

Análise de agrupamento e método AMMI em divergência genética de trigo

Damallys de Assis Oliveira¹, Anderson Cristiano Neisse², Kuang Hongyu³

1 Introdução

Ensaio com várias combinações de locais e ambientes distintos, são conduzidos anualmente em várias partes do mundo por instituições de melhoramento de plantas. Os objetivos são: identificar cultivares superiores para uma região específica para serem cultivados em safras futuras nesses mesmos ambientes e avaliar sua estabilidade antes de seu lançamento comercial, compreender o comportamento ambiental da região específica e, em particular, determinar se esta região pode ser subdividida em diferentes locais.

Alguns métodos procuram identificar os genótipos com menor contribuição para a interação entre genótipos e ambientes ($G \times E$), chamados genótipos estáveis e que poderiam ser recomendados para toda a população de ambientes, desde que mostrem também um desempenho médio desejável. Numerosos métodos têm sido usados ao longo do tempo em busca do entendimento da interação $G \times E$.

A análise de agrupamento apresenta a finalidade de reunir, por algum critério de classificação, os genótipos em grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre os grupos, sendo adequada para identificar os genótipos divergentes e com maior probabilidade de sucesso nos cruzamentos (DUDA et. al, 2001; LIU et. al, 2013).

Dentre as novas análises estatísticas propostas para a interpretação da interação $G \times E$, o modelo AMMI (*Additive Main effects and Multiplicative Interaction*) vem ganhando grande visibilidade pelo maior número de interpretações disponíveis. Os parâmetros da interação $G \times E$ são estimados a partir da decomposição por valores singulares (DVS) da matriz de interação $G \times E$ (AMMI), os padrões de respostas de genótipos e ambientes podem ser visualizado graficamente usando *biplots*.

A combinação de agrupamentos na realização das análises AMMI consiste em técnicas que permitem obter uma solução que seja o consenso entre os dados empregados, de modo a fornecer soluções robustas e de melhor qualidade.

2 Metodologia

2.1 Modelo AMMI

O modelo AMMI permite um melhor detalhamento da somas de quadrados da interação ($SQ_{G \times E}$) trazendo consequentemente vantagens na análise desse feito, em que combina a análise de variância e a decomposição por valor singular em um único modelo, sob os pressupostos de que os efeitos principais (genótipo e ambiente) que são de natureza aditiva e interação.

O AMMI ganhou grande visibilidade nos últimos anos por ser útil para qualquer conjunto de dados provenientes de experimentos com dois fatores de classificação cruzada. E pelo fato de explicar num único gráfico os efeitos da interação $G \times E$ de forma eficiente, baseando-se na aproximação da DVS. A matriz GE é a matriz de interação entre os genótipos e os ambientes, (matriz de resíduo dos efeitos principais), em que cada elemento $(ge)_{ij}$ de **GE** são dados por (GAUCH, 2013):

$$(ge)_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

em que Y_{ij} é a média das repetições do genótipo i no ambiente j , com $i = 1, 2, \dots, g$ e $j = 1, 2, \dots, e$; $\bar{Y}_{i.}$ é a média do genótipo i ; $\bar{Y}_{.j}$ é a média do ambiente j e $\bar{Y}_{..}$ é a média geral do experimento.

¹ Graduada pelo Departamento de Estatística, UFMT. email: damallys@gmail.com.

² Mestrando em Estatística Aplicada e Biometria, UFV. email: a.neisse@gmail.com

³ Professor Adjunto do Departamento de Estatística, UFMT. email: prof.kuang@gmail.com.

Existem várias técnicas para atribuir os graus de liberdade a um modelo AMMI, um dos procedimentos usuais consiste em determinar os graus de liberdade associados à cada parcela da $SQ_{G \times E}$, ou seja, associada a λ_k^2 , relacionada a cada membro da família de modelos AMMI, obtém-se o quadrado médio (QM) correspondente a cada parcela (ou modelo), em seguida, é obtido um teste F avaliando-se a significância de cada componente em relação ao $QM_{\text{Erro médio}}$. Isso resulta num quadro de análise de variância semelhante ao tradicional, com desdobramento para fonte de variação da interação $G \times E$ (GAUCH, 2013; HONGYU *et al.*, 2015).

3 Resultados e Discussão

Considerando que a produtividade de trigo é uma variável quantitativa discreta, os dados originais da matriz de interação $G \times E$ foram submetidos à análise de agrupamento, para tornar a sua distribuição mais apropriada para a realização dos modelos AMMI e GGE *Biplot*.

Para a análise dos dados, utiliza-se inicialmente o procedimento hierárquico e após aglomeração não-hierárquica. Assim utilizou-se a distância de correlação da matriz obtida pela matriz de interação $G \times E$. A partir disso foi escolhido um medóide para cada grupo, para que pudesse representar o conjunto na análise. Os grupos formados estão apresentados nas Figuras 1 e 2, que representam os dois dendrogramas para genótipo e ambientes respectivamente.

Figura 1. Dendrograma obtido pelo emprego do medóide com base na distância das correlações dos 50 genótipos.

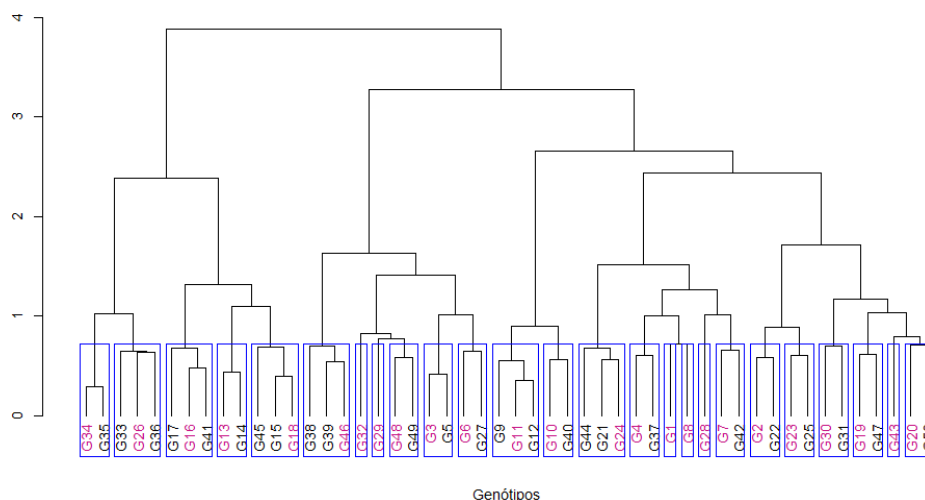
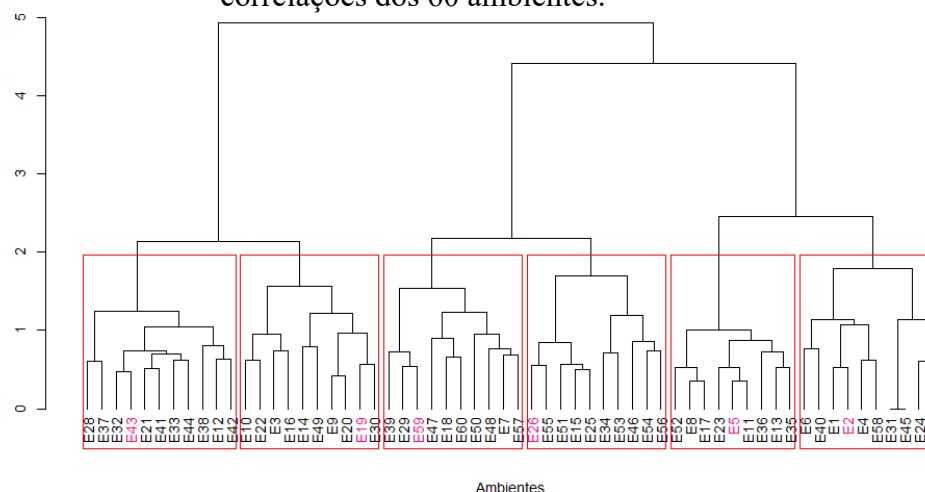


Figura 2. Dendrograma obtido pelo emprego do medóide com base na distância das correlações dos 60 ambientes.



Na Figura 1 os retângulos azuis representam os 25 aglomerados genótipos formados e também possui os genótipos de roxo que são os medóides de cada grupo de agrupamento. As caixas de vermelho na Figura 2, delimita os 6 agrupamentos ambientais escolhidos, os ambientes medóides são marcados de rosa. A escolha de grupos levou em conta a melhor interpretação visual e que pudessem manter ao mesmo tempo mais do que 70% da variabilidade nos dois componentes

principais, a fim de manter o máximo de genótipos e ambientes quanto possível ao ganhar vantagens na análise visual.

A Tabela 1 apresenta a análise de variância conjunta dos dados, bem como o desdobramento da interação $G \times E$, considerando 25 genótipos em 6 ambientes com 2 blocos (repetições), referente à produtividade de trigo (t/ha). Verifica-se, ao nível de 1% de significância ($p < 0,01$), que os efeitos de genótipos (G) e a interação $G \times E$ são significativos. O efeito ambiente (E) foi significativo ao nível de 5% de significância ($p < 0,05$).

Tabela 1. Análise de variação conjunta para um conjunto de dados com 25 genótipos e 6 ambientes em 2 blocos, e a decomposição das somas de quadrados da interação $G \times E$.

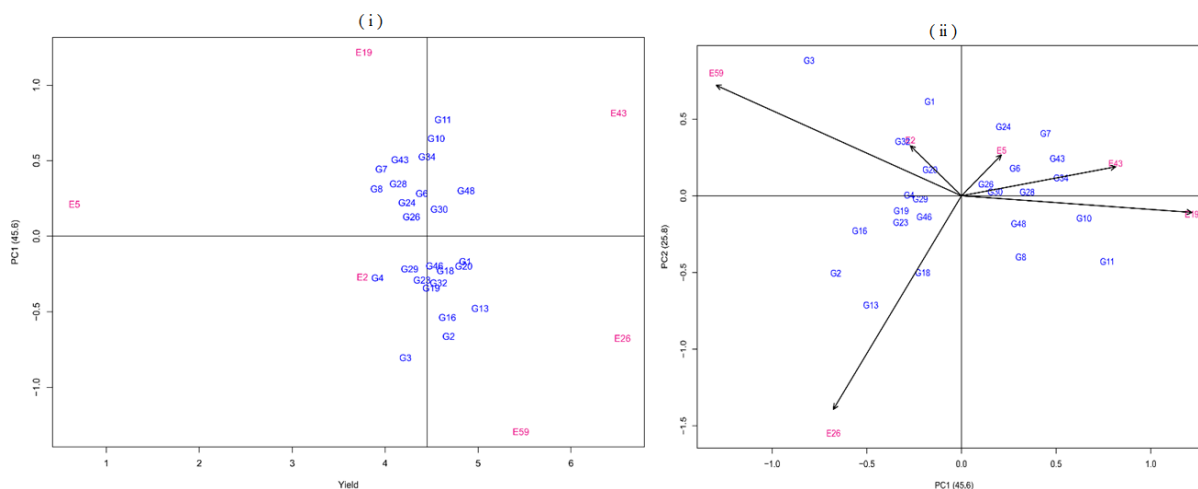
Fonte de variação	g.l	SQ	QM	F	valor-p
Bloco/Ambiente	6	6,99	1,164	5,717	0,0028
Ambiente (E)	5	1251,47	250,294	214,955	0,0158
Genótipo (G)	24	26,60	1,108	5,441	<0,0001
Interação	120	84,94	0,708	3,475	<0,0001
PC1	28	38,69	1,382	6,780	<0,0001
PC2	26	21,89	0,842	4,130	<0,0001
PC3	24	12,69	0,529	2,600	<0,0001
PC4	22	7,62	0,346	1,700	0,0345
PC5	20	4,05	0,202	0,990	0,4778
Resíduos	144	29,33	0,204	-	-
Total	299	1.399,33	252.569,51	-	-

As somas de quadrados dos efeitos genótipos (G), ambiente (E) e a interação ($G \times E$), corresponde a 89,88%, 1,91% e 6,10% respectivamente da soma de quadrados total. Esses resultados indicam que os genótipos apresentam comportamento diferenciado nos ambientes, justificando o estudo mais aprofundado sobre o comportamento dos genótipos para identificar suas magnitudes de interação com os ambientes. O coeficiente de variação (CV%) foi de 10,04% conferindo boa precisão ao experimento, já que o genótipo é altamente influenciado pelo ambiente.

O biplot AMMI1 (Figura 3 (i)) contém uma das principais variantes de genótipos e ambientes (eixo do gráfico) e uma das maiores proporções da interação entre o eixo e as estruturas (vertical do gráfico). Genótipos e estilos de trabalho são (sem uma contribuição para $SQ_{G \times E}$) que são os seguintes pontos de vista próximos de zero.

O biplot AMMI2 (Figura 3 (ii)) é utilizado para distinguir entre a média, efeitos principais ou de interação, ou seja, visualizado apenas para os efeitos multiplicativos da interação contidos nos dois primeiros PC's e dispersos em ambas as ordenadas do gráfico. Para os pesquisadores, isso corresponde à distinção fundamental e importante entre adaptações amplas e estreitas, que têm implicações bastante diferentes para os mega-ambientes.

Figura 3. *Biplots* da análise AMMI1 (Yield vs. PC1 (i)) e AMMI2 (PC1 vs. PC2 (ii)) para dados de produtividade de trigo (t/ha) com 25 genótipos (G) e 6 ambientes (E).



4 Conclusões

Todas as análises se mostraram eficientemente, pois foram capazes de explorar a variabilidade presentes em dados de MET devido à interação $G \times E$. Além disso, complementando um do outro, foi possível ganhar confiabilidade na análise. Os dados provenientes do 17^o Ensaio de Rendimento de Trigo Semi-Árido (SAWYT), do CIMMYT o AMMI explicou 71,4% de variabilidade nos dois primeiros componentes. Os genótipos G13 e o G2 são mais próximos da definição do "genótipo ideal".

Referências Bibliográficas

- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. International Journal of Computational Intelligence and Applications. IMPERIAL COLLEGE PRESS, v. 1, p. 335–339, 2001.
- GAUCH, H. G. A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Science*, v.53, n.5, p.1860-1869, 2013.
- HONGYU, K. Comparação entre os modelos os modelos ammi e gge biplot para os dados de ensaios multi-ambientais. *Revista Brasileira de Biometria.*, v.33, p. 139–155, 04 2015.
- LIU, Y.; ZHONGMOU, L.; XIONG, H.; GAO, X.; XU, J.; WU, Z. Understanding and enhancement of internal clustering validation measures. *IEEE transactions on cybernetics*, Institute of Electrical and Electronics Engineers, Inc., 345 E. 47 th St. NY NY 10017-2394 United States, v. 43, n. 3, p. 982–994, 2013.
- YAN, W.; HUNT, L. A; SHENG, Q.; SZLAVNICS, Z. Cultivar evaluation and mega-environment investigation based on the gge biplot. *Crop Science*, Crop Science Society of America, v. 40, n. 3, p. 597–605, 2000.
- YAN, W.; HOLLAND, J. B. A heritability-adjusted gge biplot for test environment evaluation. *Euphytica*, Springer, v. 171, n. 3, p. 355–369, 2010.

YAN, W. Gge biplot vs. ammi graphs for genotype-by-environment data analysis. Journal of the Indian Society of Agricultural Statistics, v. 65, n. 2, p. 181–193, 2011.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.