

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CORRELAÇÃO CANÔNICA PARA IDENTIFICAÇÃO
DE CARACTERES IMPORTANTES NA SELEÇÃO
PRECOCE DE TUBÉRCULOS DE BATATA

TESE DE DOUTORADO

Maria Helena Rigão

Santa Maria, RS, Brasil

2009

**CORRELAÇÃO CANÔNICA PARA IDENTIFICAÇÃO DE
CARACTERES IMPORTANTES NA SELEÇÃO PRECOCE DE
TUBÉRCULOS DE BATATA**

por

Maria Helena Rigão

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em
Agronomia, da Universidade Federal de Santa Maria (UFSM, RS), como
requisito parcial para obtenção do grau de
Doutor em Agronomia

Orientador: Prof. Lindolfo Storck

Santa Maria, RS, Brasil

2009

**Universidade Federal de Santa Maria
Centro de Ciências Rurais
Programa de Pós-Graduação em Agronomia**

A Comissão Examinadora, abaixo assinada,
aprova a Tese de Doutorado

**CORRELAÇÃO CANÔNICA PARA IDENTIFICAÇÃO DE CARACTERES
IMPORTANTES NA SELEÇÃO PRECOCE DE TUBÉRCULOS DE BATATA**

elaborada por
Maria Helena Rigão

Como requisito parcial para obtenção do grau de
Doutor em Agronomia

COMISSÃO EXAMINADORA:

Lindolfo Storck, Dr.
(Presidente/Orientador)

Thomas Newton Martin, Dr. (UTFPR)

Adriano Mendonça Souza, Dr. (UFSM)

Dilson Antônio Bisognin, Dr. (UFSM)

Alberto Cargnelutti Filho, Dr. (UFSM)

Santa Maria, 26 de fevereiro de 2009.

A

MIZ TLI TLAN

Àqueles que contribuíram na realização deste trabalho e, principalmente, ao Comitê de Orientação, Prof. Dilson Antônio Bisognin, Prof. Sidinei José Lopes e Prof. Lindolfo Storck, orientador, meus sinceros agradecimentos.

*Todos os caminhos são os mesmos; não conduzem a lugar algum.
Então, a pergunta é: esse caminho tem um coração?
Se tiver, o caminho é bom, se não tiver, não possui importância alguma.*

Carlos Castañeda

RESUMO

Tese de Doutorado
Programa de Pós-Graduação em Agronomia
Universidade Federal de Santa Maria, RS, Brasil

CORRELAÇÃO CANÔNICA PARA IDENTIFICAÇÃO DE CARACTERES IMPORTANTES NA SELEÇÃO PRECOCE DE TUBÉRCULOS DE BATATA

AUTOR: Maria Helena Rigão

ORIENTADOR: Lindolfo Storck

Santa Maria, 26 de fevereiro de 2009.

A cultura da batata é mundialmente reconhecida como uma importante fonte de alimento no combate a fome, por suas qualidades nutricionais, bem como uma alternativa ao emprego rural e ingressos financeiros. No Brasil a batata caracteriza-se por uma grande dependência de cultivares de origem européia, motivo pelo qual, muitas universidades e empresas dedicam-se a pesquisa, na área de melhoramento, na tentativa de obter cultivares de qualidade e adaptada as nossas condições. Uma estratégia utilizada pelos melhoristas é a seleção precoce, por reduzir custos e trabalho para manutenção e multiplicação de clones. O objetivo deste trabalho foi testar a correlação canônica, como uma ferramenta estatística capaz de auxiliar na seleção precoce de clones de batata, identificando os caracteres, altamente relacionados, entre tubérculos plantados e sua produção. Para isto realizaram-se três ensaios, em três diferentes épocas, na área experimental do Departamento de Fitotecnia da Universidade Federal de Santa Maria, utilizando 10 progênies de batata. Os caracteres utilizados, para a realização da análise de correlação canônica, para os tubérculos plantados e sua respectiva produção foram: comprimento, maior diâmetro, menor diâmetro e massa. A correlação canônica identificou que há uma relação entre caracteres de tubérculos plantados e produzidos. A maior associação ocorre em comprimento de tubérculos, sugerindo que, ao plantar-se tubérculos mais longos obtém-se, também, tubérculos mais alongados.

Palavras-chave: *Solanum tuberosum* L., melhoramento, análise multivariada

ABSTRACT

Doctoral Thesis
Graduate Program of Agronomy
Federal University of Santa Maria, RS, Brazil

CANONICAL CORRELATION FOR IDENTIFICATION OF MAJOR CHARACTERS IN EARLY SELECTION OF TUBERS OF POTATOES

AUTHOR: Maria Helena Rigão

ADVISER: Lindolfo Storck

Santa Maria, February 26th 2009.

The potato is recognized worldwide as an important source of food to decrease hunger, because of its nutritional traits, as well as an alternative to rural employment and financial income. In Brazil, potato is characterized by high dependence of cultivars of European origin, for this reason, many universities and companies have contributed to breeding researches, in an attempt to achieve cultivars with quality and adaptation it to our conditions. A strategy used by breeders is the early selection, to reduce costs and effort for clones maintenance and multiplication. The objective of this work was to test the canonical correlation, as a statistical tool to assist in the early selection of potatoes clones, to identify characters, related to tubers planted and their production. For this, there three experiments in three different seasons were conducted, in the experimental area Fitotecnia Department of the Federal University of Santa Maria, with 10 potatoes progenies. The characters used for the analysis of canonical correlation were the tubers planted and their production: length, larger diameter, smaller diameter and mass. The canonical correlation identified shows that there is a relationship between characters of the tubers planted and produced. The largest association occurs with tuber length, suggesting that if it were planted long tubers, you get, more elongated tubers as well.

Key-words: *Solanum tuberosum* L., breeding, multivariate analysis

LISTA DE FIGURAS

Figura 01 - Logo do Ano Internacional da Batata	21
Figura 02 - Fluxograma para desenvolvimento de novas cultivares	25
Figura 03 - Dendograma de similaridade pelo método do vizinho mais próximo	65

LISTA DE TABELAS

Tabela 01- Produção mundial de batatas de 1994 a 2006 em milhões de toneladas.....	17
Tabela 02- Consumo anual de batatas por pessoa (kg).....	18
Tabela 03- Importações brasileiras de batata semente - Maio de 2002 a maio de 2003.....	18
Tabela 04- Área plantada, produção e rendimento de batata - Brasil - Safra 2006 - 2007.....	19
Tabela 05- Dados do cultivo da batata no Rio Grande do Sul 2000-2006.....	20
Tabela 06- Composição de uma batata de aproximadamente 150 g	20
Tabela 07- Análise da variância multivariada no delineamento inteiramente casualizado	46
Tabela 08- Descrição dos parentais das progênes e número de tubérculos usados na época 1 de plantio (14/03/2006)	49
Tabela 09- Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 1	52
Tabela 10- Correlação linear entre as variáveis dos tubérculos plantados na época 1	52
Tabela 11- Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 1	53
Tabela 12- Correlação linear entre as variáveis dos tubérculos colhidos na época 1.....	53
Tabela 13- Correlação linear entre as variáveis dos grupos G1 e G2 na época 1	54
Tabela 14- Coeficientes canônicos do primeiro par canônico na época 1	55
Tabela 15- Número de tubérculos plantados e colhidos de cada progênie nas épocas 1 e 2.....	55
Tabela 16- Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 2	56
Tabela 17- Correlação linear entre as variáveis dos tubérculos plantados na época 2.....	56
Tabela 18- Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 2	57
Tabela 19- Correlação linear entre as variáveis dos tubérculos colhidos na época 2.....	58
Tabela 20- Correlação linear entre as variáveis dos grupos G1 e G2 na época 2	59
Tabela 21- Coeficientes canônicos do primeiro par canônico na época 2	59
Tabela 22- Número de tubérculos plantados e colhidos de cada progênie nas épocas 1, 2 e 3	60

Tabela 23- Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 3	61
Tabela 24- Correlação linear entre as variáveis dos tubérculos plantados na época 3	61
Tabela 25- Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 3	62
Tabela 26- Correlação linear entre as variáveis dos tubérculos colhidos na época 3.....	62
Tabela 27- Correlação linear entre as variáveis dos grupos G1 e G2 na época 3	63
Tabela 28- Coeficientes canônicos do primeiro par, na época 3 (outono).....	63
Tabela 29- Medidas de dissimilaridade entre as nove progênies, em relação aos caracteres, comprimento, maior diâmetro, menor diâmetro e massa dos tubérculos colhidos na época 3 com base na distância de Mahalanobis	64

LISTA DE SIGLAS

ABBA - Associação Brasileira da Batata
ANOVA - Análise de Variância
ASBAT - Associação dos Produtores de Batata da Quarta Colônia Italiana
CIP - Centro Internacional de la Papa
CPACT - Centro de Pesquisa Agropecuária de Clima Temperado da EMBRAPA
DHE - Distingüibilidade, Homogeneidade e Estabilidade
EMATER - Empresa de Assistência Técnica e Extensão Rural
EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária
EPAGRI - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina
EPAMIG - Empresa de Pesquisa Agropecuária de Minas Gerais
ESALQ - Escola Superior de Agricultura “Luiz de Queiroz”
FAO - Food and Agriculture Organization of the United Nations
FEPAGRO - RS - Fundação Estadual de Pesquisa Agropecuária do Rio Grande do Sul
IBGE - Instituto Brasileiro de Geografia e Estatística
IAC - Instituto Agrônômico de Campinas
IAPAR - Instituto Agrônômico do Paraná
IAS - Instituto Agrônômico do Sul
MANOVA - Análise de Variância Multivariada
SAA - RS - Secretaria da Agricultura e Abastecimento do RS
UEPG - Universidade Estadual de Ponta Grossa
UFLA - Universidade Federal de Lavras
UFSM - Universidade Federal de Santa Maria
UNESP - Universidade Estadual Paulista
UPF - Universidade de Passo Fundo
UTFPR - Universidade Tecnológica Federal do Paraná
VCU - Valor de Cultivo e Uso.

SUMÁRIO

1 INTRODUÇÃO	14
2 REVISÃO BIBLIOGRÁFICA	16
2.1 Cultura da batata.....	16
2.2 Melhoramento genético.....	23
2.3 Considerações sobre técnicas estatísticas univariadas e multivariadas	27
2.4 Correlações.....	31
2.4.1 Coeficiente de correlação linear de Pearson	32
2.4.2 Correlação parcial.....	33
2.4.3 Análise de regressão	34
2.4.4 Análise de trilha	36
2.4.5 Correlação canônica	36
2.5 Análise de agrupamento.....	40
2.6 Análise de variância multivariada	44
3 METODOLOGIA	48
4 RESULTADOS E DISCUSSÃO.....	51
4.1 Primeira época de plantio.....	51
4.2 Segunda época de plantio.....	55
4.3 Terceira época de plantio	60
5 CONCLUSÃO	68
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	69
ANEXO.....	76

1 INTRODUÇÃO

Sabe-se que nesse século, entre tantos desafios que deverão ser enfrentados pela raça humana, podem ser destacados os problemas climáticos, a superpopulação de algumas áreas, a subnutrição e, em alguns casos, a fome. Não cabe aqui discutir as causas desses problemas, mas sim, contribuir, dentro das possibilidades, com um projeto maior, de oferecer mais alimento e em condições mais saudáveis, para minimizar os problemas de falta de alimento, senão da humanidade, pelo menos de nossa região.

A batata é fonte de alimento, emprego rural, ingressos financeiros e ainda, um dos alimentos mais completos, devido a proteínas de alta qualidade, vitaminas e sais minerais, além de proporcionar energia oriunda dos carboidratos. Em termos nutricionais, 100 gramas de batata suprem cerca de 10% das necessidades recomendadas de proteína para uma criança, ou 10% da demanda de tiamina, niacina, vitamina B6, ácido fólico e 50% da vitamina C para um adulto (HAASE, 2007).

O ano de 2008 foi declarado o Ano Internacional da Batata, com objetivo de criar uma consciência sobre a importância da batata e da agricultura em geral, através de questões de interesse mundial como a fome, a pobreza e os perigos que corre o meio ambiente e, nas palavras do Diretor Geral da FAO, Jacques Diouf: “A batata está em primeira linha na luta contra a fome e a pobreza no mundo”. (POTATO, 2008).

Os programas brasileiros de melhoramento genético da batata são de fundamental importância, tanto na perspectiva de que esse produto assumirá um papel importante na nutrição humana, como uma fonte de alimento adequada, sustentável e de alta qualidade para as populações; mas também como uma busca de novas alternativas as cultivares importadas (PEREIRA, 2003; SILVA et al., 2006). Quando se fala em melhoramento genético da batata, PINTO (2001) destaca que existem vários caracteres que podem ser melhorados geneticamente, um dos mais importantes é, sem dúvida alguma, a produtividade de tubérculos, assim como a aparência dos tubérculos, resistência às doenças fúngicas, bacterianas e viróticas, resistência aos insetos e outros. Ou seja, é praticamente impossível obter uma nova cultivar contendo todos os caracteres desejáveis. Assim, os programas de melhoramento devem, em primeiro lugar, definir quais características devem ser priorizadas.

Hoje, no Brasil, os programas de melhoramento de batata concentram-se na EMBRAPA (Embrapa Hortaliças e Clima Temperado) e nas Universidades, com a participação de outras empresas. A Universidade Federal de Santa Maria desenvolve um Programa de Genética e Melhoramento da Batata, com o objetivo de desenvolver novas cultivares, que atendam a indústria de processamento e o consumo de mesa, com características mínimas de qualidade de tubérculo, resistentes a requeima e que sejam altamente adaptadas as condições edafocológicas do Brasil (UFSM/BATATA, 2007).

O melhoramento genético, para a obtenção de novas cultivares, é uma tarefa de longo prazo, uma alternativa para acelerar este processo é realizar a seleção precoce, ou seja, a seleção nas primeiras gerações após a hibridação. A vantagem da seleção precoce é poder agilizar o programa e encurtar os ciclos de seleção, desde que clones superiores da progênie possam ser adequadamente identificados nas primeiras gerações clonais. O descarte do maior número possível de clones com caracteres indesejáveis, na primeira geração clonal, também é recomendado, para reduzir custos e o trabalho para a manutenção e multiplicação destes materiais (BISOGNIN & DOUCHES, 2002; BISOGNIN, 2003).

Integrando-se a este contexto, de contribuir para que o Programa de Genética e Melhoramento da Batata da Universidade Federal de Santa Maria obtenha cultivares, de acordo com a suas propostas, em um menor tempo possível, realizou-se este trabalho, com o objetivo de, utilizando a correlação canônica, identificar caracteres importantes, que possam ser utilizados para a seleção precoce de clones de batata.

2 REVISÃO BIBLIOGRÁFICA

2.1 Cultura da batata

A história da batata começa há uns 8.000 anos, perto do lago Titicaca, na Cordilheira dos Andes, América do Sul. As comunidades que habitavam esta região começaram a domesticar as plantas silvestres de batata, que havia em abundância ao redor do lago. Os Incas, que utilizaram métodos agrícolas avançados, deram especial importância à produção de milho, mas a batata foi decisiva para a segurança alimentícia de seu império. A batata e, principalmente, um produto elaborado com a batata dessecada e congelada, chamado “chuño” foi um dos principais alimentos usado na alimentação de oficiais, soldados e escravos, bem como reserva, para casos de emergência, quando ocorriam perdas de colheita (POTATO, 2008).

Há, no continente americano, aproximadamente, 200 espécies de batatas silvestres, mas foi na região central dos Andes que os agricultores conseguiram selecionar e melhorar a primeira variedade do tubérculo. O que hoje se conhece como batata, gênero *Solanum* e espécie *tuberosum* contém apenas um fragmento da diversidade genética das sete espécies reconhecidas de batata e as 5.000 variedades, que continuam sendo cultivadas nos Andes, (CIP, 2008).

A invasão espanhola, em 1532, colocou um fim a civilização Inca, mas não a batata, pois, em toda a história andina, a batata, em todas as suas formas, tem sido “o alimento do povo”. Os conquistadores, quando de regresso a Europa, levaram a *Solanum tuberosum*. Há registro do cultivo de batata nas Ilhas Canárias em 1565 e, em 1573 na península Ibérica. Daí seguiu para Roma, Viena e, em 1597, já se cultivava batatas em Londres, que a seguir chegou à França e depois aos Países Baixos (POTATO, 2008). Com a “época dos descobrimentos” da Europa os marinheiros levavam os tubérculos para serem consumidos durante suas viagens. Assim, a batata chegou à Índia, a China e ao Japão no início do século XVII.

A batata foi bem aceita na Irlanda, porém um ataque de requeima (*Phytophthora infestans*) causou a “Fome da Batata”, ocorrida entre 1845 e 1849, reduzindo em 2,5 milhões de habitantes, a população irlandesa. Destes, 1,5 milhões morreram de fome e um milhão

imigrou, principalmente para os EUA. Este desastre ocorreu devido à **dependência, pela** maioria da população, de uma única espécie. A catástrofe irlandesa levou a busca de cultivares mais produtivas e resistentes a enfermidades. Melhoristas da Europa e América do Norte utilizaram germoplasma importado do Chile e desenvolveram muitas das cultivares, que formam, hoje, a base da produção de batata nestas duas regiões (POTATO, 2008).

O colonialismo e a imigração europeia levaram a batata para Bengala, Egito, Marrocos, Nigéria, Austrália, Argentina e Brasil. Desde o Cáucaso, a batata chegou à Turquia, da Rússia foi para a China e daí para a Coreia. Finalmente, no século XX, a batata converteu-se em um alimento mundial. Desde a década de 1960, o cultivo da batata tem-se estendido nos países em desenvolvimento. Só na Índia e na China, o total da produção passou de 16 milhões de toneladas em 1960 para quase 100 milhões em 2006 (POTATO, 2008). Atualmente, mesmo com a redução da produção de batatas pelos países desenvolvidos, ainda há muito espaço para sua produção nos países em desenvolvimento, onde seu consumo é menos de uma quarta parte, do que representa nos países desenvolvidos, como mostra a Tabela 01.

Tabela 01 - Produção mundial de batatas de 1994 a 2006 em milhões de toneladas

Países	1994	1996	1998	2000	2002	2004	2006
Desenvolvidos	168,69	193,59	169,25	182,04	163,58	171,79	155,25
Em desenvolvimento	102,38	117,71	131,41	146,51	152,41	157,77	159,12
Mundo	271,07	311,31	300,67	328,55	315,98	329,56	314,37

Fonte: FAO (2008)

No ano de 2007 os Países Baixos obtiveram o maior rendimento, entre os países produtores de batata, com uma produtividade de 44,72 t/ha, seguidos da França, com 43,24 t/ha. O Brasil, entretanto, classificou-se na posição 45º, com uma produtividade de 23,71 t/ha. (FAO, 2008).

O consumo anual da batata por pessoa, no mundo, é na ordem de 27 kg, porém, no Brasil, é de apenas 15 kg, sendo que, em países como a Polónia e Bolívia chega a 190 kg e 100 kg, respectivamente (SHIMOYAMA, 2006). O peruano consome 80 kg de batatas por ano, o europeu, 93 kg; enquanto que os bielorrussos consomem 338 kg ao ano (PAPAS ANDINAS, 2008; FRUIT-INFORM, 2008) (Tabela 02).

Tabela 02 - Consumo anual de batatas por pessoa (kg)

País	Consumo anual por pessoa (kg)
Bielo-Rússia	338
Polônia	190
Quirguistão	152
Rússia	142
Ucrânia	141
Armênia	131
Reino Unido	114
Irlanda	101
Bolívia	100
Peru	80
França	55
Itália	40
Brasil	15

Fonte: Shimoyama, 2006; Papas Andinas, 2008; Fruit-inform, 2008

Quase 70% da batata-semente básica, de alta sanidade, utilizada no Brasil é importada, dependência que já foi bem maior em outras décadas. O País gasta por ano cerca de 5 milhões de dólares com a importação anual deste produto, o que corresponde entre 150 a 200 mil caixas. O custo com a importação é alto, cada caixa de batata-semente de mais ou menos 30 kg custa em média de 25 dólares, dependendo da variedade (DIAS, 2006). Mais da metade das importações (52,50%) vem da Holanda, seguido do Canadá, de onde se compra 21,94% de batatas-semente importadas (Tabela 03).

Tabela 03 - Importações brasileiras de batata semente - Maio de 2002 a maio de 2003

País	Quantidade (t)	Porcentagem (%)
Holanda	1435,76	52,50%
Canadá	600,00	21,94%
Chile	587,76	21,49%
Escócia	75,00	2,74%
França	32,25	1,18%
Alemanha	3,75	0,14%
Total	2734,52	100,00%

Fonte: ABBA (2008)

Atualmente o Brasil chega a produzir, por ano, em três safras, cerca de 3.300.000 toneladas de batata, em uma área cultivada anual de, aproximadamente, 140.000 hectares. Os estados de Minas Gerais, São Paulo, Paraná e Rio Grande do Sul são os que possuem maior área de cultivo, 87,09% da área nacional, porém, Goiás e Bahia destacam-se em rendimento da cultura, com uma produtividade de 40,12 t/ha e 39,13 t/ha, respectivamente. Mesmo com a redução das lavouras, a produção continua crescendo, graças a investimentos em tecnologia, variedades mais produtivas e, particularmente, na qualidade das sementes (Tabela 04).

Tabela 04 - Área plantada, produção e rendimento de batata - Brasil - Safra 2006-2007

Estado produtor	Área plantada (ha)	Produção colhida (t)	Rendimento (t/ha)
Minas Gerais	40.700	1.126.155	27,67
São Paulo	32.070	751.626	23,44
Paraná	27.651	611.330	22,11
Rio Grande do Sul	24.307	386.393	15,89
Santa Catarina	7.379	101.454	13,75
Bahia	7.015	274.495	39,13
Goiás	3.280	131.593	40,12
Espírito Santo	484	7.961	16,45
Paraíba	330	2.890	8,76
Distrito Federal	2	60	30,00
Brasil	143.218	3.393.957	23,70

Fonte: IBGE (2008)

O Rio Grande do Sul, que foi um dos pioneiros na introdução desta cultura no país e já possuiu uma área muito maior, destinada ao seu cultivo, teve, devido a conjunturas econômicas e sociais, uma redução, na ordem de 63%, em sua área de plantio, no período de 1976 a 2006 (FIOREZE, 2006). Porém, mesmo aumentando de produtividade na ordem de 123% (Tabela 05), manteve-se muito abaixo da média nacional (23,70 t/ha) (Tabela 04).

Tabela 05 – Dados do cultivo da batata no Rio Grande do Sul 2000-2006.

Ano	Área (ha)	Produção (t)	Rendimento (t/ha)
2000	39.364	391.183	9,94
2001	37.079	384.563	10,37
2002	35.933	384.115	10,69
2003	29.451	313.157	10,63
2004	25.959	294.912	11,36
2005	23.614	284.129	12,03
2006	24.118	332.581	13,79
2007	19.643	386.396	19,67

Fonte: IBGE (2008)

A batata é um dos alimentos mais consumidos do mundo, não somente pela composição nutricional, mas pela versatilidade gastronômica, bem como, pelo baixo preço de comercialização de seus tubérculos (ABBA, 2008; PAPAS ANDINAS, 2008). A Tabela 06 mostra os componentes de uma batata de aproximadamente 150 g.

Tabela 06 – Composição de uma batata de aproximadamente 150 g.

Componentes	Valores
Energia	150 cal
Proteínas	3,7 g
Lipídios (gordura)	0 g
Carboidratos	23 g
Fibras	27 g
Sódio	5 mg
Potássio	729 mg

Fonte: ABBA (2008)

Nos Andes peruanos, onde começou a história da batata, localiza-se o Centro Internacional da Batata (Centro Internacional de la Papa - CIP), que ajuda as comunidades campestres a restabelecer a produção de 1.200 variedades andinas em uma área de 12.000 ha. Essa diversidade genética será a base das novas variedades adaptadas às necessidades mundiais (CIP, 2008).

Dada a importância da cultura da batata, a Assembléia Geral da ONU, em 22 de dezembro de 2005, declarou, 2008, o Ano Internacional da Batata, com o objetivo de criar uma consciência da importância da batata, como alimento nos países em desenvolvimento e

promover a pesquisa e desenvolvimento de sistemas de produção de batata, a fim de contribuir com o cumprimento dos Objetivos de Desenvolvimento do Milênio das Nações Unidas (POTATO, 2008).

A Figura 01 mostra o Logo do Ano Internacional da Batata. O desenho, criado pelo italiano Giancarlo de Pol, mostra, na parte superior, uma bandeja com batatas colhidas e, na parte inferior, dentro da terra, um tubérculo dourado, representando a íntima ligação da humanidade com a agricultura. O lema “Tesouro enterrado” destaca a importância, freqüentemente descuidada, que tem a batata para a população rural, a economia e a segurança alimentícia do planeta. O Logo foi criado em árabe, chinês, espanhol, francês, inglês e russo (POTATO, 2008).

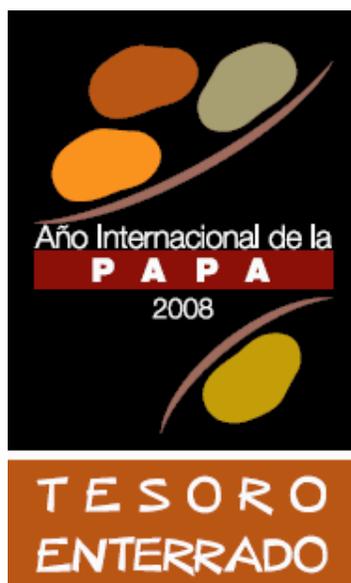


Figura 01 – Logo do Ano Internacional da Batata

A história da cultura da batata no Brasil começa no século XIX, quando chegou, trazida da Europa, pelos imigrantes que vieram para o Sul. O melhoramento genético, propriamente dito, iniciou no Brasil em 1942, no município de Rio Grande - RS, na Estação Experimental de Horticultura, da Secretaria de Agricultura - RS. Este programa contribuiu para o desenvolvimento de diversas cultivares, que tiveram boa aceitação no Rio Grande do Sul. No entanto, o programa foi finalizado devido à infestação de muchadeira (*Ralstonia solanacearum*) no campo experimental daquela estação (PEREIRA, 2003).

Ainda na década de 40, iniciaram-se dois outros programas, concomitantemente, sendo um em Pelotas - RS, no extinto Instituto Agronômico do Sul (IAS), atualmente Embrapa Clima Temperado, e outro em Campinas - SP, no Instituto Agronômico de Campinas (IAC). Ambos os programas lançaram diversas cultivares. Entre elas, a Baronesa, lançada em 1955 pelo IAS, que obteve o maior êxito e, ainda hoje, predomina nas lavouras de batata do Rio Grande do Sul. A cultivar Aracy, lançada pelo IAC, teve alguma repercussão em São Paulo (PEREIRA, 2003).

Programas de outras instituições: Embrapa Hortaliças, Instituto Agronômico do Paraná (IAPAR), Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (EPAGRI), iniciados posteriormente, também lançaram cultivares no mercado (PEREIRA, 2003).

Hoje, os programas de melhoramento de batata concentram-se na Embrapa (Embrapa Hortaliças e Clima Temperado), com a participação de empresas privadas e de empresas estaduais de P&D (EPAGRI, IAPAR) e nas Universidades (Universidade Federal de Santa Maria - UFSM - RS e Universidade de Passo Fundo - UPF - RS). Há também programas conduzidos pelo IAC-SP e pela Universidade Federal de Lavras - UFLA - MG. O programa de melhoramento da UFLA tem produzido informações sobre mudanças meteorológicas durante o processo de condução de populações segregantes, além de informações básicas sobre resistência ao calor. A EPAMIG está reiniciando seu programa de melhoramento de batata. Trabalhos sobre resistência a insetos também estão sendo implementados na Embrapa Hortaliças, em conjunto com a Universidade Estadual Paulista – UNESP - *Campus* de Jaboticabal, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ - SP, Universidade Estadual de Ponta Grossa – UEPG - PR, e com a Empresa de Assistência Técnica e Extensão Rural do RS - EMATER-RS.

A Universidade Federal de Santa Maria desenvolve, em seu Departamento de Fitotecnia, dois programas relativos à cultura da batata, sendo eles: (i) Programa de Genética e Melhoramento da Batata, com o objetivo de desenvolver novas cultivares, que atendam a indústria de processamento e o consumo de mesa, com características mínimas de qualidade de tubérculo, resistentes a requeima e que sejam altamente adaptadas às condições edafocológicas do Brasil (UFSM/BATATA, 2007). Este programa já possui vários trabalhos concluídos, publicados e sendo colocados em prática, pelos destinatários de sua produção, em suas diversas linhas de pesquisa; (ii) Programa Integrado de Produção de Batata-Semente, cujo objetivo é atender a demanda total de batata-semente dos produtores da região central do

RS. Este programa é uma cooperação técnica entre Universidade Federal de Santa Maria (UFSM), Empresa de Assistência Técnica e Extensão Rural do RS (EMATER), Fundação Estadual de Pesquisa Agropecuária do RS (FEPAGRO), Centro de Pesquisa Agropecuária de Clima Temperado (CPACT) da EMBRAPA, Secretaria da Agricultura e Abastecimento do RS (SAA) e Associação dos Produtores de Batata da Quarta Colônia Italiana (ASBAT) (www.ufsm.br/batata).

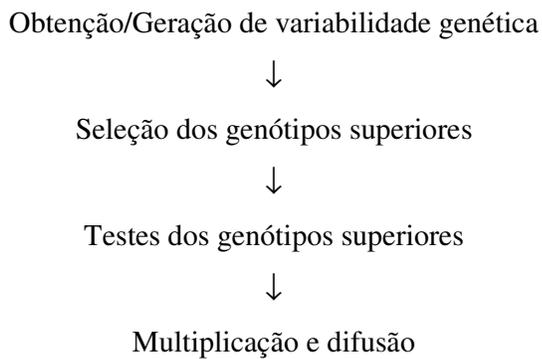
2.2 Melhoramento genético

O melhoramento genético, para a obtenção de novas cultivares, é uma tarefa de longo prazo e requer a dedicação intensiva de pesquisadores, que atuam em várias áreas, tais como: genética, fisiologia, estatística, botânica, bioquímica, fitopatologia, entomologia; que devem trabalhar como equipe, para facilitar a identificação das cultivares mais promissoras a serem colocadas à disposição dos produtores (BORÉM, 2001).

Uma etapa importante, em programas de melhoramento, é a escolha de genitores porque, clones superiores da progênie podem ser multiplicados e mantidos indefinidamente. Esta escolha deve ser efetuada de acordo com os objetivos de cada programa, sendo necessário conhecer algumas de suas características, tais como a capacidade de transferência de caracteres de interesse e a habilidade ou capacidade de cruzamento. É interessante que os genitores sejam geneticamente distantes, possibilitando o aumento da heterose que poderá ser mantida pela propagação clonal (TAI & YOUNG, 1984; BISOGNIN, 2003).

Definidos os genitores, os programas, em geral, iniciam-se com hibridações controladas, obtendo-se sementes botânicas que são semeadas em casa de vegetação. A partir de cada semente botânica, é obtido um clone, que é diferente dos demais, e que pertence ao conjunto de clones provenientes de um cruzamento, que constitui uma família clonal. Os novos clones originados constituem a geração plântula. Estas plantas normalmente são cultivadas em vasos pequenos, produzindo poucos tubérculos de tamanho reduzido. Na geração de plântula geralmente são descartados os piores clones, principalmente para caracteres de alta herdabilidade, como a aparência, formato de tubérculos, profundidade de gemas, número de tubérculos, entre outros (NEELE et al., 1989; GONZÁLEZ et al., 2002). Porém, a seleção para caracteres relativos à produção de tubérculos é pouco eficiente nesta fase (TAI, 1975; PINTO et al., 1994; AMARO, 2002). A partir da geração de plântula são

obtidos os tubérculos que constituirão a primeira geração clonal e, com as sucessivas multiplicações, as gerações clonais 2, 3, e assim por diante. ALVES et al. (1999) sugerem a seguinte seqüência para um programa de melhoramento de plantas de propagação vegetativa:



O programa de genética e melhoramento da batata, da Universidade Federal de Santa Maria, com o objetivo de desenvolver clones de qualidade e resistentes a requeima, utiliza o esquema de melhoramento mostrado na Figura 02 (COSTA, 2007).

A seleção realizada nas primeiras gerações após a hibridação, em programas de melhoramento de batata, é denominada de seleção precoce. A vantagem da seleção precoce é poder agilizar o programa e encurtar os ciclos de seleção, desde que, clones superiores da progênie possam ser adequadamente identificados nas primeiras gerações clonais. O descarte do maior número possível de clones com caracteres indesejáveis, na primeira geração clonal, também é recomendado, para reduzir custos e o trabalho para a manutenção e multiplicação dos mesmos (BISOGNIN, 2003).

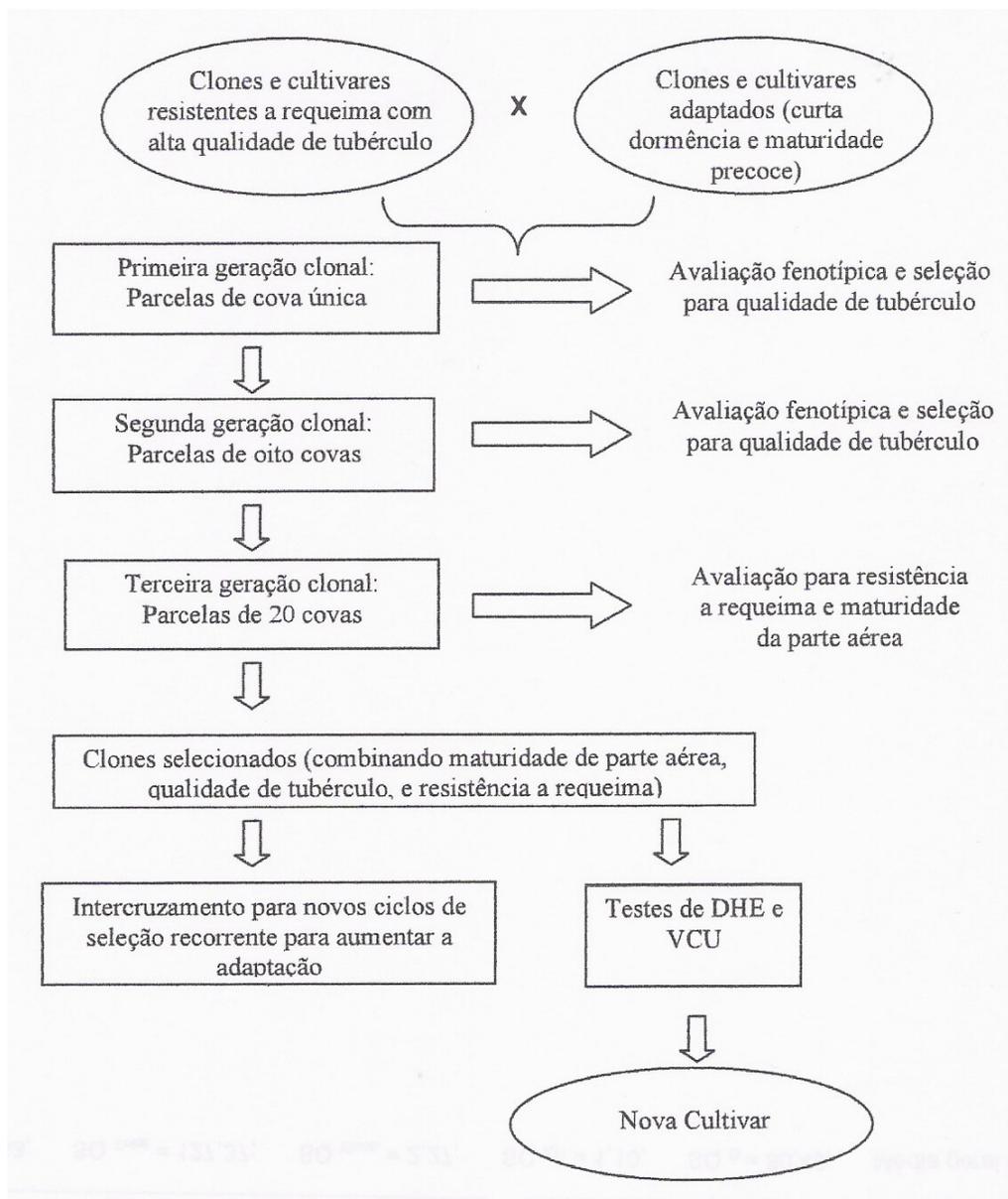


Figura 02 – Fluxograma para desenvolvimento de novas cultivares resistentes a requeima, combinando alta qualidade de tubérculo com maturidade precoce de parte aérea, adaptado de Costa (2007). DHE: Distinguibilidade, Homogeneidade e Estabilidade; VCU: Valor de Cultivo e Uso.

Uma prática muito empregada pelos melhoristas de batata é a seleção visual, muitas vezes denominada de “preferência do melhorista”. Para TAI (1975), este tipo de seleção apresenta baixa eficiência. BROWN (1987) recomenda a seleção visual nos primeiros estágios, apenas para identificação de cruzamentos com alto potencial de produzirem

cultivares. A proposta seria de identificar quais famílias apresentariam maiores médias de notas, variando de 1 a 9, as quais seriam dadas por melhoristas. Somente as famílias com maiores médias iriam fornecer materiais para posteriores avaliações e as demais seriam descartadas. Para caracteres específicos, PEREIRA (2000) recomenda o uso da glico-fita com eficiência satisfatória para eliminação de clones com alto teor de açúcares redutores nas duas primeiras gerações clonais.

A seleção realizada nas gerações iniciais, de acordo com LAMPERT (2004), principalmente para caracteres de baixa herdabilidade, geralmente é pouco eficiente, devendo-se adiar a seleção para as gerações posteriores, em que serão realizados experimentos com maior número de repetições. O autor acrescenta que, apenas os melhores genótipos devem ser multiplicados e avançados, devido ao custo relacionado com a manutenção de enorme quantidade de tubérculos de vários clones, além da necessidade de área e mão-de-obra para instalação dos ensaios.

Um estudo da herança do formato de tubérculo, realizado por DE JONG & BURNS (1993), indica que o controle é realizado por um gene de grande efeito e um número desconhecido de modificadores, em que o formato redondo é dominante sobre o longo. O formato dos tubérculos é um caracter de alta herdabilidade. Quando o genótipo possui o formato predominante desejável, os principais cuidados para se evitar a produção de tubérculos deformados seriam o suprimento ideal de água durante o enchimento dos mesmos e o adequado preparo do solo.

Para avaliar a viabilidade da seleção precoce para caracteres importantes para o processamento da batata na forma de fritura, AMARO (2002) utilizou índices de formato do tubérculo (relação comprimento/diâmetro maior e diâmetro maior/diâmetro menor). Nas estimativas das herdabilidades no sentido amplo, ele encontrou 70,4% para o índice comprimento/diâmetro maior e 58,0% para o índice diâmetro maior/diâmetro menor. As estimativas dos coeficientes de correlação simples entre as gerações de plântula x clonal 1 (C1) foram de 0,51 para comprimento/diâmetro maior em clones e 0,77 em famílias, sendo ambos os valores significativos, indicando a possibilidade de se realizar a seleção precoce. Para o índice de formato diâmetro maior/diâmetro menor, as correlações foram baixas, indicando à impossibilidade de se praticar a seleção precoce para esse caracter. Portanto, ele sugere que a seleção precoce para os caracteres estudados deve ser realizada de maneira branda e negativa, isto é, eliminando-se apenas os clones muito indesejáveis.

A influência da interação genótipo x ambiente, na expressão de diversos caracteres de batata, vem sendo descritos em vários trabalhos. Os autores relatam que as características relacionadas com a produção e qualidade em batata são influenciadas por fatores genéticos, mas com uma forte interação com o ambiente. Alguns desses caracteres são diretamente relacionados com a qualidade de processamento, como teor de açúcares redutores e coloração de chips, são influenciadas não só pelo ambiente de cultivo, mas, também, por fatores relacionados à fisiologia pós-colheita, dificultando o processo de obtenção de materiais que possuam boas características agronômicas e de qualidade industrial (PINTO, 2003; PEREIRA, 2003; AUGUSTIN, 2007; SIMON, 2005).

Hoje, o desenvolvimento de novas cultivares de batata recebe, cada vez mais, contribuição das técnicas biotecnológicas, que são importantes na limpeza clonal, na indexação de doenças e no armazenamento de germoplasma *in vitro*. Porém, os métodos baseados na biotecnologia não substituirão os de melhoramento tradicional e o melhorista deve examinar criticamente a aplicação de todas as metodologias quanto aos aspectos econômicos e a eficiência das mesmas (PEREIRA, 2003). Outra importante contribuição, para se obter o progresso genético no melhoramento de plantas, é a necessidade de estabelecer critérios de seleção artificial que sejam eficientes, portanto, análises estatísticas, que venham auxiliar na escolha da melhor estratégia de seleção tornam-se importantes ferramentas para o melhor entendimento das relações genéticas entre os caracteres (CARVALHO et al., 2001).

2.3 Considerações sobre técnicas estatísticas univariadas e multivariadas

Estabelecer relações, encontrar ou propor leis explicativas é papel próprio da ciência. Para isso é necessário controlar, manipular e medir as variáveis, que são consideradas importantes ao entendimento do fenômeno analisado. Um aspecto essencial neste processo é a avaliação estatística das informações. Quando se faz ciência, a partir da análise de variáveis isoladas, para, a seguir, fazer inferências sobre a realidade, o pesquisador geralmente utiliza-se da estatística univariada. Esta simplificação tem vantagens e desvantagens. Quando um fenômeno depende de muitas variáveis, geralmente este tipo de análise falha, pois, não basta conhecer informações estatísticas isoladas, é necessário, também, conhecer a totalidade das informações fornecida pelo conjunto das variáveis. Na análise univariada, as relações existentes entre as variáveis não são percebidas e, assim, efeitos antagônicos ou sinérgicos

entre variáveis, dificultam a interpretação do fenômeno, a partir das variáveis consideradas. Porém, no caso restrito de variáveis independentes entre si, é possível, com razoável segurança, interpretar um fenômeno complexo, usando as informações estatísticas de poucas variáveis. As informações estatísticas mais relevantes neste tipo de análise são as medidas de tendência central e de dispersão dos dados.

A denominação "Análise Multivariada" corresponde a um grande número de métodos e técnicas, que utilizam, simultaneamente, todas as variáveis, na interpretação teórica do conjunto de dados obtidos. Ela apresenta uma visão global do fenômeno, pois estuda as relações entre as variáveis e como seus efeitos determinam o comportamento destas variáveis e a totalidade do fenômeno. São muitas as técnicas de produção de uma análise multivariada e, cada uma, com sua finalidade específica. Sendo assim, a técnica a ser aplicada, vai depender daquilo que o pesquisador pretende extrair em relação aos seus dados. Algumas técnicas são utilizadas para prever, outras para otimizar, outras para resumir, cada uma com seus respectivos métodos e aplicabilidades. HAIR et al. (1998) dizem que a análise multivariada auxilia na formulação de questões complexas, de forma específica e precisa, possibilitando a condução de pesquisas teóricas significativas.

Embora, na cronologia de CORDEIRO (2009), as primeiras idéias sobre análise de dados, utilizando múltiplas respostas, tenham sido já utilizadas por Pearson em 1901, Fisher em 1928, Hotelling em 1931, Wilks em 1932 e Bartlett em 1937, as técnicas multivariadas só começaram a ser aplicadas com maior frequência, nas últimas décadas, com o avanço da tecnologia computacional e o grande número de programas computacionais estatísticos, com módulos de análise multivariada implementados. Isto mostra, na visão de MOITA NETO (2004), um paradoxo, a complexidade matemática pode ser substituída pela simplicidade didática, através do uso de *software* estatístico, sendo possível pensar estatística, sem ser estatístico. Esta aparente facilidade esbarra em alguns problemas de ordem prática. Um deles é a dificuldade do usuário em definir qual o método adequado, para solucionar o seu problema, outro, e talvez mais delicado, é que, a falta do conhecimento teórico matemático, das premissas para a utilização de certas técnicas, pode levar o pesquisador a inferências pouco confiáveis. Neste sentido, HAIR et al. (1998), recomendam um exame detalhado e rigoroso dos dados, evitando a violação dos pressupostos dos modelos utilizados, pois, em qualquer caso, os procedimentos multivariados estimam o modelo multivariado e produzem resultados, mesmo quando as suposições são severamente transgredidas.

Ao utilizar técnicas multivariadas necessita-se testar duas vezes as pressuposições, primeiro para as variáveis separadas, de modo semelhante aos testes de suposição para análise univariada e, a seguir, para as variáveis que atuam coletivamente nos modelos multivariados HAIR et al. (1998). A seguir serão comentadas algumas premissas ao uso de técnicas estatísticas multivariadas.

a) Sensibilidade ao tamanho da amostra

Aliado a um correto procedimento de amostragem, o tamanho da amostra deve ser adequado, para permitir generalizações dos resultados, os quais podem ser verificados quanto a significância dos testes. HAIR et al. (1998) recomendam, no mínimo, cinco vezes mais observações, que o número de variáveis a serem analisadas, podendo diferir, de acordo com cada técnica.

b) Influência de dados extremos

Observações atípicas, extremas ou *outliers* são aqueles valores muito diferentes dos outros. Podem ser causados por: erro na entrada de dados ou na codificação; observações extraordinárias, para as quais o pesquisador não tem explicação; ou eventos inesperados. Eles podem alterar o resultado das análises e também causar violação da normalidade dos dados. O pesquisador deve examinar os dados, em busca de observações atípicas, e eliminá-las da análise, se possível, pois seu impacto sobre os resultados gerais será desproporcional (HAIR et al., 1998). Em dados univariados, uma simples observação dos mesmos, ou um gráfico *Box-Plot*, identificará as observações extremas. Em dados multidimensionais uma observação é considerada *outlier*, se está muito distante das outras, no espaço p -dimensional, definido por estas variáveis e pode ser identificado em gráficos de dispersão tridimensionais, juntamente com a distância de Mahalanobis e gráficos do tipo *Q-Q plot* (MINGOTI, 2005).

c) Linearidade das relações

De modo geral os modelos lineares prevêem valores que se ajustam a uma linha reta. A simples observação de gráficos, como os diagramas de dispersão, é um método para a verificação da linearidade (HAIR et al., 1998).

d) Normalidade

Uma das premissas mais importantes, para a utilização de técnicas multivariadas, é a pressuposição da normalidade dos dados, para as variáveis métricas (HAIR et al., 1998), embora, JOHNSON & WICHERN (1992) alertem que, para dados reais, a presença de variáveis com distribuição normal multivariada exata, dificilmente ocorre. Assim, a densidade normal é, freqüentemente, uma aproximação útil à verdadeira distribuição da população. Uma transformação de dados adequada é uma alternativa para a violação do pressuposto de normalidade multivariada de dados (JOHNSON & WICHERN, 1992). Se uma variável é normal multivariada, também é normal univariada, embora a recíproca nem sempre seja verdadeira. O diagnóstico mais simples, da normalidade univariada, é a observação do histograma, em que os dados observados seguem uma distribuição aproximadamente normal, porém, mais utilizados e mais confiáveis são os testes estatísticos, como, por exemplo, os testes Lilliefors e Kolmogorov-Smirnov, disponíveis na maioria dos *softwares* estatísticos (HAIR et al., 1998). A normalidade multivariada pode ser verificada pelo gráfico de probabilidade Q-Q *plot*. Quando a normalidade *p*-variada se ajusta bem aos dados amostrais, este gráfico deve resultar em uma nuvem de pontos próximos a uma reta. Curvas diferentes da reta revelam ausência de normalidade (MINGOTI, 2005).

e) Homocedasticidade

Esta suposição relaciona-se, primeiramente, a relações de dependência entre variáveis. Refere-se à suposição de que as variáveis dependentes devem ter mesma variância, ao longo do domínio das variáveis preditoras. A verificação da homogeneidade de variâncias univariada pode ser realizada pelos testes Cochran ou Bartlett. No caso multivariado, o teste mais utilizado, para verificar a igualdade das matrizes de variância/covariância, é o teste de Box, que é uma generalização do teste de igualdade de variâncias de Bartlett. O teste de Box é muito sensível a desvios da normalidade, assim, deve-se, primeiro, verificar a normalidade, para depois executá-lo. A falta de homogeneidade pode ser contornada através de transformações de dados, semelhantes às empregadas para obter a normalidade dos dados (HAIR et al., 1998).

f) Multicolinearidade

Duas ou mais variáveis são colineares, se possuem relação exata, ou seja, se um dos vetores é uma combinação linear dos outros. A correlação exata raramente ocorre, mas correlações fortes já são perigosas. Além da observação da magnitude da correlação entre as variáveis, outra sugestão, para identificar o grau de multicolinearidade, é análise dos autovalores associados a matriz de correlação. Um autovalor nulo, ou muito pequeno, indica a existência de multicolinearidade. Quando a multicolinearidade aumenta, a habilidade de definir quaisquer efeitos das variáveis diminui. Também se observa, que alguns estimadores atingem valores muito altos, evidenciando uma estimativa pouco confiável (HAIR et al., 1998). A correção recomendada é retirar uma, ou mais, variáveis independentes, altamente correlacionadas. Outra alternativa, sugerida em CRUZ & CARNEIRO (2003), é retirar as variáveis correspondentes aos maiores valores, localizados nos dois últimos autovetores, associados a matriz das correlações.

Além das considerações anteriores, sobre as premissas, para o uso de técnicas univariada ou multivariada, deve-se considerar o efeito da taxa do erro tipo I (α), nos dois casos. Apenas lembrando, o erro tipo I é a probabilidade de se rejeitar a hipótese nula (H_0), quando ela é verdadeira. A utilização de testes estatísticos separados, para cada variável, provoca um erro tipo I global muito elevado. Por exemplo, na realização de 5 testes univariados, admitindo-se que sejam independentes, cada um com uma significância de 0,05, a probabilidade de se rejeitar, pelo menos uma hipótese nula é: $1 - P(\text{não rejeitar nenhuma das hipótese nulas}) = 1 - (0,95 \times 0,95 \times 0,95 \times 0,95 \times 0,95) = 0,227 \cong 23\%$, que é um valor inaceitável (HAIR et al., 1998). Assim, as decisões tomadas, a partir de resultados de muitos testes univariados, ou de um teste multivariado, poderão ser muito diferentes, ou mesmo, contraditórias, devendo o pesquisador estar atendo, na hora de decidir que tipo de análise realizará.

2.4 Correlações

Em produção vegetal, assim como em outras áreas de estudo, muitas vezes se faz necessário medir a existência e/ou a intensidade da relação entre caracteres. Dentre as

metodologias disponíveis, as mais utilizadas são: coeficiente de correlação de Pearson; coeficiente de correlação parcial; regressão simples; regressão múltipla; análise de trilha e correlação canônica. Nenhuma dessas ferramentas é hierarquicamente mais importante ou informativa do que a outra, apenas, para sua utilização, devem ser considerado o tipo de associação entre os caracteres que se deseja medir ou relacionar e os requisitos exigidos na utilização de cada método.

2.4.1 Coeficiente de Correlação Linear de Pearson

É a medida que dá o grau de intensidade entre duas variáveis aleatórias contínuas X e Y, com distribuição normal bivariada e que possuam uma relação linear de causa e efeito entre si. O coeficiente de correlação de Pearson, representado por ρ , é adimensional e pode variar de -1 até 1. O valor do coeficiente igual a zero revela a ausência de relação linear entre as variáveis. O sinal positivo indica uma relação direta entre as variáveis, enquanto que, o sinal negativo aponta para uma associação inversa entre elas, isto é, valores altos de uma variável estão associados a valores baixos da outra variável, e vice-versa. Como os dados, para o cálculo da correlação entre variáveis, normalmente são retirados de uma amostra aleatória, o coeficiente ρ , pode ser estimado por:

$$r_{xy} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \quad \mathbf{01}$$

Sendo:

$\text{cov}(X,Y)$ = covariância entre as variáveis aleatórias X e Y;

$\text{var}(X)$ = variância da variável aleatória X;

$\text{var}(Y)$ = variância da variável aleatória Y.

Para a realização do teste de significância da correlação linear de Pearson utiliza-se a estatística t_c :

$$t_c = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad \mathbf{02}$$

Com r_{xy} = correlação estimada e n = número de elementos da amostra. Aceita-se a hipótese nula ($H_0: \rho = 0$) se $|t_c| < t_{(\alpha; n-2)}$ (COSTA NETO, 2002).

Um coeficiente de correlação significativo é, apenas, um indicativo de associação entre variáveis, não fornecendo subsídio algum da relação de causa e efeito. Da mesma forma,

coeficientes de correlação baixo não representam falta de associação entre caracteres, mas sim a inexistência de relação casual (CARVALHO et al., 2004).

2.4.2 Correlação parcial

Ao investigar a relação entre três variáveis X_1 , X_2 e X_3 , pode-se calcular o coeficiente de correlação de Pearson para cada par de variáveis. Essas correlações são representadas por r_{12} , r_{13} e r_{23} . O r_{12} , por exemplo, mede a correlação total existente entre as variáveis X_1 e X_2 , englobados os efeitos que X_3 possa ter causado sobre o comportamento das mesmas. Assim, é possível, que a utilização do coeficiente de correlação simples, apenas entre duas variáveis, sem considerar o efeito que a terceira possa provocar sobre elas, traga um resultado impreciso. A correlação parcial, então, é a medida de correlação utilizada quando se deseja considerar a relação existente entre duas variáveis, descontando o efeito da terceira. Isso equivale, por exemplo, a calcular a correlação entre X_1 e X_2 , mantendo X_3 constante, que se anota por $r_{12.3}$ e pode ser estimado por:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad \mathbf{03}$$

De modo análogo esse conceito pode ser estendido para mais variáveis (CRUZ et al., 1994).

Para a utilização desta análise é necessário observar, que os dados sigam uma distribuição normal multivariada, mais as outras suposições da correlação linear de Pearson (CARVALHO et al., 2004).

O teste de significância é similar ao utilizado para a correlação linear de Pearson. A estatística utilizada é:

$$t_c = \frac{r \sqrt{n - n_x - 2}}{\sqrt{1 - r^2}} \quad \mathbf{04}$$

Com: r = correlação parcial estimada entre duas variáveis; n = número de observações da qual o coeficiente de correlação simples foi calculado; n_x = número de variáveis

independentes. Aceita-se a hipótese nula ($H_0: \rho = 0$) se $|t_c| < t_{(\alpha; n-n_x-2)}$ (CARVALHO et al., 2004).

2.4.3 Análise de regressão

A análise de regressão e a correlação são duas técnicas estreitamente relacionadas, que envolvem uma forma de estimação. Essas técnicas analisam dados amostrais, para mostrar como duas ou mais variáveis estão relacionadas, uma com a outra, em certa população. Enquanto a correlação dá o número que resume a magnitude, ou o grau de relacionamento entre as variáveis, a análise de regressão fornece um modelo matemático, que descreve esse relacionamento. Esse modelo pode ser utilizado para estimar ou prever valores futuros de uma variável, quando se conhecem, ou se supõe conhecidos, os valores de outras variáveis (COSTA NETO, 2002).

a) Análise de regressão linear simples

É uma tentativa de estabelecer, como modelo de relação entre duas variáveis, uma linha reta. Na regressão linear, os valores de Y são preditos com base em valores dados, ou conhecidos, de X. Assim, dados n pares de duas variáveis X_i e Y_i ($i=1, \dots, n$) e admitindo-se que Y é uma função linear de X, a função que se deseja obter será da forma:

$$Y = \alpha + \beta X + \varepsilon \quad 05$$

Pode-se, então, estabelecer um modelo estatístico da forma:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad 06$$

onde Y_i é a variável dependente, X_i é a variável independente, α e β são os parâmetros e ε_i representa a influência de outros fatores e o erro na medição da variável Y.

Para a utilização do modelo de regressão linear simples há a necessidade de que sejam observados os seguintes pressupostos:

1. A relação entre X e Y é linear;
2. Os valores de X são fixados (controlados), isto é, X não é uma variável aleatória;

3. Os erros têm distribuição normal, com média zero e variância comum σ^2 , $\varepsilon_i \sim N(0, \sigma^2)$;
4. Os erros não possuem autocorrelação, $\text{cov} [\varepsilon_i, \varepsilon_j] = 0$, para $i \neq j$, e $i, j = 1, \dots, n$.

A partir dos valores amostrais obtidos é possível estimar os parâmetros α e β , utilizando o método dos mínimos quadrados, e obtendo-se o estimador de Y:

$$\hat{Y} = A + B X \quad \mathbf{07}$$

Segue-se o teste de hipótese para validar a estimativa do parâmetro de β e o cálculo do intervalo de confiança para o modelo estimado.

No caso do teste confirmar que o modelo é linear, é importante calcular o coeficiente de determinação (r^2) para verificar a qualidade de estimação do modelo (COSTA NETO, 2002).

b) Análise de regressão linear múltipla

A regressão múltipla envolve três ou mais variáveis. Há ainda uma única variável dependente, porém duas ou mais variáveis independentes. A teoria é uma extensão da análise de regressão simples. Aqui, também, a análise tem por objetivo encontrar um modelo, que possa ser utilizado para prever valores de Y, para valores dados das diversas variáveis independentes. O objetivo das diversas variáveis independentes é melhorar o modelo na sua capacidade de predição, comparando-o com o modelo linear simples.

O modelo de regressão linear múltiplo, com k variáveis independentes e β_j ($j=0, \dots, k$) parâmetros, pode ser escrito como:

$$Y_I = \beta_0 + \beta_1 X_{I1} + \beta_2 X_{I2} + \dots + \beta_k X_{Ik} + \varepsilon_I \quad I=1,2,\dots, N \quad \mathbf{08}$$

Onde $\varepsilon_i \sim N(0, \sigma^2)$

Esse modelo pode ser escrito na forma matricial como:

$$Y = X\beta + \varepsilon \quad \mathbf{09}$$

Sendo Y a matriz ($n \times 1$) das observações aleatórias, X a matriz ($n \times k$) das variáveis independentes, β é matriz ($k \times 1$) dos coeficientes de regressão e ε é a matriz ($n \times 1$) dos erros aleatórios.

A estimação dos parâmetros pode ser realizada pelo método dos mínimos quadrados ou de máxima verossimilhança. Após a estimação dos parâmetros, há a necessidade de testar a significância da regressão e realizar um teste individual para cada coeficiente, a fim de validar o modelo. O cálculo do coeficiente de determinação múltiplo (R^2), dá a medida do ajuste da regressão (HAIR, 1998).

2.4.4 Análise de trilha

A análise de trilha estuda os efeitos diretos e indiretos de caracteres sobre uma variável básica, onde as estimativas são obtidas por meio de equações de regressão, em que as variáveis são anteriormente padronizadas. No caso de ser considerado apenas um modelo casual, a análise de trilha é apenas uma análise de regressão parcial padronizada. Entretanto, quando estão envolvidas várias inter-relações complexas, a análise de trilha constitui-se em uma expansão da regressão múltipla, da forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_I X_I + \varepsilon, \quad I=1,2,\dots, N \quad 10$$

Após a padronização e estimação dos coeficientes pelo método dos mínimos quadrados, pode ser escrita como:

$$Y = P_1 X_1 + P_2 X_2 + P_3 X_3 + \dots + P_I + P_e \varepsilon, \quad I=1,2,\dots, N, \quad 11$$

$$Y = P_1 X_1 + P_2 X_2 + \dots + P_n X_n + P_e E$$

Onde:

$$P_e = \sqrt{1 - R^2}$$

Os coeficientes P_i são chamados de coeficientes de trilha. Por este modelo estimam-se os efeitos diretos e indiretos das variáveis explicativas sobre a variável base (CRUZ et al., 1994).

2.4.5 Correlação canônica

A correlação canônica pode ser vista como uma extensão da regressão múltipla. Ela permite avaliar as inter-relações entre dois conjuntos de variáveis, compostos por combinações lineares dos vários caracteres que os constituem. A equação básica pode ser expressa por:

$$Y_1 + Y_2 + Y_3 + \dots + Y_p = X_1 + X_2 + X_3 + \dots + X_q \quad 12$$

O princípio da correlação canônica é desenvolver uma combinação linear, em cada um dos conjuntos de variáveis, de modo que a ela seja maximizada. Nesse método não existe distinção entre variável independente e dependente, existem somente dois conjuntos de variáveis, em que se busca a máxima correlação entre ambos (MORRISON, 1978).

Assim, o objetivo da correlação canônica é determinar uma combinação linear para cada grupo de variáveis (dependentes e independentes) que maximize a correlação entre os dois grupos. Funções adicionais, não correlacionadas com a primeira, podem ser calculadas, sendo limitadas pelo número de variáveis do menor grupo (CRUZ & REGAZZI, 1994).

Para expressar a correlação canônica determina-se a combinação linear entre X e Y,

$$U = A_1X_1 + A_2X_2 + \dots + A_MX_M \quad 13$$

$$V = B_1Y_1 + B_2Y_2 + \dots + B_NY_N \quad 14$$

de modo que a correlação $Corr(U, V)$ seja maximizada.

Supondo que X seja uma matriz $m \times p$ e Y $n \times q$, então $C = cov(X, Y)$ ou seja, separando a matriz C em quatro partes tem-se:

$$C = \begin{bmatrix} \sum_{pxp}^{11} & \sum_{pxq}^{12} \\ \sum_{qxp}^{21} & \sum_{qxq}^{22} \end{bmatrix} \quad 15$$

As covariâncias entre variáveis de diferentes conjuntos, uma variável de X e outra de Y estarão contidas em Σ_{12} ou Σ_{21} . Analisar as covariâncias em Σ_{12} ou Σ_{21} pode ser trabalhoso, ainda mais se p e q forem grandes. Porém, o principal objetivo da correlação canônica é resumir as associações entre X e Y em função de algumas poucas correlações escolhidas, ao invés das pxq correlações (CRUZ & REGAZZI, 1994).

A combinação linear é uma forma simples de resumir o conjunto de variáveis. Se

$$U = A'X \quad 16$$

$$V = B'Y \quad 17$$

$$Var(U) = a' Cov(X)a = a' \sum_{11} a \quad 18$$

$$Var(V) = b' Cov(Y)b = b' \sum_{22} b \quad 19$$

$$Cov(U, V) = a' Cov(X, Y)b = a' \sum_{12} b \quad 20$$

a correlação canônica busca a determinação dos vetores a e b , tais que,

$$Corr(U, V) = \frac{a' \sum_{12} b}{\sqrt{a' \sum_{11} a} \sqrt{b' \sum_{22} b}} \quad 21$$

seja a maior possível.

Existirão $\min(p, q) - 1$ pares de variáveis canônicas independentes do par de correlação máxima, que irão expressar a variância total dos dois grupos de variáveis (JOHNSON & WICHERN, 1992).

Para calcular este máximo

$$\max_{a, b} Corr(U, V) = \rho_1 \quad 22$$

restrito pela combinação linear (primeiro par de variáveis canônicas):

$$U_1 = e_1' \sum_{11}^{-1/2} X = a_1' \quad 23$$

$$V_1 = f_1' \sum_{22}^{-1/2} Y = b_1' \quad 24$$

Neste caso, ρ_1 é o autovalor de

$$\sum_{11}^{-1/2} \sum_{12} \sum_{22}^{-1} \sum_{21} \sum_{11}^{-1/2} \quad 25$$

e e_1 é seu respectivo autovetor.

Supondo A uma matriz quadrada de ordem k e v um vetor $k \times 1$, pode-se mostrar que a equação $Av = \lambda v$ tem k soluções, sendo λ um escalar. Cada solução é dada por um par formado por um escalar λ_i e um vetor v_i . O escalar é chamado de autovalor de A e o vetor de autovetor de A (JOHNSON & WICHERN, 1992).

O k -ésimo par de variáveis canônicas pode ser descrito como:

$$U_k = e_k' \sum_{11}^{-1/2} X \quad 26$$

$$V_k = f_k' \sum_{22}^{-1/2} Y \quad 27$$

maximizando a correlação

$$Corr(U_k, V_k) = \rho_k \quad 28$$

O teste de significância para todas as correlações canônicas é:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0, s = \min \{p, q\}$$

$$H_1 : \rho_i \neq 0 \text{ para algum } i; 1 \leq i \leq s$$

A estatística do teste é:

$$\chi^2 = - [n - 1 - 0,5(p + q + 1)] \ln \left[\prod_{i=1}^s (1 - r_i^2) \right] \quad 29$$

Que segue a distribuição $\chi^2_{(\alpha; pq)}$.

Sendo:

n = número de observações;

p = número de variáveis do grupo X;

q = número de variáveis do grupo Y;

r_i^2 = correlação canônica, ao quadrado, da equação a ser testada.

Se a hipótese é rejeitada faz-se o teste:

$$H_0 : \rho_k > 0 \text{ e } \rho_{k+1} = \rho_{k+2} = \dots = \rho_s = 0, s = \min \{p, q\} \text{ por meio de}$$

$$\chi^2 = - [n - 1 - 0,5(p + q + 1)] \ln \left[\prod_{i=k+1}^s (1 - r_i^2) \right] \quad 30$$

Que segue a distribuição $\chi^2_{[\alpha; (p-k)(q-k)]}$.

Em produção vegetal pode-se destacar alguns trabalhos que foram realizados utilizando a correlação canônica. Por exemplo, TAVARES et al. (1999) utilizaram a análise de correlação canônica com o objetivo de estudar as relações entre os principais fatores de produção em pimentão (peso e número de frutos) e os caracteres do fruto. Concluíram que esta correlação foi significativa e que, o número total de frutos foi o caracter que apresentou maior efeito positivo sobre o peso total de frutos.

COIMBRA et al. (2000), com o objetivo de estudar as relações entre os componentes primários e secundários do rendimento de grãos no feijão, também utilizaram a análise de correlação canônica. Os autores concluíram que os componentes primários e secundários do rendimento de grãos de feijão preto não são independentes. A principal conclusão de associação é que plantas com maior estatura e maior ciclo reprodutivo influenciam negativamente no número de grãos por legume.

Para verificar a existência e a intensidade de associação entre as características químicas, físicas e dimensionais das fibras da madeira em clones de *Eucalyptus grandis* e *Eucalyptus saligna*, TRUGILHO et al. (2003) utilizaram a correlação canônica. Foi

encontrada uma associação significativa entre os grupos de variáveis, indicando que a madeira de alto teor de lignina insolúvel e baixo teor de cinzas está associada com a madeira que possui elevada contração radial e tangencial e alta densidade básica.

2.5 Análise de agrupamento

A análise de agrupamento ou análise de conglomerados ou, ainda, análise de *cluster*, consiste em método estatístico multivariado, com o objetivo de agrupar elementos, de uma amostra ou população, de modo que esses grupos possuam semelhança entre as características, ou variáveis, que neles foram medidas. Assim, grupos heterogêneos possuem elementos diferentes em relação às mesmas características. Essa técnica apresenta a vantagem de reduzir o espaço multidimensional, a uma medida de distância entre os elementos medidos, representando-os em um espaço bidimensional, muito mais simplificado. Como resultado da análise de agrupamento, tem-se o dendograma, que apresenta o arranjo entre os objetos em uma escala de distância. Este arranjo mostra apenas a afinidade entre os grupos, não definindo nenhuma ordenação entre eles (JOHNSON & WICHERN, 1992).

Na análise de agrupamento os valores das distâncias entre os objetos são organizados na forma de matrizes. Não há necessidade da estimação de parâmetros, daí o seu caráter não-probabilístico. Deste modo, sua interpretação é, muitas vezes, subjetiva. A representação gráfica de similaridade ou dissimilaridade entre os objetos e, mais especificamente, entre grupos de objetos afins, contribui como forte critério de decisão. Assim, o reagrupamento de um conjunto de dados, ou observações, em subgrupos homogêneos é o objetivo principal desta análise, definindo-se, uma maior homogeneidade dentro do subconjunto e maior heterogeneidade em relação a outros subconjuntos (JOHNSON & WICHERN, 1992).

As distâncias aplicáveis em níveis de mensuração não-nominais ou quantitativas podem ser divididas em distâncias métricas, associadas exclusivamente com a medida vetorial das variáveis mensuradas, não incluindo nenhuma medida de variação; e distâncias estatísticas, incluindo medidas de variação. As distâncias métricas são definidas, como a menor distância entre os objetos, que são representados pelas realizações das variáveis em um espaço multidimensional. A variabilidade dentro e entre as variáveis mensuradas é desconsiderada. De modo geral, as distâncias métricas são utilizadas nos casos em que somente uma unidade é utilizada.

A distância euclidiana é uma distância métrica, utilizada para a representação dos pontos, na estrutura de similaridade. Ela representa o menor espaço entre dois pontos, sendo uma extensão do teorema de Pitágoras para o caso multidimensional. Sejam dois indivíduos no espaço p-dimensional, representados pelos vetores $X_1^t = [X_{11} \ X_{12} \ \dots \ X_{1p}]$ e $X_2^t = [X_{21} \ X_{22} \ \dots \ X_{2p}]$, a distância euclidiana entre eles será:

$$d(X_1, X_2) = \sqrt{(X_1 - X_2)^t (X_1 - X_2)} \quad 31$$

As distâncias estatísticas levam em consideração a inclusão de uma medida de variabilidade entre as variáveis avaliadas. O uso de distâncias estatísticas reduz os efeitos de escala, já que estas distâncias ponderam as diferenças entre objetos, pelo efeito da variação no atributo, reduzindo consideravelmente os possíveis efeitos de escala, que tornam as distâncias métricas não indicadas em casos em que variáveis de diferentes unidades são utilizadas (JOHNSON & WICHERN, 1992).

Uma distância estatística muito utilizada é a distância de Mahalanobis, definida por:

$$d(X_1, X_2) = \sqrt{(X_1 - X_2)^t S^{-1} (X_1 - X_2)} \quad 32$$

Sendo $X_1^t = [X_{11} \ X_{12} \ \dots \ X_{1p}]$ e $X_2^t = [X_{21} \ X_{22} \ \dots \ X_{2p}]$, vetores de medidas correspondentes a dois indivíduos no espaço p-dimensionais e S^{-1} a inversa da matriz de variância/covariância amostral destes indivíduos (JOHNSON & WICHERN, 1992).

Nos procedimentos de classificação, as técnicas de conglomerados, ou *cluster*, são frequentemente classificadas em dois tipos: hierárquicos e não-hierárquicos.

Os métodos hierárquicos apresentam como resultados séries de agrupamentos em uma escala de afinidade, partindo do pressuposto de que o conjunto de dados é um único grande conjunto e cada objeto forma ou formará um subconjunto próprio.

Os métodos hierárquicos subdividem-se em aglomerativos ou divisivos.

As técnicas hierárquicas aglomerativas descrevem a orientação do agrupamento, partindo do princípio de que cada objeto é um agrupamento natural, posteriormente reunindo-se a outros de maior afinidade através de fusões de n objetos, que sucessivamente são reunidos até formar o supra-agrupamento, que é o conjunto de objetos como um todo. Devido

a propriedade aglomerativa é possível construir um gráfico chamado dendograma ou dendrograma, que representa a história do agrupamento (MINGOTI, 2005).

O dendograma é um gráfico, em forma de árvore. O eixo vertical indica o nível de similaridade (ou dissimilaridade) e o eixo horizontal indica os elementos amostrais numa ordem conveniente, relacionada a história do grupo. As linhas verticais, partindo dos elementos amostrais, têm altura correspondente ao nível em que os elementos foram considerados semelhantes, isto é, a distância do agrupamento ou o nível de similaridade (MINGOTI, 2005).

Ligação simples, ou método do vizinho mais próximo, é o procedimento que utiliza a distância mínima entre dois objetos de um conjunto n , de grupos distintos como sendo a distância entre os grupos. O próximo grupo é representado pela menor distância entre o primeiro grupo determinado e o objeto mais próximo a este. Os passos seguem-se até o encadeamento de todos os objetos em um único agrupamento, este com diferentes arranjos de objetos em um dado nível da escala de distâncias.

Ligação completa ou método do vizinho mais distante é exatamente oposto ao da ligação simples, em que no primeiro passo considera-se a distância entre dois grupos como sendo a distância entre os objetos de maior distância, estes definindo grupos polarizados. Com a redução das distâncias entre os grupos e objetos, estes passam a formar agrupamentos com menor distância, encadeando-se.

Ligação média trata de uma variação dos procedimentos descritos anteriormente, sendo que neste, a distância entre dois grupos é representada pela média da distância entre todos os pares de objetos pertencentes a cada grupo. Vários algoritmos são propostos para a condução deste procedimento.

O método de Ward é baseado na redução da informação resultante, dada a inclusão de um conjunto de objetos em um grupo. Esta redução de informação é determinada pela soma total do quadrado do erro de cada objeto, em função da média do grupo a que este, supostamente, pertença. Esta regra de inclusão envolve todos os pares possíveis, sendo definidos como pertencente a um dado grupo o objeto que contribua o mínimo com o aumento da soma de quadrado do erro (MINGOTI, 2005).

Os procedimentos hierárquicos divisivos descrevem a orientação do agrupamento a partir de um supra-agrupamento, representado pelo conjunto de objetos, que é dividido em agrupamentos subseqüentes, de menor afinidade, até o retorno ao objeto.

Dentre os métodos divisivos, têm-se o método da distância média subdivisora, que consiste na divisão do conjunto de dados em dois grupos em um número de combinações 2^{n-1} . Define-se, então, qual dos pares apresenta a maior dissimilaridade. A este, o subdivisor, são adicionados seqüencialmente os objetos de maior dissimilaridade, até que os grupos possam ser polarizados em torno deste subdivisor.

Os procedimentos não-hierárquicos ou particionais resultam em um arranjo dos objetos em um número de agrupamentos pré-definido. Estes procedimentos podem ser do tipo seriais, nos quais um objeto é alocado por vez, ou do tipo simultâneo, em que todos os objetos são alocados ao mesmo tempo. Os métodos de partição definem uma posição definitiva para os objetos no decorrer da sua condução, primando exclusivamente pelos critérios estabelecidos no início destas, no caso a determinação do número de grupos. Algumas técnicas que representam este método são baseadas em propriedades da matriz de soma de quadrados da análise de variância (MINGOTI, 2005).

Outras técnicas baseiam-se na matriz de soma de quadrados e produtos da análise de variância multivariada, composta de uma submatriz de efeito entre os tratamentos (B) e outra matriz de efeito dentro dos tratamentos (W) (JOHNSON & WICHERN, 1992).

A técnica de análise de agrupamento, conforme CRUZ & CARNEIRO (2003), pode ser utilizada no início de programas de melhoramento, como auxílio na definição de estratégias de trabalho. Sabe-se que a variabilidade genética, criada ou espontânea, é o ponto de partida de qualquer programa de melhoramento. A análise de agrupamento pode sugerir a escolha de parentais mais dissimilares, associados a alta variabilidade dos caracteres de interesse do pesquisador, levando a geração de uma população segregante com alta probabilidade de apresentar características de interesse, reduzindo tempo e custos em um programa de melhoramento. Assim, esta técnica foi escolhida por SILVA et al. (2008) para reduzir o número de caracteres necessários à classificação das progênies de batata, sendo que, o método também indicou que o formato de tubérculo foi o caracter com maior eficiência para o estudo de dissimilaridade. OLIVEIRA et al. (2000) para avaliar a divergência genética entre clones de batata-doce, de várias regiões brasileiras, também fizeram uso da análise de agrupamento. Enquanto, VIDIGAL et al. (1997), trabalhando com cultivares de mandioca, a utilizaram, na etapa inicial de um programa de melhoramento, para avaliar a contribuição de algumas características morfoagronômicas para a variabilidade total.

2.6 Análise de variância multivariada (MANOVA)

Nos experimentos mais comuns, cada parcela fornece o valor de uma só variável a ser estudada pelos métodos estatísticos tradicionais. Mas há casos em que cada parcela do experimento dá observações relativas a duas ou mais variáveis a serem analisadas. Se essas variáveis são independentes, pode-se realizar uma análise para cada uma delas, de forma independente. Porém, o mais comum, é que haja uma correlação entre essas variáveis. Nestes casos uma análise univariada é inadequada, pois seus resultados serão comprometidos pelo grau da correlação existente entre essas variáveis. Em casos como estes uma análise multivariada é o recomendado, pois essa análise levará em consideração a estrutura de correlação existente entre as variáveis envolvidas na análise (GOMES, 1985).

A análise de variância, quando aplicada a todas as variáveis, simultaneamente, é denominada de análise de variância multivariada (MANOVA). Da mesma forma que a análise de variância univariada (ANOVA), ela tem o propósito de fazer uma análise de variação dos dados, entre ou dentro de tratamentos, para tirar conclusões sobre possíveis diferenças entre as médias destes tratamentos. Recomenda-se o uso da análise de variância multivariada, quando existe uma estreita correlação entre as variáveis, caso contrário, as análises univariadas de cada variável, resolvem o problema. Enquanto a ANOVA testa a hipótese de igualdade das médias de tratamentos, MANOVA testa a igualdade entre os vetores de médias, das múltiplas variáveis dependentes (HAIR, 1998).

Os pressupostos, para a utilização da MANOVA, é de que os dados sigam uma distribuição normal multivariada, os tratamentos devem ser independentes e com mesma matriz de variância e covariância. No teste de hipóteses para a igualdade de médias, na MANOVA, a hipótese nula a ser testada é a da igualdade das médias de tratamento, para um conjunto de p variáveis, simultaneamente (HAIR, 1998).

Supondo que se deseja comparar g tratamentos, sendo:

$X_{11}, X_{12}, \dots, X_{1g}$, vetor de dados do tratamento 1

$X_{21}, X_{22}, \dots, X_{2g}$, vetor de dados do tratamento 2

.....

$X_{g1}, X_{g2}, \dots, X_{g^g}$, vetor de dados do tratamento g

O modelo teórico para o delineamento inteiramente casualizado é

$$\tilde{X}_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad 33$$

com:

$i = 1, 2, \dots, g$, tratamentos;

$j = 1, 2, \dots, n_i$, repetições no tratamento i ;

\tilde{X}_{ij} = vetor de observações do tratamento i na repetição j ;

μ = vetor de médias geral;

τ_i , = efeito do i -ésimo tratamento;

ε_{ij} = erro.

O modelo estimado será:

$$\tilde{X}_{ij} = \bar{X} + (\bar{X}_i - \bar{X}) + (\tilde{X}_{ij} - \bar{X}_i)$$

Observação = estimativa da média geral + estimativa do efeito de tratamento + erro

sendo:

\tilde{X}_{ij} = vetor de observações do tratamento i na repetição j ;

\bar{X} = vetor de médias geral;

\bar{X}_i = vetor de médias do i -ésimo tratamento.

Desenvolvendo o modelo estimado encontram-se as matrizes, soma de quadrados e produtos total (T), soma de quadrados e produtos de tratamentos (B) e soma de quadrados e produtos dos erros (E), definidas a seguir.

$$T = \sum_{i=1}^g \sum_{j=1}^{n_i} (\tilde{X}_{ij} - \bar{X})(\tilde{X}_{ij} - \bar{X})^T \quad 34$$

$$B = \sum_{i=1}^g \left[n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T \right] \quad 35$$

$$E = \sum_{i=1}^g \sum_{j=1}^{n_i} (\tilde{X}_{ij} - \bar{X}_i)(\tilde{X}_{ij} - \bar{X}_i)^T \quad 36$$

$$T = B + E \quad 37$$

Para uma MANOVA, no delineamento inteiramente casualizado, o quadro da análise da variância será como o da Tabela 7.

Tabela 07 – Análise da variância multivariada no delineamento inteiramente casualizado

Causas de variação	Graus de liberdade	Matrizes
Tratamentos	$g - 1$	B
Erro	$n - g$	E
Total	$n - 1$	$T = B + E$

Sendo, g = número de tratamentos e $n = \sum_{i=1}^g n_i$, B, E e T as matrizes definidas anteriormente em 34, 35, 36 (JOHNSON & WICHERN, 1992).

Alguns dos critérios utilizados, para avaliar as diferenças entre os tratamentos na MANOVA são: traço de Hotelling-Lawley, traço de Pillai, maior raiz de Roy e lambda de Wilks. A seguir descreve-se a estatística referente ao critério Lambda de Wilks, que é o mais encontrado nas análises de variância multivariada (JOHNSON & WICHERN, 1992).

O critério de Wilks ou Lambda de Wilks é definido pela estatística:

$$\Lambda = \frac{\det E}{\det(B + E)} \quad 39$$

As matrizes B e E estão definidas em 35 e 36, respectivamente.

A aproximação da distribuição F é:

$$\left(\frac{1 - \Lambda^{\frac{1}{t}}}{\Lambda^{\frac{1}{t}}} \right) \left(\frac{rt - 2f}{pq} \right) \cap F(pq; rt - 2f) \quad 40$$

Onde:

p = número de variáveis;

q = número de graus de liberdade de tratamento;

v = número de graus de liberdade do erro;

$$r = v - \frac{p - q + 1}{2};$$

$$t = \sqrt{\frac{p^2 q^2 - 4}{p^2 + q^2 - 5}}, \quad \text{se } p^2 + q^2 - 5 > 0;$$

$t = 1$, nos outros casos;

$$f = \frac{pq - 2}{4}.$$

Devido as suas características e com a disponibilidade de programas computacionais mais amigáveis, alguns pesquisadores já utilizam a MANOVA na análise de seus dados. VIEIRA et al. (2007), estudando a leucena, submetida a diferentes espaçamentos e alturas de corte, optaram pelo uso da MANOVA, ao invés de parcelas subdivididas, por considerar que as medidas tomadas ao longo do tempo, na mesma unidade experimental são correlacionadas entre si. LEDO et al. (2003) utilizaram MANOVA para estimar a heterose em várias características, simultaneamente. Enquanto que, VIANA et al. (2007), no início de seu trabalho, submeteram as características de frutos de maracujazeiro-amarelo a MANOVA e, a partir das informações geradas, utilizaram outras técnicas multivariadas para avaliar a influência da planta doadora de pólen, sobre as características dos frutos produzidos.

3 METODOLOGIA

Foram realizados três ensaios com batatas, na área experimental do Departamento de Fitotecnia da Universidade Federal de Santa Maria, na cidade de Santa Maria, Rio Grande do Sul. Época 1 em latitude 29°43' 28"S, longitude 53°43'11"W e 102 m de altitude, época 2 em latitude 29°43'29"S, longitude 53°43'17"W e 105 m de altitude e época 3 em latitude 29°43'35"S, longitude 53°43'18"W e 107 m de altitude. O solo pertence a unidade de mapeamento Santa Maria, classificado como Brunizem Hidromórfico pelo Sistema de Brasileiro de Classificação de Solos (EMBRAPA,1999). O clima da região, segundo a classificação de Köeppen (MORENO, 1961) é do tipo Cfa – temperado chuvoso, com chuvas bem distribuídas ao longo do ano e subtropical do ponto de vista térmico.

Os tubérculos de batata utilizados foram progênies, resultantes de 10 cruzamentos do Programa de Genética e Melhoramento da Batata da UFSM (Tabela 08). Estes tubérculos correspondem à primeira geração clonal dos cruzamentos. O plantio da primeira época ocorreu em 14 de março de 2006. Plantou-se 100 tubérculos, amostrados nas dez progênies relacionadas na Tabela 08, conforme um delineamento inteiramente ao acaso. Casualizou-se um tubérculo por repetição (cova), devidamente identificado, em quatro filas, com 80cm de distância entre as filas e 33cm entre plantas.

O manejo da cultura, incluindo preparo do solo, adubação, tratos culturais e colheita realizaram-se segundo as recomendações técnicas de BISOGNIN (1996). Antes do plantio, avaliou-se em cada tubérculo os caracteres: comprimento, maior diâmetro, menor diâmetro e massa, constituindo-se, assim, o primeiro grupo de caracteres (G1) para a realização da análise de correlação canônica.

A colheita dos tubérculos, de cada cova com produção, ocorreu em 15 de junho de 2006, obtendo-se um total de 261 tubérculos em 71 covas. Dos tubérculos colhidos avaliou-se os mesmos caracteres do tubérculo-semente, com estimação das médias entre os tubérculos de cada cova, formando os cinco caracteres do segundo grupo (G2): número de tubérculos por cova, comprimento (média), maior diâmetro (média), menor diâmetro (média) e massa (média). Após as análises, escolheu-se três tubérculos de cada cova, quando isto foi possível, um de tamanho grande, outro de tamanho pequeno e outro de tamanho intermediário, para a

próxima etapa do experimento. Os tubérculos foram tratados com 30 ppm de ácido giberélico e etanol, para quebra de dormência (BENEDETTI et al., 2005) e armazenados a 25°C até o plantio.

Tabela 08 – Descrição dos parentais das progênes e número de tubérculos usados na época 1 de plantio (14/03/2006).

Progênes	Parental feminino	Parental masculino	N. de tubérculos
A504	Michigan Purple	SMIJ319-1	03
A508	Asterix	SMIJ456-4Y	17
A509	SMIG274-3	A091-1	10
A511	H101-2	SMIG274-3	11
A514	E226-5	SMIG274-3	10
A517	Torridon	BI865-2	10
A518	Torridon	B0288-17	11
A519	H054-3	Tollocan	08
A520	Stirling	MSH419-1	10
A701	Monalisa	Ciclaman	10
Soma	-	-	100

Com os dados obtidos, das medidas dos caracteres dos tubérculos, antes do plantio (G1) e pós-colheita (G2), realizou-se análises das medidas descritivas de ambos os grupos (G1 e G2). Também, se procedeu à análise da correlação linear de Pearson (STEEL et al., 1997), estudo da multicolinearidade entre os caracteres de cada grupo (CRUZ & REGAZZI, 1994), avaliação dos pressupostos de normalidade pelo teste de Lilliefors (SPRENT & SMEETON, 2007), homogeneidade de variância pelo teste de Cochran e teste de Bartlett (STEEL et al., 1997) e normalidade multivariada pela análise do gráfico Q-Q *plot* (HAIR, 1998). Após, foi aplicada a análise de variância, análise de variância multivariada (MANOVA), e análise de correlação canônica entre os grupos (CRUZ & REGAZZI, 1994).

Para o cálculo da correlação canônica, denotou-se as medidas dos tubérculos plantados (variáveis do G1), como X (x_1, x_2, x_3, x_4); e as medidas dos tubérculos colhidos (variáveis do G2) como Y (y_2, y_3, y_4, y_5), conforme a metodologia descrita em CRUZ & REGAZZI (1994). Pelo método, as variáveis X e Y são reescritas, como uma combinação linear: $U = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$ e $V = b_2y_2 + b_3y_3 + b_4y_4 + b_5y_5$. A correlação canônica consiste em determinar a combinação linear entre U e V de modo que a correlação $\text{Corr}(U,V)$ seja maximizada. Denotando-se R_{xx} , R_{yy} e R_{xy} as matrizes das correlações amostrais, a primeira

correlação canônica (r_1), correspondente ao primeiro par canônico, é a raiz quadrada o primeiro autovalor ($r_1 = \sqrt{\lambda_1}$), solução da equação $|R_{xy}R_{yy}^{-1}R'_{xy} - \lambda R_{xx}| = 0$. De modo análogo, encontram-se as demais correlações canônicas. Os coeficientes de ponderação dos pares canônicos são os autovetores associados aos respectivos autovalores. Para testar a hipótese de nulidade da ausência de correlação entre as variáveis, emprega-se uma estatística de qui-quadrado (CRUZ & REGAZZI, 1994).

Em 25 de agosto de 2006 realizou-se o plantio de 161 tubérculos, selecionados da colheita da primeira época, mantendo-se a representatividade e variabilidade das dez progênes iniciais. Em 06 de dezembro de 2006 foram colhidos 628 tubérculos em 85 covas. Novamente, avaliou-se os caracteres do tubérculo-semente (plantado em cada cova) e os mesmos caracteres dos tubérculos colhidos, constituindo os dados de G1 e G2 da segunda época. Os procedimentos foram os mesmos realizados no experimento da época um, bem como todas as análises estatísticas.

Na terceira época, o plantio foi 15 de março de 2007, com 59 tubérculos selecionados dentre os colhidos na segunda época. Para tanto, selecionou-se seis tubérculos representativos de cada uma das progênes iniciais do trabalho (primeira época). O critério de seleção foi o mesmo realizado nas etapas anteriores, um tubérculo pequeno, um grande e um intermediário, de cada cova com produção, quando isto foi possível. A exceção ocorreu com o genótipo A509, do qual só havia a disponibilidade de 5 tubérculos. Em 04 de junho de 2007 procedeu-se a colheita, obtendo-se 191 tubérculos nas 53 covas com produção. Novamente, os caracteres do tubérculo mãe (plantado em cada cova) e os mesmos caracteres dos tubérculos colhidos foram avaliados, constituindo os dados de G1 e G2 da terceira época.

Os dados da terceira época também foram submetidos à análise de agrupamento. Para a realização desta análise calculou-se a matriz das distâncias de Mahalanobis, entre todos os pares de progênes. O método de agrupamento utilizado foi o do vizinho mais próximo. Neste método identificam-se as progênes mais similares, na matriz das distâncias, formando agrupamentos de progênes mais próximas, que podem ser visualizadas em um diagrama chamado dendograma (CRUZ & REGAZZI, 1994).

As análises estatísticas foram realizadas com os aplicativos computacionais Genes (CRUZ, 2001) e SAEG (2007).

4 RESULTADOS E DISCUSSÃO

Com os resultados obtidos nas três épocas de plantio, em que foi realizado o experimento, utilizaram-se métodos estatísticos multivariados para identificar qual, ou quais, caracteres medidos nos tubérculos, antes do plantio, mais se relacionam com caracteres dos tubérculos produzidos. Essa relação poderá servir como mais uma alternativa para a seleção precoce de clones de batata.

4.1 Resultados da primeira época de plantio

Dos 100 tubérculos plantados, apenas 71 produziram, devido, possivelmente, a dormência dos tubérculos-semente. Realizou-se a análise dos dados correspondentes aos tubérculos que tiveram alguma produção, pois, para a análise de correlação canônica, necessita-se relacionar os dados de cada tubérculo plantado, com os de sua respectiva produção.

As estatísticas dos quatro caracteres avaliados nos tubérculos plantados (comprimento, maior diâmetro, menor diâmetro e massa), em cujas covas obtiveram-se produção de tubérculos, estão descritos na Tabela 09 e representam os caracteres do grupo 1 (G1), para a análise de correlação canônica. Destaca-se a medida do CV, para massa de tubérculo (82,17%), por ser o maior, em relação ao CV dos outros caracteres dos tubérculos plantados. A amplitude entre as medidas é grande, pois se deseja trabalhar com tubérculos, de dimensões distintas.

A análise de variância (ANOVA) indicou que existe diferença significativa, pelo teste F, entre as médias das 10 progênies estudadas, para os caracteres, comprimento, maior diâmetro e menor diâmetro, a um nível de significância de 5%. O caracter massa, pelo mesmo teste, não foi significativo, isto é, não há diferença significativa entre as médias das 10 progênies, para massa de tubérculo plantado (Tabela 09).

A análise de variância multivariada (MANOVA) entre os caracteres do G1 foi significativa, pelo critério Lambda de Wilks, com um nível de 5% de significância. Por este

critério conclui-se que há diferença significativa entre os vetores de médias, das 10 progênes em estudo.

Tabela 09 - Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 1 (n = número de tubérculos, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G1	n	Média	Mínimo	Máximo	DP	CV (%)	T L
x1= comprimento (mm)	71	36,07*	15,40	61,16	11,60	32,16	H ₀
x2= maior diâmetro (mm)	71	27,41*	13,31	52,39	7,95	29,00	H ₀
x3= menor diâmetro (mm)	71	23,72*	12,38	43,26	6,35	26,77	H ₀
x4= massa (g)	71	17,01*	1,50	75,20	13,98	82,19	H ₁

* Diferenças entre médias de genótipos significativas pelo teste F (5%); ns = diferenças não significativas;

H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

As correlações lineares entre as variáveis do G1 (Tabela 10) são todas altas e significativas, a mais forte ocorre entre maior diâmetro e menor diâmetro (0,9711), seguida da correlação entre maior diâmetro e massa (0,9382). Na análise de multicolinearidade, utilizando o critério número de condição (NC) (CRUZ & CARNEIRO, 2003), encontrou-se NC = 134, assim, decidiu-se manter todas as variáveis de G1 para a realização da análise de correlação canônica.

Tabela 10 - Correlação linear entre as variáveis dos tubérculos plantados na época 1

Variáveis	x1	x2	x3
x1= comprimento	1	-	-
x2= maior diâmetro	0,7902*	1	-
x3= menor diâmetro	0,7569*	0,9711*	1
x4= massa	0,8484*	0,9382*	0,9264*

* Significativo pelo teste t em nível de 5% de probabilidade

As estatísticas descritivas, das variáveis avaliadas nos tubérculos produzidos, nas 71 covas da época 1, estão na Tabela 11, e representam os caracteres do grupo 2 (G2). Observa-se que o número máximo de tubérculos produzidos por cova foi 18. Os demais caracteres tiveram suas dimensões aumentadas, em relação aos valores correspondentes dos tubérculos plantados (Tabela 9), o que é esperado. Apenas os valores do CV mantiveram-se muito próximos aos valores obtidos para os tubérculos plantados (Tabela 09). O maior CV (80,34%)

continua sendo o do caracter massa, seguido do CV correspondente a número de tubérculo por cova (71,50%).

A ANOVA indicou que não existe diferença significativa entre as médias das 10 progênes, para os caracteres, número de tubérculos por cova e comprimento. Para os demais, maior diâmetro, menor diâmetro e massa, existem diferenças significativas entre as médias, a um nível de significância de 5%, entre os tubérculos produzidos (Tabela 11).

Tabela 11 - Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 1 (n = número de covas, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G2	n	Média	Mínimo	Máximo	DP	CV (%)	T L
y1= n tubérculos/cova	71	3,67 ^{ns}	1,00	18,00	2,63	71,66	H ₁
y2= comprimento (mm)	71	48,73 ^{ns}	23,81	102,97	15,12	31,03	H ₀
y3= maior diâmetro (mm)	71	38,57*	19,56	76,80	11,85	30,72	H ₁
y4= menor diâmetro (mm)	71	32,63*	3,81	50,10	9,02	27,64	H ₀
y5= massa (g)	71	51,22*	5,40	193,00	41,15	80,34	H ₁

* Diferenças entre médias de genótipos significativas pelo teste F (5%); ns = diferenças não significativas;

H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

A MANOVA, entre os caracteres do G2, também foi significativa, pelo critério Lambda de Wilks, com um nível de 5% de significância. Por este critério conclui-se que há diferença significativa entre os vetores de médias, das 10 progênes em estudo.

Um dos pressupostos, para a utilização das correlações, é que a associação entre os caracteres de interesse seja linear (HAIR, 1998). Os diagramas de dispersão entre todas as variáveis de G1 e de G2 mostraram que, a relação entre o caracter, número de tubérculos por cova, não é linear com nenhum dos outros caracteres de G2, assim, ele foi retirado do cálculo da correlação linear e, conseqüentemente, da análise da correlação canônica.

Tabela 12 - Correlação linear entre as variáveis dos tubérculos colhidos na época 1

Variáveis do G2	y2	y3	y4
y2= comprimento	1		
y3= maior diâmetro	0,9014*	1	
y4= menor diâmetro	0,8190*	0,9120*	1
y5= massa	0,9074*	0,9555*	0,8748*

* Significativo pelo teste t em nível de 5% de probabilidade

Da mesma forma que no G1, as correlações lineares entre os caracteres do G2 (Tabela 12) foram todas significativas e altas, sendo mais forte a correlação linear entre os caracteres maior diâmetro e massa (0,9555), seguido da correlação entre maior diâmetro e menor diâmetro (0,9120). Nos testes de multicolinearidade encontrou-se $NC = 94$, isto é, colinearidade fraca entre as variáveis (CRUZ & CARNEIRO, 2003). Assim, mantêm-se todos os caracteres (y_2 , y_3 , y_4 e y_5) do G2, para a análise de correlação canônica.

A estimativa de produtividade para a colheita da época 1 foi obtida pelo produto: número médio de tubérculo por cova (3,67), massa média por tubérculo (51,22g), número de covas por hectare (37.500) e a proporção de covas com produção (71/100), ou seja, 5005 kg/ha. Esta produtividade é baixa, possivelmente, devido a dormência dos tubérculos-semente (29% não produziram) e por ainda não terem sido selecionados para produtividade.

Apenas 3, das 16 correlações lineares entre G1 e G2 são significativas (Tabela 13). Se comparadas com as correlações das Tabelas 10 e 12, observa-se que são bem mais fracas. Apenas o caracter menor diâmetro, dos tubérculos produzidos, possui correlação significativa com os caracteres maior diâmetro, menor diâmetro e massa, relativos aos tubérculos plantados.

Tabela 13 - Correlação linear entre as variáveis dos grupos G1 e G2 na época 1

G2↓	G1→	x1=comprimento	x2=maior diâmetro	x3=menor diâmetro	x4=massa
y2= comprimento		0,1620	0,0437	0,0534	0,1084
y3= maior diâmetro		0,0755	0,1449	0,1423	0,1602
y4= menor diâmetro		0,1970	0,2501*	0,2469*	0,2435*
y5= massa		0,1005	0,1331	0,1306	0,1486

* Significativo pelo teste t em nível de 5% de probabilidade

A análise de correlação canônica realizada entre as variáveis dos G1 e do G2 indicou, que apenas o primeiro par canônico foi significativo, a um nível de 1% de probabilidade, com uma correlação de $r_1 = 0,6782$. Isto indica que existe uma associação entre os caracteres dos tubérculos plantados e a sua produção. Estes resultados (Tabela 14) sugerem que tubérculos menos longos (-0,5060) e de caracter maior diâmetro, maiores (0,6577) produziram tubérculos, em média, menos longo (-0,7029) e com caracter maior diâmetro, maior (0,6971).

Tabela 14 - Coeficientes canônicos do primeiro par canônico na época 1 (outono).

Coeficientes canônicos das variáveis do G 1	
x1 = comprimento	-0,5060
x2 = maior diâmetro	0,6577
x3 = menor diâmetro	-0,3087
x4 = massa	0,1312
Coeficientes canônicos das variáveis do G 2	
y2 = comprimento	-0,7029
y3 = maior diâmetro	0,6971
y4 = menor diâmetro	-0,1087
y5 = massa	0,0901
r1 = correlação canônica	0,6782

4.2 Resultados da segunda época de plantio

Dos 261 tubérculos produzidos na época 1, selecionou-se um tubérculo de tamanho pequeno, um de tamanho grande e outro de tamanho intermediário, de cada cova, com o objetivo de manter a maior diferença possível entre os clones e a representatividade das 10 progênies iniciais. Os tubérculos foram tratados com 30 ppm de ácido giberélico e etanol, para quebra de dormência (BENEDETTI et al., 2005) e armazenados a 25°C até o plantio. Em 25 de agosto de 2006 os 161 tubérculos, em condições, foram plantados (Tabela 15).

Tabela 15 - Número de tubérculos plantados e colhidos, de cada progênie, nas épocas 1 e 2

Progênies	N. de tubérculos da época 1		N. de tubérculos da época 2	
	Plantados em 14/03/2006	Colhidos em 15/06/2006	Plantados em 25/08/2006	Colhidos em 06/12/2006
A504	3	5	4	12
A508	17	70	39	169
A509	10	17	13	11
A511	11	3	2	20
A514	10	26	22	108
A517	10	25	15	105
A518	11	47	21	61
A519	08	13	9	45
A520	10	30	19	48
A 701	10	25	17	49
Total	100	261	161	628

As estatísticas dos caracteres avaliados nos tubérculos plantados (comprimento, maior diâmetro, menor diâmetro e massa), em cujas covas obtiveram-se produção de tubérculos (Tabela 16), representam os caracteres do G1. Apenas 85 das 161 covas plantadas resultaram em produção (53%). Da mesma forma que na época 1, o CV de massa de tubérculo (85,43%) é muito maior do que o CV dos outros caracteres, que se mantiveram em torno dos 33%.

Tabela 16 - Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 2 (n = número de tubérculos, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G1	n	Média	Mínimo	Máximo	DP	CV (%)	T L
x1= comprimento (mm)	85	56,07 ^{ns}	19,55	109,17	20,18	35,99	H ₀
x2= maior diâmetro (mm)	85	45,17 ^{ns}	17,89	81,37	14,88	32,94	H ₀
x3= menor diâmetro (mm)	85	37,89 ^{ns}	14,17	60,28	11,37	30,01	H ₀
x4= massa (g)	85	72,73 ^{ns}	3,60	282,80	62,14	85,44	H ₁

ns = diferenças não significativas; H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

A ANOVA não foi significativa, para nenhum dos caracteres, isto é, para todos os caracteres, não houve diferença significativa entre as médias das 10 progênies analisadas. Entretanto a MANOVA, utilizando-se o critério Lambda de Wilks, com um nível de significância de 5%, indica que existe diferença significativa entre os vetores de médias destas progênies.

Todas as correlações lineares entre os caracteres de G1 foram fortes e significativas (Tabela 17). A correlação mais alta foi entre maior diâmetro e menor diâmetro (0,9622), seguida da correlação entre maior diâmetro e massa (0,9269). No estudo da multicolinearidade entre os caracteres encontrou-se NC = 112 (CRUZ & CARNEIRO, 2003), possibilitando o uso de todos eles na análise de correlação canônica.

Tabela 17 - Correlação linear entre as variáveis dos tubérculos plantados na época 2

Variáveis do G1	x1	x2	x3
x1= comprimento	1		
x2= maior diâmetro	0,8537*	1	
x3= menor diâmetro	0,8383*	0,9622*	1
x4= massa	0,8743*	0,9269*	0,8945*

* Significativo pelo teste t em nível de 5% de probabilidade de erro.

As estatísticas descritivas dos caracteres avaliados, nos 85 tubérculos que produziram, estão na Tabela 18. O número máximo de tubérculos produzido por cova foi 27, sendo este, o caracter com o maior CV (67,62%). As médias de todos os caracteres (Tabela 18) diminuíram sensivelmente, em relação aos mesmos caracteres do G1 (Tabela 16), sendo que, massa de tubérculos plantados, passou de 72,73g (Tabela 16) para 25,05g (Tabela 18), nos tubérculos colhidos. Como consequência, os valores dos desvios-padrão e CV, aumentaram. Estas alterações podem ter sido consequência do excesso de chuvas no período.

A ANOVA foi significativa, a um nível de 5%, apenas para comprimento de tubérculo. A análise não encontrou diferença significativa, entre as médias das 10 progênies, para os demais caracteres (Tabela 17). A MANOVA, através do critério Lambda de Wilks, indica que há diferença significativa entre os vetores de médias das progênies.

Tabela 18 - Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 2 (n = número de covas, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G2	n	Média	Mínimo	Máximo	DP	CV (%)	T L
y1= n tubérculos/cova	85	7,32 ^{ns}	2,00	27,00	4,95	67,62	H ₀
y2= comprimento (mm)	85	42,39*	18,44	71,04	10,82	25,53	H ₀
y3= maior diâmetro (mm)	85	30,34 ^{ns}	15,34	51,85	6,45	21,28	H ₀
y4= menor diâmetro (mm)	85	26,18 ^{ns}	14,64	43,84	5,38	20,57	H ₁
y5= massa (g)	85	25,05 ^{ns}	1,30	81,27	15,39	61,46	H ₁

* Diferenças entre médias de genótipos significativas pelo teste F (5%); ns = diferenças não significativas;

H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

A estimativa de produtividade, para a colheita da época 2, foi obtida pelo produto: número médio de tubérculo por cova (7,32), massa média por tubérculo (25,05g), número de covas por hectare (37.500) e a proporção de covas com produção (85/161), ou seja, 3630 kg/ha, equivalente a 72% da produtividade obtida na época anterior. Neste período ocorreu uma precipitação de 364mm (ANEXO), nos 60 dias após o plantio, período este, em que, de acordo com PEREIRA & DANIELS (2003), qualquer estresse afeta o rendimento final. Aliado a isto, o rendimento também é afetado pela radiação solar que, conforme BISOGNIN et al. (2008a), é a variável ambiental determinante das diferenças de rendimento, nas condições de cultivo de primavera e outono, nesta região. Também contribuiu para a baixa produtividade, o fato de que, um terço dos tubérculos selecionados para o plantio, eram de tamanho relativamente menor. Verificou-se, também, neste segundo experimento que,

somente 53% dos tubérculos plantados, produziram, enquanto que, na época 1, 71% dos tubérculos produziram. Este fato pode ser explicado pelos diferentes níveis de dormência dos tubérculos, uma vez que, a interação entre clones, quebra de dormência e safra é significativa (BENEDETTI et al., 2005) e que, tubérculos produzidos durante a primavera têm menor dormência do que os produzidos durante o outono (BISOGNIN et al., 2008b). A chuva do período também pode ser responsável por esta redução.

Número de tubérculo por cova, não fez parte das análises, por não ser linear (HAIR, 1998) com nenhum dos outros caracteres. A correlação linear entre os caracteres do G2 (Tabela 19) foi alta e significativa para todos. A maior correlação ocorreu entre maior diâmetro e menor diâmetro (0,9614). Na análise de multicolinearidade encontrou-se $NC = 113$ (CRUZ & CARNEIRO, 2003), mantendo-se todas as variáveis (y2, y3, y4, y5), para a análise de correlação canônica.

Tabela 19 - Correlação linear entre as variáveis dos tubérculos colhidos na época 2

Variáveis do G2	y2	y3	y4
y2= comprimento	1		
y3= maior diâmetro	0,8482*	1	
y4= menor diâmetro	0,8241*	0,9614*	1
y5= massa	0,8934*	0,9263*	0,8902*

* Significativo pelo teste t em nível de 5% de probabilidade

As correlações entre as variáveis correspondentes aos tubérculos plantados (G1) e sua produção (G2) (Tabela 20), com duas únicas exceções, são todas significativas. A variável comprimento, do G1, é a que apresenta as mais altas correlações com as variáveis do G2, sendo: 0,3823 com menor diâmetro; 0,3608 com maior diâmetro; 0,3592 com comprimento e 0,3544 com massa. Comparando-se estes resultados (Tabela 20) com os da época 1 (Tabela 13), observa-se que aqui, 14 de 16 correlações são significativas, enquanto que, na época 1, apenas 3 das 16 correlações foram significativas.

Tabela 20 - Correlação linear entre as variáveis dos grupos G1 e G2 na época 2

G2↓	G1→	x1=comprimento	x2=maior diâmetro	x3=menor diâmetro	x4=massa
y2= comprimento		0,3592*	0,2362*	0,2291*	0,1941
y3= maior diâmetro		0,3608*	0,2984*	0,3132*	0,2367*
y4= menor diâmetro		0,3823*	0,2910*	0,3092*	0,2285*
y5= massa		0,3544*	0,2540*	0,2641*	0,2050

* Significativo pelo teste t em nível de 5% de probabilidade H_0 = segue distribuição normal; H_1 = não segue distribuição normal.

A análise de correlação canônica (Tabela 21) realizada entre os caracteres dos grupos G1 e G2, para a época 2 de plantio, indicou que, apenas o primeiro par canônico foi significativo, a um nível de 1% de probabilidade de erro, com correlação de $r_1 = 0,4968$. Esta correlação mostra que os caracteres dos dois grupos não são independentes e é possível encontrar uma associação que influencia as relações entre os mesmos. Pelos coeficientes canônicos correspondentes ao primeiro par canônico é possível observar que tubérculos mais compridos (0,7698) e com menor massa (-0,6264), são determinantes da produção de tubérculos com o caracter menor diâmetro maior (0,6962) e caracter maior diâmetro menor (-0,6649), nesta época de plantio.

Tabela 21 - Coeficientes canônicos do primeiro par canônico na época 2 (primavera).

Coeficientes canônicos das variáveis do G 1	
x1= comprimento	0,7698
x2 = maior diâmetro	0,1017
x3 = menor diâmetro	0,0103
x4 = massa	-0,6264
Coeficientes canônicos das variáveis do G 2	
y2= comprimento	0,2351
y3= maior diâmetro	-0,6649
y4 = menor diâmetro	0,6962
y5 = massa	0,1337
r1 = correlação canônica	0,4968

4.3 Resultados da terceira época de plantio

Dos 628 tubérculos produzidos na época 2, foram selecionados e avaliados 59 tubérculos, plantados em 15 de março de 2007. Esses tubérculos foram plantados novamente em covas individuais. Cada conjunto de três covas consecutivas, contendo aleatoriamente um tubérculo de tamanho pequeno, outro de tamanho médio e outro de tamanho grande, representantes de cada uma das 10 progênes do início do experimento, constituiu uma unidade experimental. O experimento foi conduzido no delineamento inteiramente casualizado com duas repetições. O critério de seleção foi a maior variabilidade possível entre os clones e a representatividade das 10 progênes que iniciaram o experimento. A Tabela 22 mostra o total de tubérculos plantados e colhidos em cada etapa do experimento, de acordo com sua progênie. Na última etapa, a progênie A509 não produziu nenhum tubérculo.

Tabela 22 - Número de tubérculos plantados e colhidos, de cada progênie, nas épocas 1, 2 e 3.

Progênes	N. de tubérculos da época 1		N. de tubérculos da época 2		N. de tubérculos da época 2	
	Plantados em	Colhidos em	Plantados em	Colhidos em	Plantados em	Colhidos em
	14/03/2006	15/06//2006	25/08/2006	06/12//2006	15/03/2007	04/06/2007
A504	3	5	4	12	6	24
A508	17	70	39	169	6	32
A509	10	17	13	11	5	0
A511	11	3	2	20	6	15
A514	10	26	22	108	6	21
A517	10	25	15	105	6	18
A518	11	47	21	61	6	25
A519	08	13	9	45	6	13
A520	10	30	19	48	6	22
A 701	10	25	17	49	6	21
Total	100	261	161	628	59	191

As medidas descritivas dos caracteres avaliados nos tubérculos plantados nesta época, a exemplo das anteriores, mostram que os valores dos CVs não são baixos, sendo o maior para a massa de tubérculo (81,31%). Isto ocorre, porque a seleção dos genótipos é realizada, buscando a maior variabilidade possível, entre os materiais disponíveis (Tabela 23).

Tabela 23 - Estatísticas descritivas das variáveis correspondentes aos tubérculos plantados na época 3 (n = número de tubérculos, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G1	n	Média	Mínimo	Máximo	DP	CV (%)	T L
x1= comprimento (mm)	53	48,03 ^{ns}	20,10	107,83	18,19	37,87	H ₀
x2= maior diâmetro (mm)	53	34,56 ^{ns}	15,90	54,56	8,71	25,20	H ₀
x3= menor diâmetro (mm)	53	30,02 ^{ns}	13,75	46,63	7,25	24,15	H ₀
x4= massa (g)	53	34,59 ^{ns}	2,00	135,80	28,13	81,32	H ₁

ns = diferenças não significativas; H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

A ANOVA mostrou que não há diferença significativa entre as médias dos caracteres das progênies plantadas (Tabela 23). Da mesma forma, a MANOVA, pelo critério Lambda de Wilks, também não foi significativa.

Todas as correlações entre as variáveis do G1 (Tabela 24) foram significativas e altas. A maior correlação ocorreu entre o maior e menor diâmetro (0,9572), seguida da correlação entre comprimento de tubérculo e massa (0,9366). A análise considerou fraca (NC = 92), a multicolinearidade colinearidade entre este grupo de variáveis, segundo critérios descritos em CRUZ & CARNEIRO (2003), assim, todos os caracteres foram usados para a análise de correlação canônica.

Tabela 24 - Correlação linear entre as variáveis dos tubérculos plantados na época 3

Variáveis do G1	x1	x2	x3
x1= comprimento	1		
x2= maior diâmetro	0,8146*	1	
x3= menor diâmetro	0,7659*	0,9572*	1
x4= massa	0,9366*	0,8846*	0,8519*

* Significativo pelo teste t em nível de 5% de probabilidade de erro.

As análises descritivas, das variáveis que compõe o grupo G2 (Tabela 25) mostram que o CV do número de tubérculo por cova é o mais alto (64,63%), relativo aos outros caracteres da mesma tabela. Isto ocorre, devido à variabilidade existente na seleção dos clones de cada genótipo e, também, porque são 10 progênies sendo testados. Todos os CVs tiveram uma redução em relação a G1 (Tabela 23), sendo que, massa teve uma redução considerável, passando de 81,31%, nos tubérculos-semente, para 61,75%, nos tubérculos colhidos.

Tabela 25 - Estatísticas descritivas das variáveis correspondentes aos tubérculos colhidos na época 3 (n = número de covas, DP = desvio padrão, CV = coeficiente de variação) e teste Lilliefors (T L).

Variáveis do G2	n	Média	Mínimo	Máximo	DP	CV (%)	T L
y1= n tubérculos/cova	53	3,77 ^{ns}	1,00	12,00	2,45	64,99	H ₀
y2= comprimento (mm)	53	47,92*	25,47	77,87	11,48	23,96	H ₀
y3= maior diâmetro (mm)	53	37,06*	20,26	58,81	8,21	22,15	H ₀
y4= menor diâmetro (mm)	53	31,53*	19,39	50,42	6,56	20,81	H ₀
y5= massa (g)	53	39,57*	4,80	122,15	24,43	61,74	H ₁

* Diferenças entre médias de genótipos significativas pelo teste F (5%); ns = diferenças não significativas;

H₀ = segue distribuição normal; H₁ = não segue distribuição normal.

A ANOVA não foi significativa, apenas, para número de tubérculos por cova, mas, o teste F mostrou, a um nível de significância de 5%, que há diferença entre as médias das progênies para os demais caracteres (Tabela 25). A MANOVA também mostrou diferença significativa, entre os vetores de médias, pelo critério Lambda de Wilks.

A estimativa de produtividade para a colheita da época 3 foi obtida pelo produto: número médio de tubérculo por cova (3,64), massa médio por tubérculo (38,06g), número de covas por hectare (37.500) e a proporção de covas com produção (51/59), ou seja, 4491 kg/ha, valor baixo, possivelmente, pelas mesmas razões da época 1.

As medidas das correlações lineares entre os caracteres do G2, (Tabela 26) foram todas altas e significativas. A maior correlação ocorreu entre maior e menor diâmetro (0,9589), seguido de maior diâmetro e massa (0,9404). No estudo da multicolinearidade entre as variáveis encontrou-se NC = 109 (CRUZ & CARNEIRO, 2003), permitindo o uso de todas as variáveis do G2, para a análise de correlação canônica. O caracter número de tubérculos por cova não foi avaliado, por não ter uma relação linear com os demais caracteres do G2.

Tabela 26 - Correlação linear entre as variáveis dos tubérculos colhidos na época 3

Variáveis do G2	y2	y3	y4
y2= comprimento	1		
y3= maior diâmetro	0,8180*	1	
y4= menor diâmetro	0,7624*	0,9589*	1
y5= massa	0,8922*	0,9404*	0,9048*

* Significativo pelo teste t em nível de 5% de probabilidade

Das 16 correlações, entre as variáveis dos tubérculos plantados e produzidos (Tabela 27), apenas 9 são significativas a 5% de probabilidade de erro. Apenas maior diâmetro e menor diâmetro, possuem correlação significativa com todos os caracteres dos tubérculos colhidos. A massa dos tubérculos plantados, que é um caracter agrônômico importante nesta cultura, não teve correlação linear significativa com nenhum dos caracteres correspondente a tubérculos produzidos.

Tabela 27 - Correlação linear entre as variáveis dos grupos G1 e G2 na época 3

G2↓	G1→	x1=comprimento	x2=maior diâmetro	x3=menor diâmetro	x4=massa
y2= comprimento		0,3462*	0,3780*	0,3628*	0,2342
y3= maior diâmetro		0,1728	0,2920*	0,3245*	0,1332
y4= menor diâmetro		0,2332	0,3483*	0,3844*	0,2130
y5= massa		0,2332	0,3196*	0,3353*	0,1819

* Significativo pelo teste t em nível de 5% de probabilidade

A análise de correlação canônica indicou, novamente, que apenas o primeiro par canônico é significativo ($r_1=0,6420$), a um nível de 1% (Tabela 28). Portanto, há uma dependência entre os caracteres dos tubérculos plantados com sua produção. A observação dos coeficientes canônicos sugere que tubérculos plantados mais longos (0,9753) e menor massa (-1,2240) produziram tubérculos mais longos (0,7939) e menor massa (-0,5227), nesta época de plantio.

Tabela 28 - Coeficientes canônicos do primeiro par, na época 3 (outono)

Coeficientes canônicos das variáveis do G1	
x1= comprimento	0,9753
x2 = maior diâmetro	0,4784
x3 = menor diâmetro	0,1065
x4 = massa	-1,2240
Coeficientes canônicos das variáveis do G2	
y2= comprimento	0,7939
y3= maior diâmetro	-0,1425
y4 = menor diâmetro	0,2758
y5 = massa	-0,5227
r1 = correlação canônica	0,6420

Dos dados obtidos dos tubérculos produzidos, na época 3 do experimento, também se realizou a análise agrupamento, para verificar se havia alguma similaridade entre as progênes avaliadas.

As medidas de dissimilaridade foram realizadas entre as nove progênes, relativas aos caracteres, comprimento, maior diâmetro, menor diâmetro e massa, com base na distância de Mahalanobis. A menor distância observada ocorreu entre as progênes A508 e A519 (3,19) e, a maior distância, entre A511 e A520 (51,06) (Tabela 29). Realizaram-se os testes que garantem a observação dos pressupostos, para a utilização da distância de Mahalanobis (CRUZ & REGAZZI, 1994).

Tabela 29 - Medidas de dissimilaridade entre as nove progênes, em relação aos caracteres, comprimento, maior diâmetro, menor diâmetro e massa dos tubérculos colhidos na época 3, com base na distância de Mahalanobis.

Progênes	A508	A511	A514	A517	A518	A519	A520	A701
A504	6,66	18,67	13,09	8,68	9,00	4,85	19,18	25,86
A508		12,06	13,15	5,23	12,23	3,19	13,61	7,19
A511			22,71	29,54	38,01	11,77	51,06	25,13
A514				13,07	7,97	19,07	30,51	33,96
A517					6,46	12,68	6,79	13,71
A518						15,64	11,12	29,35
A519							18,81	14,15
A520								15,16

Com estas medidas de dissimilaridades entre as progênes e utilizando o método hierárquico do vizinho mais próximo, construiu-se o dendograma (Figura 03), ou diagrama da árvore, para agrupar as progênes. O grupo inicial é formado pelos genótipos mais próximos (A508 e A5119), calculam-se então as distâncias entre este grupo e os demais. Estas associações, ou agrupamentos formam o dendograma. Pela análise do dendograma não é possível observar nenhum grupo de progênes mais próximos, que se destaque. Com os caracteres utilizados, para a elaboração deste dendograma, não é possível identificar grupos similares. Este resultado dá maior confiabilidade às conclusões obtidas na correlação canônica, ou seja, a correlação existente entre os caracteres dos tubérculos plantados e dos produzidos, não é comprometida por serem clones similares.

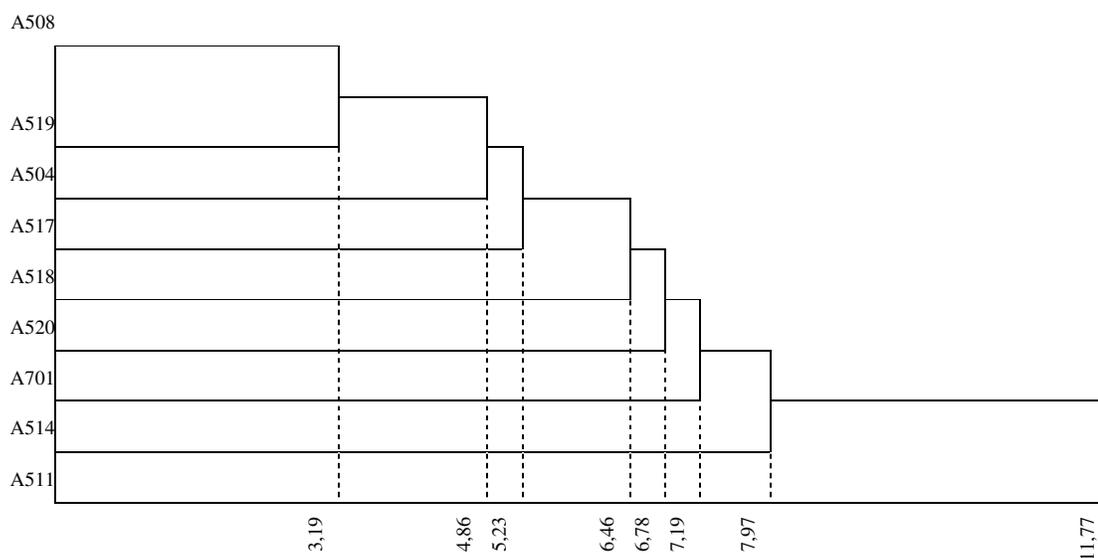


Figura 03 – Dendrograma de similaridade pelo método do vizinho mais próximo

Em alguns caracteres ocorreu a observação de valores extremos, causando a falta de aproximação com a distribuição normal, conforme indicam as Tabelas 09, 11, 16, 18, 23 e 25. Estes valores extremos são decorrência do critério de seleção dos tubérculos, uma vez que, com o objetivo de verificar se as dimensões dos tubérculos plantados, realmente, estavam relacionadas com sua produção, independente de outros fatores, optou-se por usar a maior variabilidade disponível entre os tubérculos de batata. Deste modo, visando o programa de melhoramento, optou-se pela realização das análises, mantendo todas as observações encontradas.

Nas três épocas (Tabelas 10, 17 e 24) as correlações entre os caracteres dos tubérculos plantados (x_1 , x_2 , x_3 , x_4) são todas altas, positivas e significativas com um nível de significância de 5%. O mesmo ocorre para as correlações lineares entre as variáveis dos tubérculos produzidos (y_2 , y_3 , y_4 , y_5), (Tabelas 12, 19 e 26).

Comparando as correlações significativas entre os tubérculos plantados e produzidos (Tabelas 13, 20 e 27) não é possível observar nenhum padrão, que se repita nas três épocas. Em cada uma delas, a relação entre tubérculos plantados e colhidos manifestou-se de forma muito distinta, por suas correlações lineares. Na época 1, apenas, maior diâmetro, menor diâmetro e massa tiveram correlação linear significativa e, somente, com o caracter menor diâmetro, ou seja, apenas, três correlações significativas. Já, na época 2, ocorreram catorze

correlações significativas, comprimento, maior diâmetro e menor diâmetro tiveram correlação significativa, com todos os caracteres correspondentes aos tubérculos produzidos. Porém, na época 3 a relação é completamente distinta, apenas maior diâmetro e menor diâmetro foram significativos com todos os caracteres dos tubérculos produzidos. Estas diferenças, apenas reafirmam os resultados de muitos trabalhos (ANDREU, 2005; BISOGNIN, et al., 2008A; LAMBERT, 2004; LOVATO, C., 1993; MENEZES, et al., 1999), mostrando que, fatores ambientais afetam, não só a produção, como também a aparência e a qualidade de tubérculos. Porém, só pela observação destas correlações, não é possível fazer uma inferência confiável. Quando variáveis, como as que se está estudando, estão correlacionadas entre si, deve-se considerar que, o efeito da colinearidade entre elas, mesmo fraco, pode levar a resultados pouco confiáveis, induzindo a conclusões equivocadas (CRUZ & CARNEIRO, 2003). Uma alternativa é utilizar a correlação canônica, que considera o efeito das correlações entre as variáveis envolvidas.

Quanto às correlações canônicas, das três épocas (uma seqüência de três gerações clonais), pode-se observar que, na época 2, o valor da correlação significativa foi o mais baixo ($r_1=0,4968$), enquanto que, na época 1 ($r_1= 0,6782$) e na terceira ($r_1=0,6420$) eles foram maiores e muito próximos.

Na época 1, os coeficientes canônicos explicam mais facilmente a associação entre comprimento e maior diâmetro, sugerindo a relação inversa entre eles, ou seja, tubérculos mais compridos e com o caracter maior diâmetro, menor, produziram tubérculos de formato semelhante. Os outros coeficientes, por serem bem menores, pouco irão contribuir para esta associação. Esta foi a época em que houve o menor número de correlações lineares significativas entre G1 e G2, três, em dezesseis (Tabela 13). Tanto a ANOVA como a MANOVA foram significativas (Tabelas 09 e 11), nesta época, com exceção do caracter comprimento (Tabela 11).

Na época 2 os coeficientes canônicos (Tabela 21) não mostram com clareza e coerência as relações entre tubérculos plantados e colhidos. Também, nesta época, é onde há um maior número de correlações lineares significativas entre G1 e G2 (Tabela 20). Há, ainda, divergência entre as significâncias de ANOVA e MANOVA (Tabelas 16 e 18). Outro dado de destaque é que, apenas nesta época, todas as dimensões dos tubérculos colhidos (Tabela 18) foram menores do que as dimensões dos tubérculos plantados (Tabela 16). Tudo isto reflete a atuação dos fatores ambientais sobre a cultura.

Na época 3, os coeficientes canônicos também mostram, de modo claro, a relação entre caracteres, embora não sejam os mesmos da época 1. Aqui, observa-se a relação inversa entre comprimento e massa, indicando que, tubérculos longos e de menor massa, produziram, também, tubérculos longos e de menor massa. Os outros coeficientes contribuem pouco para a interpretação, por serem bem menores. Nesta época apenas nove, das dezesseis correlações lineares entre G1 e G2 são significativa (Tabelas 27). A ANOVA e a MANOVA não foram significativas para G1 (Tabela 23), embora tenham sido significativas para G2 (Tabela 25).

Embora não tenha sido observada uma coerência entre as três épocas (seqüência de três gerações clonais), quanto aos resultados da análise de correlação canônica, o fato de haver significância para o primeiro par canônico (nas três épocas) indica que existe uma relação entre os caracteres de tubérculos plantados e os caracteres de sua produção, que poderão contribuir na prática do melhoramento (seleção precoce pelos caracteres dos tubérculos-semente). A repetição do estudo com outros materiais genéticos e/ou em ambientes mais favoráveis é sugerida, para elucidar e ampliar o assunto proposto, uma vez que, na literatura pesquisada não se encontrou, ainda, trabalhos utilizando a correlação canônica, para estudar a associação entre este tipo de caracteres em batata. Da mesma forma que um resultado obtido, com base em um só estudo (material genético e ambiente), mesmo que este seja benéfico, não deve ser considerado como sendo regra permanente, o contrário também é verdadeiro.

5 CONCLUSÕES

Nas três primeiras gerações clonais de batata estudadas, a análise de correlação canônica mostrou que existe uma associação entre os caracteres de tubérculos plantados e produzidos. A maior associação ocorre entre comprimento, sugerindo que, ao plantar-se tubérculos mais longos obtém-se, também, tubérculos mais alongados.

Embora exista uma relação significativa entre os caracteres dos tubérculos plantados e sua produção, apenas a correlação canônica, não parece ser suficiente, para ser utilizada como ferramenta para a seleção precoce de clones de batata, considerando-se o material genético utilizado e apenas nas três primeiras gerações.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABBA. Disponível em: <www.abbabatatabrasileira.com.br> Acesso em: 08/08/2008.

ALVES, S. J. et al. Melhoramento genético de plantas de propagação vegetativa. In: DESTRO D. ; MONTALVÁN, R. **Melhoramento genético de plantas**. Londrina: UEL, p. 345-367, 1999.

ANDREU, A.A. Associação entre características agronômicas da batata nos plantios de primavera e outono no Rio Grande do Sul. **Ciência e Agrotecnologia**, Lavras, v. 29, n. 5, p. 925-929, set/out 2005.

AUGUSTIN, L. **Interação genótipo x ambiente e variabilidade genética de caracteres agronômicos e qualidade de processamento em batata**. 2007. 123 p. Tese (Doutorado em Agronomia) – Universidade de Passo Fundo, Passo Fundo, RS, 2007.

AMARO, G. B. **Seleção precoce de clones de batata para caracteres do tubérculo**. 2002. 73 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG, 2002.

BENEDETTI, M. et al. Quebra de dormência de minitubérculos de batata. **Ciência Rural**. Santa Maria, v.35, n.1, p. 31-38, jan/fev 2005.

BISOGNIN, D. A. (Coord). **Recomendações técnicas para o cultivo da batata no Rio Grande do Sul e Santa Catarina**. Santa Maria: Universidade Federal de Santa Maria. 1996.

BISOGNIN, D. A.; DOUCHES, D. S. Early generation selection for potato tuber quality in progenies of lateblight resistant parents. **Euphytica**, v. 127, p. 1-9, 2002.

BISOGNIN, D. A. Melhoramento da batata para resistência às doenças. In: PEREIRA, A. da SILVA, DANIELS, J. **O cultivo da batata na região sul do Brasil**. Embrapa Clima Temperado, Brasília, DF: Embrapa Informação Tecnológica, p. 125–142, 2003.

BISOGNIN, D. A. et al. Desenvolvimento e rendimento de clones de batata na primavera e no outono. **Pesquisa Agropecuária Brasileira**. Brasília, v.43, n.6, p. 699-705, jun 2008a.

BISOGNIN, D. A. et al. Envelhecimento fisiológico de tubérculos de batata produzidos durante o outono e a primavera e armazenados em diferentes temperaturas. **Bragantia**, Campinas, v.67, n.1, p. 59-65, 2008b.

BORÉM, A. **Melhoramento de plantas**. 3. ed. Viçosa: UFV, 500 p. 2001.

BROWN, J. A comparison between single plant plots and five plant plots for the initial selection stage of a potato breeding programme. **Euphytica**, Wageningen, v. 36, n. 3, p. 711-718, 1987.

CARVALHO, F. I. F. et al. **Estimativas e implicações da herdabilidade como estratégia de seleção**. Pelotas: Ed. Universitária da UFPel, 99 p. 2001.

CARVALHO, F. I. F. et al. **Estimativas e implicações da correlação no melhoramento vegetal**. Pelotas: Ed. Universitária da UFPel, 142 p. 2004.

CIP – Centro Internacional de la papa. Disponível em: <www.cipotato.org> Acesso em: 08/09/2008.

COIMBRA, J. L. M. et al. Correlações canônicas: II - análise do rendimento de grãos de feijão e seus componentes. **Ciência Rural**, Santa Maria, v. 30, n. 1, p. 31-35, 2000.

CORDEIRO, G. M. **Cronologia de alguns conceitos e fatos importantes da estatística**.

ABE – Associação Brasileira de Estatística. Disponível em:

<<http://www.ime.usp.br/~juvencio/cronologiajaneiro02.pdf>> .Acesso em: 05/01/2009.

COSTA, L. C. **Desenvolvimento de clones de batata de alta qualidade de tubérculo a partir de genitores resistentes a requeima**. 2007. 79 p. Tese (Doutorado em Agronomia) – Universidade Federal de Santa Maria, Santa Maria, RS, 2007.

COSTA NETO, P. L. O. **Estatística**. 2. ed. São Paulo: Edgard Blücher, 280 p. 2002.

CRUZ, C. D. et al. Estudos sobre divergência genética. III. Comparação de técnicas multivariadas. **Revista Ceres**, Viçosa, v.41, p. 191-201, 1994.

CRUZ, C.D., REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 390 p. 1994.

CRUZ, C. D. **Programa Genes**: versão Windows – Aplicativo computacional em genética e estatística. Viçosa: Editora UFV, 648 p. 2001.

CRUZ, C. D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, v. 2, 585 p. 2003.

DE JONG, H.; BURNS, V. J. Inheritance of tuber shape in cultivated diploid potatoes. **American Potato Journal**, Orono, v. 70, n. 3, p. 267-283, Mar 1993.

DIAS, J. A. C. S. **Produção de mini-tubérculos de batata-semente**. 2006. Disponível em: <www.infobibos.com/artigos/2006_3/minituberculos/index.htm> Acesso em: 08/09/2008

EMBRAPA - Centro Nacional de Pesquisa de Solos (Rio de Janeiro, RJ). Sistema Brasileiro de classificação de solos. Brasília: Embrapa - SPI, 412 p. 1999.

FAO - Food and Agriculture Organization. Disponível em: <www.fao.org> Acesso em: 08/09/2008.

FIGUEIREDO, C. A bataticultura na agricultura familiar: da crise a uma nova perspectiva. In: IX Reunião técnica de pesquisa e extensão da cultura da batata da região sul, 2006. Santa Maria. **Anais ...** Santa Maria: UFSM, p. 2-26, 2006.

FRUIT-INFORM. **Ukraine remains among five leading global potato producers.** Disponível em: <www.lol.org.ua:80/eng/vegetables/showart.php?id=63425> Acesso em 25/09/2008.

GOMES, F. P. **Curso de Estatística Experimental.** 11. ed., São Paulo: Livraria Nobel, 467 p. 1985.

GONZÁLEZ, M. E. et al. **Criterios de selección en primera generación clonal cuando no sea posible cuantificar el rendimiento.** 2002. Disponível em: <www.redepapa.org/documentosred2.html>. Acesso em: 08/09/2008.

HAASE, N. U. The nutritional value of potatoes. **Potato Research.** v. 50, p. 415-417, 2007.

HAIR, J. F., et al. **Multivariate data analysis.** 5. ed. New Jersey: Prentice-Hall, 730 p. 1998.

IBGE – Instituto Brasileiro de Geografia e Estatística. Disponível em: <www.ibge.gov.br>. Acesso em: 08/09/2008.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** 3. ed. New Jersey: Prentice-Hall, 503 p. 1992.

LAMBERT, E. S. **Estratégias para o melhoramento da batata para condições tropicais.** 2004. 142 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG, 2004.

LEDO, C. A. S. et al. Análise de variância multivariada para os cruzamentos dialélicos. **Ciência e Agrotecnologia,** Lavras. v. 27, n. 6, p. 1214-1221, nov/dez 2003.

LOVATO, C. Influência do ambiente no desenvolvimento da batata. **Ciência Rural,** Santa Maria, v. 23, n. 1, p. 101-106, jan/abr 1993.

MENEZES, C. B. et al. Avaliação de genótipos de batata (*solanum tuberosum* l.) nas safras “das águas” e de inverno no sul de Minas Gerais. **Ciência e Agrotecnologia,** Lavras, v.23, n.4, p. 776-783, out/dez 1999.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: Ed. UFMG, 295 p. 2005.

MOITA NETO, J. M. Estatística multivariada. Uma visão didática-metodológica. **Crítica na Rede**. Filosofia da ciência. 09 maio 2004. Disponível em: <www.criticanarede.com/cien_estatistica>. Acesso em: 05/01/2009.

MORENO, J.A. **Clima do Rio Grande do Sul**. Porto Alegre: Secretaria da Agricultura, 41p. 1961.

MORRISON, D. F. **Multivariate statistical methods**. 2. ed. Tokyo: McGraw Hill, 415 p. 1978.

NEELE, A. E. F., et al. Optimising visual selection in early clonal generations of potato based on genetic and economic considerations. **Theoretical and Applied Genetics**. v.78, n. 5, p. 665-671, Nov 1989.

OLIVEIRA, A. C. B. de et al. Avaliação da divergência genética em batata-doce por procedimentos multivariados. **Acta Scientiarum**, Maringá, v. 22, n. 4, p. 895-900, 2000.

PAPAS ANDINAS. Disponível em: <www.papasandinas.org> Acesso em: 08/09/2008.

PEREIRA, A. S. Melhoramento genético da Batata. In: Simpósio de atualização e melhoramento de plantas, 4., 2000, Lavras. **Anais...** Lavras: UFLA, 2000. p. 41-50, 2000.

PEREIRA, A. S. Desenvolvimento de cultivares nacionais de batata. **Batata Show**. n.7, Ano 3, jul 2003.

PEREIRA, A. S.; DANIELS, J. **O cultivo da batata na região sul do Brasil**. Brasília, DF: Embrapa Informações Tecnológicas, 2003.

PINTO, C. A. B. P.; VALVERDE, V. I. R.; ROSSI, M. S. Eficiência da seleção nas primeiras gerações clonais em batata (*Solanum tuberosum* L.). **Pesquisa Agropecuária Brasileira**, Brasília, v. 29, n. 5, p. 771-778, maio 1994.

PINTO, C. A. B. P. Desenvolvimento de novas cultivares de batata. **Batata Show**. n. 2, Jul 2001.

PINTO, C.A.B.P. Seleção de famílias e seleção recorrente em batata. In: Congresso Brasileiro de Melhoramento de Plantas, 2., 2003, Porto Seguro. **Anais...** Porto Seguro: Embrapa Mandioca e Fruticultura, 2003.

POTATO 2008. Disponível em: <www.potato2008.org>. Acesso em: 08/09/2008.

SAEG - Sistema para Análises Estatísticas, Versão 9.1: Fundação Arthur Bernardes - UFV - Viçosa, 2007.

SHYMOYAMA, N. A cadeia produtiva da batata no Brasil. IX Reunião técnica de pesquisa e extensão da cultura da batata da região sul, 2006, Santa Maria. **Anais...** Santa Maria: UFSM, p. 80-87, 2006.

SILVA E. C. et al. Produção de minitubérculos a partir de brotos de cultivares de batata em diferentes combinações de substratos. **Horticultura Brasileira**, Campinas, v. 24, n. 2, p. 241–244, abr/jun 2006.

SILVA, G. O. et al. Importância de caracteres na dissimilaridade de progênies de batata em gerações iniciais de seleção. **Bragantia**, Campinas, v.67, n.1, p. 141-144, 2008.

SIMON, G. A. **Interação famílias por ambientes e seleção de clones de batata resistentes à pinta preta e tolerantes ao calor**. 2005. 314 p. Tese (Tese em Agronomia) – Universidade Federal de Lavras – MG, 2005.

SPRENT, P.; SMEETON, N.C. **Applied nonparametric statistical methods**. 4th ed. Boca Raton: Chapman & Hall, 530 p. 2007.

STEEL, R.G.D.; TORRIE, J.H.; DICKEY, D.A. **Principles and procedures of statistics: a biometrical approach**. 3.ed. New York: McGraw Hill Book, 666 p.1997.

TAI, G. C. C. Effectiveness of visual selection for early clonal generation seedlings of potato. **Crop Science**, Madison, v. 15, n. 1, p. 15-18, jan/feb 1975.

TAI, G.C.C.; YOUNG, D. A. Early generation selection for important agronomic characteristics in a potato breeding program. **American Potato Journal**. v. 61, p. 419-434, 1984.

TAVARES, M. et al. Efeitos diretos e indiretos e correlações canônicas para caracteres relacionados com a produção de pimentão. **Bragantia**, Campinas, v. 58, n.1, p. 41-47, 1999.

TRUGILHO P. F., et al. Correlação canônica das características químicas e físicas da madeira de clones de *Eucalyptus grandis* e *Eucalyptus saligna*. **Cerne**, Lavras, v.9, n.1, p. 84-94, 2003.

UFSM/BATATA. Disponível em <www.ufsm.br/batata> Acesso em 25/09/2007.

VIANA, A. P., et al. Polinização seletiva em maracujazeiro amarelo (*Passiflora edulis* f. *flavicarpa*) monitorada por vetores canônicos. **Ciência Rural**, Santa Maria, v.37, n.6, p. 1627-1633, nov/dez 2007.

VIDIGAL, M. C. et al. Divergência genética entre cultivares de mandioca por meio de estatística multivariada. **Bragantia**, Campinas, v. 56 n. 2, p. 263-271, 1997.

VIEIRA, F. T. P. A. et al. Uma abordagem multivariada em experimento silvipastoril com *Leucaena leucocephala* (Lam.) de Wit no agreste de Pernambuco. **Ciência Florestal**, Santa Maria, v. 17, n. 4, p. 333-342, out/dez 2007.

ANEXO

Época 1		Época 2		Época 3	
Plantados em	Colhidos em	Plantados em	Colhidos em	Plantados em	Colhidos em
14/03/2006	15/06//2006	25/08/2006	06/12//2006	15/03/2007	04/06/2007
100 tubérculos	71 covas	161 tubérculos	85 covas	59 tubérculos	53 covas
5005 kg/ha (estimativa)		3630 kg/ha (estimativa)		5025 kg/ha (estimativa)	

Precipitação (mm) na época 1 (14/03/2006 a 15/06/2006)

Dias	março/2006	abril/2006	maio/2006	junho/2006
01 a 10	2,9	14,3	0,0	9,0
11 a 20	29,0	19,8	55,8	7,8
21 a 31	48,3	34,5	25,9	61,0
Total	80,2	68,6	81,7	77,8

Precipitação (mm) na época 2 (25/08/2006 a 06/12/2006)

Dias	agosto/2006	setembro/2006	outubro/2006	novembro/2006	dezembro/2006
01 a 10	39,6	58,8	50,5	47,3	1,1
11 a 20	27,7	84,2	148,1	59,3	20,0
21 a 31	8,3	11,2	10,3	28,0	63,1
Total	75,6	154,2	208,9	134,6	84,2

Precipitação (mm) na época 3 (15/03/2007 a 04/06/2007)

Dias	março/2007	abril/2007	maio/2007	junho/2007
01 a 10	142,0	54,5	41,6	60,3
11 a 20	2,1	15,4	52,5	50,1
21 a 31	29,5	52,1	8,7	25,2
Total	173,6	122,0	102,8	135,6