PROPOSTA DE APLICAÇÃO DA ANÁLISE DE VARIÂNCIA COM MODELO AUTOREGRESSIVO PARA ANÁLISE DE VARIEDADES DE TRIGO (*Triticum aestivum L.)*

DIOGO FRANCISCO ROSSONI[1], MARIELE VILELA BERNARDES PRADO[2], JAIR ROCHA DO PRADO[3]

RESUMO

Para a realização da ANOVA, os erros devem ser independentes e identicamente distribuídos segundo uma normal. Logo, caso isso não ocorra ou haja autocorrelação espacial entre os dados, a ANOVA é seriamente afetada. Assim, é importante o estudo e utilização de métodos de análise para experimentos em campo que levem em consideração a autocorrelação espacial. Com o conhecimento das posições relativas das amostragens (dados referenciados), a variabilidade espacial passa a ser utilizada como um fator positivo, identificando interações importantes nas conclusões experimentais.

Para esse estudo propõe-se uma metodologia que considera a dependência espacial nas análises: a análise de variância com modelo autoregressivo (ANOVA-AR). Comparando-se a ANOVA-AR com a ANOVA clássica, aplicados a dados da produção de variedades de trigo, constatou-se melhora na análise e na explicação das variações do experimento, através da redução do Quadrado Médio do Erro (QME), da redução da Soma Quadrada dos Blocos e da atribuição de 3,6% da variação ao parâmetro ρ.

Palavras-chaves: Dependência espacial, modelo autoregressivo, ANOVA-AR

INTRODUÇÃO

A análise de variância (ANOVA) é utilizada como forma de comparar as médias de tratamentos obtidas através de experimento. Para executarmos uma ANOVA, deve-se respeitar os princípios básicos da experimentação: repetição, aleatorização e controle local. Segundo Banzatto e Kronka (1992) o princípio de repetição consiste na repetição do experimento básico e tem por finalidade propiciar a obtenção de uma estimativa do erro experimental. O princípio de aleatorização tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer das unidades experimentais. É suposto que nenhum tratamento é favorecido por estar em um local com melhores condições de campo. Pode-se considerar também que a aleatorização é usada para satisfazer a condição de independência das observações. O controle local tem o intuito de dividir um ambiente heterogêneo em sub-ambientes homogêneos, tornando o delineamento experimental mais eficiente, pela redução do erro experimental.

Para a realização da ANOVA, os erros devem ser independentes e identicamente distribuídos segundo uma normal. Logo, caso isso não ocorra ou haja autocorrelação espacial entre os dados, a ANOVA é seriamente afetada. Assim, é importante o estudo e utilização de métodos de análise para experimentos em campo que levem em consideração a autocorrelação espacial.

[1] Mestrando em Estatística e Experimentação Agropecuária, DEX/UFLA diogo.rossoni@gmail.com
[2] Mestranda em Estatística e Experimentação Agropecuária, DEX/UFLA mari_bernardes@yahoo.com.br
[3] Mestrando em Estatística e Experimentação Agropecuária, DEX/UFLA jairmat1@yahoo.com.br

MATERIAL E MÉTODOS

Os dados foram coletados no Centro de Pesquisa Agrícola da Universidade do Estado de Montana-USA e são referentes a produção de trigo (*Triticum aestivum L.)*, com um total de 102 observações georeferenciadas, com 34 variedades de trigo organizadas em um DBC com 3 blocos.

A Análise de Variância com modelo autoregressivo (ANOVA_AR) utiliza o modelo AR, mais conhecido como SAR (espacial autoregressivo ou *spatial lag model*), que descreve a variação espacial no vetor resposta $Y_{n\times 1}$ e é definido por:

$$Y = \rho WY + X\beta + \varepsilon \qquad (0.1)$$

em que:

Y é um vetor $n\times 1$ de valores observados;
ρ é um parâmetro espacial autoregressivo;
W é uma matriz $n\times n$ com atribuições de peso da vizinhança espacial;
X é uma matriz $n\times p$ das observação das variáveis independentes;
β é um vetor $p\times 1$ dos parâmetros;
ε é um vetor $n\times 1$ dos erros inerentes a cada observação.

A matriz W é obtida através da multiplicação de duas outras matrizes D e C $(W = D\times C)$. A matriz C de dimensões $n\times n$ é binária, e descreve a vizinhança das parcelas experimentais. Cada elemento c_{ij} dessa matriz pode ser definido por:

$$c_{ij} = \begin{cases} 1 \text{ se } c_{ij} \text{ é adjacente a } c_{i+1,j}, \\ 1 \text{ se } c_{ij} \text{ é adjacente a } c_{i-1,j}, \\ 1 \text{ se } c_{ij} \text{ é adjacente a } c_{i,j+1}, \\ 1 \text{ se } c_{ij} \text{ é adjacente a } c_{i,j-1}, \\ 0 \text{ caso contrário.} \end{cases}$$

A matriz D é uma matriz diagonal com os elementos $1/k_i$, em que k_i é a soma dos valores da linha i da matriz C

Estimação do parâmetro ρ

A estimação do parâmetro ρ do modelo AR será pelo método de máximo verrosimilhança (MV). A MV é baseada no principio de selecionar valores para os parâmetros que maximizem a probabilidade de ter obtido os dados observados. Smirnov e Anselin (2001) propuseram diversas técnicas para a estimação de ρ, como decomposição de autovalores, método de decomposição matricial através da decomposição de Cholesky e aproximação de funções.

Neste estudo utilizou-se a decomposição de autovalores.

Decomposição de autovalores

A solução original da estimação por MV de um modelo autoregressivo espacial, originalmente proposta por Ord (1975), consiste em explorar a decomposição do Jacobiano $|I - \rho W|$ em termos dos autovalores ω_i (com $i = 1, 2, ..., N$) da matriz W, dada por:

$$\left|I-\rho W\right|=\prod_{i=1}^{n}\left(1-\rho\omega_i\right) \tag{0.2}$$

ou

$$\ln\left|I-\rho W\right|=\sum_{i=1}^{n}\ln\left(1-\rho\omega_i\right) \tag{0.3}$$

A partir da equação (0.2) ou (0.3) obtém-se um polinômio que não têm solução única e que deve ser resolvido iterativamente.

Análise de variância de modelo autoregressivo (ANOVA-AR)

A análise de variância de modelo autoregressivo (ANOVA-AR), foi proposta por Long (1996). A idéia básica consiste em transformar observações ditas autocorrelacionadas em observações não-correlacionadas. Para tanto, após obtermos a estimativa ρ deve-se proceder o ajuste dos dados observados através da equação (0.4)

$$Y_{adj}=Y-\left(\hat{\rho}WY-\hat{\rho}\beta_0\right) \tag{0.4}$$

em que:

Y é um vetor $n\times1$ de valores observados;

$\hat{\rho}$ é a estimativa do parâmetro espacial autoregressivo;

W é uma matriz $n\times n$ com atribuições de peso da vizinhança espacial;

β_0 é a média dos valores observados;

Y_{adj} é um vetor $n\times1$ de valores ajustados;

Com os valores de Y_{adj} obtém-se uma nova tabela de Análise de Variância. Esta tabela é utilizada como base para a construção da ANOVA-AR (Tabela 1).

Tabela 1: Análise de variância de modelo autoregressivo (ANOVA-AR)

FV	GL	SQ	QM	F
ρ	1	$SQ\rho$	-----	-----
Parâmetros	p	SQR_{adj}	$QMR_{adj}=SQR_{adj}/p$	$QMR_{adj}/QM\mathrm{E}_{adj}$
Resíduo	$n-p-2$	$SQ\mathrm{E}_{adj}$	$QM\mathrm{E}_{adj}$	
Total	$n-1$	SQT		

O valor da Soma Quadrada da fonte de variação ρ $(SQ\rho)$ é obtido através da diferença entre a Soma de Quadrado Total (SQT) da Análise de Variância dos dados não ajustados e a Soma de Quadrado Total Ajustada (SQT_{adj}) da Análise de Variância com dados ajustados pela equação (0.4), isto é:

$$SQ\rho=SQT-SQT_{adj}$$

O Quadrado Médio do Erro também deve ser ajustado por:

$$QME_{adj}=\frac{SQE_{adj}}{(n-p-2)}$$

em que:
n é o número de observações;
p é o número de parâmetros.

Todas as análises estatísticas foram feitas utilizando o software R.

Para a construção das matrizes *W, C e D,* utilizou-se uma vizinhança adjacente de distancia 1.

RESULTADOS E DISCUSSÃO

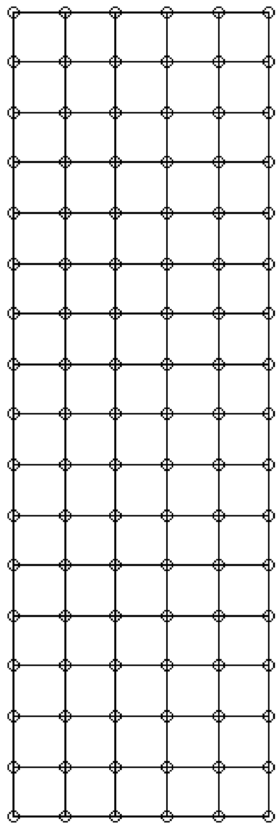

Figura 1: Vizinhança adjacente entre as amostras

Pode-se observar na Figura 1, que a vizinhança em diagonal não foi considerada devido a distância estipulada.

A matriz W de dimensões 102 x 102 será expressa de forma resumida.

$$W = \begin{pmatrix} 0 & 1/2 & 0 & \cdots & & 0 \\ 1/3 & 0 & 1/3 & & & 0 \\ \vdots & & & & & \\ & & & & 0 & 1/3 \\ & & & & 1/2 & 0 \end{pmatrix}_{102\times 102}$$

Após obter os autovalores da matriz *W* calculou-se o valor do coeficiente de autocorrelação ρ , tendo sido obtido o valor de *0,17545*.

Com este valor foi possível transformar os dados correlacionados em dados não correlacionados através da equação (0.4)

Tabela 2: Análise de variância clássica (ANOVA)

	GL	*SQ*	*QM*	*F-value*	*p-value*
bloco	2	312510	156255	1,4	0,2552
cultivar	33	15605390	472891	4,2	0,0003
Resíduos	66	7395029	112046		

Tabela 3: Análise de variância com modelo autoregressivo (ANOVA-AR)

	GL	*SQ*	*QM*	*F-value*	*p-value*
ρ	1	831795			
bloco	2	195808	97904	0,945763	0,393662
cultivar	33	15556620	471413	5	0,000001
Resíduos	65	6728706	103518,6		

CONCLUSÃO

Comparando a ANOVA com a ANOVA-AR pode-se peceber que houve uma pequena redução na variabilidade do experimento expressa através da redução do quadrado médio do erro (QME). A variabilidade que era atribuída ao fator bloco teve uma redução de 37%.

O parâmetro de autocorrelação, ρ , explicou 3,56% da variabilidade do experimento.

REFERÊNCIAL BIBLIOGRÁFICO

BANZATTO D.A.; KRONKA S.N. **Experimentação agrícola.** Jaboticabal, FUNEP, 1992. 246p.

LONG,D.S. **Spatial statistics for analysis of variance of agronomic field trials.** P 251-278. Practical Handbook of Spatial Statistics, 1996.

ORD, J.K.; **Estimation methods for models of spatial interaction.** J. Amer. Statist. Assoc. v.70, 1975, p. 120-126.

R DEVELOPMENT CORE TEAM (2010). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

SMIRNOV O.; ANSELIN L.; **Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach.** Computacional Statistics & Data Analysis, v. 35, 2001, p. 301-319.