Local 2 -1 -1
    Proc*Local 2 0 0 -1 0 0 -1 0 0 ;
contrast 'Proc CT: Local CT vs outros'
    Local -1 2 -1
    Proc*Local 0 -1 0 0 2 0 0 -1 0 ;
contrast 'Proc IC: Local IC vs outros'
    Local -1 -1 2
    Proc*Local 0 0 -1 0 0 -1 0 0 2 ;

estimate 'Proc CB: Local CB vs outros'
    Local 2 -1 -1
    Proc*Local 2 0 0 -1 0 0 -1 0 0 /divisor=2;
estimate 'Proc CT: Local CT vs outros'
    Local -1 2 -1
    Proc*Local 0 -1 0 0 2 0 0 -1 0 /divisor=2;
estimate 'Proc IC: Local IC vs outros'
    Local -1 -1 2
    Proc*Local 0 0 -1 0 0 -1 0 0 2 /divisor=2;
RUN;

```