

ANÁLISE DE CORRESPONDÊNCIA – UMA FERRAMENTA ÚTIL NA INTERPRETAÇÃO DE MAPAS DE PRODUTIVIDADE⁽¹⁾

J. R. P. CARVALHO⁽²⁾, S. R. VIEIRA⁽³⁾ & R. C. C. P. MORAN⁽⁴⁾

RESUMO

O objetivo do presente trabalho foi apresentar a Análise de Correspondência como um método estatístico multivariado que pode ajudar na interpretação de mapas de produtividade. Para tal, levantou-se a hipótese de a utilização de mapas de produtividade, para delimitar regiões homogêneas no campo, depender da estabilidade espacial e temporal desta variável, tornando-se preocupante seu uso indiscriminado como ferramenta de informação preditiva. Quatro mapas de produtividade foram usados neste trabalho, relativos a amostras coletadas de componentes de produção para arroz em 1994, milho em 1995, algodão em 1996 e milho em 1998, em uma malha com 110 pontos, espaçados de 10 x 10 m, da estação experimental do Instituto Agrônomo de Campinas em Votuporanga (SP). Os dados originais de produtividade foram categorizados e analisados por meio da Análise de Correspondência Simples e Múltipla. Concluiu-se que existe grande dissimilaridade entre as culturas, ou classes de culturas, medidas por meio da distância euclidiana no plano.

Termos de indexação: variabilidade espacial e temporal, agricultura de precisão, análise exploratória, dados categorizados.

⁽¹⁾ Recebido para publicação em setembro de 2000 e aprovado em dezembro de 2001.

⁽²⁾ Pesquisador da Embrapa Informática Agropecuária. Universidade Estadual de Campinas – UNICAMP, Barão Geraldo - Br. Caixa Postal 6041, CEP 13083-970 Campinas (SP). E-mail: jrui@cnptia.embrapa.br

⁽³⁾ Pesquisador do Centro de Solos e Recursos Agroambientais, Instituto Agrônomo – IAC. Caixa Postal 28, CEP 13001-970 Campinas (SP). E-mail: sidney@cec.iac.br

⁽⁴⁾ Professora do Instituto de Matemática, Estatística e Computação Científica, UNICAMP. E-mail: regina@ime.unicamp.br

SUMMARY: *ANALYSIS OF CORRESPONDENCE – A USEFUL TOOL TO INTERPRET YIELD MAPPING*

The objective of this work was to present the Analysis of Correspondence as a multivariate statistical method that can help with yield mapping interpretation. There is a concern about its indiscriminated use as a tool of predictive information, without taking into account the spatial and temporal variability of the yield. The hypothesis that the use of yield mapping is strictly related to the variability of the components was studied. Four yield maps were used corresponding to collected yield samples for rice in 1994, corn in 1995, cotton in 1996 and corn in 1998, in a grid of 110 points, 10 x 10 m spaced, at the experimental station of the Instituto Agrônomo, at Votuporanga (SP), Brazil. The original yield data were categorised and analyzed using Simple and Multiple Correspondence Analysis. A great dissimilarity between the cultures or classes of cultures was found as measured by the Euclidean Distance in the plan.

Index terms: spatial and temporal variability, precision agriculture, exploratory analysis, categorical data.

INTRODUÇÃO

A agricultura tradicional, influenciada pelo processo de globalização da economia, vem sendo desafiada na obtenção de níveis de produtividade que sejam compatíveis internacionalmente. Entretanto, as medidas que vêm sendo adotadas sugerem a utilização de áreas de grande extensão, as quais são, geralmente, consideradas homogêneas. Neste sentido, utiliza-se o conceito da necessidade média para aplicação de insumos (Capelli, 1999), ou seja, os mesmos níveis de fertilizantes, defensivos, água, etc. – são aplicados em toda a área, sem considerar as suas necessidades específicas e localizadas, resultando em uma produtividade não-uniforme para a lavoura, além do fato de acarretar problemas ambientais e ecológicos.

As tecnologias e as práticas emergentes da agricultura de precisão desafiam o uso da média como valor representativo de uma situação. O manejo da variabilidade é o principal foco de sua atenção. Duas formas principais de variabilidade são imediatamente detectadas: espacial e temporal. A variação espacial é a variação vista no campo, alterada pelos diferentes fatores que a compõem. A variação temporal é a variação que ocorre quando se comparam mapas de produtividade de um ano para outro.

O uso de GPS (Global Positioning Systems), combinado com dados de produção agrícola, permite elaborar mapas de produtividade. Esses mapas constituem uma ferramenta de diagnóstico muito importante para a agricultura de precisão, pois promovem o uso mais eficiente das informações, possibilitando, desta maneira, a administração da variabilidade no campo (Emmott et al., 1997).

A comparação de mapas de colheita para diferentes anos é uma etapa muito importante no

processo de identificação de zonas de manejo. Todavia, o que determina similaridade ou dissimilaridade entre esses mapas permanece subjetivo, enquanto não se tiver um critério para tal. Seriam algumas poucas áreas que aparecem moldadas semelhantemente no mapa? Seriam fatores como seleção de cor, número de classes e pontos de ruptura temáticos? Seriam a presença, ausência ou controle da variabilidade fatores de influência em similaridade?

O objetivo deste trabalho foi apresentar a análise de correspondência como um método estatístico multivariado que pode ajudar na interpretação de mapas de produtividade. Com isto, pretendeu-se estudar a hipótese de a utilização de mapas de produtividade para identificação de zonas de manejo depender da variabilidade de rendimento no ano e entre anos, tornando-se, desta maneira, preocupante seu uso indiscriminado, sem considerar as variações espaciais e temporais inerentes.

MATERIAL E MÉTODOS

Na área experimental do Instituto Agrônomo – IAC, localizada no Núcleo de Agronomia da Noroeste em Votuporanga (SP), em Latossolo Vermelho-Amarelo arenoso, foram coletadas amostras de componentes de produtividade para arroz em 1994 – A4, milho em 1995 – M5, algodão em 1996 – A6 e milho em 1998 – M8, em uma malha com 110 pontos, espaçados de 10 x 10 m (Figura 1). Em cada ponto, foram coletadas as partes aéreas em 5 m², onde também foi medida a produtividade.

Os mapas de produtividade, ou mapa de isolinhas, para as culturas dos quatro anos foram obtidos pela interpolação dos valores em qualquer posição no

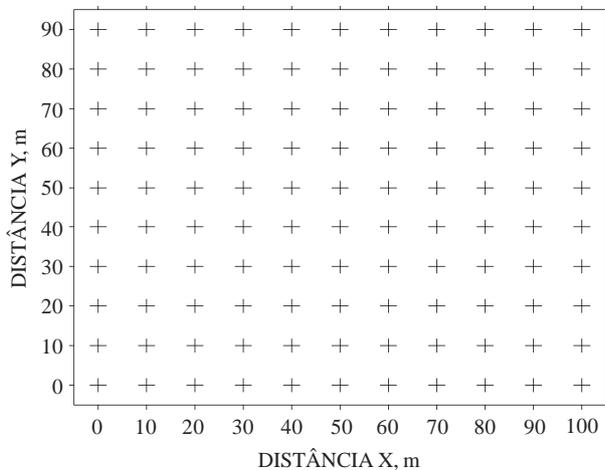


Figura 1. Esquema de amostragem.

campo de estudo, sem tendência e com variância mínima, por meio do método de interpolação chamado krigagem. Segundo Vieira et al. (1983), a precisão da localização das isolinhas entre dois pontos é extremamente dependente da densidade de pontos por área e, conseqüentemente, da distância entre os pontos.

O método estatístico usualmente utilizado para avaliar diferenças de médias de duas amostras de dados é o teste t de Student. Entretanto, as condições para que este teste seja apropriadamente aplicado são de que as amostras devem ser independentes e normalmente distribuídas. O problema é que estas condições raramente se aplicam em mapas de produtividade, considerando a dependência espacial e temporal das observações.

Outra forma de efetuar a comparação de mapas é por meio do índice KAPPA (K) de concordância. Esta medida estatística foi introduzida por psicólogos e adaptada como medida de concordância para mapas, por Congalton & Mead (1983). Este índice testa se a associação entre mapas é estatisticamente significativa ou casual. Ele também permite a comparação de dados com estruturas definidas (Rosenfield & Fitzpatrick-Lins, 1986). O índice KAPPA varia de 0 a 1 em que 0 indica que os resultados acontecem totalmente ao acaso e 1 indica concordância perfeita e é definido por:

$$K = \left\{ n \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i}) \right\} / \left\{ n^2 - \sum_{i=1}^r (x_{i+} * x_{+i}) \right\}$$

em que r é o número de linhas em uma tabela de classificação cruzada; x_{ii} o número de combinações na diagonal; x_{i+} o total de observações na linha i; x_{+i} o total de observações na coluna i, e n o número total de observações.

Em resumo, existem diversas maneiras de tratar o problema em questão. Há também os métodos de classificação cruzada, coeficiente de similaridade de

Qui-Quadrado e Cramer (IDRISI, 1999), dentre outros. Entretanto, a maioria deles depende de condições preestabelecidas que, normalmente, não se verificam para dados temporais e espacialmente dependentes.

A Análise de Correspondência é uma técnica multivariada para análise exploratória de dados categorizados. A preocupação com a análise destes tipos de dados já vem desde o começo do século. Contudo, esta técnica difundiu-se somente na década de 60, quando Jean-Paul Benzécri definiu um método mostrando suas propriedades algébricas e geométricas, denominado "Analyse Factorielle des Correspondences" (Benzécri, 1992). Ela converte uma matriz de dados não-negativos em um particular tipo de gráfico que exhibe as linhas e colunas da matriz como pontos de um espaço vetorial de dimensão menor que a original, de maneira que as relações entre as linhas, entre as colunas e entrelinhas e colunas possam ser interpretadas.

Exatamente por estar interessada em estudar a correspondência entre variáveis, é que esta técnica recebeu o nome de Análise de Correspondência. Sua geometria e álgebra fazem com que pertença a uma família de técnicas de disposição gráfica que são baseadas em aproximação de uma matriz por outra de posto menor, por meio da decomposição em valores singulares. Ou seja, o objetivo desta análise é achar um subespaço que melhor ajuste o conjunto (nuvem) de pontos no espaço euclidiano. Este ajuste é feito pelo método de quadrado mínimo ponderado onde a distância euclidiana generalizada (ponderada) é utilizada em um sistema de massas pontuais (Greenacre & Hastie, 1987).

Um dos pontos relevantes desta técnica é que não há exigência de normalidade para a resposta estudada. Em conseqüência, os testes estatísticos inferenciais não são aqui utilizados, estando a solução sugerida pela distribuição gráfica de seus resultados, o que constitui a base da escola francesa para as análises multivariadas (Benzécri, 1992).

A Análise de Correspondência Simples parte de uma matriz de dados representados por uma tabela de contingência. O desenvolvimento do algoritmo e a sua geometria fornecem as regras básicas para a interpretação. O algoritmo adaptado de Greenacre (1984) e Barioni Jr. (1995) segue os seguintes passos:

Seja N uma tabela de contingência com I categorias de linhas e J categorias de colunas, com elementos não negativos, de modo que a soma de cada linha ou coluna não seja nula. Seja n_{ij} o número de freqüências observadas pela interseção da i-ésima categoria da variável A com a j-ésima categoria da variável B, logo $N = [n_{ij}]_{I \times J}$ é a matriz de freqüências absolutas. A matriz N pode transformar-se em uma matriz de freqüências relativas expressa por: $P = (1/n) N$, em que P denomina-se matriz de correspondência.

Em função da matriz de correspondência P, define-se o i-ésimo perfil linha como o vetor:

$a_i = (p_{i1}/p_i, p_{i2}/p_i, \dots, p_{ij}/p_i)^T$, representando cada vetor a_i ($i=1, \dots, I$) uma distribuição multinomial, condicionada à i -ésima categoria da variável A. Conseqüentemente, a matriz do perfil linha é definida por: $R = D_r^{-1} P$ e a matriz do perfil coluna por: $C = D_c^{-1} P'$, sendo D a matriz diagonal de frequência relativa a cada linha ou a cada coluna. Cada perfil (linha ou coluna) é afetado por um peso p_i para $i = 1, \dots, I$ ou p_j para $j = 1, \dots, J$, proporcional aos respectivos totais de linha ou de coluna dos dados originais. Estes pesos são denominados massas, sendo $r_i = p_i = n_i/n$ para $i = 1, \dots, I$ é a massa da i -ésima linha e $c_j = p_j = n_j/n$ para $j = 1, \dots, J$ é a massa da j -ésima coluna.

O centróide da linha, ou coluna, de uma tabela de contingência indica geometricamente a posição média dos perfis (linha ou coluna), como se fosse o centro de gravidade ou ponto de equilíbrio da matriz de dados. Como cada perfil (linha ou coluna) está associado a uma massa, define-se como centróide a média ponderada dos perfis (linha ou coluna), ou seja, o centróide do perfil linha $Ce_{(r)} = \sum_i r_i a_i = R'r = c$ e o centróide do perfil coluna $Ce_{(c)} = \sum_j c_j b_j = C'c = r$.

Cada perfil linha a_i define um ponto no espaço das linhas R^J ; logo, neste espaço, acomoda-se um conjunto de pontos representando os I perfis linha. O fato de os elementos de cada perfil linha somarem 1 indica que a nuvem está restrita a uma região, no subespaço de dimensão $J-1$ do espaço das colunas, conhecida como simplexo. A interpretação do espaço do perfil coluna é feita de forma análoga. O formato do conjunto de pontos, ou o aparecimento de pontos isolados, pode indicar a existência ou não de associações entre pontos. Essas associações são medidas através das distâncias entre os pontos, utilizando a distância Qui-Quadrado (χ^2) que é uma distância euclidiana ponderada entre os perfis linha e o centróide Ce , sendo definida por: $D^2\{a_i, Ce\} = (R - Ce'1)^T D_{Ce}^{-1} (R - Ce'1)$, em que D_{Ce}^{-1} é a matriz de ponderação com diagonal representada pelo inverso das coordenadas do centróide. Se a métrica $RD_{Ce}^{-1/2}$ com suas respectivas massas for aplicada, obtém-se a distância euclidiana clássica.

De acordo com a natureza dos dados por analisar, o conjunto de pontos terá maior ou menor dispersão em relação ao centróide. A dispersão pode ser calculada como a média dos desvios dos vários pontos ao centróide. O quantificador normalmente usado da dispersão dos pontos dessa nuvem no espaço R^J é chamado de inércia total do perfil linha.

Na interpretação de mapas de produtividade, além de estar interessado na similaridade dos mapas, é interessante o estudo das similaridades de classes ou áreas dentro dos mapas. Essas diferenças entre classes ou áreas são observáveis pela investigação das associações existentes entre as categorias dentro de cada classe e, ou, entre as categorias de diferentes classes. Ela segue o mesmo objetivo da Análise Discriminante Canônica de

Fisher, ou seja, o de maximizar a variabilidade entre classes, representada na Análise de Correspondência pela inércia entre classes.

O tratamento matemático para a execução da Análise de Correspondência em classes segue o mesmo princípio para a Análise de Correspondência Múltipla. Enquanto na Análise de Correspondência Simples a tabela de contingência é usada, na Análise de Correspondência Múltipla, usa-se a tabela de incidência ou tabela Burt. Na Análise de Correspondência Múltipla, todas as associações entre pares de variáveis são analisadas, bem como a associação entre uma variável e ela mesma. A inclusão das associações de uma variável com ela mesma, o que não acontece na Análise de Correspondência Simples, é que permite ser a Análise de Correspondência Múltipla eficiente em sua parte computacional.

Muitas técnicas multivariadas baseiam-se nas análises na matriz de dados originais ou em alguma forma transformada destes dados. Na maioria das vezes, é mais apropriado escolher a forma transformada, para que valores de cada variável não dependam de possíveis unidades arbitrárias de medida.

Para tornar as variáveis de produtividade comparáveis, as 110 amostras de cada uma das quatro variáveis - A4, M5, A6 e M8 foram classificadas em cinco grupos de aproximadamente igual tamanho, de conformidade com seus valores de produtividade, tendo sido atribuídos a elas números correspondentes às suas posições na classificação conjunta do grupo. Os grupos foram definidos pelo módulo Proc Rank do software SAS (SAS, 1998). Assim, para a variável A4, por exemplo, a menor produtividade pertencerá ao grupo zero e a maior ao grupo quatro, respeitando suas posições originais.

Para estudar se existe similaridade entre classes de cada variável, a matriz de dados antes composta de quatro variáveis (A4, M5, A6, M8) foi transformada em 20 variáveis, e, para cada variável de produtividade, cinco variáveis foram criadas, correspondentes aos grupos 0,1,2,3 e 4. Assim sendo, as novas variáveis foram definidas, como se segue:

A40, A41, A42, A43, A44 \Rightarrow Produtividade de arroz de 1994 nas classes 0,1,2,3,4;

M50, M51, M52, M53, M54 \Rightarrow Produtividade de milho de 1995 nas classes 0,1,2,3,4;

A60, A61, A62, A63, A64 \Rightarrow Produtividade de algodão de 1996 nas classes 0,1,2,3,4;

M80, M81, M82, M83, M84 \Rightarrow Produtividade de milho de 1998 nas classes 0,1,2,3,4.

A matriz de dados para aplicação da Análise de Correspondência tem a dimensão de 20 colunas por 110 linhas e é composta de 0 e 1's, denominada matriz de incidência, indicando presença ou ausência de cada nova variável criada para cada amostra coletada.

RESULTADOS E DISCUSSÃO

Os mapas de produtividades para as culturas de arroz, milho, algodão e milho, para os anos de 1994, 1995, 1996 e 1998, estão apresentados nas figuras 2, 3, 4 e 5, respectivamente. Os mesmos grupos anteriormente obtidos foram usados para subdividir os dados de produtividade. Apesar de ser essa produtividade medida na mesma área experimental e abranger culturas e anos diferentes, observaram-se grandes diferenças numéricas entre os diversos grupos para os quatro mapas, indicando que as variabilidades, espacial e temporal, exerceram influência nos mapas de produtividade.

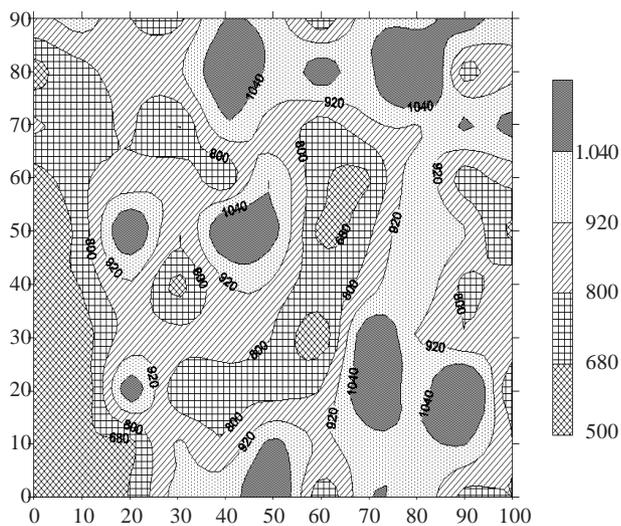


Figura 2. Mapa de produtividade de arroz (A4) em kg ha^{-1} - Votuporanga, 1994.

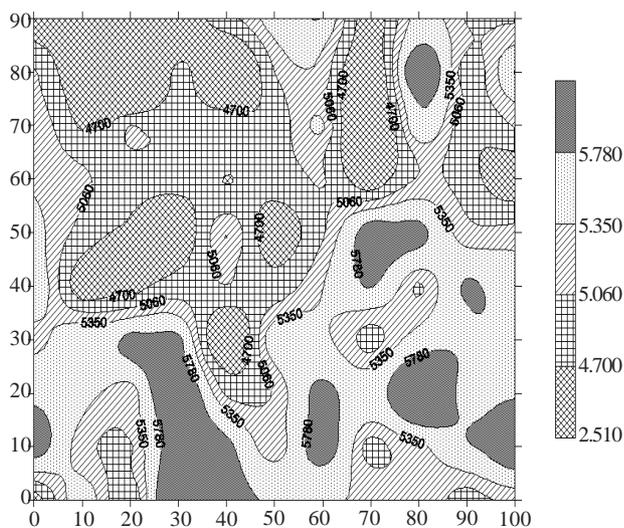


Figura 3. Mapa de produtividade de milho (M5) em kg ha^{-1} - Votuporanga, 1995.

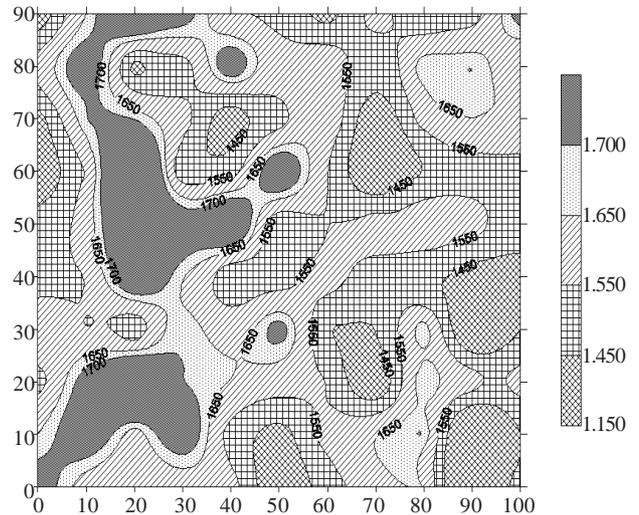


Figura 4. Mapa de produtividade de algodão (A6) em kg ha^{-1} - Votuporanga, 1996.

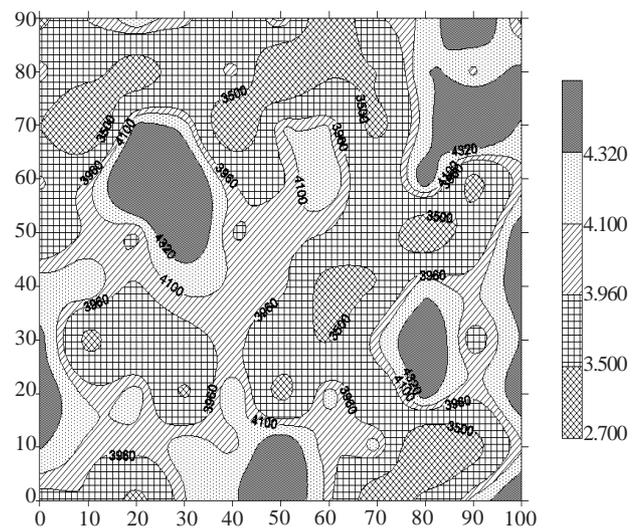


Figura 5. Mapa de produtividade de milho (M8) em kg ha^{-1} - Votuporanga, 1998.

Para verificar a similaridade entre os mapas de produtividade, utilizou-se a Análise de Correspondência Simples por meio do módulo Proc Corresp do software SAS (SAS, 1998) (Quadros 1 e 2).

Para analisar um conjunto de dados e utilizar a Análise de Correspondência como instrumento, foram usados gráficos para visualizar as relações entre as variáveis. Ao representar um conjunto de pontos, não foram observadas as freqüências absolutas de cada linha, mas, sim, o perfil de uma linha ao longo das categorias das colunas. Ao introduzir os perfis, padronizaram-se os pontos, reduzindo, assim, a informação à proporção de indivíduos em cada linha.

Quadro 1. Decomposição da Inércia e do Qui-Quadrado. Análise de Correspondência Simples para as variáveis de produtividade

Valor singular	Inércia principal	Qui-quadrado	%	7	14	21	28	35
				-----+	-----+	-----+	-----+	-----+
0,378	0,143	127,294	36,92	*****				
0,371	0,138	122,725	35,59	*****				
0,326	0,106	94,780	27,49	*****				
Total	0,388	344,799		Graus de liberdade = 327				

Quadro 2. Resumo de estatísticas para as variáveis de produtividade. Análise de correspondência simples

Variável	Eixo1	Eixo2	C1	C2	Qr1	Qr2
A4	1,565	0,656	0,602	0,105	0,815	0,138
M5	-0,999	1,281	0,246	0,405	0,346	0,548
A6	0,201	-1,216	0,010	0,375	0,016	0,565
M8	-0,745	-0,666	0,140	0,112	0,228	0,176

Eixo 1 e Eixo 2 referem-se, respectivamente, à decomposição da inércia para o primeiro e segundo eixo; C1 e C2 referem-se, respectivamente, aos pontos que melhor explicam a inércia para o primeiro e segundo eixo; Qr1 e Qr2 referem-se, respectivamente, à qualidade de representação dos pontos para o primeiro e segundo eixo.

Para obter os eixos, observaram-se os percentuais acumulados das inércias principais e, com base na percentagem explicada pela inércia, calculou-se o número de eixos necessários para a representação dos pontos.

Como pode ser visto no quadro 1, o total da estatística de Qui-Quadrado é 344,79, a qual é uma medida de associação entre as linhas e colunas nas três dimensões da tabela de dados transformada, obtida pela decomposição dos valores singulares. Cerca de 37 % da estatística de Qui-Quadrado e Inércia é explicada pelo primeiro eixo vetorial, enquanto cerca de 36 % é explicada pelo segundo eixo vetorial.

Neste trabalho, os pontos foram representados em dois eixos, indicando ser a associação entre linhas e colunas bidimensional, com 72 % da inércia explicada. Como a menor dimensão do quadro 1 foi a terceira, não houve perda de informação quando somente duas dimensões foram consideradas.

Na interpretação dos eixos, verificou-se que a inércia dos pontos decompôs-se em cada eixo. São os componentes da inércia que formam os coeficientes que indicam o quanto cada ponto contribuiu para a determinação da direção dos eixos e o quanto cada ponto está representado em cada eixo indicado pelos coeficientes c1 e c2, como pode ser visto no quadro 2.

No mesmo quadro, a qualidade de sua representação, indicada pelos coeficientes qr1 e qr2, também pode ser vista. A qualidade da representação é o coeficiente que dá o quanto cada ponto está, ou não, bem representado no subespaço considerado. Os resultados dos quadros 1 e 2 possibilitam a geração e a interpretação do gráfico bidimensional – Produtividade, para as culturas (Figura 6), definido pelas coordenadas dos dois primeiros eixos vetoriais.

Cada ponto representa o perfil de produtividade de cada uma das culturas ao longo das 110 amostras. Assim sendo, M5 representa o perfil da produtividade do milho em 95 ao longo das 110 amostras, A4 o perfil de arroz para 94, M8 o perfil de milho para 98 e A6 o perfil de algodão para 1996, respectivamente. Como a Análise de Correspondência aloca os pontos em um espaço euclidiano, as projeções dos pontos sobre os eixos permitem tirar conclusões sobre suas similaridades ou dissimilaridades.

Analisando a figura 6 e o quadro 2, percebe-se que a variável A4 foi a que mais contribuiu para o primeiro eixo vetorial com 60 % da inércia explicada. A variável M8 contribuiu com 14 % para este mesmo eixo. A distância de suas projeções sobre os eixos mostra grande dissimilaridade entre elas. O primeiro eixo vetorial revela um claro contraste entre as culturas de arroz e algodão com o milho.

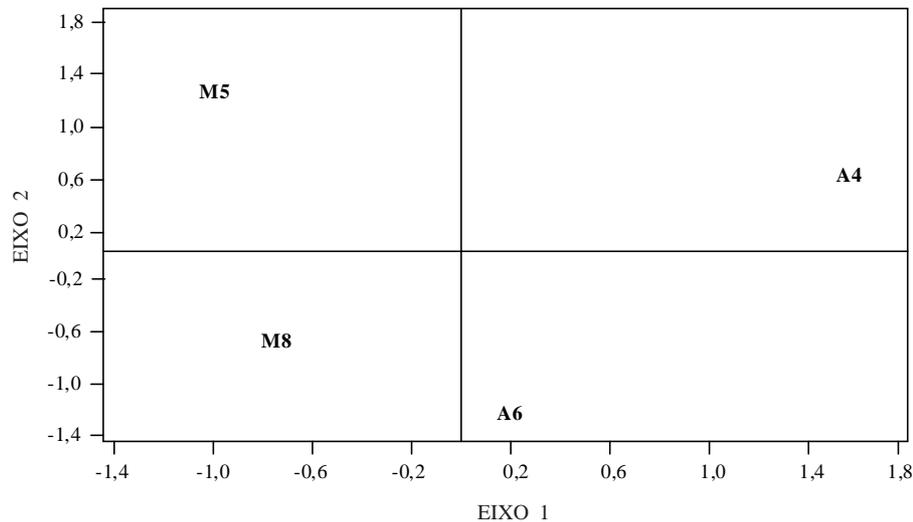


Figura 6. Produtividade para as culturas - Produtividade de arroz 94 (A4) / Produtividade de milho 95 (M5) / Produtividade de algodão 96 (A6) / Produtividade de milho 98 (M8).

No segundo eixo vetorial, a variável M5 contribuiu com 41 % e a A6 com 38 % para a inércia explicada. Suas distâncias apresentam-se bastante dissimilares. Neste eixo, estão relacionados os anos de 1994 e 1995 com 1996 e 1998, evidenciando, portanto, um contraste entre os primeiros anos com os últimos.

Os mapas de produtividade para arroz em 94 e milho em 98 mostraram-se completamente distintos. O mesmo aconteceu para milho 95 e algodão 96, ou seja, a variabilidade espacial e a temporal influenciaram a variável produtividade dentro e entre anos.

Assim, conclui-se que não são somente as exigências nutricionais das diferentes culturas podem alterar os mapas de produtividade, mas também as variações espaciais e temporais são causas de sua alteração, tornando-se um motivo maior de preocupação o seu uso indiscriminado.

A Análise de Correspondência simples, numa tabela de incidência, é semelhante à Análise de Correspondência Múltipla. Como era esperado, os valores das contribuições e qualidade de representação vistos no quadro 3 são bem inferiores aos da Análise de Correspondência Simples. Como o plano de amostragem continuou com 110 pontos, mas o número de variáveis aumentou de quatro para vinte, a capacidade de justificar a inércia para o mesmo número de eixos foi menor.

O quadro 3 apresenta as coordenadas das variáveis para os dois eixos e suas contribuições e qualidade de representação: as variáveis A40, M50, M51, M53, M54, A63, A64, M81, M83 e M84 mostraram maior contribuição para o primeiro eixo, enquanto as variáveis A41, A42, A43, A44, M52, A60, A61, A62, M80 e M82 para o segundo eixo.

Estes resultados possibilitam a geração e a interpretação de um gráfico bidimensional – Produtividade por Classes, conforme pode ser visto na figura 7, definido pelas coordenadas nos dois eixos ortogonais.

O ponto A60 plotado na figura 7, por exemplo, é a melhor representação da distribuição condicionada, no plano, a ter a menor produtividade na variável A6, portanto, refere-se somente às localizações com alguma dimensão na classe zero. O vetor A60, como perfil linha, descreve a distribuição condicional das classes de produtividade ao longo do tempo por cultura. Seguem-se conclusões similares para os demais pontos.

Interpretando a figura 7 com a ajuda do quadro 3, verifica-se que a distância das projeções das classes M50, M81, A64 e M54 no primeiro eixo são bastante próximas, indicando que elas são similares. O mesmo acontece para as classes M53, M84, A40, M83, M51 e, com menos rigor, para a A63. Entretanto, os dois grupos são bastantes dissimilares, pois suas projeções encontram-se em quadrantes opostos em relação ao eixo.

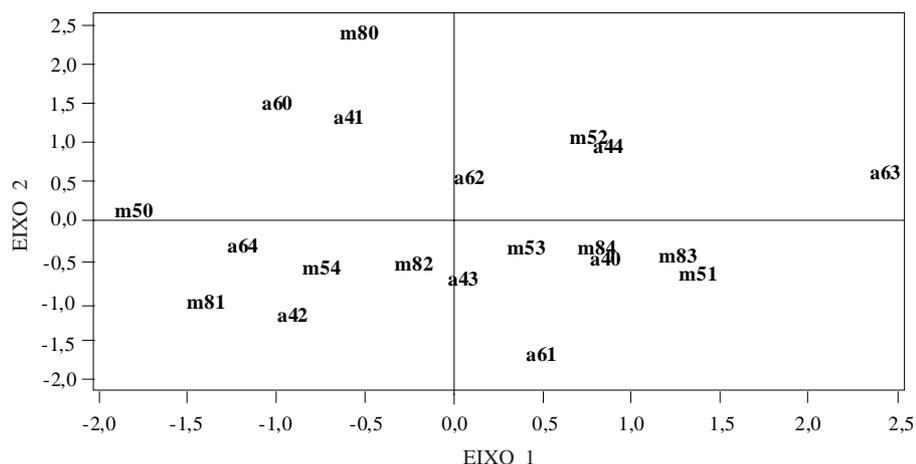
Em relação ao segundo eixo, as classes M80, A60, A41, M52, A44 e A62 apresentam pouca distância quando de sua projeção sobre o eixo, podendo ser consideradas similares. O mesmo acontece, em relação ao segundo eixo para as classes A61, A42, A43 e M82; entretanto, estes dois grupos são dissimilares. Ou seja, para cada eixo, existem grupos compostos por variáveis que são similares dentro dos grupos e dissimilares entre grupos.

No primeiro eixo, as classes 0 e 4 para a produtividade de milho 95 (M50 e M54) são similares entre si e similares com a classe 1 da produtividade

Quadro 3. Resumo das estatísticas para as variáveis de produtividade divididas por classes. Análise de correspondência múltipla

Classe	Eixo1	Eixo2	C1	C2	Qr1	Qr2
A40	0,855	-0,413	0,034	0,008	0,067	0,015
A41	-0,581	1,395	0,016	0,092	0,030	0,173
A42	-0,913	-1,140	0,055	0,085	0,115	0,176
A43	0,051	-0,662	0,001	0,016	0,001	0,030
A44	0,902	1,076	0,040	0,057	0,078	0,109
M50	-1,812	0,203	0,164	0,002	0,318	0,003
M51	1,382	-0,591	0,095	0,017	0,185	0,033
M52	0,779	1,128	0,031	0,066	0,062	0,127
M53	0,393	-0,269	0,007	0,003	0,014	0,006
M54	-0,759	-0,535	0,028	0,014	0,055	0,027
A60	-0,980	1,553	0,034	0,087	0,063	0,155
A61	0,504	-1,616	0,015	0,160	0,032	0,321
A62	0,099	0,594	0,001	0,023	0,001	0,047
A63	2,437	0,666	0,162	0,012	0,282	0,020
A64	-1,156	-0,247	0,079	0,003	0,160	0,007
M80	-0,529	2,451	0,012	0,273	0,024	0,505
M81	-1,364	-0,956	0,101	0,049	0,201	0,096
M82	-0,211	-0,487	0,002	0,010	0,003	0,019
M83	1,246	-0,375	0,081	0,007	0,159	0,014
M84	0,821	-0,335	0,035	0,005	0,069	0,011

Eixo 1 e Eixo 2 referem-se, respectivamente, à decomposição da inércia para o primeiro e segundo eixo; C1 e C2 referem-se, respectivamente, aos pontos que melhor explicam a inércia para o primeiro e segundo eixo; Qr1 e Qr2 referem-se, respectivamente, à qualidade de representação dos pontos para o primeiro e segundo eixo.

**Figura 7. Produtividade por classes - Produtividade de arroz 94 (A40-A44) / Produtividade de milho 95 (M50-M54) / Produtividade de algodão 96 (A60-A64) / Produtividade de milho 98 (M80-M84).**

de milho 98 (M81) e classe 4 de algodão 96 (A64). O segundo grupo neste eixo mostra similaridades entre as classes 3 e 4 para a produtividade de milho 98 (M83 e M84) e entre as classes 1 e 3 para produtividade de milho 95 (M51 e M53), bem como com as classes 0 de arroz 94 (A40) e 3 de algodão 96 (A63). Entretanto, ambos os grupos são bastante dissimilares.

No segundo eixo, existe clara similaridade para as classes 1 e 4 da produtividade de arroz 94 (A41 e A44) e classes 0 e 2 de algodão 96 (A60 e A62) e elas são similares com a classe 0 da produtividade de milho 98 (M80) e classe 2 de milho 95 (M52). No segundo grupo, as classes 2 e 3 da produtividade de arroz 94 (A42 e A43) são similares entre si e com as classes 1 de algodão 96 (A61) e 2 de milho 98 (M82).

Comparando os eixos, percebeu-se que, apesar de serem dissimilares, as produtividades de milho 95 (M5) e 98 (M8) têm a maioria de suas classes, contribuindo para o primeiro eixo, enquanto as classes de produtividade de arroz 94 e algodão 96 contribuem em maior número para o segundo eixo. Com isto verifica-se que as variações espaciais e temporais, apesar de influenciarem a produtividade de milho para 95 e 98, têm maior influência na produtividade de arroz 94 e algodão 96.

CONCLUSÕES

1. As exigências nutricionais das diferentes culturas alteraram os mapas de produtividade; entretanto, as variações espaciais e temporais foram também causas de sua alteração. Quando as variáveis produtividade de arroz 94, milho 95, algodão 96 e milho 98 foram comparadas, por meio da Análise de Correspondência Simples, a distância de suas projeções sobre os eixos mostrou-se bastante dissimilar, indicando que os mapas de produtividade sofreram influência de variações espaciais e temporais.

2. Quando as culturas foram categorizadas em grupos de produtividade, encontraram-se classes de culturas similares dentro de grupos e dissimilares entre grupos, mostrando que as variações espaciais e temporais influenciaram os mapas de produtividade, tanto no conjunto como quando subdivididos em classes.

3. Considerando a influência das variações espaciais e temporais, torna-se temeroso o uso indiscriminado de mapas de produtividade sem que haja um modelo de produtividade definido. Este modelo seria construído ao produzir mapas de produtividade em uma mesma área durante certo número de anos, caracterizando as possíveis variações.

LITERATURA CITADA

- BARIONI Jr., W. Análise de correspondência na identificação dos fatores de risco associados à diarreia e à performance de leitões na fase de lactação. Piracicaba, Escola Superior de Agricultura Luiz de Queiroz, 1995. 97p. (Tese de Mestrado)
- BENZÉCRI, J.P. Correspondence analysis handbook. New York, Marcel Decker, 1992. 665p. (Statistics: Textbooks and Monographs, 125)
- CAPELLI, N.L. Agricultura de precisão – novas tecnologias para o processo produtivo [online]. [S.l.: s.n., 1999] [citado em 28 jul. 2000]. Disponível na Internet: <<http://www.cria.org.br/gip/gipap/capelli.htm>>
- CONGALTON, R.G. & MEAD, R.A.A. Quantitative method to test for consistency and correctness in photointerpretation. *Photogr. Eng. Sens.*, 49:69-74, 1983.
- EMMOTT, A.; HALL, J. & MATTHEWS, R. Precision farming applied to plantation agriculture [online]. [S.l.: s.n., 1997] [citado em 10 fev. 2000]. Disponível na Internet: <<http://www.silsoe.cranfield.ac.uk/cpf/>>
- GREENACRE, M.J. Theory and applications of correspondence analysis. London, Academic Press, 1984. 364p.
- GREENACRE, M.J. & HASTIE, T. The geometric interpretation of correspondence analysis. *J. Am. Statist. Assoc.*, 82:437-447, 1987.
- IDRISI WWW TUTORIAL. Map pairs analysis tools II [online]. [S.l.: s.n., 1999] [citado em 29 ago. 2000]. Disponível na Internet: <<http://www.sbg.ac.at/geo/idrisi/wwwtutor/mptools2.htm>>
- ROSENFELD, G.H. & FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. *Photogr. Eng. Remot. Sens.*, 52:223-227, 1986.
- SAS INSTITUTE. SAS/STAT users guide: version 6. 4.ed. Cary, 1998. v.1. 943p.
- VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R. & BIGGAR, J.W. Geostatistical theory and application to variability of some agronomical properties. *Hilgardia*, 51:1-75, 1983.

