



ADILSON SILVA DA COSTA

**SIMILARIDADE DE FORMULAÇÕES DE BLENDS DE
CAFÉS DE DIFERENTES QUALIDADES UTILIZANDO A
ANÁLISE DE CORRESPONDÊNCIA GENERALIZADA**

**LAVRAS – MG
2018**

ADILSON SILVA DA COSTA

**SIMILARIDADE DE FORMULAÇÕES DE BLENDS DE CAFÉS DE DIFERENTES
QUALIDADES UTILIZANDO A ANÁLISE DE CORRESPONDÊNCIA
GENERALIZADA**

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 a obtenção do título de Mestre.

Prof. Dr. Marcelo Ângelo Cirillo
Orientador

**LAVRAS-MG
2018**

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

Costa, Adilson Silva da.

Similaridade de formulações de blends de cafés de diferentes
qualidades utilizando a análise de correspondência generalizada /
Adilson Silva da Costa. - 2018.

59 p. : il.

Orientador: Marcelo Ângelo Cirillo.

Dissertação (Mestrado acadêmico) - Universidade Federal de
Lavras, 2018.

Bibliografia.

1. Blends. 2. Cafés especiais. 3. Índice de gravidade. I. Cirillo,
Marcelo Ângelo. II. Título

ADILSON SILVA DA COSTA

**SIMILARIDADE DE FORMULAÇÕES DE BLENDS DE CAFÉS DE DIFERENTES
QUALIDADES UTILIZANDO A ANÁLISE DE CORRESPONDÊNCIA
GENERALIZADA**

**SIMILARITY OF COFFEE BLENDS FORMULATIONS OF DIFFERENT
QUALITIES USING CORRESPONDENCE ANALYSIS**

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 a obtenção do título de Mestre.

APROVADA em 31 de agosto de 2018.

Profa. Dra. Carla Regina Guimarães Brighenti	UFSJ-MG
Profa. Dra. Evelise Roman Corbalan Gois Freire	UFLA-MG
Profa. Dra. Patrícia de Siqueira Ramos	UNIFAL-MG

Prof. Dr. Marcelo Ângelo Cirillo
Orientador

**LAVRAS-MG
2018**

Aos meus pais, João Teodoro da Costa (*in memoriam*) e Maria Belmira Silva Costa.

DEDICO.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me iluminado neste caminho cheio de espinhos e obstáculos.

Aos meus pais, João Teodoro da Costa (*in memoriam*) e Maria Belmira Silva Costa, que sempre me ensinaram o valor do conhecimento, apesar de suas baixas escolaridades, pelo apoio durante toda a minha vida e pelo esforço para garantir este momento.

Aos meus familiares por me darem força, sabendo o quanto é difícil a carreira acadêmica.

À minha prima Natali James pela amizade, risadas, companheirismo, por ouvir minhas lamentações, em momentos de desespero e por sempre me apoiar com palavras positivas e motivando-me sempre.

A todos os professores da Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA que me ajudaram, de alguma forma, no desenvolvimento deste trabalho.

À Universidade Federal de Lavras e ao Departamento de Estatística (DES) e a seus funcionários.

Ao CNPQ pela concessão da bolsa de mestrado, para a realização deste trabalho, em especial, ao meu orientador Marcelo Angelo Cirillo que me orientou nesta dissertação com paciência, disposição, compreensão e respeitando as minhas limitações.

Em especial, ao amigo Edimar Fornazari Garcia que sempre me ajudou quando precisei.

Aos colegas da UFLA que não hesitaram em me ajudar em algum momento de dificuldade, Joaquim Francisco Mazunga, Allana Lívia, Paulo Cesar Ossani, Claudio Chipenete e Lourenço Manuel.

A todos os amigos que, apesar de não serem aqui citados, contribuíram direta ou indiretamente com este trabalho, meu muito obrigado.

RESUMO

Dada a elevada demanda do mercado consumidor, cada vez mais exigente, em relação a padrões de qualidade e certificações, as atividades inerentes à industrialização de cafés e seus derivados destacam-se na formulação de *blends* com cafés de diferentes variedades e qualidades. A questão é que cada variedade apresenta particularidades quanto à composição química e fisiológica, temperatura e forma de processamento que são fatores que caracterizam os atributos sensoriais, sabor, acidez e amargor. Portanto torna-se complexa avaliar a qualidade sensorial dos *blends* em relação a esses atributos. Dada essa motivação, este trabalho tem por objetivo propor uma metodologia que permita estudar a similaridade entre *blends* de cafés formados pela composição de cafés de grão arábica, canéfora e comercial. A metodologia proposta é dada, em função da técnica de análise de correspondência generalizada, a qual permite avaliar diferentes métricas, além da proposição de um índice de gravidade que expresse o risco de um avaliador atribuir um nota errada em um avaliação sensorial. Concluiu-se que a métrica de Hellinger apresentou resultados, relacionados aos índices de gravidade, superiores à métrica Qui-Quadrado que justifica avaliar o erro de mensuração, supostamente a ser cometido pelo avaliador.

Palavras-chave: Blends. Cafés especiais. Índice de gravidade.

ABSTRACT

Due to a highly demanding consuming market with respect to quality standards and certificates, the formulation of blends with different varieties and qualities of coffee stands out among the activities related to coffee processing and industrialization. Each coffee variety is particular regarding its chemical composition, physiology, temperature, and processing procedures, which are factors that characterize sensory attributes such as flavor, acidity, and bitterness. Therefore, assessing the sensory quality of blends considering these attributes is complex. This study aimed at proposing a methodology to study the similarities among coffee blends from arabica, canephora, and commercial grains. The methodology is based on Correspondence Analysis, which allows the evaluation of different measurements, and proposes a gravity index that determines the risk of an evaluator to provide wrong scores at sensory analyses. Regarding gravity indices, the Hellinger distance presented higher results than the Chi-squared test, which justifies the evaluation of measurement errors likely made by the evaluator.

Keywords: Blends. Specialty coffees. Gravity indexes.

SUMÁRIO

1	INTRODUÇÃO	9
2	REFERENCIAL TEÓRICO	11
2.1	Aspectos gerais da análise de correspondência	11
2.2	Operações matriciais e definições intrínsecas à análise de correspondência simples	12
2.3	A distância de Hellinger e a relação com a análise de correspondência simples	17
2.4	Escore de Risco	20
2.4.1	Construção do Índice para dois grupos (k=2)	21
2.4.2	Construção do índice para três grupos (k = 3)	22
2.4.3	Construção do índice para k grupos (k >3)	24
2.5	Blends de café e os principais aspectos na qualidade e formulação	25
3	METODOLOGIA	28
3.1	Descrição dos experimentos e obtenção das medidas da análise de correspondência generalizada	28
3.2	Simulação de tabelas de contingência com dependência entre os níveis categóricos	31
4	RESULTADOS E DISCUSSÕES	33
4.1	Estudo preliminar da explicação da inércia na análise de correspondência utilizando as distâncias Qui-Quadrado e Hellinger	33
4.2	Estudos de simulação do comportamento das métricas de Qui-Quadrado e Hellinger	34
4.3	Aplicação da análise de correspondência utilizando as distâncias Qui-Quadrado e Hellinger	36
4.4	Índice de gravidade quanto à definição da escala a ser interpretada nos mapas perceptuais	42
5	CONCLUSÕES	46
	REFERÊNCIAS	47
	ANEXO A – DECOMPOSIÇÃO DE VALORES SINGULARES ENVOLVIDA NA RESTITUIÇÃO DA VARIAÇÃO AMOSTRAL EXPLICADA PELOS EIXOS	52
	ANEXO B - DEMONSTRAÇÃO DA LIMITAÇÃO DA DISTÂNCIA DE HELLINGER POR UMA PROPRIEDADE QUE DEFINE SUA AMPLITUDE	54
	ANEXO C - TABELAS DE FREQUÊNCIA UTILIZADA NA REALIZAÇÃO DAS ANÁLISES DE CORRESPONDÊNCIA	56
	ANEXO D – ROTINA UTILIZADA NO SOFTWARE R PARA A OBTENÇÃO DOS RESULTADOS DE SIMULAÇÃO	57

1 INTRODUÇÃO

Segundo relatam Costa, Guilhoto e Imori (2013), no que concerne à importância dos setores agroindustriais, dada a posição do Brasil na produção e comercialização de cafés no cenário mundial, nota-se que a cafeicultura apresenta uma importante contribuição econômica e social.

Nesse contexto, com uma certa relevância, destaca-se o agronegócio café, no qual, dentre as inúmeras espécies exploradas comercialmente, no Brasil, inserem-se os cafés *Coffea arabica* e *Coffea canephora*, respectivamente, representando, aproximadamente, 75% e 25% da produção nacional (RIBEIRO et al., 2014).

No tocante à formulação dos *blends* entre os cafés arábica e canéfora, ainda, persistem críticas quanto à composição desses cafés, decorrentes de alterações causadas na composição química. Entretanto, em virtude da constante oscilação de preços na cafeicultura, o uso desse café em *blends* com arábica tem como maior justificativa a redução de custos, uma vez que o canéfora é mais barato que os grãos arábica (MENDES, 2005). Ivoglo et al. (2008) observaram que a elaboração de um *blend* proporciona ao produto final uma expressiva capacidade de competição no mercado, tendo em vista o maior rendimento industrial, bem como os menores preços médios em sua comercialização.

Decorrente da aplicabilidade na indústria de torração, para que a qualidade dos *blends* seja aprimorada, outros fatores como, por exemplo, o método de processamento a ser utilizado deverá ser considerado, haja vista que os compostos voláteis que conferem o aroma e o sabor da bebida do café são influenciados e, conseqüentemente, a qualidade da bebida é afetada (BORÉM et al., 2008; FRANCA; MENDONÇA; OLIVEIRA, 2005; MAARSE; VISSCHER, 1996).

Diante da possibilidade da formação de inúmeros *blends*, a partir das misturas dos cafés das espécies arábica e canéfora, em diferentes concentrações a serem realizadas pelas torrefadoras, uma diversidade de cafés poderão ser gerados em novos produtos. A diferenciação desses produtos envolve um dos programas da ABIC (Associação Brasileira da Indústria do Café), com o lançamento de certificações, como parte do Programa de Qualidade do Café (PQC), com o propósito de diferenciar os produtos, no que tange à consistência do processo de fabricação, com vista a ampliar o consumo de café. Assim, o setor produtivo e o mercado estabelecem uma relação comercial transparente, pautada na livre escolha por parte do consumidor, na agregação de valores e diferenciação de preços em função da qualidade do produto.

Tendo por base essa motivação, a agregação de valores à qualidade e à produção de *blends* dar-se-á na proposta do aprimoramento e/ou de novas metodologias estatísticas que sejam mais informativas, nhoque diz respeito à adição correta e precisa da proporção dos cafés arábica e canéfora, na formação de *blends*, relacionando com a percepção sensorial dos consumidores, haja vista que boa parte da população brasileira não tem conhecimento específico e técnico da formulação de *blends* de cafés especiais com canéfora.

Desta forma, em síntese, este trabalho teve por objetivo estudar os principais aspectos sensoriais, em relação à formulação de *blends* de cafés de diferentes qualidades, considerando dados categorizados ordinais e diferentes métricas Qui-Quadrado e Hellinger a serem incorporadas na análise de correspondência, em conjunto com a formulação de índices de gravidades, que agreguem informações para detectar similaridades entre os *blends* com diferentes composições.

2 REFERENCIAL TEÓRICO

2.1 Aspectos gerais da análise de correspondência

A análise de correspondência é uma técnica multivariada exploratória, desenvolvida para analisar dados categóricos e sua principal vantagem consiste em reproduzir mapas, nos quais a similaridade entre os níveis das variáveis categóricas poderá ser identificada. Com esse propósito, consideram-se os dados estruturados em tabelas de contingência de dupla e/ou múltiplas entradas, originando, respectivamente, a análise de correspondência simples e múltipla (BENZÉCRI, 1992).

Ressalta-se que, nessa última representação, a tabela de contingência é obtida por uso de variáveis indicadoras e o produto cruzado entre elas origina a tabela BURT (BENZÉCRI, 1992), equivalente a uma tabela de frequências. Desta forma, em ambas as situações, toda a álgebra é dada, em função do número I de linhas e J colunas, de modo que o número máximo de dimensões, interpretados como componentes, nos quais os mapas poderão assumir, é da ordem de $T = \min(I, J) - 1$ (GREENACRE; BLASIUS, 2006). Dentre os resultados, o principal a ser destacado são as coordenadas, vistas como escores, e medidas corroborativas à qualidade de representação gráfica, tais como, correlação e contribuição de variável em relação aos eixos.

Quanto à variabilidade amostral, a dispersão dos pontos referente ao centroide é mensurada pela inércia e sua dispersão nos escores. Segundo menciona Greenacre (2007), a principal forma a ser distribuída consiste em um procedimento simétrico, isto é, a inércia é distribuída de forma equilibrada pelas coordenadas dos níveis das variáveis representadas no sentido das “linhas” e níveis da outra variável, representada no sentido das “colunas”. Para isso, parte-se do pressuposto de que os dados sejam corrigidos pela distância Qui-Quadrado, a ser vista com maiores detalhes na Seção 2.3.

Uma forma alternativa, para a obtenção de novas coordenadas, pode ser encontrada por diferentes métricas. Assim, Cuadras, Cuadras e Greenacre (2006) descreveram um procedimento, no qual se considera a decomposição da inércia por meio da substituição da distância Qui-Quadrado pela distância de Hellinger. Segundo os autores, a adoção dessa nova distância preserva a principal característica da análise de correspondência tradicional, caracterizada pelo princípio da equivalência distribucional.

Segundo menciona Costa (2016), entende-se que, ao considerar a unificação de duas linhas que apresentem os mesmo perfis, ao utilizar a distância qui-quadrado, a unificação dos

perfis semelhantes poderá ser substituída por um única linha que represente a soma desses perfis, sem, contudo, deturpar as distâncias entre as colunas da tabela, ou seja, as distâncias são preservadas mesmo quando os perfis são unificados.

Mediante o exposto, procede-se à descrição das operações matriciais e à definição dos principais conceitos, a serem utilizadas na análise de correspondência simples, a partir dos dados organizados em uma tabela de contingência.

2.2 Operações matriciais e definições intrínsecas à análise de correspondência simples

Na literatura concernente à análise de correspondência, nota-se que essa técnica apresenta uma linguagem bastante peculiar. Tais definições e conceitos são derivados da física e, para que sejam compreendidos, necessariamente, exige-se que os dados sejam estruturados em uma tabela de contingência. Logo, seja um arranjo de uma tabela de contingência, considerando duas variáveis categóricas A e B, respectivamente, com I e J categorias, conforme ilustrado na Tabela 1.

Tabela 1 - Exemplo de uma tabela de contingência com I linhas e J colunas.

		B					
A	1	2	...	j	...	J	Total
1	a_{11}	a_{12}	...	a_{1j}	...	a_{1J}	a_{1+}
2	a_{21}	a_{22}	...	a_{2j}	...	a_{2J}	a_{2+}
⋮	⋮	⋮		⋮		⋮	⋮
I	a_{i1}	a_{i2}		a_{ij}	...	a_{iJ}	a_{i+}
⋮	⋮	⋮	...	⋮	...	⋮	⋮
I	a_{I1}	a_{I2}		a_{Ij}	...	a_{IJ}	a_{I+}
Total	a_{+1}	a_{+2}	...	a_{+j}	...	a_{+J}	a_{++}

Fonte: Dados do autor (2018)

Dadas as linhas I e J colunas, resultantes da classificação das n observações de uma amostra ou população, a notação descrita na Tabela 1 resume que a_{ij} corresponde à frequência observada na i-ésima categoria da variável A e j-ésima categoria da variável B; a_{i+} refere-se aos totais marginais das linhas da categoria A, obtidos por

$$a_{i+} = \sum_{j=1}^J a_{ij}, i = 1, 2, \dots, I; \quad (1)$$

a_{+j} representa os totais marginais referentes aos níveis categóricos da variável B, obtidos conforme a expressão (2) e, por fim, naturalmente, o total geral dado em (3).

$$a_{+j} = \sum_{i=1}^I a_{ij}, j = 1, 2, \dots, J \quad (2)$$

$$a_{++} = \sum_{i=1}^I \sum_{j=1}^J a_{ij}, i = 1, 2, \dots, I ; j = 1, 2, \dots, J \quad (3)$$

Tendo por base a expressões (1), (2) e (3) mencionadas anteriormente, torna-se possível obter a matriz de correspondência, na qual cada célula é transformada em uma proporção (4), assim, temos então p_{ij} correspondente a cada elemento a_{ij} , conforme encontra-se descrito na Tabela 2.

$$p_{ij} = \frac{a_{ij}}{a_{++}}, i=1, \dots, I; j=1, \dots, J. \quad (4)$$

Tabela 2 - Matriz de correspondência.

		B					
A	1	2	...	j	...	J	Total
1	p_{11}	p_{12}	...	p_{1j}	...	p_{1J}	p_{1+}
2	p_{21}	p_{22}	...	p_{2j}	...	p_{2J}	p_{2+}
⋮	⋮	⋮	...	⋮	...	⋮	⋮
i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{iJ}	p_{i+}
⋮	⋮	⋮	...	⋮	...	⋮	⋮
I	p_{I1}	p_{I2}	...	p_{Ij}	...	p_{IJ}	p_{I+}
Total	p_{+1}	p_{+2}	...	p_{+j}	...	p_{+J}	$p_{++}=1$

Fonte: Dados do autor (2018)

Os totais de cada linha, referentes às categorias da variável A e a cada coluna, referentes às categorias da variável B, são dados pelas expressões:

$$P_{i+} = \sum_{j=1}^J P_{ij} \quad (5)$$

$$P_{+j} = \sum_{i=1}^I P_{ij} \quad (6)$$

As proporções marginais dadas pelas expressões (5) e (6) são organizadas em vetores e, dada a nomenclatura da análise de correspondência, são denominados perfil médio de linha (7) e coluna (8), ou centroide de linha e coluna, dados por \tilde{r} e \tilde{c} , respectivamente:

$$\tilde{r} = \tilde{p}_i = (p_{1+}, p_{2+}, \dots, p_{i+}) = \left(\frac{a_{1+}}{a_{++}}, \frac{a_{2+}}{a_{++}}, \dots, \frac{a_{i+}}{a_{++}} \right) \quad (7)$$

$$\tilde{c} = p_j = (p_{+1}, p_{+2}, \dots, p_{+j}) = \left(\frac{a_{+1}}{a_{++}}, \frac{a_{+2}}{a_{++}}, \dots, \frac{a_{+j}}{a_{++}} \right) \quad (8)$$

Os elementos de um centroide refletem a importância relativa de uma determinada categoria, em relação às demais e, uma vez que são calculados em relação ao total da tabela, são chamados de massas. Portanto, caso alguma variável, associada a algum componente desses vetores, tenha massas (proporções) maiores, em relação às demais, o centroide é influenciado por essa variável e tal fato resultará em mapas perceptuais assimétricos.

Outra forma de organizar os vetores, sendo nomeados como perfis “linha” e “coluna”, são descritos nas expressões (9) e (10), exemplificados para a i -ésima linha e j -ésima coluna.

$$\tilde{r}_i = \left(\frac{p_{i1}}{p_{i+}}, \frac{p_{i2}}{p_{i+}}, \dots, \frac{p_{iJ}}{p_{i+}} \right) \quad (9)$$

$$\tilde{c}_j = \left(\frac{p_{1j}}{p_{+j}}, \frac{p_{2j}}{p_{+j}}, \dots, \frac{p_{ij}}{p_{+j}} \right) \quad (10)$$

A partir dos vetores do perfil linha e coluna, numericamente, pode-se verificar o princípio da equivalência distribucional, em que dois perfis semelhantes podem ser combinados em um único perfil, com massa igual à soma das massas individuais, pertinentes aos perfis considerados na combinação.

Em outras palavras, exemplificando pelos perfis de linhas r_i e $r_{i'}$, com massas a_{i+}/a_{++} e $a_{i'+}/a_{++}$, dessa forma $r_i \approx r_{i'}$ e, naturalmente, $a_{i+} \approx a_{i'+}$ e $a_{ij} \approx a_{ij'}$. Logo, somando os elementos dos perfis de linha, para cada coluna, temos, então, um novo perfil de linha com os elementos $[(r_{i1}+r_{i'1}); \dots; (r_{ij}+r_{ij'})]$, com respectivas massas $(a_{i+}a_{i'+})/a_{++}$.

A métrica do espaço é a distância euclidiana, assim, a distância entre dois perfis linhas é obtida em (11) e, de forma análoga, a distância entre dois perfis colunas é dada em (12):

$$d_{(i,i')}^2 = \sum_{j=1}^J \frac{1}{p_{+j}} \left(\frac{p_{ij}}{p_{i+}} - \frac{p_{ij'}}{p_{i'+}} \right)^2 \quad (11)$$

$$d_{(j,j')}^2 = \sum_{i=1}^I \frac{1}{p_{i+}} \left(\frac{p_{ij}}{p_{+j}} - \frac{p_{ij'}}{p_{+j'}} \right)^2 \quad (12)$$

Pelo fato de que a métrica do espaço é a distância Qui-Quadrado, de modo a satisfazer a propriedade da equivalência distribucional (GREENACRE, 1992), tem-se que a importância desse resultado é que a inércia dos perfis linhas e colunas seja a mesma. Logo a diferença dos perfis linhas ou colunas e seus respectivos centroides é contemplada por essa distância, por meio de uma correção da matriz de correspondência P , originando uma nova matriz de dados, definida em

$$\mathbf{M} = [\mathbf{m}_{ij}] = \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}}. \quad (13)$$

Em função dessa matriz, aplica-se a decomposição dos valores singulares, assumindo a matriz de covariância dos perfis, representadas por \mathbf{MM}^t , referindo-se à matriz de covariância associada aos perfis colunas e $\mathbf{M}^t\mathbf{M}$ à matriz de covariância associada aos perfis linhas sendo \mathbf{M}^t a transposta de \mathbf{M} .

Os resultados obtidos com essa decomposição são de extrema importância, na análise de correspondência, uma vez que os autovalores quantificam a informação relativa à inércia e

sua distribuição nos eixos, dados como componentes que restituem a variabilidade amostral, explicada pelo mapas perceptuais. Assim, temos que a inércia total (IT) é quantificada por (14), podendo ser obtida pela estatística do teste qui-quadrado, usualmente conhecida e n definido como o número total de observações.

$$IT = \sum_{i=1}^I \lambda_i^2 = \frac{\chi^2}{n}, \text{ em que} \quad (14)$$

λ_i corresponde aos autovalores não nulos de $\mathbf{M}^t\mathbf{M}$. Logo a informação da inércia restituída a cada eixo é dada a seguir.

$$\frac{\lambda_i}{IT}. \quad (15)$$

Os autovetores normalizados de $\mathbf{M}^t\mathbf{M}$ e $\mathbf{M}\mathbf{M}^t$, obtidos na decomposição dos valores singulares, representados respectivamente por \mathbf{U} e \mathbf{V} , são aplicados no procedimento da distribuição da inércia, na obtenção das coordenadas padronizadas, por meio das expressões (16) e (17), resultantes dos escores das linhas L e das colunas C.

$$\mathbf{L} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{U} \quad (16)$$

$$\mathbf{C} = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{V}, \quad (17)$$

em que \mathbf{D}_r representa a matriz diagonal dos vetores massa linha e \mathbf{D}_c representa a matriz diagonal dos vetores massa coluna, definidos por:

$$\mathbf{D}_r = \begin{pmatrix} p_{1+} & 0 & \cdots & 0 \\ 0 & p_{2+} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{I+} \end{pmatrix} \quad (18)$$

$$D_C = \begin{pmatrix} p_{+1} & 0 & \cdots & 0 \\ 0 & p_{+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{+J} \end{pmatrix} \quad (19)$$

Com base nesses resultados, novas coordenadas das linhas e das colunas, conhecidas como coordenadas principais, são dadas, respectivamente, por:

$$Y = D_r^{-\frac{1}{2}} U^t C \quad (20)$$

$$Z = D_c^{-\frac{1}{2}} V^t L \quad (21)$$

2.3 A distância de Hellinger e a relação com a análise de correspondência simples

A introdução da distância de Hellinger, aplicada ao caso discreto na análise de correspondência como uma nova métrica competidora com a distância Qui-Quadrado, foi proposta por Rao (1995). Assim, mantendo a notação dada anteriormente, a distância de Hellinger, para dois perfis linhas (i, i') e colunas (j, j'), é definida como:

$$d_H^2(i, i') = \sum_{j=1}^J \left(\sqrt{\frac{p_{ij}}{p_{+i}}} - \sqrt{\frac{p_{i'j}}{p_{+i'}}} \right)^2 \quad (22)$$

$$d_H^2(j, j') = \sum_{i=1}^I \left(\sqrt{\frac{p_{ij}}{p_{+j}}} - \sqrt{\frac{p_{ij'}}{p_{+j'}}} \right)^2 \quad (23)$$

Convém ressaltar que a conexão dessa distância com a análise de correspondência pode ser compreendida, ao reportar um resultado dado por Dias (1988), enunciado da seguinte forma: Suponha que F seja uma função de distribuição discreta com probabilidade dada pelo vetor $P=(p_1, p_2, \dots, p_j)$. Contextualizando para a análise de correspondência, podemos associar essas probabilidades, por exemplo, aos níveis categóricos da variável descrita no sentido coluna. Portanto a distribuição empírica, obtida em J observações, é representada por S_n em:

$$S_n = \left(\frac{a_{1j}}{a_{j+}}, \frac{a_{2j}}{a_{j+}}, \dots, \frac{a_{jJ}}{a_{j+}} \right), \quad (24)$$

Dessa forma, aplicando a distância de Hellinger, tem-se por definição:

$$d_H^2(F, S_n) = \sum_{j=1}^J \left(\sqrt{\frac{a_j}{a_{j+}}} - \sqrt{p_j} \right)^2. \quad (25)$$

A relação dessa quantidade com a distância Qui-Quadrado é pautada em uma demonstração feita por Dias (1988), cujo principal resultado mostra-se a seguir.

$$\left(\sqrt{\frac{a_j}{a_{j+}}} - \sqrt{p_j} \right)^2 = \frac{(a_j - a_{j+}p_j)^2}{a_{j+}^2 p_j} \left(1 + \sqrt{\frac{a_j}{a_{j+}p_j}} \right)^{-2}. \quad (26)$$

Assim, é possível notar que a distância Qui-Quadrado é incorporada na distância de Hellinger, portanto pode-se afirmar-se que:

$$\chi^2 = \sum_{j=1}^J \frac{(a_j - a_{j+}p_j)^2}{a_{j+}p_j} \text{ resulta em } d_H^2(F, S_n) = \frac{\chi^2}{n} \quad (27)$$

sendo essa quantidade a inércia total, obtida na decomposição de valores singulares, envolvida na restituição da variação amostral explicada pelos eixos. A demonstração desse resultado se encontra no Anexo A.

Entre as vantagens que a distância de Hellinger proporciona, destaca-se que essa distância é limitada por uma propriedade que define sua amplitude. Para melhor entendimento, dentro da análise de correspondência, define-se P e Q como sendo:

$$P = \left(\frac{a_{11}}{a_{1+}}, \dots, \frac{a_{1I}}{a_{1+}} \right) \text{ e } Q = \left(\frac{a_{11}}{a_{+1}}, \dots, \frac{a_{1I}}{a_{+I}} \right). \quad (28)$$

Com base nesses vetores, Stjerneman (1983) demonstrou que:

$$d_H^2(P,Q) \leq \Delta(P,Q) \leq \sqrt{d_H^2(P,Q)(2-d_H^2(P,Q))} \leq \sqrt{2}d_H(P,Q)$$

Pela primeira desigualdade,

$$\begin{aligned} d_H^2(P,Q) &= \frac{1}{2} \sum_i |\sqrt{p_i} - \sqrt{q_i}| |\sqrt{p_i} - \sqrt{q_i}| \leq \frac{1}{2} \sum_i |\sqrt{p_i} - \sqrt{q_i}| (\sqrt{p_i} + \sqrt{q_i}) \\ &\leq \frac{1}{2} \sum_i |p_i - q_i| = \Delta(P,Q) \end{aligned}$$

Pelas duas últimas desigualdades,

$$\begin{aligned} \Delta^2(P,Q) &= \frac{1}{4} \left(\sum_{i \in [n]} |p_i - q_i| \right)^2 = \frac{1}{4} \sum_{i \in [n]} \left((\sqrt{p_i} - \sqrt{q_i}) (\sqrt{p_i} + \sqrt{q_i}) \right)^2 \\ &\leq \frac{1}{4} \left(\sum_{i \in [n]} (\sqrt{p_i} - \sqrt{q_i})^2 \right) \left(\sum_{i \in [n]} (\sqrt{p_i} + \sqrt{q_i})^2 \right) \end{aligned}$$

Por Cauchy- Schwarz:

$$\leq \frac{1}{2} d_H^2(P,Q) \cdot \left(2 + 2 \sum_{i \in [n]} \sqrt{p_i} \sqrt{q_i} \right) \leq d_H^2(P,Q) \cdot (2 - d_H^2(P,Q)) \leq \sqrt{2} d_H(P,Q)$$

Maiores detalhes desse resultado são dados no Anexo B. Sobre a perspectiva de avaliar a distância de Hellinger como uma métrica, Cabral, Frery e Ramires (2014) enunciam três suposições, dadas em (29) – (30) e (31). Assim, definindo os vetores de um modo geral, para os casos discretos por $U = \{u_1, \dots, u_k\}$, $W = \{w_1, \dots, w_k\}$ e $T = \{t_1, \dots, t_k\}$, temos:

$$d_H^2(U,W) \geq 0 \text{ e } d_H^2(U,W) = 0 \Leftrightarrow U=W \quad (29)$$

$$d_H^2(U,W) = 0 \Leftrightarrow U=W \quad (30)$$

$$d_H^2(U,W) = d_H^2(W,U)$$

$$d_H^2(U,W) \leq d_H^2(U,T) + d_H^2(T,W) \quad (31)$$

Uma particularidade que justifica a distância de Hellinger, em relação a outras medidas, é que, caso alguma distância viole uma das suposições mencionadas, origina-se o conceito de medidas de divergência (EVREN; TUNA, 2012). A saber, exemplificamos, para o caso discreto, que a medida Kullback-Leibler, definida em (32) é classificada como uma medida de divergência, considerando a discrepância entre U e W, para uma amostra X_1, \dots, X_n .

$$d_{KL}^2 = \sum_x u(x) \log \frac{u(x)}{w(x)} \quad (32)$$

Segundo Evren e Tuna (2012), essa medida é sempre positiva, porém não satisfaz a suposição (30). Além do mais, Príncipe (2010) alerta que, no caso contínuo, quando $w(x)=0$, ocorrem problemas de instabilidade, logo uma alternativa, para esse problema, pode ser dada pelo uso da distância de Hellinger.

No tocante à aplicação da distância de Hellinger, na análise de correspondência, Jaeger et al. (2014) mencionam que o uso da distribuição Qui-Quadrado entre os perfis tem como principal a deficiência em situações em que baixas massas são observadas. Nesse caso, a similaridade entre as variáveis são imprecisas, em razão do forte impacto na estatística Qui-Quadrado, por resultar em baixas frequências esperadas.

Quanto ao exposto, a distância de Hellinger é uma alternativa viável de ser usada. Meyners, Castura e Carr (2013) apontam que, em geral, os programas estatísticos ainda ignoram o uso dessa distância, na análise de correspondência e, como consequência, muitas análises de similaridades poderão ser questionadas.

2.4 Escore de Risco

O escore de risco, conforme menciona Nakano (2010), tem por objetivo associar um índice de gravidade, que não se restringe a probabilidades, mas, sim, à definição das categoriais, mais especificamente ao risco de um indivíduo ser classificado, em uma categoria que represente uma maior gravidade, tendo como referência a ocorrência de um evento esperado, previamente definido.

Nesse cenário, pode-se conjecturar uma situação, na qual há três possíveis desfechos, para um paciente em um determinado hospital. Desta forma, definem-se os eventos por:

- A – Alta do paciente, atribui-se a uma categoria menos grave.
- B – Permanecer no ambulatório, associa-se a uma categoria intermediária.
- C – Ir para a UTI, associa-se a uma categoria mais grave.

Para tal situação, o escore de risco proporciona um índice de gravidade que classifica um indivíduo com um risco mais grave em relação a outro paciente, de forma independente e que cada paciente possa apresentar uma probabilidade de ser classificado.

A construção desse índice, em consonância com a análise de correspondência, parte do pressuposto de que assumir um vetor de probabilidades \mathbf{P} , no qual cada componente poderá ser especificado pelos perfis das variáveis categóricas ordinais, dispostas em linhas e/ou colunas. Tal índice é obtido, a partir da distância euclidiana entre o ponto \mathbf{P} e a melhor situação possível, na qual o grupo de menor gravidade é classificado com probabilidade 1.

Com essa breve explanação, segue-se ao procedimento matemático, para a construção do índice, considerando diferentes números de k categorias.

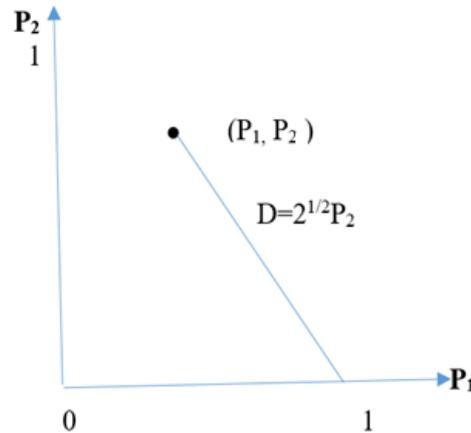
2.4.1 Construção do Índice para dois grupos ($k=2$)

Nesse caso, supõe-se que existe o problema de classificação entre dois grupos, G_0 e G_1 , em que G_1 indica o grupo de maior gravidade. Seja \mathbf{P} um vetor de probabilidades, seja o vetor dado por $(P_1; P_2) = (1 - P_2, P_2)$, em que P_2 seja proporcional à distância euclidiana entre $\mathbf{i} = (1; 0)$ e $\mathbf{P} = (P_1; P_2)$, isto é, a razão entre a distância de $\mathbf{P} = (P_1, P_2)$ à melhor situação possível $\mathbf{i} = (1,0)$, assim, a distância euclidiana entre esses pontos é calculada por (33).

$$\frac{D\{(1;0);(P_1;P_2)\}}{D\{(1;0);(0;1)\}} = \frac{\sqrt{(1-P_1)^2 + P_2^2}}{\sqrt{2}} = \frac{P_2\sqrt{2}}{\sqrt{2}} = P_2 \quad (33)$$

O resultado (33) permite interpretar que $D\{(1;0); (P_1;P_2)\}$ corresponde ao percurso completo entre o ponto observado e a melhor situação possível, passando por todos os pontos melhores do que $\mathbf{P} = (P_1; P_2)$ (NAKANO, 2010). Geometricamente, isso pode ser observado na Figura 1.

Figura 1 - Distância percorrida pelo índice de gravidade para $k=2$ grupos.



Fonte: Nakano (2010)

Em síntese, para $k=2$ grupos, o índice de gravidade denominado por IG_2 é dado por:

$$IG_2 = \frac{D\{(1;0);(P_1;P_2)\}}{D\{(1;0);(0;1)\}} = P_2 \quad (34)$$

2.4.2 Construção do índice para três grupos ($k = 3$)

Considerar agora o caso em que se tem três grupos ordenados por gravidade de G_0 a G_2 , sendo G_0 o grupo de menor gravidade, G_1 o grupo de gravidade intermediária e G_2 o de maior gravidade. Analogamente ao caso de dois grupos, considere agora os vetores $\mathbf{P} = (P_1; P_2; P_3)$ e $\mathbf{i} = (1; 0; 0)$, referindo-se ao vetor da melhor situação possível.

Preliminarmente, calcula-se a distância percorrida entre os pontos \mathbf{P} e \mathbf{i} , de forma a contemplar todos os pontos que sejam melhores que \mathbf{P} e piores que \mathbf{i} . Assim, a distância euclidiana entre \mathbf{P} e $\mathbf{P}_1 = (P_1; P_2 + P_3; 0)$, que contempla todos os pontos melhores que \mathbf{P} e piores que \mathbf{P}_1 , é dada em:

$$D_1 = D\{(p_1; p_2 + p_3; 0), (p_1; p_2; p_3)\} = \sqrt{p_3^2 + p_3^2} = p_3 \sqrt{2} \quad (35)$$

Adicionando a distância euclidiana entre \mathbf{P}_1 e $\mathbf{i} = (1;0;0)$, contemplando todos os pontos piores que \mathbf{i} e melhores do que \mathbf{P}_1 , resulta na expressão:

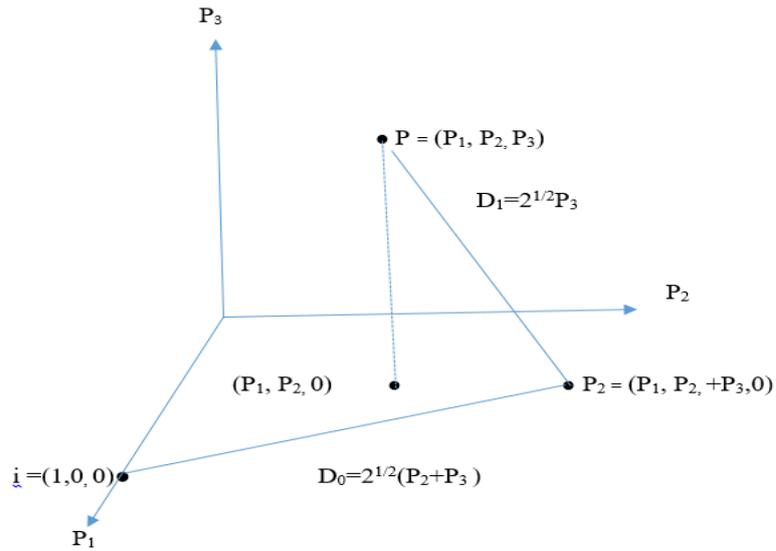
$$\begin{aligned}
D_0 &= D\left\{(1;0;0);(P_1;P_2+P_3;0)\right\} = \sqrt{(1-P_1)^2 + (P_2+P_3)^2} \\
&= \sqrt{(P_2+P_3)^2 + (P_2+P_3)^2} = \sqrt{2}(P_2+P_3).
\end{aligned} \tag{36}$$

Desta forma, a distância total entre \mathbf{P} e \mathbf{i} resume-se em D .

$$D = D_0 + D_1 = \sqrt{2}(P_2+2P_3). \tag{37}$$

Geometricamente, a trajetória percorrida entre os pontos é ilustrada na Figura 2.

Figura 2 - Distância percorrida pelo índice de gravidade para $k = 3$ grupos.



Fonte: Nakano (2010)

Considerando este caso de $k = 3$ grupos, o valor máximo possível a ser atingido é $2\sqrt{2}$, a distância percorrida total entre $(1;0;0)$ e $(0;0;1)$. Dessa forma, pode-se considerar como índice de gravidade a função **IG3** dada por:

$$IG_3 = \frac{D}{D\{(0;0;1);(0;1;0)\} + D\{(1;0;0);(0;1;0)\}} = \frac{\sqrt{2}(P_2+2P_3)\sqrt{2}}{2\sqrt{2}} = \frac{P_2+2P_3}{2} \tag{38}$$

2.4.3 Construção do índice para k grupos (k > 3)

Seguindo a extensão dos resultados mencionados anteriormente, a construção do índice de gravidade, para uma situação que envolve um número de grupos $k > 3$, é dada por iniciar com a categoria G_0 a $G_{(k-1)}$. Enfatizamos que G_0 indica o grupo de menor gravidade. Os demais grupos são definidos ordenadamente, segundo o aumento da gravidade e, por fim, $G_{(k-1)}$ é considerado o grupo de maior gravidade.

Como mostrado anteriormente, o objetivo inicial é calcular a distância percorrida entre os pontos $\mathbf{P} = (p_1, p_2, \dots, p_{k-1})$ e $\mathbf{i} = (1, 0, \dots, 0)$, de forma a contemplar todos os pontos que sejam melhores que \mathbf{P} e piores que \mathbf{i} . Para isso, calculamos a distância entre \mathbf{P} e $\mathbf{P}_{(k-2)}$ cujo resultado é dado em (39).

$$\begin{aligned} D_{(k-2)} &= D \left\{ \left(p_1, p_2, \dots, p_{(k-2)} + p_{(k-1)}; 0 \right); \left(p_1, p_2, \dots, p_{(k-2)}; p_{(k-1)} \right) \right\} \\ &= \sqrt{p_{(k-1)}^2 + p_{(k-1)}^2} = \sqrt{2} p_{(k-1)} \end{aligned} \quad (39)$$

O resultado entre a distância de $\mathbf{P}_{(k-2)}$ e $\mathbf{P}_{(k-3)}$ é dado a seguir:

$$\begin{aligned} D_{(k-3)} &= D \left\{ \left(p_1; p_2; \dots; p_{(k-3)} + p_{(k-2)} + p_{(k-1)}; 0; 0 \right); \left(p_1; p_2; \dots; p_{(k-3)}; p_{(k-2)} + p_{(k-1)}; 0 \right) \right\} \\ &= \sqrt{\left(p_{(k-2)} + p_{(k-1)} \right)^2 + \left(p_{(k-2)} + p_{(k-1)} \right)^2} = \sqrt{2} \left(p_{(k-2)} + p_{(k-1)} \right). \end{aligned} \quad (40)$$

E, assim por diante até a distância de \mathbf{P}_1 e \mathbf{i} , dado na expressão (41)

$$\begin{aligned} D_0 &= D \left\{ \left(1; 0; \dots; 0 \right); \left(p_1, p_2 + p_3 + \dots + p_{(k-2)} + p_{(k-1)}, \dots, 0; 0 \right) \right\} \\ &= \sqrt{(1-p_1)^2 + \left(\sum_{i=1}^{k-1} p_i \right)^2} = \sqrt{\left(\sum_{i=1}^{k-1} p_i \right)^2 + \left(\sum_{i=1}^{k-1} p_i \right)^2} \\ &= \sqrt{2} \sum_{i=1}^{k-1} p_i = \sqrt{2} \left(p_1 + p_2 + \dots + p_{(k-1)} \right). \end{aligned} \quad (41)$$

Logo, a distância total considerada é dada por:

$$D = \sum_{r=0}^{k-2} D_r = \sqrt{2} \sum_{j=1}^{k-1} j p_j = \sqrt{2} (p_2 + 2p_2 + (k-1)p_{k-1}) \quad (42)$$

Considerando o caso de k grupos ($k > 3$), o valor máximo possível atingido por D_r é verificado em $r = 0, 1, \dots, k-2$. Assim, a distância total entre $(1; 0; 0; \dots; 0)$ e $(0; 0; \dots; 0; 1)$ é obtida por :

$$\max \{D\} = \sum_{r=0}^{k-2} \max \{D_r\} = \sum_{r=0}^{k-2} \sqrt{2} = (k-1)\sqrt{2} \quad (43)$$

Logo, para o caso geral de k ($k > 3$) grupos, o índice de gravidade, denotado por IG_k , é finalizado, conforme expressão:

$$IG_k = \frac{D}{\max \{D\}} = \frac{\sqrt{2} (p_1 + 2p_2 + \dots + (k-1)p_{(k-1)})}{(k-1)\sqrt{2}} = \frac{p_1 + 2p_2 + \dots + (k-1)p_{(k-1)}}{(k-1)} \quad (44)$$

Dado esse procedimento metodológico, Nakano (2010) menciona algumas notas importantes que carecem de ser discutidas.

Nota 1: O índice agrega pesos maiores, para os grupos de maior gravidade, dando peso zero ao grupo G_0 de menor gravidade. O índice é dividido pela distância máxima possível, simplesmente, para padronizá-lo numa escala 0 a 1, que pode também ser representada por 0 a 100%.

Nota 2: Em vez de um índice de gravidade (IG), poderia haver o interesse em definir um Índice de Qualidade (IQ). O índice de qualidade é o inverso do IG e ele é definido a partir da distância (euclidiana) entre o ponto \mathbf{P} e a pior situação possível. Com um desenvolvimento análogo ao índice de gravidade definido em (44), pode-se expressar o IQ por:

$$IQ_k = 1 - IG_k. \quad (45)$$

2.5 Blends de café e os principais aspectos na qualidade e formulação

Com a elevada demanda do consumo de café, seja no mercado comercial, em termos de ações ou commodities, bem como no aumento do consumo, surge um espaço para que novos produtos sejam introduzidos em uma escala industrial. Com essa motivação, destacam-se os

blends de cafés de diferentes variedades, em que os grãos são torrados separadamente e misturados apenas após a moagem. Isso acontece, porque cada tipo de café tem um ponto particular de torra.

Dentre as inúmeras variedades de cafés, que podem participar da composição de um *blend*, notoriamente, tem-se a mistura de grãos de *C.arábica*, caracterizada por resultar em bebidas aromáticas e mais ácidas, com um sabor diferenciado, em relação aos grãos de *C.canéfora*, no qual a bebida é mais amarga resultando em bebidas mais encorpadas. (FIGUEIREDO, 2013). Contudo fatores externos, tais como formas de processamento podem maximizar determinadas características desejadas que atendam a perfis de diferentes consumidores.

No caso do processamento via úmida, observa-se que a ocorrência de fermentações indesejáveis é reduzida, quando comparada ao processamento via seca, no qual, em geral, observam-se maiores teores de açúcares quando comparado ao processamento por via úmida (AMORIM et al., 2009; BORÉM, 2008; GONZALÉZ-RIOS et al., 2007).

Em contexto mais desafiador à pesquisa, podemos ressaltar questões que envolvem a incorporação de diferentes variedades dos cafés mencionados, em consonância com os fatores externos que influenciam a qualidade e a aceitação dos *blends*. Diante do exposto, alguns relatos na literatura têm enfatizado esse problema. Estudos com *blends*, obtidos por Lima Filho et al. (2011) e Mendes (2005), contendo proporções de até 60% de conilon natural e descascados, são recomendáveis à formulação de *blends* com café arábica sem comprometer a qualidade da bebida final.

Evidentemente, tornam-se questionáveis quais variáveis refletem a qualidade da bebida, uma vez que a composição fisiológica dos grãos, procedência, ou seja, composição química e física da bebida, bem como aspectos sensoriais dos consumidores, ainda requerem pesquisas mais direcionadas. Alguns exemplos da influência desses aspectos são descritos a seguir.

A composição físico-química define a qualidade final da bebida, sendo responsável por suas características sensoriais, e a acidez dos grãos está relacionada com a qualidade do produto, a intensidade a vários fatores como os níveis de fermentação que ocorre nos grãos, condições climáticas, durante a colheita e secagem, estágios de maturação, local de origem e forma de preparo (PIMENTA, 2003; SIQUEIRA; ABREU, 2006).

Convém ressaltar que a qualidade intrínseca dos grãos de café, ou seja, sua constituição química é que determinará a qualidade diferenciada de um café especial. Após a torra, os compostos químicos dos grãos resultarão na formação de sabores e aromas que

caracterizarão a bebida. Portanto não se deve confundir a qualidade do café com a preferência do consumidor, já que a qualidade intrínseca dos grãos é determinada pela interação dos fatores genéticos, ambientais e do processamento por fatores socioeconômico-culturais e pelos conhecimentos específicos relacionados com a bebida do café, além da preferência do consumidor (GIOMO; BORÉM, 2011).

O ponto ideal de maturação para a colheita é o estágio cereja em que a composição química do fruto se encontra desenvolvida e equilibrada com o máximo potencial de expressão de qualidade de bebida.

O pH indica mudanças nos frutos como o processo indesejável de fermentação que ocorre na pré e pós-colheita resultando em defeitos no produto. De acordo com Siqueira e Abreu (2006), o pH ideal deve estar entre 4,95 e 5,20, tornando as características sensoriais do café mais aceitáveis pelo consumidor, sem excesso de amargor ou acidez, sendo desejável no café um maior teor de extrato aquoso, tanto pelo ponto de vista do rendimento industrial, como pela contribuição para assegurar o corpo da bebida necessário em cafés tipo expresso.

3 METODOLOGIA

Em concordância com os objetivos propostos, a metodologia foi desenvolvida nas seguintes etapas: 3.1- Descrição dos experimentos e obtenção das medidas da análise de correspondência generalizada e 3.2- Simulação de tabelas de contingência com dependência entre os níveis categóricos.

3.1 Descrição dos experimentos e obtenção das medidas da análise de correspondência generalizada

Os *blends* foram constituídos por cafés de diferentes qualidades e procedência, assim, consideraram-se, na formalização dos *blends*, os cafés arábica especial (CAE), mais especificamente, o café Bourbon amarelo, Canéfora (CC) e comercial torrado e moído (CAC), nas proporções descritas na Tabela 3.

Tabela 3 - Proporções utilizadas na composição dos *blends* formados pelos cafés: bourbon amarelo (CAE), arábica comercial torrado e moído (CAC) e canéfora (CC).

Amostra	CAE	CAC	CC
1	1,000	0,000	0,000
2	0,670	0,330	0,000
3	0,340	0,330	0,330
4	0,500	0,500	0,000
5	0,500	0,000	0,500
6	0,340	0,660	0,000
7	0,340	0,000	0,660
8	0,000	1,000	0,000
9	0,000	0,000	1,000

Fonte: Dados do autor (2018)

Seguindo essas especificações, tem-se que os *blends* formados pela composição de dois cafés foram numerados por *blends* com misturas binárias, os cafés puros denominados como mistura pura e, por fim, a caracterização dos *blends* com os três tipos de cafés foi caracterizada por mistura ternária.

Em se tratando do café arábico especial considerado neste trabalho, por ser uma das espécies mais requisitadas no mercado de cafés especiais, dadas suas características particulares, de corpo médio e baixa acidez, procedeu-se com a realização de quatro repetições diferenciadas pelo tipo de processamento e concentrações 7% e 10% m/v (35g/500 mL).

Dadas as misturas (TABELA 3), o preparo das amostras foi feito utilizando água potável a 93°C sem adição de açúcar. O tempo de extração foi de 4 minutos, utilizando como método de preparo a filtração. Dessa maneira, foram evitados quaisquer riscos relacionados a reações alérgicas ou ao aumento da taxa de glicose nos indivíduos avaliadores pertencentes às amostras de consumidores comuns, respeitando-se as normas de higiene impostas pelo comitê de ética sob o protocolo de CAAE: 14959413.1.0000.5148.

Cada repetição foi executada em sessões distintas com intervalos de 24 horas, em função do número excessivo de avaliações. O grupo de indivíduos avaliadores, em um experimento de teste, foi composto por cinco provadores qualificados, no sentido de serem considerados aptos à diferenciação das amostras nos experimentos sensoriais, decorrente da experiência em degustação de experimentos.

Cada avaliador degustou cerca de 20 mL de bebida preparada, a partir dos *blends* formulados, na temperatura de, aproximadamente, 65°C, servidos em copos descartáveis. Após a degustação de cada *blend*, o avaliador registrou sua avaliação em fichas apropriadas.

Os *blends* foram avaliados recebendo notas que variaram de 0 a 10, para as características qualitativas da bebida: corpo, acidez; sabor, amargor e nota final. A característica nota final representou a impressão global da qualidade descrita pelo avaliador.

Com base nesses dados, considerando esses atributos, tabelas de contingência foram construídas, assumindo os *blends* definidos como Puros, Binários e Ternários e, para os atributos sensoriais avaliados em sabor, acidez e amargor, as notas obtidas foram categorizadas em quatro classes C_1, \dots, C_4 , conforme sugere o layout descrito na Tabela 4. As tabelas de contingência obtidas encontram-se no ANEXO C.

Tabela 4 - Layout da tabela de contingência estruturada pela frequência de notas, a_{ij} ($i=1,\dots,3$ e $j=1,\dots,4$) dos *blends* em relação às categorias C_1,\dots,C_4 computadas nos quatro experimentos.

	C_1	C_2	C_3	C_4
<i>blends</i>	[<3]	[3-5]	[5-7]	[>7]
Pura (P)	a_{11}	a_{12}	a_{13}	a_{14}
Binária (B)	a_{21}	a_{22}	a_{23}	a_{24}
Ternária (T)	a_{31}	a_{32}	a_{33}	a_{34}

Fonte: Dados do autor (2018)

Considerando a estrutura de dados (TABELA 5) e a expressão (46) definida por Naito (2007), na qual a decomposição de valores singulares, para que a análise de correspondência contemple as métricas Qui-Quadrado e Hellinger, adotaram-se valores específicos para os parâmetros (α,β) .

$$D_1^{\frac{1}{2}} (I-\beta)\mathbf{1}\mathbf{t}^t \left\{ \frac{1}{\alpha(1+\beta)} \left(D_1^{-1} Q D_c^{-1} \right)^\alpha - \mathbf{1}\mathbf{1}\beta \right\} \left((I-(1-\beta)\mathbf{1}\mathbf{c}^t)^t D_c^{\frac{1}{2}} \right) \quad (46)$$

Assim, em particular, seguindo recomendação do autor, fixou-se $(\alpha=1,\beta=0)$ e $(\alpha=0,5,\beta=1)$, de modo que os escores da análise de correspondência foram obtidos, respectivamente, para as métricas da distância Qui-Quadrado e de Hellinger. Desta forma, os mapas perceptuais foram construídos e D_1 referiu-se à raiz quadrada da matriz diagonal das massas marginais, obtidas pelo perfil linha, D_c de forma análoga, às massas para o perfil coluna; Q à matriz de proporção corrigida pela distância Qui-Quadrado e, por fim, I à matriz identidade.

Com a obtenção dos escores, computou-se a correlação de cada nível categórico em relação aos eixos, conforme a expressão (47). Assim tem-se cor_{ik} , correspondente à correlação dos níveis das variáveis *blends* e cor_{jk} aos níveis da variável nota. Em ambas as expressões, a identificação dos eixos foi feita por $k=1$ e 2 . Já f_{i1},\dots, f_{ik} e f_{j1},\dots, f_{jk} referem-se aos escores obtidos, ao identificador do eixo, conforme descrito na Seção 2.2.

$$\begin{aligned} \text{cor}_{ik} &= \frac{f_{ik}^2}{d_i^2} \text{ em que, } d_i^2 = f_{i1}^2 + \dots + f_{ik}^2 \\ \text{cor}_{jk} &= \frac{f_{jk}^2}{d_j^2} \text{ em que, } d_j^2 = f_{j1}^2 + \dots + f_{jk}^2 \end{aligned} \quad (47)$$

Dadas essas especificações, o script na linguagem R, para a obtenção dos resultados, encontra-se no ANEXO D.

Posteriormente, para cada variável representada na tabela de contingência, obteve-se o índice de gravidade, seguindo a metodologia descrita nas seções (2.4.2) e (2.4.3). O propósito foi sugerir qual mapa perceptual é mais adequado, para representação geométrica dos escores, que agregue uma interpretação, pautada por referenciais teóricos, que justifique a similaridade dos *blends* com as categorias de notas sugeridas. Para a construção desse índice, consideraram-se as correlações de cada nível categórico com o 1º eixo, como probabilidades exigidas na obtenção dos índices, uma vez que a metodologia considera como melhor situação o ponto $i=(0,0,1)$, portanto as componentes que formam as coordenadas do ponto P não podem ser maiores que 1.

Em relação ao desempenho da utilização das métricas Qui-Quadrado e Hellinger, em tabelas de contingência com dependência dos níveis categóricos, procedeu-se com um estudo de simulação, descrito na seção 3.2.

3.2 Simulação de tabelas de contingência com dependência entre os níveis categóricos

O procedimento de simulação Monte Carlo, envolvido nesse processo, considerou 1000 realizações. As tabelas de contingência representadas pela matriz $X_{I \times C}$ tiveram a i -ésima linha ($i=1,2,3$) fixada, as frequências observadas Y_{ij} ($j=1,2,3,4$) foram simuladas, seguindo a distribuição binomial correlacionada, ou seja, $Y_{ij} \sim BC(n, \pi, \rho)$ definida por:

$$p(Y|n_j, \pi_j, \rho) = \binom{n_j}{y_j} \pi_j^{y_j} (1-\pi)^{n_j-y_j} (1-\rho) IA_1(y_j) + \pi^{y_j/n_j} (1-\pi_j)^{n_j-y_j/n_j} \rho IA_2(y_j), \quad (48)$$

sendo $A_1 = 0, 1, \dots, n_j$, $A_2 = 0, n_j$, $y_j = 0, \dots, n_j$ e $0 \leq \rho \leq 1$. Seguindo o procedimento dado por Cirillo e Ramos (2014), o vetor de variáveis aleatórias $\mathbf{Y} = (y_1, y_2, \dots, y_j)$, do qual cada componente representou o número de ocorrências na categoria j , para $j=1,2,3,4$, associado a

um vetor $\pi_j=(\pi_1, \pi_2, \dots, \pi_j)$, sendo π_j corresponde à probabilidade de sucesso da binomial e ρ à taxa de mistura das distribuições binomial (n, π) com probabilidade $(1-\rho)$, é uma distribuição Bernoulli modificada, representada por B e $M(\pi)$, assumindo 0 ou n valores com probabilidade ρ .

Seguindo as especificações, os valores paramétricos, utilizados na geração das tabelas de contingência, encontram-se resumidos na (TABELA 5). Para cada cenário, computou-se a inércia, em relação aos três primeiros eixos, considerando as métricas estudadas neste trabalho.

Tabela 5 - Valores paramétricos utilizados como cenário para gerar as tabelas de contingência à aplicação da análise de correspondência de dimensão $L=3$, e $C=4$.

Proporção da Binomial (π_j)	Grau de correlação (ρ)
	0,2
0,2	0,5
	0,8
	0,2
0,5	0,5
	0,8
	0,2
0,9	0,5
	0,8

Fonte: Dados do autor (2018)

4 RESULTADOS E DISCUSSÕES

4.1 Estudo preliminar da explicação da inércia na análise de correspondência utilizando as distâncias Qui-Quadrado e Hellinger

Em concordância com as tabelas de frequências descritas no Anexo A, das quais as contagens das notas foram feitas, considerando as pontuações obtidas nas 5 repetições, isto é, notas dadas pelos provadores em cada experimento, em relação aos cafés puros (P), as misturas compostas por duas (B) e três (T) variedades, em função das notas de classificação, realizou-se a análise de correspondência considerando as métricas de Qui-Quadrado e Hellinger. Desta forma, inicialmente, os resultados foram descritos na Tabela 6, para cada atributo sensorial avaliado, sendo úteis na validação dos mapas perceptuais e interpretados como a porcentagem de variação amostral obtida pela decomposição da inércia.

Tabela 6 - Porcentagem de variação amostral obtida na análise de correspondência, utilizando a distância Qui-Quadrado e de Hellinger, a ser restituída nos três eixos principais, para os atributos sensoriais: Sabor, Acidez e Amargor.

Distância Qui-Quadrado						
	Sabor		Acidez		Amargor	
	autovalores	acum. (%)	autovalores	acum. (%)	autovalores	acum. (%)
Eixo 1	0,6891	53,23	0,5330	61,86	0,5556	87,70
Eixo 2	0,5216	93,52	0,2671	92,86	0,0780	100
Eixo 3	0,0838	100	0,0614	100	3,975e17	0
Distância de Hellinger						
	Sabor		Acidez		Amargor	
	autovalores	acum. (%)	autovalores	acum. (%)	autovalores	acum. (%)
Eixo 1	0,3837	55,25	0,3498	60,61	0,3262	91,80
Eixo 2	0,2386	89,61	0,1509	84,76	0,028	100
Eixo 3	0,0721	100	0,0764	100	0,0013	0

Fonte: Dados do autor (2018)

Os resultados evidenciados na Tabela 6 indicam que, de um modo geral, para todos os atributos, nota-se que a variação amostral, explicada pela inércia total restituída nos dois primeiros eixos, cujos valores foram superiores a 80%, confirmaram que o estudo da

similaridade entre os níveis categóricos das variáveis *blends* e notas pode ser representado. Em relação à comparação das métricas, notam-se as diferenças, em relação ao uso em um mapa perceptual, representado por dois eixos bidimensionais.

A distância de Hellinger, sendo esta a principal competidora à distribuição Qui-quadrado e, embora a variação amostral explicada pelos dois eixos tenha sofrido uma pequena redução, tal fato não inviabiliza o estudo das associações entre os níveis categóricos das variáveis em um mapa com dois eixos.

Convém destacar que, para o atributo amargor, ambas as métricas concentraram, praticamente, toda a variação amostral explicada por um único eixo. Tal situação sugere que o princípio da independência entre as variáveis categóricas, exigido na realização da análise de correspondência, foi violado. Diante do exposto, torna-se cabível um estudo do comportamento dessas métricas em relação a cenários que envolvam diferentes graus de correlação entre as variáveis categóricas. Com esse propósito, seguem os resultados descritos na Tabela 7.

4.2 Estudos de simulação do comportamento das métricas de Qui-Quadrado e Hellinger

Em diferentes cenários, caracterizados pelo modelo binomial correlacionado de tal forma que baixas, moderadas e elevadas proporções e correlações são contemplados, os resultados encontrados, na Tabela 8, tendo como foco a inércia restituída pelo primeiro eixo, evidencia que, conforme se aumenta o grau de correlação entre as variáveis, a métrica Qui-Quadrado tende a sofrer um maior impacto, ao passo que a métrica de Hellinger reflete resultados mais robustos.

Tabela 7 - Simulação da decomposição da inércia da análise de correspondência, considerando as métricas definidas pela distância Qui-quadrado e Hellinger, em uma tabela de contingência estruturada em 3 linhas e 4 colunas.

Proporção da binomial (π_j)	Grau de correlação ρ	Distância Qui-Quadrado			Distância Hellinger		
		1°	2°	3°	1°	2°	3°
0,2	0,2	0,328	0,089	1.151e-17	0,268	0,105	3.150e-02
	0,5	0,984	0,169	2.803e-17	0,358	0,147	0,035
	0,8	3,401	0,197	4.824e-17	0,484	0,136	0,021
0,5	0,2	0,506	0,245	0,113	0,199	0,121	0,068
	0,5	0,817	0,416	0,223	0,298	0,163	0,104
	0,8	1,198	0,544	0,279	0,392	0,185	0,111
0,9	0,2	0,419	0,159	0,080	0,290	0,726	0,065
	0,5	0,664	0,271	0,214	0,140	0,093	0,066
	0,8	0,864	0,304	0,203	0,173	0,098	0,071

Fonte: Dados do autor (2016)

Nota-se que a inércia sendo distribuída de forma mais equilibrada nos eixos implica uma distribuição mais simétrica pelos escores das linhas e/ou colunas, de tal forma que o estudo das associações entre as variáveis seja mais adequado ao utilizar essa distância em situações de dependência entre os níveis categóricos. Evidentemente que, para uma baixa explicação da inércia nos dois primeiros eixos, a construção de mapas perceptuais bidimensionais não é propícia para ambas as métricas.

Em termos práticos, a importância desses resultados conduz à realização da análise de correspondência, sem a necessidade de uma inferência estatística preliminar, usualmente, dada pelo teste Qui-Quadrado, sob a hipótese de independência entre as variáveis categóricas. Além do mais, a utilização desses testes é criticada, em situações de baixas frequências marginais, as quais resultam em frequências esperadas pequenas.

Assim, naturalmente, o teste Qui-Quadrado é criticado, porém tal argumentação não é enfatizada, quando esse teste é aplicado à análise de correspondência, ou quando se propõe a utilizar essa distância na matriz de probabilidades \mathbf{P} , a ser utilizada para obter as matrizes de covariâncias dos perfis “linhas” ou “colunas”.

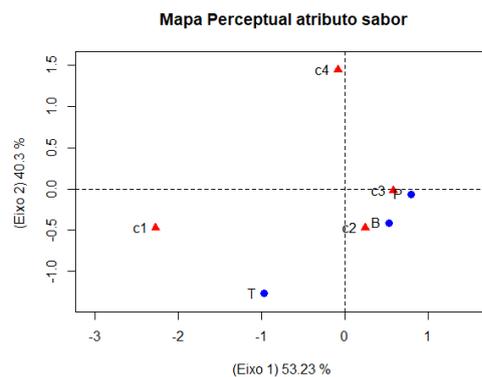
Referente a esse problema, os resultados robustos observados na métrica de Hellinger conduzem à afirmação de que o uso do teste Qui-Quadrado pode ser descartado, portanto torna-se promissora a realização da análise de correspondência com a perspectiva da distância de Hellinger a ser utilizada como fundamento da análise de correspondência.

4.3 Aplicação da análise de correspondência utilizando as distâncias Qui-Quadrado e Hellinger

Em se tratando da aplicação das métricas, em relação à avaliação sensorial dos *blends* em função das notas, mesmo com as críticas mencionadas em relação ao uso da distribuição Qui-quadrado, na análise de correspondência, convém expor uma análise comparativa das associações entre as duas abordagens dessa análise, uma vez que outros fatores não contemplados neste trabalho, porém conhecidos na literatura, como, por exemplo, propriedades físicas e químicas, ponto de torração, temperatura características individuais a cada consumidor, interferem na preferências dos *blends*.

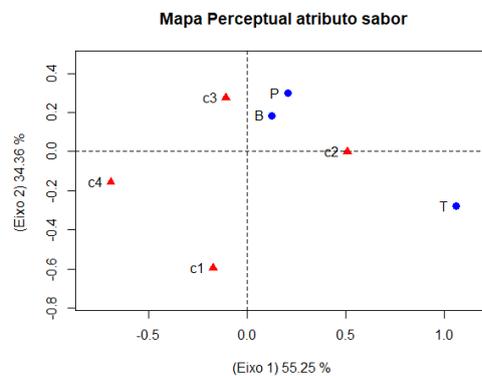
Sob esse enfoque, seguem as associações, representadas nos mapas perceptuais de cada atributo. Inicialmente, nas Figuras 3 e 4, seguem as associações para o atributo sabor.

Figura 3 - Mapa perceptual para o atributo sabor, considerando a métrica Qui-Quadrado.



Fonte: Dados do autor (2018)

Figura 4 - Mapa perceptual para o atributo sabor, considerando a métrica de Hellinger.



Fonte: Dados do autor (2018)

Em termos comparativos, as associações representadas pelas métricas Qui-Quadrado (FIGURA 3) e Hellinger (FIGURA 4), em geral, apresentaram um comportamento similar, em relação à inércia, ou seja, nota-se que os mapas foram assimétricos com influência dos perfis referentes aos cafés puros (P) e *blends* com composição binária (B), em conjunto com moderadas categorias de notas [3-5] (C₂) e [5-7] (C₃).

Diante do exposto, sendo menos rigoroso em relação à análise das correlações de cada perfil com os eixos, há evidências para afirmar que essas categorias citadas podem sugerir uma associação ou padrão, em relação à preferência dos *blends* e/ou perfil sensorial dos provadores. Todavia um resultado de interesse e digno de ser ressaltado, em ambos os mapas perceptuais, referiu-se aos *blends* com composição ternária (T) e às categorias com notas extremas [<3] (C₁) e [>7] (C₄). Para ter maior subsídio, em relação ao estudo da similaridade e/ou heterogeneidade desses perfis, recorre-se às seguintes argumentações científicas, pautadas nas referências descritas a seguir.

Segundo mencionam Alessandrini et al. (2008), a torração resulta em alterações físicas/químicas que influenciam as complexas reações que ocorrem, principalmente, na cor, aroma e sabor do café. Nesse contexto, ao considerar um *blend* formado pela composição de três cafés, sendo um deles o café comercial torrado e moído, torna-se coerente que seu sabor seja diferenciado, em algum aspecto aos *blends*, que não consideram esse tipo de café em sua composição, ou seja, cafés puros (P) ou misturas binárias (B).

Evidentemente, isso implica que a variedade e espécie do café, a ser considerada na análise da preferência dos consumidores, em relação aos *blends*, é relevante e carece de ser enfatizada e o equilíbrio entre corpo, acidez, doçura e grau de torração deverão ser aspectos relevantes na análise da qualidade de um *blend*.

A diferenciação entre as categorias de notas extremas, isto é, C₁ e C₄, supostamente são compreendidas por um perfil sensorial associado a fatores externos. Nesse contexto, Andrade e Lopes (2015) mencionam, por exemplo, que consumidores de um poder aquisitivo mais elevado tendem a preferir cafés puros. Assim, justifica-se pelos resultados neste trabalho notas inferiores ao *blends* avaliados.

Em termos estatísticos, os resultados descritos na Tabela 9 evidenciam que, em ambos os mapas, a qualidade de representação dos *blends* ternários (T) é fortemente caracterizada no eixo 2, ao passo que a representação das categorias C₁ e C₄ são contrastantes em relação aos eixos obtidos em cada métrica. Assim, a nomeação dos eixos torna-se questionável, em relação às categorias de notas, de modo a não fornecer informação suficiente, para seleção de

um mapa perceptual, que represente a similaridade entre os *blends* e a avaliação sensorial do atributo sabor, nas classes das notas categorizadas.

Tabela 8 - Coordenadas e correlação com os eixos da análise de correspondência considerando as métricas Qui-Quadrado e Hellinger para o atributo sabor.

Distância Qui-Quadrado				
Perfis Linha				
<i>Blends</i>	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,8005	-0,0614	0,3431	0,0021
Binária (B)	0,5305	-0,4084	0,1507	0,0940
Ternária(T)	-0,9721	-1,2662	0,5060	0,9038
Perfis coluna				
[<3] (c1)	-2,2728	-0,4720	0,9258	0,0881
[3-5] (c2)	0,2466	-0,4700	0,0109	0,0873
[5-7] (c3)	0,5889	-0,0179	0,0621	0,0001
[>7] (c4)	-0,0791	1,4434	0,0011	0,8243
Distância de Hellinger				
Perfis Linha				
<i>Blends</i>	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,1548	-0,0803	0,1782	0,0430
Binária (B)	0,2317	-0,0452	0,3994	0,136
Ternária(T)	0,2383	0,3762	0,4223	0,9434
Perfis Coluna				
[<3] (c1)	0,4118	0,3040	0,4122	0,7398
[3-5] (c2)	-0,1972	0,1328	0,0901	0,1412
[5-7] (c3)	-0,0560	-0,1219	0,007	0,1189
[>7] (c4)	0,4601	-0,004	0,4904	0,0002

Fonte: Dados do autor (2018)

Em relação ao atributo acidez, os mapas perceptuais obtidos encontram-se ilustrados nas Figuras 5 e 6. Para esse atributo, nota-se que a métrica teve muita influência, em relação ao deslocamento do centroide, de forma a resultar em mapas assimétricos. Um resultado perceptível é verificado na formação de tendências entre os níveis categóricos. Assim, percebe-se, ao utilizar a métrica de Hellinger (FIGURA 6), a existência de uma relação linear entre os níveis categóricos da variável *blends* e uma tendência polinomial para os níveis da variável nota.

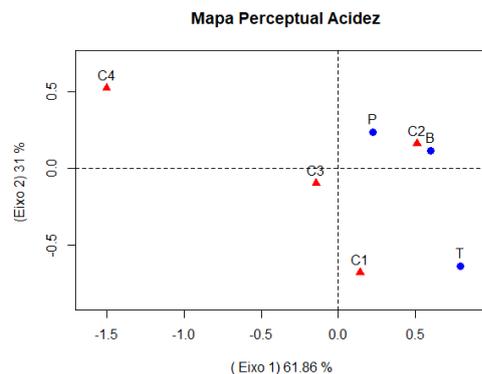
Em se tratando da similaridade entre as categorias de notas, com os níveis categóricos dos *blends*, há evidências científicas discutidas por Galli e Barbas (2004) de que a acidez corresponde à sensação de gosto primário, sendo um dos principais atributos a impactar a avaliação sensorial referente à qualidade da bebida. Neste trabalho, por se tratar de *blends*

formados por proporções de cafés arábica e canéfora, o grau de acidez das duas espécies é bem distinto.

Em geral, a bebida do café arábica é mais ácida do que a bebida do canéfora (GINZ et al., 2000). Nesse contexto, as associações entre os cafés puros (P) e os *blends* binários (B) confirmam esse efeito, quando o mapa perceptual foi construído na métrica Qui-Quadrado, ao passo que, na distância de Hellinger, dadas as tendências observadas, as interpretações das associações tornam-se duvidosas.

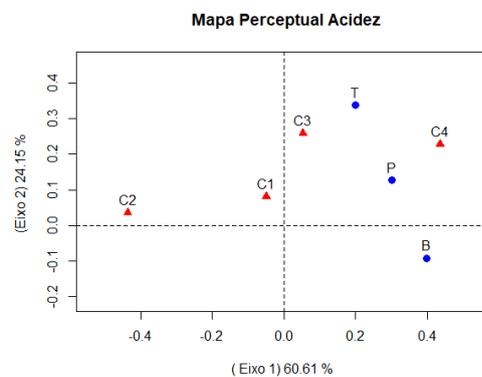
Outro fator relevante, na seleção do mapa perceptual com a métrica adequada para a interpretabilidade das associações, é corroborada pelas misturas ternárias, assim, na métrica Qui-Quadrado (FIGURA 5), percebe-se que a mistura ternária (T), por caracterizar um bebida mais ácida, está associada a categorias de notas mais inferiores c_1 , sendo essa associação indicando maior qualidade no eixo 1 (TABELA 9).

Figura 5 - Mapa perceptual para o atributo Acidez, considerando a métrica Qui-quadrado.



Fonte: Dados do autor (2018)

Figura 6 - Mapa perceptual para o atributo Acidez, considerando a métrica Hellinger.



Fonte: Dados do autor (2018)

Tabela 9 - Coordenadas e correlação com os eixos obtidos na análise de correspondência considerando as métricas Qui-Quadrado e Hellinger para o atributo acidez.

Distância Qui-Quadrado				
Perfis Linha				
<i>Blends</i>	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,2249	0,2380	0,0491	0,1190
Binária (B)	0,5970	0,1147	0,3461	0,0274
Ternária(T)	0,7892	0,6375	0,6048	0,8534
Perfis coluna				
[<3] (c1)	0,1396	-0,6787	0,0076	0,5974
[3-5] (c2)	0,5096	0,1609	0,1017	0,0336
[5-7] (c3)	-0,1433	-0,0961	0,0080	0,0120
[>7] (c4)	-1,5016	0,5247	0,8827	0,3570
Distância de Hellinger				
Perfis Linha				
<i>Blends</i>	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,3003	0,1274	0,3131	0,1169
Binária (B)	0,3978	-0,916	0,5492	0,0604
Ternária(T)	0,1992	0,3381	0,1378	0,8227
Perfis Coluna				
[<3] (c1)	-0,0485	0,0826	0,0061	0,0529
[3-5] (c2)	-0,4359	0,0356	0,4934	0,0010
[5-7] (c3)	0,0537	0,2586	0,0075	0,5261
[>7] (c4)	0,4358	0,2285	0,4930	0,4110

Fonte: Dados do autor (2018)

Em relação ao atributo amargor, as associações entre os níveis categóricos das notas e dos *blends* são ilustradas nas Figuras 7 e 8. Notoriamente, observa-se que, para esse atributo, ambas as métricas diferenciam a mistura ternária (T), associada às notas extremas c_1 e c_4 . Contudo a discussão dos resultados, no tocante ao amargor, é bastante complexa, quanto às limitações da análise de correspondência, na qual as tabelas de contingência geradas não foram estratificadas, em relação aos fatores ao tipo de processamento e /ou proporções de *blends* de arábica e canéfora, sendo este, comercialmente, denominado por conillon.

Tais informações são relevantes, na discriminação dos *blends*, em relação às avaliações sensoriais. Lima Filho et al. (2015), em um estudo anterior, avaliaram as características sensoriais e aceitação sensorial de café expresso produzido com *blends* de café arábica e conillon em proporções de até 60% de conillon na mistura final.

Neste estudo, foram verificadas maior intensidade de aroma caramelizado nas bebidas de café 100% arábica descascado e maior intensidade de gosto amargo nas bebidas, à medida que se aumentou o teor de café conillon no *blend*. Vale ressaltar que a sensação de amargor é

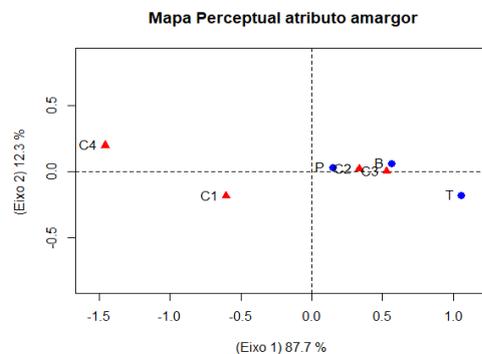
atribuída, em parte, à cafeína, e as duas espécies diferem substancialmente nesse aspecto, ou seja, o arábica é mais suave e ácido (seco), o conilon é amargo e mais marcante.

Siqueira e Abreu (2006) apontam que o pH ideal seja entre 4,95 a 5,20, implicando características sensoriais dos cafés mais aceitáveis pelo consumidor, sem excesso de amargor ou acidez.

Em função dessas argumentações, nota-se, por meio das Figuras 7 e 8, que a composição dos *blends* binários (B) apresentou características similares aos cafés puros (P), no entanto não indicaram uma maior preferência, dada a similaridade com as notas intermediárias C₂ e C₃. Tal combinação supostamente pode ser explicada pela concentração de cafeína, a qual resulta em cafés com sabor amargo, bastante característico e, segundo Monteiro e Trugo (2005), influenciam o sabor e aroma da bebida do café.

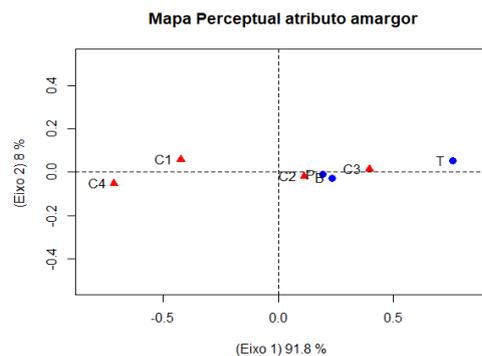
Assim, entendemos que os resultados evidenciados, para o atributo amargor, são, de certa forma, correlacionados com os resultados observados para o atributo sabor, incluindo as importâncias das variáveis categóricas em relação à representatividade dos eixos (TABELA 10).

Figura 7 - Mapa perceptual para o atributo amargor, considerando a métrica qui-quadrado.



Fonte: Dados do autor (2018)

Figura 8 - Mapa perceptual para o atributo amargor, considerando a métrica de Hellinger.



Fonte: Dados do autor (2018)

Tabela 10 - Coordenadas e correlação com os eixos obtidos na análise de correspondência considerando as métricas Qui-Quadrado e Hellinger para o atributo amargor.

Distância Qui-Quadrado				
Perfis linha				
Blends	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,1488	0,0322	0,0151	0,0280
Binária (B)	0,5644	0,0621	0,2183	0,1040
Ternária(T)	1,0576	-0,1792	0,7664	0,8678
Perfis Coluna				
[<3] (c1)	-0,6026	-0,1835	0,1261	4.632e-01
[3-5] (c2)	0,3365	0,0190	0,0393	5.013e-03
[5-7] (c3)	0,5305	0,0023	0,0977	7.382e-05
[>7] (c4)	-1,456	0,1966	0,7368	5.316e-01
Distância de Hellinger				
Blends	Coord. (eixo 1)	Coord. (eixo 2)	Cor (eixo 1)	Cor (eixo 2)
Pura (P)	0,1943	-0,0111	0,0569	0,0338
Binária (B)	0,2339	0,0281	0,0825	0,2182
Ternária(T)	0,7556	0,0520	0,8607	0,7479
Perfis Coluna				
[<3] (c1)	-0,4220	0,0576	0,2088	0,4916
[3-5] (c2)	0,1130	-0,0199	0,0150	0,0585
[5-7] (c3)	0,3955	0,0129	0,1834	0,0249
[>7] (c4)	-0,7111	0,0535	0,5928	0,4250

Fonte: Dados do autor (2018)

4.4 Índice de gravidade quanto à definição da escala a ser interpretada nos mapas perceptuais

Os resultados anteriores evidenciaram que os mapas perceptuais construídos com as métricas dadas pelas distâncias de Qui-Quadrado e Hellinger conduziram a diferentes interpretações, no tocante à identificação das similaridades dos *blends* com as classes de notas, de modo a exigir do pesquisador um conhecimento da procedência das espécies dos cafés, bem como aspectos fisiológicos e químicos que constituem cada bebida, para que a discriminação seja feita corretamente.

Decorrente desse problema, torna-se complexa a seleção de um mapa perceptual, além de escolher a métrica Qui-Quadrado ou Hellinger a ser utilizada na análise de correspondência. Com esse enfoque, mantendo o propósito de agregar informações estatísticas, que corroborem com essa escolha, sugerem-se os índices de gravidade, nos quais, em sua construção, requerem a especificação das coordenadas ($P_1, P_2, \dots, 1-P_1, P_2, \dots, P_{k-1}$) e, para a aplicação proposta no trabalho, o primeiro índice foi obtido em função de $k=3$.

Nessa perspectiva, os resultados descritos na Tabela 11 correspondem às coordenadas necessárias, para o cálculo do IG_3 , em uma escala ordinal, interpretada como uma ordem de gravidade, de tal forma que, quanto maiores os componentes envolvidos na mistura, maior a dificuldade em discriminar corretamente o *blend* em relação à atribuição das notas para cada atributo sensorial avaliado.

Tabela 11 - Resultados da classificação, em uma escala ordinal de dificuldade do provador, para avaliar incorretamente a discriminação do *blend* para cada atributo sensorial, considerando a cor(eixo 1)_x massa linha para cada *blend*.

Blend	Classificação da gravidade	Definição do ponto P	Métrica	Atributo sensorial		
				Sabor	Acidez	Amargor
Mistura Ternária (T)	Forte	P ₃	Qui-Quadrado	0,055	0,0665	0,8784
			Hellinger	0,0464	0,0151	0,9384
Mistura Binária (B)	Moderada	P ₂	Qui-Quadrado	0,0828	0,1963	0,7209
			Hellinger	0,2196	0,3020	0,4784
Mistura Pura (P)	Frac	P ₁	Qui-Quadrado	0,1122	0,0162	0,8716
			Hellinger	0,0588	0,1033	0,8379

Fonte: Dados do autor (2018)

Obs: a correlação em amargor é igual a $1 - \text{coord}(\text{sabor}) - \text{coord}(\text{acidez})$.

Seguindo o mesmo procedimento metodológico, porém, considerando as classes de notas definidas na escala ordinal, obtiveram-se as coordenadas necessárias, para a obtenção do índice IG_4 , sendo esse interpretado como índice da gravidade, no ato de atribuir erroneamente uma nota, em uma classe inferior. Os resultados encontram-se descritos na Tabela 12.

Tabela 12 - Resultados da classificação, em uma escala ordinal, referente ao ato de atribuir erroneamente uma nota, em uma classe inferior, em relação ao atributo sensorial, considerando as coordenadas obtidas por $\text{cor}(\text{eixo } 1)_X$ massa coluna para cada categoria.

Classes de Nota	Classificação da gravidade	Definição do ponto P	Métrica	Atributo sensorial		
				Sabor	Acidez	Amargor
c ₄	Muito Forte	P ₄	Qui-Quadrado	0,0001	0,0490	0,9508
			Hellinger	0,0343	0,0273	0,9383
c ₃	Forte	P ₃	Qui-Quadrado	0,0293	0,0022	0,9685
			Hellinger	0,0425	0,0020	0,955
c ₂	Moderada	P ₂	Qui-Quadrado	0,0042	0,0593	0,9365
			Hellinger	0,0027	0,2878	0,7094
c ₁	Fracá	P ₁	Qui-Quadrado	0,0914	0,0001	0,9485
			Hellinger	0,0272	0,0001	0,9727

Fonte: Dados do autor (2018)

Dados os resultados descritos nas Tabelas 11 e 12, os índices de gravidade calculados permitiram obter informações que auxiliam na escolha de uma métrica, a ser utilizada na análise de correspondência, sugerindo, naturalmente, uma interpretação que respeite a escala ordinal adotada.

Desta forma, os resultados descritos na Tabela 13 foram interpretados, de modo que quanto maior o índice, maior a evidência de que a métrica utilizada seja adequada, no sentido de que as similaridades detectadas sejam mais restritas ou próximas da escala ordinal pré-estabelecida, na construção das tabelas de frequências e, principalmente, por avaliar o risco em relação às mensurações incorretas que o avaliador poderá cometer.

Tabela 13 - Índices de gravidade obtidos, em função dos pontos P utilizados na comparação das métricas, adaptadas nas análises de correspondência para o estudo das similaridades.

Métrica	IG ₃	IG ₄
Sabor		
Qui-Quadrado	0,0964	0,0785
Hellinger	0,1562	0,0534
Acidez		
Qui-Quadrado	0,164	0,0418
Hellinger	0,166	0,1939
Amargor		
Qui-Quadrado	1,241	1,909
Hellinger	1,287	1,752

Fonte: Dados do autor (2018)

A gravidade dada em relação à escala ordinal, definida pelo número de componentes que constituem um *blend* puro (P), binário (B) e ternário (T), sintetizada em IG_3 evidenciam que, para todos os atributos, a métrica de Hellinger apresenta maior gravidade, na discriminação dos *blends*, em relação aos atributos. Esse resultado tem um maior destaque, ao comparar os mapas perceptuais para o atributo amargor (FIGURA 8) em que, praticamente, todos os *blends* ficaram localizados paralelos ao eixo 1

Em relação à interpretação do índice, dada a escala ordinal adotada na construção das classes, os resultados sintetizados IG_4 evidenciam, novamente, que a distância de Hellinger apresentou uma maior gravidade em relação ao ato de especificar incorretamente uma avaliação para os atributos sabor e acidez. Entretanto tal resultado não é verificado de forma mais expressiva, em relação ao atributo amargor. Os índices, para as métricas, foram próximos, corroborando com as mesmas características observadas nos mapas perceptuais (FIGURAS 7 e 8)

5 CONCLUSÕES

Em relação aos estudos de simulação, considerando a dependência entre os níveis categóricos, a métrica de Hellinger é recomendada por ter proporcionado um melhor desempenho na distribuição da inércia total, de forma mais equilibrada.

Em relação aos índices de gravidade obtidos quanto às escalas ordinais, representadas pela composição dos *blends* e classe de notas, sugere-se que a métrica de Hellinger apresentou resultados superiores à métrica Qui-Quadrado, que justifica avaliar o erro de mensuração a ser cometido pelo avaliador.

REFERÊNCIAS

AITCHISON, J. **The statistical analysis of compositional data**. London: Chapman and Hall, 1986. 416 p.

AITCHISON, J.; GREENACRE, M. J. Biplots of compositional data. **Journal of the Royal Statistical Society Applied Statistics**, Oxford, v. 51, n. 4, p. 375-392, Oct. 2002.

ALESSANDRINI, L. et al. Near infrared spectroscopy: an analytical tool to predict coffee roasting degree. **Chimica Acta**, Amsterdam, v. 625, n. 1, p. 95-102, Sept. 2008.

AMORIM, A. C. L. et al. Green and roasted arábica coffees differentiated by ripeness, process and cup quality via electrospray ionization mass spectrometry fingerprinting. **Journal of the Brazilian Chemical Society**, São Paulo, v. 20, n. 2, p. 313-321, 2009.

ANDRADE, D. C. T.; LOPES, D. C. O comportamento de compra do consumidor de café orgânico no sul de Minas. **Revista Agrogeoambiental**, Pouso Alegre, v. 7, n. 1, p. 75-86, mar. 2015.

BENZÉCRI, J.-P. **Correspondence analysis handbook**. 2. ed. London: Taylor & Francis, 1992. 665 p.

BORÉM, F. M. et al. Qualidade do café natural e despulpado após secagem em terreiro e com altas temperaturas. **Ciência e Agrotecnologia**, Lavras, v. 32, n. 5, p. 1609-1615, se./out. 2008.

BORÉM, F. M. Processamento do café. In: _____. (Ed.). **Pós-colheita do café**. Lavras: Ed. UFLA, 2008. p. 127-158.

BOTTAZZI, D. et al. A numerical approach for the analysis of the coffee roasting process. **Journal of Food Engineering**, Essex, v. 112, n. 3, p. 243-252, Oct. 2012.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 8, de 11 de junho de 2003. Regulamento técnico de identidade e de qualidade para a classificação do café beneficiado grão cru. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 13 jun. 2003. Seção 1, p. 22-29.

BRAZIL SPECIALTY COFFEE ASSOCIATION - BSCA. **Café**: BSCA apresenta novo mapa das origens produtoras do Brasil. Varginha: BSCA, 2014. Disponível em: <<http://bsca.com.br/noticia.phd?id=242>>. Acesso em: 24 jul. 2017.

BRAZIL SPECIALTY COFFEE ASSOCIATION - BSCA. **O que são cafés especiais**. Varginha: BSCA, 2014. Disponível em: <<http://bsca.com.br/cafés-especiais.phd>>. Acesso em: 11 jul. 2017.

CABRAL, R. S.; FRERY, A. C.; RAMIRES, J. A. Variability Analysis of complex networks measures based on stochastic distances. **Physica A: statistical mechanics and its applications**, Oxford, v. 415, p. 73-86, Dec. 2014.

CAFÉS especiais: produtores de café estão ressuscitando o bourbon amarelo. **Valor Econômico**, São Paulo, jul. 2007. Disponível em: <<http://www.revistacafeicultura.com.br/index.php?tipo=ler&mat=11657>>. Acesso em: 30 jul. 2017.

CIRILLO, M. A.; RAMOS, P. S. Goodness-of-fit tests for modified multinomial logit model. **Chilean Journal of Statistics**, Valparaíso, v. 5, n. 1, p. 73-85, Apr. 2014.

COSTA, A. L. A. **Análise de correspondência simples com novos escores e o uso da análise de correspondência múltipla em dados composicionais de granulometria de grãos de café**. 2016. 94 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2016.

COSTA, C. C.; GUILHOTO, J. J. M.; IMORI, D. Importância dos setores agroindustriais na geração de renda e emprego para a economia brasileira. **Revista de Economia e Sociologia Rural**, Brasília, v. 51, n. 4, p. 797-814, out./dez. 2013.

CUADRAS, C. M.; CUADRAS, D.; GREENACRE, M. J. A comparison of different methods for representing categorical data. **Communications in Statistics. Part B. Simulation and Computation**, New York, v. 35, n. 2, p. 447-459, 2006.

DIAS, R. **Estimação por mínima distância de hellinger**. 1988. 93 p. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas, Campinas, 1988.

DUARTE, G. S.; PEREIRA, A. A.; FARAH, A. Chlorogenic acids and other relevant compounds in Brazilian coffees processed by semi-dry and wet post-harvesting methods. **Food Chemistry**, London, v. 118, n. 3, p. 851-855, Feb. 2010.

EUGÊNIO, M. H. A. **Análise sensorial química e perfil de voláteis de cafés especiais das quatro regiões cafeeiras do Estado de Minas Gerais**. 2016. 25 p. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras, 2016.

EVREN, A.; TUNA, E. On some properties of goodness of fit measures based on statistical entropy. **International Journal of Recent Research and Applied Studies**, Essex, v. 13, n. 1, p. 142-205, 2012.

FIGUEIREDO, L. P. **Abordagem sensorial e química da expressão de genótipos de Bourbon em diferentes ambientes**. 2013. 127 p. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras, 2013.

_____. **Perfil sensorial e químico de genótipos de cafeeiro Bourbon de diferentes origens geográficas**. 2010. 81 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras, 2010.

FRANCA, A. S.; MENDONÇA, J. C. F.; OLIVEIRA, S. D. Composition of green and roasted coffees of different cup qualities. **LWT - Food Science and Technology**, London, v. 38, n. 7, p. 709-715, Nov. 2005.

GALLI, V.; BARBAS, C. Capillary electrophoresis for the analysis of short-chain organic acids in coffee. **Journal of Chromatography. A**, Amsterdam, v. 1032, n. 1/2, p. 299-304, Apr. 2004.

GINZ, M. et al. Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. **European Food Research and Technology**, Berlin, v. 211, n. 6, p. 404-410, Nov. 2000.

GIOMO, G. S.; BORÉM, F. M. Cafés especiais no Brasil: opção pela qualidade. **Informe Agropecuário**, Belo Horizonte, v. 32, n. 261, p. 7-16, mar./abr. 2011.

GONZALÉZ-RIOS, O. et al. Impacto of “ecological” post-harvest processing on coffee aroma: Roasted coffee. **Journal of Food Composition and Analysis**, San Diego, v. 20, n. 3/4, p. 297-307, May 2007.

GREENACRE, M. Correspondence analysis in medical research. **Statistical Methods in Medical Research**, Sevenoaks, v. 1, n. 1, p. 97-117, 1992.

_____. **Correspondence analysis in practice**. 2. ed. Orlando: CRC Press, 2007. 296 p.

GREENACRE, M.; BLASIUS, J. **Multiple correspondence analysis and related methods**. London: Chapman & Hall, 2006. 880 p.

GUEDES, T. A. et al. Seleção de variáveis categóricas utilizando análise de correspondência e análise de procrustes. **Acta Scientiarum**, Maringá, v. 21, n. 4, p. 861-868, 1999.

IVOGLO, M. G. et al. Divergência genética entre progênies de café robusta. **Bragantia**, Campinas, v. 67, n. 4, p. 823-831, out./dez. 2008.

JAEGER, S. R. et al. Comparison of check-all-that-apply and forced-choice yes/no question format for sensory characterisation. **Food Quality and Preference**, Barking, v. 35, p. 32-40, July 2014.

LIMA FILHO, T. et al. Composição físico-química e qualidade sensorial de café conilon produzido no Estado do Espírito Santo e submetido a diferentes formas de processamento. **Semina: ciências agrárias**, Londrina, v. 34, n. 4, p. 1723-1730, jul./ago. 2015.

_____. Perfil sensorial e aceitabilidade de bebida de café tipo expresso preparadas a partir de blends de café arábica e conilon. **Enciclopédia Biosfera**, Goiânia, v. 7, n. 12, p. 1-17, 2011.

LINGLE, T. R. **The coffee cupper's handbook: systematic guide to the sensory evaluation of coffee's flavor**. 7. ed. Long Beach: Specialty Coffee Association of American, 2011. 66 p.

LOPES, D. C.; ANDRADE, D. C. T. O comportamento de compra do consumidor de café orgânico no sul de Minas. **Revista Agrogeoambiental**, Pouso Alegre, v. 7, n. 1, p. 75-86, mar. 2015.

MAARSE, H.; VISSCHER, C. A. **Volatiles compounds in foods: quantitative and qualitative data**. 7. ed. The Netherlands: TNO Nutrition and Food Research Institute, 1996. 2350 p.

MALTA, M. R.; CHAGAS, S. J. R. M. Processamento e qualidade do café produzido na região do cerrado mineiro. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 6., 2009, Vitória. **Resumos...** Brasília: Embrapa Café, 2010. p. 1-5.

- MENDES, L. C. **Estudos para determinação das melhores formulações de blends de café arábica (*C. arabica*) com café robusta (*C. canephora* Conilon) para uso no setor de cafés torrados e moídos e de cafés espresso.** 2005. 186 p. Tese (Doutorado em Tecnologia de Alimentos) – Universidade Estadual de Campinas, Campinas, 2005.
- MEYNERS, M.; CASTURA, J. C; CARR, B. T. Existing and new approaches for the analysis of CATA data. **Food Quality and Preference**, Barking, v. 30, n. 2, p. 309-319, Dec. 2013.
- MONTEIRO, M. A. M. et al. Perfil sensorial da bebida café (*Coffea arabica* L.) determinado por análise tempo-intensidade. **Ciência e Tecnologia de Alimentos**, Campinas, v. 25, n. 4, p. 772-780, out./dez. 2005.
- MONTEIRO, M. C.; TRUGO, L. C. Determinação de compostos bioativos em amostras comerciais de café torrado. **Química Nova**, São Paulo, v. 28, n. 4, p. 637-641, jul./ago. 2005.
- NAITO, S. D. N. P. **Análise de correspondência generalizada.** 2007. 119 p. Dissertação (Mestrado em Bioestatística) - Universidade de Lisboa, Lisboa, 2007.
- NAKANO, E. Y. **Soluções bayesianas para alguns problemas clássicos com dados discretos.** 2010. 111 p. Tese (Doutorado em Ciências) - Universidade de São Paulo, São Paulo, 2010.
- NAVARINI, L.; RIVETTI, D. Water quality for Espresso coffee. **Food Chemistry**, London, v. 122, n. 2, p. 424-428, Sept. 2010.
- PEZZOPANE, J. R. M. et al. Escala para avaliação de estágios fenológicos do cafeeiro arábica. **Bragantina**, Campinas, v. 62, n. 3, p. 499-505, 2003.
- PIMENTA, C. J. **Qualidade do café.** 3. ed. Lavras: Ed. UFLA, 2003. 304 p.
- PINTO, N. A. V. D. et al. Avaliação de componentes químicos de padrões de bebida para preparo do café expresso. **Ciência e Agrotecnologia**, Lavras, v. 26, n. 4, p. 826-829, jul./ago. 2002.
- PRINCIPE, J. C. **Information theoretic learning: renyi's entropy and kernel perspectives.** New York: Springer-Verlag New York, 2010. 448 p.
- RAO, C. R. A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. **Questiío**, Barcelona, v. 19, n. 1/3, p. 23-63, 1995.
- RIBEIRO, B. B. et al. Avaliação química e sensorial de Blends de *Coffea canephora* Pierre e *Coffea arabica* L. **Coffee Science**, Lavras, v. 9, n. 2, p. 178-186, abr./jun. 2014.
- RIBEIRO, D. E. **Interação genótipo e ambiente na composição química e qualidade sensorial de cafés especiais em diferentes formas de processamento.** 2013. 62 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal de Lavras, Lavras, 2013.

RODARTE, M. P. **Análise sensorial, química e perfil de constituintes voláteis de cafés especiais**. 2008. 147 p. Tese (Doutorado em Ciência dos Alimentos) -Universidade Federal de Lavras, Lavras, 2008.

SAES, A. M. Do vinho ao café; aspectos sobre a política de diferenciação. **Informações Econômicas**, Piracicaba, v. 36, n. 2, p. 7-19, fev. 2006.

SANTOS, E. N. F. **Álgebra intervalar na análise de correspondência**: um estudo de caso em testes de aceitação sensorial com erros de medidas. 2012. 101 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) Universidade Federal de Lavras, Lavras, 2012.

SIQUEIRA, H. H.; ABREU, C. M. P. Composição físico-química e qualidade do café submetido a dois tipos de torração e com diferentes formas de processamento. **Ciência e Agrotecnologia**, Lavras, v. 30, n. 1, p. 112-117, jan./fev. 2006.

SOUZA, M. C. M.; SAES, M. S. M.; OTANI, M. N. Pequenos agricultores familiares e sua inserção no mercado de cafés especiais; uma abordagem preliminar. **Informações Econômicas**, São Paulo, v. 32, n. 11, p. 16-26, nov. 2002.

SPECIALTY COFFEE ASSOCIATION OF AMERICA - SCAA. **Specialty coffe facts & figures**. New York: SCAA, 2012. 2 p.

STEERNEMAN, T. On the total variation and Hellinger distance between signed measures; an application to product measures. **Proceedings of the American Mathematical Society**, Providence, v. 88, n. 4, p. 684-688, Aug. 1983.

TALLIS, G. M. The use of a generalized multinomial distribution in the estimation of correlation in discrete data. **Journal of the Royal Statistical Society. Series B**, Oxford, v. 24, n. 2, p. 530-534, 1962.

TAUNAY, A. de E. **História do café no Brasil**. Rio de Janeiro: Departamento Nacional do Café, 1939. v. 3, 474 p.

TOLEDO, J. L. B. de; BARBOSA, A. T. **Classificação e degustação do café**. Rio de Janeiro: SEBRAE, 1998. 91 p.

VILAS-BOAS, B. M. et al. Seleção de extratos e tempo de extração para a determinação de açúcares em café torrado. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 5, p. 1169-1173, set./out. 2001.

ANEXO A – DECOMPOSIÇÃO DE VALORES SINGULARES ENVOLVIDA NA RESTITUIÇÃO DA VARIAÇÃO AMOSTRAL EXPLICADA PELOS EIXOS

Demonstração de um resultado feito por Dias (1988) em que o autor indicou alguns passos e, por meio disso, desenvolvemos algebricamente os produtos notáveis para a diferença de quadrados na obtenção do resultado.

Suponha que F seja uma função de distribuição discreta com probabilidade dada pelo vetor $P = (P_1, P_2, \dots, P_J)$.

A distribuição empírica obtida em J observações é representada por

$$S_n = \left(\frac{a_{1j}}{a_{j+}}, \frac{a_{2j}}{a_{j+}}, \dots, \frac{a_{jJ}}{a_{j+}} \right)$$

A distância de Hellinger por definição

$$\begin{aligned} d_H^2(F, S_n) &= \sum_{j=1}^J \left(\sqrt{\frac{a_j}{a_{j+}}} - \sqrt{p_j} \right)^2 = \sum_{j=1}^J \left(\frac{a_j}{a_{j+}} - 2\sqrt{\frac{a_j}{a_{j+}}} \cdot \sqrt{p_j} + p_j \right) \\ &= 2 - 2 \sum_{j=1}^J \sqrt{\frac{a_j}{a_{j+}}} \cdot \sqrt{p_j} = 2 \left(1 - \sum_{j=1}^J \sqrt{\frac{a_j}{a_{j+}}} \cdot \sqrt{p_j} \right) \end{aligned}$$

Temos que

$$\begin{aligned} \left(\sqrt{\frac{a_j}{a_{j+}}} - \sqrt{p_j} \right) &= \left(\frac{\sqrt{a_j} - \sqrt{a_{j+}} \sqrt{p_j}}{\sqrt{a_{j+}}} \right) \cdot \left(\frac{\sqrt{a_j} + \sqrt{a_{j+}} \sqrt{p_j}}{\sqrt{a_j} + \sqrt{a_{j+}} \sqrt{p_j}} \right) \\ &= \frac{(a_j - a_{j+} p_j)}{\sqrt{a_{j+}}} \cdot \frac{1}{(\sqrt{a_j} + \sqrt{a_{j+}} \sqrt{p_j})} \cdot \frac{(\sqrt{a_j} \sqrt{p_j})}{(\sqrt{a_{j+}} \sqrt{p_j})} \\ &= \frac{(a_j - a_{j+} p_j) \cdot (\sqrt{a_j} \sqrt{p_j})}{(\sqrt{a_j} \cdot \sqrt{a_{j+}} \sqrt{p_j}) \cdot (\sqrt{a_j} + \sqrt{a_{j+}} \sqrt{p_j})} \\ &= \frac{(a_j - a_{j+} p_j)}{(\sqrt{a_{j+}} \sqrt{p_j})} \cdot \left(\frac{\sqrt{a_j} + \sqrt{a_{j+}} \sqrt{p_j}}{\sqrt{a_j} \sqrt{p_j}} \right)^{-1} \\ &= \frac{(a_j - a_{j+} p_j)}{(\sqrt{a_{j+}} \sqrt{p_j})} \cdot \left(\frac{\sqrt{a_j}}{\sqrt{a_{j+}} \sqrt{p_j}} + \frac{\sqrt{a_{j+}} \sqrt{p_j}}{\sqrt{a_{j+}} \sqrt{p_j}} \right)^{-1} \\ &= \frac{(a_j - a_{j+} p_j)}{(\sqrt{a_{j+}} \sqrt{p_j})} \cdot \left(\frac{\sqrt{a_j}}{\sqrt{a_{j+}} \sqrt{p_j}} + 1 \right)^{-1} \end{aligned}$$

$$\text{Portanto } \left(\frac{\sqrt{a_j}}{a_{j+}} - \sqrt{p_j} \right)^2 = \frac{(a_j - a_{j+} p_j)^2}{a_{j+}^2 p_j} \cdot \left(\frac{\sqrt{a_j}}{\sqrt{a_{j+}} \sqrt{p_{j+}}} + 1 \right)^{-2}$$

$$\text{Suponha } \chi^2 = \frac{(a_j - a_{j+} p_j)^2}{(a_{j+} p_j)}$$

$$\text{Logo } \left(\frac{\sqrt{a_j}}{\sqrt{a_{j+}}} - \sqrt{p_j} \right)^2 = \frac{\chi^2}{a_{j+}} \cdot \left(\frac{\sqrt{a_j}}{\sqrt{a_{j+}} \sqrt{p_j}} + 1 \right)^{-2}$$

$$\text{Ou seja, } d_H^2(F, S_n) = \frac{\chi^2}{a_{j+}} = \frac{\chi^2}{n}$$

Obs:

$$\text{Note que } \sum_{j=1}^J \frac{a_j}{a_{j+}} + \sum_{j=1}^J p_j = 1 + 1 = 2$$

**ANEXO B - DEMONSTRAÇÃO DA LIMITAÇÃO DA DISTÂNCIA DE HELLINGER
POR UMA PROPRIEDADE QUE DEFINE SUA AMPLITUDE**

Definindo os vetores $P = \left(\frac{a_1}{a_{i+}}, \dots, \frac{a_1}{a_{i+}} \right)$ e $Q = \left(\frac{a_1}{a_{j+}}, \dots, \frac{a_j}{a_{j+}} \right)$

$$d_H^2(P, Q) \leq \Delta(P, Q) \leq \sqrt{d_H^2(P, Q)(2 - d_H^2(P, Q))} \leq \sqrt{2}d_H(P, Q)$$

$$|\sqrt{p_i} - \sqrt{q_i}| \leq |\sqrt{p_i} + \sqrt{q_i}|$$

$$|\sqrt{p_i} - \sqrt{q_i}| |\sqrt{p_i} - \sqrt{q_i}| \leq |\sqrt{p_i} + \sqrt{q_i}| |\sqrt{p_i} - \sqrt{q_i}|$$

$$(\sqrt{p_i} - \sqrt{q_i})^2 \leq |p_i - q_i|$$

$$d_H^2(P, Q) = \frac{1}{2} \sum_i (\sqrt{p_i} - \sqrt{q_i})^2 \leq \frac{1}{2} \sum_i |\sqrt{p_i} - \sqrt{q_i}| |\sqrt{p_i} - \sqrt{q_i}|$$

$$\leq \frac{1}{2} \sum_i |p_i - q_i| = \Delta(P, Q)$$

$$\text{Logo } \Delta^2(P, Q) = \left(\frac{1}{2} \sum_i |p_i - q_i| \right)^2 = \frac{1}{4} \left(\sum_i |\sqrt{p_i} - \sqrt{q_i}| |\sqrt{p_i} + \sqrt{q_i}| \right)^2$$

Usando a desigualdade de Cauchy Schwarz

$$\Delta^2(P, Q) \leq \left(\frac{1}{2} \sum_i |p_i - q_i|^2 \right) \cdot \frac{1}{2} \left(\sum_i |\sqrt{p_i} + \sqrt{q_i}|^2 \right)$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot \frac{1}{2} \left(\sum_i (p_i + 2\sqrt{p_i} \cdot \sqrt{q_i} + q_i) \right)$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot \frac{1}{2} \left(\sum_i p_i + \sum_i q_i + 2 \sum_i \sqrt{p_i} \cdot \sqrt{q_i} \right)$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot \left[\frac{1}{2} \left(2 + 2 \sum_i \sqrt{p_i} \cdot \sqrt{q_i} \right) \right]$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot \left[\frac{1}{2} (2 + 2(1 - d_H^2(P, Q))) \right]$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot 2 \left[\frac{1}{2} (1 + 1(1 - d_H^2(P, Q))) \right]$$

$$\Delta^2(P, Q) \leq d_H^2(P, Q) \cdot [(2 - d_H^2(P, Q))]$$

$$\Delta(P, Q) \leq \sqrt{d_H^2(P, Q) \cdot [(2 - d_H^2(P, Q))]}$$

Pela seguinte relação $d_H^2(P, Q) = 1 - F(P, Q)$ temos que

$$-d_H^2(P, Q) = -1 + F(P, Q)$$

$$2 - d_H^2(P, Q) = 2 - 1 + F(P, Q)$$

$$2 - d_H^2(P, Q) = 1 + F(P, Q)$$

$$2 - d_H^2(P, Q) = 1 + 1$$

$$2 - d_H^2(P, Q) = 2$$

Logo

$$\Delta(P, Q) \leq \sqrt{d_H^2(P, Q)2} \leq \sqrt{2}d_H$$

**ANEXO C - TABELAS DE FREQUÊNCIA UTILIZADA NA REALIZAÇÃO DAS
ANÁLISES DE CORRESPONDÊNCIA**

Tabela 1 - Frequência de notas para o atributo sabor.

Blends	C ₁	C ₂	C ₃	C ₄	Totais	Massa(Linha)
	[<3]	[3-5]	[5-7]	[>7]		
Pura (P)	1	6	2	3	12	0,33
Binária (B)	0	7	13	0	20	0,55
Ternária (T)	1	1	2	0	4	0,11
Totais	2	14	17	3	36	
Massa(Coluna)	0,05	0,38	0,47	0,08		

Tabela 2 - Frequência de notas para o atributo acidez.

Blends	C ₁	C ₂	C ₃	C ₄	Totais	Massa(Linha)
	[<3]	[3-5]	[5-7]	[>7]		
Pura (P)	1	8	1	2	12	0,33
Binária (B)	2	11	7	0	20	0,55
Ternária (T)	0	2	2	0	4	0,11
Totais	3	21	10	2	36	
Massa(Coluna)	0,08	0,58	0,27	0,05		

Tabela 3 - Frequência de notas para o atributo amargor.

Blends	C ₁	C ₂	C ₃	C ₄	Totais	Massa(Linha)
	[<3]	[3-5]	[5-7]	[>7]		
Pura (P)	3	2	5	2	12	0,33
Binária (B)	1	13	6	0	20	0,55
Ternária (T)	0	3	1	0	4	0,11
Totais	4	18	12	2	36	
Massa(Coluna)	0,11	0,5	0,33	0,05		

ANEXO D – ROTINA UTILIZADA NO SOFTWARE R PARA A OBTENÇÃO DOS RESULTADOS DE SIMULAÇÃO

```

rm(list=ls(all=T)) # limpa a memória do sistema
dados <- read.table ( )
dados
# ##### Marginais ##### #
tcol = apply(dados,2,sum) # total coluna #
tcol
tlin = apply(dados,1,sum) # total linha #
tlin
# ##### Matriz dos perfis ##### #
# ##### Perfil Coluna ##### #
Pcol = matrix(0, nrow(dados),ncol(dados))
for (j in 1:length(tcol)) {
  auxcol = tcol[j]
  for (i in 1:length(tlin)) {
    Pcol[i,j] = dados[i,j]*(1/auxcol)
  }
}
Pcol
# ##### Perfil linha ##### #
Plin = matrix(0, nrow(dados),ncol(dados))
for (i in 1:length(tlin)) {
  auxlin = tlin[i]
  for (j in 1:length(tcol)) {
    Plin[i,j] = dados[i,j]*(1/auxlin)
  }
}
Plin
# #####
P = as.matrix(dados*1/sum(dados)) # ## Prop ## #
plin = apply(P,1,sum) # total linha #
pcol = apply(P,2,sum) # total coluna #
# ##### Aplicando a correção qui-quadrado ##### #
Q = matrix(0,nrow(dados),ncol(dados))
for (i in 1:length(plin)) {

```

```

for (j in 1:length(pcol)) {
  Q[i,j] = (P[i,j]-(plin[i]*pcol[j])/sqrt(plin[i]*pcol[j]))}
# ##### Análise de correspondência Generalizada ##### #
alfa = 1; beta = 0
plin = as.matrix(plin)
plin_vector <- as.vector(plin)
pcol = as.matrix(pcol)
pcol_vector = as.vector(pcol)
T1 = diag(sqrt(plin_vector))%*%diag(rep(1,nrow(dados))) - (1-
beta)*(diag(rep(nrow(dados),1))%*%matrix(1,nrow(dados),1)%*%t(plin))
T2 = (1/(alfa*(1+beta)))*abs((diag(1/plin_vector)%*%Q%*%diag(1/pcol_vector)))^alfa
T3 = matrix(1,nrow(dados),1)%*%t(matrix(1,ncol(dados),1))*beta
T4 = t(diag(rep(1,ncol(dados)))-(1-beta)*(matrix(1,ncol(dados),1)%*%t(pcol))%*%
diag(pcol_vector)
aux1 = T1%*%(T2-T3)
ACG = aux1%*%T4
# ##### Aplicação da decomposição de valores singulares ##### #
S <- svd(ACG)
av = S$d
# ##### Coordenadas principais de linhas ##### #
U = as.matrix(S$u)
A = diag(1/sqrt(plin_vector))%*%U%*%diag(S$d)
A = A[,1:2]
##### correlação dos perfis linhas com cada eixo #####
d1A = sum(A[,1]^2) ; d2A = sum(A[,2]^2)
Acor_1 = as.matrix((A[,1]^2)/d1A) ; Acor_2 = as.matrix((A[,2]^2)/d2A)
##### Coordenadas principais de colunas ##### #
V = as.matrix(S$v)
B = diag(1/sqrt(pcol_vector))%*%V%*%diag(S$d)
B = B[,1:2]
##### correlação dos perfis linhas coluna de cada eixo ##### #
d1B = sum(B[,1]^2) ; d2B = sum(B[,2]^2)
Bcor_1 = (B[,1]^2)/d1B ; Bcor_2 = (B[,2]^2)/d2B
##### organização dos dados ##### #
RES_linha <- cbind(A,Acor_1,Acor_2)
colnames (RES_linha) <- c("coord_1","coord_2","cont_rel_1","cont_rel_2")
colnames(RES_linha)

```

```

RES_col <- cbind(B,Bcor_1,Bcor_2)
colnames (RES_col) <- c("coord_1","coord_2","cont_rel_1","cont_rel_2")
colnames(RES_col)
# ##### Resultados finais ##### #
Av ; RES_linha ; RES_col

# ##### Construção dos gráficos INÍCIO - Graficos ##### #
Data.row = RES_linha[,1:2] ; Data.col = RES_col[,c(1,2)]
Color = T
Title = "Titulo"
Names.Axis.X = "Eixo X" ##### paste("Eixo X", round( ), "%", step="")### nao ler####
Names.Axis.Y = "Eixo Y"
Names.Label = c(rownames(dados), colnames(dados))
# Names.Label = c("A", "30", rep("2",5))
Data = rbind(RES_linha[,1:2],RES_col[,c(1,2)])
Coord.X = c(min(Data[,1])-0.1, max(Data[,1])+0.1)
Coord.Y = c(min(Data[,2])-0.1, max(Data[,2])+0.1)
plot(0,0, # cria grafico para as coordenadas principais das linhas
     xlab = Names.Axis.X, # Nomeia Eixo X
     ylab = Names.Axis.Y, # Nomeia Eixo Y
     main = Title, # Titulo
     asp = 1, # Aspecto do Grafico
     # pch = 17, # Formato dos pontos
     cex = 0, # Tamanho dos pontos
     xlim = Coord.X, # Dimensao para as linhas do grafico
     ylim = Coord.Y # Dimensao para as colunas do grafico
     # col = ifelse(Color,"red","black")) # Cor dos pontos
abline(h = 0, v=0, cex = 1.5, lty=2) # cria o eixo central

## Pontos linhas
Points (Data.row, pch = 16, cex = 1.2, col = ifelse(Color,"blue","black")) # adiciona ao grafico as
coordenadas principais das colunas
## Pontos colunas
Points (Data.col, pch = 17, cex = 1.2, col = ifelse(Color,"red","black")) # adiciona ao grafico as
coordenadas principais das colunas
## Nomes da coordenadas
Text (Data, cex=1, pos=3, Names.Label) # Coloca os nomes dos pontos das coordenadas principais
das linhas

```