



CAMILLA MARQUES BARROSO

**PROPOSTA DE ANÁLISE ESTATÍSTICA DE
TESTES TRIANGULARES CONSIDERANDO
A DISTRIBUIÇÃO BINOMIAL**

**LAVRAS - MG
2013**

CAMILLA MARQUES BARROSO

**PROPOSTA DE ANÁLISE ESTATÍSTICA DE TESTES TRIANGULARES
CONSIDERANDO A DISTRIBUIÇÃO BINOMIAL**

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

Orientador
Dr. Júlio Sílvio de Sousa Bueno Filho

**LAVRAS - MG
2013**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Barroso, Camilla Marques.

Proposta de análise estatística de testes triangulares considerando a distribuição binomial / Camilla Marques Barroso. - Lavras : UFLA, 2013.

62 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.

Orientador: Júlio Sílvio de Sousa Bueno Filho.

Bibliografia.

1. Limiar de detecção. 2. Máxima verossimilhança. 3. Modelo Thurstoniano. I. Universidade Federal de Lavras. II. Título.

CDD - 519.24

CAMILLA MARQUES BARROSO

**PROPOSTA DE ANÁLISE ESTATÍSTICA DE TESTES TRIANGULARES
CONSIDERANDO A DISTRIBUIÇÃO BINOMIAL**

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

APROVADA em: 10 de abril de 2013.

Dr. Edwin Moisés Marcos Ortega

ESALQ

Dr. Augusto Ramalho de Morais

UFLA

Dr. Júlio Sílvio de Sousa Bueno Filho
Orientador

**LAVRAS - MG
2013**

À minha mãe, Fátima

DEDICO

AGRADECIMENTOS

A Deus, pela vida;

À minha querida mãe Fátima, pelo exemplo, pelo amor e pela dedicação na boa educação;

Aos meus irmãos, Renan e Louise, pelo companheirismo;

A todos os meus familiares e amigos, pelo apoio e carinho;

Ao meu noivo, Rodrigo, pelo amor e parceria;

Ao meu orientador Júlio Sílvio de Sousa Bueno Filho, pelos conhecimentos e esclarecimentos intelectuais e morais confiados a mim, pela paciência e compreensão;

Aos professores pelas importantes contribuições nesta dissertação, especialmente aos professores, Dr. Edwin Moisés Marcos Ortega, Dr. Augusto Ramalho de Moraes e Dra. Thelma Sáfadi, por aceitarem colaborar com seus conhecimentos neste Projeto;

A todos funcionários pela colaboração e carinho;

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade de cursar o mestrado;

À CAPES, pela concessão da bolsa de estudos, tornando financeiramente possível a realização do mestrado;

A todos que de alguma forma contribuíram, mesmo inconscientemente, para a realização deste trabalho.

”A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.”

Albert Einstein

RESUMO

Testes triangulares são métodos utilizados em análise sensorial para determinar se existem diferenças entre amostras. O provador recebe três amostras codificadas, sendo duas iguais e uma diferente, devendo avaliar as amostras na ordem fornecida e identificar a amostra diferente. Experimentos como estes são analisados usando aproximações normais, no entanto pode-se representar a habilidade em discriminar determinado atributo como a variável latente de um modelo thurstoniano. Objetivou-se com este projeto comparar duas formas de análise, considerando distribuição binomial e aproximação normal, em um teste triangular para adulterantes de café. Os tratamentos foram arrançados em esquema fatorial, considerando dois fatores: tipos de adulterante e percentagem de adulteração. O delineamento experimental foi em blocos casualizados com vinte provadores, sendo cada um considerado um bloco de controle local. Ajustaram-se modelos de regressão para os efeitos das percentagens em cada tipo de adulterante. Para a análise usando distribuição binomial é especificada a função de verossimilhança e especificado o algoritmo completo de obtenção das estimativas numéricas de máxima verossimilhança. Foram calculadas estimativas pontuais e por intervalo para as curvas de regressão e os limiares de detecção. O ajuste considerando o modelo binomial resulta em intervalos de confiança mais estreitos e alterou testes de hipóteses sobre os limiares de detecção. A metodologia está implementada em R e pode ser usada rotineiramente por cientistas da área de alimentos.

Palavras-chave: Limiar de detecção. Máxima verossimilhança. Modelo thurstoniano.

ABSTRACT

Triangular discrimination tests are a widely used method in sensory analysis to detect differences between food samples. The tester proves three coded samples, being two equal and one different. He must pick up the different one. Usually this kind of experiments are analysed under normal approximation. However, we can represent the ability to spot differences as a latent variable in a Thurstonian model. The objective of this dissertation is to compare both ways to analyse a triangular test for coffee adulterants. Treatment structure is a factorial scheme with two factors: type of adulterant and percentage of adulteration. Experimental units where the samples assigned to 20 people, being each person a block (randomized blocks design). Regression models were fitted to the factorial effects. For the analysis using binomial distribution, full algorithm to yield Maximum likelihood estimates is presented. Point and interval estimates for regression models and the threshold of detecting ability were worked out. Binomial fitting result in narrower confidence regions and changed some hypothesis on detection threshold. The whole method is implemented in R and can be used as routine analysis for food scientists.

Keywords: Thresholds. Maximum likelihood. Thurstonian model.

SUMÁRIO

1	INTRODUÇÃO.....	10
2	REFERENCIAL TEÓRICO.....	11
2.1	Análise Sensorial.....	11
2.1.1	Tipos de Teste Sensorial.....	12
2.2	Teste Triangular.....	13
2.2.1	Procedimento.....	14
2.2.2	Análise usual.....	14
2.3	Modelos Lineares Generalizados.....	15
2.4	Modelos Thurstonianos.....	15
2.4.1	Modelo Thurstoniano para um teste triangular	16
2.4.2	Aproximação normal.....	17
2.4.3	Análise considerando distribuição binomial.....	18
2.5	Estimação.....	19
2.6	Seleção de modelos.....	20
2.7	Limiar de detecção.....	20
3	METODOLOGIA.....	22
3.1	Descrição dos dados do exemplo real.....	22
3.2	Análise supondo normalidade.....	22
3.2.1	Modelo Linear.....	23
3.3	Análise supondo resposta binomial.....	24
3.4	Comparar as análises.....	24
4	RESULTADOS E DISCUSSÃO.....	25
4.1	Análise de modelo fixo supondo normalidade.....	25
4.2	Análise de modelo fixo supondo resposta binomial.....	29
4.3	Ajuste de curvas de regressão.....	35
4.4	Limiares de detecção.....	37
5	CONCLUSÃO.....	43
	REFERÊNCIAS.....	44
	APÊNDICE.....	46

1 INTRODUÇÃO

Os testes triangulares são métodos utilizados em análise sensorial para determinar se existe pequena diferença entre duas ou mais amostras. São técnicas rápidas e usadas quando essas diferenças são pequenas, mas precisam ser detectadas. Esse procedimento se caracteriza como um tipo de teste de discriminação no qual são apresentadas ao provador três amostras codificadas, sendo duas iguais, do mesmo tratamento e outra de um segundo tratamento supostamente diferente, cabendo ao provador avaliar as amostras na ordem fornecida e identificar a “diferente” (KEMP; HOLLOWOOD; HORT, 2009).

Na análise usual para dois tratamentos, determina-se, para um dado número de provadores, o número de respostas corretas. Esse valor pode ser comparado a valores críticos tabelados para testes triangulares, que representam o número mínimo de respostas corretas para que se possa concluir que existe diferença significativa entre as amostras, levando em consideração o nível de significância.

O teste triangular apresenta como vantagem, a menor probabilidade de acertar ao acaso e diferenciar as amostras de maneira global e, como desvantagens, ser pouco prático com número elevado de amostras (tratamentos) e poder ser afetado pela fadiga sensorial (do provador).

Neste teste, a habilidade em discriminar determinado atributo não pode ser diretamente medida ou observada. Uma alternativa é empregar modelos que relacionam um conjunto de variáveis observáveis com um conjunto de variáveis não observáveis, chamadas variáveis latentes. Um exemplo são os modelos Thurstonianos que são utilizados para investigar a existência e medir essas estruturas latentes, como itens de um questionário.

O objetivo com este projeto é comparar a análise estatística de um teste triangular usando a distribuição binomial com a análise supondo distribuição normal, especialmente quanto às estimativas dos limiares de detecção.

2 REFERENCIAL TEÓRICO

Esta seção aborda a teoria dos testes sensoriais, em particular, testes triangulares. Em seguida é feita uma revisão sobre modelos lineares generalizados, em especial modelos Thurstonianos para estes tipos de teste. O conceito de limiar de detecção é introduzido com o intuito de comparar, através deste, formas de análise de dados advindos de testes triangulares como análise considerando normalidade e considerando distribuição binomial.

2.1 Análise Sensorial

A análise sensorial é definida pela Associação Brasileira de Normas Técnicas (ABNT, 1993) como a disciplina científica usada para evocar, medir, analisar e interpretar reações das características dos alimentos e materiais como são percebidas pelos sentidos da visão, olfato, gosto, tato e audição, ou seja, é uma ciência usada para medir utilizando os sentidos humanos um determinado atributo ou a característica global do produto e quantificá-la.

Segundo (BIEDRZYCKI, 2009), a análise sensorial é muito utilizada com diferentes objetivos, como no desenvolvimento de novos produtos, na modificação de produtos existentes no mercado, na avaliação da validade de um produto, na comparação de um produto com produtos concorrentes, na identificação da preferência dos consumidores, na melhoria da qualidade de um produto, entre outros.

No passado, a produção de alimentos, muitas vezes dependia da sensibilidade sensorial de um único provador que estava encarregado da produção ou tomava decisões sobre as mudanças do processo para garantir que o produto tivesse as características desejáveis. Esta foi a tradição histórica de mestres cervejeiros, provadores de vinho, de produtos lácteos e inspetores de outros alimentos, que atuaram como árbitros na avaliação sensorial (HEYMANN; LAWLESS, 2003).

A análise moderna substituiu esses provadores individuais por painéis de pessoas e a utilização de métodos específicos que formaram os experimentos planejados. Essa mudança ocorreu por várias razões. Primeiro, foi reconhecido que os julgamentos de um painel são, em geral, mais confiáveis do que o julgamento

de um único indivíduo e isso implicava menos risco uma vez que o único provador poderia de alguma forma estar indisponível para tomar decisões. Em segundo lugar, um único provador pode ou não refletir o que os consumidores realmente esperam de um produto (HEYMANN; LAWLESS, 2003).

2.1.1 Tipos de Teste Sensorial

Uma preocupação quando se escolhe um método sensorial é assegurar que esse método é adequado para responder às questões que estão sendo analisadas sobre o produto em teste. Por esta razão, os testes são geralmente classificados de acordo com a sua principal finalidade (HEYMANN; LAWLESS, 2003).

Os métodos sensoriais são agrupados em afetivos e analíticos. Nos métodos afetivos, é possível utilizar pessoas sem treinamento prévio, pois buscam-se respostas resultantes de estímulos e reações espontâneas do indivíduo ao degustar e avaliar o produto. Tem como finalidade avaliar a aceitação e preferência dos consumidores em relação a um ou mais produtos.

Nos métodos analíticos é necessária a seleção ou treinamento da equipe sensorial para realização das análises e também é exigida uma avaliação objetiva, ou seja, não são consideradas as preferências dos membros da equipe. Estes testes, por sua vez, também se dividem em dois grupos: testes descritivos e discriminativos.

O método sensorial descritivo envolve técnicas que permitem a avaliação da intensidade dos atributos sensoriais de produtos. Por outro lado, os testes de discriminação devem ser usados quando se quer determinar se existem pequenas diferenças entre duas amostras. Neste caso não importam qualidades específicas de cada amostra, mas a diferença entre elas. É possível que duas amostras sejam quimicamente diferentes na formulação, mas os seres humanos não percebam esta diferença.

As indústrias quando reformulam um produto, usando diferentes ingredientes, podem não querer que o consumidor detecte diferenças. Por exemplo, um fabricante de sorvete pode pretender substituir o sabor baunilha usado em seu sorvete por um sabor mais barato de baunilha. No entanto, ele não quer que o con-

sumidor perceba diferença no produto. Um teste de discriminação corretamente executado, indicando que as duas formulações de sorvetes não são sensivelmente diferentes permitiria que a empresa fizesse a substituição com o risco reduzido (HEYMANN; LAWLESS, 2003).

Estes testes podem também ser utilizados quando uma alteração de processamento é feita esperando que não afetaria as características sensoriais do produto. Em ambos os casos o objetivo do teste é não rejeitar a hipótese nula, ou seja, aceitar que os produtos não são estatisticamente diferentes. No entanto, quando uma empresa reformula um produto para fazer uma "nova versão melhorada", então o teste de discriminação pode ser utilizado para indicar que as duas formulações são diferentes. Neste caso o objetivo da discriminação é rejeitar a hipótese nula, ou seja, os produtos não são estatisticamente iguais. Se os dados indicam que as duas formulações são diferentes, então a indústria deve fazer outro teste que indique que a formulação "nova" é percebida como melhor pelo consumidor alvo (HEYMANN; LAWLESS, 2003).

Se a diferença entre as amostras é muito grande e, portanto, óbvia, testes de discriminação não são úteis. Em outras palavras, testes de discriminação são mais úteis quando as diferenças entre as amostras são sutis. Entretanto, essas diferenças sutis tornam o risco de erro do tipo II mais provável, ou seja, não rejeitar a hipótese nula quando a mesma na realidade é falsa (HEYMANN; LAWLESS, 2003).

2.2 Teste Triangular

O teste triangular deve ser usado quando o objetivo é determinar se existe diferença sensorial entre dois produtos. Tem a sua utilização limitada no caso de produtos parecidos e que se encontram muito confusos. Mas, este método é eficiente em situações como:

- a) determinar se existem diferenças nos produtos resultante de mudança nos ingredientes, processamento, embalagem ou armazenamento;
- b) determinar se existe uma diferença global, em que nenhum atributo específico pode ser identificado como tendo sido afetado;
- c) para selecionar provadores;

2.2.1 Procedimento

Segundo (KEMP; HOLLOWOOD; HORT, 2009), o teste triangular se resume em apresentar para cada provador, três amostras codificadas, sendo duas iguais e uma diferente. O provador examina cada produto e seleciona o diferente. A probabilidade de acertar ao acaso é $1/3$. Determina-se, para um dado número de provadores, o número de respostas corretas a partir do qual afirma-se se existe ou não diferença entre as amostras.

Geralmente, são usados de 20 a 40 provadores no teste. Os provadores devem, de preferência, conhecer o método triangular e o produto em teste, especialmente porque a memória de gosto desempenha um papel importante no procedimento (CARR; CIVILLE; MEILGAARD, 2006).

Se possível, deve-se oferecer as amostras simultaneamente, no entanto, amostras que são encorpadas (deixam um gosto) ou que mostram pequenas diferenças na aparência podem ser oferecidas sequencialmente, sem invalidar o teste.

Em delineamentos balanceados, prepara-se um número igual das seis possíveis combinações das três amostras (ABB, BAA, AAB, BBA, ABA, BAB) e apresenta-se de forma aleatória aos provadores. É comum, no entanto, montar ensaios com números diferentes de provador por combinação de produtos, devido a problemas de agendamento.

2.2.2 Análise usual

Para avaliação dos resultados obtidos, (ROESSLER et al, 1978) têm sugerido uma série de tabelas que facilitam determinar se uma diferença estatística entre duas amostras foi detectada em um teste de discriminação.

Dessa forma, em um teste triangular para dois tratamentos, determina-se, para um dado número de provadores, o número de respostas corretas. Esse valor é comparado com um valor crítico tabelado que representa o número mínimo de respostas corretas para que se possa concluir que existe diferença significativa entre as amostras, levando em consideração o nível de significância (ROESSLER et al, 1978).

2.3 Modelos Lineares Generalizados

Em geral, o uso de modelos lineares clássicos, não são adequados para analisar dados de proporções, pois as pressuposições do modelo não são atendidas. Uma alternativa é a utilização da teoria de modelos lineares generalizados, sendo a distribuição binomial, um caso particular (MCCULLAGH; NELDER, 1989).

Os modelos lineares generalizados formulados por Nelder e Wedderburn (1972), mostraram ser possível unir vários modelos estatísticos em uma só classe na qual os dados não precisam ter distribuição normal e a média não é necessariamente uma combinação linear dos parâmetros, mas que alguma função da média o é.

Esses modelos podem ser definidos pela seguinte estrutura:

- a) Componente aleatório: uma distribuição de probabilidade, pertencente à família exponencial, para a variável resposta;
- b) Componente sistemático: um conjunto de variáveis explicativas descrevendo a estrutura linear do modelo;
- c) Função de ligação: uma função que liga o componente aleatório e o sistemático (MCCULLAGH; NELDER, 1989).

2.4 Modelos Thurstonianos

Um modelo Thurstoniano é essencialmente uma função que associa uma variável contínua a uma variável discreta. Em geral esta associação decorre do uso de funções de ligação que devolvem a probabilidade de ocorrência da variável discreta para níveis da variável contínua (variável latente).

Alguns fatores como habilidade em discriminar determinado atributo em análise sensorial não podem ser diretamente medidos ou observados. Uma alternativa para lidar com essa questão é empregar modelos que relacionam variáveis observáveis com variáveis latentes. Esses modelos são empregados quando o interesse é medir fatores não observáveis, que possuem um significado intuitivo para todos, porém que na maioria das vezes não podem ser medidos diretamente. Exemplos disso são os modelos Thurstonianos, que são modelos lineares generalizados

utilizados para investigar a existência e medir essas estruturas latentes, como itens de um questionário (BROCKHOFF; CHRISTENSEN, 2010).

2.4.1 Modelo Thurstoniano para um teste triangular

Suponha que em um teste triangular, os provadores recebam duas amostras de um produto A e uma de um produto B e seja solicitado a eles identificar a amostra diferente.

Os provadores darão as respostas baseados em suas percepções, embora a variabilidade de percepção possa levá-los a uma resposta incorreta no que diz respeito à real diferença entre os produtos (CARR; CIVILLE; MEILGAARD, 2006).

Em um teste triangular, os modelos Thurstonianos assumem que as mudanças nessa percepção seguem uma distribuição de probabilidade (normal ou logística, por exemplo). O quantil desta distribuição indica o nível de percepção correspondente a determinada probabilidade de discriminar corretamente. Quando dois produtos têm magnitudes sensoriais muito diferentes, eles não serão confundidos porque não há sobreposição em suas distribuições de percepção (CHRISTENSEN, 2012).

A Figura 1 apresenta os resultados de dois provadores em um teste triangular. No primeiro ensaio, o provador selecionou a amostra b porque perceptivelmente é mais distante de ambas as amostras a_1 e a_2 do que as amostras a_1 e a_2 são uma da outra. Nesse ensaio o provador dá uma resposta correta porque a sua percepção é consistente com a real diferença entre os produtos. No segundo ensaio, o provador seleciona a amostra a_1 porque, perceptivelmente, é mais distante das amostras a_2 e b do que a_2 e b são uma da outra. Nesse ensaio, o provador dá uma resposta incorreta porque a percepção do provador é inconsistente com a real diferença entre os produtos. Em ambos ensaios, o provador aplica a regra de decisão correta. Entretanto, devido à natureza probabilística da percepção, em um ensaio a resposta é correta e em outro a resposta é incorreta (CARR; CIVILLE; MEILGAARD, 2006).

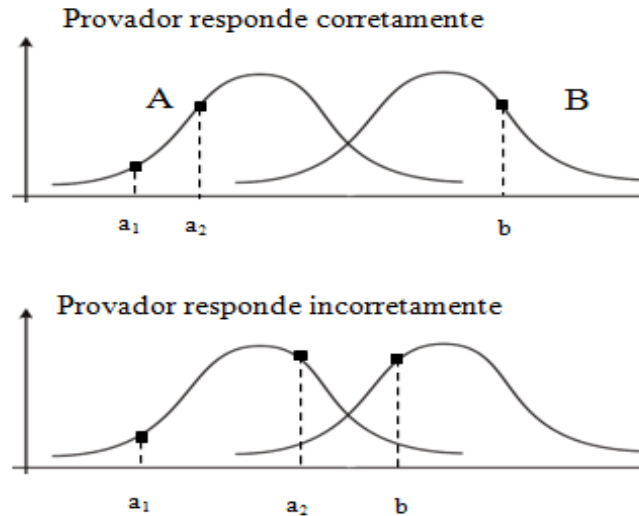


Figura 1 Respostas corretas e incorretas em um teste triangular

Situações desse tipo não são raras em experimentos que envolvem análise sensorial, daí a necessidade de se aprimorar as técnicas de avaliação de dados com essas características.

2.4.2 Aproximação Normal

Pode-se interpretar os testes triangulares como tendo uma resposta binária, ou seja, uma distribuição Bernoulli para cada ensaio.

Porém, pode-se também supor uma aproximação, considerando a proporção de respostas corretas em cada teste, ou seja, supor que essas proporções sigam uma distribuição normal. Testes de normalidade podem ser usados para analisar essa suposição.

Neste caso usa-se a análise de variância (ANAVA) para verificar se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente.

Considerando o modelo linear adequado, pode-se aplicar a análise de variância para verificar os efeitos dos tratamentos. A validade da análise de variância

depende de algumas pressuposições, que particularmente no caso dos modelos lineares, faz-se suposições de distribuição normal dos erros, aditividade dos efeitos do modelo e homogeneidade das variâncias. Nos casos em que a pressuposição de normalidade não é satisfeita, uma alternativa que vem sendo adotada é o uso da transformação de Box e Cox, que fizeram uma ampla abordagem sobre transformação de dados, a qual transforma os dados com a finalidade de se obter normalidade ou normalidade aproximada (BOX; COX, 1964).

2.4.3 Análise considerando distribuição binomial

Uma modelagem mais direta do modelo Thurstoniano é dada quando se usa a distribuição binomial (BROCKHOFF; CHRISTENSEN, 2010).

A função de verossimilhança dos n ensaios se refere a um experimento com b provadores para avaliar τ tratamentos e pode ser representada por:

$$L(\boldsymbol{\pi}, \mathbf{y}) = \prod_{i=1}^n p(y_i/\pi_i) = \prod_{i=1}^n \binom{m}{y_i} \pi_i^{y_i} (1-\pi_i)^{m-y_i}$$

em que

y_i : número de respostas corretas

m : número de tentativas

$$\pi_i = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1+e^{-\eta_i}} \right)$$

$\eta_i = \mathbf{x}_i^T \boldsymbol{\theta}$ a variável latente

\mathbf{x}_i = linhas da matriz de delineamento

$$\boldsymbol{\theta} = \left[\boldsymbol{\beta}', \boldsymbol{\tau}' \right]'$$

$\boldsymbol{\beta}$ = vetor dos efeitos de bloco

$\boldsymbol{\tau}$ = vetor dos efeitos de tratamento

Aplicando o operador natural à função de verossimilhança, obtém-se a função log-verossimilhança dada por:

$$\ell = \sum_{i=1}^n \log \binom{m}{y_i} + \sum_{i=1}^n y_i \log \pi_i + \sum_{i=1}^n (m - y_i) \log(1 - \pi_i)$$

As derivadas primeiras da função de verossimilhança em relação aos vetores de parâmetros podem ser obtidas da seguinte forma, sendo b representando blocos (provadores) e τ representando tratamentos:

$$\frac{\partial \ell}{\partial \mathbf{b}} = \frac{\partial \ell}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \mathbf{b}}$$

e

$$\frac{\partial \ell}{\partial \boldsymbol{\tau}} = \frac{\partial \ell}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \boldsymbol{\tau}}$$

As derivadas segundas da função de verossimilhança em relação aos vetores de parâmetros podem ser obtidas como se segue:

$$\frac{\partial^2 \ell}{\partial \mathbf{b}^2} = \frac{\partial}{\partial \mathbf{b}} \left[\frac{\partial \ell}{\partial \mathbf{b}} \right]$$

e

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} = \frac{\partial}{\partial \boldsymbol{\tau}} \left[\frac{\partial \ell}{\partial \boldsymbol{\tau}} \right]$$

Essas derivadas são usadas para calcular as soluções para os parâmetros usando o algoritmo de convergência de Newton-Raphson.

2.5 Estimação

Para os modelos lineares generalizados considera-se a estimação por máxima verossimilhança, sendo a função escore total para o parâmetro $\boldsymbol{\theta}$ dada por

$$\mathbf{U}_{\boldsymbol{\theta}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

em que

$$\ell = \sum_{i=1}^n \log \binom{m}{y_i} + \sum_{i=1}^n y_i \log \pi_i + \sum_{i=1}^n (m - y_i) \log(1 - \pi_i)$$

A estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ do vetor de parâmetros $\boldsymbol{\theta} = [\mathbf{b}', \boldsymbol{\tau}']'$ é calculada igualando-se $\mathbf{U}_{\boldsymbol{\theta}}$ a zero. Em geral, as equações $\mathbf{U}_{\boldsymbol{\theta}} = 0$ não são lineares e tem que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson (MCCULLAGH; NELDER, 1989).

Tal processo é definido expandindo-se a função escore $\mathbf{U}_{\boldsymbol{\theta}}$ em série de Taylor em torno de um valor inicial $\boldsymbol{\theta}^{(0)}$, de modo que

$$\mathbf{U}_{\boldsymbol{\theta}} \simeq \mathbf{U}_{\boldsymbol{\theta}}^{(0)} + U'_{\boldsymbol{\theta}}{}^{(0)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}),$$

em que \mathbf{U}'_{θ} corresponde a primeira derivada de \mathbf{U}_{θ} em relação a θ . Assim, repetindo o procedimento acima, chega-se ao processo iterativo (MCCULLAGH; NELDER, 1989):

$$\theta^{(i+1)} = \theta^{(i)} + [(-\mathbf{U}'_{\theta})^{-1}]^{(i)} \mathbf{U}_{\theta}^{(i)}$$

em que $i = 0, 1, 2, \dots$ representa o número de iterações.

2.6 Seleção de modelos

A escolha do modelo apropriado é extremamente importante na análise de dados. Procura-se o modelo que envolva o mínimo de parâmetros possível e que explique bem o comportamento da variável resposta.

O método da razão de verossimilhanças consiste em comparar o máximo valor da verossimilhança, restrito por $H_0 : \theta = 0$, com o máximo valor da verossimilhança sob a hipótese $H_1 : \theta \neq 0$, que não é restrita, por meio da razão de verossimilhanças:

$$LR = \frac{\max_{H_0} L(\cdot)}{\max_{H_1} L(\cdot)}$$

(WALD, 1943) mostrou que, sob H_0 ,

$$-2\ln(LR) \sim \chi^2(v)$$

para n grande, sendo v o número de parâmetros estimados sob H_1 menos o número de parâmetros estimados sob H_0 e considerando \log o logaritmo natural.

2.7 Limiar de detecção

Uma das primeiras características sensoriais a ser medida foi o limiar absoluto. O limiar absoluto ou de detecção é o nível abaixo do qual nenhuma sensação é produzida por um estímulo e acima do qual uma sensação chega à consciência do provador (PENG; JAEGER; HAUTUS, 2012).

Na prática, algumas complicações surgem em tentar aplicar a ideia de limite. Primeiro verifica-se que existe uma variabilidade no ponto em que os provadores mudam a sua resposta. Existe variabilidade mesmo dentro de um único indivíduo. Em uma sequência de ensaios, mesmo dentro da mesma sessão experimental, o ponto no qual uma pessoa muda a sua ou as suas respostas serão diferentes. Há diferenças entre os indivíduos, especialmente em relação à sensibilidade a sabor e cheiro. Isto levou ao estabelecimento de regras comuns para a definição de um limiar, tais como o nível em que a detecção ocorre 50% das tentativas (PENG; JAEGER; HAUTUS, 2012).

3 METODOLOGIA

Esta seção apresenta a descrição dos dados usados neste trabalho, os métodos de análise considerando normalidade e considerando distribuição binomial e formas de comparar ambas análises.

3.1 Descrição dos dados do exemplo real

Os dados¹ utilizados nesse trabalho advindos de um teste triangular são provenientes de um experimento conduzido no esquema fatorial 3 x 6 para avaliar a adição de diferentes tipos de adulterantes em amostras de café. Os fatores estudados foram: tipos de adulterantes em três níveis (casca de café, milho e palha de café) e proporção de adulterante em seis níveis (1%, 10%, 20%, 30%, 40% e 50%). O delineamento experimental foi em blocos casualizados, sendo que cada um dos vinte provadores constituiu um bloco. Durante o teste, possivelmente devido a problemas de agendamento dos provadores, houve a ocorrência de parcelas perdidas.

Cada provador teve três chances para cada um dos dezoito tratamentos no teste triangular. Dessa forma, obteve-se a proporção de acertos para cada observação.

3.2 Análise supondo normalidade

No teste triangular realizado, foram dadas três chances a cada provador em cada ocasião de prova, para cada tratamento. Cada provador deveria avaliar todos os tratamentos. Em seguida, registrou-se os números de acertos e de erros e foi descontada a porcentagem de acertos ao acaso (1/3), calculando-se a proporção de acertos corrigida (PAC) para cada observação por meio de:

$$PAC = \frac{\text{acertos}}{\text{acertos} + \text{erros}} - \frac{1}{3}$$

¹TAVARES, K. M. Lavras: UFLA/DCA, 2012

3.2.1 Modelo Linear

Considerando que os dados obtidos são provenientes de um experimento no delineamento em blocos casualizados sendo os tratamentos e os blocos de efeitos fixos, e ainda, que os erros possuem distribuição normal de média zero e variância constante, o vetor das observações pode ser descrito como tendo distribuição normal de média $X\tau + Zb$ e variância $I\sigma_e^2$, sendo representado na forma matricial por:

$$Y \sim N(X\tau + Zb, I\sigma_e^2)$$

em que

Y: vetor dos valores observados;

X: matriz de delineamento dos tratamentos;

τ : efeitos de tratamentos;

Z: matriz de delineamento dos blocos;

b: efeitos de blocos.

Aplicou-se a análise de variância aos valores obtidos de PAC para verificar os efeitos dos tratamentos. Ajustaram-se modelos de regressão para os efeitos do fatorial e, em seguida, estimaram-se os coeficientes do modelo de regressão mais bem ajustado.

Construiu-se então, gráficos das curvas de regressão estimadas, estabelecendo no eixo das abcissas a percentagem de adulterante e no eixo das ordenadas, a proporção de acertos estimada. Traçou-se uma linha horizontal na proporção 0,25 (valor de referência em análises sensoriais) e encontrou-se a percentagem do adulterante associado. Esse valor corresponde à solução pontual para os limiares de detecção para cada adulterante, ou seja, o valor a partir do qual o adulterante é percebido. Os valores exatos podem ser encontrados igualando-se as equações de regressão a 0,25.

Considerando *EP* as estimativas dos erros padrões das estimativas dos parâmetros, pode-se calcular a estatística para avaliação dos coeficientes do modelo de regressão, testando a hipótese $H_0 : \theta_i = 0$ vs $H_a : \theta_i \neq 0$, sendo $\hat{\theta}_i$ as estimativas dos coeficientes do modelo quadrático.

$$t_{calc} = \frac{\hat{\theta}_i - \theta_i}{EP}$$

Construíram-se intervalos de confiança para as proporções de acertos e isto permite construir intervalos de confiança aproximados para os limiares de detecção.

3.3 Análise supondo resposta binomial

A função de verossimilhança é dada por:

$$L(\boldsymbol{\pi}, \mathbf{y}) = \prod_{i=1}^n p(y_i | \pi_i) = P(y_i | \pi_i) = \prod_{i=1}^n \binom{m}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m - y_i}$$

sendo:

\mathbf{y} = vetor dos valores observados $\pi_i = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1 + e^{-\eta_i}} \right)$

$\eta_i = \mathbf{x}_i^T \boldsymbol{\theta}$ a variável latente

\mathbf{x}_i = linhas da matriz de delineamento

$\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\tau}']'$

$\boldsymbol{\beta}$ = vetor dos efeitos de bloco

$\boldsymbol{\tau}$ = vetor dos efeitos de tratamento

Na análise de verossimilhança binomial é especificada a função de verossimilhança. A seguir, calculam-se o gradiente e a Hessiana, que são então utilizados para estimar os coeficientes de regressão e, em seguida, ajustar as curvas para cada adulterante.

Foram construídas curvas com os intervalos de confiança para a proporção de detecção e, essas curvas foram utilizadas para aproximar intervalos de confiança para o limiar de detecção.

3.4 Comparar as análises

Os melhores modelos de regressão em cada metodologia foram comparados quanto à precisão das estimativas por meio de gráficos das curvas de regressão estimadas e, em especial, quanto à estimação do limiar de detecção de adulterante através de intervalos de confiança.

4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os principais resultados obtidos considerando normalidade e considerando distribuição binomial.

4.1 Análise de modelo fixo supondo normalidade

Supondo normalidade considera-se o modelo:

$$E [PAC_{ijk}] = \mu + b_k + A_i + P_j + (AP)_{ij}$$

sendo:

PAC_{ijk} é o valor da proporção de acertos na parcela que recebeu o adúlterante i , na percentagem de adúlterante j , provador k ;

μ = constante experimental comum;

b_k = efeito do provador k (bloco), para $k = 1, \dots, 20$;

A_i = efeito do tipo de adúlterante i , para $i = 1, \dots, 3$;

P_j = efeito da percentagem de adúlterante j , para $j=1, \dots, 6$;

$(AP)_{ij}$ = efeito da interação entre os tipos x percentagem de adúlterantes.

A transformação Box-Cox é uma das possíveis formas de contornar o problema de dados que não obedecem os pressupostos da análise de variância, como normalidade dos dados. Considerando a percentagem de acertos corrigida, encontramos $\lambda=0,95$. O intervalo para λ inclui o valor $\lambda = 1$, sendo, portanto, desnecessária qualquer transformação.

A Figura 2 apresenta os valores da log-verossimilhança em relação a λ , quando se utilizou a transformação de Box-Cox. Observa-se que o máximo da função log-verossimilhança ocorreu próximo de "1".

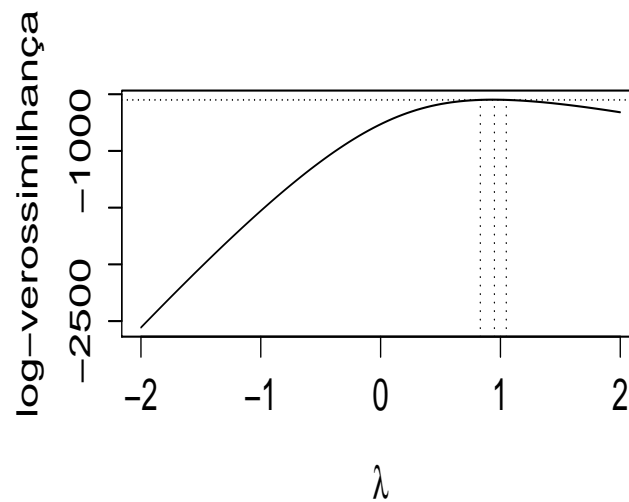


Figura 2 Perfil de log-verossimilhança em função do parâmetro λ da transformação Box-Cox

Os dados de cada provador supõe-se serem independentes, ou seja, cada resposta de um provador não tem qualquer relação com a de outro. Outro requisito necessário para que a análise de variância fosse realizada, homogeneidade de variâncias, foi examinado e atendido.

Neste trabalho foi utilizado um esquema fatorial, estudando o efeito de dois fatores, tipo de adulterante e percentagem de adulterante, simultaneamente. O delineamento experimental foi em blocos casualizados, sendo cada provador considerado um bloco. Realizou-se a análise de variância para testar os efeitos de blocos e tratamentos, sendo que para identificar as diferenças significativas foi feito o desdobramento do número de graus de liberdade de tratamentos conforme mostra a Tabela 1.

Tabela 1 Análise de variância da proporção de acertos corrigida

Fonte de variação	GL	SQ	QM	valor F	Pr
Blocos	19	3,59	0,19	2,19	0,0033*
Tratamentos	(17)	13,63	0,80	9,28	< 0,0001*
Adulterante (A)	2	1,32	0,66	7,63	0,0006*
Porcentagem (P)	5	10,11	2,02	23,37	< 0,0001*
A X P	10	2,21	0,22	2,55	0,0057*
Resíduo	300	25,95	0,09		
Total	336	43,18			

'*' O teste é significativo no nível de 1% de probabilidade

Pode-se observar por meio da análise de variância 1 que houve diferença significativa no nível de 1% de probabilidade entre os adulterantes. O mesmo resultado foi obtido para porcentagem e para a interação. Logo, os adulterantes e as porcentagens apresentam efeitos diferentes sobre a proporção de respostas corretas. No caso da significância para a interação, pode-se inferir que os efeitos dos tipos de adulterantes sobre a proporção de respostas corretas, dependem da concentração ou porcentagem destes adulterantes.

Para avaliar o efeito das porcentagens de adulterantes sobre a proporção de acertos para cada tipo de adulterante, utilizou-se o modelo completo com o modelo de regressão quadrática com intercepto comum e a análise de variância foi não-significativa ($p > 0,19$), demonstrando que os dois modelos não diferem estatisticamente. Assim, verifica-se que o modelo de regressão quadrática teve um bom ajuste e pode ser usado para estudar o comportamento da variável dependente em relação às variáveis explicativas. Esse modelo pode ser representado como:

$$E [PAC_{ijk}] = \beta_0 + \beta_{1ij}x_j + \beta_{2ij}x_j^2 + b_k$$

$$V [PAC_{ijk}] = \sigma^2$$

sendo os índices i, j, k indicando adulterantes, porcentagem e provedores, respectivamente e b_k representando provedor k.

A análise de variância para regressão está apresentada na Tabela 2:

Tabela 2 Análise de variância da percentagem de acertos corrigida, com avaliação do efeito de regressão linear e quadrática para cada adulterante

Fonte de variação	GL	SQ	QM	valor F	Pr
Blocos	19	3,59	0,19	2,19	0,0033*
Tratamentos	(17)	13,63	0,80	9,28	< 0,001*
Reg. Linear	3	11,40	3,80	43,40	< 0,001*
Casca	1	1,35	1,35	15,59	< 0,001*
Milho	1	2,45	2,45	28,29	< 0,001*
Palha	1	0,39	0,39	4,51	0,0345*
Reg. Quad	3	0,77	0,25	2,94	0,0333*
Casca	1	0,19	0,19	2,22	0,1300
Milho	1	0,67	0,67	7,77	0,0056*
Palha	1	0,02	0,02	0,18	0,6700
Desvios	11	1,46	0,13	1,53	0,5300
Resíduo	300	25,95	0,09		
Total	336	43,18			

'*' O teste é significativo no nível de 5% de probabilidade

O teste F foi significativo para regressão linear e quadrática ($p < 0,05$), indicando que é possível estabelecer uma relação entre a percentagem de acertos corrigida e adulterante-percentagem. Deve-se escolher a equação de regressão de mais alto grau que foi significativa. Logo, é possível determinar uma equação do segundo grau. O teste foi não-significativo para desvios de regressão, indicando que não existe regressão significativa maior que 2º grau, ou seja, os termos maiores que 2º grau não contribuem de modo significativo, sendo que, o abandono desses termos não resulta em falta de ajuste.

O coeficiente de determinação para o modelo ajustado é dado por:

$$R^2 = \frac{SQRL + SQRQ}{SQTrat} \simeq 0,89$$

O ajuste do modelo de regressão quadrática, considerando aduterante e percentagem, explicou 89% da variação ocorrida na proporção de acertos corrigida devida aos afeitos de tratamentos.

A Tabela 3 apresenta as estimativas dos coeficientes do modelo quadrático, as estimativas dos erros-padrões, valor de t e valor-p associados para testar a significância dos coeficientes. Verifica-se que para os aduterantes casca e palha não houve efeito significativo do coeficiente quadrático.

Tabela 3 Estimativa dos coeficientes de regressão linear e quadrática, erro padrão, valor calculado da estatística t e respectivo valor p, para estudo da proporção de acertos corrigida em cada tipo de aduterante

Aduterante	Regressão	Estimativa	Erro Padrão	t-valor	$Pr(t)$
casca	β_1	0.0190	0.0048	3.948	< 0.0001*
	β_2	-0.0002	0.0001	-1.490	0.1371
milho	β_1	0.0252	0.0047	5.319	< 0.0001*
	β_2	-0.0003	0.0000	-2.789	0.0056*
palha	β_1	0.0099	0.0047	2.124	0.0345*
	β_2	-0.0000	0.0000	-0.422	0.6733

'*' O teste é significativo ao nível de 5% de probabilidade

4.2 Análise de modelo fixo supondo resposta binomial

A função de verossimilhança é dada por:

$$L(\boldsymbol{\pi}, \mathbf{y}) = \prod_{i=1}^n P(y_i | \pi_i) = P(y_i | \Pi_i) = \prod_{i=1}^n \binom{m}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m - y_i}$$

sendo:

$$\pi_i = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1 + e^{-\eta_i}} \right)$$

$\eta_i = \mathbf{x}_i \boldsymbol{\theta}$ a variável latente

\mathbf{x}_i = linhas da matriz de delineamento

$$\boldsymbol{\theta} = [\mathbf{b}', \boldsymbol{\tau}']'$$

\mathbf{b} = vetor dos efeitos de bloco (b-1 efeitos)

$\boldsymbol{\tau}$ = vetor dos (v-1) efeitos de tratamento

Aplicando o operador logaritmo natural à função de verossimilhança, obtém-se a função log-verossimilhança dada por:

$$\ell = \sum_{i=1}^n \log \binom{m}{y_i} + \sum_{i=1}^n y_i \log \pi_i + \sum_{i=1}^n (m - y_i) \log(1 - \pi_i)$$

Calcula-se a derivada da log-verossimilhança em relação ao vetor de blocos da seguinte forma:

$$\frac{\partial \ell}{\partial \mathbf{b}} = \frac{\partial \ell}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \mathbf{b}}$$

Tem-se que:

$$\frac{\partial \ell}{\partial \pi_i} = \sum_{i=1}^n y_i \frac{1}{\pi_i} - \sum_{i=1}^n (m - y_i) \frac{1}{1 - \pi_i}$$

e

$$\frac{\partial \pi_i}{\partial \mathbf{b}} = \frac{2}{3} (-1) \{1 + e^{-(x_i \boldsymbol{\tau} + z_i b)}\}^{-2} e^{-(x_i \boldsymbol{\tau} + z_i b) (-z_i)}$$

$$\frac{\partial \pi_i}{\partial \mathbf{b}} = \frac{2}{3} z_i \frac{e^{-(x_i \boldsymbol{\tau} + z_i b)}}{[1 + e^{-(x_i \boldsymbol{\tau} + z_i b)}]^2}$$

$$\frac{\partial \pi_i}{\partial \mathbf{b}} = \frac{2}{3} z_i \frac{e^{-(x_i \boldsymbol{\tau} + z_i b)}}{[1 + e^{-(x_i \boldsymbol{\tau} + z_i b)}]} \frac{1}{[1 + e^{-(x_i \boldsymbol{\tau} + z_i b)}]}$$

$$\frac{\partial \pi_i}{\partial \mathbf{b}} = \frac{2}{3} z_i \frac{3(1 - \pi_i)}{2} \frac{3\pi_i - 1}{2}$$

$$\frac{\partial \pi_i}{\partial \mathbf{b}} = \frac{z_i}{2} (1 - \pi_i) (3\pi_i - 1)$$

Logo,

$$\frac{\partial \ell}{\partial \mathbf{b}} = \left[\sum_{i=1}^n y_i \frac{1}{\pi_i} - \sum_{i=1}^n (m - y_i) \frac{1}{1 - \pi_i} \right] \left[\frac{z_i}{2} (1 - \pi_i) (3\pi_i - 1) \right]$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n \left[\frac{y_i z_i (1-\pi_i)(3\pi_i-1)}{2\pi_i} - \frac{z_i(m-y_i)(1-\pi_i)(3\pi_i-1)}{2(1-\pi_i)} \right]$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n \left\{ (3\pi_i - 1) \left[\frac{y_i z_i (1-\pi_i)}{2\pi_i} - \frac{z_i(m-y_i)}{2} \right] \right\}$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n \left\{ \frac{(3\pi_i-1)}{2\pi_i} [z_i (y_i - m\pi_i)] \right\}$$

$$\frac{\partial \ell}{\partial \mathbf{b}} = \sum_{i=1}^n \left[\frac{(3\pi_i-1)z_i}{2} \left(\frac{y_i}{\pi_i} - m \right) \right]$$

A derivada da log-verossimilhança em relação ao vetor de tratamentos é dada por:

$$\frac{\partial \ell}{\partial \boldsymbol{\tau}} = \frac{\partial \ell}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \boldsymbol{\tau}}$$

Tem-se que:

$$\frac{\partial \ell}{\partial \pi_i} = \sum_{i=1}^n y_i \frac{1}{\pi_i} - \sum_{i=1}^n (m - y_i) \frac{1}{1-\pi_i}$$

e

$$\frac{\partial \pi_i}{\partial \boldsymbol{\tau}} = \frac{2}{3} (-1) \{1 + e^{-(x_i \tau + z_i b)}\}^{-2} e^{-(x_i \tau + z_i b)} (-x_i)$$

$$\frac{\partial \pi_i}{\partial \boldsymbol{\tau}} = \frac{2}{3} x_i \frac{e^{-(x_i \tau + z_i b)}}{[1 + e^{-(x_i \tau + z_i b)}]^2}$$

$$\frac{\partial \pi_i}{\partial \boldsymbol{\tau}} = \frac{2}{3} x_i \left[\frac{3(1-\pi_i)}{2} \right] \left[\frac{3\pi_i-1}{2} \right]$$

$$\frac{\partial \pi_i}{\partial \boldsymbol{\tau}} = \frac{x_i}{2} (1 - \pi_i)(3\pi_i - 1)$$

$$\text{Logo, } \frac{\partial \ell}{\partial \boldsymbol{\tau}} = \left[\sum_{i=1}^n y_i \frac{1}{\pi_i} - \sum_{i=1}^n (m - y_i) \frac{1}{1-\pi_i} \right] \left[\frac{x_i}{2} (1 - \pi_i)(3\pi_i - 1) \right]$$

$$\frac{\partial \ell}{\partial \boldsymbol{\tau}} = \sum_{i=1}^n \left[\frac{x_i y_i (1-\pi_i)(3\pi_i-1)}{2\pi_i} - \frac{x_i(m-y_i)(1-\pi_i)(3\pi_i-1)}{2(1-\pi_i)} \right]$$

$$\frac{\partial \ell}{\partial \boldsymbol{\tau}} = \sum_{i=1}^n [(3\pi_i - 1)] \left\{ (3\pi_i - 1) \left[\frac{x_i y_i (1-\pi_i)}{2\pi_i} - \frac{x_i(m-y_i)}{2} \right] \right\}$$

$$\frac{\partial \ell}{\partial \boldsymbol{\tau}} = \sum_{i=1}^n \left\{ (3\pi_i - 1) \left[\frac{x_i y_i - \pi_i x_i y_i - \pi_i m x_i + \pi_i x_i y_i}{2\pi_i} \right] \right\}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left[(3\pi_i - 1) \left(\frac{x_i y_i}{2\pi_i} - \frac{x_i m}{2} \right) \right] \\ \frac{\partial \ell}{\partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left[\frac{(3\pi_i - 1)}{2\pi_i} x_i y_i - \frac{(3\pi_i - 1)x_i m}{2} \right] \\ \frac{\partial \ell}{\partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left\{ \frac{x_i}{2} \left[\frac{(3\pi_i - 1)y_i}{\pi_i} - (3\pi_i - 1)m \right] \right\} \\ \frac{\partial \ell}{\partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left[\frac{x_i}{2} (3\pi_i - 1) \left(\frac{y_i}{\pi_i} - m \right) \right]\end{aligned}$$

A derivada segunda da log-verossimilhança em relação ao vetor de bloco é dada por:

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left\{ \frac{z_i^2 e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})}}{[1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})}]} \left(\frac{y_i}{\pi_i} - m \right) - \frac{2}{3} \frac{y_i z_i e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})} (3\pi_i - 1) z_i}{\pi_i^2 (1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})})^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left\{ x_i^2 \frac{3(1-\pi_i)}{2} \frac{3\pi_i - 1}{2} \left(\frac{y_i}{\pi_i} - m \right) - \frac{y_i z_i^2 (3\pi_i - 1) 3(1-\pi_i)(3\pi_i - 1)}{3\pi_i^2 4} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left\{ \frac{3z_i^2 y_i (1-\pi_i)(3\pi_i - 1)}{4\pi_i} - \frac{3z_i^2 m (1-\pi_i)(3\pi_i - 1)}{4} - \frac{z_i^2 y_i (3\pi_i - 1)^2 (1-\pi_i)}{4\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left\{ \frac{9z_i^2 y_i \pi_i (1-\pi_i)(3\pi_i - 1) - 9z_i^2 m \pi_i^2 (1-\pi_i)(3\pi_i - 1) - 3z_i^2 y_i (1-\pi_i)(3\pi_i - 1)^2}{12\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left\{ \frac{3z_i^2 (1-\pi_i)(3\pi_i - 1) [3y_i \pi_i - 3m \pi_i^2 - y_i (3\pi_i - 1)]}{12\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left[\frac{z_i^2 (1-\pi_i)(3\pi_i - 1) (y_i - 3m \pi_i^2)}{4\pi_i^2} \right] \\ \frac{\partial^2 \ell}{\partial \mathbf{b}^2} &= \sum_{i=1}^n \left[\frac{z_i^2 (1-\pi_i)(3\pi_i - 1) (y_i - 3m \pi_i^2)}{4\pi_i^2} \right]\end{aligned}$$

A derivada segunda da log-verossimilhança em relação ao vetor de tratamentos é dada por:

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left\{ \frac{x_i^2 e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})}}{[1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})}]} \left(\frac{y_i}{\pi_i} - m \right) - \frac{2}{3} \frac{y_i x_i e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})} (3\pi_i - 1) x_i}{\pi_i^2 (1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})})^2} \right\} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left\{ x_i^2 \frac{3(1-\pi_i)}{2} \frac{3\pi_i - 1}{2} \left(\frac{y_i}{\pi_i} - m \right) - \frac{y_i x_i^2 (3\pi_i - 1) 3(1-\pi_i)(3\pi_i - 1)}{3\pi_i^2 4} \right\}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left\{ \frac{3x_i^2 y_i (1-\pi_i)(3\pi_i-1)}{4\pi_i} - \frac{3x_i^2 m(1-\pi_i)(3\pi_i-1)}{4} - \frac{x_i^2 y_i (3\pi_i-1)^2 (1-\pi_i)}{4\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left\{ \frac{9x_i^2 y_i \pi_i (1-\pi_i)(3\pi_i-1) - 9x_i^2 m \pi_i^2 (1-\pi_i)(3\pi_i-1) - 3x_i^2 y_i (1-\pi_i)(3\pi_i-1)^2}{12\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left\{ \frac{3x_i^2 (1-\pi_i)(3\pi_i-1) [3y_i \pi_i - 3m \pi_i^2 - y_i (3\pi_i-1)]}{12\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left[\frac{x_i^2 (1-\pi_i)(3\pi_i-1)(y_i - 3m \pi_i^2)}{4\pi_i^2} \right] \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\tau}^2} &= \sum_{i=1}^n \left[\frac{x_i^2 (1-\pi_i)(3\pi_i-1)(y_i - 3m \pi_i^2)}{4\pi_i^2} \right]\end{aligned}$$

A derivada mista é definida da seguinte forma:

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \mathbf{b} \partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left\{ \frac{z_i y_i x_i e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})} \pi_i - (3\pi_i - 1) \frac{1}{3} e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})} x_i y_i z_i}{(1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})})^2 \pi_i^2} - \frac{m z_i e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})} x_i}{(1 + e^{-(x_i \boldsymbol{\tau} + z_i \mathbf{b})})^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b} \partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left\{ \frac{x_i y_i z_i 3(1-\pi_i)(3\pi_i-1)^2}{4(3\pi_i-1)} - \frac{x_i y_i z_i (1-\pi_i)(3\pi_i-1)^2}{4\pi_i^2} - \frac{m x_i z_i 3(1-\pi_i)(3\pi_i-1)^2}{4(3\pi_i-1)} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b} \partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left\{ \frac{3x_i y_i z_i (1-\pi_i)(3\pi_i-1) \pi_i^2 - x_i y_i z_i (1-\pi_i)(3\pi_i-1)^2 - 3\pi_i^2 m x_i z_i (1-\pi_i)(3\pi_i-1)}{4\pi_i^2} \right\} \\ \frac{\partial^2 \ell}{\partial \mathbf{b} \partial \boldsymbol{\tau}} &= \sum_{i=1}^n \left\{ \frac{x_i z_i}{4\pi_i^2} (1-\pi_i)(3\pi_i-1) [3y_i \pi_i^2 - y_i (3\pi_i-1) - 3\pi_i^2 m] \right\}\end{aligned}$$

Foi construída uma função no R para calcular as soluções de Newton-Raphson para modelo fixo com blocos e tratamentos. O pseudocódigo relacionado abaixo traz os principais passos desta função.

Passo 1: Atribuir valores iniciais arbitrários para $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = [\mathbf{b}', \boldsymbol{\tau}']'$$

Passo 2: Calcular

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta}$$

Passo 3: Calcular

$$\boldsymbol{\pi} = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1+e^{-\boldsymbol{\eta}}} \right)$$

Passo 4: Calcular o gradiente dado por:

$$\boldsymbol{\nabla} = \mathbf{X}' \left[\frac{(3\boldsymbol{\pi}-1)}{2} \left(\frac{y}{\boldsymbol{\pi}} - m \right) \right]$$

Passo 5: Definir

$$\mathbf{V} = \text{diag}[(1 - \boldsymbol{\pi})(3\boldsymbol{\pi} - 1)(y - 3m\boldsymbol{\pi}^2)(1/(4\boldsymbol{\pi}^2))]$$

Passo 6: Calcular a Hessiana

$$\mathbf{H} = \mathbf{X}' \mathbf{V} \mathbf{X}$$

Passo 7: Atualizar $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} + \left[(-\mathbf{H})^{-1} \right]^{(i)} \boldsymbol{\nabla}^{(i)}$$

Passo 8: Repetir os passos 2 a 7 até que

$$\boldsymbol{\theta}^{(i+1)} - \boldsymbol{\theta}^{(i)} < \omega$$

Em que ω é o valor especificado para o critério de convergência (sendo que foi utilizado um valor de $\omega = 0,0001$)

Os valores estimados de $\boldsymbol{\theta}$ foram usados para calcular as log-verossimilhanças e, em seguida, comparou-se o modelo completo, que inclui todos tratamentos, com o modelo de regressão linear sem intercepto comum pelo teste de razão de verossimilhanças (Tabela 4).

Tabela 4 Teste da razão de verossimilhanças

-2log(Verosimilhança)	GL	valor-p
11,138658	12	0,5170745

O teste mostrou que os dois modelos testados não diferem estatisticamente, confirmando que o modelo de regressão linear teve um bom ajuste e pode ser usado para estudar o comportamento da variável dependente em relação às variáveis explicativas.

As estimativas dos parâmetros para análise dos coeficientes do modelo linear (Tabela 5) mostram que para o adulerante casca houve um maior acréscimo na proporção de acertos e menor acréscimo com palha.

Tabela 5 Estimativas dos coeficientes de regressão, erro padrão e respectivo valor calculado da estatística t, para estudo da proporção de acertos em cada tipo de adulerante

Adulerante	Regressão	Estimativa	Erro padrão	t	p-valor
casca	β_0	-2,8951	0,5690	-5,0879*	< 0.01*
	β_1	0,1311	0,0213	6,1659*	< 0.01*
milho	β_0	-1,8052	0,4216	-4,2813*	< 0.01*
	β_1	0,1075	0,0174	6,1596*	< 0.01*
palha	β_0	-2,1701	0,5080	-4,2721*	< 0.01*
	β_1	0,0617	0,0153	4,0417*	< 0.01*

'*' O teste é significativo ao nível de 5 % de probabilidade

4.3 Ajuste de curvas de regressão

Considerando normalidade, o modelo de regressão quadrática foi o que apresentou melhor ajuste. A representação gráfica da proporção de acertos em função da percentagem dos adulerantes, considerando intercepto comum, mostra uma tendência crescente com aumento da percentagem de adulerante (Figura 3).

Por outro lado, considerando-se o modelo binomial, o modelo de regressão linear com interceptos diferentes foi o melhor ajustado e sua representação gráfica encontra-se na Figura 4. É preciso notar que o intercepto comum é uma propriedade intrínseca de modelos binomiais, não havendo necessidade de prover para isto n modelos de regressão. Em ambos os gráficos o eixo das abcissas representa a concentração de adulerante e no eixo das ordenadas está a proporção de acertos, embora no caso da aproximação normal tenha-se descontado a proporção de acertos casuais (1/3).

À medida que se aumentam as concentrações desses adulerantes, aumenta-se a proporção de acertos, ou seja, mais facilmente detectáveis eles serão. Ambas curvas apresentaram um bom ajuste e a regressão linear considerando o modelo binomial se assemelha à regressão quadrática considerando aproximação normal. Neste caso, para comparar entre os dois modelos deve-se tomar alguns aspectos específicos que permitam distingui-los. Optou-se por comparar os modelos quanto às bandas de confiança produzidas para as respectivas curvas e sua projeção no

limiar de detecção.

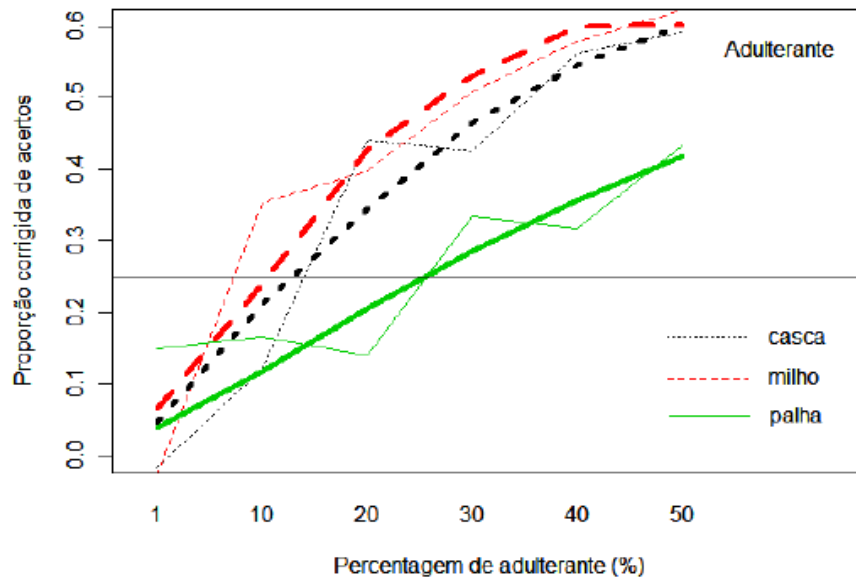


Figura 3 Representação gráfica da proporção de acertos e ajuste pela equação de regressão (quadrática) sob aproximação normal em função da porcentagem dos adulterantes para cada tipo de adulterante

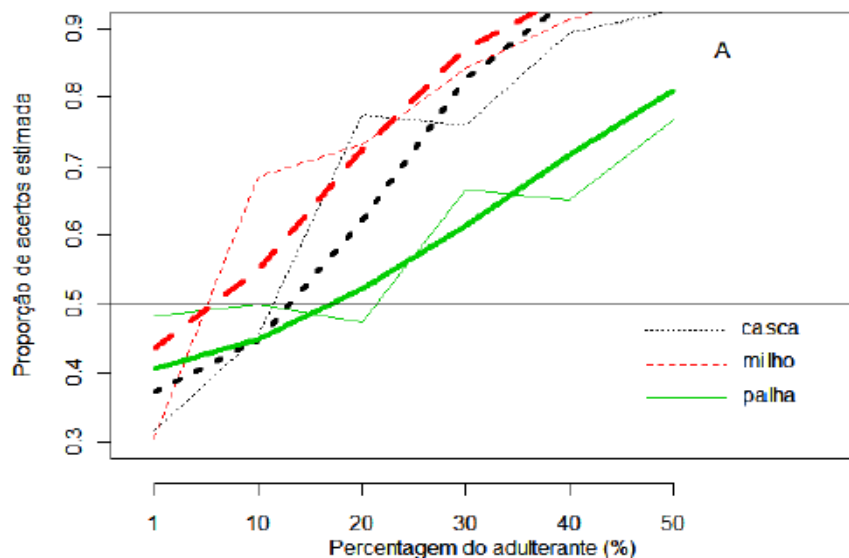


Figura 4 Representação gráfica da proporção de acertos e ajuste pela equação de regressão (linear) sob modelo binomial em função da porcentagem dos adulterantes para cada tipo de adulterante

4.4 Limiares de detecção

A partir das curvas estimadas considerando normalidade pode-se traçar uma linha horizontal na proporção 0,25 (valor de referência em análises sensoriais) e encontrar a concentração de adulterante correspondente. Esse valor corresponde à solução pontual para os limiares de detecção para cada adulterante, ou seja, o valor a partir do qual o adulterante é percebido. Os valores exatos podem ser encontrados igualando a equação de regressão a 0,25:

$$y = ax^2 + bx + c = 0,25$$

$$ax^2 + bx - 0,25 = 0$$

Considerando as respectivas estimativas dos coeficientes de regressão para cada adulterante (casca, milho e palha, respectivamente) obtém-se as equações abaixo a serem resolvidas:

$$\begin{aligned}
 \text{casca} &: -0,0002x^2 + 0,0190x - 0,25 = 0 \\
 \text{milho} &: -0,0003x^2 + 0,0252x - 0,25 = 0 \\
 \text{palha} &: -0,00002x^2 + 0,0099x - 0,25 = 0
 \end{aligned}$$

Dessa forma obtém-se as soluções pontuais para os limiares de detecção e a partir das curvas de regressão estimadas encontrou-se intervalos de confiança para as proporções e então, determinou-se intervalos de confiança aproximados para os limiares de detecção, como mostra a Tabela 6.

Tabela 6 Valores médios e limites inferiores e superiores dos limiares de detecção considerando aproximação normal

	LI (% adulterante)	média (% adulterante)	LS (% adulterante)
casca	3,9	15,8	25,4
milho	2,7	11,5	18,7
palha	8,2	26,7	> 50

Percebe-se que os intervalos para cada adulterante se sobrepõe na proporção 0,25, ou seja, não é possível diferir entre os adulterantes quanto ao limiar de detecção.

O mesmo foi feito para distribuição binomial considerando a proporção 0,5:

$$\pi = \frac{1}{3} + \frac{2}{3} \left(\frac{1}{1+e^{-\eta_i}} \right) = \frac{1}{2} \Rightarrow \eta_i = -\ln 3$$

Assim, obtém-se as equações para cada um dos adulterantes:

$$\begin{aligned}
 \text{casca} &: -2,8951 + 0,1311x = -\ln 3 \\
 \text{milho} &: -1,8052 + 0,1075x = -\ln 3 \\
 \text{palha} &: -2,1701 + 0,0617x = -\ln 3
 \end{aligned}$$

Em seguida obtém-se as soluções pontuais para os limiares de detecção e através das curvas de regressão estimadas determinou-se os intervalos de confiança

para esses limiares, como apresentado na Tabela 7.

Tabela 7 Intervalos de confiança para os limiares de detecção considerando distribuição binomial (aproximação numérica)

	LI (% adulterante)	média (% adulterante)	LS (% adulterante)
casca	11,9	13,7	14,7
milho	3,2	6,8	7,2
palha	13,1	17,4	19,1

Percebe-se que o intervalo de confiança para o limiar de detecção do adulterante milho não se sobrepõe aos outros intervalos, considerando a proporção 0,5, ou seja, o adulterante milho foi diferente de casca e palha em relação aos limiares de detecção.

Talvez mais útil que a apresentação na forma de tabela seja observar gráficos destes intervalos de confiança, apresentados com as bandas de confiança na altura do limiar de detecção Figura 5 e Figura 6.

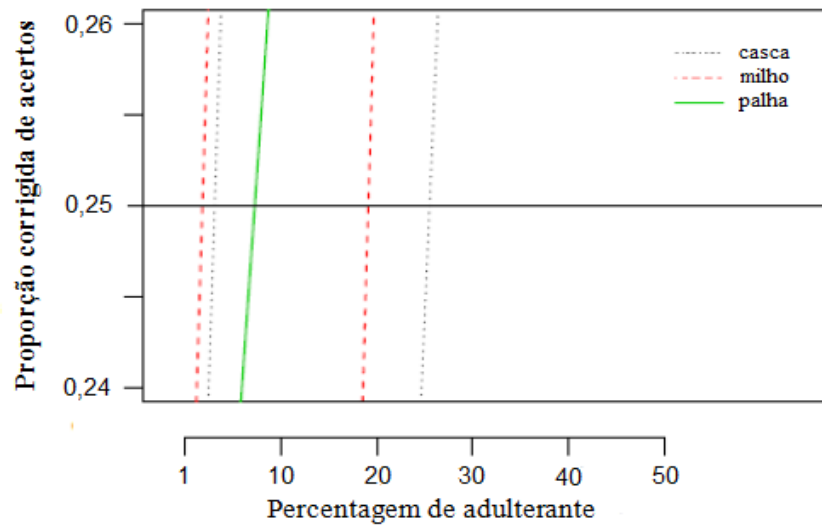


Figura 5 Representação gráfica da proporção de acertos e ajuste pela equação de regressão (quadrática) sob aproximação normal em função da porcentagem dos adulterantes para cada tipo de adulterante

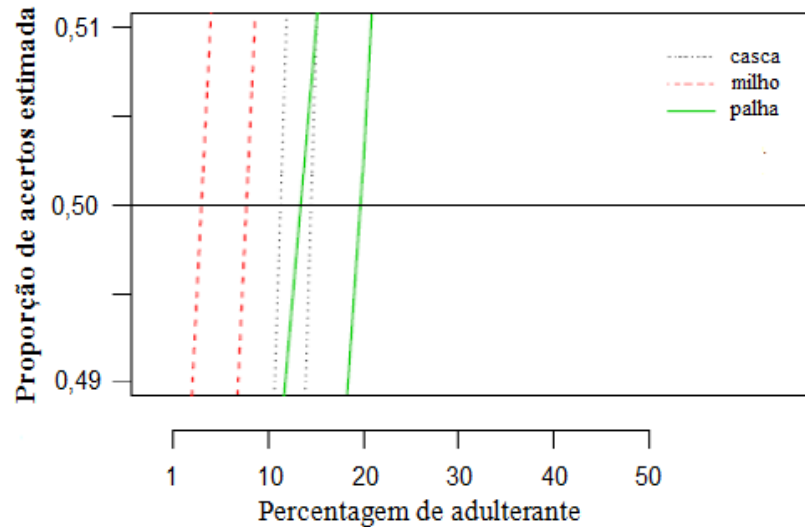


Figura 6 Representação gráfica da proporção de acertos e ajuste pela equação de regressão (linear) sob modelo binomial em função da percentagem dos adulterantes para cada tipo de adulterante

É importante notar no entanto que as bandas de confiança para o modelo binomial são mais estreitas que as do modelo normal. Como resultado, foi mais fácil discriminar os adulterantes quanto aos limiares de detecção, ou seja, embora ambas as curvas de regressão sejam bem ajustadas, a distribuição binomial foi mais eficiente, com maior poder de discriminação, especialmente quanto aos limiares de detecção.

Nesse estudo considerou-se o modelo fixo (que no caso de variáveis contínuas é justificado pela aleatorização, (HINKELMANN; KEMPTHORNE, 2007). No entanto, este modelo não leva em conta agregações de observações do mesmo provador (e tampouco potenciais correlações entre provadores). Pode-se estar superestimando a precisão das estimativas. Em estudos futuros, pretende-se considerar uma distribuição para os provadores, ou seja, considerar um modelo misto, com distribuição para as unidades experimentais e para um dos fatores. Consideramos que em ambas as formas de análise esta deve ser a solução mais adequada.

As rotinas de análise desenvolvidas no pacote estatístico R, embora não muito amigáveis, podem ser usadas em análises de rotina em laboratórios de análise sensorial de alimentos. Pretende-se também investigar mais formas de análise em maior número de experimentos (testes, por exemplo, em que são oferecidas cinco amostras e o provador deve identificar as duas diferentes).

5 CONCLUSÃO

A análise de testes triangulares considerando a distribuição binomial é mais precisa que a aproximação normal e pode ser utilizada na rotina de pesquisadores da área de ciência de alimentos.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12806**: análise sensorial de alimentos e bebidas. Terminologia. Rio de Janeiro, 1993.

BIEDRZYCKI, A. **Aplicação da avaliação sensorial no controle de qualidade em uma indústria de produtos cárneos**. 2009. 64 p. Monografia (Graduação em Engenharia de Alimentos)-Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/17622>>. Acesso em: 12 dez. 2012.

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of Royal Statistical Society. B.**, Oxford, v.26, n.2, p. 211-252, 1964. Disponível em: <<http://www.jstor.org/discover/10.2307/2984418?uid=2&uid=4&sid=21102092274751>>. Acesso em: 14 nov. 2012.

BROCKHOFF, P. B.; CHRISTENSEN, R. H. B. Thurstonian models for sensory discrimination tests as generalized linear models. **Food Quality and Preference**, Barking, v.21, n.3, p. 330-338, 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950329309000585>>. Acesso em: 12 nov. 2012.

CARR, B. T.; CIVILLE, G. V.; MEILGAARD, M. C. **Sensory evaluation techniques**. 4nd ed. Boca Raton: CRC Press, 2006. 448 p.

CHRISTENSEN, R. H. B. **Sensometrics**: Thurstonian and statistical model. Kongens Lyngby: DTU Informatics, 2012. 388 p. Disponível em: <http://orbit.dtu.dk/fedora/objects/orbit:111007/datastreams/file_b3b75800dc2a-4489ad98b2363833ec8a/content>. Acesso em: 20 jul. 2012.

HEYMANN, H; LAWLESS, H. T. **Sensory evaluation of food: principles and practices**. 2nd ed. New York: Springer, 2003. 587 p.

HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments: introduction to experimental design**. 2nd ed. New York: J. Wiley, 2007. v. 1, 631 p.

JORGENSEN, B. **The theory of linear models**. New York: Chapman & Hall, 1993. 226 p.

KEMP, S.E.; HOLLOWOOD, T.; HORT, J. **Sensory Evaluation: a practical handbook**. 1nd ed. London: J. Wiley, 2009. 196 p.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**, 2nd ed. London: Taylor & Francis, 1989. v.37, 511 p.

MORETTI, C. L.; SARGENT, S. A. Alteração de sabor e aroma em tomates causada por impacto. **Scientia agricola**. Piracicaba, v.57, n.3, p. 385-388, 2000. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162000000300002>. Acesso em: 08 jul. 2012.

NELDER, J.A.; WEDDERBURN, R. W. M. Generalized linear model. **Journal of the Royal Statistical Society , A.**, London, v.135, n.3. p. 370-384, 1972. Disponível em: <<http://www.jstor.org/discover/10.2307/2344614?uid=2&uid=4&sid=21102092741611>>. Acesso em: 24 jun. 2012.

PENG, M.; JAEGER, S. R.; HAUTUS, M. J. Determining odour detection thresholds: Incorporating a method-independent definition into the implementation of ASTM E679. **Food Quality and Preference**, Barking, v.25, n.2, p. 95-104, 2012.

ROESSLER, E. B.; PANGBORN, R. M.; SIDEL, J. L.; STONE, H. Expanded statistical tables for estimating significance in paired-preference, paired difference, duo-trio and triangle tests. **Journal of Food Science**, Chicago, v.43, n.3, p. 940-943, 1978.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. **Transactions of the American Mathematical Society**, Providence, v.54, n.3. p. 426-482, 1943.

WEISBERG, S. **Applied linear regression**. 2nd ed. New York: J. Wiley, 1985. 324 p.

APÊNDICE

Conjunto de dados advindos do teste triangular

Provador	Adulterante	Perc	Acertos	Erros
1	casca	1	0	3
1	casca	10	3	0
1	casca	20	3	0
1	casca	30	3	0
1	casca	40	2	1
1	casca	50	3	0
1	milho	1	1	2
1	milho	10	2	1
1	milho	20	2	1
1	milho	30	2	1
1	milho	40	3	0
1	milho	50	3	0
1	palha	10	1	2
1	palha	20	1	2
1	palha	30	0	2
1	palha	40	2	1
1	palha	50	1	2
2	casca	1	1	2
2	casca	10	1	2
2	casca	20	3	0
2	casca	30	3	0
2	casca	40	2	0
2	casca	50	3	0
2	milho	1	1	2
2	milho	10	1	2
2	milho	20	2	1
2	milho	30	3	0
2	milho	40	3	0
2	milho	50	3	0
2	palha	1	1	0
2	palha	10	2	0
2	palha	20	2	0
2	palha	30	1	1
2	palha	40	2	0
2	palha	50	3	0

3	casca	1	1	2
3	casca	10	3	0
3	casca	20	3	0
3	casca	30	3	0
3	casca	40	3	0
3	casca	50	3	0
3	milho	1	1	2
3	milho	10	1	0
3	milho	20	2	0
3	milho	30	3	0
3	milho	40	3	0
3	milho	50	3	0
3	palha	1	1	2
3	palha	10	1	2
3	palha	20	1	2
3	palha	30	3	0
3	palha	40	2	1
3	palha	50	2	1
4	casca	1	0	3
4	casca	10	2	1
4	casca	20	3	0
4	casca	30	3	0
4	casca	40	3	0
4	casca	50	3	0
4	milho	1	1	2
4	milho	10	3	0
4	milho	20	1	0
4	milho	30	3	0
4	milho	40	3	0
4	milho	50	1	0
4	palha	1	1	2
4	palha	10	3	0
4	palha	20	3	0
4	palha	30	3	0
4	palha	40	3	0
4	palha	50	3	0
5	casca	1	0	3
5	casca	10	2	1
5	casca	20	3	0

5	casca	30	3	0
5	casca	40	2	1
5	casca	50	3	0
5	milho	1	1	2
5	milho	10	3	0
5	milho	20	3	0
5	milho	30	2	0
5	milho	40	3	0
5	milho	50	2	0
5	palha	1	2	1
5	palha	10	1	2
5	palha	20	1	2
5	palha	30	2	1
5	palha	40	1	2
5	palha	50	3	0
6	casca	1	1	1
6	casca	10	1	1
6	casca	20	2	1
6	casca	30	2	0
6	casca	40	3	0
6	casca	50	1	0
6	milho	1	1	2
6	milho	30	3	0
6	milho	40	1	1
6	milho	50	1	0
6	palha	1	0	3
6	palha	10	2	0
6	palha	20	2	0
6	palha	30	2	0
6	palha	40	1	1
6	palha	50	2	0
7	casca	1	1	1
7	casca	10	2	1
7	casca	20	1	1
7	casca	30	1	2
7	casca	40	2	0
7	casca	50	2	1
7	milho	1	0	3
7	milho	10	2	1

7	milho	20	0	2
7	milho	30	2	1
7	milho	40	3	0
7	milho	50	1	1
7	palha	1	1	2
7	palha	10	3	0
7	palha	20	3	0
7	palha	30	1	2
7	palha	40	2	1
7	palha	50	3	0
8	casca	1	1	2
8	casca	10	1	1
8	casca	20	3	0
8	casca	30	1	1
8	casca	40	2	0
8	casca	50	2	0
8	milho	1	0	2
8	milho	10	2	1
8	milho	20	3	0
8	milho	30	2	0
8	milho	40	2	0
8	milho	50	2	0
8	palha	1	2	0
8	palha	10	0	2
8	palha	20	0	2
8	palha	30	2	1
8	palha	40	2	0
8	palha	50	3	0
9	casca	1	0	1
9	casca	10	0	2
9	casca	30	1	1
9	casca	40	1	0
9	casca	50	2	0
9	milho	1	2	1
9	milho	20	2	0
9	milho	30	2	0
9	milho	40	2	0
9	milho	50	3	0
9	palha	1	2	1

9	palha	10	1	0
9	palha	20	1	0
9	palha	30	2	1
9	palha	40	2	0
9	palha	50	1	0
10	milho	10	1	2
10	milho	20	1	2
10	milho	40	1	0
10	milho	50	2	0
10	palha	1	2	1
10	palha	10	1	2
10	palha	20	1	2
10	palha	30	2	0
10	palha	40	3	0
10	palha	50	1	1
11	casca	1	2	1
11	casca	10	1	1
11	casca	20	2	1
11	casca	30	2	0
11	casca	40	3	0
11	casca	50	3	0
11	milho	1	1	2
11	milho	10	2	1
11	milho	20	2	1
11	milho	30	3	0
11	milho	40	2	1
11	milho	50	2	1
11	palha	1	0	3
11	palha	10	1	2
11	palha	20	1	2
11	palha	30	2	1
11	palha	40	1	2
11	palha	50	3	0
12	casca	1	2	1
12	casca	10	0	2
12	casca	20	2	1
12	casca	30	3	0
12	casca	40	2	0
12	casca	50	3	0

12	milho	1	1	1
12	milho	10	2	1
12	milho	20	2	1
12	milho	30	2	0
12	milho	40	3	0
12	milho	50	2	0
12	palha	1	2	1
12	palha	10	2	1
12	palha	20	2	1
12	palha	30	0	2
12	palha	40	3	0
12	palha	50	3	0
13	casca	1	1	1
13	casca	10	0	1
13	casca	20	2	1
13	casca	30	0	1
13	casca	40	1	0
13	milho	1	0	2
13	milho	10	1	1
13	milho	20	2	1
13	milho	30	2	0
13	milho	40	1	0
13	milho	50	3	0
13	palha	1	1	2
13	palha	10	1	1
13	palha	20	0	2
13	palha	30	2	0
13	palha	40	3	0
13	palha	50	1	0
14	casca	1	1	1
14	casca	10	1	0
14	casca	40	2	0
14	casca	50	3	0
14	milho	1	1	2
14	milho	10	2	1
14	milho	20	3	0
14	milho	30	1	0
14	milho	40	2	1
14	milho	50	1	0

14	palha	1	1	2
14	palha	10	0	2
14	palha	20	0	2
14	palha	30	1	1
14	palha	40	0	1
14	palha	50	2	1
15	casca	1	2	1
15	casca	10	0	3
15	casca	20	1	2
15	casca	30	2	1
15	casca	40	0	3
15	casca	50	3	0
15	milho	1	1	2
15	milho	10	2	1
15	milho	20	3	0
15	milho	30	2	1
15	milho	40	3	0
15	milho	50	3	0
15	palha	1	1	2
15	palha	10	2	1
15	palha	20	2	1
15	palha	30	1	1
15	palha	40	1	1
15	palha	50	2	1
16	casca	1	1	1
16	casca	10	1	2
16	casca	20	0	1
16	casca	30	3	0
16	casca	40	3	0
16	casca	50	3	0
16	milho	1	0	1
16	milho	10	3	0
16	milho	20	3	0
16	milho	30	2	0
16	milho	40	3	0
16	milho	50	2	0
16	palha	1	1	1
16	palha	10	3	0
16	palha	20	3	0

16	palha	30	1	2
16	palha	40	1	2
16	palha	50	1	0
17	casca	1	1	1
17	casca	10	1	2
17	casca	20	3	0
17	casca	30	2	1
17	casca	40	2	1
17	casca	50	0	1
17	milho	10	1	1
17	milho	30	0	1
17	palha	1	1	0
17	palha	10	0	1
17	palha	20	0	1
17	palha	30	1	0
17	palha	40	0	1
17	palha	50	0	1
18	casca	1	0	2
18	casca	10	0	1
18	casca	20	2	0
18	casca	30	1	0
18	casca	40	2	0
18	casca	50	2	0
18	milho	1	0	1
18	milho	10	3	0
18	milho	20	1	1
18	milho	30	3	0
18	milho	40	1	1
18	milho	50	3	0
18	palha	1	1	2
18	palha	10	0	3
18	palha	20	0	3
18	palha	30	1	1
18	palha	40	1	1
18	palha	50	1	1
19	casca	1	0	1
19	casca	20	2	0
19	casca	30	0	1
19	casca	40	1	0

19	casca	50	2	0
19	milho	1	1	0
19	milho	10	1	2
19	milho	20	0	1
19	milho	30	2	1
19	milho	40	1	0
19	milho	50	3	0
19	palha	1	1	2
19	palha	10	0	3
19	palha	20	0	3
19	palha	30	2	0
19	palha	40	1	1
19	palha	50	0	1
20	casca	1	0	3
20	casca	10	2	1
20	casca	20	2	1
20	casca	30	3	0
20	casca	40	3	0
20	casca	50	3	0
20	milho	1	1	2
20	milho	10	2	1
20	milho	20	3	0
20	milho	30	1	2
20	milho	40	3	0
20	milho	50	3	0
20	palha	1	1	2
20	palha	30	2	0
20	palha	40	2	0
20	palha	50	2	0

Comandos utilizados no programa R considerando aproximação normal

```
rm(list=ls(all=T))
```

Contaminantes de cafe:

```
dados <- read.table("KatianeCafe.csv",header=TRUE)
attach(dados)
str(dados)
```

```
Provador <- factor(Provador)
Porcentagem <- factor(Perc)
PAC <- Acertos/(Acertos+Erros)-1/3
```

```
library(MASS)
```

```
PACc <- -min(PAC)+PAC+0.01
par(mar=c(2,2,2,2))
y<- PACc~Provador+Adulterante*Porcentagem
box <- boxcox(y, ylab="log-verossimilhança")
```

```
library(stats)
lambda <- box$lambda[which(box$lambda == max(box$lambda))]
```

Modelo completo:

```
modelo.c <- lm(PAC ~ Adulterante*Porcentagem + Provador)
X <- model.matrix(~ Adulterante*Porcentagem + Provador)
anova(modelo.c)
summary(modelo.c)
coef(modelo.c)
```

Modelo regressao:

```
modelo.r <- lm(PAC ~ Adulterante:(I(Perc)+I(Perc^2))+Provador)
Xr <- model.matrix(~ Adulterante:(I(Perc)+I(Perc^2)) + Provador)
solve(t(Xr)%*%(Xr))
anova(modelo.r)
summary(modelo.r)
coef(modelo.r)
```

Comparando modelo completo com regressao:

```
anova(modelo.c,modelo.r)
p <- predict(modelo.r, interval="confidence")
pr <- predict(modelo.r)
b <- solve(t(Xr)%*%Xr)%*%(t(Xr)%*%pr)
media <- p[,1]
LI <- p[,2]
LS <- p[,3]
```

Recalculando e imprimindo os limiares de percepcao:

```
theta <- coef(modelo.r)[21:26]
```

Coeficientes:

```
$delta <- theta[1:3]^2+theta[4:6]$
sdelta <- sqrt(delta)
```

Solucao pontual para os limiares:

```
$sol <- as.real(-theta[1:3]+sdelta)/(2*theta[4:6])$
```

Testando a igualdade entre os limiares de percepcao:

```
par(mfrow=c(1,1))
interaction.plot(Perc,Adulterante,LI,col=1:3,ylim=c(0,1),lwd=1,
                 fixed=TRUE,xtick=TRUE,
                 xlab="Percentagem de adulterante (\%)",
                 ylab="Proporcao corrigida de acertos")
par(new=TRUE)
interaction.plot(Perc,Adulterante,LS,col=1:3,ylim=c(0,1),lwd=1,
                 fixed=TRUE,xtick=TRUE,
                 xlab="Percentagem do adulterante (\%)",
                 ylab="Proporcao corrigida de acertos")
abline(h=.25)
```

Curvas de regressao:

```
interaction.plot(Perc,Adulterante,fitted(modelo.c),col=1:3,
                ylim=c(0,0.6),lwd=1,fixed=TRUE,
                xlab="Percentagem de adulterante (%)",
                ylab="Proporçao corrigida de acertos")
par(new=TRUE)
interaction.plot(Perc,Adulterante,media,col=1:3,ylim=c(0,0.6),
                lwd=4,fixed=TRUE,
                xlab="Percentagem de adulterante (%)",
                ylab="Proporçao corrigida de acertos")
abline(h=.25)
```

Comandos utilizados no programa R considerando distribuição binomial

```

rm(list=ls(all=T))
library(MASS)
source("funcoesCamila.R")

##### Contaminantes de cafe #####

dados <- read.table("KatianeCafe.csv",header=TRUE)
attach(dados)
str(dados)
Bloc <- factor(Provador)
A <- factor(Adulterante)
P <- factor(Perc)

Xb <- model.matrix(~ -1+Bloc)
Xt <- model.matrix(~ -1+A:P)
Xr <- model.matrix(~ -1+A/(I(Perc)))
Xr2 <- model.matrix(~ -1+A/(I(Perc)+I(Perc^2)))

#Acertos/(Acertos+Erros)
Media <- predict(lm(Acertos/(Acertos+Erros)~A:P+Bloc))
m <- Acertos+Erros
y <- Acertos
#mean(y/m)
taut <- 0*(1:dim(Xt)[2])
taur <- 0*(1:(dim(Xr)[2]))
taur2 <- 0*(1:(dim(Xr2)[2]))
beta <-0*(1:dim(Xb)[2])
thetat <- c(beta,taut)
thetar <- c(beta,taur)
thetar2 <- c(beta,taur2)

Xmt <- as.matrix(cbind(Xb,Xt))
Xmr <- as.matrix(cbind(Xb,Xr))
Xmr2 <- as.matrix(cbind(Xb,Xr2))

X <- Xmt

```

```

theta <- thetat
saidat <- NR.Camila(X,y,m,theta)
thetat <- saidat$theta
eta <- as.real(X%*%thetat)
pival <- (1+2/(1+exp(-eta)))/3
lVt <- sum(y*log(pival) + (m-y)*log(1-pival))

```

```

X <- Xmr
theta <- thetar
saidar <- NR.Camila(X,y,m,theta)
thetar <- saidar$theta

```

```

intBlocos <- mean(thetar[1:20])
(log(1/3)-thetar[21:23])/thetar[24:26]
eta <- as.real(X%*%thetar)
pival <- (1+2/(1+exp(-eta)))/3
lVr <- sum(y*log(pival) + (m-y)*log(1-pival))

```

```

X <- Xmr2
theta <- thetar2
saidar2 <- NR.Camila(X,y,m,theta)
thetar2 <- saidar2$theta
eta <- as.real(X%*%thetar2)
pival <- (1+2/(1+exp(-eta)))/3
lVr2 <- sum(y*log(pival) + (m-y)*log(1-pival))

```

```
Lambda <- 2*(lVt-lVr)
```

```

#lamb<- -2*log(lVr/lVt)
#lamb

```

```

glt <- dim(Xmt) [2]
glr <- dim(Xmr) [2]
glr2 <- dim(Xmr2) [2]

```

```

Diferenca_veros <- c(lVt-lVr,lVr2-lVr)
GL <- c(glt-glr,glr2-glr)

pval <- 1-pchisq(Diferenca_veros,GL)

cbind(Diferenca_veros,GL,pval)

thetar[21:26]

Cov <- -saidar$Hi
Cov[21:26,21:26]

cbind(thetar[21:26],sqrt(diag(Cov[21:26,21:26])),
      thetar[21:26]/sqrt(diag(Cov[21:26,21:26])))

IC <- predito(Xr,thetar[21:26],Cov[21:26,21:26])

media <- IC$pim
piLI <- IC$piLI
piLS <- IC$piLS

cbind(media,piLI,piLS)

# Curvas de regressao

interaction.plot(Perc,A,Media,fixed=TRUE,ylim=c(0.3,0.9),
                xtick=TRUE,
                xlab="Percentagem do adúlterante (%)",
                ylab="Proporção de acertos estimada",col=1:3)

par(new=TRUE)

interaction.plot(Perc,A,media,fixed=TRUE,ylim=c(0.3,0.9),
                xtick=TRUE,lwd=4,
                xlab="Percentagem do adúlterante (%)",
                ylab="Proporção de acertos estimada",col=1:3)

abline(h=.5)

```

```

# Limiares de deteccao

interaction.plot(Perc,A,piLI, fixed=TRUE, ylim=c(0,1), xtick=TRUE,
                xlab="Percentagem de adulterante (%)",
                ylab="Proporçao de acertos estimada", col=1:3)

par(new=TRUE)
interaction.plot(Perc,A,piLS, fixed=TRUE, ylim=c(0,1), xtick=TRUE,
                xlab="Percentagem de adulterante (%)",
                ylab="Proporçao de acertos estimada", col=1:3)

abline(h=.5)

# function to calculate pi
modelo <- function(theta){
  eta <- as.real(X%%theta)
  picalc <- (1+2/(1+exp(-eta)))/3
  return(picalc)
}

# funcao para calcular IC limiares

predito <- function(X,theta,Cov){
  Xtheta <- as.real(X%%theta)
  Vd <- diag(X %% Cov %% t(X))
  LI <- Xtheta-1.96*Vd
  LS <- Xtheta+1.96*Vd
  pim <- 1/3+(2/3)/(1+exp(-Xtheta))
  piLI <- 1/3+(2/3)/(1+exp(-LI))
  piLS <- 1/3+(2/3)/(1+exp(-LS))
  return(list(pim=pim,piLI=piLI,piLS=piLS))
}

# funcao para calcular Newton Raphson
NR.Camila <- function(X,y,m,theta){
  theta.old <- theta
  crit <- 1
  cont <- 0

```

```

while(crit > 0.0001){
  theta.old <- theta
  pic <- modelo(theta)
  pim <- (3*pic-1)*(y/pic-m)/2
  # Gradiente
  grad <- t(X)%*%pim
  # Hessiana
  V <- diag(as.vector((1-pic)*(3*pic-1)*(y-3*m*pic^2)*(1/(4*pic^2))))
  H <- t(X)%*%V%*%X
  Hi <- ginv(H)
  theta <- theta.old - as.real(Hi %*% grad)
  crit <- as.real(t(theta-theta.old)%*%(theta-theta.old))
  cont <- cont+1
}
return(list(theta=theta, Hi=Hi))
}

# funcao para calcular log-verossimilhanca
lVero <- function(X,y,m,theta){
  eta <- as.real(X%*%theta)
  pival <- (1+2/(1+exp(-eta)))/3
  lVero <- sum(y*log(pival) + (m-y)*log(1-pival))
  return(lVero=lVero)
}

```