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

**Caracterização da estrutura de interação genótipo e ambiente
utilizando modelo AMMI e W-AMMI por meio de Biplot**

Welinton Yoshio Hirai

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

**Piracicaba
2019**

Welinton Yoshio Hirai
Licenciado em Matemática

**Caracterização da estrutura de interação genótipo e ambiente
utilizando modelo AMMI e W-AMMI por meio de Biplot**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

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

Piracicaba
2019

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

Hirai, Welinton Yoshio

Caracterização da estrutura de interação genótipo e ambiente
utilizando modelo AMMI e W-AMMI por meio de Biplot / Welinton Yoshio
Hirai. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. –
– Piracicaba, 2019 .
65 p.

Dissertação (Mestrado) – – USP / Escola Superior de Agricultura “Luiz
de Queiroz”.

1. Multivariada 2. Melhoramento de plantas 3. Viticultura 4. Heteroge-
neidade 5. Valores ausentes . I. Título.

AGRADECIMENTOS

Primeiramente devo agradecer ao meus dois pilares de vida, minha família, que sempre me apoiaram a cada escolha que eu fizesse, mesmo com nossas dificuldades, fizeram de tudo para que eu pudesse crescer. E para minha namorada, que sempre foi minha melhor amiga, companheira e amor, esteve ao meu lado a cada momento, seja feliz ou triste, uma relação que sempre foi um conjunto de confiança, admiração e sacrifício, esteve acompanhando todos os meus passos e evolução, cuidando para que não me perdesse no meio do caminho.

Aos meus novos e velhos amigos, aos novos que fiz em Piracicaba, pessoas dedicadas que conheci dentro do departamento, mesmo sendo um grupo tão heterogêneo, respeitamos-nos como pessoa e acreditamos no potencial profissional de cada um, além das maravilhosas tarde de café que fazem o dia mais feliz. Em especial a minha turma de mestrado de 2017, que mostraram grande esforço em relação a cada dificuldade que passamos. E aos velhos amigos que continuo tendo contato, grandes amizades que construí no período da graduação e que sempre vou levar comigo.

A todos os professores do Programa de Pós-graduação em Estatística e Experimentação Agronômica (PPGEEA) que contribuíram na minha formação acadêmica, em especial ao professor Carlos Tadeu dos Santos Dias, meu orientador que auxiliou neste trabalho compartilhando seu conhecimento e experiência, confiando no meu esforço profissional. E a professora Sônia Maria De Stefano Piedade, que sempre admirei como educadora, pesquisadora e pessoa, me mostrou que o melhor conhecimento é aquele que podemos compartilhar e utilizar para ajudar outras pessoas.

A todos os funcionários da PPGEEA, que sempre estão a disposição para ajudar e auxiliar todos os alunos, fazendo sempre seu melhor para o programa crescer, em especial ao Jorge Alexandre Wiendl pelas tardes de conversas, ensinamentos e risadas.

Agradecer aos dois pesquisadores que forneceram os dados para as análises. A Dra. Mara Fernandes Moura do IAC (Instituto Agronômico de Campinas) que tive o prazer de conhece-la pessoalmente, e ao Dr. Fernando Antonio Souza de Aragão da EMBRAPA em Fortaleza.

E o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) que concedeu a bolsa, sem a qual seria inviável o desenvolvimento de pesquisas no Brasil.

”A perfeição não está diretamente conectada ao sacrifício,
mais sim com o prazer no processo.”

Gustavo Cruz

”Stop trying to fill your head with science... for to fill
your heart with love is enough.”

Richard Feynman

SUMÁRIO

Resumo	7
Abstract	8
Lista de Figuras	9
Lista de Tabelas	10
1 Introdução	11
2 Revisão de Literatura	13
2.1 Análise de grupos de experimentos	13
2.1.1 ANOVA para grupos de experimentos em um delineamento inteiramente casualizado	13
2.1.1.1 Estimativa dos parâmetros	14
2.1.1.2 Soma de quadrados	15
2.1.1.3 Quadros da ANOVA	16
2.1.2 ANOVA para grupos de experimentos em um delineamento casualizado em blocos	17
2.1.2.1 Estimativa dos parâmetros	18
2.1.2.2 Soma de quadrados	19
2.1.2.3 Quadros da ANOVA	20
2.2 Transformação de Box-Cox	21
2.3 Análise de Componentes Principais (ACP)	21
2.3.1 Decomposição em Valores Singulares (DVS)	23
2.3.2 Biplot	24
2.4 Interação Genótipo e Ambiente (IGA)	25
2.5 Modelo AMMI	26
2.5.1 Critério de seleção da família de modelos AMMI	28
2.6 AMMI ponderado (W-AMMI)	28
2.7 Imputação Múltipla	30
3 Material e Métodos	33
3.1 Descrição da 1ª análise - Dados IAC	33
3.1.1 O experimento	33
3.1.1.1 Os porta-enxertos	33
3.1.1.2 Os locais	33
3.1.1.3 A análise	34
3.2 Descrição da 2ª análise - Dados EMBRAPA	34
4 Resultados e Discussões da 1ª análise - Dados IAC	35
4.1 Análises iniciais	35
4.1.1 Para a variável	35
4.1.2 Fatores individuais	35
4.1.3 Interação entre os fatores	37
4.2 Análise para grupos de experimentos	39
4.2.1 Transformação da variável resposta	42
4.3 Modelo AMMI	45
5 Resultados e Discussões da 2ª análise - Dados EMBRAPA	49
5.1 Análises iniciais	49
5.1.1 Para a variável	49
5.1.2 Fatores individuais	49
5.2 Análise para grupos de experimentos	51

5.3	Modelo AMMI	53
5.3.1	Matriz GE	53
5.3.2	Modelo AMMI sem imputação	54
5.3.3	Modelo AMMI com imputação	55
5.3.4	Modelo W-AMMI com imputação	57
6	Conclusões	59
6.1	Pesquisas Futuras	59
	Referências	61
	Apêndices	65

RESUMO

Caracterização da estrutura de interação genótipo e ambiente utilizando modelo AMMI e W-AMMI por meio de Biplot

A estatística é uma ferramenta muito importante na área de melhoramento genético devido a necessidade de se analisar, em determinadas espécies, características de adaptabilidade e estabilidade. Uma medida que ajuda o pesquisador nas avaliações destes comportamentos é a interação genótipo $\times$ ambiente (IGA). Existem inúmeras metodologias que ajudam na caracterização deste efeito, e um destes métodos é o modelo AMMI (*Additive Main effects and Multiplicative Interaction*), em que os efeitos são estimados utilizando ANOVA (Análise de Variância) e a estrutura da interação é caracterizada por meio de ACP (Análise de Componentes Principais). Entretanto, uma pressuposição necessária para o modelo é a homogeneidade de variâncias, e caso este pressuposto não aconteça, foi proposto uma generalização do modelo AMMI, o W-AMMI (*Weighted AMMI*) que utiliza um método de DVS (Decomposição em Valores Singulares) ponderado. Com isto, o trabalho teve como objetivo avaliar a IGA por meio dos modelos AMMI e W-AMMI através de gráficos *Biplot's*. Utilizou-se 2 conjuntos de dados, em que o primeiro experimento foi realizado no Instituto Agrônomo de Campinas, afim de avaliar um híbrido de uva (SR 0.501-17) enxertada sobre 4 porta-enxertos (IAC 766, IAC 572, IAC 571-6 e IAC 313), em 2 municípios do estado de São Paulo (Votuporanga e Jundiaí) nos anos de 2012, 2013 e 2014. O segundo experimento foi realizado pela EMBRAPA - Fortaleza com o objetivo de caracterizar o fruto do melão a partir de 92 famílias, em 3 diferentes ambientes. Na primeira análise, mesmo com o fator da interação da ANOVA conjunta não sendo significativa, prosseguiu-se com a abordagem do modelo AMMI afim de caracterizar o comportamento de estabilidade que os porta-enxertos apresentam nos diferentes ambientes. Por conta da alta heterogeneidade entre os ambientes, observou-se que o modelo W-AMMI apresentou melhor comportamento para a caracterização da IGA. A análise para o segundo experimento apresentou dados faltantes, e desta forma, foi utilizado o método de imputação baseado na DVS livre de distribuição. Por falta de heterogeneidade nos 3 ambientes, constatou-se que o modelo W-AMMI apresentou comportamento parecido com o AMMI para a descrição da IGA. Conclui-se que mesmo em casos que haja independência entre os fatores de genótipos e ambientes, seria viável ao pesquisador utilizar o modelo AMMI como um complemento na análise, devido a complexidade multivariada que este fator pode apresentar. Além disto, para experimentos com homogeneidade de variância o modelo W-AMMI não apresenta melhora na caracterização, evidenciando desta forma, o objetivo da metodologia.

Palavras-chave: Multivariada; Melhoramento de plantas; Viticultura; Heterogeneidade; Valores ausentes

ABSTRACT

Characterization of structure of genotype and environment interaction using AMMI and W-AMMI models through Biplot

Statistic is a very important approach in field of quantitative genetics due to necessity analyse, in determinate species, characteristics of the adaptability and stability. One measure that helps the researcher assessing this behavior is the Genotype $\times$ Environment Interaction (GEI). There many methodologies that help characterization this effect, and one of there methods is the AMMI (Additive Main effects and Multiplicative Interaction) model, where the effects are estimated using ANOVA (Analysis of Variance) and the stucture of interaction is characterized for PCA (Principal Component Analysis). However, a assumption necessity for the model is the homogeneity of variance, and for this, was proposed a generalization of AMMI model, the W-AMMI (Weighted AMMI) that using SVD (Singular Value Decomposition) weighted. In this work, a objective was evaluated the GEI through the AMMI and W-AMMI models using Biplot's graphs. It was analyzed two data sets, fist experiment was design in IAC (Institute Agronomic of Campinas), that evaluate a hybrid grape (SR 0.501-17) on four rootstocks (IAC 766, IAC 572, IAC 571-6 e IAC 313), in two city of state of São Paulo (Votuporanga and Jundiaí), in the years 2012, 2013 and 2014. The second experiment was carried by EMBRAPA-Fortaleza with aim of characterizing the melon fruit from 92 families, in 3 differents enviroments. In the first analysis, even with interaction factor of ANOVA not was significative, continued with approach of AMMI model, in order to characterize the stability behavior the rootstocks present in the different environments. Due to heterogeneity among the environments, was observed that W-AMMI model presented better behavior for description of the IGA. The analysis for the second experimet presented missing values, and was used the imputation method based on DVS free of distribution. Due to lack of heterogeneiry in the enviroments, it was observed that W-AMMI model presented similar with AMMI, for description of the GEI. Finally, was concluded the even in cases whose factor of genotype and environments being independence, would be feasible for the researcher use the AMMI model for complement in the analysis, because the multivariate complexity that this factor can present. In addition, for experiments with homogeneity of variance, the W-AMMI model does not present improvement in the characterization, thus evidencing the objeive of the methodology.

Keywords: Multivariate; Plant breeding; Viticulture; Heterogeneity; Missing values

LISTA DE FIGURAS

2.1	Representações gráficas dos valores de linhas e colunas das matrizes $\mathbb{A}$ e $\mathbb{B}$ respectivamente.	25
2.2	Gráficos representativos da adaptabilidade e estabilidade resposta de 3 tipos de genótipos em 5 ambientes distintos	26
2.3	Fluxograma do método de DVS ponderada	29
3.1	Mapa de calor para contagens de números de repetições de cada família em cada ambiente .	34
4.1	<i>Box plot</i> (I) e Histograma (II) para a variável de produção (kg/ha).	35
4.2	<i>Box plot</i> dos níveis de porta-enxertos (I) e anos por locais (II)	36
4.3	Gráficos univariados do valor médio (I) e desvio-padrão (II) em relação a porta-enxertos . .	36
4.4	Gráficos univariados do valor médio (I) e desvio-padrão (II) em relação a anos por locais . .	37
4.5	<i>Box plot</i> da interação entre porta-enxertos e anos por locais	38
4.6	Gráfico de interação entre os nível de um fator dentro dos níveis do outro fator, em relação aos valores médios de produção (kg/ha)	39
4.7	Gráfico de barras dos valores de quadrado médio residual para cada ambiente	40
4.8	<i>Box plot</i> (I), histograma (II) e Q-Q normal (III) dos resíduos estudantizados	42
4.9	Gráficos de dispersão dos valores ajustados (I) e das ocasiões (II), pelos resíduos estudantizados	42
4.10	Gráficos entre o valor λ em relação ao Log da função de verossimilhança antes (I) e depois (II) da transformação	43
4.11	<i>Box plot</i> (I), histograma (II) e Q-Q normal (III) dos resíduos estudantizados com a variável reposta transformada	44
4.12	Gráficos de dispersão dos valores ajustados (I) e das ocasiões (II), pelos resíduos estudantizados com a variável reposta transformada	44
4.13	Biplot para a matriz de interação GE	46
4.14	Biplot para a matriz de interação GE, com a variável resposta transformada	47
4.15	Biplot para a matriz de interação GE utilizando o modelo W-AMMI	48
5.1	<i>Box plot</i> (I) e Histograma (II) para a variável peso	49
5.2	Gráfico de dispersão para os valores média de peso em relação aos fatores aditivos	50
5.3	Gráfico de dispersão para os valores de desvio-padrão de peso em relação aos fatores aditivos	50
5.4	<i>Box plot</i> para a variável peso em relação aos ambientes	51
5.5	<i>Box plot</i> para a variável peso em relação as famílias	51
5.6	<i>Box plot</i> (I), histograma (II) e Q-Q normal (III) dos resíduos estudantizados.	53
5.7	Gráficos de dispersão dos valores ajustados (I) e das ocasiões (II), pelos resíduos estudantizados	53
5.8	Gráficos da matriz de interação GE com valores ausentes	54
5.9	Biplot para a matriz de interação GE com valores ausentes	55
5.10	Gráficos da matriz de interação GE com valores imputados	56
5.11	Biplot para a matriz de interação GE com valores imputados	57
5.12	Gráficos da matriz de interação GE com valores imputados utilizando a DVS ponderada . .	58
5.13	Biplot para a matriz de interação GE do modelo W-AMMI	58

LISTA DE TABELAS

2.1	Esperança dos Quadrados Médios para grupos de experimentos em Delineamento Inteiramente Causalizado	16
2.2	Quadro da ANOVA para grupos de experimentos, com o fator local sendo de efeito fixo e aleatório	17
2.3	Esperança dos Quadrados Médios para grupos de experimentos em Delineamento Causalizado em Blocos	20
2.4	Quadro da ANOVA para grupos de experimentos, com o fator local sendo de efeito fixo e aleatório	20
4.1	Tabela para valores médios e desvios-padrões (kg/ha), de porta-enxerto dentro de anos por locais	38
4.2	Quadro da ANOVA individual para cada ambiente	40
4.3	Quadro da ANOVA conjunta	41
4.4	Quadro da ANOVA para grupos de experimentos, com a transformação da variável resposta	43
4.5	Valores estimados da matriz GE, provenientes do modelo da ANOVA conjunta	45
4.6	Valores estimados da matriz GE, provenientes do modelo da ANOVA conjunta com a variável resposta transformada	45
4.7	Quadro de seleção de modelos AMMI utilizando o critério de Gollob	46
4.8	Matriz W de ponderação	47
4.9	Nova matriz gerada a partir do processo iterativo de DVS ponderado	48
5.1	Quadro da ANOVA individual para cada ambiente	52
5.2	Quadro da ANOVA conjunta	52
5.3	Tabela com as informações dos valores ausentes	54
5.4	Quadro de seleção de modelos AMMI utilizando o critério de Gollob	55
5.5	Informações dos valores imputados	56
5.6	Quadro de seleção de modelos AMMI utilizando o critério de Gollob	56

1 INTRODUÇÃO

Em virtude das evoluções tecnológicas e computacionais estamos cada vez mais avançando nas pesquisas gerando resultados que antes não eram possíveis, e um fator que propicia esse desenvolvimento é a facilidade de trabalhar conjuntamente com inúmeras variáveis vindas de grandes conjuntos de dados. Entretanto, analisá-los a partir de estatísticas univariada ou bivariada torna-se massivo e trabalhoso, desta forma, a estatística multivariada fornece grande contribuição, uma vez que sua aplicabilidade utiliza métodos matemáticos somados às avaliações estatísticas, sendo capaz de caracterizar, agrupar, discriminar, e até reduzir as informações contidas em banco de dados.

Em estudos de melhoramento genético, têm-se a necessidade de investigar os diferentes comportamentos que determinados genótipos podem produzir em ambientes diversos, tanto em função da adaptabilidade, como da estabilidade. Para isso é necessário medir as causas e efeitos naturais que ocorrem devido à interação entre os genótipos e ambientes (IGA). Essas análises podem ajudar o pesquisador a reduzir o excessivo custo de avaliações genéticas futuras, pois as informações geradas podem prever o comportamento genotípico.

Na literatura tem-se muitas metodologias estatísticas diferentes para avaliações do comportamento genético com abordagem univariada tais como: SHUKLA (1972) que utiliza os componentes de variância para as avaliações, LIN e BINNS (1988) que mensura a resposta genotípica com a resposta máxima para cada ambiente e encontra a distância entre as duas métricas, TAI (1971) por meio de regressão linear simples. Todavia, devido a complexidade estrutural da IGA a maior desvantagem desses métodos ocorre devido a ausência da possibilidade de captar a resposta multidimensional (FLORES *et al.*, 1998).

Desta forma, MANDEL (1971) desenvolveu um modelo estatístico que analisa dois fatores aditivos, e a partir das estimativas dos erros particiona o fator multiplicativo utilizando técnicas de álgebra linear. GAUCH e ZOBEI (1988), utilizam desta metodologia para propor um modelo para avaliação genética, que obtêm as estimativas dos efeitos aditivos de genótipos e ambientes à partir da metodologia de análise de variância (ANOVA), e o efeito multiplicativo da IGA pela análise de componentes principais (ACP), os modelos AMMI (*Additive Main effects and Multiplicative Interaction*).

A ACP é uma metodologia da álgebra linear que calcula novas variáveis não correlacionadas (índices) geradas por combinações lineares das variáveis originais (MANLY e ALBERTO, 2017). Com base nesses índices, reduz-se a informação para uma dimensão menor, porém representando o comportamento essencial das variáveis originais, não considerando o ruído ou redundância que existente nos dados. Além disto, para uma melhor interpretação dos resultados, foi proposto por KUNO (1971) o Biplot, um gráfico bidimensional formado por eixos cartesianos, produzido pelos índices avaliados com maiores informações, e vetores que representam as grandezas das variáveis do experimento.

Entretanto, nas avaliações da IGA pode haver casos em que os ambientes são significativamente heterogêneos, ou os dados apresentam observações atípicas que podem ser causadas, por fatores não previstos pelo pesquisador. De acordo com EDWARDS e JANNINK (2006), uma importante pressuposição na análise de experimentos multi-ambientais é a homogeneidade dos erros e das variâncias da IGA. Para RODRIGUES *et al.* (2016), o método de mínimos quadrados é sensível à presença de valores atípicos, resultando numa interpretação equivocada e numa possível decisão incorreta. Caso haja a violação dessas pressuposições, HADASCH *et al.* (2018) afirmam que o método de estimação generalizada de mínimos quadrados se torna mais apropriado do que o método de mínimos quadrados ordinários.

Em virtude disto foi proposta uma generalização para o modelo AMMI, que pondera as estimativas considerando a presença de heterogeneidade das variâncias e valores atípicos. O modelo AMMI Ponderado (W-AMMI - *Weighted AMMI*), proposto por RODRIGUES *et al.* (2014), que ajusta os efeitos principais pelo método de mínimos quadrados, porém usa um algoritmo de ponderação da decomposição de valor singular (DVS) com posto reduzido para estimar o efeito multiplicativo de interação.

O trabalho tem como objetivo caracterizar a estrutura de interação genótipo e ambiente utilizando os modelos AMMI e o W-AMMI por meio dos *Biplot's*, verificando se há uma resposta significativa na estimação dos efeitos aditivos e multiplicativos, dada a heterogeneidade entre ambientes.

2 REVISÃO DE LITERATURA

2.1 Análise de grupos de experimentos

Em certos experimentos biológicos, há necessidade de se verificar como determinados tratamentos respondem em diferentes locais, isto por conta das características ambientais que acontecem, como exemplos: tipo do solo, clima da região, histórico do local, tipo de vegetação, entre outros (BANZATTO e KRONKA, 2006). Assim a estatística experimental oferece uma metodologia chamada Análise de Variância (ANOVA) para grupos de experimentos que consegue estimar efeitos dos fatores individuais e da interação, provocados pelo tratamento aplicado no experimento em diferentes locais.

Esta análise é realizada geralmente por delineamentos simples e bem planejados pois nos diferentes locais é necessário estar com a mesma estrutura (CAMPOS, 1983). Para a sequência da realização desta análise, primeiramente é efetuada a ANOVA individual de cada local que leva em consideração apenas o fator de tratamento dentro do modelo (ou o fator de bloco, caso seja necessário no experimento), em seguida realiza-se a ANOVA conjunta levando em consideração o efeito de local como um novo fator no modelo estatístico, e em virtude disso a interação entre eles.

Os fatores apresentados no modelo, podem ser analisados como dois tipos de efeitos. Os efeitos fixos, em que os níveis de um determinado fator, são atribuídos a um conjunto finito e particular dos dados levados ao interesse do pesquisador, ou os efeitos aleatórios, em que os níveis de um fator, se apresentam como um conjunto infinito que é estabelecido à partir de uma amostra aleatória oriunda da distribuição de probabilidade que os dados se originam.

2.1.1 ANOVA para grupos de experimentos em um delineamento inteiramente casualizado

Sendo um conjunto de dados com $i = 1, \dots, I$ (tratamentos), $j = 1, \dots, J$ (locais) e $k = 1, \dots, R$ (repetições) realizado em um delineamento inteiramente casualizado, desta forma temos que o modelo estatístico para a ANOVA individual é dado por (PIMENTEL, 1990; BANZATTO e KRONKA, 2006; MONTGOMERY *et al.*, 2012):

$$Y_{ik} = \mu + \tau_i + \epsilon_{ik}, \quad (2.1)$$

sendo que:

Y_{ik} : observação do i -ésimo tratamento e k -ésima repetição;

μ : constante inerente à parametrização do modelo (usualmente média geral);

τ_i : efeito do i -ésimo tratamento;

ϵ_{ik} : erro experimental do i -ésimo tratamento e k -ésima repetição, assumindo-se $\epsilon_{ik} \stackrel{IID}{\sim} \mathbb{N}(0, \sigma^2)$

Em seguida é feita a ANOVA conjunta, levando em consideração o local como fator no modelo estatístico (BANZATTO e KRONKA, 2006; MONTGOMERY *et al.*, 2012; PIMENTEL, 1990), desta forma procedendo-se que os efeitos de tratamento e local é fixo, temos que:

$$Y_{ijk} = \mu + \tau_i + \lambda_j + (\tau\lambda)_{ij} + \epsilon_{ijk}, \quad (2.2)$$

sendo que:

Y_{ijk} : observação do i -ésimo tratamento, j -ésimo local e k -ésima repetição;

μ : constante inerente a parametrização do modelo (usualmente média geral);

τ_i : efeito do i -ésimo tratamento;

λ_j : efeito do j -ésimo local;

$(\tau\lambda)_{ij}$: efeito da interação do i -ésimo tratamento e j -ésimo local;

ϵ_{ijk} : erro experimental do i -ésimo tratamento, j -ésimo local e k -ésima repetição, assumindo-se $\epsilon_{ijk} \stackrel{IID}{\sim}$

$$\mathbb{N}(0, \sigma^2)$$

considerando o caso em que o efeito de tratamento é fixo, todavia de local sendo aleatório, os fatores de local e interação do modelo anterior 2.2 seguem como:

$$\lambda_j \stackrel{IID}{\sim} \mathbb{N}(0, \sigma_L^2) \quad \text{e} \quad (\tau\lambda)_{ij} \stackrel{IID}{\sim} \mathbb{N}(0, \sigma_{TL}^2) \quad (2.3)$$

2.1.1.1 Estimativa dos parâmetros

Supondo que as pressuposições são atendidas, o erro (ϵ) tende a normalidade, possui valor esperado nulo, variância constante e independente (MONTGOMERY *et al.*, 2012), os resíduos podem ser encontrados pela diferença entre os valores observados pelos esperados,

$$\begin{aligned} \epsilon_{ijk} &= Y_{ijk} - E[Y_{ijk}] \\ &= Y_{ijk} - (\mu + \tau_i + \lambda_j + (\tau\lambda)_{ij}) \end{aligned} \quad (2.4)$$

desta forma, utilizando-se o método de mínimos quadrados (MOOD, 1963), para encontrar as estimativas dos parâmetros $\hat{\mu}$, $\hat{\tau}$, $\hat{\lambda}$ e $(\hat{\tau}\lambda)$, dado pela soma dos resíduos ao quadrado

$$\begin{aligned} Z(Y; \mu, \tau, \lambda, (\tau\lambda)) &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R \epsilon_{ijk}^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - (\mu + \tau_i + \lambda_j + (\tau\lambda)_{ij}))^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - (\tau\lambda)_{ij})^2 \end{aligned} \quad (2.5)$$

minimizando a função da soma de quadrados (Z), e realizando as derivadas parcial em relação a cada parâmetro, temos:

$$\left\{ \begin{aligned} \frac{\partial Z(Y; \mu, \tau, \lambda, (\tau\lambda))}{\partial \mu} &= -2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - (\tau\lambda)_{ij}) & \text{(I)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, (\tau\lambda))}{\partial \tau} &= -2 \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - (\tau\lambda)_{ij}) & \text{(II)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, (\tau\lambda))}{\partial \lambda} &= -2 \sum_{i=1}^I \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - (\tau\lambda)_{ij}) & \text{(III)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, (\tau\lambda))}{\partial (\tau\lambda)} &= -2 \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - (\tau\lambda)_{ij}) & \text{(IV)} \end{aligned} \right. \quad (2.6)$$

$$\left\{ \begin{array}{l}
-2\left(\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R Y_{ijk} - IJR\hat{\mu} - JR \sum_{i=1}^I \hat{\tau}_i - IR \sum_{j=1}^J \hat{\lambda}_j - R \sum_{i=1}^I \sum_{j=1}^J (\hat{\tau}\lambda)_{ij}\right) = 0 \quad \text{(I)} \\
-2\left(\sum_{j=1}^J \sum_{k=1}^R Y_{ijk} - JR\hat{\mu} - JR\hat{\tau}_i - R \sum_{j=1}^J \hat{\lambda}_j - R \sum_{j=1}^J (\hat{\tau}\lambda)_{ij}\right) = 0 \quad \text{(II)} \\
-2\left(\sum_{i=1}^I \sum_{k=1}^R Y_{ijk} - IR\hat{\mu} - R \sum_{i=1}^I \hat{\tau}_i - IR\hat{\lambda}_j - R \sum_{i=1}^I (\hat{\tau}\lambda)_{ij}\right) = 0 \quad \text{(III)} \\
-2\left(\sum_{k=1}^R Y_{ijk} - R\hat{\mu} - R\hat{\tau}_i - R\hat{\lambda}_j - R(\hat{\tau}\lambda)_{ij}\right) = 0 \quad \text{(IV)}
\end{array} \right. \quad (2.7)$$

Igualando-os a zero chega-se num sistema homogêneo 2.7, entretanto, chega-se a um sistema que não possui solução única. Tratando-se como equações normais não determinadas para o sistema ter solução única, são impostas seguintes restrições:

$$\sum_{i=1}^I \hat{\tau}_i = \sum_{j=1}^J \hat{\lambda}_j = \sum_{i=1}^I (\hat{\tau}\lambda)_{ij} = \sum_{j=1}^J (\hat{\tau}\lambda)_{ij} = \sum_{i=1}^I \sum_{j=1}^J (\hat{\tau}\lambda)_{ij} = 0$$

com isso, dispondo de notações matemáticas para somas das médias apresentadas no Apêndice 6.1, tem-se

De (I), chega-se $\hat{\mu} = \bar{Y}...$

Com (I) e (II), chega-se $\hat{\tau}_i = \bar{Y}_{i..} - \bar{Y}...$

Com (I) e (III), chega-se $\hat{\lambda}_j = \bar{Y}_{.j.} - \bar{Y}...$ (2.8)

Com (I), (II), (III), (IV) chega-se $(\hat{\tau}\lambda)_{ij} = \bar{Y}_{ij.} - \bar{Y}... - (\bar{Y}_{i..} - \bar{Y}...) - (\bar{Y}_{.j.} - \bar{Y}...) = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...$

2.1.1.2 Soma de quadrados

A soma de quadrados total corrigida é dada por,

$$\begin{aligned}
& \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \bar{Y}...) ^2 = \\
& = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R [Y_{ijk} + (\bar{Y}_{i..} - \bar{Y}_{i..}) + (\bar{Y}_{.j.} - \bar{Y}_{.j.}) + (\bar{Y}_{ij.} - \bar{Y}_{ij.}) + (\bar{Y}... - \bar{Y}...) - \bar{Y}...]^2 = \\
& = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R [(\bar{Y}_{i..} - \bar{Y}...) + (\bar{Y}_{.j.} - \bar{Y}...) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...) + (Y_{ijk} - \bar{Y}_{ij.})]^2 = \quad (2.9) \\
& = JR \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}...)^2 + IR \sum_{j=1}^J (\bar{Y}_{.j.} - \bar{Y}...)^2 + \\
& + R \sum_{i=1}^I \sum_{j=1}^J (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...)^2 + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \bar{Y}_{ij.})^2
\end{aligned}$$

observa-se que os duplos produtos são todos nulos, e desta forma a expressão da soma de quadrados total pode ser particionada em soma de quadrados para tratamento, local, interação tratamento e local e resíduo, dado por $SQT_{Total} = SQT_{Trat} + SQT_{Local} + SQT \times L + SQRes$

Pela correção de Bessel (FAREBROTHER, 1999), pode-se rescrever as somas de quadrados, de forma que:

$$\begin{aligned}
 SQT_{Total} &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R Y_{ijk}^2 - \frac{Y_{...}^2}{IJR} \\
 SQT_{Trat} &= \frac{1}{JR} \sum_{i=1}^I Y_{i..}^2 - \frac{Y_{...}^2}{IJR} \\
 SQT_{Local} &= \frac{1}{IR} \sum_{j=1}^J Y_{.j.}^2 - \frac{Y_{...}^2}{IJR} \\
 SQT_{L} &= \frac{1}{R} \sum_{i=1}^I \sum_{j=1}^J Y_{ij.}^2 - \frac{Y_{...}^2}{IJR} \\
 SQT \times L &= SQT_{L} - SQT_{Trat} - SQT_{Local}
 \end{aligned} \tag{2.10}$$

E a soma de quadrados dos resíduos da análise conjunta, é dada pela soma dos quadrados dos resíduos das análises individuais, $SQRes_{[Local_1]} + SQRes_{[Local_2]} + \dots + SQRes_{[Local_J]}$ (BANZATTO e KRONKA, 2006)

2.1.1.3 Quadros da ANOVA

Com isto, tem-se que a tabela dos valores esperados dos quadrados médios, para local como efeito fixo ou aleatório (Tabela 2.1), e o quadro da ANOVA para grupos de experimentos (Tabela 2.2), com o valores F para cada caso.

Tabela 2.1. Esperança dos Quadrados Médios para grupos de experimentos em Delineamento Inteiramente Causalizado

$E[QM]$	Local: efeito fixo Tratamento: efeito fixo	Local: efeito aleatório Tratamento: efeito fixo
$E[QMLocal]$	$\sigma^2 + \frac{IR \sum_j \lambda_j^2}{J-1}$	$\sigma^2 + R\sigma_{TL}^2 + IR\sigma_L^2$
$E[QMTrat]$	$\sigma^2 + \frac{JR \sum_i \tau_i^2}{I-1}$	$\sigma^2 + R\sigma_{TL}^2 + \frac{JR \sum_i \tau_i^2}{I-1}$
$E[QMT \times L]$	$\sigma^2 + \frac{R \sum_i \sum_j (\tau\lambda)_{ij}^2}{(I-1)(J-1)}$	$\sigma^2 + R\sigma_{TL}^2$
$E[QMRes]$	σ^2	σ^2

Tabela 2.2. Quadro da ANOVA para grupos de experimentos, com o fator local sendo de efeito fixo e aleatório

Fontes de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	Local: Efeito fixo	Local: Efeito aleatório
Locais	(J-1)	$SQLocal$	$QMLocal = \frac{SQLocal}{J-1}$	$F_{calc} = \frac{QMLocal}{QMRes}$	$F_{calc} = \frac{QMTrat}{QMT \times L}$
Tratamentos	(I-1)	$SQTrat$	$QMTrat = \frac{SQTrat}{I-1}$	$F_{calc} = \frac{QMTrat}{QMRes}$	$F_{calc} = \frac{QMTrat}{QMT \times L}$
Interação ($T \times L$)	(I-1)(J-1)	$SQT \times L$	$QMT \times L = \frac{SQT \times L}{(I-1)(J-1)}$	$F_{calc} = \frac{QMT \times L}{QMRes}$	$F_{calc} = \frac{QMT \times L}{QMRes}$
Resíduo (Res)	IJ(R-1)	$SQRes$	$QMRes = \frac{SQRes}{IJ(R-1)}$	—————	—————
Total	(IJR-1)	$SQTotal$	—————	—————	—————

2.1.2 ANOVA para grupos de experimentos em um delineamento casualizado em blocos

Obtém-se a ANOVA, para um delineamento casualizado em blocos (DCB) para um conjunto de dados com $i = 1, \dots, I$ (tratamentos), $j = 1, \dots, J$ (locais) e $k = 1, \dots, R$ (blocos), tem-se que o modelo estatístico para a ANOVA individual é dado por (PIMENTEL, 1990; BANZATTO e KRONKA, 2006; MONTGOMERY *et al.*, 2012):

$$Y_{ik} = \mu + \tau_i + b_k + \epsilon_{ik}, \quad (2.11)$$

sendo que:

Y_{ik} : observação do i -ésimo tratamento e k -ésimo bloco;

μ : constante inerente à parametrização do modelo (usualmente média geral);

τ_i : efeito do i -ésimo tratamento;

b_k : efeito do k -ésimo bloco;

ϵ_{ik} : erro experimental do i -ésimo tratamento e k -ésimo bloco, em que assume-se $\epsilon_{ik} \stackrel{IID}{\sim} N(0, \sigma^2)$

Em seguida é realizada a ANOVA conjunta considerando o efeito de tratamentos, e local como fixo e o efeito de bloco como aleatório (controle dos possíveis efeitos externos contidos na área experimental):

$$Y_{ijk} = \mu + \tau_i + \lambda_j + b_{k(j)} + (\tau\lambda)_{ij} + \epsilon_{ijk}, \quad (2.12)$$

sendo que:

Y_{ijk} : observação do i -ésimo tratamento, j -ésimo local e k -ésimo bloco;

μ : constante inerente à parametrização do modelo (usualmente média geral);

τ_i : efeito do i -ésimo tratamento;

λ_j : efeito do j -ésimo local;

$b_{k(j)}$: efeito do k -ésimo bloco dentro do j -ésimo local, em que assume-se $b_{k(j)} \stackrel{IID}{\sim} N(0, \sigma_B^2)$;

$(\tau\lambda)_{ij}$: efeito da interação do i -ésimo tratamento e j -ésimo local;

ϵ_{ijk} : erro experimental do i -ésimo tratamento, j -ésimo local e k -ésimo bloco, em que assume-se $\epsilon_{ijk} \stackrel{IID}{\sim} N(0, \sigma^2)$

Reciprocamente ao caso da ANOVA para grupos de experimentos em DIC, também podemos verificar quando o local é de efeito aleatório, a partir do modelo 2.12 assume-se que local segue aproximadamente uma distribuição normal e, com isto, a interação também

$$\lambda_j \stackrel{IID}{\sim} N(0, \sigma_L^2) \quad \text{e} \quad (\tau\lambda)_{ij} \stackrel{IID}{\sim} N(0, \sigma_{TL}^2) \quad (2.13)$$

2.1.2.1 Estimativa dos parâmetros

Supondo que as pressuposições da ANOVA foram atendidas, o resíduo é dado pela diferença entre o valor observado pelo esperado (Equação 2.14)

$$\begin{aligned}\epsilon_{ijk} &= Y_{ijk} - E[Y_{ijk}] \\ &= Y_{ijk} - (\mu + \tau_i + \lambda_j + b_{k(j)} + (\tau\lambda)_{ij})\end{aligned}\quad (2.14)$$

desta forma, utilizando o método de mínimos quadrados, é calculado a soma dos resíduos ao quadrado e esta é chamada de função da soma dos resíduos ao quadrado (Equação 2.15)

$$Z(Y; \mu, \tau, \lambda, b, (\tau\lambda)) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R \epsilon_{ijk}^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij})^2 \quad (2.15)$$

com isso encontra-se os estimadores $\hat{\mu}$, $\hat{\tau}$, $\hat{\lambda}$, $\hat{b}$ e $(\hat{\tau\lambda})$, de modo que minimize a função (Z), após isto calcula as derivadas parciais em relação a cada parâmetro

$$\left\{ \begin{aligned} \frac{\partial Z(Y; \mu, \tau, \lambda, b, (\tau\lambda))}{\partial \mu} &= -2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij}) & \text{(I)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, b, (\tau\lambda))}{\partial \tau} &= -2 \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij}) & \text{(II)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, b, (\tau\lambda))}{\partial \lambda} &= -2 \sum_{i=1}^I \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij}) & \text{(III)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, b, (\tau\lambda))}{\partial b} &= -2 \sum_{i=1}^I (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij}) & \text{(IV)} \\ \frac{\partial Z(Y; \mu, \tau, \lambda, b, (\tau\lambda))}{\partial (\tau\lambda)} &= -2 \sum_{k=1}^R (Y_{ijk} - \mu - \tau_i - \lambda_j - b_{k(j)} - (\tau\lambda)_{ij}) & \text{(V)} \end{aligned} \right. \quad (2.16)$$

e igualado-os ao valor nulo levando à um sistema de equações homogêneo

$$\left\{ \begin{aligned} -2 \left(\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R Y_{ijk} - IJR\hat{\mu} - JR \sum_{i=1}^I \hat{\tau}_i - IR \sum_{j=1}^J \hat{\lambda}_j - I \sum_{j=1}^J \sum_{k=1}^R \hat{b}_{k(j)} - R \sum_{i=1}^I \sum_{j=1}^J (\hat{\tau\lambda})_{ij} \right) &= 0 & \text{(I)} \\ -2 \left(\sum_{j=1}^J \sum_{k=1}^R Y_{ijk} - JR\hat{\mu} - JR\hat{\tau}_i - R \sum_{j=1}^J \hat{\lambda}_j - \sum_{j=1}^J \sum_{k=1}^R \hat{b}_{(k)j} - R \sum_{j=1}^J (\hat{\tau\lambda})_{ij} \right) &= 0 & \text{(II)} \\ -2 \left(\sum_{i=1}^I \sum_{k=1}^R Y_{ijk} - IR\hat{\mu} - R \sum_{i=1}^I \hat{\tau}_i - IR\hat{\lambda}_j - I \sum_{k=1}^R \hat{b}_{(k)j} - R \sum_{i=1}^I (\hat{\tau\lambda})_{ij} \right) &= 0 & \text{(III)} \\ -2 \left(\sum_{i=1}^I -I\hat{\mu} - \sum_{i=1}^I \hat{\tau}_i - I\hat{\lambda}_j - I\hat{b}_{(k)j} - \sum_{i=1}^I (\hat{\tau\lambda})_{ij} \right) &= 0 & \text{(IV)} \\ -2 \left(\sum_{k=1}^R Y_{ijk} - R\hat{\mu} - R\hat{\tau}_i - R\hat{\lambda}_j - \sum_{k=1}^R \hat{b}_{(k)j} - R(\hat{\tau\lambda})_{ij} \right) &= 0 & \text{(V)} \end{aligned} \right. \quad (2.17)$$

Encontra-se novamente um sistema de equações normais não determinado, em que para o próximo passo são impostas as seguintes restrições:

$$\sum_{i=1}^I \hat{\tau}_i = \sum_{j=1}^J \hat{\lambda}_j = \sum_{i=1}^I (\hat{\tau}\hat{\lambda})_{ij} = \sum_{j=1}^J (\hat{\tau}\hat{\lambda})_{ij} = \sum_{i=1}^I \sum_{j=1}^J (\hat{\tau}\hat{\lambda})_{ij} = \sum_{k=1}^R \hat{b}_{(k)j} = \sum_{j=1}^J \sum_{k=1}^R \hat{b}_{(k)j} = 0$$

com isso, chegamos que dispondo de notações matemáticas apresentadas no Apêndice 6.1

De (I), tem-se $\hat{\mu} = \bar{Y}_{...}$

Com (I) e (II), chega-se que $\hat{\tau}_i = \bar{Y}_{i..} - \bar{Y}_{...}$

Com (I) e (III), chega-se que $\hat{\lambda}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$ (2.18)

Com (I), (III) e (IV), chega-se que $\hat{b}_{(k)j} = \bar{Y}_{.jk} - \bar{Y}_{.j.}$

Com (I), (II), (III), (IV) tem-se $(\hat{\tau}\hat{b})_{ij} = \bar{Y}_{ij.} - \bar{Y}_{...} - (\bar{Y}_{i..} - \bar{Y}_{...}) - (\bar{Y}_{.j.} - \bar{Y}_{...}) = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

2.1.2.2 Soma de quadrados

A soma de quadrados total corrigida é dada por (MONTGOMERY, 2013):

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \bar{Y}_{...})^2 &= \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R [Y_{ijk} + \bar{Y}_{i..} - \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{.j.} + (\bar{Y}_{ij.} - \bar{Y}_{ij.}) + (\bar{Y}_{.jk} - \bar{Y}_{.jk}) + (\bar{Y}_{.j.} - \bar{Y}_{.j.}) + \\ &+ (\bar{Y}_{...} - \bar{Y}_{...}) - \bar{Y}_{...}]^2 = \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R [(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) + \\ &+ (\bar{Y}_{.jk} - \bar{Y}_{.j.}) + (Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{.jk} + \bar{Y}_{.j.})]^2 = \\ &= JR \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}_{...})^2 + IR \sum_{j=1}^J (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + R \sum_{i=1}^I \sum_{j=1}^J (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + \\ &+ I \sum_{j=1}^J \sum_{k=1}^R (\bar{Y}_{.jk} - \bar{Y}_{.j.})^2 + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R (Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{.jk} + \bar{Y}_{.j.})^2 \end{aligned} \quad (2.19)$$

como os duplos produtos são todos nulo a expressão da soma de quadrados total pode ser particionada por: $SQT_{Total} = SQT_{Trat} + SQT_{Local} + SQT \times L + SQT_{Bloco} + SQT_{Res}$. Como visto na ANOVA para grupos de experimentos em um delineamento inteiramente casualizado, pela correção de Bessel (FAREBROTHER, 1999), as somas de quadrados de total, tratamentos, locais, interação tratamentos e locais e para resíduos, continuam similares às anteriores 2.10.

Agora para a soma de quadrados de blocos dentro de local, observa-se que o cálculo é dado pela diferença entre a soma de quadrados de blocos pela soma de quadrados de local,

$$SQT_{Bloco}[Local] = \frac{1}{I} \sum_{j=1}^J \sum_{k=1}^R Y_{.jk}^2 - \frac{1}{IR} \sum_{j=1}^J Y_{.j.}^2 \quad (2.20)$$

ou por BANZATTO e KRONKA (2006) a $SQBloco[Local]$, pode ser calculado também utilizando as somas de quadrados encontradas de cada ANOVA individual

$$SQBloco[Local] = SQBloco[Local_1] + SQBloco[Local_2] + \dots + SQBloco[Local_J] \quad (2.21)$$

Porém, na ANOVA para grupos de experimentos não há interesse verificar se o efeito de blocos dentro dos locais é significativamente diferente, dado que, como o bloco é um princípio de controle local dos efeitos não controlados contidos na área experimental. Desta forma, cada local têm efeitos particulares, porém não deve ser descartado do modelo este fator, pois aumentaria os graus de liberdade do resíduo inflacionado com o efeito externos do diferentes locais.

2.1.2.3 Quadros da ANOVA

Novamente, é observado dois casos, sendo quando local é de efeito fixo e aleatório, conduzindo para a esperança dos quadrados médios (Tabela 2.3) e o quadro da ANOVA para grupos de experimentos (Tabela 2.4).

Tabela 2.3. Esperança dos Quadrados Médios para grupos de experimentos em Delineamento Causalizado em Blocos

$E[QM]$	Local: efeito fixo Tratamento: efeito fixo	Local: efeito aleatório Tratamento: efeito fixo
$E[QMBloco[Local]]$	$\sigma^2 + \frac{I \sum_j^J \sum_k^R b_{(k)j}^2}{J(R-1)}$	$\sigma^2 + IJ\sigma_B^2$
$E[QMLocal]$	$\sigma^2 + \frac{IR \sum_j^J \lambda_j^2}{J-1}$	$\sigma^2 + IJ\sigma_B^2 + R\sigma_{TL}^2 + IR\sigma_L^2$
$E[QMTrat]$	$\sigma^2 + \frac{JR \sum_i^I \tau_i^2}{I-1}$	$\sigma^2 + R\sigma_{TL}^2 + \frac{JR \sum_i^I \tau_i^2}{I-1}$
$E[QMT \times L]$	$\sigma^2 + \frac{R \sum_i^I \sum_j^J (\tau\lambda)_{ij}^2}{(I-1)(J-1)}$	$\sigma^2 + R\sigma_{TL}^2$
$E[QMRes]$	σ^2	σ^2

Tabela 2.4. Quadro da ANOVA para grupos de experimentos, com o fator local sendo de efeito fixo e aleatório

Fontes de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	Local: Efeito fixo	Local: Efeito aleatório
$Bloco[Local] (B.[L.])$	$J(R-1)$	$SQB.[L.]$	$QMB.[L.] = \frac{SQB.[L.]}{J(R-1)}$	_____	_____
Locais	$(J-1)$	$SQLocal$	$QMLocal = \frac{SQLocal}{J-1}$	$F_{calc} = \frac{QMLocal}{QMRes}$	F_{calc}^*
Tratamentos	$(I-1)$	$SQTrat$	$QMTrat = \frac{SQTrat}{I-1}$	$F_{calc} = \frac{QMTrat}{QMRes}$	$F_{calc} = \frac{QMTrat}{QMT \times L}$
Interação $(T \times L)$	$(I-1)(J-1)$	$SQT \times L$	$QMT \times L = \frac{SQT \times L}{(I-1)(J-1)}$	$F_{calc} = \frac{QMT \times L}{QMRes}$	$F_{calc} = \frac{QMT \times L}{QMRes}$
Resíduo (Res)	$J(I-1)(R-1)$	$SQRes$	$QMRes = \frac{SQRes}{J(I-1)(R-1)}$	_____	_____
Total	$(IJR-1)$	$SQTotal$	_____	_____	_____

$* : F_{calc} = \frac{QMT_{rat} + QMRes}{QMB.[L.] + QMT \times L} \sim \mathbb{F}_{\nu_1, \nu_2}$. Em que ν_1, ν_2 são graus de liberdade encontrados pelo método de SATTERTHWAITE (1946):

$$\nu_1 = \frac{(QMT_{rat} + QMRes)^2}{\frac{(QMT_{rat})^2}{(I-1)} + \frac{(QMRes)^2}{J(I-1)(R-1)}} \quad \nu_2 = \frac{(QMB.[L.] + QMT \times L)^2}{\frac{(QMB.[L.])^2}{(J(R-1))} + \frac{(QMT \times L)^2}{(I-1)(J-1)}} \quad (2.22)$$

2.2 Transformação de Box-Cox

Tem casos em que as pressuposições da ANOVA não são satisfeitas. Uma das hipóteses para isto acontecer é que a variável resposta apresenta irregularidades como a não aditividade, heterogeneidade e não normalidade, assim BOX e COX (1964) propõe um método que encontra família de transformações a partir de um determinado parâmetro λ , e esta é aplicada na variável resposta y por meio de uma função monótona y^λ apresentada como

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \ln(y), & \text{se } \lambda = 0 \end{cases} \quad (2.23)$$

o critério para encontrar o valor para λ que melhor transforme a variável resposta é obtida pelo método de estimação da Máxima Verossimilhança, sendo que logaritmo da função de verossimilhança é dada por:

$$L(\lambda) = -\frac{n}{2} \ln(\hat{\sigma}_\lambda^2) + (\lambda - 1) \sum \ln(y) \quad (2.24)$$

sendo que $\hat{\sigma}_\lambda^2$ é a estimativa da variância para a variável transformada.

2.3 Análise de Componentes Principais (ACP)

Em muitos casos, tem-se a necessidade de compreender as características de diferentes variáveis conjuntamente, entretanto estas análises podem se tornar de difícil interpretação, caso seja utilizado as metodologias de estatística univariada ou bivariada, em virtude dos inúmeros espaços de dimensões que as interpretações podem apresentar. Para isso, a estatística utiliza fundamentos da álgebra linear com o objetivo de auxiliar na caracterização conjunta de múltiplas variáveis, reduzindo o seu espaço de dimensão sem perder informações essenciais do conjunto de variáveis originais. Uma da metodologia utilizada com está finalidade é a Análise de Componentes Principais, onde são gerados índices (novas variáveis) que são combinações lineares, não correlacionadas entre si, das variáveis originais. Estes novos índices possuem a característica de descrever a estrutura de variância-covariância conjunta das variáveis originais utilizando “poucas” dimensões (JOHNSON e WICHERN, 2007). Nesta seção são formalizadas as técnicas de ACP, Decomposição de Valores Singulares (DVS) e o método gráfico Biplot.

Antes da descrição da ACP, serão apresentados alguns conceitos algébricos básicos para a aplicabilidade da técnica multivariada. Segundo ANTON e RORRES (2012) tem-se pela Definição 2.3.1 que autovetores e autovalores são transformações lineares de uma determinada matriz ($\mathbb{X}$) com a característica de preservarem propriedades geométricas descritivas. Este valores são encontrados como apresentado no Teorema 2.3.1.

Definição 2.3.1 (ANTON e RORRES, 2012) Se $\mathbb{X}$ for uma matriz $p \times p$, então um vetor não nulo $\vec{v} \in \mathbb{R}^p$ é denominado **autovetor** de $\mathbb{X}$ se $\mathbb{X}\vec{v}$ for um múltiplo escalar de $\vec{v}$, isto é,

$$\mathbb{X}\vec{v} = \lambda\vec{v}$$

com algum escalar λ . O escalar λ é denominado **autovalor** de $\mathbb{X}$, e dizemos que $\vec{v}$ é um **autovetor associado a λ** .

Teorema 2.3.1 (Equação característica de $\mathbb{X}$) (ANTON e RORRES, 2012) Se $\mathbb{X}$ for uma matriz $p \times p$, então λ é um autovalor de $\mathbb{X}$ se, e somente se, λ satisfaz a equação

$$\det(\lambda \mathbb{I} - \mathbb{X}) = \det(\mathbb{X} - \lambda \mathbb{I}) = 0$$

Inicialmente para as diversas variáveis que serão analisadas, é necessária uma padronização nos dados, isto tem o propósito de igualar as diferentes unidades de medida que as variáveis apresentam, caso contrário, o primeiro índice gerado pela ACP pode se sobressair expressivamente em relação aos demais, devido a alta variância de algumas variáveis originais. Dadas p variáveis $(X_1, X_2, \dots, X_p)$ compostas por n repetições, é aplicada a padronização dada pela Equação 2.25, de modo que, as observações x_{ij} gerem novos valores x_{ij}^* , e estes apresentem como característica variâncias unitárias e médias nulas.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \text{ em que } i = 1, \dots, n \quad (2.25)$$

$$j = 1, \dots, p$$

temos que:

x_{ij} : observação da i -ésima repetição e j -ésima variável;

$\bar{x}_j$: média da j -ésima variável;

s_j : desvio padrão da j -ésima variável;

Por meio deste procedimento, a matriz de variância-covariância ($\Sigma_{p \times p}$) das variáveis padronizadas se iguala a matriz de correlação dos dados originais ($C_{p \times p}$), e são calculados os autovetores ($\vec{v}_j = [a_{j1}, a_{j2}, \dots, a_{jp}]^T$) com seus correspondentes autovalores (λ_j).

Os componentes principais são gerados a partir da combinação linear dos autovetores com as variáveis originais (Equação 2.26). Estas combinações são restritas, de modo que, a soma ao quadrado dos elementos do autovetor seja unitária ($a_{j1}^2 + a_{j2}^2 + \dots + a_{jp}^2 = 1$), isto faz com que nenhuma variável original apresente dependência linear com a demais evitando uma possível inflação da variância de Z_j associado à uma determinada variável X_j .

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Z_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (2.26)$$

A variância destes índices é expressada pelos seus respectivos autovalores $\lambda_j = \text{Var}[Z_j]$, com a propriedade de que os valores sejam ordenados de forma decrescentes. Este padrão faz com que, poucos índices sejam suficientes para descrever o comportamento conjunto das variáveis originais.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \equiv \text{Var}[Z_1] \geq \text{Var}[Z_2] \geq \dots \geq \text{Var}[Z_p] \geq 0$$

Propriedade dos autovalores:

1. $\sum_{j=1}^p \lambda_j = \text{traço}(\mathbb{C})$
2. $\prod_{j=1}^p \lambda_j = |\mathbb{C}|$

2.3.1 Decomposição em Valores Singulares (DVS)

Um teorema muito importante na álgebra linear é a DVS, no qual uma matriz qualquer pode ser fatorada em um produto de três matrizes $\mathbb{X}_{m \times n} = \mathbb{U}\Phi\mathbb{V}^T$, de tal forma que, $\mathbb{U}$ e $\mathbb{V}$ são matrizes ortogonais e Φ uma matriz diagonal. Muitas áreas utilizam esta metodologia afim de diminuir a quantidade de informações que uma matriz pode apresentar, como por exemplo: em processamento de dados para compressão de informações que uma imagem apresenta, afim de reduzir seu espaço de armazenamento (ANTON e RORRES, 2012), outro exemplo seria em pesquisas biológicas, utilizado para diminuir as dimensões apresentadas por um sistemas de equações que descreve comportamentos ecológicos (LEGENDRE e LEGENDRE, 2003).

Teorema 2.3.2 (Decomposição em Valores Singulares) (ANTON e RORRES, 2012) Se $\mathbb{X}$ for uma matriz $m \times n$ de posto k , então $\mathbb{X}$ pode ser fatorada como

$$\mathbb{X} = \mathbb{U}\Phi\mathbb{V}^T = \left[\begin{array}{cccc|cccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_m \end{array} \right] \left[\begin{array}{cccc|cccc} \sigma_1 & 0 & \cdots & 0 & & & & \\ 0 & \sigma_2 & \cdots & 0 & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & \\ 0 & 0 & \cdots & \sigma_k & & & & \\ \hline & & & & \mathbf{0}_{(m-k) \times k} & & & \end{array} \right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array} \right] \quad (2.27)$$

em que $\mathbb{U}$, Φ e $\mathbb{V}$ têm tamanhos $m \times m$, $m \times n$ e $n \times n$, respectivamente e satisfazem as condições:

- (i) $\mathbb{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$ diagonaliza ortogonalmente a matriz simétrica $\mathbb{X}^T\mathbb{X}$;
- (ii) As entradas diagonais não nulas de Φ são $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \dots, \sigma_k = \sqrt{\lambda_k}$, nomeadas de valores singulares de $\mathbb{X}$, sendo que $\lambda_1, \lambda_2, \dots, \lambda_k$ são os autovalores não nulos de $\mathbb{X}^T\mathbb{X}$ associados aos vetores coluna de $\mathbb{V}$;
- (iii) Os vetores coluna de $\mathbb{V}$ (autovetores de $\mathbb{X}^T\mathbb{X}$) são ordenados de tal modo que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$;
- (iv) $\mathbf{u}_i = \frac{\mathbb{X}\mathbf{v}_i}{\|\mathbb{X}\mathbf{v}_i\|} = \frac{1}{\sigma_i}\mathbb{X}\mathbf{v}_i \quad (i = 1, 2, \dots, k)$;
- (v) $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ é uma base ortonormal de $\text{col}(\mathbb{X})$;
- (vi) $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_m\}$ é uma extensão de $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ a uma base ortonormal de R^m ;

Os vetores $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ são denominados vetores singulares à esquerda de $\mathbb{X}$, e $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ os vetores singulares à direita de $\mathbb{X}$.

2.3.2 Biplot

Como descrito em componentes principais, devido as suas propriedades pode-se visualizar a estrutura multivariada de uma determinada matriz $p \times p$ utilizando poucos índices. Para facilitar a visualização e interpretação dos resultados conjuntamente, uma estrutura p dimensional pode ser representada por meio representação de 2 dimensões. Esta visualização bi-dimensional de uma estrutura multidimensional é chamada de Biplot.

Seja então uma matriz $\mathbb{X}$, com posto k ($k > 2$), como visto anteriormente pode-se escrever esta matriz a partir de um produto de três matrizes utilizando o método da DVS expresso como $\mathbb{X} = \mathbb{U}\Phi\mathbb{V}^T$, de acordo com KUNO (1971) pode-se encontrar uma aproximação da matriz original com base num produto de duas matrizes de posto 2, como visto abaixo

$$\mathbb{X} = \mathbb{U}\Phi\mathbb{V}^T = \mathbb{U}\Phi^{c/2}\Phi^{(1-c)/2}\mathbb{V}^T = \mathbb{A}\mathbb{B}^T \quad (2.28)$$

sendo que:

c : valor arbitrário de escala com valores $0 \leq c \leq 1$, em relação à matriz linha $\mathbb{U}$ ou matriz coluna $\mathbb{V}$, onde usualmente é calculado com o valor $1/2$, de modo que, particione por igual as duas matrizes;

$\mathbb{A} = \mathbb{U}\Phi^{c/2}$: representação das linhas da matriz $\mathbb{X}$ com dimensão $n \times l$;

$\mathbb{B}^T = \Phi^{(1-c)/2}\mathbb{V}^T$: representação das colunas da matriz $\mathbb{X}$ com dimensão $l \times p$, sendo que por operações de álgebra de matrizes pode-se reescrever como $\mathbb{B} = \mathbb{V}\Phi^{(1-c)/2}$;

Em seguida, são definindo quantos componentes principais são necessários para descrever a estrutura da matriz original, este número de componentes é representado pelo índice l . Tomando $l = 2$, são considerados dois componentes principais para descrever o comportamento das p variáveis originais,

assim obtêm-se as matrizes $\mathbb{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{bmatrix}$ e $\mathbb{B}^T = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \end{bmatrix}$ no qual os valores de linhas e

colunas das matrizes, são as coordenadas no plano cartesiano:

$$(x, y) = \{(a_{11}, a_{12}), (a_{21}, a_{22}), \dots, (a_{n1}, a_{n2}), (b_{11}, b_{21}), (b_{12}, b_{22}), \dots, (b_{1p}, b_{2p})\} \quad (2.29)$$

Com finalidade de ilustração, seja considerado $n = 5$, $p = 4$ e $l = 2$, pode-se observar as respectivas coordenadas a partir da Figura 2.1. II. Além disto, pode ocorrer de apenas a primeira componente principal ser significativa para a caracterização, utiliza-se a média das colunas da matriz $\mathbb{X}$ como coordenadas do eixo da abscissa (Figura 2.1. I)

Figura 2.1. Representações gráficas dos valores de linhas e colunas das matrizes $\mathbb{A}$ e $\mathbb{B}$ respectivamente.

2.4 Interação Genótipo e Ambiente (IGA)

Para estudos de melhoramento genético é importante analisar o comportamento de um conjunto de genótipos em ambientes diferentes, por conta dos fatores particulares que cada local pode apresentar devido às condições climáticas, tipos de solo, histórico do local, entre outros. Entretanto, segundo GREGORIUS e NAMKOONG (1986) há casos em que as respostas dos efeitos genotípicos e ambientais não são capazes de serem descritas separadamente, levando assim a necessidade de utilizar um termo que representa conjuntamente a contribuição destes efeitos. Esse termo, que será analisado, é chamado de interação genótipo e ambiente (IGA), e estas repostas podem ser separadas em dois tipos a interação cruzada (qualitativa) e a não cruzada (quantitativa) (KANG, 2002).

A interação cruzada acontece quando o efeito genético muda relativamente em relação aos efeitos dos ambientes, expondo assim comportamentos diferentes apresentando melhores respostas em alguns locais, porém baixas em outros (Figura 2.4.I) como por exemplo: maior peso, melhor produtividade, maior rigidez, entre outros. Este tipos de melhora tem como possibilidades as condições de adaptabilidade que os genótipos indica em determinados ambientes.

A interação não cruzada acontece quando o efeito genético permanece constante em relação aos diferentes ambientes, apresentando assim pouca variabilidade (Figura 2.4.II), esta resposta leva a uma possível estabilidade genética, representando características homogêneas de genótipos em um conjunto de ambientes. Segundo KANG *et al.* (2004) a estabilidade pode ser caracterizada em 3 tipos: Do tipo 1, nomeada como estabilidade biológica, que acontece quando a variabilidade entre os ambientes é pequena. Do tipo 2, ou chamada de estabilidade agrônômica, quando a resposta dos ambientes é paralela a média de todos os genótipos analisados. E do tipo 3, quando a regressão genotípica nos ambientes apresentam um quadrado médio residual pequeno.

Para o pesquisador, este processo que diferencia a adaptabilidade da estabilidade é necessária, dado que, um caso implica a não perda de produção em diferentes ecossistemas, enquanto o outro caso pode providenciar uma vantagem significativa em ambientes específicos (WADE *et al.*, 1999).

Figura 2.2. Gráficos representativos da adaptabilidade e estabilidade resposta de 3 tipos de genótipos em 5 ambientes distintos

Baseando-se nos modelos de ANOVA para grupos de experimentos (Equação 2.2 e 2.12), podemos então analisar o comportamento de genótipo e ambiente representado pela equação (PERKINS e JINKS, 1968) para o caso de um DIC e DBC, com $i = 1, \dots, G$ genótipos, $j = 1, \dots, E$ ambientes, e $k = 1, \dots, R$ repetições ou blocos:

$$\begin{aligned}
 \text{[DIC]} \quad Y_{ijk} &= \mu + g_i + e_j + (ge)_{ij} + \epsilon_{ijk} \\
 \text{[DBC]} \quad Y_{ijk} &= \mu + g_i + e_j + (ge)_{ij} + b_{k(j)} + \epsilon_{ijk}
 \end{aligned} \tag{2.30}$$

sendo que:

Y_{ijk} : observação do i -ésimo genótipo, j -ésimo ambiente e k -ésima repetição/bloco;

μ : constante inerente à parametrização do modelo (usualmente média geral);

g_i : efeito do i -ésimo genótipo;

e_j : efeito do j -ésimo ambiente;

$b_{k(j)}$: efeito do k -ésimo bloco dentro do j -ésimo ambiente;

$(ge)_{ij}$: efeito da interação do i -ésimo genótipo e j -ésimo ambiente;

ϵ_{ijk} : erro experimental do i -ésimo genótipo, j -ésimo ambiente e k -ésima repetição/bloco, assumindo-se

$$\epsilon_{ijk} \stackrel{IID}{\sim} N(0, \sigma^2)$$

2.5 Modelo AMMI

Segundo GAUCH (2013) o modelo AMMI é útil na análise de experimentos multi-ambientais, pois tem como propósitos: entender a estrutura de interação genótipos $\times$ ambientes e melhorar a acurácia das seleções e recomendações genéticas.

Primeiro é realizado a ANOVA, descrito pelos Modelos 2.30, sendo estimados as partes aditiva e multiplicativa dos efeitos genéticos e ambientais, em seguida a estrutura da IGA é descrita a partir

da ACP. Com o método de mínimos quadrados, tem-se que as estimativas do efeito de interação é representada por uma matriz 2.31 possuindo G linhas (genótipos) e E colunas (ambientes).

$$\mathbf{GE} = \begin{bmatrix} (\hat{ge})_{11} & (\hat{ge})_{12} & \dots & (\hat{ge})_{1E} \\ (\hat{ge})_{21} & (\hat{ge})_{22} & \dots & (\hat{ge})_{2E} \\ \vdots & \vdots & \vdots & \vdots \\ (\hat{ge})_{i1} & (\hat{ge})_{i2} & \dots & (\hat{ge})_{iE} \\ \vdots & \vdots & \vdots & \vdots \\ (\hat{ge})_{G1} & (\hat{ge})_{G2} & \dots & (\hat{ge})_{GE} \end{bmatrix} \quad (2.31)$$

Aplicado a DVS na matriz $\mathbf{GE}$, no qual $l \leq \min(G-1, E-1)$ é o número de autovalores gerados, e ordenados por $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_n^2$. Assim pode-se descrever o comportamento multi-ambiental da interação utilizando os componentes principais iniciais, da a propriedade descrita na Seção 2.3, permitindo a seleção de um modelo parcimonioso e com boa capacidade de variabilidade original, dado as restrições de ortonormalidade e ortogonalidade.

$$1. \sum_i \gamma_{ik} = 0 \quad \text{e} \quad \sum_j \delta_{jk} = 0$$

$$2. \sum_i \gamma_{ik}^2 = 1 \quad \text{e} \quad \sum_j \delta_{jk}^2 = 1$$

3. Para $k \neq k'$ temos

$$\begin{aligned} \sum_i \gamma_{i1} \gamma_{ik} &= 0 & \text{e} & \quad \sum_j \delta_{j1} \delta_{jk} = 0 \\ \sum_i \gamma_{i2} \gamma_{ik} &= 0 & \text{e} & \quad \sum_j \delta_{j2} \delta_{jk} = 0 \\ \vdots & & & \vdots \\ \sum_i \gamma_{i(k-1)} \gamma_{ik} &= 0 & \text{e} & \quad \sum_j \delta_{j(k-1)} \delta_{jk} = 0 \end{aligned}$$

Desta forma, é separada a parte determinística dada por $k = 1, \dots, K$ autovalores que contém a maior variabilidade do efeito de IGA, da parte residual (ruído) $k = K + 1, \dots, n$ autovalores com a menor variabilidade do efeito de IGA (Equação 2.32).

$$(\hat{ge})_{ij} = \underbrace{\sum_{k=1}^K \lambda_k \gamma_{ik} \delta_{jk}}_{\text{Parte determinística}} + \underbrace{\sum_{k=K+1}^n \lambda_k \gamma_{ik} \delta_{jk}}_{\text{Parte residual}} = \sum_{k=1}^K \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ij} \quad (2.32)$$

sendo que:

λ_k : raiz quadrada do k -ésimo autovalor (valor singular) da matriz $\mathbf{GE}$;

γ_{ik} : i -ésimo elemento do vetor coluna $\vec{\gamma}_k$ associado a λ_k ;

δ_{jk} : j -ésimo elemento do vetor linha $\vec{\delta}_k$ associado a λ_k ;

$$Y_{ijr} = \mu + g_i + e_j + \sum_{k=1}^K \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ge} + \epsilon_{ijr} \quad (2.33)$$

sendo que:

Y_{ijr} : resposta da r -ésima repetição, no i -ésimo genótipo e j -ésimo ambiente;

μ : constante inerente á parametrização do modelo (usualmente média geral);

g_i : efeito do i -ésimo genótipo;

e_j : efeito do j -ésimo ambiente;

λ_k : raiz quadrada do k -ésimo autovalor da matriz $\mathbf{GE}$;

γ_{ik} : i -ésimo elemento do vetor coluna $\vec{\gamma}_k$ associado a λ_k ;

δ_{jk} : j -ésimo elemento do vetor linha $\vec{\delta}_k$ associado a λ_k ;

ρ_{ij} : ruído da parte multiplicativa;

ϵ_{ijr} : erro médio experimental do i -ésimo genótipo, j -ésimo ambiente e r -ésima repetição, em que assume-se $\epsilon_{ger} \stackrel{IID}{\sim} \mathbb{N}(0, \sigma^2)$

2.5.1 Critério de seleção da família de modelos AMMI

É necessário encontrar o número de termos do modelos AMMI $AMMI_k$, $k = 1, \dots, K$, que são significativos na parte determinística. PIEPHO (1995), faz um estudo de simulação utilizando 4 critérios para seleção das famílias de modelos AMMI (Critério de Gollob, F_{GH1} , F_{GH2} e F_R), e conclui que o teste F_R é melhor e mais robusto para casos de não normalidade e de heterocedasticidade dos dados.

Graus de liberdade dado por GOLLOB (1968), é um dos métodos mais utilizado, por conta da simplicidade de cálculo, sendo a diferença entre o número de parâmetros estimados em relação ao número de restrições da DVS. Para o cálculo dos $AMMI_k$ são utilizados G elementos do vetor colunas $\vec{\gamma}_k$, e E elementos do vetor linhas $\vec{\delta}_k$, ambos associados à 1 valor singular de λ_k , levando à $(G + E + 1)$. Para tal, foram necessárias as restrições de ortonormalidade e ortogonalidade, chegando à $(2 + 2 + 2(k - 1))$, com isto temos:

$$GL_{AMMI_k} = (G + E + 1) - (2 + 2 + 2(k - 1)) = (G + E + 1) - (2 + 2k) = G + E - 1 - 2k \quad (2.34)$$

2.6 AMMI ponderado (W-AMMI)

O modelo AMMI estima os parâmetros pelo método de mínimos quadrados e estrutura o efeito multiplicativo a partir do método de DVS, levando em consideração todas as pressuposições da ANOVA, como a normalidade, independência e variância constante dos erros.

Os delineamentos realizados em grupos de experimentos são conduzidos em diferentes locais com características biológicas e climatológicas particulares, podendo gerar valores amostrais heterogêneos resultando em heterocedasticidade dos erros e superestimação dos valores explicativos da IGA (EDWARDS e JANNINK, 2006).

Para estes casos, RODRIGUES *et al.* (2014) propõem uma generalização do modelo AMMI, o W-AMMI (*Weighted AMMI*), em que são ajustados os efeitos aditivos utilizando o método de mínimos quadrados ponderados e o efeito multiplicativo com o método de decomposição em valor singular ponderado proposto por SREBRO e JAAKKOLA (2003), ambos a partir de uma matriz $\mathbb{W}$, que leva em consideração os números de repetições e os componentes de variância, encontrados em cada local.

A seguir serão descritos os passos necessários para o cálculo do modelo W-AMMI:

Passo 1: Estimar os componentes de variância utilizando o modelo ANOVA para cada ambiente individualmente σ_j^2 , em que $j = 1, \dots, J$ (locais). Os modelos são decorrentes ao delineamento experimental utilizado.

Passo 2: Construir a matriz $\mathbb{W}$ de ponderação, constituindo-se por $i = 1, \dots, I$ como os valores

dos genótipos e $j = 1, \dots, J$ os valores dos ambientes, com dimensão $(I \times J)$

$$\mathbb{W} = \begin{bmatrix} \frac{1/\sigma_{e_1}^2}{m_{\sigma_e^2}} & \frac{1/\sigma_{e_2}^2}{m_{\sigma_e^2}} & \dots & \frac{1/\sigma_{e_J}^2}{m_{\sigma_e^2}} \\ \frac{1/\sigma_{e_1}^2}{m_{\sigma_e^2}} & \frac{1/\sigma_{e_2}^2}{m_{\sigma_e^2}} & \dots & \frac{1/\sigma_{e_J}^2}{m_{\sigma_e^2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1/\sigma_{e_1}^2}{m_{\sigma_e^2}} & \frac{1/\sigma_{e_2}^2}{m_{\sigma_e^2}} & \dots & \frac{1/\sigma_{e_J}^2}{m_{\sigma_e^2}} \end{bmatrix} \circ \begin{bmatrix} \frac{n_{rep1,1}}{m_n} & \frac{n_{rep1,2}}{m_n} & \dots & \frac{n_{rep1,J}}{m_n} \\ \frac{n_{rep2,1}}{m_n} & \frac{n_{rep2,2}}{m_n} & \dots & \frac{n_{rep2,J}}{m_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{repI,1}}{m_n} & \frac{n_{repI,2}}{m_n} & \dots & \frac{n_{repI,J}}{m_n} \end{bmatrix} \quad (2.35)$$

sendo que:

$\sigma_{e_j}^2$: é o componente de variância do erro calculado para o j -ésimo ambiente;

$m_{\sigma_e^2}$: valor máximo da expressão $1/\sigma_{e_j}^2$ ($\max(1/\sigma_{e_j}^2)$), considerando o j -ésimo ambiente;

$n_{rep_{i,j}}$: é o número de repetição para o i -ésimo genótipos no j -ésimo ambiente;

m_n : valor máximo de repetição no conjunto de dados ($\max(n_{rep_{i,j}})$), considerando os i -ésimos genótipos e j -ésimos ambientes;

Passo 3: Calcular a estimativa dos parâmetros a partir da matriz $\mathbb{W}$, por meio do método de mínimos quadrados ponderado, utilizando o modelo que contém os fatores aditivos e multiplicativos, para assim encontrar a matriz de IGA.

Passo 4: Aplicar a DVS ponderada (*Weighted low-rank SVD*) calculado por método iterativo, como apresentado na Figura 2.6.

Figura 2.3. Fluxograma do método de DVS ponderada

sendo:

W : matriz de ponderação ($I \times J$) apresentando valores $0 \leq W_{i,j} \leq 1$;

$\mathbb{1}$: matriz de 1's ($I \times J$);

Z : matriz alvo do método de DVS ponderado, que para o caso do modelo AMMI a GE;

$\circ$: produto de Hadamard;

t : número de iterações;

X : uma aproximação de baixo posto, com $rank(X) = N$, pode-se inicializar $X^{(0)} = Z$ ou $X^{(0)} = 0$.

Observa-se que o critério de parada do processo iterativo é obtido quando

$$\sum_{j=1}^J \sum_{i=1}^I [(x_{ij}^{t+1} - x_{ij}^t)^2] \rightarrow 0$$

que é uma modificação da abordagem proposta por EASTMENT e KRZANOWSKI (1982) que estima a quantidade de autovalores ótimos para ser utilizado em uma análise de componente principais. No final temos novas matrizes U_N , D_N e V_N que descrevem o comportamento ponderado da matriz alvo (Z), por meio da nova matriz $Z^* \approx U_N D_N V_N$.

Desta forma, a partir da Z^* é realizados novamente os cálculo para o critério de seleção dos modelos AMMI, e encontrando quantos biplot's significativos são necessário para a descrever a IGA.

2.7 Imputação Múltipla

O método de imputação múltipla desenvolvido por KRZANOWSKI (1988); BERGAMO (2007), é utilizado para dados em que haja necessidade de avaliar dois fatores que são arranjos de forma matricial. Além disso, não tem-se a necessidade de supor uma distribuição na variável resposta. O método consiste na utilização do teorema da DVS, como já citado, em uma matriz qualquer X com $dim(n, p)$ que pode ser decomposta, da seguinte forma

$$X_{(n,p)} = U_{(n,p)} \Phi_{(p,p)} V_{(p,p)}^T \quad (2.36)$$

Suponha-se então uma matriz X_{miss} com m valores ausentes denotados por $(x_{miss}^{[1]}, x_{miss}^{[2]}, \dots, x_{miss}^{[m]})$ em que cada valor tem sua respectiva posição representada por $i_{miss}^{[k]}$ nas linhas, $j_{miss}^{[k]}$ nas colunas e $[1], [2], \dots, [m]$ índice sequencial do valor ausente.

$$X_{miss} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ x_{miss}^{[1]} & x_{22} & x_{23} & x_{24} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{miss}^{[2]} & x_{34} & \dots & x_{3p} \\ x_{41} & x_{42} & x_{43} & x_{44} & \dots & x_{miss}^{[m]} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.37)$$

No primeiro processo iterativo da imputação múltipla, é calculado os valores médios $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$ e desvios padrão $(s_1, s_2, \dots, s_p)$ em relação às colunas da matriz X_{miss} desconsiderando os valores ausentes. Os valores médios são utilizados como uma imputação inicial para os valores ausentes como mostrado na matriz (2.38) abaixo, também mais à frente esses dois cálculo serão utilizados para a “despadronização”

da matriz.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ \bar{x}_1 & x_{22} & x_{23} & x_{24} & \dots & x_{2p} \\ x_{31} & x_{32} & \bar{x}_2 & x_{34} & \dots & x_{3p} \\ x_{41} & x_{42} & x_{43} & x_{44} & \dots & \bar{x}_p \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.38)$$

O método de imputação utilizando DVS será realizado de modo individual para cada valor, com isto, inicialmente tem-se o primeiro valor ausente $x_{miss}^{[1]}$. Desta forma utilizando a matriz $\mathbb{X}_{miss}$ contendo os valores médios, são geradas duas matrizes (2.39) sendo que:

$\mathbb{X}_{-i_{miss}}$: matriz em que é retirada a linha $i_{miss}^{[1]}$ que está localizada o valor ausente;

$\mathbb{X}_{-j_{miss}}$: matriz em que é retirada a coluna $j_{miss}^{[1]}$ que está localiza da o valor ausente;

$$\mathbb{X}_{-i_{mis}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ x_{31} & x_{32} & \bar{x}_2 & x_{34} & \dots & x_{3p} \\ x_{41} & x_{42} & x_{43} & x_{44} & \dots & \bar{x}_p \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \mathbb{X}_{-j_{mis}} = \begin{bmatrix} x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ x_{22} & x_{23} & x_{24} & \dots & x_{2p} \\ x_{32} & \bar{x}_2 & x_{34} & \dots & x_{3p} \\ x_{42} & x_{43} & x_{44} & \dots & \bar{x}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.39)$$

Com estas matrizes é realizada uma padronização nas observações utilizando as médias e desvios padrões calculadas inicialmente desconsiderando os valores ausentes. Reproduzindo novas observações x_{ij}^* em que as colunas tenham variância unitária e média nula.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \text{ em que } i = 1, \dots, n \quad (2.40)$$

$$j = 1, \dots, p$$

temos que:

x_{ij} : observação da i -ésima linha e j -coluna;

$\bar{x}_j$: média da j -ésima coluna;

s_j : desvio padrão da j -ésima coluna;

Em seguida é realizada a DVS para cada uma dessas matrizes, gerando seis matrizes,

$$\begin{aligned} \mathbb{X}_{-i_{miss}}^* &= \mathbb{U}_{-i_{miss}} \mathbf{\Phi}_{-i_{miss}} \mathbb{V}_{-i_{miss}}^T, \text{ com as respectivas dimensões:} \\ \dim(n-1, p) &= \dim(n-1, p) \dim(p, p) \dim(p, p) \\ \mathbb{X}_{-j_{miss}}^* &= \mathbb{U}_{-j_{miss}} \mathbf{\Phi}_{-j_{miss}} \mathbb{V}_{-j_{miss}}^T, \text{ com as respectivas dimensões:} \\ \dim(n, p-1) &= \dim(n, p-1) \dim(p-1, p-1) \dim(p-1, p-1) \end{aligned} \quad (2.41)$$

sendo que $\dim$ é dimensão da respectiva matriz, dada por seu número de linhas e colunas;

A partir disto, pode-se encontrar uma nova matriz $\mathbb{Y}_{imp_1}^*$ em que o valor ausente é calculado utilizando o teorema da D.V.S.:

$$\mathbb{Y}_{imp_1}^* = \mathbb{U}_{-j_{miss}} \mathbf{\Phi}_{-j_{miss}}^{(\frac{a}{b})} \mathbf{\Phi}_{-i_{miss}}^{(\frac{b-a}{b})} \mathbb{V}_{-i_{miss}}^T \quad (2.42)$$

sendo que:

imp_{u_1} : termo que denota o primeiro processo iterativo da imputação, nota-se que o número de coeficientes fica dependente da quantidade de imputações que serão realizados;

a, b : valores arbitrários para ponderar o expoente da multiplicação entre as matrizes diagonais de valores singulares. BERGAMO (2007) utiliza cinco processos iterativos de imputação, com valores $a = 8, 9, 10, 11, 12$ e $b = 20$;

Suponha um produto de matrizes $(\mathbb{A}\mathbb{B})$, o número de colunas da primeira matriz $(\mathbb{A})$ deve ser igual ao número de linhas da segunda matriz $(\mathbb{B})$, entretanto pode-se observar que para o cálculo da Equação 2.42 as matrizes possuem dimensões $dim(n, p-1)$, $dim(p-1, p-1)$, $dim(p, p)$ e $dim(p, p)$ respectivamente. Para isto, na matriz diagonal $\Phi_{-i_{miss}}$ descarta-se o último autovalor apresentando assim $dim(p-1, p-1)$, e por consequência para matriz $\mathbb{V}_{-i_{miss}}^T$ é retirada a última coluna ficando com $dim(p-1, p)$.

O próximo passo é realizar a “despadronização” utilizando os valores calculados de média e desvios padrão, $y_{ij} = y_{ij}^* \cdot s_j + \bar{x}_j$, gerando a matriz $\mathbb{Y}_{imp_{u_1}}^{[1]}$ que apresenta o valor $y_{imp_{u_1}}^{[1]}$ localizado na $i_{miss}^{[1]}$ -ésima linha e $j_{miss}^{[1]}$ -ésima coluna da matriz.

Com isto, é realizado reciprocamente o primeiro passo iterativo para os demais valores $(x_{miss}^{[2]}, \dots, x_{miss}^{[m]})$, de forma que, tem-se um vetor $y_{imp_{u_1}} = (y_{imp_{u_1}}^{[1]}, y_{imp_{u_1}}^{[2]}, \dots, y_{imp_{u_1}}^{[m]})$ e estes são utilizados para substituir na matriz $\mathbb{X}_{miss}$ (2.43), como visto abaixo.

$$\mathbb{X}_{miss} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ y_{imp_{u_1}}^{[1]} & x_{22} & x_{23} & x_{24} & \dots & x_{2p} \\ x_{31} & x_{32} & y_{imp_{u_1}}^{[2]} & x_{34} & \dots & x_{3p} \\ x_{41} & x_{42} & x_{43} & x_{44} & \dots & y_{imp_{u_1}}^{[m]} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.43)$$

Calculado todos os valores do primeiro processo de imputação, segue-se para o segundo passo do processo iterativo, em que novamente é calculado a média e desvio padrão em relação à coluna, porém agora considerando a matriz $\mathbb{X}_{miss}$ composta pelos valores do vetor $y_{imp_{u_1}}$. Em seguida é realizando novamente todos os passos da imputação gerando assim, novos vetores $y_{imp_{u_2}} = (y_{imp_{u_2}}^{[1]}, y_{imp_{u_2}}^{[2]}, \dots, y_{imp_{u_2}}^{[m]})$, agora para a segunda imputação.

Neste passo os m -ésimo valores ausente da matrix $\mathbb{X}_{miss}$ serão substituídos pela médias dos 2 processos iterativo $(y_{imp_{u_1}}, y_{imp_{u_2}})$

Com isto, ao fim dos M processo iterativos de imputação, obtemos o vetor $\hat{y}_{imp_u}$ que é dado pelos valores médios de cada m -ésimo índice. Sendo utilizados para substituir os valores ausentes da matriz $\mathbb{X}_{miss}$ como indicado abaixo

$$\begin{aligned} \hat{y}_{imp_u}^{[1]} &= \sum_{k=1}^M \frac{y_{imp_{u_k}}^{[1]}}{k} \\ \hat{y}_{imp_u}^{[2]} &= \sum_{k=1}^M \frac{y_{imp_{u_k}}^{[2]}}{k} \\ &\vdots \\ \hat{y}_{imp_u}^{[m]} &= \sum_{k=1}^M \frac{y_{imp_{u_k}}^{[m]}}{k} \end{aligned} \quad \mathbb{X}_{miss} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \dots & x_{1p} \\ \hat{y}_{imp_u}^{[1]} & x_{22} & x_{23} & x_{24} & \dots & x_{2p} \\ x_{31} & x_{32} & \hat{y}_{imp_u}^{[2]} & x_{34} & \dots & x_{3p} \\ x_{41} & x_{42} & x_{43} & x_{44} & \dots & \hat{y}_{imp_u}^{[m]} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (2.44)$$

3 MATERIAL E MÉTODOS

3.1 Descrição da 1ª análise - Dados IAC

3.1.1 O experimento

O conjunto de dados foi proveniente de um Projeto Financiado pela FAPESP de número 2012/00292-9, o qual foi realizado no Instituto Agronômico de Campinas (IAC), e teve como objetivo verificar o desempenho do híbrido de uva branca (SR 0.501-17), sobre 4 diferentes porta-enxertos (IAC 766, IAC 572, IAC 571-6 e IAC 313) em 2 municípios do estado de São Paulo (Votuporanga e Jundiaí), ambos nos anos de 2012, 2013 e 2014.

Destinado para vinificação, o híbrido SR 0.501-17 foi desenvolvido na Estação Experimental de São Roque, pelo pesquisador Wilson Corrêa Ribas, e tem ótimas características produtivas e enológicas, possuindo bagas pequenas, redondas e com sementes, com um sabor suave moscatel e ciclo de maturação precoce.

3.1.1.1 Os porta-enxertos

Os porta-enxertos foram desenvolvidos no IAC, e possuem as seguintes características:

- IAC 766 “Campinas” proveniente do cruzamento entre “Ripária do Traviú” com a espécie de videira tropical *Vitis caribaea*. Muito utilizado em São Paulo e norte de Paraná, sendo vigoroso apresentando perfeita adaptação às condições ambientais do Sudeste do Brasil (DE SOUSA e MARTINS, 2002).
- IAC 572 “Jales” proveniente do cruzamento da *Vitis caribaea* e a variedade de porta-enxerto Riparia x Rupestris 101-14. Sendo muito utilizado nas regiões Oeste e Noroeste do estado de São Paulo (DE SOUSA e MARTINS, 2002).
- IAC 571-6 “Jundiaí” proveniente do cruzamento de *Vitis caribaea* e *Pirovano 57*, mesmo sendo pouco utilizado, apresenta ótimo vigor e adaptação às regiões mais quentes, tanto em solos argilosos com arenosos. Suas estacas enraízam com facilidade e tem demonstrado boa tolerância a pérola-da-terra (DE SOUSA e MARTINS, 2002).
- IAC 313 “Tropical” proveniente do cruzamento entre o porta-enxerto “Golia” e a espécie de videira tropical *V. cinerea*. Apresenta grande vigor vegetativo, e adapta-se às condições sub-tropicais e tropicais do Brasil, apresenta um bom comportamento em áreas infestadas por nematoides (DE SOUSA e MARTINS, 2002).

3.1.1.2 Os locais

- Três experimentos foram desenvolvidos em Votuporanga, no Centro de Seringueira e Sistemas Agroflorestais do Instituto Agronômico com latitude e longitude de 20°15’S e 50°30’O e altitude de 483m. Segundo EMBRAPA (2013), o solo é classificado como Argissolo Vermelho Amarelo. O município apresenta temperatura média mensal de 23.6°C, precipitação pluvial anual de 1.312mm e estações chuvosas de Dezembro a Março. Conhecido o mínimo risco de geadas durante o período de Junho a Novembro e a baixa precipitação pluvial no meses de Agosto e Setembro, então apresenta condições favoráveis para o processo de produção e maturação da uva. As datas de poda foram em 05/07/2012, 18/07/2013 e 17/07/2014, e a colheita em 14/11/2012, 12/12/2013 e 26/11/2014.
- Outros três experimentos foram realizados em Jundiaí, sendo localizado na região Leste do estado de São Paulo, o experimento foi realizado no Centro APTA de Frutas, sendo 23°17’S e 46°9’O,

com altitude de 700 a 900m. Solo classificado, de acordo com EMBRAPA (2013), como Cambissolo Vermelho Distrófico. Apresentando uma precipitação pluvial anual de 1.4 mm, temperatura média de 19.5°C e umidade relativa do ar de 70,6%. O inverno é ameno, porém sujeito a ocorrência de geadas, com baixa precipitação pluvial. Por ter um verão quente e seco, pode ocasionar doenças fúngicas como antracnose, míldio, mancha das folhas, ferrugem, oídio e podridão dos cachos. As podas ocorreram nos dias 02/08/2012, 01/08/2013 e 22/07/2014, e a colheita em 12/12/2012, 30/12/2013 e 18/12/2014.

3.1.1.3 A análise

Foram considerados 6 ambientes, vinda das combinações dos 2 municípios, Votuporanga e Jundiaí, em relação aos 3 anos que ocorreram os experimentos, onde foram realizados por meio do delineamento ao acaso em bloco, sendo 4 blocos em Votuporanga e 6 blocos em Jundiaí. As análises utilizaram a característica de produção de cachos, dados em quilos por hectare (kg/ha) obtida pela massa total dos cachos de cada planta no momento da colheita, com o propósito de caracterizar o desempenho do híbrido SR 0.501-17 sobre diferentes porta-enxertos em dois ambientes.

3.2 Descrição da 2ª análise - Dados EMBRAPA

O experimento foi efetuado pela EMBRAPA - Fortaleza, afim de caracterizar a estrutura do melão com 92 famílias (genótipos), realizado em 3 diferentes ambientes. Por conta de fatores externos não controlados, ocorreu a ausência da germinação de 3 famílias: 52, 55 e 79, desta maneira foram desconsiderados na análise. Em razão das condições locais serem homogêneas o experimento foi estruturado por meio de um DIC com 3 repetições. Certas atenções devem ser tomada pois em determinados ambientes, são apresentados quantidades menores de repetições ou até valores ausentes. Para visualizar como estão os dados, foi apresentado na Figura 3.1 os números de repetições (escalas de cores de cinza) em cada família (eixo y) no diferentes ambientes (eixo x), no qual observa-se que as famílias 86 no ambiente 1; as famílias 5, 70 no ambiente 2 e as famílias 31, 47, 92 e 95 no ambiente 3 apresentam dados faltantes, que na Figura 3.1 são representados por quadrados em brancos.

Figura 3.1. Mapa de calor para contagens de números de repetições de cada família em cada ambiente

4 RESULTADOS E DISCUSSÕES DA 1ª ANÁLISE - DADOS IAC

4.1 Análises iniciais

4.1.1 Para a variável

Antes de considerar os fatores de genótipos e ambientes, foram realizadas avaliações iniciais por meio dos gráficos indicado na Figura 4.1 I e II em relação a característica de produção (kg/ha), visto que foi utilizada como variável resposta para as análises estatísticas. Dentro do *box plot* (Figura 4.1. I) foram indicadas as dispersões das observações por meio das enumerações, e na legenda os valores das medidas de posição (mediana e média), e de dispersão (variância e desvio padrão). Observou-se um comportamento assimétrico à direita, sendo a média (quadrado cinza) maior que a mediana (linha vertical preta). Algumas observações ultrapassaram o limite superior considerado os possíveis valores atípicos.

Figura 4.1. *Box plot* (I) e Histograma (II) para a variável de produção (kg/ha).

4.1.2 Fatores individuais

Em seguida foram apresentados *box plot's* (Figura 4.2) e gráficos univariados (Figura 4.3 e 4.4) ambos para a variável de produção em relação aos fatores individuais de porta-enxertos e anos por locais. Nas Figuras 4.3 e 4.4, as medidas de média (I) e desvio-padrão (II) dos níveis de cada fator são representadas por linhas horizontais menores, e o valor geral foi indicado por uma linha horizontal maior.

Observa-se nos *box plot's* para porta-enxerto (Figura 4.2.I) que todos os níveis tendem a uma assimetria à direita com alguns pontos atípicos. Nota-se no porta-enxerto IAC 766, que a observação 69 se apresentou afastado das demais, e isto acabou influenciando em um alto valor de desvio-padrão (Figura 4.3. II) em relação aos demais. Os valores médios do fator porta-enxertos aparecem próximos de um intervalo de 2500 a 3200 (Figura 4.3. I).

Figura 4.2. Box plot dos níveis de porta-enxertos (I) e anos por locais (II)

Figura 4.3. Gráficos univariados do valor médio (I) e desvio-padrão (II) em relação a porta-enxertos

Em relação ao fator de anos por locais, na Figura 4.2.II o ambiente de 2012 Jundiaí foi o único que apresentou característica de uma distribuição assimétrica à esquerda evidenciando que alguns valores se apresentam superiores a média. O ambiente 2012 Votuporanga indicou uma alta dispersão para os dados (Figura 4.2.II), e grandes valores de média e de desvio-padrão (Figura 4.4). Nota-se que para os dois primeiros anos (2012 e 2013), a cidade de Votuporanga apresentou maiores valores de média e desvio-padrão, em relação à Jundiaí (Figura 4.4. I), no entanto em 2014 houve uma queda nos valores

de produção deixando-o próximo de Jundiaí (Figura 4.2. II).

Figura 4.4. Gráficos univariados do valor médio (I) e desvio-padrão (II) em relação a anos por locais

4.1.3 Interação entre os fatores

Foram realizados estudos em relação a interação entre os fatores, utilizando uma tabela com os valores de média e desvio-padrão (Tabela 4.1) em relação à cada nível de porta-enxerto dentro de cada nível do fator anos por locais, totalizando 24 *box plot*'s em que no eixo y são os níveis de anos por locais, e as cores são os níveis de porta-enxertos (Figura 4.5), e um gráfico (Figura 4.6) no qual foram calculados os valores médios dos níveis de porta-enxertos (retas 1, 2, 3, 4) dentro de cada nível de anos por locais (eixo x), e para os níveis de anos por locais (retas 1, 2, 3, 4, 5, 6) dentro de cada nível de porta-enxertos (eixo x).

Na Figura 4.5 pode-se observar que o ambiente 2012 Votuporanga apresentou maiores valores médios dentro dos porta-enxertos IAC 313, IAC 571-6 e IAC 766 em relação ao IAC 572 (Tabela 4.1). Também são indicados na Figura 4.5 e Tabela 4.1 que em 2014, ambas as cidades apresentam comportamentos próximos, tanto em relação as médias e desvios-padrão dos porta-enxertos.

Na Figura 4.6.I pode-se observar que os maiores valores são encontrados em Votuporanga 2012 e 2013 representados por um “pico” ao longo das retas (níveis de porta-enxerto), e que depois tornam-se estável como pode-se observa em 2014 Votuporanga. Além disto, na Figura 4.6. II as cores das retas foram definidas, de modo que, possam ser identificadas por locais diferentes (Vermelho - Votuporanga e Preto - Jundiaí). Notou-se um comportamento diferente apenas para a reta de 2012 Jundiaí, que cresceu para o porta-enxerto IAC 572, enquanto para os demais ambientes diminuíram.

Tabela 4.1. Tabela para valores médios e desvios-padrões (kg/ha), de porta-enxerto dentro de anos por locais

Porta-Enxerto	Ano Local	Média	Desvio-padrão
IAC 313	2012 Jundiai	2829,81	184,71
IAC 571-6	2012 Jundiai	2001,78	452,07
IAC 572	2012 Jundiai	2509,24	334,17
IAC 766	2012 Jundiai	1798,61	778,80
IAC 313	2012 Votuporanga	5265,51	1657,20
IAC 571-6	2012 Votuporanga	5514,23	1746,94
IAC 572	2012 Votuporanga	3954,89	1531,38
IAC 766	2012 Votuporanga	6663,15	2440,81
IAC 313	2013 Jundiai	1358,25	603,00
IAC 571-6	2013 Jundiai	1196,50	548,79
IAC 572	2013 Jundiai	783,00	172,21
IAC 766	2013 Jundiai	891,25	234,33
IAC 313	2013 Votuporanga	3734,05	1128,80
IAC 571-6	2013 Votuporanga	3814,73	1092,29
IAC 572	2013 Votuporanga	3505,19	792,03
IAC 766	2013 Votuporanga	3823,92	2003,40
IAC 313	2014 Jundiai	2349,17	681,47
IAC 571-6	2014 Jundiai	2109,04	774,59
IAC 572	2014 Jundiai	1792,19	183,92
IAC 766	2014 Jundiai	2047,94	601,85
IAC 313	2014 Votuporanga	2345,84	764,63
IAC 571-6	2014 Votuporanga	2471,42	791,55
IAC 572	2014 Votuporanga	1592,04	543,61
IAC 766	2014 Votuporanga	2275,21	645,s4

Figura 4.5. Box plot da interação entre porta-enxertos e anos por locais

Figura 4.6. Gráfico de interação entre os nível de um fator dentro dos níveis do outro fator, em relação aos valores médios de produção (kg/ha)

4.2 Análise para grupos de experimentos

Em seguida foi realizada ANOVA individual (Equação 2.11) levando em consideração cada ambiente, com os resultados indicados na Tabela 4.2 notou-se que, apenas para os níveis de porta-enxertos em 2012 Jundiaí se apresentaram estatisticamente diferentes, com um valor p aproximadamente de 0,0020. Os valores de quadrado médio residual para cada ANOVA individual podem ser visualizadas por meio do gráfico de barras (Figura 4.7), no qual observou-se que 2012_Votuporanga indicou um alto valor em relação aos demais, sendo uma suspeita aos altos valores de desvio-padrão que foram encontrados neste local (Figura 4.4).

O próximo passo foi realizar a ANOVA conjunta (Tabela 4.3) considerando os Ano Local como um fator no modelo (Equação 2.12), sendo que, foi concluído pelo teste F que o fator da interação não foi significativa. Apenas para o efeito de Ano Local mostrou-se significativo, evidenciando uma possível diferença entre os contrastes dos níveis deste fator.

Tabela 4.2. Quadro da ANOVA individual para cada ambiente

Fontes de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F_{cal}	$valor-p$
ANOVA individual para: 2012 Jundiaí					
Blocos	3	2.170.470,45	723.490,15	9,31	0,0040**
Porta-Enxerto	3	2.655.556,03	885.185,34	11,39	0,0020**
Resíduo	9	699.594,89	77.732,77		
ANOVA individual para: 2012 Votuporanga					
Blocos	5	5.909.255,35	1.181.851,07	0,27	0,9200 ^{ns}
Porta-Enxerto	3	22.228.784,15	7.409.594,72	1,72	0,2056 ^{ns}
Resíduo	15	64.594.672,04	4.306.311,47		
ANOVA individual para: 2013 Jundiaí					
Blocos	3	346.264,50	115.421,50	0,55	0,6630 ^{ns}
Porta-Enxerto	3	851.042,50	283.680,83	1,34	0,3208 ^{ns}
Resíduo	9	1.901.768,00	211.307,56		
ANOVA individual para: 2013 Votuporanga					
Blocos	5	20.226.785,23	4.045.357,05	3,96	0,0172*
Porta-Enxerto	3	396.679,15	132.226,38	0,13	0,9411 ^{ns}
Resíduo	15	15.314.310,80	1.020.954,05		
ANOVA individual para: 2014 Jundiaí					
Blocos	3	339.931,65	113.310,55	0,25	0,8578 ^{ns}
Porta-Enxerto	3	628.165,22	209.388,41	0,47	0,7129 ^{ns}
Resíduo	9	4.041.392,58	449.043,62		
ANOVA individual para: 2014 Votuporanga					
Blocos	5	3.578.537,09	715.707,42	1,78	0,1780 ^{ns}
Porta-Enxerto	3	2.801.291,50	933.763,83	2,32	0,1167 ^{ns}
Resíduo	15	6.037.390,02	402.492,67

Figura 4.7. Gráfico de barras dos valores de quadrado médio residual para cada ambiente

Tabela 4.3. Quadro da ANOVA conjunta

Fontes de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	$valor-p$
Bloco[Ano Local]	24	32.571.244,26	——	——	——
Porta-Enxerto	3	9.496.402,06	3.165.467,35	2,46	0,07 ^{ns}
Ano Local	5	243.557.177,84	48.711.435,57	37,88	<0,00**
Porta-Enxerto×Ano Local	15	20.065.116,49	1.337.674,43	1,04	0,43 ^{ns}
Resíduo	72	92.589.128,33	1.285.960,12		

Previamente foram realizadas análises gráficas do comportamento dos resíduos estudentizados encontrados pelo modelo da ANOVA conjunta, por meio do: *box plot* (Figura 4.8. I), histograma (Figura 4.8. II) e o gráfico Q-Q normal (Figura 4.8. III) que utiliza os quantis observados pelos quantis teóricos da distribuição normal. Nos 3 gráficos, nota-se a presença de um prolongam na calda da distribuição dos resíduos, indicando a presença de alguns valores atípicos nos dados.

Outros gráficos utilizados na literatura para verificar o ajuste do modelo foram construídos com os valores ajustados pelo resíduos estudentizados (Figura 4.9. I), e das sequências de ocasiões dos dados pelo resíduos estudentizados (Figura 4.9. II). Na Figura 4.9. I, nota-se um afinilamento à esquerda, indicando que o ajuste do modelo não foi o indicado para a descrição dos dados, visto que algumas observações apresentam valores altos dos resíduos. Na Figura 4.9. II observou-se que estes mesmos valores se afastam do conjunto de observações.

Com o auxílio do software R (R CORE TEAM, 2017), foram utilizados 2 testes estatísticos para os resíduos estudentizados do modelo da ANOVA conjunta. O teste de Shapiro-Wilk (ROYSTON, 1982), que tem como hipótese nula (H_0) à normalidade dos resíduos, e que no presente trabalho apresentou um valor p aproximado de 0, e o teste de Breusch-Pagan (BREUSCH e PAGAN, 1979) com a hipótese nula (H_0) de homogeneidade das variâncias dos resíduos, e também apresentou um valor p aproximado de 0,01. Desta forma, conjuntamente com as análises gráficas e os testes, indicaram que o modelo da ANOVA conjunta não foi bom para descrever o comportamento dos dados, uma suspeita seria por conta da presença de alguns valores atípicos.

Figura 4.8. Box plot (I), histograma (II) e Q-Q normal (III) dos resíduos estudantizados

Figura 4.9. Gráficos de dispersão dos valores ajustados (I) e das ocasiões (II), pelos resíduos estudantizados

4.2.1 Transformação da variável resposta

Em seguida foi realizada uma transformação na variável resposta (produção), utilizando o método de Box-cox (BOX e COX, 1964). Conforme o gráfico gerado (Figura 4.10. I) obteve-se o valor de $\lambda = 0,16$ que indica que as pressuposições da ANOVA para o modelo conjunto não foram satisfeitas. Desta forma, aplicou-se a função monótona (Equação 2.23) na variável resposta, e foi gerado novamente o

gráfico (Figura 4.10. II) no qual o λ apresentou valor 1, supondo que as pressuposições do modelo foram atendidas para a variável resposta transformada.

Figura 4.10. Gráficos entre o valor λ em relação ao Log da função de verossimilhança antes (I) e depois (II) da transformação

Para a ANOVA conjunta com transformação da variável resposta (Tabela 4.4), tem-se que além do fator anos por locais ser significativo, agora o fator de porta-enxertos também foi estatisticamente significativo à um nível de 0,05. Além disto, pode-se observar que o comportamento dos resíduos estudentizados (Figura 4.11) após a transformação se apresentaram aceitáveis segundo as pressuposições, principalmente os possíveis valores atípicos que se aproximaram à nuvem de pontos (Figura 4.12). Indicando que o modelo de ANOVA conjunta após a transformação foi bem ajustado aos dados.

Tabela 4.4. Quadro da ANOVA para grupos de experimentos, com a transformação da variável resposta

Fontes de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p
Bloco[Ano Local]	24	45,77	1,91	—	—
Porta-Enxerto	3	13,38	4,46	3,10	0,03**
Ano Local	5	373,42	74,68	51,88	<0,00**
Porta-Enxerto $\times$ Ano Local	15	22,48	1,50	1,04	0,43 ^{ns}
Resíduo	72	103,65	1,44

Figura 4.11. Box plot (I), histograma (II) e Q-Q normal (III) dos resíduos estudentizados com a variável reposta transformada

Figura 4.12. Gráficos de dispersão dos valores ajustados (I) e das ocasiões (II), pelos resíduos estudentizados com a variável reposta transformada

O teste de Shapiro-Wilk para os resíduos do modelo completo indicou uma possível adequação dos dados por apresentar um valor de aproximadamente 0,37, e com isto, não rejeita-se a hipótese de nulidade, considerando assim, que os resíduos tendem a um comportamento similar a distribuição normal. Entretanto, o teste de Breuch-Pagan ainda continua sendo significativo, com valor p aproximado para 0,01, indicando uma não homogeneidade dos resíduos. Todavia, para o experimento analisado, os ambientes

foram considerados como as combinações entre anos por locais, e a pressuposição seria que o fator ano influencia numa dependência entre experimentos. Entretanto os modelos AMMI e W-AMMI estudados no trabalho não incorpora a dependência dos dados, desta forma continuou-se as análises utilizando a combinação entre anos por locais.

Todavia, o presente trabalho teve como objetivo caracterizar o comportamento do fator multiplicativo (interação) utilizando os modelos AMMI e W-AMMI, sendo que, mesmo não apresentando significância, a análise foi realizada afim de encontrar indícios de estabilidade dos porta-enxertos durante os diferentes níveis de Ano Locais. Além disto, o modelo W-AMMI apresenta bons resultados à alta heterogeneidade de ambientes (RODRIGUES *et al.*, 2014), uma vantagem que pode ajudar na análise deste experimento devido o alto valor de QM_{res} que foi apresentado.

4.3 Modelo AMMI

A partir do modelo da ANOVA conjunta foi encontrada a matriz GE dos valores estimados para o efeito da interação genótipos e ambientes, na qual a Tabela 4.5 apresentam valores sem transformação, e na Tabela 4.6 com a transformação. As somas das colunas, e das linhas multiplicadas pelos números de repetições referente a cada local (R_j) resultam a valores nulos, isto evidencia que a estimação foi realizada pelo método de mínimos quadrados.

Tabela 4.5. Valores estimados da matriz GE, provenientes do modelo da ANOVA conjunta

	2012 Jundiai	2012 Votuporanga	2013 Jundiai	2013 Votuporanga	2014 Jundiai	2014 Votuporanga	Σ
IAC 313	374,47	-254,41	130,52	-155,90	104,11	4,24	0
IAC 571-6	-380,56	67,30	41,77	-2,23	-63,02	202,81	0
IAC 572	706,26	-912,66	207,64	267,60	199,49	-97,20	0
IAC 766	-700,18	1099,78	-379,93	-109,48	-240,57	-109,84	0
$\Sigma \times R_j$	0	0	0	0	0	0	$\Sigma = 0$

Tabela 4.6. Valores estimados da matriz GE, provenientes do modelo da ANOVA conjunta com a variável resposta transformada

	2012 Jundiai	2012 Votuporanga	2013 Jundiai	2013 Votuporanga	2014 Jundiai	2014 Votuporanga	Σ
IAC 313	0,51	-0,33	0,37	-0,27	0,09	-0,06	0
IAC 571-6	-0,56	0,03	0,16	0,01	-0,17	0,34	0
IAC 572	0,96	-0,60	-0,15	0,43	0,15	-0,47	0
IAC 766	-0,92	0,89	-0,38	-0,17	-0,07	0,19	0
$\Sigma \times R_j$	0	0	0	0	0	0	$\Sigma = 0$

Nos valores da Tabela 4.5 e 4.6 utilizou-se o métodos da DVS, e desta forma, realizou-se o cálculo para seleção de modelos AMMI pelo critério de Gollob (Tabela 4.7), observou-se que mesmo a IGA apresentando uma estrutura multivariada com $3 \leq \min(4 - 1, 6 - 1)$ autovalores, apenas o primeiro componente principal foi necessário para descrever a estrutura do fator multiplicativo, para os casos com e sem a transformação na variável reposta. Uma evidência a este resultados, poderia ser devido a ausência da interação evidenciando o efeito da estabilidade que os porta-enxertos apresentam no diferentes ambientes.

Pode-se observar com a Tabela 4.7 que o modelo AMMI sem a transformação resultou numa proporção explicativa de 93,58%, enquanto para o o modelo AMMI com a transformação obteve-se uma proporção de 84,53%, o que ocasionou numa perca da explicação que a 1ª componente principal. Isto pode ser relacionado a estabilidade da variação que a transformação faz BOX e Cox (1964).

Tabela 4.7. Quadro de seleção de modelos AMMI utilizando o critério de Gollob

AMMI	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p	Proporção dos autovalores (λ_k)	Proporção Acumulada
Sem transformação na variável resposta							
AMMI ₁	7	21757352,89	3108193,27	2,42	0,03	93,58	93,58
AMMI ₂	5	1003035,61	200607,12	0,16	0,98	4,31	97,89
AMMI ₃	3	490410,11	163470,04	0,13	0,94	2,11	100,00
Com transformação na variável resposta							
AMMI ₁	7	23,63	3,38	2,34	0,03	84,53	84,53
AMMI ₂	5	3,13	0,63	0,43	0,82	11,20	95,74
AMMI ₃	3	1,19	0,40	0,28	0,84	4,26	100,00

Os biplot's (Figura 4.13 e 4.14) utilizaram apenas a primeira componente principal no eixo y e os valores médios dos fatores de porta-enxertos e anos por locais no eixo x. Para o *Biplot* do AMMI sem transformação, temos que o ambiente 2012 Votuporanga apresenta-se em sentido oposto e com maior comprimento, em relação aos demais, isto por conta dos altos valores de desvios-padrão e média (Tabela 4.1). Os porta-enxertos estão todos próximos ao eixo da média geral representando a estabilidade em relação aos diferentes ambientes, a ausência de interação que ocorreu em relação à ANOVA. Pode-se observar que o IAC 766 se apresentou próximo de 2012 Votuporanga, supondo uma relação a este ambiente, no qual pode-se notar nas análises iniciais com a Figura 4.6, o mesmo aconteceu para o porta-enxerto IAC 572 que apresenta melhor valores de produção apenas para o ambiente 2012 Jundiá.

Para o biplot com a variável resposta transformada (Figura. 4.14) no eixo da primeira componente principal um lado é composto pelos ambientes 2012 Votuporanga e 2014 Votuporanga, evidenciando uma tendência dos porta enxertos IAC 766 e IAC 571-6, e o lado oposto aos demais ambientes com os porta-enxertos de IAC 313 e IAC 572. Entretanto o comportamento da IGA apresentou-se similar em relação ao biplot 4.13

Figura 4.13. Biplot para a matriz de interação GE

Figura 4.14. Biplot para a matriz de interação GE, com a variável resposta transformada

Em seguida foi utilizado o modelo W-AMMI para caracterizar o mesmo comportamento da IGA, porém ponderando a alta heterogeneidade entre os ambientes. Foi calculada a matriz de ponderação W (Tabela 4.8), utilizando as estimativas dos componentes de variância pelo pacote estatístico lme4, e encontrando os números de repetições/blocos para cada ambientes como indicado na Seção 2.6. Esta matriz W foi utilizada no método de mínimos quadrados ponderados que calculou as estimativas para o efeito multiplicativo e em seguida aplicou-se a DVS ponderada, gerando uma nova matriz GE (Tabela 4.9).

A primeira diferença que pode-se notar do biplot para o modelo W-AMMI (Figura 4.15) é a proporção explicada da primeira componente principal sendo maior que as outras, porém sem perda de participação de algum ambiente, como aconteceu no biplot AMMI em relação a 2014 Votuporanga (Figura 4.13) que aparecia concorrente ao eixo x. Neste biplot o ambiente 2012 Jundiai ficou oposto aos demais com os porta-enxertos IAC 313 e IAC 572 mais relacionados a ele.

Tabela 4.8. Matriz W de ponderação

	2012 Jundiai	2012 Votuporanga	2013 Jundiai	2013 Votuporanga	2014 Jundiai	2014 Votuporanga
IAC 313	0,67	0,02	0,25	0,09	0,13	0,19
IAC 571-6	0,67	0,02	0,25	0,09	0,13	0,19
IAC 572	0,67	0,02	0,25	0,09	0,13	0,19
IAC 766	0,67	0,02	0,25	0,09	0,13	0,19

Tabela 4.9. Nova matriz gerada a partir do processo iterativo de DVS ponderado

	2012 Jundiai	2012 Votuporanga	2013 Jundiai	2013 Votuporanga	2014 Jundiai	2014 Votuporanga
IAC 313	167,93	-460,95	-76,02	-362,44	-102,44	-202,30
IAC 571-6	-223,22	224,64	199,11	155,11	94,31	360,15
IAC 572	307,16	-1311,76	-191,46	-131,49	-199,61	-496,30
IAC 766	-251,88	1548,08	68,37	338,82	207,73	338,46

Figura 4.15. Biplot para a matriz de interação GE utilizando o modelo W-AMMI

5 RESULTADOS E DISCUSSÕES DA 2ª ANÁLISE - DADOS EMBRAPA

5.1 Análises iniciais

5.1.1 Para a variável

A variável que se utilizou para as análises foi o peso por ser uma característica importante para a avaliação do desempenho do melão. Desta forma, foram realizadas descrições paramétricas (média, variância) e gráficas em relação à variável. Nos gráficos de *box plot* (Figura 5.1. I) e histograma (Figura 5.1. II) foi observado que a variável apresenta um comportamento assimétrico a direita, em que a média (quadrado cinza) possui maior valor do que a mediana. Há alguns valores dispersos à direita da nuvem de pontos (possíveis pontos atípicos) porém as medidas de dispersão (desvio-padrão e variância) foram relativamente baixa.

Figura 5.1. *Box plot* (I) e Histograma (II) para a variável peso

5.1.2 Fatores individuais

Em seguida foram feitas avaliações em relação aos fatores aditivos: ambientes e famílias individualmente. Para melhorar a visualização dos valores de média e desvio-padrão da variável, em relação aos níveis do fator de ambiente e de família, foram dispersos pontos (F) em um plano cartesiano no qual os valores no eixo y são as medidas, e no eixo x as ordens de famílias. Os ambientes são representados por uma linha horizontal contínua e a média geral por uma linhas tracejada. Pode-se perceber na Figura 5.2 que a média está entre um intervalo aproximadamente de 0 à 2,5. Algumas famílias possuem valores médios altos, como a F2, F63 e F88 se apresentando bem acima em relação ao demais. Os ambientes possuem valores médios próximos entre si, e da média geral.

Reciprocamente foi gerado o mesmo gráfico descritivo para os desvios (Figura 5.3), em que observa-se os valores de desvio-padrão não ultrapassaram de 0,8. Os ambientes se apresentam um pouco maiores que a maioria das famílias, além disto notou-se que algumas possuem valores um pouco acima dos demais como a F53, F65, F78 e F80.

Figura 5.2. Gráfico de dispersão para os valores média de peso em relação aos fatores aditivos

Figura 5.3. Gráfico de dispersão para os valores de desvio-padrão de peso em relação aos fatores aditivos

Nos *box plot's* (Figura 5.4) para os níveis do fator ambiente, todos apresentaram uma distribuição assimétrica à direita, com alguns valores atípicos. No caso dos níveis do fator famílias (Figura 5.5) são apresentadas distribuições assimétrica tanto à direita como à esquerda.

Figura 5.4. *Box plot* para a variável peso em relação aos ambientes

Figura 5.5. *Box plot* para a variável peso em relação as famílias

5.2 Análise para grupos de experimentos

Foi realizado uma ANOVA para cada ambiente individualmente indicado na Tabela 5.1. Para os 3 ambientes as famílias apresentaram-se significativas (rejeitando a hipótese nula) evidenciando que, devido as condições experimentais impostas pelo experimento existe um contraste de médias das famílias

que são estatisticamente diferentes entre si. Pode-se notar que nenhum dos $QM_{residual}$ para os três ambientes possuem valores muito elevados.

Tabela 5.1. Quadro da ANOVA individual para cada ambiente

Fontes de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p
ANOVA individual para: Ambiente 1					
Famílias	90	40,32	0,45	4,26	<0,0000**
Resíduo	159	16,71	0,11		
ANOVA individual para: Ambiente 2					
Famílias	89	27,38	0,31	5,74	<0,0000**
Resíduos	85	4,55	0,05		
ANOVA individual para: Ambiente 3					
Famílias	87	32,22	0,37	3,41	<0,0000**
Resíduo	80	8,68	0,11		

No quadro da ANOVA conjunta (Tabela 5.2) a interação entre genótipos $\times$ ambientes foi significativo a 5%, e em virtude disto, rejeita-se a hipótese nula indicando que existe uma dependência entre os efeitos de famílias e ambientes, este resultado evidencia que estes fatores não podem ser estudados separadamente, fazendo com que seja necessário estender o estudo utilizando modelo AMMI.

Tabela 5.2. Quadro da ANOVA conjunta

Fontes de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p
Ambientes	2	11,02	5,51	59,62	<0,0000**
Famílias	91	79,27	0,87	9,43	<0,0000**
Ambientes $\times$ Famílias	175	20,64	0,12	1,28	0,0306**
Resíduo	324	29,94	0,09		

Antes de fazer os cálculos para o modelo AMMI, foi estudada a distribuição dos resíduos estudentizados gerados pelo modelo da ANOVA conjunta. Inicialmente visualizamos o comportamento dos resíduos com os gráficos de *box plot* (Figura 5.6. I), histograma (Figura 5.6. II) e o gráfico Q-Q (Figura 5.6. III). Notou-se que os resíduos estudentizados apresentaram “caldas” nos dois extremos, podendo indicar presença de alguns valores atípicos, ou uma possível falta de ajuste no modelo. Nos gráficos dos resíduos estudentizados em relação aos valores ajustados (Figura 5.7. I), e a ordem das observações (Figura 5.7. II), notou-se a presença de apenas um valor (observação 141) que possa influenciar na falta de ajuste do modelo.

Como os valores de desvio padrão (Figura 5.3) e média (Figura 5.2) se apresentaram próximos, optou-se à prosseguir ao modelo AMMI, sem aplicar a transformação de Box-Cox na variável resposta.

foi gerado um gráfico (Figura 5.8) para facilitar a visualização da matriz, sendo que as linhas tracejadas perpendiculares ao eixo x são as 92 linhas da matriz (95 famílias) e cada sub-gráfico representa as colunas da matriz (3 ambientes).

Nota-se que possuem sete valores ausentes (linhas tracejada vermelha), sendo identificadas as informações de quais famílias, ambientes e as posições que pertencem na matriz GE (Tabela 5.3). Isto aconteceu devido à algumas famílias não possuírem observações em determinados ambientes como apresentado na seção de descrição do experimento (Seção 3.2).

Tabela 5.3. Tabela com as informações dos valores ausentes

Família	Ambiente	Linha	Coluna
86	1	83	1
5	2	5	2
70	2	68	2
31	3	31	3
47	3	47	3
92	3	89	3
95	3	92	3

Figura 5.8. Gráficos da matriz de interação GE com valores ausentes

Assim, foram realizadas duas abordagens utilizando o modelo AMMI. A primeira retirando as famílias que possuíam valores ausentes, e utilizando a matriz GE com apenas 85 linhas. A segunda abordagem foi aplicar a metodologia de imputação (Seção 2.7) calculando valores para preencher a matriz.

5.3.2 Modelo AMMI sem imputação

Sem as famílias com valores ausentes foi realizado o critério de seleção de Gollob (Tabela 5.4), no qual observou-se significativos a utilização de 2 componentes principais para descrever o comportamento da IGA. O AMMI₂ apresenta um alto valor da proporção acumulada (98,64), uma hipótese para

isto ter acontecido, seria devido ao experimento ter apresentado apenas 3 ambientes para a avaliação, caracterizando poucos eixos no espaço dimensional que seria combinado para gerar novos índices.

Com o biplot (Figura. 5.9), observou-se que a maior parte das famílias apresentam-se próximas à origem indicando estabilidade em relação aos diferentes ambientes (são relacionadas por iguais). Porém algumas possuem características de adaptabilidade, como a família 65, em relação ao ambiente 3 e inversa aos ambientes 1 e 2. Destacam-se também, valores das famílias 60, 17, 67 e 88 que se apresentam próximas ao ambiente 1. A família 94 entre os dois ambientes 1 e 2 considerando uma boa em relação a eles e a família 24 entre os ambientes 1 e 3.

Tabela 5.4. Quadro de seleção de modelos AMMI utilizando o critério de Gollob

AMMI	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p	Proporção dos autovalores (λ_k)	Proporção Acumulada
AMMI ₁	92	17,35	0,19	2,04	0,00	59,00	59,00
AMMI ₂	90	11,66	0,13	1,40	0,02	39,65	98,64
AMMI ₃	88	0,40	0,00	0,05	1,00	1,36	100,00

Figura 5.9. Biplot para a matriz de interação GE com valores ausentes

5.3.3 Modelo AMMI com imputação

Em seguida, foi realizado o método da imputação múltipla direto na matriz de interação GE como descrita na Seção 2.7, gerando os valores imputados indicados na Tabela 5.5 com as informações de família, ambientes e posição da matriz. Foi observado o comportamento destes valores imputados (linhas tracejada vermelha) no gráfico da matriz GE (Figura 5.10), em relação aos ambientes (sub-gráficos) e as demais famílias (linhas tracejadas).

Tabela 5.5. Informações dos valores imputados

Família	Ambiente	Linha	Coluna	Valores Imputados
86	1	83	1	-0,29
5	2	5	2	-0,04
70	2	68	2	0,01
31	3	31	3	-0,17
47	3	47	3	0,02
92	3	89	3	-0,04
95	3	92	3	-0,10

Figura 5.10. Gráficos da matriz de interação GE com valores imputados

Novamente pelo critério de Gollob (Tabela 5.6), foi significativo a utilização do modelo AMMI₂ (2 componentes principais) que descrevem 97,74 da proporção acumulada da IGA, as famílias que possuem valores imputados estão representado da cor vermelha no Biplot (Figura 5.11). Pode-se observar que as famílias 70, 47, 92 e 95 se apresentaram próximas a origem indicando uma estabilidade em relação aos 3 diferentes ambientes. A família 86 se apresentou mais relacionada ao eixo do ambiente 2 indicando uma adaptabilidade, as famílias 5 e 31 apresentaram valores próximo ao eixo do ambiente 1. Mesmo com a adição dos valores imputados, não surgiram diferenças nas demais famílias em relação ao AMMI₂ sem imputação.

Tabela 5.6. Quadro de seleção de modelos AMMI utilizando o critério de Gollob

AMMI	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F_{cal}	valor p	Proporção dos autovalores (λ_k)	Proporção Acumulada
AMMI ₁	92	17,89	0,19	2,10	0,00	57,77	57,77
AMMI ₂	90	12,37	0,14	1,49	0,01	39,96	97,74
AMMI ₃	88	0,70	0,01	0,09	1,00	2,26	100,00

Figura 5.11. Biplot para a matriz de interação GE com valores imputados

5.3.4 Modelo W-AMMI com imputação

Foi realizado a metodologia do modelo W-AMMI, afim de verificar o comportamento dos valores imputados na matriz GE. Como descrito na Seção 2.6 do modelo W-AMMI, estimou-se os componentes de variância em relação à cada ambiente a partir do pacote lme4, considerando família como de efeito aleatório, e quantificou-se os números de repetições para cada família nos diferentes ambientes, gerando a matriz W de ponderação (Equação 2.35).

O passo seguinte seria estimar os efeitos aditivos e multiplicativos utilizando o método de mínimos quadrados ponderados, entretanto como o conjunto de dados apresentou valores ausentes, a matriz de resíduos possui menos observações do que a matriz com valores imputados. Em virtude disto para o método da DVS ponderada, a matriz alvo utilizada foi a calculada pelo método de imputação múltipla (Figura 5.10). O processo iterativo foi realizado 315 vezes até o critério de parada $\epsilon = 1 \times 10^{-10}$ gerando uma nova matriz (Figura 5.12).

O modelo W-AMMI não apresenta melhoras significativas na caracterização da estrutura da IGA, como pôde ser observado na Figura 5.13. Apenas o valor da família 65 se estabilizou em relação ao ambiente 3 se aproximando da origem. Muito devido a falta de heterogeneidade entre os ambientes, que apresentam valores de desvio-padrão próximos.

Figura 5.12. Gráficos da matriz de interação GE com valores imputados utilizando a DVS ponderada

Figura 5.13. Biplot para a matriz de interação GE do modelo W-AMMI

6 CONCLUSÕES

I. Primeira análise

Mesmo com a não significância da interação na ANOVA conjunta, a utilização do modelo AMMI ajudou na descrição da estrutura genótipos $\times$ ambientes utilizando apenas a primeira componente principal em função da média dos fatores aditivos.

A proporção explicativa da primeira componente principal do modelo W-AMMI foi maior do que os modelos AMMI e AMMI com a variável resposta transformada (Box-Cox), além disto, não houve perda de informação dos índices gerados pelo método de DVS. Isto ajuda na avaliação do comportamento multivariado que existe na interação genótipo $\times$ ambiente.

II. Segunda análise

A adição de valores imputados, utilizando a técnica de imputação múltipla livre de distribuição, para a análise do modelo AMMI não apresentou diferenças no comportamento dos genótipos relacionados aos diferentes ambientes.

O modelo W-AMMI não ofereceu melhoras consideráveis quando os ambientes não possuem alta heterogeneidades entre si. Os valores imputados também não sofreram alterações.

III. Conclusões gerais

Mesmo em experimentos cuja interação não seja significativa, seria interessante para o pesquisador utilizar a abordagem do modelo AMMI, devido a complexidade que a interação genótipos e ambientes pode apresentar.

O modelo W-AMMI é uma abordagem apropriada apenas quando há casos de alta heterogeneidade entre os ambientes, uma vez que, aplicando a transformação na variável resposta torna-se difícil a discussão dos resultados devido a conversão nas unidades de medidas.

6.1 Pesquisas Futuras

Ideias sobre pesquisas futuras:

- Estudar a dependência entre os diferentes anos;
- Adicionar um novo fator (ex. anos) para a metodologia do modelo AMMI e W-AMMI, em virtude de melhorar a descrição do comportamento de genótipos $\times$ ambientes $\times$ anos;
- Estudar o comportamento da imputação múltipla livre de distribuição, quando há presença de valores discrepantes;
- Utilizar outras abordagens de imputação, verificando se a estrutura da matriz de interação GE sofre alterações;
- Aprimorar o modelo AMMI utilizando mais de uma variável resposta na análise;

REFERÊNCIAS

- ANTON, H. e C. RORRES, 2012 *Álgebra Linear com aplicações*. Bookman, 10th edition.
- BANZATTO, D. A. e S. D. N. KRONKA, 2006 *Experimentação Agrícola*. Funep, Jaboticabal, fourth edition.
- BERGAMO, G. C., 2007 *Imputação múltipla livre de distribuição utilizando a decomposição por valor singular em matriz de interação*. Estatística e experimentação agrônômica, Escola Superior de Agricultura "Luiz de Queiroz".
- BOX, G. E. P. e D. R. COX, 1964 An Analysis of Transformations. Source J. R. Stat. Soc. Ser. B **26**: 211–252.
- BREUSCH, T. S. e A. R. PAGAN, 1979 A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* pp. 1287–1294.
- CAMPOS, H. D., 1983 *Estatística aplicada à experimentação com cana-de-açúcar*.
- DE SOUSA, J. S. I. e F. P. MARTINS, 2002 *Viticultura brasileira: principais variedades e suas características*. Fealq.
- EASTMENT, H. T. e W. J. KRZANOWSKI, 1982 Cross-Validatory Choice of the Number of Components from a Principal Component Analysis. *Technometrics* **24**: 73–77.
- EDWARDS, J. W. e J.-L. JANNINK, 2006 Bayesian Modeling of Heterogeneous Error and Genotype $\times$ Environment Interaction Variances. *Crop Sci.* **46**: 820.
- EMBRAPA, S., 2013 Sistema brasileiro de classificação de solos. Cent. Nac. Pesqui. Solos Rio Janeiro .
- FAREBROTHER, R. W., 1999 *Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900*. Springer Science & Business Media.
- FLORES, F., M. MORENO, e J. CUBERO, 1998 A comparison of univariate and multivariate methods to analyze G $\times$ E interaction. *F. Crop. Res.* **56**: 271–286.
- GAUCH, H. G. e R. W. ZOBEL, 1988 Predictive and postdictive success of statistical analyses of yield trials*. *Theor Appl Genet* **76**: 1–10.
- GAUCH, H. G. J., 2013 A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Sci.* **53**: 1860–1869.
- GOLLOB, H. F., 1968 A Statistical Model which Combines Features of Factor Analytic and Analysis of Variance Techniques. *Psychometrika* **33**: 73–115.
- GREGORIUS, H.-R. e G. NAMKOONG, 1986 Joint analysis of genotypic and environmental effects. *Theor Appl Genet* **72**: 413–422.
- HADASCH, S., J. FORKMAN, W. A. MALIK, e H. P. PIEPHO, 2018 Weighted Estimation of AMMI and GGE Models. *J. Agric. Biol. Environ. Stat.* .
- JOHNSON, R. e D. WICHERN, 2007 *Applied multivariate statistical analysis.. PrenticeHall International*.
- KANG, M. S., 2002 *Quantitative genetics, genomics, and plant breeding*. CABI.
- KANG, M. S., V. T. PRABHAKARAN, e R. B. MEHRA, 2004 Plant Breeding: Mendelian to Molecular Approaches. pp. 535–572, Springer Netherlands, Dordrecht.

- KRZANOWSKI, W. J., 1988 Missing value imputation in multivariate data using the singular value decomposition of a matrix. *Biometrical Lett.* **25**: 31–39.
- KUNO, R. G., 1971 The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika* **58**: 453–467.
- LEGENDRE, P. e L. LEGENDRE, 2003 Numerical ecology: third english edition. *Developments in environmental modelling* **20**.
- LIN, C.-S. e M. R. BINNS, 1988 A superiority measure of cultivar performance for cultivar $\times$ location data. *Canadian Journal of Plant Science* **68**: 193–198.
- MANDEL, J., 1971 A New Analysis of Variance Model for Non-Additive Data. *Technometrics* **13**: 1–18.
- MANLY, B. F. J. e J. A. N. ALBERTO, 2017 *Multivariate Statistical Methods - A Primer*. CRC Press, fourth edition.
- MONTGOMERY, D. C., 2013 *Design and analysis of experiments*. John wiley & sons.
- MONTGOMERY, D. C., E. A. PECK, e G. G. VINING, 2012 *Introduction to linear regression analysis*, volume 821. John Wiley & Sons.
- MOOD, A. M., 1963 *Introduction to the Theory of Statistics*. McGraw-Hill, United States of America.
- PERKINS, J. M. e J. L. JINKS, 1968 Environmental and genotype-environmental components of variability III. Multiple lines and crosses. *Heredity (Edinb)*. **23**: 339.
- PIEPHO, H. P., 1995 Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theor Appl Genet* **90**: 438–443.
- PIMENTEL, G. F., 1990 *Curso de estatística experimental* .
- R CORE TEAM, 2017 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RODRIGUES, P. C., M. MALOSETTI, H. G. J. GAUCH, e F. A. VAN EEUWIJK, 2014 A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Sci.* **54**: 1555–1570.
- RODRIGUES, P. C., A. MONTEIRO, e V. M. LOURENÇO, 2016 A robust AMMI model for the analysis of genotype-by-environment data. *Bioinformatics* **32**: 58–66.
- ROYSTON, J., 1982 An extension of shapiro and wilk's w test for normality to large samples. *Applied statistics* pp. 115–124.
- SATTERTHWAITE, F. E., 1946 An approximate distribution of estimates of variance components. *Biometrics Bull.* **2**: 110–114.
- SHUKLA, G. K., 1972 Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity (Edinb)*. **29**: 237.
- SREBRO, N. e T. JAAKKOLA, 2003 Weighted Low-Rank Approximations. In *Twent. Int. Conf. Mach. Learn.*, edited by Fawcett T. e Mishra N., pp. 720–727, Washington DC.
- TAI, G. C., 1971 Genotypic stability analysis and its application to potato regional trials. *Crop science* **11**: 184–190.

WADE, L., C. McLAREN, L. QUINTANA, D. HARNPICHITVITAYA, S. RAJATASEREEKUL, A. SARAWGI, A. KUMAR, H. AHMED, A. SINGH, R. RODRIGUEZ, J. SIOPONGCO, e S. SARKARUNG, 1999 Genotype by environment interactions across diverse rainfed lowland rice environments. *F. Crop. Res.* **64**: 35–50.

APÊNDICES

Notações matemáticas para somas das médias

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^R Y_{ijk} = Y_{...} \quad \text{e, temos que} \quad \frac{Y_{...}}{IJR} = \bar{Y}_{...}$$

$$\sum_{j=1}^J \sum_{k=1}^R Y_{ijk} = Y_{i..} \quad \text{e, temos que} \quad \frac{Y_{i..}}{JR} = \bar{Y}_{i..} \quad \text{onde} \quad i = 1, \dots, I$$

(6.1)

$$\sum_{i=1}^I \sum_{k=1}^R Y_{ijk} = Y_{.j.} \quad \text{e, temos que} \quad \frac{Y_{.j.}}{IR} = \bar{Y}_{.j.} \quad \text{onde} \quad j = 1, \dots, J$$

$$\sum_{k=1}^R Y_{ijk} = Y_{ij.} \quad \text{e, temos que} \quad \frac{Y_{ij.}}{R} = \bar{Y}_{ij.} \quad \text{onde} \quad i = 1, \dots, I \text{ e } j = 1, \dots, J$$