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

**Interação tripla genótipos $\times$ locais $\times$ anos: um teste estatístico para
verificar a contribuição de cada fator**

Mirian Fernandes Carvalho Araújo

Tese apresentada para obtenção do título de Doutora em
Ciências. Área de concentração: Estatística e Experimen-
tação Agronômica

**Piracicaba
2017**

Mirian Fernandes Carvalho Araújo
Licenciada em Matemática

**Interação tripla genótipos $\times$ locais $\times$ anos: um teste estatístico para
verificar a contribuição de cada fator**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

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

Tese apresentada para obtenção do título de Doutora em
Ciências. Área de concentração: Estatística e Experimen-
tação Agronômica

Piracicaba
2017

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

Araújo, Mirian Fernandes Carvalho

Interação tripla genótipos $\times$ locais $\times$ anos: um teste estatístico para verificar a contribuição de cada fator / Mirian Fernandes Carvalho Araújo. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2017 .

93 p.

Tese (Doutorado) – – USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Interação $G \times L \times A$ 2. Contribuição de genótipos, locais e anos 3. Teste F modificado I. Título.

DEDICATÓRIA

*Dedico esse trabalho ao meu esposo **Lúcio Borges de Araújo**, que esteve ao meu lado em todo o tempo, auxiliando nos resultados deste trabalho e sendo suporte em nossa casa.*

*A minha filha **Isabel** que enche nosso lar de amor e ao nosso filho **Natan** que ainda no ventre já nos alegra com sua existência.*

"Se o Senhor não edificar a casa, em vão trabalham os que a edificam... Eis que os filhos são herança da parte do Senhor" Sl 127.1a,3a

AGRADECIMENTOS

A Deus, autor e consumidor da minha fé, por sua eterna fidelidade e misericórdia. Sem Ele, nada podemos fazer.

Ao professor Dr. Carlos Tadeu dos Santos Dias, pela amizade, exemplo de vida, paciência e pela orientação na elaboração deste trabalho.

A ESALQ/USP pela estrutura física e humana disponível.

A Universidade Federal de Uberlândia/UFU pelo afastamento para conclusão do trabalho e a todos os colegas pelo apoio sempre presente.

A Capes pela concessão de bolsas de estudo.

Ao conselho do programa de Pós-graduação em Estatística e Experimentação Agronômica.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP pelo aprendizado e excelente formação em todo o tempo.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, as secretárias Solange de Assis Paes Sabadin e Luciane Brajão, pelos auxílios permanentes.

Aos colegas de estudo do doutorado pelos momentos de estudos e dedicação nas disciplinas, em especial à amiga Valiana Teodoro por me ajudar com documentos e socorrer sempre que preciso, você foi minha presença quando não pude comparecer, muito obrigada.

Aos meus pais, Jair e Lindomar, por todo incentivo desde a infância a fim de obter uma boa formação. A minha irmã Léia e meu cunhado Diego pela amizade e acolhida sempre agradável.

Aos amigos mais chegados que irmãos, D. Edna, Evani, Lucimara, Eliseu e Dirce por todo apoio e hospitalidade, vocês fazem parte da minha família.

A todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	13
2 REVISÃO BIBLIOGRÁFICA	15
2.1 A interação genótipos $\times$ ambientes	15
2.2 Avaliação da interação $G \times E$	16
2.3 Análise de Variância - ANOVA	17
2.4 Decomposição da interação $G \times L \times A$	21
2.4.1 Decomposição da soma dos graus de liberdade da interação $G \times L \times A$ com os graus de liberdade dos fatores G , $G \times L$ e $G \times A$	22
2.4.2 Decomposição da soma dos graus de liberdade da interação $G \times L \times A$ com os graus de liberdade dos fatores L , $L \times G$ e $L \times A$	23
2.4.3 Decomposição da soma dos graus de liberdade da interação $G \times L \times A$ com os graus de liberdade dos fatores A , $A \times G$ e $A \times L$	24
2.5 Análise <i>multiway</i>	26
2.5.1 Modelos PARAFAC	27
2.5.2 Modelos de Tucker	29
2.5.2.1 Modelos Tucker3	29
2.5.2.2 Propriedades do modelo Tucker3	30
2.5.2.3 Modelos de Tucker2	31
2.5.2.4 Modelos de Tucker1	32
2.5.3 Relações entre os modelos de componentes de três entradas	33
2.5.3.1 Hierarquia dos modelos PARAFAC e Tucker3	33
2.6 Graus de liberdade dos modelos <i>multiway</i>	34
2.7 Dimensionalidade de um modelo de Tucker	34
2.7.1 Procedimento <i>Dif Fit</i> de Timmerman-Kiers	35
2.7.2 Análise residual	36
2.7.3 Critério <i>st</i> de Ceulemans-Kiers	36
2.8 Dimensionalidade de um modelo PARAFAC	37
2.8.1 Procedimento de dividir ao meio (<i>Split-half</i>)	37
2.8.2 Consistência do núcleo	37
2.9 Estabilidade do modelo e poder preditivo por validação	38
2.10 <i>Biplot</i>	39
2.10.1 Decomposição em valores singulares	39
2.10.2 <i>Biplot</i> padrão	40
2.11 <i>Joint plot</i>	41
2.12 Distribuição qui-quadrado não-central	42
3 MATERIAL E MÉTODOS	45
3.1 Características dos dados	45
3.2 Gráfico <i>tripplot</i>	45
3.3 Análise de dados admitindo três entradas	46
3.3.1 Análise de variância conjunta	46
3.4 Estudo de simulação	47

3.5	Proposta de teste estatístico para a contribuição de genótipos, locais e anos para a interação tripla	48
4	RESULTADOS	53
4.1	Gráfico <i>triplo</i>	53
4.1.1	Estudo de simulação	54
4.2	Aplicação em dados reais	78
4.2.1	Análise de variância conjunta admitindo três entradas	78
5	DISCUSSÃO	81
5.1	Aspectos computacionais	81
5.2	PESQUISAS FUTURAS	81
	REFERÊNCIAS	83
	APÊNDICES	87

RESUMO

Interação tripla genótipos $\times$ locais $\times$ anos: um teste estatístico para verificar a contribuição de cada fator

O presente trabalho teve por objetivos propor um método alternativo para testar a contribuição de cada genótipo, local e ano para a interação genótipos $\times$ locais $\times$ anos por meio de um teste F , em dados provenientes de experimentos agrônômicos e implementar uma rotina computacional para a realização da análise de dados, mediante a contribuição de cada genótipo, local e ano para a interação. O estudo avalia treze genótipos e nove ambientes (combinação de três locais e três anos). Simulou-se as somas de quadrados das fatias horizontais (genótipos), verticais (locais) e frontais (anos) do arranjo de três entradas de interação genótipos $\times$ locais $\times$ anos ($G \times L \times A$) gerando 500, 1000 e 5000 experimentos para verificar a distribuição empírica. Os resultados indicaram um ajuste à distribuição qui-quadrado não-central para as fatias horizontais, verticais e frontais do arranjo de três entradas de interação $G \times L \times A$, verificados também pelo teste de Kolmogorov-Smirnov e o gráfico QQplot. Na aplicação do teste F proposto, identificou-se os genótipos, locais e anos que contribuem mais para a interação genótipos $\times$ locais $\times$ anos, sendo de fácil aplicação e interpretação.

Palavras-chave: Interação $G \times L \times A$; Contribuição de genótipos, locais e anos; Teste F modificado

ABSTRACT

Triple interaction genotypes $\times$ locations $\times$ years studies: A statistical test for factor contribution analysis

The present study objectives is to demonstrate an alternative method for test factor contribution on a triple interaction system segregating environment effects (locations $\times$ years) and genotypes effect. The analysis consists of two parts: (1) a F test, applied in agronomical experiments for means calculation; and (2) the implementation of a computer routine for simulating experiments and check results with the proposed method. The triple interaction (genotypes $\times$ locations $\times$ years) was studied in thirteen genotypes and nine environments (a combination of three locations and three years), in a block design with three repetitions. The sum of squares within horizontal slices (genotypes), vertical slices (locations) and frontal slices (years) of three way array the genotypes $\times$ locations $\times$ years ($G \times L \times A$) were simulated, generating 500, 1000 and 5000 experiments to test the empirical distribution. The results indicates a non-central chi-squared distribution for horizontal, vertical and frontal slices in the three way array $G \times L \times A$, also was tested using the Kolmogorov-Smirnov test and QQplot plot. The application of F test identified the genotypes, locations and years showing the relative contribution of these factors for the triple interaction (genotypes $\times$ locations $\times$ years), getting an easy routine and fast interpretation method.

Keywords: $G \times L \times A$ interaction; Factor contribution; Modified F test

LISTA DE FIGURAS

2.1	Performance de dois genótipos em dois ambientes em que (a) sem interação, (b) interação $G \times E$ quantitativa, (c) interação $G \times E$ qualitativa	16
2.2	Particionamento de um arranjo de três entradas em fatias horizontais, verticais e frontais	26
2.3	Decomposição de um arranjo de três entradas propostas por Harshman (1970) e Carrol e Chang (1970) . .	27
2.4	Modelo PARAFAC com R componentes	28
2.5	Representação gráfica do modelo Tucker3	30
2.6	Representação gráfica do modelo Tucker2	32
2.7	Representação gráfica do modelo Tucker1, com primeiro modo reduzido	33
2.8	Modelo PARAFAC escrito como um modelo de Tucker3	34
2.9	Representação de dois marcadores de objetos e um marcador de variáveis em um <i>biplot</i>	40
4.1	Gráfico <i>tripplot</i> para a produtividade de feijão (ton/ha) representando a interação e mostrando as relações entre genótipos, locais e anos	53
4.2	Histogramas para as SQ dos genótipos 1, 2 e 3 com 500 experimentos simulados	54
4.3	Histogramas para as SQ dos genótipos 4, 5 e 6 com 500 experimentos simulados	55
4.4	Histogramas para as SQ dos genótipos 7, 8 e 9 com 500 experimentos simulados	55
4.5	Histogramas para as SQ dos genótipos 10, 11 e 12 com 500 experimentos simulados	56
4.6	Histograma para a SQ do genótipo 13 com 500 experimentos simulados	56
4.7	Histogramas para as SQ dos genótipos 1, 2 e 3 com 1000 experimentos simulados	57
4.8	Histogramas para as SQ dos genótipos 4, 5 e 6 com 1000 experimentos simulados	57
4.9	Histogramas para as SQ dos genótipos 7, 8 e 9 com 1000 experimentos simulados	58
4.10	Histogramas para as SQ dos genótipos 10, 11 e 12 com 1000 experimentos simulados	58
4.11	Histograma para a SQ do genótipo 13 com 1000 experimentos simulados	59
4.12	Histogramas para as SQ dos genótipos 1, 2 e 3 com 5000 experimentos simulados	59
4.13	Histogramas para as SQ dos genótipos 4, 5 e 6 com 5000 experimentos simulados	60
4.14	Histogramas para as SQ dos genótipos 7, 8 e 9 com 5000 experimentos simulados	60
4.15	Histogramas para as SQ dos genótipos 10, 11 e 12 com 5000 experimentos simulados	61
4.16	Histograma para a SQ do genótipo 13 com 5000 experimentos simulados	61
4.17	Histogramas para as SQ dos locais 1, 2 e 3 com 500 experimentos simulados	62
4.18	Histogramas para as SQ dos locais 1, 2 e 3 com 1000 experimentos simulados	63
4.19	Histogramas para as SQ dos locais 1, 2 e 3 com 5000 experimentos simulados	63
4.20	Histogramas para as SQ dos anos 1, 2 e 3 com 500 experimentos simulados	64
4.21	Histogramas para as SQ dos anos 1, 2 e 3 com 1000 experimentos simulados	65
4.22	Histogramas para as SQ dos anos 1, 2 e 3 com 5000 experimentos simulados	65
4.23	Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 500 experimentos simulados	67
4.24	Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 500 experimentos simulados	68
4.25	Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 500 experimentos simulados	68
4.26	Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 500 experimentos simulados . . .	69
4.27	Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 500 experimentos simulados	69
4.28	Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 1000 experimentos simulados	70
4.29	Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 1000 experimentos simulados	70
4.30	Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 1000 experimentos simulados	71
4.31	Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 1000 experimentos simulados . .	71
4.32	Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 1000 experimentos simulados	72
4.33	Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 5000 experimentos simulados	72

4.34	Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 5000 experimentos simulados	73
4.35	Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 5000 experimentos simulados	73
4.36	Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 5000 experimentos simulados . .	74
4.37	Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 5000 experimentos simulados	74
4.38	Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 500 experimentos simulados	75
4.39	Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 1000 experimentos simulados	75
4.40	Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 5000 experimentos simulados	76
4.41	Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 500 experimentos simulados	76
4.42	Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 1000 experimentos simulados	77
4.43	Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 5000 experimentos simulados	77

LISTA DE TABELAS

2.1	Representação dos dados médios de g genótipos avaliados em l locais e k anos para um caráter genérico Y	18
2.2	Esquema da ANOVA de g genótipos avaliados em l locais e em a anos dispostos em um delineamento aleatorizado em blocos	22
2.3	Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com G	23
2.4	Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com L	24
2.5	Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com A	26
3.1	Caracterização dos ambientes experimentais	45
3.2	Esquema da ANOVA para experimentos de um mesmo grupo de g genótipos avaliados em l locais e a anos	47
3.3	Esquema da ANOVA com teste F para obtenção de genótipos que contribuem significativamente para a interação $G \times L \times A$	49
3.4	Esquema da ANOVA com teste F para obtenção de locais que contribuem significativamente para a interação $G \times L \times A$	50
3.5	Esquema da ANOVA com teste F para obtenção de anos que contribuem significativamente para a interação $G \times L \times A$	51
4.1	Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos genótipos para 500, 1000 e 5000 experimentos simulados	66
4.2	Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos locais para 500, 1000 e 5000 experimentos simulados	66
4.3	Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos anos para 500, 1000 e 5000 experimentos simulados	67
4.4	ANOVA para um conjunto de dados com 13 genótipos avaliados em 3 locais, 3 anos com 3 blocos .	78
4.5	Teste F para obtenção de genótipos que contribuem significativamente para a interação $G \times L \times A$	78
4.6	Teste F para obtenção de locais que contribuem significativamente para a interação $G \times L \times A$	79
4.7	Teste F para obtenção de anos que contribuem significativamente para a interação $G \times L \times A$	79

1 INTRODUÇÃO

Em um experimento, ao utilizar os mesmos genótipos em diferentes ambientes, pode ocorrer mudanças nos genótipos devido ao ambiente, ou seja, ocorre o que se chama de interação genótipos $\times$ ambientes. Para os melhoristas, esta interação interfere de forma negativa na seleção dos melhores genótipos. Entretanto, com a ausência de interação, um único experimento forneceria resultados universais, uma vez que uma variedade se adaptaria à maioria dos ambientes.

Os experimentos multi-ambientais são aqueles conduzidos em vários anos a fim de obter um bom genótipo e conseqüentemente, recomendar um cultivar. Assim, a interação genótipos $\times$ ambientes é um desafio na melhora de cultivares e a análise de dados desses experimentos multi-ambientais é de extrema importância para o melhoramento genético de plantas (ARAUJO, 2009).

O objetivo inicial de um experimento multi-ambiental é identificar cultivares superiores, ou seja, aquelas que mantêm um comportamento estável ao mudar de local e/ou ano. Na prática, avalia o rendimento de um genótipo em vários locais e anos. Um segundo objetivo de dados multi-ambientais é investigar as relações entre os locais e anos testados. Um experimento multi-ambiental é avaliado por vários anos, ou seja, tem-se a interação genótipos $\times$ locais $\times$ anos e os dados são dispostos em uma tabela ou arranjo de três entradas, em que cada entrada refere-se a genótipos, locais e anos.

Um dos interesses do pesquisador é verificar se ao longo dos anos há uma estrutura comum em relação aos anos e qual resposta os genótipos apresentam. Por exemplo, alguns genótipos podem ter respostas positivas em alguns locais, e da mesma forma, alguns locais podem estar mais associados com alguns genótipos do que outros por alguns anos.

Os pesquisadores têm combinado os fatores locais e anos devido a falta de uma ferramenta adequada para o estudo da interação genótipos $\times$ locais $\times$ anos. Segundo VARELA *et al.* (2006), em alguns casos, esta combinação de locais e anos gera uma perda da informação no ajuste de modelo de duas entradas e na estimação dos efeitos da interação. Assim, um teste estatístico F é necessário para permitir o desdobramento e a interpretação da interação genótipos $\times$ locais $\times$ anos (ARAUJO *et al.*, 2012).

Os objetivos do presente trabalho são propor um método alternativo para testar a contribuição de cada genótipo, local e ano, considerados de efeitos fixos, para a interação genótipos $\times$ locais $\times$ anos por meio de um teste F , em dados provenientes de experimentos agrônômicos e implementar uma rotina computacional para a realização da análise de dados, mediante a contribuição de cada genótipo, local e ano para a interação.

2 REVISÃO BIBLIOGRÁFICA

2.1 A interação genótipos $\times$ ambientes

Na condução de um experimento, os genótipos analisados podem sofrer influência, positiva ou negativa, quando comparados em diferentes ambientes. Em um ambiente, várias características podem ser avaliadas, tais como: tipos de solo, níveis de fertilidade, temperaturas e práticas culturais (DAS, 2006) e os genótipos podem apresentar performances relativas diferentes nos ambientes. Essa mudança de performance relativa de genótipos em diferentes ambientes é denominada como interação genótipos $\times$ ambientes ($G \times E$, em que E significa *environments*).

Para CHAVES (2001), a interação $G \times E$ é resultado da resposta diferencial dos genótipos à variação ambiental e não deve ser considerada como um problema, mas como um fenômeno biológico natural, que quando bem conhecido, será melhor aproveitado no processo de seleção. As ferramentas estatísticas juntamente com o conhecimento biológico da interação leva a melhores informações e a entender a complexidade da interação. Logo, os genótipos que interagem positivamente com os ambientes podem fazer a diferença entre um bom e um ótimo cultivar (DUARTE e VENCovsky, 1999).

Segundo FALCONER (1987), a interação $G \times E$ é uma fonte de variação fenotípica, que na maioria dos casos é inseparável da variância ambiental. Um experimento não deve ser realizado apenas em um ambiente, pois segundo CROSSA (1990) poderá haver uma superestimação dos ganhos genéticos, ou seja, para verificar a significância da interação de genótipos com ambientes, é necessário repetir o experimento várias vezes.

A interação pode ser simples ou complexa. A interação simples mostra que a presença de genótipos são adaptados em vários ambientes, e a interação complexa mostra que determinados genótipos adaptam-se a apenas alguns ambientes o que gera uma complicação ao pesquisador, ao recomendar uma cultivar. Segundo ALLARD e BRADSHAW (1964), as variáveis ambientais podem ser previsíveis como as características gerais do clima que ocorrem de maneira sistemática ou estejam sob o controle do homem, ou imprevisíveis que são as variações climáticas que não estão sob o controle do homem.

Os melhoristas de plantas entendem que a interação $G \times E$ têm um importante significado para a obtenção de variedades superiores. Entretanto, para identificar genótipos superiores surgem obstáculos no processo de seleção e na própria recomendação de cultivares, pois a resposta de um genótipo, em comparação a outro, é variável, sendo tais variações decorrentes da mudança ambiental (OLIVEIRA *et al.*, 2003); (KANG, 1998).

A interação $G \times E$ foi estudada pela primeira vez por FISHER e MACKENZIE (1923). Desde então, muitos trabalhos têm sido feitos para análises estatísticas da interação $G \times E$, seja por estatísticos, agrônomos, melhoristas ou geneticistas.

Quando se detecta algum tipo de interação, vários métodos podem ser utilizados para sua interpretação, entre eles, incluem estudos de regressão e métodos envolvendo variáveis ambientais externas.

A regressão foi utilizada pela primeira vez por YATES e COCHRAN (1938) na análise de grupos de experimentos por apresentarem interação $G \times E$. Neste caso, o grau de associação entre diferenças varietais pôde ser verificado pelo cálculo da regressão dos rendimentos das variedades isoladas sobre os rendimentos médios de todas as variedades, mostrando assim que a regressão explicava grande parte da interação em uma série de experimentos de cevada. O método de regressão foi também usado por PERKINS e JINKS (1968) para estimação de parâmetros em um modelo genético aplicado à biometria.

Outro método usado na investigação da interação é por meio da análise de variância (ANOVA) em que os dados são dispostos em uma tabela de dupla entrada. Por envolver vários experimentos, é possível determinar a magnitude da interação por meio da razão do quadrado médio da interação ($QM_{G \times E}$) pelo quadrado médio do resíduo (QM_{Res}). Entretanto, somente a detecção de significância

para a interação não esclarece as implicações que estas possam ter sobre o melhoramento, de forma que, estudos de detalhamento deste componente da variação são em geral necessários.

No melhoramento de plantas, o mais importante da interação $G \times E$ é aquela do tipo cruzada, pois implica em mudanças no comportamento de genótipos nos ambientes. Quando a interação não é cruzada, genótipos com médias superiores podem ser recomendados para todos os ambientes. A Figura 2.1 (a) apresenta dois genótipos em dois ambientes com ausência de interação, (b) os dois genótipos tem interação apenas quantitativa em relação aos ambientes, já em (c) a interação é qualitativa, ou seja, ocorre mudanças nas respostas dos genótipos ao passar de um ambiente para outro.

Figura 2.1. Performance de dois genótipos em dois ambientes em que (a) sem interação, (b) interação $G \times E$ quantitativa, (c) interação $G \times E$ qualitativa

2.2 Avaliação da interação $G \times E$

É preciso boa compreensão dos passos envolvidos no delineamento adotado, bem como, sua condução, análise e interpretação para que a interação seja bem avaliada. Segundo ARAUJO (2009), algumas análises da interação $G \times E$ não se baseiam em um experimento delineado para este fim, como o caso da interação com locais e anos. O que os pesquisadores fazem é utilizar dados de genótipos e linhas experimentais que foram avaliados por meio de locais e anos como parte de um programa de melhoramento genético com outras finalidades, como por exemplo, experimentos conduzidos para comparar genótipos e determinar o tamanho adequado da amostra. Uma desvantagem do procedimento é que tanto genótipos como os experimentos podem não ser uma amostra aleatória de genótipos avaliados. É preferível utilizar amostra aleatória dos genótipos para que os objetivos do estudo sejam maiores e amplamente alcançados, ao contrário se os considerarem fixos (ANNICCHIARICO, 2002).

Para obtenção das estimativas dos efeitos das interações entre genótipos, locais e anos é preciso conduzir o experimento em dois ou mais locais e anos. Neste caso, a interação $G \times E$ na análise de variância pode ser dividida em componentes devido a genótipos $\times$ locais ($G \times L$), genótipos $\times$ anos ($G \times A$) e genótipos $\times$ locais $\times$ anos ($G \times L \times A$). Quando os locais não são escolhidos aleatoriamente de todos os possíveis locais na área, podem ser considerados de efeito fixo. Há pesquisadores que adotam efeito aleatório para locais pois dizem não ter o controle sobre as condições climáticas que podem ocorrer nos locais durante algum ano. Tal razão, justifica o uso de efeitos aleatórios também para anos.

É necessário repetir o experimento pelo menos duas vezes em cada local e em cada ano para obter uma estimativa do erro experimental que será usado para verificar a significância da interação e quanto maior for o número de repetições, melhor será a estimativa do erro experimental. Deve-se considerar na interpretação todas as causas de variação significativas e avaliar a importância prática da variação observada entre os valores médios.

Uma baixa interação $G \times L$ refere-se à consistência da performance de genótipos em diferentes locais. A consistência do desempenho dos genótipos em diferentes anos é indicada pela interação $G \times A$. E a interação $G \times L \times A$ refere-se à consistência entre genótipos para cada combinação de anos e locais e sua significância indica que a performance relativa entre genótipos não foi a mesma em cada combinação de locais e anos. Entretanto, para todas as interações citadas anteriormente é preciso examinar os seus valores médios para determinar se a significância da interação é devida à mudanças de classificação entre genótipos ou à mudanças nas diferenças entre os genótipos, sem haver variação na classificação.

A não significância da interação de genótipos com locais, anos ou com a combinação locais e anos, indica que um teste em local durante um determinado ano poderia ser suficiente para identificar genótipos com potencial genético superior. Se um genótipo tem melhor performance em um local e ano, este pode ser também superior a outros locais e anos (ARAUJO, 2009).

Os pesquisadores não se preocupam tanto quando a interação $G \times E$ não são referentes às mudanças na classificação de performance entre os genótipos. Nestas condições, por exemplo, um experimento em um local determina que certo ano poderia ser usado para identificar genótipos superiores. Mesmo com variação das magnitudes de superioridade, o mesmo genótipo pode ser superior em todos locais e anos. Quando a interação $G \times E$ é significativa e envolve mudanças na performance entre genótipos, os pesquisadores levam em consideração a extensão das mudanças de classificação e o impacto no melhoramento genético. Mas é bom que outros pesquisadores avaliem este mesmo conjunto de dados e verifiquem se as formas do procedimento e resultados são as mesmas ou não. Segundo DAS (2006), as opções para os pesquisadores difere para cada tipo de interação:

- i) $G \times L$: Quando há flutuações na posição de performance dos genótipos nos locais, isso sugere um desenvolvimento de genótipos para diferentes locais utilizando programas e teste de seleção independentes. Os programas de melhoramento independentes são apropriados se as diferenças entre os locais for devido ao tipo de solo ou outros fatores que são consistentes de ano para ano. Diferenças temporais entre locais associadas com condições climáticas não usuais não justificam o uso de programas independentes.
- ii) $G \times A$: Programas de melhoramento independentes para diferentes anos não são aconselháveis. É preciso identificar os genótipos que apresentaram uma performance superior por meio dos anos. Uma redução no tempo de melhoria genético pode ser feito se substituir os anos por múltiplos locais em um ano. Tal substituição é efetiva somente quando as diferenças nas condições climáticas entre os locais são comparáveis às diferenças entre os anos.
- iii) $G \times L \times A$: Havendo modificações nas posições dos genótipos com a interação locais $\times$ anos, o pesquisador deve identificar genótipos com performance médias superiores sobre os locais e anos. Pode-se, por exemplo, ter um experimento em que as interações $G \times A$ e $G \times L$ não sejam significativas, enquanto que a interação $G \times L \times A$ é significativa, ou seja, a interação parece estar associada à uma condição específica que provocou uma variação na classificação dos genótipos entre certas combinações de locais $\times$ anos. Espera-se que um determinado genótipo que apresentou alta performance média em determinado ano, continue a apresentar boa performance no ano seguinte, mas não necessariamente o melhor naquela época em particular.

2.3 Análise de Variância - ANOVA

Supondo-se que g genótipos foram avaliados em l locais e em a anos dispostos em um delineamento em blocos aleatorizados, sendo bloco dentro de locais e de anos. O modelo mais simples e comum para a análise estatística de um conjunto de dados é dado por:

$$Y_{ijk r} = \mu + b_r(l_j a_k) + g_i + l_j + a_k + (gl)_{ij} + (ga)_{ik} + (la)_{jk} + (gla)_{ijk} + \varepsilon_{ijk r} \quad (2.1)$$

sendo que:

$Y_{ijk r}$: é a resposta do i -ésimo genótipo no j -ésimo local e k -ésimo ano no r -ésimo bloco;

μ : é uma constante comum às respostas (média geral);

$b_r(l_j a_k)$: é o efeito do r -ésimo bloco dentro do j -ésimo local e do k -ésimo ano ($r = 1, 2, \dots, R$);

g_i : é o efeito do i -ésimo genótipo ($i = 1, 2, \dots, G$);

l_j : é o efeito do j -ésimo local ($j = 1, 2, \dots, L$);

a_k : é o efeito do k -ésimo ano ($k = 1, 2, \dots, A$);

$(gl)_{ij}$: é o efeito da interação específica do i -ésimo genótipo com o j -ésimo local;

$(ga)_{ik}$ é o efeito da interação específica do i -ésimo genótipo com o k -ésimo ano;

$(la)_{jk}$ é o efeito da interação específica do j -ésimo local com o k -ésimo ano;

$(gla)_{ijk}$ é o efeito da interação específica do i -ésimo genótipo com o j -ésimo local com o k -ésimo ano; e

$\varepsilon_{ijk r}$: é o erro experimental associado ao i -ésimo genótipo com o j -ésimo local com o k -ésimo ano e no r -ésimo bloco, assumido independente e $\varepsilon_{ijk r} \sim N(0, \sigma^2)$.

Esse modelo de análise de variância (ANOVA) é particularmente utilizado quando locais diferem ao longo dos anos e também pode ser usado para testar anos dentro de cada local. Os fatores são fixos ou aleatórios, normalmente quando os investigadores controlam o nível de um fator, o fator é fixo. Quando o investigador amostra aleatoriamente os níveis de um fator em uma população, o fator é aleatório. A forma básica de organização dos dados para análise da interação genótipos $\times$ locais $\times$ anos está representado na Tabela 2.1, na qual estão representadas as médias dos g tratamentos (genótipos) sobre r repetições, derivadas de um experimento em delineamento apropriado. Quando todos os genótipos são avaliados em todos os locais e anos, com o mesmo número de repetições por experimento, diz-se que os dados são balanceados.

Tabela 2.1. Representação dos dados médios de g genótipos avaliados em l locais e k anos para um caráter genérico Y

Genótipos	Locais							
	1		...		L			
	Anos		...		Anos			
	1	...	A	...	1	...	A	Média ($\bar{Y}_{i..}$)
1	Y_{111}	...	Y_{12a}	...	Y_{1l1}	...	Y_{1la}	$\bar{Y}_{1..}$
2	Y_{211}	...	Y_{22a}	...	Y_{2l1}	...	Y_{2la}	$\bar{Y}_{2..}$
$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
G	Y_{g11}	...	Y_{g2a}	...	Y_{gl1}	...	Y_{gla}	$\bar{Y}_{g..}$
Médias ($\bar{Y}_{.jk}$)	$\bar{Y}_{.11}$	...	$\bar{Y}_{.1a}$	...	$\bar{Y}_{.l1}$	...	$\bar{Y}_{.la}$	$\bar{Y}_{...}$

A análise de variância feita sob o modelo (2.1) com base nos dados da Tabela 2.1 é feita por meio do cálculo dos efeitos principais pelas marginais da Tabela 2.1 e as interações como desvios do

em que $(\mathbf{X}'\mathbf{X})^-$ é uma inversa generalizada, uma vez que o sistema é indeterminado, admitindo-se as condições marginais:

$$\begin{aligned} \sum_{i=1}^G g_i &= \sum_{j=1}^L l_j = \sum_{k=1}^A a_k = \sum_{i=1}^G \sum_{j=1}^L (gl)_{ij} = \sum_{i=1}^G \sum_{k=1}^A (ga)_{ik} = \\ &= \sum_{j=1}^L \sum_{k=1}^A (la)_{jk} = \sum_{i=1}^G \sum_{j=1}^L \sum_{k=1}^A (gla)_{ijk} = 0. \end{aligned} \quad (2.2)$$

Assim, a solução do sistema de equações normais é:

$$\begin{aligned} \hat{\mu} &= \bar{Y}_{....}; \\ \hat{g}_i &= \bar{Y}_{i...} - \bar{Y}_{....}; \\ \hat{l}_j &= \bar{Y}_{.j..} - \bar{Y}_{....}; \\ \hat{a}_k &= \bar{Y}_{..k.} - \bar{Y}_{....}; \\ (\hat{gl})_{ij} &= \bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....}; \\ (\hat{ga})_{ik} &= \bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....}; \\ (\hat{la})_{jk} &= \bar{Y}_{.jk.} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y}_{....}; \\ (\hat{gla})_{ijk} &= \bar{Y}_{ijk.} - \bar{Y}_{ij..} - \bar{Y}_{i.k.} - \bar{Y}_{.jk.} + \bar{Y}_{i...} + \bar{Y}_{.j..} + \bar{Y}_{..k.} - \bar{Y}_{....}. \end{aligned}$$

Esses são estimadores não viesados dos parâmetros, sujeitos às condições marginais dadas acima. Se essas condições não forem impostas sobre os parâmetros, então os estimadores dos parâmetros individuais são viesados.

A matriz de interações $\mathbf{GLA}_{(g \times l \times a)} = [(\hat{gla}_{ijk})]$, é dada como segue:

$$\mathbf{GLA}_{(g \times l \times a)} = \begin{pmatrix} \hat{gla}_{111} & \hat{gla}_{121} & \cdots & \hat{gla}_{1l1} & \hat{gla}_{112} & \hat{gla}_{122} & \cdots & \hat{gla}_{1l2} & \cdots & \hat{gla}_{11a} & \hat{gla}_{12a} & \cdots & \hat{gla}_{1la} \\ \hat{gla}_{211} & \hat{gla}_{221} & \cdots & \hat{gla}_{2l1} & \hat{gla}_{212} & \hat{gla}_{222} & \cdots & \hat{gla}_{2l2} & \cdots & \hat{gla}_{21a} & \hat{gla}_{22a} & \cdots & \hat{gla}_{2la} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{gla}_{g11} & \hat{gla}_{g21} & \cdots & \hat{gla}_{gl1} & \hat{gla}_{g12} & \hat{gla}_{g22} & \cdots & \hat{gla}_{gl2} & \cdots & \hat{gla}_{g1a} & \hat{gla}_{g2a} & \cdots & \hat{gla}_{gla} \end{pmatrix}$$

Desta forma, as somas de quadrados dos efeitos principais e efeitos das interações, são obtidas por:

$$\begin{aligned} SQ_{Total} &= \frac{1}{R} \sum_{i=1}^G \sum_{j=1}^L \sum_{k=1}^A Y_{ijk}^2 - \frac{Y_{....}^2}{RGLA} \text{ com } GLAR - 1 \text{ graus de liberdade;} \\ SQ_{Bd.LA} &= \frac{1}{R} \sum_{i=1}^G \sum_{j=1}^L \sum_{k=1}^A Y_{ijk}^2 - \frac{1}{GR} \sum_{j=1}^L \sum_{k=1}^A Y_{.jk.}^2 \text{ com } LA(R-1) \text{ graus de liberdade;} \\ SQ_G &= \frac{1}{RLA} \sum_{i=1}^G Y_{i...}^2 - \frac{Y_{....}^2}{RGLA} \text{ com } G-1 \text{ graus de liberdade;} \\ SQ_L &= \frac{1}{RGA} \sum_{j=1}^L Y_{.j..}^2 - \frac{Y_{....}^2}{RGLA} \text{ com } L-1 \text{ graus de liberdade;} \\ SQ_A &= \frac{1}{RGL} \sum_{k=1}^A Y_{..k.}^2 - \frac{Y_{....}^2}{RGLA} \text{ com } A-1 \text{ graus de liberdade;} \\ SQ_{GL} &= \frac{1}{RA} \sum_{i=1}^G \sum_{j=1}^L Y_{ij..}^2 - \frac{Y_{....}^2}{RGLA} - SQ_G - SQ_L \text{ com } (G-1)(L-1) \text{ graus de liberdade;} \\ SQ_{GA} &= \frac{1}{RL} \sum_{i=1}^G \sum_{k=1}^A Y_{i.k.}^2 - \frac{Y_{....}^2}{RGLA} - SQ_G - SQ_A \text{ com } (G-1)(A-1) \text{ graus de liberdade;} \end{aligned}$$

$$SQ_{LA} = \frac{1}{RG} \sum_{j=1}^L \sum_{k=1}^A Y_{.jk}^2 - \frac{Y_{....}^2}{RGLA} - SQ_L - SQ_A \text{ com } (L-1)(A-1) \text{ graus de liberdade;}$$

$SQ_{GLA} = SQ_{Total} - SQ_G - SQ_L - SQ_A - SQ_{GL} - SQ_{GA} - SQ_{LA}$ com $(G-1)(L-1)(A-1)$ graus de liberdade;

$SQ_{Res} = SQ_{Total} - SQ_G - SQ_L - SQ_A - SQ_{GL} - SQ_{GA} - SQ_{LA} - SQ_{GLA} - SQ_{Bd.LA}$ com $LA(G-1)(R-1)$ graus de liberdade.

Segundo CHAVES (2001), é preciso definir a natureza fixa ou aleatória dos efeitos do modelo em estudo para realização do teste F . Se os efeitos principais são fixos, a interação será fixa. Se pelo menos um dos fatores for aleatório, a interação será aleatória.

As hipóteses testadas no modelo (2.1) são:

- i) $H_0 : g_i = 0$, para todo i
 $H_1 : g_i \neq 0$, para algum i
- ii) $H_0 : l_j = 0$, para todo j
 $H_1 : l_j \neq 0$, para algum j
- iii) $H_0 : a_k = 0$, para todo k
 $H_1 : a_k \neq 0$, para algum k
- iv) $H_0 : (gl)_{ij} = 0$, para todo i, j
 $H_1 : (gl)_{ij} \neq 0$, para algum i, j
- v) $H_0 : (ga)_{ik} = 0$, para todo i, k
 $H_1 : (ga)_{ik} \neq 0$, para algum i, k
- vi) $H_0 : (la)_{jk} = 0$, para todo j, k
 $H_1 : (la)_{jk} \neq 0$, para algum j, k
- vii) $H_0 : (gla)_{ijk} = 0$, para todo i, j, k
 $H_1 : (gla)_{ijk} \neq 0$, para algum i, j, k

2.4 Decomposição da interação $G \times L \times A$

O efeito da interação genótipos $\times$ locais $\times$ anos ($G \times L \times A$) visa verificar, estatisticamente, o efeito do genótipo na presença dos locais e anos ou o efeito do local na presença dos genótipos e anos, ou ainda o efeito do ano na presença dos genótipos e locais. Quando se verifica, significativamente, evidências na presença da interação do genótipo com o local e/ou ano, conclui-se que há um efeito diferenciado do genótipo na presença de pelo menos um local e/ou ano ou vice-versa. Neste caso, o procedimento adequado para interpretar o efeito da interação $G \times L \times A$ é por meio de sua decomposição, que consiste em estudar o efeito de cada fator na presença de cada nível dos outros fatores.

Quando o efeito da interação é significativo e, também, verifica-se significância para algum efeito principal, deve-se interpretar e concluir somente os resultados referentes ao efeito da interação, ignorando os resultados obtidos para os efeitos principais (ARAUJO, 2008). Este fato é justificado por meio das hipóteses testadas para os efeitos principais, que sob a condição de interação significativa perde o seu significado. O quadro da ANOVA é dado como mostrado na Tabela 2.2.

Com a presença de significância para os tratamentos, deve-se fazer o desdobramento dos $t - 1$ graus de liberdade de tratamentos, para estudar os efeitos principais de genótipos (G), locais (L) e anos (A), bem como das interações entre os fatores, $G \times L$, $G \times A$, $L \times A$ e $G \times L \times A$.

Tabela 2.2. Esquema da ANOVA de g genótipos avaliados em l locais e em a anos dispostos em um delineamento aleatorizado em blocos

F.V.	GL	SQ	QM	F
G	$G - 1$	SQ_G	SQ_G/GL_G	QM_G/QM_{Res}
L	$L - 1$	SQ_L	SQ_L/GL_L	QM_L/QM_{Res}
A	$A - 1$	SQ_A	SQ_A/GL_A	QM_A/QM_{Res}
G×L	$(G - 1)(L - 1)$	SQ_{GL}	SQ_{GL}/GL_{GL}	QM_{GL}/QM_{Res}
G×A	$(G - 1)(A - 1)$	SQ_{GA}	SQ_{GA}/GL_{GA}	QM_{GA}/QM_{Res}
L×A	$(L - 1)(A - 1)$	SQ_{LA}	SQ_{LA}/GL_{LA}	QM_{LA}/QM_{Res}
G×L×A	$(G - 1)(L - 1)(A - 1)$	SQ_{GLA}	SQ_{GLA}/GL_{GLA}	QM_{GLA}/QM_{Res}
Trat	$t - 1$	SQ_{Trat}	SQ_{Trat}/GL_{Trat}	QM_{Trat}/QM_{Res}
Bd.LA	$LA(R - 1)$	$SQ_{Bd.LA}$	$SQ_{Bd.LA}/GL_{Bd.LA}$	—
Resíduo	$LA(G - 1)(R - 1)$	SQ_{Res}	SQ_{Res}/GL_{Res}	
Total	$GLAR - 1$	SQ_{Total}		

F.V.: Fonte de variação; GL: Graus de liberdade; SQ: Soma de quadrados; QM: Quadrado médio;

Trat: Tratamento; Bd.LA: Bloco dentro do local e do ano.

2.4.1 Decomposição da soma dos graus de liberdade da interação G×L×A com os graus de liberdade dos fatores G, G×L e G×A

Essa decomposição corresponde a obter o efeito do fator G dentro do fator L e dentro do fator A, representado por G(LA). Suponha que o delineamento seja aleatorizado em blocos. Assim, o modelo adotado é:

$$Y_{ijk r} = \mu + b_r(l_j a_k) + g_{i(jk)} + l_j + a_k + (la)_{jk} + \varepsilon_{ijk r}$$

em que:

$g_{i(jk)}$: é o efeito do i -ésimo genótipo na presença do j -ésimo local e k -ésimo ano.

O $g_{i(jk)}$ é definido como:

$$g_{i(jk)} = \mu_{i(jk)} - \mu_{.jk}, \text{ tal que } \sum_{i=1}^G g_{i(jk)} = 0, \text{ para todo } j \text{ e } k.$$

Sendo que $\mu_{i(jk)}$ refere-se à média populacional do i -ésimo genótipo na presença do j -ésimo local e k -ésimo ano e $\mu_{.jk}$ refere-se à média populacional do j -ésimo local e k -ésimo ano.

As hipóteses testadas são:

- i) $H_{01} : l_j = 0$, para todo j
 $H_{a1} : l_j \neq 0$, para algum j
- ii) $H_{02} : a_k = 0$, para todo k
 $H_{a2} : a_k \neq 0$, para algum k
- iii) $H_{03} : g_{i(jk)} = 0$, para todo i na presença de j e k
 $H_{a3} : g_{i(jk)} \neq 0$, para algum i na presença de j e k

A soma de quadrados do genótipo dentro do local e do ano é dada por:

$$SQ_{G(L_j A_k)} = \frac{1}{R} \sum_{i=1}^G Y_{i(jk)}^2 - \frac{Y_{.jk}^2}{RG}$$

com $(G - 1)$ graus de liberdade associados, sendo que

$$Y_{i(jk)} = \sum_{r=1}^R Y_{i(jk)r} \text{ o total do } i\text{-ésimo genótipo dentro do } j\text{-ésimo local e do } k\text{-ésimo ano;}$$

$Y_{.jk.} = \sum_{i=1}^G \sum_{r=1}^R Y_{ijk r}$ o total do j -ésimo local e k -ésimo ano.
Verifica-se que,

$$F_{(G(L_j A_k))} = \frac{QM_{G(L_j A_k)}}{QM_{Res}} \sim F_{((G-1), LA(G-1)(R-1))}, \text{ para todo } j \text{ e } k \text{ sob } H_{03}.$$

Ao nível α de significância, rejeita-se H_{03} quando,

$$F_{(G(L_j A_k))} = \frac{QM_{G(L_j A_k)}}{QM_{Res}} \geq F_{(\alpha, (G-1), LA(G-1)(R-1))}, \text{ para } j = 1, \dots, L \text{ e } k = 1, \dots, A.$$

O esquema da ANOVA, com aplicação do teste F , considerando a decomposição dos $(G-1)(L-1)(A-1)$ graus de liberdade da interação $G \times L \times A$ com $(G-1)$ graus de liberdade dos genótipos, para verificar o efeito de $G(LA)$ é dado na Tabela 2.3.

Tabela 2.3. Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com G

F.V.	GL	SQ	QM	F
Bd.LA	$LA(R-1)$	$SQ_{Bd.LA}$	$QM_{Bd.LA}$	
L	$(L-1)$	SQ_L	QM_L	
A	$(A-1)$	SQ_A	QM_A	
$L \times A$	$(L-1)(A-1)$	SQ_{LA}	QM_{LA}	
$G(L_1 A_1)$	$(G-1)$	$SQ_{G(L_1 A_1)}$	$QM_{G(L_1 A_1)}$	$QM_{G(L_1 A_1)} / QM_{Res}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$G(L_j A_k)$	$(G-1)$	$SQ_{G(L_j A_k)}$	$QM_{G(L_j A_k)}$	$QM_{G(L_j A_k)} / QM_{Res}$
$G(LA)$	$LA(G-1)$	$SQ_{G(LA)}$	$QM_{G(LA)}$	
Resíduo	$LA(G-1)(R-1)$	SQ_{Res}	QM_{Res}	
Total	$GLAR-1$	SQ_{Total}		

Têm-se que a $SQ_{G(LA)} = \sum_{k=1}^A \sum_{j=1}^L SQ_{G(L_j A_k)}$, com $LA(G-1)$ graus de liberdade e, observa-se que $SQ_{G(LA)} = SQ_{G \times L \times A} + SQ_{G \times L} + SQ_{G \times A} + SQ_G$, com $[(G-1)(L-1)(A-1) + (G-1)(L-1) + (G-1)(A-1) + (G-1)] = [LA(G-1)]$ graus de liberdade. Vê-se claramente por esta decomposição que o efeito não é puro, no que diz respeito à interação, pois à interação adiciona-se $SQ_{G \times L}$, $SQ_{G \times A}$ e SQ_G .

2.4.2 Decomposição da soma dos graus de liberdade da interação $G \times L \times A$ com os graus de liberdade dos fatores L , $L \times G$ e $L \times A$

Essa decomposição corresponde a obter o efeito do fator L dentro do fator G e dentro do fator A , representado por $L(GA)$. Suponha que o delineamento seja aleatorizado em blocos. Assim, o modelo adotado é:

$$Y_{ijk r} = \mu + b_r(l_j a_k) + g_i + l_{j(ik)} + a_k + (ga)_{ik} + \varepsilon_{ijk r}$$

em que:

$l_{j(ik)}$: é o efeito do j -ésimo local na presença do i -ésimo genótipo e k -ésimo ano.

O $l_{j(ik)}$ é definido como:

$$l_{j(ik)} = \mu_{j(ik)} - \mu_{i.k}, \text{ tal que } \sum_{j=1}^L l_{j(ik)} = 0, \text{ para todo } i \text{ e } k.$$

Sendo que $\mu_{j(ik)}$ refere-se à média populacional do j -ésimo local na presença do i -ésimo genótipo e k -ésimo ano e $\mu_{i.k}$ refere-se à média populacional do i -ésimo genótipo e k -ésimo ano.

As hipóteses testadas são:

- i) $H_{01} : g_i = 0$, para todo i
 $H_{a1} : g_i \neq 0$, para algum i
- ii) $H_{02} : a_k = 0$, para todo k
 $H_{a2} : a_k \neq 0$, para algum k
- iii) $H_{03} : l_{j(i,k)} = 0$, para todo j na presença de i e k
 $H_{a3} : l_{j(i,k)} \neq 0$, para algum j na presença de i e k

A soma de quadrados do local dentro do genótipo e do ano é dada por:

$$SQ_{L(G_i A_k)} = \frac{1}{R} \sum_{j=1}^L Y_{(i)j(k)}^2 - \frac{Y_{i.k}^2}{RL}$$

com $(L-1)$ graus de liberdade associados, sendo que

$$Y_{(i)j(k)} = \sum_{r=1}^R Y_{(i)j(k)r} \text{ o total do } j\text{-ésimo local dentro do } i\text{-ésimo genótipo e do } k\text{-ésimo ano;} \\ Y_{i.k} = \sum_{j=1}^L \sum_{r=1}^R Y_{ijkr} \text{ o total do } i\text{-ésimo genótipo e } k\text{-ésimo ano.}$$

Verifica-se que,

$$F_{(L(G_i A_k))} = \frac{QM_{L(G_i A_k)}}{QM_{Res}} \sim F_{((L-1), LA(G-1)(R-1))}, \text{ para todo } i \text{ e } k \text{ sob } H_{03}.$$

Ao nível α de significância, rejeita-se H_{03} quando,

$$F_{(L(G_i A_k))} = \frac{QM_{L(G_i A_k)}}{QM_{Res}} \geq F_{(\alpha, (L-1), LA(G-1)(R-1))}, \text{ para } i = 1, \dots, G \text{ e } k = 1, \dots, A.$$

O esquema da ANOVA, com aplicação do teste F , considerando a decomposição dos $(G-1)(L-1)(A-1)$ graus de liberdade da interação $G \times L \times A$ com $(L-1)$ graus de liberdade dos locais, para verificar o efeito de $L(GA)$ é dado na Tabela 2.4.

Tabela 2.4. Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com L

F.V.	GL	SQ	QM	F
Bd.LA	$LA(R-1)$	$SQ_{Bd.LA}$	$QM_{Bd.LA}$	
G	$(G-1)$	SQ_G	QM_G	
A	$(A-1)$	SQ_A	QM_A	
$G \times A$	$(G-1)(A-1)$	SQ_{GA}	QM_{GA}	
$L(G_1 A_1)$	$(L-1)$	$SQ_{L(G_1 A_1)}$	$QM_{L(G_1 A_1)}$	$QM_{L(G_1 A_1)} / QM_{Res}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$L(G_i A_k)$	$(L-1)$	$SQ_{L(G_i A_k)}$	$QM_{L(G_i A_k)}$	$QM_{L(G_i A_k)} / QM_{Res}$
$L(GA)$	$GA(L-1)$	$SQ_{L(GA)}$	$QM_{L(GA)}$	
Resíduo	$LA(G-1)(R-1)$	SQ_{Res}	QM_{Res}	
Total	$GLAR-1$	SQ_{Total}		

Têm-se que a $SQ_{L(GA)} = \sum_{k=1}^A \sum_{i=1}^G SQ_{L(G_i A_k)}$, com $GA(L-1)$ graus de liberdade e, observa-se que $SQ_{L(GA)} = SQ_{G \times L \times A} + SQ_{L \times G} + SQ_{L \times A} + SQ_L$, com $[(G-1)(L-1)(A-1) + (G-1)(L-1) + (L-1)(A-1) + (L-1)] = [GA(L-1)]$ graus de liberdade. Vê-se claramente por esta decomposição que o efeito também não é puro, no que diz respeito à interação, pois à interação adiciona-se $SQ_{L \times G}$, $SQ_{L \times A}$ e SQ_L .

2.4.3 Decomposição da soma dos graus de liberdade da interação $G \times L \times A$ com os graus de liberdade dos fatores A , $A \times G$ e $A \times L$

Essa decomposição corresponde a obter o efeito do fator A dentro do fator G e dentro do fator L , representado por $A(GL)$. Suponha que o delineamento seja aleatorizado em blocos. Assim, o modelo adotado é:

$$Y_{ijk r} = \mu + b_r(l_j a_k) + g_i + l_j + a_{k(ij)} + (gl)_{ij} + \varepsilon_{ijk r}$$

em que:

$a_{k(ij)}$: é o efeito do k -ésimo ano na presença do i -ésimo genótipo e j -ésimo local.

O $a_{k(ij)}$ é definido como:

$$a_{k(ij)} = \mu_{k(ij)} - \mu_{ij.}, \text{ tal que } \sum_{k=1}^A a_{k(ij)} = 0, \text{ para todo } i \text{ e } j.$$

Sendo que $\mu_{k(ij)}$ refere-se à média populacional do k -ésimo ano na presença do i -ésimo genótipo e j -ésimo local e $\mu_{ij.}$ refere-se à média populacional do i -ésimo genótipo e j -ésimo local.

As hipóteses testadas são:

- i) $H_{01} : g_i = 0$, para todo i
 $H_{a1} : g_i \neq 0$, para algum i
- ii) $H_{02} : l_j = 0$, para todo j
 $H_{a2} : l_j \neq 0$, para algum j
- iii) $H_{03} : a_{k(ij)} = 0$, para todo k na presença de i e j
 $H_{a3} : a_{k(ij)} \neq 0$, para algum j na presença de i e j

A soma de quadrados do ano dentro do genótipo e do local é dada por:

$$SQ_{A(G_i L_j)} = \frac{1}{R} \sum_{k=1}^A Y_{(ij)k.}^2 - \frac{Y_{ij..}^2}{RA}$$

com $(A - 1)$ graus de liberdade associados, sendo que

$Y_{(ij)k.} = \sum_{r=1}^R Y_{(ij)kr}$ o total do k -ésimo ano dentro do i -ésimo genótipo e do j -ésimo local;

$Y_{ij..} = \sum_{k=1}^A \sum_{r=1}^R Y_{ijk r}$ o total do i -ésimo genótipo e j -ésimo local.

Verifica-se que,

$$F_{(A(G_i L_j))} = \frac{QM_{A(G_i L_j)}}{QM_{Res}} \sim F_{((A-1), LA(G-1)(R-1))}, \text{ para todo } i \text{ e } j \text{ sob } H_{03}.$$

Ao nível α de significância, rejeita-se H_{03} quando,

$$F_{(A(G_i L_j))} = \frac{QM_{A(G_i L_j)}}{QM_{Res}} \geq F_{(\alpha, (A-1), LA(G-1)(R-1))}, \text{ para } i = 1, \dots, G \text{ e } j = 1, \dots, L.$$

O esquema da ANOVA, com aplicação do teste F , considerando a decomposição dos $(G - 1)(L - 1)(A - 1)$ graus de liberdade da interação $G \times L \times A$ com $(A - 1)$ graus de liberdade dos anos, para verificar o efeito de $A(GL)$ é dado na Tabela 2.5.

Têm-se que a $SQ_{A(GL)} = \sum_{j=1}^L \sum_{i=1}^G SQ_{A(G_i L_j)}$, com $GL(A - 1)$ graus de liberdade e, observa-se que $SQ_{A(GL)} = SQ_{G \times L \times A} + SQ_{A \times G} + SQ_{A \times L} + SQ_A$, com $[(G - 1)(L - 1)(A - 1) + (G - 1)(L - 1) + (G - 1)(A - 1) + (G - 1)] = [GL(A - 1)]$ graus de liberdade. Vê-se claramente por esta decomposição que o efeito também não é puro, no que diz respeito à interação, pois à interação adiciona-se $SQ_{A \times G}$, $SQ_{A \times L}$ e SQ_A .

Tabela 2.5. Esquema da ANOVA com teste F obtida pela decomposição de $G \times L \times A$ com A

F.V.	GL	SQ	QM	F
Bd.LA	$LA(R-1)$	$SQ_{Bd.LA}$	$QM_{Bd.LA}$	
G	$(G-1)$	SQ_G	QM_G	
L	$(L-1)$	SQ_L	QM_L	
$G \times L$	$(G-1)(L-1)$	SQ_{GL}	QM_{GL}	
$A(G_1L_1)$	$(A-1)$	$SQ_{A(G_1L_1)}$	$QM_{A(G_1L_1)}$	$QM_{A(G_1L_1)}/QM_{Res}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$A(G_iL_j)$	$(A-1)$	$SQ_{A(G_iL_j)}$	$QM_{A(G_iL_j)}$	$QM_{A(G_iL_j)}/QM_{Res}$
$A(GL)$	$GL(A-1)$	$SQ_{A(GL)}$	$QM_{A(GL)}$	
Resíduo	$LA(G-1)(R-1)$	SQ_{Res}	QM_{Res}	
Total	$GLAR-1$	SQ_{Total}		

2.5 Análise *multiway*

Na análise de dados multivariados, os dados são dispostos em tabelas de dupla entrada ou numa matriz. No entanto, para uma ampla variedade de dados a estrutura mais apropriada seria uma tabela ou arranjo de três entradas.

Em arranjos de duas entradas é usual fazer uma diferenciação entre as partes do arranjo, como linhas e colunas. Tal distinção também é feita para arranjos de três entradas e uma divisão adequada é: fatias horizontais, verticais e frontais (Figura 2.2). É possível fatiar um arranjo de três dimensões $\mathbf{X}$ ($I \times J \times K$) de três maneiras diferentes, o símbolo “ $_$ ” indica que $\mathbf{X}$ é um arranjo de três entradas. A primeira maneira origina as fatias horizontais $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_I$, todas de dimensões ($J \times K$). A segunda maneira origina as fatias verticais $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J$, todas de dimensões ($I \times K$). E a última maneira origina as fatias frontais $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$, com todas as dimensões ($I \times J$). É conveniente usar esta notação, mas às vezes pode ser confusa, por exemplo, não se sabe se $\mathbf{X}_2$ é a segunda fatia horizontal, vertical ou frontal. Com o intuito de evitar tal ambiguidade, utiliza-se, por exemplo, $\mathbf{X}_{i=2}$ para a segunda fatia horizontal.

Figura 2.2. Particionamento de um arranjo de três entradas em fatias horizontais, verticais e frontais

Os trabalhos de análises *multiway* tiveram origem na psicometria e o trabalho pioneiro surgiu em meados do século XX e terminou por volta de 1980, quando os mais importantes modelos *multi-way* e os seus algoritmos foram introduzidos.

As primeiras ideias em análises *multiway* foram publicadas por CATTELL (1944), CATTELL (1952). O princípio da parcimônia de THURSTONE (1935) afirma que uma estrutura simples deve ser encontrada para descrever uma matriz de dados ou uma matriz de correlações com a ajuda de fatores. Já em uma análise simultânea de várias matrizes, CATTELL (1944) propôs utilizar o princípio do perfil paralelo proporcional que afirma que um conjunto comum de fatores pode ser encontrado e podem ser ajustados com diferentes dimensões para muitas matrizes ao mesmo tempo. Isso é o mesmo que encontrar um arranjo de três entradas.

O trabalho mais importante de Cattell, segundo SMILDE *et al.* (2004) foi o de 1952 em que o autor define o arranjo *multiway*. Cattell define objeto, circunstância/tempo, atributo, escala e observador como as cinco entradas para um arranjo *multiway* idealizado e, por razões práticas, reduziu-as a um arranjo de três entradas com pessoas, atributos e circunstâncias.

A decomposição de um arranjo ou uma matriz de três entradas foi apresentado primeiramente por TUCKER (1963), TUCKER (1964), TUCKER (1966). Nestes trabalhos, a decomposição procurava encontrar matrizes de carga **A**, **B** e **C** e um arranjo núcleo **G** de três entradas, que foram introduzidos com um exemplo numérico hipotético de 12 indivíduos, 9 tratamentos e 5 observadores. LEVIN (1965) em seu artigo mostrou uma semelhança entre o núcleo do modelo Tucker e a matriz de autovalores na decomposição em valor singular.

Outros modelos de três entradas foram introduzidos de forma independente por CARROLL e CHANG (1970) definido por CANDECOMP (*Canonical Decomposition*) e por HARSHMAN (1970) que definiu por PARAFAC (*Parallel Factor Analysis*). No entanto, a modelagem que envolve estes dois modelos é a mesma. A ideia básica do modelo é a utilização dos mesmos fatores que descrevem a variação em diversas matrizes simultaneamente, embora com diferentes coeficientes de ponderação para cada matriz. Esta é exatamente a ideia do perfil paralelo proporcional de CATTELL (1944).

Um modelo PARAFAC com três entradas consiste em determinar matrizes **A**, **B** e **C** que possuem o mesmo número de fatores (Figura 2.3). Este modelo geralmente produz eixos de coordenadas originais (o que significa que não há liberdade de rotação na orientação dos vetores de carga).

Figura 2.3. Decomposição de um arranjo de três entradas propostas por Harshman (1970) e Carrol e Chang (1970)

2.5.1 Modelos PARAFAC

O modelo de componentes com três entradas (PARAFAC) e o modelo CANDECOMP são aproximadamente os mesmos e serão aqui, nomeados de PARAFAC que segundo SANCHEZ e KOWALSKI (1990) podem ser chamados de decomposição trilinear. O modelo PARAFAC utiliza diferentes notações, sendo as mais usadas as que apresentam somatórios e componentes simultâneos e as menos usadas são as que apresentam produto de Kronecker, produto tensorial (produto de Hadamard) e produtos de Khatri-Rao (RAO e MITRA, 1971), (SCHOTT, 1997).

O modelo PARAFAC surge a partir da generalização da decomposição em valor singular. Um modelo de duas entradas para a matriz **X** ($I \times J$), com elementos x_{ij} , com base na sua decomposição em valor singular ($\mathbf{X} = \mathbf{AGB}'$) truncada em R componentes é:

$$x_{ij} = \sum_{r=1}^R a_{ir} g_{rr} b_{jr} + e_{ij}; \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad r = 1, \dots, R \quad (2.3)$$

em que:

a_{ir} : é o elemento que está na i -ésima linha e r -ésima coluna da matriz de autovetores **A** ($I \times R$);

g_{rr} : é o elemento que está na r -ésima linha e r -ésima coluna da matriz de autovalores $\mathbf{G}$;

b_{jr} : é o elemento que está na j -ésima linha e r -ésima coluna da matriz de autovetores $\mathbf{B}$ ($J \times R$);

e_{ij} : é o elemento que está na i -ésima linha e j -ésima coluna da matriz residual, com variação não explicada pelo modelo com R componentes.

O modelo (2.3) na forma matricial é dado por:

$$\mathbf{X} = \mathbf{A}\mathbf{G}\mathbf{B}' + \mathbf{E}. \quad (2.4)$$

Uma matriz de três entradas $\underline{\mathbf{X}}$ ($I \times J \times K$), com elementos x_{ijk} , a forma generalizada para um modelo PARAFAC é:

$$x_{ijk} = \sum_{r=1}^R a_{ir}b_{jr}c_{kr} + e_{ijk} \quad (2.5)$$

em que:

R : é o número de componentes usados no modelo PARAFAC;

a_{ir} : é o elemento que está na i -ésima linha e r -ésima coluna da matriz de componentes $\mathbf{A}$;

b_{jr} : é o elemento que está na j -ésima linha e r -ésima coluna da matriz de componentes $\mathbf{B}$;

c_{kr} : é o elemento que está na k -ésima linha e r -ésima coluna da matriz de componentes $\mathbf{C}$;

e_{ijk} : é o elemento que está na i -ésima linha, j -ésima coluna e k -ésimo tubo do arranjo residual, com a variação não explicada pelo modelo com R componentes.

O modelo (2.4) é trilinear: fixando dois parâmetros (por exemplo, a e b) e x_{ijk} é expresso como uma função linear dos demais parâmetros (por exemplo, c). Na Figura 2.4 apresenta uma descrição do modelo.

Figura 2.4. Modelo PARAFAC com R componentes

Uma outra forma de reescrever o modelo (2.5) é em termos das matrizes $\mathbf{X}_k$, em que $\mathbf{X}_k$ é a k -ésima fatia frontal ($I \times J$) do arranjo de três entradas $\underline{\mathbf{X}}$ ($I \times J \times K$):

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k = c_{k1}\mathbf{a}_1\mathbf{b}_1' + \dots + c_{kR}\mathbf{a}_R\mathbf{b}_R' + \mathbf{E}_k \quad (2.6)$$

em que

$\mathbf{D}_k$: é uma matriz diagonal com a k -ésima linha de $\mathbf{C}$;

$\mathbf{a}_r$ e $\mathbf{b}_r$ são as r -ésimas colunas de $\mathbf{A}$ e $\mathbf{B}$ respectivamente;

$\mathbf{E}_k$ é um termo residual.

Cada $\mathbf{X}_k$ é modelada usando os mesmos componentes, mas com diferentes ponderações, representados por $\mathbf{D}_k$. Assim, todas as fatias $\mathbf{X}_k$ são modeladas com perfis paralelos e proporcionais $d_{k1}\mathbf{a}_1\mathbf{b}'_1, \dots, d_{kR}\mathbf{a}_R\mathbf{b}'_R$. Desta forma, é possível escrever o modelo utilizando uma notação de componentes simultâneos. Pela simetria do modelo (2.6), há três formas completamente equivalentes de escrever o modelo PARAFAC, que são descritos em termos das fatias frontais, horizontais e verticais, respectivamente:

$$\begin{aligned}\mathbf{X}_k &= \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}', \quad k = 1, \dots, K \\ \mathbf{X}_i &= \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}', \quad i = 1, \dots, I \\ \mathbf{X}_j &= \mathbf{A}\mathbf{D}_j(\mathbf{B})\mathbf{C}', \quad j = 1, \dots, J\end{aligned}\tag{2.7}$$

em que as matrizes $\mathbf{D}_k(\mathbf{C})$, $\mathbf{D}_i(\mathbf{A})$ ou $\mathbf{D}_j(\mathbf{B})$ são interpretadas com operadores que extrai as k -ésimas ou i -ésimas ou j -ésimas colunas das matrizes de cargas apropriadas ($\mathbf{C}$, $\mathbf{A}$ e $\mathbf{B}$, respectivamente).

Quando as matrizes de componentes não tem colunas proporcionais, o modelo PARAFAC fornece estimativas únicas com uma condição necessária para o modelo, ou seja, as estimativas de $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ não podem ser modificadas sem mudar o resíduo (não tem liberdade de rotação) (HARSHMAN¹, 1972 apud KROONENBERG (2008)). Esta propriedade é conhecida como propriedade dos eixos intrínsecos porque a base de vetores de orientação encontrados são únicos.

Os parâmetros $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ podem ser estimados com diferentes algoritmos. Os fatores são estimados simultaneamente, ao contrário da análise de componentes principais, em que os componentes podem ser estimados um de cada vez. Isto ocorre porque os componentes no modelo PARAFAC são não ortogonais e, portanto, dependem um do outro. A estimação dos componentes do modelo PARAFAC de forma sequencial como em análise de componentes principais, fornece resultados diferentes quando comparados com a estimativa de componentes simultâneos, e a aproximação sequencial não fornece uma solução de mínimos quadrados.

2.5.2 Modelos de Tucker

Ledyard Tucker foi um dos pioneiros na análise *multiway*. Em 1964 e 1966, Tucker propôs uma série de modelos conhecidos como análise de componentes principais de N -entradas ou modelos Tucker. KROONENBERG e DE LEEUW (1980) e KROONENBERG (1983) apresentam um extenso estudo dos modelos de Tucker.

2.5.2.1 Modelos Tucker3

A partir da equação (2.3), uma possível generalização do modelo de componentes principais para dados de duas entradas é a utilização de uma matriz núcleo não diagonal $\tilde{\mathbf{G}}$, o que resulta no seguinte modelo:

$$x_{ij} = \sum_{p=1}^P \sum_{q=1}^Q a_{ip} \tilde{g}_{pq} b_{jq} + e_{ij}\tag{2.8}$$

ou, em notação matricial,

$$\mathbf{X} = \mathbf{A}\tilde{\mathbf{G}}\mathbf{B}' + \mathbf{E}\tag{2.9}$$

em que “ $\sim$ ” na parte superior do $\mathbf{G}$ indica as diferentes matrizes centrais das equações (2.9) e (2.4), e $\tilde{g}_{pq}$ é um elemento de $\tilde{\mathbf{G}}$. Ao contrário do modelo da equação (2.3), não há nenhuma exigência no modelo (2.8) para que $\mathbf{A}$ e $\mathbf{B}$ tenham o mesmo número de componentes, p e q podem assumir diferentes valores até P e Q e $\tilde{\mathbf{G}}$ seja de dimensão $(P \times Q)$, fazendo com que o número de componentes seja diferente nos dois

¹HARSHMAN, R.A. Determination and proof of minimum uniqueness conditions for PARAFAC1. **UCLA Working Papers in Phonetic**, Ann Arbor, v.22, p.30-44, 1972.

modos. A matriz núcleo não diagonal $\tilde{\mathbf{G}}$ significa explicitamente que no modelo existem interações entre os fatores. Esta é uma propriedade importante dos modelos Tucker em geral. Na análise de componentes principais tradicional, vetores de cargas só interagem aos pares. Para exemplificar, o segundo vetor de escores interage com o segundo vetor de cargas pela magnitude definida pelo segundo valor singular. No modelo (2.8) todos os vetores podem interagir. Por exemplo, o primeiro vetor de escores interage com o terceiro vetor de carga com uma magnitude definida pelo elemento $\tilde{g}_{13}$.

O modelo (2.8) pode ser generalizado para um arranjo de três entradas $\underline{\mathbf{X}}$, com elementos x_{ijk}

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \quad (2.10)$$

em que

e_{ijk} é um elemento do arranjo $\underline{\mathbf{E}}$ ($I \times J \times K$);

a_{ip} , b_{jq} e c_{kr} são elementos típicos das matrizes de cargas $\mathbf{A}$ ($I \times P$), $\mathbf{B}$ ($J \times Q$) e $\mathbf{C}$ ($K \times R$), respectivamente;

g_{pqr} é um elemento típico do arranjo núcleo $\underline{\mathbf{G}}$ ($P \times Q \times R$).

Este é o modelo Tucker3 de $\underline{\mathbf{X}}$ (P, Q, R), em que a notação (P, Q, R) é usada para indicar que o modelo tem P , Q , R fatores em três entradas diferentes. A Figura 2.5 mostra graficamente este modelo.

Figura 2.5. Representação gráfica do modelo Tucker3

Com uma matrização adequada para os arranjos $\underline{\mathbf{X}}$ e $\underline{\mathbf{G}}$ com relação ao primeiro modo e a notação do produto de Kronecker, o modelo Tucker3 pode ser escrito como (SMILDE *et al.*, 2004):

$$\mathbf{X} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})' + \mathbf{E} \quad (2.11)$$

em que

$\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_K]$ é uma matrização de $\underline{\mathbf{X}}$ com dimensão $(I \times JK)$ e $\mathbf{X}_k$ como definido na equação (2.6);

$\mathbf{E}$ é definido similarmente;

$\mathbf{G} = [\mathbf{G}_1 \mathbf{G}_2 \dots \mathbf{G}_R]$ é uma matrização do arranjo núcleo $\underline{\mathbf{G}}$ com dimensão $(P \times QR)$, em que $\mathbf{G}_r$ é a r -ésima fatia frontal de dimensão $(P \times Q)$ de $\underline{\mathbf{G}}$;

$\otimes$ é o produto de kronecker entre as matrizes $\mathbf{C}$ e $\mathbf{B}$.

2.5.2.2 Propriedades do modelo Tucker3

O modelo Tucker3 tem liberdade rotacional, facilmente observado a partir do modelo (2.11):

$$\mathbf{X} = \mathbf{A}\mathbf{G}(\mathbf{C}' \otimes \mathbf{B})' + \mathbf{E}$$

$$\begin{aligned}
&= \mathbf{A} \mathbf{T}_A \mathbf{T}_A^{-1} \mathbf{G} (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E} \\
&= \tilde{\mathbf{A}} \tilde{\mathbf{G}} (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}
\end{aligned} \tag{2.12}$$

em que $\mathbf{T}_A$ é uma matriz ortonormal qualquer, $\tilde{\mathbf{A}} = \mathbf{A} \mathbf{T}_A$ e $\tilde{\mathbf{G}} = \mathbf{T}_A^{-1} \mathbf{G}$. A transformação da matriz de cargas $\mathbf{A}$ pode ser definida de forma semelhante para $\mathbf{B}$ e $\mathbf{C}$, usando $\mathbf{T}_B$ e $\mathbf{T}_C$, respectivamente. Assim, como há liberdade de rotação, o componente de ortogonalidade das matrizes pode ser obtido sem nenhum custo para o ajuste, definindo as matrizes $\mathbf{T}_A$, $\mathbf{T}_B$ e $\mathbf{T}_C$ apropriadamente. Portanto, o modelo Tucker3 não fornece matrizes de componentes únicos devido a liberdade de rotação.

É conveniente construir as matrizes de componentes ortogonais, ou seja, $\mathbf{A} \mathbf{A}' = \mathbf{I}_{(I \times I)}$, $\mathbf{B} \mathbf{B}' = \mathbf{I}_{(J \times J)}$ e $\mathbf{C} \mathbf{C}' = \mathbf{I}_{(K \times K)}$. A soma dos quadrados dos elementos de um arranjo núcleo associado com a combinação de certos fatores representa a quantidade de variação explicada pela combinação de fatores em diferentes entradas. Se $\underline{\mathbf{X}}$ ($I \times J \times K$) é um arranjo de três entradas modelado por um modelo Tucker3 (P, Q, R) e se $\hat{\underline{\mathbf{X}}}$ representa a parte ajustada de $\underline{\mathbf{X}}$, então segue que (KROONENBERG, 1984):

$$\begin{aligned}
\|\underline{\mathbf{X}}\|^2 &= \|\hat{\underline{\mathbf{X}}}\|^2 + \|\underline{\mathbf{E}}\|^2 \\
\|\hat{\underline{\mathbf{X}}}\|^2 &= \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}^2
\end{aligned} \tag{2.13}$$

em que $\underline{\mathbf{E}}$ é um arranjo dos resíduos e $\|\cdot\|$ é a norma de Euclides, definida como: $\|\underline{\mathbf{X}}\| = \sqrt{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$, para um arranjo de três entradas $\underline{\mathbf{X}}$ ($I \times J \times K$). Em outras palavras, a equação (2.13) significa que a variação de $\underline{\mathbf{X}}$ é dividida em uma variação não explicada ($\underline{\mathbf{E}}$) e uma variação explicada pelo modelo ($\hat{\underline{\mathbf{X}}}$). Além disso, a soma de quadrados ajustada pode ser dividida em partes relacionadas a cada combinação de componentes em diferentes direções.

A liberdade de rotação do modelo Tucker3 também pode ser utilizada para rotacionar o arranjo núcleo para uma estrutura simples, como também é comum em análises de duas entradas. A imposição de restrições $\mathbf{A} \mathbf{A}' = \mathbf{I}_{(I \times I)}$, $\mathbf{B} \mathbf{B}' = \mathbf{I}_{(J \times J)}$ e $\mathbf{C} \mathbf{C}' = \mathbf{I}_{(K \times K)}$ não é suficiente para a obtenção de uma única solução. Para obter estimativas únicas dos parâmetros, além de impor que as matrizes de carga sejam ortogonais, $\mathbf{A}$ deverá conter também os autovetores de $\mathbf{X}(\mathbf{C} \mathbf{C}' \otimes \mathbf{B} \mathbf{B}') \mathbf{X}'$ correspondendo aos autovetores desta mesma matriz. Restrições semelhantes poderiam ser colocadas em $\mathbf{B}$ e $\mathbf{C}$.

2.5.2.3 Modelos de Tucker2

Nos modelos Tucker3 (P, Q, R) de um arranjo $\underline{\mathbf{X}}$ ($I \times J \times K$) todos três modos são reduzidos, ou seja, usualmente $P < I$, $Q < J$ e $R < K$. Também existem modelos em que apenas dois dos três modos são reduzidos, neste caso são chamados de modelos de Tucker2. Isto origina três modelos especiais, dependendo de qual modo é reduzido. Suponha que um modelo de Tucker3 é ajustado para $\underline{\mathbf{X}}$ ($I \times J \times K$), mas $\mathbf{C}$ é escolhida para ser a matriz identidade $\mathbf{I}$ de dimensão $K \times K$. Assim, não há interesse na redução no terceiro modo porque a base não é modificada. Então, o modelo de Tucker2 pode ser escrito como:

$$\mathbf{X}_{(K \times IJ)} = \mathbf{I} \mathbf{G}_{(K \times PQ)} (\mathbf{B} \otimes \mathbf{A})' + \mathbf{E} = \mathbf{G}_{(K \times PQ)} (\mathbf{B} \otimes \mathbf{A})' + \mathbf{E} \tag{2.14}$$

em que

$\mathbf{G}_{(K \times PQ)}$ é uma matriz ($K \times PQ$) e esta é uma versão matrizada própria para o arranjo núcleo “estendido” $\underline{\mathbf{G}}$ ($P \times Q \times K$);

$\mathbf{X}_{(K \times IJ)}$ é uma versão matrizada própria para $\underline{\mathbf{X}}$.

Na notação de somatório o modelo de Tucker2 é:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q a_{ip} b_{jq} g_{pqk} + e_{ijk} \tag{2.15}$$

em que x_{ijk} , a_{ip} , b_{jq} são elementos de $\underline{\mathbf{X}}$ ($I \times J \times K$), $\mathbf{A}$ ($I \times P$) e $\mathbf{B}$ ($J \times Q$), respectivamente. Além disso, g_{pqk} é um elemento do arranjo núcleo estendido $\underline{\mathbf{G}}$ ($P \times Q \times K$). Ao comparar as equações (2.15) com (2.10), verifica-se a ausência de um somatório. Isto ocorre devido a redução de um dos modos. A Figura 2.6 é a representação do modelo de Tucker2 que quando comparada com a Figura 2.5, nota-se que a matriz $\mathbf{C}$ agora está ausente, ou seja, $\mathbf{C} = \mathbf{I}$.

Figura 2.6. Representação gráfica do modelo Tucker2

A equação (2.14) mostra que o modelo de Tucker2 também tem liberdade de rotação, pois $\mathbf{G}$ pode ser pós-multiplicado $\mathbf{U} \otimes \mathbf{V}$ e $(\mathbf{B} \otimes \mathbf{A})'$ pré-multiplicada por $(\mathbf{U} \otimes \mathbf{V})^{-1}$ resultando em $(\mathbf{B}(\mathbf{U}')^{-1} \otimes \mathbf{A}(\mathbf{V}')^{-1})'$ sem alterar o ajuste. Por isso, as matrizes dos componentes $\mathbf{A}$ e $\mathbf{B}$ podem ser ortogonais sem mudança no ajuste.

2.5.2.4 Modelos de Tucker1

Como mencionado no modelo de Tucker3, três modos são reduzidos e no modelo de Tucker2 têm dois modos reduzidos. Assim, o modelo de Tucker1 tem apenas um modo reduzido. Existem três modelos diferentes de Tucker1 para um arranjo $\underline{\mathbf{X}}$ ($I \times J \times K$), dependendo do modelo que foi reduzido. No modelo de Tucker2 na Figura 2.6, o terceiro modo não foi reduzido, em substituição de $\mathbf{C}$ por $\mathbf{I}$. Se o modo $\mathbf{B}$ também for substituído por $\mathbf{I}$, ambos o segundo e o terceiro modo não são reduzidos. O modelo de Tucker1 resultante é:

$$\mathbf{X}_{(K \times IJ)} = \mathbf{I} \mathbf{G}_{(K \times JP)} (\mathbf{I} \otimes \mathbf{A})' + \mathbf{E} \quad (2.16)$$

em que $\mathbf{G}_{(K \times JP)}$ tem dimensão $(K \times JP)$ e somente o terceiro modo é reduzido. O modelo (2.16) é modificado se $\underline{\mathbf{X}}$ for matricizada da forma:

$$\mathbf{X}_{(I \times JK)} = \mathbf{A} \mathbf{G}_{(P \times JK)} + \mathbf{E} \quad (2.17)$$

em que

$\mathbf{X}_{(I \times JK)}$ é a matrização de $\underline{\mathbf{X}}$ de dimensão $(I \times JK)$;

$\mathbf{G}_{(P \times JK)}$ é uma matrização de $\underline{\mathbf{G}}$ de dimensão $(P \times JK)$.

Se $\mathbf{X}_{(I \times JK)}$ for substituída por $\mathbf{X}$, $\mathbf{A}$ por $\mathbf{T}$ e $\mathbf{G}_{(P \times JK)}$ por $\mathbf{P}'$, então obtém-se uma solução pela análise de componentes principais (ACP) usual. O modelo de Tucker1 é exatamente uma ACP na matrização apropriada de $\underline{\mathbf{X}}$. Desta forma, os algoritmos para encontrar $\mathbf{A}$ e $\mathbf{G}_{(P \times JK)}$ estão disponíveis, como por exemplo, a decomposição por valores singulares.

Na notação de somatório o modelo de Tucker1 é:

$$x_{ijk} = \sum_{p=1}^P a_{ip} g_{pjk} + e_{ijk} \quad (2.18)$$

em que x_{ijk} , a_{ip} , g_{pjk} e e_{ijk} são elementos de $\underline{\mathbf{X}}$, componente da matriz $\mathbf{A}$ ($I \times P$), elemento do arranjo núcleo estendido $\underline{\mathbf{G}}$ ($P \times J \times K$) e matriz de erros $\underline{\mathbf{E}}$, respectivamente. Em comparação com a equação (2.10), dois somatórios estão ausentes. Isto se deve ao fato de reduzir dois modos.

A equação (2.17) mostra que o modelo de Tucker1 também tem liberdade de rotação. Desta forma, $\mathbf{A}$ pode ser escolhida ortogonal sem perda na proporção de variabilidade total explicada pelo modelo. A Figura 2.7 apresenta graficamente o modelo de Tucker1. Como no modelo de Tucker2, há diferentes modelos de Tucker1 para certo conjunto de dados dependendo de qual modo será reduzido.

Figura 2.7. Representação gráfica do modelo Tucker1, com primeiro modo reduzido

2.5.3 Relações entre os modelos de componentes de três entradas

Os modelos de três entradas, bem como suas propriedades que já foram apresentadas, pode ser uma difícil escolha do melhor modelo para um principiante. É importante saber, que para determinadas situações, alguns modelos se adequam melhor e para isso é preciso entender bem suas diferenças e similaridades.

O primeiro passo é compreender como os modelos PARAFAC e Tucker3 estão relacionados. No modelo PARAFAC os eixos são únicos enquanto que no modelo Tucker3 isto não ocorre devido a liberdade de rotação. O modelo Tucker3 pode ser transformado (rotacionado) e simplificado para se assemelhar a um modelo PARAFAC, havendo pouca ou nenhuma perda de ajuste. As propriedades estatísticas dos dados, os ruídos e erros sistemáticos são uma importante regra na escolha do modelo.

Em determinada aplicação, nem sempre encontra-se uma resposta clara ou até mesmo um modelo adequado de três entradas. Pode acontecer de um ou mais modelos precisarem ser ajustados para se encontrar este melhor modelo. Em muitos casos, os modelos apresentam-se muito bons e sua seleção é baseada em critérios práticos como a precisão de algoritmo ou velocidade de ajuste. Assim, faz-se necessário bom conhecimento das opções de modelos de três entradas com suas propriedades e relações, a fim de fazer escolhas objetivas dos modelos adequados para cada situação.

2.5.3.1 Hierarquia dos modelos PARAFAC e Tucker3

O modelo PARAFAC geral de R -componentes de um arranjo $\underline{\mathbf{X}}$ ($I \times J \times K$) é dado pela equação (2.5). Acrescentando o termo g_{pqr} ao somatório com $g_{pqr} = 1$ caso $p = q = r$ e $g_{pqr} = 0$, caso contrário, tem-se que o modelo (2.5) pode ser reescrito como a equação (2.10) que é o modelo de Tucker3 de $\underline{\mathbf{X}}$. Logo, o modelo PARAFAC pode ser entendido como um modelo de Tucker3 restrito (Figura 2.8). Por esta razão, há uma hierarquia entre os modelos PARAFAC e Tucker3. Um modelo PARAFAC com R -componentes sempre tem um ajuste pior que o modelo de Tucker3 (R, R, R) ou apresenta mesmo ajuste. No entanto, isto não significa que o modelo de Tucker3 (R, R, R) seja o preferido. Na Figura 2.8 a linha tracejada no arranjo $\underline{\mathbf{G}}$, refere-se a diagonal de um cubo, chamada de superdiagonal.

Nota-se intuitivamente que o modelo de Tucker3 também pode ser representado como um modelo PARAFAC restrito, embora use muito mais componentes.

Figura 2.8. Modelo PARAFAC escrito como um modelo de Tucker3

2.6 Graus de liberdade dos modelos *multiway*

Segundo KROONENBERG (2008), os autores Weesie e Van Houwelingen² (1983) foram os primeiros que atribuíram graus de liberdade aos ajustes dos modelos de três entradas. Eles discutiram os graus de liberdade para o modelo Tucker3, mas o princípio pode ser estendido aos modelos de Tucker2 e PARAFAC. TIMMERMAN e KIERS (2000) apresentam a seguinte definição para os graus de liberdade (df):

$$df = n^{\circ} \text{ de observações} - n^{\circ} \text{ de parâmetros livres} \quad (2.19)$$

com o número de parâmetros livres, f_p ,

$$\begin{aligned} f_p = & n^{\circ} \text{ de parâmetros independentes na matriz de componentes observada} \\ & + n^{\circ} \text{ de elementos no arranjo núcleo} \\ & - n^{\circ} \text{ de indeterminação rotacional para matrizes de componente (Tucker) ou} \\ & - n^{\circ} \text{ de restrições de comprimento unitário (PARAFAC)} \end{aligned} \quad (2.20)$$

O número de observações é calculado diretamente, ou seja, conta-se apenas o número de observações. Em arranjos de três entradas o número de observações é IJK . Entretanto, ao calcular graus de liberdade, deve-se levar em consideração a regra do produto máximo. Tal regra se baseia “em casos em que o tamanho de uma das entradas seja maior que o produto das outras duas, devem ser feitos ajustes especiais (para o cálculo do número de parâmetros livres), isso porque segundo KIERS e HARSHMAN (1997) em tais casos o modo maior pode ser reduzido consideravelmente sem perda de informação para os componentes das outras duas entradas e para o arranjo núcleo. Especificamente, quando $I > JK$, os dados podem ser reduzidos a um conjunto de dados de dimensão $JK \times J \times K$. Então, em tais casos, no cálculo do número de parâmetros livres f_p , I deve ser substituído por JK ” (CEULEMANS e KIERS, 2006). Entretanto, uma consequência disto está, por exemplo, que para ajustar perfeitamente um arranjo de dados $I \times J \times K$ com $I > JK$, precisa-se de que JK , J , K componentes para as três entradas ajustarem perfeitamente aos dados. Assim, no cálculo dos graus de liberdade, isto deve ser levado em consideração.

Para os modelos *multiway*, o número de parâmetros livres é o seguinte:

$$\begin{aligned} \text{PARAFAC } f_p &= I \times R + J \times R + K \times R + R - 3R \\ \text{Tucker2 } f_p &= I \times P + J \times Q + P \times Q \times K - P^2 - Q^2 \\ \text{Tucker3 } f_p &= I \times P + J \times Q + K \times R + P \times Q \times R - P^2 - Q^2 - R^2. \end{aligned} \quad (2.21)$$

2.7 Dimensionalidade de um modelo de Tucker

Na seleção de dimensionalidade de um modelo de Tucker é importante observar que nem todas as combinações de dimensões são manejáveis. Isto deve-se a cada passo do algoritmo, pois a matriz

²WEESIE, J.; VAH HOUWELINGEN, H. **GEPCAM user's manual**: Generalized principal components analysis with missing values. Utrecht: University of Utrecht, Institute of Mathematical Statistics, 1983. 47p. Technical report.

para a qual os autovalores são calculados tem o posto igual ao produto do número de componentes dos outros modos. Caso um posto seja menor que o modo atual, há falha no algoritmo pois terá uma solução redundante, ou seja, existe outra solução que explica a mesma variabilidade, mas com um número menor de componentes. Um outro aspecto a ser notado é que o posto do arranjo núcleo é o posto do arranjo de três entradas ortogonal, e este pode ser ortogonal somente quando o produto dos números de componentes forem maiores ou iguais ao número de componentes no modo restante (WANSBEEK e VERHEES, 1989).

2.7.1 Procedimento *DifFit* de Timmerman-Kiers

Dado um modelo de Tucker3, TIMMERMAN e KIERS (2000) sugeriu um procedimento para selecionar a dimensionalidade semelhante ao *Scree plot* na análise de componentes principais. Inicialmente (1° filtro) selecionam-se todos os possíveis modelos de Tucker3 que satisfazem a seguinte condição proposta por KRUSKAL (1989):

$$\begin{aligned} P &\leq QR \\ Q &\leq PR \\ R &\leq PQ \end{aligned} \tag{2.22}$$

O próximo passo (2° filtro), consiste em selecionar, dentro de uma determinada classe de modelos de Tucker3, com o mesmo número total de componentes $S = P + Q + R$, o modelo que tem a maior proporção da soma de quadrados ajustada, SQ_S , ou equivalentemente, a menor soma de quadrados residual ou deviance. Para calcular classes com S diferente, calcula-se $dif_S = SQ_S - SQ_{S-1}$. Leva-se em consideração apenas os dif_S que são consecutivamente (em ordem crescente) maiores. TIMMERMAN e KIERS (2000) definiram ainda um valor de relevância (*salience value*): $b_S = dif_S / dif_{S*}$, em que dif_{S*} é o maior valor depois de dif_S . Assim, seleciona-se o modelo que têm o valor de b_S mais alto, sendo este procedimento conhecido de critério *DifFit*.

Neste mesmo trabalho, os autores propuseram um ponto de corte para dif_S , não levando em consideração aqueles modelos que apresentam valores abaixo deste limite. O dif_S deve ser maior que a proporção média da variabilidade explicada, tomada sobre todos possíveis valores de S (tal proporção dada por $\|\mathbf{Z}\|/S_{min}$, em que $S_{min} = \min(I; JK) + \min(J; IK) + \min(K; IJ) - 3$. Isto é equivalente ao critério de Kaiser da análise de componentes principais (JOHNSON e WICHERN, 1998); (BARROSO e ARTES, 2003), em que os autovalores devem ser maiores que o valor médio da variância explicada por cada componente.

A visualização do procedimento de Timmerman-Kiers é feita pelo *Scree plot multiway*, construído a partir da versão *Scree plot*, no qual as somas de quadrados residuais para cada dimensionalidade são colocadas em gráficos contra a soma S do número de componentes em cada uma das entradas. O procedimento de *DifFit* é essencialmente uma maneira para definir o número ótimo S no *Scree plot multiway*, e assim, a dimensionalidade S ótima do modelo é obtida quando a variação da explicação entre dimensionalidades consecutivas passa ser pequena. TIMMERMAN e KIERS (2000) apresentaram este procedimento para o modelo de Tucker3, mas trabalha da mesma forma para outros modelos de Tucker.

Na proposta inicial destes autores, a soma de quadrados residual para cada modelo era calculado ajustando o modelo de Tucker3 para cada dimensionalidade fixa por meio do algoritmo ALS (mínimos quadrados alternados). Entretanto, KIERS e KINDEREN (2003) mostraram que é suficiente aproximar os modelos por um único cálculo, usando o método original de mínimos quadrados de Tucker, e calcular a soma de quadrados residual para os modelos em consideração usando o arranjo central, o que promoverá uma grande economia no tempo gasto com os ajustes dos modelos.

2.7.2 Análise residual

Uma alternativa para o procedimento de Timmerman-Kiers é avaliar a soma de quadrados residual junto com os graus de liberdade. É preciso construir um novo gráfico com a soma de quadrados residual d de cada modelo contra os graus de liberdade df .

Nos gráficos de análise residual, a avaliação do modelo pode ser facilitada desenhando linhas indiretas entre os espaços dos modelos com uma constante $k = d/df$. Se a origem está incluída no gráfico, estas linhas podem ser vistas como um critério para retirar a origem com um declive k . Se $k = 1$, então cada parâmetro adicionado ou cada grau de liberdade perdido conduz ao ganho de uma unidade na soma de quadrados residual. Se dois modelos estão ligados por uma linha que tem um declive íngreme, que consiste em ter um k grande, então pela troca de alguns graus de liberdade pode-se adquirir uma redução grande na soma de quadrados residual, ou em outras palavras, um modelo ajustando-se consideravelmente melhor. Mas caso o declive seja muito baixo, com k pequeno, muitos graus de liberdade são sacrificados para que a redução nos resíduos seja pequena ou ter um pequeno aumento no ajuste. É possível usar a parcimônia preferindo modelos com bom ajuste e poucos parâmetros.

Esta análise residual foi projetada para ajudar na escolha dos modelos de Tucker que diferem em dimensionalidade, mas é possível incluir modelos de outras classes. Pode ser uma vantagem incluir os modelos de Tucker2 e modelos apropriados de PARAFAC nesta análise residual. Em modelos de PARAFAC não há uma maneira correta para determinar o número de graus de liberdade, mas tendo decidido sobre o número de graus de liberdade, tais modelos podem ser incluídos no gráfico da análise residual (KROONENBERG, 2008).

2.7.3 Critério st de Ceulemans-Kiers

CEULEMANS e KIERS (2006) propuseram um método de seleção de modelos em gráficos de análise residual. O critério de seleção é conhecido por critério st definido como:

$$st_i = \frac{\frac{d_{i-1} - d_i}{df_{i-1} - df_i}}{\frac{d_i - d_{i+1}}{df_i - df_{i+1}}}, \quad \text{com } i = 1, \dots, n \quad (2.23)$$

em que d_i é o resíduo, df_i é o número de graus de liberdade do modelo i e n é o número máximo de modelos que se pode ajustar. Vale ressaltar que tal critério se baseia na soma de quadrados residual e graus de liberdade e não na soma de quadrados do ajuste e número de parâmetros livres.

O procedimento para seleção do modelo consiste nos passos seguintes:

1. Determinar os valores de df e d de todas as soluções das quais deseja escolher;
2. Observar df para cada valor de $i = 1, \dots, n$, retendo somente a melhor solução em termos do ajuste;
3. Ordenar as n soluções pelos valores de df e denotá-los por s_i ($i = 1, \dots, n$);
4. Excluir todas as soluções s_i para qual uma solução s_j ($j < i$) existe tal que $d_j < d_i$;
5. Considerar consecutivamente todos os grupos de soluções iguais adjacentes. Excluir a solução mediana, se seu valor estiver acima ou na linha que conecta aos seus vizinhos no gráfico da análise residual;
6. Repetir o passo 5 até que nenhuma solução possa ser excluída;
7. Determinar os valores de st das soluções obtidas, de acordo com a equação (2.23);
8. Selecionar a solução com o menor valor de st .

Resumindo, os passos de 1 a 6 determinam quais modelos devem fazer parte da avaliação do st e os passos 7 e 8 encontram o melhor ângulo $\phi_{i-1,i+1}$ entre as linhas que conectam o i -ésimo modelo com seu anterior ($i - 1$), com o modelo subsequente ($i + 1$).

Os autores advertem que os passos do critério não devem ser seguidos rigorosamente e critérios que levam em consideração um bom conhecimento do assunto também podem ser usados para seleção, de forma que um modelo pouco inferior ao indicado pelo método às vezes pode ser escolhido por possuir boas qualidades de interpretação.

2.8 Dimensionalidade de um modelo PARAFAC

A seleção da dimensionalidade de um modelo PARAFAC é bem mais simples que a seleção da dimensionalidade de um modelo de Tucker, pois todos os modos tem o mesmo número de componentes. Em estudos práticos, como na psicologia, o número máximo de componentes nos modelos PARAFAC é de dois ou três, mas em alguns modelos da física com aplicações em química podem ser ajustados com grandes quantidades de componentes. MARCO *et al.* (2005) mostraram que são necessários seis componentes no modelo PARAFAC para o estudo de substâncias químicas em flores de hibisco. MURPHY *et al.* (2006) indicaram um modelo PARAFAC com nove componentes no estudo de conteúdo químico na troca de água em um lastro de navios.

2.8.1 Procedimento de dividir ao meio (*Split-half*)

Uma forma de determinar a dimensionalidade de um modelo PARAFAC é avaliar a estabilidade da solução, dividindo o conjunto de dados ao meio e executar uma análise separada em ambas as partes. Havendo uma solução ótima, esta deve aparecer nas duas análises.

É preciso advertir que deve haver objetos suficientes no modo que estará sendo dividido, ou seja, ambas as partes devem ser grandes o suficiente para minimizar a influência individual de dados específicos. Definir o quanto “grande o suficiente” é complicado, pois vai depender muito do nível de ruído e da qualidade da estrutura dos dados. HARSHMAN e DE SARBO (1984) discutiram e demonstraram o procedimento de dividir ao meio com detalhes e exemplos ilustrativos. Os autores também sugeriram fazer duas (ou mais) divisões ortogonais. Para análises de dados dividido ao meio, o requisito é que os dados sejam divididos aleatoriamente em quatro partes mais ou menos iguais, por exemplo, A, B, C e D, que são combinados em quatro novos conjuntos de dados: (A + B e C + D) e (A + C e B + D). KIERS e MECHELEN (2001) apresentam uma aproximação semelhante para o modelo de Tucker3.

2.8.2 Consistência do núcleo

BRO (1998) propôs uma aproximação para selecionar o número de componentes do modelo PARAFAC, conhecido por princípio de consistência do núcleo, para avaliar quantos componentes no modelo PARAFAC são necessários para descrever os dados, e em BRO e KIERS (2003) este princípio é nomeado de CONCORDIA ou diagnóstico de consistência do núcleo. O método consiste em avaliar quão distante está a ordem do núcleo derivada do ajuste de um modelo PARAFAC de um modelo ideal, ou seja, de um arranjo núcleo superdiagonal $\mathbf{G}$.

Um arranjo núcleo calculado a partir dos componentes de um modelo PARAFAC é uma superdiagonal se para um modelo de três entradas somente g_{111} , g_{222} , $g_{333} \dots$ tem valores consideráveis e todos os outros elementos de arranjo são próximos a zero. Na aproximação de BRO (1998), o arranjo núcleo é padronizado de tal forma que os elementos da superdiagonal sejam iguais a 1 (arranjo superidentidade).

Em um modelo de três entradas PARAFAC, a discrepância do arranjo ideal é dado por:

$$CONCORDIA = 1 - \frac{\sum_{p=1}^S \sum_{q=1}^S \sum_{r=1}^S (g_{pqr} - i_{pqr})^2}{\sum_{p=1}^S \sum_{q=1}^S \sum_{r=1}^S (i_{pqr})^2}. \quad (2.24)$$

em que i_{pqr} é um elemento de um arranjo superidentidade, de modo que $i_{pqr} = 1$ se $p = q = r$ e $i_{pqr} = 0$, caso contrário.

O diagnóstico de consistência de núcleo terá um valor muito baixo (negativo) quando os componentes dentro de uma entrada forem altamente correlacionados. Caso não se deseje isso, os i_{pqr} podem ser substituídos por g_{pqr} no denominador, e a medida é chamada consistência do núcleo normalizada. Os autores questionam se isto faria diferença em relação a decisão sobre o número de componentes retidos no modelo, mas com relação a consistência do núcleo, para valores próximos de 1, significa que o modelo é informativo. Pode-se usar também o grau de superdiagonalidade para o mesmo propósito, uma vez que essa medida é igual à soma de quadrados dos elementos da superdiagonal dividida pela soma total de quadrados dos elementos de núcleo. Todas as três medidas serão iguais a 1 no caso de um arranjo superdiagonal.

2.9 Estabilidade do modelo e poder preditivo por validação

Na construção de modelo é preciso saber responder se o modelo encontrado será adequado em novas amostras. Uma ideia a seguir é dividir o conjunto em duas partes e ajustar um modelo em uma das partes e depois comparar com os dados que não foram envolvidos na estimação. Este procedimento é conhecido por validação. Se as diferenças forem pequenas, pode-se dizer que as estimativas de parâmetro tem bom poder de predição. Vale ressaltar, que para tal aproximação nada é feito para as insuficiências da amostra original, entretanto tem-se uma boa informação de quão bom o modelo se mostrará em outras amostras da mesma população (KROONENBERG, 2008).

Para avaliar o poder preditivo de um modelo pode-se também usar o procedimento de reamostragem *Jackknife*. RIU e BRO (2003) retiraram uma fatia completa de cada vez, e seu objetivo era usar jackknife para estimar os erros padrão e procurar pontos discrepantes, e não avaliar o poder preditivo do modelo. Os autores LOUWERSE *et al.* (1999) reformularam esta proposta baseado em EASTMENT e KRZANOWSKI (1982), no qual também removeram fatias inteiras e desenvolveram uma maneira sofisticada para combinar os resultados da retirada da fatia, de forma que o poder preditivo possa ser avaliado. Outra proposta é remover cada observação, ou vários dados ao mesmo tempo, sendo que estas observações são consideradas como perdidas. Os autores usaram um algoritmo EM para estimar os dados indicados como perdido, e compararam os valores estimados com os valores observados.

O poder preditivo de um modelo geralmente é avaliado pela soma de quadrados dos erros preditivos (*PRESS*), sendo calculada comparando os valores de todos os dados originais com os seus valores estimados, com base nos modelos sem os dados em questão. Assim, para o caso de três entradas, x_{ijk} é comparado com $\hat{x}_{ijk}^{PQR}$ e o poder preditivo, $PRESS_{PQR}$, é dado por:

$$PRESS_{PQR} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{x}_{ijk}^{PQR} - x_{ijk})^2. \quad (2.25)$$

Além disso, LOUWERSE *et al.* (1999) desenvolveram a estatística $W_{PQR, (P-1)QR}$, para medir a diferença relativa do $PRESS_{PQR}$ entre dois modelos que diferem em um único componente (apresentada para a primeira entrada). A estatística W é definida como:

$$W_{(P-1)QR, PQR} = \frac{\frac{PRESS_{(P-1)QR} - PRESS_{PQR}}{df_{(P-1)QR} - df_{PQR}}}{\frac{PRESS_{PQR}}{df_{PQR}}} \quad (2.26)$$

em que df são os graus de liberdade do modelo. Para evitar valores negativos, que podem ocorrer com a redução dos graus de liberdade sem a diminuição suficiente no *PRESS*, pode-se introduzir um limite inferior de zero.

LOUWERSE *et al.* (1999) também sugeriram uma forma de minimizar o número de modelos que tem que ser avaliados, ou seja, encontrar modelos com baixos valores de *PRESS*. A estratégia usada é que dado um modelo com um número específico de componentes, o próximo conjunto de modelos a ser avaliado serão aqueles com um componente a mais em cada entrada. O melhor destes modelos é tomado como uma base para o próximo passo. Este processo será interrompido se o *PRESS* aumentar ou não diminuir suficientemente (por exemplo, menor que 10^{-6}).

2.10 Biplot

Biplot é uma representação gráfica em que as linhas e colunas de uma matriz são apresentadas em um gráfico com duas ou três-dimensões, sendo que a construção do *biplot* se baseia na decomposição em valores singulares da matriz de dados (GABRIEL, 1971). A importância deste gráfico é que possibilita a inspeção visual da posição de uma unidade amostral relativa a outra, bem como a importância relativa de cada uma das variáveis à posição de qualquer unidade e também a correlação entre as variáveis. Ainda é possível visualizar como as unidades amostrais se agrupam e quais variáveis contribuem para sua posição dentro dessa representação.

2.10.1 Decomposição em valores singulares

Suponha uma matriz $\mathbf{A}$ com informação de I objetos em J variáveis. A decomposição em valor singular (DVS) da matriz $\mathbf{A}$ é definida como:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}' \quad (2.27)$$

que pode ser escrita em notação de somatórios da seguinte forma

$$a_{ij} = \sum_{s=1}^P \lambda_s \mathbf{u}_{is} \mathbf{v}_{js} \quad (2.28)$$

em que $P = \min(I, J)$, ou seja, geralmente são necessários P termos para reproduzir exatamente a matriz original $\mathbf{A}$. Os escalares λ_s são os valores singulares organizados em ordem decrescente, $\mathbf{u}_s$ é um vetor de objetos e $\mathbf{v}_s$ é um vetor de variáveis. Em ambos os casos, os vetores são ortonormais. $\mathbf{U}$ e $\mathbf{V}$ são matrizes que contêm os vetores $\mathbf{u}_s$ e $\mathbf{v}_s$ em suas colunas, respectivamente.

Para encontrar uma aproximação de menor dimensão para $\mathbf{A}$, deve-se minimizar a distância entre a matriz de aproximação $\hat{\mathbf{A}}$. Esta distância entre os elementos das duas matrizes, $\mathbf{A} = (a_{ij})$ e $\hat{\mathbf{A}} = (\hat{a}_{ij})$, é definido como:

$$d(\mathbf{A}, \hat{\mathbf{A}}) = \sqrt{\sum_{i=1}^I \sum_{j=1}^J (a_{ij} - \hat{a}_{ij})^2}, \quad (2.29)$$

e ECKART e YOUNG (1936) mostraram que a melhor aproximação Q -dimensional de mínimos quadrados da matriz $\mathbf{A}$ é obtida pela decomposição em valor singular de $\mathbf{A}$, somando somente os Q ($Q < P$) primeiros termos da equação (2.28).

Os primeiros Q vetores $\mathbf{u}_s$ e $\mathbf{v}_s$, com Q usualmente dois ou três, são usados como coordenadas para o gráfico de representação dos dados. Estes podem ser combinados em diferentes formas com os valores singulares λ_s , como mostrado a seguir:

$$\hat{x}_{ij} = \sum_{s=1}^Q u_{is}(\lambda_s v_{js}) = \sum_{s=1}^Q y_{is} z_{js}, \quad (2.30)$$

$$\hat{x}_{ij} = \sum_{s=1}^Q (u_{is} \lambda_s^{1/2}) (\lambda_s^{1/2} v_{js}) = \sum_{s=1}^Q y_{is}^* z_{js}^*, \quad (2.31)$$

em que y e z são as coordenadas dos objetos e das variáveis, sendo que somente as coordenadas principais das variáveis são escalonadas e y^* e z^* são as coordenadas dos objetos e das variáveis, em que coordenadas principais são simetricamente escalonadas.

2.10.2 Biplot padrão

Um *biplot* padrão é representado por uma tabela entre objetos e variáveis $\mathbf{A}$, decomposta no produto $\mathbf{Y}\mathbf{Z}'$, em que $\mathbf{Y}$ é de dimensão $I \times Q$ e $\mathbf{Z}$ é de dimensão $J \times Q$. Com uma decomposição de duas-dimensões, cada elemento $\hat{x}_{ij}$ de $\hat{\mathbf{X}}$ pode ser escrito como:

$$\hat{x}_{ij} = y_{i1} z_{j1} + y_{i2} z_{j2} \quad (2.32)$$

que é o produto interno dos vetores linhas (y_{i1}, y_{i2}) e (z_{j1}, z_{j2}) . Um *biplot* é obtido representando cada linha como um ponto Y_i com coordenadas (y_{i1}, y_{i2}) e cada coluna como um ponto Z_j com coordenadas (z_{j1}, z_{j2}) em um gráfico de duas dimensões. Tais pontos são geralmente chamados de marcadores de linhas e marcadores de colunas, respectivamente. Se escrever Y_i'' para a projeção ortogonal de Y_i no segmento OZ_j , θ_{ij} para o ângulo entre os segmentos OY_i e OZ_j , e ainda escrever $\|OZ_j\|^2$ para o comprimento do vetor OZ (representação geométrica na Figura 2.9), tem-se que:

$$\hat{x}_{ij} = \|OZ_j\| \|OY_i\| \cos(\theta_{ij}) = \|OZ_j\| \|OY_i''\|. \quad (2.33)$$

Figura 2.9. Representação de dois marcadores de objetos e um marcador de variáveis em um *biplot*

A equação (2.33) mostra que $\hat{x}_{ij}$ é proporcional ao comprimento de OY_i'' . Assim, a relação ou a interação entre dois objetos com a mesma variável pode ser avaliada comparando os comprimentos das

projeções dentro daquela variável. Ademais, a relação ou interação entre um vetor objeto $0Y_i$ e a variável $0Z_j$ é positivo se possuem um ângulo agudo e negativo se possuem um ângulo obtuso. Quando a projeção do marcador Y_i dentro do vetor de variável $0Z_j$ coincide com a origem, $\hat{x}_{ij}$ é igual a zero e o objeto tem aproximadamente o valor médio, para esta variável, se os dados x_{ij} foram centrados. Um valor positivo para $\hat{x}_{ij}$ indica que o i -ésimo objeto tem um alto escore na variável j , com relação ao escore médio desta, e valores negativos indicam que o i -ésimo objeto tem um escore relativamente baixo na variável i .

2.11 Joint plot

Um *joint plot* na análise *multiway* é semelhante a um *biplot* padrão e todos os princípios de interpretação do *biplot* padrão podem ser utilizados. Foi usado pela primeira vez na ciência social e psicologia para componentes de Tucker e ele tem algumas propriedades especiais (KROONENBERG, 1983). A diferença nesta construção é que o *joint plot* é construído como um *biplot* para dois fatores dada a matriz de componentes do modelo Tucker3 referente ao terceiro fator (modo) ou fator de referência (modo de referência). Cada *joint plot* é construído usando diferentes fatias do arranjo núcleo. As fatias são feitas para cada componente do modo de referência. Cada fatia contém o poder de ligação ou os pesos para os componentes dos modos apresentados no gráfico. Os coeficientes no componente associado ao modo de referência pondera inteiramente o *joint plot* por seus valores, de forma que os *joint plot* são pequenos para os pequenos valores no componente e grande para aqueles com grandes coeficientes.

O ponto inicial para a construção de um *joint plot* após ajustar um modelo de Tucker3, é obter uma matriz $\Delta_r = \mathbf{A}\mathbf{G}_r\mathbf{B}' = \mathbf{A}_r^*\mathbf{B}_r^{*'} de dimensão $I \times J$, com $r = 1, 2, \dots, R$ ou uma matriz $\Delta_k = \mathbf{A}\mathbf{H}_k\mathbf{B}' = \mathbf{A}_k^*\mathbf{B}_k^{*'} de dimensão $I \times J$, com $k = 1, 2, \dots, K$ após ajustar um modelo de Tucker2. Para cada fatia do núcleo, $\mathbf{G}_r$ (ou $\mathbf{H}_k$), é necessário construir um *joint plot* para a matriz de componentes $\mathbf{A}^* (J \times P)$ e $\mathbf{B}^* (J \times Q)$.$$

Segundo KROONENBERG (1994), a construção de um *joint plot* é feita como segue. A fatia do arranjo núcleo $\mathbf{G}_r (P \times Q)$ é decomposta via decomposição em valor singular como

$$\mathbf{G}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r'$$

e os vetores singulares $\mathbf{U}_r$ e $\mathbf{V}_r'$ são combinados com as matrizes $\mathbf{A}$ e $\mathbf{B}$, respectivamente, e a matriz diagonal $\mathbf{\Lambda}_r$ com os valores singulares é dividido entre as duas matrizes de forma que:

$$\mathbf{A}_r^* = \left(\frac{I}{J}\right)^{1/4} \mathbf{A} \mathbf{U}_r \mathbf{\Lambda}_r^{1/2} \quad (2.34)$$

$$\mathbf{B}_r^* = \left(\frac{J}{I}\right)^{1/4} \mathbf{B} \mathbf{V}_r \mathbf{\Lambda}_r^{1/2} \quad (2.35)$$

As colunas das matrizes de componentes ajustadas referem-se aos eixos do *joint plot*. Caso a matriz $\mathbf{G}_r$ (ou $\mathbf{H}_k$) não seja quadrada, o seu posto é geralmente, $M = \min(P, Q)$, e somente M *joint plot* podem ser apresentados. O procedimento completo é a rotação de cada matriz de componentes para uma matriz ortonormal, seguido por um alongamento (ou encolhimento) dos componentes rotacionados. O tamanho do alongamento (ou encolhimento) dos eixos é regulado pela raiz quadrada de $\lambda_{mm}^{(r)}$ e pela raiz quarta de (J/I) . Vale ressaltar que se existe uma grande diferença na variabilidade explicada pelos eixos, ou seja, entre $(\lambda_{mm}^r)^2$ e $(\lambda_{m'm'}^r)^2$, pode ocorrer uma dispersão visual considerável no gráfico, pois os coeficientes dos componentes são multiplicados por $(\lambda_{mm}^r)^{1/2}$.

Como $\mathbf{A}_r^* \mathbf{B}_r^{*'} = \Delta_r$, cada elemento δ_{ij}^r é igual ao produto interno de $\mathbf{a}_i^* \mathbf{b}_j^{*'}$, e isto proporciona um alongamento na ligação entre a i -ésima linha da matriz de componentes $\mathbf{A}$ e a j -ésima linha da matriz de componentes $\mathbf{B}$, controlado pela r -ésima fatia do arranjo núcleo. Exibindo simultaneamente os dois modos em um gráfico, podem ser obtidas conclusões visuais sobre as relações entre eles. O espaçamento

e a ordem das projeções dos objetos em uma variável correspondem ao tamanho do produto interno entre eles e assim, a importância relativa daquela variável para os objetos.

O *joint plot* oferece como vantagem que a interpretação das relações de variáveis e objetos podem ser feitas diretamente, não é necessário envolver os eixos dos componentes ou seus rótulos. Outra vantagem é que por meio da fatia do arranjo central $\mathbf{G}_r$ (ou $\mathbf{H}_k$), os eixos das coordenadas *joint plot* são escalonados de acordo com a importância relativa, de forma que visualmente uma impressão correta da dispersão dos componentes é criada. Entretanto, ao escalonar simetricamente os componentes (como descrito anteriormente), as distâncias entre os objetos não são aproximações da distância Euclidiana, nem os ângulos entre as variáveis representam correlações. O *joint plot* para o modelo de Tucker3 é útil na investigação do significado dos objetos com respeito às variáveis explicitamente, dado um componente do terceiro modo. Para o modelo de Tucker2, o *joint plot* provê a informação sobre as relações entre objetos e variáveis dadas a um nível do terceiro modo (KROONENBERG, 2008).

Em relação a interpretação de um *joint plot* (VARELA *et al.*, 2006), suponha um gráfico que é projetado sobre o r -ésimo componente principal da terceira entrada tal que, no *joint plot* aparecem todos os níveis das duas primeiras entradas. Em seguida, a partir de matrizes $\mathbf{C}$ (matriz das componentes principais da terceira entrada), selecione os níveis deste fator com maior peso no r -ésimo componente (positivos ou negativos), tais valores que determinam os níveis da terceira entrada. Suponha que a matriz $\mathbf{C}$ tem um valor positivo e elevado associado ao k -ésimo nível da terceira entrada, então proximidades entre os níveis da primeira e segunda entrada (por exemplo, o i -ésimo nível do primeiro fator e o j -ésimo nível do segundo fator) indicam que a interação tripla entre o i -ésimo nível da primeira entrada, j -ésimo nível da segunda entrada e k -ésimo nível da terceira entrada é positiva. Em compensação, se o i -ésimo nível do primeiro fator estiver muito distante do j -ésimo nível do segundo fator, indica que a interação tripla associada com i -ésimo nível da primeira entrada, j -ésimo nível da segunda entrada e k -ésimo nível da terceira entrada é negativa.

Suponha que a matriz $\mathbf{C}$ tem um valor negativo associado ao k -ésimo nível do terceiro fator, então proximidades entre os níveis do primeiro e segundo fator (por exemplo, i -ésimo nível do primeiro fator e j -ésimo nível do segundo fator) no *joint plot* indicam que a interação tripla entre o i -ésimo nível da primeira entrada, j -ésimo nível da segunda entrada e k -ésimo nível da terceira entrada é negativa. Em compensação, se o i -ésimo nível do primeiro fator estiver muito distante do j -ésimo nível do segundo fator, indica que a interação tripla associada ao i -ésimo nível da primeira entrada, j -ésimo nível da segunda entrada e k -ésimo nível da terceira entrada é positiva.

Em geral, os níveis de uma entrada localizado no centro do *joint plot* são considerados um conjunto que tem um desempenho médio em todos os outros modos.

2.12 Distribuição qui-quadrado não-central

Em probabilidade e estatística, se $U_1, U_2, \dots, U_v$ são variáveis aleatórias independentes e normalmente distribuídas, e $\delta_1, \delta_2, \dots, \delta_v$ são constantes, então a distribuição de

$$\sum_{j=1}^v (U_j + \delta_j)^2 \quad (2.36)$$

depende de $\delta_1, \delta_2, \dots, \delta_v$ somente através das suas somas de quadrados. Tal distribuição é conhecida como qui-quadrado (χ^2) não-central com v graus de liberdade e parâmetro de não-centralidade $\theta = \sum_{j=1}^v \delta_j^2$, representada por $\chi_v^2(\theta)$.

A função densidade probabilidade é expressa como (JOHNSON *et al.*, 1995):

$$f(u; v, \theta) = \frac{\exp\{-1/2(\theta + u)\}}{2^{v/2}} \sum_{j=0}^{\infty} \left(\frac{\theta}{4}\right)^j \frac{u^{(v/2)+j-1}}{j! \Gamma(1/2v + j)},$$

em que $\Gamma[.]$ refere-se à função gama.

Uma estimativa do parâmetro de não-centralidade é apresentada por HOGG e CRAIG (1978) como:

$$\hat{\theta} = \frac{\sum_{j=1}^v U_j^2}{\sigma^2}$$

3 MATERIAL E MÉTODOS

3.1 Características dos dados

Os dados a serem utilizados são relativos a experimentos com 13 genótipos de feijão que foram conduzidos em 9 experimentos distintos constituídos pelos anos agrícolas de 2000/2001, 2001/2002 e 2005/2006, nos municípios de Dourados e Aquidauana no estado de Mato Grosso do Sul, sendo que os experimentos foram instalados na época das águas (Dourados) e também na época das secas (Dourados e Aquidauana). Cada local é constituído de município e uma época de instalação, conforme descrito na Tabela 3.1. Cada experimento foi instalado em delineamento em blocos ao acaso, com 3 blocos em cada experimento.

Tabela 3.1. Caracterização dos ambientes experimentais

Município	Época	Local ¹	Ano agrícola
Dourados	“das águas”	L1	2000/2001 (A1)
Dourados	“das secas”	L2	2000/2001 (A1)
Aquidauana	“das secas”	L3	2000/2001 (A1)
Dourados	“das águas”	L1	2001/2002 (A2)
Dourados	“das secas”	L2	2001/2002 (A2)
Aquidauana	“das secas”	L3	2001/2002 (A2)
Dourados	“das águas”	L1	2005/2006 (A3)
Dourados	“das secas”	L2	2005/2006 (A3)
Aquidauana	“das secas”	L3	2005/2006 (A3)

¹O fator local consiste na combinação de municípios com épocas

Para cada um dos genótipos, em cada um dos ambientes, foram avaliadas as seguintes variáveis respostas: número médio de vagens por planta (VAG): medido na colheita durante o processo de arranquio; número médio de sementes por vagem (SEM): obtido na colheita durante o processo de trilha; massa de 100 sementes (MCS): medida após a colheita e expresso em gramas; produtividade de grãos (PROD): medida após a colheita e expressa em ton/ha.

Neste trabalho, será considerado apenas a variável produtividade de grãos.

3.2 Gráfico *tripplot*

Sejam as matrizes **A**, **B** e **C** com duas colunas cada. O produto de elementos por elementos dessas matrizes é um arranjo de três entradas **Z**, desde que **A**, **B** e **C** tenha posto igual a 2. Tais matrizes podem ser representadas por um gráfico de duas dimensões, semelhante ao *biplot* proposto por GABRIEL (1971). Este gráfico de duas dimensões é chamado de *tripplot*, pois representa o efeito de três fatores. Com a representação das matrizes **A**, **B** e **C** no gráfico *tripplot*, é possível visualizar e comparar todos os elementos do arranjo **Z** por meio de linhas, colunas e tubos.

Em pesquisa experimental de três entradas é possível visualizar todos os dados em um *tripplot*. A representação de um conjunto de dados de três entradas genótipos $\times$ locais $\times$ anos em um *tripplot*, pressupõe encontrar suas três matrizes de componentes **A**, **B** e **C**. Para obter o arranjo de três entradas com três matrizes de componentes é possível por meio da decomposição PARAFAC (item 2.5.1), também chamada de decomposição por valores singulares generalizada por outros autores como SMILDE *et al.* (2004), que é essencial para o processo inverso de multiplicação de matrizes.

Sejam g genótipos testados em l locais e a anos dispostos em um arranjo de três entradas **Z** de dimensão $g \times l \times a$. A decomposição PARAFAC de **Z** gera P componentes principais, sendo P o posto do arranjo **Z**. Cada componente principal é composto por: uma combinação da importância dos genótipos

(a_{ir}) , uma combinação da importância dos locais (b_{jr}) e uma combinação da importância dos anos (c_{kr}) . Assim, cada elemento de $\underline{\mathbf{Z}}$ é recuperado por

$$z_{ijk} = \sum_{p=1}^P a_{ip} b_{jp} c_{kp}, \quad (3.1)$$

em que

a_{ip} : é o elemento que está na i -ésima linha e p -ésima coluna da matriz de componentes $\mathbf{A}$ do modelo PARAFAC;

b_{jp} : é o elemento que está na j -ésima linha e p -ésima coluna da matriz de componentes $\mathbf{B}$ do modelo PARAFAC;

c_{kp} : é o elemento que está na k -ésima linha e p -ésima coluna da matriz de componentes $\mathbf{C}$ do modelo PARAFAC.

Segundo ARAUJO (2009), para representar os dados de genótipos $\times$ locais $\times$ anos em um *tripplot* q -dimensional, deve-se selecionar matrizes $\mathbf{A}$ de dimensão $(g \times q)$, $\mathbf{B}$ de dimensão $(l \times q)$ e $\mathbf{C}$ de dimensão $(a \times q)$. Ao submeter um conjunto de dados de três entradas a um modelo PARAFAC, as matrizes resultantes $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$, na teoria, podem ser representadas por meio de um *tripplot* q -dimensional, mas apenas os *tripplots* bidimensionais tem fácil visualização e interpretação. Um *tripplot* bidimensional pode ser construído usando os dois primeiros componentes principais, e tal *tripplot* é suficiente quando os primeiros dois componentes principais têm uma aproximação suficiente para a matriz $\underline{\mathbf{Z}}$.

3.3 Análise de dados admitindo três entradas

3.3.1 Análise de variância conjunta

Com o objetivo de verificar se existe a interação entre genótipos, locais e anos, realiza-se uma análise de variância conjunta que envolve o estudo de todos os genótipos em todos os locais e todos os anos, sendo que em cada local utiliza-se um delineamento como o aleatorizado em blocos. Assume-se que o efeito de genótipos é fixo, o efeito de locais e o efeito de anos também como fixos, obtendo os efeitos das interações duplas (genótipos $\times$ locais, genótipos $\times$ anos e locais $\times$ anos) e tripla (genótipos $\times$ locais $\times$ anos) como fixas. Essa análise será representada pelo seguinte modelo matemático:

$$Y_{ijk r} = \mu + b_r(l_j a_k) + g_i + l_j + a_k + (gl)_{ij} + (ga)_{ik} + (la)_{jk} + (gla)_{ijk} + \varepsilon_{ijk r} \quad (3.2)$$

em que:

$Y_{ijk r}$: é o valor observado do i -ésimo genótipo no j -ésimo local no k -ésimo ano e no r -ésimo bloco, com $i = 1, 2, \dots, G$, $j = 1, 2, \dots, L$, $k = 1, \dots, A$ e $r = 1, 2, \dots, R$;

μ : é uma constante comum às respostas;

$b_r(l_j a_k)$: é o efeito do r -ésimo bloco dentro do j -ésimo local e do k -ésimo ano;

g_i : é o efeito do i -ésimo genótipo;

l_j : é o efeito do j -ésimo local;

a_k : é o efeito do k -ésimo ano;

$(gl)_{ij}$: é o efeito da interação do i -ésimo genótipo com o j -ésimo local;

$(ga)_{ik}$: é o efeito da interação do i -ésimo genótipo com o k -ésimo ano;

$(la)_{jk}$: é o efeito da interação do j -ésimo local com o k -ésimo ano;

$(gla)_{ijk}$: é o efeito da interação do i -ésimo genótipo com o j -ésimo local com o k -ésimo ano;

$\varepsilon_{ijk r}$: é erro experimental associado ao i -ésimo genótipo, no j -ésimo local, no k -ésimo ano e no r -ésimo bloco assumido ser independente e $\varepsilon_{ijk r} \sim N(0, \sigma^2)$.

Na Tabela 3.2 apresenta-se o esquema da análise de variância para o modelo (3.2), com graus de liberdade e esperanças dos quadrados médios, de acordo com KUTNER *et al.* (2005).

Tabela 3.2. Esquema da ANOVA para experimentos de um mesmo grupo de g genótipos avaliados em l locais e a anos

Fonte de Variação	GL	E[QM]
Bd.LA	$LA(R-1)$	—
G	$G-1$	$\sigma^2 + lar\phi_G$
L	$L-1$	$\sigma^2 + gar\phi_L$
A	$A-1$	$\sigma^2 + glr\phi_A$
G×L	$(G-1)(L-1)$	$\sigma^2 + ar\phi_{GL}$
G×A	$(G-1)(A-1)$	$\sigma^2 + lr\phi_{GA}$
L×A	$(L-1)(A-1)$	$\sigma^2 + gr\phi_{LA}$
G×L×A	$(G-1)(L-1)(A-1)$	$\sigma^2 + r\phi_{GLA}$
Resíduo	$LA(G-1)(R-1)$	σ^2
Total	$GLAR-1$	

$\text{Bd.LA: Bloco dentro de local e de ano; } \phi_G = \frac{\sum_{i=1}^G g_i^2}{G-1}; \phi_L = \frac{\sum_{j=1}^L l_j^2}{L-1}; \phi_A = \frac{\sum_{k=1}^A a_k^2}{A-1}$
 $\phi_{GL} = \frac{\sum_{i=1}^G \sum_{j=1}^L (gl)_{ij}^2}{(G-1)(L-1)}; \phi_{GA} = \frac{\sum_{i=1}^G \sum_{k=1}^A (ga)_{ik}^2}{(G-1)(A-1)}; \phi_{LA} = \frac{\sum_{j=1}^L \sum_{k=1}^A (la)_{jk}^2}{(L-1)(A-1)}$
 $\phi_{GLA} = \frac{\sum_{i=1}^G \sum_{j=1}^L \sum_{k=1}^A (gla)_{ijk}^2}{(G-1)(L-1)(A-1)}.$

3.4 Estudo de simulação

Inicialmente, estudou-se por simulação, utilizando o software R CORE TEAM (2017), o comportamento das somas de quadrados das fatias horizontais, verticais e frontais da matriz de interação genótipos × locais × anos (G×L×A), visando a aplicação de um teste para os efeitos de genótipos, locais e anos.

Supõe-se que $\mathbf{Y} = [y_{ij}]$ é a tabela de dupla entrada para um conjunto de dados escolhido e que se ajuste a análise de variância (ANOVA) dada pelo modelo 3.2.

Considere uma barra sobre a letra para denotar média. Dessa forma, os estimadores dos parâmetros do modelo (3.2) são:

$$\begin{aligned}
 \hat{\mu} &= \bar{Y}_{....} & \hat{g}_i &= \bar{Y}_{i...} - \bar{Y}_{....} & \hat{l}_j &= \bar{Y}_{.j..} - \bar{Y}_{....} \\
 \hat{a}_k &= \bar{Y}_{...k} - \bar{Y}_{....} & (\hat{gl})_{ij} &= \bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....} \\
 (\hat{ga})_{ik} &= \bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....} & (\hat{la})_{jk} &= \bar{Y}_{.jk.} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y}_{....} \\
 (\hat{gla})_{ijk} &= \bar{Y}_{ijk.} - \bar{Y}_{ij..} - \bar{Y}_{i.k.} - \bar{Y}_{.jk.} + \bar{Y}_{i...} + \bar{Y}_{.j..} + \bar{Y}_{..k.} - \bar{Y}_{....}
 \end{aligned}$$

Assim, com base nos dados originais é possível estimar $\hat{\mu}$, $\hat{g}_i$, $\hat{l}_j$, $\hat{a}_k$, $(\hat{gl})_{ij}$, $(\hat{ga})_{ik}$, $(\hat{la})_{jk}$ e $(\hat{gla})_{ijk}$ gerando números aleatórios de uma distribuição normal com média 0 e variância σ^2 para o termo $\varepsilon_{ijk r}$ (σ^2 é também estimado pelo QM_{Res}). Pode-se também gerar repetições que imitem o padrão dos dados originais, mas os quais têm embutido uma estrutura de variabilidade dos mesmos. Geraram-se

500, 1000 e 5000 experimentos para a variável resposta produtividade de grãos, dos quais obteve-se as somas de quadrados dos elementos das fatias horizontais, verticais e frontais do arranjo de três entradas de interação $G \times L \times A$.

Utilizou-se a função “*fitdistr*” para estimar o parâmetro de não-centralidade da distribuição qui-quadrado não-central, a função “*ks.test*” para verificar a aderência das somas de quadrados a essa distribuição por meio do teste de Kolmogorov-Smirnov e por último, construiu-se o gráfico QQplot para visualização do ajuste das somas de quadrados das fatias horizontais, verticais e frontais da matriz de interação $G \times L \times A$ à distribuição qui-quadrado não-central.

3.5 Proposta de teste estatístico para a contribuição de genótipos, locais e anos para a interação tripla

Seja a matriz de interação $\mathbf{GLA}_{(g \times l \times a)} = (\widehat{gla}_{ijk})$ dada da seguinte forma:

$$\mathbf{GLA}_{(g \times l \times a)} = \begin{pmatrix} \widehat{gla}_{111} & \widehat{gla}_{121} & \cdots & \widehat{gla}_{1l1} & \widehat{gla}_{112} & \widehat{gla}_{122} & \cdots & \widehat{gla}_{1l2} & \cdots & \widehat{gla}_{11a} & \widehat{gla}_{12a} & \cdots & \widehat{gla}_{1la} \\ \widehat{gla}_{211} & \widehat{gla}_{221} & \cdots & \widehat{gla}_{2l1} & \widehat{gla}_{212} & \widehat{gla}_{222} & \cdots & \widehat{gla}_{2l2} & \cdots & \widehat{gla}_{21a} & \widehat{gla}_{22a} & \cdots & \widehat{gla}_{2la} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \widehat{gla}_{g11} & \widehat{gla}_{g21} & \cdots & \widehat{gla}_{gl1} & \widehat{gla}_{g12} & \widehat{gla}_{g22} & \cdots & \widehat{gla}_{gl2} & \cdots & \widehat{gla}_{g1a} & \widehat{gla}_{g2a} & \cdots & \widehat{gla}_{gla} \end{pmatrix}$$

A hipótese testada para verificar a contribuição de cada genótipo para a interação $G \times L \times A$ é como segue:

- i) H_{01} : o i -ésimo genótipo não contribui para a interação $G \times L \times A$, para todo $i = 1, \dots, G$
 H_{a1} : o i -ésimo genótipo contribui para a interação $G \times L \times A$, para algum $i = 1, \dots, G$

Seja $G_i(G \times L \times A)$ a representação do efeito da contribuição do i -ésimo genótipo para a interação. A soma de quadrados devido ao efeito desse genótipo é obtida por:

$$SQ_{G_i(G \times L \times A)} = \sum_{j=1}^L \sum_{k=1}^A (\widehat{gla}_{ijk})^2 \quad (i = 1, \dots, G),$$

com $\frac{(G-1)(L-1)(A-1)}{G}$ graus de liberdade associados e o quadrado médio correspondente é obtido dividindo-se cada soma de quadrados pelo respectivo número de graus de liberdade, ou seja,

$$QM_{G_i(G \times L \times A)} = \frac{SQ_{G_i(G \times L \times A)}}{\frac{(G-1)(L-1)(A-1)}{G}} = \frac{G(SQ_{G_i(G \times L \times A)})}{(G-1)(L-1)(A-1)}, \quad i = 1, \dots, G.$$

Observa-se que a $SQ_{G_1(G \times L \times A)} + \cdots + SQ_{G_i(G \times L \times A)} = SQ_{G \times L \times A}$, com $(G-1)(L-1)(A-1)$ graus de liberdade.

Tem-se que a soma de quadrados devido ao efeito de genótipos dividido pela variância segue distribuição qui-quadrado não-central com parâmetro de não-centralidade θ_1 , ou seja:

$$\frac{SQ_{G_i(G \times L \times A)}}{\sigma^2} \sim \chi^2_{\frac{(G-1)(L-1)(A-1)}{G}}(\theta_{1i})$$

$$\text{em que } \theta_{1i} = \frac{\sum_{j=1}^L \sum_{k=1}^A (\widehat{gla}_{ijk})^2}{\sigma^2}.$$

Assim, verifica-se que:

$$F^*_{(G_i(G \times L \times A))} = \frac{QM_{G_i(G \times L \times A)}}{QM_{Res}} \sim F_{([(G-1)(L-1)(A-1)] / G, LA(G-1)(R-1))} \text{ para todo genótipo } i \text{ sob } H_{01}.$$

Seja a hipótese $H_0 : \theta_{1i} = 0$ versus $H_a : \theta_{1i} > 0$, que é equivalente a hipótese H_{01} versus H_{a1} , tem-se que a estatística $F^*_{(G_i(G \times L \times A))}$ sob H_0 tem distribuição F central com $[(G-1)(L-1)(A-1)] / G$ e $LA(G-1)(R-1)$ graus de liberdade (MOSER, 1996).

Ao nível α de significância, rejeita-se H_0 ou H_{01} quando,

$$F_{(G_i(G \times L \times A))}^* = \frac{QM_{G_i(G \times L \times A)}}{QM_{Res}} \geq F_{(\alpha, [(G-1)(L-1)(A-1)]/G, LA(G-1)(R-1))} \text{ para } i = 1, \dots, G.$$

Por meio do teste F , aplicado a ANOVA, levando em consideração a decomposição dos $(G-1)(L-1)(A-1)$ graus de liberdade da interação $G \times L \times A$ com $[(G-1)(L-1)(A-1)]/G$ graus de liberdade para os genótipos, resulta no esquema mostrado na Tabela 3.3.

Tabela 3.3. Esquema da ANOVA com teste F para obtenção de genótipos que contribuem significativamente para a interação $G \times L \times A$

Fonte de Variação	GL	QM	F
Bd.LA	$LA(R-1)$	$QM_{Bd.LA}$	
G	$G-1$	QM_G	
L	$L-1$	QM_L	
A	$A-1$	QM_A	
$G \times L$	$(G-1)(L-1)$	QM_{GL}	
$G \times A$	$(G-1)(A-1)$	QM_{GA}	
$L \times A$	$(L-1)(A-1)$	QM_{LA}	
$G \times L \times A$	$(G-1)(L-1)(A-1)$	QM_{GLA}	
$G_1(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{G}$	$QM_{G_1(G \times L \times A)}$	$\frac{QM_{G_1(G \times L \times A)}}{QM_{Res}}$
...	...	...	...
$G_g(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{G}$	$QM_{G_g(G \times L \times A)}$	$\frac{QM_{G_g(G \times L \times A)}}{QM_{Res}}$
Resíduo	$LA(G-1)(R-1)$	QM_{Res}	
Total	$GLAR-1$		

Bd.LA: Bloco dentro de local e de ano; QM: Quadrado médio;

$G_i(G \times L \times A)$: é o efeito da contribuição do i -ésimo genótipo dentro da interação, com $i = 1, \dots, G$.

A hipótese testada para verificar a contribuição de cada local para a interação $G \times L \times A$ é como segue:

- ii) H_{02} : o j -ésimo local não contribui para a interação $G \times L \times A$, para todo $j = 1, \dots, L$
 H_{a2} : o j -ésimo local contribui para a interação $G \times L \times A$, para algum $j = 1, \dots, L$

Seja $L_j(G \times L \times A)$ a representação do efeito da contribuição do j -ésimo local para a interação. A soma de quadrados devido ao efeito desse local é obtida por:

$$SQ_{L_j(G \times L \times A)} = \sum_{i=1}^G \sum_{k=1}^A (\widehat{gla}_{ijk})^2 \quad (j = 1, \dots, L),$$

com $\frac{(G-1)(L-1)(A-1)}{L}$ graus de liberdade associados e o quadrado médio correspondente é obtido dividindo-se cada soma de quadrados pelo respectivo número de graus de liberdade, ou seja,

$$QM_{L_j(G \times L \times A)} = \frac{SQ_{L_j(G \times L \times A)}}{\frac{(G-1)(L-1)(A-1)}{L}} = \frac{L(SQ_{L_j(G \times L \times A)})}{(G-1)(L-1)(A-1)}, \quad j = 1, \dots, L.$$

Observa-se que a $SQ_{L_1(G \times L \times A)} + \dots + SQ_{L_j(G \times L \times A)} = SQ_{G \times L \times A}$, com $(G-1)(L-1)(A-1)$ graus de liberdade.

Tem-se que a soma de quadrados devido ao efeito de locais dividido pela variância segue distribuição qui-quadrado não-central com parâmetro de não-centralidade θ_2 , ou seja:

$$\frac{SQ_{L_j(G \times L \times A)}}{\sigma^2} \sim \chi_{\frac{(G-1)(L-1)(A-1)}{L}}^2(\theta_{2j})$$

em que $\theta_{2j} = \frac{\sum_{i=1}^G \sum_{k=1}^A (\widehat{gla}_{ijk})^2}{\sigma^2}$.

Assim, verifica-se que:

$$F_{(L_j(G \times L \times A))}^* = \frac{QM_{L_j(G \times L \times A)}}{QM_{Res}} \sim F_{([(G-1)(L-1)(A-1)]/L, LA(G-1)(R-1))} \text{ para todo local } j \text{ sob } H_{02}.$$

Seja a hipótese $H_0 : \theta_{2j} = 0$ versus $H_a : \theta_{2j} > 0$, que é equivalente a hipótese H_{02} versus H_{a2} , tem-se que a estatística $F_{(L_j(G \times L \times A))}^*$ sob H_0 tem distribuição F central com $[(G-1)(L-1)(A-1)]/L$ e $LA(G-1)(R-1)$ graus de liberdade (MOSER, 1996).

Ao nível α de significância, rejeita-se H_0 ou H_{02} quando,

$$F_{(L_j(G \times L \times A))}^* = \frac{QM_{L_j(G \times L \times A)}}{QM_{Res}} \geq F_{(\alpha, [(G-1)(L-1)(A-1)]/L, LA(G-1)(R-1))} \text{ para } j = 1, \dots, L.$$

Por meio do teste F , aplicado a ANOVA, levando em consideração a decomposição dos $(G-1)(L-1)(A-1)$ graus de liberdade da interação $G \times L \times A$ com $[(G-1)(L-1)(A-1)]/L$ graus de liberdade para os locais, resulta no esquema mostrado na Tabela 3.4.

Tabela 3.4. Esquema da ANOVA com teste F para obtenção de locais que contribuem significativamente para a interação $G \times L \times A$

Fonte de Variação	GL	QM	F
Bd.LA	$LA(R-1)$	$QM_{Bd.LA}$	
G	$G-1$	QM_G	
L	$L-1$	QM_L	
A	$A-1$	QM_A	
$G \times L$	$(G-1)(L-1)$	QM_{GL}	
$G \times A$	$(G-1)(A-1)$	QM_{GA}	
$L \times A$	$(L-1)(A-1)$	QM_{LA}	
$G \times L \times A$	$(G-1)(L-1)(A-1)$	QM_{GLA}	
$L_1(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{L}$	$QM_{L_1(G \times L \times A)}$	$\frac{QM_{L_1(G \times L \times A)}}{QM_{Res}}$
$\dots$	$\dots$	$\dots$	$\dots$
$L_l(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{L}$	$QM_{L_l(G \times L \times A)}$	$\frac{QM_{L_l(G \times L \times A)}}{QM_{Res}}$
Resíduo	$LA(G-1)(R-1)$	QM_{Res}	
Total	$GLAR-1$		

$L_j(G \times L \times A)$: é o efeito da contribuição do j -ésimo local dentro da interação, com $j = 1, \dots, L$.

A hipótese testada para verificar a contribuição de cada ano para a interação $G \times L \times A$ é como segue:

- iii) H_{03} : o k -ésimo ano não contribui para a interação $G \times L \times A$, para todo $k = 1, \dots, A$
 H_{a3} : o k -ésimo ano contribui para a interação $G \times L \times A$, para algum $k = 1, \dots, A$

Seja $A_k(G \times L \times A)$ a representação do efeito da contribuição do k -ésimo ano para a interação. A soma de quadrados devido ao efeito desse ano é obtida por:

$$SQ_{A_k(G \times L \times A)} = \sum_{i=1}^G \sum_{j=1}^L (\widehat{gla}_{ijk})^2 \quad (k = 1, \dots, A),$$

com $\frac{(G-1)(L-1)(A-1)}{A}$ graus de liberdade associados e o quadrado médio correspondente é obtido dividindo-se cada soma de quadrados pelo respectivo número de graus de liberdade, ou seja,

$$QM_{A_k(G \times L \times A)} = \frac{SQ_{A_k(G \times L \times A)}}{\frac{(G-1)(L-1)(A-1)}{A}} = \frac{A(SQ_{A_k(G \times L \times A)})}{(G-1)(L-1)(A-1)}, \quad k = 1, \dots, A.$$

Observa-se que a $SQ_{A_1(G \times L \times A)} + \dots + SQ_{A_k(G \times L \times A)} = SQ_{G \times L \times A}$, com $(G-1)(L-1)(A-1)$ graus de liberdade.

Tem-se que a soma de quadrados devido ao efeito de anos dividido pela variância segue distribuição qui-quadrado não-central com parâmetro de não-centralidade θ_3 , ou seja:

$$\frac{SQ_{A_k(G \times L \times A)}}{\sigma^2} \sim \chi^2_{\frac{(G-1)(L-1)(A-1)}{A}}(\theta_{3k})$$

$$\text{em que } \theta_{3k} = \frac{\sum_{i=1}^G \sum_{j=1}^L (\widehat{gla}_{ijk})^2}{\sigma^2}.$$

Assim, verifica-se que:

$$F_{(A_k(G \times L \times A))}^* = \frac{QM_{A_k(G \times L \times A)}}{QM_{Res}} \sim F_{([(G-1)(L-1)(A-1)]/A, LA(G-1)(R-1))} \text{ para todo local } k \text{ sob } H_{03}.$$

Seja a hipótese $H_0 : \theta_{3k} = 0$ versus $H_a : \theta_{3k} > 0$, que é equivalente a hipótese H_{03} versus H_{a3} , tem-se que a estatística $F_{(A_k(G \times L \times A))}^*$ sob H_0 tem distribuição F central com $[(G-1)(L-1)(A-1)]/A$ e $LA(G-1)(R-1)$ graus de liberdade (MOSER, 1996).

Ao nível α de significância, rejeita-se H_0 ou H_{03} quando,

$$F_{(A_k(G \times L \times A))}^* = \frac{QM_{A_k(G \times L \times A)}}{QM_{Res}} \geq F_{(\alpha, [(G-1)(L-1)(A-1)]/A, LA(G-1)(R-1))} \text{ para } k = 1, \dots, A.$$

Por meio do teste F , aplicado a ANOVA, levando em consideração a decomposição dos $(G-1)(L-1)(A-1)$ graus de liberdade da interação $G \times L \times A$ com $[(G-1)(L-1)(A-1)]/A$ graus de liberdade para os anos, resulta no esquema mostrado na Tabela 3.5.

Tabela 3.5. Esquema da ANOVA com teste F para obtenção de anos que contribuem significativamente para a interação $G \times L \times A$

Fonte de Variação	GL	QM	F
Bd.LA	$LA(R-1)$	$QM_{Bd.LA}$	
G	$G-1$	QM_G	
L	$L-1$	QM_L	
A	$A-1$	QM_A	
$G \times L$	$(G-1)(L-1)$	QM_{GL}	
$G \times A$	$(G-1)(A-1)$	QM_{GA}	
$L \times A$	$(L-1)(A-1)$	QM_{LA}	
$G \times L \times A$	$(G-1)(L-1)(A-1)$	QM_{GLA}	
$A_1(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{A}$	$QM_{A_1(G \times L \times A)}$	$\frac{QM_{A_1(G \times L \times A)}}{QM_{Res}}$
...	...	...	...
$A_a(G \times L \times A)$	$\frac{(G-1)(L-1)(A-1)}{A}$	$QM_{A_a(G \times L \times A)}$	$\frac{QM_{A_a(G \times L \times A)}}{QM_{Res}}$
Resíduo	$LA(G-1)(R-1)$	QM_{Res}	
Total	$GLAR-1$		

$A_k(G \times L \times A)$: é o efeito da contribuição do k -ésimo ano dentro da interação, com $k = 1, \dots, A$.

A proposta de graus de liberdade igualitários para genótipos, locais e anos, vai de encontro à proposta de GAUCH JR (1992) para obtenção dos graus de liberdade dos componentes multiplicativos de um modelo AMMI. Este autor diz que esta definição de graus de liberdade é apropriada para dados experimentais que têm os efeitos de interação e erros experimentais. Os dados experimentais são ruidosos (ou seja, a resposta é igual a padrão mais ruído). A definição tem os mesmos princípios da definição de MANDEL (1971), que enfatiza na quantidade de ruído recuperado em cada componente do modelo AMMI, mas esta definição tem uma aplicabilidade mais ampla. Em MANDEL (1971), os graus de liberdade são definidos para ser tal que o quadrado médio do ruído resultante seja uma estimativa do erro experimental

aleatório para efeitos aditivos e multiplicativos do modelo AMMI, quer na presença ou ausência de efeitos aditivos e de interação.

Sob as pressuposições usuais da ANOVA, os erros são distribuídos independente dos tratamentos, assim, todas as parcelas da decomposição da interação tripla tem o mesmo ruído esperado. Como consequência, os graus de liberdade $[(G - 1)(L - 1)(A - 1)]$ da interação tripla podem ser decomposto em $[(G - 1)(L - 1)(A - 1)]/G$ graus de liberdade para avaliar a contribuição de genótipos, $[(G - 1)(L - 1)(A - 1)]/L$ graus de liberdade para avaliar a contribuição de locais e $[(G - 1)(L - 1)(A - 1)]/A$ graus de liberdade para avaliar a contribuição de anos.

4 RESULTADOS

4.1 Gráfico *triplot*

A interpretação de um gráfico *triplot* quanto à interação tripla como por exemplo, genótipos $\times$ locais $\times$ anos é feita observando a magnitude e o sinal dos escores de genótipos, locais e anos dos eixos. Desta forma, escores baixos, próximos de zero, são característicos dos genótipos, locais e anos que contribuíram pouco ou quase nada para a interação, caracterizando-se como estáveis (DUARTE e VENCovsky, 1999). Portanto, no gráfico *triplot* serão considerados como estáveis os genótipos, locais e anos que estiverem próximos a origem, ou seja, escores próximos de zero.

Observa-se na Figura 4.1 que os genótipos 2, 3, 4, 6, 7, 8 e 10, são os que estão mais próximos da origem, portanto possíveis genótipos estáveis, sendo necessário aplicar um teste estatístico para verificar quais contribuem ou não para a interação tripla. Em relação aos genótipos 5, 9 e 12 são os que estão mais distantes da origem, sendo estes os possíveis genótipos que contribuem para a interação genótipos $\times$ locais $\times$ anos. Ainda, com relação aos genótipos, é possível construir alguns grupos de genótipos que apresentam características semelhantes: Grupo 1 - genótipos 5 e 9; Grupo 2 - genótipos 2, 4 e 10; Grupo 3 - genótipos 3, 6, 7 e 8; Grupo 4 - genótipos 1, 11 e 13; Grupo 5 - genótipo 12.

Levando em consideração os locais, nota-se que possivelmente o local 1 (município de Dourados na época das águas) é um local estável, pois está próximo a origem e os locais 2 (município de Dourados na época das secas) e 3 (município de Aquidauana na época das secas) são os que mais contribuem para a interação $G \times L \times A$, sendo que os locais não apresentam uma formação de grupos, o que indica que cada local tem características próprias.

Em relação aos anos, percebe-se que os anos 1 (2000/2001) e 2 (2001/2002) possivelmente contribuíram para a interação $G \times L \times A$ e o ano 3 (2005/2006) foi um ano estável por apresentar valores próximos a origem. Não houve também formação de grupos dentro do fator ano, pois cada ano também tem características próprias.

Figura 4.1. Gráfico *triplot* para a produtividade de feijão (ton/ha) representando a interação e mostrando as relações entre genótipos, locais e anos

4.1.1 Estudo de simulação

As Figuras de 4.2 a 4.16 apresentam os histogramas dos 500, 1000 e 5000 experimentos simulados para as somas de quadrados (SQ) das fatias horizontais (genótipos) do arranjo de três entradas de interação $G \times L \times A$. As linhas contínuas representam o ajuste de uma distribuição qui-quadrado não-central teórica e as barras verticais a distribuição empírica dos dados de somas de quadrados simulados para genótipos. Observa-se pelas Figuras 4.2 a 4.16 que as distribuições das somas de quadrados dos genótipos estão bem próximas de uma distribuição qui-quadrado não-central com 3,69 graus de liberdade e ainda, a medida que o número de experimentos simulados aumenta, essa aproximação se torna melhor.

Figura 4.2. Histogramas para as SQ dos genótipos 1, 2 e 3 com 500 experimentos simulados

Figura 4.3. Histogramas para as SQ dos genótipos 4, 5 e 6 com 500 experimentos simulados

Figura 4.4. Histogramas para as SQ dos genótipos 7, 8 e 9 com 500 experimentos simulados

Figura 4.5. Histogramas para as SQ dos genótipos 10, 11 e 12 com 500 experimentos simulados

Figura 4.6. Histograma para a SQ do genótipo 13 com 500 experimentos simulados

Figura 4.7. Histogramas para as SQ dos genótipos 1, 2 e 3 com 1000 experimentos simulados

Figura 4.8. Histogramas para as SQ dos genótipos 4, 5 e 6 com 1000 experimentos simulados

Figura 4.9. Histogramas para as SQ dos genótipos 7, 8 e 9 com 1000 experimentos simulados

Figura 4.10. Histogramas para as SQ dos genótipos 10, 11 e 12 com 1000 experimentos simulados

Figura 4.11. Histograma para a SQ do genótipo 13 com 1000 experimentos simulados

Figura 4.12. Histogramas para as SQ dos genótipos 1, 2 e 3 com 5000 experimentos simulados

Figura 4.13. Histogramas para as SQ dos genótipos 4, 5 e 6 com 5000 experimentos simulados

Figura 4.14. Histogramas para as SQ dos genótipos 7, 8 e 9 com 5000 experimentos simulados

Figura 4.15. Histogramas para as SQ dos genótipos 10, 11 e 12 com 5000 experimentos simulados

Figura 4.16. Histograma para a SQ do genótipo 13 com 5000 experimentos simulados

As Figuras de 4.17 a 4.19 apresentam os histogramas dos 500, 1000 e 5000 experimentos simulados para as somas de quadrados (SQ) das fatias verticais (locais) do arranjo de três entradas de interação $G \times L \times A$. As linhas contínuas representam o ajuste de uma distribuição qui-quadrado não-central teórica e as barras verticais a distribuição empírica dos dados de somas de quadrados simulados para locais. Observa-se pelas Figuras 4.17 a 4.19 que as distribuições das somas de quadrados dos locais estão bem próximas de uma distribuição qui-quadrado não-central com 16 graus de liberdade e ainda, a medida que o número de experimentos simulados aumenta, essa aproximação se torna mais evidente.

Figura 4.17. Histogramas para as SQ dos locais 1, 2 e 3 com 500 experimentos simulados

Figura 4.18. Histogramas para as SQ dos locais 1, 2 e 3 com 1000 experimentos simulados

Figura 4.19. Histogramas para as SQ dos locais 1, 2 e 3 com 5000 experimentos simulados

As Figuras de 4.20 a 4.22 apresentam os histogramas dos 500, 1000 e 5000 experimentos simulados para as somas de quadrados (SQ) das fatias frontais (anos) do arranjo de três entradas de interação $G \times L \times A$. As linhas contínuas representam o ajuste de uma distribuição qui-quadrado não-central teórica e as barras verticais a distribuição empírica dos dados de somas de quadrados simulados para anos. Observa-se pelas Figuras 4.20 a 4.22 que as distribuições das somas de quadrados dos anos estão bem próximas de uma distribuição qui-quadrado não-central com 16 graus de liberdade.

Figura 4.20. Histogramas para as SQ dos anos 1, 2 e 3 com 500 experimentos simulados

Figura 4.21. Histogramas para as SQ dos anos 1, 2 e 3 com 1000 experimentos simulados

Figura 4.22. Histogramas para as SQ dos anos 1, 2 e 3 com 5000 experimentos simulados

Nas Tabelas 4.1, 4.2 e 4.3 são apresentadas as estimativas de máxima verossimilhança dos parâmetros de não-centralidade da distribuição de qui-quadrado não-central. Nestas mesmas tabelas são apresentados os valores do nível descritivo p do teste de Kolmogorov-Smirnov para verificar se as somas de quadrados das fatias horizontais, verticais e frontais do arranjo de três entradas de interação $G \times L \times A$ seguem distribuição qui-quadrado não-central com 3,69, 16 e 16 graus de liberdade, respectivamente.

Observa-se na Tabela 4.1 que os valores estimados dos parâmetros de não-centralidade, para cada genótipo dentro da interação, não diferem a medida que se aumenta o número de simulações, mostrando assim que um número de 500 simulações foi suficiente para se estimar o parâmetro de não-centralidade da distribuição qui-quadrado não-central. No entanto, o $\hat{\theta}$ varia de um genótipo para outro, indicando assim, que cada genótipo deve ser testado com uma particular distribuição F não-central, no sentido de verificar sua contribuição para a interação. Ainda nesta tabela, observa-se pelo valor- p que quase todas somas de quadrados devido aos genótipos dentro da interação genótipos $\times$ locais $\times$ anos seguem uma distribuição qui-quadrado não-central. Este fato não foi observado, ao nível de 5% de significância, somente para o genótipo 7 e com 5000 simulações, estando perfeitamente dentro da margem de erro do teste.

Tabela 4.1. Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos genótipos para 500, 1000 e 5000 experimentos simulados

	500		1000		5000	
	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p
G_1	26,01	0,7179	25,76	0,8884	25,08	0,3323
G_2	20,12	0,8847	20,31	0,3297	19,82	0,5669
G_3	15,21	0,7850	15,82	0,3587	16,01	0,3226
G_4	24,11	0,9215	24,01	0,4087	24,17	0,2784
G_5	79,30	0,2401	80,16	0,5473	79,97	0,2670
G_6	15,70	0,7726	16,28	0,7141	16,24	0,0856
G_7	32,10	0,8859	32,59	0,2024	32,89	0,0264
G_8	26,41	0,9417	27,19	0,4210	26,64	0,1475
G_9	52,04	0,9253	52,74	0,6146	52,89	0,1700
G_{10}	8,27	0,2385	8,45	0,4388	8,25	0,1478
G_{11}	46,60	0,6940	46,16	0,9771	46,52	0,1030
G_{12}	84,03	0,4484	83,66	0,6911	83,30	0,2840
G_{13}	32,21	0,5964	32,88	0,5936	32,75	0,3648

Da Tabela 4.2, têm-se as mesmas conclusões para as somas de quadrados devido a locais dentro da interação genótipos $\times$ locais $\times$ anos, ou seja, que os valores dos parâmetros de não-centralidade, para cada local dentro da interação, não difere a medida que se aumenta o número de simulações, mostrando assim que um número de 500 simulações foi um bom número para se estimar o parâmetro de não-centralidade da distribuição qui-quadrado não-central e que todas as somas de quadrados devido aos locais dentro da interação genótipos $\times$ locais $\times$ anos seguem uma distribuição qui-quadrado não-central.

Tabela 4.2. Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos locais para 500, 1000 e 5000 experimentos simulados

	500		1000		5000	
	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p
L_1	178,87	0,9882	182,86	0,8756	181,84	0,6334
L_2	324,13	0,9268	325,95	0,4101	324,72	0,7937
L_3	246,76	0,8133	247,37	0,1627	247,01	0,9293

Pela Tabela 4.3, as conclusões também são as mesmas para as somas de quadrados devido a anos dentro da interação genótipos $\times$ locais $\times$ anos, ou seja, que os valores dos parâmetros de não-centralidade, para cada ano dentro da interação, também não difere a medida que se aumenta o número de simulações,

mostrando assim que um número de 500 simulações foi suficiente para se estimar o parâmetro de não-centralidade da distribuição qui-quadrado não-central e que todas as somas de quadrados devido aos anos dentro da interação genótipos $\times$ locais $\times$ anos seguem uma distribuição qui-quadrado não-central.

Tabela 4.3. Estimativas dos parâmetros de não-centralidade ($\hat{\theta}$) e valor- p do teste de Kolmogorov-Smirnov para aderência à distribuição qui-quadrado não-central da soma de quadrados dos anos para 500, 1000 e 5000 experimentos simulados

	500		1000		5000	
	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p	$\hat{\theta}$	valor- p
A_1	325,96	0,9996	327,90	0,9696	327,16	0,9611
A_2	177,35	0,8922	179,26	0,4595	178,60	0,6347
A_3	246,51	0,8835	248,97	0,9591	247,86	0,3172

As mesmas observações feitas anteriormente com relação a distribuição de probabilidades das somas de quadrados devido a genótipos, locais e anos dentro da interação genótipos $\times$ locais $\times$ anos, também podem ser verificadas através dos gráficos QQplot nas Figuras de 4.23 a 4.43. Observa-se que em praticamente todos os gráficos os pontos estão dispostos próximo de linha reta, o que leva a concluir que estas somas de quadrados apresentam um bom ajuste à distribuição qui-quadrado não-central com parâmetro de não-centralidade θ , até mesmo a soma de quadrados referente ao genótipo 7 que havia sido rejeitado pelo teste de Kolmogorov-Smirnov com 5000 simulações. Alguns gráficos apresentam valores fora da linha nas extremidades, o que não interfere para conclusão do teste, uma vez que a maior importância se deve a valores menores que estão dentro da linha reta.

Figura 4.23. Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 500 experimentos simulados

Figura 4.24. Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 500 experimentos simulados

Figura 4.25. Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 500 experimentos simulados

Figura 4.26. Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 500 experimentos simulados

Figura 4.27. Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 500 experimentos simulados

Figura 4.28. Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 1000 experimentos simulados

Figura 4.29. Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 1000 experimentos simulados

Figura 4.30. Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 1000 experimentos simulados

Figura 4.31. Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 1000 experimentos simulados

Figura 4.32. Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 1000 experimentos simulados

Figura 4.33. Gráficos QQplot qui-quadrado para as SQ dos genótipos 1, 2 e 3 com 5000 experimentos simulados

Figura 4.34. Gráficos QQplot qui-quadrado para as SQ dos genótipos 4, 5 e 6 com 5000 experimentos simulados

Figura 4.35. Gráficos QQplot qui-quadrado para as SQ dos genótipos 7, 8 e 9 com 5000 experimentos simulados

Figura 4.36. Gráficos QQplot qui-quadrado para as SQ dos genótipos 10, 11 e 12 com 5000 experimentos simulados

Figura 4.37. Gráfico QQplot qui-quadrado para a SQ do genótipo 13 com 5000 experimentos simulados

Figura 4.38. Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 500 experimentos simulados

Figura 4.39. Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 1000 experimentos simulados

Figura 4.40. Gráficos QQplot qui-quadrado para as SQ dos locais 1, 2 e 3 com 5000 experimentos simulados

Figura 4.41. Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 500 experimentos simulados

Figura 4.42. Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 1000 experimentos simulados

Figura 4.43. Gráficos QQplot qui-quadrado para as SQ dos anos 1, 2 e 3 com 5000 experimentos simulados

4.2 Aplicação em dados reais

4.2.1 Análise de variância conjunta admitindo três entradas

A Tabela 4.4 corresponde a ANOVA efetuada com os dados observados de 13 genótipos e 9 ambientes (combinação de 3 locais e 3 anos). Verifica-se que o efeito de blocos dentro de locais e de anos, o efeito de genótipos, o efeito de locais e o efeito de anos são não significativos ao nível de 5% de significância, enquanto que os efeitos das interações duplas (genótipos $\times$ locais, genótipos $\times$ anos e locais $\times$ anos) e interação tripla (genótipos $\times$ locais $\times$ anos) são significativos ao nível de 5% de significância e suas somas de quadrados (SQ) correspondem a 10,11%, 9,15%, 15,14% e 8,72%, respectivamente, da soma de quadrados total.

Tabela 4.4. ANOVA para um conjunto de dados com 13 genótipos avaliados em 3 locais, 3 anos com 3 blocos

Fonte de Variação	GL	SQ	QM	F	valor- p
B d. LA	18	0,70	0,04	1,00	0,4608
G	12	31,60	2,63	2,02	0,0622
L	2	65,21	32,60	4,11	0,1071
A	2	13,02	6,51	0,82	0,5030
G $\times$ L	24	21,19	0,88	2,32	0,0065
G $\times$ A	24	19,17	0,80	2,10	0,0143
L $\times$ A	4	31,71	7,91	203,28	<0,0001
G $\times$ L $\times$ A	48	18,27	0,38	9,74	<0,0001
Resíduo	216	8,51	0,04		
Total	350	209,40			

Através do teste F , aplicado a ANOVA, levando em consideração a decomposição dos $(G - 1)(L - 1)(A - 1) = (13 - 1)(3 - 1)(3 - 1) = 48$ graus de liberdade da interação G $\times$ L $\times$ A com $[(G - 1)(L - 1)(A - 1)/G] = [(13 - 1)(3 - 1)(3 - 1)/13] \approx 3,69$ graus de liberdade para os genótipos, resulta na Tabela 4.5. Observa-se que somente para o genótipo 10 não se rejeita H_0 ao nível de 5% de significância, ou seja, este genótipo não contribui significativamente para a interação genótipos $\times$ locais $\times$ anos. Os demais genótipos são significativos e contribuem para a interação. Desta forma, pode-se afirmar que tais genótipos apresentam respostas heterogêneas quanto à locais e anos.

Tabela 4.5. Teste F para obtenção de genótipos que contribuem significativamente para a interação G $\times$ L $\times$ A

Fonte de Variação	GL	SQ	QM	F	valor- p
$G_1(G \times L \times A)$	3,69	0,9992	0,2706	6,8677	0,0001
$G_2(G \times L \times A)$	3,69	0,7791	0,2110	5,3547	0,0006
$G_3(G \times L \times A)$	3,69	0,6276	0,1700	4,3138	0,0030
$G_4(G \times L \times A)$	3,69	0,9492	0,2571	6,5239	0,0001
$G_5(G \times L \times A)$	3,69	3,1467	0,8522	21,6281	<0,0001
$G_6(G \times L \times A)$	3,69	0,6350	0,1720	4,3649	0,0027
$G_7(G \times L \times A)$	3,69	1,2898	0,3493	8,8653	<0,0001
$G_8(G \times L \times A)$	3,69	1,0453	0,2831	7,1847	<0,0001
$G_9(G \times L \times A)$	3,69	2,0858	0,5649	14,3365	<0,0001
$G_{10}(G \times L \times A)$	3,69	0,3299	0,0893	2,2673	0,0682
$G_{11}(G \times L \times A)$	3,69	1,8140	0,4913	12,4682	<0,0001
$G_{12}(G \times L \times A)$	3,69	3,2853	0,8898	22,5808	<0,0001
$G_{13}(G \times L \times A)$	3,69	1,2819	0,3472	8,8109	<0,0001

Agora, levando em consideração a decomposição dos mesmos 48 graus de liberdade da interação G $\times$ L $\times$ A com $[(G - 1)(L - 1)(A - 1)/L] = [(13 - 1)(3 - 1)(3 - 1)/13] = 16$ graus de liberdade para os locais

com aplicação do teste F na ANOVA, resulta na Tabela 4.6. Nota-se que todos os locais contribuem significativamente para a interação tripla $G \times L \times A$ ao nível de 5% de significância, o que leva a perceber também respostas heterogêneas aos locais para essa variável.

Tabela 4.6. Teste F para obtenção de locais que contribuem significativamente para a interação $G \times L \times A$

Fonte de Variação	GL	SQ	QM	F	valor- p
$L_1(G \times L \times A)$	16	4,334	0,2709	6,8743	<0,0001
$L_2(G \times L \times A)$	16	7,9403	0,4963	12,5945	<0,0001
$L_3(G \times L \times A)$	16	5,9945	0,3747	9,5082	<0,0001

Por último, levando em consideração a decomposição dos mesmos 48 graus de liberdade da interação $G \times L \times A$ com $[(G - 1)(L - 1)(A - 1)/A] = [(13 - 1)(3 - 1)(3 - 1)/13] = 16$ graus de liberdade para os anos com aplicação do teste F na ANOVA, resulta na Tabela 4.7. Observa-se que todos os anos contribuem significativamente para a interação ao nível de 5% de significância.

Tabela 4.7. Teste F para obtenção de anos que contribuem significativamente para a interação $G \times L \times A$

Fonte de Variação	GL	SQ	QM	F	valor- p
$A_1(G \times L \times A)$	16	7,9984	0,4999	12,6866	<0,0001
$A_2(G \times L \times A)$	16	4,2625	0,2664	6,7609	<0,0001
$A_3(G \times L \times A)$	16	6,0079	0,3755	9,5294	<0,0001

5 DISCUSSÃO

Nesse trabalho foi proposto um método alternativo ao *tripplot* para testar a contribuição de cada genótipo, local e ano, considerados de efeitos fixos, para a interação genótipos $\times$ locais $\times$ anos por meio de um teste F . Para identificar genótipos, locais e anos estáveis, utilizou-se o gráfico *tripplot* que é de fácil entendimento e traz ao pesquisador informações reais sobre a interação tripla. Entretanto, essa informação não tem significância em relação a quais genótipos, locais e anos contribuem mais para a interação. Verificou-se no estudo de simulação das somas de quadrados das fatias horizontais, verticais e frontais do arranjo de três entradas genótipos $\times$ locais $\times$ anos que estas seguem distribuição qui-quadrado não central. Na aplicação em dados reais do teste proposto, foi possível detectar os genótipos, locais e anos que contribuem mais para a interação tripla genótipos $\times$ locais $\times$ anos, sendo de fácil aplicação e interpretação.

Além disso, há uma diferença entre os resultados do teste F e os obtidos pelo gráfico *tripplot* (Figura 4.1), por exemplo, genótipos que eram candidatos a serem estáveis pelo gráfico apresentaram uma contribuição significativa para a interação. Estas diferenças se devem ao fato de que no *tripplot* está representado 73,37% da soma de quadrados da interação tripla (que pode mascarar alguns resultados importantes) e o teste proposto é baseado em 100%. Assim, de uma forma geral, recomenda-se o uso do gráfico *tripplot* associado ao teste F .

5.1 Aspectos computacionais

O gráfico *tripplot* foi construído segundo a rotina apresentada por ARAUJO (2009). Utilizou-se o software R CORE TEAM (2017) para a simulação dos dados, para estimar o parâmetro de não-centralidade da distribuição qui-quadrado não-central, verificar a aderência à distribuição e implementação da rotina computacional para a realização da análise de dados, mediante a contribuição de cada genótipo, local e ano para a interação. Todos os programas feitos no R estão nos apêndices.

5.2 PESQUISAS FUTURAS

Aplicação do teste F para verificar a contribuição dos níveis dos fatores para mudança na natureza dos níveis, aleatórios e/ou fixos.

REFERÊNCIAS

- ALLARD, R. e A. BRADSHAW, 1964 Implications of genotype-environmental interactions in applied plant breeding. *Crop science* **4**: 503–508.
- ANNICCHIARICO, P., 2002 *Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations*. Number 174, Food & Agriculture Org.
- ARAUJO, L., 2009 *Seleção e análise dos modelos PARAFAC e Tucker e gráfico triplot com aplicação em interação tripla*. Ph.D. thesis, Escola Superior de Agricultura Luiz de Queiroz.
- ARAUJO, M., 2008 *Teste estatístico para contribuição de genótipos e ambientes na matriz de interação GE*. Ph.D. thesis, Universidade de São Paulo.
- ARAUJO, M., L. ARAUJO, P. FARIA, e C. DIAS, 2012 Statistical test for genotype and environment contribution in the genotypes x environments interaction matrix. *Interciencia* **37**: 515–519.
- BARROSO, L. e R. ARTES, 2003 *Análise multivariada*. Lavras: SEAGRO.
- BRO, R., 1998 *Multi-way analysis in the food industry: models, algorithms, and applications*. Ph.D. thesis, Københavns UniversitetKøbenhavns Universitet, LUKKET: 2012 Det Biovidenskabelige Fakultet for Fødevarer, Veterinærmedicin og NaturressourcerFaculty of Life Sciences, LUKKET: 2012 Institut for FødevarevidenskabDepartment of Food Science, LUKKET: 2012 Kvalitet og TeknologiQuality & Technology.
- BRO, R. e H. A. KIERS, 2003 A new efficient method for determining the number of components in parafac models. *Journal of chemometrics* **17**: 273–286.
- CARROLL, J. e J.-J. CHANG, 1970 Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika* **35**: 283–319.
- CATTELL, R., 1944 parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika* **9**: 267–283.
- CATTELL, R., 1952 The three basic factor-analytic research designstheir interrelations and derivatives. *Psychological bulletin* **49**: 499.
- CEULEMANS, E. e H. KIERS, 2006 Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British journal of mathematical and statistical psychology* **59**: 133–150.
- CHAVES, L., 2001 Interação de cultivares com ambientes. Recursos genéticos e melhoramento - plantas pp. 673–713.
- CROSSA, J., 1990 Statistical analyses of multilocation trials. *Advances in agronomy* **44**: 55–85.
- DAS, L., 2006 *Genetics and Plant Breeding*. New Age International.
- DUARTE, J. e R. VENCovsky, 1999 *Interação genótipo × ambiente: Uma Introdução à análise AMMI*. Sociedade Brasileira de Genética.
- EASTMENT, H. e W. KRZANOWSKI, 1982 Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* **24**: 73–77.
- ECKART, C. e G. YOUNG, 1936 The approximation of one matrix by another of lower rank. *Psychometrika* **1**: 211–218.

- FALCONER, D., 1987 *Introdução à genética quantitativa*. UFV.
- FISHER, R. e W. MACKENZIE, 1923 Studies in crop variation. ii. the manurial response of different potato varieties. *The Journal of Agricultural Science* **13**: 311–320.
- GABRIEL, K., 1971 The biplot graphic display of matrices with application to principal component analysis. *Biometrika* pp. 453–467.
- GAUCH JR, H., 1992 *Statistical analysis of regional yield trials: AMMI analysis of factorial designs..* Elsevier Science Publishers.
- HARSHMAN, R., 1970 Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in phonetics* **16**: 1–84.
- HARSHMAN, R. e W. DE SARBO, 1984 An application of parafac to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. *Research methods for multimode data analysis* pp. 602–642.
- HOGG, R. e A. CRAIG, 1978 *Introduction to mathematical statistics*. Macmillan Publishing.
- JOHNSON, N., S. KOTZ, e N. BALAKRISHNAN, 1995 *Continuous univariate distributions, vol. 2*. Wiley, New York,.
- JOHNSON, R. e D. WICHERN, 1998 *Applied multivariate statistical analysis*. Prentice Hall.
- KANG, M., 1998 Using genotype-by-environment interaction for crop cultivar development. *Advances in agronomy* **62**: 199–252.
- KIERS, H. e R. HARSHMAN, 1997 Relating two proposed methods for speedup of algorithms for fitting two-and three-way principal component and related multilinear models. *Chemometrics and Intelligent Laboratory Systems* **36**: 31–40.
- KIERS, H. e A. KINDEREN, 2003 A fast method for choosing the numbers of components in tucker3 analysis. *British Journal of Mathematical and Statistical Psychology* **56**: 119–125.
- KIERS, H. e I. MECHELEN, 2001 Three-way component analysis: Principles and illustrative application. *Psychological methods* **6**: 84.
- KROONENBERG, P., 1983 *Three-mode principal component analysis: Theory and applications*, volume 2. DSWO press.
- KROONENBERG, P., 1984 Three-mode principal component analysis illustrated with an example from attachment theory. *Research Methods for Multimode Data Analysis* pp. 64–103.
- KROONENBERG, P., 1994 The tuckals line: A suite of programs for three-way data analysis. *Computational Statistics & Data Analysis* **18**: 73–96.
- KROONENBERG, P., 2008 *Applied multiway data analysis*. John Wiley & Sons.
- KROONENBERG, P. e J. DE LEEUW, 1980 Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **45**: 69–97.
- KRUSKAL, J., 1989 Rank, decomposition, and uniqueness for 3-way and n-way arrays. *Multiway Data Analysis* pp. 8–18.
- KUTNER, M., C. NACHTSHEIM, e J. NETER, 2005 *Applied linear regression models*. McGraw-Hill/Irwin.

- LEVIN, J., 1965 Three-mode factor analysis. *Psychological Bulletin* **64**: 442–452.
- LOUWERSE, D., A. SMILDE, e H. A. KIERS, 1999 Cross-validation of multiway component models. *Journal of chemometrics* **13**: 491–510.
- MANDEL, J., 1971 A new analysis of variance model for non-additive data. *Technometrics* **13**: 1–18.
- MARCO, P., M. LEVI, I. SCARMINIO, R. POPPI, e M. TREVISAN, 2005 Exploratory analysis of simultaneous degradation of anthocyanins in the calyces of flowers of the hibiscus sabdariffa species by parafac model. *Analytical sciences* **21**: 1523–1527.
- MOSER, W., 1996 *Linear models: a mean model approach*. Elsevier.
- MURPHY, K., G. RUIZ, W. DUNSMUIR, e T. D. WAITE, 2006 Optimized parameters for fluorescence-based verification of ballast water exchange by ships. *Environmental science & technology* **40**: 2357–2362.
- OLIVEIRA, A., J. DUARTE, e J. PINHEIRO, 2003 Emprego da análise ammi na avaliação da estabilidade produtiva em soja. *Pesq. agropec. bras., Brasília* **38**: 357–364.
- PERKINS, J. e J. JINKS, 1968 Environmental and genotype-environmental components of variability. *Heredity* **23**: 339–356.
- R CORE TEAM, 2017 *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- RAO, C. e S. MITRA, 1971 *Generalized inverse of matrices and its applications*. Wiley New York.
- RIU, J. e R. BRO, 2003 Jack-knife technique for outlier detection and estimation of standard errors in parafac models. *Chemometrics and Intelligent Laboratory Systems* **65**: 35–69.
- SANCHEZ, E. e B. KOWALSKI, 1990 Tensorial resolution: a direct trilinear decomposition. *Journal of Chemometrics* **4**: 29–45.
- SCHOTT, J., 1997 *Matrix analysis for statistics*. John Wiley & Sons.
- SMILDE, A., R. BRO, e P. GELADI, 2004 Multi-way analysis with applications in the chemical sciences.
- THURSTONE, L., 1935 *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* University of Chicago Press.
- TIMMERMAN, M. e H. KIERS, 2000 Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British journal of mathematical and statistical psychology* **53**: 1–16.
- TUCKER, L., 1963 Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change* pp. 122–137.
- TUCKER, L., 1964 The extension of factor analysis to three-dimensional matrices. *Contributions to mathematical psychology* pp. 110–127.
- TUCKER, L., 1966 Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**: 279–311.
- VARELA, M., J. CROSSA, J. RANE, A. JOSHI, e R. TRETHOWAN, 2006 Analysis of a three-way interaction including multi-attributes. *Crop and Pasture Science* **57**: 1185–1193.
- WANSBEEK, T. e J. VERHEES, 1989 Models for multidimensional matrices in econometrics and psychometrics. *Multiway data analysis* pp. 543–552.

YATES, F. e W. COCHRAN, 1938 The analysis of groups of experiments. The Journal of Agricultural Science **28**: 556–580.

APÊNDICES

Apêndice A

Códigos realizados no R para simulação e aderência

```
#####
# Etapa 1- Simulação
#####
ef<-read.table("efeitos.txt",h=F)
attach(ef)
V1
set.seed(100001)
r=3
y=3
l=3
g=13
sigma=sqrt(0.0394)      # ajustar para o QMRes do modelo
b0=V1[1]
blocos=array(V1[2:28], dim=c(r,l,y)) #(bloco, local, ano)
GEN= V1[29:41]
LOC=c(0.000, 0.000, 0.000)
GL=array(V1[45:83], dim=c(g,l))
YEAR=c(0.000, 0.000, 0.000)
GY=array(V1[87:125],dim=c(g,y))
LY=array(rep(0,(l*y)), dim=c(l,y))
GLY= array(V1[135:251],dim=c(g,l,y))
cena1=NULL
nexp=1000
for (exp in 1:nexp){
  for (i in 1:g){
    for (j in 1:l){
      for (k in 1:y){
        for (h in 1:r){
          resp=b0+blocos[h,j,k]+GEN[i]+LOC[j]+YEAR[k]+GL[i,j]+GY[i,k]+LY[j,k]+GLY[i,j,k]+
            +rnorm(1,0,sigma)
          cena1=rbind(cena1,c(exp,i,j,k,h,round(resp,4)))
        }}}
      label<-c("Exp", "GEN", "LOC", "YEAR", "BLOCO", "Resp")
      colnames(cena1)<-label
      write.table(cena1, "cena1_500.txt", sep="\t")
    }
  }
}

#####
# FIM da ETAPA1
#####

#####
# Etapa 2- Obtenção SQ
#####
rm( list=c("cena1", "GEN", "LOC", "YEAR") )
source("functions.r") # carregar funções do arquivo "functions";
dados<-read.table("cena1_500.txt",h=T)
names(dados)
SQGEN<-matrix(NA, ncol=nexp, nrow=g)
SQLOC<-matrix(NA, ncol=nexp, nrow=l)
SQANO<-matrix(NA, ncol=nexp, nrow=y)
for(i in 1:nexp){
```



```

dados1<-dados[(1+(i-1)*351):(i*351),]
attach(dados1)
YEAR <-as.factor(YEAR)
LOC <-as.factor(LOC)
GEN <-as.factor(GEN)
BLOCO <- as.factor(BLOCO)
mod <- aov(Resp~YEAR*LOC*GEN + BLOCO%in%(LOC:YEAR))
QMRes<-anova(mod)[9,3]
Int_Tripla<-tripla(GEN,LOC,YEAR,BLOCO,Resp)
SQGEN[,i]<-Gen(Int_Tripla,3,QMRes)[,2]
SQLOC[,i]<-Loc(Int_Tripla,3,QMRes)[,2]
SQANO[,i]<-Ano(Int_Tripla,3,QMRes)[,2]
}
SQGEN<-t(SQGEN)
SQLOC<-t(SQLOC)
SQANO<-t(SQANO)
write.table(SQGEN, "SQGEN.txt", sep="\t")
write.table(SQLOC, "SQLOC.txt", sep="\t")
write.table(SQANO, "SQANO.txt", sep="\t")

#####
# FIM da ETAPA2
#####

#####
# Etapa 3 - Avaliação da simulação
#####
require(MASS)
require(xtable)

#####
# Para Genótipos
#####
dados<-read.table("SQGEN.txt",h=T)
dados<-as.matrix(dados)
n<-g # ajustar
g.l<-(g-1)*(l-1)*(y-1)/n
valores_Gen<-matrix(,n,2)
for(j in 1:n){
x<-dados[,j]/(sigma^2)
est.lamb<-fitdistr(x,"chi-squared",df=g.l,list(ncp=mean(x)), lower=0.01)
valores_Gen[j,]<- c(est.lamb$est[[1]],ks.test(x,"pchisq",g.l,est.lamb$est[[1]])$p)
}
label_gen<-paste0(rep("GEN", g)," ",1:g)
colnames(valores_Gen)<-c("nao centralidade", "valor p")
rownames(valores_Gen)<-label_gen
round(valores_Gen,4)
print.xtable(xtable(valores_Gen,dig=4), type="html", file="teste_ks_gen.html")

##Histogramas
nplot=5 # ajustar
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("HIST_GEN", k,".jpeg", sep=""))
par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Gen[j,1]

```

```

hist(x,freq=FALSE,main="",sub=label_gen[j],xlab="",ylab="")
yy <- sort(rchisq(500, df=g.l,ncp=nc))
points(yy,dchisq(yy,g.l,nc),type="l")
}
dev.off()
}
dev.off()
dev.off()
dev.off()
dev.off()
dev.off()
dev.off()

##qqplot
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("QQPLOT_GEN", k, ".jpeg", sep=""))
par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Gen[j,1]
yy <- qchisq(ppoints(length(x)), df=g.l,ncp=nc)
qqplot(x, yy,xlab="Quantis amostrais",ylab="Quantis teóricos");
title(sub = label_gen[j]); abline(0,1)
}
dev.off()
}
dev.off()
dev.off()
dev.off()
dev.off()
dev.off()
dev.off()

#####
# Para Locais
#####
dados<-read.table("SQLLOC.txt",h=T)
dados<-as.matrix(dados)
n<-l # ajustar
g.l<-(g-1)*(l-1)*(y-1)/n
valores_Loc<-matrix(,n,2)
for(j in 1:n){
x<-dados[,j]/(sigma^2)
est.lamb<-fitdistr(x,"chi-squared",df=g.l,list(ncp=mean(x)), lower=0.01)
valores_Loc[j,]<- c(est.lamb$est[[1]],ks.test(x,"pchisq",g.l,est.lamb$est[[1]])$p)
}
label_Loc<-paste0(rep("LOC", 1)," ",1:1)
colnames(valores_Loc)<-c("nao centralidade", "valor p")
rownames(valores_Loc)<-label_Loc
valores_Loc
print.xtable(xtable(valores_Loc,dig=4), type="html", file="teste_ks_loc.html")

##Histogramas
nplot=1 # ajustar
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("HIST_LOC", k, ".jpeg", sep=""))

```

```

par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Loc[j,1]
hist(x,freq=FALSE,main="",sub=label_Loc[j],xlab="",ylab="")
yy <- sort(rchisq(500, df=g.l,ncp=nc))
points(yy,dchisq(yy,g.l,nc),type="l")
}
dev.off()
}
dev.off()

##qqplot
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("QQPLOT_LOC", k, ".jpeg", sep=""))
par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Loc[j,1]
yy <- qchisq(ppoints(length(x)), df=g.l,ncp=nc)
qqplot(x, yy, xlab="Quantis amostrais",ylab="Quantis teóricos");
title(sub = label_Loc[j]); abline(0,1)
}
dev.off()
}
dev.off()

#####
# Para Anos
#####
dados<-read.table("SQANO.txt",h=T)
dados<-as.matrix(dados)
n<-y # ajustar
g.l<-(g-1)*(l-1)*(y-1)/n
valores_Ano<-matrix(n,n,2)
for(j in 1:n){
x<-dados[,j]/(sigma^2)
est.lamb<-fitdistr(x,"chi-squared",df=g.l,list(ncp=mean(x)), lower=0.01)
valores_Ano[j,]<- c(est.lamb$est[[1]],ks.test(x,"pchisq",g.l,est.lamb$est[[1]])$p)
}
label_Ano<-paste0(rep("ANO", y)," ",1:y)
colnames(valores_Ano)<-c("nao centralidade", "valor p")
rownames(valores_Ano)<-label_Ano
valores_Ano
print.xtable(xtable(valores_Ano,dig=4), type="html", file="teste_ks_ano.html")

##Histogramas
nplot=1 # ajustar
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("HIST_ANO", k, ".jpeg", sep=""))
par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Ano[j,1]
hist(x,freq=FALSE,main="",sub=label_Ano[j],xlab="",ylab="")
yy <- sort(rchisq(500, df=g.l,ncp=nc))

```

```

points(yy,dchisq(yy,g.l,nc),type="l")
}
dev.off()
}
dev.off()

##qqplot
for (k in 0:(nplot-1)){
X11()
jpeg(filename=paste("QQPLOT_ANO", k, ".jpeg", sep=""))
par(mfrow = c(1,3))
for(j in (1+3*k):(3+3*k)){
x<-dados[,j]/(sigma^2)
nc=valores_Ano[j,1]
yy <- qchisq(ppoints(length(x)), df=g.l,ncp=nc)
qqplot(x, yy, xlab="Quantis amostrais",ylab="Quantis teóricos");
title(sub = label_Ano[j]); abline(0,1)
}
dev.off()
}
dev.off()

```

Apêndice B

Funções em R para teste F

```

#####
# estimar interação tripla                                     #
#                                                             #
# como usar: Int_Tripla<-tripla(GEN,AMB,ANO,REP,Resp) #
# GEN: vetor com os níveis dos genótipos                #
# AMB: vetor com os níveis dos ambientes                #
# ANO: vetor com os níveis dos anos                    #
# REP: vetor com os níveis das repetições ou blocos    #
# Resp: vetor com a variável resposta                  #
#####

tripla<-function (GEN,AMB,ANO,REP,PROD){
ANO <-as.factor(ANO)
AMB <-as.factor(AMB)
GEN <-as.factor(GEN)
REP <- as.factor(REP)
y<-length(levels(ANO))
l<-length(levels(AMB))
g<-length(levels(GEN))
b<-length(levels(REP))
m<-mean(PROD)
Ym<-tapply(PROD,ANO,mean)
Lm<-tapply(PROD,AMB,mean)
Gm<-tapply(PROD,GEN,mean)

#efeitos principais
EY<-Ym-m
EL<-Lm-m
EG<-Gm-m

#efeitos duplos

```

```

GLm<-tapply(PROD,list(GEN,AMB),mean)
GYm<-tapply(PROD,list(GEN,ANO),mean)
LYm<-tapply(PROD,list(AMB,ANO),mean)
EGL<-matrix(NA,g,l)
for(i in 1:g){
for(j in 1:l){
EGL[i,j]=GLm[i,j]-Gm[i]-Lm[j]+m
}}
EGY<-matrix(NA,g,y)
for(i in 1:g){
for(j in 1:y){
EGY[i,j]=GYm[i,j]-Gm[i]-Ym[j]+m
}}
ELY<-matrix(NA,l,y)
for(i in 1:l){
for(j in 1:y){
ELY[i,j]=LYm[i,j]-Lm[i]-Ym[j]+m
}}
ELY

#efeitos triplos
GLYm<-tapply(PROD,list(GEN, AMB, ANO),mean)
EGLY<-array(NA,c(g,l,y))
for(i in 1:g){
for(j in 1:l){
for (k in 1:y){
EGLY[i,j,k]=GLYm[i,j,k]-GLm[i,j]-GYm[i,k]-LYm[j,k]+Gm[i]+Lm[j]+Ym[k]-m
}}}
return(EGLY)
}

#####
# Teste F cada nivel do fator genótipo #
# # #
# Gen(Int_Tripla,nrep,QMRes) #
# Int_Tripla: arranjo com efeitos da interação tripla #
# nrep: n. de repetições #
# QMRes: Quadrado Médio do Resíduo da ANOVA #
#####
Gen<-function(Int_Tripla,nrep,QMRes){
g<-dim(Int_Tripla)[1]
l<-dim(Int_Tripla)[2]
y<-dim(Int_Tripla)[3]
gl_g<-(g-1)*(l-1)*(y-1)/g
glRes<-l*y*(g-1)*(nrep-1)
Gen_i<-matrix(NA,nrow=g,ncol=5)
for (i in 1:g){
SQ<-nrep*sum(Int_Tripla[i,,]^2)
QM<-SQ/gl_g
F<-QM/QMRes
p<-pf(F,gl_g,glRes,lower.tail = FALSE)
Gen_i[i,]<-cbind(round(gl_g,2),round(SQ,4),round(QM,4),round(F,4),round(p,4))
}
colnames(Gen_i)<-c("GL","SQ","QM","F", "valor p")
return(as.data.frame(Gen_i))
}

#####

```

```

# Teste F cada nivel do fator Local                                     #
#                                                                       #
# Loc(Int_Tripla,nrep,QMRes)                                           #
# Int_Tripla: arranjo com efeitos da interação tripla #
# nrep: n. de repetições                                             #
# QMRes: Quadrado Médio do Resíduo da ANOVA                         #
#####
Loc<-function(Int_Tripla,nrep,QMRes){
  g<-dim(Int_Tripla)[1]
  l<-dim(Int_Tripla)[2]
  y<-dim(Int_Tripla)[3]
  gl_g<-(g-1)*(l-1)*(y-1)/l
  glRes<-l*y*(g-1)*(nrep-1)
  Loc_j<-matrix(NA,nrow=l,ncol=5)
  for (j in 1:l){
    SQ<-nrep*sum(Int_Tripla[,j]^2)
    QM<-SQ/gl_g
    F<-QM/QMRes
    p<-pf(F,gl_g,glRes,lower.tail = FALSE)
    Loc_j[j,]<-cbind(round(gl_g,2),round(SQ,4),round(QM,4),round(F,4),round(p,4))
  }
  colnames(Loc_j)<-c("GL","SQ","QM","F", "valor p")
  return(as.data.frame(Loc_j))
}

#####
# Teste F cada nivel do fator Ano                                     #
#                                                                       #
# Ano(Int_Tripla,nrep,QMRes)                                           #
# Int_Tripla: arranjo com efeitos da interação tripla #
# nrep: n. de repetições                                             #
# QMRes: Quadrado Médio do Resíduo da ANOVA                         #
#####
Ano<-function(Int_Tripla,nrep,QMRes){
  g<-dim(Int_Tripla)[1]
  l<-dim(Int_Tripla)[2]
  y<-dim(Int_Tripla)[3]
  gl_g<-(g-1)*(l-1)*(y-1)/l
  glRes<-l*y*(g-1)*(nrep-1)
  Ano_k<-matrix(NA,nrow=y,ncol=5)
  for (k in 1:y){
    SQ<-nrep*sum(Int_Tripla[,k]^2)
    QM<-SQ/gl_g
    F<-QM/QMRes
    p<-pf(F,gl_g,glRes,lower.tail = FALSE)
    Ano_k[k,]<-cbind(round(gl_g,2),round(SQ,4),round(QM,4),round(F,4),round(p,4))
  }
  colnames(Ano_k)<-c("GL","SQ","QM","F", "valor p")
  return(as.data.frame(Ano_k))
}

```