

UNIVERSIDADE FEDERAL DE ALFENAS

JOSIANE DOS SANTOS ALVES

**Otimização do tamanho amostral na análise da qualidade de sementes de soja: uma
abordagem bayesiana**

Alfenas/MG

2019

JOSIANE DOS SANTOS ALVES

**OTIMIZAÇÃO DO TAMANHO AMOSTRAL NA ANÁLISE DA QUALIDADE DE
SEMENTES DE SOJA: UMA ABORDAGEM BAYESIANA**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas - MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.

Orientador: Prof. Dr. Luiz Alberto Beijo.

Coorientador: Prof. Dr. Fabrício Goecking Avelar.

Alfenas/MG

2019



JOSIANE DOS SANTOS ALVES

**“OTIMIZAÇÃO DO TÁMANHO AMOSTRAL NA ANÁLISE DA QUALIDADE DE
SEMENTES DE SOJA: UMA ABORDAGEM BAYESIANA”**

A Banca Examinadora, abaixo assinada, aprova a
Dissertação apresentada como parte dos requisitos para
a obtenção do título de Mestre em Estatística Aplicada
e Biometria pela Universidade Federal de Alfenas.
Área de Concentração: Estatística Aplicada e
Biometria

Aprovado em: 16 de dezembro de 2019.

Prof. Dr. Luiz Alberto Beijo

Instituição: UNIFAL-MG

Assinatura: _____

Dr. Thiago Taglialegna Salles

Assinatura: _____

Profª. Dra. Natália da Silva M. Fonseca

Instituição: UNIFAL-MG

Assinatura: _____

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas

A474o Alves, Josiane dos Santos
Otimização do tamanho amostral na análise da qualidade de sementes de soja: abordagem bayesiana / Josiane dos Santos Alves. – Alfenas/MG, 2020. 76f. : il. –

Orientador: Luiz Alberto Beijo.
Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Alfenas, 2020.
Bibliografia.

1. Inferência bayesiana. 2. Fator de Bayes. 3. Intervalo de credibilidade. 4. Controle de qualidade de sementes. I. Beijo, Luiz Alberto. II. Título.

CDD-519.5

*Dedico esse trabalho à Deus,
nosso pai supremo, que esteve
comigo em todos os momentos,
me dando forças para continuar
mesmo diante dos desafios, eu nada
seria sem a fé no nosso senhor.*

AGRADECIMENTOS

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

Agradeço a todas as pessoas que contribuíram para minha titulação e para meu crescimento pessoal. Sou o resultado da confiança e da força de cada um de vocês:

À Deus, em primeiro lugar, que me deu saúde, forças para prosseguir, sabedoria, sempre me guiando em todos os momentos para a conclusão deste trabalho.

À minha família, minha mãe Tereza, meu pai Vantuir, minhas irmãs Cláudia e Elaine e meus sobrinhos, Júlia e Bruno.

Aos diretores do Grupo Grão de Ouro, pela oportunidade, pela compreensão no tempo em que estive fora para os estudos e por terem cedido os dados e o laboratório para análise.

À Rosiane Campos, pela prestatividade, gratuidade e companheirismo, me ajudando nas minhas funções na empresa em que trabalho, enquanto estive ausente nos horários de estudo. Sem você eu não teria conseguido.

À todos os meus amigos, em especial, Eliane Corsini e Maurício Vilela, que estiveram ao meu lado, não só na etapa desse estudo, mas em todos os momentos que se passaram desde que nos conhecemos.

Aos amigos de trabalho, que sempre me passaram palavras de motivação, em especial à Liana, Marcelinho, Paula, Rafael e Willian, pela ajuda na amostragem e montagem do experimento.

Aos alunos/amigos do Programa de Pós Graduação em Estatística Aplicada e Biometria, em especial ao Luiz Henrique e ao Matheus, por me ajudarem com os experimentos.

Ao Prof. Dr. Luiz Alberto Beijo, meu orientador e ao Prof. Dr. Fabrício Goecking Avelar, meu coorientador, pela ajuda, ensinamentos, dedicação, paciência, por compartilharem comigo seus saberes, pelo aprendizado e por contribuírem para minha evolução.

Ao Governo Federal e à Universidade Federal de Alfenas, a todos os professores e servidores do Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Enfim, agradeço a todos que contribuíram para que eu pudesse chegar até aqui e concretizar o sonho da minha dissertação.

RESUMO

Para assegurar a qualidade da semente repassada ao produtor, as revendas de sementes têm adotado um controle de qualidade interno. Conforme recomendação das Regras para Análise de Sementes (RAS), para lotes acima de 60 sacos, orienta-se retirar amostras de 30 sacos, que são furados utilizando-se um calador, onde tal feito pode causar insatisfação ou rejeição por parte do cliente. Além disso, quanto maior a quantidade amostrada, maior o custo e os resíduos gerados com as análises. Logo, faz-se necessário estudos para a minimização da quantidade de sacos furados com a retirada de amostra, sem prejudicar as decisões quanto à utilidade do lote analisado. Para se tomar decisões com base na amostra utiliza-se o processo de inferência, a teoria bayesiana permite, por tratar o parâmetro de interesse de forma aleatória, uma interpretação mais realística do fenômeno estudado. Diante desses fatos, o presente estudo teve como objetivo verificar, utilizando a abordagem bayesiana, com qual tamanho amostral pode-se inferir sobre a porcentagem de germinação de sementes de soja, sem alterar o critério de decisão quanto a aceitação ou não do lote analisado. Para o experimento foram selecionados os três principais fornecedores de sementes de soja no ano de 2018, de uma revenda localizada na cidade de Alfenas. Utilizou-se na análise, *priori* não informativa e dois conjuntos de dados como *priori* informativa. Para avaliar o efeito da redução da amostra, dos 30 sacos analisados, foram selecionadas de forma aleatória, 5000 subamostras para cada tamanho amostral ($n_s = 28, 26, 24, 22, 20, 18, 16, 14, 12, 10, 8, 6, 4$). A decisão de rejeitar ou não o lote foi baseada nos limites e na amplitude do intervalo de credibilidade de 95% e no log Fator de Bayes. Diante dos resultados, observou-se que o uso da *priori* informativa, apresentou uma redução maior no tamanho amostral em comparação com o uso da *priori* não informativa, para a maioria dos lotes. Pode-se concluir que utilizando um tamanho amostral maior ou igual a 14 sacos, não se altera a decisão tomada comparativamente ao uso de amostra de tamanho 30 sacos, tende a reduzir a insatisfação por parte do produtor, bem como a diminuição dos gastos e dos resíduos gerados com as análises.

Palavras Chave: Inferência bayesiana, Fator de Bayes, Intervalo de credibilidade, Controle de qualidade de sementes.

ABSTRACT

To ensure the quality of the seed passed on to the producer, seed resellers have adopted an internal quality control. As recommended by the Rules for Seed Analysis (RAS), for lots above 60 bags, it is advisable to take samples from 30 bags, which are punctured using a nozzle, where this can cause dissatisfaction or rejection by the customer. In addition, the greater the amount sampled, the greater the cost and waste generated from the analyzes. Therefore, studies are needed to minimize the number of punctured bags with the withdrawal of samples, without jeopardizing decisions regarding the usefulness of the analyzed batch. In order to make decisions based on the sample, the inference process is used, the Bayesian theory allows, by treating the parameter of interest at random, a more realistic interpretation of the studied phenomenon. Given these facts, the present study aimed to verify, using the Bayesian approach, with which sample size one can infer about the germination percentage of soybean seeds, without changing the decision criterion as to whether or not to accept the analyzed lot. For the experiment, the three main soybean seed suppliers in 2018 were selected, from a reseller located in the city of Alfenas. In the analysis, non-informative textit priori and two data sets were used as informative textit priori. To assess the effect of reducing the sample, of the 30 bags analyzed, 5000 subsamples were selected at random for each sample size ($n_s = 28, 26, 24, 22, 20, 18, 16, 14, 12, 10, 8, 6, 4$). The decision to reject the lot or not was based on the limits and the amplitude of the 95 % credibility interval and the Bayes Factor log. In view of the results, it was observed that the use of the informative textit priori, presented a greater reduction in the sample size in comparison with the use of the non-informative textit priori, for most lots. It can be concluded that using a sample size greater than or equal to 14 bags, the decision made compared to the use of a sample of size 30 bags does not change, tends to reduce the dissatisfaction on the part of the producer, as well as the reduction of expenses and costs. waste generated from the analyzes.

Key words: Bayesian Inference, Bayes Factor, Credibility Range, Seed Quality Control.

LISTA DE TABELAS

Tabela 1 –	Intensidade de amostragem, adaptado de RAS (BRASIL, 2009, pag. 26)	24
Tabela 2 –	Tolerâncias máximas admitidas entre os resultados das repetições do mesmo teste, adaptado de RAS (BRASIL, 2009, pag. 383)	26
Tabela 3 –	Tabela de evidências de Jeffreys (1961) para o Log do fator de Bayes .	37
Tabela 4 –	Nomenclaturas dos lotes de sementes de soja analisados.	43
Tabela 5 –	Valores do tamanho amostral, dos quartis dos dados e da distribuição <i>à priori</i> , dos hiperparâmetros, da média dos dados e da média <i>à priori</i> para os Fornecedores 1, 2 e 3.	48
Tabela 6 –	Resumo dos resultados obtidos com a análise dos 12 lotes com <i>Priori</i> Não Informativa (PNI), <i>Priori</i> Informativa com dados do Termo de Conformidade (PITC) e <i>Priori</i> Informativa com dados das análises do laboratório (PIL).	62
Tabela 7 –	Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos cultivares dentro dos Fornecedores 1 e 2.	71
Tabela 8 –	Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos lotes dentro dos cultivares do Fornecedor 1.	72
Tabela 9 –	Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos lotes dentro dos cultivares do Fornecedor 2.	73

LISTA DE FIGURAS

Figura 1 –	Soja (<i>Glycine max</i>).	20
Figura 2 –	Fluxograma do controle de qualidade da soja.	22
Figura 3 –	Retirada de amostra de semente acondicionada.	44
Figura 4 –	Montagem do experimento, papel de germinação e placa contadora de sementes.	44
Figura 5 –	Montagem do experimento, rolinhos de papel a serem colocados para germinação.	45
Figura 6 –	Câmara Germinadora BOD.	45
Figura 7 –	Sementes após 5 dias na germinadora BOD, à esquerda - sementes vigorosas e à direita - sementes não vigorosas.	46
Figura 8 –	Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G e J), Log do Fator de Bayes (B, E, H e K) e Amplitude (C, F, I e L), para os diferentes tamanhos amostrais do Fornecedor 1, Cultivar A, lotes 01 e 02.	51
Figura 9 –	Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G e J), Log do Fator de Bayes (B, E, H e K) e Amplitude (C, F, I e L), para os diferentes tamanhos amostrais do Fornecedor 1, Cultivar B, lotes 03 e 04.	53
Figura 10 –	Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M, e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 2, Cultivar C, lotes 05 e 06.	55
Figura 11 –	Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 2, Cultivar D, lotes 07 e 08.	57
Figura 12 –	Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I e L, O e R), para os diferentes tamanhos amostrais do Fornecedor 3, Cultivar E, lotes 09 e 10.	59

Figura 13 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 3, Cultivar F, lotes 11 e 12. 61

LISTA DE SÍMBOLOS

$O(\mathbf{H}_0, \mathbf{H}_1 \mathbf{X})$	- Chance <i>à posteriori</i> em favor de H_0 .
$O(\mathbf{H}_0, \mathbf{H}_1)$	- Chance <i>à priori</i> em favor de H_0 .
$p(\boldsymbol{\theta} \mathbf{X})$	- Distribuição <i>à posteriori</i> para o vetor de parâmetros $\boldsymbol{\theta}$.
$p(\boldsymbol{\theta})$	- Distribuição <i>à priori</i> para o vetor de parâmetros $\boldsymbol{\theta}$
$\beta(a_0, b_0)$	- Distribuição Beta parâmetros a_0 e variância b_0 .
$p(\boldsymbol{\pi})$	- Distribuição de probabilidade Beta para o vetor de parâmetro $\boldsymbol{\pi}$.
$BF_{0,1}$	- Fator de Bayes à favor da hipótese - H_0
$BF_{1,0}$	- Fator de Bayes contra a hipótese - H_0
$p(\mathbf{X} \boldsymbol{\theta})$	- Função de verossimilhança do parâmetro $\boldsymbol{\theta}$
H_1	- Hipótese alternativa.
H_0	- Hipótese nula.
$\log BF_{0,1}$	- Log do Fator de Bayes.
\log	- Logaritmo na base 10.
$E(\boldsymbol{\pi} \mathbf{X})$	- Média <i>à posteriori</i> da distribuição Beta.
$\boldsymbol{\pi}$	- Parâmetro da proporção.
$\boldsymbol{\theta}$	- Parâmetro de um modelo
n	- Tamanho amostral
\mathbf{X}	- Vetor de distribuições da variável explanatória
\mathbf{x}	- Vetor de observações da variável explanatória
$\boldsymbol{\theta}$	- Vetor de parâmetros de um modelo

LISTA DE ABREVIATURAS E SIGLAS

- BOD** - *Biochemical Oxygen Demand*
- CQ** - Controle de Qualidade
- CV** - Coeficiente de Variação
- CQI** - Controle de Qualidade Interno
- DBC** - Delineamento em Blocos Casualizado
- DIC** - Delineamento Inteiramente Casualizado
- ICr** - Intervalo de Credibilidade
- PIB** - Produto Interno Bruto
- PIL** - *Priori* Informativa com dados do laboratório
- PITC** - *Priori* Informativa com dados do Termo de Conformidade
- PNI** - *Priori* não informativa
- RAS** - Regras para análise de sementes
- TC** - Termo de Conformidade
- 1AL01** - Fornecedor 1, Cultivar A, Lote 01
- 1AL02** - Fornecedor 1, Cultivar A, Lote 02
- 1BL03** - Fornecedor 1, Cultivar B, Lote 03
- 1BL04** - Fornecedor 1, Cultivar B, Lote 04
- 2CL05** - Fornecedor 2, Cultivar C, Lote 05
- 2CL06** - Fornecedor 2, Cultivar C, Lote 06
- 2DL07** - Fornecedor 2, Cultivar D, Lote 07
- 2DL81** - Fornecedor 2, Cultivar D, Lote 08
- 3EL09** - Fornecedor 3, Cultivar E, Lote 09
- 3EL10** - Fornecedor 3, Cultivar E, Lote 10
- 3FL11** - Fornecedor 3, Cultivar F, Lote 11

3FL12 - Fornecedor 3, Cultivar F, Lote 12

SUMÁRIO

1	INTRODUÇÃO	17
1.1	OBJETIVOS	18
1.1.1	Objetivos Específicos	19
2	REFERENCIAL TEÓRICO	20
2.1	CULTIVO DA SOJA	20
2.2	CONTROLE DE QUALIDADE DE SEMENTES	21
2.2.1	Amostragem de sementes	23
2.2.2	Teste de germinação	24
2.2.3	Tolerância do teste baseada na amplitude	25
2.3	ESTATÍSTICA EXPERIMENTAL	26
2.3.1	Princípios básicos da experimentação	27
2.3.2	Delineamentos experimentais	28
2.4	ESTUDOS DE DETERMINAÇÃO DE TAMANHO AMOSTRAL	28
2.5	ABORDAGEM BAYESIANA	31
2.5.1	Informações <i>à priori</i>	32
2.5.2	Distribuição <i>à posteriori</i> para a proporção	33
2.5.3	Intervalo de credibilidade	35
2.5.4	Fator de Bayes	35
2.5.5	Aplicações em estudos experimentais	37
3	MATERIAL E MÉTODOS	41
3.1	INFORMAÇÕES PARA MONTAGEM DO EXPERIMENTO	41
3.2	DADOS UTILIZADOS NA ANÁLISE	43
3.3	Distribuição <i>à posteriori</i> para a proporção	46
3.3.1	Distribuições <i>à priori</i>	47
3.3.2	Avaliação da precisão e da tomada de decisão	48
4	RESULTADOS E DISCUSSÃO	50
5	CONCLUSÕES	65
6	SUGESTÕES DE TRABALHOS FUTUROS	66
	REFERÊNCIAS	67

1 INTRODUÇÃO

O agronegócio tem sido um dos principais setores que influencia no PIB nacional, gerando cerca de 37% de todos os empregos do país, fazendo com que o Brasil ocupe grande destaque na exportação de grãos (LIMA, POZO E FREITAS, 2017). Segundo projeções do Ministério da Agricultura e Pecuária, será o principal exportador de soja entre 2022 e 2032, com participação no mercado mundial de mais de 44% do volume de exportação (Companhia Nacional de Abastecimento (CONAB), 2019).

O setor agropecuário apresenta elevado potencial econômico, criando assim inúmeras oportunidades de negócios às empresas que apresentam uma proposta de valor, que vá de encontro com as reais necessidades do produtor rural. O mercado de insumos agrícolas está cada vez mais competitivo, portanto, para que as empresas se mantenham na concorrência, além de um preço acessível, faz-se necessário ofertar produtos e serviços de qualidade, que atendam e muitas vezes surpreendam os clientes diante de suas necessidades e desejos.

Nesse sentido, destacam-se as empresas que possuem um Controle de Qualidade (CQ), que é o conjunto de procedimentos realizados por profissionais capacitados, destinados a assegurar que os ensaios necessários e relevantes sejam executados, assegurando que o produto não seja disponibilizado para uso e venda até que o mesmo cumpra com as especificações pré-estabelecidas (KRZYZANOWSKI *et al.*, 2008).

O CQ se torna um bom aliado no mercado de sementes, é importante e necessário o acompanhamento em todas as fases do sistema de produção e comercialização, buscando superar limitações impostas pelos diversos fatores que podem afetar a qualidade (ROCHA *et al.* 2017). Tais fatores podem alterar a porcentagem de germinação até a sua chegada ao estoque da revenda, que são, umidade, temperatura, quebra, entre outros. O uso de sementes de baixa qualidade compromete a produtividade da lavoura, causando sérios prejuízos financeiros ao produtor, gerando insatisfação e a possível perda do cliente.

Para assegurar a qualidade da semente repassada ao produtor, as revendas têm adotado um Controle de Qualidade Interno (CQI), realizando a verificação da porcentagem de germinação das sementes garantida pelo fornecedor. As normas para testes de germinação seguidas pelos fornecedores de sementes são regidas pelas Regras para Análise de Sementes (RAS) (BRASIL 2009). De acordo com as instruções, é necessário retirar amostras de uma porcentagem pré-estabelecida da quantidade de sacos de cada lote.

A maioria dos lotes recebidos nas vendas possuem acima de 60 sacos e conforme a RAS é necessário retirada de amostra de 30 sacos por lote recebido. Para tal feito, utiliza-se um calador simples para furar os sacos e retirar as amostras. O furo ocasionado no saco é tampado com um adesivo destinado para esse fim, no entanto, os sacos utilizados no experimento são os mesmos que serão comercializados ao produtor.

Della Lucia *et al.* (2007), explica que o efeito das características da embalagem influencia significativamente, na aceitação e intenção de compra do consumidor, ou seja, o fato do saco ser furado, muitas das vezes pode causar insatisfação ou rejeição por parte do cliente ao receber o produto. Além disso, um tamanho amostral maior pode elevar os custos e os resíduos gerados com as análises. Nesse sentido, são necessários estudos que minimizem a quantidade de sacos furados para retirada de amostra.

As análises das propriedades germinativas das sementes são realizadas por meio de experimentos. A estatística experimental é uma importante ferramenta para tais avaliações, trata-se de uma ciência que utiliza a experimentação para analisar, avaliar, comparar e medir características de um determinado tratamento ou produto, na busca de tomada de decisão (GOMES, 2009).

Geralmente, os testes realizados nas análises estatísticas de dados experimentais são fundamentados na teoria frequentista, no entanto a teoria bayesiana possui uma forma de inferência estatística que utiliza informação *à priori* à coleta de dados, que junto com a função de verossimilhança traz todas as informações sobre o parâmetro de interesse, permitindo, na maioria das vezes, chegar a um resultado com um bom nível de precisão com um menor tamanho amostral (CARVALHO; BEIJO e MUNIZ, 2017).

Diante destes fatos, o presente estudo teve como objetivo utilizar a teoria bayesiana na análise dos dados, com diferentes tipos de *priori*, buscando verificar se reduzir o tamanho amostral, influencia significativamente na precisão da estimação da resposta porcentagem de germinação de sementes de soja e na aceitação ou não do lote analisado.

1.1 OBJETIVOS

O estudo teve como objetivo averiguar se o uso da inferência bayesiana, com diferentes tipos de *priori*, possibilita a verificação da porcentagem de germinação de sementes de soja,

com um tamanho amostral menor, sem grandes perdas na precisão.

1.1.1 Objetivos Específicos

Os objetivos gerais foram divididos nos seguintes objetivos específicos:

- avaliar diferentes tipos de estruturas de distribuição *à priori* na inferência bayesiana para análise de germinação de sementes de soja;
- utilizar o intervalo de credibilidade e a amplitude do intervalo de credibilidade de 95% e o Log do Fator de Bayes para avaliação da precisão e tomada de decisão;
- avaliar se a redução na quantidade de sacos de sementes de soja amostrados, para verificação da porcentagem de germinação garantida pelo fornecedor, influencia significativamente na precisão;
- selecionar o tamanho amostral ótimo dentre os tamanhos $n_s = (4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28)$.

2 REFERENCIAL TEÓRICO

Nesta seção, serão apresentadas as revisões de literatura relacionadas ao cultivo da soja, controle de qualidade, regras para análise de sementes e a teoria bayesiana.

2.1 CULTIVO DA SOJA

A soja (*Glycine max*) (Figura 1) é uma leguminosa rica em elementos que trazem benefícios para a saúde do ser humano e de animais, como aminoácidos, proteínas, fósforo, potássio, ferro, zinco, cálcio e vitaminas B e E (BEHRENS e SILVA, 2004).



Figura 1 – Soja (*Glycine max*).

Fonte: <http://ruralpecuaria.com.br>, acesso em 16/03/2020.

A soja faz parte do conjunto de atividades agrícolas com maior destaque no mercado mundial e no período entre os anos agrícolas 2000/01 e 2013/14, a soja e o milho foram as culturas que apresentaram maior crescimento, tanto em consumo quanto em produção. É a cultura com maior produção entre os grãos do país e a maior área cultivada, com previsão de atingir na safra 2018/19, em torno de 114 milhões de toneladas e área plantada de 35,8 milhões de hectares (CONAB, 2019). O complexo soja, composto pela soja em grãos e seus derivados, como óleo e farelo de soja, foi o principal produto exportado em 2017, representando U\$31,7

bilhões e cerca de 33% das exportações do agronegócio (KRZYZANOWSKI; FRANÇA NETO e HENNING, 2018).

Conforme Hirakuri e Lazarotto (2014), cerca de 90% dos grãos consumidos são direcionados ao processo de esmagamento, gerando farelo e óleo de soja, em uma proporção próxima a (80/20), sem considerar as perdas. O principal produto gerado nesse processo é o farelo de soja, que junto com o milho, constituirá matéria-prima essencial para a fabricação de rações.

França Neto *et al.* (2007), Krzyzanowski *et al.* (2008), Scheeren *et al.* (2010), Amorim *et al.* (2011) e Rocha *et al.* (2017), destacam a importância de pesquisas dirigidas para a utilização de sementes de soja de alta qualidade, que possibilitam aumento da produtividade.

2.2 CONTROLE DE QUALIDADE DE SEMENTES

Defini-se Controle de Qualidade (CQ) como o conjunto de procedimentos realizados por profissionais capacitados, destinados a assegurar que os ensaios necessários e relevantes sejam executados, assegurando que o produto não seja disponibilizado para uso e venda, até que o mesmo cumpra com as especificações pré-estabelecidas (KRZYZANOWSKI *et al.*, 2008).

O CQ é um setor de extrema importância no mercado de sementes. A utilização de sementes vigorosas, de alta qualidade, associada a boas práticas de semeadura contribuem para o sucesso da lavoura e para que máximas produtividades sejam alcançadas. Sementes vigorosas são aquelas que possuem propriedades que determinam o seu potencial para uma emergência rápida e uniforme e o desenvolvimento de plântulas normais sob ampla diversidade de condições de ambiente (FRANÇA NETO *et al.* 2007).

As empresas sementeiras, que são empresas responsáveis pela produção e preparo de sementes, obrigatoriamente, fazem a verificação do vigor da semente comercializada, ou seja, todos os lotes devem conter um documento denominado "Termo de Conformidade (TC)", onde se atesta a porcentagem de germinação que aquele lote apresentou em análises laboratoriais. A análise das propriedades germinativas de sementes é fundamentada por um documento do Ministério da Agricultura denominado Regras para Análise de Sementes (RAS) (BRASIL, 2009).

De acordo com Krzyzanowski; França-Neto e Henning (2018), alguns fatores podem

influenciar uma deterioração, que é qualquer modificação degenerativa da semente, após ter atingido sua qualidade fisiológica máxima. Esses fatores são, principalmente, umidade, temperatura, quebra, condições e tempo de armazenamento, causando a perda da qualidade germinativa, conforme Rocha *et al.* (2017), as sementes ou grãos podem sofrer alterações em sua composição química em razão do ambiente de transporte e estocagem. Portanto, é aconselhável a verificação da qualidade germinativa da semente, em todo processo de produção e comercialização.

Na maioria das situações, a semente é comercializada para revenda e essa funciona como intermediadora para a comercialização entre as empresas sementeiras e os produtores rurais. Para assegurar a qualidade da semente repassada ao produtor, as revendas têm adotado um Controle de Qualidade Interno (CQI), buscando verificar a porcentagem de germinação do lote. Tal verificação traz vantagens tanto para o produtor quanto para a revenda.

As empresas sementeiras obrigatoriamente devem seguir a RAS nas análises das propriedades germinativas das sementes. Como as revendas desejam realizar as análises nos mesmos padrões das empresas sementeiras, elas também seguem a RAS, no entanto, não obrigatoriamente, por se tratar apenas de um CQI, conforme o fluxograma apresentado na Figura 2:

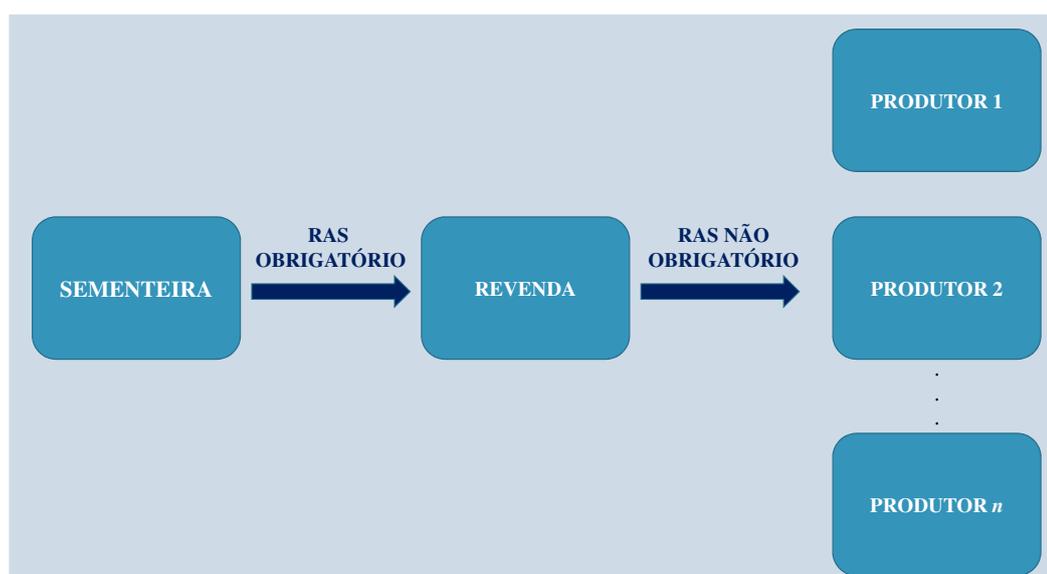


Figura 2 – Fluxograma do controle de qualidade da soja.

Fonte: do autor.

2.2.1 Amostragem de sementes

Em muitas situações, torna-se impossível ou inviável analisar todos os elementos de uma população, principalmente por fatores éticos, por envolver populações que tendem ao infinito ou populações que não possibilitam a análise sobre todos os elementos por envolver o consumo do objeto em si. O processo de amostragem consiste na obtenção de amostras adequadas para análise de um todo. Uma amostra representativa carrega consigo as características da população (MORENTIN E BUSSAD, 2004).

O processo para se obter uma amostragem adequada para análise de sementes é descrito por RAS (BRASIL, 2009), que explica que o lote de sementes deve ser disposto de tal forma que possua no mínimo duas faces expostas, com espaçamentos entre pilhas e entre pilhas e paredes, que permitam a amostragem representativa do mesmo. O lote para a retirada da semente para análise, deve ser o mais homogêneo possível e caso haja suspeita de ser excessivamente heterogêneo, o laboratório deverá realizar testes de heterogeneidade. O encarregado da amostragem deve solicitar ao proprietário das sementes, ou seu representante, informações completas sobre o lote em questão.

As amostras devem ser retiradas do lote por meio de caladores, que são instrumentos apropriados para esse fim. Os caladores para sementes acondicionadas em sacos devem ter o comprimento mínimo aproximado ao da diagonal dessas embalagens, com o diâmetro variando de 1,25-2,50 cm e deve ser inserido diagonalmente nos sacos, até atingir a massa de sementes, num ângulo de 30 graus. Estando no interior da massa, deve ser girado algumas vezes ou levemente agitado até que fique completamente cheio de sementes. Em seguida, deve-se despejar as sementes, em um recipiente apropriado. Após a retirada do calador, deve-se fechar o orifício do saco com uma fita adesiva destinada para esse fim.

Para sementes acondicionadas, um número suficiente de sacos é apresentado na Tabela 1.

Tabela 1 – Intensidade de amostragem, adaptado de RAS (BRASIL, 2009, pag. 26)

Nº de sacos por lote	Nº de amostras simples
1 - 4	3 amostras de cada saco
5 - 8	2 amostras de cada saco
9 - 15	1 amostra de cada saco
16 - 30	15 amostras no total
31 - 59	20 amostras no total
60 ou mais	30 amostras no total

Fonte: do autor.

As amostras devem ser embaladas e identificadas de maneira a especificar o lote ao qual elas pertencem. A embalagem individual a ser usada para a amostra deve ser de material resistente, para não se romper durante o envio ao laboratório. As embalagens individuais devem ser acondicionadas de maneira a evitar danos durante o transporte, sendo preservadas contra o excesso de calor, umidade e contaminação.

O responsável pela tomada das amostras deve remetê-las, sem demora, ao Laboratório de Análise de Sementes. Quando não for possível realizar a análise após a retirada da amostra, elas deverão ser mantidas em locais climatizados de tal maneira que as alterações na qualidade da semente como dormência, grau de umidade e porcentagem de germinação sejam as mínimas possíveis. Após a análise, as sementes que sobraram deverão ser mantidas em arquivo por um período equivalente a validade do teste de germinação, que geralmente é correspondente a validade do lote.

2.2.2 Teste de germinação

Para verificar a viabilidade do lote, realiza-se o teste de germinação, verificando a porcentagem de germinação que o lote apresenta. A RAS (Brasil, 2009) define germinação de sementes em teste de laboratório como a emergência e desenvolvimento das estruturas essenciais do embrião, demonstrando sua aptidão para produzir uma planta normal sob condições favoráveis de campo. Uma plântula é considerada normal quando mostra potencial para continuar seu desenvolvimento e da origem a plantas normais, quando desenvolvidas sob condições favoráveis. Nos testes de laboratório a porcentagem de germinação de sementes corresponde à proporção do número de sementes que produziu plântulas classificadas como normais.

As sementes são colocadas em substratos, que são papéis que apresentam capacidade de retenção de água, resistência, não devendo conter substâncias tóxicas em quantidades que possam causar dano às raízes das plântulas. Os tipos de papel comumente utilizados como substrato são o mata-borrão, o papel toalha e o de filtro. Para que se calcule a quantidade de água a ser adicionada é conveniente utilizar a relação volume de água (mL) por peso do substrato (g), adicionando um volume de água em quantidade equivalente à 2 a 3 vezes o peso do substrato.

Para que as sementes sejam colocadas em quantidade correta e com espaçamento adequado, utiliza-se o contador de sementes, que é uma placa com tamanho próximo aos dos papéis de germinação, perfurada com 50 orifícios, de tamanho e forma semelhantes ao das sementes a serem contadas. As sementes são distribuídas sobre a placa superior e verifica-se se todos os orifícios contêm apenas uma semente. Após a verificação a placa é retirada.

O equipamento para germinação utilizado é uma câmara germinadora BOD (do inglês, *Biochemical Oxygen Demand*), que é uma câmara que proporciona simulações de temperatura e luz para as condições ideais de germinação de sementes. A temperatura ajustada na análise é de 25°C constantes.

2.2.3 Tolerância do teste baseada na amplitude

Para que o resultado de um teste de germinação possa ser considerado satisfatório e válido para emissão do resultado, é preciso que a variação entre as porcentagens de germinação das sementes esteja dentro das tolerâncias máximas permitidas. Para se fazer essa verificação, determina-se a média das repetições, em seguida localiza-se esse valor na Tabela apresentada na página 383 das RAS (BRASIL, 2009) e verifica-se a respectiva tolerância máxima permitida. Se a variação entre as porcentagens de germinação das repetições for inferior ou igual a essa tolerância, a média representará o resultado do teste de germinação.

O cálculo da tolerância do teste de germinação é válido para repetições de 100 sementes, obtidas ao acaso da mesma amostra de trabalho. Se necessário, agrupar sub-repetições de 50 para formar repetições de 100 sementes. Recomenda-se utilizar 2,5% de probabilidade.

A tolerância do teste aumenta a medida que a porcentagem de germinação diminui, para médias de germinação de 51% à 99%, conforme a Tabela 2. Logo, diante da Tabela 2

apresentada, considerando-se um valor máximo para a amplitude igual a 7%, abrangeria médias de germinação entre 4% e 97% e a tolerância máxima permitida seria atendida.

Tabela 2 – Tolerâncias máximas admitidas entre os resultados das repetições do mesmo teste, adaptado de RAS (BRASIL, 2009, pag. 383)

51 - 99 (%)	Menor que 51%	Amplitude (%)
99	2	5
98	3	6
97	4	7
96	5	8
95	6	9
94	7	10
93	8	10
92	9	11
91	10	11
.	.	.
.	.	.
.	.	.
50	49	20
51	50	20

Fonte: do autor.

As análises das propriedades germinativas das sementes são realizadas por meio de experimentos, sendo as técnicas utilizadas para o planejamento, execução e análise de experimentos, fundamentadas na Estatística Experimental.

2.3 ESTATÍSTICA EXPERIMENTAL

Conforme Fachin (2004), o método experimental consiste na manipulação das variáveis, de maneira pré-determinada e seus efeitos são suficientemente controlados e conhecidos pelo pesquisador e tem o objetivo de descobrir conexões causais. Na experimentação, faz-se necessário abandonar as próprias opiniões e as opiniões alheias, considerando os resultados como eles se apresentam.

As técnicas experimentais se aplicam a diferentes áreas de estudo, como na agronomia, medicina, engenharia, psicologia e tem sido usada como base do progresso do conhecimento nas áreas científicas consistindo em estabelecer hipóteses científicas, planejar experimentos para verificação dessas hipóteses, executar o experimento, registrar os dados de interesse, analisar os dados gerando conhecimento sobre a hipótese testada e estabelecer uma conclusão (GOMES,

2009).

Grande parte das técnicas de experimentação atuais foram formalizadas pelo estatístico Ronald A. Fischer (1890 - 1962), que as desenvolveu na Estação de Agricultura de Rothamstead, na Inglaterra, onde o estudo agrícola da experimentação, deu origem a vários termos técnicos. Logo percebeu-se sua utilidade em diversas áreas (VIEIRA, 1989).

Conforme Porto (2014), para se obter confiança e credibilidade em uma experimentação, faz-se necessário o planejamento da pesquisa. O pesquisador deve ter bem definido seus objetivos e o conhecimento técnico específico não estatístico sobre o problema deve ser valorizado.

2.3.1 Princípios básicos da experimentação

Gomes (2009) explica que ao se realizar um experimento, podem existir efeitos de fatores não controlados, os quais podem alterar os resultados obtidos. Tais efeitos recebem o nome de variação aleatória ou erro experimental e podem ser tão significativos a ponto de influenciar nos resultados do experimento, fazendo com que um tratamento seja considerado melhor ou pior por simples acaso, por ter sido favorecido por esses fatores não controlados. Portanto, ao se planejar um experimento, alguns princípios devem ser atendidos. Tais princípios recebem o nome de repetição, casualização e controle local. Ao considera-los, o pesquisador diminui o erro experimental gerado na análise e mantém a aleatoriedade dos dados.

Repetição

A ideia em experimentação é comparar grupos e não unidades. As unidades experimentais do mesmo grupo recebem o nome de repetições ou réplicas e tem por finalidade propiciar a obtenção de uma estimativa do erro experimental. Do ponto de vista estatístico é desejável que os experimentos tenham um grande número de repetições, porém em alguns casos, fatores éticos e/ou a disposição de recursos limitam a quantidade de repetições (GOMES, 2009).

Casualização

O princípio da casualização consiste na distribuição aleatória dos tratamentos nas parcelas e se baseia no sorteio das unidades experimentais, para que as diferenças entre dois

grupos sejam explicadas apenas pelo tratamento, quando os grupos são iguais. O princípio assegura a validade da estimativa do erro experimental, tornando-os independentes (GOMES, 2009).

Controle local

Porto (2014) explica que para se realizar um experimento, as unidades experimentais devem ser homogêneas e em várias situações é difícil isso acontecer. O controle local consiste em subdividir unidades heterogêneas em subunidades homogêneas, buscando reduzir o erro experimental e dar mais confiabilidade ao delineamento. É um princípio que agrupa unidades experimentais tão uniformes quanto possível, que constituirão um bloco. A variação dentro dos blocos deve ser a menor possível (GOMES, 2009).

2.3.2 Delineamentos experimentais

Ramalho; Ferreira e Oliveira (2012) definem delineamento experimental como a forma em que os tratamentos são atribuídos as unidades experimentais. São procedimentos que apresentam alternativas para proceder à casualização. Os principais delineamentos experimentais são Delineamentos Inteiramente Casualizado (DIC), Delineamentos em quadrado latino e Delineamentos em Blocos Casualizados (DBC).

O DBC é considerado o tipo mais importante e mais utilizado na experimentação, por permitir a análise dos dados, mesmo que as unidades experimentais apresentem heterogeneidade. O controle local nesse tipo de experimento é feito repartindo-se a área experimental ou o material experimental heterogênea(o) em subáreas homogêneas denominadas bloco. Cada bloco recebe todos os tratamentos através de sorteio e devem ser o mais uniforme possível (GOMES, 2009).

2.4 ESTUDOS DE DETERMINAÇÃO DE TAMANHO AMOSTRAL

Nesta seção, serão apresentados alguns estudos envolvendo otimização do tamanho amostral, usando teoria Frequentista, com o intuito de evidenciar a importância para algumas áreas, em diminuir o tamanho amostral, por viabilidade e/ou necessidade. A literatura

apresenta poucos trabalhos relacionados a otimização do tamanho amostral na análise de qualidade germinativa de sementes de soja, tanto para teoria frequentista, quanto para bayesiana.

Cargnellutti Filho e Gonçalves (2011) utilizaram a teoria frequentista para avaliarem 28 genótipos de soja por meio de cinco experimentos, com o intuito de estimar o número de repetições necessárias na predição do valor real dos genótipos. Os autores afirmaram que tal estimativa é importante por minimizar recursos humanos e financeiros e por aproveitar melhor as informações existentes. Foram avaliados os caracteres de produtividade e de morfologia. Utilizou-se a análise de variância, estimou-se o coeficiente de repetibilidade e calculou-se o número de repetições necessário na predição do valor real dos genótipos. Experimentos com três repetições identificam genótipos de soja superiores com 17,52% (altura de inserção de primeira vagem) a 61,07% (número de nós por planta) de exatidão no prognóstico de seu valor real. Foi concluído que para melhorar a precisão da avaliação de caracteres em plantas de soja, necessita-se utilizar maior número de repetições.

Através de um experimento, Estefanel *et al.* (1984) objetivaram determinar o tamanho da amostra de plantas de soja, capaz de estimar a média com um intervalo de confiança máximo de 10% do valor da média, com probabilidade de erro de 5%, para onze características agronômicas da soja, além de verificar a influência dos fatores densidade de planta, espaçamento entre fileiras, ciclo de cultivar, época de semeadura e nível de fertilidade do solo, sobre o tamanho da amostra. A pesquisa foi realizada visando minimizar os custos com as análises e determinar um tamanho amostral confiável para informações mais precisas. Foi concluído que o tamanho da amostra é influenciado significativamente pelos fatores densidade de planta, ciclo da cultivar, época de semeadura e níveis de fertilidade do solo e não se pode utilizar o mesmo tamanho de amostra para estimar, com a mesma precisão, diferentes características agronômicas da soja.

Num experimento a campo, Cargnellutti Filho *et al.* (2015) determinaram o tamanho amostral para a estimação da média de caracteres morfológicos e produtivos de aveia preta (*avena sativa*), com erros de estimação fixados em 2%, 4%, 6%, 8%, 10%, 15% e 20% da média, com grau de confiança de 95% e verificando-se a variabilidade do tamanho de amostra entre caracteres e entre épocas de avaliação. Os autores concluíram que 47 plantas são suficientes para a estimação da média dos caracteres morfológicos e produtivos, para um erro de estimação máximo de 20% da média e grau de confiança de 95%. Concluíram também que

essas combinações de erro de estimação (precisão) e tamanho de amostra podem ser utilizadas por pesquisadores para o planejamento de seus experimentos, de acordo com a precisão desejada.

Toebe *et al.* (2017) conduziram um experimento, para determinar o tamanho da amostra necessário para a estimação da média de caracteres, assumindo-se erros de estimação iguais a 1, 2, 3, ..., 20% da média, além de avaliar as relações lineares existentes entre caracteres de plantas de *Crotalaria spectabilis*, que é utilizada como cultura de cobertura ou de adubação verde e apresentam elevada capacidade de produção de biomassa e fixação de nitrogênio. Os caracteres avaliados nas plantas foram: altura da planta, número de ramos produtivos, massa seca da parte aérea, massa seca de vagens, número de vagens, peso médio de cada vagem, número de sementes, número médio de sementes por vagem, massa de cem grãos e peso total de sementes. Após a determinação do tamanho de amostra, foram construídos histogramas de frequência e gráficos de dispersão entre os caracteres avaliados. Foi destacado que o tamanho amostral é inversamente proporcional ao erro permitido *à priori* pelo pesquisador, sendo maior de acordo com o aumento da variabilidade dos dados e da confiabilidade desejada. Os autores concluíram que, em *Crotalaria spectabilis*, são necessárias 121 plantas para a estimação da média da maioria dos caracteres, com erro de estimação máximo de 10% da média.

Santos *et al.* (2010) desenvolveram um estudo com o objetivo de estimar o tamanho da amostra, para alface (*Lactuca sativa*) cultivada em diferentes ambientes. Foram avaliadas a fitomassa fresca e a fitomassa seca. Para cada variável, em cada linha de cultivo, foram estimadas a média, a variância e o coeficiente de variação (CV). Nos casos em que as variâncias das linhas foram homogêneas, calculou-se um CV% médio para cada ambiente e este foi utilizado para o cálculo do tamanho de amostra. Nos casos em que as variâncias foram heterogêneas, utilizou-se o maior CV% do ambiente para o cálculo do tamanho da amostra. Foram realizados testes F, ao nível de significância de 5%, para comparar as variâncias entre os ambientes de cultivo, dentro de cada variável e cada ano de experimentação. Estimou-se o tamanho ideal de amostra para cada linha de cultivo. Os autores destacaram que a determinação do tamanho da amostra é uma importante estratégia no momento do planejamento do experimento para reduzir o erro experimental, ou seja, quanto maior o tamanho da amostra, maior a precisão. No entanto, salientam também que amostras grandes aumentam os gastos e o tempo para realização da amostragem. Concluiu-se que o tamanho de amostra para alface varia em função do ambiente, da época de cultivo, da cultivar utilizada e da

variável observada. Para a semiamplitude do intervalo de confiança de 10%, admitiu-se amostrar 30, 40 e 33 plantas na linha de cultivo para estufa, túnel e campo, respectivamente para a fitomassa fresca das plantas. Para a fitomassa seca das plantas é possível amostrar, na linha, 27, 29 e 32 plantas, respectivamente, para estufa, túnel e campo.

Existem vários procedimentos estatísticos que tratam da análise dos dados (GOMES, 2009). Uma alternativa para se trabalhar com amostras menores é a aplicação da inferência bayesiana, visto que não existem pressuposições acerca do tamanho amostral (BRIGHENTI; RESENDE E BRIGHENTI, (2011); CARVALHO; BEIJO e MUNIZ, (2017)). A teoria bayesiana apresenta métodos de estimação, como os intervalos de credibilidade e métodos de decisão como os testes de hipótese e o fator de Bayes.

2.5 ABORDAGEM BAYESIANA

De acordo com Rossi (2011), a teoria bayesiana foi desenvolvida por Thomas Bayes em meados do século XVIII, apresentada como uma nova metodologia de análise de dados, fundamentada em probabilidades condicionais, baseada principalmente em um conhecimento *à priori* em relação as incertezas envolvidas no estudo. Enquanto no modelo frequentista, o parâmetro é um vetor escalar ou um vetor desconhecido, porém fixo, no modelo Bayesiano, é considerado um escalar ou vetor aleatório.

Alguns estudos como o de Harold Jeffreyes's (1981 - 1989), evidenciaram a importância da utilização do método Bayesiano em problemas científicos, porém estes estudos exigiam a resolução de cálculos complexos, aos quais a falta de recursos computacionais na época, fizeram com que o enfoque Bayesiano não fosse utilizado. No entanto, esse não é mais considerado um problema, pois, surgiram vários programas para efetuar as computações estatísticas (BERGER, 1985).

De acordo com Paulino; Turkman e Murteira (2003), a utilização de informação *à priori* em inferência bayesiana requer a especificação de uma distribuição *à priori* para a quantidade de interesse θ (parâmetro). Esta distribuição deve representar probabilisticamente o conhecimento que se tem sobre θ antes da realização do experimento.

A distribuição *à posteriori* de um parâmetro θ contém toda a informação probabilística a respeito deste parâmetro e sua função de densidade *à posteriori* é a melhor descrição do

processo de inferência.

Considere uma quantidade de interesse desconhecida θ não observável. A informação de que dispomos sobre θ , resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ , dada por, $p(x | \theta)$. O teorema de Bayes é o modelo utilizado para determinar este aumento de informação (GELMAN *et al.* 2014), sendo assim, tem-se:

$$p(\theta | x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x | \theta)p(\theta)}{p(x)} = \frac{p(x | \theta)p(\theta)}{\int p(\theta, x)d\theta} \quad (2.1)$$

Sendo o denominador somente a constante de normalização, pode-se simplificar a expressão escrevendo que a distribuição *à posteriori* é proporcional ao produto entre a verossimilhança $p(x | \theta)$ e a *priori* $p(\theta)$, sendo a verossimilhança uma função dos parâmetros de um modelo estatístico que permite inferir sobre o seu valor a partir de um conjunto de observações (BERGER, 1985). Realizando as manipulações algébricas necessárias e mantendo somente os termos que envolvem o parâmetro de interesse θ , obtém-se:

$$p(\theta | x) \propto p(x | \theta).p(\theta) \quad (2.2)$$

2.5.1 Informações *à priori*

De acordo com Gelman *et al.* (2014), a utilização de informação *à priori* na análise bayesiana requer a especificação de uma distribuição *à priori* para a quantidade de interesse θ . Esta distribuição deve representar probabilisticamente o conhecimento que se tem sobre θ antes da realização do experimento, possuída por alguém, que se identifica como o especialista (perito, expert) do problema concreto (seja ele o pesquisador, o estatístico ou outrem) e contém elementos subjetivos, no entanto, fundamentados por fontes objetivas (dados históricos do problema ou de problemas análogos, fatos).

A partir do conhecimento que se tem sobre θ , pode-se definir uma família paramétrica de densidades. Assim, a distribuição *à priori* é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros

indexadores da família de distribuições *à priori* são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse. Existem diferentes formas de especificação da distribuição *à priori* (BERGER, 1985).

a) **Priori Não Informativa**

Segundo Stone (2013), a *priori* não informativa refere-se a especificação de distribuições *à priori* quando se espera que a informação dos dados seja dominante, reconhecendo a necessidade de alguma forma de análise que, em algum sentido, consiga captar esta noção de uma *priori* que tenha um efeito mínimo, relativamente aos dados, na inferência final. Tal análise pode ser pensada como um ponto de partida quando não se consegue fazer uma elicitación detalhada do verdadeiro conhecimento *à priori*.

A ideia de não informação *à priori* que se pode ter é pensar em todos os possíveis valores de θ como igualmente prováveis. Fazendo $p(\theta) \propto k$ para θ variando em um subconjunto que compreenda θ significa que nenhum valor particular tem preferência.

b) **Priori Informativa**

Conforme Gelman *et al.* (2014), a *priori* informativa é dada quando a distribuição assume algum grau de conhecimento acerca do parâmetro, privilegiando alguma parte do espaço paramétrico. Pode ser classificada como pouco informativa ou vaga, informativa e muito informativa. Deve-se manter um cuidado com *priori* muito informativa, para não conflitar com a verossimilhança.

Quanto a estrutura, a ***priori conjugada*** é dada quando as distribuições *à priori* e *à posteriori* pertencem a mesma classe de distribuições e assim a atualização do conhecimento que se tem de θ envolve apenas uma mudança nos hiperparâmetros. Neste caso, o aspecto sequencial do método Bayesiano pode ser explorado definindo-se apenas a regra de atualização dos hiperparâmetros, uma vez que as distribuições permanecem as mesmas, as distribuições *à priori* conjugadas podem ser informativas ou não informativas (GELMAN *et al.* 2014).

2.5.2 Distribuição *à posteriori* para a proporção

Kinas e Andrade (2010) recomendam a distribuição *Beta*, frequentemente utilizada para estimar a proporção, uma vez que essa distribuição está definida no intervalo [0,1]. Seja o parâmetro de interesse a proporção, π ($0 \leq \pi \leq 1$, **desconhecido**), de indivíduos portadores

de certa característica em uma população. Para o i -ésimo indivíduo tem-se que X_i segue uma distribuição Bernoulli(π), sendo π a probabilidade de ter a característica.

Selecionada uma amostra aleatória de tamanho n da população, supõe-se que se observe $t =$ número de elementos portadores da característica de interesse na amostra. Então, dado π , a distribuição de probabilidade de T é Binomial com parâmetros n e π , isto é,

$$P(T = t | \pi) = \binom{n}{t} \pi^t (1 - \pi)^{n-t}, t = 0, 1, \dots, n. \quad (2.3)$$

Assumindo então que, *à priori*, a distribuição de probabilidade de π seja uma *Beta* com parâmetros a_0 e b_0 conhecidos ($a_0 > 0, b_0 > 0$), isto é:

$$p(\pi) = \frac{1}{\beta(a_0, b_0)} \pi^{a_0-1} (1 - \pi)^{b_0-1}, \quad (2.4)$$

em que $\beta(a_0, b_0)$ é a função matemática $\frac{\Gamma(a_0)\Gamma(b_0)}{\Gamma(a_0+b_0)}$ tal que $\Gamma(n) = (n - 1)!$.

À partir da equação 2.2, temos que:

$$p(\pi | t) \propto (\pi^t (1 - \pi)^{n-t}) (\pi^{a_0-1} (1 - \pi)^{b_0-1}), \quad (2.5)$$

ou seja,

$$p(\pi | t) \propto \pi^{(a_0+t)-1} (1 - \pi)^{(b_0+n-t)-1}, \quad (2.6)$$

o que implica em uma distribuição *à posteriori* conjugada (FINK, (1997), COOK, (2019), BERGER, (1985)), isto é, também é uma função densidade de probabilidade *Beta* porém com parâmetros $a_0 + t$ e $b_0 + n - t$ dada por:

$$\pi | t \sim \text{Beta}(\alpha^* = a_0 + t, \beta^* = b_0 + n - t), \quad (2.7)$$

em que $t = \sum_{i=1}^n x_i$ é o total de sucessos observados na amostra. Sua média *à posteriori* é dada por:

$$E(\pi | T) = \frac{a_0 + t}{a_0 + b_0 + n}. \quad (2.8)$$

Como distribuição *a priori* não informativa para π , considera-se que π segue uma distribuição $Beta(a_0, b_0)$, com parâmetros $a_0 = b_0 = 1$, ou seja, nenhum local do espaço paramétrico é privilegiado (BERGER, 1985).

2.5.3 Intervalo de credibilidade

As definições de intervalo de credibilidade são feitas por Kinas e Andrade (2010), onde explicam que a forma mais adequada de expressar a informação que se tem sobre um parâmetro é através de sua distribuição *à posteriori*. No entanto, quando estima-se um parâmetro através de um único valor numérico, toda a informação presente na distribuição *à posteriori* é resumida através deste número. Portanto, é importante associar alguma informação sobre o quão precisa é a especificação deste número.

Definição: O Intervalo de Credibilidade (ICr) ou intervalo de confiança Bayesiano, baseado na distribuição *à posteriori* é um intervalo de credibilidade de $100(1 - \alpha)\%$, ou nível de credibilidade (ou confiança) $1 - \alpha$, para θ se

$$\int_{-\infty}^a p(\theta | x) d\theta = \int_b^{\infty} p(\theta | x) d\theta = \frac{\alpha}{2}. \quad (2.9)$$

O intervalo de credibilidade de 95% ($ICr_{95\%}$) para θ é o intervalo delimitado pelos percentis 2,5% ($\theta_{[2,5\%]}$) e 97,5% ($\theta_{[97,5\%]}$) da distribuição *à posteriori* $p(\theta | x)$ para θ .

A interpretação do intervalo de credibilidade de 95% é que existe 95% de probabilidade de que o verdadeiro parâmetro esteja contido no intervalo calculado.

2.5.4 Fator de Bayes

Hogg *et al.* (2005) explicam que uma das formas de realizar testes de Hipóteses e tomar decisão fundamentando-se na distribuição *à posteriori* é baseando-se em algum critério padronizado como o Fator de Bayes.

Sejam duas hipóteses, H_0 conhecida como hipótese nula e H_1 conhecida como hipótese alternativa e uma sequência de observações, $X = (x_1, x_2, \dots, x_n)$, com $n \geq 1$, tal que estes dados são provenientes do modelo (BERGER, 1985). A partir dos dados existentes, é obtida a

probabilidade *à posteriori* ($\pi(\theta | X)$), com o interesse de verificar as hipóteses

$$\begin{cases} H_0 : \theta \in \theta_0 \\ H_1 : \theta \in \theta_1 \end{cases}$$

em que $\theta_0 \cup \theta_1 = \Theta$ e $\theta_0 \cap \theta_1 = \emptyset$.

Logo, pode-se calcular as distribuições condicionais:

$$P(\theta \in \theta_0) \text{ e } P(\theta \in \theta_1), \quad (2.10)$$

que representam as probabilidades condicionais das hipóteses H_0 e H_1 , respectivamente, serem verdadeiras. Uma regra de decisão simples é aceitar H_0 se $P(\theta \in \theta_0) \geq P(\theta \in \theta_1)$, e caso contrário aceitar H_1 .

Segundo Wetzels e Wagenmakers (2012), o Fator de Bayes é uma medida da evidência provida pelos dados a favor de um modelo estatístico. Em favor de H_0 é dado por:

$$BF_{0,1} = \frac{O(H_0, H_1 | x)}{O(H_0, H_1)}, \quad (2.11)$$

em que:

$$O(H_0, H_1 | x) = \frac{P(H_0 | x)}{P(H_1 | x)}, \quad (2.12)$$

é denominada chance *à posteriori* em favor de H_0 ,

$$O(H_0, H_1) = \frac{P(H_0)}{P(H_1)}, \quad (2.13)$$

é denominada chance *a priori* em favor de H_0 .

Da mesma forma, temos que

$$BF_{1,0} = \frac{O(H_0, H_1)}{O(H_0, H_1 | x)}. \quad (2.14)$$

Aplicando o logaritmo na base 10 no valor do fator de Bayes, a fim de obter números com intervalos menores, obtém-se uma medida conhecida como peso de evidência, que pode ser utilizada para tomadas de decisão:

$$\log BF_{0,1}(x) = \log O(H_0, H_1 | x) - \log O(H_0, H_1), \quad (2.15)$$

ou

$$\log BF_{1,0}(x) = \text{Log}O(H_0, H_1) - \log O(H_0, H_1 | x). \quad (2.16)$$

Pela equação 2.16, observa-se que

$$\log BF_{1,0}(x) = (\log BF_{0,1}(x))^{-1} = \frac{1}{\log BF_{0,1}(x)}. \quad (2.17)$$

Jeffreys (1961) sugeriu interpretar o fator de Bayes através de uma calibragem, ou seja, sugeriu dividir os possíveis valores do cálculo em quatro intervalos. A Tabela 3 mostra a divisão destes intervalos.

Tabela 3 – Tabela de evidências de Jeffreys (1961) para o Log do fator de Bayes

Valor Assumido do Log BF	Evidência em favor de H_0
0 ~ 0,5	Fraca
0,5 ~ 1	Substancial
1 ~ 2	Forte
> 2	Decisiva

Fonte: Jeffreys, 1961.

Quando obtém-se valores negativos, tem-se que a evidência é a favor de H_1 .

2.5.5 Aplicações em estudos experimentais

Nesta seção, serão apresentados alguns estudos experimentais com análises utilizando a teoria bayesiana.

Brigenti; Rezende e Brigenti (2011) realizaram um estudo com o objetivo de estimar através da metodologia sequencial bayesiana, o parâmetro π (probabilidade de sucesso) da distribuição da proporção, aplicando esse método de estimação em dados reais de infestação do psilídeo no alecrim do campo (*Baccharis dracunculifolia*). A estimação da proporção de plantas atingidas por psilídeos, em um contexto estatístico torna-se um objeto de pesquisa, no qual, a utilização de métodos inferenciais Bayesianos agrega informações precisas e plausíveis

de serem atualizadas, que auxiliarão na tomada de decisão. Foram geradas sequencialmente amostras de ensaios de Bernoulli, seguindo os valores paramétricos fixados em $\pi = 0,1$; $0,3$; $0,7$; e $0,9$, mantendo-se um truncamento no procedimento com $n=1000$. Após o processo de simulação da população, foram calculados o risco imediato e o risco esperado de acordo com cada valor obtido sequencialmente, adotando-se como hiperparâmetros da *priori* Beta os valores 1, 3 e 5, tanto para o parâmetro α , quanto para o parâmetro β . Quando $\alpha = \beta = 1$, a densidade de Beta se reduz à uniforme no intervalo $(0,1)$, sendo uma distribuição não informativa e quando $\alpha = 3$ ou $\beta = 3$ e $\alpha = 5$ e $\beta = 5$, tem-se respectivamente, as *priori* de menor e maior grau de informatividade. Nesse tipo de método, deve-se parar a amostragem na observação em que o risco esperado a *posteriori* for maior que o risco imediato. Verificou-se que na área de alecrim do campo avaliada, o processo foi interrompido após avaliação de 10 plantas, obtendo-se uma proporção de infestação de psilídeos igual a 28,57%. Concluiu-se que a técnica de estimação sequencial bayesiana foi mais adequada para estimação da proporção quando o valor populacional está mais afastado dos valores extremos.

A isoterma de Langmur é um modelo estatístico de adsorção (RODRIGUES; NEVES e ANGELICA, 2014), no qual Carvalho; Beijo e Muniz (2017) utilizaram a teoria bayesiana no ajuste do modelo em simulação de dados de adsorção de metais traço, considerando as distribuições à *priori* Informativa e não informativa. Os autores explicaram que estudos de adsorção de metais traços em laboratório enfrentam uma limitação na quantidade de amostra disponível devido à geração de materiais contaminados, os quais ocasionam uma elevação de custos, principalmente devido à necessidade de tratamento e descarte específicos dos resíduos tóxicos que são gerados. Portanto, os autores ajustaram a isoterma de Langmuir sob enfoque Bayesiano, avaliando, via simulação, a acurácia e a precisão das estimativas obtidas sob diferentes tamanhos amostrais. Os resultados indicaram que o método Bayesiano é eficiente e as estimativas obtidas com uso das distribuições *a priori* informativas dos parâmetros apresentaram maiores precisão e acurácia mesmo em tamanhos amostrais inferiores, comparativamente aquelas obtidas com a distribuição *a priori* não informativa. Os autores evidenciaram que o uso da inferência bayesiana, com distribuição *a priori* informativa mostra-se uma alternativa viável e de grande potencial para contornar a limitação prática acerca do tamanho amostral, de modo que seja possível obterem-se estimativas precisas utilizando-se menor quantidade de amostra.

Com o objetivo de selecionar genótipos de feijão-caupi (*Vigna unguiculata*) que reúnam

alta adaptabilidade e estabilidade fenotípicas, Teodoro *et al.* (2015) utilizou a teoria bayesiana para analisar dados de quatro experimentos, conduzidos em delineamento de blocos ao acaso, em que a produtividade de grãos de 20 genótipos de feijão-caupi semiprostrado foi avaliada no Estado do Mato Grosso do Sul. As distribuições *à priori* pouco informativas, foram representadas utilizando-se distribuições de probabilidade com grande variância (Distribuição Normal) e para as distribuições *à priori* informativas, adotou-se o conceito de metanálise, com uso de informações de trabalhos anteriores. A comparação entre as distribuições *à priori* foi realizada por meio do fator de Bayes. Foi identificado que a abordagem bayesiana proporciona maior acurácia na seleção de genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. Com base na *priori* informativa, foram selecionados os genótipos classificados como de adaptabilidade específica a ambientes favoráveis e a ambientes desfavoráveis.

A metodologia bayesiana foi utilizada por Martins Filho *et al.* (2008) para ajustar o modelo não-linear logístico para dados de crescimento de duas cultivares de feijoeiro, “Neguinho” (*Phaseolus vulgaris*) e “Carioca” (*Phaseolus vulgaris* Pinto Group). Os autores explicaram que, quando se dispõe de poucas observações longitudinais, os métodos iterativos, muitas vezes, produzem estimativas irrealistas para os parâmetros. Isso pode levar à confecção de curvas de crescimento atípicas. Além disso, quando se trata de comparações de curvas provenientes de tratamentos diferentes, na maioria das vezes, a distribuição dos estimadores dos parâmetros de modelos não-lineares não seguem distribuições gaussianas. Desse modo torna-se complexo o processo de formulação de testes estatísticos, pois não serão atendidas pressuposições relacionadas à utilização da teoria assintótica. Por outro lado, a inferência bayesiana tem se mostrado eficaz em reduzir o número de estimativas viesadas, mesmo utilizando poucas informações e a estimação por intervalo apresenta-se mais precisa em relação à obtida pela metodologia clássica. Os autores concluíram que a metodologia bayesiana permitiu comparar curvas de crescimento de cultivares de feijoeiro de forma eficiente, sem utilizar aproximações assintóticas, demonstrando que a cultivar “Carioca” apresenta um incremento em altura maior, ou seja, é mais precoce que a cultivar “Neguinho”.

O método de Eberhart e Russell é um método baseado em modelos de regressão linear simples. Nascimento *et al.* (2011) objetivaram com seu trabalho, propor uma abordagem bayesiana do método, para avaliar a adaptabilidade e a estabilidade fenotípica de genótipos de alfafa (*Medicago sativa*), bem como avaliar a eficiência da utilização de distribuições *à priori*

informativas e pouco informativas. As informações *à priori* pouco informativas foram representadas por distribuições de probabilidade com grande variância (Distribuição Normal) e as informativas, utilizou-se informações de trabalhos anteriores. Para comparar as distribuições *à priori*, utilizou-se o Fator de Bayes. Os autores concluíram que a abordagem bayesiana aplicada à estudos de adaptabilidade e estabilidade, por meio de modelos de regressão linear simples é eficiente para a classificação dos genótipos e com base nos resultados de seleção de modelos via fator de Bayes, o uso de distribuições *à priori* informativas é superior à qualidade de ajuste, em comparação as distribuições *à priori* não informativas.

Winter *et al.* (2006) utilizou a teoria bayesiana para obter as estimativas de correlações genéticas e fenóticas para características de peso em diferentes idades de codornas de corte. foram pesadas 1022 aves individualmente, utilizadas para a avaliação de desempenho produtivo aos 7 dias, 14 dias, 28 dias e 42 dias. Foram calculados os coeficientes de variação, os intervalos de credibilidade e as regiões de alta densidade a 90% para as correlações genéticas entre as características analisadas. Os intervalos de credibilidade das estimativas de correlações genéticas entre os pesos aos 28 e 42 dias e aos 14 e 28 dias foram os que apresentaram as menores amplitudes de variação, o que pode ser confirmado pelos menores coeficientes de variação (7,40 e 14,09%), o que conclui-se que as estimativas destes parâmetros foram mais precisas.

Existem diversos contextos e áreas que se pode aplicar a metodologia bayesiana. A utilização de informação prévia a respeito do parâmetro de interesse pode fazer com que as estimativas *à posteriori* sejam mais coerentes com a realidade da variável de interesse (ROSSI, 2011).

No entanto, não foram encontrados estudos envolvendo proporção de sementes de soja e teoria bayesiana.

3 MATERIAL E MÉTODOS

As sementes de soja para análise foram obtidas em uma revenda de sementes e insumos agrícolas, localizada na cidade de Alfenas-MG. A empresa possui um controle de qualidade interno para testar e verificar as tecnologias e produtos oferecidos pelos fornecedores. Uma parte das análises consiste em verificar se as sementes a serem revendidas aos produtores apresentam a qualidade germinativa necessária, para uma boa produção. Foram utilizados resultados das análises do ano anterior à montagem do experimento, para identificar se existe diferença significativa dos lotes dentro dos cultivares, e dos cultivares dentro das empresas, que seguem a seguir.

3.1 INFORMAÇÕES PARA MONTAGEM DO EXPERIMENTO

As sementes vêm separadas nas notas fiscais por lote. Anexo a nota fiscal, vem um relatório denominado “Termo de Conformidade (TC)”, trata-se de um documento obrigatório, conforme definições do ministério da agricultura, ou seja, todo lote recebido pela revenda possui esse documento. Neste termo consta a porcentagem de germinação do lote, garantida pelo fornecedor. Existem fatores que podem alterar essa qualidade germinativa, até sua chegada ao estoque da revenda, que são, umidade, temperatura, quebra, tempo, condições de armazenamento, entre outros. Foram constatados alguns casos em que a porcentagem de germinação no plantio foi bem menor que a garantida pelo fornecedor, gerando insatisfação do cliente. Portanto, toda semente comprada por essa empresa, é verificada, buscando-se obter nos testes, no mínimo, em média 80% de germinação.

Com o objetivo de verificar se existe diferença significativa dos lotes dentro dos cultivares e dos cultivares dentro das empresas para definições de montagem do experimento, foram selecionados dois principais fornecedores de sementes de soja no ano de 2017, aos quais denominaremos Fornecedor 1 e Fornecedor 2. Para cada fornecedor foram selecionados ao acaso três cultivares, totalizando seis, aos quais chamaremos de Cultivar A, Cultivar B, ..., Cultivar F. Foram considerados 7 lotes do Cultivar A, 15 lotes do Cultivar B, 11 lotes do Cultivar C, 10 lotes do Cultivar D, 2 lotes do Cultivar E e 11 lotes do Cultivar F, que eram lotes disponíveis.

Mediante os resultados das análises feitas no ano de 2017, realizou-se o teste de Kruskal-Wallis (KW), que é uma análise de variância não paramétrica que pode ser usada para comparar várias amostras independentes e não exige as suposições de normalidade da variável, nem homogeneidade de variâncias entre os tratamentos (CAMPOS, 1983), com o objetivo de verificar se existem diferenças significativas entre as médias do atributo vigor, dos seguintes quesitos: cultivares dentro de cada fornecedor e lotes dentro de cada cultivar. Além disso foram calculados o 1º quantil (Q_1), a mediana, a média, o 3º quantil (Q_3) e o coeficiente de variação (CV).

As hipóteses são:

$$\begin{cases} H_0 : \pi_1 = \pi_2 = \dots = \pi_k \\ H_1 : \pi_1, \pi_2, \dots, \pi_k, \text{ pelo menos um } \pi_i \text{ é diferente, } i = 1, \dots, k \end{cases}$$

Os resultados referentes às médias de cultivares dentro de cada fornecedor, o 1º quantil, a mediana, a média, o 3º quantil, o coeficiente de variação e o resultado do Teste de Kruskal-Wallis dos cultivares dentro dos Fornecedores 1 e 2 encontram-se na 7 do Apêndice A, o qual descreve que não existe diferença significativa entre as médias do atributo vigor, ou seja, as médias do atributo vigor dos cultivares do Fornecedor 1 e dos cultivares do Fornecedor 2 são estatisticamente iguais.

Os resultados referentes às médias dos lotes dentro de cada cultivar, o 1º quantil, a mediana, a média, o 3º quantil, o coeficiente de variação e o valor do Teste de Kruskal-Wallis encontram-se na Tabela 8 do Apêndice B para o Fornecedor 1 e na Tabela 9 do Apêndice B para o Fornecedor 2. Em ambas tabelas (8 e 9), os resultados indicam que existe diferença significativa em pelo menos uma das médias do atributo vigor dos lotes, dentro dos cultivares, ou seja, pelo menos uma média do atributo vigor dos lotes em cada cultivar é estatisticamente diferente, tanto para o Fornecedor 1, quanto para o Fornecedor 2.

Logo, conclui-se que não existe diferença significativa dos cultivares dentro dos fornecedores e que existe diferença significativa dos lotes dentro de cada cultivar. Essas informações foram úteis para definições de montagem do experimento, ou seja, identifica-se que os lotes não são homogêneos dentro dos cultivares e cada lote deverá ser analisado individualmente.

3.2 DADOS UTILIZADOS NA ANÁLISE

Para o experimento da soja, foram selecionados os três principais fornecedores de sementes para a revenda no ano de 2018, aos quais chamaremos de Fornecedor 1, Fornecedor 2 e Fornecedor 3. Para cada fornecedor foram selecionados ao acaso dois cultivares, totalizando seis, aos quais chamaremos de Cultivar A, Cultivar B, ..., Cultivar F. Para cada cultivar, foram selecionados, aleatoriamente, dois lotes, totalizando doze, aos quais chamaremos de Lote 01, Lote 02, ..., Lote 12. Definiremos as nomenclaturas, conforme a Tabela 4.

Tabela 4 – Nomenclaturas dos lotes de sementes de soja analisados.

Fornecedor	Cultivar	Lote	Nomenclatura
1	A	L01	1AL01
1	A	L02	1AL02
1	B	L03	1BL03
1	B	L04	1BL04
2	C	L05	2CL05
2	C	L06	2CL06
2	D	L07	2DL07
2	D	L08	2DL08
3	E	L09	3EL09
3	E	L10	3EL10
3	F	L11	3FL11
3	F	L12	3FL12

Fonte: do autor.

Todo o processo de amostragem e de realização do experimento foi feito conforme recomendações das RAS (BRASIL, 2009). A armazenagem do produto é feita em depósito, por empilhamento dos sacos e separadas por lote. Para a retirada da amostra, utilizou-se um calador e colheu-se amostras de 30 sacos por lote, totalizando 360 amostras, sendo coletados cerca de 100 gramas de sementes por saco. As sementes foram colocadas em sacos de papel, destinados para esse fim, sendo um saco de papel para cada saco do lote e foram mantidas em câmara fria até as datas selecionadas para o experimento, que foram realizados entre 15 e 60

dias após a retirada das amostras, nos períodos entre setembro e novembro de 2018 (Figura 3).



Figura 3 – Retirada de amostra de semente acondicionada.

Fonte: do autor.

O experimento foi conduzido no Laboratório de análises da revenda e foi realizado em Delineamento em Blocos Casualizados. Colocava-se 6 amostras de cada um dos 12 lote por dia, totalizando 72 amostras. Foram necessários cinco dias não consecutivos para colocar o experimento e cinco dias não consecutivos para tirá-lo, sendo os dias considerados os blocos, o que totaliza as 360 amostras coletadas. da amostra de 100 gramas coletada de cada saco, foram selecionadas à esmo 50 sementes, as quais foram colocadas em papéis de germinação, umedecidos com água destilada, com a ajuda de um contador de sementes (Figura 4).



Figura 4 – Montagem do experimento, papel de germinação e placa contadora de sementes.

Fonte: do autor.

Em seguida fez-se um rolo com os papéis de germinação, identificando-os de acordo com o lote. Posteriormente, os rolinhos foram colocados em um saco plástico, para manter a

umidade, sendo que cada saco plástico recebeu 6 rolos, totalizando 12 sacos plásticos, conforme pode ser observado na Figura 5.

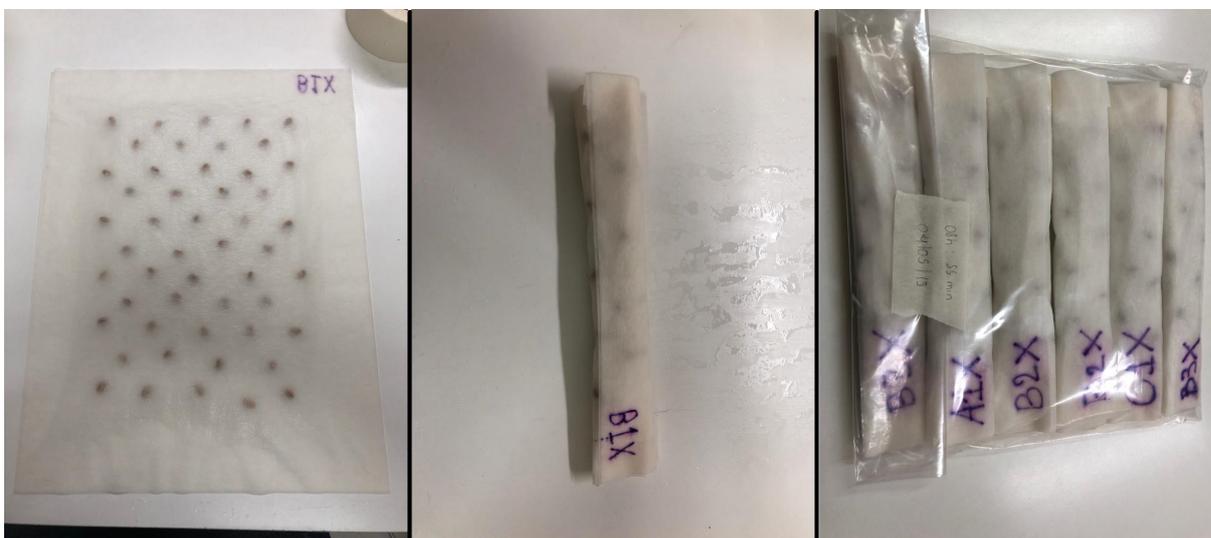


Figura 5 – Montagem do experimento, rolinhos de papel a serem colocados para germinação.

Fonte: do autor.

A partir daí as sementes foram colocados na câmara germinadora BOD, conforme apresentado na Figura 6.



Figura 6 – Câmara Germinadora BOD.

Fonte: do autor.

A avaliação foi realizada após 5 dias, ou aproximadamente 120 horas. Observou-se cada semente e anotou-se aquelas que apresentaram características vigorosas; germinadas; normais; anormais; mortas e dormentes, conforme Figura 7. As sementes consideradas na análise foram as que apresentaram características vigorosas, que é o interesse da revenda. Todo o processo foi acompanhado pelo profissional responsável pelas análises de sementes do laboratório da revenda.



Figura 7 – Sementes após 5 dias na germinadora BOD, à esquerda - sementes vigorosas e à direita - sementes não vigorosas.

Fonte: do autor.

3.3 Distribuição *à posteriori* para a proporção

A proporção de germinação (π) está em um intervalo limitado $(0, 1)$ e a distribuição Beta foi utilizada como distribuição *à priori* para π . Logo, selecionada uma amostra aleatória de tamanho n e observando o número de elementos portadores da característica de interesse na amostra (t) e assumindo a distribuição *à priori* uma Beta com hiperparâmetros a_0 e b_0 conhecidos ($a_0 > 0, b_0 > 0$), tem-se uma distribuição *à posteriori* conjugada, com distribuição Beta de parâmetros $(a_0 + t)$ e $(b_0 + n - t)$, dada por:

$$p(\pi | t) \propto \pi^{(a_0+t)-1} (1 - \pi)^{(b_0+n-t)-1}. \quad (3.1)$$

3.3.1 Distribuições *à priori*

Como distribuição *à priori* não informativa para π , assumiu-se que $\pi \sim \text{Beta}(\alpha_0, \beta_0)$, com $\alpha_0 = \beta_0 = 1$, conforme descrito na seção 2.5.2.

Como distribuição *à priori* informativa, utilizou-se as informações dos Termos de Conformidade (TC) descritos na seção 2.2, recebidos juntos às notas fiscais dos lotes entregues no ano de 2018. Foram utilizados os termos de 56 lotes do Fornecedor 1, 93 lotes do Fornecedor 2 e 24 lotes do Fornecedor 3, que eram os termos disponíveis para análise.

Como outra forma de *priori* informativa, utilizou-se os resultados obtidos no laboratório de análises de sementes da revenda do ano de 2017. A empresa possuía dados de análise de duas empresas sementeiras, das consideradas no estudo, o Fornecedor 2 e o Fornecedor 3. Foram utilizados os resultados da análise de 91 lotes do Fornecedor 2 e 58 lotes do Fornecedor 3, que eram dados disponíveis.

Para gerar os hiperparâmetros das *prioris* informativas, foram calculados o 1º e o 3º quartil dos dados, para cada fornecedor. Em seguida, fixou-se o valor da média e simulou-se o valor de α_0 até que se encontrassem valores de β_0 tais que o 1º e o 3º quartil da distribuição *à priori*, se aproximassem do 1º e 3º quartil dos dados

$$\bar{\pi} = \frac{\alpha_0}{\alpha_0 + \beta_0} \Rightarrow \beta_0 = \frac{\alpha_0}{\bar{\pi}} - \alpha_0.$$

O tamanho amostral (n), os valores dos quartis dos dados e da distribuição *à priori*, dos hiperparâmetros, da média dos dados e da média *à priori*, se encontram relacionados na Tabela 5:

Tabela 5 – Valores do tamanho amostral, dos quartis dos dados e da distribuição *à priori*, dos hiperparâmetros, da média dos dados e da média *à priori* para os Fornecedores 1, 2 e 3.

Dados TC					Informações <i>à priori</i> TC				
Fornecedor	n	Q_1	Q_3	Média	α_0	β_0	Q_1	Q_3	Média
1	56	0,89	0,93	0,9114	98,0	9,57	0,89	0,93	0,9110
2	93	0,93	0,97	0,9453	72,0	4,19	0,93	0,96	0,9450
3	24	0,88	0,90	0,8950	270,0	33,37	0,88	0,90	0,8900

Dados Laboratório					Informações <i>à priori</i> Laboratório				
Fornecedor	n	Q_1	Q_3	Média	α_0	β_0	Q_1	Q_3	Média
1	–	–	–	–	–	–	–	–	–
2	91	0,78	0,87	0,8155	33,5	7,35	0,78	0,86	0,8200
3	58	0,83	0,89	0,8630	55,0	8,95	0,83	0,89	0,8600

Fonte: do autor.

3.3.2 Avaliação da precisão e da tomada de decisão

Para avaliar o efeito da redução da amostra, dos 30 sacos analisados, foram selecionadas de forma aleatória, 5000 subamostras para cada tamanho amostral ($n_s = 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28$) e o processo foi repetido três vezes. Em seguida, para cada caso foram construídos Intervalos de Credibilidade de 95% e calculado o Fator de Bayes a favor de H_0 :

$$\begin{cases} H_0 : \pi \geq 80\% \\ H_1 : \pi < 80\% \end{cases}$$

Se $\pi \geq 80\%$, então o lote é aceito e se $\pi < 80\%$ então o lote é recusado. Para o teste de hipóteses foi adotada escala de evidência Jeffreys, sendo considerada forte ou decisiva em favor de H_0 quando $\log BF_{01} \geq 1$. Quando obtêm-se $\log BF_{01} < 1$, considera-se que não existem evidências à favor de H_0 , conforme seção 2.5.4.

Foram analisadas as amplitudes das estimativas intervalares de 95% dadas por

$\pi_{97,5\%} - \pi_{2,5\%}$ e definiu-se 7% o máximo permitido conforme descrito na seção 2.2.3. Foram confeccionados gráficos exibindo os resultados das análises.

Portanto, o critério de decisão quanto ao tamanho amostral ótimo será baseado nos critérios acima, o limite inferior do intervalo de credibilidade deve estar acima de 80% para os lotes aceitos, devem existir evidências fortes ou decisivas à favor de H_0 , quando $\pi \geq 80\%$ e a amplitude do intervalo de credibilidade deve ser menor que 7%.

Todas as análises estatísticas e gráficos, foram realizadas e construídos respectivamente, utilizando-se o software R (R CORE TEAM, 2018).

4 RESULTADOS E DISCUSSÃO

Na Figura 8, observam-se os resultados das análises do Fornecedor 1, Cultivar A, lotes 1AL01, 1AL02. Todos os critérios pré estabelecidos foram satisfeitos com tamanho amostral 10 sacos, tanto com *priori* não informativa (Figuras 8A, 8B e 8C), quanto com *priori* informativa (Figuras 8D, 8E e 8F) utilizando-se TC, para o lote 1AL01. Para o lote 1AL02, os critérios pré estabelecidos foram satisfeitos com tamanho amostral 8 sacos com *priori* não informativa (Figuras 8G, 8H e 8I) e com tamanho amostral 6 sacos com *priori* informativa (Figuras 8J, 8K e 8L) utilizando-se os dados do TC.

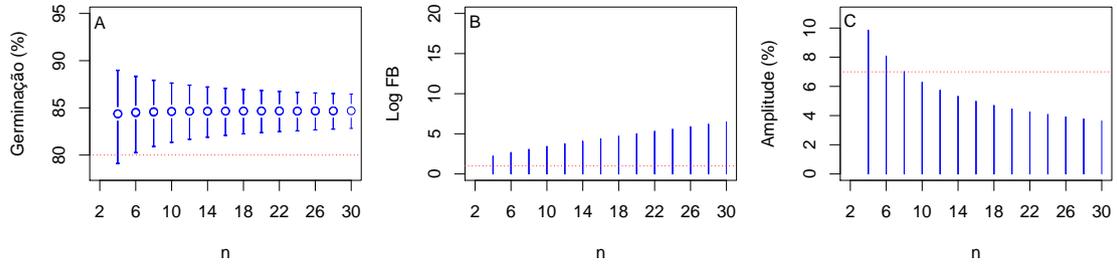
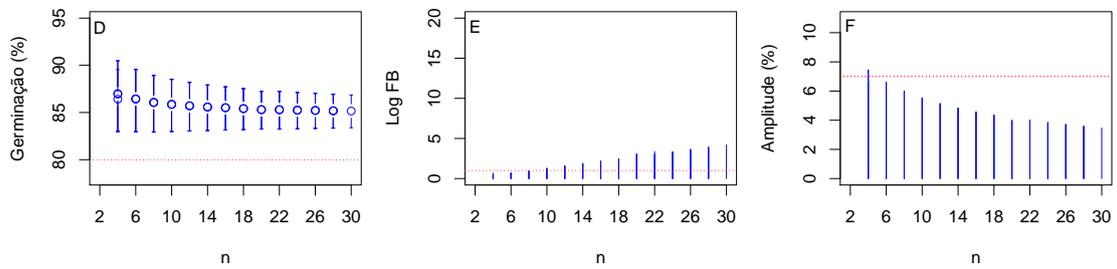
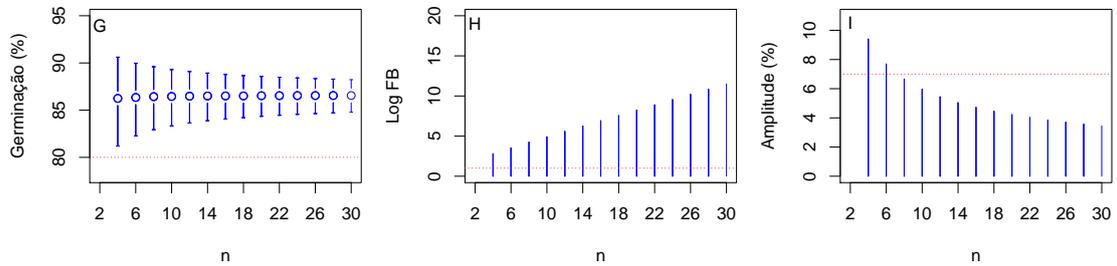
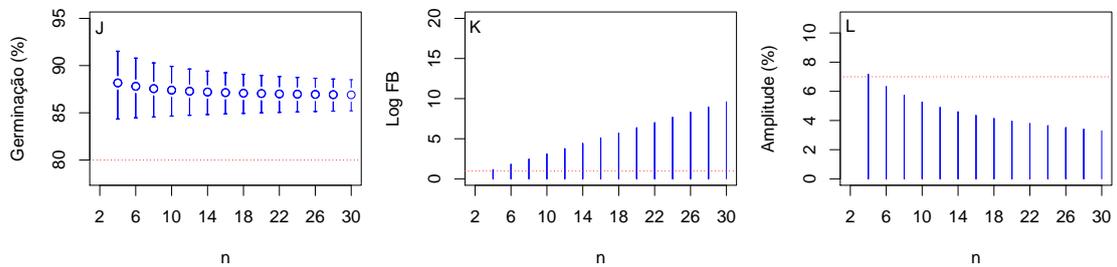
(a) *Priori* não informativa Lote 1AL01(b) *Priori* dados do TC Lote 1AL01(c) *Priori* não informativa Lote 1AL02(d) *Priori* dados do TC Lote 1AL02

Figura 8 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G e J), Log do Fator de Bayes (B, E, H e K) e Amplitude (C, F, I e L), para os diferentes tamanhos amostrais do Fornecedor 1, Cultivar A, lotes 01 e 02.

Fonte: do autor.

Na Figura 9, observam-se os resultados das análises do Fornecedor 1, Cultivar B, lotes 1BL03 e 1BL04. Os critérios de decisão e precisão foram atendidos com tamanho amostral 10 sacos para o lote 1BL03 (Figuras 9A, 9B e 9C) com *priori* não informativa e 14 sacos (Figuras 9D, 9E e 9F) com *priori* informativa utilizando os dados do TC. Nas Figuras 9G, 9H e 9I, pode-se observar que com tamanho amostral 8 sacos atende-se todos os critérios pré estabelecidos com *priori* não informativa e nas Figuras 9J, 9K e 9L, pode-se observar que acontece o mesmo com tamanho amostral 6 sacos para o lote 1BL04.

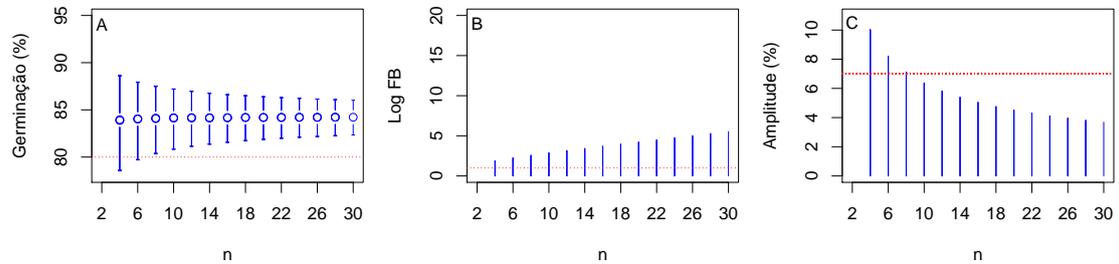
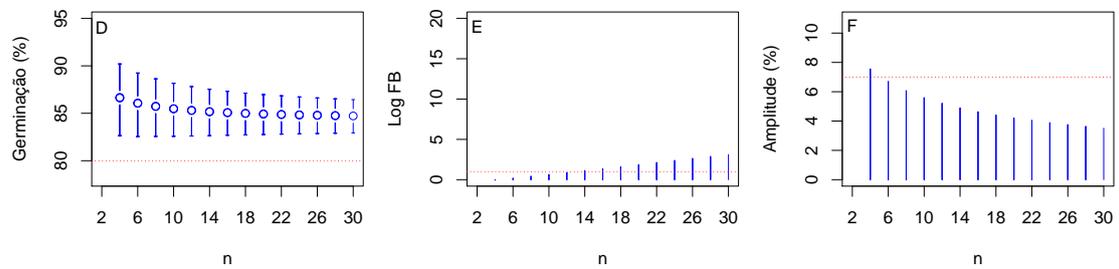
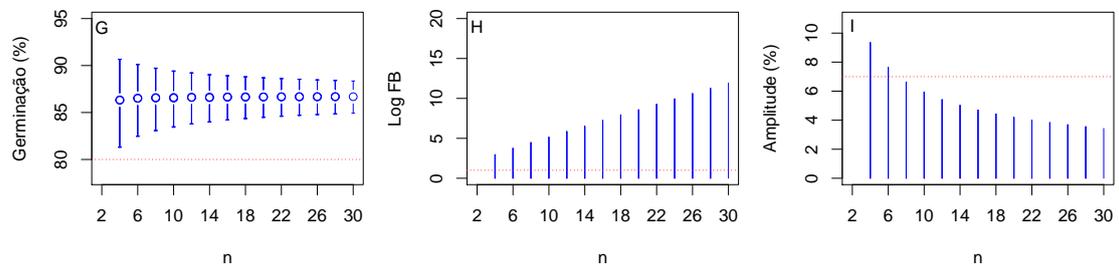
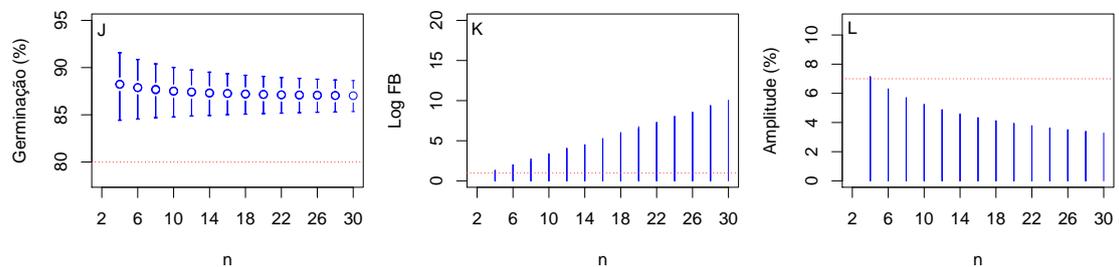
(a) *Priori* não informativa Lote 1BL03(b) *Priori* dados do TC Lote 1BL03(c) *Priori* não informativa Lote 1BL04(d) *Priori* dados do TC Lote 1BL04

Figura 9 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G e J), Log do Fator de Bayes (B, E, H e K) e Amplitude (C, F, I e L), para os diferentes tamanhos amostrais do Fornecedor 1, Cultivar B, lotes 03 e 04.

Fonte: do autor.

Os resultados do Fornecedor 2, Cultivar C, lotes 2CL05, 2CL06 encontram-se na Figura 10. Para o lote 2CL05, para amostras de tamanho iguais ou superiores a 8 sacos, os critérios pré estabelecidos são atendidos com *priori* não informativa (Figuras 10A, 10B e 10C), tamanho amostral iguais ou superiores a 4 sacos com *priori* informativa utilizando-se as informações do TC (Figuras 10D, 10E e 10F) e tamanho amostral iguais ou superiores a 6 sacos utilizando-se as informações do laboratório (Figuras 10G, 10H e 10I). Para o lote 2CL06, os critérios pré estabelecidos são atendidos com tamanho amostral iguais ou superiores a 6 sacos com *priori* não informativa (Figuras 10J, 10K e 10L) e *priori* informativa (Figuras 10P, 10Q e 10R) utilizando os dados do laboratório e com tamanho amostral iguais ou superiores a 4 sacos utilizando como *priori* informativa, os dados do TC (Figuras 10M, 10N e 10O).

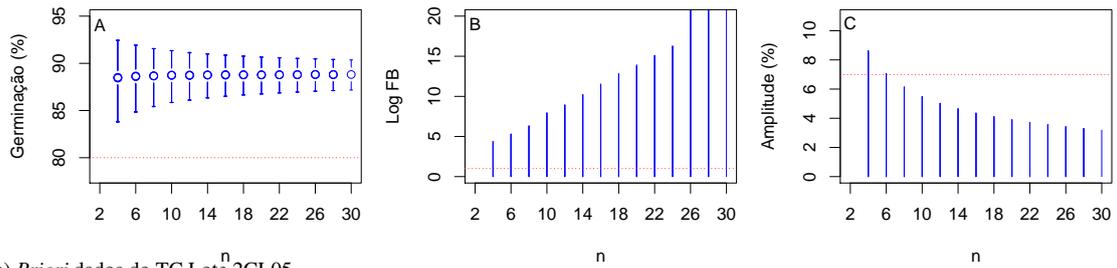
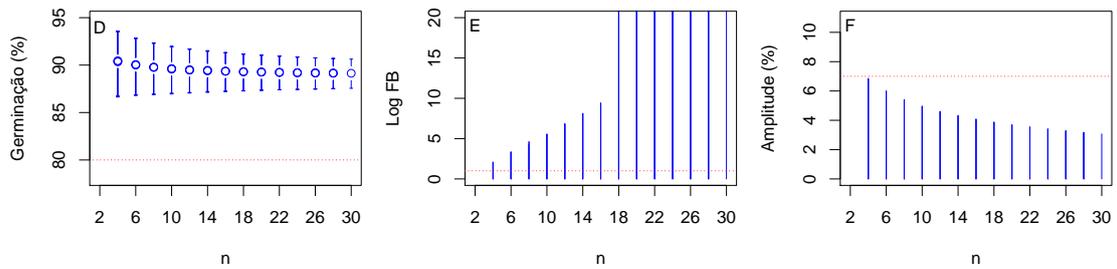
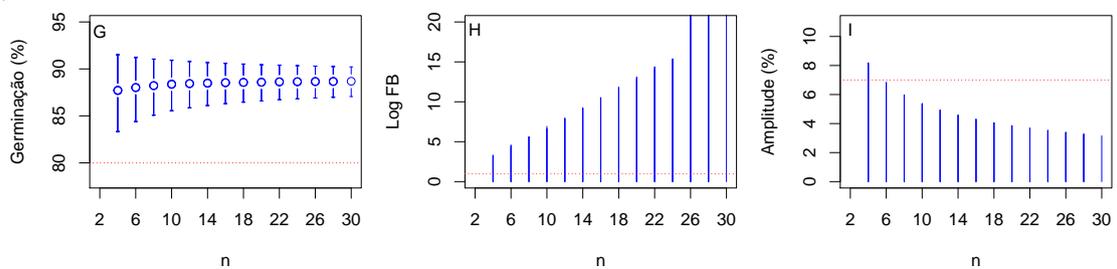
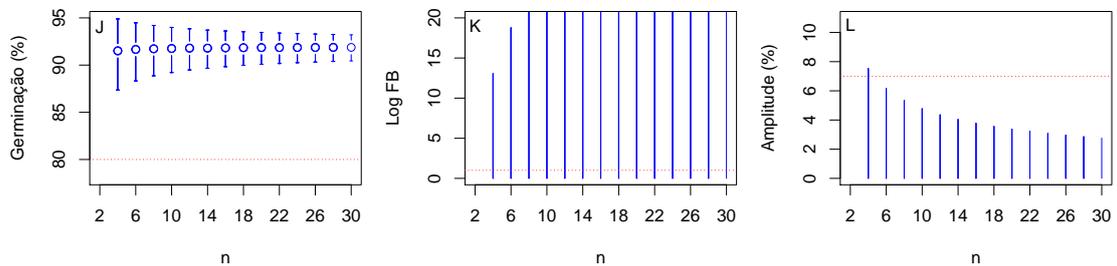
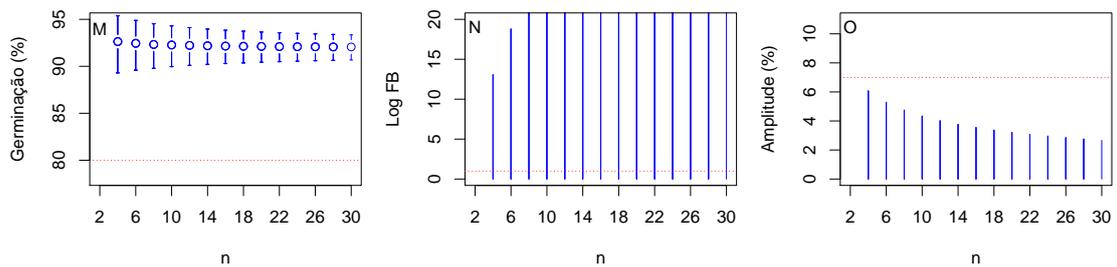
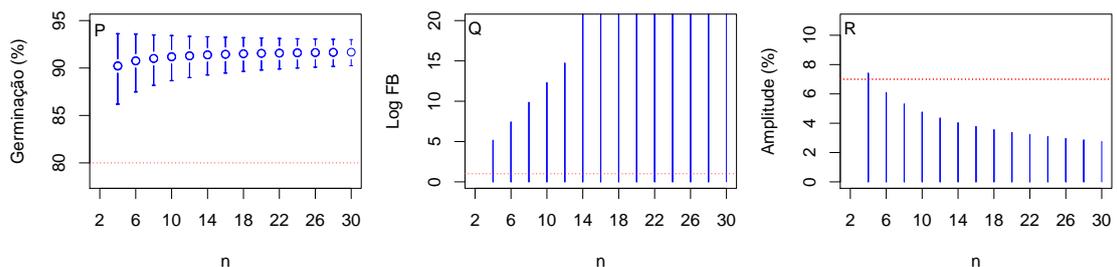
(a) *Priori* não informativa Lote 2CL05(b) *Priori* dados do TC Lote 2CL05(c) *Priori* dados do laboratório Lote 2CL05(d) *Priori* não informativa Lote 2CL06(e) *Priori* dados do TC Lote 2CL06(f) *Priori* dados do laboratório Lote 2CL06

Figura 10 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M, e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 2, Cultivar C, lotes 05 e 06.

Fonte: do autor.

Observa-se na Figura 11, os resultados das análises do Fornecedor 2, Cultivar D, lotes 2DL07 e 2DL08. No caso do lote 2DL07 os critérios pré estabelecidos foram atendidos com tamanho amostral 6 sacos com *priori* não informativa (Figuras 11A, 11B e 11C) e *priori* informativa utilizando os dados do laboratório (Figuras 11G, 11H e 11I) e com tamanho amostral 4 sacos utilizando como *priori* informativa, os dados do TC (Figura 11D, 11E e 11F). No caso do lote 2DL08 manteve-se os mesmos tamanhos amostrais que no caso do lote 2DL07 com *priori* não informativa (6 sacos, Figuras 14J, 14K e 14L), com *priori* informativa utilizando os dados do TC (4 sacos, Figuras 11M, 11N e 11O) e com *priori* informativa utilizando os dados do laboratório (6 sacos, Figuras 11M, 11N e 11O).

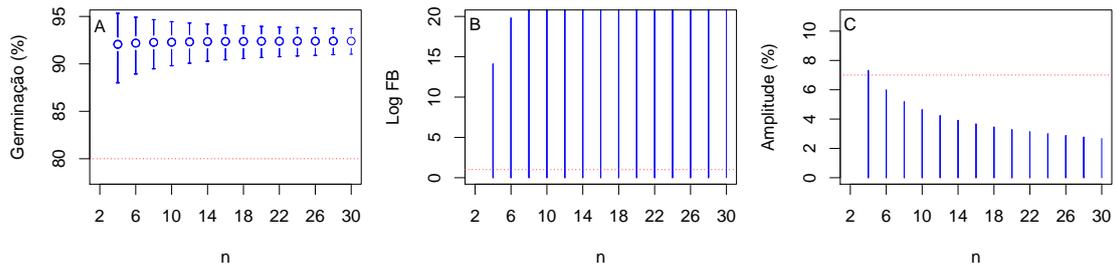
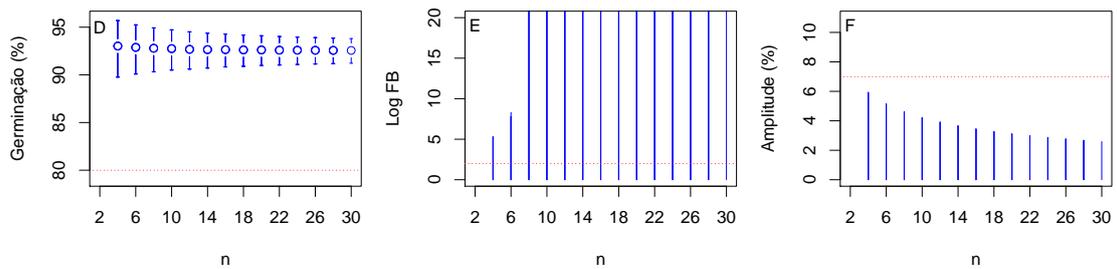
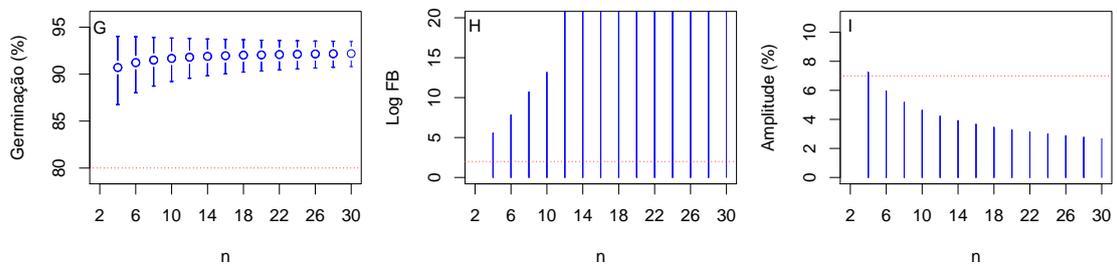
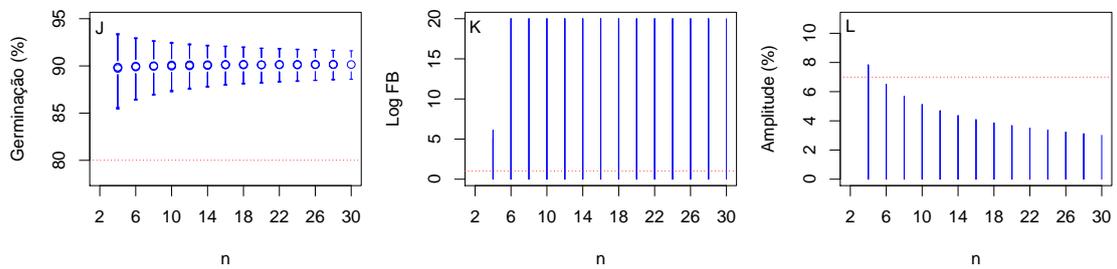
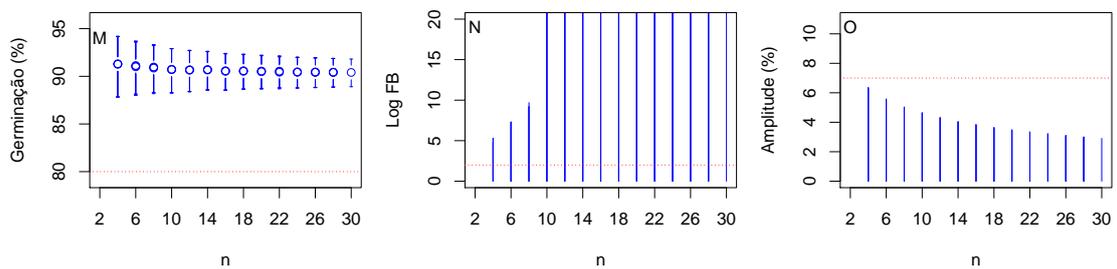
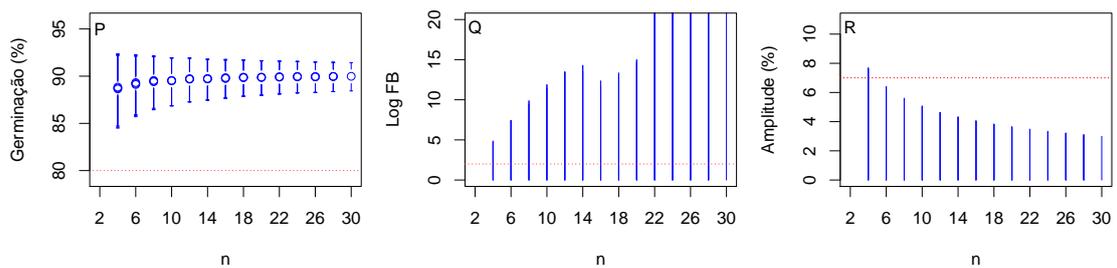
(a) *Priori* não informativa Lote 2DL07(b) *Priori* dados do TC Lote 2DL07(c) *Priori* dados do laboratório Lote 2DL07(d) *Priori* não informativa Lote 2DL08(e) *Priori* dados do TC Lote 2DL08(f) *Priori* dados do laboratório Lote 2DL08

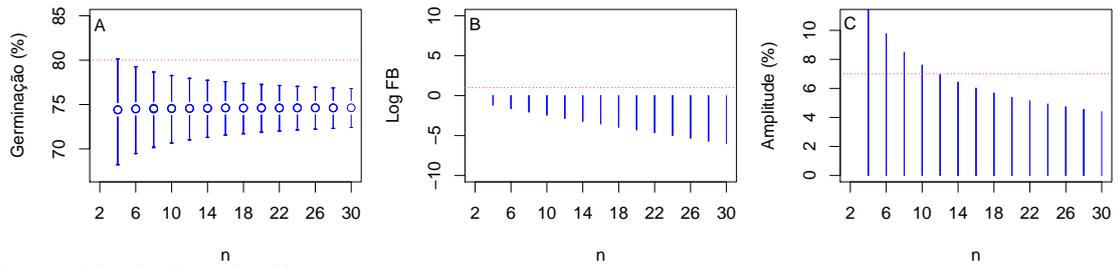
Figura 11 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 2, Cultivar D, lotes 07 e 08.

Fonte: do autor.

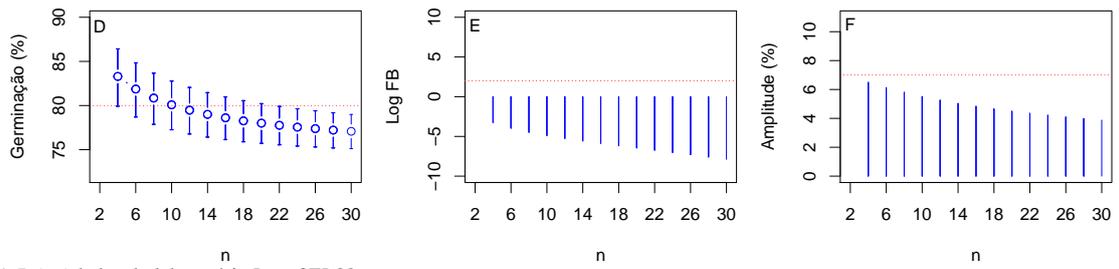
Na Figura 12, apresentam-se os resultados das análises do Fornecedor 3, Cultivar E, lotes 3EL09, 3EL10. O lote 3EL09 germinou abaixo de 80%, ou seja, trata-se de um lote que não seria aceito pela revenda. Para amostras de tamanho iguais ou superiores a 12 sacos, todos os critérios pré estabelecidos são atendidos com *priori* não informativa (Figura 12A, 12B e 12C). No caso da *priori* informativa com os dados do TC, para nenhum tamanho amostral analisado o lote seria aceito, logo, com tamanhos amostrais maiores ou iguais a 4 sacos atende-se todos os critérios pré estabelecidos (Figuras 12D, 12E e 12F). Utilizando a *priori* informativa com os dados do laboratório, com tamanhos amostrais maiores ou iguais a 12 sacos, os critérios pré estabelecidos são atendidos (Figuras 12G, 12H e 12I).

Para o lote 3EL10, todos os critérios estabelecidos foram atendidos para amostras de tamanho maior ou igual a 10 sacos com *priori* não informativa (Figuras 12J, 12K e 12L), e com *priori* informativa usando dados do TC (Figuras 12P, 12Q e 12R). Com 8 sacos, os critérios estabelecidos são atendidos quando se utiliza *priori* informativa com dados do laboratório (Figuras 12M, 12N e 12O).

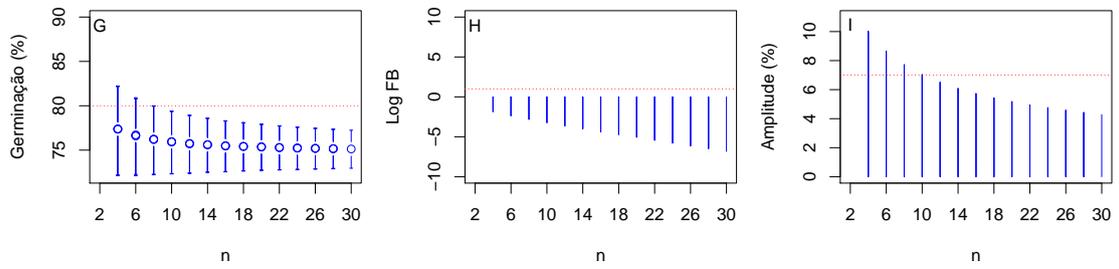
(a) *Priori* não informativa Lote 3EL09



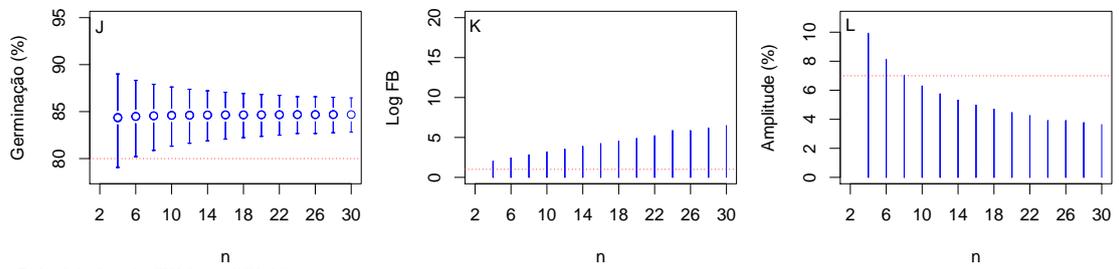
(b) *Priori* dados do TC Lote 3EL09



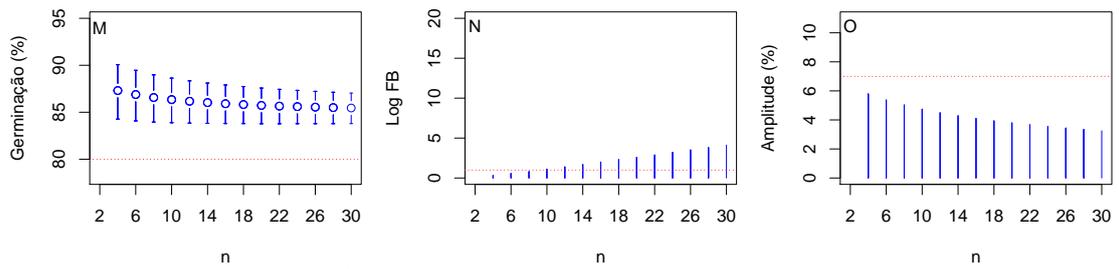
(c) *Priori* dados do laboratório Lote 3EL09



(d) *Priori* não informativa Lote 3EL10



(e) *Priori* dados do TC Lote 3EL10



(f) *Priori* dados do laboratório Lote 3EL10

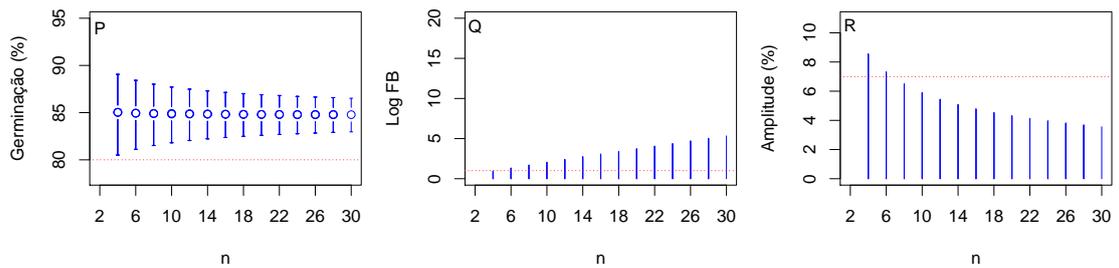


Figura 12 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I e L, O e R), para os diferentes tamanhos amostrais do Fornecedor 3, Cultivar E, lotes 09 e 10.

Fonte: do autor.

As análises do Fornecedor 3, Cultivar F, Lotes 3FL11 e 3FL12, apresentam-se na Figura 13. Para o lote 3FL11, com *priori* não informativa (Figura 13A, 13B e 13C), todos os critérios estabelecidos são atendidos com tamanho amostral 8 sacos. Em ambas as *priori* informativas, os critérios estabelecidos são atendidos com tamanho amostral 4 sacos (Figuras 13D, 13E, 13F, 13G, 13H e 13I).

Para o Lote 3FL12, com *priori* não informativa todos os critérios estabelecidos são atendidos com tamanho amostral 8 sacos (Figura 13J, 13K e 13L). Para ambas as *priori* informativas, utilizando tamanho amostral maiores ou iguais a 6 sacos, os critérios pré estabelecidos são atendidos (Figuras 13M, 13N, 13O e Figuras 13P, 13Q e 13R).

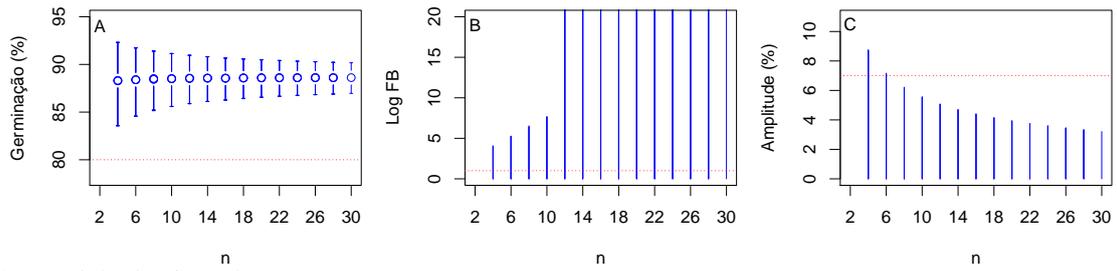
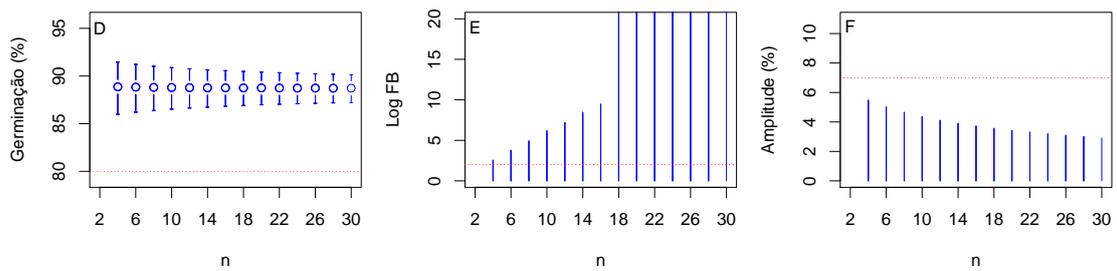
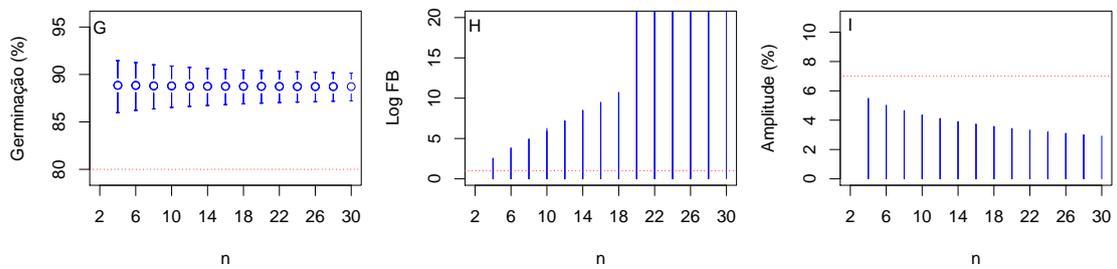
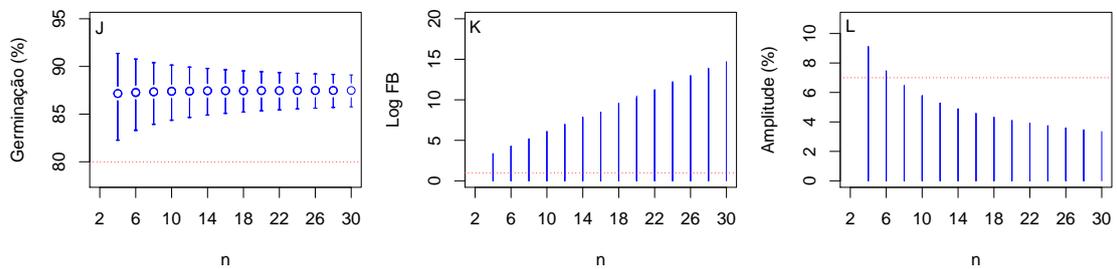
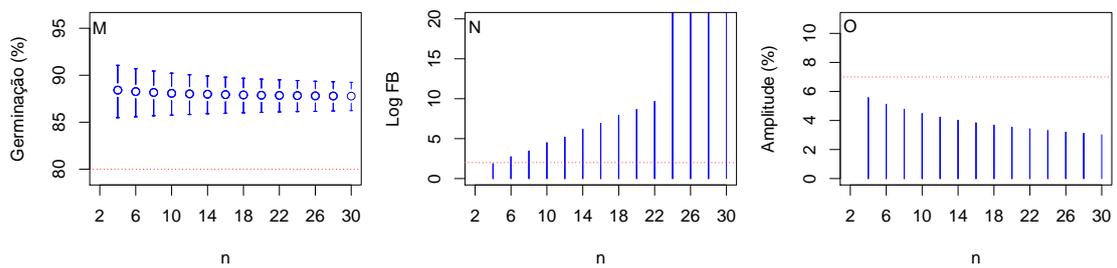
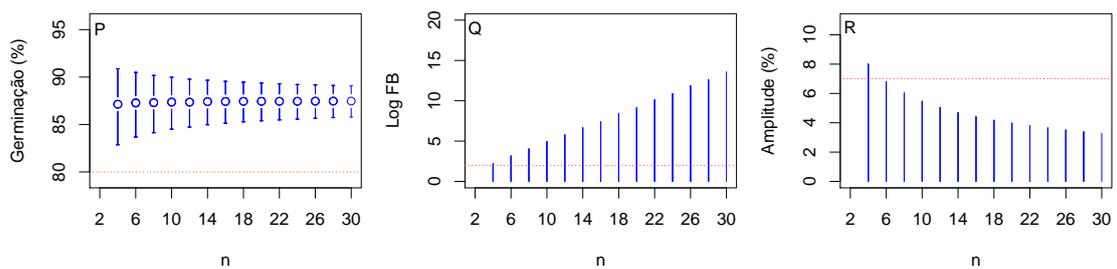
(a) *Priori* não informativa Lote 3FL11(b) *Priori* dados do TC Lote 3FL11(c) *Priori* dados do laboratório Lote 3FL11(d) *Priori* não informativa Lote 3FL12(e) *Priori* dados do TC Lote 3FL12(f) *Priori* dados do laboratório Lote 3FL12

Figura 13 – Proporção de sementes de soja germinadas com intervalos de credibilidade de 95% (A, D, G, J, M e P), Log do Fator de Bayes (B, E, H, K, N e Q) e Amplitude (C, F, I, L, O e R), para os diferentes tamanhos amostrais do Fornecedor 3, Cultivar F, lotes 11 e 12.

Fonte: do autor.

Observa-se na Tabela 6 que, com o uso da *Priori* Não Informativa (PNI) e da *Priori* Informativa com dados do Laboratório (PIL), considerando a análise de todos os lotes, reduziu-se de 30 para 12 sacos o tamanho amostral. Com o uso da *Priori* Informativa com dados do Termo de Conformidade (PITC), considerando a análise de todos os lotes, reduziu-se de 30 para 14 sacos o tamanho amostral.

Tabela 6 – Resumo dos resultados obtidos com a análise dos 12 lotes com *Priori* Não Informativa (PNI), *Priori* Informativa com dados do Termo de Conformidade (PITC) e *Priori* Informativa com dados das análises do laboratório (PIL).

Lote	PNI	PITC	PIL
1	10	10	—
2	8	6	—
3	10	14	—
4	8	6	—
5	8	4	6
6	6	4	6
7	6	4	6
8	6	4	6
9	12	4	12
10	10	10	8
11	8	4	4
12	8	6	6
Média	8,33	6,33	6,5

Fonte: do autor.

A PITC, em 9 dos 12 lotes analisados, reduziu em pelo menos 2 sacos o tamanho amostral, quando comparado com a PNI. Dos 8 lotes analisados com PIL, 4 apresentaram redução de pelo menos 2 sacos no tamanho amostral, comparado com a PNI. Comparando a PITC com a PIL, observa-se que em 5 dos 8 lotes analisados, a PITC obteve melhor desempenho na redução do tamanho amostral.

Quando utilizou-se PNI, o tamanho amostral mínimo foi 6 sacos, o máximo foi de 12 sacos e a média foi de 8,33 sacos por lote. Três lote obtiveram precisão desejada com tamanho amostral 6 sacos, cinco lotes com tamanho amostral 8 sacos, três lotes com tamanho amostral

10 sacos e um lote com tamanho amostral 12 sacos.

Quando utilizou-se PITC, o tamanho amostral mínimo foi 4 sacos, o máximo foi 14 sacos e a média foi de 6,33 sacos por lote. Seis lotes analisados mantiveram precisão desejada com tamanho amostral 4 sacos, três lotes com tamanho amostral 6 sacos, dois lotes com tamanho amostral 10 sacos e 1 lote com tamanho amostral 14 sacos.

Quando utilizou-se PIL, o tamanho amostral mínimo foi 4 sacos, o máximo foi 12 sacos e a média foi de 6,5 sacos por lote. Um lote analisado manteve precisão desejada com tamanho amostral 4 sacos, cinco lotes com tamanho amostral 6 sacos, um lote com tamanho amostral 8 sacos e um lote com tamanho amostral 12 sacos.

Diante dos resultados, considerou-se que a melhor priori utilizada no estudo foi a PIL, pois, considerando a análise de todos os lotes, reduziu de 30 para 12 sacos o tamanho amostral, o que equivale a uma redução de 60%. Os dados desta *priori* são registrados mais próximo da sementeira, o que vem de encontro com o objetivo do controle de qualidade, de proporcionar sementes de boa qualidade ao produtor, como evidenciado no estudo de Scheeren *et al.* (2010), que verificou a necessidade de se utilizar um teste de vigor antes da sementeira, para uma verificação das informações constantes nos termos de conformidade.

O uso da priori informativa em ambas estratégias apresentou uma redução maior no tamanho amostral em comparação com o uso da priori não informativa para todos os lotes. Os resultados corroboram com o estudo de Carvalho; Beijo e Muniz (2017), que utilizaram a teoria bayesiana no ajuste do modelo em simulação de dados de adsorção de metais traço, considerando as distribuições à priori informativa e não informativa. Os resultados indicaram que as estimativas obtidas com uso das distribuições à priori informativas dos parâmetros apresentaram maior precisão, mesmo em tamanhos amostrais inferiores, comparativamente aquelas obtidas com a distribuição à priori não informativa. A inferência bayesiana mostra-se uma alternativa viável e de grande potencial para contornar a limitação prática acerca do tamanho amostral, de modo que seja possível obterem-se estimativas precisas utilizando-se menor quantidade de amostra.

Observa-se com a análise dos gráficos do log do FB nas Figuras de 8 a 13 (B, E, H, K, N e Q), que o tamanho amostral é diretamente proporcional ao valor encontrado para o log do FB, ou seja a medida que se aumenta o tamanho amostral, aumenta-se também o valor do fator de Bayes. Os resultados estão de acordo com o estudo feito por Schonbrodt e Wagenmakers (2017) que também verificaram que a medida que se aumenta o tamanho amostral, aumenta-se

a força da evidência e que se pode assumir tamanhos amostrais mínimos de acordo com o peso da evidência desejada.

A amplitude do intervalo de credibilidade é utilizada em muitos estudos para determinação de tamanhos amostrais ótimos. Essas combinações de erro de estimação (precisão) e tamanho de amostra são eficientes para o planejamento de experimentos, de acordo com a precisão desejada. Observa-se com a análise dos gráficos, nas Figuras de 8 à 13 (C, F, I, L, O e R), que a medida que se aumenta o tamanho amostral, diminui-se a amplitude do intervalo de credibilidade, como comprovado no estudo de Cargnelluti Filho *et al.* (2015), onde determinou-se o tamanho amostral para a estimação da média de caracteres morfológicos e produtivos de aveia preta *avena sativa*, com erros de estimação fixados em 2%, 4%, 6%, 8%, 10%, 15% e 20% da média, com grau de confiança de 95%. Os autores concluíram que a opção por menores tamanhos de amostra proporciona maiores erros de estimação.

O lote 3EL09 germinou abaixo de 80%, ou seja, tratá-se de um lote que não seria aceito pela revenda. As sementes analisadas não apresentavam danos visíveis a olho nu, ainda assim apresentou porcentagem de germinação abaixo do mínimo estabelecido. Os resultados corroboram com os estudos de Maeda *et al.* (1977), que verificaram que embora os testes de germinação apresentem maiores índices nas sementes que eram aparentemente livres de danificações, este fato não impediu que sementes aparentemente boas apresentassem necrose no teste de germinação em laboratório, demonstrando que mesmo com bom aspecto, elas apresentavam injúrias não visíveis a olho nu.

Analisando o lote 3EL09 com *priori* informativa com os dados do TC, na Figura 6D, pode-se observar que a informação à *priori* elevou os valores das médias, para os tamanhos amostrais menores ou iguais a 22 sacos, o fato se deve pela média à *priori* ser de 89% para esse Fornecedor, conforme descrito na seção 3.3.1. No entanto, para nenhum tamanho amostral analisado o lote seria aceito, não alterando o critério de decisão, o que se confirma pelo log do Fator de Bayes. Verifica-se a importância de se utilizar mais de um método teórico para análise da precisão e tomada de decisão, pois no caso do estudo de Brigenthi; Rezende e Brigenthi (2011), que objetivaram estimar através da metodologia sequencial bayesiana, a probabilidade de sucesso da distribuição da proporção, aplicando esse método de estimação em dados reais de infestação do psilídeo no alecrim do campo, os autores verificaram que proporções próximas a 0 ou 1, tiveram muita influência das *priori* informativas adotadas e resultaram em estimativas distantes da proporção real.

5 CONCLUSÕES

O uso da inferência bayesiana, com diferentes tipos de *priori*, possibilita a verificação da porcentagem de germinação de sementes de soja, com um tamanho amostral menor, sem grandes perdas na precisão. As informações da distribuição *à priori* influenciam na redução do tamanho amostral.

O tamanho amostral é diretamente proporcional ao valor encontrado para o Log do FB e inversamente proporcional à amplitude do intervalo de credibilidade de 95%.

A distribuição *à priori* fundamentada em resultados obtidos no laboratório de análises de sementes do ano anterior apresentou melhores resultados, utilizando um tamanho amostral maior ou igual a 12 sacos, não ocorre elevada redução na precisão e não se altera a decisão tomada de quando se adota amostra de tamanho 30 sacos.

O tamanho amostral mínimo depende do fornecedor, da *priori* e do lote analisado. Considerando a análise feita sob todos os lotes, utilizando um tamanho amostral maior ou igual a 14 sacos, não ocorre elevada redução na precisão e não se altera a decisão tomada quando se adota amostra de tamanho 30 sacos. Logo, reduz-se no mínimo, em torno de 50% a quantidade de sacos furados para retirada de amostra, independente da *priori* adotada.

6 SUGESTÕES DE TRABALHOS FUTUROS

O teste de tetrazólio é um teste bioquímico que tem como principal objetivo determinar a viabilidade das sementes e identificar danos mecânicos, danos por umidade e/ou danos por percevejos.

As revendas de sementes têm adotado o teste de tetrazólio para identificar possíveis danos causados a semente, buscando sempre ofertar aos produtores de grãos, sementes de qualidade. A RAS recomenda que o teste deve ser realizado em 400 sementes por lote, subdivididas em quatro repetições de 100 sementes ou oito repetições de 50, retiradas ao acaso de uma amostra representativa da amostra submetida. Opcionalmente, o teste poderá ser realizado em duas repetições de 100 sementes ou quatro de 50, totalizando 200 sementes.

Novamente há uma limitação no tamanho amostral, pois a análise de um lote demanda cerca de 1 hora e 30 minutos do profissional responsável. Como o volume de lotes recebido pelas revendas geralmente é grande, a redução da quantidade de sementes para análise no teste de tetrazólio mantendo uma precisão aceitável e não mudando a resposta viabilidade da semente analisada, seria de grande contribuição e poderia reduzir os custos com as análises.

Além disso, alguns estudos como o de Damasceno *et al.* (2015), evidenciaram correlações entre teste de germinação e teste de tetrazólio, portanto, sugere-se que estudos futuros utilizem a inferência bayesiana com as mesmas metodologias e novas metodologias para análise, buscando implantar melhorias no processo do teste de tetrazólio.

REFERÊNCIAS

- AMORIM, F. A. HAMAWAKI, O. T. SOUSA, L. B. LANA, R. M. Q. HAMAWAKI, C. D. L. Época de semeadura no potencial produtivo de soja em Uberlândia - MG. **Ciências Agrárias**, Londrina, v. 32, n. 1, p. 1793-1802, 2011.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Regras para análise de sementes**, Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília-DF, 399 p. 2009.
- BEHRENS, J. H. SILVA, M. A. A. P. Atitude do Consumidor em Relação à Soja e Produtos Derivados, **Ciência e Tecnologia de Alimentos**, Campinas-SP, v. 24, n. 3, p. 431-439, 2004.
- BERGER, J. O. **Statistical Decision Theory and Bayesian Analysis**, Springer-Verlag, second edition, 617 p. New York, 1985.
- BRIGHENTI, C. R. G. RESENDE, M. BRIGHENTI, D. M. Estimativa Sequencial Bayesiana Aplicada à Proporção de Infestação de Psilídeos em Alecrim do Campo, **Revista Brasileira de Biometria**, São João Del Rei - MG, v. 29, n. 2, p. 342-354, 2011.
- CAMPOS, H. **Estatística experimental não-paramétrica**, 4. ed. Piracicaba: ESALQ-USP, 350 p. 1983.
- CARGNELLUTI FILHO, A. GONÇALVES, E. C. P. Estimativa do número de repetições para a avaliação de caracteres de produtividade e de morfologia em genótipos de soja, **Comunicata Scientiae**, Santa Maria - RS, v. 2, n. 1, p. 25-33, 2011.
- CARGNELLUTI FILHO, A. TOEBE, M. ALVES, B. M. BURIN, C. SANTOS, G. O. FACCO, G. NEU, I. M. M. Dimensionamento amostral para avaliar caracteres morfológicos e produtivos de aveia preta em épocas de avaliação, **Ciência Rural**, Santa Maria - RS, v. 45, n° 1, p. 9-13, 2015.
- CARVALHO, D. T.; BEIJO, L. A.; MUNIZ, J. A. Uma abordagem Bayesiana da Isoterma de Langmuir. **Revista Brasileira de Biometria**, Lavras - MG, v. 35, n. 2, p. 376-401, 2017.
- Companhia Nacional de Abastecimento (CONAB), ISSN 2318-6852, **Acompanhamento da safra brasileira grãos**, v. 6, Safra 2018/19, N. 6 - Sexto levantamento, Brasília-DF, p. 1-145. março 2019.
- COOK, J. D. Conjugate prior relationships, disponível em https://www.johndcook.com/blog/conjugate_prior_diagram/#beta, acesso em 28/12/2019.
- DAMASCENO, L. F. JUNGHANS, T. G. SILVA, J. J. AUD, F. F. **Correlação entre o teste de tetrazólio e o teste de germinação na aferição da viabilidade das sementes de Passiflora edulis**, In 9ª Jornada Científica - Embrapa Mandioca e Fruticultura, 2015, Cruz das Almas - BA.
- DELLA LUCIA, S. M. MINIM, V. P. R. SILVA, C. H. O. MINIM, L. A. Fatores da embalagem de café orgânico torrado e moído na intenção de compra do consumidor, **Ciência e**

Tecnologia de Alimentos, Viçosa - MG, v. 27, n. 3, p. 485-491, 2007.

ESTEFANEL, V. SACCOL, A. V. SCHNEIDER, F. M. BURIOL, G. A. HELDWEIN, A. B. MANFRON, P. A. Tamanho da amostra para estimar características agronômicas da soja, **Centro de Ciências Rurais**, Santa Maria-RS, v. 14, n. 3-4, p. 221-229, 1984.

FACHIN, O. **Fundamentos de Metodologia Científica** Ed Saraiva. 206 p. 2004,

FINK, D. A Compendium of Conjugate Priors, Environmental Statistics Group, Department of Biology, Montana State University, Bozeman, MT 59717, May 1997. disponível em <https://www.johndcook.com/CompendiumOfConjugatePriors.pdf>, acesso em 28/12/2019.

FRANÇA NETO, J. B. KRZYZANOWSKI, F. C. PÁDUA, G. P. HENNING, A. A. **Tecnologia da produção de semente de soja de alta qualidade**. EMBRAPA Soja Circular Técnica 40, 12 p. Londrina/PR, 2007.

GELMAN, A. CARLIN, J. B. STERN, H. S. DUNSON, D. B. VEHTARI, A. RUBIN D. B. **Bayesian Data Analysis** Chapman and Hall, 3ª ed. 639 p. 2014.

GOMES, F. P. **Curso de Estatística Experimental** Piracicaba-SP: FEALQ, 2009, vol. 15, 451 p.

HIRAKURI, M. H. LAZZARATTO, J. J. **O Agronegócios da Soja nos Contextos Mundial e Brasileiro**, Embrapa, Londrina-PR, 2014.

HOGG, R. V. CRAIG, A. T. **Introduction to mathematical statistics**. 6. ed. Upper Saddle River: Prentice Hall, 2005.

JEFFREYS, H. **Theory of Probability**. 3ª ed. Oxford University, 1961.

KINAS, P. G. ANDRADE, H. A. **Introdução a Análise Bayesiana (com R)**, maisQnada, 258 p. Porto Alegre-RS, 2010.

KRZYZANOWSKI, F. C. FRANÇA NETO, J. B. HENNING, A. A. COSTG, N. P. **O controle de qualidade agregando valor à semente de soja - Série sementes**, EMBRAPA Soja Circular técnica 54, ISSN 1516-7860, Londrina-PR, 2008.

KRZYZANOWSKI, F. C. FRANÇA NETO, J. B. HENNING, A. A. **A alta qualidade da sementes de soja: fator importante para a produção da cultura**, EMBRAPA Soja Circular técnica 136, ISSN 2176-2864, Londrina-PR, 2018.

LIMA, J. G. POZO, O. V. C. FREITAS, R. R. Startups no agronegócio brasileiro: uma revisão sobre as potencialidades do setor. **Brazilian Journal of Production Engineering (BJPE)**, São Mateus - ES, v. 3, n. 1, p. 107-121, 2017.

MAEDA, J. A. MIRANDA, M. A. C. ARKCOLL, D. ZINK, E. Influência de diversos fatores externos sobre a qualidade da semente de soja. **Bragantia**, Campinas - SP, v. 36, n. 17, p. 147-154, 1977.

MARTINS FILHO, S. SILVA, F. F. CARNEIRO, A. P. S. MUNIZ, J. A. Abordagem Bayesiana das curvas de crescimento de duas cultivares de feijoeiro **Ciência Rural**, Santa

Maria - RS, v. 38, n. 6, p. 1516-1521, 2008.

NASCIMENTO, M. SILVA, F. F., SAFADI, T., NASCIMENTO, A. C. C. FERREIRA, R. P. CRUZ, C. D. Abordagem bayesiana para avaliação da adaptabilidade e estabilidade de genótipos de alfafa, **Pesquisa Agropecuária Brasileira**, Brasília-DF, v. 46, n. 1, p. 26-32, 2011.

MORENTIN, P. A. BUSSAD, W. O. **Estatística Básica**, 5^a. Ed., Editora Saraiva, São Paulo-SP, 2004.

PORTO, W. A. C. **A utilização da estatística experimental em indústrias da grande Salvador e seu ensino**, Tese (Doutorado em Estatística e Experimentação Agropecuária), Universidade Federal de Lavras (UFLA), Lavras-MG, 245 p. 2014.

PAULINO, D. C. TURKMAN, M. A. A. MURTEIRA, B. **Estatística Bayesiana**. Lisboa: FCG, 2003. 446 p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

RAMALHO, M. A. P. FERREIRA, D. F. OLIVEIRA, A. C. **Experimentação em Genética e Melhoramento de Plantas**, Lavras-MG, UFLA, 3 Ed. 328 p. 2012,

ROCHA, G. C. RUBIO NETO, A. CRUZ, S. J. S. CAMPOS, G. W. B. CASTRO, A. C. O. SIMON, G. A. Qualidade Fisiológica de sementes de soja tratadas e armazenadas **Revista Científica**, Ilha Solteira - SP, v. 1, n. 5, 2017.

RODRIGUES, G. M. A. NEVES, R. F. ANGELICA, R. S. Beneficiamento de uma argila tipo paligorskita da bacia de S. Luis-Grajaú, região de Alcântara, MA, e sua utilização como adsorvente de fósforo. **Cerâmica**, Belem - PA, n. 60, p. 117-126, 2014.

ROSSI, R. M. **Introdução aos métodos bayesianos na análise de dados zootécnicos com uso do WinBUGS**, Eduem, 245 p. Maringá-PR, 2011.

SCHONBRODT, F. D. WAGENMAKERS, E. J. Bayes factor design analysis: Planning for compelling evidence, **Psychon Bull Rev**, Amsterdam (Germany), n. 25, p. 128–142, 2017.

SANTOS, D. HAESBAERT, F. M. PUHL, O. J. SANTOS, J. R. A. LUCIO, A. D. Suficiência Amostral para alface cultivada em diferentes ambientes, **Ciência Rural**, Santa Maria-RS, v. 40, n. 4, 2010.

STONE, J. V. **Bayes' rule: A tutorial introduction to Bayesian analysis**. [s.l]: Subtel, 2013.

SCHEEREN, B. R. PESKE, S. T. SCHUCH, L. O. B. BARROS, A. C. A. Qualidade fisiológica e produtiva se semente de soja **Revista Brasileira de Sementes**, Pelotas-RS, v. 23, n. 3, p. 35-41, 2010.

TEODORO, P. E. NASCIMENTO, M. TORRES, F. E. BARROSO, L. M. A SAGRILO, E. Perspectiva Bayesiana na seleção de genótipos de feijão-caupi em ensaios de valor de cultivo e uso, **Pesquisa agropecuária brasileira**, Brasília - DF, v. 50, n. 10, p. 878-885, 2015.

TOEBE, M. BANDEIRA, C. T. FORTES, S. K. G. CARVALHO, J. O. TARTAGLIA, F. L. TAMBARA, A. L. MELO, P. J. Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*, **Bragantia**, Campinas-SP, v. 76, n. 1, p. 45-53, 2017.

VIEIRA, S. HOFFMANN, R. **Estatística Experimental**, São Paulo-SP: ATLAS, 1989, 179 p.

WETZELS, R. WAGENMAKERS, E. J. A default Bayesian hypothesis test for correlations and partial correlations. **Psychonomic Bulletin & Review**, Amsterdam (Germany), v. 19, n. 6, p. 1057-1064, 2012.

WINTER, E. M. W. ALMEIDA, M. I. M. OLIVEIRA, E. G. MARTINS, E. N. NATEL, A. S. SUREK, D. Aplicação do método Bayesiano na estimação de correlações genéticas e fenótipos de peso em codornas de corte em várias idades, **Revista Brasileira de Zootecnia**, Curitiba - PR, v. 35, n. 4, p. 1684-1690, 2006.

APÊNDICES

APÊNDICE A - Resultados referentes às médias de cultivares dentro de cada empresa, o 1º quantil, a mediana, a média, o 3º quantil, o coeficiente de variação e o resultado do Teste de Kruskal-Wallis (KW) dos cultivares dentro dos fornecedores 1 e 2.

Tabela 7 – Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos cultivares dentro dos Fornecedores 1 e 2.

Fornecedor 1						
Cultivar	Q_1	Mediana	Q_3	Média	CV	valor p KW
A	0,800	0,850	0,849	0,900	8,017	0,2217
B	0,780	0,840	0,829	0,880	10,162	
C	0,760	0,840	0,805	0,880	14,371	
Fornecedor 2						
Cultivar	Q_1	Mediana	Q_3	Média	CV	p-valor KW
D	0,840	0,880	0,881	0,9050	5,987	0,0879
E	0,850	0,880	0,870	0,900	4,759	
F	0,860	0,880	0,865	0,900	7,839	

Fonte: do autor.

APÊNDICE B - Resultados referentes às médias dos lotes dentro de cada cultivar, o 1º quantil, a mediana, a média, o 3º quantil, o coeficiente de variação e o valor do Teste de Kruskal-Wallis (KW) para os lotes dentro dos cultivares da Empresa A e da Empresa B.

Tabela 8 – Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos lotes dentro dos cultivares do Fornecedor 1.

Fornecedor 1						
Cultivar A						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,915	0,930	0,930	0,945	2,776	0,0027
2	0,880	0,880	0,885	0,885	1,130	
3	0,870	0,890	0,885	0,905	3,860	
4	0,810	0,830	0,845	0,865	8,055	
5	0,840	0,850	0,865	0,875	4,377	
6	0,795	0,800	0,800	0,805	2,041	
7	0,720	0,730	0,735	0,745	2,605	
Cultivar B						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,795	0,810	0,805	0,820	2,379	0,0019
2	0,650	0,700	0,735	0,785	18,101	
3	0,855	0,870	0,875	0,890	3,904	
4	0,695	0,730	0,750	0,785	10,777	
5	0,890	0,910	0,910	0,930	4,575	
6	0,875	0,890	0,905	0,920	5,812	
7	0,755	0,780	0,785	0,810	5,649	
8	0,650	0,690	0,680	0,720	7,204	
9	0,805	0,840	0,845	0,880	8,935	
10	0,870	0,900	0,905	0,935	6,599	
11	0,820	0,830	0,830	0,840	1,391	
12	0,810	0,860	0,830	0,880	9,123	
13	0,855	0,870	0,880	0,895	4,910	
14	0,825	0,850	0,840	0,865	5,143	
15	0,850	0,870	0,860	0,880	3,289	
Cultivar C						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,875	0,900	0,900	0,925	4,057	< 0,0001
2	0,680	0,720	0,780	0,890	15,677	
3	0,895	0,910	0,905	0,920	2,116	
4	0,880	0,890	0,895	0,905	2,140	
5	0,855	0,860	0,865	0,870	2,909	
6	0,875	0,890	0,890	0,905	2,901	
7	0,830	0,850	0,845	0,865	4,042	
8	0,860	0,890	0,880	0,910	6,691	
9	0,855	0,870	0,865	0,880	2,214	
10	0,850	0,870	0,860	0,880	3,289	
11	0,915	0,930	0,930	0,945	2,776	

Tabela 9 – Resumo dos dados, 1º quantil, mediana, média, 3º quantil, coeficiente de variação e valor do Teste de KW dos lotes dentro dos cultivares do Fornecedor 2.

Fornecedor 2						
Cultivar D						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,895	0,910	0,910	0,925	2,837	0,0002
2	0,830	0,850	0,850	0,870	4,898	
3	0,820	0,830	0,830	0,840	1,391	
4	0,875	0,880	0,880	0,885	1,856	
5	0,860	0,870	0,870	0,880	1,327	
6	0,895	0,920	0,925	0,950	4,794	
7	0,900	0,910	0,915	0,925	2,093	
8	0,955	0,970	0,970	0,985	2,662	
9	0,815	0,830	0,830	0,845	3,111	
10	0,810	0,830	0,825	0,845	4,140	
Cultivar E						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,815	0,840	0,845	0,870	5,248	0,1081
2	0,880	0,890	0,895	0,905	2,140	
Cultivar F						
Lote	Q_1	Mediana	Média	Q_3	CV	p-valor KW
1	0,855	0,870	0,865	0,880	2,214	0,0153
2	0,870	0,890	0,885	0,905	3,859	
3	0,865	0,890	0,885	0,905	3,859	
4	0,775	0,820	0,815	0,860	6,454	
5	0,815	0,820	0,820	0,825	1,991	
6	0,810	0,820	0,835	0,845	7,152	
7	0,820	0,840	0,852	0,880	4,453	
8	0,860	0,880	0,864	0,880	5,653	
9	0,900	0,900	0,900	0,900	0,000	
10	0,800	0,810	0,810	0,820	1,426	
11	0,875	0,880	0,880	0,885	1,856	

Fonte: do autor.

APÊNDICE C - Rotinas utilizadas para análise dos dados

Rotina reamostragem

Nessa rotina faz-se a leitura dos dados com a quantidade de sementes germinadas por saco. Em seguida, para cada tamanho amostral (28, 26, ..., 4), seleciona-se 5000 subamostras por sorteio, com o comando `sample`. Dessa forma, calcula-se o limite inferior, o limite superior, a média, a amplitude, o log do Fator de Bayes, a probabilidade de ser maior que 80% e o desvio padrão para a média e para a mediana. Em seguida há um comando para salvar os resultados.

```
setwd("Leitura dos dados")
dados=c(38,44,44,45,43,44,47,48,39,45,44,34,46,46,45,35,36,42,38,
48,46,42,44,43,44,41,34,34,45,47) #dados de germinação de um
lote
nsubamostra=4 #tamanho da subamostra
#reamostrando os dados da amostra
ni = length(dados)
NS<-5000 #num.simulação
NR<-7 #num.respostas (colunas)
z <- matrix(0,NS,NR,
dimnames=list(c(),c("Pest","LIICr","LSICr","Amplitude","Prob>0.8",
"FTB","DP"))) #Inicialização do vetor z
#Fazendo o looping
for (i in 1:NS) #Loop para obter
mat.auxi<-mat.or.vec(0,1)
mat.auxi<-dados
N=length(mat.auxi)
sortelem=sample(N,nsubamostra) ; #sorteando o 1º elemento entre
1 e k)
amostra<-mat.auxi[sortelem] #pega os elementos do vetor pop que
estão nas posições
Tsuce<-sum(amostra) #total de sucesso em todas as amostras
Tsuba<- length(amostra)*50
#total de sementes em todas as amostras
```

```

alfa0=98 #hiperparâmetro
beta0= 9.574 #hiperparâmetro
a1=alfa0+Tsuce
b1=beta0+Tsuba-Tsuce
p0h0=pbeta(0.80,alfa0,beta0,lower.tail = F)
p0h1=pbeta(0.80,alfa0,beta0,lower.tail = T)
R0h0=p0h0/p0h1
z[i,1] <- a1/(a1+b1) #média
z[i,2] <- qbeta(0.025,a1,b1) #Limite Inferior Cr95%
z[i,3] <- qbeta(0.975,a1,b1) #Limite Superiro Cr95%
z[i,4] <- z[i,3]-z[i,2] #Aplitude ICr
z[i,5] <- pbeta(0.80,a1,b1,lower.tail = F) #Prob se ser maior
ou igual que 80%
z[i,6] <- log((z[i,5]/(1-z[i,5])/R0h0), 10) #log do Fator Bayes
z[i,7] <-sqrt((a1*b1)/((a1+b1+1)*(a1+b1)^2)) #Desvio padrão

resultado=matrix(0,NR,NS)
resultado= round(z,4)
cbind(mean(z[,1]),mean(z[,2]),mean(z[,3]),mean(z[,4]),mean(z[,5]),mean(z[,6]),mean(z[,7]))
#resultados para a média
cbind(median(z[,1]),median(z[,2]),median(z[,3]),median(z[,4]),median(z[,5]),median(z[,6]),median(z[,7]))
#resultados para a mediana
#Salvando simulação em xls
write.table(resultado,"EmprA-Cult1-Lotex,n=4,NS=1.xls")

```

Rotina para construção dos gráficos do intervalo de credibilidade, log do fator de Bayes e amplitude

```

#GRÁFICO 1ALO01
dados<-read.table("dados-Lote-A1X.txt",h=T)
dados
attach(dados)
LI<-LI*100
LS<-LS*100

```

```

MEDIA<-MEDIA*100
AMPL<-AMPLITUDE*100
N<-N
FB<-FB-media
summary(FB)

# Gráfico da proporção
par(mfrow = c(1,3), cex=1.25, oma = c(1,1,1,1), mar=c(4,4,1,1))
library(gplots)
plotCI(x=N, y=MEDIA, uiw=LS-MEDIA,
liw=MEDIA-LI,xaxt="n",xlim=c(2,30),
ylim=c(78,95), type="b",
col="blue", lwd=1,main=, xlab="n", ylab="Germinação (%)")
axis(1,at=c(2,6,10,14,18,22,26,30))
text(2,94,"D")
abline(h=80,lty=3,lwd=1,col="red")

# Gráfico do Log do Fator de Bayes
plot(N,FB,xaxt="n",xlim=c(2,30),
ylim=c(0,20),type="h",col="blue",
main=, xlab="n", ylab="Log FB")
axis(1,at=c(2,6,10,14,18,22,26,30))
text(2,19,"E")
abline(h=2,lty=3,lwd=1,col="red")

# Gráfico da Amplitude
plot(N,AMPL,xaxt="n",xlim=c(2,30),
ylim=c(0,11),type="h",col="blue",
main=, xlab="n", ylab="Amplitude (%)")
axis(1,at=c(2,6,10,14,18,22,26,30))
text(2,10.5,"F")
abline(h=7,lty=3,lwd=1,col="red")

```