



**REJANE CORRÊA DA ROCHA**

**TÉCNICAS ESTATÍSTICAS E DE ANÁLISE  
DE IMAGENS RADIOGRÁFICAS NA  
AVALIAÇÃO DA QUALIDADE DE  
SEMENTES DE MAMONA**

**LAVRAS-MG**

**2011**

**REJANE CORRÊA DA ROCHA**

**TÉCNICAS ESTATÍSTICAS E DE ANÁLISE DE IMAGENS  
RADIOGRÁFICAS NA AVALIAÇÃO DA QUALIDADE DE  
SEMENTES DE MAMONA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientadora  
Dra. Thelma Safadi

Coorientadora  
Dra. Maria Laene Moreira de Carvalho

**LAVRAS- MG  
2010**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Rocha, Rejane Corrêa da.

Técnicas Estatísticas e de análise de imagens radiográficas na  
avaliação da qualidade de sementes de mamona / Rejane Corrêa da.

Rocha. – Lavras : UFLA, 2010.

104 p. : il.

Tese (Doutorado) Universidade Federal de Lavras, 2010.

Orientador: Thelma Sáfydi.

Bibliografia.

1. *Ricinus communis*. 2. Inferência bayesiana. 3. Análise  
discriminante. 4. Teste de raios-X. I. Universidade Federal de  
Lavras. II. Título.

CDD - 519.5

**REJANE CORRÊA DA ROCHA**

**TÉCNICAS ESTATÍSTICAS E DE ANÁLISE DE IMAGENS  
RADIOGRÁFICAS NA AVALIAÇÃO DA QUALIDADE DE  
SEMENTES DE MAMONA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 26 de janeiro de 2010.

Dra. Maria Laene Moreira de Carvalho	UFLA
Dr. André Vital Saúde	UFLA
Dr. Renato Mendes Guimarães	UFLA
Dr. Renato Ribeiro de Lima	UFLA

Dra. Thelma Safadi  
Orientadora

**LAVRAS - MG  
2010**

Aos meus pais, Antonio e Célia.

Ao meu marido Clayton e aos meus filhos Vinícius e Giovana.

Dedico.

## **AGRADECIMENTOS**

Agradeço a Deus por ser minha força maior neste caminho, que sempre me ajudou e que está sempre comigo.

Aos meus pais, ao meu marido Clayton, aos meus filhos, Vinícius e Giovana, aos meus irmãos, às minhas cunhadas, à minha sogra Tininha e à minha “fiel escudeira” Cida pelo apoio, confiança e carinho.

À Universidade Federal de Lavras, em especial ao Departamento de Ciências Exatas, pela oportunidade de concretização deste trabalho.

Ao professora Dra. Thelma Sáfy, pela orientação, ensinamentos, amizade, conselhos, apoio e confiança durante os últimos seis anos.

Aos professores Maria Laene Moreira de Carvalho pelos ensinamentos e André Vital Saúde pela colaboração.

A todos os professores que contribuíram para o enriquecimento dos conhecimentos indispensáveis a este trabalho.

Aos funcionários do Departamento de Ciências Exatas, em especial à Selminha, Edila e Josi pela eficiência e amizade.

Ao Professor Júlio Marcos Filho (ESALq) pela colaboração na realização do teste de raios-X. Aos colegas dos LAS-UFLA pela colaboração na montagem dos experimentos, em especial a Heloísa.

Aos meus amigos de doutorado. Em especial à Luciane, Verônica e Graziela por sempre estarem prontas a ajudar, pelo apoio durante esta caminhada e pela sincera amizade que demonstraram para comigo.

Ao CNPq, pelo financiamento do projeto.

**MUITO OBRIGADO!!!**

## RESUMO

Foram implementados métodos para serem utilizados na avaliação da qualidade das sementes de mamona, utilizando-se técnicas de análise bayesiana e de análise discriminante associadas às técnicas de análise de imagens radiográficas das mesmas. Foram utilizadas 800 sementes de mamona, da cultivar IAC-226, safra 2009, identificadas individualmente. Essas sementes foram divididas em duas amostras de 400 sementes cada, sendo a primeira submetida ao teste de raios-X sem nenhum tipo de preparo e a segunda pré tratada com a substância contraste gasosa. A amostra sem preparo, também foi submetida aos testes de germinação e vigor. Foi proposto um método para estimar parâmetros bayesianos utilizados no estudo potencial de viabilidade de sementes de mamona para as amostras sem preparo e com contraste. Três avaliadores independentes fizeram a análise visual das radiografias das sementes. Para encontrar os estimadores bayesianos foi utilizado o amostrador de Gibbs, sendo as amostras da distribuição conjunta geradas por meio de métodos computacionais de simulação. A convergência das cadeias foi verificada por meio dos testes de Geweke, Raftery e Lewis e Heidelberg e Welch. Foram simuladas amostras de tamanhos diferentes para comparar amostras de valores observados. O método bayesiano mostrou-se eficiente na estimação dos parâmetros e das médias condicionais *a posteriori* e na comparação entre as amostras de diferentes tamanhos não houve diferenças entre os resultados. Os resultados da amostra pré tratada com contraste foram mais precisos do que nas sem preparo, mas com a desvantagem desse ser destrutivo, uma vez que a submissão das sementes à exposição com gases tóxicos afeta o processo de germinação. Também foram propostos métodos de classificação do potencial de viabilidade e do potencial de germinação de lotes de sementes de mamona, utilizando-se as medidas de área (em pixels) das sementes da amostra sem preparo, calculadas por meio do processamento automático de imagens, utilizando técnicas de análise discriminante. Para os dois métodos propostos foi ajustado o modelo linear de Fisher e calculados a função e o escore discriminante. O método proposto para a classificação do potencial de viabilidade dos lotes de semente de mamona mostrou-se eficiente, com uma proporção de acertos de 88%. A proporção de classificações corretas para o método de classificação do potencial de germinação foi de 62% , sendo esse resultado relevante para a área de tecnologia de sementes, uma vez que associou características físicas e filológicas das sementes em um único método.

Palavras-chave: Análise de imagens radiográficas. Análise bayesiana. Análise discriminante. Qualidade de sementes de mamona.

## ABSTRACT

Aiming to evaluate the quality of castor seeds, we implemented methods based in Bayesian analysis and discriminant analysis techniques associated with analysis of radiographic images of those seeds. We used 800 seeds of castor bean from cultivar IAC-226, harvested in 2009, identified individually, and divided into two samples of 400 seeds each. The first one was subjected to X-ray test without any preparation and the second one pre-treated with the gaseous substance contrast. The sample without preparation was also subjected to tests of germination and vigor. We proposed a Bayesian method to estimate the parameters usually applied in analysis to measure the potential feasibility of castor seeds for the samples with and without contrast. Three independent experts proceeded to the visual analysis of seeds radiographs. To find the Bayes estimators we used the Gibbs sampler, where the samples of the conjoint distribution were generated using methods of computational simulation. The convergence of the ranges was verified by tests of Geweke, Raftery and Lewis and Heidelberg and Welch. We simulated different sizes of samples to compare samples of observed values. The Bayesian method was efficient in the estimation of the parameters and the conditional means a posteriori, and we did not find different results when comparing samples of a variety of sizes. The results of the sample pretreated with contrast were more accurate than in the unprepared one, but with the disadvantage of being destructive, since the exposition of the seeds to toxic gases affects the germination process. We also proposed methods for classification of potential viability and germination potential of seed lots of castor oil, using measures of area (in pixels) of the seed without preparing the sample. Those measures were calculated by means of automatic processing of images using discriminant analysis techniques. For the two methods proposed we adjusted and calculated the Fisher linear function and the discriminant score. The proposed method to classify the potential viability of the seed lots of castor beans was efficient, with a hit ratio of 88%. Concerning to the method to classify the germination potential of the seeds we found 62% of correct classifications. This is a relevant result to the science field of seed technology, since it allowed us to associate physical and philological characteristics of the seeds in a single method.

**Keywords:** Analysis of radiographic images. Bayesian analysis. Discriminant analysis. Quality of castor seeds.

## SUMÁRIO

	<b>CAPÍTULO 1</b>	<b>Introdução geral</b>	10
	<b>CAPÍTULO 2</b>	<b>Análise bayesiana do potencial de viabilidade de sementes de mamona por meio de análise radiográfica</b>	12
<b>1</b>	<b>INTRODUÇÃO</b>		14
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>		16
<b>2.1</b>	<b>Teste dos raios-X</b>		16
<b>2.2</b>	<b>Inferência bayesiana</b>		18
<b>2.2.1</b>	<b>Distribuições <i>a priori</i></b>		19
<b>2.2.2</b>	<b>Distribuições <i>a posteriori</i> e estimação</b>		21
<b>2.2.3</b>	<b>Métodos de Monte Carlo via cadeias de Markov (MCMC)</b>		22
<b>2.2.4</b>	<b>Amostrador de Gibbs</b>		23
<b>2.2.5</b>	<b>CrITÉrios de convergência</b>		24
<b>3</b>	<b>MATERIAL E MÉTODOS</b>		26
<b>3.1</b>	<b>Obtenção das imagens radiográficas</b>		26
<b>3.2</b>	<b>Avaliação visual das imagens radiográficas das sementes de mamona</b>		27
<b>3.3</b>	<b>Estudo do potencial de viabilidade das sementes</b>		29
<b>3.3.1</b>	<b>Método das frequências relativas</b>		29
<b>3.3.2</b>	<b>Método bayesiano via simulação Monte Carlo em cadeias de Markov (MCMC)</b>		33
<b>3.4</b>	<b>Simulação de amostras de diferentes tamanhos</b>		36
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>		38
<b>4.1</b>	<b>Método das frequências relativas</b>		38
<b>4.2</b>	<b>Método bayesiano via simulação Monte Carlo em cadeias de Markov (MCMC)</b>		39
<b>4.3</b>	<b>Simulação de amostras de diferentes tamanhos</b>		44
<b>5</b>	<b>CONCLUSÃO</b>		50
	<b>REFERÊNCIAS</b>		51
	<b>CAPÍTULO 3</b>	<b>Análise de dados obtidos em imagens radiográficas de sementes de mamona por meio de processamento automático de imagens</b>	55
<b>1</b>	<b>INTRODUÇÃO</b>		57
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>		59
<b>2.1</b>	<b>Testes de avaliação da qualidades do lote de sementes</b>		59
<b>2.1.1</b>	<b>Teste de germinação</b>		59
<b>2.1.2</b>	<b>Teste de vigor</b>		60
<b>2.2</b>	<b>Teste dos raios-X</b>		61
<b>2.3</b>	<b>Análise automática de imagens</b>		62
<b>2.4</b>	<b>Análise discriminante</b>		63

<b>3</b>	<b>MATERIAL E MÉTODOS</b> .....	67
<b>3.1</b>	<b>Análise visual de imagens</b> .....	68
<b>3.2</b>	<b>Análise automática de imagens</b> .....	70
<b>3.3</b>	<b>Análise discriminante</b> .....	71
<b>3.3.1</b>	<b>Regras de classificação</b> .....	73
<b>3.3.2</b>	<b>Estimação das probabilidades incorretas</b> .....	75
<b>3.3.3</b>	<b>Estimação da probabilidade global de acerto</b> .....	77
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	79
<b>4.1</b>	<b>Análise descritiva dos dados</b> .....	79
<b>4.2</b>	<b>Análise discriminante</b> .....	83
<b>4.2.1</b>	<b>Análise discriminante para amostras classificadas por meio da análise visual</b> .....	83
<b>4.2.2</b>	<b>Análise discriminante para amostras classificadas por meio do teste de germinação</b> .....	87
<b>5</b>	<b>CONCLUSÃO</b> .....	90
	<b>REFERÊNCIAS</b> .....	91
	<b>CONSIDERAÇÕES FINAIS</b> .....	95
	<b>ANEXOS</b> .....	96

## **CAPÍTULO 1**

### **Introdução geral**

O Brasil é um dos países no mundo com maior potencial para produção de combustível de origem vegetal (biodiesel). Por ser um produto biodegradável, derivado de fontes renováveis, o biodiesel é uma alternativa viável para a substituição do petróleo, que traz um impacto significativo na qualidade do meio ambiente.

Dentre as fontes de matéria prima utilizadas para o biodiesel a mamoneira se destaca por ser uma planta rústica e adaptada a diversas condições edafoclimáticas do Brasil. As sementes da mamona possuem um alto rendimento de óleo de excelente qualidade e, ainda, após o processo de extração, gera subprodutos que podem ser aplicados para outras finalidades.

Um dos entraves à produção de mamona em larga escala é a disponibilidade de sementes de alta qualidade. Atualmente, os métodos para avaliação da qualidade dos lotes de sementes mais utilizados são os testes de germinação e de vigor. Esses são baseados em plântulas e têm as desvantagens de consumirem um período de tempo relativamente amplo, serem destrutivos e produzirem resultados que variam entre laboratórios e entre analistas.

Um método que vem sendo utilizado na avaliação da qualidade dos lotes de semente é o teste de raios-X o qual, tem por finalidade diferenciar sementes bem formadas das sementes vazias, com danos mecânicos ou com ataque de insetos. Vem se destacando por ser rápido, preciso e simples. Além disso, tem a vantagem de ser não destrutivo, já que as sementes utilizadas, durante o teste, não são afetadas em sua qualidade pelas baixas dosagens de

raios-X. Assim, essas sementes podem, posteriormente, serem utilizadas para fins de semeadura, permitindo estabelecer relações entre os danos observados nas imagens radiográficas e o potencial físico e fisiológico.

Uma das limitações do teste de raios-X é a subjetividade na análise das imagens radiográficas, pois, assim como nos testes baseados em plântulas, sua interpretação depende do analista. Essa subjetividade poderia ser minimizada com o uso de técnicas de processamento automático de imagens, pois as imagens radiográficas das sementes seriam analisadas por meio de softwares.

Na literatura podem-se encontrar, nas mais diversas áreas do conhecimento, várias publicações que utilizam as técnicas de análise bayesiana, análise discriminante e processamento digital de imagens. A associação dessas técnicas com as técnicas, utilizadas na área de tecnologia de sementes, como os teste de raios-X, germinação e vigor, poderiam contribuir de forma significativa para o avanço dos métodos de avaliação da qualidade dos lotes de sementes de mamona, bem como, de outras culturas.

Este estudo foi realizado com o objetivo de implementar métodos para serem utilizados na avaliação da qualidade das sementes de mamona, considerando, técnicas de análise bayesiana e de análise discriminante associadas às técnicas de análise de imagens radiográficas das mesmas. Para tal, o referido estudo foi estruturado em dois capítulos. No capítulo dois é proposto um método para calcular estimadores bayesianos utilizados no estudo potencial de viabilidade de sementes de mamona, mediante a análise visual das radiografias das mesmas, feitas por três avaliadores independentes. No capítulo três foram propostos métodos de classificação do potencial de viabilidade e do potencial de germinação de lotes sementes de mamona, com base nas medidas de área das sementes calculadas por meio do processamento automático de imagens, utilizando técnicas de análise discriminante.

## CAPÍTULO 2

### **Análise bayesiana do potencial de viabilidade de sementes de mamona por meio de análise radiográfica**

#### **RESUMO**

Foi feito o estudo o potencial de viabilidade de sementes de mamona a partir da análise visual das radiografias das mesmas, feitas por três avaliadores independentes, utilizando métodos bayesianos. Foram utilizadas 800 sementes de mamona, da cultivar IAC-226, safra 2009, identificadas individualmente. Essas sementes foram divididas em duas amostras de 400 sementes cada, sendo a primeira submetida ao teste de raios-X sem nenhum tipo de preparo e a segunda pré tratada com a substância contraste gasosa. A partir dos resultados do teste de raios-X foram calculadas as probabilidades da semente ter originado uma plântula normal, da semente que originou uma plântula normal ter sido avaliada como potencialmente inviável e da semente que não germinou ou originou uma plântula não normal ter sido avaliada como potencialmente inviável, utilizando o método da frequência relativa e métodos bayesianos. Para encontrar os estimadores bayesianos foi utilizado o amostrador de Gibbs, sendo as amostras da distribuição conjunta geradas por meio de métodos computacionais de simulação. A convergência das cadeias foi verificada por meio dos testes de Geweke, Raftery e Lewis e Heidelberg e Welch. Foram simuladas amostras de tamanhos diferentes para comparar as estimativas dos parâmetros obtidos por meio de métodos bayesianos com os valores obtidos a partir das amostras observadas. O método bayesiano mostrou-se eficiente na estimação dos parâmetros e das médias condicionais a *posteriori*, sendo os resultados desses próximos dos estimados pelo método da frequência relativa. Na comparação entre as amostras de diferentes tamanhos não houve diferenças entre os resultados. Concluiu-se que os métodos bayesianos podem ser de grande interesse prático para obtenção de boas inferências na área de tecnologia de sementes, pois eles podem se incorporar à opinião de especialistas da área e também podem ser utilizados no estudo da viabilidade de sementes de outras espécies.

Palavras-chave: Análise de imagens. Teste de raios-X. Inferência bayesiana.

## ABSTRACT

In this section, we observed the potential feasibility of castor seeds from the visual analysis of its radiographs, made by three independent experts using Bayesian methods. We used 800 seeds of castor bean from cultivar IAC-226, harvested in 2009, identified individually, and divided into two samples of 400 seeds each. The first one was subjected to X-ray test without any preparation and the second sample pre-treated with the gaseous substance contrast. After that, using the relative frequency and Bayesian methods, we calculated the likelihood of the following possible outcomes seed which germinate or originated an irregular non standard seedling have been evaluated as potentially feasible. To find the Bayes estimators we used the Gibbs sampler, where the samples of the conjoint distribution were generated using methods of computational simulation. The convergence of the ranges was verified by tests of Geweke, Raftery and Lewis and Heidelberg and Welch. The simulation of different sizes of samples aimed to compare the parameters and values estimated using Bayesian methods and observed samples, respectively. The Bayesian method was efficient in the estimation of the parameters and the conditional means a posteriori, which is close to those estimated by relative frequency. Comparing the samples of different sizes we did not observe significant differences among the results. We concluded that the Bayesian methods can be applied in a large spectrum of attempts to obtain good inferences in the science field of seed technology gave that they can incorporate both quantitative and qualitative attributes such as the opinions of specialists. Furthermore, they can be used to study the viability of seeds of other species.

Keywords: Image analysis. X-ray test. Bayesian inference.

## 1 INTRODUÇÃO

A cultura de mamona vem se expandindo cada vez mais, não só no Brasil, mas mundialmente, pelo seu potencial de produção de óleo de excelente qualidade, para ser usado tanto na indústria como para suprir a demanda de biodiesel.

A disponibilidade de sementes de boa qualidade a serem utilizadas para o plantio é um dos grandes entraves para a expansão da cultura. Atualmente os testes mais utilizados para avaliação da qualidade dos lotes são os testes de germinação e vigor. Esses testes são destrutivos e consomem um período de tempo relativamente amplo. Esse período varia entre as diferentes espécies, sendo que para mamona é de 14 a 21 dias (BRASIL, 2009). Além disso, segundo Marcos Filho (2005), testes baseados em desempenho de plântulas produzem resultados mais variáveis que os desejados, pois a interpretação desses dependem da experiência e sensibilidade dos analistas.

Novas técnicas não-destrutivas para avaliar as características fisiológicas das sementes podem satisfazer a demanda dos agricultores, das indústrias de sementes. Essa demanda é focada principalmente na produção e no armazenamento de sementes de alta qualidade sob as normas de uma agricultura sustentável e de precisão.

Visando maior precisão e rapidez de execução da avaliação da qualidade de sementes diversos métodos tem sido desenvolvidos. Dentre eles, o método de raios-X se destaca por ser um método rápido e não destrutivo, mas tem a limitação dessa avaliação ser subjetiva. Esse método já é recomendado pela Associação Internacional de Análise de Sementes (INTERNATIONAL SEED TESTING ASSOCIATION - ISTA, 2004) com a finalidade de diferenciar sementes bem formadas das vazias, com danos mecânicos ou com ataque

de insetos. No entanto essa metodologia pode ser refinada e deve ser adaptada para as diferentes espécies.

As técnicas de análise bayesiana tem sido pouco utilizadas na área de análise de imagens de sementes. No entanto, podem ser ferramentas úteis para analisar dados obtidos na avaliação do potencial de viabilidade das sementes de mamona. Dessa forma, podem contribuir para o avanço dos testes de avaliação da qualidade dos lotes de sementes de mamona, bem como, de outras culturas.

O objetivo nesse capítulo foi estimar parâmetros para a avaliação do potencial de viabilidade de sementes de mamona por meio de inferência bayesiana, a partir das análises das radiografias dessas sementes feitas por três avaliadores independentes.

## 2 REFERENCIAL TEÓRICO

### 2.1 Teste dos raios-X

O uso da radiografia por meio dos raios-X de baixa energia é recomendado pela ISTA (2004) para determinar a qualidade das sementes. É considerado um método rápido e não destrutivo e tem como finalidade básica determinar sementes cheias, vazias, com danos mecânicos e com ataque de insetos. Segundo Bino, Aartase e Burg (1993), quando os raios-X passam por uma semente, a radiação é absorvida em vários graus, dependendo da espessura, densidade e composição da semente e do comprimento de onda da radiação. Cria-se, assim, uma imagem permanente em um filme radiográfico.

Segundo Simak (1991), dependendo da finalidade, três diferentes métodos radiográficos podem ser usados para análise de sementes: radiografia direta, de contraste ou tridimensional. No caso da radiografia direta, as sementes podem ser utilizadas secas, sem qualquer tipo de preparação especial como corte ou escarificação, ou seja, sem danificar o material analisado. A escolha do método apropriado, bem como a interpretação correta das imagens radiográficas, é de extrema importância. Uma das limitações desse método é a subjetividade na análise, pois essa depende da experiência e sensibilidade do avaliador, podendo levar a erros de interpretação.

O teste de raios-X foi utilizado, primeiramente, por Simak e Gustafsson (1953) na avaliação da qualidade de sementes de espécies florestais. Mas relatos na literatura durante as últimas duas décadas demonstram que essa técnica é eficiente e capaz de melhorar a informação sobre a produção de sementes de boa qualidade.

Argerich e Bradford (1989), Carvalho et al. (1999), Han, Bowers e Dodd (1992), e Liu et al. (1997) utilizaram o método de raios-X no acompa-

nhamento do desenvolvimento de sementes. Escasinas e Hill(1998) e Giradin, Chavagnat e Bockstaller (1993), na detecção de danos internos específicos. Battisti et al.(2000) na visualização de injúrias mecânicas e de danos por insetos. Giradin, Chavagnat e Bockstaller (1993), Liu et al. (1997), Oliveira (2004) e Oliveira, Carvalho e Davide (2003), na visualização de danos decorrentes de outros fatores adversos pré e pós-colheita. Simak, Bergen e Henrinson (1989) na análises de sementes não germinadas no final do teste de germinação. Cravioto et al.(2002), na seleção da fração de sementes puras. Liu et al. (1993), em estudos de morfologia e sementes submetidas ao “*priming*.”

Atualmente, análise de imagens por meio de raios-X, ainda, vem sendo amplamente utilizada, em vários segmentos da área de tecnologia de sementes.

Nassif e Cícero (2006) verificaram a possibilidade de se utilizar o teste de raios-X na detecção de anormalidades em embriões de sementes de acerola.

Socolowski e Cicero (2008) avaliaram as características morfológicas dos embriões de sementes de *Tecoma stans* L. Juss. ex Kunth pertencentes a diferentes classes de massa, por meio da análise de imagens obtidas pelo uso dos raios X, correlacionando-as com a germinação e morfologia das plântulas.

Mondo et al. (2009) avaliaram sementes de feijão da variedade cultivada IAC-Carioca submetidas a diferentes números e pressões de impactos mecânicos, pela análise de imagens fotográficas e de raios X.

Carvalho, Carvalho e Davide (2009) verificaram a ocorrência de alterações morfológicas internas, por meio de raios-X e os efeitos dessas alterações na qualidade de sementes de seis espécies de *Lauraceae* encontradas no sul de Minas Gerais.

Pinto et al. (2009) utilizaram técnica de análise de imagem por meio do teste de raios-X para identificar danos mecânicos em sementes de soja de diferentes lotes da cultivar BRS184.

Carvalho et al. (2009) avaliaram a eficiência da utilização do teste de raios X na detecção de danos internos, utilizando dez lotes de diferentes níveis de qualidade.

Santos et al. (2009) investigaram a aplicação da técnica de raios-X para avaliar a qualidade de sementes do mamão brasileiro (híbrido UENF/CALIMAN 01) e no estabelecimento de um protocolo experimental.

## 2.2 Inferência bayesiana

A inferência bayesiana é o processo de encontrar um modelo de probabilidade para um conjunto de dados e resumir o resultado por uma distribuição de probabilidades sobre os parâmetros do modelo e sobre quantidades não observadas, tais como predição para novas observações (GELMAN et al., 2000).

Em outras palavras, admite-se que as incertezas são passíveis de serem descritas coerentemente por uma distribuição de probabilidade. Assim, um procedimento de estimação do parâmetro  $\theta$  consiste em descrever a incerteza inicial do pesquisador sobre  $\theta$ , por meio da distribuição *a priori*,  $\pi(\theta)$ , e em seguida, combina-se essa informação com aquela proveniente dos dados  $X$ , resumida na função de verossimilhança. O resultado desse procedimento é uma distribuição de probabilidade,  $\pi(\theta|X)$ , chamada de distribuição *a posteriori* e a inferência sobre  $\theta$  é feita por meio da caracterização dessa distribuição (MIGON; SOUZA; SCHIMIDT, 2008).

O processo de inferência bayesiana é baseado no Teorema de Bayes e pode ser representado por:

$$\pi(\theta|X) = \frac{L(\theta|X)\pi(\theta)}{\int_{\Theta} L(X|\theta)\pi(\theta)d\theta} \quad (1)$$

em que

- a)  $\pi(\theta|X)$  é a distribuição *a posteriori*;
- b)  $\pi(\theta)$  é a distribuição *a priori*;
- c)  $L(\theta|X)$ , é a função de verossimilhança de  $\theta$  correspondente à amostra aleatória observada, dada por

$$L(\theta|X) = \prod_{i=1}^n f(x_i|\theta)$$

em que  $x_1, x_2, \dots, x_n$  é uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ , com  $\theta \in \Theta$ , sendo  $\Theta$  o espaço paramétrico.

O denominador de (1) é uma constante com relação a  $\theta$ . Portanto o Teorema de Bayes pode ser representado como:

$$\pi(\theta|X) \propto L(\theta|X)\pi(\theta), \quad (2)$$

em que  $\propto$  representa proporcionalidade.

### 2.2.1 Distribuições *a priori*

A distribuição *a priori*,  $\pi(\theta)$ , vem sendo usada para descrever uma informação sobre os parâmetros desconhecidos antes que se possa avaliar os dados em questão, isto é, expressa o conhecimento que o pesquisador tem a respeito do problema, sendo muitas vezes baseada simplesmente na sua expectativa subjetiva deste (BOX; TIAO, 1992).

Quando o pesquisador tem alguma informação prévia sobre o estudo em questão ele pode usar uma *priori* informativa. Segundo Migon, Souza e Schimidt (2008), um dos grandes desafios em inferência bayesiana é a especificação da distribuição *a priori* dos parâmetros. Uma maneira de especificar

a priori é por meio do uso de formas funcionais, ou seja, o conhecimento a priori de  $\theta$  é usado para associar uma família de distribuições paramétricas. Dentre as classes de prioris baseadas em formas funcionais, a priori conjugada é a mais utilizada. A classe de prioris conjugadas é definida a seguir.

**Definição 1** *Assumindo que foi observada uma amostra aleatória de tamanho  $n$ ,  $X = x_1, x_2, \dots, x_n$ , de  $\pi(X|\theta)$ , uma classe de distribuições a priori  $\mathcal{P}$  é dita conjugada à família de distribuições amostrais  $\mathcal{F} = \{\pi(X|\theta), \theta \in \Theta\}$ , se para todo  $\pi \in \mathcal{F}$  e  $\pi(\theta) \in \mathcal{P}$  então  $\pi(\theta|X) \in \mathcal{P}$ .*

Migon e Gamerman (1999) mostram as famílias de distribuições conjugadas. A seguir será mostrada mais detalhadamente a família de distribuições beta conjugada a Bernoulli, que foi utilizada como distribuição a priori neste estudo.

**Teorema 1** *Suponha que  $\theta$  seja a probabilidade de sucesso em determinado ensaio de Bernoulli,  $\pi(X|\theta) \sim \text{Ber}(\theta)$ . Se  $\theta$  for associado a uma distribuição a priori na família de distribuições beta,  $\pi(X|\theta) \sim \text{Beta}(\alpha, \beta)$  então a distribuição a posteriori de  $\theta$ ,  $\pi(\theta|X)$ , será a distribuição Beta com parâmetros  $(\alpha + T)$  e  $(n - T + \beta)$ , em que  $T = \sum_{i=1}^n x_i$ , isto é,  $\pi(\theta|X) \sim \text{Beta}(\alpha + n, \beta + T)$ .*

**Demonstração 1** *Por meio do teorema de Bayes, tem-se*

$$\begin{aligned} \pi(\theta|X) &\propto L(\theta|X)\pi(\theta) \\ &\propto \theta^T(1-\theta)^{n-T}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \underbrace{\theta^{\alpha+T-1}(1-\theta)^{n-T+\beta-1}}_{\text{Beta}(\alpha+T, n-T+\beta)} \end{aligned} \quad (3)$$

Para Migon, Souza e Schimidt (2008) a principal vantagem de se usar

*prioris* conjugadas é a possibilidade de reconhecer a distribuição *a posteriori* de  $\theta$  em (1), olhando apenas para núcleo da densidade resultante, ou seja, não é necessário calcular a integral do denominador de (1). Entretanto, é comum o núcleo da *posteriori* não apresentar forma conhecida e assim a integral em (1) não apresenta forma analítica fechada. Para resolver essa integral é necessário aproximar a distribuição *a posteriori* de interesse por meio de métodos numéricos.

Se há pouca ou nenhuma informação para incorporar a *priori*, é desejável que a *priori* tenha o mínimo de influência possível sobre a distribuição *a posteriori*. Nesse caso, considera-se uma distribuição não informativa. Uma classe bastante conhecida de *prioris* não informativas é a introduzida por Jeffreys (1961)

### 2.2.2 Distribuições *a posteriori* e estimação

A distribuição *a posteriori*  $\pi(\theta|X)$  contém toda informação sobre o parâmetro a ser estimado,  $\theta$ . Geralmente é útil resumir, de alguma maneira, a informação contida na *posteriori*. Pode-se estimar  $\theta$  por meio de uma única quantidade, estimação pontual, ou associando medidas de incerteza, estimação por intervalo (MIGON; SOUZA; SCHIMIDT, 2008).

Inferir sobre a informação contida na *posteriori* por meio de um único valor de  $\theta$ , resulta em resumir ao extremo a informação disponível. É importante obter o quão é precisa sua informação sobre  $\theta$ . Uma forma de fazer isso é fornecer uma região de valores de  $\theta$ , que é tão pequena quanto possível, mas contém o máximo de probabilidade *a posteriori* possível. Esta região, para um nível de credibilidade  $(1 - \alpha)$ , é chamada de intervalo de credibilidade e é definida por:

**Definição 2** Seja  $\theta$  uma quantidade desconhecida, definida em  $\Theta$ .  $C \in \Theta$  é

uma região de  $100(1 - \alpha)\%$  de credibilidade para  $\theta$  se  $P(\theta \in C|X) \geq 1 - \alpha$ .

Assim, quanto menor for o intervalo, mais concentrada é a distribuição do parâmetro, ou seja, o tamanho do intervalo informa sobre a dispersão de  $\theta$ . É possível construir uma infinidade de intervalos, utilizando a definição (2), mas é de maior interesse aquele com menor comprimento possível. Os intervalos, com comprimentos mínimos, são obtidos tomando-se os valores de  $\theta$  com maior densidade *a posteriori* denominados intervalos de credibilidade com máxima densidade *a posteriori*, ou intervalos HPD.

### 2.2.3 Métodos de Monte Carlo via cadeias de Markov (MCMC)

Os métodos computacionais de Monte Carlo via Cadeias de Markov (MCMC) têm sido largamente usados na estatística bayesiana, possibilitando simular amostras de uma determinada densidade *a posteriori*  $\pi(\theta|X)$ , cuja geração direta é difícil ou complicada (GAMERMAN; LOPES, 2006).

Uma cadeia de Markov é um processo estocástico cujos dados são gerados a partir de uma distribuição de probabilidade, em que o valor gerado em um estado atual só depende do valor gerado no passo anterior, caracterizando, assim, dependência dos valores gerados.

A idéia básica do método é construir uma cadeia de Markov com distribuição de equilíbrio igual à de interesse. Nessa cadeia, cada estado pode ser atingido, tomando-se por base qualquer outro com um número finito de iterações. Após um número suficientemente grande de iterações, a cadeia converge para uma distribuição de equilíbrio, dando origem à amostra da distribuição de interesse, que pode ser usada nas inferências.

Existem vários métodos para a construção da cadeia de Markov. Dentre eles, os mais utilizados são o método de Metropolis-Hastings e o amos-

trador de Gibbs. O Amostrador de Gibbs é indicado quando a expressão da condicional completa *a posteriori* tem a forma conhecida e o algoritmo de Metropolis-Hastings é indicado quando essa não tem forma conhecida. Gaman e Lopes (2006) apresentam uma descrição detalhada dos métodos de simulação baseados nos métodos de Monte Carlo via Cadeias de Markov.

#### 2.2.4 Amostrador de Gibbs

O amostrador de Gibbs foi proposto por Geman e Geman (1984) e difundido por Gelfand e Smith (1990). É um esquema iterativo de amostragem de uma cadeia de Markov, cujo núcleo de transição é formado por distribuições condicionais completas (GAMERMAN; LOPES, 2006).

Para descrever o algoritmo assume-se que a distribuição  $\pi(\theta|x)$  correspondente a distribuição *a posteriori*, em que  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  e  $p$  é o número de parâmetros. E, também, que as densidades condicionais completas *a posteriori*  $\pi(\theta_i|\theta_{-i}, x)$  estão disponíveis, em que  $i = 1, 2, \dots, p$ ,  $\theta_i$  é a  $i$ -ésima componente do vetor de parâmetros  $\theta$  e  $\theta_{-i} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ , ou seja, é o vetor  $\theta$  retirando-se a  $i$ -ésima componente (PAULINO; TURKAMAN; MURTEIRA, 2003).

O algoritmo do Amostrador de Gibbs é descrito a seguir.

- a) Iniciar o contador de iterações da cadeia  $t = 1$  e estabelecer os valores iniciais  $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)$
- b) Obter um novo valor  $\theta^t = (\theta_1^t, \theta_2^t, \dots, \theta_p^t)$  por meio de  $\theta^{t-1}$  e sucessivas gerações de valores:

$$\begin{aligned}\theta_1^{(t)} &\sim \pi\left(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}\right) \\ \theta_2^{(t)} &\sim \pi\left(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}\right) \\ &\vdots\end{aligned}$$

$$\theta_p^{(t)} \sim \pi\left(\theta_p | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{p-1}^{(t)}\right)$$

- c) Muda-se o contador para  $t + 1$  e volta-se ao passo (b), até que a convergência seja alcançada.

À medida que o número  $t$  de iterações aumenta, a sequência de valores gerados se aproxima da distribuição de equilíbrio, ou seja, da densidade marginal desejada para cada parâmetro, quando se assume que a convergência foi atingida (GAMERMAN; LOPES, 2006). Uma vez que a convergência tenha sido atingida, para retirar a influência dos valores iniciais, descartam-se as  $k$  primeiras observações (*burn-in*). Para eliminar a correlação de valores consecutivos  $\theta^{(t)}$  e  $\theta^{(t+1)}$ , toma-se uma observação a cada  $l$  observações (*thin*). Obtém-se, assim, uma cadeia de tamanho  $L$  que converge para distribuição de equilíbrio.

### 2.2.5 Critérios de convergência

Para a avaliação da convergência da cadeia de Markov ao estado de equilíbrio, vários testes foram propostos. Dentre eles, estão os testes de Geweke (1992), Heidelberger e Welch (1993) e Raftery e Lewis (1992). Esses testes estão implementados no pacote BOA (*Bayesian Output Analysis*) do software R (R DEVELOPMENT CORE TEAM, 2009).

Pelo método de Geweke (1992), propõe-se o diagnóstico de convergência para cadeias de Markov, baseado no teste de igualdade de médias da primeira e última parte da cadeia de Markov, geralmente dos primeiros 10% e dos últimos 50%. Ele se baseia em técnicas de análise espectral.

O método de Raftery e Lewis (1992) é baseado na acurácia de estimação do quantil. O método fornece as estimativas do *burn-in*, que é o número de iterações que devem ser descartadas, o número de iterações que devem ser

computadas e, a distância mínima de uma iteração à outra para se obter a sub-amostra aproximadamente independente (*thin*). Segundo os autores, se o fator de dependência for maior que 5, pode-se concluir que a cadeia apresenta problemas de convergência.

O método de Heidelberger e Welch (1993), por meio de testes estatísticos, testa a hipótese nula de estacionariedade da amostra gerada. Se a hipótese nula é rejeitada para um dado valor, o teste é repetido depois de descartadas os primeiros 10% das iterações. Se a hipótese é novamente rejeitada, outros 10% são descartados após o descarte dos 10% primeiros. Este processo é repetido até que se tenha uma proporção de 50% dos valores iniciais descartados. Se a hipótese for novamente rejeitada ou o teste não conseguir ser realizado há uma indícios de falha da estacionariedade, indicando a necessidade de um número maior de iterações. Se o teste for satisfatório, o número inicial de iterações descartadas é indicado como o tamanho do *burn-in* (NOGUEIRA; SAFADI; FERREIRA, 2004).

### 3 MATERIAL E MÉTODOS

Os dados foram obtidos, a partir de uma amostra de três quilos de sementes de mamona da cultivar IAC-226, recém-colhidas (safra 2009).

Considerando essa amostra, foram selecionadas, aleatoriamente, 800 sementes e essas foram submetidas ao teste de raios-X. Dessas 800 sementes, 400 sementes foram submetidas ao teste, sem nenhum preparo prévio, e as outras 400 foram pré tratadas com a substância contraste gasosa, clorofórmio ( $CHCl_3$ ).

No tratamento com a substância contraste, as 400 sementes foram colocadas em um vidro com capacidade de um litro e o agente clorofórmio foi colocado na parte superior da tampa, com vedação de tule, para que não houvesse contato com as sementes. Conforme Gordon, Gosling e Wang (1991) as sementes foram expostas a esse tratamento, durante quatro horas e, logo após, as sementes foram imediatamente submetidas ao teste de raios-X.

Todos os procedimentos descritos abaixo foram realizados, separadamente, para amostra de 400 sementes, sem preparo prévio e para amostra de 400 sementes pré tratadas com substância contraste.

#### 3.1 Obtenção das imagens radiográficas

Para obtenção das imagens radiográficas, as sementes de mamona foram arranjadas em placas de isopor e identificadas uma a uma. Em cada placa, 50 sementes de mamona foram distribuídas na mesma posição, com a carúncula voltada para baixo, em 5 linhas e 10 colunas, igualmente espaçadas. As sementes foram encaixadas em alvéolos e fixadas com fitas adesivas transparentes, para que não saíssem da posição. A semente identificada como 1 foi

posicionada na placa 1, na primeira linha e primeira coluna. A identificada como 2 foi a semente da placa 1, linha 1 e coluna 2 e a como 50 foi a posicionada na linha 5 e coluna 10 da mesma placa. Na placa 2, foram colocadas as sementes identificadas de 51 a 100 e, assim, sucessivamente, até completar o número necessário de sementes utilizadas no experimento.

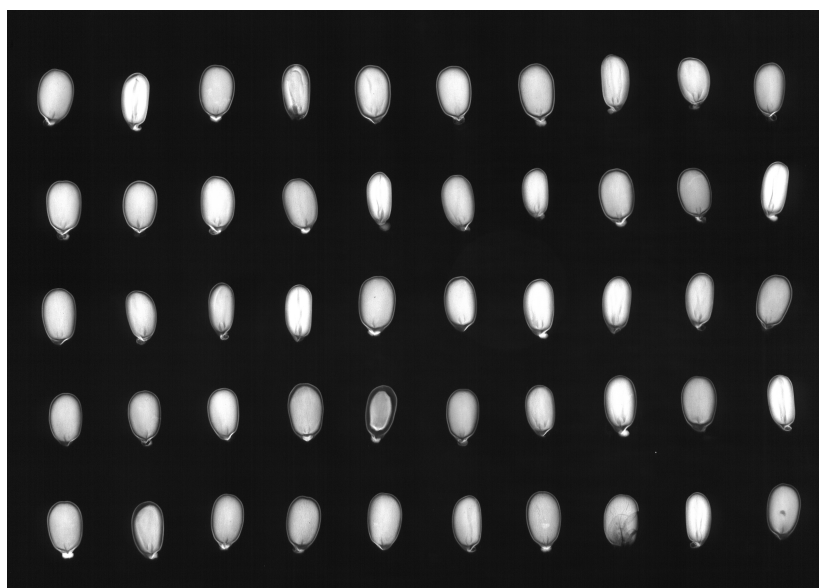


Figura 1 Radiografia das sementes de mamona

Após a preparação, as placas foram colocadas sobre o filme MR 2000-1 Kodak e submetidas, durante um minuto, à radiação a uma distância de 35cm da fonte. Foi utilizado aparelho de raios-X, modelo 43085-N, da marca Faxitron-HP. As imagens das sementes nas radiografias (Figura 1) foram analisadas visualmente.

### 3.2 Avaliação visual das imagens radiográficas das sementes de mamona

As imagens radiográficas de cada uma das sementes foram analisadas por três avaliadores independentes e classificadas conforme a anatomia interna das mesmas. Para cada semente classificada como potencialmente viável foi

associado o valor "0" e, como potencialmente inviável, o valor "1".

A avaliação como potencialmente viável ou inviável foi baseada conforme critério estabelecido por Carvalho, Alves e Oliveira (2010), que classificaram as sementes de acordo com a morfologia interna visualizada na radiografia e submetidas aos testes de germinação, emergência e taxa de crescimento de plântula. Sementes classificadas como completas e opacas (Figura 2A) tiveram maior porcentagem de germinação, seguidas pelas classificadas como completa e manchada e parcialmente cheia e opaca, Figura 2B e Figura 2C, respectivamente. Sementes classificadas como parcialmente cheias e translúcidas (Figura 2D), parcialmente cheias e com deformação do embrião (Figura 2E), parcialmente cheia e manchadas (Figura 2F), apresentaram um índice de germinação intermediário, e aquelas classificadas como vazia tiveram a menor germinação (Figura 2G).

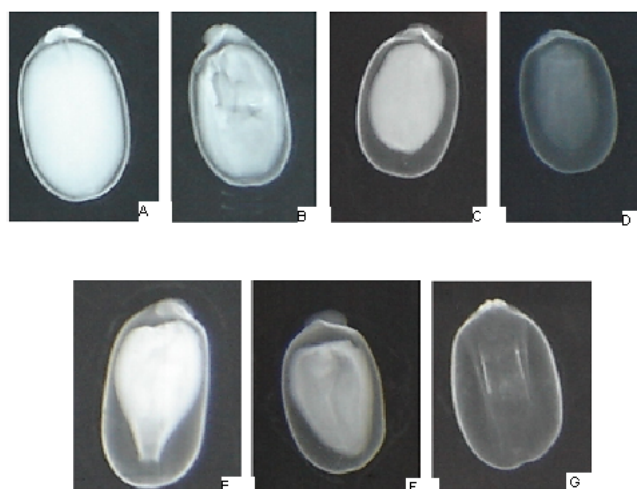


Figura 2 Radiografias de sementes de mamona classificadas em sete categorias conforme a sua morfologia interna: A - Cheia e opaca, B - Cheia e manchada, C - Parcialmente cheia e opaca, D - Parcialmente cheia e translúcida, E - Parcialmente cheia com defeito no embrião, F - Parcialmente cheia e manchada, G - Vazia. UFLA - Lavras, 2007

Assim, as sementes cujas imagens eram semelhantes às das Figuras(2A), (2B) ou (2C) foram consideradas como potencialmente viáveis, e das Figuras

(2D), (2E), (2F) ou (2G) como potencialmente inviáveis.

### **3.3 Estudo do potencial de viabilidade das sementes**

Baseado em estudos de Achcar e Ruffino Neto (2003) que propuseram uma metodologia para estimar parâmetros bayesianos no estudo da prevalência de tuberculose, por meio da análise de imagens de três exames de raios-X do tórax feita por um avaliador, uma adaptação dessa metodologia foi proposta nesse estudo. Três avaliadores independentes analisaram as mesmas imagens radiográficas de sementes de mamona. Essas análises foram utilizadas na estimação de parâmetros bayesianos para o estudo potencial fisiológico das sementes de mamona.

#### **3.3.1 Método das frequências relativas**

Para análise do potencial de viabilidade de sementes de mamona, submetidas à avaliação visual das imagens radiográficas feitas por três especialistas independentes, foram considerados quatro possíveis resultados (0, 1, 2 ou 3) para o número de avaliações como potencialmente inviáveis. O resultado foi “0”, quando nenhum dentre três avaliadores considerarem a imagem radiográfica da semente como potencialmente inviável. Foi “1” quando um dentre os três avaliadores, foi “2”, quando dois dentre três e foi “3” quando os três avaliadores consideraram a imagem radiográfica da semente como potencialmente inviável.

Considerando que 400 sementes foram analisadas pelos avaliadores, para cada um dos quatro possíveis resultados, o número de ocorrências é dado por  $(n_i)$ . Assim, a distribuição de frequências, para o número de classificações como potencialmente inviáveis ( $X$ ), é mostrada na (Tabela 1).

Tabela 1 Distribuição de frequência para os resultados das avaliações das sementes como potencialmente inviáveis

X	Número de Ocorrências	Frequência Relativa
0	$n_0$	$n_0/400$
1	$n_1$	$n_1/400$
2	$n_2$	$n_2/400$
3	$n_3$	$n_3/400$
Total	400	1

X: número de classificações como potencialmente inviáveis

Supondo independência entre os três avaliadores, pode-se pensar em duas situações com distribuições de probabilidade, definindo-se:

- $N$  como o evento que representa a semente que originou plântula normal;
- $p$  a probabilidade da semente ter originado uma plântula normal;
- $p_1$  a probabilidade da semente que originou uma plântula normal, tenha sido avaliada como potencialmente inviável;
- $p_2$  a probabilidade da semente, que não germinou ou não originou uma plântula normal, tenha sido avaliada como potencialmente inviável.

Na primeira situação,  $X$  é a variável aleatória que representa o número de avaliações consideradas como potencialmente inviáveis entre as sementes que originaram plântulas normais ( $N$ ). A distribuição condicional de  $X$  em relação a  $N$  é dada por uma binomial  $b(3, p_1)$ , tal que

$$P(X = x|N) = \binom{3}{x} p_1^x (1 - p_1)^{3-x} \quad (4)$$

em que  $x = (0, 1, 2, 3)$ ,  $p_1$  é probabilidade da semente ser inviável dentre as que originaram plântulas normais e  $(1 - p_1)$  é a probabilidade da semente ser viável dentre as que originaram plântulas normais.

Na segunda situação, a variável aleatória  $X$  representa o número de avaliações consideradas inviáveis entre as sementes que originaram plântulas

não normais ( $\bar{N}$ ), isto é, sementes que não germinaram ou originaram plântulas anormais. A distribuição condicional de  $X$  em relação a  $\bar{N}$  é dada por uma binomial  $b(3, p_2)$ , tal que

$$P(X = x|\bar{N}) = \binom{3}{x} p_2^x (1 - p_2)^{3-x} \quad (5)$$

em que  $x = (0, 1, 2, 3)$ ,  $p_2$  é a probabilidade de ser inviável dentre as que originaram plântulas não normais e  $(1 - p_2)$  é a probabilidade da semente ser viável dentre as que originaram plântulas não normais.

Sendo  $p$  a probabilidade da semente ter originado uma plântula normal,  $P(N)$ , então  $(1 - p)$  é a probabilidade da semente de ter originado uma plântula não normal,  $P(\bar{N})$ . Assim, a probabilidade não condicional da variável aleatória  $X$  é dada por:

$$\begin{aligned} P(X = x) &= P(X = x, N) + P(X = x, \bar{N}) \\ &= P(X = x|N)P(N) + P(X = x|\bar{N})P(\bar{N}). \end{aligned} \quad (6)$$

Reescrevendo 6 tem-se:

$$P(X = x) = p \binom{3}{x} p_1^x (1 - p_1)^{3-x} + (1 - p) \binom{3}{x} p_2^x (1 - p_2)^{3-x} \quad (7)$$

em que  $x = 0, 1, 2, 3$ .

Logo, tem-se o sistema:

$$\begin{cases} P(X = 0) = p(1 - p_1)^3 + (1 - p)(1 - p_2)^3, & \text{se } x = 0; \\ P(X = 1) = 3pp_1(1 - p_1)^2 + 3(1 - p)p_2(1 - p_2)^2, & \text{se } x = 1; \\ P(X = 2) = 3pp_1^2(1 - p_1) + 3(1 - p)p_2^2(1 - p_2), & \text{se } x = 2; \\ P(X = 3) = pp_1^3 + (1 - p)p_2^3, & \text{se } x = 3. \end{cases} \quad (8)$$

Utilizando o teorema de Bayes, tem-se que a probabilidade da semente originar uma plântula normal, dado o número de avaliações como potencialmente inviáveis, pode ser escrita como:

$$P(N|X = x) = \frac{P(X = x|N)P(N)}{P(X = x)} \quad (9)$$

Substituindo (4) e (6) em (9) tem-se que

$$P(N|X = x) = \frac{pp_1^x(1 - p_1)^{3-x}}{pp_1^x(1 - p_1)^{3-x} + (1 - p)p_2^x(1 - p_2)^{3-x}} \quad (10)$$

em que  $x = (0, 1, 2, 3)$  e  $P(N|X = x)$  é a probabilidade a *posteriori* da semente originar uma plântula normal, dado que foram  $x$  avaliações como potencialmente inviáveis dentre os 3 avaliadores.

Na estimação dos parâmetros pontuais desconhecidos  $p$ ,  $p_1$  e  $p_2$  foi utilizado o sistema de equações obtido, tomando-se por base (8) e das frequências relativas (Tabela 1) dos resultados do números de avaliações como potencialmente inviáveis.

Substituindo as estimativas dos parâmetros  $p$ ,  $p_1$  e  $p_2$  em (10), obtém-se as estimativas das probabilidades condicionais,  $P(N|X = x)$ . Multiplicando essas estimativas pelo número de ocorrências para cada uma das possíveis classificações como potencialmente inviáveis (Tabela 1), obtém-se o valor esperado de plântulas normais a partir das avaliações dos especialistas.

Além do método da frequência relativa, outros métodos podem ser utilizados para se obter as estimativas para  $p$ ,  $p_1$  e  $p_2$ . Dentre esses métodos, destaca-se o método bayesiano, via simulação Monte Carlo em Cadeias de Markov (MCMC), por apresentar grande precisão e simplicidade na análise de dados.

### 3.3.2 Método bayesiano via simulação Monte Carlo em cadeias de Markov (MCMC)

Considerando que a variável aleatória  $X$  (número de avaliações como potencialmente inviáveis), tem-se uma distribuição de probabilidade dada por uma mistura de duas distribuições binomiais:

$$P(X = x) = \sum_{j=1}^2 \lambda_j \binom{3}{x} p_j^x (1 - p_j)^{3-x} \quad (11)$$

em que  $x = (0, 1, 2, 3)$ ;  $\lambda_j$  é a probabilidade da plântula originada ser realmente normal e  $\lambda_1 + \lambda_2 = 1$ ;  $p_1 = P(X = x|N)$  e  $p_2 = P(X = x|\bar{N})$ .

Para uma amostra aleatória de  $n$  sementes a função de verossimilhança para  $\lambda_1$ ,  $p_1$  e  $p_2$  é dada por:

$$L(\lambda_1, p_1, p_2) = \prod_{i=1}^n \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1 - p_j)^{3-x_i}. \quad (12)$$

Como os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  variam entre 0 e 1, para a análise bayesiana do modelo serão consideradas as seguintes distribuições *a priori*:

$$\begin{aligned} \lambda_1 &\sim \text{Beta}(a, b) \\ p_1 &\sim \text{Beta}(c_1, d_1) \\ p_2 &\sim \text{Beta}(c_2, d_2) \end{aligned} \quad (13)$$

em que  $a, b, c_1, d_1, c_2$  e  $d_2$  são os hiperparâmetros conhecidos das *prioris* e assume-se que  $\lambda_1$ ,  $p_1$  e  $p_2$  são independentes.

Quando não se tem informação *a priori* sobre os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$ , assume-se uma *priori* não informativa. Nesse caso, igualam-se os hiperparâmetros  $a = b = 1$  e considera-se como distribuição *a priori* a distribuição uniforme,  $U(0, 1)$ .

Usando o teorema de Bayes (equação 2) tem-se que a distribuição a *posteriori* conjunta para  $\lambda_1$ ,  $p_1$  e  $p_2$  dado  $\mathbf{x}' = (x_1, x_2, \dots, x_n)$  pode ser escrita por:

$$\begin{aligned} \pi(\lambda_1, p_1, p_2 | X = x) &\propto \lambda_1^{a-1} (1 - \lambda_1)^{b-1} \\ &\quad p_1^{c_1-1} (1 - p_1)^{d_1-1} p_2^{c_2-1} (1 - p_2)^{d_2-1} \\ &\quad \times \left[ \prod_{i=1}^n \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1 - p_j)^{3-x_i} \right]. \end{aligned} \quad (14)$$

Para facilitar o algoritmo e eliminar o produto de somatórios na distribuição a *posteriori* conjunta, Tanner e Wong (1987) introduziram variáveis artificiais ou latentes. Definiu-se  $m'_i = (m_{1i}, m_{2i})$ , em que  $i = 1, \dots, n$ ,  $m_{1i} + m_{2i} = 1$  e  $m_{1i}$  tem uma distribuição de Bernoulli com a probabilidade de sucesso  $h_{1i}$  dada por:

$$h_{1i} = \frac{\lambda_1 \binom{3}{x_i} p_1^{x_i} (1 - p_1)^{3-x_i}}{\sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1 - p_j)^{3-x_i}} \quad (15)$$

em que  $i = 1, \dots, n$ .

Assim, obteve-se a distribuição conjunta de  $m'_i$  dados  $\lambda_1$ ,  $p_1$ ,  $p_2$  e  $x_i$  é:

$$\pi(m'_i | \lambda_1, p_1, p_2, x_i) \propto h_{1i}^{m_{1i}} (1 - h_{1i})^{m_{2i}} \quad (16)$$

isto é,

$$\pi(m'_1, m'_2, \dots, m'_n | \lambda_1, p_1, p_2, x_i) \propto \prod_{i=1}^n h_{1i}^{m_{1i}} (1 - h_{1i})^{m_{2i}} \quad (17)$$

em que  $i = 1, \dots, n$ . Como  $m_{1i} + m_{2i} = 1$ , então

$$\begin{aligned}
\pi(m'_1, m'_2, \dots, m'_n | \lambda_1, p_1, p_2, x_i) &\propto \lambda_1^{r_1} (1 - \lambda_1)^{r_2} \\
& p_1^{s_1} (1 - p_1)^{3r_1 - s_1} p_2^{r_2} (1 - p_2)^{3r_2 - s_2} \\
& \times \left[ \prod_{i=1}^n \sum_{j=1}^2 \lambda_j \binom{3}{x_i} p_j^{x_i} (1 - p_j)^{3 - x_i} \right]
\end{aligned} \tag{18}$$

em que  $r_1 = \sum_{i=1}^n m_{1i}$ ;  $r_2 = \sum_{i=1}^n m_{2i}$ ;  $s_1 = \sum_{i=1}^n m_{1i}x_i$  e  $s_2 = \sum_{i=1}^n m_{2i}x_i$ .

Combinando-se (14) e (18) tem-se:

$$\begin{aligned}
\pi(\lambda_1, p_1, p_2 | x, m) &\propto \lambda_1^{a+r_1-1} (1 - \lambda_1)^{b+r_2-1} p_1^{c_1+s_1-1} (1 - p_1)^{d_1+3r_1-s_1-1} \\
& p_2^{c_2+r_2-1} (1 - p_2)^{d_2+3r_2-s_2-1}
\end{aligned} \tag{19}$$

Logo, a partir de (19), são encontradas as distribuições condicionais a *posteriori*:

$$\begin{aligned}
\lambda_1 | p_1, p_2, x, m &\sim \text{Beta}(a + r_1, b + r_2) \\
p_1 | \lambda_1, p_2, x, m &\sim \text{Beta}(c_1 + s_1, d_1 + 3r_1 - s_1) \\
p_2 | \lambda_1, p_1, x, m &\sim \text{Beta}(c_2 + s_2, d_2 + 3r_2 - s_2)
\end{aligned} \tag{20}$$

Para encontrar amostras dos parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  para a distribuição a *posteriori* marginal foi utilizado o amostrador de Gibbs, descrito em (2.2.4). Foram considerados como valores iniciais para os parâmetros  $\lambda_1^0$ ,  $p_1^0$  e  $p_2^0$  as estimativas  $p$ ,  $p_1$  e  $p_2$ , respectivamente, calculadas por meio do método das frequências relativas (3.3.1). Essas amostras foram geradas pelas das condicionais completas a *posteriori* dadas em (20), conforme o algoritmo:

- a) Dado o valor inicial  $\lambda_1^0$ ,  $p_1^0$  e  $p_2^0$ , gerou-se  $n$  observações da distribuição de Bernoulli com probabilidade de sucesso  $h_{1i}$  dada em 15;

- b) Iniciou-se o contador  $t = 1$ ;
- c) Gerou-se  $\lambda_1^t$ ,  $p_1^t$  e  $p_2^t$  a partir das distribuições condicionais a *posteriori*:
- $\pi(\lambda_1 | p_1, p_2, x, m)$ ;
  - $\pi(p_1 | \lambda_1, p_2, x, m)$ ;
  - $\pi(p_2 | \lambda_1, p_1, x, m)$ ;
- d) Incrementou-se o contador ( $t + 1$ ) e repete-se (c) até que a convergência seja alcançada.

A implementação do amostrador de Gibbs foi feita no software R (R DEVELOPMENT CORE TEAM, 2009) e verificação da convergência da cadeia foi feita utilizando-se o pacote BOA do mesmo software. As rotinas com a implementação do amostrador de Gibbs são apresentadas no ANEXO A.

### 3.4 Simulação de amostras de diferentes tamanhos

Foram simuladas amostras de sementes de tamanhos 200 e 500 considerando os dados das amostras observadas. A amostra observada tinha 400 sementes, analisadas por três avaliadores diferentes, mas o conjunto de dados tinha  $n_i$  ocorrências dos possíveis  $i$  resultados, com  $i = 0, 1, 2, 3$ , conforme descrito na Seção (3.3.1).

Para o cálculo do número de ocorrências, para cada um dos possíveis resultados da amostra simulada, tomou-se o produto do tamanho dessa pela frequência relativa da amostra observada (Tabela 1). Assim, por exemplo, na amostra simulada, de tamanho 500 o número de dados iguais a “0” foi  $(500n_0/400)$ , iguais a “1” foi  $(500n_1/400)$ , “2” foi  $(500n_2/400)$  e “3” foi  $(500n_3/400)$ .

Para cada amostra simulada, foi feita análise bayesiana do potencial

de viabilidade, conforme os procedimentos da Seção (3.3.2). Os resultados dessas análises foram comparados com os resultados das amostras observadas.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Método das frequências relativas

A distribuição de frequência do número das sementes potencialmente inviáveis, classificadas conforme Seção (3.3), para o teste de raios-X, aplicados na amostra de 400 sementes sem nenhum preparo e na amostra das 400 sementes preparadas com contraste é apresentado na Tabela 2.

Tabela 2 Distribuição de frequências para as amostras de 400 sementes sem nenhum preparo e de 400 sementes preparadas com contraste classificadas por três avaliadores e probabilidades condicionais e valores esperados calculados pelo método frequentista

	Sem preparo				Com contraste			
	Análise Visual		MFR		Análise Visual		MFR	
X	$F_i$	$F_r$	P(N X)	$E(N)_i$	$F_i$	$F_r$	P(N X)	$E(N)_i$
0	291	0,73	0,999	286,36	318	0,79	0,991	317,74
1	62	0,16	0,762	47,22	46	0,12	0,928	42,69
2	30	0,08	0,086	2,58	16	0,04	0,119	1,91
3	17	0,04	0,003	0,05	20	0,05	0,001	0,04

X: número de classificações como potencialmente inviáveis;  $F_i$ : Frequência Absoluta;  $F_r$ : Frequência Relativa;  $E(N)_i = P(N|X) \times F_i$ : número esperado de plântulas normais; MFR: método da frequência relativa

Utilizando-se as frequências relativas (Tabela 2) resolveu-se o sistema (8). Os parâmetros estimados  $p$ ,  $p_1$  e  $p_2$  para as sementes sem preparo foram, respectivamente, 0,8455, 0,05176, e 0,6496 e para as sementes preparadas com contraste foram, respectivamente, 0,9059, 0,04863 e 0,8096. Com base nesses parâmetros foram calculadas, utilizando (6), as probabilidades condicionais das plântulas originadas serem normais, dado o número de avaliações como potencialmente inviáveis. Também foram calculados os valores esperados de plântulas normais  $E(N)_i$ , dados pelo produto entre a probabilidade condicional das plântulas originadas serem normais e as frequências absolutas, para cada uma das possíveis classificações para as duas amostras de 400 sementes. Essas probabilidades condicionais e o número esperado de plântulas

normais são mostrados na Tabela 2.

O método da frequência relativa, utilizado para estimar o número esperado de plântulas normais (Tabela 2), gerou boas estimativas, tanto para as amostras de sementes sem preparo como para as com contraste, pois, se três avaliadores classificaram as sementes como potencialmente inviáveis, espera-se que o número de plântulas normais geradas seja próximo de zero, e, se o número de classificações como potencialmente inviável for igual a zero, espera-se que o número de plantas normais seja próximo do número de classificações iguais a zero ( $F_0$ ).

Uma desvantagem desse método consiste na dificuldade da obtenção do desvio padrão, associado às estimativas pontuais, para os parâmetros  $p$ ,  $p_1$  e  $p_2$ . Isso ocorre, pois, o modelo utilizado neste estudo é dado por uma mistura de binomiais e, segundo Titterton, Smith e Marcov (1985), esse tipo de modelo, usualmente, apresenta problemas, para estimar médias e variâncias, usando o método dos momentos e da máxima verossimilhança.

#### **4.2 Método bayesiano via simulação Monte Carlo em cadeias de Markov (MCMC)**

Para efetuar a análise bayesiana do potencial de viabilidade de plântulas normais das amostras de 400 sementes sem nenhum preparo e de 400 sementes preparadas com contraste, foram considerados os dados da avaliação das radiografias por três especialistas independentes, apresentados na Tabela 2. Conforme metodologia descrita em (3.3.2), para se encontrar amostras da distribuição conjunta, utilizou-se o amostrador de Gibbs. Os valores iniciais para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  para as sementes sem preparo foram, respectivamente, 0,8455, 0,05176, e 0,6496 e para as sementes preparadas com contraste foram, respectivamente, 0,9059, 0,04863 e 0,8096, calculados pelo

método das frequências relativas. E os hiperparâmetros da distribuição *priori*:  $a = b = c_1 = d_1 = c_2 = d_2 = 1^1$ . No procedimento de simulação, para cada iteração do amostrador de Gibbs, foram seguidos os passos:

- a) Geram-se valores  $m_{1i}$  da distribuição de Bernoulli com probabilidade de sucesso  $h_{1i}$ , dada em (15).
- b) A partir dos valores de  $m_{1i}$  gerados em (i.), calcula-se os valores de  $r_1 = \sum_{i=1}^n m_{1i}$ ;  $r_2 = \sum_{i=1}^n m_{2i}$ ;  $s_1 = \sum_{i=1}^n m_{1i}x_i$  e  $s_2 = \sum_{i=1}^n m_{2i}x_i$ .
- c) Calculou-se os parâmetros  $\lambda_1^t$ ,  $p_1^t$  e  $p_2^t$  das distribuições condicionais a *posteriori*, dadas por  $Beta(1 + r_1, 1 + r_2)$ ,  $Beta(1 + s_1, 1 + 3r_1 - s_1)$  e  $Beta(1 + s_2, 1 + 3r_2 - s_2)$ , respectivamente.

Foram geradas 1000 amostras com 20000 iterações cada. Para cada amostra, das 20000 iterações geradas, as primeiras 1000 foram descartadas e foi tomada uma a cada 19 iterações, obtendo-se uma cadeia com 1000 dados para cada uma das amostras com 400 sementes sem preparo e com contraste.

Tabela 3 Resultados dos testes para verificar a convergência da cadeia

Testes	Raftery e Lewis	Heidlbeger e Welch	Geweke		
Parâmetros	FD	E	HW	z	p
Sem preparo					
$\lambda_1$	1,327	Sim	Sim	-0,56	0,57
$p_1$	2,115	Sim	Sim	-0,16	0,87
$p_2$	1,326	Sim	Sim	-0,10	0,92
Com contraste					
$\lambda_1$	0,953	Sim	Sim	-0,26	0,79
$p_1$	1,124	Sim	Sim	-0,99	0,31
$p_2$	1,123	Sim	Sim	-0,54	0,51

**FD**: fator de dependência; **E**: estacionária; **HW**: Half-width Test; **z**: z-score; **p**: p-valor

Na Tabela 3 são apresentados os resultados dos testes de Raftery e Lewis, Heidelberg e Welch e Geweke. A convergência foi verificada pelo teste Raftery e Lewis <sup>2</sup>, pois o fator de dependência (FD) foi menor que “5” para

<sup>1</sup>Com os hiperpâmetros iguais a 1, tem-se uma distribuição Uniforme,  $U(0,1)$

<sup>2</sup>O teste foi realizado ao nível 5% de acurácia

todos os parâmetros. No teste de Heidelberg e Welch, em que todas as cadeias passaram pelos testes de estacionariedade e no teste de Geweke a convergência foi verificada por meio do *valor-p* obtido, que foi maior que o nível de significância pré-fixado de 5%.

Pela forma característica da densidade a *posteriori*, Figura 3, pode-se acreditar que as estimativas dos parâmetros são verdadeiramente, as médias das distribuições a *priori*.

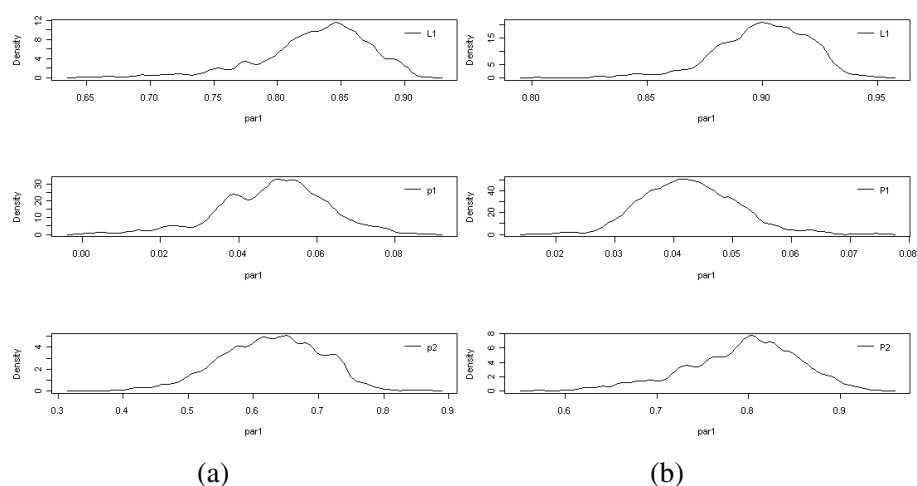


Figura 3 (a) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 400 sementes sem preparo. (b) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 400 sementes com contraste

Na Tabela 4, observa-se a média, desvio padrão a *posteriori* e os intervalos de credibilidade para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$ . Verificou-se que médias a *posteriori* para os parâmetros, aproximam-se dos resultados obtidos no método da frequência relativa. Foi, relativamente, fácil estimar os intervalos de credibilidade, o que é uma vantagem em relação ao método das frequências relativas. Os comprimentos dos intervalos de credibilidade para as amostras sem preparo e com contraste, tiveram uma boa precisão, uma vez que, foram calculados, com base na análise de três avaliadores diferentes.

Na estimação das probabilidades de originar uma plântula normal, ( $\lambda_1$ ), e de não originar uma plântula normal, dado que foi avaliado como potencialmente inviável ( $p_2$ ), o método do contraste foi mais preciso. Isso se deve ao fato de que a visualização de danos fica mais evidente nas radiografias das sementes submetidas à substância contraste. No entanto, esse método tem a desvantagem de ser destrutivo, uma vez que submissão das sementes à exposição com gases tóxicos afeta o processo germinação (GORDON; GOSLING; WANG, 1991).

Tabela 4 Estimativas da média a posteriori, desvio padrão a posteriori e intervalo de credibilidade

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade*	
			LI	LS
<b>Sem preparo</b>				
$\lambda_1$	0,8264	0,0454	0,73064	0,9023
$p_1$	0,0477	0,0145	0,0184	0,0755
$p_2$	0,6214	0,0783	0,4722	0,7758
<b>Com contraste</b>				
$\lambda_1$	0,9013	0,0197	0,8617	0,9369
$p_1$	0,0428	0,0078	0,0284	0,0593
$p_2$	0,7912	0,0617	0,6703	0,9117

\* ao nível de 95% de probabilidade; LI: limite inferior; LS: limite superior

Para encontrar as estimativas das probabilidades *a posteriori*,  $P(N|X = x)$ , dadas por 10, foram utilizados estimadores de Monte Carlo baseados nas médias condicionais *a posteriori* das 1000 amostras simuladas pelo amostrador de Gibbs, essas estimativas são dadas por:

$$\hat{P}(N|X = x) = \frac{\sum_{s=1}^{1000} \frac{\lambda_1^{(s)} p_1^{(s)x} (1 - p_1^{(s)})^{3-x}}{[\lambda_1^{(s)} p_1^{(s)x} (1 - p_1^{(s)})^{3-x} + (1 - \lambda_1^{(s)}) p_2^{(s)x} (1 - p_2^{(s)})^{3-x}]}}{1000} \quad (21)$$

Na Tabela 5 são apresentadas as estimativas bayesianas para as médias das condicionais *a posteriori* ( $P(N|X = x)$ ), o desvio padrão para as médias das condicionais *a posteriori* e intervalos de credibilidade para  $P(N|X = x)$  das amostras de 400 sementes sem preparo e de 400 sementes com contraste.

Também são apresentados o número esperado de plântulas normais,  $(E(N)_i)$ , para as duas amostras avaliadas, que é dado pelo produto entre as frequências absolutas (Tabela 1) e as médias a *posteriori* e o intervalo de credibilidade para o número esperado de plântulas normais.

Observa-se que as médias das condicionais a *posteriori* obtidas pelo método bayesiano são próximas das estimativas obtidas por meio do método da frequência relativa (Tabela 2). Além disso os intervalos de credibilidade tanto para médias das condicionais a *posteriori*, quanto para o número esperado de plântulas normais foram precisos, uma vez que foi incorporado ao método a análise de três avaliadores.

Assim, os métodos bayesianos podem ser úteis para inferir na qualidade dos lotes de sementes, pois além de obter boas inferências para os parâmetros, permitem incorporar a opinião de especialistas na área de tecnologia de sementes, o que pode levar a resultados bastante precisos e diminuir o erro intra avaliador da análise visual por meio das radiografias das sementes.

Tabela 5 Estimativas bayesianas para a média e desvio padrão das condicionais a *posteriori*, intervalo de credibilidade a 95% de probabilidade para média condicional a *posteriori*, o número esperado de plântulas normais para cada uma das classificações como potencialmente inviáveis, intervalos de credibilidade para o número esperado de plântulas normais

X	Média	SD	IC para Média		$E(N)_i$	IC para $E(N)_i$	
			LI	LS		LI	LS
<b>Sem preparo</b>							
0	0,9845	0,00060	0,9836	0,9850	284,521	284,272	284,665
1	0,6979	0,00951	0,6836	0,7059	43,269	42,384	43,764
2	0,0859	0,00251	0,0884	0,0895	2,577	2,650	2,686
3	0,0029	0,00008	0,0028	0,0031	0,049	0,048	0,052
<b>Com Contraste</b>							
0	0,9984	0,0001	0,9983	0,9984	317,501	317,459	317,513
1	0,9019	0,00120	0,9001	0,9030	41,492	41,4055	41,538
2	0,1249	0,00071	0,1237	0,1254	1,999	1,9782	2,006
3	0,0016	0,00002	0,0002	0,0016	0,033	0,0032	0,033

X: número de classificações como potencialmente inviáveis;  $E(N)_i$ : número esperado de plântulas normais; **IC para Média**: Intervalo de credibilidade a 95% de probabilidade para média da condicional a posteriori; **IC para  $E(N)_i$** : Intervalo de credibilidade a 95% de probabilidade para o valor esperado de plântulas normais; LI: limite inferior; LS: limite superior

### 4.3 Simulação de amostras de diferentes tamanhos

Para comparar as estimativas dos parâmetros obtidas, utilizando-se métodos bayesianos, para diferentes tamanhos de amostra, foram considerados os valores das frequências relativas (Tabela 2). Foram simuladas amostras de sementes de tamanhos 200 sem preparo e com contraste e de 500 sem preparo e com contraste. A distribuição de frequências para as amostras simuladas está apresentada na Tabela 6.

Para as amostras simuladas de 200 sementes sem preparo e com contraste e 500 sementes sem preparo e com contraste foram geradas 1000 amostras com 20000 iterações cada. Para cada amostra, das 20000 iterações geradas, as primeiras 9000 foram descartadas e foi tomada uma a cada 11 iterações, obtendo-se uma cadeia com 1000 dados para cada uma das amostras com 200 sementes sem preparo.

Tabela 6 Distribuição de frequências para as amostras simuladas de 200 sementes sem nenhum preparo e preparadas com contraste e 500 sementes sem nenhum preparo e preparadas com contraste

X	200 sementes		500 sementes	
	Sem preparo	Com contraste	Sem preparo	Com contraste
0	146	159	364	398
1	31	23	78	57
2	15	8	38	20
3	9	10	21	25

X: número de classificações como potencialmente inviáveis

A convergência da cadeia foi verificada por meio dos testes de Raftery e Lewis, Heidelberg e Welch e Geweke (Tabela 7).

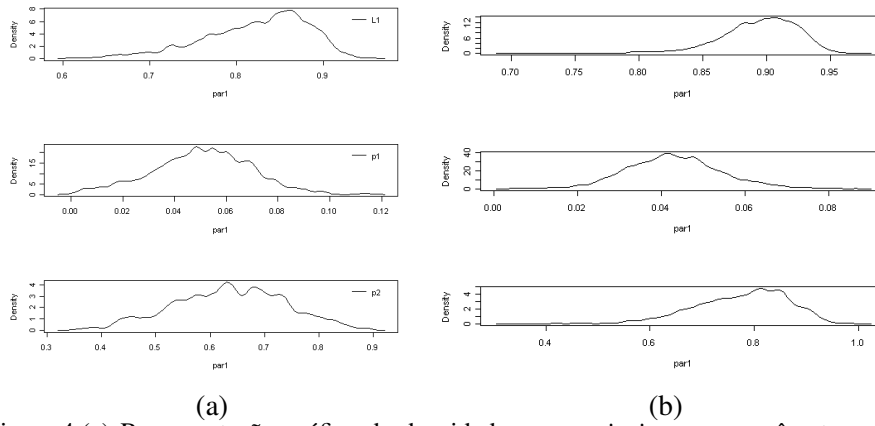
Tabela 7 Resultados dos testes para verificar a convergência da cadeia

Amostras	Testes Parâmetros	Raftery e Lewis	Heidlbeger e Welch	Geweke	
		<b>FD</b>	<b>E</b>	<b>HW</b>	<b>p-valor</b>
200 sementes sem preparo	$\lambda_1$	1,838	Sim	Sim	0,437
	$p_1$	1,124	Sim	Sim	0,768
	$p_2$	1,219	Sim	Sim	0,744
200 sementes com contraste	$\lambda_1$	1,327	Sim	Sim	0,059
	$p_1$	2,115	Sim	Sim	0,329
	$p_2$	2,969	Sim	Sim	0,169
500 sementes sem preparo	$\lambda_1$	1,124	Sim	Sim	0,228
	$p_1$	1,327	Sim	Sim	0,231
	$p_2$	1,365	Sim	Sim	0,446
500 sementes com contraste	$\lambda_1$	1,123	Sim	Sim	0,067
	$p_1$	1,219	Sim	Sim	0,489
	$p_2$	1,229	Sim	Sim	0,933

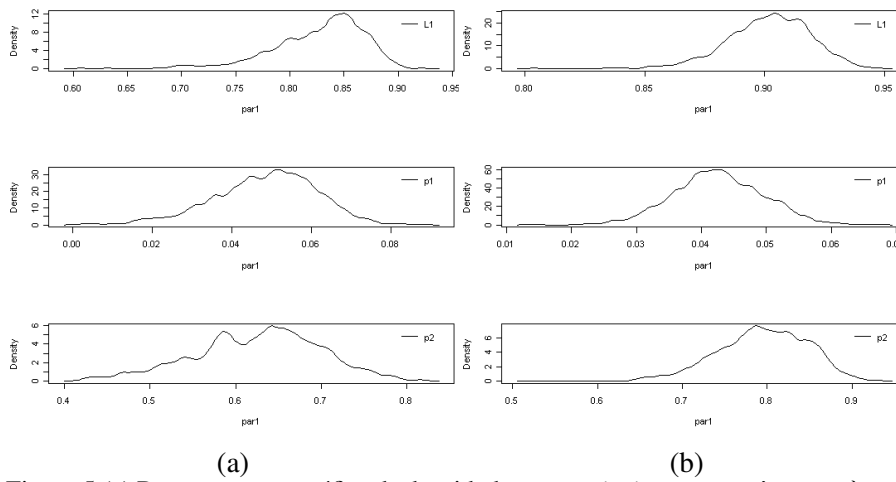
**FD:** fator de dependência; **E:** estacionária; **HW:** Half-width Test

Pela análise visual das Figuras 4 e 5, pode-se observar que pela forma característica das densidades *a posteriori* representa as estimativas dos parâmetros das amostras simuladas. Essa observação pode ser confirmada pelas estimativas apresentadas na Tabela 8.

Comparando as estimativas das médias *a posteriori* das amostras simuladas (Tabela 8) com as estimativas das amostras de 400 sementes sem preparo e com contraste, pode-se observar que não houve diferenças significativas entre essas. Essa observação, foi confirmada, por meio do teste t-Student, pois,



(a) (b)  
 Figura 4 (a) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 200 sementes sem preparo. (b) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 200 sementes com contraste



(a) (b)  
 Figura 5 (a) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 500 sementes sem preparo (b) Representação gráfica da densidade a *posteriori* para os parâmetros  $\lambda_1$ ,  $p_1$  e  $p_2$  da cadeia de 500 sementes com contraste

Tabela 8 Estimativas da média a posteriori, desvio padrão a posteriori e intervalo de credibilidade

Amostra	Parâmetros	Média	Desvio-padrão	IC*	
				LI	LS
200 sementes sem preparo	$\lambda_1$	0,813	0,062	0,681	0,913
	$p_1$	0,047	0,019	0,009	0,082
	$p_2$	0,620	0,103	0,415	0,798
200 sementes com contraste	$\lambda_1$	0,893	0,032	0,831	0,949
	$p_1$	0,043	0,0116	0,020	0,066
	$p_2$	0,774	0,093	0,590	0,925
500 sementes sem preparo	$\lambda_1$	0,827	0,0407	0,737	0,889
	$p_1$	0,047	0,0125	0,021	0,070
	$p_2$	0,622	0,0709	0,465	0,749
500 sementes com contraste	$\lambda_1$	0,901	0,017	0,868	0,933
	$p_1$	0,043	0,0070	0,030	0,057
	$p_2$	0,794	0,053	0,692	0,893

\*Intervalo de Credibilidade ao nível de 95% de probabilidade; LI: limite inferior; LS: limite superior

todos os o p-valores foram menores que 0,05.

As estimativas bayesianas para a média e desvio padrão das condicionais a *posteriori*, intervalo de credibilidade, para média condicional a *posteriori*, o número esperado de plântulas normais para cada uma das classificações como potencialmente inviáveis e intervalos de credibilidade, para o número esperado de plântulas normais para as amostras simuladas sem preparo e com contraste, são apresentadas na Tabela 9.

Comparando as médias condicionais a *posteriori* das amostras simuladas sem preparo e com preparo com as médias condicionais *posteriori* das amostras de 400 sementes sem preparo e com contraste (Tabela 5), por meio do teste t-Student, verificou-se que não existem diferenças entre as médias das amostras, pois os p-valores para todas as comparações foram menores que 0,05.

A ISTA (2004) recomenda para o teste de raios-X a utilização de 400 sementes. Como não houve diferenças significativas entre as estimativas das

Tabela 9 Estimativas bayesianas para a média e desvio padrão das condicionais a *posteriori*, intervalo de credibilidade a 95% de probabilidade para média condicional a *posteriori*, o número esperado de plântulas normais para cada uma das classificações como potencialmente inviáveis, intervalos de credibilidade para o número esperado de plântulas normais

X	Média	SD	IC Média		$E(N)_i$	IC $E(N)_i$	
			LI	LS		LI	LS
Amostra de 200 sementes sem preparo							
0	0,9807	0,0011	0,9796	0,9821	143,195	143,020	143,387
1	0,6741	0,0182	0,6627	0,6916	20,896	20,542	21,440
2	0,0927	0,0027	0,0905	0,0979	1,391	1,358	1,469
3	0,0034	0,0001	0,0032	0,0035	0,031	0,030	0,032
Amostra de 200 sementes com contraste							
0	0,9970	0,00020	0,9967	0,9972	158,5230	158,488	158,555
1	0,8666	0,00596	0,8586	0,8720	19,9306	19,748	20,056
2	0,1273	0,00519	0,1190	0,1321	1,0184	0,953	1,057
3	0,0019	0,00006	0,0017	0,0019	0,0187	0,018	0,019
Amostra de 500 sementes sem preparo							
0	0,9858	0,00024	0,9855	0,9859	358,830	358,730	358,881
1	0,6973	0,00806	0,6886	0,7044	54,386	53,709	54,949
2	0,0766	0,00229	0,0746	0,0791	2,912	2,835	3,006
3	0,0027	0,00010	0,0023	0,0028	0,057	0,048	0,059
Amostra de 500 sementes com contraste							
0	0,9985	0,00008	0,9984	0,9986	397,403	397,363	397,443
1	0,9039	0,00246	0,9017	0,9072	51,522	51,397	51,710
2	0,1209	0,00256	0,1183	0,1226	2,418	2,366	2,452
3	0,0016	0,00004	0,0015	0,0016	0,039	0,038	0,040

X: número de classificações como potencialmente inviáveis;  $E(N)_i$ : número esperado de plântulas normais; **IC Média**: Intervalo de credibilidade a 95% de probabilidade para média da condicional a *posteriori*; **IC  $E(N)_i$** : Intervalo de credibilidade a 95% de probabilidade para o valor esperado de plântulas normais; LI: limite inferior; LS: limite superior

amostras de 400 sementes e das amostras simuladas de tamanho 200 e 500, em futuras aplicações do método bayesiano para análise de radiografias de sementes de mamona poderão ser utilizadas amostras com 200 sementes, sem que haja prejuízo nas análises. Além disso, haverá economia de material e redução do tempo de análise.

## 5 CONCLUSÃO

A análise radiográfica das sementes de mamona pode ser utilizada na estimativa da viabilidade do lote de sementes.

O método bayesiano é eficiente na estimação das médias a *posteriori* dos parâmetros e das médias condicionais a *posteriori* para as possíveis classificações feitas pelos especialistas no estudo da viabilidade de sementes de mamona considerando de imagens radiográficas. Os intervalos de credibilidade, para o número esperado de plântulas normais, foi preciso. Não houve diferenças entre as médias a *posteriori* das amostras de 400 sementes sem preparo e com contraste para as médias das amostras simuladas.

Métodos bayesianos são de grande interesse prático, para obtenção de boas inferências na área de tecnologia de sementes. Com a vantagem de, além disso, poderem ser utilizados no estudo da viabilidade de sementes de outras espécies.

O método da frequência relativa apresentou bons resultados para os parâmetros e para as condicionais (P(N|X)). A desvantagem na utilização desse método em relação ao método bayesiano, consiste na dificuldade de estimar o desvio padrão, para o modelo de mistura de binomiais por meio do método da máxima verossimilhança.

## REFERÊNCIAS

- ACHCAR, J.; RUFFINO NETTO, A. Estudo da prevalência da tuberculose: uso de métodos bayesianos. **Revista Brasileira de Epidemiologia**, São Paulo, v. 6, n. 4, p. 380-387, set./dez. 2003.
- ARGERICH, C.; BRADFORD, K. The effects of priming and ageing on seed vigour in tomato. **Journal Experimental Botany**, London, v. 40, n. 5, p. 599-607, May 1989.
- BATTISTI, A. et al. Detection and evaluation of seed damage of cypress, *Cupressus sempervirens* A., in Italy. **Seed Science and Technology**, Zürich, v. 28, n. 3, p. 729-738, June 2000.
- BINO, R.; AARTASE, J.; BURG, W. van der. Non-destructive x-ray analysis of *Arabidopsis* embryo mutants. **Seed Science Research**, Wallingford, v. 3, n. 3, p. 167-170, Sept. 1993.
- BOX, G.; TIAO, G. **Bayesian inference in statistical analysis**. New York: J. Wiley, 1992. 603p.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Regra para análise de sementes**. Brasília, 2009. 399 p.
- CARVALHO, L. R.; CARVALHO, M. L. M.; DAVIDE, A. C. Utilização do teste de raios X na avaliação da qualidade de sementes de espécies florestais de Lauraceae. **Revista Brasileira de Sementes**, Brasília, v. 31, n. 4, p. 57-66, jul./ago. 2009.
- CARVALHO, M. L. M.; ALVES, R. A.; OLIVEIRA, L. M. Radiographic analysis in castor bean seeds (*Ricinus communis* L.). **Revista Brasileira de Sementes**, Brasília, v. 32, n. 1, p. 170-175, jan. 2010.
- CARVALHO, M. L. M. et al. Pre-harvest stress cracks in maize (*Zea mays* L.) kernels as characterized by visual, x-ray and low temperature scanning electron microscopical analysis: effect on kernel quality. **Seed Science Research**, Wallingford, v. 9, n. 3, p. 227-236, Sept. 1999.
- CARVALHO, M. L. M. et al. Teste de raios x na avaliação da qualidade de sementes de abóbora. **Revista Brasileira de Sementes**, Brasília, v. 31, n. 2, p. 221-227, mar./abr. 2009.

CRAVIOTO, R. et al. Description of pure seed fraction of oat through usual evaluations and radiographic images. **Pesquisa Agropecuária Brasileira**, Brasília, v. 37, n. 8, p. 1183-1188, ago. 2002.

ESCASINAS, A.; HILL, M. Stress cracking in maize (*Zea mays* L.) seeds. **Annals of Tropical Research**, Baybay Leyte, v. 10, n. 3/4, p. 182-192, Dec. 1988.

GAMERMAN, D.; LOPES, H. **Markov chain Monte Carlo**: stochastic simulation for bayesian inference. London: Chapman & Hall, 2006. 236 p. (Texts in Statistical Science Series).

GELFAND, A.; SMITH, A. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, New York, v. 85, n. 410, p. 398-409, June 1990.

GELMAN, A. et al. **Bayesian data analysis**. Boca Raton: Chapman & Hall, 2000. 526 p.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 6, n. 6, p. 721-741, Aug. 1984.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. et al. (Ed.). **Bayesian statistics**. Oxford: Oxford University, 1992. p. 169-193.

GIRADIN, P.; CHAVAGNAT, A.; BOCKSTALLER, C. Determination des caractéristiques des semences de maïs grâce à la radiographie aux rayons x. **Seed Science and Technology**, Zürich, v. 21, n. 3, p. 545-551, June 1993.

GORDON, A.; GOSLING, P.; WANG, B. **Tree and shrub seed handbook**. 4th ed. Zürich: ISTA, 1991. 190 p.

HAN, Y.; BOWERS, S.; DODD, R. Nondestructive detection of split-pit peaches. **Transactions of ASAE**, Saint Joseph, v. 35, n. 6, p. 2063-2067, Nov./Dec. 1992.

HEIDELBERG, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v. 31, n. 6, p. 1109-1144, Nov./Dec. 1993.

INTERNATIONAL SEED TESTING ASSOCIATION. **International rules for Seed Testing Association**. Zürich, 2004. 174 p.

JEFFREYS, H. **Theory probability**. 3rd ed. Oxford: Oxford University, 1961. 470 p.

LIU, Y. et al. Amounts of nuclear DNA and internal morphology of gibberellin-and abscisic acid-deficient tomato (*Lycopersicon esculentum* Mill.) seeds during maturation, imbibition and germination. **Annals of Botany**, London, v. 79, n. 2, p. 164-168, Feb. 1997.

\_\_\_\_\_. X-ray studies on change in embryo and endosperm morphology during priming and imbibition of tomato seeds. **Seed Science Research**, Wallingford, v. 3, n. 3, p. 171-178, Sept. 1993.

MARCOS FILHO, J. **Fisiologia de sementes de plantas cultivadas**. Piracicaba: FEALQ, 2005. 495 p.

MIGON, H.; GAMERMAN, D. **Statistical inference: an integrated approach**. London: Arnold, 1999. 272 p.

MIGON, H.; SOUZA, A.; SCHIMIDT, A. **Modelos hierárquicos e aplicações**. Águas de São Pedro: ABE, 2008. 279p.

MONDO, V. H. V. et al. Avaliação de danos mecânicos em sementes de feijão por meio da análise de imagens. **Revista Brasileira de Sementes**, Brasília, v. 31, n. 2, p. 27-35, mar./abr. 2009.

NASSIF, D. S. P.; CICERO, S. M. Avaliação de sementes de acerola por meio de raios-x. **Revista Brasileira Fruticultura**, Cruz das Almas, v. 28, n. 3, p. 542-545, maio/jun. 2006.

NOGUEIRA, D.; SAFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência para o método de Monte Carlo via cadeias de Markov. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 65, n. 224, p. 59-88, jul./dez. 2004.

OLIVEIRA, L. M. **Avaliação da qualidade de sementes de *Tabebuia Serratifolia* Vahl Nich. e *T. impetiginosa* (Martius Ex A. P. de Candolle Standley) envelhecidas natural e artificialmente**. 2004. 160 p. Tese (Doutorado em Fitotecnia) - Universidade Federal de Lavras, Lavras, 2004.

OLIVEIRA, L. M.; CARVALHO, M. L. M.; DAVIDE, A. C. Utilização do teste de raios-x na avaliação da qualidade de sementes de canafístula (*Peltophorum dubium* (sprengel) Taubert.) **Revista Brasileira de Sementes**, Brasília, v. 25, n. 1, p. 116-120, jan./fev. 2003.

PAULINO, C.; TURKAMAN, A.; MURTEIRA, B. **Inferência bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446p.

PINTO, T. L. F. et al. An assessment of mechanical and stink bug damage in soybean seed using X-ray analysis test. **Seed Science and Technology**, Zürich, v. 37, n. 1, p. 110-120, Apr. 2009.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing, reference index**. Version 2.9.0. Vienna: R Foundation for Statistical Computing, 2009. Disponível em: <<http://www.r-project.org>>. Acesso em: 20 out. 2009.

RAFTERY, A.; LEWIS, S. How many iterations in the gibbs sampler? In: BERNARDO, J. et al. (Ed.). **Bayesian statistics**. Oxford: Oxford University, 1992. p. 763-774.

SANTOS, S. A. et al. X-ray technique application in evaluating the quality of papaya seeds. **Seed Science and Technology**, Zürich, v. 37, n. 3, p. 776-780, June 2009.

SIMAK, M. Tree and shrub seed handbook. In: GORDON, A. G.; GOSLING, P. G.; WANG, B. S. P. (Ed.). **Testing of forest tree and shrub seeds by X-radiography**. Zürich: ISTA, 1991. p. 1-28.

SIMAK, M.; BERGTEN, U.; HENRINKSSON, G. Evaluation of ungerminated seeds at the end of germination test by radiography. **Seed Science and Technology**, Zürich, v. 17, n. 2, p. 361-369, 1989.

SIMAK, M.; GUSTAFSSON, A. X-ray photography and sensitive in forest tree species. **Hereditas**, Lund, v. 39, p. 458-468, 1953.

SOCOLOWSKI, F.; CICERO, S. M. Caracterização morfológica de embriões por imagens de raios x e relação com a massa e a qualidade fisiológica de sementes de *Tecoma stans* L. Juss. ex Kunth (Bignoniaceae). **Revista Brasileira de Sementes**, Brasília, v. 30, n. 2, p. 200-208, mar./abr. 2008.

TANNER, M.; WRONG, W. The calculation of posterior distributions by data augmentation. **Journal of the American Statistical Association**, New York, v. 82, n. 398, p. 528-550, June 1987.

TITTERRINGTON, D.; SMITH, A.; MAKOV, U. **Statistical analysis of finite mixture distributions**. New York: J. Wiley, 1985. 243 p.

### **CAPÍTULO 3**

#### **Análise de dados obtidos em imagens radiográficas de sementes de mamona por meio de processamento automático de imagens**

##### **RESUMO**

Para implementar métodos de classificação dos potenciais de viabilidade e de germinação de lotes de sementes de mamona, a partir das medidas de área das sementes calculadas por meio do processamento automático de imagens, foram utilizadas técnicas de análise discriminante. Os dados foram obtidos de uma amostra de 400 sementes de mamona, da cultivar IAC-226, recém colhidas (safra 2009). Essas sementes foram identificadas individualmente e submetidas aos a análise radiográfica e a testes de germinação e vigor. As radiografias das sementes foram analisadas por duas técnicas de análise de imagem, a visual e a automática. Na análise visual, feita por três avaliadores diferentes, as sementes foram classificadas conforme seu potencial de viabilidade. Na análise automática foram calculadas medidas de área das sementes a partir de técnicas de processamento automático de imagens. Foram propostos dois métodos para classificação das sementes a partir das áreas obtidas por processamento automático, utilizando técnicas de análise discriminante. No primeiro método a classificação foi baseada na análise radiográfica e, no segundo, no potencial de germinação e vigor. Para os dois métodos propostos foi ajustado o modelo linear de Fisher e calculada a função e o escore discriminante. Para avaliar a qualidade do ajuste foi utilizado o método da validação cruzada. O método proposto para a classificação da análise radiográfica mostrou-se eficiente, com uma proporção de acertos de 88%. A proporção de classificações corretas para o método de classificação do potencial de germinação foi de 62%, resultado relevante para área de tecnologia de sementes. Os métodos propostos para a classificação podem ser uma alternativa viável para avaliar a qualidade de lotes de sementes de mamona.

Palavras-chave: Análise de imagens. Análise discriminante. Teste de raios-X.

## ABSTRACT

Aiming to implement methods for classifying the viability and germination potential of seed lots of castor, from the seed area measurements, we used discriminant analysis techniques. The seed area measurements to calibrate that analysis were calculated by means of automatic image proceedings. Data base were obtained from a sample of 400 seeds of castor bean cultivar IAC-226, freshly harvested (2009 harvest). These seeds were individually identified and submitted to the X-ray analysis and tests of germination and vigour. Radiographs of seeds were analyzed by both visual and automatic techniques of image analysis. In visual analysis, carried out by three different experts the seeds were classified according to their potential viability. The proceedings to calculate the area of the seeds were based on automatic analysis, using techniques for automatic processing of images. We proposed two methods for classification of seeds from the areas obtained by automatic processing, using techniques of discriminant analysis. In both methods, the classification was based on radiographic analysis and on germination and vigour, respectively, adjusting and calculating Fisher linear function and the discriminant score. The validation cross approach assessed the quality of function fit. The method developed to classify the outcomes from radiographic analysis was efficient, with a hit ratio of 88 %. On the other hand, the method for classifying the potential for germination had the proportion of correct classifications of 62 %, result relevant for the seeds technology area. The proposed methods for classification can be a valid and viable alternative while assessing the quality of batches of castor seeds.

Keywords: Image analysis. Discriminant analysis. X-ray test.

## 1 INTRODUÇÃO

A cultura de mamona vem se expandindo, em virtude de seu potencial de produção de óleo de excelente qualidade para ser usado tanto na indústria como para suprir a demanda de biodiesel. A qualidade e disponibilidade de sementes utilizadas para produção da planta é um dos grandes entraves para a expansão dessa cultura.

O teste mais utilizado, atualmente, para detectar diferenças entre lotes é o teste de germinação o qual é realizado sob condições ideais de emergência e detecta a baixa qualidade dos lotes em estágio avançado de deterioração, já que a perda da capacidade germinativa é a última consequência desse processo.

Além do teste de germinação, os testes de vigor são também utilizados no controle de qualidade por várias empresas produtoras de sementes com o intuito de detectar diferenças mais sutis na qualidade dos lotes de sementes e, conseqüentemente, garantir um nível mínimo de desempenho das sementes destinadas à comercialização.

Segundo Marcos Filho (2005), muitos testes consomem período de tempo relativamente amplo e produzem resultados mais variáveis que os desejados, tanto entre laboratórios como entre analistas. Isso ocorre, porque a interpretação de testes, baseados no desempenho de plântula,s mostra-se relativamente subjetiva, pois depende da experiência e sensibilidade do analista.

Mais recentemente, diversos métodos de avaliação da qualidade de sementes têm sido desenvolvidos, visando, principalmente, à maior precisão e rapidez de execução. O teste de raios-X, por exemplo, é recomendado pela Associação Internacional de Análise de Sementes (INTERNATIONAL SEED TESTING ASSOCIATION, 2004), com a finalidade de diferenciar sementes bem formadas das vazias, com danos mecânicos ou com ataque de insetos. É

um método preciso, simples e rápido, tendo a grande vantagem de ser não destrutivo, já que as sementes utilizadas durante o teste não são afetadas em sua qualidade pelas dosagens de raios-X. Posteriormente, essas sementes podem ser utilizadas para fins de semeadura, permitindo estabelecer relações entre os danos observados nas imagens radiográficas e o potencial físico e fisiológico.

Uma das limitações do teste de raios-x é a subjetividade na análise das imagens radiográficas, pois, assim como nos testes baseados em plântulas, a interpretação depende do analista. Essa subjetividade poderia ser minimizada com o uso de técnicas de análise automática digital de imagens.

As técnicas de análise automática de imagens vêm sendo muito utilizadas em várias áreas do conhecimento, como na medicina, para detectar diferenças sutis nas imagens, não perceptíveis aos olhos humanos. Essas técnicas têm sido associadas aos métodos estatísticos, como a análise discriminante, para classificar padrões.

A implementação de um método, que associe os resultados dos testes de germinação, de vigor e de raios-X com técnicas de processamento automático de imagens e técnicas de análise de discriminante, poderiam ser extremamente úteis para se inferir a respeito da qualidade dos lotes sementes, uma vez que estariam sendo envolvidas as características físicas e fisiológicas inerentes a esses lotes.

Neste capítulo objetivou-se implementar métodos de classificação do potencial de viabilidade e do potencial de germinação de lotes sementes de mamona, a partir das medidas de área das sementes calculadas por meio do processamento automático de imagens, utilizando técnicas de análise discriminante.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Testes de avaliação da qualidades do lote de sementes**

A estimativa do potencial fisiológico das sementes pode ser mais eficiente quando as informações sobre a qualidade são obtidas pela combinação de testes que avaliem diferentes características como a germinação e o vigor. Por esse motivo, dificilmente apenas um teste é, na maioria das vezes, suficiente para indicar, com certa precisão, o potencial de desempenho das sementes nas diferentes condições de campo.

#### **2.1.1 Teste de germinação**

Bewley (1997), Bewley e Black (1994), Come e Tissauoi (1973) e Egley (1999) enfatizam que germinação é o processo que compreende de eventos celulares e metabólicos que se iniciam com a absorção da água por sementes quiescentes e culminam com o alongamento do eixo embrionário. No entanto, em tecnologia de sementes, inclui-se a fase de crescimento da plântula normal que apresenta todas estruturas essenciais para a formação de uma planta produtiva (BRASIL, 2009).

Segundo Machado e Cicero (2003), o teste de germinação determina a porcentagem de sementes capazes de germinar normalmente, com base nas características morfológicas das plântulas. É conduzido sob condições ótimas e é padronizada para cada espécie avaliada.

A intensidade da resposta da semente ao ambiente é variável entre as espécies. Dessa forma, as condições de execução dos testes para análises de sementes são estabelecidas nas Regras para Análise de Sementes (BRASIL, 2009) as quais indicam o substrato, a temperatura, os limites de tempo, luz ou

outros tratamentos especiais para condução do teste.

Relatos sobre a metodologia, para condução do teste de germinação em sementes de mamona, são encontrados na literatura. Carneiro e Pires (1983) verificaram a influência de temperatura e substratos na capacidade germinativa de sementes de mamona. Santos et al. (2007) adequaram a metodologia do teste de germinação para avaliação da qualidade de sementes de mamona. Lima et al. (2007) caracterizaram a qualidade fisiológica de sementes de mamona da cultivar Guarani, provenientes do campo experimental da EMBRAPA, Clima Temperado, Pelotas, RS.

### **2.1.2 Teste de vigor**

O vigor é um dos aspectos mais importantes na análise de qualidade de sementes. Marcos Filho (1994) o descreve como o reflexo de um conjunto de características ou propriedades que determinam seu potencial fisiológico, isto é, o comportamento das sementes expostas a diferentes condições ambientais.

Os testes de vigor foram desenvolvidos com objetivo de avaliar diferenças de vigor entre lotes de sementes, as quais não são possíveis detectar com a utilização do teste de germinação (CHEROBINI, 2006). Nesses testes são avaliados aspectos como o desenvolvimento das plântulas, tolerância a condições adversas, uniformidade e velocidade de emergência e o aspecto físico do lote, que inclui densidade e morfologia interna e externa.

Normalmente, nos testes de germinação são realizadas duas contagens uma, no início, e outra, no final do teste. Para sementes de mamona a primeira contagem é realizada no sétimo dia e a segunda no décimo quarto (BRASIL, 2009). Alguns testes de vigor podem ser realizados conjuntamente com o teste de germinação, dentre eles está o da primeira contagem (KRZYZANOWSKY; VIEIRA R.D.; FRANÇA NETO, 1999).

Para Nakawaga (1999), lotes que apresentam maior porcentagem de plântulas normais, na primeira contagem, podem ser considerados mais vigorosos. E, para Association of Official Seed Analysts - AOSA (1983), sementes que originam plântulas com maiores valores de comprimento de parte aérea e pesos de matéria verde ou seca, num mesmo período de tempo, são consideradas mais vigorosas .

O crescimento das plântulas, na maioria da vezes, é visível. Contudo, são necessários valores numéricos para separar as mais vigorosas das menos vigorosas. A determinação do comprimento médio das plântulas normais ou de suas estruturas essenciais é realizado tendo em vista que as amostras que apresentam os maiores valores médios são as mais vigorosas (CHEROBINI, 2006).

Além dos testes tradicionais para avaliação da qualidade das sementes, outros testes têm sido desenvolvidos com o objetivo de proporcionar maior rapidez e precisão nos resultados, como os relacionados à análise de imagens.

## **2.2 Teste dos raios-X**

O uso da radiografia por meio dos raios-X de baixa energia é recomendado pela ISTA (2004) para determinar a qualidade das sementes. É considerado um método rápido e não destrutivo e tem como finalidade básica determinar sementes cheias, vazias, com danos mecânicos e com ataque de insetos. Segundo Bino, Aartase e Burg (1993), quando os raios-X passam por meio de uma semente, a radiação é absorvida em vários graus, dependendo da espessura, densidade e composição da semente e do comprimento de onda da radiação. Cria-se, assim, uma imagem permanente em um filme radiográfico.

Atualmente análise de imagens por meio de raios-X ainda vem sendo amplamente utilizada em vários segmentos da área de tecnologia de sementes.

Carvalho, Carvalho e Davide (2009) verificaram a ocorrência de alterações morfológicas internas por meio de raios-X e os efeitos dessas alterações na qualidade de sementes de seis espécies de *Lauraceae* encontradas no sul de Minas Gerais. Pinto et al. (2009) utilizaram técnica de análise de imagem por meio do teste de raios-X, para identificar danos mecânicos em sementes de soja de diferentes lotes da cultivar BRS184. Santos et al. (2009) investigaram a aplicação da técnica de raios-X, para avaliar a qualidade de sementes do mamão brasileiro (híbrido UENF/CALIMAN 01) e no estabelecimento de um protocolo experimental.

### **2.3 Análise automática de imagens**

O uso de processamento automático de imagens já tem sido bastante explorado em áreas como o do processamento de imagens de microscopia (WU; MERCHANT; CASTLEMAN, 2008), robótica e visão computacional (GONZALEZ; WOODS, 2000).

A morfologia matemática é uma linha de estudos em processamento de imagens que explora formas na imagem (DOUGHERTY; LOTUFO, 2003; SOILLE, 1999). As imagens de raios-X de sementes apresentam elementos com formas muito bem definidas, como pode ser observado na Figura 6 . Imagens com tais características são uma excelente aplicação para o processamento por morfologia matemática.

Um processo de análise automática de imagens pode ser dividido em quatro etapas: aquisição, filtragem ou pré-processamento, segmentação e análise. Esse processo varia conforme a aplicação, para mais detalhes, sugere-se a leitura de Dougherty e Lotufo (2003).

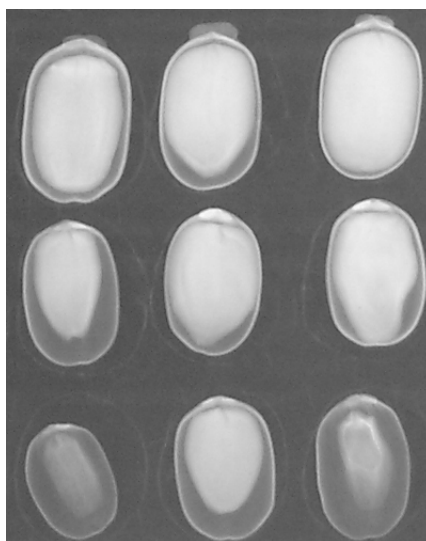


Figura 6 Parte de uma imagem de raios-X de sementes de mamona

#### 2.4 Análise discriminante

Um procedimento de classificação ou simplesmente classificador é algum método formal capaz de decidir, com base em informações fornecidas, a que grupo ou população um determinado objeto pertence. Geralmente, constrói-se uma regra de classificação, com base em uma amostra de dados e a utiliza para classificar novos objetos. Em estatística, o procedimento de análise discriminante está entre os mais utilizados e difundidos na literatura para classificação de padrões (Anderson (2003); Johnson e Wichern (2008); Lachenbruch e Goldstain (1975); Mardia, Kent e Bibby (1979)). As técnicas de análise discriminante vêm sendo utilizadas em diversas áreas. Por exemplo, na área de análise de imagens, Khoury Junior et al. (2005) avaliaram, por meio de técnicas de análise discriminante, a capacidade de classificar defeitos em tábuas de eucalipto, utilizando imagens digitais. Na área de tecnologia de sementes, Kataoka (2009) utilizou análise discriminante para avaliar o nível de qualidade de lotes de sementes de nabo forrageiro.

Segundo Mingotti (2005) para se aplicar a análise discriminante é necessário que os grupos, para os quais cada elemento amostral pode ser classificado sejam predefinidos, ou seja, conhecidos *a priori*, considerando suas características gerais. Esse conhecimento permite a elaboração de uma regra matemática, chamada de regra de discriminação, que é utilizada, para classificar novos elementos amostrais nos grupos já existentes. Mas é certo que todo o processo de tomada de decisões traz consigo um possível erro de decisão. Logo, deve-se construir uma regra de classificação que minimize o número de classificações incorretas, ou seja, o erro de afirmar que um elemento amostral pertence a uma população quando, na realidade, ele pertence a outra. Além disso, é importante construir uma regra que minimize o custo de uma classificação incorreta.

Considerando  $g$  populações ou grupos  $\Pi_1, \dots, \Pi_g$ ,  $g \geq 2$ . Supondo-se que associada com cada população  $\Pi_j$  há uma função densidade de probabilidade (fdp)  $f_j(x)$  em  $\mathbf{R}_p$ , tal que se um indivíduo pertence à população  $\Pi_j$ , então ele possui fdp  $f_j(x)$ . Com a análise discriminante objetiva-se alocar um indivíduo para uma dessas  $g$  populações com base em suas medidas  $\mathbf{x} = [x_1, x_2, \dots, x_p]$ , mensuradas no espaço  $p$ -dimensional. Deseja-se definir uma regra que permita dizer a que população  $\Pi_j$ ,  $j = 1, \dots, g$ , um determinado indivíduo é mais provável pertencer. Uma regra discriminante  $\lambda$  corresponde a uma divisão do  $\mathbf{R}_p$  em regiões disjuntas  $\mathbf{R}_1, \dots, \mathbf{R}_g$  ( $\bigcup \mathbf{R}_j = \mathbf{R}_p$ ). A regra  $\lambda$  é definida por alocar  $x$  para  $\Pi_j$  se  $x \in \mathbf{R}_j$ , para  $j = 1, \dots, g$ . A discriminação será mais precisa se  $\Pi_j$  tem sua probabilidade concentrada em  $\mathbf{R}_j$ , para cada  $j$ .

Para Casella e Berger (2002), quando os parâmetros são conhecidos é possível construir uma regra de classificação que minimize a chance de se classificar um elemento amostral incorretamente, pois será possível usar o princípio da máxima verossimilhança. Nesse caso, deve-se alocar uma observação

de  $\mathbf{x} = [x_1, x_2, \dots, x_p]$  para uma das populações  $\Pi_1, \dots, \Pi_g$ , isto é, deve-se alocar uma observação de  $\mathbf{x}$  para uma  $\Pi_j$  tal que

$$f_j(x) = \max f_i(x). \quad (1)$$

Na prática, na maioria das vezes os parâmetros populacionais não são conhecidos. Quando esses parâmetros não são conhecidos, mas a forma da distribuição da população é conhecida, é possível estimar os parâmetros, considerando as amostras de treinamento. Nessa situação, aplica-se a regra da máxima verossimilhança amostral, em que uma observação de  $x$  é alocada em uma das  $\Pi_j$  populações, de tal forma que,

$$\hat{f}_j(\mathbf{x}) = \max \hat{f}_i(\mathbf{x}). \quad (2)$$

Fisher (1938) propôs uma regra de discriminação quando as distribuições são completamente desconhecidas e as matrizes de variâncias e covariâncias são iguais, isto é,  $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$ . O método de Fisher consiste em encontrar uma função linear, que maximize a razão entre a soma dos quadrados entre grupos e a soma dos quadrados dentro dos grupos.

Uma vez que a função discriminante for calculada, uma observação pode ser alocada em uma das  $g$  populações com base em seu “escore discriminante” e, como na amostra, a população de origem de cada observação é conhecida, será possível identificar quais deles foram classificados corretamente ou não. Assim pode ser calculada a taxa estimada de erro, que é a fração das observações incorretamente classificadas, sendo uma estimativa da taxa de erro verdadeira.

Segundo Ferreira (2008), o desempenho de uma regra de classificação pode ser mensurado pelas taxas de erro ou probabilidades de classificações incorretas. Essas probabilidades de classificações incorretas devem ser esti-

madam a partir da regra de classificação obtidas das amostras de treinamento. Na literatura vários métodos foram propostos para estimar as probabilidades de classificações incorretas, dentre eles se destacam: o método da ressubstituição e método da validação cruzada (JOBSON,1998; RENCHER, 2002).

### 3 MATERIAL E MÉTODOS

Os dados foram obtidos a partir de uma amostra de 400 sementes de mamona da cultivar IAC-226, recém-colhidas (safra 2009) sem nenhum tipo de preparo. Essas sementes foram identificadas individualmente e submetidas aos testes de raios-X, germinação e vigor.

Para obtenção das imagens radiográficas, as sementes de mamona foram arranjadas em placas de isopor e identificadas uma a uma. Em cada placa, 50 sementes de mamona foram distribuídas na mesma posição, com a carúncula voltada para baixo, em 5 linhas e 10 colunas, igualmente espaçadas. As sementes foram encaixadas em alvéolos e fixadas com fitas adesivas transparentes, para que não saíssem da posição. A semente identificada como 1 foi posicionada na placa 1, na primeira linha e primeira coluna. A identificada como 2, foi a semente da placa 1, linha 1 e coluna 2 e a como 50 foi a posicionada na linha 5 e coluna 10 da mesma placa. Na placa 2 foram colocadas as sementes identificadas de 51 a 100 e assim sucessivamente, até completar o número necessário de sementes utilizadas no experimento.

Após a preparação, as placas foram colocadas sobre o filme MR 2000-1 Kodak e submetidas, durante um minuto, à radiação a uma distância de 35cm da fonte. Foi utilizado aparelho de raios-X modelo 43085-N da marca Faxitron-HP. Foi feita a análise visual e automática das imagens radiográficas das sementes.

Nessa mesma amostra de 400 sementes foram realizados os testes de germinação e vigor. Para execução desses testes, a amostra foi semeada em bandejas de poliestireno de alto impacto (PSAI), em células individualizadas, contendo substrato areia lavada e esterilizada. Foram utilizadas bandejas da marca Star Pack, com dimensões 545mm por 280mm, compostas por 32 célu-

las individualizadas, dispostas em 4 linhas e 8 colunas.

As bandejas contendo as sementes foram colocadas em uma câmara de crescimento, com a iluminação controlada e temperatura de  $25^{\circ}C$ , conforme recomendado por Brasil (2009). Cada célula foi avaliada e regada diariamente, durante 14 dias, no mesmo horário.

No 14<sup>o</sup> dia foram identificadas as células cujas sementes germinaram (originaram plântulas) e não germinaram. Nesse mesmo dia, as plântulas foram retiradas das células e foram medidos o diâmetro, a altura e a massa fresca da parte aérea dessas.

As plântulas foram levadas para estufa com circulação de ar, a uma temperatura  $80^{\circ}C$ , durante 24 horas, e após esse período, foi medida a sua massa seca dessas.

Esses dados foram utilizados para verificar a correlação entre as medidas de vigor e os dados obtidos pela análise das imagens radiográficas das sementes.

### **3.1 Análise visual de imagens**

As imagens radiográficas de cada uma das sementes foram analisadas por três avaliadores independentes e classificadas conforme a anatomia interna das mesmas. Para cada semente classificada, como potencialmente viável, foi associado o valor "0" e como potencialmente inviável, o valor "1".

A avaliação, como potencialmente viável ou inviável, foi baseada conforme critério estabelecido por Carvalho, Alves e Oliveira (2010), que classificaram as sementes de acordo com a morfologia interna visualizada na radiografia e submetidas aos testes de germinação, emergência e taxa de crescimento de plântula. Sementes classificadas como completas e opacas (Fi-

gura 7A) tiveram maior porcentagem de germinação, seguidas pelas classificadas como completa e manchada e parcialmente cheia e opaca, Figura 7B e Figura 7C, respectivamente. Sementes classificadas como parcialmente cheias e translúcidas (Figura 7D), parcialmente cheias e com deformação do embrião (Figura 7E), parcialmente cheia e manchadas (Figura 7F), apresentaram um índice de germinação intermediário, e aquelas classificadas como vazia tiveram a menor germinação (Figura 7G).

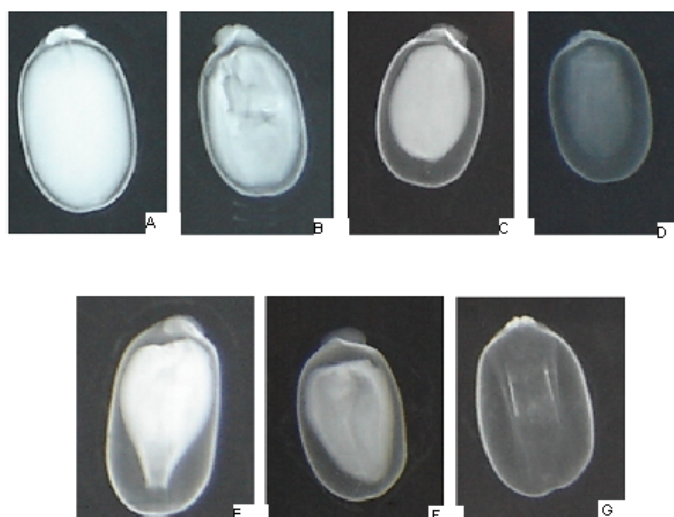


Figura 7 Radiografias de sementes de mamona classificadas em sete categorias conforme a sua morfologia interna: A - Cheia e opaca, B - Cheia e manchada, C - Parcialmente cheia e opaca, D - Parcialmente cheia e translúcida, E - Parcialmente cheia com defeito no embrião, F - Parcialmente cheia e manchada, G - Vazia. UFLA - Lavras, 2007

Assim, as sementes cujas imagens eram semelhantes às das Figuras(7A), (7B) ou (7C) foram consideradas como potencialmente viáveis, e das Figuras (7D), (7E), (7F) ou (7G) como potencialmente inviáveis.

### 3.2 Análise automática de imagens

Baseados na análise visual das radiografias das sementes, testes de raios-X, em que as principais medidas utilizadas para inferir sobre a qualidade do lote são, a inspeção visual das áreas cheia e vazia da semente, a má formação na proximidade do eixo embrionário da semente e a opacidade da área cheia da semente, foram obtidas as seguintes medidas de área por meio do processamento automático:

- a) AT: área da semente toda;
- b) AC: área da parte cheia;
- c) AV: área da parte vazia;
- d) RCAT: razão área da parte cheia sobre área da semente toda;
- e) RVT: razão área do espaço vazio sobre área da semente toda;
- f) AVI: área 1/3 da semente ao redor do eixo embrionário (área vital);
- g) AVIC: área da parte cheia da área vital;
- h) AVIV: área da parte vazia da área vital;
- i) RAVIC: razão da área da parte cheia sobre área vital;
- j) RAVIV: razão área espaço vazio da área vital sobre área vital;
- k) MEDNC: média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda;
- l) SDNC: desvio padrão amostral dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda.

O procedimento usado para a segmentação automática de imagens de raios-X das sementes de mamona foi feito pelo colaborador André Vital Saúde<sup>1</sup> e está descrito detalhadamente no ANEXO B. Foi utilizado o software Python (PYTHON SOFTWARE FOUNDATION, 2009) e conjunto de ferramentas, a toolbox de morfologia matemática SDC Morphology Toolbox (SDC MORPHOLOGY TOOLBOX FOR MATLAB, 2009).

As medidas dos valores absolutos das áreas foram dadas em números de pixels, havendo necessidade de transformações para a obtenção da unidade de medida para área. Essas transformações podem gerar distorção nas medidas, pois dependem da resolução em que as imagens radiográficas das sementes de mamona foram digitalizadas. Para evitar distorções nas medidas de área, para as análises estatísticas, foram utilizadas as razões entre as áreas das partes cheia e vazia e área total. Também foram usadas nessas análises estatísticas as médias dos níveis de cinza, pois essas representam a densidade dos tecidos da semente (BRASIL, 2009).

Para selecionar sementes de mamona, como potencialmente viáveis ou inviáveis, a partir das medidas de área calculadas por meio da análise automática das imagens radiográficas dessas, utilizou-se a técnica de análise de discriminante.

### 3.3 Análise discriminante

A classificação *a priori* dos grupos aos quais as medidas de área das sementes pertencem foi baseada nos resultados de dois testes aplicados na amostra de sementes sem preparo. Assim, a mesma amostra foi classificada por dois critérios, sendo o primeiro baseado na análise visual das radiografias das

---

<sup>1</sup>Professor Adjunto do Departamento de Ciências da Computação, Universidade Federal de Lavras (UFLA), saude@dcc.ufla.br

sementes de mamona e o segundo, no teste de germinação.

Para a classificação, baseada na análise visual das radiografias da amostra de sementes de mamona sem preparo, foi dividida em dois grupos, correspondentes a 1 e 2, conforme as análises das imagens radiográficas das sementes feitas por três especialistas independentes (3.1). Assim, as sementes que foram classificadas no grupo “1”, tiveram zero ou uma avaliação como potencialmente inviáveis dentre as três possíveis. As sementes classificadas no grupo “2”, tiveram duas ou três avaliações como potencialmente inviáveis dentre as três possíveis.

Para a classificação, baseada no teste de germinação, a amostra de 400 sementes de mamona sem preparo foi dividida em dois grupos, correspondentes a 1 e 2, sendo que as sementes classificadas no grupo “1” foram aquelas que não germinaram e as classificadas no grupo “2” foram as que germinaram.

Todos os procedimentos descritos a seguir, foram aplicados para classificações *a priori* baseadas na análise visual das radiografias e no teste de germinação.

A determinação das regras de classificação dependem das densidades de probabilidade da populações serem conhecidas ou não e da das variâncias serem homogêneas ou não. Assim, antes de determinar as regras de classificação, foi testada se amostras classificadas *a priori* tinham distribuição normal multivariada, utilizando-se a função *mshapiro.test* do pacote *mvnrmtest*, do software R (R DEVELOPMENT CORE TEAM, 2009). Essa função se baseia em buscar uma combinação linear das  $p$  variáveis originais e aplicar o teste de Shapiro-Wilk nessa nova variável.

Também, foi testado por meio do Teste de Box-M a homogeneidade das variâncias. Esse teste foi proposto por Box (1949) como uma modificação ao teste proposto por Barlett (1947). O interesse foi testar para  $g \geq 2$

populações a hipótese:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g \quad (3)$$

Considerando amostras aleatórias das  $g$  populações normais  $p$ -variadas de tamanho  $n_i$ ,  $N_p(\mu_i, \Sigma_i)$  e  $n = \sum_{i=1}^g n_i$ , a estatística do teste de Box's M dada por:

$$\chi^2 = \left[ \left( \sum_{i=1}^g \frac{1}{\nu_i} - \frac{1}{n-g} \right) \left( \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right) \right] \times \left[ \sum_{i=1}^g \ln|S_i| - (n-g)|S_c| \right] \quad (4)$$

possui distribuição assintótica qui-quadrado com  $f = (g-1)p(p+2)/2$  graus de liberdade sendo  $S_i$  e  $S_c$  estimadores não viesados das matrizes de covariâncias definidos por

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^I (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^\top \quad (5)$$

em que  $x_{ij} \in \mathbf{R}_p$  corresponde a  $j$ -ésima unidade da  $i$ -ésima população e

$$S_c = \frac{1}{n-1} (n_i - 1) S_i. \quad (6)$$

A rotina do teste de Box's M foi implementada no software R (R DEVELOPMENT CORE TEAM, 2009).

### 3.3.1 Regras de classificação

Considerando duas populações, suponha que se tenha uma amostra aleatória de  $n_1$  elementos da população 1 ( $\Pi_1$ ) e uma amostra aleatória de  $n_2$  elementos da população 2 ( $\Pi_2$ ), sendo as duas amostras independentes entre si. Nesse caso, os vetores de parâmetros das médias,  $\mu_1$  e  $\mu_2$  serão estimados pelos vetores de médias amostrais obtidos das amostras 1 e 2, respectivamente, isto é,

$$\bar{\mathbf{x}}_1 = \frac{\sum_{j=1}^{n_1} \mathbf{x}_{1j}}{n_1} \text{ e } \bar{\mathbf{x}}_2 = \frac{\sum_{j=1}^{n_2} \mathbf{x}_{2j}}{n_2}.$$

As matrizes de  $\Sigma_1$  e  $\Sigma_2$ , são estimadas pelas matrizes de covariâncias amostrais de cada grupo separadamente, ou seja,

$$\mathbf{S}_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)^\top \text{ e}$$

$$\mathbf{S}_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)^\top$$

Segundo Mingotti (2005), quando as matrizes de covariâncias são diferentes entre si, isto é,  $\Sigma_1 \neq \Sigma_2$ , mas a distribuição da população é conhecida, a função discriminante a ser utilizada é a função quadrática. Assumindo-se que as populações 1 e 2,  $\Pi_1$  e  $\Pi_2$ , são normalmente distribuídas e que parâmetros populacionais foram estimados a partir de amostras de treinamento, a função discriminante quadrática foi dada por

$$\hat{f}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_1)^\top |\mathbf{S}_1|^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) - (\mathbf{x} - \bar{\mathbf{x}}_2)^\top |\mathbf{S}_2|^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) + [\ln|\mathbf{S}_1| - \ln|\mathbf{S}_2|] \quad (7)$$

em que  $\bar{\mathbf{x}}_1$  é o vetor de médias estimadas da população  $\Pi_1$ ,  $\bar{\mathbf{x}}_2$  é o vetor de médias estimadas da população  $\Pi_2$  e  $\mathbf{S}_1$  é matriz de covariância amostral da população  $\Pi_1$  e  $\mathbf{S}_2$  é matriz de covariância amostral da população  $\Pi_2$ .

A regra de alocação foi baseada na razão de verossimilhança, isto é,

$$\lambda(\mathbf{x}) = \frac{f.d.p.\Pi_1}{f.d.p.\Pi_2}, \quad (8)$$

em que f.d.p é a função de densidade e probabilidade. Para  $\mathbf{S}_1 \neq \mathbf{S}_2$ , tem-se

$$-2\ln\lambda(\mathbf{x}) = -2\ln\lambda(\mathbf{x}) \left( \frac{\mathbf{S}_1}{\mathbf{S}_2} \right) + \left[ \left( \frac{\mathbf{x} - \bar{\mathbf{x}}_1}{\mathbf{S}_1} \right) - \left( \frac{\mathbf{x} - \bar{\mathbf{x}}_2}{\mathbf{S}_2} \right) \right] \quad (9)$$

Assim o escore discriminante foi dado por  $-2\ln\lambda(\mathbf{x})$  e um elemento amostral com o vetor de observações  $\mathbf{x}$  foi alocado em  $\Pi_1$  se  $-2\ln\lambda(\mathbf{x})$  foi

menor que zero, em  $\Pi_2$  se  $-2\ln\lambda(\mathbf{x})$  foi maior que zero e em qualquer uma das duas populações se  $-2\ln\lambda(\mathbf{x})$  foi igual a zero.

Fisher (1938) propôs uma regra de discriminação, quando as distribuições são completamente desconhecidas e as matrizes de variâncias e covariâncias são iguais, isto é,  $\Sigma_1 = \Sigma_2$ . Para o caso de duas populações em que os parâmetros foram estimados pelas de amostras de treinamento, a função discriminante de Fisher foi dada por

$$\hat{f}(\mathbf{x}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top |\mathbf{S}|^{-1} \mathbf{x} - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (10)$$

em que  $S$  é matriz de covariância amostral combinada ( $S_{pp}$ ), definida por

$$\mathbf{S}_{pp} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{(n_1 + n_2 - 2)}. \quad (11)$$

em que  $n_1$  é o número de elementos da amostra 1 e  $n_2$  é o número de elementos da amostra 2.

Para Ferreira (2008), quando os custos e as probabilidades *a priori* forem idênticos na duas populações,  $\mathbf{x}$  deve ser alocado na população  $\Pi_1$  caso

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top |\mathbf{S}|^{-1} \mathbf{x} \geq \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (12)$$

e na população  $\Pi_2$  caso contrário. Nesse caso,  $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top |\mathbf{S}|^{-1} \mathbf{x}$  é denominada função discriminante linear de Fisher e  $\frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$  é o escore discriminante.

### 3.3.2 Estimação das probabilidades incorretas

Na avaliação da função de classificação foram considerados dois tipos de erro. O erro 1 ocorreu quando o elemento amostral que pertencia a  $\Pi_1$  foi classificado pela regra como pertencendo a  $\Pi_2$ . O erro 2 ocorreu quando um elemento de  $\Pi_2$  foi classificado como sendo de  $\Pi_1$ . As probabilidades de ocor-

rências dos erros 1 e 2 foram denotadas por  $P(2|1)$  e  $P(1|2)$ , respectivamente. Os métodos utilizados na estimação das probabilidade de ocorrências de erros foram os da ressubstituição e da validação cruzada.

### Método da Ressubstituição

Nesse método, os escores de cada elemento amostral observado das populações  $\Pi_1$  e  $\Pi_2$  são calculados, sendo a regra de discriminação utilizada para classificar os  $n = n_1 + n_2$  elementos da amostra conjunta. Assim, os mesmos elementos participam da estimação da regra de classificação e da estimação dos erros de classificação (MINGOTTI, 2005). Na Tabela 10 é apresentada as frequências genéricas para as classificações corretas e incorretas.

Tabela 10 Distribuição do frequências genéricas para os erros de classificações

		População classificada pela regra		
		1	2	Total
População de origem	1	$n_{11}$	$n_{12}$	$n_1$
	2	$n_{21}$	$n_{22}$	$n_2$

em que  $n_{ij}$  é o número de elementos da população de origem  $i$  e classificadas em  $j$ . Quando  $i = j$  os elementos foram classificados corretamente e quando  $i \neq j$  os elementos foram classificados incorretamente. Assim, as probabilidades de ocorrência dos erros 1 e 2 são dados, respectivamente, por

$$\hat{P}(2|1) = \frac{n_{12}}{n_1} \quad \text{e} \quad \hat{P}(1|2) = \frac{n_{21}}{n_2}. \quad (13)$$

Segundo Johnson e Wichern (2008), esse método consistente mas viciado e tende a subestimar as verdadeiras probabilidades de ocorrências de erros. No entanto, pode servir como uma etapa inicial de avaliação, uma vez que se essas probabilidades foram muito elevadas, é sinal de que a regra de classificação deve ser reformulada. Para contornar esse problema, Lauchebruch e Mickey (1968) sugerem usar um procedimento de validação cruzada (*cross*

*validation*).

### **Método da validação cruzada**

A utilização desse método, para o caso de duas populações, consistiu nos seguintes passos:

- a) retirou-se um vetor de observações da amostra conjunta e utilizaram-se  $(n_1 + n_2 - 1)$  elementos amostrais restantes para construir a função de discriminação;
- b) utilizou-se a regra de discriminação construída em (3.3.1) para classificar o elemento que ficou de fora, verificando se a regra de discriminação classificou corretamente esse elemento;
- c) retornou-se o elemento amostral que foi retirado da amostra e retirou-se outro, repetindo-se os passos (a) e (b), até que todos os elementos  $n = (n_1 + n_2)$  da amostra conjunta fossem testados.

Testada toda a amostra conjunta, as frequências dos erros de classificação, foram distribuídas como na Tabela 10 e calculou-se as probabilidades de ocorrência de erros (13).

Para Mingotti (2005) as estimativas desse método são aproximadamente não viciadas e melhores que as do método da ressubstituição.

### **3.3.3 Estimação da probabilidade global de acerto**

Estimou-se a probabilidade global de acerto da função discriminante dada por:

$$\hat{p}(\text{acerto}) = \frac{n_{11} + n_{22}}{n_1 + n_2} \quad (14)$$

em que  $n_1$  é o número de observações do amostra 1,  $n_2$  é o número de observações do amostra 2,  $n_{11}$  é o número de observações classificadas corretamente no amostra 1 e  $n_{22}$  é o número de observações classificadas corretamente no amostra 2.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise descritiva dos dados

Na Tabela 11 está apresentada a distribuição do número de avaliações como potencialmente inviáveis segundo o grupo de sementes que germinou e não germinou.

Tabela 11 Distribuição do número de avaliações como potencialmente inviáveis segundo o grupo de sementes que germinou e não germinou

Grupo	X				Total
	0	1	2	3	
Não germinou	101	5	6	9	121
Germinou	269	5	2	3	279
Total	370	10	8	12	400

X: número de classificações como potencialmente inviáveis

Relacionando os resultados da análise visual com os resultados do teste de germinação, observou-se que, das 370 sementes que tiveram o número de classificações iguais a “0”, na análise visual, os três avaliadores analisaram a semente como potencialmente viável, mas 101 não germinaram. Essa taxa de erro de avaliação de 27,3% deve ser considerada, na utilização da imagem radiográfica para avaliação da qualidade das sementes, no entanto, vale a pena salientar que nesse tipo de análise são avaliadas apenas as características físicas da semente. Para uma semente gerar uma plântula normal, devem, também, ser levadas em consideração as características fisiológicas.

Observou-se, também, que considerando as sementes que tiveram o número de avaliações iguais a “0”, na análise visual e germinaram e as que tiveram o número de avaliações igual a “3”, na análise visual e não germinaram foram as classificadas corretamente. As proporções dos índices de classificação para as sementes que germinaram e não germinaram são 72% e 75%, respectivamente. Essas proporções de acerto comprovam a subjetividade da

análise visual, uma vez que essa foi baseada em análises de avaliadores diferentes.

A porcentagem de germinação das sementes sem preparo foi igual de 70%, valor inferior ao padrão nacional de comercialização de sementes de mamona que é de 85% (BRASIL, 2005)..

Estatísticas descritivas relativas às medidas do diâmetro, da massa seca e da altura da parte aérea da plântulas originadas são apresentados na Tabela 12.

Tabela 12 Estatísticas descritivas das medidas do diâmetro, altura e peso seco das plântulas originadas

	<b>Diâmetro (mm)</b>	<b>Altura (cm)</b>	<b>Massa Seca (g)</b>
Mínimo	2,540	4,000	0,036
1ºQuartil	3,240	8,700	0,118
Mediana	3,570	9,700	0,133
Média	3,564	9,796	0,131
3ºQuartil	3,900	10,800	0,148
Máximo	4,960	15,600	0,252
Desvio padrão	0,461	1,854	0,002

Foram calculadas, com base, no processamento digital de imagens, doze medidas de área cujas as estatísticas descritivas são apresentadas na Tabela 13. Para as análises estatísticas, foram utilizadas as razões entre as áreas cheia e vazia e área total, as áreas cheia e vazia da parte vital e área total da parte vital e a média dos níveis de cinza.

Para verificar a relação existente entre os dados obtidos pela análise automática de imagens e as medidas obtidas no teste de vigor, foi calculada a matriz de correlação desses dados (Tabela 14).

Os coeficientes de correlação entre o diâmetro, altura e peso seco das plântulas originadas e a razão entre a área cheia e área total foram, respectivamente, 0,135, 0,069 e 0,242, indicando uma baixa correlação entre essas

Tabela 13 Estatísticas descritivas das medidas de áreas das sementes de mamona obtidos por meio do processamento automático de imagens

	AT	AC	AV	RCAT	RVT	AVI
Mínimo	51263	38938	716	0.5248	0.01032	15769
1º Quartil	77094	58184	10640	0.7388	0.13003	21724
Mediana	82973	62698	12610	0.7569	0.15303	23042
Média	82681	62556	12816	0.7574	0.15387	23069
3º Quartil	88892	67555	14850	0.7804	0.17103	24547
Máximo	104355	86400	30326	0.9344	0.37832	28791
D.P.	8855,47	7217,55	4080,47	0,0047	0045	218,05
	AVIC	AVIV	RAVIC	RAVIV	MEDNC	SDNC
Mínimo	2674	0	0.1274	0.1400	137.1	378.8
1º Quartil	13960	3850	0.6235	0.1741	189.9	1128.2
Mediana	15220	5160	0.6584	0.2250	202.3	1322.7
Média	15334	5109	0.6654	0.2204	201.0	1294.9
3º Quartil	16786	6161	0.7090	0.2617	214.0	1485.3
Máximo	24986	15709	0.9437	0.7484	240.2	2276.8
D.P.	2229,52	1848,34	0,080	0,076	18,77	282,74

AT: área da semente toda; AC: área da parte cheia; AV: área da parte vazia; RCAT: razão área da parte cheia sobre área da semente toda; RVT: razão área do espaço vazio sobre área da semente toda; AVI: área 1/3 da semente ao redor do eixo embrionário (área vital); AVIC: área da parte cheia da área vital; AVIV: área da parte vazia da área vital; RAVIC: razão da área da parte cheia sobre área vital; RAVIV: razão área espaço vazio da área vital sobre área vital; MEDNC: média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda; SDNC: desvio padrão amostral dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda

Tabela 14 Matriz de covariâncias das medidas de área e das medidas obtidas no teste de vigor

	DIM	ALT	PS	RCAT	RVT	RAVIC	RAVIV
ALT	0,192						
PS	0,204	0,212					
RCAT	0,135	0,069	0,242				
RVT	-0,133	-0,061	-0,212	-0,990			
RAVIC	0,075	0,075	0,183	0,861	-0,839		
RAVIV	-0,071	-0,075	-0,165	-0,862	0,855	-0,994	
MEDNC:	0,046	-0,032	0,068	0,047	-0,038	0,013	-0,006

DIM: diâmetro; ALT: altura; PS: peso seco; RCAT: razão área da parte cheia sobre área da semente toda; RVT: razão área do espaço vazio sobre área da semente toda; RAVIC: razão da área da parte cheia sobre área vital; RAVIV: razão área espaço vazio da área vital sobre área vital; MEDNC: média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda

medidas. Segundo Krzyzanowsky, Vieira e França Neto (1999), a matéria seca de plântulas é uma medida de vigor do lote das sementes. Assim, esperava-se uma alta correlação entre a razão da área cheia e área total da semente com a massa seca de plântulas, pois a massa seca reflete a reserva da semente e essa é representada pela área cheia da imagem.

Tabela 15 Estatísticas descritivas para as medidas de área em relação ao grupo que germinou e não germinou

	<b>Medidas</b>	<b>Média</b>	<b>DP</b>	<b>Mínimo</b>	<b>Máximo</b>
Não Germinou	RCAT	0,7472	0,05387	0,5921	0,9343
	RVT	0,1635	0,05197	0,0103	0,3116
Germinou	RAVIC	0,6428	0,0946	0,1274	0,9437
	RAVIV	0,2422	0,0909	0,14100	0,7484
	MEDNC	172,9170	68,6405	2,2584	240,1933
	RCAT	0,7551	0,03896	0,7384	0,8538
Germinou	RVT:	0,1562	0,0372	0,1341	0,3030
	RAVIC	0,6634	0,0676	0,6249	0,8702
	RAVIV	0,2223	0,0635	0,1794	0,4105
	MEDNC	183,721	58,907	187,5426	232,9538

DP: desvio padrão; RCAT: razão área da parte cheia sobre área da semente toda; RVT: razão área do espaço vazio sobre área da semente toda; RAVIC: razão da área da parte cheia sobre área vital; RAVIV: razão área espaço vazio da área vital sobre área vital; MEDNC: média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda

Na Tabela 15 são apresentadas as medidas descritivas das razões das áreas das semente e a média dos níveis de cinza para os grupos de sementes que não germinaram e germinaram. Comparando as médias das razões das áreas dos grupos das sementes que germinaram e as que não germinaram, por meio do Teste-t (Tabela 16), verificou-se que não existem diferenças significativas entre essas, ao nível de 5% de probabilidade. Esses resultados mostram que a razão entre áreas pode não ser uma medida determinante na classificação dos grupos de sementes que germinaram e não germinaram.

Na comparação entre as médias dos níveis de cinza<sup>2</sup> com o grupo de

<sup>2</sup>a medida *médias de níveis de cinza* para facilitar a leitura será substituída por *níveis de cinza*

Tabela 16 Resultados do teste t-Student na comparação da média das razões entre áreas e níveis de cinza em relação à germinação

<b>Medidas</b>	<b>t</b>	<b>p-valor</b>
RCAT	-0,37	0,702
RVT	0,44	0,664
RAVIC	-0,62	0,533
RAVIV	0,71	0,4784
MEDNC	-12,61	0,0007

RCAT: razão área da parte cheia sobre área da semente toda; RVT: razão área do espaço vazio sobre área da semente toda; RAVIC: razão da área da parte cheia sobre área vital; RAVIV: razão área espaço vazio da área vital sobre área vital; MEDNC: média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda

sementes que germinaram e não germinaram, por meio do teste t-Student, constatou-se que houve diferenças entre essas, ao nível de 5% de probabilidade, indicando que a opacidade é uma medida importante para diferenciação entre as sementes que germinaram e não germinaram.

## 4.2 Análise discriminante

A mesma amostra de 400 sementes sem preparo foi classificada por dois diferentes critérios. No primeiro critério foram utilizados, os resultados da análise visual. No segundo, os resultados do teste de germinação. Para os dois critérios, os dados relativos à razão entre áreas e médias de níveis de cinzas foram classificados por meio da técnica de análise discriminante. Os resultados dessas análises são apresentados a seguir.

### 4.2.1 Análise discriminante para amostras classificadas por meio da análise visual

Para classificações *a priori* dos grupos, por meio da análise visual, foi utilizado o seguinte critério:

- a) **Grupo 1:** das sementes classificadas como potencialmente viáveis, isto é, tiveram 0 ou 1 classificações com potencialmente inviáveis dentre as três possíveis;
- b) **Grupo 2:** das sementes classificadas como potencialmente inviáveis, isto é, tiveram 2 ou 3 classificações com potencialmente inviáveis dentre as três possíveis.

Para verificar a hipótese igualdade das matrizes de covariâncias entre os grupos, foi aplicado o teste de Box's M, a 5% de significância, obtendo-se o valor-p= 0. Essa hipótese foi rejeitada, indicando o uso da função quadrática.

Foi aplicado o teste de Shapiro Wilk multivariado para cada um dos grupos, a 5% de significância, obtendo-se p-valor = 0 para ambos, indicando a utilização da regra linear de Fisher.

De acordo com Johnson e Wichern (2008), a regra quadrática é mais afetada pela falta de normalidade do que a regra linear de Fisher é afetada pela falta de igualdade das matrizes de covariâncias. Mingotti (2005) recomenda ajustar os modelos linear de Fisher e quadrático nos casos em que alguma pressuposição é violada, ficando com aquele que resultar em menores erros de classificação. Caso ambos deem resultados semelhantes, opta-se pelo modelo linear, uma vez que a matriz de covariâncias é estimada com um número maior de observações.

Assim, foram ajustados os dois modelos, o linear de Fisher e o quadrático, obtendo-se erros globais de classificação de 12% e 10,3%, respectivamente. Como esses erros foram semelhantes, conforme a recomendação de Mingotti (2005), optou-se pelo modelo linear de Fisher.

Para avaliar a função de classificação de Fisher, foi utilizado o método da ressubstituição e método da validação cruzada (Tabela 17). As probabi-

lidades de classificações incorretas, para o método da ressubstituição foram  $\hat{P}(2|1) = 0,107$  e  $\hat{P}(1|2) = 0,35$  e, para método da validação cruzada, foram  $\hat{P}(2|1) = 0,113$  e  $\hat{P}(1|2) = 0,35$ .

Pode-se observar que a probabilidade estimada de classificações incorretas de sementes potencialmente viáveis como inviáveis, foi nos dois métodos aproximadamente 11%, enquanto a probabilidade de classificações incorretas de sementes inviáveis como viáveis foi de 35%. Analisando esses resultados, pode-se afirmar que a regra de classificação é mais adequada para identificar as sementes potencialmente viáveis.

Tabela 17 Resultados da classificação pela regra de Fisher e para avaliação da classificação pelo método da validação cruzada para os grupos classificados *a priori* pelos critérios do teste de raios-X e do teste de germinação

<b>Método da Ressubstituição</b>				
<b>População Classificada</b>	<b>Análise Visual</b>		<b>Germinação</b>	
	<b>População Origem</b>		<b>População Origem</b>	
	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>
1	339	7	70	107
2	41	13	51	172
Total	380	20	121	279
Número de acertos	339	13	70	172
Proporção de acertos	0,892	0,65	0,58	0,62

<b>Método da validação cruzada</b>				
<b>População Classificada</b>	<b>Análise Visual</b>		<b>Germinação</b>	
	<b>População Origem</b>		<b>População Origem</b>	
	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>
1	337	7	66	109
2	43	13	55	170
Total	380	20	121	279
Número de acertos	337	13	66	170
Proporção de acertos	0,887	0,650	0,55	0,61

A função linear de Fisher estimada foi dada por:

$$\begin{aligned} \hat{f}(x) = & 316,73RCAT + 265,51RVT - 223,29RAVIC \\ & - 221,19RAVIV + 0,001MEDNC \end{aligned} \quad (15)$$

em que  $\hat{f}$  é a função discriminante de Fisher, RCAT é razão área da parte cheia sobre área da semente toda, RVT é a razão área do espaço vazio sobre área da semente toda, RAVIC é a razão da área da parte cheia sobre área vital, RAVIV é a razão área espaço vazio da área vital sobre área vital, MEDNC é a média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda.

A constante de delimitação da região de classificação foi dada por 81,86. Ou seja, uma semente cujas medidas fossem substituídas na função discriminante e o resultado fosse maior ou igual que 81,86, seria classificada no grupo das potencialmente viáveis e, se o resultado fosse menor que 81,86, seria classificada no grupo das potencialmente inviáveis.

A proporção geral de classificações corretas foi de 88%. Assim, de uma maneira geral, a regra de classificação pode ser considerada adequada, uma vez que os grupos foram classificados por meio da análise visual e as medidas utilizadas foram obtidas por meio da análise automática de imagens radiográficas, sendo essas medidas baseadas nas características físicas das sementes.

O método proposto nesta seção para a classificação da viabilidade dos lotes de sementes, utilizando análise automática de imagens, pode ser alternativa ao da análise visual. Uma vantagem desse método proposto é que ele seria menos subjetivo, uma vez que as classificações de outras sementes dependeriam apenas das medidas obtidas via análise automática de imagens. Mas, como ele é baseado somente em aspectos físicos da semente, deve ser combinado com testes que avaliem diferentes características como a germinação e o vigor.

#### 4.2.2 Análise discriminante para amostras classificadas por meio do teste de germinação

Para classificações *a priori* dos grupos por meio do teste de germinação foi utilizado o seguinte critério:

- a) **Grupo 1:** das sementes que não germinaram;
- b) **Grupo 2:** das sementes que germinaram, isto é, que originaram plântulas normais.

Verificou-se por meio do teste de Box's M, a 5% de significância, que as matrizes de covariâncias não eram homogêneas ( $p = 0$ ). Foi aplicado o teste de Shapiro Wilk multivariado para as amostras dos grupos 1 e 2, a 5% de significância, obtendo-se p-valor 0,00003 e 0,00063, respectivamente, indicando a utilização da regra linear de Fisher.

Pelos mesmos motivos justificados em (4.2.1), foram ajustados os modelos linear de Fisher e quadrático, obtendo-se erros de classificação de 39% e 36%, respectivamente, optando-se pelo modelo linear de Fisher.

Na avaliação da função de classificação foram utilizados os métodos da ressubstituição e da validação cruzada (Tabela 17). As probabilidades de classificações incorretas para o método da ressubstituição foram  $\hat{P}(2|1) = 0,32$  e  $\hat{P}(1|2) = 0,28$  e para método da validação cruzada foram  $\hat{P}(2|1) = 0,35$  e  $\hat{P}(1|2) = 0,29$ . Observou-se que a probabilidade estimada de classificações incorretas para os dois grupos foi aproximadamente 30%, indicando que a regra de classificação não é adequada para identificar as sementes que germinaram e não germinaram.

A função linear de Fisher estimada foi dada por:

$$\begin{aligned} \hat{f}(x) = & -96,61RCAT - 105,12RVT + 67,92RAVIC \\ & + 77,1RAVIV - 0,003MEDNC \end{aligned} \quad (16)$$

em que  $\hat{f}$  é a função discriminante de Fisher, RCAT é razão área da parte cheia sobre área da semente toda, RVT é a razão área do espaço vazio sobre área da semente toda, RAVIC é a razão da área da parte cheia sobre área vital, RAVIV é a razão área espaço vazio da área vital sobre área vital, MEDNC é a média dos níveis de cinza dos pontos pertencentes à parte cheia da semente toda.

A constante de delimitação da região de classificação foi dada por -27,28044. Assim, uma semente cujo score for maior ou igual que -27,28044 seria classificada no grupo das sementes que não germinaram e com menor seria classificada no grupo das que germinaram.

O proporção geral de classificações corretas foi de 61%. Como esse método foi baseado em características físicas (áreas) e fisiológicas (germinação), uma das hipóteses que justifica esse baixo índice de classificações corretas é, que, além das características físicas, outras características tais como a presença de dormência ou mesmo de microorganismos nas sementes, afetam o processo de germinação. Uma outra hipótese é que quatro das cinco variáveis preditoras foram baseadas nas medidas da razão entre áreas e essas ficaram muito confusas, quando separadas nos grupos das sementes que germinaram e não germinaram, como é mostrado na Tabela 15 e discutido na seção (4.1). Assim, para aumentar o poder discriminação, seria necessário acrescentar outras variáveis, além das cinco observadas, ou substituí-las por outras.

Comparando os índices de classificações corretas para as sementes que germinaram e não germinaram, considerando os resultados do método proposto nessa seção e com os resultados da análise visual e o teste germinação (seção 4.1), observou-se:

- a) para as sementes que foram classificadas no grupo das que germinaram, a proporção de classificação correta, utilizando a análise discriminante foi de 62% e, na análise do visual, foi de 72%;
- b) para as sementes que, foram classificadas no grupo das que não germinaram, o índice de classificação correta, utilizando a análise discriminante foi de 58% enquanto na análise visual foi de 75%.

Assim, apesar da subjetividade, as classificações dos avaliadores, baseadas na análise radiográfica foram mais eficientes do que as baseadas na análise automática de imagens, levando a indícios de que a pequena variação entre medidas de área, obtida por meio da análise automática de imagens podem ter afetado os resultados.

## 5 CONCLUSÃO

O método proposto para a classificação do potencial de viabilidade dos lotes de sementes, utilizando análise automática de imagens e técnicas de análise discriminante, mostrou-se eficiente, com um índice de classificações corretas de 88%. Esse método pode ser uma alternativa viável à análise visual de imagens, com a vantagem de ser menos subjetivo.

O índice de classificações corretas para o método de classificação do potencial de germinação, utilizando análise automática de imagens e técnicas de análise discriminante, foi de 62%. Esse resultado foi relevante, para área de tecnologia de sementes, uma vez foram associadas características físicas e filológicas das sementes em um único método.

Para implementação de um método mais eficiente para classificações, baseadas na associação de características físicas (medidas baseadas na análise automática) e de características fisiológicas (testes de germinação e vigor), seriam necessárias mais variáveis que discriminassem melhor esses grupos.

**REFERÊNCIAS**

ANDERSON, T. W. **Applied multivariate statistical analysis**. 3rd ed. New Jersey: J. Wiley, 2003. 767 p.

ASSOCIATION OF OFFICIAL SEEDS ANALYSTS. **Seed vigor testing handbook**. Lincoln, 1983. 88p. (Contribution, 32).

BARLLET, M. S. Multivariate analysis. **Journal of the Royal Statistical Society - Series B**, London, v. 9, n. 2, p. 176-197, 1947.

BEWLEY, J. Seed germination and dormancy. **The Plant Cell**, Rockville, v. 9, n. 7, p. 1055-1066, July 1997.

BEWLEY, J.; BLACK, M. **Seeds: physiology of development and germination**. 2nd ed. New York: Plenum, 1994. 445p.

BINO, R.; AARTASE, J.; BURG, W. van der. Non-destructive x-ray analysis of *Arabidopsis* embryo mutants. **Seed Science Research**, Wallingford, v. 3, n. 3, p. 167-170, Sept. 1993.

BOX, G. E. P. A general distribution theory for a class of likelihood criteria germination. **Biometrika**, London, v. 6, p. 317-346, 1949.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Padrões para produção e comercialização de sementes de mamona**. Brasília, 2005. Disponível em: <<http://extranet.agricultura.gov.br/sislegis-consulta/consultarLegislacao.do>>. Acesso em: 5 jan. 2010.

\_\_\_\_\_. **Regra para análise de sementes**. Brasília, 2009. 399p.

CARNEIRO, J.; PIRES, J. Influência da temperatura e do substrato na germinação de sementes de mamona. **Revista Brasileira de Sementes**, Brasília, v. 5, n. 3, p. 127-131, 1983.

CARVALHO, L. R.; CARVALHO, M. L. M.; DAVIDE, A. C. Utilização do teste de raios X na avaliação da qualidade de sementes de espécies florestais de Lauraceae. **Revista Brasileira de Sementes**, Brasília, v. 31, n. 4, p. 57-66, jul./ago. 2009.

- CARVALHO, M. L. M.; ALVES, R. A.; OLIVEIRA, L. M. Radiographic analysis in castor bean seeds (*Ricinus communis* L.). **Revista Brasileira de Sementes**, Brasília, v. 32, n. 1, p. 170-175, jan. 2010.
- CASELLA, G.; BERGER, R. **Statistical inference**. Belmont: Wadsworth, 2002. 660 p.
- CHEROBINI, E. **Avaliação da qualidade de sementes e mudas de espécies florestais**. 2006. 103 p. Dissertação (Mestrado em Engenharia Florestal) - Universidade Federal de Santa Maria, Santa Maria, 2006.
- COME, D.; TISSAUOI, T. Interrelated effects of imbibition, temperature and oxygen on the seed germination. In: HEYDECKER, W. (Ed.). **Seed ecology**. London: Butterworth, 1973. p. 157-158.
- DOUGHERTY, E.; LOTUFO, R. **Hands-on morphological image processing**. Bellingham: International Society for Optical Engine, 2003. 272 p. (SPIE Tutorial Texts in Optical Engineering, TT59).
- EGLEY, G. Reflections in my career in weed seed germination research. **Seed Science Research**, Wallingford, v. 9, n. 1, p. 3-12, Feb. 1999.
- FERREIRA, D. F. **Estatística multivariada**. Lavras: Ed. da UFLA, 2008. 662p.
- FISHER, R. The statistical utilization of multiple measurements. **Annals of Eugenics**, Cambridge, v. 8, p. 376-386, 1938.
- GONZALEZ, R.; WOODS, R. **Processamento digital de imagens**. São Paulo: E. Blücher, 2000. 509 p.
- INTERNATIONAL SEED TESTING ASSOCIATION. **International rules for Seed Testing Association**. Zürich, 2004. 174 p.
- JOBSON, J. D. **Applied multivariate data analysis**. New York: Springer-Verlag, 1998. 731 p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6th ed. New Jersey: Pearson, 2008. 767 p.
- KATAOKA, V. **Testes estatísticos na validação de métodos de ensaios de proficiência interlaboratoriais em sementes de nabo forrageiro**. 2009. 244 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2009.

KHOURY JUNIOR, J. et al. Análise discriminante paramétrica para reconhecimento de tábuas de eucalipto utilizando imagens digitais. **Revista Árvore**, Viçosa, MG, v. 29, n. 2, p. 299-309, mar./abr. 2005.

KRZYZANOWSKY, F.; VIEIRA R.D., FRANÇA NETO, J. **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, 1999. 109 p.

LACHENBRUCH, P. A.; GOLDSTAIN, M. Discriminant analysis. **Biometrics**, Arlington, v. 35, n. 1, p. 69-85, Mar. 1979.

LAUCHEBRUCH, P.; MICKEY, M. Estimation of error rates in discriminant analysis. **Technometrics**, Washington, v. 27, n. 2, p. 189-198, Mar. 1968.

LIMA, M. et al. Caracterização da qualidade fisiológica de sementes de mamona cultivar guarani. **Revista Brasileira de Biociências**, Porto Alegre, v. 5, n. 2, p. 675-677, jul. 2007.

MACHADO, C. F.; CICERO, S. M. Aroeira-branca [*Lithraea molleoides* (Vell.) Engl. - Anacardiaceae] seed quality evaluation by the X-ray test. **Scientia Agricola**, Piracicaba, v. 60, n. 2, p. 393-397, 2003.

MARCOS FILHO, J. **Fisiologia de sementes de plantas cultivadas**. Piracicaba: FEALQ, 2005. 495 p.

\_\_\_\_\_. Teste de envelhecimento acelerado. In: VIEIRA, R.; CARVALHO, N. (Ed.). **Teste de vigor em sementes**. Jaboticabal: FUNEP, 1994. p. 133-149.

MARDIA K. V.; KENT, J. T.; BIBBY, J. **Multivariate analysis**. London: Academic, 1979. 502 p.

MINGOTTI, S. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Ed. da UFMG, 2005. 297p.

NAKAWAGA, J. Testes de vigor baseados na avaliação de plântulas. In: VIEIRA, R.; CARVALHO, N. (Ed.). **Testes de vigor em sementes**. Jaboticabal: FUNEP, 1994. p. 49-85.

PINTO, T. L. F. et al. An assessment of mechanical and stink bug damage in soybean seed using X-ray analysis test. **Seed Science and Technology**, Zürich, v. 37, n. 1, p. 110-120, Apr. 2009.

PYTHON SOFTWARE FOUNDATION. **Python programming language**. Disponível em: <<http://docs.python.org/3.1/>>. Acesso em: 20 out 2009.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing, reference index**. Version 2.9.0. Vienna: R Foundation for Statistical Computing, 2009. Disponível em: <<http://www.r-project.org>>. Acesso em: 20 out. 2009.

RENCHE, A. C. **Methods of multivariate analysis**. New York: J. Wiley, 2002. 708 p.

SANTOS, D. et al. Teste de germinação em sementes de mamona. In: CONGRESSO BRASILEIRO DE MAMONA, 2., 2007, Aracaju. **Anais...** Aracaju: UFSE, 2006. 1 CD-ROM.

SANTOS, S.A. et al. X-ray technique application in evaluating the quality of papaya seeds. **Seed Science and Technology**, Zürich, v. 37, n. 3, p. 776-780, June 2009.

SDC MORPHOLOGY TOOLBOX FOR MATLAB. **Morphological image processing for MATLAB**. Version 1.6. Naperville: SDC Information Systems, 2008. Disponível em: <<http://www.mmorph.com/>>. Acesso em: 20 out 2009.

SOILLE, P. **Morphological image analysis**. New York: Springer-Verlag, 1999. 391 p.

WU, Q.; MERCHANT, F.; CASTLEMAN, K. **Microscope image processing**. São Paulo: Elsevier Academic, 2008. 548 p.

## CONSIDERAÇÕES FINAIS

Os métodos bayesianos se mostraram eficientes no estudo da viabilidade de sementes. Incorporar a opinião de três avaliadores pode ser bastante útil para diminuir a subjetividade desse teste.

Os métodos propostos para classificar as sementes a partir da análise de imagens, utilizando técnicas de análise discriminante, são uma alternativa viável para avaliação da qualidade de lotes de mamona.

Uma proposta futura de estudo é associar as medidas de áreas obtidas por meio de análise de imagens a outros testes utilizados na área de tecnologia de sementes, para que se possa obter um teste rápido na avaliação da qualidade dos lotes de sementes de mamona, bem como para outras espécies.

## ANEXOS

**ANEXO A – Rotina no software R para gerar as cadeias MCMC via  
Amostrador de Gibbs**

```

#Programa 400 sementes sem preparo x=read.table("Normal.txt",h=T)
#Funções para o amostrador de Gibbs
ProbSus=function(lambda1,p1,p2,x){
n=nrow(x)
x=as.vector(x)
h1=x
for (i in 1:n)
{xi=x[i,1] a=dbinom(xi, 3, p1) b=dbinom(xi,3,p2) h1[i,]=(lambda1*a)/(lambda1*a
+(1-lambda1)*b)}
return(h1)}
#Bernouli
require(Rlab)
Bern=function(h1,x){
m1=x
m2=x
n=nrow(x)
for (i in 1:n){
m1[i,]=rbern(1, h1[i,])
m2[i,]=1-m1[i,]}
return(m1,m2)}
#Função Amostrador de Gibbs# Gibbs=function(Lambda1,P1,P2){
Lambda1[1]=lambda1
P1[1]=p1
P2[1]=p2
for (j in 2:int){
h=ProbSus(lambda1,p1,p2,x)
m=Bern(h,x)
A=m$m1
A=sum(A)
A=1+A
B=m$m2
B=sum(B)
B=1+B
lambda1=rbeta(1,A,B)
}
}

```

```

C=(m$m1)*x
C=sum(C)
C=C+1
D=1+3*(A-1)-(C-1)
p1=rbeta(1,C,D)
E=(m$m2)*x
E=sum(E)
E=E+1
F=1+3*(B-1)-(E-1)
p2=rbeta(1,E,F)
Lambda1[j]=lambda1
P1[j]=p1
P2[j]=p2}
return(Lambda1,P1,P2)}
#Função gera caedia Cadeia=function(AG){
for (k in 1:INT){
j=JUMP*(k-1)+BUR
CLambda1[k]=AG$Lambda1[j]
CP1[k]=AG$P1[j]
CP2[k]=AG$P2[j]}
return(CLambda1,CP1,CP2)}
#Valores iniciais (calculados de Sistema no Maple)
lambda1=0.8454497922
p1=0.0517556854
p2=0.6496900703
#Inicializando a cadeia reduzida INT=1000 CLambda1=rep(0,INT)
CP1=rep(0,INT)
CP2=rep(0,INT)
BUR=1000 #Burn-in #
JUMP=19 # SALTO
ns= 1000
#Matriz de armazenamento de dados
MCLambda1=matrix(0:0, nrow=INT, ncol=ns)
MCP1=matrix(0:0, nrow=INT, ncol=ns)
MCP2=matrix(0:0, nrow=INT, ncol=ns)
COND=matrix(0:0, nrow=4, ncol=ns)
SDCOND=matrix(0:0, nrow=4, ncol=ns)
#NS SIMULAÇÕES DE GIBBS# int=20000 # Número de interações da ca-
deia completa
Lambda1=rep(0,int)#Inicializando o vetor do parametro Lambda1
P1=rep(0,int)# do parametro P1
P2=rep(0,int) # do parametro P2
#NÚMERO DE SIMULAÇÕES#

```

```

for (v in 1:ns){
AG=Gibbs (Lambda1,P1,P2) #Cadeia completa
Cad= Cadeia(AG) #Cadeia Reduzida
MCLambda1[,v]=Cad$CLambda1
MCP1[,v]=Cad$CPI
MCP2[,v]=CadM$CP2
}
# Gerando as Condicionais Estimada P(N|X)
H1=rep(0,INT)
A=rep(0,INT)
B=rep(0,INT)
S=rep(0,4)
S2=rep(0,4)
C=rep(0,4)
for (w in 1:ns)
{ for (k in 1:4){
X=k-1;
S[k]=0
for (z in 1:INT) {
A[z]=MCP1[z,w] $\hat{X}^*(1-MCP1[z,w])^{(3-X)}$ 
B[z]=MCP2[z,w] $\hat{X}^*(1-CP2[z])^{(3-X)}$ 
H1[z]=(MCLambda1[z,w]*A[z])/(MCLambda1[z,w]*A[z] +(1-MCLambda1[z,w])*B[z])
S[k]=S[k]+H1[z] }
COND[k,w]=S[k]/INT #Media da condicional
}}
# CONDICIONAIS X=x de todas as simulações}
#Termino da geração de 1000 amostras
#Sumario das Condicionais das 1000 amostras
X0=X1=X2=X3=rep(0,ns)
X0=COND[1,]
X1=COND[2,]
X2=COND[3,]
X3=COND[4,]
#Fim

```

## ANEXO B – Medição de características pelo processamento digital de imagens

André Vital Saúde <sup>1</sup>

Nesse anexo descreveu-se detalhadamente o procedimento computacional usado para a segmentação automática de imagens de raios-X de sementes de mamona. Para cada semente de mamona são detectados três objetos: a semente inteira, parte cheia e a cavidade vazia da semente.

A partir desses três objetos, é fácil obter as medidas de área da parte, da cavidade vazia e da semente inteira, utilizadas nos experimentos. Estes objetos são também usados como máscaras para se obter média e desvio padrão dos níveis de cinza no parte cheia ou na cavidade.

### Noções Básicas

Todo o processamento das imagens de sementes foi realizado com o uso de ferramentas da morfologia matemática. Foi introduzido alguns conceitos de morfologia matemática, sendo recomendado que se consulte Dougherty e Lotufo (2003) para uma visão mais aprofundada.

A operação fundamental da morfologia matemática é a erosão. A *erosão* do conjunto  $A$  pelo conjunto  $B$  é denotada por  $A \ominus B$  e é definida por

$$A \ominus B = \{x : B_x \subset A\},$$

onde  $B_x$  é a translação do conjunto  $B$  até o ponto  $x$ .

---

<sup>1</sup>Professor Adjunto do Departamento de Ciências da Computação, Universidade Federal de Lavras (UFLA), saude@dcc.ufla.br

A operação dual da erosão é a dilatação. A *dilatação* do conjunto  $A$  por  $B$  é denotada por  $A \oplus B$  e definida por

$$A \oplus B = (A^c \ominus \widetilde{B})^c,$$

onde  $A^c$  é o complementar do conjunto  $A$ , e  $\widetilde{B}$  é a reflexão do conjunto de pontos  $B$  em torno da origem do sistema de coordenadas.

A *abertura* da imagem  $A$  pela imagem  $B$  é denotada por  $A \circ B$  e é definida como uma composição de erosão e dilatação por

$$A \circ B = (A \ominus B) \oplus B.$$

A operação dual da abertura é o fechamento, que é definido como uma dilatação seguida de uma erosão. O *fechamento* de  $A$  por  $B$  é denotado por  $A \bullet B$  e definido por

$$A \bullet B = (A \oplus B) \ominus B.$$

Chama-se de reconstrução morfológica de um objeto  $A$  a partir de um objeto  $B$  e um elemento estruturante  $E$ , a aplicação da seguinte operação infinitas vezes

$$B = \min\{A, B \oplus E\}$$

onde  $B$  é o resultado da reconstrução. Na prática não é necessário aplicar a operação infinitas vezes. O processo converge em um  $B$  que não mais se modifica, e pode ser interrompido.

Além do fechamento morfológico, foi utilizado neste trabalho o fechamento de buracos. Um buraco é um conjunto de pontos de fundo, totalmente

cercado por pontos do objeto. O *fechamento de buracos* transforma todos os buracos em objeto.

Finalmente, além dos operadores morfológicos, este trabalho utiliza também a conhecida técnica de Otsu para se fazer *threshold* automático (GONZALEZ; WOODS, 2000). A técnica detecta automaticamente o melhor valor  $t$  para se segmentar automaticamente uma imagem em objeto e fundo pela simples aplicação de um limiar (*threshold*)  $t$  à imagem.

### Segmentação da imagem

Cada imagem de raio-X apresentava 50 sementes bem organizadas espacialmente. Assim foi feita uma divisão automática da imagem em 50 imagens contendo apenas uma semente cada uma.

A Figura 8 ilustra os resultados parciais do processo, sendo que a Figura 8(a) é a imagem original. Para cada imagem, foram realizados os passos descritos a seguir.

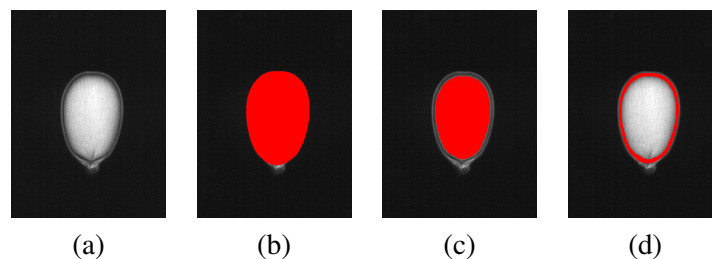


Figura 8 Segmentação da semente. Em a) a imagem original; em b) a semente inteira é destacada; em c) o miolo é destacado; e em d) a concavidade vazia é destacada.

**Passo 1:** Partindo da imagem ilustrada na Figura 8(a), a semente inteira é segmentada. São realizados na seqüência:

- a) um *threshold* em 40, para gerar uma primeira segmentação;

- b) uma abertura por uma cruz, para remover pontos isolados da imagem;
- c) um fechamento de buracos, para que o objeto obtido seja único e sem buracos;
- d) uma abertura por um disco de raio 75. A abertura por um grande disco remove irregularidades nas bordas e garante que a extremidade inferior da semente também seja removida. O resultado é exibido na Figura 8(b).

**Passo 2:** Novamente partindo da Figura 8(a), é segmentado da parte cheia da semente. Para isto, são realizados na seqüência:

- a) uma abertura seguida de um fechamento por um disco de raio 5;
- b) uma abertura por um disco de raio 20;
- c) sobre o resultado da abertura, é feito um *threshold* em 150;
- d) uma erosão por um disco de raio 20 resulta em um objeto que se encontra seguramente interno da parte cheia;
- e) é feita uma reconstrução morfológica da imagem original a partir do objeto obtido no passo anterior, resultando em uma imagem de níveis de cinza que contém apenas os pixels da parte cheia da semente, sendo pretos os pontos restantes;
- f) da parte cheia passa por um *threshold* automático pelo método de Otsu, e é seguido de uma abertura por um disco de raio 32 (divisão inteira  $75/2$ ). O resultado da segmentação da parte cheia é ilustrado na Figura 8(c).

**Passo 3:** A cavidade vazia da semente é obtido a partir das imagens das Figuras 8(b) e 8(c). São realizados os seguintes passos na seqüência:

- a) uma erosão da Figura 8(b) por um disco de raio 7, removendo o envólucro mais externo da semente;

- d) da imagem obtida no passo anterior, é subtraída a imagem da Figura 8(b), que é da parte cheia. O resultado da subtração é a cavidade vazia, ilustrada na Figura 8(d).

Finalmente, as medidas de área, médias e desvios podem ser realizadas.

**REFERÊNCIAS**

DOUGHERTY, E.; LOTUFO, R. **Hands-on morphological image processing**. Bellingham: International Society for Optical Engine, 2003. 272 p (SPIE Tutorial Texts in Optical Engineering, TT59).

GONZALEZ, R.; WOODS, R. **Processamento digital de imagens**. São Paulo: E. Blücher, 2000. 509 p.