



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**ANÁLISE DE CLUSTER E BOOTSTRAP
NO ESTUDO DA INTERAÇÃO
GENÓTIPO $\times$ AMBIENTE**

João Vitor Ribeiro Silva

Uberlândia-MG
2018

João Vitor Ribeiro Silva

**ANÁLISE DE CLUSTER E BOOTSTRAP
NO ESTUDO DA INTERAÇÃO
GENÓTIPO $\times$ AMBIENTE**

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

Orientador: Lúcio Borges de Araújo

**Uberlândia-MG
2018**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

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

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Lúcio Borges de Araújo

Priscila Neves Faria

Janser Moura Pereira

**Uberlândia-MG
2018**

AGRADECIMENTOS

Gostaria de agradecer aos meus pais, José Geraldo e Maria Aparecida, por todo o apoio, amor, conselhos. Vocês foram essenciais para que eu pudesse conquistar mais uma etapa. Obrigado por toda a ajuda, por estarem sempre ao meu lado e por fazer o possível e impossível para que eu pudesse chegar até aqui. Essa conquista é apenas mais uma de nossas conquistas!

Agradeço a minha irmã Débora, por ter sido meu apoio durante toda a graduação. Você sempre foi o meu exemplo, obrigado pelas conversas, pelos conselhos, companhia e, acima de tudo, por sempre estar do meu lado.

As minhas irmãs Bárbara e Livia, agradeço por todo o exemplo e apoio. Agradeço a toda a minha família pelas energias positivas e por sempre confiarem na minha capacidade de conseguir realizar os meus sonhos.

As pessoas que conheci durante a graduação, não tenho palavras para agradecer por todos esses anos de companheirismo.

Em especial, queria agradecer ao William Costa, Lia Rezende Bruno Albernaz e Julienderson Costa por serem as melhores pessoas que eu poderia conhecer, obrigado por todo o apoio e por sempre estarem ao meu lado. Vocês são muito especiais para mim.

Aos meus colegas e amigos, vocês foram parte dessa conquista. Obrigado Rafael Esposti, Amanda Londero, Matheus Júlio e todos os demais pela companhia, conselhos e amizade.

Ao meu orientador, Prof. Dr. Lúcio Borges de Araújo, agradeço por todos os conselhos, por anos de orientação e pela confiança em mim. Você com certeza foi um grande mestre, que eu vou levar para a vida toda.

Agradeço a todos os professores, técnicos e servidores que fizeram parte dessa conquista. Sem vocês, nada disso seria possível. Obrigado por anos de ensinamentos e conselhos.

Agradeço aos colegas de estágio, por todo o desenvolvimento e crescimento que obtive ao lado de vocês. Bruna Costa, Kênia Oliveira, Aline Silva, Maria Clara Soares e Suemar Avelar, muito obrigado.

A todos que, de alguma forma fizeram parte dessa conquista, eu só devo a vocês o meu muito obrigado!

RESUMO

A experimentação agrícola é uma técnica estatística que está se tornando cada dia mais reconhecida e utilizada, pois existe a necessidade de produzir mais alimentos em menos tempo/espço. Estudos mostram que a interação entre genótipos e ambientes é importante nesse tipo de análise, pois existem genótipos que são favorecidos em certos ambientes, o que traz um aumento na produção desses genótipos.

Nesse trabalho serão apresentados o uso de uma técnica de experimentação agrícola chamada AMMI (*additive main effects and multiplicative interaction model*) cujo objetivo é analisar a interação entre genótipos e ambientes utilizando o método *Bootstrap* para fazer uma reamostragem dos dados e obter informações mais precisas.

Além da análise AMMI, será realizada uma análise de *Cluster* para verificar se os ambientes que estão divididos em regiões são homogêneos dentro dos grupos e heterogêneos entre os mesmos.

Palavras-chave: Experimentação agrícola, AMMI, cluster analysis.

ABSTRACT

Agricultural experimentation is a statistical technique that is becoming more and more recognized and used since there is a need to produce more food in less time/space.

Studies show that the interaction between genotypes and environments is important in this type of analysis, because there are genotypes that are favored in certain environments, which brings an increase in the production of these genotypes.

In this paper we will present the use of an agricultural experimentation technique called AMMI (additive main effects and multiplicative interaction model) whose objective is to analyze the interaction between genotypes and environments using the Bootstrap method to make a resampling of the data and obtain more accurate information.

In addition to the AMMI analysis, a Cluster analysis will be performed to verify if the environments that are divided into regions are homogeneous within the groups and heterogeneous among them.

Keywords: Agricultural experimentation, AMMI, cluster analysis.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Metodologia	5
3 Resultados	9
3.1 Análise de Cluster	9
3.2 Análise AMMI	10
3.2.1 Reamostragem Bootstrap	12
4 Conclusões	17
Referências Bibliográficas	19

LISTA DE FIGURAS

3.1	Biplot dos genótipos e ambientes	11
3.2	Região de confiança de 95% para os genótipos	13
3.3	Região de confiança de 95% para os ambientes	14

LISTA DE TABELAS

2.1	ANAVA conjunta para a análise da interação $G \times E$	6
2.2	Nível de classificação de estabilidade dos genótipos e ambientes	8
3.1	Divisão dos ambientes por região	9
3.2	Análise descritiva das regiões para a produtividade em $kg.ha^{-1}$	9
3.3	Classificação dos ambientes conforme análise de <i>Cluster</i>	10
3.4	Análise descritiva para as novas regiões obtidas na análise de <i>Cluster</i> para a produtividade em $kg.ha^{-1}$	10
3.5	ANAVA conjunta para a análise AMMI	11
3.6	Decomposição da interação $G \times E$ utilizando o método de Gollob	11
3.7	Análise descritiva dos genótipos para a produtividade em $kg.ha^{-1}$	12
3.8	Nível de estabilidade dos genótipos	13
3.9	Análise descritiva dos ambientes para a produtividade em $kg.ha^{-1}$	14
3.10	Nível de estabilidade dos ambientes	15

1. INTRODUÇÃO

No Brasil, o milho é a segunda cultura mais produzida, perdendo apenas para a soja. Além disso, o Brasil é o segundo maior exportador de milho do mundo, sendo os Estados Unidos o maior exportador.

O milho é uma das culturas mais antigas do mundo, alimento para a subsistência humana, principal insumo para a produção de alguns animais (como suínos e aves) e, segundo a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) é muito importante economicamente, por suas diversas formas de utilização.

O economista inglês Thomas Robert Malthus em seu livro intitulado “Ensaio sobre o princípio da população” [13], desenvolveu a uma teoria demográfica conhecida como malthusianismo. Nessa teoria, o autor diz que a população mundial cresce seguindo uma progressão geométrica, ou seja, cresce exponencialmente, enquanto a produção de alimentos cresce seguindo uma progressão aritmética, o que ocasionaria na falta de alimentos e fome em todo o mundo.

Após esse fato, viu-se uma necessidade em aumentar a produção de alimentos para suprir todas as necessidades da população. Segundo anunciado pela Organização das Nações Unidas para a Alimentação e a Agricultura [7], a produção de alimentos deveria aumentar em 70% nos próximos 40 anos. Com isso, os estudos de cultivares geneticamente modificados foram evoluindo por meio de experimentação agrícola multi-ambientais, ganhando uma maior importância.

Uma análise proposta por [8] em 1988, realizava a decomposição da soma de quadrados ($SQ_{G \times E}$) em eixos, separados em padrão e ruído, em que os primeiros eixos (eixos de menor grau) eram compostos por todo o padrão da interação genótipo $\times$ ambiente ($G \times E$) e os últimos eixos continham somente o ruído. Essa análise foi chamada de modelo de efeitos aditivos principais e interação multiplicativa (AMMI, do inglês, *Additive Main Effects and Multiplicative Interaction*).

Em alguns casos, existe uma necessidade em generalizar os resultados para condições gerais de regiões/ambientes na avaliação de diversos genótipos de cultivares. Pesquisadores, na maioria das vezes, desejam encontrar um genótipo que, combinado com um ambiente possua uma interação positiva e resulte em um maior rendimento, visando aumentar a produção de alimentos. Porém, essa interação, também chamada de $G \times E$, pode fazer com que os melhores genótipos em alguns ambientes, não sejam os melhores em outros, conforme [3].

Para [1] a interação resulta da resposta diferente dos genótipos em ambientes distintos, por isso, é importante que existam diferenças entre os ambientes, para que possa resultar um

desempenho diferente para cada local

A significância dessa interação $G \times E$ é verificada apenas quando o experimento é repetido múltiplas vezes, pois realizar o experimento uma única vez poderá superestimar os ganhos genéticos [2].

Existem diversos métodos para a realização da análise da interação $G \times E$, como a análise de variância (ANOVA), análise de componentes principais (ACP), regressão linear (RL), zoneamento agrônomo e avaliação de padrões de dissimilaridade, porém esses métodos possuem uma baixa eficiência [6].

Um método que tem se destacado dentre os demais é o modelo AMMI, pois todas as outras análises, como análise de variância (ANOVA), análise de componentes principais (ACP) e regressão linear (RL), são consideradas sub-casos da análise AMMI pois, se tivermos apenas os efeitos aditivos, a análise AMMI é reduzida a uma ANOVA, se tivermos apenas os efeitos multiplicativos, a análise AMMI é reduzida a uma ACP e se tivermos apenas uma relação entre genótipo e ambiente, a análise AMMI é reduzida a uma RL [14].

O modelo AMMI é um método utilizado para generalizar os dados de genótipos e ambientes provenientes de uma tabela de dupla entrada e engloba a ANOVA para os efeitos aditivos, como os genótipos e ambientes e também para o efeito multiplicativo, ou seja, a interação $G \times E$. É realizado um desdobramento da soma de quadrados da interação $G \times E$ ($SQ_{G \times E}$) para retirar resíduos adicionais e, por fim, os dados são aproximados a uma matriz de interação, o que permite uma interpretação dos dados sem ruídos.

Existem diversos métodos para fazer o desdobramento da interação $G \times E$ e definir o modelo AMMI final, porém o método proposto por [9] que utiliza a decomposição em valor singular (DVS) é um método simples, que garante um número maior de fatores multiplicativos com um menor rigor nos graus de liberdade e, consequentemente, explica melhor o padrão em razão da interação $G \times E$ [12].

Após a escolha do modelo final, um gráfico biplot é obtido para realizar as análises visualmente, o que permite uma interpretação dos dados mais detalhada. Segundo [4], uma das virtudes do método AMMI está na possibilidade de representar os efeitos da interação para cada genótipo e ambiente em um único gráfico.

Com a falta de precisão na identificação de estabilidade dos genótipos e ambientes por parte do gráfico biplot [6] e o desconhecimento da distribuição de probabilidade do modelo AMMI, viu-se a necessidade de utilização de um outro método para realizar a essa análise e, para isso, foi utilizada a reamostragem *Bootstrap*.

A metodologia de reamostragem *Bootstrap* proposta por [5] consiste em dividir a amostra em sub-amostras de tamanho menor e reamostrar essas sub-amostras quantas vezes julgar necessário, afim de obter uma amostra maior com um nível de incerteza menor e melhorar a qualidade do modelo [11].

Para a realização de comparações entre grupos, a análise de agrupamentos (ou análise de *Cluster*) é uma análise simples, de fácil aplicação e com resultados de fácil interpretação. A análise de *Cluster* é uma estatística multivariada cujo objetivo é realizar a separação de grupos

de dados, em que esses grupos sejam homogêneos e, cada grupo, seja o mais heterogêneo possível do outro grupo [10].

Os objetivos desse trabalho são realizar uma reamostragem *Bootstrap* para avaliar as incertezas do modelo AMMI e as estabilidades dos genótipos e ambientes, além de realizar uma análise de *Cluster* para verificar se os ambientes em cada região tem um desempenho homogêneo em relação a interação com os genótipos.

2. METODOLOGIA

Os dados a serem utilizados são dados de milho amarelo (*Zea mays*), obtidos de uma multinacional na área de sementes, que contam com 3 repetições de 9 genótipos em 7 ambientes, provenientes de 4 regiões. Os plantios ocorreram na safra de 2016 – 2017 em diferentes locais no Brasil e a produtividade dos materiais foi medida em quilograma por hectare (*kg/ha*).

Uma análise de *cluster* hierárquico será realizada para verificar se as regiões utilizadas são heterogêneas entre si (e homogêneas dentro das mesmas) quanto a interação $G \times E$ nesses locais. A análise de *cluster* utiliza da distância euclidiana para analisar a distância entre os pontos para que possa ser calculado o índice de similaridade entre os pontos.

A distância euclidiana ao quadrado entre dois pontos $a = (a_1, \dots, a_n)$ e $b = (b_1, \dots, b_n)$ pode ser calculada por:

$$X^2_{ab} = \sum_{j=1}^n (d_{aj} - d_{bj})^2 \quad (2.1)$$

em que:

d_{aj} é a j -ésima coordenada de a ;

d_{bj} é a j -ésima coordenada de b .

Para isso, foi utilizado o método de K-means, que visa agrupar as observações em k grupos, estes, com a menor distância interna. Esse método funciona em 4 etapas:

1ª Etapa : são gerados k centroides aleatórios;

2ª Etapa : são calculadas todas as distâncias entre todas as observações e os centroides e assim, as observações cuja distância é mínima com o centroide, é atribuída àquele grupo;

3ª Etapa : com todos os grupos formados, é feito um novo cálculo para o centroide, utilizando a média dos valores de cada grupo;

4ª Etapa : são repetidas as etapas 2 e 3 até que os centroides se mantenham estáticos.

Assim, os ambientes serão classificados entre os k grupos criados, de acordo com a distância de cada uma das observações, onde os grupos não necessariamente possuem o mesmo tamanho.

Para a análise AMMI, o primeiro passo é realizar uma análise de variância conjunta (ANAVA) para a interação $G \times E$. A ANAVA baseia-se na decomposição da variação total da variável resposta em efeitos principais (Genótipo e Ambiente), efeitos multiplicativos (interação $G \times E$) e erro experimental, sendo que, em cada ambiente, teve-se um delineamento em blocos

casualizados. O modelo AMMI pode ser descrito por:

$$Y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + b_{k(j)} + \varepsilon_{ijk} \quad (2.2)$$

sendo

Y_{ijk} : é o valor da produtividade observada no i-ésimo genótipo, no j-ésimo ambiente e no k-ésimo bloco, com $i=1,\dots,g$; $j=1,\dots,e$; $k=1,\dots,r$;

μ : é a média geral;

g_i : é o efeito do i-ésimo genótipo;

e_j : é o efeito do j-ésimo ambiente;

$(ge)_{ij}$: é o efeito da interação do genótipo i com o ambiente j;

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

ε_{ijk} : é o erro experimental associado ao i-ésimo genótipo, no j-ésimo ambiente e no k-ésimo bloco, independente e com distribuição normal com média 0 e variância σ^2 .

A Tabela 2.1 mostra como seria o primeiro passo para fazer a análise AMMI, ou seja, a ANAVA conjunta para os dados.

Tabela 2.1: ANAVA conjunta para a análise da interação $G \times E$

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrados Médios	Teste F
Blocos	SQ_{Bloco}	$e(r-1)$	$QM_{Bloco} = \frac{SQ_{Bloco}}{e(r-1)}$	$\frac{QM_{Bloco}}{QM_{Erro}}$
Genótipos (G)	SQ_G	$(g-1)$	$QM_G = \frac{SQ_G}{(g-1)}$	$\frac{QM_G}{QM_{Erro}}$
Ambientes (E)	SQ_E	$(e-1)$	$QM_E = \frac{SQ_E}{(e-1)}$	$\frac{QM_E}{QM_{Erro}}$
Interação $G \times E$	$SQ_{G \times E}$	$(g-1)(e-1)$	$QM_{G \times E} = \frac{SQ_{G \times E}}{(g-1)(e-1)}$	$\frac{QM_{G \times E}}{QM_{Erro}}$
Erro	SQ_{Erro}	$e(g-1)(r-1)$	$QM_{Erro} = \frac{SQ_{Erro}}{eg(n-1)}$	-
Total	SQ_{Total}	$(n-1)$	-	-

em que a variação total da variável resposta pode ser medida por meio das somas de quadrados definidas para cada uma das componentes:

$$SQ_G = e \sum_i (\hat{g}_i)^2 \quad (2.3)$$

$$SQ_E = g \sum_j (\hat{e}_j)^2 \quad (2.4)$$

$$SQ_{G \times E} = \sum_{i,j} (\hat{ge}_{ij})^2 \quad (2.5)$$

$$SQ_{ERRO} = \sum_{i,j,k} \hat{\varepsilon}_{ijk}^2 \quad (2.6)$$

$$SQ_{TOTAL} = SQ_G + SQ_E + SQ_{G \times E} + SQ_{ERRO} \quad (2.7)$$

sendo:

r : é o número de repetições;

n : é o tamanho da amostra;

SQ_G : é a soma de quadrados dos genótipos;

SQ_E : é a soma de quadrados dos ambientes;

SQ_{ERRO} : é a soma de quadrados dos erros;

$SQ_{G \times E}$: é a soma de quadrados da interação;

SQ_{TOTAL} : é a soma de quadrados total;

e : é o número de ambientes;

g : é o número de genótipos;

$\hat{g}_i$: é o efeito do i -ésimo genótipo;

$\hat{e}_j$: é o efeito do j -ésimo ambiente;

$\hat{\varepsilon}_{ij}$: é o erro experimental associado ao i -ésimo genótipo, do j -ésimo ambiente no k -ésimo bloco;

$\hat{g}e_{ij}$: é o termo efeito da interação do i -ésimo genótipo com o j -ésimo ambiente.

Com a interação $G \times E$ significativa, o próximo passo é fazer a decomposição da $SQ_{G \times E}$, utilizando a decomposição em valor singular e o método de Gollob, que é o método mais simples utilizado na análise AMMI e permite a captação de todo padrão da interação $G \times E$. Através do método de Gollob, é possível obter a matriz de efeitos $\widehat{GE}$.

Dessa matriz de efeitos $\widehat{GE}$ foram feitas 400 reamostragens *bootstrap* em relação aos genótipos (com reposição nas linhas) e 400 reamostragens *bootstrap* em relação aos ambientes (com reposição nas colunas), obtendo as matrizes $\widehat{GE}_{kl}^*$, em que k representa o genótipo (g) ou ambiente (e); $l = 1, \dots, 400$.

A fim de obter a matriz de resíduos corrigidos, a matriz $\widehat{GE}_{kl}^*$ foi centrada em relação a média, linha e coluna por:

$$\overline{ge}_{ij}^* = \hat{ge}_{ij}^* - \hat{ge}_{i.}^* - \hat{ge}_{.j}^* + \hat{ge}_{..}^* \quad (2.8)$$

em que:

$\overline{ge}_{ij}^*$ é o resíduo corrigido do i -ésimo genótipo no j -ésimo ambiente;

$\hat{ge}_{ij}^*$ é o valor residual observado no i -ésimo genótipo no j -ésimo ambiente;

$\hat{ge}_{i.}^*$ é a média dos resíduos do i -ésimo genótipo;

$\hat{ge}_{.j}^*$ é a média dos resíduos do j -ésimo ambiente;

$\hat{ge}_{..}^*$ é a média geral dos resíduos.

Assim que os resíduos são centralizados, é possível calcular o efeito *bootstrap* para os genótipos e para os ambientes. Esse efeito é calculado utilizando a decomposição em valor singular (DVS), onde o efeito *bootstrap* da interação é decomposto em 3 partes: $\overline{GE}_{ij}^* = U * S * V'$, sendo U a matriz com os vetores *bootstrap* singular-coluna γ , S a matriz com os autovalores *bootstrap* e V a matriz com os vetores *bootstrap* singular-linha. Assim, os efeitos do genótipo e

ambiente serão:

$$\overline{G}^* = US^{1/2} \quad (2.9)$$

$$\overline{E}^* = S^{1/2}V' \quad (2.10)$$

em que: $\overline{G}^*$ é o efeito *bootstrap* para os genótipos e $\overline{E}^*$ é o efeito *bootstrap* para os ambientes.

Os efeitos calculados também podem ser chamados de escores. Esses escores são notas atribuídas a cada genótipo e ambiente que serão utilizados para obter a região de confiança em que cada um se encontra no plano cartesiano $IPCA_1 \times IPCA_2$ (em que $IPCA_i$ indica a i-ésima componente principal da interação) com o objetivo de definir se esse genótipo ou ambiente está ou não em uma região de estabilidade.

A região de confiança de 95% para a média dos escores, também chamada de Coeficiente Bootstrap de Estabilidade (CBE), pode ser calculada baseada no quadrado da distância de mahalanobis e é dada por:

$$CBE = [(\overline{\theta}_k^* - \theta_k)' S_{\theta_k}^{-1} (\overline{\theta}_k^* - \theta_k) \leq \frac{2(B-1)}{(B-2)} F_{(2, B-2)\alpha}] B^{-1} \quad (2.11)$$

sendo para cada genótipo e ambiente:

B o número de amostras *bootstrap*;

$\overline{\theta}_k^*$ o vetor $[\overline{IPCA}_1; \overline{IPCA}_2]$;

S_{θ_k} a matriz de covariâncias.

Assim, quando a região de confiança engloba o ponto $(0, 0)$, pode-se dizer que a variável estudada é estável. Quando a variável é estável, essa estabilidade pode ser classificada em 4 níveis, de acordo com o CBE, conforme segue na Tabela 2.2.

Tabela 2.2: Nível de classificação de estabilidade dos genótipos e ambientes

Estimativa	Nível de Estabilidade
$CBE \leq Q_3$	Consistente
$Q_2 \leq CBE \leq Q_3$	Alta
$Q_1 \leq CBE \leq Q_2$	Média
$CBE < Q_1$	Baixa

em que:

CBE varia entre 0 e 1;

Q_i representa o i-ésimo quartil no intervalo.

Todas as análises foram feitas utilizando o software R version 3.4.4, com um nível de significância de 5%.

3. RESULTADOS

3.1 ANÁLISE DE CLUSTER

A primeira análise a ser realizada é a análise de *Cluster*, com o objetivo de verificar se existe uma homogeneidade entre as regiões dos ambientes analisados. Os 7 ambientes estão divididos em 4 regiões, que podem ser observadas na Tabela 3.1.

Tabela 3.1: Divisão dos ambientes por região

Região	Ambiente
R01	E01
	E03
R02	E05
	E07
R03	E04
	E06
R04	E02

Para caracterizar essas regiões, foi realizada uma análise descritiva simples, contendo os mínimos, máximos e média de cada uma das regiões. Esses resultados podem ser observados na Tabela 3.2. Pode-se observar que a região 04 é a mais produtiva e a região 03 a menos produtiva.

Tabela 3.2: Análise descritiva das regiões para a produtividade em $kg.ha^{-1}$

Região	Min	Max	Média
R01	1739,59	10577,01	6190,75
R02	4153,79	9468,62	7023,34
R03	1794,21	9242,36	5640,28
R04	5399,48	12030,57	9095,73

Utilizando a análise de *Cluster* e o método de k-means com 4 grupos, obtem-se os ambientes classificados conforme a Tabela 3.3, onde pode ser observado que as regiões não são as mesmas dos dados originais, com excessão da R04.

Tabela 3.3: Classificação dos ambientes conforme análise de *Cluster*

Classificação	Ambiente
$\hat{R}01$	E01
$\hat{R}04$	E02
$\hat{R}03$	E03
$\hat{R}03$	E04
$\hat{R}03$	E05
$\hat{R}02$	E06
$\hat{R}02$	E07

Assim, os novos grupos formados pela análise de *Cluster* são: $\hat{R}01$: E01; $\hat{R}02$: E06 e E07; $\hat{R}03$: E03, E04 e E05; $\hat{R}04$: E02 e a análise descritiva para esses novos grupos pode ser observada na Tabela 3.4. Sendo que $\hat{R}02$ é a região menos produtiva e $\hat{R}04$ a região mais produtiva.

Tabela 3.4: Análise descritiva para as novas regiões obtidas na análise de *Cluster* para a produtividade em $kg.ha^{-1}$

Região	Min	Max	Média
$\hat{R}01$	5068,97	10577,01	7376,69
$\hat{R}02$	1794,21	8606,35	5449,06
$\hat{R}03$	1739,59	9468,62	6477,98
$\hat{R}04$	5399,48	12030,57	9095,73

Ou seja, mesmo os ambientes pertencendo a mesma região geográfica no Brasil, os ambientes que pertencem a mesma região tiveram efeito heterogêneo na interação com os genótipos e assim foram agrupados em *Clusters* diferentes durante a análise. Isso pode ser devido à diferenças existentes dentro das regiões que afetam os ambientes, como o clima, altitude ou condições meteorológicas, por exemplo.

3.2 ANÁLISE AMMI

Para a análise AMMI, foram utilizados 9 genótipos e 7 ambientes, estes, provenientes de 4 regiões e contando com 3 repetições cada, ou seja, esse experimento contou com um total de 189 observações.

O primeiro passo para a realização da análise AMMI é realizar a ANAVA e o resultado pode ser observado na Tabela 3.5.

Tabela 3.5: ANAVA conjunta para a análise AMMI

FV	GL	SQ	QM	F	Valor p
Ambiente	6	445895639	74315940	13,58	<0,0001
Bloco	14	76587217	5470515	5,14	<0,0001
Genótipo	8	81395552	10174444	9,57	<0,0001
Interação G×E	48	85337847	1777872	1,67	0,01
Erro	112	119108361	1063468	-	-

Com a interação G×E sendo significativa ($F = 1,67$ e valor $p = 0,01$), é possível realizar a decomposição da interação e realizar a análise AMMI.

A decomposição da interação G×E e a contribuição de cada IPCA pode ser observada na Tabela 3.6.

Tabela 3.6: Decomposição da interação G×E utilizando o método de Gollob

FV	GL	SQ	QM	F	Valor p
Interação G×E	48	85337847	1777872	1,67	0,01
IPCA1	13	52782729	4060210.0	3,82	0,0000
IPCA2	11	16460959	1496450.8	1,41	0,1781
IPCA3	9	8219811	913312.4	0,86	0,5631
IPCA4	7	3922239	560319.9	0,53	0,8102
IPCA5	5	2418264	483652.7	0,45	0,8125
IPCA6	3	1533844	511281.5	0,48	0,6968

Observando a decomposição da interação G×E, é possível observar que apenas a primeira decomposição foi significativa, porém para a construção do gráfico biplot foram utilizados os eixos IPCA1 e IPCA2, que representam 81,1% da variação da interação G×E. O biplot é construído com o IPCA1 no eixo X e IPCA2 no eixo Y, como segue na Figura 3.1.

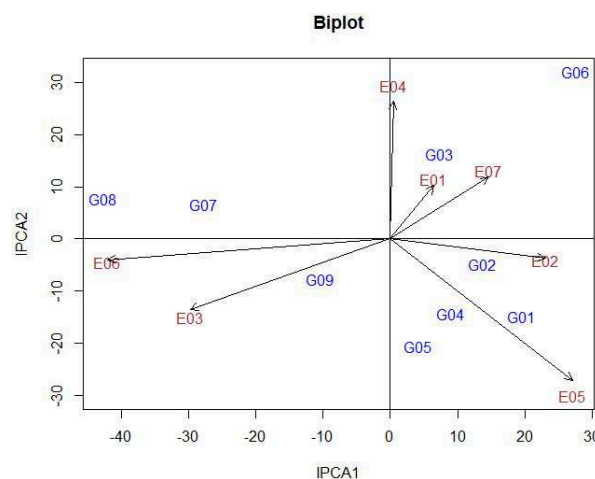


Figura 3.1: Biplot dos genótipos e ambientes

A interpretação do gráfico biplot pode ser feita utilizando o ângulo entre os genótipos e ambientes.

O cosseno do ângulo entre as variáveis expressa a magnitude da associação entre as mesmas, podendo essa associação ser positiva ou negativa. Quanto mais próximo de 1 ou -1, maior será a associação entre o genótipo e ambiente.

Assim, é possível observar que os ambientes E01, E04 e E07 teriam uma associação positiva com os genótipos G03 e G06, enquanto o ambiente E03 teria uma associação negativa com esses genótipos. Os genótipos G01, G02, G04 e G05 teriam uma associação positiva com os ambientes E02 e E05, e uma associação negativo nos ambientes E03 e E04. Os ambientes E03 e E06 teriam uma associação positiva apenas para os genótipos G07, G08 e G09.

3.2.1 REAMOSTRAGEM BOOTSTRAP

Utilizando a metodologia de reamostragem *Bootstrap*, os genótipos e ambientes foram reamostrados 400 vezes com a finalidade de obter uma estimativa mais precisa para os estimadores estatísticos e uma melhor qualidade no modelo.

A princípio, uma análise descritiva foi realizada para os genótipos em relação ao rendimento e pode ser observada na Tabela 3.7.

Tabela 3.7: Análise descritiva dos genótipos para a produtividade em $kg.ha^{-1}$

Genótipo	N	Min	Média	Mediana	Max	Desvio Padrão
G01	21	1794,21	7297,06	7958,43	11131,71	2294,39
G02	21	1789,79	7176,79	7598,57	11373,60	2379,45
G03	21	2175,17	6476,05	7055,90	9747,78	2088,03
G04	21	2737,11	5955,84	6128,39	8325,63	1785,30
G05	21	1888,12	5772,65	5732,95	9427,31	2017,75
G06	21	1739,59	7158,77	7846,67	12030,57	2644,65
G07	21	3054,04	6186,02	6478,82	9394,29	1591,65
G08	21	2738,43	6345,22	6400,99	8494,04	1293,86
G09	21	4683,68	7808,79	7666,67	11295,80	1612,53
TOTAL	189	1739,59	6686,35	6811,57	12030,57	2073,54

A partir da Tabela 3.7 é possível notar que o genótipo 09 é, em média, o mais produtivo dentre todos os genótipos, com uma produtividade de aproximadamente $7800kg.ha^{-1}$ e o genótipo 05 é o que possui a menor produtividade média, de $5772kg.ha^{-1}$.

Utilizando a reamostragem, foi gerada uma região de confiança de 95% para analisar a estabilidade dos genótipos, como pode ser observado na Figura 3.2.

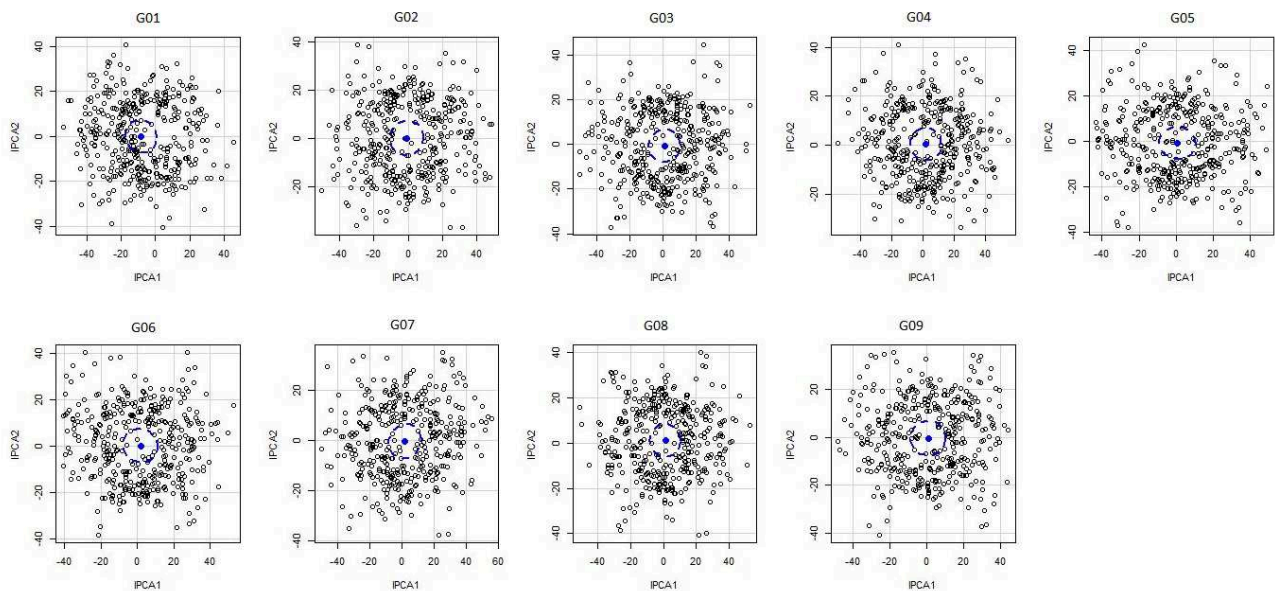


Figura 3.2: Região de confiança de 95% para os genótipos

Um genótipo é dito estável quando a região de confiança contém o vetor nulo $(0, 0)$. Analisando a Figura 3.2, é possível observar que apenas o genótipo G01 foi considerado instável nessa análise.

Assim, utilizando o CBE é possível classificar os outros genótipos quando ao nível de estabilidade e esse resultado pode ser observado na Tabela 3.8.

Tabela 3.8: Nível de estabilidade dos genótipos

Genótipo	Nível de estabilidade
G02	Baixa
G03	Baixa
G04	Baixa
G05	Baixa
G06	Baixa
G07	Baixa
G08	Baixa
G09	Baixa

Como pode ser observado, todos os genótipos estáveis foram classificados com baixa estabilidade, assim, deve-se ter cautela na indicação desses genótipos.

O mesmo procedimento feito para os genótipos será realizado para os ambientes. A análise descritiva para os ambientes em relação ao rendimento pode ser encontrada na Tabela 3.9.

Tabela 3.9: Análise descritiva dos ambientes para a produtividade em $kg.ha^{-1}$

Ambiente	N	Min	Média	Mediana	Max	Desvio Padrão
E01	27	5068,97	7376,69	7598,57	10577,01	1360,42
E02	27	5399,48	9095,73	9171,26	12030,57	1598,01
E03	27	1739,59	5004,80	5275,86	8941,61	1877,22
E04	27	5103,04	7207,27	7313,94	9242,36	1058,50
E05	27	4153,79	7221,86	7386,69	9468,62	1350,33
E06	27	1794,21	4073,29	4222,47	6180,30	1344,63
E07	27	4911,43	6824,83	6727,06	8606,35	1122,19
TOTAL	189	1739,59	6686,35	6811,57	12030,57	2073,54

Aqui pode-se observar que o ambiente cuja produtividade média foi a maior nesse conjunto de dados, foi o ambiente 02, com produtividade de mais de $9000kg.ha^{-1}$, enquanto o ambiente que teve a menor produtividade média foi o ambiente 06, com produtividade de $4073kg.ha^{-1}$.

Afim de analisar a estabilidade dos ambientes, foi gerada uma região de confiança de 95%, conforme pode ser observado na Figura 3.3.

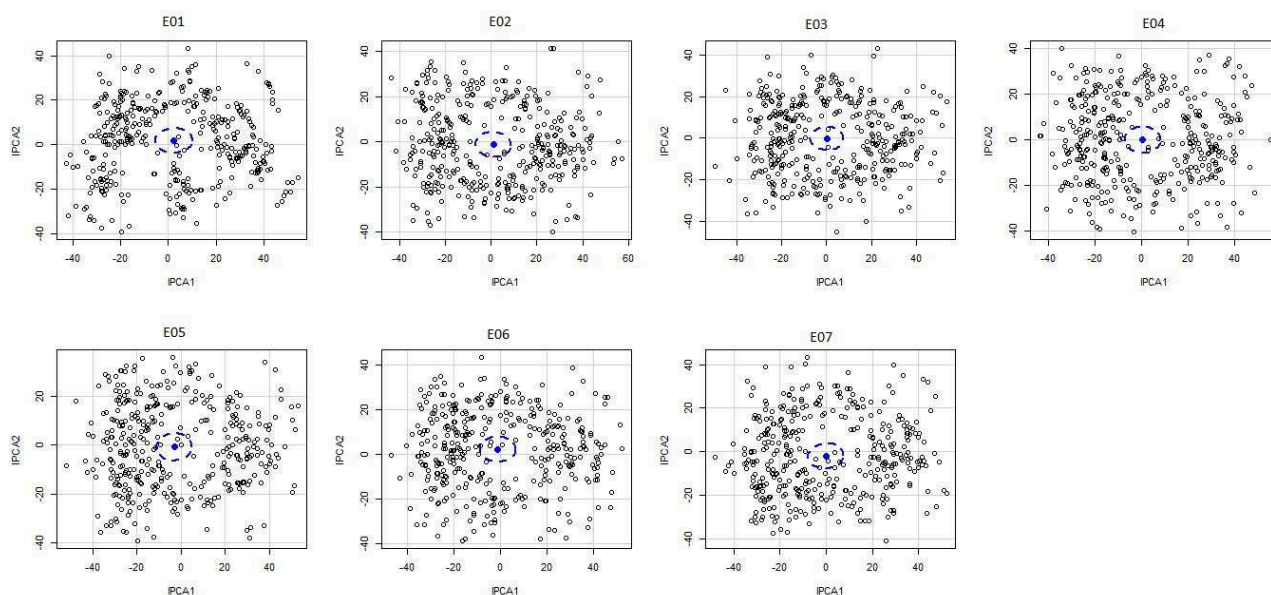


Figura 3.3: Região de confiança de 95% para os ambientes

Observando a Figura 3.3, é possível observar que todos os ambientes foram considerados estáveis, ou seja, a região de confiança de todos os genótipos continham o vetor nulo (vetor $(0,0)$). Assim, os ambientes podem ser classificados quanto ao nível de estabilidade e os resultados pode ser analisados na Tabela 3.10.

Tabela 3.10: Nível de estabilidade dos ambientes

Ambiente	Nível de estabilidade
E01	Baixa
E02	Baixa
E03	Baixa
E04	Baixa
E05	Baixa
E06	Baixa
E07	Baixa

Assim como nos genótipos, os ambientes também foram 100% classificados com estabilidade baixa, que também pede cautela com as indicações de quais ambientes podem ser bons para os genótipos.

De modo geral, os genótipos e ambientes apresentaram baixa estabilidade, ou seja, nenhum dos genótipos pode ser utilizado em todos os ambientes. Isso se deve ao fato de que esses materiais fazem parte de um programa de melhoramento e, portanto, era esperado que houvessem adaptações de genótipos a certos ambientes.

4. CONCLUSÕES

A análise de Cluster mostrou que os ambientes que, originalmente são definidos dentro das regiões geográficas no Brasil, não obtiveram a interação $G \times E$ homogênea dentro dos grupos e heterogênea entre os grupos, visto que os ambientes foram classificados em grupos diferente dos originais. Isso pode ser devido a diferentes características dentro das regiões, fazendo com que os ambientes não fossem homogêneos dentro das regiões.

A análise AMMI permitiu uma decomposição satisfatória da soma de quadrados da interação $G \times E$ ($SQ_G \times E$), em que 2 dos eixos foram utilizadas na criação do gráfico biplot, representando 81,1% da variação da interação e permitindo uma interpretação visual dos resultados da interação.

A reamostragem *bootstrap* fez-se necessária para ter uma maior precisão da informação. Essa reamostragem de 400 vezes para os genótipos e 400 vezes para os ambientes, permitiu uma identificação de uma região de confiança de 95% para a interação e, conseqüentemente, uma identificação da estabilidade dessas observações.

A partir dessa estabilidade, também foi possível identificar em qual dos 4 níveis de estabilidade cada uma das observações seria classificada, com uma precisão maior. Todos os ambientes e genótipos foram classificados com baixa estabilidade, o que requer certo cuidado ao fazer indicações desses genótipos e/ou ambientes.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] CHAVES, J. L.: *Interação de cultivares com ambientes*. Recursos genéticos e melhoramento - plantas, pp. 673–713, 2001.
- [2] CROSSA, J.: *Statistical Analyses of multilocation trials. Advances in Agronomy*. New York, 44:55–85, 1990.
- [3] DIAS, C. T. S. e. a.: *A metodologia AMMI: com aplicação ao melhoramento genético*. Piracicaba: ESALQ/USP, 2014.
- [4] DUARTE, J. B.; VENCOSKY, R.: *Interação genótipos x ambientes. Uma introdução à análise AMMI*. Sociedade Brasileira de Genética. Ribeirão Preto - SP, 1999.
- [5] EFRON, B.: *Bootstrap methods: another look at jackknife*. Annals of Statistics, 7:1–26, 1979.
- [6] FARIA, P. N.: *Utilização de técnicas multivariadas na análise da divergência genética via modelo AMMI com reamostragem "bootstrap"*. Tese de Doutorado, 2012.
- [7] Food e United Nations, A. O. of the: *Representante da FAO Brasil apresenta cenário da demanda por alimentos*, 2017. <http://www.fao.org/brasil/noticias/detail-events/en/c/901168/>, acessado em 02/06/2018.
- [8] GAUCH, H. G.: *Model selection and validation for yield trials with interaction*. Biometrics, Washington, 44:705–715, 1988.
- [9] GOLLOB, H. F.: *A statistical model which combines features of factor analytic and analysis of variance techniques*. Psychometrika, 33:73–115, 1968.
- [10] KHATTREE, R.; NAIK, D. N.: *Multivariate data reduction and discrimination with SAS software*. New York: BBU and John Wiley Sons, p. 574p, 2000.
- [11] LAVORANTI, O. J.: *Estabilidade e adaptabilidade fenotípica através da reamostragem "Bootstrap" no modelo AMMI*. Tese de Doutorado, 2003.
- [12] MAIA, M. C. C. e. a.: *AMMI-Bootstrap no estudo da interação genótipo por ambiente em soja*. Rev. Mat. Est., 24:7–24, 2006.
- [13] MALTHUS, T. R.: *Ensaio sobre o princípio da população*. 1798.

- [14] ZOBEL, R.W; WRIGHT, M. G. H.: *Statistical analysis of a yield trial*. Agronomy Journal, 80:388–393, 1988.