

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

**UTILIZAÇÃO DE TÉCNICAS DE ANÁLISE DE AGRUPAMENTO
DO RISCO DE GEADA NO ESTADO DO PARANÁ PARA A CULTURA
DO MILHO SAFRINHA**

ROGÉRIO MENDONÇA MARTINS

Tese apresentada à Faculdade de Ciências Agronômicas da Universidade Estadual Paulista - Câmpus de Botucatu, para obtenção do título de Doutor em Agronomia - Área de Concentração em Energia na Agricultura.

BOTUCATU-SP
2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CÂMPUS DE BOTUCATU

**UTILIZAÇÃO DE TÉCNICAS DE ANÁLISE DE AGRUPAMENTO
DO RISCO DE GEADA NO ESTADO DO PARANÁ PARA A CULTURA
DO MILHO SAFRINHA**

ROGÉRIO MENDONÇA MARTINS

Orientadora: Prof. Dra. Sheila Zambello de Pinho

Co-Orientador: Dr. Sérgio Luiz Gonçalves

Tese apresentada à Faculdade de Ciências Agronômicas da Universidade Estadual Paulista - Câmpus de Botucatu, para obtenção do título de Doutor em Agronomia – Área de Concentração em Energia na Agricultura.

BOTUCATU- SP
2008

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO - SERVIÇO TÉCNICO DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP - FCA LAGEADO - BOTUCATU (SP)

Martins, Rogério Mendonça, 1968-
M383u Utilização de técnicas de análise de agrupamento do risco de geada no Estado do Paraná para a cultura do milho safrina / Rogério Mendonça Martins. - Botucatu : [s.n.], 2008.
viii, 85 f. : il. color., gráfs., tabs.

Tese (Doutorado)-Universidade Estadual Paulista, Faculdade de Ciências Agrônomicas, Botucatu, 2008
Orientador: Sheila Zambello de Pinho
Co-orientador: Sérgio Luiz Gonçalves
Inclui bibliografia

1. Milho. 2. Temperatura. 3. Milho - Cultura. 4. Geada.

I. Pinho, Sheila Zambello de. II. Gonçalves, Sérgio Luiz. III. Universidade Estadual Paulista "Júlio de Mesquita Filho" (Campus de Botucatu). Faculdade de Ciências Agrônomicas. IV. Título.

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CAMPUS DE BOTUCATU
CERTIFICADO DE APROVAÇÃO

TÍTULO: "UTILIZAÇÃO DE TÉCNICAS DE ANÁLISE DE AGRUPAMENTO DO
RISCO DE GEADA NO ESTADO DO PARANÁ PARA A CULTURA DO
MILHO SAFRINHA"

ALUNO: ROGÉRIO MENDONÇA MARTINS

ORIENTADOR: PROFA. DRA. SHEILA ZAMBELLO DE PINHO

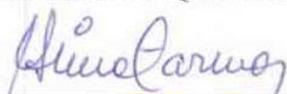
Aprovado pela Comissão Examinadora



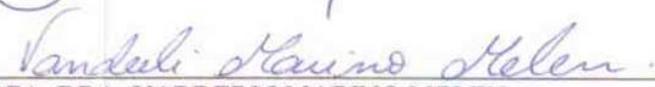
PROFA. DRA. SHEILA ZAMBELLO DE PINHO



PROFA. DRA. LIDIA RAQUEL DE CARVALHO



PROFA. DRA. MARISTELA SIMÕES DO CARMO



PROFA. DRA. VARDERLI MARINO MELEN



PROF. DR. VANDIR MEDRI

Data da Realização: 30 de abril de 2008.

À Deus, meu eterno protetor.

A minha mãe Dirce, ao meu pai Antonio (*in memoriam*),
e minhas irmãs Denise e Eliane

DEDICO

AGRADECIMENTOS

A Deus por ter me concedido força e coragem para enfrentar todas as dificuldades e concluir este trabalho.

A Profa. Dra. Sheila Zambello de Pinho pela orientação, ensinamentos, incentivo e amizade.

Ao pesquisador Dr. Sérgio Luiz Gonçalves pelas orientações, incentivo e amizade.

A minha mãe Dirce Mendonça Martins, pelo apoio e incentivo no decorrer do curso e irmãs Denise Mendonça Martins e Eliane Mendonça Martins.

A todos meus amigos, que sempre me apoiaram e incentivaram.

Em especial, a minha amiga Maria Dirce Bossa Vale por todo apoio dado nas horas difíceis e pela revisão gramatical.

Aos amigos do Departamento de Estatística, pelo constante incentivo, amizade e apoio.

As amigas Ana Vergínia, Jacinta e Vanderli, pela agradável convivência no curso de pós-graduação.

Ao Sr. Edmirson Borrozzino, técnico em Meteorologia - Setor de Agrometeorologia do IAPAR - por ter fornecido os dados das estações agrometeorológicas.

Ao amigo Luiz Carlos Fabrício de Melo pela cooperação e ajustes finais.

À Prof^a. Corina Maria Tedeschi Busnardo pela versão do resumo na língua inglesa.

A todos que direta ou indiretamente colaboraram para a realização deste trabalho.

SUMÁRIO

LISTA DE QUADROS.....	VII
LISTA DE FIGURAS.....	VIII
RESUMO.....	1
ABSTRACT	2
1 INTRODUÇÃO	3
2 REVISÃO DE LITERATURA.....	7
2.1 O Milho Safrinha no Paraná	7
2.2 Experimentação Agronômica	10
2.3 Análise de Agrupamento.....	15
2.4 Medidas de Similaridade e Dissimilaridade	18
2.5 Coeficientes de Similaridade e Dissimilaridade para dados quantitativos	19
2.5.1 Distância Euclidiana	20
2.5.1.1 Distância Euclidiana Média	21
2.5.1.2 Distância Euclidiana Ponderada	21
2.5.1.3 Distância Generalizada de Mahalanobis	22
2.5.1.4 Distância de Minkowski	22
2.5.2 Coeficiente Geral de Gower	23
2.5.3 Coeficiente de Similaridade de Cattell.....	23
2.5.4 Coeficiente de Camberra	24
2.5.5 Coeficiente de Correlação.....	24
2.5.6 Coeficientes de dissimilaridade ou critério de semelhança.....	24
2.6 Classificação das Técnicas de Agrupamento.....	25
2.6.1 Técnica Hierárquica.....	26
2.6.1.1 Técnica hierárquica aglomerativa	27
2.6.1.1.1 Vizinho mais Próximo (Single Linkage)	27
2.6.1.1.2 Vizinho mais distante (Complete Linkage)	28
2.6.1.1.3 Distância média (Average Linkage)	28
2.6.1.2 Técnica hierárquica divisiva	28
2.6.2 Técnica não hierárquica	29
2.6.3 Técnicas de validação dos agrupamentos	29
2.6.3.1 Gráfico de silhueta.....	30
2.6.3.2 Gráfico de perfil.....	31
2.7 Outras Aplicações de Análise de Agrupamento.....	31

2.8	Resumo da revisão de literatura.....	35
3	MATERIAL E MÉTODOS.....	39
3.1	Material.....	39
3.2	Métodos.....	41
3.2.1	Medidas descritivas para amostras multivariadas.....	42
3.2.2	Análise de Agrupamento.....	43
3.2.2.2	Algoritmo de agrupamento.....	43
3.2.2.2.1	Método do vizinho mais próximo (“Single Linkage Method” – SLM).....	44
3.2.2.2.2	Método do vizinho mais distante (“Complete Linkage Method” – CLM).....	45
3.2.2.2.3	UPGMA (Unweighted Pair-Group Method Using Arithmetic Averages).....	45
3.3	Definição do número de grupos.....	46
3.3.1	Dendograma.....	46
3.3.2	Análise do comportamento do nível de fusão.....	46
3.4	Validação e Interpretação dos agrupamentos.....	47
3.4.1	Gráfico de Perfil.....	47
3.4.2	Programas Computacionais.....	47
4	RESULTADOS E DISCUSSÃO.....	49
4.1	Método do Vizinho mais Próximo.....	50
4.2	Método do Vizinho mais Distante.....	54
4.3	Método não ponderado de agrupamento aos pares por médias aritméticas.....	57
4.4	Comportamento dos agrupamentos com relação aos três algoritmos.....	60
	CONCLUSÕES.....	61
	REFERÊNCIAS BIBLIOGRÁFICAS.....	63
	APÊNDICE.....	71

LISTA DE QUADROS

	Página
Quadro 1 – Resumo da revisão da literatura	35
Quadro 2 – Risco médio de geada do florescimento ao enchimento de grãos do milho safrinha, semeado em 10 diferentes épocas de 06/jan a 06/abr em 22 estações meteorológicas do Estado do Paraná	40

LISTA DE FIGURAS

	Página
Figura 1	Dendograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar, extraído dos dados originais pelo “Single Linkage Method” – Método do vizinho mais próximo..... 51
Figura 2	Mapeamento das 22 estações, utilizando o método do vizinho mais próximo..... 52
Figura 3	Gráfico de perfil para o vizinho mais próximo..... 53
Figura 4	Dendograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar extraído dos dados originais pelo “Complete Linkage Method” – Método do vizinho mais distante..... 54
Figura 5	Mapeamento das 22 estações, utilizando o método do vizinho mais distante..... 55
Figura 6	Gráfico de perfil para o vizinho mais distante. 56
Figura 7	Dendograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar obtido dos dados originais pelo “UPGMA” (Unweighted Pair – Group Method Using Arithmetic Averages Linkage Method) – Método não ponderado de agrupamento aos pares por médias aritméticas..... 57
Figura 8	Mapeamento das 22 estações, utilizando o método não ponderado de agrupamento aos pares por médias aritméticas..... 58
Figura 9	Gráfico de perfil para o método não ponderado de agrupamento aos pares por médias aritméticas..... 59

UTILIZAÇÃO DE TÉCNICAS DE ANÁLISE DE AGRUPAMENTO DO RISCO DE GEADA NO ESTADO DO PARANÁ PARA A CULTURA DO MILHO SAFRINHA.

Botucatu, 2008. 85f. Tese (Doutorado em Agronomia – Área de Concentração em Energia na Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Autor: ROGÉRIO MENDONÇA MARTINS

Orientadora: Sheila Zambello de Pinho

Co-orientador: Sérgio Luiz Gonçalves

RESUMO

O presente trabalho tornou-se relevante por verificar as áreas aptas para o cultivo do milho safrinha no Estado do Paraná, tendo como objetivo oferecer uma metodologia que possa contribuir para compreensão da variabilidade agrometeorológica desse Estado, fornecendo informações anuais por decêndios em 22 regiões analisadas por meio de banco de dados históricos de temperatura, diagnosticando as áreas homogêneas para identificar as regiões propícias ao cultivo do milho safrinha. Para atingir este objetivo trabalhou-se com a técnica de análise de agrupamento, por meio de um conjunto de dados fornecido pelo IAPAR – Londrina. Na análise utilizou-se a técnica hierárquica aglomerativa e três métodos de agrupamento. A série histórica constitui-se do vizinho mais próximo, vizinho mais distante e método não ponderado aos pares de médias aritméticas. Como síntese dos agrupamentos, os resultados mostraram a formação de quatro grupos para o método do vizinho mais próximo e vizinho mais distante, formando cinco grupos para o método não ponderado aos pares de médias. Observou-se no gráfico de perfil que nas dez simulações houve um risco maior de geada para as simulações mais tardias. Através dos agrupamentos identificou-se as localidades com a mesma característica de temperatura e as simulações ofereceram um embasamento para a melhor época do plantio.

Palavras-Chave: milho (*Zea Mays*), temperatura, análise de agrupamento, técnica hierárquica aglomerativa.

THE USE OF FROST RISK CLUSTER ANALYSIS TECHNIQUES IN THE STATE OF PARANÁ FOR CORN WINTER CROPS. Botucatu,2008, 85 pages. Thesis (PhD in Agronomy- Major Area: Energy and Agriculture) - Faculdade de Ciências Agronômicas , Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Author: ROGÉRIO MENDONÇA MARTINS

Advisor: Sheila Zambello de Pinho

Co-advisor: Sérgio Luiz Gonçalves

ABSTRACT

This work became relevant for verifying the favorable areas for the cultivation of winter corn in the State of Paraná, offering a methodology which led to a better understanding of the agrometeorological variability in the State, providing annual information by decennials, in 22 regions analyzed by a temperature historical data base, diagnosing the homogeneous areas to identify favorable ones for the cultivation of winter corn. To reach this objective, this study adopted the cluster analysis technique through data from IAPAR – Londrina. During the analysis, the agglomerative (bottom-up) hierarchical technique and three clusters methods were used. The historical series were constituted by the nearest neighbor, the farthest neighbor and the non-weighted method to the pairs of arithmetic means. As like the clusters' synthesis, the nearest and farthest neighbors' method results showed the development of 4 groups, resulting in 5 groups for the non-weighted method to the pairs of means. The profile graph showed that in all ten simulations there was greater risk of frost in the simulations conducted the latest. Through clustering, locations with the same temperature characteristic were identified, and the simulations provided a basis for best sowing period.

Key words: corn (*Zea Mays*) , temperature, cluster analysis, agglomerative (bottom-up) hierarchical technique

1 INTRODUÇÃO

A crise econômica e financeira ocorrida no Brasil, no final dos anos 80, mudou radicalmente o destino da seguridade agrícola brasileira. O Ministério da Agricultura e do Abastecimento em 1993, identificou uma série de problemas operacionais no Programa de garantia da Atividade Agropecuária – PROAGRO, mostrando a sua vulnerabilidade em função dos altos índices de perdas na agricultura em virtude de secas e chuvas na colheita, levando agricultores à falência, pois o acúmulo da dívida passou de 700 milhões de reais. (GONÇALVES, 2004).

Decorrente desses fatores surgiu, em 1995, o Projeto de Redução de Riscos Climáticos na Agricultura, passo importante para o Zoneamento Agrícola do Brasil, destacando-se como de fundamental importância a determinação da época de semeadura proporcionando simultaneidade de períodos em que as condições são mais favoráveis com os estádios críticos da cultura.

Existe nas diferentes culturas, um período crítico específico em relação à exigência hídrica. Ramos & Mundstak (1986), mostram que o milho, em especial, necessita de disponibilidade hídrica para a sua frutificação uma vez que, após o espigamento, a

planta diminui o seu crescimento vegetativo para proceder ao enchimento de grãos, momento este que a disponibilidade hídrica se torna relevante por estar diretamente envolvida na distribuição de fotoassimilados.

É racional, compreender a importância da profissionalização do mercado de milho safrinha, pois o que antes era visto como uma mera oportunidade de venda, hoje é, para muitos agricultores, um mercado profissional rentável.

O problema, em questão, é que alguns produtores plantam o milho safrinha sem o prévio conhecimento das condições climáticas de sua região e nesse caso, acabam tendo poucas possibilidades de sucesso. O bom resultado do cultivo depende de um adequado planejamento da lavoura começando pela escolha da área que será destinada ao plantio da safrinha de milho.

Numa visão histórica, sabe-se que países em desenvolvimento, como o Brasil, dependem fortemente da agricultura, sofrem com as condições extremas de precipitação que provocam grandes perdas de produção agrícola prejudicando sensivelmente toda a área econômica. Esses eventos meteorológicos extremos associados aos efeitos econômicos causam graves conseqüências na sociedade, por serem causados por anomalias extensas e altamente prejudiciais. Uma forma de minimizar esses problemas é dispor de diagnóstico para tais fatos.

Para o desenvolvimento das plantas existe um limite térmico onde há um desenvolvimento de suas funções e seus processos fisiológicos. Todas as etapas, desde a semeadura até seu florescimento, exigem um somatório térmico que é calculado através de um índice chamado de graus-dia, onde leva-se em conta o acúmulo térmico da temperatura inferior e superior dentro dos limites que a planta se desenvolve. Daí a necessidade de se conhecer a época de plantio analisando todo o ciclo da cultura, procurando prever as condições ambientais em todas as suas fases fenológicas.

Diante disso, desenvolveu-se o zoneamento agrícola do Estado do Paraná onde são repassadas aos agricultores todas as informações referentes ao plantio de suas culturas com o objetivo de reduzir os riscos climáticos envolvidos na exploração agrícola, por meio da indicação do plantio, somente nas áreas climaticamente aptas e, em épocas que tenham condições de clima adequado.

O clima paranaense tem um grande impacto na produção de grãos e, para que essa cultura se amplie, sem gerar prejuízos econômicos para o agricultor, se faz necessário o conhecimento das estações climáticas do estado.

No contexto climático, há certa limitação em relação à representatividade de informações pontuais obtidas em experimentos de campo indicando a necessidade de se desenvolverem estudos em séries históricas de diversas estações meteorológicas, objetivando delimitar áreas semelhantes quanto ao comportamento dos parâmetros climáticos, a fim de que se possa quantificar os riscos envolvidos na comercialização das culturas, uma vez que o estado ocupa uma faixa de transição com acentuada variabilidade climática.

Para que isso ocorra de forma efetiva e satisfatória, este trabalho objetiva, ainda que estudos climáticos empreendidos no Estado do Paraná tenham sido realizados, oferecer uma metodologia para melhorar a compreensão da variabilidade agrometeorológica ao fornecer informações gerais, em nível anual, relacionado a estudos mais detalhados por decêndios (período de dez dias) possibilitando uma leitura mais clara sobre as diferenças de temperatura nas 22 regiões analisadas nesse estado.

Especificamente, este estudo vem contemplar a organização do banco de dados históricos de temperatura no Estado do Paraná para diagnosticar as áreas homogêneas desse conjunto de dados, com isso indicando as áreas propícias ao cultivo do milho safrinha, contribuindo, também, para a implantação do sistema de monitoramento agroclimático e para a avaliação do cronograma de atividades agrícolas do estado e, principalmente, demonstrar a eficiência dos três algoritmos da técnica hierárquica aglomerativa para os agrupamentos.

Desse modo, o problema que a análise de agrupamento pretende resolver segundo Rosa Neto (2006) é:

Dada uma amostra de n objetos, cada um deles medidos segundo p variáveis, procurar um esquema de classificação que agrupe os objetos em g grupos. Devem ser determinados também o número e as características desses grupos (EVERITT, 2001).

Hair (2005), afirma que a análise de agrupamento é um instrumento útil e importante para a análise de dados em diversas situações, pois permite ao pesquisador

aplicá-las em qualquer conjunto de dados e, a escolha do método do agrupamento depende do objetivo da avaliação ao agregar os objetos.

Diante desse contexto, este trabalho tem por objetivos: 1) diagnosticar quais áreas aptas ao cultivo do milho para o risco de geada e melhorar a compreensão da variabilidade agrometeorológica fornecendo informações gerais, em nível anual e por decêndio na cultura do milho safrinha; 2) comparação dos métodos multivariados para averiguar o agrupamento das regiões com base nas temperaturas caracterizando as mais propícias ao risco de geadas no Paraná e, portanto, menos indicadas para o cultivo do milho safrinha.

2 REVISÃO DE LITERATURA

2.1 O Milho Safrinha no Paraná

O milho tem seu lugar de destaque no Brasil pela sua fácil adaptação e produção no país. O aumento de produtividade é o resultado do avanço tecnológico que vem contribuindo para a redução de perdas relacionadas aos riscos climáticos. Para a cultura do milho a água é fator determinante na produção, principalmente na fase da germinação e nos períodos de floração e enchimento de grãos. Semente de má qualidade, manejo inadequado de fertilizantes e equipamentos, também são fatores que colocam em risco a semeadura do milho.

A seleção da época, em que as condições climáticas dominantes indicam disponibilidade de água para os períodos críticos, se faz relevante para o sucesso da produção de grãos que muito depende das variações do clima. Esta variável não controlável, o

clima, é um dos mais importantes componentes de produção da cultura, afirmam Sans et. al, 2001.

Desde a década de 70, o Estado do Paraná vem se consagrando como o maior produtor nacional de milho, cuja produção corresponde a mais de 25% do total produzido no país. A área plantada ficou estável nos últimos anos, sendo que o aumento da produção foi possibilitado pelos ganhos de produtividade (WREGGE et. al, 1998).

O Paraná tem cerca de 230 mil produtores, que são contemplados com uma produção que gira em torno de 8 milhões de toneladas ao ano, gerando aproximadamente 70 mil empregos no campo sem computar a mão de obra utilizada na indústria. Daí, a importância da espécie no contexto sócio econômico do Estado. O cultivo do milho está presente em todas as regiões do Estado, por ser a espécie de fácil adaptação, todavia, nas regiões Sul, Sudeste e Norte se concentram a produção de milho e as atividades pecuaristas. Na safrinha a produção paranaense atinge 45% da produção nacional.

O milho safrinha é uma cultura desenvolvida de janeiro a abril, normalmente, após a soja precoce, e, em alguns locais, após o milho de verão e o feijão das águas, portanto, não tem um período pré-fixado para seu plantio, como o milho de safra normal, que é plantado no início das chuvas. Isso torna a sua produtividade bastante prejudicada pelo regime das chuvas e por fortes limitações de radiação solar e temperatura na fase final de seu ciclo. Por ele ser plantado após uma cultura de verão, a semeadura depende da época do plantio da cultura antecessora e de seu ciclo. Dessa forma, o planejamento do plantio do milho safrinha se inicia com a cultura de verão, com o intuito de liberar a área o mais cedo possível.

Tudo isso vem torná-la uma cultura de alto risco, uma vez que o período de chuva encontra-se no fim, provocando uma inconstante espacial e temporal, o que leva a uma variabilidade de produção.

O potencial produtivo do milho safrinha pode ser comprometido se ocorrer geadas, baixa temperatura e déficit hídrico. O plantio fora da época propicia afeta o seu desenvolvimento. Isso se dá, porque parte do ciclo transcorre em períodos onde a soma térmica é baixa abalando o processo metabólico da planta. Quando a planta é submetida à temperatura noturna abaixo de 5° C, leva em média 48 horas para recuperar a taxa de fixação de CO₂, acarretando a má formação dos grãos. A germinação é prejudicada pelas temperaturas

do solo abaixo de 10° C e superior 42° C, sendo que a temperatura entre 25 a 30° C propicia a germinação das sementes e emergência das plântulas (OLIVEIRA, 2003).

É relevante conhecer a época do plantio analisando todo o ciclo da cultura, prevendo as condições ambientais em suas fases fenológicas, pois a maior dificuldade encontrada diz respeito às variações ambientais não previsíveis, que correspondem aos fatores ambientais altamente variáveis, tanto espacialmente como de forma temporal (precipitação, temperatura, vento, etc). Portanto, a época ideal para semeadura refere-se ao período em que a cultura tem maior probabilidade de desenvolver-se em condições edafoclimáticas favoráveis.

Os melhores resultados em grãos são obtidos nos plantios após a colheita da soja, principalmente as de ciclo precoce, tornando-se vantajoso pelo aproveitamento do residual de fertilidade deixado pela cultura da soja como também por evitar a ocorrência das doenças de colmo e de grãos (PEIXOTO, 2002).

Gonçalves et al. (2002) realizaram um estudo compilando informações e conhecimentos e aplicando a análise de séries de dados em solos e relevo procurando delimitar com maior segurança, as áreas com características adequadas e os períodos mais favoráveis ao cultivo da “safrinha”, visando a redução de riscos relacionados com os fatores climáticos, contribuindo para que o potencial produtivo dos híbridos seja alcançado, com o mínimo de perda possível. O estudo teve como objetivo classificar a definição de zonas homogêneas quanto a épocas de plantio, tendo como norte a análise detalhada dos riscos de geadas e deficiência hídrica no período de fixação da cultura até a maturação.

O Paraná apresenta oscilações no clima, seja no regime térmico, seja no regime hídrico. A insuficiência ou má distribuição de precipitação pluvial ao longo da estação de crescimento das culturas pode ser considerada como a principal causa das frustrações das safras agrícolas (FARIA et al, 1993 apud GONÇALVES et al, 2006).

De acordo com Farias et al. (1993), não só a deficiência hídrica determina o insucesso do empreendimento, mas também a quantificação das safras, aumento da produção de grãos e melhoria das operações agrícolas (apud GONÇALVES, 2006). O zoneamento é uma das ferramentas a disposição do agricultor para garantir mais estabilidade na produção.

Segundo levantamento do IBGE (2006), sobre o resultado da produção agrícola de cereias, o Paraná contribuí com 19,8% na safra nacional com uma produção de

23,96 milhões de toneladas de grãos na safra de verão e inverno. Confirmando-se, com este resultado, a posição histórica que o estado mantém na liderança da produção de grãos no país.

De acordo com a SEAB – Pr – (2007), as geadas do fim de maio e início de junho causaram perdas de 900 mil toneladas na safrinha de milho, cuja produção esperada era de 5,7 milhões de toneladas de grãos, sendo reduzida para 4,8 milhões de toneladas.

Em 2008, o Estado deverá plantar uma área recorde de safrinha, estimada em 1,6 milhão de hectares, crescimento de 11% na comparação com o ano passado que foi de 1,44 milhões de hectares (SAMORA, 2007). Enfatiza, ainda, que o aumento do plantio se deve a alta de preços pelo consumo interno e pelas exportações. Esse plantio já teve início no Paraná, atingindo 3,3% da área prevista. De acordo com as condições climáticas os agricultores poderão ter uma colheita com volume recorde na segunda safra estimado em 6,6 milhões de toneladas.

O Ministério da Agricultura estima que 44 milhões de toneladas de milho em 2007 e 2008, serão consumidas pelos brasileiros, apontando para uma alta de 3,5 milhões de toneladas na comparação com o ano-safra anterior. O Brasil deverá produzir 53,3 milhões de toneladas de milho em 2007 e 2008.

2.2 Experimentação Agronômica

A agricultura depende diretamente das inter-relações do sistema solo-planta-atmosfera, onde a água é imprescindível à sobrevivência das plantas. Doorenbos & Kassan (1979), “afirmam que o clima é um dos fatores importantes na determinação dos requerimentos hídricos para crescimento e produção das culturas”. Quando elas se encontram em fase de crescimento, necessitam de água para o estabelecimento da cultura entre o florescimento e a formação de grãos.

Fornasiere Filho (1992) apud Maluf (1992) relata que o milho é uma cultura que pertence ao grupo das plantas C4, e deve ser cultivado em latitudes que variam de

50° N a 42° S e, nessa ampla área territorial destaca-se a multiplicidade de ambientes. Nesse contexto, o Paraná encontra condições favoráveis ao seu cultivo destacando-se como um dos principais produtores nacionais.

Gonçalves et al. (1997) utilizaram séries de dados de temperatura de 29 estações do Instituto Agrônômico do Paraná, onde foram calculadas as probabilidades empíricas de ocorrência de temperaturas superiores a 30° C no período do florescimento do feijoeiro, na safra “das águas”, para cinco épocas de semeadura, simuladas no décimo quinto dia dos meses de julho a novembro. Para mapear o estado em diferentes faixas de risco fez-se o cálculo de correlações lineares significativas entre os níveis de risco, latitude e altitude. Observou-se que nas regiões localizadas ao Sul do Estado, a probabilidade de ocorrência de altas temperaturas foi sempre menor que 20%, para todas as datas de semeadura simuladas. Nas outras regiões, mais ao Norte e ao Oeste pode-se observar um maior risco entre 40 e 60% para semeaduras efetuadas a partir de 15 de setembro, indicando haver restrições para o cultivo nas épocas de semeaduras tardias.

Lozada e Angelocci (1999) estudaram as observações de duração de subperíodos e do ciclo semeadura-colheita de um híbrido de milho, em ensaio de épocas de semeadura, para estimar valores de temperatura-base para o subperíodo da semeadura à floração e para comparar as estimativas da duração desse subperíodo através do uso de graus-dia, com ou sem correção para o efeito da disponibilidade hídrica do solo. Incluem-se nessas comparações as estimativas feitas com uso dos valores da temperatura-base obtidos no trabalho e com uso do valor de 10° C citado com frequência na literatura.

Gadioli et al. (2000) avaliaram o efeito da temperatura média do ar no rendimento de grãos com a finalidade de prever a época de ocorrência dos estádios fenológicos da cultura de milho em função da soma calórica. Realizaram um experimento de campo utilizando três híbridos, com diferentes exigências térmicas, em três diferentes épocas de semeadura. Para estimativa do rendimento de grãos foram determinados seus diferentes componentes, onde se observou decréscimo do rendimento com aumento da temperatura em função das épocas de semeadura utilizadas. A fenologia foi associada ao número relativo de graus-dia, onde se observou uma relação linear. Há uma temperatura média do ar correspondente ao máximo rendimento de grãos, e os estádios fenológicos da cultura podem

ser previstos através do conhecimento prévio da série histórica de temperatura da localidade e da soma calórica requerida para o ponto de maturidade fisiológica.

Em termos de conceituação teórica Gerage (2000), coloca que as maiores flutuações de produtividade ocorrem na “safrinha”, em função de que nesse período de cultivo, janeiro a agosto, as lavouras ficam sujeitas às variações climáticas mais acentuadas. Por ser o milho bastante vulnerável às variações térmicas e à distribuição das chuvas, os efeitos climáticos são visíveis no desempenho da cultura.

Mesmo sendo possível a obtenção de produtividade razoável em condições adversas, os maiores rendimentos se obtêm sob condições climáticas adequadas, evitando assim, que possam ocorrer prejuízos significativos. Cultivado em temperaturas abaixo de 10° C o crescimento do milho fica estacionado retardando o ciclo e expondo a planta às pragas e doenças.

Já, a temperatura acima de 30°C estimula aumento no processo de respiração, resultando em ganhos insignificantes na matéria seca acumulada pela planta. Vale saber, a respeito desses extremos térmicos, que cada híbrido ou variedade acumula um total de graus-dia ou unidades de calor, relativamente constantes durante a fase de desenvolvimento, independente da região em que é cultivado. (AYOADE, 2004).

Em Gonçalves et al. (2002), vê-se que existe uma gama de informações e conhecimentos advindos de estudos realizados com a cultura no Paraná, que permitem juntamente com análise de séries de dados meteorológicos, solos e relevo, delimitar com segurança as áreas mais propícias e os períodos mais favoráveis ao cultivo da safrinha, porém os riscos climáticos não se encontram totalmente definidos no Estado.

Devido a sua localização geográfica, o Estado do Paraná, apresenta grandes variações de relevo, com altitudes de zero a 1.300 m, segundo Caviglione et al. (2000) e grande diversidade de solos (Empresa Brasileira de Pesquisa Agropecuária, 1981), com reflexos diretos sobre a temperatura, evapotranspiração e o balanço hídrico local. Por isso, são encontrados climas distintos (INSTITUTO AGRONÔMICO DO PARANÁ) e microclimas com condições apropriadas para cultivo de várias espécies. As temperaturas diminuem do norte para o sul, com exceção do litoral. Quando ocorrem geadas no Estado do Paraná, nos períodos críticos da cultura do milho safrinha, sua produção fica sensivelmente prejudicada, bem como sua qualidade.

Com o objetivo de se verificar a ocorrência de geadas no Estado do Paraná, tem-se utilizado para os cálculos climáticos da ocorrência de geadas, temperaturas mínimas do ar no abrigo 3°C. Procurou-se estudar séries históricas de temperatura mínima de abrigo 0°C, temperatura de relva 0°C, temperatura mínima de abrigo 3°C e geadas observadas nas 28 estações da rede do Instituto Agrônômico do Paraná - IAPAR. Foram definidos os períodos com maior probabilidade de ocorrência da primeira e última geada nas estações que representam as diferentes regiões do Estado. Também foram calculadas as probabilidades de persistência de geada para os dias seguintes. Os resultados indicaram que na maior parte (89%) das estações meteorológicas, o método de se adotar temperatura mínima do ar no abrigo 3°C para determinar a ocorrência de geada não tem boa precisão. A comparação do número de geadas com base nos critérios T_{\min} 3°C e observação visual, mostrou que em 71% das estações o número de geadas observadas é significativamente maior que as determinadas pelo critério T_{\min} 3°C, em 18% ocorre o contrário e em 11% os resultados são comparáveis, (KIM et al, 2003).

Gonçalves (2004) utilizou um modelo de simulação de balanço hídrico, que usa séries históricas de dados diários de evapotranspiração máxima e precipitação para determinar os riscos de deficiência hídrica no período do pré-florescimento ao enchimento de grãos da cultura do milho, selecionando as melhores épocas de semeadura para a região dos Campos Gerais do Paraná.

Assad et al. (2004) relatam que a partir das indicações do último relatório (International Panel of Climatic Change) - IPCC, foram feitas várias simulações e avaliados os impactos que um aumento na temperatura média do ar de 1° C, 3° C e 5,8° C e um incremento de 15% na precipitação pluvial teriam na potencialidade da cafeicultura brasileira, definida pelo atual zoneamento agroclimático do café (*Coffea arabica* L.) nos estados de Goiás, Minas Gerais, São Paulo e Paraná. Os resultados indicam uma redução de área apta para a cultura superior a 95% em Goiás, Minas Gerais e São Paulo, e de 75% no Paraná.

Cardoso et al. (2004) investigaram os efeitos da época de semeadura no estabelecimento e rendimento do milho safrinha para Londrina – PR, mediante simulações de longo período com o modelo CERES. Os tratamentos consistiram na combinação de 12 épocas de semeadura, uma em cada decêndio, de janeiro a abril, sob dois cenários: sem limitação hídrica (produção potencial) e com limitação hídrica, considerando-se uma cultivar

de ciclo precoce, XL-520, durante 24 anos. Os resultados mostraram decréscimos de rendimentos de até 38% da produtividade potencial e de 44% da produtividade sob restrição hídrica à medida que se retardou a semeadura, devido à coincidência dos períodos críticos de desenvolvimento com condições subótimas de radiação solar, temperatura e disponibilidade hídrica. Observaram-se aumentos na duração do ciclo de 120 a 140 dias quando o milho foi semeado em janeiro e de 160 a 170 dias para semeadura em abril. Observou-se que há risco de geadas de 4% para os cultivos semeados em final de março e 8% para os semeados em abril.

Streck e Alberto (2006) avaliaram, por meio de estudo numérico, a existência de impacto da possível mudança climática sobre o rendimento das culturas do trigo, soja e milho em Santa Maria, RS. Foram criados cenários de mudança climática, dobrando-se a quantidade de CO₂, com diferentes aumentos de temperatura do ar, com aumento e sem aumento de precipitação pluvial. O rendimento das três culturas foi simulado com modelos matemáticos disponíveis na literatura. Concluiu-se que a mudança climática, projetada pela simulação, influenciou o rendimento de grãos de trigo, soja e milho. O aumento de 2, 3 e 6^o C na temperatura do ar pode anular os efeitos benéficos do aumento de CO₂ no rendimento de trigo, soja e milho, respectivamente.

Cargnelutti Filho et al. (2006) trabalharam com o objetivo de verificar, com dados de temperatura mínima média decendial do ar (T_m) de 41 municípios do Estado do Rio Grande do Sul, de 1945 à 1974, se a T_m pode ser estimada em função da altitude, latitude e longitude. Para cada um dos 36 decêndios do ano, realizaram-se análise de correlação, análise de trilha das variáveis causais – altitude, latitude e longitude – sobre o efeito T_m, e estimaram-se os parâmetros do modelo das equações de regressão linear múltipla, pelo método passo a passo, com teste para saída de variáveis, considerando T_m como variável dependente e altitude, latitude e longitude como variáveis independentes. Na validação dos modelos de estimativa da T_m, usou-se o coeficiente de correlação linear de Pearson, entre a T_m estimada e a T_m observada em dez municípios do Estado, com dados da série de observações meteorológicas de 1975 a 2004. A temperatura mínima média decendial do ar pode ser estimada pelas coordenadas geográficas em qualquer local e decêndio, no Estado do Rio Grande do Sul. A altitude e latitude explicam melhor a variação da T_m.

Assis et al. (2006) utilizaram um procedimento estocástico com o objetivo de estimar a produtividade potencial da cultura do milho de acordo com valores

médios diários de temperatura do ar e radiação solar. Foram fornecidos pela Estação Agrometeorológica de Piracicaba (ESALQ/USP) valores de temperatura média diária do ar medidos durante 86 anos e da radiação solar global diária durante 25 anos. Os valores correspondentes à produtividade potencial foram simulados 1000 vezes, em cada data de semeadura considerada. A metodologia de estimação permitiu definir a ordem de grandeza da produtividade potencial do milho a determinada localidade, com base nos dados de temperatura e radiação solar.

Gonçalves et al. (2002) avaliaram através de análise de regressão Linear Múltipla risco de geada em todo ciclo da cultura do milho “safrinha”, com atenção especial ao período compreendido entre o florescimento e a maturação, bem como os de deficiência hídrica no estabelecimento e no florescimento, para identificar zonas homogêneas climaticamente e indicar as melhores épocas de semeadura para o Estado do Paraná. O risco de geada, nos trinta dias que antecederam a maturação, foi considerado o de maior restrição ao cultivo da “safrinha” de milho. Este fator, associado à combinação de menor risco de deficiência hídrica, permitiu a identificação de oito regiões distintas, com períodos de semeadura variando de janeiro a agosto.

2.3 Análise de Agrupamento

Hair et al. (2005) definem a Análise de Agrupamentos como sendo um grupo de técnicas proveniente de uma das áreas da estatística conhecida por Análise Multivariada cuja finalidade primária é juntar objetos com base nas características que eles possuem.

A análise de agrupamento é um conjunto de técnicas, que utilizada sob determinadas condições nos dados existentes, contribui para a formação de grupos naturais ou homogêneos.

A maioria dos autores considera a análise de agrupamento como o conjunto de técnicas que permite dividir os dados, que normalmente apresentam componentes com observações multivariadas ou multidimensionais, em grupos naturais.

A divisão em grupos ou classes facilita a compreensão das observações e o desenvolvimento das teorias científicas que vêm utilizando essas técnicas em importantes áreas científicas para a identificação de padrões de comportamento nos dados indicados, auxiliando com isso o processo de descoberta do conhecimento. Elas são utilizadas ainda na fase exploratória da pesquisa, onde a falta de hipóteses a priori sobre as observações e o desconhecido número de grupos, permitem que as mesmas auxiliem na organização dos dados em estruturas significativas de fácil interpretação.

Dessa forma Bussab et al. (1990) descreve:

Dada uma amostra de n objetos, cada um deles medido segundo p variáveis, procura-se um esquema de classificação que agrupe os objetos em k grupos.

Ao gerarem grupos para o pesquisador, este por ter pleno conhecimento sobre o problema consegue identificar os agrupamentos semelhantes.

Para isso os objetos deverão ser mensurados nas diferentes variáveis fornecendo uma matriz de dados de n objetos por p variáveis, a qual será trabalhada através de algoritmos para se obter grupos homogêneos ou heterogêneos (FREI, 2006).

Nesse mesmo vértice, Hair et. al (2005) afirmam que a análise de agrupamento também é denominada por Taxonomia Numérica.

Segundo Messetti (2000), a representação gráfica, denominada por dendograma ou fenograma, apresenta um visual de grande utilidade que pode ser utilizado para a classificação, comparação e discussão descritiva dos agrupamentos. No gráfico, o eixo das abscissas representa as unidades taxonômicas operacionais (OTU's) e o eixo das ordenadas, os valores dos níveis de similaridade, associação ou de correlação obtidos através da metodologia aplicada.

Os agrupamentos são feitos com base nas medidas de distância (coeficiente de similaridade e dissimilaridade) calculada entre os indivíduos (elementos).

Como pré-requisito para aplicar essas técnicas, além da existência dos dados no conjunto inicial, é necessário definir qual o critério a ser utilizado na definição da proximidade entre os objetos. Na classificação, os objetos são agregados a grupos ou classes pré-definidas, e nos agrupamentos os grupos surgem durante o processo. Por conseguinte, a análise de agrupamento também pode ser utilizada para auxiliar uma classificação que não tenha sido aplicada ainda devido à falta dos grupos, e, características dos mesmos.

Quando da utilização das técnicas de análise de agrupamento, não é possível determinar antecipadamente as variáveis dependentes e independentes, ao contrário, as técnicas permitem examinar as relações de interdependência entre todo o conjunto de variáveis, nesse ponto sendo similar a análise fatorial, porém diferenciando-se da mesma por tratar-se dos objetos. À análise fatorial trabalha com as variáveis, buscando reduzir o conjunto das mesmas através da criação de fatores que medirão aspectos em comum (HAIR et al, 2005).

Sokal e Sneath (1963), apud Messetti (2000), contribuíram de forma grandiosa para essa metodologia no livro “Principles of Numerical Taxonomy”, voltado para a área biológica. A partir dessa época houve uma revolução nos trabalhos científicos centrados em dois motivos. O primeiro deve-se ao avanço tecnológico: as técnicas, que na época eram consideradas inviáveis pela grande quantidade de cálculos, passaram a ser viáveis. O segundo diz respeito à importância da classificação biológica no meio científico (taxonomia numérica), em que ocorre a junção de várias informações sobre o mesmo indivíduo.

De acordo com Anderberg (1973), a análise de agrupamento foi empregada como uma ferramenta eficaz na investigação científica. Um de seus papéis mais úteis é gerar hipóteses sobre a estrutura da categoria. Os resultados da análise de agrupamento contribuem diretamente para o desenvolvimento do sistema de classificação.

2.4 Medidas de Similaridade e Dissimilaridade

Dois aspectos de particular importância têm de ser considerados na análise de agrupamento, que são a medida de similaridade adotada para formar os grupos e a escolha do método de agrupamento.

Segundo Mingoti (2005), para que se possa proceder ao agrupamento de elementos, é necessário que se decida a priori a medida de similaridade ou dissimilaridade que será utilizada, sendo que tipos diferentes de medidas produzirão um determinado tipo de agrupamento.

Segundo Hair et al (2005), a similaridade entre objetos é uma medida de correspondência ou semelhança entre objetos a serem agrupados.

Quando se trabalha com Análise Fatorial, estabelece-se uma matriz de correlação entre variáveis, um processo semelhante ocorre em Análise de Agrupamentos onde as características que determinam a similaridade são primeiramente especificadas. Isso vem demonstrar que qualquer objeto pode ser comparado a qualquer outro através da medida de similaridade.

Bussab et. al (1990) orientam que nas medidas de similaridade pode-se observar que quanto maior o valor observado mais parecidos são os objetos e nas medidas de dissimilaridades quanto maior o valor observado menos parecidos serão os objetos. Temos, então, o coeficiente de correlação como um exemplo de similaridade, enquanto que a distância euclidiana é um exemplo de dissimilaridade.

Embora a similaridade entre objetos possa ser medida de diversas formas, três métodos dominam as aplicações de análise de agrupamentos.

- a) Medidas correlacionais que representam a similaridade pela correspondência de padrões ao longo das características (variáveis).
- b) Medidas de distância que na verdade é uma medida de dissimilaridade, em que os valores maiores denotam menos similaridade.

- c) Medidas de associação usadas para comparar objetos cujas características são medidas apenas em termos não-métricos.

Dependendo de seus objetivos e do tipo de dados esses métodos podem representar uma perspectiva única de similaridade (HAIR, 2005).

Mardia et al. (1979) apresentaram as seguintes propriedades de dissimilaridade. Sejam P e Q dois pontos de um conjunto de indivíduos de interesse. Uma função real $d(P,Q)$ torna-se uma distância quando possui as seguintes propriedades:

- i) $d(P,Q) = d(Q,P)$ simetria;
- ii) $d(P,Q) \geq 0$;
- iii) $d(P,P) = 0$;
- iv) $d(P,Q) = 0$ se e somente se $P=Q$;
- v) $d(P,Q) \leq d(P,R) + d(R,Q)$ desigualdade triangular

O termo de dissimilaridade apareceu em função de que à medida que $d(P,Q)$ cresce, P e Q tornam-se cada vez mais dissimilares. Há que se considerar que o direcionamento para a similaridade ocorre em sentido contrário, ou seja, menor $d(P,Q)$, maior similaridade entre P e Q.

Maxwell (1977) enfatiza que o primeiro estágio em muitos métodos da análise de agrupamento é a conversão ($n \times p$) da matriz dos dados, X, ($n \times n$) em uma matriz de similaridade ou de dissimilaridade inter-individual.

2.5 Coeficientes de Similaridade e Dissimilaridade para dados quantitativos

Para esta metodologia existe um conceito fundamental que é a escolha de um critério que determina a distância entre os objetos.

Após dispor os dados em forma de matriz, onde cada coluna representa as variáveis de interesse e linhas representando os elementos o próximo passo é a escolha de uma medida que identifique o quanto dois indivíduos são parecidos. Tem-se assim o

coeficiente de similaridade ao qual será gerada uma matriz D_{ij} de similaridade pelo qual se verifica a distância entre os elementos i e i' .

Em Bussab et al. (1990), são discutidos alguns coeficientes de similaridade e dissimilaridade para dados quantitativos.

2.5.1 Distância Euclidiana

Este é o coeficiente de dissimilaridade mais conhecido e utilizado para indicar a proximidade entre objetos. É, simplesmente, a distância geométrica entre dois objetos em um espaço multidimensional. A idéia básica é considerar cada observação como um ponto em um espaço euclidiano e, desse modo, calcular o coeficiente que representará a distância física entre os pontos (AZAMBUJA, 2005).

Seja $X = (x_{ij})$ a matriz de dados, onde x_{ij} representa o i -ésimo elemento de j -ésimo variável, com $i=1,2,\dots,n$ e $j=1,2,\dots,p$. Desta forma cada vetor coluna representa uma variável e cada vetor linha um determinado elemento (OTU).

A medida mais conhecida para indicar a proximidade entre dois OTU i e i' é a distância Euclidiana, dada por:

$$D_{ii'} = \left[\sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right]^{1/2}$$

ou em linguagem matricial :

$$D_{ii'} = \left[(x_i - x_{i'})' (x_i - x_{i'}) \right]^{1/2}$$

onde x_i e $x_{i'}$ são vetores relativos às respostas no i e i' ésimo elemento do estudo, respectivamente (MESSETTI, 2000).

2.5.1.1 Distância Euclidiana Média

Uma outra medida derivada da distância Euclidiana muito utilizada é a Distância Euclidiana Média, expressa pela raiz quadrada da divisão entre o somatório do quadrado das diferenças, pelo número de variáveis envolvidas.

Pode-se observar que o valor da distância Euclidiana aumenta quando novas variáveis são incorporadas às originais. Uma maneira possível de contornar esse problema é dividir esse valor pela raiz quadrada do número de caracteres, isto é:

$$\Delta_{ii'} = \frac{1}{\sqrt{p}} d_{ii'}$$

Essa distância é apenas um reescalonamento da distância anterior, possuindo as mesmas propriedades e, portanto, produzindo os mesmos resultados se submetidos às técnicas de análise de agrupamentos. Esse coeficiente possui uma propriedade interessante, garante que essa distância pode ser utilizada na ausência de dados para algumas coordenadas (“missing values”).

2.5.1.2 Distância Euclidiana Ponderada

Segundo Rosa Neto (2006), deriva-se da distância euclidiana onde está associada a uma questão freqüente em análise de agrupamento que é a ponderação das variáveis, ou seja, dar peso para as variáveis que o pesquisador julgar mais importante. Assim, pode-se criar uma matriz diagonal B de ponderação para as variáveis x_1, x_2, \dots, x_p com respectivos pesos $w_{11}, w_{22}, \dots, w_{pp}$ e a distância define-se:

$$D(x_l - x_k) = [(x_l - x_k)' B (x_l - x_k)]^{1/2}$$

2.5.1.3 Distância Generalizada de Mahalanobis

Para Mingoti (2005), a distância de Mahalanobis entre dois objetos x_l e x_k é um caso particular da distância Euclidiana ponderada:

$$D(x_l, x_k) = [(x_l - x_k)' V^{-1} (x_l - x_k)]^{1/2}$$

onde, calcula-se a inversa da matriz de variâncias e covariâncias V^{-1} , entre as variáveis utilizadas fornecendo assim a distância de Mahalanobis.

Valor Absoluto

Em vez dos desvios quadráticos, é muito comum o uso do valor absoluto:

$$D_{i,i'} = \sum_{j=1}^p w_j |x_{ij} - x_{i'j}|$$

onde os w_j 's representam as ponderações para as variáveis. Os valores mais usados são os $w_j = 1$ ou $w_j = \frac{1}{p}$. Essa medida é conhecida como métrica "**city-block**".

2.5.1.4 Distância de Minkowski

Constitui-se numa generalização da distância Euclidiana e é dada por:

$$D_{i,i'} = \left[\sum_{j=1}^p w_j |x_{ij} - x_{i'j}|^k \right]^{\frac{1}{k}}$$

onde os w_j 's representam as ponderações para as variáveis e k é um inteiro qualquer.

Deve ser observado que se $k = 1$, tem-se o valor absoluto e se $k = 2$ e $w_j = 1$ para $j = 1, 2, 3, \dots, p$, a distância Euclidiana.

2.5.2 Coeficiente Geral de Gower

Esse coeficiente proposto por Gower (1971) é baseado na proporção da variação relativa à maior discrepância possível.

$$D_{ii'} = -\log_{10} \left[1 - \frac{1}{p} \sum_{j=1}^p \frac{|x_{ij} - x_{i'j}|}{\max(x_j) - \min(x_j)} \right]$$

onde temos dentro do somatório um tipo de padronização das variáveis.

2.5.3 Coeficiente de Similaridade de Cattell

O coeficiente de similaridade de Cattell é dado

$$C_{ii'} = \frac{2 \left[p - \frac{2}{3} \right] - D_{ii'}^2}{2 \left[p - \frac{2}{3} \right] + D_{ii'}^2}$$

onde $D_{i,i'}^2$ é a distância Euclidiana com variáveis padronizadas (z_i e $z_{i'}$).

Uma outra forma dessa medida é devida a Cattell, Couter e Comark (1971), onde é feita uma outra correção,

$$C_{ii'} = \frac{\sqrt{2p} - D_{ii'}^2}{\sqrt{2p} + D_{ii'}^2}$$

2.5.4 Coeficiente de Camberra

Fornece a proporção da diferença em relação ao total

$$D_{ii'} = \frac{1}{p} \sum_{j=1}^p \frac{|x_{ij} - x_{i'j}|}{x_{ij} + x_{i'j}}$$

2.5.5 Coeficiente de Correlação

Segundo Messetti (2000), Sokal & Sneath (1963) utilizam como coeficiente de similaridade para caracterizar as relações entre os caracteres, o coeficiente momento produto de Pearson definido por:

$$R_{i,i'} = \frac{\sum_{j=1}^p (x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})}{\sqrt{\left[\sum_{j=1}^p (x_{ij} - \bar{x}_i)^2 \right] \left[\sum_{j=1}^p (x_{i'j} - \bar{x}_{i'})^2 \right]}}$$

onde $\bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij}$ e $\bar{x}_{i'} = \frac{1}{p} \sum_{j=1}^p x_{i'j}$.

2.5.6 Coeficientes de dissimilaridade ou critério de semelhança

A escolha do critério para avaliar a distância entre os objetos é um conceito fundamental para a técnica de análise de agrupamento. Este critério quantifica o quanto os objetos são semelhantes ou dessemelhantes.

No coeficiente de similaridade quanto mais semelhantes são os indivíduos maior é o valor do coeficiente; já no coeficiente de dissimilaridade quanto menos semelhantes os indivíduos maior o valor do coeficiente.

Como exemplo de coeficiente de parença, pode-se citar o coeficiente de correlação como sendo uma medida de similaridade, já que, quanto maior for o coeficiente de correlação entre dois grupos de dados, mais parecidos serão eles. A distância euclidiana, por sua vez, é uma medida de dissimilaridade, pois quanto maior for a distância euclidiana entre dois objetos, mais dissimilares eles serão.

Essa distinção, no entanto, não é fundamental, na medida em que é possível, de maneira geral, estabelecer uma medida de dissimilaridade a partir de uma de similaridade e vice-versa.

A similaridade é uma medida de igualdade entre dois objetos. Tipicamente a similaridade entre dois objetos x e y assume valores entre 0 e 1, onde 0 expressa que dois objetos não são similares, enquanto que 1 expressa máxima similaridade. Geralmente são consideradas as seguintes condições para definir a similaridade.

- * $0 \leq s(x, y) \leq 1$
- * $s(x, y) = 1$
- * $s(x, y) = s(y, x)$

A dissimilaridade é a medida de diferença entre dois objetos, onde existem várias maneiras possíveis de se obter essa medida. Quando o objeto é o agrupamento de dados, as medidas são baseadas no coeficiente de correlação de Pearson onde é medido o nível de relacionamento entre duas variáveis (AZAMBUJA, 2005).

2.6 Classificação das Técnicas de Agrupamento

Existe uma série de técnicas encontradas na literatura que requerem do usuário um prévio conhecimento das propriedades dos algoritmos disponíveis

para a escolha da técnica de agrupamento, onde muitos algoritmos têm sido propostos para a análise de agrupamento.

Para o início do processo são definidos os elementos desejados para a aplicação da análise, bem como os critérios que irão definir as semelhanças entre eles.

Os dados são dispostos em uma matriz, onde as colunas representam as variáveis de interesse e as linhas representam os elementos.

Quando as variáveis não estão na mesma unidade de medida existe a necessidade de uma transformação (padronização) a fim de que elas contribuam igualmente para a similaridade entre os indivíduos.

Frei (2006) coloca que a padronização evita que as unidades escolhidas para mensurar as variáveis possam afetar a similaridade entre os objetos. Assim, as variáveis contribuem, de forma mais igualitária, para a similaridade entre os objetos.

A padronização se dá pela subtração da média, de cada observação, dividido pelo respectivo desvio-padrão.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

em que x_j e s_j representam, respectivamente, a média e o desvio-padrão dos valores da variável j .

Segundo Mingoti (2005) as técnicas de agrupamento ou cluster são frequentemente classificadas em dois tipos: técnicas hierárquicas e não-hierárquicas, sendo que as hierárquicas são classificadas em aglomerativas e divisivas.

2.6.1 Técnica Hierárquica

Os grupos construídos de forma hierárquica, produzem uma estrutura do tipo árvore.

Everitt et al. (2001) salientam que os métodos hierárquicos dão forma à espinha dorsal da análise de agrupamento na prática. Estão eles amplamente disponíveis e de fácil utilização nos softwares estatísticos. As escolhas que o investigador necessita fazer no decorrer da análise são: a medida de proximidade, o método de agrupamento e, freqüentemente, o número de agrupamentos.

2.6.1.1 Técnica hierárquica aglomerativa

Considera-se cada objeto como sendo o grupo inicial, e por meio de uniões sucessivas, chega-se a um único grupo, ou seja, todos os objetos pertençam a um grupo ao final do algoritmo.

A técnica do método hierárquico aglomerativo é de fácil entendimento e pode ser representada através de dendogramas, em que o primeiro passo é calcular as medidas de dissimilaridade (ou similaridade) entre todos os pares possíveis de elementos e, assim, formar os grupos por processos aglomerativos ou divisivos.

Aldenderfer e Blashfield (1984) relatam que, no ano de 1973, de todos os seus artigos publicados usando a análise de agrupamento, dois terços deles usavam o método aglomerativo hierárquico pelo fato de ser um método de mais fácil compreensão e simples ligação.

2.6.1.1.1 Vizinho mais Próximo (Single Linkage)

Esse método tem seu procedimento iniciado com a procura dos dois objetos mais similares na matriz de similaridade, ou seja, a distância entre os grupos é definida como sendo aquela entre os objetos mais parecidos entre esses grupos.

2.6.1.1.2 Vizinho mais distante (Complete Linkage)

Após agrupar os dois indivíduos mais semelhantes, de menor distância, verifica-se a distância máxima deste primeiro grupo para os objetos restantes (FREI, 2006).

2.6.1.1.3 Distância média (Average Linkage)

Este método consiste em agrupar os dois objetos mais semelhantes e na seqüência utilizar a média aritmética das distâncias dos objetos de cada grupo para confeccionar a nova matriz de distâncias.

Utiliza-se a similaridade média dos indivíduos ou grupo que se pretende unir a um grupo já existente.

2.6.1.2 Técnica hierárquica divisiva

Considera-se um grande grupo que contém todas as observações e a partir de separações sucessivas, as observações mais diferentes entre si são separadas e transformadas em grupos menores, reproduz-se o procedimento até cada elemento ficar isolado num único agrupamento.

Meyer (2002) afirma que todos os indivíduos estão num mesmo grupo, que se divide em um ou mais grupos, os quais se subdividem sucessivamente até o final do processo.

Aldenderfer & Blashfield (1984) definem que ao contrário dos métodos aglomerativos hierárquicos os métodos iterativos e divisivos não foram muito usados nem examinados pelo fato de não serem tão compreendidos.

2.6.2 Técnica não hierárquica

Nesses métodos o usuário tem que ter especificado previamente o número “k” de grupos desejado. A partição de n elementos em k grupos (cluster) é efetuada diretamente na matriz de dados.

O algoritmo mais utilizado na literatura é o método das k-médias onde cada elemento amostral é recolocado a menor distância cujo vetor de médias é o mais próximo do vetor de valores observados para o respectivo elemento.

Frei (2006) define o processo de k-médias como:

- Separar os n objetos em k grupos, de forma aleatória;
- Calcular os centróides (médias) de cada grupo;
- Percorrer o conjunto de objetos, associando-o ao agrupamento cujo centróide está mais próximo e recalculando o centróide do agrupamento que recebe o novo objeto e do agrupamento que perdeu o objeto.
- Repetir o passo anterior até que nenhuma reassociação tenha lugar.

2.6.3 Técnicas de validação dos agrupamentos

A validação do resultado de um agrupamento, em geral, é feita com base em índices estatísticos, que julgam, de maneira qualitativa, o mérito das estruturas encontradas.

Um índice quantifica alguma informação a respeito da qualidade de um agrupamento, no qual sua aplicação é dada pelo critério de validação (FACELI e CARVALHO e SOUTO, 2005).

Com este objetivo algumas técnicas são propostas.

2.6.3.1 Gráfico de silhueta

O gráfico de silhueta mede a qualidade dos *cluster* com base na proximidade entre os padrões de um *cluster* e na distância dos padrões de um *cluster* ao mais próximo. As silhuetas mostram quais padrões estão bem situados dentro dos seus *clusters* e quais estão fora de um *cluster* apropriado.

Sunaga (2006) considera um objeto i pertencente ao grupo A . Então a dissimilaridade média de i em relação a todos os outros objetos de A pode ser denotada por $a(i)$. Considerando um grupo diferente C calcula-se então, a dissimilaridade média de i em relação a todos os objetos de C , que será denotada por $d(i, C)$. Após calcular $d(i, C)$ para todos os grupos $C \neq A$, seleciona-se a menor delas, aqui denotada por $b(i)$, conforme a fórmula:

$$b(i) = \min d(i, C), C \neq A$$

Este número representa a dissimilaridade de i em relação ao seu grupo vizinho. Assim o valor da silhueta no ponto i é dado por:

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}$$

A silhueta é seguida de $(-1 \leq s(i) \leq 1)$. Se o valor da silhueta é próximo a 1, significa que os objetos foram bem agrupados. Se o valor da silhueta é 0, significa que um outro grupo próximo deve ser a melhor opção. A função objetivo representa a média de $s(i)$ para $i = 1, 2, \dots, N$. Nesse caso, o melhor número de grupos ocorre quando o valor de $s(i)$

médio for máximo.

2.6.3.2 Gráfico de perfil

Barroso e Artes (2003) apresentam o gráfico de perfil como procedimento para verificar a qualidade dos agrupamentos formados e enriquecer a interpretação. No eixo das abscissas indicam-se as variáveis e no eixo das ordenadas, as escalas de medidas. Cada média é representada por um ponto nos eixos cartesianos. Unindo-se os pontos, observam-se os perfis dos grupos.

2.7 Outras Aplicações de Análise de Agrupamento

Cruz (1990) enfatiza três estudos de técnicas multivariadas, como procedimentos eficientes no melhoramento de plantas onde se pode verificar que as técnicas de análise de agrupamento, análise de componentes principais e análise de variáveis canônicas são eficazes para predizer fenômenos genéticos baseando-se simultaneamente em um complexo de variáveis de importância genética. Estes procedimentos levam a um enriquecimento das informações disponíveis, fazendo com que seja permitida uma melhor identificação de unidades seletivas e também uma eficiência em estudos de predições genéticas.

Bernardes (1998) trabalhou com dados diários de chuva para 154 postos de períodos heterogêneos para o Estado do Paraná, onde seu objetivo principal foi determinar regiões pluviométricamente homogêneas para um período de 23 anos. Para as análises, foram utilizadas técnicas da estatística convencional (média, desvio padrão, coeficiente de variação) e da análise multivariada (análise de agrupamento). Os resultados de

todas as etapas foram mapeados utilizando recursos do geoprocessamento. Foram encontradas cinco regiões pluviometricamente homogêneas, que foram comparadas com as áreas de concentração das culturas de soja, trigo, milho e feijão e tiveram seu perfil pluviométrico descrito em relação às épocas de recomendação de semeadura e desenvolvimento dessas culturas.

Nery, Baldo e Martins (1998) estudaram 26 séries de temperatura mínima, com dados diários, obtidos junto ao Instituto Agrônômico do Paraná (IAPAR), Londrina, Pr, onde utilizaram o método multivariado de agrupamentos, para caracterização dessas temperaturas no Estado do Paraná. Na obtenção do corte do dendograma (determinação dos grupos homogêneos), buscou-se gerar grupos homogêneos utilizando o relevo do estado. Foram obtidos ainda, algumas estatísticas (média, desvio-padrão, coeficiente de variação) para entender os processos geradores desses grupos. Os resultados apresentaram cinco áreas homogêneas de temperatura mínima, sendo que as regiões central e sul do estado mostraram as menores temperaturas para o período estudado.

Fonseca e Silva (1999), com o objetivo de testar as técnicas de análise multivariada e da medida de divergência genética representada pela distância generalizada de Mahalanobis na seleção de descritores e na identificação de duplicidades de acessos de feijão (*Phaseolus vulgaris* L.), utilizaram cinquenta acessos do Banco Ativo de Germoplasma (BAG-Feijão), da Embrapa - Centro Nacional de Pesquisa de Arroz e Feijão (CNPAP). Esses acessos foram avaliados em junho de 1993, utilizando-se delineamento experimental em blocos casualizados, com duas repetições. Dez descritores com características quantitativas e fenológicas foram analisados por meio de variáveis canônicas e distância de Mahalanobis. Todos os caracteres foram importantes na descrição do germoplasma. A técnica de agrupamento pela distância generalizada de Mahalanobis mostrou-se viável e eficaz na identificação de duplicidades do feijoeiro, podendo ser utilizada rotineiramente no Banco de Germoplasma.

Scheeren et al. (2000) desenvolveram um trabalho com o objetivo de agrupamento de unidades amostrais similares em razão dos fatores físicos e dos nutrientes do solo, das acículas e da serapilheira de um povoamento de *Araucária angustifolia* (Bert.) O. Ktze. Isso foi realizado mediante a comparação sistemática de uma distância estatística e da obtenção de funções discriminantes, que permitiram classificar novas unidades amostrais nos

grupos formados, para a Floresta Nacional de Canela (RS). Foram utilizadas as técnicas de análise estatística multivariada de análise de cluster e análise discriminante. Por meio de análise de cluster, tomando-se por base uma matriz de doze casos (árvores) por 84 colunas (variáveis), foram obtidos três grupos, sendo estes distintos da classificação natural de sítios. Pela análise discriminante, foram determinadas duas funções discriminantes, que tomaram por base oito variáveis brutas, como suficientes para classificar novos indivíduos nos grupos obtidos pela análise de cluster.

Benin et al. (2003), com o objetivo de estabelecer relação entre as diferentes técnicas multivariadas utilizadas em estimativas de divergência genética, trabalharam com um experimento no ano de 2001, quando doze cultivares de aveia foram testados para sete caracteres agrônômicos, utilizando-se o delineamento experimental de blocos casualizados com quatro repetições. A distância Euclidiana e de Mahalanobis evidenciaram baixa correlação (0,529) e, quando utilizadas na confecção de dendogramas, não possibilitaram a formação de agrupamentos similares. As análises gráficas por componentes principais e variáveis canônicas também apresentaram dispersões distintas. Foi possível a identificação de genótipos dissimilares e com média elevada, que poderão ser utilizados com grande probabilidade de sucesso em hibridações dirigidas em aveia.

Diniz et al. (2003) utilizaram dados de temperatura com o objetivo de determinar as regiões homogêneas de temperaturas máxima e mínima para o Estado do Rio Grande do Sul. Para atingir este objetivo, foram aplicados quatro métodos de agrupamento a um conjunto de dados de temperaturas máxima e mínima de 40 estações meteorológicas do Rio Grande do Sul. O método adotado foi o da ligação completa com a obtenção de 4 regiões representativas dos pontos de vista climatológico e fisiológico do Estado. A climatologia das regiões foi feita pelo cálculo da média das séries temporais das estações contidas em cada região homogênea. Este procedimento permitiu, assim, obter a estrutura térmica regional.

Metz e Monard (2005) apresentaram uma metodologia para análise de clusters em cada nível da hierarquia, com o objetivo de aplicar esta metodologia de explicação dos clusters gerados a partir de algoritmos de AM (Aprendizagem de Máquina) não supervisionado, mais especificamente de clustering hierárquico, para obter uma descrição simbólica dos agrupamentos e, conseqüentemente, facilitar a compreensão e utilização do conhecimento extraído.

Keller Filho et al. (2005) realizaram um estudo com o objetivo de identificar no Brasil, regiões homogêneas quanto à distribuição de probabilidades de chuva e assim contribuir para estudos de riscos climáticos na agricultura. As regiões foram delimitadas mediante aplicação da análise de agrupamento hierárquica, com variáveis classificatórias definidas pela proporção de pântadas secas e por medidas de posição, escala e forma das distribuições de frequências da quantidade de chuva. A análise de agrupamento permitiu identificar 25 zonas pluviometricamente homogêneas em todo o território brasileiro.

Fonseca et al. (2006) avaliaram a divergência genética entre 32 clones de café conilon (*Coffea canephora* Pierre ex Frohener) componentes de três variedades clonais melhoradas, com vistas à identificação dos mais dissimilares, para o estabelecimento de programas de cruzamentos dirigidos. A divergência genética foi avaliada por procedimentos multivariados: distância generalizada de Mahalanobis, método de agrupamento de otimização de Tocher e técnica de variáveis canônicas. Sete caracteres foram avaliados em experimento conduzido em Marilândia, ES. Os genótipos ES 92, ES 25 e ES 22 são os mais divergentes, sendo os dois últimos os mais indicados para cruzamento com os demais, tendo em vista aliarem divergência genética a um bom desempenho produtivo.

Arriel et al. (2006) compararam diferentes técnicas multivariadas na caracterização de 35 genótipos de gergelim mediante 769 marcadores de *random amplified polymorphic DNA* (RAPD). As distâncias genéticas foram obtidas pelo complemento aritmético do coeficiente de Jaccard e agrupadas pelos métodos hierárquicos do vizinho mais próximo, do vizinho mais distante, das médias aritméticas não ponderadas (UPGMA), do método de otimização de Tocher e análises de coordenadas principais. O agrupamento dos genótipos foi alterado em função dos diferentes métodos usados. Adotando-se a mesma distância genética (0,36) como valor de corte, diferenciaram-se quatro grupos no método do vizinho mais próximo, 13 para o vizinho mais distante, 11 no UPGMA e quatro no Tocher. Entre os métodos hierárquicos, o UPGMA apresentou o melhor ajuste das distâncias originais e estimadas (CCC= 0,89).

Albuquerque et al. (2006) propuseram uma sistemática para o estudo e interpretação da estabilidade dos métodos de análise de agrupamento, através de vários algoritmos de agrupamento em dados de vegetação. Utilizaram-se dados provenientes de levantamento na Mata da Silvicultura, da Universidade Federal de Viçosa em Viçosa, MG.

Para a análise de agrupamento, foram estimadas as matrizes de distância de Mahalanobis com base nos dados originais e via reamostragem “bootstrap”, bem como aplicados os métodos da ligação simples, ligação completa e média das distâncias, do centróide, da mediana e de Ward. Para a detecção de associação entre os métodos, foi aplicado o teste Qui-Quadrado (χ^2) a 1 e 5% de probabilidade. Para os diversos métodos de agrupamento foi obtida a correlação cofenética. Os resultados de associação dos métodos foram semelhantes. Verificou-se também que os métodos são coincidentes, exceto os métodos do centróide e Ward e os métodos do centróide e mediana, em comparação com o de Ward, respectivamente, com base nas matrizes de Mahalanobis a partir dos dados originais e “bootstrap”. A sistemática proposta é promissora para o estudo e interpretação da estabilidade dos métodos de análise de agrupamento.

Messetti (2007) avaliou a divergência genética de 12 populações de girassol do Banco de Germoplasma da EMBRAPA /Soja de Londrina por meio de técnicas multivariadas. O modelo experimental constituiu-se do delineamento em blocos casualizados envolvendo 12 variedades de girassol avaliadas sob cinco caracteres morfoagronômicos. Com base nas estimativas da distância Mahalanobis e distância euclidiana foi realizada a análise de agrupamento adotando-se três algoritmos hierárquicos. Para determinar o número de grupos adotou-se o dendograma, a análise do nível de fusão e a análise do comportamento de similaridade. Pode-se concluir pela existência da divergência genética, detectando-se quatro grupos geneticamente diferentes e caracterizados pelos escores médios.

2.8 Resumo da revisão de literatura

QUADRO 1 – Resumo da Revisão de Literatura

Análise	Autor	Cultura	Observações
	Sokal & Sneath 1963		Principles of Numerical Taxonomy
	Cormack 1971		Coefficiente de similaridade de Cattel

Análise	Autor	Cultura	Observações
	Gower 1971		Coefficient of similarity
Análise Multivariada	Anderberg 1973		
	Maxwell 1977		Multivariate analysis in behavioral research
	Doorenbos & Kassam 1979		
	Mardia et al. 1979		Multivariate analysis
	Cattell 1980		Coeficiente de similaridade de Cattell
Análise Multivariada	Aldenderfer & Blashfiel 1984		
	Ramos & Milho Mundstock 1986		
Agrupamentos	Cruz 1990	Milho	
	Bussab 1990		Introdução à análise de agrupamento
	Maluf 1992	Milho	
	Fornasieri Filho 1992	Milho	
	Gonçalves et al. 1997		Correlação linear entre os níveis de risco de geada
	Bernardes 1998	Soja, trigo, milho e feijão	
	Nery, Baldo & Martins 1998		Método multivariado de agrupamento
	Aurélio et al. 1999		Análise de distância entre pares de objetos
	Fonseca & Silva 1999	Feijão	Dist. Mahalanobis
	Lozada & Angeloni 1999	Milho	Uso de graus-dia
	Messetti 2000	Girassol	Dist. Mahalanobis Técnicas hierárquicas
	Meyer 2000	Milho	Coefic. Associação Técnicas hierárquicas
	Scheeren et al. 2000	Araucária angustifolia (Bert.)	Análise de Cluster Análise Discriminante

Análise	Autor	Cultura	Observações
	Gadionli et al. 2000	Milho	
	Gerage 2000	Milho	
	Everitti et al. 2001		Cluster Analysis
	Sans et al. 2001	Milho	
	Meyer, 2002	Milho	Coefic. de Jaccard Coefic. de Sorensen-Dice Coefic. Anderberg e Ochiai
Regressão Linear Múltipla	Gonçalves 2002	Milho	
	Peixoto 2002	Milho	
	Benin et. al. 2003	Aveia	Dist. Euclideana Dist. Mahalanobis
	Diniz et al. 2003		Método da Ligação Completa
	Barroso & Artes 2003		Gráfico de Perfil
	Kim et al 2003		Probabilidade de ocorrência de geadas
	Oliveira 2003	Milho	
	Assad et al. 2004	Café	
	Ayoade 2004		Introdução a Climatologia
	Cardoso et al. 2004	Milho	
	Gonçalves 2004	Milho	Distância euclidiana
	Azambuja, 2005		Algoritmos e Indicadores de Similaridade
	Faceli, Carvalho & Souto 2005		Validação de um agrupamento
	Metz & Monard 2005		Clustering hierárquico
	Thadeu Filho et al. 2005		
Análise Multivariada	Hair et al. 2005		
	Mingoti 2005		Análise de dados através de métodos de estatística multivariada
	Magalhães & Lima 2005		Noções de probabilidade e estatística

Análise	Autor	Cultura	Observações
	Keller Filho et al. 2005		Análise de agrupamento hierárquica
	Rosa Neto 2006		Algoritmos aglomerativos para análise de agrupamento
	Assis et al. 2006	Milho	
	Fonseca et al. 2006	Café conilon	Dist Mahalanobis Método de Tocher Téc. Variáveis Canônicas
	Arriel et al. 2006	Gergelim	Coefficiente de Jaccard Método vizinho mais próximo, vizinho mais distante UPGMA Método de Tocher Análise de coordenadas principais
	Albuquerque et al. 2006	Mata da Silvicultura	Dist. Mahalanobis Ligação simples, Ligação completa e médias das distâncias, da mediana e de Ward
	Gonçalves et al. 2006	Milho	
	Sunaga 2006		Método de k-média Validação estatística
Análise de Agrupamento	Frei 2006		Introdução a análise de agrupamento
	Streck & Alberto 2006	Trigo, soja e milho	
Regressão Linear Múltipla	Cargnelutti et al. 2006		
	Messetti 2007	Girassol	Dist. Mahalanobis Dist. Euclidiana
	SEAB – Pr 2007	Milho	

3 MATERIAL E MÉTODOS

3.1 Material

Com o objetivo de diagnosticar quais áreas aptas ao cultivo do milho para o risco de geada e melhorar a compreensão da variabilidade agrometeorológica fornecendo informações gerais, em nível anual e por decênio na cultura do milho safrinha onde trabalhou-se com uma série de 21 anos sendo utilizadas 22 estações agrometeorológicas do IAPAR, identificando os períodos extremos de temperatura para o período homogêneo de 1986 a 2006, de forma a conhecer os limites de variação temporal.

O fato de estar trabalhando apenas com 22 estações se justifica por ter esta uma série completa de 21 anos, já que outras estações teriam sido desativadas e mesmo criadas recentemente, inviabilizando a análise dos dados em uma série temporal que contemple as mesmas estações.

Os dados de temperatura máxima e temperatura mínima foram fornecidos pelo IAPAR-Londrina, onde foram calculados os graus dias por meio da expressão

$$GD = (T_{\max} + T_{\min})/2 - T_b$$

em que GD são os graus-dia ou unidades térmicas; T_{\max} a temperatura máxima do dia; T_{\min} a temperatura mínima do dia, e T_b a temperatura base igual a 10°C, que é a condição necessária da quantidade diária de calor para que o milho safrinha floresça.

Através de simulação observou-se também a contagem de um intervalo de 5 dias após a semente ter sido lançada no solo e, a partir daí, pode-se medir quantos dias foram necessários para o florescimento do milho obedecendo o total da soma de no mínimo 800° dia para o seu florescimento de acordo com fórmula para o cálculo dos graus-dia e literatura.

O procedimento para se verificar a soma dos 800° dias foi feito medindo-se a temperatura média diária de uma série de 21 anos ocorrida em cada estação. Em seguida foram realizadas 10 simulações sendo que a 1ª iniciada em 06 de janeiro conforme (Quadro 2). Para cada estação as datas variavam conforme o florescimento. Os 30 e 60 dias foram divididos em decêndios dos quais tirou-se a média do risco diário (média decencial).

A partir do momento em que o milho atingiu 800° C de calor pôde-se verificar qual foi exatamente a data de florescimento, obedecendo as datas de simulação e, a partir daí, verificando o período crítico onde mediu-se a temperatura diária para diagnosticar quais são as regiões não favoráveis ao cultivo do milho safrinha, ou seja, regiões cuja temperatura esteja igual ou menor que 1° C, sendo esta temperatura não favorável ao cultivo do milho de acordo com a literatura.

O risco médio de geada foi obtido de forma decencial, ou seja, a condição necessária para diagnosticar-se o risco de haver em um período de dez dias, no mínimo, temperatura que satisfizesse esta condição ($\geq 1^\circ \text{C}$).

QUADRO 2 – Risco médio (%) de geada do florescimento ao enchimento de grãos do milho safrinha, semeado em 10 diferentes épocas de 06/jan a 06/abr para 22 estações meteorológicas das regiões no Estado do Paraná.

ESTAÇÕES	6/jan	16/jan	26/jan	5/fev	15/fev	25/fev	7/mar	17/mar	27/mar	6/abr
Bandeirantes	0.00	0.00	0.00	0.00	0.00	0.16	0.87	1.51	1.98	1.90
Bela Vista do Paraíso	0.00	0.00	0.00	0.00	0.00	0.40	1.11	1.59	2.38	1.98
Cambará	0.00	0.00	0.00	0.48	1.11	2.14	2.94	4.68	4.76	3.65
Ibiporã	0.00	0.00	0.00	0.00	0.00	0.16	0.87	1.51	2.46	2.30
Joaquim Távora	0.00	0.00	0.00	0.24	0.87	1.90	2.94	5.56	5.95	5.24
Londrina	0.00	0.00	0.00	0.00	0.00	0.40	1.27	2.94	3.89	3.10
Paranavaí	0.00	0.00	0.00	0.00	0.00	0.32	1.03	2.70	3.57	3.25
Umuarama	0.00	0.00	0.00	0.00	0.00	0.00	1.03	3.17	3.25	3.25
Clevelândia	2.30	7.22	10.79	15.08	22.14	31.90	34.13	22.78	10.63	4.92
Fernandes Pinheiro	0.79	2.06	3.97	6.98	13.02	22.06	28.97	18.97	6.51	3.81
Francisco Beltrão	0.79	0.79	3.73	7.86	10.63	18.17	28.97	31.43	22.30	9.68
Guarapuava	4.60	11.03	15.32	20.24	26.35	32.30	31.27	21.98	10.63	4.52
Guaraqueçaba	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.11	0.24
Laranjeiras do Sul	0.00	0.16	1.90	3.65	7.30	14.05	21.35	21.27	15.32	6.90
Morretes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nova Cantu	0.00	0.00	0.00	0.63	1.75	4.76	7.54	12.54	14.29	11.27
Palmas	6.67	15.32	19.84	27.62	40.40	44.05	29.68	10.08	2.54	0.00
Palotina	0.00	0.00	0.16	1.67	3.17	6.90	12.06	22.86	31.11	30.00
Pato Branco	0.00	1.35	3.17	7.38	13.10	26.19	30.63	17.94	5.71	1.51
Pinhais	1.51	6.11	9.52	14.52	20.95	28.17	25.32	12.54	4.05	1.35
Planalto	0.87	0.87	2.46	7.14	9.76	14.52	20.56	25.24	29.37	24.92
Telemaco Borba	0.00	1.03	4.60	6.03	10.24	15.95	21.27	25.71	18.41	6.67

Fonte: Iapar - Londrina

3.2 Métodos

A fim de verificar a existência de diferenças de temperatura entre as regiões, bem como as diferentes datas de simulação, um estudo preliminar foi feito através de

uma análise exploratória multivariada dos dados para se obter medidas descritivas de posição da estrutura multivariada, representadas pelos vetores de médias, referente a 10 simulações para as 22 estações meteorológicas.

3.2.1 Medidas descritivas para amostras multivariadas

A teoria necessária para o desenvolvimento do método está baseada na literatura de Mingoti (2005), onde o estudo apresenta medidas descritivas para amostras multivariada dos dados. Da análise exploratória constarão as medidas descritivas de posição e variabilidade da estrutura multivariada representadas pelos vetores de médias necessários para o desenvolvimento do método.

Medidas de posição para um conjunto de dados: Considere uma variável X com observações representadas por x_1, x_2, \dots, x_n . A média desse conjunto é a soma dos valores dividida pelo número total de observações, isto é:

$$\bar{X}_{\text{obs}} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Vetor de médias amostrais. O vetor de médias μ será estimado pelo vetor \bar{X} de médias amostrais, em que \bar{X}_j representa a média da j -ésima variável com $j = 1, 2, \dots, p$ dada por:

$$\bar{X}_{p \times 1} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \dots \\ \bar{X}_p \end{bmatrix} \quad \text{sendo} \quad \bar{X}_j = \sum_{j=1}^p X_p \quad j = 1, 2, \dots, p.$$

3.2.2 Análise de Agrupamento

A análise de agrupamento (cluster analysis) é um termo usado para descrever diversas técnicas numéricas, cujo objetivo fundamental é classificar os valores de uma matriz de dados, sob estudos, em grupos na conversão em uma matriz de dissimilaridade.

A técnica classificatória pode ser utilizada quando se deseja explorar as similaridades entre indivíduos ou entre variáveis definindo-os em grupos. No procedimento todas as variáveis são medidas em cada indivíduo e em todos os indivíduos para cada variável. A similaridade entre elementos é obtida através de coeficientes específicos para cada tipo de variável (discreta, contínua, binária, etc).

3.2.2.2 Algoritmo de agrupamento

Algoritmo é uma seqüência de passos que permite que um problema possa ser resolvido no computador de maneira automática e repetitiva. É preciso também, definir como os dados serão processados e armazenados no computador, onde sua solução está baseada em dois pontos: a seqüência de passos e a forma como os dados serão armazenados.

Uma coleção de diferentes algoritmos que agrupam objetos é chamada análise de agrupamento.

Um bom algoritmo de agrupamento caracteriza-se pela produção de classes de alta qualidade, nas quais a similaridade intra-classe é alta e a inter-classe é baixa. A qualidade do resultado depende do método utilizado para medir esta similaridade, normalmente baseado em análise de distância entre pares de objetos, e da habilidade em descobrir algum ou todos os padrões escondidos (AURÉLIO, 1999).

Para a formação dos grupos na classificação em métodos hierárquicos e não hierárquicos são utilizados os algoritmos. Os métodos hierárquicos aglomerativos são formados através de uma matriz de similaridade, onde se procuram os dois objetos mais

semelhantes, a partir daí o par de objeto é agrupado formando um único indivíduo. Uma nova matriz de similaridade é formada onde se identifica novamente o par mais semelhante para formar um novo grupo. O procedimento se repete até que todos os objetos fiquem em um só grupo.

Na formação dos conglomerados serão necessários alguns métodos de maior utilização pelos softwares estatísticos e também pela sua diferença quanto à função usada para o cálculo repetido dos coeficientes de similaridade entre novos agrupamentos.

Para formalizar esta etapa considera-se os agrupamentos l e k contendo r_l e r_k elementos, em que esses números são maiores ou, no mínimo, iguais a 1. Se os agrupamentos l e k se unem, isto é indicado como (l, k) com $r_{l,k} = r_l + r_k$ indivíduos.

Para aplicar a técnica de análise de agrupamento será utilizado o método hierárquico aglomerativo que tem como função operar a matriz de similaridade, à partir de cada objeto como sendo grupo inicial.

Os algoritmos descritos abaixo foram utilizados para a formação dos agrupamentos.

3.2.2.2.1 Método do vizinho mais próximo (“Single Linkage Method” – SLM)

Neste método as ligações entre elementos (estações) e grupos ou entre grupos são feitas por ligações simples entre pares de objetos.

Uma população P_j candidata-se a um agrupamento quando apresenta uma distância a este agrupamento igual à sua menor distância com relação aos membros do agrupamento. A distância entre dois agrupamentos l e k será dada por:

$$d_{l,k} = \min_{\substack{l \in L \\ k \in K}} d_{l,k}$$

3.2.2.2.2 Método do vizinho mais distante (“Complete Linkage Method” – CLM)

Bussab et al (1990) enfatizam que a similaridade entre dois grupos é definida como aquela apresentada pelos elementos de cada grupo que menos se parecem, ou seja, formam-se todos os pares com um membro de cada grupo, e a similaridade entre os grupos é definida pelo par que menos se parece.

Uma população P_j candidata-se a um agrupamento quando apresenta uma distância em relação a este agrupamento igual à sua maior distância com relação aos membros do agrupamento. A distância entre dois agrupamentos l e k será dada por:

$$d_{L,K} = \max_{\substack{l \in L \\ k \in K}} d_{l,k}$$

3.2.2.2.3 UPGMA (Unweighted Pair-Group Method Using Arithmetic Averages)

Método não ponderado de agrupamento aos pares por médias aritméticas.

Este método consiste em atribuir pesos iguais a cada indivíduo do grupo e calcula a similaridade média de um indivíduo que pretende se juntar ao grupo já existente.

Define-se a distância entre duas populações como a média entre os valores individuais das populações de um dos grupos com os de outro grupo.

$$d_{l,k} = \frac{1}{r_l r_k} \sum_{\substack{l \in L \\ k \in K}} d_{l,k}$$

3.3 Definição do número de grupos

É de grande importância saber como se deve proceder para a escolha do número g de grupos para definir a partição do conjunto de dados analisados.

Ao usarmos diferentes algoritmos temos o surgimento de diferentes agrupamentos. Isso leva a conclusão de que não existe qualquer procedimento de seleção padrão e objetivo. Para definir o número de grupos foi adotado como critério, o bom senso.

A seguir decidiu-se trabalhar com dendograma para obter os agrupamentos.

3.3.1 Dendograma

É a forma gráfica mais usada para representar o resultado final dos diversos agrupamentos. Resultante de uma análise estatística de determinados dados pode-se representá-lo na técnica hierárquica aglomerativa onde assemelha-se aos ramos de uma árvore que se vão dividindo em sucessivos ramos.

A escala vertical indica o nível de similaridade, quanto mais elevado o índice mais heterogênea as partes agrupadas. No eixo horizontal, são marcadas as estações meteorológicas em uma ordem conveniente das regiões.

3.3.2 Análise do comportamento do nível de fusão

À medida que o procedimento passa do estágio k para $(k+1)$, e a similaridade entre os grupos que estão sendo agrupados vai decrescendo e, conseqüentemente, a distância entre os grupos vai aumentando, pode-se apresentar um gráfico do número de grupos “versus” o nível de distância (fusão) do agrupamento em cada estágio do processo. Se

visualmente ocorrem “*pontos de salto*”, isto indica o momento ideal de parada do algoritmo, isto é, o número de grupos final. Quando o segmento de reta no gráfico se estabiliza, ou não apresenta variações significativas, tem-se a partição ótima. Caso existam vários pontos de salto, sugere-se aplicar outro procedimento (MINGOTTI, 2005).

3.4 Validação e Interpretação dos agrupamentos

Para o contexto de agrupamento, quando se tem uma solução “ótima”, o pesquisador deve ter muito cuidado na validação e na garantia de significância prática da solução final. A validação é a garantia de que a solução de agrupamento seja representativa e que os grupos realmente sejam diferentes entre si.

3.4.1 Gráfico de Perfil

Na validação dos resultados do agrupamento foi utilizado o gráfico de perfil onde se verificou a qualidade dos agrupamentos na interpretação dos resultados.

A análise do comportamento se deu pela média de temperatura de cada agrupamento ao longo de toda a série.

3.4.2 Programas Computacionais

Para o cumprimento de todas as etapas deste trabalho na análise dos dados foram utilizados os seguintes softwares.

Microsoft Office Excel 2003 – Análises univariadas; cálculos de médias.

Statistica 6.0 – Análise de agrupamento; dendogramas para os três algoritmos da técnica hierárquica aglomerativa.

Tabwin32 - Versão 2.1.

4 RESULTADOS E DISCUSSÃO

Com base nas estimativas das distâncias euclidianas entre as 22 estações, a partir dos dados da Tabela 1, foi realizada a análise de agrupamentos pelos três algoritmos da técnica hierárquica. A utilização dessas distâncias e esses algoritmos justificam-se por serem estes muito utilizados em trabalhos científicos na área agronômica envolvendo diversas culturas.

Nesta etapa, foi efetuado um estudo do risco de geada em 22 estações agrometeorológicas do Estado do Paraná, onde se trabalhou com três algoritmos da técnica hierárquica de agrupamento através de dendogramas com o objetivo de averiguar as regiões mais propícias ao risco de geadas no Paraná e, portanto, menos indicadas para o cultivo do milho safrinha.

4.1 Método do Vizinho mais Próximo

Observa-se na Figura 1, com o corte na distância euclidiana verificou-se a formação de 4 grupos de estações agrometeorológicas quanto ao risco de geadas no período de 06 de Janeiro a 06 de Abril para as 22 estações agrometeorológicas do Paraná.

No grupo (1), encontram-se as cidades de Nova Cantu, Joaquim Távora, Cambará, Morretes, Guaraqueçaba, Umuarama, Paranaíba, Londrina, Ibiporã, Bela Vista do Paraíso e Bandeirantes.

No grupo (2), encontram-se as cidades de Telêmaco Borba, Laranjeiras do Sul, Francisco Beltrão, Pinhais, Pato Branco, Fernandes Pinheiro, Guarapuava e Clevelândia.

No grupo (3), encontram-se as cidades de Planalto e Palotina.

No grupo (4), encontra-se a estação de Palmas, considerada a região mais fria do Estado do Paraná.

Pela sua localização geográfica, embora as cidades de Morretes e Guaraqueçaba estejam localizadas no litoral, seu clima se assemelha ao norte e noroeste do estado, fato este, que se encontra no primeiro grupo.

Fazendo-se o mapeamento do dendograma forma-se a Figura 2 onde se pode ter uma melhor visualização das regiões com climas homogêneos.

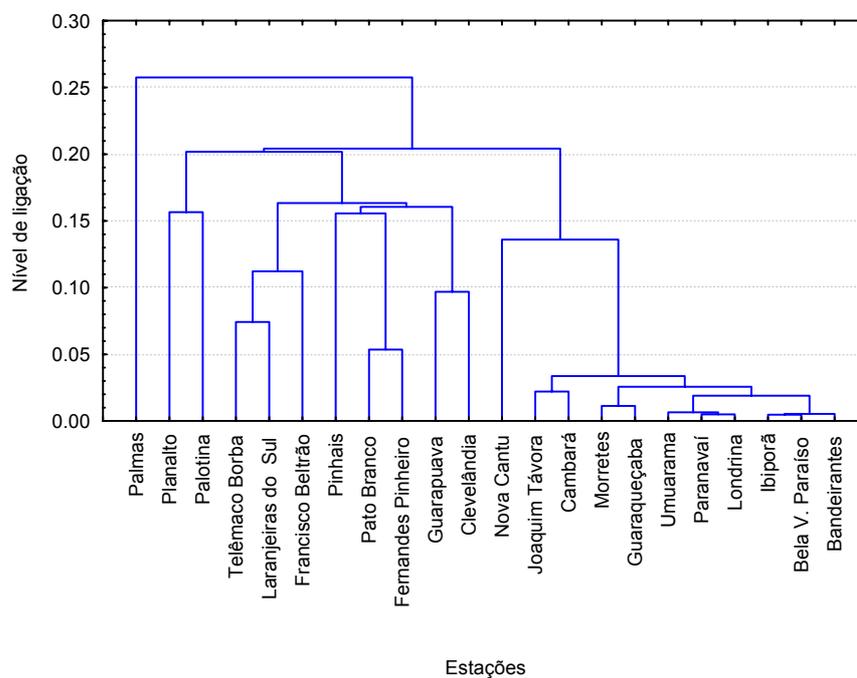


Figura 1 – Dendrograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar, extraído dos dados originais pelo “Single Linkage Method” – Método do vizinho mais próximo.

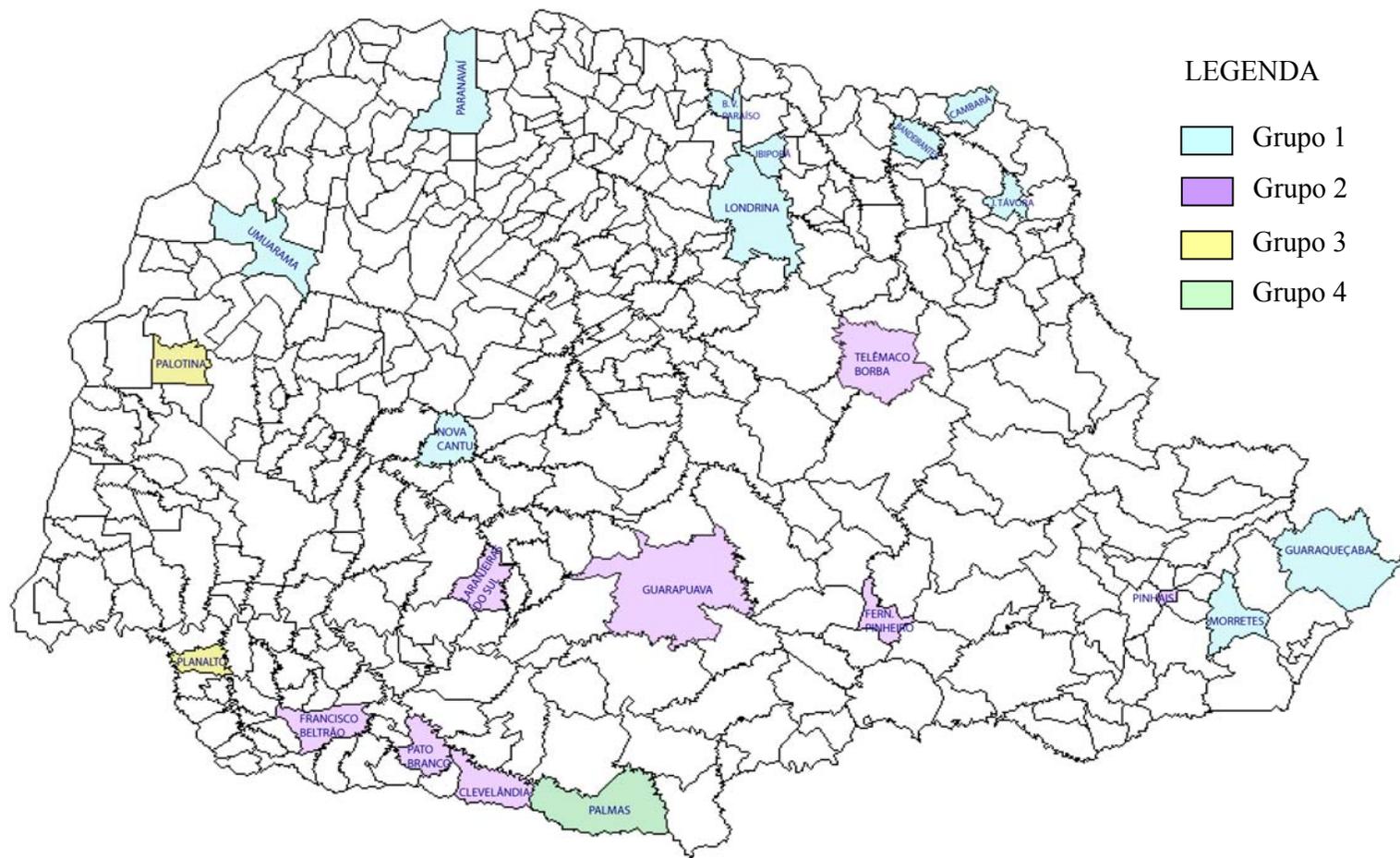


Figura 2 – Mapeamento das 22 estações, utilizando o método do vizinho mais próximo.

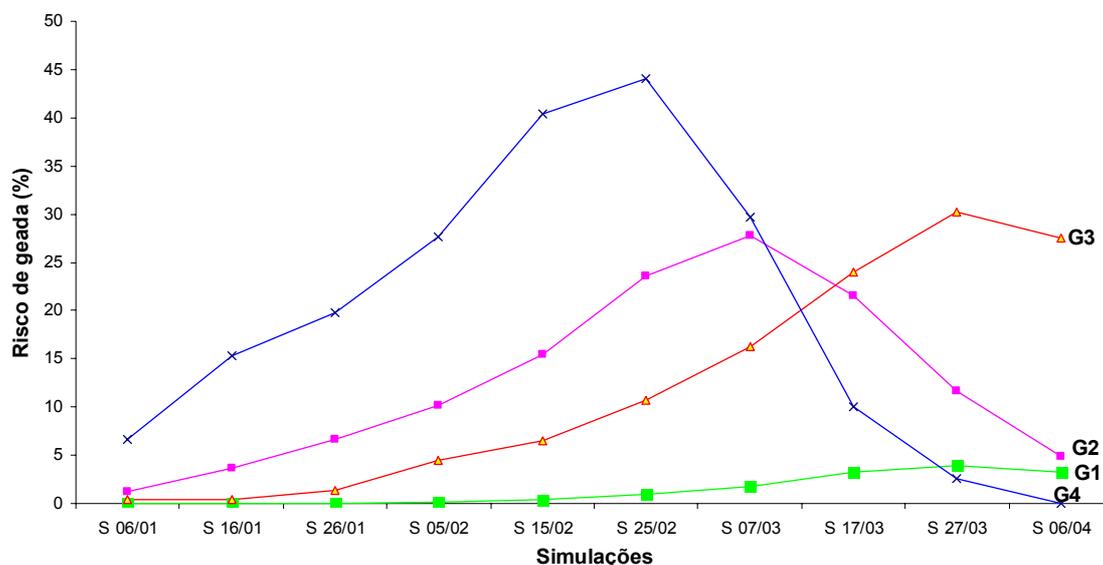


Figura 3 – Gráfico de perfil para o vizinho mais próximo.

Para a validação dos resultados do agrupamento utilizou-se o gráfico de perfil que apresenta a média dos grupos obedecendo as dez simulações para as 22 estações agrometeorológicas, no qual pode-se verificar no agrupamento (4) que a data de semeadura com maior risco ocorre em 25/02.

Outro aspecto interessante é que a data de semeadura em 06/01 oferece um risco mínimo para os agrupamentos (1), (2) e (3), isto é, para o agricultor, a melhor data para o plantio. Com isso, pode-se concluir que quanto maior a demora no plantio maior será o risco de geada.

O agrupamento (3) diferencia-se do agrupamento (2) e (4), bem como do agrupamento (1).

O agrupamento (2) tem comportamento semelhante ao agrupamento (3) que são as cidades de Palotina e Planalto, com uma oscilação de mais ou menos 20% menor e com um descêndio de atraso. Observou-se, também, que nos dois últimos descêndios no grupo (3) aumentou o risco de geadas.

No agrupamento (4) composto apenas pela cidade de Palmas, pode-se observar que, embora seja uma região mais fria, existe um pico na série, nos períodos entre a quinta e sexta simulação, voltando a série com um comportamento que se assemelha as demais regiões, embora com temperaturas mais baixas.

4.2 Método do Vizinho mais Distante

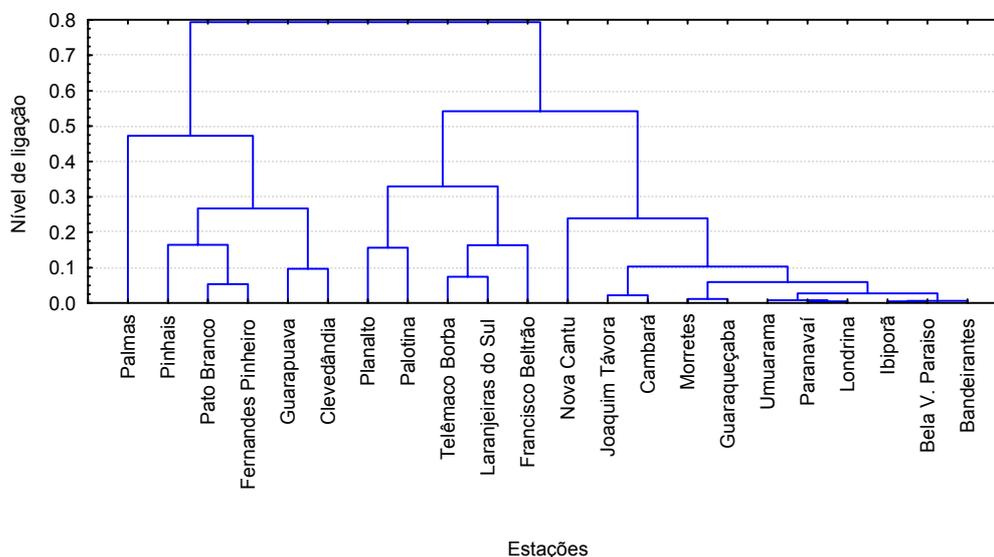


Figura 4 – Dendrograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar extraído dos dados originais pelo “Complete Linkage Method” – Método do vizinho mais distante.

Observa-se, como no algoritmo anterior, que o algoritmo do vizinho mais distante, também agrega a formação de 4 grupos de estações agrometeorológicas quanto ao risco de geadas no período de 06 de Janeiro a 06 de Abril para as 22 estações agrometeorológicas do Paraná.

Já no grupo (1), temos as cidades de Nova Cantu, Joaquim Távora, Cambará, Morretes, Guaraqueçaba, Umuarama, Paranaíba, Londrina, Ibiporã, Bela Vista do Paraíso e Bandeirantes.

No grupo (2), temos as cidades de Pinhais, Pato Branco, Fernandes Pinheiro, Guarapuava e Clevedândia, cujas temperaturas são semelhantes.

No grupo (3), observam-se as cidades de Planalto, Palotina, Telêmaco Borba, Laranjeiras do Sul e Francisco Beltrão.

No grupo (4), encontra-se a estação de Palmas, considerada a região mais fria do Estado do Paraná, e cuja disposição no dendrograma também ficou isolada das demais regiões devido às temperaturas mais baixas.

Fazendo-se o mapeamento do dendrograma forma-se a Figura 4 onde se tem uma melhor visualização das regiões com climas homogêneos.

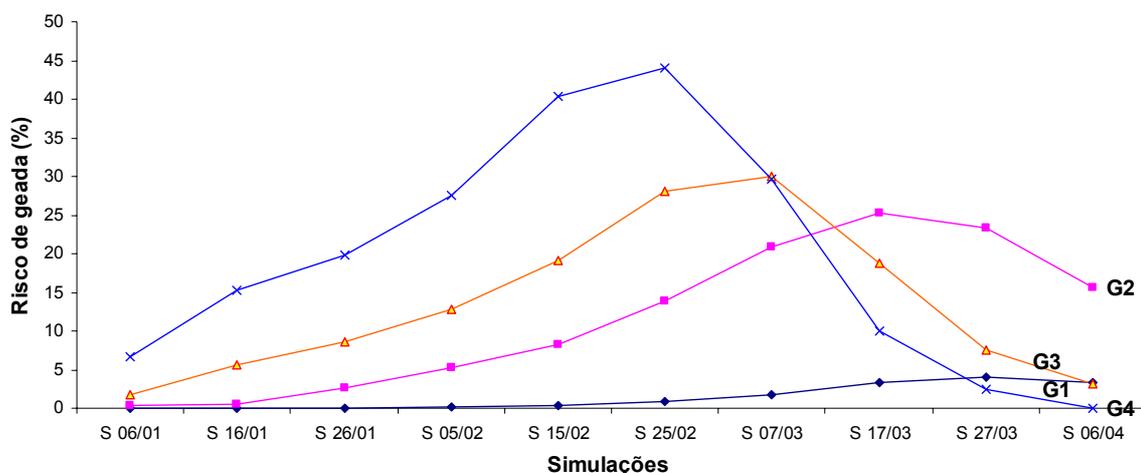


Figura 6 – Gráfico de perfil para o vizinho mais distante.

Conforme o período de semeadura observou-se que o gráfico de perfil demonstrou o comportamento dos agrupamentos nas dez simulações realizadas.

Para o agrupamento (1) as regiões agrupadas foram as mesmas que a do algoritmo do vizinho mais próximo.

Vale ressaltar que fazem parte deste agrupamento as cidades de Morretes e Guaraqueçaba, e embora estas regiões estejam situadas mais ao litoral do estado, se assemelham com as demais regiões ao norte e noroeste do estado.

Os agrupamentos (2), (3) e (4) até a quinta simulação, apresentam a mesma tendência no risco da geada.

No entanto, observa-se no agrupamento (2), que a cidade de Pinhais, embora esteja contida no agrupamento (2), se localiza mais para o sul do estado e no agrupamento (3), a cidade de Telêmaco Borba se distancia das demais regiões que fazem parte do agrupamento. Isso se verifica perfeitamente no mapa do estado do Paraná.

No agrupamento (4), formado pela cidade de Palmas, por ser considerada a cidade mais fria do Estado, pode-se observar que o maior pico na série foi na sétima simulação em 07/03 diferente do algoritmo do vizinho mais próximo onde o maior pico se deu na sexta simulação em 25/02.

4.3 Método não ponderado de agrupamento aos pares por médias aritméticas

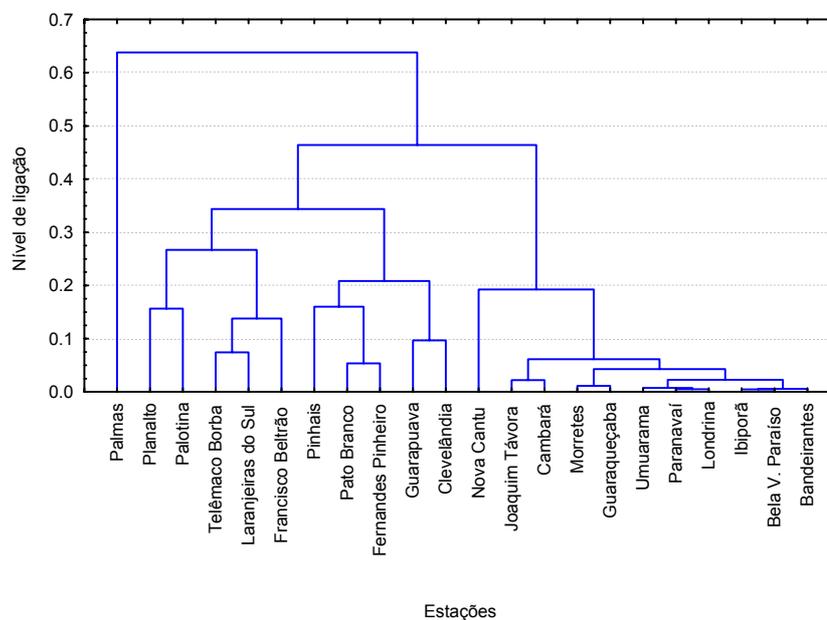


Figura 7 – Dendrograma resultante da análise de agrupamento das 22 estações agrometeorológicas do Iapar obtido dos dados originais pelo “UPGMA” (Unweighted Pair – Group Method Using Arithmetic Averages Linkage Method) – Método não ponderado de agrupamento aos pares por médias aritméticas.

No algoritmo não ponderado de agrupamento aos pares por médias aritméticas pode-se observar a inclusão de um quinto grupo de estações agrometeorológicas quanto ao risco de geadas, no período de 06 de Janeiro a 06 de Abril para as 22 estações agrometeorológicas do Paraná.

Já no grupo (1), encontram-se as cidades de Nova Cantu, Joaquim Távora, Cambará, Morretes, Guaraqueçaba, Umuarama, Paranavaí, Londrina, Ibiporã, Bela Vista do Paraíso e Bandeirantes. Neste último grupo há de se observar a semelhança com o quarto grupo do algoritmo do vizinho mais próximo.

No grupo (2), observam-se as cidades de Pinhais, Pato Branco, Fernandes Pinheiro, Guarapuava e Clevelândia.

O grupo (3) é composto pelas cidades de Telêmaco Borba, Laranjeiras do Sul e Francisco Beltrão.

No grupo (4), encontram-se as cidades de Planalto e Palotina com temperaturas semelhantes.

No grupo (5), assim como nos demais algoritmos, encontramos a cidade de Palmas separada das demais cidades, indicando uma região com temperaturas muito diferentes das demais regiões.

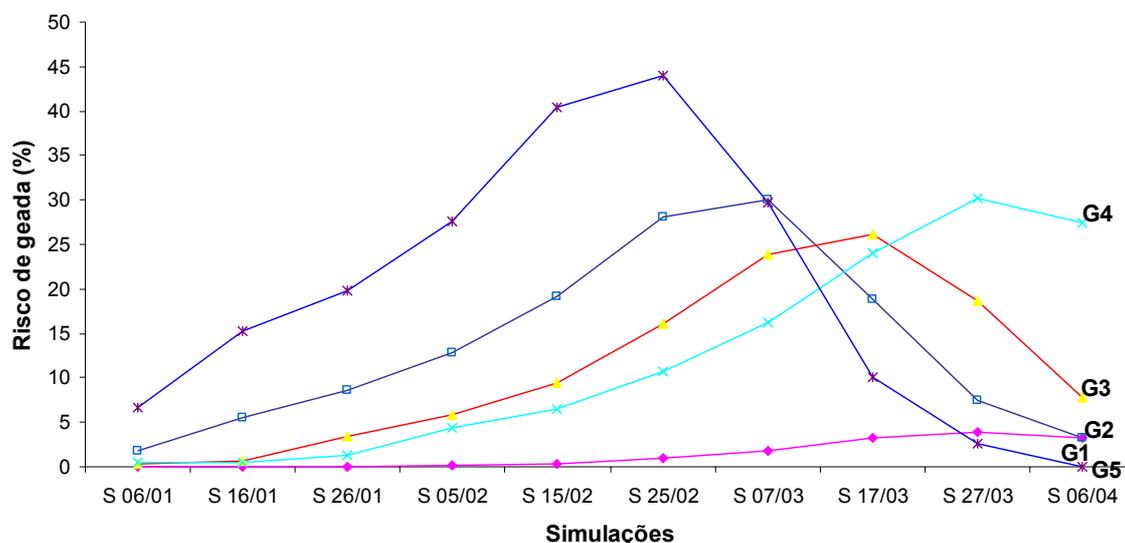


Figura 9 – Gráfico de perfil para o método não ponderado de agrupamento aos pares por médias aritméticas.

Assim como nos algoritmos do vizinho mais próximo e vizinho mais distante no agrupamento (1) participam as mesmas cidades. Pode-se observar que o comportamento desta série praticamente não se altera até a quarta simulação, indicando um bom período para o cultivo do milho safrinha.

Os agrupamentos (2) e (3), também se assemelham ao algoritmo do vizinho mais distante, ficando inalterada a série até a terceira simulação em 26/01.

Há de se observar que no algoritmo do método não ponderado houve a formação de um quinto agrupamento constituído pela cidade de Palmas.

Na sexta e sétima simulação houve um pico no risco médio de geada chegando a quase 20%. Observou-se, também, que só depois da oitava simulação em 17/03 as temperaturas ficaram mais próximas aos demais agrupamentos.

4.4 Comportamento dos agrupamentos com relação aos três algoritmos

Em relação aos três algoritmos é possível observar que o grupo (1) manteve-se inalterado, constituído com as mesmas cidades.

Já no grupo (2), verifica-se que o algoritmo do vizinho mais próximo consegue agregar também as cidades que integram o grupo (3) do algoritmo do método ponderado de agrupamento aos pares de médias aritméticas que são as cidades de Telêmaco Borba, Laranjeiras do Sul e Francisco Beltrão.

No grupo (3) do algoritmo do vizinho mais próximo o agrupamento se deu pela cidade de Planalto e Palotina que é semelhante ao grupo (4) do método não ponderado de agrupamento aos pares por médias aritméticas.

Pode-se constatar ainda que, o único agrupamento que ficou isolado foi composto pela cidade de Palmas identificada pelo agrupamento (4) nos algoritmos do vizinho mais próximo e do vizinho mais distante, onde, a mesma cidade é classificada como grupo (5) no algoritmo do método não ponderado de agrupamento aos pares por médias aritméticas.

CONCLUSÕES

Com a aplicação da análise de agrupamento considerando os métodos do vizinho mais próximo, vizinho mais distante e método não ponderado de agrupamento aos pares por médias aritméticas aos dados de temperatura das 22 estações agrometeorológicas, pode-se diagnosticar diferenças entre os agrupamentos.

Na análise de agrupamento os gráficos multivariados auxiliam na interpretação e definição final quanto ao número de grupos, indicando no agrupamento do método não ponderado aos pares por médias aritméticas o isolamento da cidade de Palmas.

Para a variabilidade temporal considerando as dez simulações verifica-se que a cidade de Palmas se diferencia das demais, apresentando sempre os maiores percentuais do risco de geada justificando assim a necessidade de um plantio mais cedo para se ter um menor risco de geada e conseqüentemente uma maior produção.

Em todos os métodos de agrupamento, as cidades ao norte do Estado são mais propícias ao cultivo do milho safrinha, sendo portanto as mais ao norte e as cidades do litoral foram classificadas no mesmo grupo, denominado de grupo 1 (G1). De

acordo com os gráficos de perfis este grupo mantém sempre o menor risco de geada em todas as datas.

Para a cidade de Palmas, que ficou sozinha em um grupo no resultado da análise de agrupamento evidencia-se que a antecipação do plantio do milho se faz necessário distanciando-se do inverno pelo fato de se ter um menor risco de geada, isto se deve pela quantidade de calor acumulada para o seu florescimento, que são 800° dias, condição esta que se pode obter em épocas de plantio cuja temperatura se encontre mais elevada.

Como demonstrado pelo gráfico de perfil, pode-se diagnosticar todo o comportamento dos grupos ao longo das dez simulações.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, M. A.; FERREIRA, R. L. C.; SILVA, J. A. A.; SANTOS, E. S.; STOSIC, B.; SOUZA, A. L. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. **Rev. Árvore**, Viçosa-MG, v.30, n.2, p.257-265, 2006.

ALDENDERFER, M. S.; BLASHFIEDL, R. K. **Series: Quantitative Applications in the Socia Sciences. Califórnia: University of Iowa, 1984. 88 p.**

ANDERBERG, M. R. **Cluster analysis for applications.** London: Academic Press, 1973. 359 p.

ARRIEL, N. H. C.; MAURO, A. O. D.; MAURO, S. M. Z. D.; BAKKE, O. A.; TREVISOLI, S. H. U.; COSTA, M. M.; CAPELOTO, A.; CORRADO, A. R. Técnicas multivariadas na determinação da diversidade genética em gergelim usado marcadores RAPD. **Pesq. Agropec. Bras**, Mai. 2006, v. 41, n.5, p. 801 – 809.

ASSAD, E. D.; PINTO, H. S.; JUNIOR, J. Z.; ÁVILA, A. M .H. Impacto das mudanças climáticas no zoneamento agroclimático do café no Brasil. **Pesq. Agropec. Bras**, Brasília, v. 39, n.11, p. 1057-1064, nov. 2004.

ASSIS, J. P.; NETO, D. D.; REICHARDT, K.; MANFRON, P. A.; MARTIN, T. N.; BONNECARRÈRE, R. A. G. Dados climáticos simulados e produtividade potencial do milho. **Pesq. Agrop. Bras.**, Brasília, v.41, n.5, p. 731-737, Mai. 2006.

AURÉLIO, M.; VELASCO, M.; LOPES, C. H. Descoberta de conhecimento e mineração de dados. 1999. Apostila – Laboratório de Inteligência Computacional Aplicada (IAC) – Departamento de Engenharia Elétrica PUC, Rio de Janeiro.

AYOADE, J. O. Introdução à Climatologia para os Trópicos. 10ª edição, berthand Brasil, 2004.

AZAMBUJA, S. **Estudo e implementação da análise de agrupamento em ambientes virtuais de aprendizagem.** Rio de Janeiro, 2005, 212 p. Dissertação (Mestrado em Informática – Núcleo de computação eletrônica) – Universidade Federal do Rio de Janeiro.

BARROSO, L. P.; ARTES, R. **Análise multivariada:** Minicurso do 10 Simpósio de estatística Aplicada A experimentação Agronômica – RBRAS, 48 reunião Anual da região Brasileira da Sociedade Internacional de biometria – SEAGRO. Lavras: UFLA, 2003.

BENIN, G.; CARVALHO, F. I. F.; OLIVEIRA, A. C.; MARCHIORO, V. S.; LORENCETTI, C.; KUREK, A. J.; SILVA, J. A. G.; CRUZ, P. J.; HARTWIG, I.; SCHMIDT, D. A. M. Comparação entre medidas de dissimilaridade e estatísticas multivariadas como critérios no direcionamento de hibridações em aveia. Ciênc. . Rural. Santa Maria, Jul-ago, 2003,, v.33, n.4, p. 657 – 662.

BERNARDES, L. R. M. **Determinação de regiões pluviométricamente homogêneas no Estado do Paraná, através de técnicas de análise multivariada.**

São Paulo, 1998, 136 p. Tese (Doutorado em Engenharia – área de concentração: engenharia de transportes) – Escola Politécnica da Universidade de São Paulo.

BUSSAB, W. O.; ANDRADE, D.F.; MYAZAKY, E.S. Introdução à análise de agrupamentos. **Associação Brasileira de Estatística**, 1990. 105p.

CARDOSO, C. O. ; FARIA, R. T.; FOLEGATTI, M. V. Simulação do rendimento e riscos climáticos para o milho safrinha em Londrina – PR, utilizando o modelo Ceres-Maize. **Rev. Agríc. Jaboticabal**, v.24, n.2, p.291-300, mai/ago. 2004.

CARGNELUTTI FILHO, A. C.; MALUF, J. R. T.; MATZENAUER, R.; STOLZ, A. P. Altitude e coordenadas geográficas na estimativa da temperatura mínima média decendial do ar no estado do Rio Grande do Sul. *Pesq. Agropec. Bras.*, Brasília, Jun. 2006, v. 41, n.6.

CATTELL, R. B. The screen test for the number of factor. **Multivariate analysis**. Chapman & Hall, 1980.

CORMACK, R. A review of classification. **Journal of the Royal Statistical Society**. Serie A, n. 134, p. 321-367, 1971.

CRUZ, C. D. **Aplicações de algumas técnicas multivariadas no melhoramento de plantas**. São Paulo, 1990. 188 p. Tese (Doutorado em agronomia – área de concentração: genética e melhoramento de plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

DINIZ, G. B.; BERLATO, M. A.; CLARKE, R. T.; FONTANNA, D. C. Identificação de regiões homogêneas de temperaturas máxima e mínima do Rio Grande do Sul. **Rev. Bras. Agro.**, Santa Maria, v.11, n.2, p. 303 – 312, 2003.

DOORENBOS, J. ; KASSAM, A. H.; **Yield response to water**. Roma: FAO, 1979. 179 p. (FAO. Irrigação and Drainage Paper, 33).

EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster Analysis**. Institute of Psychiatry, kings College London, Fourth Edition, 2001, 229 p.

FACELI, K.; CARVALHO, A.C. P. L. F.; SOUTO, M. C. P.: Validação de algoritmos de agrupamento. Relatórios técnicos do ICMC, São Carlos – SP, Fev. 2005. nº 254.

FONSECA, A. F. A.; SEDIYAMA, T.; CRUZ, C. D.; SAKAIYAMA, N. S.; FERRÃO, M. A. G.; FERRÃO, R. G.; BRAGANÇA, S. M. Divergência genética em café conilon. **Pesq. Agropec. Bras.**, Abr. 2006, v.41, n.4, p. 599-605

FONSECA, J. R.; SILVA, H. T. Identificação de duplicidades de acessos de feijão por meio de técnicas multivariadas. **Pesq. Agropec. Bras.**, Mar. 1999, Vol.34 n.3.

FORNASIERI FILHO, D. **A cultura do milho**. Jaboticabal: FUNEP, 1992. 273p.

FREI, F. Introdução à análise de agrupamentos. Editora UNESP, 2006.

GADIOLI, J. L.; NETO, D. D.; GARCIA, AL G.; BASANTA, MA. D. V.; Temperatura do ar, rendimento de grãos de milho e caracterização fenológica associada à soma calórica. **Scientia agrícola**. Piracicaba, v. 57, n.3, p. 377-383, Jul/Set. 2000.

GERAGE, A. C. A Cultura do Milho no Paraná. In: **Milho – Estratégias de Manejo Para a Região Sul**, P. 23-32, 2000.

GONÇALVES, S. L. **Riscos de deficiência hídrica e épocas de semeadura de milho (*Zea Mays L.*) na região dos Campos Gerais do Paraná**. Maringá, 2004, 87 p. Tese (Doutorado em Agronomia - UEM – Universidade Estadual de Maringá).

GONÇALVES, S. L.; BERTONHA, A. ; CARAMORI, P. H.; GONÇALVES, A. C. A. Risco de deficiência hídrica e épocas de semeadura de milho na região dos Campos Gerais do Paraná. **Revista Bras. Agrometeorologia**, Piracicaba – SP, 2006.

GONÇALVES, S. L., CARAMORI, P.H., WREGE. M. S., SHIOGA. P., GERAGE. A. C., Épocas de semeadura do milho “safrinha”, no Estado do Paraná, com menores riscos climáticos. **Revista Acta Scientiarum Agronomy**. Maringá, v. 24, n.5, p. 1287-1290, 2002.

GONÇALVES, S. L.; WREGE, M. S.; CARAMORI, P. H.; MARIOT. E. J.; NETO, M. A. Probabilidade de ocorrência de temperaturas superiores a 30° C no florescimento do feijoeiro (*Phaseolus vulgaris L.*), cultivado na safra das águas no Estado do Paraná. **Rev. Bras. Agrometeorologia**, Santa Maria, v.5, n.1, p. 99-107, 1997.

GOWER, J. C. A general coefficient of similarity and some of its properties. **Biometrics**, v.27, p.857-74, 1971.

HAIR, J. F. Jr.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. Análise multivariada de dados. 5ª edição, Editora BOOKMAN, 2005.

KELLER FILHO, T. K.; ASSAD, E. D.; LIMA, R. S. R. Regiões pluviometricamente homogêneas no Brasil. **Pesq. Agro. Bras.** Vol. 40 n. 4 Brasília, Abr. 2005.

IBGE ; PR mantém liderança na produção de grãos, diz IBGE. 2006, Disponível em: <
<http://www.tribunadonorte.com/?page=teste&dados=aWRfbm90cz0xNjI3MSZwYWdpbmE9UEFSQU7B>> Acesso em: 10 de Agosto de 2007.

LOZADA, B. I.; ANGELOCCI, L. R.; Efeito da temperatura do ar e da disponibilidade hídrica do solo na duração de subperíodos e na produtividade de um híbrido de milho (*Zea mays*). **Revista Bras. Agrometeorologia**, Santa Maria, v.7, n.1, p. 37-43, 1999.

KIM, I.; GRODZKI, L. ; CARAMORI. P. H.; VISSOTTO, S. Análise da ocorrência de geadas no Estado do Paraná. **Rev. Bras. Agro.**, Santa Maria, v. 11, n.1, p. 99 – 106, 2003.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. London: Chapman and Hall, 1979. 164 p.

MAXWELL, A. E. **Multivariate analysis in behavioral research**. London: Chapman and Hall, 1977. 164 p.

MESSETTI, A. V. L.; **Estudo da semelhança de genótipos de girassol (*Helianthus annuus* L.) com o uso da distância generalizada de Mahalanobis na análise de agrupamento**. Botucatu - SP: 2000, 86 p. Dissertação (Mestrado em agronomia – área de concentração: energia na agricultura) – Universidade Estadual Paulista “Júlio de Mesquita Filho”.

MESSETI, A. V. L.; **Utilização de técnicas multivariadas na avaliação da divergência genética de populações de girassol (*Helianthus annuus L.*)** Botucatu – SP: 2007. 87 p. Tese (Doutorado em agronomia – área de concentração: energia na agricultura) – Universidade Estadual Paulista “Júlio de Mesquita Filho”.

METZ, J.; MONARD, M. C. XXX Congresso da Sociedade Brasileira de Computação. Clustering hierárquico: uma metodologia para auxiliar na interpretação dos clusters. São Leopoldo-RS, 2005.

MEYER, A. S., **Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes.** Piracicaba – SP: 2002. 106 p. (Dissertação em agronomia – área de concentração: Estatística e Experimentação Agronômica) – ESALQ/USP.

MINGOTI, S. S. Análise de dados através de métodos de estatística multivariada. Editora UFMG, 2005.

NERY, J. T.; BALDO, M. C.; MARTINS, M. L. O. F. Variabilidade da temperatura mínima no Estado do Paraná. **Rev. Bras. Agrometeorologia**, 1998, v.6, n.2, p. 207 – 212

OLIVEIRA, E. F., Fatores que garantem o sucesso da safrinha de milho. VI Seminário Nacional de Milho Safrinha, IAPAR, Londrina, 2003.

PEIXOTO, C. M.; Safrinha programada garante o sucesso. Informativo Pioneer, 2002, ano VIII, n. 15.

RAMOS, L. R. M.; MUNDSTOCK, C. M. Acúmulo de material seca em três híbridos de milho no período de formação dos grãos. **Pesquisa Agropecuária Brasileira**; Brasília, v.21, nº 3, p. 265 – 270, 1986.

ROSA NETO, E. A. **Algoritmos aglomerativos para análise de agrupamento: exemplos e aplicações.** Monografia apresentada como parte dos requisitos necessários para conclusão do curso de Matemática – Habilitação Bacharelado. UEL – Universidade Estadual de Londrina. 86 p.

SANS, L. M. A.; ASSAD, E. D., GUIMARÃES, D. P., AVELLAR, G. Zoneamento de riscos climáticos para a cultura de milho na região Centro-Oeste do Brasil e para o estado de Minas Gerais. **Rev. Bras. Agrometeorologia**, Passo Fundo, v.9, n.3, p. 527- 535, 2001.

SCHEEREN, L. W.; GEHRARDT, E. J.; FINGER, C. A. G.; LONGHI, S. J.; SCHNEIDER, P. R. Agrupamento de unidades amostrais de *Araucaria angustifolia* (Bert.) O. Ktze. Em função de variáveis do solo, da serapilheira e das acículas, na região de canela, **Rev. Ciênc. Florestal**, 2000, vol. 10 n. 2, p. 39 – 57.

SEAB – Secretaria da Agricultura do Paraná; Milho safrinha contabiliza prejuízo no Paraná. 2007, Disponível em: http://www2.faepp.com.br/noticias/exibe_noticia.php?id=1060&formatar=impressao > Acesso em: 21 de Agosto de 2007.

STATISTICA FOR WINDOWS. **General conventions and statistics I**, volume I. **Graphics**, volume 2. **Statistics II**, volume 3. StatSoft technical Support. 1994. USA.

STRECK, N. A.; ALBERTO, C. M.; Estudo numérico do impacto da mudança climática sobre o rendimento de trigo, soja e milho. **Pesq. Agropec. Bras.** V.41, n.9, Set. 2006.

SOKAL, R.R.; SNEATH, P.H.A **Principles of Numerical Taxonomy**. San Francisco, W.H. Freeman, 1963. 450 p.

SAMORA, R.; Milho safrinha supera em área a safra normal no Paraná. 2007, Disponível em: < <http://www.globo.com/economia/mat/2007/04/02/295183920.asp> > Acesso em: 15 de Abril de 2007.

SUNAGA, D. Y.; **Aplicação de técnicas de validação estatística e biológica em agrupamento de dados de expressão gênica.** Curitiba – PR: 2006. 228 p. Dissertação (Mestrado em Informática Aplicada) – Pontifícia Universidade Católica do Paraná.

WREGE, M. S; GONÇALVES, S. L.; CARAMORI, P.H.; VASCONCELLOS, M.E.C.; OLIVEIRA, D.; ABUCARUB NETO, M.; BORROZZINO, E.& CAVIGLIONE, J.H. Risco de deficiência hídrica na cultura do feijoeiro durante a safra das águas no Estado do Paraná. **Revista de Agrometeorologia.** Santa Maria, v.5, n.1, p. 51-59, 1998.

APÊNDICE

Datas de Simulação

Bandeirantes			
Semeadura	Florescimento	30d	60d
6/jan	26/fev	28/mar	27/abr
16/jan	6/mar	5/abr	5/mai
26/jan	18/mar	17/abr	17/mai
5/fev	28/mar	27/abr	27/mai
15/fev	8/abr	8/mai	7/jun
25/fev	19/abr	19/mai	18/jun
7/mar	1/mai	31/mai	30/jun
17/mar	16/mai	15/jun	15/jul
27/mar	2/jun	2/jul	1/ago
6/abr	17/jun	17/jul	16/ago

Bela Vista do Paraíso			
Semeadura	Florescimento	30d	60d
6/jan	1/mar	31/mar	30/abr
16/jan	10/mar	9/abr	9/mai
26/jan	21/mar	20/abr	20/mai
5/fev	31/mar	30/abr	30/mai
15/fev	11/abr	11/mai	10/jun
25/fev	22/abr	22/mai	21/jun
7/mar	4/mai	3/jun	3/jul
17/mar	20/mai	19/jun	19/jul
27/mar	5/jun	5/jul	4/ago
6/abr	21/jun	21/jul	20/ago

Cambará			
Semeadura	Florescimento	30d	60d
6/jan	26/fev	28/mar	27/abr
16/jan	8/mar	7/abr	7/mai
26/jan	19/mar	18/abr	18/mai
5/fev	29/mar	28/abr	28/mai
15/fev	9/abr	9/mai	8/jun
25/fev	21/abr	21/mai	20/jun
7/mar	3/mai	2/jun	2/jul
17/mar	18/mai	17/jun	17/jul
27/mar	4/jun	4/jul	3/ago
6/abr	20/jun	20/jul	19/ago

Ibiporã			
Semeadura	Florescimento	30d	60d
6/jan	27/fev	29/mar	28/abr
16/jan	9/mar	8/abr	8/mai
26/jan	19/mar	18/abr	18/mai
5/fev	30/mar	29/abr	29/mai
15/fev	9/abr	9/mai	8/jun
25/fev	19/abr	19/mai	18/jun
7/mar	2/mai	1/jun	1/jul
17/mar	16/mai	15/jun	15/jul
27/mar	2/jun	2/jul	1/ago
6/abr	18/jun	18/jul	17/ago

Joaquim Távora			
Semeadura	Florescimento	30d	60d
6/jan	26/fev	28/mar	27/abr
16/jan	8/mar	7/abr	7/mai
26/jan	19/mar	18/abr	18/mai
5/fev	30/mar	29/abr	29/mai
15/fev	9/abr	9/mai	8/jun
25/fev	21/abr	21/mai	20/jun
7/mar	4/mai	3/jun	3/jul
17/mar	20/mai	19/jun	19/jul
27/mar	7/jun	7/jul	6/ago
6/abr	25/jun	25/jul	24/ago
Paranavaí			
Semeadura	Florescimento	30d	60d
6/jan	26/fev	28/mar	27/abr
16/jan	8/mar	7/abr	7/mai
26/jan	19/mar	18/abr	18/mai
5/fev	29/mar	28/abr	28/mai
15/fev	9/abr	9/mai	8/jun
25/fev	21/abr	21/mai	20/jun
7/mar	3/mai	2/jun	2/jul
17/mar	18/mai	17/jun	17/jul
27/mar	4/jun	4/jul	3/ago
6/abr	20/jun	20/jul	19/ago

Londrina			
Semeadura	Florescimento	30d	60d
6/jan	28/fev	30/mar	29/abr
16/jan	10/mar	9/abr	9/mai
26/jan	20/mar	19/abr	19/mai
5/fev	31/mar	30/abr	30/mai
15/fev	10/abr	10/mai	9/jun
25/fev	22/abr	22/mai	21/jun
7/mar	5/mai	4/jun	4/jul
17/mar	21/mai	20/jun	20/jul
27/mar	8/jun	8/jul	7/ago
6/abr	26/jun	26/jul	25/ago
Umuarama			
Semeadura	Florescimento	30d	60d
6/jan	25/fev	27/mar	26/abr
16/jan	6/mar	5/abr	5/mai
26/jan	16/mar	15/abr	15/mai
5/fev	27/mar	26/abr	26/mai
15/fev	6/abr	6/mai	5/jun
25/fev	18/abr	18/mai	17/jun
7/mar	1/mai	31/mai	30/jun
17/mar	15/mai	14/jun	14/jul
27/mar	1/jun	1/jul	31/jul
6/abr	17/jun	17/jul	16/ago

Clevelândia			
Semeadura	Florescimento	30d	60d
6/jan	14/mar	13/abr	13/mai
16/jan	25/mar	24/abr	24/mai
26/jan	6/abr	6/mai	5/jun
5/fev	19/abr	19/mai	18/jun
15/fev	3/mai	2/jun	2/jul
25/fev	24/mai	23/jun	23/jul
7/mar	20/jun	20/jul	19/ago
17/mar	20/jul	19/ago	18/set
27/mar	14/ago	13/set	13/out
6/abr	31/ago	30/set	30/out

Fernandes Pinheiro			
Semeadura	Florescimento	30d	60d
6/jan	11/mar	10/abr	10/mai
16/jan	22/mar	21/abr	21/mai
26/jan	2/abr	2/mai	1/jun
5/fev	15/abr	15/mai	14/jun
15/fev	28/abr	28/mai	27/jun
25/fev	17/mai	16/jun	16/jul
7/mar	10/jun	10/jul	9/ago
17/mar	6/jul	5/ago	4/set
27/mar	3/ago	2/set	2/out
6/abr	23/ago	22/set	22/out

Francisco Beltrão			
Semeadura	Florescimento	30d	60d
6/jan	3/mar	2/abr	2/mai
16/jan	13/mar	12/abr	12/mai
26/jan	24/mar	23/abr	23/mai
5/fev	5/abr	5/mai	4/jun
15/fev	17/abr	17/mai	16/jun
25/fev	2/mai	1/jun	1/jul
7/mar	23/mai	22/jun	22/jul
17/mar	17/jun	17/jul	16/ago
27/mar	12/jul	11/ago	10/set
6/abr	4/ago	3/set	3/out

Guarapuava			
Semeadura	Florescimento	30d	60d
6/jan	15/mar	14/abr	14/mai
16/jan	27/mar	26/abr	26/mai
26/jan	7/abr	7/mai	6/jun
5/fev	20/abr	20/mai	19/jun
15/fev	5/mai	4/jun	4/jul
25/fev	27/mai	26/jun	26/jul
7/mar	22/jun	22/jul	21/ago
17/mar	21/jul	20/ago	19/set
27/mar	14/ago	13/set	13/out
6/abr	30/ago	29/set	29/out

Guaraqueçaba			
Semeadura	Florescimento	30d	60d
6/jan	25/fev	27/mar	26/abr
16/jan	7/mar	6/abr	6/mai
26/jan	17/mar	16/abr	16/mai
5/fev	29/mar	28/abr	28/mai
15/fev	8/abr	8/mai	7/jun
25/fev	20/abr	20/mai	19/jun
7/mar	3/mai	2/jun	2/jul
17/mar	17/mai	16/jun	16/jul
27/mar	3/jun	3/jul	2/ago
6/abr	18/jul	17/ago	16/set

Laranjeiras do Sul			
Semeadura	Florescimento	30d	60d
6/jan	8/mar	7/abr	7/mai
16/jan	19/mar	18/abr	18/mai
26/jan	30/mar	29/abr	29/mai
5/fev	11/abr	11/mai	10/jun
15/fev	24/abr	24/mai	23/jun
25/fev	9/mai	8/jun	8/jul
7/mar	29/mai	28/jun	28/jul
17/mar	19/jun	19/jul	18/ago
27/mar	10/jul	9/ago	8/set
6/abr	31/jul	30/ago	29/set

Morretes			
Semeadura	Florescimento	30d	60d
6/jan	25/fev	27/mar	26/abr
16/jan	7/mar	6/abr	6/mai
26/jan	17/mar	16/abr	16/mai
5/fev	29/mar	28/abr	28/mai
15/fev	9/abr	9/mai	8/jun
25/fev	20/abr	20/mai	19/jun
7/mar	3/mai	2/jun	2/jul
17/mar	17/mai	16/jun	16/jul
27/mar	2/jun	2/jul	1/ago
6/abr	18/jun	18/jul	17/ago

Nova Cantu			
Semeadura	Florescimento	30d	60d
6/jan	27/fev	29/mar	28/abr
16/jan	10/mar	9/abr	9/mai
26/jan	20/mar	19/abr	19/mai
5/fev	31/mar	30/abr	30/mai
15/fev	11/abr	11/mai	10/jun
25/fev	23/abr	23/mai	22/jun
7/mar	7/mai	6/jun	6/jul
17/mar	23/mai	22/jun	22/jul
27/mar	10/jun	10/jul	9/ago
6/abr	27/jun	27/jul	26/ago

Palmas			
Semeadura	Florescimento	30d	60d
6/jan	19/mar	18/abr	18/mai
16/jan	31/mar	30/abr	30/mai
26/jan	13/abr	13/mai	12/jun
5/fev	26/abr	26/mai	25/jun
15/fev	16/mai	15/jun	15/jul
25/fev	15/jun	15/jul	14/ago
7/mar	19/jul	18/ago	17/set
17/mar	19/ago	18/set	18/out
27/mar	7/set	7/out	6/nov
6/abr	23/set	23/out	22/nov

Palotina			
Semeadura	Florescimento	30d	60d
6/jan	24/fev	26/mar	25/abr
16/jan	6/mar	5/abr	5/mai
26/jan	17/mar	16/abr	16/mai
5/fev	28/mar	27/abr	27/mai
15/fev	7/abr	7/mai	6/jun
25/fev	19/abr	19/mai	18/jun
7/mar	2/mai	1/jun	1/jul
17/mar	19/mai	18/jun	18/jul
27/mar	7/jun	7/jul	6/ago
6/abr	26/jun	26/jul	25/ago

Pato Branco			
Semeadura	Florescimento	30d	60d
6/jan	14/mar	13/abr	13/mai
16/jan	25/mar	24/abr	24/mai
26/jan	6/abr	6/mai	5/jun
5/fev	19/abr	19/mai	18/jun
15/fev	3/mai	2/jun	2/jul
25/fev	24/mai	23/jun	23/jul
7/mar	20/jun	20/jul	19/ago
17/mar	20/jul	19/ago	18/set
27/mar	14/ago	13/set	13/out
6/abr	31/ago	30/set	30/out

Pinhais			
Semeadura	Florescimento	30d	60d
6/jan	17/mar	16/abr	16/mai
16/jan	28/mar	27/abr	27/mai
26/jan	10/abr	10/mai	9/jun
5/fev	23/abr	23/mai	22/jun
15/fev	9/mai	8/jun	8/jul
25/fev	1/jun	1/jul	31/jul
7/mar	30/jun	30/jul	29/ago
17/mar	28/jul	27/ago	26/set
27/mar	22/ago	21/set	21/out
6/abr	8/set	8/out	7/nov

Planalto			
Semeadura	Florescimento	30d	60d
6/jan	26/fev	28/mar	27/abr
16/jan	9/mar	8/abr	8/mai
26/jan	20/mar	19/abr	19/mai
5/fev	31/mar	30/abr	30/mai
15/fev	11/abr	11/mai	10/jun
25/fev	23/abr	23/mai	22/jun
7/mar	8/mai	7/jun	7/jul
17/mar	25/mai	24/jun	24/jul
27/mar	13/jun	13/jul	12/ago
6/abr	2/jul	1/ago	31/ago

Telemaco Borba			
Semeadura	Florescimento	30d	60d
6/jan	6/mar	5/abr	5/mai
16/jan	17/mar	16/abr	16/mai
26/jan	28/mar	27/abr	27/mai
5/fev	9/abr	9/mai	8/jun
15/fev	21/abr	21/mai	20/jun
25/fev	6/mai	5/jun	5/jul
7/mar	25/mai	24/jun	24/jul
17/mar	18/jun	18/jul	17/ago
27/mar	13/jul	12/ago	11/set
6/abr	6/ago	5/set	5/out

Vizinho mais próximo

vizinho mais distante

método não ponderado de agrup.

ESTAÇÕES	Grupo	ESTAÇÕES	Grupo	ESTAÇÕES	Grupo
Bandeirantes	G1	Bandeirantes	G1	Bandeirantes	G1
Bela Vista do Paraíso	G1	Bela Vista do Paraíso	G1	Bela Vista do Paraíso	G1
Cambará	G1	Cambará	G1	Cambará	G1
Ibiporã	G1	Ibiporã	G1	Ibiporã	G1
Joaquim Távora	G1	Joaquim Távora	G1	Joaquim Távora	G1
Londrina	G1	Londrina	G1	Londrina	G1
Paranavaí	G1	Paranavaí	G1	Paranavaí	G1
Umuarama	G1	Umuarama	G1	Umuarama	G1
Clevelândia	G2	Clevelândia	G3	Clevelândia	G2
Fernandes Pinheiro	G2	Fernandes Pinheiro	G3	Fernandes Pinheiro	G2
Francisco Beltrão	G2	Francisco Beltrão	G2	Francisco Beltrão	G3
Guarapuava	G2	Guarapuava	G3	Guarapuava	G2
Guaraqueçaba	G1	Guaraqueçaba	G1	Guaraqueçaba	G1
Laranjeiras do Sul	G2	Laranjeiras do Sul	G2	Laranjeiras do Sul	G3
Morretes	G1	Morretes	G1	Morretes	G1
Nova Cantu	G1	Nova Cantu	G1	Nova Cantu	G1
Palmas	G4	Palmas	G4	Palmas	G5
Palotina	G3	Palotina	G2	Palotina	G4
Pato Branco	G2	Pato Branco	G3	Pato Branco	G2
Pinhais	G2	Pinhais	G3	Pinhais	G2
Planalto	G3	Planalto	G2	Planalto	G4
Telemaco Borba	G2	Telemaco Borba	G2	Telemaco Borba	G3

Risco de geada para as 22 estações agrometeorológicas à partir de simulações pelo método do vizinho mais próximo.

	S 06/01	S 16/01	S 26/01	S 05/02	S 15/02	S 25/02	S 07/03	S 17/03	S 27/03	S 06/04
G1	0.00%	0.00%	0.00%	0.12%	0.34%	0.93%	1.78%	3.29%	3.97%	3.29%
G2	1.25%	3.72%	6.63%	10.22%	15.47%	23.60%	27.74%	21.58%	11.70%	4.92%
G3	0.44%	0.44%	1.31%	4.41%	6.47%	10.71%	16.31%	24.05%	30.24%	27.46%
G4	6.67%	15.32%	19.84%	27.62%	40.40%	44.05%	29.68%	10.08%	2.54%	0.00%

Risco de geada para as 22 estações agrometeorológicas à partir de simulações pelo método do vizinho mais distante.

	S 06/01	S 16/01	S 26/01	S 05/02	S 15/02	S 25/02	S 07/03	S 17/03	S 27/03	S 06/04
G1	0.00%	0.00%	0.00%	0.12%	0.34%	0.93%	1.78%	3.29%	3.97%	3.29%
G2	0.82%	2.06%	2.96%	4.34%	6.69%	9.10%	9.44%	8.07%	6.28%	4.17%
G3	0.89%	2.24%	3.16%	4.73%	7.34%	9.98%	10.45%	8.66%	6.79%	5.41%
G4	0.35%	1.04%	1.98%	3.62%	5.88%	10.25%	13.59%	12.95%	9.97%	7.12%

Risco de geada para as 22 estações agrometeorológicas à partir de simulações pelo método não ponderado de agrupamento aos pares por médias aritméticas.

	S 06/01	S 16/01	S 26/01	S 05/02	S 15/02	S 25/02	S 07/03	S 17/03	S 27/03	S 06/04
G1	0.00%	0.00%	0.00%	0.12%	0.34%	0.93%	1.78%	3.29%	3.97%	3.29%
G2	0.82%	2.06%	2.96%	4.34%	6.69%	9.10%	9.44%	8.07%	6.28%	4.17%
G3	0.89%	2.24%	3.16%	4.73%	7.34%	9.98%	10.45%	8.66%	6.79%	5.41%
G4	0.35%	1.04%	1.98%	3.62%	5.88%	10.25%	13.59%	12.95%	9.97%	7.12%
G5	1.51%	3.99%	6.03%	9.08%	13.41%	18.68%	19.57%	13.64%	7.18%	3.69%

Matriz de dissimilaridade entre estações agrometeorológicas

$$A = \begin{bmatrix} 0,00 & 0,01 & 0,06 & 0,01 & 0,07 & 0,03 & 0,02 & 0,02 & 0,60 & 0,43 & 0,52 & 0,63 & 0,03 & 0,36 & 0,03 & 0,21 & 0,77 & 0,48 & 0,46 & 0,48 & 0,51 & 0,41 \\ 0,01 & 0,00 & 0,05 & 0,00 & 0,07 & 0,02 & 0,02 & 0,02 & 0,59 & 0,43 & 0,51 & 0,62 & 0,03 & 0,35 & 0,04 & 0,20 & 0,77 & 0,47 & 0,45 & 0,48 & 0,50 & 0,41 \\ 0,06 & 0,05 & 0,00 & 0,05 & 0,02 & 0,03 & 0,04 & 0,04 & 0,55 & 0,39 & 0,46 & 0,59 & 0,08 & 0,30 & 0,09 & 0,15 & 0,74 & 0,43 & 0,42 & 0,44 & 0,46 & 0,36 \\ 0,01 & 0,00 & 0,05 & 0,00 & 0,07 & 0,02 & 0,02 & 0,02 & 0,59 & 0,43 & 0,51 & 0,63 & 0,03 & 0,35 & 0,04 & 0,20 & 0,77 & 0,47 & 0,46 & 0,48 & 0,50 & 0,41 \\ 0,07 & 0,07 & 0,02 & 0,07 & 0,00 & 0,05 & 0,05 & 0,05 & 0,55 & 0,38 & 0,45 & 0,59 & 0,10 & 0,30 & 0,10 & 0,14 & 0,74 & 0,41 & 0,42 & 0,45 & 0,44 & 0,35 \\ 0,03 & 0,02 & 0,03 & 0,02 & 0,05 & 0,00 & 0,00 & 0,01 & 0,58 & 0,42 & 0,50 & 0,62 & 0,05 & 0,34 & 0,06 & 0,18 & 0,76 & 0,45 & 0,45 & 0,47 & 0,48 & 0,39 \\ 0,02 & 0,02 & 0,04 & 0,02 & 0,05 & 0,00 & 0,00 & 0,01 & 0,59 & 0,42 & 0,50 & 0,62 & 0,05 & 0,34 & 0,06 & 0,18 & 0,77 & 0,45 & 0,45 & 0,47 & 0,49 & 0,40 \\ 0,02 & 0,02 & 0,04 & 0,02 & 0,05 & 0,01 & 0,01 & 0,00 & 0,59 & 0,42 & 0,50 & 0,62 & 0,05 & 0,34 & 0,06 & 0,19 & 0,77 & 0,45 & 0,45 & 0,47 & 0,49 & 0,40 \\ 0,60 & 0,59 & 0,55 & 0,59 & 0,55 & 0,58 & 0,59 & 0,59 & 0,00 & 0,20 & 0,27 & 0,10 & 0,61 & 0,32 & 0,61 & 0,49 & 0,33 & 0,54 & 0,19 & 0,16 & 0,40 & 0,28 \\ 0,43 & 0,43 & 0,39 & 0,43 & 0,38 & 0,42 & 0,42 & 0,42 & 0,20 & 0,00 & 0,22 & 0,27 & 0,44 & 0,16 & 0,45 & 0,33 & 0,47 & 0,44 & 0,05 & 0,16 & 0,34 & 0,17 \\ 0,52 & 0,51 & 0,46 & 0,51 & 0,45 & 0,50 & 0,50 & 0,50 & 0,27 & 0,22 & 0,00 & 0,33 & 0,54 & 0,16 & 0,54 & 0,35 & 0,58 & 0,33 & 0,24 & 0,33 & 0,20 & 0,11 \\ 0,63 & 0,62 & 0,59 & 0,63 & 0,59 & 0,62 & 0,62 & 0,62 & 0,10 & 0,27 & 0,33 & 0,00 & 0,64 & 0,38 & 0,64 & 0,53 & 0,26 & 0,58 & 0,26 & 0,18 & 0,44 & 0,34 \\ 0,03 & 0,03 & 0,08 & 0,03 & 0,10 & 0,05 & 0,05 & 0,05 & 0,61 & 0,44 & 0,54 & 0,64 & 0,00 & 0,38 & 0,01 & 0,23 & 0,77 & 0,50 & 0,47 & 0,49 & 0,53 & 0,43 \\ 0,36 & 0,35 & 0,30 & 0,35 & 0,30 & 0,34 & 0,34 & 0,34 & 0,32 & 0,16 & 0,16 & 0,38 & 0,38 & 0,00 & 0,38 & 0,20 & 0,60 & 0,31 & 0,20 & 0,29 & 0,24 & 0,07 \\ 0,03 & 0,04 & 0,09 & 0,04 & 0,10 & 0,06 & 0,06 & 0,06 & 0,61 & 0,45 & 0,54 & 0,64 & 0,01 & 0,38 & 0,00 & 0,24 & 0,77 & 0,51 & 0,47 & 0,49 & 0,54 & 0,44 \\ 0,21 & 0,20 & 0,15 & 0,20 & 0,14 & 0,18 & 0,18 & 0,19 & 0,49 & 0,33 & 0,35 & 0,53 & 0,23 & 0,20 & 0,24 & 0,00 & 0,72 & 0,28 & 0,37 & 0,42 & 0,31 & 0,25 \\ 0,77 & 0,77 & 0,74 & 0,77 & 0,74 & 0,76 & 0,77 & 0,77 & 0,33 & 0,47 & 0,58 & 0,26 & 0,77 & 0,60 & 0,77 & 0,72 & 0,00 & 0,79 & 0,45 & 0,32 & 0,67 & 0,57 \\ 0,48 & 0,47 & 0,43 & 0,47 & 0,41 & 0,45 & 0,45 & 0,45 & 0,54 & 0,44 & 0,33 & 0,58 & 0,50 & 0,31 & 0,51 & 0,28 & 0,79 & 0,00 & 0,48 & 0,54 & 0,16 & 0,31 \\ 0,46 & 0,45 & 0,42 & 0,46 & 0,42 & 0,45 & 0,45 & 0,45 & 0,19 & 0,05 & 0,24 & 0,26 & 0,47 & 0,20 & 0,47 & 0,37 & 0,45 & 0,48 & 0,00 & 0,16 & 0,38 & 0,21 \\ 0,48 & 0,48 & 0,44 & 0,48 & 0,45 & 0,47 & 0,47 & 0,47 & 0,16 & 0,16 & 0,33 & 0,18 & 0,49 & 0,29 & 0,49 & 0,42 & 0,32 & 0,54 & 0,16 & 0,00 & 0,43 & 0,29 \\ 0,51 & 0,50 & 0,46 & 0,50 & 0,44 & 0,48 & 0,49 & 0,49 & 0,40 & 0,34 & 0,20 & 0,44 & 0,53 & 0,24 & 0,54 & 0,31 & 0,67 & 0,16 & 0,38 & 0,43 & 0,00 & 0,22 \\ 0,41 & 0,41 & 0,36 & 0,41 & 0,35 & 0,39 & 0,40 & 0,40 & 0,28 & 0,17 & 0,11 & 0,34 & 0,43 & 0,07 & 0,44 & 0,25 & 0,57 & 0,31 & 0,21 & 0,29 & 0,22 & 0,00 \end{bmatrix}$$

MATERIAL DIDÁTICO PARA PROCEDIMENTO DOS TRÊS ALGORITMOS DO MÉTODO HIERÁRQUICO AGLOMERATIVO (Frei, 2006. 111 p.)

Técnica hierárquica aglomerativa

Considera-se cada objeto como sendo o grupo inicial, e por meio de uniões sucessivas, chega-se a um único grupo, ou seja, todos os objetos pertencem a um grupo ao final do algoritmo.

	A	B
A	4	16
B	16	14
D = C	10	14
D	14	10
E	8	16

	A	B	C	D	E
A	0				
B	12.2	0			
D ₁ = C	6.3	6.0			
D	11.7	4.5	5.6	0	
E	4.0	8.2	2.8	8.5	0

O primeiro passo é verificar a distância mínima entre dois objetos, na matriz D_1 , dada por:

$$\text{Min}(d_{ij}) = d_{CE} = \mathbf{2.8} \text{ (intersecção entre a coluna 3 e linha 5)}$$

Desta forma, os objetos C e E formam o grupo (CE). Seguindo os passos descritos anteriormente, necessita-se obter as distâncias entre os objetos do grupo (CE) e os objetos restantes. Neste ponto, o método *Single Linkage* fica caracterizado, ou seja, as distâncias entre o grupo (CE) e os demais deve ser as “menores”. Assim, tem-se:

$$d_{(CE)A} = \min \{ d_{CA}, d_{EA} \} = \min \{ 6.3, 4.0 \} = 4.0$$

$$d_{(CE)B} = \min \{ d_{CB}, d_{EB} \} = \min \{ 6.0, 8.2 \} = 6.0$$

$$d_{(CE)D} = \min \{ d_{CD}, d_{ED} \} = \min \{ 5.6, 8.5 \} = 5.6$$

Excluindo-se as linhas e colunas na matriz D_1 correspondentes aos objetos C e E, e adicionando uma linha e coluna, correspondentes às menores distâncias dos objetos ao grupo (CE), obtém-se uma nova matriz D_2 .

$$D_2 = \begin{array}{c} (CE) \\ (CE) \\ A \\ B \\ D \end{array} \begin{array}{cccc} & A & B & D \\ \left[\begin{array}{cccc} 0 & & & \\ 4.0 & 0 & & \\ 6.0 & 12.2 & 0 & \\ 5.6 & 11.7 & 4.5 & 0 \end{array} \right] \end{array}$$

O mesmo procedimento efetuado na matriz D_1 deve ser realizado na matriz D_2 , isto é, procura-se a menor distância. Assim, no exemplo apresentado, obteríamos $d_{(CE)A} = 4.0$. Dessa forma, o objeto A seria alocado no grupo (CE). Novamente obtém-se a menor distância do grupo para os objetos restantes, ou seja:

$$d_{(CEA)B} = \min \{ d_{(CE)B}, d_{AB} \} = \min \{ 6.0, 12.2 \} = 6$$

$$d_{(CEA)D} = \min \{ d_{(CE)D}, d_{AD} \} = \min \{ 5.6, 11.7 \} = 5.6$$

$$D_3 = \begin{array}{c} (CEA) \\ (CEA) \\ B \\ D \end{array} \begin{array}{ccc} & B & D \\ \left[\begin{array}{ccc} 0 & & \\ 6.0 & 0 & \\ 5.6 & 4.5 & 0 \end{array} \right] \end{array}$$

A distância mínima é encontrada para o par B e D, $d_{BD} = 4.5$; dessa maneira, obtém-se um novo grupo (BD). Pode-se observar que, nesse momento, o processo forneceu dois grupos distintos (CEA) e (BD). Finalmente, analisam-se as distâncias:

$$d_{(CEA)(BD)} = \min \{ d_{(CEA)B}, d_{(CEA)D} \} = \min \{ 6.0; 5.6 \} = 5.6$$

A matriz de distância D_4 indica a união desses dois grupos em um único grupo, o qual irá aglutinar todos os objetos (ABCDE).

$$D_4 = \begin{matrix} & (CEA) & (BD) \\ \begin{matrix} (CEA) \\ (BD) \end{matrix} & \begin{bmatrix} 0 & \\ 5.6 & 0 \end{bmatrix} \end{matrix}$$

Observando-se as expressões usadas no exemplo, para definir sucessivamente as distâncias mínimas, pode-se definir a regra básica para o método *Single Linkage*

Definição: $d_{(ij)k} = \min \{ d_{ik}, d_{jk} \}$ no qual k é um terceiro objeto.

É possível representar, em dendrograma, os grupos obtidos. Veja a Figura 1.

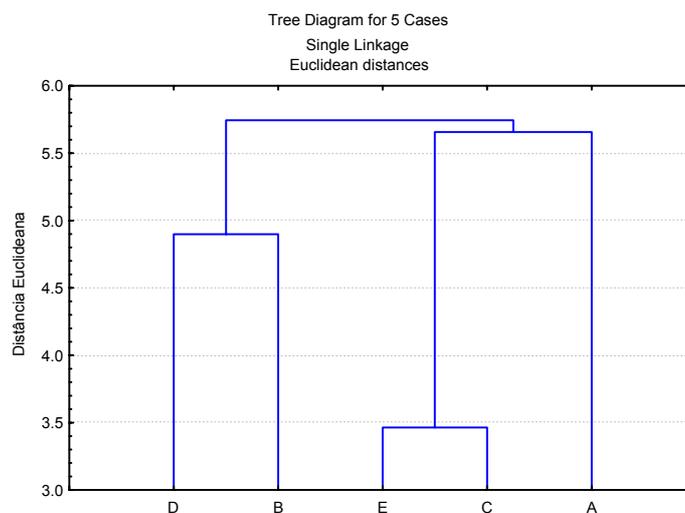


Figura 1 . Dendrograma pelo Método do Vizinho mais Próximo

Através do dendrograma, pode-se notar claramente a formação dos grupos ((CE) A) e (BD).

Vizinho mais Distante (Complete Linkage)

Ex: Utilizando-se a mesma matriz de distância do exemplo anterior para ilustrar os passos deste método.

$$D_1 = \begin{matrix} & A & B & C & D & E \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{bmatrix} 0 & & & & \\ 12.2 & 0 & & & \\ 6.3 & 6.0 & & & \\ 11.7 & 4.5 & 5.6 & 0 & \\ 4.0 & 8.2 & 2.8 & 8.5 & 0 \end{bmatrix} \end{matrix}$$

A Distância mínima entre dois objetos é:

$\text{Min}(d_{ij}) = d_{CE} = 2.8$ (intersecção entre a coluna 3 e a linha 5), e os objetos C e E formam o grupo (CE). Seguindo os passos descritos anteriormente, necessita-se obter as distâncias entre o grupo (CE) e os demais objetos. Nessa etapa, encontra-se a diferença entre os dois métodos: *Single e Complete Linkage*.

$$d_{(CE)A} = \max \{ d_{CA}, d_{EA} \} = \max \{ 6.3, 4.0 \} = 6.3$$

$$d_{(CE)B} = \max \{ d_{CB}, d_{EB} \} = \max \{ 6.0, 8.2 \} = 8.2$$

$$d_{(CE)D} = \max \{ d_{CD}, d_{ED} \} = \max \{ 5.6, 8.5 \} = 8.5$$

Verifica-se, abaixo, a modificação sofrida pela matriz de distância anterior:

$$D_2 = \begin{matrix} & (CE) & A & B & C \\ \begin{matrix} (CE) \\ A \\ B \\ D \end{matrix} & \begin{bmatrix} 0 & & & \\ 6.3 & 0 & & \\ 8.2 & 12.2 & 0 & \\ 8.5 & 11.7 & 4.5 & 0 \end{bmatrix} \end{matrix}$$

O próximo passo é, novamente, obter a menor distância. O valor é obtido pela intersecção da coluna representada pelo objeto B com a linha representada pelo objeto D. Essa distância é igual a 4.5 ($d(BD) = 4.5$), o que indica a formação de um novo grupo: objetos B e D (BD). Na próxima etapa, tem-se:

$$d_{(CE)BD} = \max \{ d_{(CE)B}, d_{(CE)D} \} = \max \{ 8.2, 8.5 \} = 8.5$$

$$d_{(BD)A} = \max \{ d_{BA}, d_{DA} \} = \max \{ 12.2, 11.7 \} = 12.2$$

obtém-se a matriz

$$D_3 = \begin{matrix} & \begin{matrix} (CE) & (BD) & A \end{matrix} \\ \begin{matrix} (CE) \\ (BD) \\ A \end{matrix} & \begin{bmatrix} 0 & & \\ 8.5 & 0 & \\ 6.3 & 12.2 & 0 \end{bmatrix} \end{matrix}$$

O próximo passo é formação do grupo (CEA), já que o menor valor da matriz de distância é 6.3. Finalmente, a última matriz de distância será obtida por:

$$d_{(CEA)BD} = \max \{ d_{(CE)(BD)}, d_{A(BD)} \} = \max \{ 12.2, 8.5 \} = 12.2$$

$$D_4 = \begin{matrix} & \begin{matrix} (CEA) & (BD) \end{matrix} \\ \begin{matrix} (CEA) \\ (BD) \end{matrix} & \begin{bmatrix} 0 & \\ 12.2 & 0 \end{bmatrix} \end{matrix}$$

e a regra básica utilizada para o cálculo das distâncias é: $d_{(ij)k} = \max \{ d_{ik}, d_{jk} \}$ em que k é um terceiro objeto.

Na seqüência, o dendrograma correspondente (Figura 2).

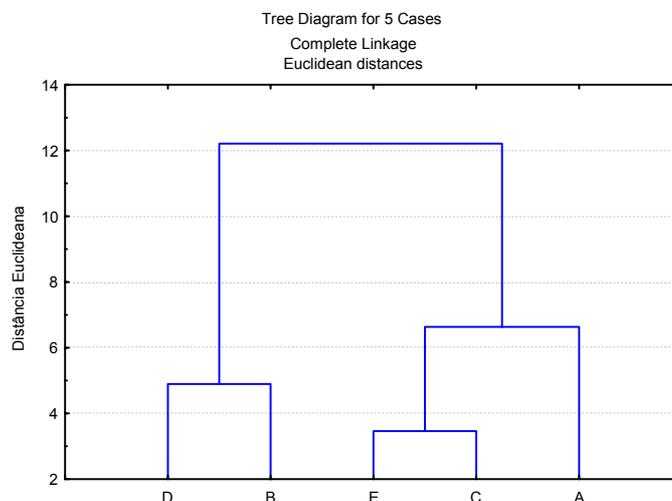


Figura 2 . Dendograma pelo Método do Vizinho mais Distante

Distância Média (Average Linkage)

Tomando-se como exemplo a matriz de distância apresentada para os métodos anteriores:

$$D_1 = \begin{matrix} & A & B & C & D & E \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{bmatrix} 0 & & & & \\ 12.2 & 0 & & & \\ 6.3 & 6.0 & & & \\ 11.7 & 4.5 & 5.6 & 0 & \\ 4.0 & 8.2 & 2.8 & 8.5 & 0 \end{bmatrix} \end{matrix}$$

Novamente tem-se como passo inicial, a obtenção do grupo (CE), o qual apresenta a distância mínima igual a 2.8. Utilizando-se a média aritmética, obtém-se as novas distâncias que irão compor a nova matriz de distância D_2 .

$$d(CE)A = \frac{1}{2} [d(C,A) + d(E,A)] = \frac{1}{2} [6.3 + 4.0] = 5.2$$

$$d(CE)B = \frac{1}{2} [d(C,B) + d(E,B)] = \frac{1}{2} [6 + 8.2] = 7.1$$

$$d(CE)D = \frac{1}{2} [d(C,D) + d(E,D)] = \frac{1}{2} [5.6 + 8.5] = 7.1$$

Portanto, a nova matriz D_2 é composta pelas distâncias:

$$D_2 = \begin{array}{c} (CE) \quad A \quad B \quad D \\ \begin{array}{c} (CE) \\ A \\ B \\ D \end{array} \begin{bmatrix} 0 & & & \\ 5.2 & 0 & & \\ 7.1 & 12.2 & 0 & \\ 7.1 & 11.7 & 4.5 & 0 \end{bmatrix} \end{array}$$

O novo grupo é formado pelos objetos B e D. Dessa forma, a fase seguinte cai, novamente, no cálculo das médias aritméticas das distâncias.

$$d(BD)A = \frac{1}{2}[d(B, A) + d(D, A)] = \frac{1}{2}[12.2 + 11.7] = 11.9$$

$$d(BD), (CE) = \frac{1}{2}[d(B, (CE)) + d(D, (CE))] = \frac{1}{2}[7.1 + 7.1] = 7.1$$

Como os resultados anteriores, tem-se a nova matriz de distâncias:

$$D_3 = \begin{array}{c} (CE) \quad (BD) \quad A \\ \begin{array}{c} (CE) \\ (BD) \\ A \end{array} \begin{bmatrix} 0 & & \\ 7.1 & 0 & \\ 5.2 & 11.9 & 0 \end{bmatrix} \end{array}$$

O objeto A se une ao grupo (CE). Finalmente, calcula-se a média para

$$d(BD), (CEA) = \frac{1}{2}[d((BD), (CE)) + d(A, (BD))] = \frac{1}{2}[7.1 + 11.9] = 9.5$$

$$D_4 = \begin{array}{c} (CEA) \quad (BD) \\ \begin{array}{c} (CEA) \\ (BD) \end{array} \begin{bmatrix} 0 & \\ 9.5 & 0 \end{bmatrix} \end{array}$$

e a regra básica utilizada para o cálculo das distâncias é: $d_{(ij)k} = \text{média} \{ d_{ik}, d_{jk} \}$ em que k é um terceiro objeto.

O dendrograma (Figura 3) referente ao procedimento acima está representado a seguir:

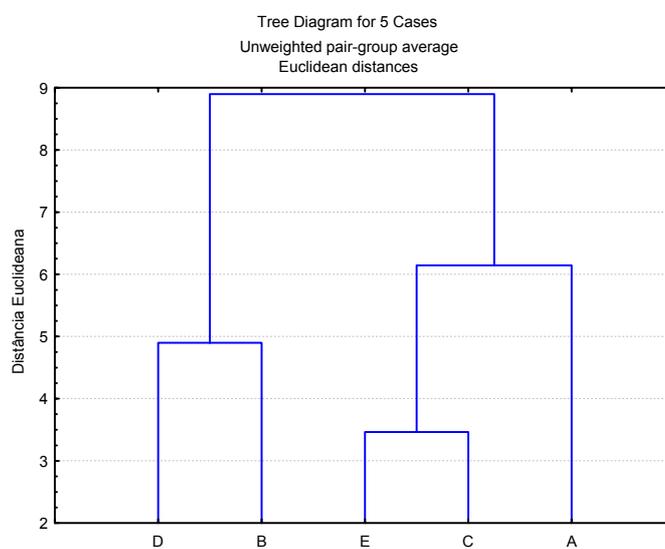


Figura 3 . Dendrograma pelo Método da Distância Média

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)