



**MARIA LUIZA CAPELLARI LEITE DA SILVA**

**AMOSTRAGEM SEQUENCIAL BAYESIANA PARA  
ESTIMAÇÃO DO VIGOR E DANOS DAS SEMENTES DE  
SOJA ATRAVÉS DO TESTE DE TETRAZÓLIO**

**LAVRAS – MG**

**2020**

**MARIA LUIZA CAPELLARI LEITE DA SILVA**

**AMOSTRAGEM SEQUENCIAL BAYESIANA PARA  
ESTIMAÇÃO DO VIGOR E DANOS DAS SEMENTES DE  
SOJA ATRAVÉS DO TESTE DE TETRAZÓLIO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para obtenção de título de Mestre.

Prof<sup>a</sup>. Dr<sup>a</sup>. Carla Regina Guimarães Brighenti  
Orientadora

**LAVRAS - MG  
2020**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Silva, Maria Luiza Capellari Leite da.

Amostragem sequencial bayesiana para estimação do vigor e danos das sementes de soja através do teste de tetrazólio / Maria Luiza Capellari Leite da Silva. - 2020.

88 p. : il.

Orientador(a): Carla Regina Guimarães Brighenti.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2020.

Bibliografia.

1. Estimação sequencial bayesiana. 2. distribuição multinomial.  
3. distribuição Dirichlet.

**MARIA LUIZA CAPELLARI LEITE DA SILVA**

**AMOSTRAGEM SEQUENCIAL BAYESIANA PARA ESTIMAÇÃO DO VIGOR  
E DANOS DAS SEMENTES DE SOJA ATRAVÉS DO TESTE DE TETRAZÓLIO**

**BAYESIAN SEQUENTIAL SAMPLING FOR VIGOR AND DAMAGE OF  
SOYBEAN SEED ESTIMATION THROUGH TETRAZOLIUM TEST.**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para obtenção de título de Mestre.

APROVADA em 29 de janeiro de 2020.

Drº Joel Augusto Muniz UFLA

Drº Peter de Matos Campos UFSJ

Profª. Drª. Carla Regina Guimarães Brighenti

Orientadora

**LAVRAS – MG**

**2020**

À Deus, por guiar-me sempre.  
Aos meus pais e irmãos pela  
confiança, amor e incentivo.

## AGRADECIMENTOS

Ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária e à Universidade Federal de Lavras, pela oportunidade de estudo, por toda ajuda e aprendizado.

À minha orientadora, Prof<sup>ª</sup>. Dr<sup>ª</sup>. Carla Regina Guimarães Brighenti, pela orientação, pelos inestimáveis aprendizados transmitidos e pela confiança em mim depositada para elaboração deste trabalho.

Ao Laboratório de Sementes da Universidade Federal de Lavras pelos dados fornecidos.

A todos os professores, amigos (as) e colegas que participaram da realização deste trabalho acadêmico, proporcionando recursos e subsídios teóricos, práticos e avaliativos para a composição deste estudo.

À Deus pela oportunidade da vida e por todas as bênçãos e glórias recebidas durante a minha existência.

Aos meus pais, Paulo Sérgio e Adailza, pela afeição, sustento, educação, paciência, incentivos aos estudos e, principalmente, pelo amor e carinho que me ofereceram, me oferecem e sempre continuarão a me oferecer. A minha gratidão eterna aos meus pais, que me ensinaram a trilhar os caminhos mais compensativos da vida.

E, aos meus irmãos, Mariane e Luiz Augusto, pelo incondicional companheirismo, amor e carinho proporcionados durante nossas vidas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”*

***Arthur Schopenhauer***

## RESUMO

A soja é uma oleaginosa de grande relevância para o Brasil por ser uma das principais culturas produzidas no país. O sucesso desta cultura depende de diversos fatores, sendo o principal deles a qualidade de sementes. Para o controle de qualidade o teste de tetrazólio tem se destacado pois além de avaliar a viabilidade de germinação e o vigor dos lotes de sementes, possibilita classificar os tipos de danos. A amostragem é uma etapa importante que deve ter seu custo reduzido. Assim, o uso da amostragem sequencial, que leva em conta a avaliação imediata de cada semente, pode permitir que a decisão sobre a viabilidade e vigor seja tomada antecipadamente, sem que todas as sementes sejam avaliadas como previsto no teste clássico por formalidade do método. Além disso, pode-se melhorar as estimativas obtidas por meio das técnicas Bayesianas, incluindo uma informação *a priori*. Desta forma, objetivou-se o estudo da aplicação dos testes sequenciais bayesianos para a verificação do vigor e das proporções de danos por categorias nas sementes de soja através do teste de tetrazólio. Utilizou-se a modelagem Binomial para estimar a proporção de vigor em lotes e a Multinomial para as categorias de danos (mecânico, por umidade ou percevejo). Em cada caso, *as prioris* conjugadas foram utilizadas sendo os parâmetros elicitados. Pode-se concluir que as quatro abordagens frequentista, bayesiana, sequencial e sequencial bayesiana, foram eficientes para realizar a estimação e tomada de decisão sobre os parâmetros, com redução do tamanho amostral. No processo sequencial bayesiano, tanto na estimação de vigor (uso da distribuição binomial), quanto de danos (uso da distribuição multinomial), foi possível reduzir o tamanho amostral. Os danos por umidade estavam presentes em 20,16% das sementes de soja avaliadas, por percevejo em 3,5% e os danos mecânicos, no caso da colheita manual, em apenas 1,92%. Obteve-se ainda 1% das sementes com mais de dois danos.

**Palavras-Chave: Estimação Sequencial Bayesiana. Distribuição Multinomial. Distribuição Dirichlet. Danos Mecânicos. Danos Por Umidade. Danos Por Percevejo.**

## ABSTRACT

Soy is an oilseed of great relevance to Brazil because it is one of the main crops produced in the country. The success of this crop depends on several factors, the main one being seed quality. For quality control, the tetrazolium test has stood out because in addition to assessing the germination viability and the vigor of the seed lots, it makes it possible to classify the types of damage. Sampling is an important step that should be reduced in cost. Thus, the use of sequential sampling, which takes into account the immediate evaluation of each seed, can allow the decision on viability and vigor to be taken in advance, without all the seeds being evaluated as foreseen in the classic test due to the formality of the method. In addition, it can improve the estimates obtained through Bayesian techniques, including a priori information. Thus, the objective was to study the application of Bayesian sequential tests to verify the vigor and damage proportions by categories in soybean seeds through the tetrazolium test. Binomial modeling was used to estimate the proportion of vigor in batches and Multinomial for damage categories (mechanical, moisture or bed bug). In each case, the combined priorities were used and the parameters were elicited. It can be concluded that the four frequentist, Bayesian, sequential and sequential Bayesian approaches, were efficient to perform the estimation and decision making on the parameters, with a reduction in the sample size. In the sequential Bayesian process, both in the estimation of vigor (use of the binomial distribution) and of damage (use of the multinomial distribution), it was possible to reduce the sample size. Moisture damage was present in 20.16% of the evaluated soybean seeds, by bed bug in 3.5% and mechanical damage, in the case of manual harvesting, in only 1.92%. It was also obtained 1% of the seeds with more than two damages.

**Keywords: Bayesian Sequential Estimation. Multinomial Distribution. Dirichlet Distribution. Mechanical Damage. Moisture Damage. Bedbug Damage.**

## LISTA DE FIGURAS

Figura 1 - Exportação Brasileira do Agronegócio por Setores – 2019.....	17
Figura 2 - Processo de alteração física, devido à oscilação da umidade da semente de soja .....	22
Figura 3 – Sementes com aspectos de deterioração por umidade coloridas pelo sal de tetrazólio .....	22
Figura 4 - Sementes com danos de percevejo coloridas pelo sal de Tetrazólio .....	23
Figura 5 - Dano mecânico por abrasão diagnosticado através do teste de Tetrazólio....	25
Figura 6 - Gráfico de aceitação sequencial.....	32
Figura 7 - Modelo de ficha com os resultados do teste de tetrazólio em sementes de soja anotados e interpretados para um lote de vigor médio .....	44
Figura 8 - Simbologia utilizada na ficha de avaliação do teste de tetrazólio .....	45
Figura 9 - Densidades das distribuições <i>a priori</i> Beta .....	55
Figura 10 - Intervalos de Credibilidade para os 12 lotes estudados, incluindo a estimativa frequentista (●). .....	58
Figura 11 - Densidades da distribuição da priori utilizada no caso sequencial bayesiano. ....	61
Figura 12 - Relatório do programa em Delphi com os resultados e os parâmetros utilizados.....	63
Figura 13 - Simplex da <i>priori</i> da Dirichlet com parâmetros (0,2456; 0,1048; 0,1496) e da <i>posteriori</i> com os parâmetros (23,2456; 42,1048; 242, 1496) .....	69

## LISTA DE TABELAS

Tabela 1 - Classificação das sementes em três classes.....	45
Tabela 2 - Classificação das sementes em cinco classes.....	45
Tabela 3 - Valores de referência do vigor das sementes. ....	49
Tabela 4 - Porcentagem das sementes vigorosas e não vigorosas.....	53
Tabela 5 - Valores dos hiperparâmetros, média (%) e variâncias (%) das <i>priori</i> .....	54
Tabela 6 - Porcentagens das médias e variâncias estimadas a partir das <i>posteriores</i> dos parâmetros de proporção de sementes vigorosas.....	56
Tabela 7 - Intervalos de Credibilidade a 95% obtidos a partir das distribuições a <i>posteriori</i> dos parâmetros de proporção de sementes vigorosas.....	58
Tabela 8 - Porcentagens de vigor dos lotes de sementes estimados pelo método sequencial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ).....	59
Tabela 9 - Valores dos hiperparâmetros para 60% e 75% de vigor com variância de 10%. .....	61
Tabela 10 - Porcentagens de vigor dos lotes de sementes estimados pelo método sequencial bayesiano, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ).....	63
Tabela 11 - Porcentagem das sementes de acordo com as três classificações: $p_1$ : dano por umidade; $p_2$ : dano por percevejo; $p_3$ : dano mecânico. ....	65
Tabela 12 - Porcentagem das sementes de acordo com as cinco classificações: $p_1$ : vigorosa; $p_2$ : com presença de dano por umidade; $p_3$ : com presença de dano por percevejo; $p_4$ : com presença de dano mecânico; $p_5$ : com presença de dois ou mais danos. .....	66
Tabela 13 – Hiperparâmetros e valores da <i>priori</i> obtidos da multinomial com três classes de danos. ....	68
Tabela 14 – Estimativas <i>a posteriori</i> da multinomial para proporção de danos em sementes de soja. ....	68

Tabela 15 - Valores da <i>priori</i> da multinomial com cinco parâmetros ( $p_1$ : vigorosa; $p_2$ : com presença de dano por umidade; $p_3$ : com presença de dano por percevejo; $p_4$ : com presença de dano mecânico; $p_5$ : com presença de 2 ou mais danos).....	70
Tabela 16 – Estimativas a <i>posteriori</i> da multinomial para proporção de danos em sementes de soja. ....	71
Tabela 17 - Valores da <i>posteriori</i> da multinomial com cinco parâmetros. ....	71
Tabela 18 - Porcentagens de danos dos lotes de sementes estimados pelo método sequencial bayesiano multinomial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ), de acordo com os três parâmetros.....	74
Tabela 19 - Porcentagens de danos dos lotes de sementes estimados pelo método sequencial bayesiano multinomial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ), de acordo com os cinco parâmetros.....	75

## SUMÁRIO

1	INTRODUÇÃO.....	15
2	REVISÃO BIBLIOGRÁFICA.....	17
2.1	A importância econômica da soja.....	17
2.2	Teste de tetrazólio.....	18
2.3	Fatores que afetam a qualidade da semente de soja.....	21
2.3.1	Danos por umidade.....	21
2.3.2	Danos por percevejos.....	23
2.3.3	Danos mecânicos.....	24
2.4	Amostragem em sementes.....	25
2.4.1	Amostragem sequencial.....	27
2.4.1.1	Teste da Razão de Probabilidade Sequencial.....	28
2.4.1.1.1	TRPS para distribuição binomial.....	30
2.4.2	Análise bayesiana.....	34
2.4.2.1	Estimação sequencial bayesiana.....	37
2.4.2.2	Teste sequencial bayesiano para distribuição Binomial.....	38
2.4.2.3	Teste bayesiano para distribuição multinomial.....	39
3	METODOLOGIA.....	43
3.1	Obtenção dos dados.....	43
3.2	Análise descritiva.....	44
3.3	Estimação frequentista.....	46
3.4	Estimação bayesiana.....	46
3.5	Estimação sequencial.....	49
3.6	Estimação sequencial bayesiana.....	50
4	RESULTADOS.....	53
4.1	Estimação do vigor de sementes de soja utilizando a distribuição Binomial... 53	

4.1.1	Abordagem frequentista da estimação do vigor .....	53
4.1.2	Abordagem bayesiana da estimação do vigor .....	54
4.1.3	Estimação sequencial do vigor .....	59
4.1.4	Estimação sequencial bayesiana do vigor .....	60
4.2	Estimação de danos em sementes de soja utilizando a distribuição Multinomial.....	64
4.2.1	Abordagem frequentista da estimação de danos .....	64
4.2.2	Abordagem bayesiana da estimação de danos .....	67
4.2.3	Estimação sequencial bayesiana de danos considerando a distribuição multinomial.....	72
5	CONCLUSÕES.....	77
	REFERÊNCIAS .....	78

## 1 INTRODUÇÃO

A soja (*Glycine max (L.) Merrill*) é o principal produto agrícola na pauta de exportações brasileiras e o maior responsável pelo aumento da colheita nacional de grãos. O sucesso de sua lavoura depende de diversos fatores, mas sem dúvida, o mais importante deles é o emprego de sementes de alta qualidade, que geram plantas de alto vigor e, portanto, um desempenho superior no campo (FRANÇA-NETO et al., 2016). Para garantir esse quesito é necessário empregar um sistema de controle de qualidade confiável, oferecendo resultados rápidos.

O processo de classificação de lotes na indústria de sementes apresenta uma séria limitação, o tempo de espera dos testes de germinação, o qual ainda não fornece informações quanto ao vigor. Desta maneira, não permite o reconhecimento dos motivos da oscilação da qualidade dos lotes de sementes, além de apresentar resultados imprecisos pela ocorrência de danos por fungos e embebição da semente (HENNING; FRANÇA NETO, 1980). Essas limitações fazem com que haja queda na produtividade e prejuízos econômicos aos produtores por tomar decisões erradas sobre o sentenciamento do lote.

O teste de tetrazólio tem se sobressaído dos demais métodos de controle de qualidade empregados na indústria de sementes no Brasil, principalmente para a cultura da soja. Ele, além de avaliar a viabilidade de germinação e o vigor dos lotes de semente, diagnostica os possíveis responsáveis pela redução da qualidade dos lotes (FRANÇA-NETO; KRZYZANOWSKI; COSTA, 1998). Esse diagnóstico apresenta grande importância sob o uso deste teste no país, pois, além de identificar os problemas que afetam a qualidade, expõe a origem desses problemas, permitindo adotar ações corretivas para garantir uma maior produtividade, além da obtenção de um diagnóstico correto.

A obtenção de amostras representativas do lote, com possível redução dos custos, sempre foi o objetivo do processo de amostragem. Uma das formas de reduzir estes custos é substituir a amostragem clássica, com tamanho fixo de sementes avaliadas, pela amostragem sequencial que possibilite o emprego de um número variável de unidades amostrais em função da qualidade do lote. A amostragem sequencial tem se mostrado mais rápida e confiável que a amostragem clássica e é indicada pela Associação Internacional de Testes em Sementes – ISTA (BÁNYAI; BARABÁS, 2002). No entanto, a amostragem sequencial ainda é pouco utilizada, devido à falta de divulgação e da necessidade de, muitas vezes, aspectos computacionais para auxiliar nos cálculos durante o processo de tomada de decisão. Pode-se ainda incluir, no plano de amostragem aspectos regionais e sazonais incorporando informações referentes a esses aspectos

por meio de técnicas bayesianas, utilizando a distribuição *a priori* (GARTHWAITE; JOLLIFFE; BYRON, 1995).

Destaca-se também que, ao se avaliar a característica de vigor das sementes, considerando apenas duas categorias, vigorosas ou não, é suficiente utilizar para análise, a distribuição binomial. Por outro lado, ao considerarmos os diferentes tipos de danos nos grãos, que possuem mais de duas categorias, a análise deve considerar a modelagem multinomial. Contudo, a análise utilizando a modelagem multinomial, tem sido pouco utilizada na área de análise de sementes. No entanto, pode ser ferramenta útil para analisar os problemas que afetam a qualidade das sementes e categorizá-los, contribuindo para o avanço dos testes de sentenciamento dos lotes de sementes de soja, bem como, qualquer outro produto com mais de duas categorias.

Desta forma, o principal objetivo deste trabalho é o estudo da aplicação dos testes sequenciais bayesianos para a verificação do vigor e de danos das sementes de soja através do teste de tetrazólio considerando as distribuições Binomial e Multinomial. A proposta abre possibilidade de minimizar o tamanho amostral utilizado, reduzindo custo de toda operação e o tempo de espera.

## 2 REVISÃO BIBLIOGRÁFICA

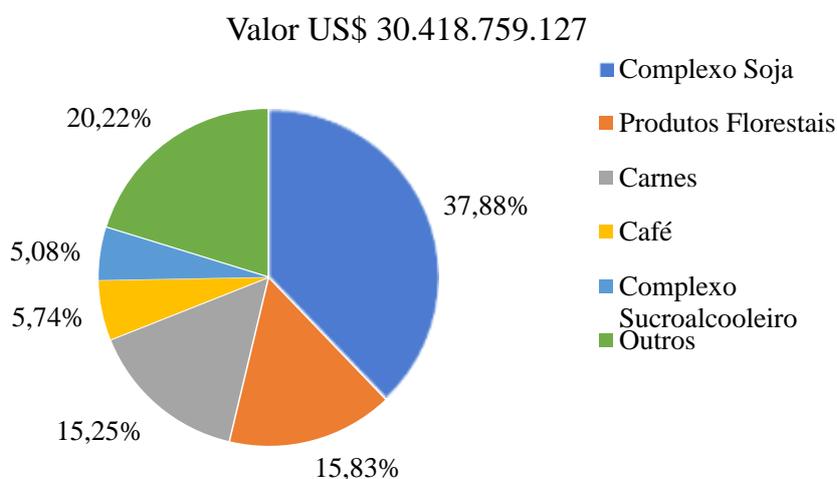
### 2.1 A importância econômica da soja

Segundo o portal da Embrapa Soja, a soja tem origem na costa leste da Ásia, principalmente ao longo do rio Yangtse, na China. Sendo hoje muito diferente de suas plantas ancestrais rasteiras (SILVEIRA; CONTE; MESQUITA, 2017).

O complexo da soja atualmente é responsável por profundas modificações no território nacional, sendo um importante elo da economia interna com a internacional. Segundo dados da Companhia Nacional de Abastecimento (CONAB) na safra 2017/2018 cerca de 116,996 milhões de toneladas de soja foram produzidas no Brasil, com uma produtividade de 3333 kg/ha. Hoje a referida oleaginosa é o principal produto agrícola na pauta de exportações brasileiras e o maior responsável pelo aumento da colheita nacional de grãos, conforme indicado na Figura 1 (MAPA, 2019).

Segundo estimativa do Departamento de Agricultura dos Estados Unidos, em sua última divulgação do quadro de oferta e demanda mundial referente ao mês de julho/2018, foi estimado para a safra mundial de soja em grãos 2018/19 que Brasil (33,52%), EUA (32,63%) e Argentina (15,86%), são responsáveis por 82,01% de toda a produção mundial de soja em grão, e a China por 61,54% de todas as importações mundiais (MAPA, 2019).

Figura 1 - Exportação Brasileira do Agronegócio por Setores – 2019



Fonte: AgroStat Brasil (MAPA, 2019).

Para dados de 2019 a expectativa é de que o Brasil será o maior produtor de soja do mundo, devido as reduções da área de produção americana em 0,71% e o aumento de quase 7% nas áreas de plantio brasileira. E na safra 2018/19, foi o maior exportador de soja do mundo, com 75 milhões de toneladas de soja em grãos exportados, aumento de 0,47%, em relação às exportações da safra 2017/18 estimada em 74,65 milhões de toneladas (CONAB, 2018).

O uso de semente de boa qualidade concede o acesso aos avanços genéticos, com as garantias de qualidade e as tecnologias de adaptação nas diversas regiões, obtendo um desempenho superior no campo, assegurando maiores produtividades (CONAB, 2016).

As sementes de soja, devido as suas características morfológicas e químicas, destacam-se por serem bastante sensíveis à ação de fatores do ambiente (MARCOS FILHO et al., 1985). A qualidade das sementes, principalmente em regiões tropicais, pode ser influenciada por diversos fatores que ocorrem em todas as etapas de produção e vários testes são propostos para avaliar sua qualidade, destacando-se o teste de tetrazólio.

A Lei nº 10.711, de 5 de agosto de 2003 regulamenta o Sistema Nacional de Sementes e Mudas, garantindo a qualidade das sementes de padrões mínimos de germinação, pureza física e varietal, bem como sanidade, exigidos por normas de produção e comercialização estabelecidas e fiscalizadas pelo governo (BRASIL, 2003).

## **2.2 Teste de tetrazólio**

De acordo com Moore (1976), a primeira divulgação do teste de tetrazólio aconteceu após a II Guerra Mundial pelo Exército Americano em decorrência de investigações das pesquisas que estavam sendo realizadas na Alemanha. Nos Estados Unidos o primeiro trabalho publicado sobre o teste de tetrazólio foi de Porter, Durrell e Romm (1947) na Universidade Estadual de Iowa.

Graças as publicações das diferentes versões dos manuais do teste de tetrazólio para a cultura da soja (HENNING; FRANÇA-NETO; COSTA, 1981; FRANÇA-NETO; PEREIRA; COSTA, 1985; FRANÇA-NETO et al., 1988; FRANÇA-NETO; KRZYZANOWSKI; COSTA, 1998) e as capacitações concedidas pelos profissionais da Embrapa Soja e de outras instituições, tem-se a certeza de que no país o teste é realizado em todos os laboratórios de análise de sementes que trabalham com a soja. Em resultado, o Brasil é líder mundial de utilização do teste de tetrazólio em nível de rotina nos dias de hoje.

O teste de tetrazólio é fundamentado através de atividades enzimáticas que catalisam as reações respiratórias, durante o ciclo de Krebs e a glicólise. A atividade dessas enzimas, principalmente a de origem do ácido málico a desidrogenase, reduzem o sal de tetrazólio (2,3,5-trifenil cloreto de tetrazólio ou TCT) nos tecidos vivos. Quando a semente é colocada na solução incolor do sal de tetrazólio ocorre a redução do composto nos tecidos vivos formando um outro composto vermelho, sendo ele estável e não-difusível, conhecido por trifênilformazan (MOORE, 1973).

Quando ocorre essa redução do sal de tetrazólio em trifênilformazan e a coloração da semente, é detectada a atividade respiratória, sinalizando a viabilidade delas. Quando não ocorre a coloração, a semente é classificada como inviável (MOORE, 1973).

Para que se tenha um resultado confiável é necessário que o analista seja bem treinado sob as técnicas do teste e tenha o conhecimento das estruturas da semente. Além disso, é necessário experiência e julgamento crítico para observar as anormalidades reveladas pelo teste de tetrazólio. A precisão do teste depende do conhecimento de todas as técnicas e procedimentos envolvidos (KRZYZANOWSKI; VIERA; FRANÇA-NETO, 1999).

Moore (1985) cita que há três objetivos básicos na avaliação das sementes:

- a) determinar sob condições ideais a capacidade de germinação de um lote de sementes;
- b) categorizar as sementes de acordo com as classes de viabilidade; e,
- c) diagnosticar as prováveis causas que resultam na perda de qualidade das sementes.

Faz-se necessário o conhecimento das diferentes classes de viabilidade para categorizar cada semente como viável, não viável e o tipos de danos que apresentam. Smith et al. (1956) estabeleceram um sistema de classificação para sementes de milho e de soja, onde as sementes viáveis são classificadas de 1 a 5 e as não viáveis de 6 a 8. Além das condições físicas é observado nesse sistema a presença, a localização e o tipo de dano. França-Neto, Pereira e Costa (1985) e França-Neto et al. (1988) modificaram esse sistema e o descreveram detalhadamente para sementes de soja.

A seguir, são apresentadas as classificações de vigor em sementes de soja:

- As classes de 1 a 3 correspondem às sementes viáveis e vigorosas:

**Classe 1:** é representada pelas sementes de mais alto vigor, ou seja, com todas as estruturas do embrião intactas e com coloração uniforme e superficial. Os tecidos do embrião estão normais e túrgidos.

**Classe 2:** é representada pelas sementes de alto vigor. São incluídas sementes com danos pequenos e superficiais, ocorrendo na superfície externa dos cotilédones. A superfície interna dos cotilédones e eixo embrionário não apresenta nenhum sinal de dano.

**Classe 3:** compreende as sementes classificadas como médio vigor. É representada por aquelas que apresentam estrias superficiais de coloração vermelho carmim forte ou áreas brancas presentes na superfície externa dos cotilédones; danos superficiais no córtex do eixo radícula hipocótilo, mas não alcançando o cilindro central. A superfície interna dos cotilédones pode apresentar pequenas áreas mais escuras correspondentes às estrias externas.

- As classes 4 e 5 correspondem às sementes apenas viáveis:

**Classe 4:** os danos são caracterizados por áreas de coloração vermelho carmim forte (tecido em estágio avançado de deterioração) ou branco leitoso (tecido morto). Os danos são visíveis na superfície interna dos cotilédones. A região vascular não está afetada, bem como a junção entre o eixo embrionário e os cotilédones. O cilindro central se apresenta intacto.

**Classe 5:** os cotilédones estão danificados severamente, mas 50% ou mais do tecido de reserva permanece viável e funcional. A região vascular, próxima ao ponto de ligação entre o eixo embrionário, está bem definida e viável. Sementes classificadas nesta classe germinarão e produzirão plântulas normais somente sob condições ideais.

- As classes de 6 a 8 correspondem às sementes que não são viáveis:

**Classe 6:** as sementes são caracterizadas pelas presenças de lesões similares às descritas para a Classe 5, mas a extensão e profundidade de tecido danificado é grande, tornando a semente não viável.

**Classe 7:** as sementes apresentam dano profundo no cilindro central. A região vascular entre o eixo embrionário e ambos os cotilédones está severamente danificada. A plúmula pode apresentar danos. Mais de 50% do tecido de reserva está deteriorado.

**Classe 8:** as sementes apresentam todas as estruturas do embrião esbranquiçadas (mortas), com tecidos flácidos e quebradiços.

### **2.3 Fatores que afetam a qualidade da semente de soja**

A produção de semente de soja de elevada qualidade é um desafio para o setor sementeiro, principalmente em regiões tropicais e subtropicais. Nessas regiões, a produção desse insumo só é possível, mediante a adoção de técnicas especiais. A utilização de sementes não certificadas apresenta menor produtividade pela menor qualidade das sementes, e ainda podem difundir patógenos provocando problemas fitossanitários, prejudicando o sistema de produção nas regiões (CONAB, 2016).

De acordo com França-Neto et al. (2016) existem quatro fundamentos que compõem a qualidade da semente de soja, são eles:

1. Qualidade fisiológica, representando uma semente com alto vigor e germinação e que resulte em adequada emergência de plântulas em campo;
2. Qualidade genética, sendo geneticamente pura, representando a cultivar que se deseja semear, sem misturas varietais;
3. Qualidade sanitária, compreendendo semente livre de outras sementes de plantas daninhas e de patógenos, sejam eles fungos, vírus, nematoides ou bactérias;
4. Qualidade física, composta por uma semente pura, livre de material inerte, como contaminantes, fragmentos de plantas, insetos, torrões e outras impurezas.

Destes fundamentos se destaca a qualidade fisiológica de sementes, que é facilmente influenciada por características de umidade, doenças, infestação de pragas e danos mecânicos (SANTOS et al., 2007). Os principais fatores que prejudicam a qualidade de semente serão apresentados a seguir, dando-se maior ênfase àqueles que reduzem a qualidade fisiológica, principalmente em regiões tropicais e subtropicais.

#### **2.3.1 Danos por umidade**

Em razão do clima predominantemente quente e úmido na região tropical, tem-se o desenvolvimento da deterioração por umidade. Sendo este um dos fatores que mais atinge a qualidade de sementes de soja (KRZYZANOWSKI; FRANÇA-NETO; COSTA, 2004).

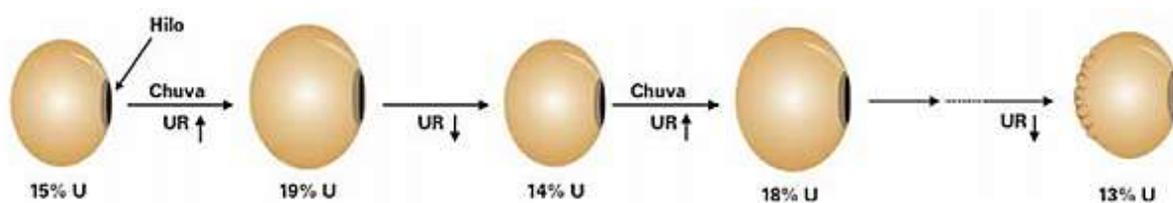
A deterioração das sementes ocorre devido à sua exposição a ciclos alternados de elevada e baixa umidade antes da colheita, em função da ocorrência de chuvas frequentes ou de

flutuações diárias de umidade relativa do ar. Essa deterioração será ainda mais intensa se tais condições estiverem associadas com condições de elevadas temperaturas (FRANÇA-NETO; HENNING, 1984).

A alteração física é ocasionada pelas sucessivas expansões e contrações do volume das sementes que ocasionam a formação de rugas nos cotilédones, na região oposta ao hilo, devido à ausência da camada tegumentar (MARCOS FILHO, 2005).

Na Figura 2 pode ser observado o processo de alterações físicas, devido à oscilação do teor de água da semente de soja em função das condições de umidade ambiental, que resultam no aparecimento de rugas na semente de soja, características da deterioração por umidade.

Figura 2 - Processo de alteração física, devido à oscilação da umidade da semente de soja



Fonte: França-Neto; Henning, 2016.

A deterioração por umidade é o dano que mais se desenvolve durante o armazenamento das sementes (MOREANO et al., 2011). A semente que sofreu chuvas em pré-colheita e apresenta esse tipo de dano, perderá a sua germinação e o seu vigor em índices acentuados durante a armazenagem, devido principalmente à evolução dos índices desse tipo de dano. O teste de tetrazólio consegue observar este tipo de dano através de manchas brancas nas sementes (Figura 3).

Figura 3 – Sementes com aspectos de deterioração por umidade coloridas pelo sal de tetrazólio



Fonte: Elaborado pelo autor.

### 2.3.2 Danos por percevejos

Outro tipo de dano que vem causando sérios prejuízos à cultura da soja é a incidência de percevejos. O rendimento e a qualidade são afetados seriamente pelos percevejos que se alimentam do grão, provocando a murcha e a má formação das vagens e grãos, resultando no retardamento do amadurecimento normal das plantas, deixando-as verdes na época da colheita (CORRÊA-FERREIRA; PANIZZI, 1999).

Os percevejos que mais atacam a soja são o pentatomídeos (família: *Pentatomidae*), que são sugadores de grãos, seguidos do percevejo-castanho ou enterrador (família: *Cydnidae*), que sugam as raízes, e os percevejo-alidídeo (família: *Alydidae*), que também sugam os grãos (CORRÊA-FERREIRA; PANIZZI, 1999).

Devido ao hábito alimentar, os percevejos causam problemas sérios à soja. Seus danos são irreversíveis, a partir de determinados níveis populacionais. Os grãos atacados ficam menores, enrugados e com cor mais escura que o normal (VILLAS BÔAS et al., 1982). Podem apresentar doenças como a mancha-fermento, causada pelo fungo *Nematospora corily Peglion*, o qual é transmitido durante a alimentação. Nos ataques iniciais, pode ocorrer abortamento de vagens (SINCLAIR, 1982).

Na Figura 4 pode ser observado a semente de soja colorida pelo sal de tetrazólio, com lesões típicas à picada de percevejo.

Figura 4 - Sementes com danos de percevejo coloridas pelo sal de Tetrazólio



Fonte: Elaborado pelo autor.

Além da redução na qualidade, na viabilidade e no vigor, as sementes de soja danificadas por percevejos sofrem alterações nos teores de proteína e de óleo. O ataque de percevejos causa retardamento da maturação (retenção foliar e haste verde), dificultando a colheita (CORRÊA-FERREIRA; PANIZZI, 1999).

O problema de percevejo na cultura da soja vem se agravando a cada safra, devido a não realização de monitoramento adequado das pragas, no aumento e no desenvolvimento de populações resistentes e nas aplicações descontroladas de produtos que desencadeiam o desequilíbrio na cultura, causando uma rápida ressurgência e níveis populacionais de percevejos e outras pragas elevados (HARGER et al., 2006; QUINTELA et al., 2006).

### **2.3.3 Danos mecânicos**

Entre as causas responsáveis pela perda de qualidade em sementes de soja, destacam-se também os danos mecânicos provocados, principalmente, durante as operações de colheita e beneficiamento. Trincas ou rachaduras situadas superficialmente são facilmente detectadas, ao passo que os danos mecânicos internos exigem exames mais detalhados para sua detecção (FLOR, 2004). Vale ressaltar que não apenas os danos grandes e visíveis, mas também danos menores ou microdanos, mesmo que invisíveis a olho nu, dependendo da sua localização, podem reduzir significativamente a qualidade das sementes (MESQUITA et al., 2002).

O conceito de dano mecânico é restrito aos distúrbios resultantes das forças destrutivas ocorridas durante a semeadura, colheita, trilha, transporte e beneficiamento. Estas injúrias não podem ser totalmente evitadas, mas sua extensão e severidade podem ser grandemente atenuadas (POPINIGIS, 1985).

A operação da colheita mecanizada é a maior fonte de danos mecânicos às sementes de soja, sejam eles imediatos ou latentes, pois durante a colheita a planta passa pelo processo de debulha, em que as sementes são separadas da palha. O dano ocorre em consequência dos impactos recebidos pela semente pelo cilindro debulhador. As quebras de pequenos fragmentos nas sementes afetam o rendimento no processo de beneficiamento, comprometendo os índices de vigor e de germinação, podendo servir de porta de entrada para patógenos (CARVALHO; NAKAGAWA, 2000).

Desta forma durante a colheita é essencial que os mecanismos de trilha estejam bem ajustados, pois é na operação de trilha das vagens que ocorre a quebra da semente de soja em decorrência do seu baixo grau de umidade. Normalmente, a semente quebra longitudinalmente na união dos cotilédones separando-os (FRANÇA-NETO; HENNING, 1984).

Outro aspecto importante a ser levado em consideração durante a colheita é o conteúdo de água da semente. Semente seca, ou seja, aquela com conteúdo abaixo de 12%, tenderá a apresentar danos mecânicos imediatos, caracterizados por fissuras, rachaduras e quebras.

Semente com conteúdo acima de 14% é mais suscetível aos danos mecânicos latentes, caracterizados por amassamentos e abrasões (FRANÇA-NETO; HENNING, 1984).

As sementes de soja possuem muita sensibilidade aos danos mecânicos. Por apresentar tegumento pouco espesso que deixa as partes vitais do eixo embrionário, praticamente, sem proteção (FRANÇA-NETO; KRZYZANOWKI; COSTA, 1998).

Dentre os testes utilizados para detectar danos mecânicos, o teste de tetrazólio tem se destacado, principalmente para soja, devido a sua rapidez e precisão. Este é capaz de identificar três tipos de danos mecânicos: rachaduras, amassamentos e abrasões (KRZYZANOWKI; VIEIRA; FRANÇA-NETO, 1999). Na Figura 5 é possível observar o dano mecânico por abrasões, devido a presença de lesões de coloração vermelho escura.

Figura 5 - Dano mecânico por abrasão diagnosticado através do teste de Tetrazólio



Fonte: Elaborado pelo autor.

## 2.4 Amostragem em sementes

De acordo com a Lei 10.711, de 5 de agosto de 2003 a amostragem de sementes tem como finalidade obter uma quantidade representativa do lote ou parte dele, quando se apresentar subdividido, objetivando a análise. Sendo ela para fins de análise, certificação e fiscalização (BRASIL, 2003).

Existem normas e procedimentos estabelecidos a partir de resultados de pesquisa e de legislação nacional e internacional que norteiam a amostragem de sementes. Para o comércio internacional devem ser seguidas as prescrições da ISTA - *International Seed Testing Assotiation*, mas para o comércio nacional existem procedimentos oficializados pelo Ministério de Agricultura, Pecuária e Abastecimento (MAPA) que estão prescritos nas Regras para Análise de Sementes (RAS) (BRASIL, 2009).

Conforme as RAS é fundamental obter amostras de tamanho adequado para os testes, na qual estejam presentes os mesmos componentes do lote de sementes e em proporções semelhantes, para a avaliação da qualidade e assim alcançar resultados uniformes e precisos (BRASIL, 2009).

A amostra média de sementes de soja que será levada para análise é de 1000 g que são extraídas do lote de 30.000 kg, sendo o lote definido como sementes identificadas dentro das tolerâncias permitidas de homogeneidade e uniformidade (BRASIL, 2009).

No ato da amostragem, o amostrador deve verificar se o lote de sementes se encontra dentro dos requisitos de rotulagem, lacre, tamanho máximo, homogeneidade e representatividade. Uma amostra será tanto mais representativa do lote à medida que aumentar o número de amostras simples, ou seja, pequenas porções de sementes retiradas de pontos do lote (CARVALHO et al., 2018).

Na prática, entretanto, um lote de sementes nunca é perfeitamente homogêneo, definindo-se como uma porção de sementes cujas partes que o compõe estejam razoável e uniformemente distribuídas por toda a sua massa. Essa uniformidade se refere a qualquer dos atributos que podem ser determinados em um exame, tais como pureza, outras sementes por número e germinação. (BRASIL, 2009). Porém, o proprietário do lote é obrigado a fornecer todas as informações relativas ao volume e mistura de cada lote amostrado. Se o estabelecimento não puder fornecer a documentação adequada para garantir a homogeneidade do lote, a amostragem deve ser recusada (CARVALHO et al., 2018).

O teste de tetrazólio requer um menor número de sementes analisadas devido as condições homogêneas que as sementes são sujeitas durante todo o preparo, o que não ocorre na realização do teste padrão de germinação (FRANÇA-NETO; KRZYZANOWSKI, 2018).

Se o fim da realização do teste for oficial, conforme estabelecido pelas RAS, utilizam-se 400 sementes para a amostragem, sendo elas subdivididas em quatro repetições de 100 sementes ou em oito repetições de 50 sementes. Ainda há a opção de realizar o teste em duas repetições de 100 sementes ou em quatro de 50, totalizando em 200 sementes, também para fins de fiscalização (BRASIL, 2009). Entretanto, se a finalidade do teste for para controle interno de qualidade ou de determinação de vigor podem ser utilizadas 100 sementes, subdivididas em duas amostras de 50 sementes em cada (MOORE, 1973; FRANÇA-NETO, 1981; FRANÇA-NETO et al., 1985, 1988, 1998, 1999).

Uma decisão sobre a qualidade do lote com base nos dados da amostra é essencialmente um teste de significância, ou seja, uma decisão relativa a uma característica de qualidade de lote

ou decisão de verdade na rotulagem sobre um lote de sementes. Nem o tamanho da amostra e nem o nível de significância são determinados pelo usuário, mas são prescritos pelas regras internacionais. Embora o aumento do tamanho da amostra melhore a confiabilidade da decisão, o aumento não é proporcional para o nível de significância (BÁNYAI; BARABÁS, 2002).

Para o controle de qualidade, é importante determinar o nível de qualidade aceitável (NQA ou  $p_1$ ) e o nível de qualidade rejeitável (NQR ou  $p_2$ ), pois com base nesses parâmetros é possível desenhar a Curva Característica de Operação (Curva CO). Tal curva mostra a probabilidade de aceitação em função do nível real de qualidade  $p$  no lote, ou seja, a fração de defeitos (BÁNYAI; BARABÁS, 2002). A curva CO esboça o desempenho de um plano para várias possíveis proporções defeituosas.

Uma vez que o lote pode apresentar qualquer proporção de defeito, pode-se descrever o comportamento deste plano com a curva característica de operação (CO), que traça a probabilidade de aceitação contra possíveis valores de proporção defeituosa (WALD, 1947).

Desta forma, a probabilidade de o lote ser aceito ( $P_a$ ) com uma proporção de defeitos  $p$ , seguindo uma distribuição binomial, é dado por:

$$P_a = P(d \leq c) = \sum_{x_i=0}^c P(d = x_i) = \sum_{d=0}^c \binom{n}{d} p^d (1-p)^{n-d}, \quad (01)$$

onde,  $d$  é o número de itens com defeito e  $c$  é o número de aceitação de defeitos no lote.

O erro tipo I ( $\alpha$ ) é chamado, na comercialização de um lote, de “risco do produtor”, pois representa a probabilidade de rejeitar um lote que deveria ser aceito por possuir um nível de qualidade igual ao NQA. À medida que  $\alpha$  aumenta o risco de rejeição de lotes com taxas defeituosas iguais a NQA aumenta, o que causa prejuízo ao produtor. Já o risco do consumidor ( $\beta$ ), é a probabilidade de aceitar um lote com um nível de qualidade rejeitável (NQR) de defeitos. À medida que NQR aumenta o risco de aceitar lotes defeituosos com taxas iguais a NQR aumenta, o que causa prejuízo ao consumidor. Risco do consumidor corresponde ao erro de tipo II.

#### 2.4.1 Amostragem sequencial

A análise sequencial foi desenvolvida durante a II Grande Guerra. A teoria formal conhecida hoje como Análise Sequencial começou em 1943 com o trabalho de A. Wald na América do norte e G. A. Barnard na Grã-Bretanha. Contribuíram também os trabalhos de Dodge e Roming (1959) e Albrecht et al. (1955) citados por Berquó (1959). No entanto, Wald

(1947) foi quem mais contribuiu com estudos sequenciais para reduzir a inspeção amostral necessária na indústria. Pode ser aplicada tanto a problemas relativos às proporções como às mensurações (WETHERILL; GLAZEBROOK, 1986). Através dela os dados são analisados à medida que são obtidos, possibilitando, na maioria dos casos, reduções apreciáveis do número total de observações que devem ser feitas para se ter conclusões estatisticamente válidas.

A amostragem sequencial é o procedimento de amostragem de aceitação mais exigente envolvendo a tomada de decisão quanto à disposição do lote ou a coleta de amostra sucessivamente à medida que cada item é coletado. Ela pode ser considerada como plano de amostragem múltipla com tamanho de amostra um e sem limite superior no número de amostras a serem coletadas. Pode ser demonstrado que a abordagem sequencial fornece uma eficiência ótima na amostragem, em relação ao tamanho médio de amostra (ASN), sendo mais baixo possível (WALD, 1947).

Planos sequenciais são frequentemente aplicados onde o tamanho da amostra é crítico, de modo que uma amostra mínima deve ser tomada. Eles são um pouco mais difíceis de administrar do que planos de amostragem múltipla, já que em aplicações específicas a quantidade de esforço de inspeção não é determinada até que a amostra seja coletada. A possibilidade de tomar uma amostra de cada vez deve existir. Além disso, o procedimento operacional requer um profissional experiente e confiável, pois é um pouco mais rígida do que a amostragem simples, dupla ou múltipla (SCHILLING; NEUBAUER, 2017).

O plano de amostragem sequencial é instrumento de fácil uso para técnicos e mesmo para agricultores mais esclarecidos, o que facilita a tomada de decisões. A sua construção é baseada no Teste Sequencial da Razão de Probabilidade (WETHERILL; GLAZEBROOK, 1986).

#### **2.4.1.1 Teste da Razão de Probabilidade Sequencial**

O Teste da Razão de Probabilidade Sequencial (TRPS) é um caso específico do teste de hipóteses, desenvolvido originalmente por A. Wald em 1943 para estudos de controle de qualidade em sistemas de manufatura. É uma generalização do teste da razão de máxima verossimilhança para análises sequenciais.

Esse teste é utilizado quando se deseja obter a aceitação ou a rejeição do desempenho de algum sistema que faça uso de requisitos estatísticos (BALLARIS, 2014). Ele apresenta

propriedades ótimas porque minimiza o número esperado de observações necessário para uma decisão. Uma vantagem adicional deste teste é que os limites para a razão de probabilidade podem ser obtidos pelo menos sem resolver qualquer problema de distribuição. Isto torna o alcance do teste mais amplo porque não se necessita restringir a certos tipos de distribuições que são comumente usados em testes usuais (GARRUTI, 1976).

O teste da razão de probabilidade procura distinguir entre duas hipóteses  $H_0$  e  $H_1$ :

$$\begin{cases} H_0: p = p_0 \text{ (a probabilidade de aceitação de } n_k \text{ defeitos na amostra)} \\ H_1: p = p_1 \text{ (a probabilidade de rejeição de } n_k \text{ defeitos na amostra)} \end{cases}$$

e a amostragem continua até que uma decisão seja tomada para uma hipótese ou para outra (GARRUTI, 1976).

De acordo com Wald (1947), considerando os dados que estão sendo analisados uma sequência de variáveis aleatórias  $x_1, x_2, x_3, \dots, x_n$ , para qualquer inteiro positivo  $n$ , quando  $H_1$  for verdadeira, a densidade de probabilidade da sequência de amostras acima definida será:

$$(i) P_{1n} = f(x_1, x_2, x_3, \dots, x_n; p_1) \quad (02)$$

E, analogamente, quando  $H_0$  for verdadeira:

$$(ii) P_{0n} = f(x_1, x_2, x_3, \dots, x_n; p_0); \quad (03)$$

Estabelecido isto, o TRPS de Wald (1947) determina que, em qualquer estágio da amostragem sequencial, onde já se coletou uma amostra de tamanho  $n$ , se:

$$(iii) \frac{f(x_1, x_2, x_3, \dots, x_n; p_1)}{f(x_1, x_2, x_3, \dots, x_n; p_0)} \geq A, \quad (04)$$

então, a amostragem deve ser cessada e deve-se rejeitar  $H_0$ ;

$$(iv) \frac{p(x_1, x_2, x_3, \dots, x_n; p_1)}{p(x_1, x_2, x_3, \dots, x_n; p_0)} \leq B, \quad (05)$$

então, a amostragem deve ser cessada e deve-se rejeitar  $H_1$ ;

$$(v) A < \frac{p(x_1, x_2, x_3, \dots, x_n; p_1)}{p(x_1, x_2, x_3, \dots, x_n; p_0)} < B, \quad (06)$$

então, continue amostrando sequencialmente, isto é, a observar o valor de  $x_{n+1}$  da variável sob estudo.

Para fins computacionais, é interessante calcular o logaritmo das desigualdades acima, o que facilita o cálculo da razão de  $P_{1n}/P_{0n}$ . As constantes A e B são escolhidas de tal forma que as probabilidades de erro  $\alpha$  e  $\beta$ , erros tipo I e II respectivamente, sejam especificados. Pode ser mostrado que são válidas as seguintes relações (WALD, 1947):

$$A = \frac{1 - \beta}{\alpha}, \quad (07) \quad B = \frac{\beta}{1 + \alpha} \quad (08)$$

O tamanho médio da amostra é uma variável aleatória, uma vez que se trata de um procedimento sequencial, o que pode muitas vezes originar problemas, pois o tamanho da amostra pode ser muito grande. No entanto, toda amostragem sequencial determinada por TSRP termina, ou seja, em algum valor de  $n$  ela cessará. Se a probabilidade é de que o procedimento terminará, o esquema é dito fechado. Se houver uma probabilidade diferente de zero de que a amostragem possa continuar indefinidamente, o esquema é chamado de aberto (ELLIS; WHITEHEAD, 1987).

Esquemas abertos são de pouca utilidade prática em comparação com esquemas fechados, e geralmente temos que os reduzir a forma fechada, colocando um limite superior ao tamanho da amostra. Esse truncamento muitas vezes dificulta a determinação de suas propriedades. Na literatura alguns autores utilizam o termo "fechado", enquanto outros utilizam o termo "truncado" para determinar a mesma ideia, isto é, a ocorrência de uma probabilidade diferente de zero. Correspondentemente, o termo "aberto", às vezes, significa não truncado (O'HAGAN; FORSTER, 2004).

#### 2.4.1.1.1 TRPS para distribuição binomial

Considerando o parâmetro proporção  $p$  de não conformidade de um lote de sementes por uma inconformidade específica de decidir entre dois valores para  $p$ : ou o lote está com uma proporção  $p_0$  de inconformidade mínima do nível de qualidade rejeitável ( $H_0$ ), ou está com a proporção  $p_1$  acima do nível de qualidade aceitável ( $H_1$ ).

$$\begin{cases} H_0: p \leq p_0 \text{ (proporção de danos abaixo ou igual ao tolerado – lote aceito)} \\ H_1: p > p_1 \text{ (proporção de danos acima do tolerado – lote rejeitado)} \end{cases}$$

O TSRP determina que em cada momento seja contado o número de sementes inviáveis. Este valor deve então ser comparado com dois limites de decisão, o  $LI$  (limite inferior) e  $LS$  (limite superior), a cada nova semente (GHOSH, 1970).

Considerando um tamanho  $n$  de amostra, o total acumulado de sementes inviáveis será  $\sum_{i=1}^k X_i \leq LS$ , onde  $X_i$  é uma variável aleatória do número de sementes com defeito, devendo-se tomar a decisão de rejeitar  $H_1$  e aceitar  $H_0$ , considerando assim o lote com a proporção de inviáveis abaixo do nível de qualidade aceitável e conseqüentemente, liberada para sua comercialização. Se,  $\sum_{i=1}^k X_i \geq LS$ , deve-se rejeitar  $H_0$  e aceitar  $H_1$ , considerando o lote inviável e, por sua vez, condenado para o plantio. No entanto, se o total acumulado estiver entre o limite inferior e superior, o analista deverá continuar amostrando (GHOSH, 1970).

A construção e avaliação de planos sequenciais para valores arbitrários de  $p_0$ ,  $p_1$ ,  $\alpha$  e  $\beta$  foram derivadas por Wald (1947) e pelo Statistical Research Group (1945). As fórmulas são as seguintes:

$$h_0 = \frac{\log \left[ \frac{(1-\beta)}{\alpha} \right]}{\log \left( \frac{p_1}{p_0} \right) + \log \left[ \frac{(1-p_0)}{(1-p_1)} \right]}; \quad (09)$$

$$h_1 = \frac{\log \left[ \frac{(1-\beta)}{\alpha} \right]}{\log \left( \frac{p_1}{p_0} \right) + \log \left[ \frac{(1-p_0)}{(1-p_1)} \right]}; \quad (10)$$

$$s = \frac{\log[(1-p_0)/(1-p_1)]}{\log \left( \frac{p_1}{p_0} \right) + \log \left[ \frac{(1-p_0)}{(1-p_1)} \right]}; \quad (11)$$

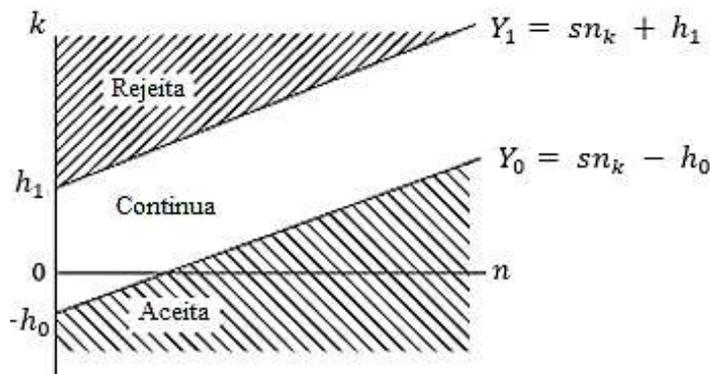
em que  $p_0$  é o nível de qualidade aceitável;  $p_1$  é o nível de qualidade rejeitável;  $\alpha$  é o erro tipo I;  $\beta$  é o erro tipo II;  $h_0$ ,  $h_1$  são interceptos; e,  $s$  é a inclinação para as retas.

Desta forma, as linhas de aceitação e rejeição são determinadas como:

$$Y_0 = sn_k - h_0 \text{ (aceitação)}, \quad (12) \quad Y_1 = sn_k + h_1 \text{ (rejeição)} \quad (13)$$

O plano é frequentemente implementado usando um gráfico, como mostrado na Figura 6, do número acumulado de defeituosos encontrados (eixo das ordenadas) em relação ao número de amostras individuais obtidas (eixo das abscissas).

Figura 6 - Gráfico de aceitação sequencial.



Fonte: Adaptado de Schilling; Neubauer, 2017.

Onde:

$n$  é o número de itens de amostra obtidos;

$k$  é o número de defeituosos encontrados pelo  $n$ -ésimo item de amostra;

$Y_1 = sn_k + h_1$ , é o limite de rejeição na  $n$ -ésima amostra;

$Y_0 = sn_k - h_0$ , é o limite de aceitação na  $n$ -ésima amostra.

Um ponto  $(p, P_a)$  na curva CO pode ser calculada (WALD, 1947) como:

$$p = \frac{1 - \left[ \frac{(1-p_1)}{(1-p_0)} \right]^x}{\left( \frac{p_1}{p_0} \right)^x - \left[ \frac{(1-p_1)}{(1-p_0)} \right]^x}, \quad (14)$$

com

$$P_a = \frac{\left[ \frac{(1-\beta)}{\alpha} \right]^x - 1}{\left[ \frac{(1-\beta)}{\alpha} \right]^x - \left[ \frac{\beta}{(1-\alpha)} \right]^x}, \quad (15)$$

em que  $p$  é a proporção de defeitos do lote,  $P_a$  é a probabilidade de aceitação e  $x$  é o número de itens avaliados.

O tamanho médio de amostra do plano fornece uma referência quanto ao tamanho esperado da amostra ( $E(N)$ ). Dada a combinação  $p$  e  $P_a$ , tem-se a equação para a  $E(N)$ :

$$E(N) = \frac{P_a \log \left[ \frac{\beta}{(1-\alpha)} \right] + (1-P_a) \log \left[ \frac{(1-\beta)}{\alpha} \right]}{p \log \left( \frac{p_1}{p_0} \right) + (1-p) \log \left[ \frac{(1-p_1)}{(1-p_0)} \right]}. \quad (16)$$

No caso de lotes muito bons ou muito ruins, a decisão pode ser tomada com base em tamanho de amostra menor. A decisão de ensaio sequencial será feita com base em um tamanho de amostra menor no mesmo nível de probabilidade. Esta é a vantagem do método de amostragem sequencial (BÁNYAI; BARABÁS, 2002).

Hoje, os *softwares* de amostragem de aceitação estão disponíveis para qualquer usuário construir um plano que atenda aos seus critérios. No entanto, deve-se notar também que um programa de planilhas pode ser usado para criar planos de amostragem e gerar planos de amostragem adequados para cada demanda. Schilling e Neubauer (2017) desenvolveram um conjunto de modelos de planos de amostragem no Microsoft Excel®. Esses modelos permitem que os usuários aprendam com mais eficiência os planos de amostragem e que programem planos mais rapidamente.

Botine (2015) desenvolveu o pacote SPRT para o *software R* para realizar o Teste de Razão de Probabilidade Sequencial de Wald de variáveis com distribuição Normal, Bernoulli, exponencial ou Poisson. Através dele é possível construir o gráfico de aceitação sequencial, assim como estabelecer regiões com truncamento. Este pacote está disponível no CRAN, o endereço de pacotes para R, e pode ser facilmente instalado.

O método de estimação sequencial não é usual por recomendações das normativas que regulamentam a amostragem por atributo. Contudo, vários autores demonstram a eficiência e rapidez da amostragem sequencial.

Ballaris (2014) em seu trabalho apresenta a utilização de plano de amostragem sequencial em sementes de soja e feijão para detecção de *Sclerotinia sclerotiorum* e conclui que o método é eficiente, desde que 800 a 1000 sementes sejam avaliadas, considerando incidências de 0,01 e 0,005%, respectivamente.

Garrut (1976) aplicou a análise sequencial para estudos de seleção de equipe de provadores em análise sensorial de alimentos, comprovando as vantagens à eficiência estatística, economia de tempo e material, além da facilidade de uso e procedimento sequencial gráfico.

Farias et al. (2001) utilizaram a amostragem sequencial baseada na lei de Taylor para o levantamento de lagartas *S. frugiperda* na cultura do milho, onde o plano de amostragem sequencial fornece o tamanho de amostra para a estimativa da densidade média populacional.

Penteado et al. (2008) objetivaram em sua pesquisa testar a técnica da amostragem sequencial para avaliação da eficiência do parasitismo de nematóide em adultos da vespa-da-

madeira, a qual se mostrou como uma alternativa viável, pois resultou em redução do tamanho da amostra, dos custos da atividade e obteve precisão nos resultados.

#### 2.4.2 Análise bayesiana

A inferência bayesiana é uma alternativa importante em relação aos procedimentos clássicos de estimação e vem recebendo reconhecimento das suas potencialidades práticas, onde a divulgação nas áreas aplicadas tem despertado interesse gradativo (KINAS, 2017). Surgiu quando Richard Price publicou a obra póstuma de Thomas Bayes, “*An essay towards solving a problem in the doctrine of chances*” em 1763.

De forma distinta à estatística frequentista, em que somente se admite probabilidade através de medidas de frequências relativas, a análise bayesiana é compreendida como a probabilidade de uma medida racional e condicional de incerteza (KINAS, 2017). A qual é fundamentada em probabilidades condicionais com um parâmetro  $\theta$ .

A informação de uma quantidade de interesse é dada pelo parâmetro  $\theta$ , que é desconhecido. A intensidade deste desconhecimento a respeito de  $\theta$  pode assumir diferentes graus de incerteza (PINHO, 2006). Desta forma, a característica essencial da inferência bayesiana é o uso da probabilidade para quantificar estas incertezas, baseando-se na probabilidade subjetiva. A qual mede o “grau de crença” que alguém deposita no acontecimento de um determinado evento do espaço amostral (SAVIAN et al., 2009; GELMAN et al., 2013).

Isto é, deve-se especificar uma distribuição de probabilidade para descrever o grau de crença sobre um parâmetro desconhecido, exigindo que os parâmetros sejam variáveis aleatórias, formalmente denominados como distribuição *a priori* (HINES et al, 2000).

A informação *a priori* é a hipótese, a qual é o valor fixado pelo pesquisador como verdadeiro para o parâmetro em estudo. Essa informação pode ter origem em fundamentos subjetivos, considerações particulares ou informações anteriores disponíveis na literatura (SORENSEN; GIANOLA, 2002).

A dificuldade principal na escolha da *priori* é que sua obtenção e quantificação são de natureza subjetiva, sendo útil sua elicitação (PAULINO et al., 2003). Existem alguns métodos de elicitções, sendo os principais (TIMPANI et al, 2015):

- Métodos estrutural: baseia-se em pesquisas diretamente relacionadas ao parâmetro em estudo;
- Métodos do histograma: processo iterativo no qual se elabora intervalos dos possíveis valores do parâmetro em estudo e avalia a probabilidade desses valores ocorrerem através de um histograma;
- Método da verossimilhança: utiliza as verossimilhanças de possíveis valores do parâmetro em estudo e elabora, por meio de uma regressão, uma função que contenha todos os possíveis pontos.

Os parâmetros indexados das distribuições *a priori* são denominados de hiperparâmetros para diferenciar dos parâmetros de interesse  $\theta$ . Esses hiperparâmetros são considerados conhecidos e traduzem a informação que se tem sobre o parâmetro  $\theta$ , antes da realização da amostra (SILVA, 2008).

Esta abordagem de hiperparâmetros, em geral, facilita a análise e é a mais utilizada para ultrapassar as eliciações das distribuições *a priori* subjetivas, de modo que o caso mais importante são as *prioris* conjugadas. O conceito consiste que as distribuições *a priori* e a *posteriori* possuam a mesma classe de distribuições, de modo que a mudança nos hiperparâmetros atualize o conhecimento que se tem sobre  $\theta$  (EHLERS, 2011).

Portanto, é possível representar a probabilidade para a ocorrência dos dois eventos pela distribuição de probabilidade conjunta e realizar as inferências em conjunto para as duas características estudadas (TIMPANI et al., 2015). Logo, o processo é baseado no teorema de Bayes e quando assume valores contínuos num dado intervalo pode ser representado por:

$$\pi(\theta|X) = \frac{L(\theta|X)\pi(\theta)}{\int L(\theta|X)\pi(\theta) d\theta}, \quad (17)$$

em que:

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

$$L(\theta|X) = \prod_{i=1}^n f(x_i|\theta), \quad (18)$$

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 teorema de Bayes é uma constante com relação a  $\theta$ , por funcionar como uma constante normalizadora de  $\pi(\theta|X)$  (PINHO, 2006). Portanto, o teorema pode ser simplificado em:

$$\pi(\theta|X) \propto L(\theta|X) \pi(\theta) \quad (19)$$

Em palavras, a distribuição *a posteriori* é proporcional ao produto da distribuição *a priori* pela verossimilhança, onde a constante de proporcionalidade é a constante normalizadora (HINES, 2000).

A forma simplificada do teorema de Bayes é útil em problemas que envolvam estimação de parâmetros, já que o denominador é apenas uma constante normalizadora. Em outras situações, como seleção de modelos, o termo possui extrema importância (EHLERS, 2011).

Conforme Box e Tiao (1992), a distribuição *a posteriori* pode ser representada como uma atualização de  $\pi(\theta|X)$ , devido a pouca informação sobre o parâmetro de interesse que a distribuição *a priori*  $\pi(\theta)$  possui. Nesta atualização está subentendido o princípio da verossimilhança, pois toda informação sobre o parâmetro está contida na distribuição amostral  $\pi(X|\theta)$  (PAULINO et al., 2003)

Assim, a densidade *a posteriori* para  $\theta$  expressa o nosso grau de crença em relação ao valor de  $\theta$ , dado o resultado da amostra (HINES, 2000).

Desta forma, a informação mais completa que se tem é a distribuição *a posteriori*. Porém, para conferir e discutir resultados, até mesmo ter conclusões de modo mais prático é necessário identificar medidas que resumem esta distribuição (GELMAN et al., 2013).

Um caso simples é a estimação pontual de  $\theta$ , em que se resume à distribuição marginal *a posteriori* por meio de um único valor,  $\hat{\theta}$ . Além disso, é importante relacionar informações sobre o quão preciso é a especificação deste valor (SAVIAN et al., 2009). De acordo com Ehlers (2011) as medidas de incertezas mais usuais são: a variância ou o coeficiente de variação para a média *a posteriori*, a medida de informação observada de Fisher para a moda *a posteriori*, e a distância entre quartis para a mediana *a posteriori*.

Contudo, o indicador mais apropriado na inferência bayesiana para essa incerteza é denominado de intervalo de credibilidade, que corresponde ao intervalo de confiança na teoria frequentista. Este intervalo é construído com base na distribuição *a posteriori* de  $\theta$ . Ou seja, é a o intervalo entre dois pontos que contém o verdadeiro valor de  $\theta$ , com probabilidade de  $(1-\alpha)$ , onde  $\alpha$  é o erro tipo I pré-estabelecido (FERNANDES et al., 2010).

Assim, quanto menor for o tamanho do intervalo mais concentrada é a distribuição de  $\theta$ . Em outras palavras, a amplitude do intervalo informa a dispersão de  $\theta$  (REIS et al., 2009).

Os intervalos de credibilidade são úteis na análise de teste de hipóteses bayesiano. Supondo que o interesse seja testar  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ . Para isso, constrói-se o intervalo de credibilidade para  $\theta$  e,  $\theta_0$  está contido no intervalo,  $H_0$  não é rejeitada. Caso contrário,  $H_0$  é rejeitada (BOLFARINE; SANDOVAL, 2001).

A partir do conhecimento sobre  $\theta$  é possível descrever sua densidade por uma forma funcional particular, por um procedimento de decisão sequencial.

#### 2.4.2.1 Estimação sequencial bayesiana

A idéia geral em análise sequencial bayesiana é que após cada observação realizada se deve comparar o risco de Bayes *a posteriori* de tomar uma decisão imediata com o risco de Bayes *a posteriori* esperado que será obtido se mais observações são tomadas (BERGER, 1985). O risco de Bayes de um procedimento sequencial  $d$  é definido por  $r(\pi, d) = E_\pi[R(\theta, d)]$ , ou seja, a esperança do risco dada *a priori*  $p(\pi)$ . O estimador de Bayes de  $\pi$  com respeito à função perda é aquele com menor risco de Bayes. No caso da função perda quadrática o estimador de Bayes para o parâmetro  $\pi$  será a média de sua distribuição atualizada.

O resultado consagrado pela análise bayesiana mostra que o menor risco de Bayes *a posteriori* é a variância da distribuição *a posteriori*, denotada por  $var_{post}(n)$  (SCHLAIFER; HOWARD, 1961). Daí,  $r_0(\pi_n, n) = var_{post}(n)$  e o menor risco de Bayes *a posteriori* esperado, se outra observação é tomada, isto é,  $r^1(\pi_n, n)$ , é a esperança desta variância, ou seja,  $r^1(\pi_n, n) = E[var_{post}(n)]$  (PHAM-GIA; TURKKAN, 1998). O tempo de parada dado em função do risco é o primeiro  $n$  para o qual  $r_0(\pi_n, n) \leq r^1(\pi_n, n)$ .

Assim, para determinar o critério de parada, é necessário calcular o risco imediato, que é dado pela variância *a posteriori* acrescida do custo de  $n$  observações, e o risco esperado, dado

pela esperança da variância *a posteriori* com o acréscimo no custo de mais uma observação. Pode-se dizer ainda que a amostragem deva continuar sempre que a esperança do risco *a posteriori* (ou a esperança da variância se mais uma observação for realizada) for menor que o risco “atual” (variância *a posteriori* com  $n$  observações) de tomar uma decisão (GARTHWAITE; JOLLIFFE; BYRON, 1995).

#### 2.4.2.2 Teste sequencial bayesiano para distribuição Binomial

A distribuição binomial descreve experimentos independentes, repetidos em condições semelhantes. Além disso, apenas dois resultados são possíveis em cada repetição: sucesso ou fracasso, sendo a probabilidade de ocorrer sucesso igual a  $p$ . Como as repetições são independentes, a probabilidade de ocorrerem  $k$  fracassos, em  $n$  repetições, é:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (20)$$

$$\mu_k = \text{Média dos } k = np \quad (21)$$

$$\sigma^2_k = \text{Variância} = np(1-p) \quad (22)$$

Portanto, a probabilidade de ocorrer, no máximo,  $k$  sementes não vigorosas, em  $n$  amostras, é:

$$F(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}, \quad (23)$$

em que,  $n$  é o número de sementes observadas e  $k$  é o total de sementes não vigorosas entre as observadas.

A distribuição Beta é caracterizada por dois parâmetros,  $a$  e  $b$ , e é frequentemente utilizada na modelagem de variáveis aleatórias contínuas com domínio entre 0 e 1, tal como as proporções e porcentagens e também usada, em estatística bayesiana, como a distribuição *a priori* conjugada de probabilidades binomiais (STURTZ; LIGGES; GELMAN, 2005). A função densidade de probabilidade da distribuição *a priori* beta com hiperparâmetros  $a$  e  $b$  é:

$$\pi(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}, \quad (24)$$

em que  $0 \leq p \leq 1$ ,  $a > 0$  e  $b > 0$ .

Sabendo que, para a distribuição *a priori* Beta ( $a, b$ ), a média e a variância são dadas por:

$$\mu_{\text{priori}} = \mu_{\text{beta}} = \frac{a}{(a+b)}, \quad (25) \quad \text{var}_{\text{priori}} = \frac{(ab)}{(a+b)^2(a+b+1)} \quad (26)$$

Então *a posteriori*  $\pi^n$  para o parâmetro será dada por:

$$\pi^n(p|k) \propto f(k|p) * p(\pi|a, b) \quad (27)$$

Para a distribuição *a posteriori* Beta ( $a', b'$ ) ou ( $a+k, b+n-k$ ), a média e a variância são dadas por:

$$\mu_{\text{posteriori}} = \frac{a+k}{(a+k+b+n-k)}, \quad (28)$$

$$\text{var}_{\text{posteriori}} = \frac{(a+k)(b+n-k)}{(a+b+n)^2(a+b+n+1)} \quad (29)$$

#### 2.4.2.3 Teste bayesiano para distribuição multinomial

Se uma resposta está limitada a um conjunto de possíveis opções ou categorias pré-estabelecidas, diz que a variável de interesse é politômica, sendo a distribuição multinomial comumente utilizada para representar tal variável (MCCULLAGH; NELDER, 1989).

Supondo que indivíduos numa população de interesse possuam uma, e apenas uma, das características  $A_1, \dots, A_k$ . Tais características podem ser, por exemplo, diagnóstico de dengue (MAFRA et al., 2010), de cárie (RIBEIRO et al., 2018), entre outras. Se a população é suficientemente grande e se uma amostra aleatória de tamanho  $m$  é sorteada, o número de indivíduos,  $x_i$ , que possuem a característica  $A_k$  é dada através da distribuição multinomial (CORDEIRO; NETO, 2004), expressa por:

$$P(x_i = x_1, \dots, x_k = x_k | m, p) = \binom{m}{x_i} p_1^{x_1} \dots p_k^{x_k} \quad (30)$$

onde,  $p_1, \dots, p_k$  são proporções populacionais de cada característica e  $\binom{m}{x_i} = \frac{m!}{x_1! \dots x_k!}$ .

De acordo com a demonstração de Casella e Berger (2010) do teorema multinomial abaixo, a distribuição marginal de qualquer componente  $n_{0i}$  da distribuição multinomial

corresponde a uma Binomial  $(n, p_i)$ , em que  $n$  é o número de tentativas e  $p_i$  é a probabilidade de obtenção de  $n_{0i}$  em cada tentativa.

**Teorema multinomial:** Seja  $m$  e  $n$  inteiros positivos. Seja  $A$  o conjunto de vetores de  $x = (x_1, \dots, x_n)$  tal que cada  $x_i$  é um inteiro não negativo  $\sum_{i=1}^n x_i = m$ . Então, para quaisquer números reais  $p_1, \dots, p_n$ ,

$$(p_1, \dots, p_n)^m = \sum_{x \in A} \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n},$$

considerando que os  $p_i$  probabilidades tais que  $\sum_{i=1}^n p_i = 1$ , tem-se que  $(p_1, \dots, p_n)^m = 1$ .

Calcula-se, então, a distribuição marginal de  $\mathbf{X}_n$  de uma distribuição multinomial. Para um valor fixo de  $x_n$ , com o objetivo de calcular distribuição marginal de  $f(x_n)$ , deve-se somar todos os possíveis valores de  $(x_1, \dots, x_{n-1})$ , ou seja, a soma sobre todos os  $(x_1, \dots, x_{n-1})$  tal que os  $x_i$  sejam todos inteiros não negativos e  $\sum_{i=1}^{n-1} x_i = m - x_n$ . Então,

$$\begin{aligned} f(x_n) &= \sum_{(x_1, \dots, x_{n-1}) \in B} \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n} \\ f(x_n) &= \sum_{(x_1, \dots, x_{n-1}) \in B} \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n} \frac{(m - x_n)! (1 - p_n)^{n - x_n}}{(m - x_n)! (1 - p_n)^{n - x_n}} \\ f(x_n) &= \frac{m!}{x_1 (m - x_n)!} p_n^{x_n} (1 \\ &\quad - p_n)^{m - x_n} \sum_{(x_1, \dots, x_{n-1}) \in B} \frac{(m - x_n)!}{x_1! \dots x_{n-1}!} \left( \frac{p_1}{1 - p_n} \right)^{x_1} \dots \left( \frac{p_{n-1}}{1 - p_n} \right)^{x_{n-1}} \end{aligned}$$

Usando o fato de que  $(x_1, \dots, x_{n-1}) = m - x_n$  e  $(p_1, \dots, p_{n-1}) = 1 - p_n$ , e ainda o teorema multinomial, pode-se ver que a somatória acima é igual a 1. Portanto, demonstra-se que a distribuição marginal de  $\mathbf{X}_n$  é uma Binomial  $(n, p_n)$ .

Assim, os equacionamentos dos parâmetros de amostragem (esperança e variância) utilizados para amostragem de questões dicotômicas são válidos para a amostragem de questões politômicas. Em função da equivalência demonstrada anteriormente, a esperança de  $x_i$  é  $np_i$  e a sua variância é  $np_i(1 - p_i)$ , que são equivalentes ao caso binomial.

A amostragem multinomial nada mais é que uma generalização da binomial para mais de duas proporções. Situações que podem ser modeladas pela probabilidade por questões de

múltipla escolha, sejam de resposta única ou múltipla, escala de Likert, escala numérica, etc. (SILVA, 2012).

A distribuição *a priori* conjugada da distribuição multinomial é a generalização multivariada da distribuição beta, conjugada da distribuição binomial, conhecida por distribuição de Dirichlet (GELMAN et al., 2013).

O modelo multinomial-Dirichlet é descrito pela seguinte forma:

Considere-se  $f_{ij}$  o número total de observações em unidades  $i = 1, \dots, I$  que por sua vez podem pertencer a uma das  $j = 1, \dots, J$  categorias, ordenadas ou não. Seja  $N_i = \sum_j f_{ij}$  o número total de observações na unidade  $i$ . O modelo mais simples assume que os dados  $f_i = \{f_{i1}, f_{i2}, \dots, f_{ij}\}$  formam distribuições multinomiais independentes de dimensão  $J$ :

$$f_i \sim \text{Multi}_J(N_i, \pi_i)$$

onde, os parâmetros de probabilidade são vetores  $\pi_i = (\pi_{i1}, \dots, \pi_{ij})'$  que modelam as frequências relativas em cada unidade  $i$  sujeitas a que  $\pi_{ij} = 1 - \sum_{j=1}^{J-1} \pi_{ij}$ .

A distribuição *a priori* conjugada consiste na distribuição de Dirichlet que é aplicada a cada conjunto de probabilidades  $I$ . Consequentemente, para cada unidade  $i$ ,  $f_i$  segue uma distribuição Multinomial, tal que:

$$p(f_{i1}, f_{i2}, \dots, f_{ij} | \pi_i) = \frac{N_i!}{f_{i1}! f_{i2}! \dots f_{ij}!} \pi_{i1}^{f_{i1}} \dots \pi_{ij}^{f_{ij}} \quad (31)$$

onde, por sua vez os vetores de probabilidade  $\pi_i$  seguem uma distribuição de Dirichlet com parâmetros  $a_{ij}$ :

$$p(\pi_{ij} | a_{ij}) \propto \prod_{j=1}^J \pi_j^{a_j-1}, \quad (32)$$

ou simplesmente:

$$\pi_i \sim \text{Dirichlet}(a_{i1}, a_{i2}, \dots, a_{ij}).$$

A distribuição *a priori* é matematicamente equivalente à verossimilhança que resulta do somatório de  $\sum_{j=1}^J a_{ij}$  observações com  $a_{ij}$  observações do  $j$ -ésimo resultado categórico (GELMAN et al., 2013). Existem várias distribuições Dirichlet não informativas que se podem empregar na modelação. Obtém-se uma densidade Uniforme quando se assume  $a_{ij} = 1$  para todo o  $i$  e  $j$ . A distribuição Uniforme atribui uma densidade igual a qualquer vetor  $\pi_i$  que obedeça a  $\sum_{j=1}^J \pi_{ij} = 1$ .

Por intermédio do teorema de Bayes (GELMAN et al.; 2013; AZEREDO LOPES; CARDOSO, 2009) as equações  $p(f_{ij}|\pi_i)$  e  $p(\pi_{ij}|a_{ij})$  produzem uma distribuição *a posteriori* Dirichlet para  $\pi_i$  com parâmetros  $a_{ij} + f_{ij}$  com:

$$\text{Valores médios:} \quad P_{ij} = E(\pi_{ij}|f_i) = \frac{a_{ij} + f_{ij}}{\sum_j (a_{ij} + f_{ij})} \quad (33)$$

$$\text{Variâncias:} \quad V(\pi_{ij}|f_i) = \frac{P_{ij}(1-P_{ij})}{[1 + \sum_j (a_{ij} + f_{ij})]} \quad (34)$$

O modelo Multinomial Dirichlet pode ser descrito sucintamente pelas seguintes equações:

$$f_i | \pi_i \sim M_J(N_i, \pi_i) \quad (35)$$

$$\pi_i | a \sim \text{Dir}(a_{i1}, a_{i2}, \dots, a_{ij}), \quad (36)$$

em que,  $a_{i1} = a_{i2} = \dots = a_{ij} = 1$ .

### 3 METODOLOGIA

#### 3.1 Obtenção dos dados

Os dados do teste de tetrazólio foram obtidos no Laboratório Central de Sementes do Departamento de Agricultura da Universidade Federal de Lavras, a partir de avaliação de sementes de soja de produtores da região do Sul de Minas Gerais – Brasil, localizada no município de Lavras/MG, a 21° 14' 43 de latitude sul e 44° 59' 59 de longitude oeste com altitude de 919 metros acima do nível do mar.

Para validação do procedimento proposto neste trabalho, foi acompanhada a análise tradicional do teste. Entre diversas análises realizadas, foram registrados os resultados de, doze lotes analisados com duas repetições de 50 sementes, totalizando 100 sementes por lote, conforme normas estabelecidas nas Regras de Análise de Sementes (BRASIL, 2009).

As sementes foram pré-condicionadas em papel toalha umedecido, permanecendo por 16 h no germinador a 25 °C. Após esse período, as sementes foram imersas em solução de sal 2,3,5 cloretos de trifênil tetrazólio e acondicionadas em câmaras incubadoras a 40 °C, por 3 h. Nessa avaliação foram retirados os tegumentos e, posteriormente com auxílio de um bisturi, as sementes foram seccionadas longitudinalmente através do centro do eixo embrionário. Em seguida os cotilédones foram separados deixando exposta a região vascular para avaliação de todos os tipos de danos (FRANÇA-NETO; KRZYZANOWSKI; COSTA, 1998). Para esta variável, as sementes recém-seccionadas foram colocadas em bandejas de plástico cobertas por uma folha de papel para germinação e avaliadas com auxílio de lupa 16 de mesa (4x) aumentando a luminosidade para a leitura.

No momento da avaliação, as características de intensidade de cor são de suma importância, permitindo a diferenciação para a classificação das sementes em relação ao vigor e viabilidade, sementes com coloração vermelho carmin apresentam tecidos vivos e vigorosos, sementes de coloração vermelho carmin forte demonstram tecidos em deterioração e sementes com cor branco leitoso apresentam tecidos mortos. Porém, vale destacar, que as partes internas dos cotilédones se apresentam descoloridas, portanto, de cor branca (FRANÇA-NETO et al., 1988).

### 3.2 Análise descritiva

Durante a avaliação de cada semente, registrou-se em ficha o tipo de dano que possa ter determinado o vigor das sementes, na Figura 7 há um modelo da ficha de avaliação do teste de tetrazólio com os resultados de níveis de vigor e as simbologias dos tipos de danos. As diferentes simbologias utilizadas na ficha de avaliação do teste de tetrazólio estão apresentadas na Figura 8.

Figura 7 - Modelo de ficha com os resultados do teste de tetrazólio em sementes de soja anotados e interpretados para um lote de vigor médio

Embrapa		FICHA DE AVALIAÇÃO DO TESTE DE TETRAZÓLIO																																																																																																																																																																																											
AMOSTRA: LOTE Nº 2				LOCAL: LONDRINA, PE																																																																																																																																																																																									
Nº DE SEMENTES TESTADAS: 2 x 50				CONC. DA SOLUÇÃO: 0,025%																																																																																																																																																																																									
DATA: 15/01/98				ANALISTA: YILMA																																																																																																																																																																																									
<table border="0"> <tr> <td>1.</td> <td>XXXXX</td> <td>     </td> <td>12</td> </tr> <tr> <td>2.</td> <td>EP LPP</td> <td>PL LPP</td> <td>PP P</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>28</td> </tr> <tr> <td>3.</td> <td>PEPEP</td> <td>PPPP</td> <td>EPPI</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>26</td> </tr> <tr> <td>4.</td> <td>PEPEP</td> <td>PI</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>12</td> </tr> <tr> <td>5.</td> <td>EPPI</td> <td>     </td> <td>6</td> </tr> <tr> <td>S.D.</td> <td>     </td> <td>0</td> </tr> <tr> <td colspan="10">Nível de vigor 1-3: 66 Viabilidade: 84</td> </tr> <tr> <td>1.</td> <td>XXXXX</td> <td>     </td> <td>8</td> </tr> <tr> <td>2.</td> <td>EPPEP</td> <td>PEPPP</td> <td>PPPI</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>24</td> </tr> <tr> <td>3.</td> <td>EP PPL</td> <td>PPPPP</td> <td>PPPPY</td> <td>PI</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>32</td> </tr> <tr> <td>4.</td> <td>PE P P P</td> <td>U</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>12</td> </tr> <tr> <td>5.</td> <td>EPPI</td> <td>     </td> <td>4</td> </tr> <tr> <td>S.D.</td> <td>     </td> <td>0</td> </tr> <tr> <td colspan="10">Nível de vigor 1-3: 64 Viabilidade: 80</td> </tr> <tr> <td>1.</td> <td>PL</td> <td>     </td> <td>4</td> </tr> <tr> <td>2.</td> <td>LE P P</td> <td>LE L</td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>     </td> <td>16</td> </tr> <tr> <td>3.</td> <td>     </td> <td>0</td> </tr> <tr> <td colspan="10">Nível de vigor 1-3: 64 Viabilidade: 80</td> </tr> </table>										1.	XXXXX								12	2.	EP LPP	PL LPP	PP P						28	3.	PEPEP	PPPP	EPPI						26	4.	PEPEP	PI							12	5.	EPPI								6	S.D.									0	Nível de vigor 1-3: 66 Viabilidade: 84										1.	XXXXX								8	2.	EPPEP	PEPPP	PPPI						24	3.	EP PPL	PPPPP	PPPPY	PI					32	4.	PE P P P	U							12	5.	EPPI								4	S.D.									0	Nível de vigor 1-3: 64 Viabilidade: 80										1.	PL								4	2.	LE P P	LE L							16	3.									0	Nível de vigor 1-3: 64 Viabilidade: 80									
1.	XXXXX								12																																																																																																																																																																																				
2.	EP LPP	PL LPP	PP P						28																																																																																																																																																																																				
3.	PEPEP	PPPP	EPPI						26																																																																																																																																																																																				
4.	PEPEP	PI							12																																																																																																																																																																																				
5.	EPPI								6																																																																																																																																																																																				
S.D.									0																																																																																																																																																																																				
Nível de vigor 1-3: 66 Viabilidade: 84																																																																																																																																																																																													
1.	XXXXX								8																																																																																																																																																																																				
2.	EPPEP	PEPPP	PPPI						24																																																																																																																																																																																				
3.	EP PPL	PPPPP	PPPPY	PI					32																																																																																																																																																																																				
4.	PE P P P	U							12																																																																																																																																																																																				
5.	EPPI								4																																																																																																																																																																																				
S.D.									0																																																																																																																																																																																				
Nível de vigor 1-3: 64 Viabilidade: 80																																																																																																																																																																																													
1.	PL								4																																																																																																																																																																																				
2.	LE P P	LE L							16																																																																																																																																																																																				
3.									0																																																																																																																																																																																				
Nível de vigor 1-3: 64 Viabilidade: 80																																																																																																																																																																																													

REPETIÇÕES	DANO MECÂNICO		UMIDADE		PERCEVEJO		DURAS	VIGOR	VIAB
	1 - 8	6 - 8	1 - 8	6 - 8	1 - 8	6 - 8			
I	32	8	76	6	12	4	0	66	84
II	28	12	76	4	14	4	0	64	80
MÉDIAS	30	10	76	5	13	4	0	65	82

Fonte: KRZYZANOWSKI, 1999.

Figura 8 - Simbologia utilizada na ficha de avaliação do teste de tetrazólio

$X$  : sem dano  
 $L$  : dano mecânico (DM)  
 $\mathcal{L}$  : dano causado por percevejo (DP)  
 $P$  : dano por umidade (DU)  
 $\mathcal{L}$  : DM + DP  
 $\mathcal{P}$  : DM + DU  
 $\mathcal{P}$  : DU + DP  
 $\mathcal{E}$  : DU + DM + DP

Fonte: KRZYZANOWSKI, 1999.

A partir dos dados registrados, foi avaliado o percentual de sementes vigorosas sob três abordagens:

- (i) percentual de sementes vigorosas ( $p$ ) e não vigorosas, seguindo um modelo Binomial;
- (ii) classificação das sementes em três categorias, considerando apenas as sementes não vigorosas e avaliando a porcentagem dos três tipos de danos, umidade, percevejo e mecânico, sendo que as sementes que apresentassem dois ou mais danos foram incluídas na categoria do dano que levou a morte dos embriões (Tabela 1).

Tabela 1 - Classificação das sementes em três classes.

$\hat{p}$	Classificação das sementes
$\hat{p}_1$	Presença de dano por umidade
$\hat{p}_2$	Presença de dano por percevejo
$\hat{p}_3$	Presença de dano mecânico

Fonte: Elaborado pelo autor.

- (iii) classificação de cinco classes considerando as sementes vigorosas, com um dos três tipos de danos e as sementes com dois ou mais danos (Tabela 2).

Tabela 2 - Classificação das sementes em cinco classes.

$\hat{p}$	Classificação das sementes
$\hat{p}_1$	Vigorosa
$\hat{p}_2$	Presença de dano por umidade
$\hat{p}_3$	Presença de dano por percevejo
$\hat{p}_4$	Presença de dano mecânico
$\hat{p}_5$	Presença de 2 ou mais danos

Fonte: Elaborado pelo autor.

### 3.3 Estimação frequentista

A análise frequentista do teste de tetrazólio foi realizada a partir da distribuição binomial a proporção de sementes vigorosas e não vigorosas, e com a distribuição multinomial a proporção de sementes conforme classificação com três (Tabela 1) e com cinco (Tabela 2) categorias.

Com as proporções observadas e estimadas foi verificado o ajustamento do conjunto de dados entre os 12 lotes com o teste qui-quadrado de Pearson (LEHMANN, 1959), dado por:

$$\chi^2 = \sum_{i=1}^i \sum_{j=1}^j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad (37)$$

em que  $O_{ij}$  e  $E_{ij}$  são, respectivamente, as frequências absolutas observadas e esperadas da  $i$ -ésima linha e da  $j$ -ésima coluna. Quando as proporções esperadas são muito próximas das observadas, o valor de  $\chi^2$  é baixo. No entanto, quando a diferença entre elas é alta há indicação de dependência entre as variáveis. Desta forma, o teste qui-quadrado confronta a hipótese nula de que as variáveis são independentes contra a hipótese alternativa de que as variáveis são dependentes, ou seja:

$$\begin{cases} H_0: \text{independência entre as variáveis} \\ H_1: \text{dependência entre as variáveis} \end{cases}$$

Sendo que, para uma tabela com  $r$  linhas e  $c$  colunas, o número de graus de liberdade para o teste do qui-quadrado é  $(r - 1)(c - 1)$ .

### 3.4 Estimação bayesiana

Na realização do teste tetrazólio, cada semente avaliada poderá ser classificada em vigorosa e não vigorosa, caracterizando um experimento de Bernoulli. Considerando,  $p$  o parâmetro de um experimento de Bernoulli, tal como a classificação em vigorosa e não vigorosa de uma semente, então a verossimilhança para a soma de ensaios de Bernoulli,  $n$ , em um lote, correspondente a uma Binomial  $K \sim Bin(n, p)$  e pode ser escrita como:

$$f(K|p) = \binom{n}{K} p^K (1 - p)^{n-K}, \quad 0 \leq p \leq 1 \quad (38)$$

Para variáveis aleatórias contínuas com domínio entre 0 e 1, tal como as proporções e porcentagens é frequentemente usada, em estatística bayesiana, como a distribuição *a priori* conjugada de probabilidades binomiais a distribuição Beta (GELMAN et al., 2013).

Conhecendo a média ( $\mu$ ) e a variância ( $\sigma^2$ ) para a distribuição *a priori* Beta ( $a, b$ ), então as estimativas de  $a$  e  $b$  pelo método dos momentos pode ser calculada da seguinte maneira:

Sabendo que:  $ab = (a + b)^2(a + b + 1)\sigma^2$ , e que  $b = \frac{a}{\mu} - a$ , tem-se que:

$$\frac{a^2}{\mu} - a^2 = \left(a + \frac{a}{\mu} - a\right)^2 \left(a + \frac{a}{\mu} - a + 1\right) \sigma^2$$

$$\frac{a^2}{\mu} - a^2 = \left(\frac{a}{\mu}\right)^2 \left(\frac{a}{\mu} + 1\right) \sigma^2$$

$$a^2 \left(\frac{1}{\mu} - 1\right) = a^2 \left(\frac{1}{\mu^2}\right) \left(\frac{a}{\mu} + 1\right) \sigma^2$$

$$\left(\frac{1}{\mu} - 1\right) \left(\frac{\mu^2}{\sigma^2}\right) = \left(\frac{a}{\mu} + 1\right)$$

$$\left(\frac{\mu(1 - \mu)}{\sigma^2}\right) = \frac{a}{\mu} + 1$$

$$a = \mu \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1\right)$$

e para o parâmetro  $b$  se tem:

$$b = a \left(\frac{1 - \mu}{\mu}\right)$$

$$b = \mu \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1\right) \left(\frac{1 - \mu}{\mu}\right)$$

$$b = (1 - \mu) \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1\right)$$

Deste modo temos a equação geral para elicitar os valores de  $a$  e  $b$  iniciais:

$$a = \mu \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1\right), \quad (39)$$

$$b = (1 - \mu) \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1\right) \quad (40)$$

O procedimento para construção das *prioris* se inicia com o fornecimento de valores elicitados de média (proporção de danos) e variância (variação dos danos) dados por:

$$\mu_{priori} = \mu_{beta} = \frac{a}{(a + b)}, \quad (41) \quad var_{priori} = \sigma^2_{beta} = \frac{(ab)}{(a + b)^2(a + b + 1)} \quad (42)$$

Neste trabalho utilizou-se duas *prioris* com valores de hiperparâmetros obtidos através da média e variância da proporção de sementes vigorosas apresentadas nos trabalhos de Schuab et al. (2007) e Lopes et al. (2011).

Para obter a distribuição *a posteriori* Beta ( $a', b'$ ) ou  $(a + q, b + n - k)$ , onde  $k$  é a soma de sementes não vigorosas, a média e a variância são dadas por:

$$\mu_{posteriori} = \frac{a + k}{(a + k + b + n - k)} \quad (43)$$

$$var_{posteriori} = \frac{(a + k)(b + n - k)}{(a + b + n)^2(a + b + n + 1)} \quad (44)$$

No caso da análise de danos das sementes de soja, tanto para o caso exposto na Tabela 1 quanto para a Tabela 2, para uso da distribuição Multinomial, a distribuição *a priori* conjugada será a generalização multivariada da distribuição Beta, conhecida por distribuição de Dirichlet (GELMAN et al., 2013).

A distribuição de Dirichlet é aplicada a cada conjunto de probabilidades  $k$ , onde os vetores de probabilidade  $p_k$  seguem uma distribuição de Dirichlet com parâmetros  $a_k$ :

$$P(p_k | a_k) \propto \prod_{k=1}^k p_k^{a_k - 1} \quad (45)$$

Da mesma forma, o procedimento para construção da distribuição *a priori* Dirichlet se inicia com o fornecimento de valores elicitados de média (proporção) e variância (variação) para cada tipo de dano dadas por:

$$\mu_{priori} = \mu_{Dir} = \frac{a_k}{\sum a_k}, \quad (46) \quad var_{priori} = \frac{a_k(a_0 - a_k)}{a_0^2(a_0 + 1)}, \quad (47)$$

onde  $a_0 = a_1 + \dots + a_k$ , de forma que quanto maior o valor de  $a$ , menor é a variância.

Por intermédio do teorema de Bayes (GELMAN et al., 2013; AZEREDO LOPES; CARDOSO, 2009) tem-se a distribuição *a posteriori* Dirichlet para  $p_k$  com parâmetros  $a_k + n_k$  com:

$$\mu_{posteriori} = \frac{a_k + n_k}{\sum_k (a_k + n_k)} \quad (48) \quad var_{posteriori} = \sigma^2_{Dir} = \frac{p_k(1 - p_k)}{[1 + \sum_k (a_k + n_k)]} \quad (49)$$

Para construção das *prioris* neste trabalho, utilizou-se valores de média e variância das proporções de danos em sementes apresentadas no trabalho de Costa et al. (2005) e nos dados da safra da soja de 2016/2017 (LORINI, 2017).

### 3.5 Estimação sequencial

Para definir as retas de aceitação e rejeição do método sequencial foram utilizados os critérios de NQA (Nível de Qualidade Aceitável) e NQR (Nível de Qualidade Rejeitável). Sabendo que o NQA é a taxa de defeitos mais elevada ou taxa de defeitos de um processo que seja considerada aceitável, descrevendo o que o plano de amostragem aceitará. E o NQR é a maior taxa de defeito ou com defeito que o consumidor está disposto a tolerar em um lote individual, descrevendo o que o plano de amostragem rejeitará. Utilizou-se dados do comércio de sementes para definir os planos de amostragem.

No Brasil, há empresas que somente colocam no mercado lotes de sementes com germinação superior a 90%, logo este será o NQA adotado para viabilidade, e para que o lote seja comercializado deve ter um mínimo de 80% de viabilidade segundo a Instrução Normativa nº 45, de 17 de setembro de 2013 (BRASIL, 2013), sendo este adotado como NQR.

Pelo quesito de vigor a Abrates - Associação Brasileira de Tecnologia de Sementes - no comitê de vigor de sementes interpreta o nível do resultado do teste através da classificação da Tabela 3 (KRZYZANOWSKI, 1999).

Tabela 3 - Valores de referência do vigor das sementes.

Níveis	
Muito alto	$\geq 85\%$
Alto	84% - 75%
Médio	74% - 60%
Baixo	59% - 50%
Muito baixo	$\leq 49\%$

Fonte: KRZYZANOWSKI, 1999.

Foram aplicadas duas proporções diferentes a fim de comparação. A primeira de 60% como NQA e 49% como NQR e a segunda de 75% de NQA e 60% de NQR, pois assim foram aceitas somente as sementes que apresentam vigor médio para cima na primeira proporção, e na segunda, vigor alto para muito alto. E para os erros  $\alpha$  e  $\beta$  foram adotados, respectivamente, os níveis de 5% e 10%.

### 3.6 Estimação sequencial bayesiana

Para estimar o parâmetro de proporção considerando o tamanho da amostra como variável aleatória na estimação sequencial, extrai-se elementos de uma amostra um por vez, e depois que cada sementes é observada, utiliza-se um critério de parada e é tomada a decisão (rejeição/aceitação de uma hipótese ou estimação de um parâmetro) ou a observação de amostras é continuada.

Para a obtenção do critério de parada na distribuição Binomial, foi considerada a função perda quadrática  $L = (p - \hat{p})^2 + C(n)$ , em que  $C(n) > 0$  é o custo de tomar uma amostra de tamanho  $n$  para estimar  $p$ . Foi calculada a esperança da variância *a posteriori* a qual permite estimar a quantidade  $r(\pi^n, n)$  que representa o menor risco de Bayes que pode ser atingido uma vez que  $\mathbf{X}^n$  foi observado.

Para decidir se deve ou não parar a amostragem, os valores de risco são comparados a cada semente avaliada e realizada nova observação até que  $r_o(\pi^n, n) \leq r^1(\pi^n, n)$ , quando a estimativa do parâmetro da proporção de viabilidade no lote testado é então calculada a partir da média da *posteriori* obtida até então.

Assim, foi necessário calcular, a cada avaliação de semente, o risco imediato que é dado pela variância *a posteriori* acrescida do custo de  $n$  observações e o risco esperado dado pela esperança da variância *a posteriori* e acréscimo no custo de mais uma observação (BRIGHENTI et al., 2011), assim:

$$r_o(\pi^n, n) = \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} + C(n) \quad (50)$$

$$r^1(\pi^n, n) = E[\text{var}_{\text{posteriori}}] + C(n + 1) \quad (51)$$

Então, o risco esperado é dado por:

$$r^1(\pi^n, n) = \left( \frac{\alpha + \beta + n}{\alpha + \beta + n + 1} \right) \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} + C(n + 1) \quad (52)$$

O procedimento para construção das *prioris*, assim como obtenção das estimativas dos hiperparâmetros foi desenvolvido no *software* R.

Para a distribuição multinomial com  $(k + 1)$  classes, de acordo com Jones (1974), a probabilidade de uma observação na  $i$ -ésima classe de  $p_i = 1, 2, \dots, k$ , e na  $(k + 1)$  pode ser  $(1 - \sum_{i=1}^k p_i)$ . Suponha que a informação *a priori* sobre  $p = (p_1, p_2, \dots, p_k)^T$

pode ser adequadamente representada por um membro do conjugado natural Dirichlet, família de distribuições com parâmetros inteiros  $m_0, n_{0i}, i = 1, 2, \dots, k$ , com densidade proporcional a:

$$\prod_{i=1}^k p_i^{n_{0i}-1} \left(1 - \sum_{i=1}^k p_i\right)^{m_0 - \sum_{i=1}^k n_{0i} - 1},$$

com  $p_i \geq 0$  e  $\sum_{i=1}^k p_i \leq 1$ .

Pelo teorema de Bayes, após  $m$  observações resultando em  $n_i$  nas  $i$ -ésimas classes, a densidade *posteriori* de  $\mathbf{p}$  será dada por uma Dirichlet, com parâmetros  $m + m_0, n_i + n_{0i}$ .

O resultado da amostragem por ser representada como um caminho de amostra, começa no ponto  $(n_0, m_0), n_0 = (n_{01}, n_{02}, \dots, n_{0k})$ , no espaço inteiro dimensional de  $(k + 1)$  e é interrompido quando o limite de parada, que tem que ser determinado, é atingido. Se  $m_0 = k + 1$  e  $n_{0i} = 1$  (*priori* uniforme para  $\mathbf{p}$ ) são tomados como a origem, então qualquer outra *priori* apropriada com parâmetros inteiros dará caminhos de amostra começando como  $(n_0 - 1, m_0 - k - 1)$ , sendo  $\mathbf{1}$  o vetor de linha unitária.

Consequentemente, *a priori* uniforme será usada para obter limites de parada. Suponha que a perda na estimativa de  $\mathbf{p}$  por  $\mathbf{d} = (d_1, d_2, \dots, d_k)^T$  tenha a forma quadrática geral  $(\mathbf{p} - \mathbf{d})^T \mathbf{K} (\mathbf{p} - \mathbf{d})$ , onde  $\mathbf{K}$  é uma matriz simétrica positiva de perda constante, em seguida, que o estimador de Bayes  $\mathbf{d}^*$  é a média da distribuição *posteriori* de  $(\mathbf{n}, m)$  com elementos  $d_i^* = (n_i + 1)/(m + k)$ , esta é também a probabilidade marginal *posteriori* que a próxima observação cai dentro da  $i$ -ésima classe. O risco de parada é dado:

$$S(\mathbf{n}, m) = \text{traço } \mathbf{K}\boldsymbol{\Sigma}, \quad (54)$$

onde  $\boldsymbol{\Sigma}$  é a matriz dispersão da distribuição *posteriori* com os elementos:

$$\text{var}(p_i) = \frac{d_i^*(1-d_i^*)}{m+k+2}, \quad \text{cov}(p_i, p_j) = -\frac{d_i^* d_j^*}{m+k+2}. \quad (55)$$

Consequentemente,

$$S(\mathbf{n}, m) = \frac{\{\sum_{i=1}^k K_{ii} d_i^* - \sum_{i,j=1}^k K_{ij} d_i^* d_j^*\}}{m + k + 2}. \quad (56)$$

O resultado geral para a função perda quadrática é dado em Owen (1970) e a média do vetor e a matriz dispersão da distribuição de Dirichlet é dado por Wilks (1950).

Para a obtenção do critério de parada na distribuição Multinomial, considerou-se um ponto  $(\mathbf{n}, m)$ , onde se tem  $B(\mathbf{n}, m)$  o risco de uma observação adicional a um custo  $c$  e prosseguir daí em diante (risco de continuação), e seja  $D(\mathbf{n}, m)$  o risco mínimo, então as equações de programação dinâmica dando a partição em pontos de parada e continuação são:

$$D(\mathbf{n}, m) = \min [S(\mathbf{n}, m), B(\mathbf{n}, m)], \quad (57)$$

$$B(\mathbf{n}, m) = c + \sum_{i=1}^k [D(\mathbf{n} + e_i, m + 1)d_i^*] + D(\mathbf{n}, m + 1) \left( 1 - \sum_{i=1}^k d_i^* \right), \quad (58)$$

$e_i$  é a linha do vetor de 1 até a  $i$ -ésima posição e zeros em outro lugar.

Estas equações são semelhantes àquelas em Freeman (1972), Jones (1974) e Lindley e Barnett (1965).

Desde  $S(\mathbf{n}, m) \rightarrow 0$ ,  $B(\mathbf{n}, m) \rightarrow c$  e  $m \rightarrow \infty$  haverá um grande valor de  $m = N^*$ , de tal forma que todos os pontos  $(\mathbf{n}, N^*)$  são pontos de parada para  $\sum_{i=1}^k n_i \leq N^*$ . As equações de programação dinâmica acima podem agora ser usadas sucessivamente para  $m \leq N^*$ , para encontrar o menor inteiro  $m$  satisfazendo:

$$B(\mathbf{n}, m) > S(\mathbf{n}, m), \quad D(\mathbf{n}, m + 1) = S(\mathbf{n}, m + 1) \text{ para todo } \mathbf{n}, \quad (59)$$

este será o tamanho máximo da amostra.

As duas condições implicam:

$$B(\mathbf{n}, m) = c + \left[ \frac{m + k + 1}{m + k + 2} \right] S(\mathbf{n}, m) \quad (60)$$

e a desigualdade reduz a:

$$(m + k + 2)^2 > \max \frac{\{\sum_{i=1}^k K_{ii} d_i^* - \sum_{i=1}^k K_i d_i^* d_j^*\}}{c}. \quad (61)$$

## 4 RESULTADOS

### 4.1 Estimação do vigor de sementes de soja utilizando a distribuição Binomial

#### 4.1.1 Abordagem frequentista da estimação do vigor

Na abordagem frequentista, a estimação da proporção de sementes vigorosas é feita utilizando-se a distribuição binomial com classificação de sementes vigorosas e não vigorosas, como definido anteriormente as sementes vigorosas são aquelas que não apresentam nenhum tipo de dano e as não vigorosas são aquelas que apresentam alguma lesão, a qual pode ter afetado o embrião ou não. Particularmente, aquelas que apresentam danos no embrião se tornam impróprias para comercialização, pois não germinarão e são classificadas como inviáveis.

Pela Tabela 4, pode-se afirmar que a média de vigor nos lotes avaliados foi de 73,42% com desvio-padrão de 10,54%. Para constatar as relevâncias das classificações foi realizado o teste qui-quadrado, o qual apresentou a estatística de teste igual a 62,66 com 11 graus de liberdade e valor-p < 0,001. Evidenciando que há diferenças entre as porcentagens de sementes vigorosas e não vigorosas entre os lotes.

Tabela 4 - Porcentagem das sementes vigorosas e não vigorosas.

Lote	1	2	3	4	5	6	7	8	9	10	11	12	Média
Vigorosas	81	54	64	62	80	78	84	75	83	60	84	76	73,42
Não Vigorosas	19	46	36	38	20	22	16	25	17	40	16	24	26,58

Fonte: Elaborado pelo autor.

Os lotes 7 e 11 foram os que apresentaram maior taxa de vigor sendo esta igual a 84%, enquanto que o lote 2 obteve o menor vigor com taxa igual a 54%.

#### 4.1.2 Abordagem bayesiana da estimação do vigor

Para estimar a proporção de sementes vigorosas dos lotes com a abordagem bayesiana, usou-se uma *priori* conjugada Beta ( $a, b$ ), sabendo que após observar a amostra a distribuição *a posteriori* é Beta ( $a + k, b + n - k$ ) onde  $k$  é o total de sementes não vigorosas.

A elicitação das duas *prioris* utilizadas foi baseada nas seguintes informações: *priori* A obtida pelos resultados apresentados por Schuab et al. (2007) ao estudarem o vigor de sementes de soja durante teste de germinação. No trabalho ele registra média de 71% de sementes vigorosas e um coeficiente de variação (CV) de 7,22. Na *priori* B os hiperparâmetros foram obtidos do trabalho de Lopes et al. (2011) que apresenta 83% de sementes vigorosas com CV de 8,65%.

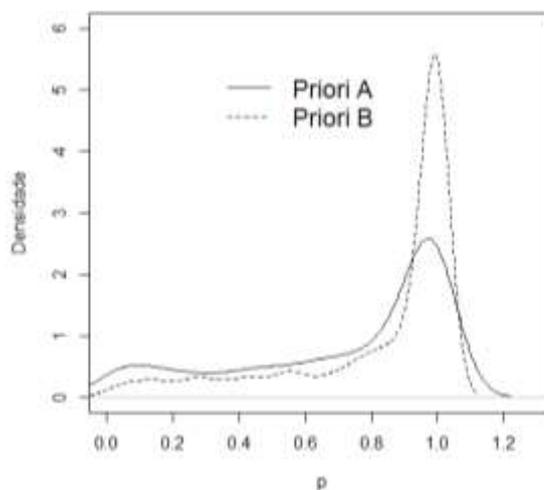
A partir das referências citadas foram elicitados os hiperparâmetros das *prioris* utilizando as expressões (39) e (40) que estão apresentadas na Tabela 5.

Tabela 5 - Valores dos hiperparâmetros, média (%) e variâncias (%) das *prioris*.

Priori	Média (%)	Variância (%)	Hiperparâmetros	
			$a$	$b$
A- Schuab et al. (2007)	71,00	10,34	0,7044	0,2877
B- Lopes et al. (2011)	83,00	7,06	0,8277	0,1695

Fonte: Elaborado pelo autor.

Com os hiperparâmetros das *prioris* obtidos, foram plotados os gráficos das duas densidades na Figura 9. Eles evidenciam a maior dispersão da *priori* A.

Figura 9 - Densidades das distribuições *a priori* Beta

Fonte: Elaborado pelo autor.

Na Tabela 6 são apresentados os resultados das distribuições *a posteriori* para cada lote. Observa-se que com o tamanho  $n = 100$ , e o uso de *priori* diferentes se obtêm resultados semelhantes na *posteriori*, indicando uma consistência nas análises. Apesar disto, percebe-se que os valores de média estimados com a *priori* B são sempre inferiores, sendo mais distantes dos valores obtidos pela abordagem frequentista apresentados na Tabela 4, evidenciando a influência da *priori* com menor variância. No entanto, a variância é sempre maior.

Tabela 6 - Porcentagens das médias e variâncias estimadas a partir das *posteriors* dos parâmetros de proporção de sementes vigorosas.

Lote	<i>Posteriori A</i> - Schuab et al. (2007)		<i>Posteriori B</i> - Lopes et al. (2011)	
	Média (%)	Variância (%)	Média (%)	Variância (%)
1	0,804892	0,001515	0,803681	0,006384
2	0,537544	0,002434	0,536347	0,002855
3	0,636562	0,002257	0,635359	0,003998
4	0,616758	0,002308	0,615557	0,003754
5	0,794990	0,001574	0,793779	0,006228
6	0,775186	0,001686	0,773977	0,005923
7	0,834597	0,001326	0,833384	0,006863
8	0,745481	0,001840	0,744273	0,005479
9	0,824695	0,001391	0,823483	0,006701
10	0,596955	0,002351	0,595754	0,003518
11	0,834597	0,001326	0,833384	0,006863
12	0,755383	0,001791	0,754174	0,005625

Fonte: Elaborado pelo autor.

Na abordagem bayesiana, já que é obtida uma distribuição para o parâmetro, é possível estabelecer um intervalo de credibilidade para o mesmo, sendo este calculado a partir dos percentis calculados na distribuição *a posteriori* com parâmetros conforme equações (28) e (29). Os resultados para os Intervalos de Credibilidade a 95% são apresentados na

Tabela 7.

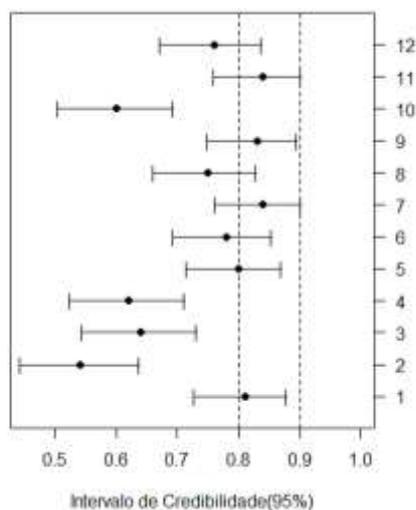
Nota-se que todos os intervalos de credibilidade calculados contêm a estimativa obtida pela abordagem frequentista. Isto indica que, nenhuma das *prioris* elicítadas teve forte influência nos resultados (Figura 10).

Tabela 7 - Intervalos de Credibilidade a 95% obtidos a partir das distribuições a *posteriori* dos parâmetros de proporção de sementes vigorosas.

Lote	Estimativa frequentista	Intervalos de Credibilidade a 95%			
		<i>Posteriori A</i>		<i>Posteriori B</i>	
		LI	LS	LI	LS
1	0,81	0,72751	0,87917	0,72863	0,88019
2	0,54	0,44435	0,63750	0,44594	0,63832
3	0,64	0,54515	0,73088	0,54656	0,73205
4	0,62	0,52459	0,71239	0,52596	0,71383
5	0,80	0,71604	0,87089	0,71759	0,87211
6	0,78	0,69397	0,85425	0,69520	0,85532
7	0,84	0,76149	0,90339	0,76274	0,90434
8	0,75	0,66116	0,82859	0,66240	0,82972
9	0,83	0,74992	0,89549	0,75116	0,89660
10	0,60	0,50431	0,69400	0,50569	0,69508
11	0,84	0,76135	0,90344	0,76278	0,90430
12	0,76	0,67215	0,83736	0,67329	0,83831

Fonte: Elaborado pelo autor.

Figura 10 - Intervalos de Credibilidade para os 12 lotes estudados, incluindo a estimativa frequentista (●).



Fonte: Elaborado pelo autor.

### 4.1.3 Estimação sequencial do vigor

Foi realizada a análise sequencial para o vigor, considerando dois limites diferentes de NQA e NQR conforme Tabela 3. Em uma primeira etapa, utilizou-se um critério de NQA de 60% e NQR de 49%, avaliando os lotes para averiguar em qual faixa de vigor se encontram, se médio para alto ou muito baixo. Na segunda etapa, usou-se um critério mais rigoroso com NQA de 75% e NQR de 60%, aceitando lotes apenas com vigor alto ou muito alto. Tais proporções foram comparadas à taxa de vigor clássica, que é a proporção de defeitos pela quantidade de amostras.

Para a construção do plano de amostragem sequencial foram utilizados os erros de produtor ( $\alpha$ ) e de consumidor ( $\beta$ ) de 0,05 e 0,10, respectivamente. Os resultados referentes a vigor dos doze lotes avaliados pelo método sequencial com seus respectivos tamanhos amostrais estão apresentados na Tabela 8.

Tabela 8 - Porcentagens de vigor dos lotes de sementes estimados pelo método sequencial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ).

Lote	$\hat{p}_{clássico}$ (%)	NQA = 60%	NQR = 49%	NQA = 75%	NQR = 60%
	(N=100)	$N_{seq}$	$\hat{p}_{seq}$ (%)	$N_{seq}$	$\hat{p}_{seq}$ (%)
1	81	23	78,26	26	80,77
2	54	>100	<sup>1</sup>	14	35,71 <sup>2</sup>
3	64	80	61,25	32	53,13 <sup>2</sup>
4	62	78	61,53	39	56,41 <sup>2</sup>
5	80	27	74,07	35	77,14
6	78	23	78,26	26	80,77
7	84	12	100,00	11	100,00
8	75	38	68,42	57	73,68
9	83	18	83,33	20	85,00
10	60	20	80,00	32	78,13
11	84	14	92,86	14	92,86
12	76	20	80,00	29	79,31

<sup>1</sup>: Não foi possível sentenciar o lote com as 100 amostras; <sup>2</sup>: Rejeitou-se os lotes.

Fonte: Elaborado pelo autor.

Nota-se que com o NQR a um nível de vigor muito baixo não se rejeitou nenhum lote, já com o nível de vigor médio rejeitou-se três dos doze lotes. No Lote 2 com o NQR a um nível muito baixo, não foi possível tomar uma decisão com as 100 amostras. Agora quando o NQR aumentou para o nível médio de vigor, conseguiu-se rejeitar o lote com 14 amostras. Holtz e Reis (2013) constataram em sementes de soja que o atraso na colheita em campos de produção de sementes afetou significativamente o vigor devido a deterioração das mesmas. Isso pode ter ocorrido com o Lote 2, que apresentou vigor de nível muito baixo. Percebe-se também houve boa aproximação entre as estimativas em todos os lotes, exceto no lote 10, em que a amostragem foi interrompida com apenas 20 sementes avaliadas para no primeiro critério e 32 no segundo critério um pouco mais rigoroso.

A redução do tamanho amostral em relação aos dois conjuntos de proporções de vigor analisados é de extrema importância, uma vez que permite uma diminuição considerável nos custos de avaliação. Ou seja, o avaliador poderá decidir se o lote será ou não aceito em menor tempo do que se considerar o método tradicional com tamanho amostral fixo de 100 sementes avaliadas e assim poderá avaliar maiores quantidades de lotes por hora trabalhada.

#### 4.1.4 Estimação sequencial bayesiana do vigor

Na análise sequencial bayesiana, para estabelecer se a amostragem deve ser interrompida ou não é necessário que os valores de risco sejam comparados a cada semente avaliada e realizada nova observação até que  $r_o(\pi^n, n) \leq r^1(\pi^n, n)$ . Desta forma, torna-se necessário o cálculo, a cada avaliação de semente, do risco imediato que é dado pela variância *a posteriori* acrescida do custo de  $n$  observações e do risco esperado dado pela esperança da variância *a posteriori* e acréscimo no custo de mais uma observação.

Entretanto, para encontrar os valores de risco imediato e esperado conforme equações (50) e (52), deve-se primeiramente encontrar os valores dos hiperparâmetros,  $a$  e  $b$ , iniciais. Para realizar o cálculo foram considerando as mesmas proporções de vigor, isto é, 60 e 75% com variância de 10% (**Erro! Fonte de referência não encontrada.**).

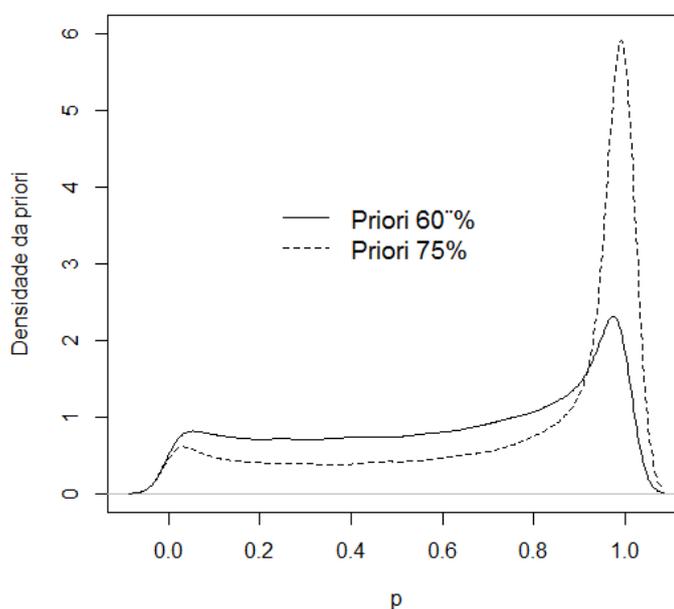
Tabela 9 - Valores dos hiperparâmetros para 60% e 75% de vigor com variância de 10%.

Vigor (%)	Hiperparâmetros	
	$a$	$b$
60	0,8400	0,5600
75	0,6563	0,2188

Fonte: Elaborado pelo autor.

Os hiperparâmetros utilizados não refletem os resultados estimados até aqui e, portanto, geram distribuição com densidades muito diferentes dos lotes avaliados. Percebe-se pela Figura 9 que há maior dispersão e deslocamento da curva de densidade quando comparada com a Figura 11.

Figura 11 - Densidades da distribuição da priori utilizada no caso sequencial bayesiano.



Fonte: Elaborado pelo autor.

Após a obtenção dos hiperparâmetros o próximo passo é o cálculo do custo de mais uma observação através da razão da variância da *priori* sob a esperança do tamanho amostral. O custo adotado foi calculado pela razão entre a variância *priori* sob a esperança do

tamanho amostral ( $E(N)$ ), permitindo que a grandeza do custo tivesse a mesma dimensão do risco, obteve-se como resultado o custo aditivo de  $10^{-4}$  por cada semente avaliada.

Considerando a avaliação da primeira semente do lote 1, sendo esta vigorosa, com um custo de  $10^{-4}$ , e com os valores dos hiperparâmetros do vigor de 60%, o risco imediato ( $r_0$ ) e o risco esperado ( $r^1$ ) são calculados a partir de:

- *a priori* Beta ( $a, b$ ) (0,84; 0,56)
- *a posteriori* ( $a + k, b + n - k$ ) = (0,84 + 1 ; 0,56 + 1 - 1) = (1,84 ; 0,56)

$$r_0(\pi^n, n) = \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} + C(n)$$

$$r_0(\pi^n, n) = \frac{(1,84)(0,56)}{(0,84 + 0,56 + 1)^2(0,84 + 0,56 + 1 + 1)} + 1 * 10^{-4}$$

$$r_0(\pi^n, n) = 0,0670118$$

$$r^1(\pi^n, n) = \left( \frac{\alpha + \beta + n}{\alpha + \beta + n + 1} \right) [Var_{posteriori}] + C(n + 1)$$

$$r^1(\pi^n, n) = \left( \frac{0,84 + 0,56 + 1}{0,84 + 0,56 + 1 + 1} \right) [0,066912] + 1 * 10^{-4}$$

$$r^1(\pi^n, n) = 0,0473318$$

Como  $r_0(0,0670118) > r^1(0,0473318)$ , deve-se continuar a amostragem. Os cálculos dos riscos devem ser realizados até que o risco esperado seja maior que o risco imediato, assim se encerra a amostragem e é realizado o sentenciamento do lote.

A estimação sequencial bayesiana foi realizada através de um aplicativo em Delphi desenvolvido por Brighenti et al. (2011) nos doze lotes. O aplicativo emite um relatório com os resultados e os parâmetros utilizados para chegar no sentenciamento dos lotes (Figura 12). Na Tabela 10 são apresentados os resultados desses relatórios com o resultado individual dos lotes.

Figura 12 - Relatório do programa em Delphi com os resultados e os parâmetros utilizados.

Estimação Sequencial Bayesiana			
Título:	Lote 1	Local:	
Data:	22/08/2019 14:04:02	Proporção Estimada:	36,12 %
Objeto de Estudo:	Soja	Tamanho Final da Amostra:	12
	Proporção:		60,00 %
	Variação:		10,00 %
	Custo :		0,0001000000
	Alfa :		0,8400000000
	Beta :		0,5600000000
	Alfa Linha :		4,8400000000
	Beta Linha :		8,5600000000
	Média a Posteriori:		0,3611940299
	Objetos Observados:		12
	Soma das Presenças:		4
	Risco Imediato (R0):		0,0172231182
	Risco Esperado (R1):		0,0173270684

Fonte: Elaborado pelo autor.

Tabela 10 - Porcentagens de vigor dos lotes de sementes estimados pelo método sequencial bayesiano, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ).

Lotes	Vigor			
	60%		75%	
	$N_{seq}$	$\hat{p}_{seq}(\%)$	$N_{seq}$	$\hat{p}_{seq}(\%)$
1	12	63,88	12	71,60*
2	12	41,49*	13	37,61*
3	12	63,88	12	63,83*
4	12	63,88	12	63,83*
5	12	63,88	12	63,83*
6	8	91,06	8	92,61
7	8	91,06	8	92,61
8	12	63,88	12	63,83*
9	11	77,10	11	77,63
10	12	71,34	12	71,60*
11	10	83,86	10	84,77
12	11	77,10	11	77,63

\*: Rejeitou-se os lotes.

Fonte: Elaborado pelo autor.

Nota-se que com um vigor de 60% há a rejeição apenas do lote 2, enquanto que com um vigor de 75% há a rejeição dos lotes 1, 2, 3, 4, 5, 8 e 10. Além disso, é possível observar que o número máximo de amostragem por lote foi de 12 sementes. O caso apresentado aqui, demonstra o valor que a escolha da priori pode ter. Os hiperparâmetros utilizados refletem lotes de sementes de soja de baixo vigor, principalmente no primeiro caso, o que não está de acordo com o resultado obtido tanto na abordagem frequentista quanto na bayesiana.

Comparando as quatro abordagens - frequentista, bayesiana, análise sequencial e análise sequencial bayesiana - utilizados para analisar os doze lotes, nota-se potencialidade no uso das mesmas, sendo que, ressalta-se a importância da escolha da distribuição *a priori*. Nas abordagens frequentista e bayesiana houve concordância entre os resultados pois a distribuição *a priori* elicitada em artigos é condizente com a amostra de lotes utilizadas. No entanto, quando adicionada na análise sequencial o método bayesiano, se observa a diminuição do tamanho amostral para o sentenciamento do lote, resultando na redução de tempo e custo do processo de amostragem, no entanto houve forte influência da distribuição *a priori* que não era compatível com os resultados frequentistas. Assim, se reforça a importância na escolha da *priori*. O fato de possibilitar incorporar informações provenientes de produtores ou mesmo de artigos científicos com dados regionais pode diminuir o tempo e custo da amostragem, no entanto, tal escolha deve ter um embasamento consistente.

## **4.2 Estimação de danos em sementes de soja utilizando a distribuição Multinomial**

### **4.2.1 Abordagem frequentista da estimação de danos**

A abordagem frequentista da estimação dos danos foi realizada inicialmente considerando apenas os três tipos de danos e assim, utilizando na análise a distribuição multinomial em três classes. Fez-se a proporção de danos correspondente a cada uma das classes, sendo o total de danos considerado 100%. Deve-se ressaltar que nestes, caso, os tamanhos amostrais diferem entre os lotes, pois eles são apenas contabilizados nas sementes danificadas. Os resultados são apresentados na Tabela 11.

Tabela 11 - Porcentagem das sementes de acordo com as três classificações:  $\hat{p}_1$ : dano por umidade;  $\hat{p}_2$ : dano por percevejo;  $\hat{p}_3$ : dano mecânico.

Lote	n	$\hat{p}_1$ (%)	$\hat{p}_2$ (%)	$\hat{p}_3$ (%)
1	19	47,37	42,10	10,53
2	46	78,26	13,04	8,70
3	36	86,11	8,33	5,56
4	38	92,11	5,26	2,63
5	20	85,00	10,00	5,00
6	22	72,73	27,27	0,00
7	16	93,75	6,25	0,00
8	25	72,00	20,00	8,00
9	17	64,71	0,00	35,29
10	40	85,00	12,50	2,50
11	16	68,75	25,00	6,25
12	24	79,17	8,33	12,50
Média (%)	26,58	77,08	14,84	8,08
Desvio-Padrão (%)	10,54	13,03	11,80	9,42

Fonte: Elaborado pelo autor.

Das 319 sementes não vigorosas obtidas nos 12 lotes, em média de 77,08% apresentaram dano por umidade, com desvio-padrão de 13,03%, enquanto que 14,84% tiveram danos por percevejo com desvio-padrão de 11,80% e 8,80% foram danificados mecanicamente com desvio-padrão de 9,42%.

Posteriormente realizou-se a análise incluindo as sementes vigorosas e incluindo em categoria separada aquelas que possuíam mais de um tipo de dano. Assim, foi utilizada a distribuição Multinomial com cinco classes, que está exibida na Tabela 12.

Tabela 12 - Porcentagem das sementes de acordo com as cinco classificações:  $\hat{p}_1$ : vigorosa;  $\hat{p}_2$ : com presença de dano por umidade;  $\hat{p}_3$ : com presença de dano por percevejo;  $\hat{p}_4$ : com presença de dano mecânico;  $\hat{p}_5$ : com presença de dois ou mais danos.

Lote	$\hat{p}_1$ (%)	$\hat{p}_2$ (%)	$\hat{p}_3$ (%)	$\hat{p}_4$ (%)	$\hat{p}_5$ (%)
1	81	7	8	2	2
2	54	35	6	4	1
3	64	29	3	2	2
4	62	35	1	1	1
5	80	14	2	1	3
6	78	16	5	0	1
7	84	15	1	0	0
8	75	18	5	2	0
9	83	11	0	6	0
10	60	32	5	1	2
11	84	11	4	1	0
12	76	19	2	3	0
Média (%)	73,42	20,17	3,50	1,91	1,00
Desvio-Padrão (%)	10,54	9,94	2,39	1,73	1,04

Fonte: Elaborado pelo autor.

Já na avaliação do vigor com a distribuição multinomial em cinco classes o dano que se sobressai é o de umidade com média de 20,17%, contra a presença de percevejo de média igual a 3,5%, danos mecânicos de média igual a 1,91% e a presença de mais de um dano com 1,0%.

Os resultados encontrados nas duas formas de classificações, de três classes e de cinco, para o dano de umidade nos lotes analisados estão de acordo com os dados encontrados na literatura (KRZYZANOWSKI; FRANÇA-NETO; COSTA, 2004; FRANÇA-NETO; HENNING, 1984).

A explicação para que os danos mecânicos nos lotes analisados sejam mais baixos que os encontrados na literatura ocorrem devido à colheita manual dos grãos ao invés do uso de máquina colhedora (COSTA et al., 1996; CARVALHO; NAKAGAWA, 2000).

O impacto dos danos por percevejo nos lotes analisados não foi considerável. Costa et al. (2005) ao estudarem os principais fatores responsáveis pela redução da qualidade de sementes de soja, constataram que a incidência de lesões por percevejo até 5% não afeta o vigor das sementes, apenas a partir de 6% há diminuição significativa da taxa de vigor.

Muller et al. (2017) demonstraram a necessidade de manejo integrado para evitar perdas de produção por ataque de percevejos-marrom somente quando o lote portou 77% de vigor com 38% de sementes com danos por percevejos.

Foi realizado o teste qui-quadrado a partir das porcentagens apresentadas, obteve-se p-valor  $< 0,001$  e  $\chi^2 = 117,77$  com 11 graus de liberdade para classificação com cinco proporções e p-valor  $< 0,001$  e  $\chi^2 = 244,67$  com 11 graus de liberdade para a classificação com três proporções. Ilustrando que os lotes são independentes.

#### 4.2.2 Abordagem bayesiana da estimação de danos

Para a abordagem bayesiana com os três tipos de danos, umidade, percevejo e mecânico, compreendendo a distribuição multinomial, usou-se *a priori* conjugada de Dirichlet  $(a_1, a_2, a_3)$ , sabendo que após observar a amostra a distribuição *a posteriori* é Dirichlet  $(a_1 + n_1, \dots, a_k + n_k)$ , onde  $n_k$  é a frequência de cada  $a$ , sendo  $k$  o número de classes da multinomial, ou seja, o número de tipos de danos.

Para a *priori* os parâmetros aplicados foram de Costa et al. (2005) obtidos no estudo do perfil dos aspectos físicos, fisiológicos e químicos de sementes de soja produzidas em seis regiões do Brasil, onde se obteve 11,17% de sementes com danos mecânicos, 4,77% com danos por percevejo e 6,8% com danos de umidade. Desta forma, utilizando a equação (46) a *priori* é a Dirichlet  $(0,2456; 0,1048; 0,1496)$ , obtida através da esperança dos danos encontrados por Costa et al. (2005), com os valores da média e variância apresentados na Tabela 13.

Tabela 13 – Hiperparâmetros e valores da *priori* obtidos da multinomial com três classes de danos.

Tipo de dano	$\alpha_k$	Média (%)	Variância (%)
Umidade	0,1496	29,91	13,98
Percevejo	0,1048	20,97	11,05
Mecânico	0,2456	49,12	16,66

Fonte: Elaborado pelo autor.

Nos lotes analisados utilizou-se a frequência dos danos mecânicos, por percevejo e por umidade para atualizar a informação *a priori* e obter cada uma das distribuições *a posteriori* conforme apresentado em (47). Na Tabela 14 são apresentadas as médias da *posteriori* para cada parâmetro.

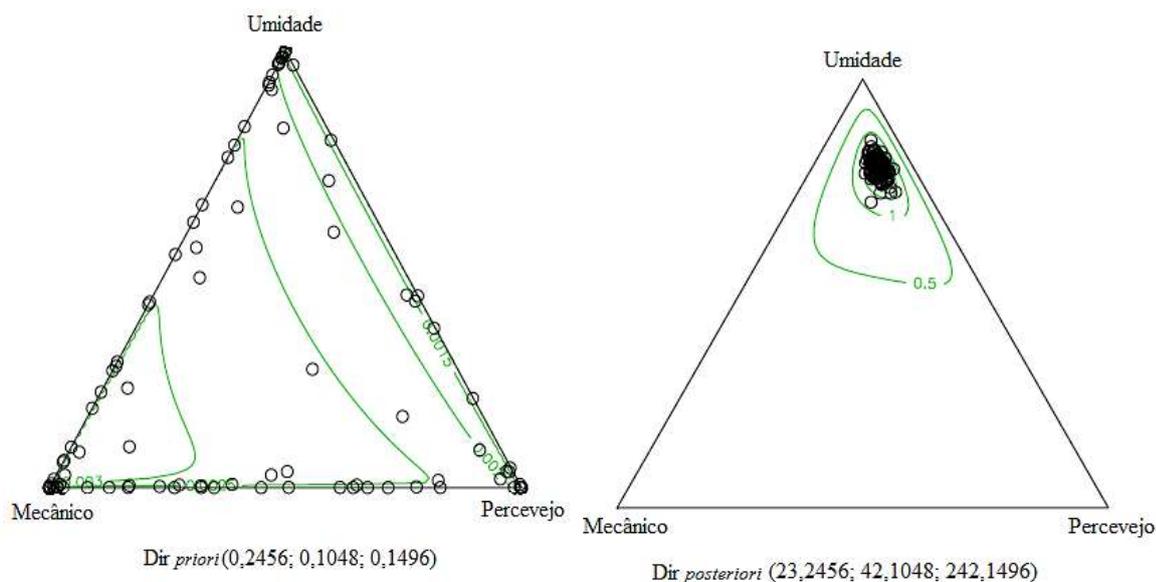
Tabela 14 – Estimativas *a posteriori* da multinomial para proporção de danos em sementes de soja.

Tipo de dano	Mecânico		Percevejo		Umidade	
	Média	Variância	Média	Variância	Média	Variância
Lote 1	0,4728	0,002444	0,4204	0,002389	0,1062	0,00093
2	0,7802	0,001681	0,1313	0,001118	0,0880	0,00079
3	0,8583	0,001192	0,0844	0,000758	0,0568	0,00052
4	0,9180	0,000738	0,0539	0,000500	0,0277	0,00026
5	0,8473	0,001269	0,1010	0,000890	0,0512	0,00048
6	0,7251	0,001954	0,2729	0,001945	0,0015	0,00001
7	0,9343	0,000602	0,0637	0,000585	0,0015	0,00001
8	0,7179	0,001985	0,2005	0,001572	0,0811	0,00073
9	0,6453	0,002244	0,0015	1,46E-05	0,3527	0,00224
10	0,8473	0,001269	0,1259	0,001079	0,0264	0,00025
11	0,6856	0,002113	0,2502	0,001839	0,0637	0,00058
12	0,7892	0,001631	0,0844	0,000758	0,1259	0,00108

Fonte: Elaborado pelo autor.

Para se ter melhor compreensão do impacto do parâmetro  $\alpha_k$  nos casos da *priori* e da *posteriori* é apresentado o simplex de dimensão 3. Ao considerarmos os danos por categoria tem-se na Figura 13, a representação da distribuição *a priori* em que se percebe os pontos mais distribuídos entre as categorias. Na segunda figura tem-se a *posteriori* considerando os dados totais de danos por categoria para construção da distribuição *a posteriori*. Nos lotes analisados a frequência dos danos mecânicos foi de 23 sementes, a de dano por percevejo de 42 sementes e de dano por umidade de 242. Estes valores foram utilizados para obtenção da *posteriori*, a qual é a Dirichlet (23,2456; 42,1048; 242,1496). É possível observar claramente que para os 12 lotes avaliados, em média, tem-se maior ocorrência de danos por umidade.

Figura 13 - Simplex da *priori* da Dirichlet com parâmetros (0,2456; 0,1048; 0,1496) e da *posteriori* com os parâmetros (23,2456; 42,1048; 242, 1496).



Fonte: Elaborado pelo autor.

Os resultados encontrados pela estimação bayesiana dos três tipos de danos estão condizentes com os valores encontrados na literatura, o qual apresentou médias de 8,08% em danos mecânicos, 14,84% em danos por percevejo e 77,08% em danos por umidade.

Para estimar a proporção das sementes nas cinco classes apresentadas na Tabela 15 com distribuição multinomial foram utilizados os dados da safra de 2016/2017 (LORINI, 2017) para obtenção dos parâmetros. Logo, a proporção para obtenção da *priori* para as

sementes vigorosas foi de 82%, para danos por umidade 3,1%, para danos por percevejo 0,7%, para danos mecânicos 4,9% e para incidência de dois ou mais danos não foi especificada, sendo utilizado então 0%. Desse modo, a *priori* conjugada é a Dirichlet (0,4520; 0,0171; 0,0039; 0,0270; 0,00) com os seguintes valores de médias e variâncias, de acordo com a Tabela 15.

Tabela 15 - Valores da *priori* da multinomial com cinco parâmetros ( $\hat{p}_1$ : vigorosa;  $\hat{p}_2$ : com presença de dano por umidade;  $\hat{p}_3$ : com presença de dano por percevejo;  $\hat{p}_4$ : com presença de dano mecânico;  $\hat{p}_5$ : com presença de 2 ou mais danos).

$\hat{p}$	$\alpha_k$	Média (%)	Variância (%)
$\hat{p}_1$	0,4520	90,41	5,78
$\hat{p}_2$	0,0171	3,42	2,20
$\hat{p}_3$	0,0039	0,77	0,51
$\hat{p}_4$	0,0270	5,40	3,41
$\hat{p}_5$	0,0000	0,00	0,00

Fonte: Elaborado pelo autor.

Com os valores registrados de cada tipo de dano em cada lote, foi atualizada a informação *a priori* e obtida a posteriori dos parâmetros em cada lote. As estimativas dos parâmetros são apresentadas na Tabela 16 a seguir.

Tabela 16 – Estimativas a *posteriori* da multinomial para proporção de danos em sementes de soja.

Tipo de dano	Vigorosa	Dano por umidade	Dano por percevejo	Dano mecânico	2 ou mais danos
Lote	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$
1	0,81047	0,06982	0,07960	0,01990	0,01990
2	0,54181	0,34843	0,05970	0,03980	0,00995
3	0,64131	0,28873	0,02985	0,01990	0,01990
4	0,62141	0,34843	0,00995	0,00995	0,00995
5	0,80052	0,13947	0,01990	0,00995	0,02985
6	0,78062	0,15937	0,04975	0,00000	0,00995
7	0,84032	0,14942	0,00995	0,00000	0,00000
8	0,75077	0,17927	0,04975	0,01990	0,00000
9	0,83037	0,10962	0,00000	0,05970	0,00000
10	0,60151	0,31858	0,04975	0,00995	0,01990
11	0,84032	0,10962	0,03980	0,00995	0,00000
12	0,76072	0,18922	0,01990	0,02985	0,00000

Fonte: Elaborado pelo autor.

Para o cálculo da *posteriori* de Dirichlet geral utilizou-se as frequências de cada parâmetro, sendo 881 sementes vigorosas, 242 com dano por umidade, 42 com danos por percevejo, 23 com danos mecânicos e 12 com dois ou mais danos. Portanto a *posteriori* é a Dirichlet (881,4520; 242,0171; 42,0039; 23,0270; 12,00) com médias e variâncias apresentadas na Tabela 17.

Tabela 17 - Valores da *posteriori* da multinomial com cinco parâmetros.

$\hat{p}$	$\alpha_k + y_k$	Média (%)	Variância (%)
$\hat{p}_1$	881,4520	73,42	0,02
$\hat{p}_2$	242,0171	20,16	16,10
$\hat{p}_3$	42,0039	3,50	3,38
$\hat{p}_4$	23,0270	1,92	1,88
$\hat{p}_5$	12,0000	1,00	0,99

Fonte: Elaborado pelo autor.

Os resultados obtidos na análise bayesiana multinomial com cinco classificações também estão condizentes, que apresentou média de 73,42% para as sementes vigorosas, 20,16% para as com danos por umidade, 3,50% para as com danos por prevejo, 1,92% para as com dano mecânico e 1,00% para as que apresentaram dois ou mais danos.

#### 4.2.3 Estimação sequencial bayesiana de danos considerando a distribuição multinomial

Para estabelecer se a amostragem deve ser continuada ou não é necessário que os valores de risco sejam comparados a cada semente avaliada e realizada nova observação até que o risco imediato ( $S(\mathbf{n}, m)$ ) seja menor ou igual ao risco esperado ( $B(\mathbf{n}, m)$ ).

Entretanto, para encontrar os valores de risco imediato e esperado, deve-se primeiramente encontrar os valores dos elementos  $d_i^*$  da distribuição *posteriori* de  $(\mathbf{n}, m)$ , dados por  $d_i^* = \frac{n_i+1}{m+k}$ , onde  $n_i$  são os valores acumulados de defeitos,  $m$  são as observações e  $k$  o número de classes. Posteriormente, obtém-se o cálculo do risco imediato, dado por:

$$S(\mathbf{n}, m) = \frac{\{\sum_{i=1}^k K_{ii}d_i^* - \sum_{i,j=1}^k K_{ij}d_i^*d_j^*\}}{m+k+2}, \quad (59)$$

onde  $\mathbf{K}$  é a matriz simétrica positiva de perda constante.

O próximo passo realizado foi o cálculo do risco esperado, para isso foi necessário saber o custo de mais uma observação. Sendo estabelecido o custo igual ao utilizado na amostragem sequencial bayesiana de vigor de  $10^{-4}$ .

A estimação sequencial bayesiana multinomial foi realizada através de uma tabela dinâmica realizada no Microsoft Excel® nos doze lotes propostos para este trabalho. Utilizando *a priori* com três classes apresentada por Costa et al. (2005) e a de cinco classes dos dados da safra da soja 2016/2017 apresentada por Lorini (2017).

Como exemplo, podemos citar a avaliação da primeira semente do lote 4, com um custo de  $10^{-3}$ , pois com o custo de  $10^{-4}$  nenhum dos doze lotes interrompeu a

amostragem. Para a *a priori* os parâmetros de Costa et al. (2005) obtidos no estudo do perfil dos aspectos físicos, fisiológicos e químicos de sementes de soja produzidas em seis regiões do Brasil, onde se obteve 11,17% de sementes com danos mecânicos, 4,77% com danos por percevejo e 6,8% com danos de umidade, o risco imediato ( $S(\mathbf{n}, m)$ ) e o risco esperado ( $(B(\mathbf{n}, m))$ ) para a multinomial com três parâmetros são calculados a partir de:

- *a priori* é a Dirichlet (0,1496; 0,1048; 0,2456);
- *a posteriori* será a Dirichlet (1,1496; 0,1048; 0,2456), considerando a semente incluída na classe 1. Tem-se:

$$d_1 = \frac{(1 + 0,1496)}{1 + 0,5} = 0,7664 \quad d_2 = \frac{(0 + 0,1048)}{1 + 0,5} = 0,0699 \quad d_3 = \frac{(0 + 0,2456)}{1 + 0,5} = 0,1637$$

$$S = \frac{0,7664 + 0,0699 + 0,1637}{1 + 3 + 2} = 0,1667$$

$$B = 0,0001 + \left[ \frac{2 + 3 + 1}{2 + 3 + 2} \right] * 0,1667 = 0,139$$

Como  $S(0,1667) > B(0,1390)$  a amostragem deve continuar. Os cálculos dos riscos devem ser realizados até que o risco esperado seja maior que o risco imediato, assim se encerra a amostragem e é realizada a estimacão.

A Tabela 18 apresenta os resultados para a multinomial com três parâmetros ( $\hat{p}_1$ : dano por umidade;  $\hat{p}_2$ : dano por percevejo;  $\hat{p}_3$ : dano mecânico) com o custo de mais uma observação de  $10^{-3}$ .

Tabela 18 - Porcentagens de danos dos lotes de sementes estimados pelo método sequencial bayesiano multinomial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ), de acordo com os três parâmetros.

Lotes	$N_{seq}$	Danos					
		$\hat{p}_1$		$\hat{p}_2$		$\hat{p}_3$	
		$N_1$	$\hat{p}_{1\ seq} (\%)$	$N_2$	$\hat{p}_{2\ seq} (\%)$	$N_3$	$\hat{p}_{3\ seq} (\%)$
1*	19	9	47,37	8	42,10	2	10,53
2	25	20	80,00	3	12,00	2	8,00
3	25	22	88,00	2	8,00	1	4,00
4	25	24	96,00	0	0,00	1	4,00
5*	20	17	85,00	2	10,00	1	5,00
6*	22	16	72,73	5	22,72	1	4,55
7*	16	15	93,75	1	6,25	0	0,00
8	25	18	72,00	5	20,00	2	8,00
9*	17	11	64,71	0	0,00	6	35,29
10	26	21	80,77	4	15,38	1	3,85
11*	16	11	68,75	4	25,00	1	6,25
12*	24	19	79,17	2	8,33	3	12,5
Média (%)		77,35		14,15		8,50	
Desvio-padrão (%)		10,20		9,08		5,47	

\*: A amostragem não foi interrompida.

Fonte: Elaborado pelo autor.

Os resultados encontrados pela estimação sequencial bayesiana multinomial com três parâmetros estão condizentes com os valores da abordagem frequentista, o qual apresentou médias de 8,08% em danos mecânicos, 14,84% em danos por percevejo e 77,08% em danos por umidade. E, conseqüentemente, com os resultados também encontrados com a estimação bayesiana.

Nota-se que com os lotes 1, 5, 6, 7, 9, 11 e 12 não é possível realizar a análise sequencial multinomial, pois o número de amostras ficou pequeno para o caso. Desta forma, observa-se que o número mínimo de amostras para interromper a amostragem dos lotes foi de 25 sementes.

A Tabela 19 apresenta os resultados para a multinomial com cinco parâmetros ( $\hat{p}_1$ : vigorosa;  $\hat{p}_2$ : com presença de dano por umidade;  $\hat{p}_3$ : com presença de dano por percevejo;  $\hat{p}_4$ : com presença de dano mecânico;  $\hat{p}_5$ : com presença de 2 ou mais danos), com o custo de mais uma observação de  $10^{-4}$ .

Tabela 19 - Porcentagens de danos dos lotes de sementes estimados pelo método sequencial bayesiano multinomial, com seus respectivos tamanhos amostrais finais ( $N_{seq}$ ), de acordo com os cinco parâmetros.

Lotes	$N_{seq}$	Danos									
		$\hat{p}_1$		$\hat{p}_2$		$\hat{p}_3$		$\hat{p}_4$		$\hat{p}_5$	
		$N_1$	$\hat{p}_{1 seq}(\%)$	$N_2$	$\hat{p}_{2 seq}(\%)$	$N_3$	$\hat{p}_{3 seq}(\%)$	$N_4$	$\hat{p}_{4 seq}(\%)$	$N_5$	$\hat{p}_{5 seq}(\%)$
1	94	77	81,91	7	7,45	6	6,38	2	2,13	2	2,13
2	94	49	52,13	35	37,23	6	6,38	3	3,19	1	1,06
3	93	57	60,64	29	31,18	3	3,23	2	2,15	1	1,08
4	93	58	61,70	32	34,41	1	1,08	1	1,08	0	0,00
5	93	74	79,72	14	15,05	2	2,15	0	0,00	2	2,15
6	93	71	75,53	16	17,20	5	5,38	0	0,00	0	0,00
7	94	79	84,04	14	14,89	1	1,06	0	0,00	0	0,00
8	94	71	75,53	16	17,02	5	5,32	2	2,13	0	0,00
9	94	78	82,98	10	10,64	0	0,00	6	6,38	0	0,00
10	94	57	60,64	30	31,91	5	5,32	1	1,06	1	1,06
11	94	78	82,98	11	11,70	4	4,26	1	1,06	0	0,00
12	94	73	77,66	18	19,15	1	1,06	2	2,13	0	0,00
Média (%)		73,27		20,69		3,49		1,88		0,67	
Desvio padrão (%)		9,40		8,69		2,04		1,24		0,73	

Fonte: Elaborado pelo autor.

Os valores estimados através da análise sequencial bayesiana multinomial com cinco parâmetros estão de acordo com os valores da análise bayesiana multinomial e, consequentemente, com os valores encontrados no experimento, de médias igual a 73,42% para as sementes vigorosas, 20,17% para as com dano de umidade, 3,50% para

as com danos por percevejo, 1,92% para as com danos mecânicos e 1% para as sementes que apresentaram dois ou mais danos.

Além disso, é possível observar que o número máximo de amostras por lote foi de 94 sementes para parar a amostragem e assim ser feito o sentenciamento. Isso indica de como o método sequencial bayesiano multinomial considerando os danos nas sementes reduz o tempo de amostragem necessário para sentenciar um lote, em comparação ao método tradicional de 100 sementes.

## 5 CONCLUSÕES

As quatro abordagens frequentista, bayesiana, sequencial e sequencial bayesiana, foram eficientes para realizar a estimação e tomada de decisão sobre os parâmetros de vigor e danos das sementes de soja.

Na tomada de decisão durante a execução do teste de tetrazólio, para a modelagem do vigor a utilização da distribuição binomial é adequada e no caso da análise de tipos de danos, em que se tem várias categorias, o uso da distribuição multinomial foi promissora.

No caso da abordagem bayesiana, tanto de amostragem fixa quanto sequencial, é importante destacar a importância da escolha da distribuição *a priori*. O uso de artigos da área de trabalho de sementes de soja para elicitar os hiperparâmetros das distribuições *a priori* mostrou-se uma alternativa promissora para melhorar a estimação dos parâmetros durante o teste de tetrazólio.

Pelo processo sequencial foi possível reduzir o número de sementes avaliadas, por outro lado, nos casos em que o critério foi muito flexível, ou muito distante dos valores obtidos pela estimação frequentista, notou-se que o tamanho da amostra foi insuficiente para tomada de decisão sobre o vigor das sementes, evidenciando a necessidade de critérios para avaliação das hipóteses consideradas.

No processo sequencial bayesiano, tanto na estimação de vigor (uso da distribuição binomial), quanto de danos em sementes (uso da distribuição multinomial), foi possível estabelecer com inclusão das informações *a priori* e redução do tamanho amostral, as proporções de vigor e danos das sementes avaliadas pelo teste de tetrazólio.

Considerando a abordagem frequentista, os danos de maior ocorrência verificados foram os ocasionados por umidade, estando presentes em aproximadamente 20,16% das sementes de soja avaliadas. Já os danos por percevejo ocorrendo em cerca de 3,5% e os danos mecânicos, no caso da colheita manual, em apenas 1,92% das sementes. Tem-se ainda 1% das sementes com mais de dois danos.

## REFERÊNCIAS

ALBRECHT, A.; GULDE, H.; MACLEAN, A.; THOMPSON, P. Continuous sampling at Minneapolis – Honeywell. **Industrial Quality Control**, v. 12, n. 3, p. 4-9, 1955.

AZEREDO LOPES, S.; CARDOSO, J. L. **Modelos de estimativa de frequência de acidentes e vítimas para dados bidireccionais de auto-estradas portuguesas** – Uma abordagem bayesiana. Relatório 5/2010 - NPTS, Laboratório Nacional de Engenharia Civil, Lisboa, Portugal, 2009.

BAYES, T; An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. **Philosophical transactions of the Royal Society of London**, n. 53, p. 370-418, 1763.

BALLARIS, A. L. Amostragem sequencial de sementes de soja e feijão na detecção de *Sclerotinia sclerotiorum*. **Journal of Seed Science**, v. 36, n. 3, 2014.

BÁNYAI, J.; BARABÁS, J. **Handbook on statistics in seed testing**. International Seed Testing Association (ISTA). 2002.

BERGER, J. O. **Statistical decision theory and bayesian analysis**. 2nd ed. New York: Springer-Verlag, 1985. 624p.

BERQUÓ, E. S. **Sobre a análise sequencial para testes de hipóteses relativas a proporções. Aplicação a problemas de medicina e de saúde pública**. São Paulo, 142p., 1959.

BOLFARINE, H.; SANDOVAL, M.C. **Introdução à inferência estatística**. Rio de Janeiro: Sociedade Brasileira de Matemática, 125p., 2001.

BOX, G. E. P.; TIAO, G.C. **Bayesian inference in statistical analysis**. New York: J. Wiley, 603p, 1992.

BRASIL. Lei n. 10.711, de 5 de agosto de 2003. Dispõe sobre o Sistema Nacional de Sementes e Mudas - SNSM, e dá outras providências. **Diário oficial da República Federativa do Brasil**, Poder executivo, Brasília, DF, 5 ago. 2003.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para Análise de Sementes. Ministério da Agricultura, Pecuária e Abastecimento. **Secretaria de Defesa Agropecuária**. Brasília, DF: Mapa/ACS, 398p, 2009.

BRASIL. CONGRESSO NACIONAL. Instrução Normativa nº 45, de 17 de setembro de 2013. Dispõe sobre os padrões para a produção e a comercialização de sementes. **Diário Oficial da União**, 2013.

BRIGHENTI, C. R. G., RESENDE, M., & BRIGHENTI, D. M. Estimativa sequencial bayesiana aplicada à proporção de infestação de psilídeos em alecrim do campo. **Rev. Bras. Biom**, v. 29, n. 2, p. 342-354, 2011.

BOTTINE, S. M., Wald's Sequential Probability Ratio Test, Package: SPRT, **CRAN**, 2015. Disponível em <<https://CRAN.R-project.org/package=SPRT>>.

CASELLA, G.; BERGER, R. L. **Inferência estatística**. Cengage Learning, v.2, p.573, 2010.

CARVALHO, N. M.; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. (Vol. 4). Jaboticabal: FUNEP, 2000.

CARVALHO, M. L. M.; PINTO, E. V R. V.; OLIVEIRA, J. A., GUIMARÃES, R. G.; SILVA, T. T. A. **Manual do amostrados de sementes**. 3 ed. Londrina: Abrates, v.1, p.143, 2018.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira**: v. 10 Safra 2017/18 - Décimo levantamento, Brasília, p. 1-178, julho 2018.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Compêndio de estudos Conab / Companhia Nacional de Abastecimento.** – v.1, Brasília: Conab, 2016.

CORDEIRO, G. M.; NETO, E. A. **Modelos paramétricos.** Pernambuco: UFRPE, 2004.

CORRÊA-FERREIRA, B. S.; PANIZZI, A. R. **Percevejos da soja e seu manejo.** Londrina: EMBRAPA-CNPSO, 1999. 45p. (EMBRAPA CNPSO. Circular Técnica, 24).

COSTA, N. P.; OLIVEIRA, M. C. N.; HENNING, A. A.; KRZYZANOWSKI, F. C.; MESQUITA, C. D. M.; TAVARES, L. C. V. Efeito da colheita mecânica sobre a qualidade da semente de soja. **Embrapa Soja-Artigo em periódico indexado (ALICE)**, 1996.

COSTA, N. P.; MESQUITA, C. M.; MAURINA, A. C.; FRANÇA NETO, J. B.; KRZYZANOWSKI, F. C.; OLIVEIRA, M. C. N.; HENNING, A. Perfil dos aspectos físicos, fisiológicos e químicos de sementes de soja produzidas em seis regiões do Brasil. **Revista Brasileira de Sementes**, v. 27, n. 2, p. 172-181, 2005.

DODGE, H. F.; ROMING, H. G. **Sampling inspection tables** - single and double sampling, 2nd ed., New York: John Wiley & Sons, 1959.

EHLERS, Ricardo S. Inferência bayesiana. **Departamento de Matemática Aplicada e Estatística, ICMC-USP**, p. 64, 2011.

ELLIS, R. H.; WHITEHEAD, J. Open, truncated and triangular sequential seed testing procedures. **Seed Science and Technology**, v. 15, p. 1-17, 1987.

FARIAS, P. R. S.; BARBOSA, J. C.; BUSOLI, A. C. Amostragem sequencial com base na lei de Taylor para levantamento de *Spodoptera frugiperda* na cultura do milho. **Scientia Agricola**, v. 58, n. 2, p. 395-399, 2001.

FERNANDES, W.; NAGHETTINI, M.; LOSCHI, R. A Bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions. **Stochastic Environmental Research and Risk Assessment**, v. 24, n. 8, p. 1127-1143, 2010.

FLOR, E. P. O.; CICERO, S. M.; FRANÇA NETO, J. D. B.; KRZYZANOWSKI, F. C. Avaliação de danos mecânicos em sementes de soja por meio da análise de imagens. **Revista Brasileira de Sementes**, v. 26, n. 1, p. 68-76, 2004.

FRANÇA-NETO, J. B. Princípios do teste de tetrazólio para a semente de soja. **Boletim LASP**, Curitiba, v.3, n.1, p.1-8, 1981.

FRANÇA-NETO, J. B.; HENNING, A. A. **Qualidades fisiológica e sanitária de sementes de soja**. Londrina: EMBRAPA-CNPSO, 39p. (EMBRAPA-CNPSO. Circular Técnica, 9), 1984.

FRANÇA-NETO, J. B.; HENNING, A. A. **Diagnóstico completo da qualidade da semente de soja** - Versão Preliminar. Londrina: EMBRAPA-CNPSO, p.9-43. 1985.

FRANÇA-NETO, J. D. B.; PEREIRA, L. G.; DA COSTA, N. P.; KRZYZANOWSKI, F. C.; HENNING, A. A. **Metodologia do teste de tetrazólio em semente de soja**. Londrina: EMBRAPA-CNPSO, Documentos 32, 58p., 1988.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C.; COSTA, N. P. **O teste de tetrazólio em sementes de soja**. Londrina: EMBRAPA-CNPSO. 72p. (EMBRAPA-CNPSO. Documentos, 116). 1998.

FRANÇA-NETO, J. B. (Ed.). **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, p. 8.5-1-8.5.28, 1999.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C., HENNING, A. A., DE PÁDUA, G. P., LORINI, I., & HENNING, F. A. **Tecnologia da produção de semente de soja de alta qualidade**. 82 p. – (Documentos / Embrapa Soja, ISSN 2176-2937; n. 380), 2016.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C. **Metodologia do teste de tetrazólio em sementes de soja**. 108p. (Documentos/Embrapa Soja, ISSN 2176-2937; n.406), 2018.

FREEMAN, P. R. Sequential estimation of the size of a population. **Biometrika**, v.59, n.1, p.9-17, 1972.

GARRUTI, R. S. **Metodologia na seleção sequencial e não sequencial de provadores para análise sensorial de alimentos e bebidas**. 1976. 220p. Tese (Doutorado). UNICAMP, Campinas, 1976.

GARTHWAITE, P. H.; JOLLIFFE, I.; BYRON, J. **Statistical inference**. 2nd ed. London: Prentice Hall, p.114-131, 1995.

GELMAN, A.; CARLIN, J.B.; STERN, H.S.; RUBIN, D.B. **Bayesian data analysis**. London: Chapman & Hall, 2013.

GHOSH, B. K. **Sequential tests of statistical hypotheses**. Massachusetts: Addison-Wesley, 454 p., 1970.

HARGER, N.; VELINI, E. D.; SILVAI, A.; ADEGAS, F. S. Averiguação das perdas por deriva nas pulverizações de agrotóxicos no Norte do Paraná. In: **Congresso Brasileiro De Soja**, 4, Londrina: Embrapa Soja, p. 155., 2006.

HENNING, A. A.; FRANÇA NETO, J. B. Problemas na avaliação da germinação de semente de soja com alta incidência de *Phomopsis sp.* **Revista Brasileira de Sementes**, Brasília, v., n.3, p.9-22, 1980.

HENNING, A. A.; FRANÇA NETO, J. B.; COSTA, N. P. Efeito da época do tratamento químico e/ou período de armazenamento sobre a qualidade fisiológica e sanitária das sementes de soja cv. Bossier e Paraná, com altos índices de *Phomopsis sp.* In: **Congresso Brasileiro de Sementes**. p. 24, 1981.

HINES, W. W.; MONTGOMERY, D. C.; GOLDSMAN, D. M. **Probabilidade e Estatística na Engenharia**. Grupo Gen-LTC, 2000.

HOLTZ, V.; REIS, E. F. Losses in mechanized harvesting soybean: A quantitative and qualitative analysis. **Revista Ceres**, v. 60, n. 3, p. 347-353, 2013.

JONES, P. W. A note on the bayesian sequential estimation of a binomial parameter, **Biometrika**, p. 642-644, 1974.

KINAS, P. G.; ANDRADE, H. A. **Introdução à análise bayesiana (com R)**. Consultor Editorial, 2017.

KRZYZANOWSKI, C. F.; VIEIRA, R. D.; FRANÇA-NETO, J. B. **Associação Brasileira de Tecnologia de Sementes**. Comitê de Vigor de Sementes. Londrina: ABRATES, 1999.

KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. B.; COSTA, N. P. **Teste do hipoclorito de sódio para semente de soja**. Londrina: Embrapa Soja (Embrapa Soja. Circular Técnica, 37), 4p., 2004.

LEHMANN, E. L. **Testing Statistical Hypotheses**. New York: John Wiley & Sons, 1959.

LINDLEY, D.V.; BARNETT, B. N. Sequential sampling two decision problems with linear losses for binomial and normal random variables, **Biometrika** **52**, 507-520, 1965.

LOPES, M. D. M.; PRADO, D.; OLIVEIRA, M.; SADER, R.; BARBOSA, R. M. Efeitos dos danos mecânicos e fisiológicos na colheita e beneficiamento de sementes de soja, **Bioscience Journal**, v. 27, n. 2, p. 230-238, 2011.

LORINI, I. **Qualidade de sementes e grãos comerciais de soja no Brasil - safra 2016/17**. Londrina: Embrapa Soja, 227p. il. (Documentos / Embrapa Soja, ISSN 2178-2937; 393), 2017.

MAFRA, A. C. C. N.; NUCCI, L. B.; CORDEIRO, R.; STEPHAN, C. Estudo caso-controlado com resposta multinomial: uma proposta de análise. **Cadernos de Saúde Pública**, v. 26, p. 451-459, 2010.

MARCOS FILHO, J.; CARVALHO, R. V.; CICERO, S. M.; DEMÉTRIO, C. G. B. Qualidade fisiológica e comportamento de sementes de soja (*Glycine max (L.) Merrill*) no armazenamento e no campo. **Anais da Escola Superior de Agricultura Luiz de Queiroz**, 42(1), 194-249, 1985.

MARCOS FILHO, J. Deterioração de sementes. **Fisiologia de sementes de plantas cultivadas**. Piracicaba: FEALQ, v. 12, p. 291-352, 2005.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2nd ed. London: Chapman & Hall, 1989.

MESQUITA, C. D. M.; COSTA, N. P.; PEREIRA, J. E.; MAURINA, A. C.; & ANDRADE, J. G. M. Perfil da colheita mecânica de soja no Brasil: safra 1998/1999. **Engenharia Agrícola**, Jaboticabal, v. 22, n. 3, p. 398-406, 2002.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO – MAPA-AGROSTAT. **Estatísticas de comércio exterior do agronegócio brasileiro**, 2019. Disponível em: <<http://indicadores.agricultura.gov.br/agrostat/index.htm>>. Acesso em: 10 jun. 2019.

MOREANO, T. B.; BRACCINI, A. L.; SCAPIM, C. A.; KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. B.; MARQUES, O. J. Changes in the effects of weathering and mechanical damage on soybean seed during storage. **Seed Science and Technology**, v. 39, p. 04-611, 2011.

MOORE, R. P. Tetrazolium staining for assessing seed quality. In: HEYDECKER, W. ed. **Seed ecology**. London: Butterworth, p. 347-366, 1973.

MOORE, R. P. Tetrazolium seed testing developments in North America. **Journal of Seed Technology**, Beltsville, v. 1, p. 17-30, 1976.

MOORE, R. P. **Handbook on tetrazolium testing**. Zurich: International Seed Testing Association, 99p., 1985.

MULLER, D.; LAMERS, N.; GNIECH, L.; SAUSEN, D.; MAMBRIN, R. B. Controle de percevejo-marrom em soja com o uso de produtos químicos e biológicos. **Anais do Congresso Internacional Das Ciências Agrárias**, v. 2, p. 1-9, Natal, Rio Grande do Norte, 2017.

O'HAGAN, A.; FORSTER, J. J. **Kendall's advanced theory of statistics**, volume 2B: Bayesian inference. Arnold, 2004.

OWEN, R. J., The optimum design of a two-factor experiment using prior information, **Ann. Math. Statist.**41, 1917-1934, 1970.

PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística bayesiana**. Editora: Fundação Calouste Gulbenkian. Lisboa, 446 p., 2003.

PENTEADO, S. R. C.; DE OLIVEIRA, E. B.; IEDE, E. T. Utilização da amostragem seqüencial para avaliar a eficiência do parasitismo de *Deladenus (Beddingia) siricidicola* (Nematoda: Neotylenchidae) em adultos de *Sirex noctilio* (Hymenoptera: Siricidae). **Ciência Florestal**, v. 18, n. 2, p. 223-231, 2008.

PHAM-GIA, T., TURKKAN, N. Distribution of the linear combination of two general beta variables and applications. **Comm. Statist. – Theory Meth.**, v. 27, 1998.

PINHO, E. M. **Estimação bayesiana para Medidas de Desempenho de Testes Diagnósticos**. Dissertação (Mestrado). Universidade Federal de São Carlos, 2006.

POPINIGIS, F. **Fisiologia de semente**. Brasília: AGIPLAN, 1985. 289p.

PORTER, R. H.; DURRELL, Mary; ROMM, H. J. The use of 2,3,5-triphenyltetrazolium-chloride as a measure of seed germinability. **Plant Physiology**, v.22, n.2, p.149, 1947.

QUINTELA, E. D., FERREIRA, S. B., GUIMARÃES, W. F. F., OLIVEIRA, L. D., OLIVEIRA, A. C., & CZEPAK, C. Desafios do Mip em grandes propriedades no Brasil Central. **Proc 4th Soybeans Congr**, p. 127-133, 2006.

REIS, R. L.; MUNIZ, J. A.; SILVA, F. F.; SAFADI, T.; AQUINO, L.H.; Abordagem bayesiana da sensibilidade de modelos para o coeficiente de endogamia. **Ciência Rural**, Santa Maria, v. 39, n. 3, p. 1752-1759, ago./dez. 2009.

RIBEIRO, G. A.; PINTO, A. C. S.; FORATORI-JUNIOR, G. A.; ORENHA, E. S.; PERES, S.; DE CARVALHO, S. H. Progressão da cárie dentária em gestantes com excesso de peso e com peso normal, no período pré e pós termo, atendidas pelo SUS. **Brazilian Oral Research**, 2018.

SANTOS, E. L.; PÓLA, J. N.; BARROS, A. S. R.; PRETE, C. E. C. Qualidade fisiológica e composição química das sementes de soja com variação na cor do tegumento. **Revista Brasileira de Sementes**, v. 29, n. 1, p. 20-26, 2007.

SAVIAN, T. V.; MUNIZ, J. A.; SÁFADI, T.; FONSECA, F. Análise bayesiana para modelos de degradabilidade ruminal. **Ciência Rural**, v. 39, n. 7, p. 2169-2177, 2009.

SCHILLING, E. G.; NEUBAUER, D. V. **Acceptance sampling in quality control**. Chapman and Hall/CRC, 2017.

SCHLAIFER, R. O.; HOWARD, R. **Applied Statistical Decision Theory**. Division of Research, Harvard Business School, Cambridge: Republished in Wiley Classic Library Series, 1961.

SCHUAB, S.R.P.; BRACCINI, A.L.; SCAPIM, C.A.; FRANÇA-NETO, J.B.; MESCHÉDE, D.K; ÁVILA, M.R. Germination test under water stress to evaluate soybean seed vigour. **Seed Science and Technology**, v. 35, n. 1, p. 187-199, 2007.

SILVA, A. H. L. Estimativa de proporções em questões politômicas. **Revista do TCU**, n. 125, p. 18-27, 2012.

SILVA, M. I. de S. **Mapeamento de QTLs em populações com pedigrees complexos**. Tese (Doutorado em Agronomia) – Universidade Federal de Lavras, Lavras, 104 p., 2008.

SILVEIRA, J. M.; CONTE, O.; MESQUITA, C. M. **Determinação de perdas na colheita de soja: copo medidor da Embrapa**. Embrapa Soja-Fôlder/Folheto/Cartilha (INFOTECA-E), 2017.

SINCLAIR, J. B. **Compendium of soybean diseases**. 2nd ed. St.Paul: The American Phytopathological Society, 104p., 1982.

SMITH, P. H.; SWEENEY, H. C.; ROONEY, J. R.; KING, K. W.; e MOORE, W. E. C. Stratifications and kinetic changes in the ingesta of the bovine rumen. **Journal of Dairy Science**, v. 39, n. 5, p. 598-609, 1956.

SORENSEN, D.; GIANOLA, D. **Likelihood, bayesian, and MCMC methods in Quantitative Genetics**. 2ª edição. New York: Springer-Verlag, 740 p., 2002.

STATISTICAL RESEARCH GROUP, Columbia University, **Sequential analysis of statistical data: applications**, AMP Report 30.2R, Columbia University, New York, 1945.

STURTZ, S.; LIGGES, U.; GELMAN, A. R2WinBUGS: a package for running WinBUGS from R, **Journal of Statistical Software**, v. 12, n. 3, p. 1–16, 2005.

TIMPANI, V. D.; DO NASCIMENTO, T. E. C. Uma breve introdução à estatística bayesiana aplicada ao melhoramento genético animal. **Embrapa Amazônia Oriental- Documentos (INFOTECA-E)**, 2015.

VILLAS BÔAS, G. L.; GAZZONI, D., FRANÇA NETO, J. B., COSTA, N., HENNING, A., & ROESSING, A. Efeito de cinco populações de percevejos sobre diversas características da soja cv. UFV-1. **Londrina: EMBRAPA-CNPSO**, 13p., 1982.

WALD, A. **Sequential analysis**. New York: John Wiley and Sons, 212p., 1947.

WILKS, S. S., *Mathematical statistics*, **J.Wiley**, New York, 1950.

WETHERILL, G. B.; GLAZEBROOK, K. D. **Sequential methods**. 3rd ed. London: Chapman and Hall, 263p., 1986.