



HERBERT STEIN PEREIRA TORRES SANTOS

**ANÁLISE DE CORRESPONDÊNCIA CANÔNICA NÃO
LINEAR COM ÊNFASE NA DESCRIÇÃO DA REDUNDÂNCIA
DA VARIABILIDADE DE DADOS SENSORIAIS DE BLENDS
DE CAFÉS COM DIFERENTES VARIEDADES.**

LAVRAS – MG

2022

HERBERT STEIN PEREIRA TORRES SANTOS

**ANÁLISE DE CORRESPONDÊNCIA CANÔNICA NÃO LINEAR COM ÊNFASE NA
DESCRIÇÃO DA REDUNDÂNCIA DA VARIABILIDADE DE DADOS SENSORIAIS
DE BLENDS DE CAFÉS COM DIFERENTES VARIEDADES.**

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 Multivariada, para a obtenção do título de Mestre.

Prof. DSc. Marcelo Ângelo Cirillo
Orientador

**LAVRAS – MG
2022**

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

Santos, Herbert Stein Pereira Torres

Análise de correspondência canônica não linear com ênfase na descrição da redundância da variabilidade de dados sensoriais de blends de cafés com diferentes variedades. / Herbert Stein Pereira Torres Santos. – Lavras : UFLA, 2022. 79 p. : il.

Dissertação (mestrado acadêmico)–Universidade Federal de Lavras, 2022.

Orientador: Prof. DSc. Marcelo Ângelo Cirillo.

Bibliografia.

1. Cafés especiais. 2. Café comercial. 3. Regressão polinomial multivariada. I. Cirillo, Marcelo Ângelo. II. Título.

HERBERT STEIN PEREIRA TORRES SANTOS

**ANÁLISE DE CORRESPONDÊNCIA CANÔNICA NÃO LINEAR COM ÊNFASE NA
DESCRIÇÃO DA REDUNDÂNCIA DA VARIABILIDADE DE DADOS SENSORIAIS
DE BLENDS DE CAFÉS COM DIFERENTES VARIEDADES.**

**NONLINEAR CANONICAL CORRESPONDENCE ANALYSIS WITH EMPHASIS
ON THE DESCRIPTION OF THE REDUNDANCY OF THE VARIABILITY OF
SENSORY DATA OF COFFEE BLENDS WITH DIFFERENT VARIETIES.**

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 Multivariada, para a obtenção do título de Mestre.

APROVADA em 25 de Abril de 2022.

Prof. DSc. Jackelya Araújo da Silva	UFPI
Prof. DSc. Augusto Maciel da Silva	UFMS
Prof. DSc. Evelise Roman Corbalan Gois Freire	UFLA

Prof. DSc. Marcelo Ângelo Cirillo
Orientador

**LAVRAS – MG
2022**

AGRADECIMENTOS

Agradeço primeiramente a Deus por me permitir viver esse momento e por me conceder determinação e perseverança para obter essa conquista.

Agradeço, também, a minha família, por seus esforços e incentivos a minha educação e contribuição no meu caráter.

Agradeço a pessoa a qual decidi compartilhar a vida, por sua cumplicidade, apoio e confiança. Obrigado Talyta.

Agradeço ao meu orientador Marcelo, que topou orientar-me nesse projeto. Por toda sua dedicação, compreensão e sabedoria, sempre me encorajando nos momentos mais árduos. Obrigado professor.

A Universidade Federal de Lavras em especial ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária e seus colaboradores, professores e coordenadores.

Agradecer a CNPq pelo incentivo a pesquisa científica através da bolsa que permitiu a chegada desse momento.

Aos amigos de longa data, assim como os conquistados durante essa caminhada por compartilharem comigo um pedacinho das suas vidas.

Aos professores Jackelya, Augusto e Evelise que aceitaram fazer parte da banca examinadora e por todas suas contribuições para este trabalho. A cada pessoa que torceu ou participou de alguma forma para este momento.

MUITO OBRIGADO!

RESUMO

A formulação de blends de café é de suma importância para a indústria cafeeira, pois proporciona ao produto final expressiva capacidade de competição de mercado e agrega atributos sensoriais que complementam a experiência de consumo. Por meio da análise de redundância e da análise de correspondência canônica é possível estudar as relações entre um conjunto de notas sensoriais e um conjunto de blends com diferentes proporções de variedade de café através de modelos de regressão linear multivariados. Contudo, não é realista assumir que tais respostas sensoriais são dadas de forma linear em relação à formulação dos blends, visto que algumas espécies de café têm maior peso na avaliação sensorial (termos quadráticos) e deve-se considerar o efeito das misturas (termo de interação). Com esta motivação, este trabalho tem como objetivo propor o uso da análise de redundância e análise de correspondência não lineares através da regressão polinomial multivariada para avaliar a aceitação de diferentes variedades de blends de café. Os blends foram formulados a partir de proporções de variedades de cafés especiais da espécie Arábica, Bourbon amarelo e Acaíá, café Conilon e de uma marca comercial de café torrado. Os blends foram avaliados recebendo notas que variaram de 0 a 10 para as características qualitativas da bebida: sabor, amargor, acidez, corpo e nota final. Resultados finais mostram que houve ganhos no percentual de variância explicada total nos modelos polinomiais em relação aos modelos clássicos.

Palavras-chave: Cafés especiais. Café comercial. Regressão polinomial multivariada. Blends. Redução de dimensionalidade. Avaliadores.

ABSTRACT

The formulation of coffee blends is of paramount importance for the coffee industry, as it provides the final product with an expressive ability to compete in the market and adds sensory attributes that complement the consumption experience. Through redundancy analysis and canonical correspondence analysis, it is possible to study the relationships between a set of sensory notes and a set of blends with different proportions of coffee variety through multivariate linear regression models. However, it is unrealistic to assume that such sensory responses are given in a linear fashion in relation to the formulation of blends, since some types of coffee have greater weight in the sensory evaluation (quadratic terms) and the effect of blends must be considered (term of interaction). With this motivation, this work aims to propose the use of redundancy analysis and nonlinear correspondence analysis through multivariate polynomial regression to evaluate the acceptance of different varieties of coffee blends. The blends were formulated from proportions of specialty coffee varieties such as Arabica, Yellow Bourbon and Acaiá, Conilon coffee and a commercial brand of roasted coffee. The blends were evaluated receiving scores that ranged from 0 to 10 for the qualitative characteristics of the drink: flavor, bitterness, acidity, body and final note. The results showed significant gains in the percentage of total explained variance in the nonlinear models in relation to the linear ones.

Keywords: Specialty coffees. Commercial coffee. Multivariate polynomial regression. Blends. Dimensionality reduction. Appraisers.

LISTA DE FIGURAS

Figura 2.1 – Fluxograma do algoritmo de análise de redundância	16
Figura 2.2 – Interpretação básica de uma representação biplot.	24
Figura 3.1 – Algoritmo de Regressão Polinomial	37
Figura 4.1 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de análise de redundância para os dados padronizados.	44
Figura 4.2 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de análise de redundância para os dados composicionais transformados pelo método isometric log-ratio.	47
Figura 4.3 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de de análise de correspondência canônica para os dados composicionais transformados pelo método qui-quadrado.	51
Figura 4.4 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de de análise de correspondência canônica para os dados composicionais transformados pelo método centered log-ratio.	54

LISTA DE TABELAS

Tabela 3.1 – Caracterização dos tratamentos utilizados na composição dos blends processados via seca.	35
Tabela 3.2 – Composição dos blends formados pelos cafés Bourbon (CEB), Comercial torrado (CT), Conilon (CC) e Acaíá (CEA)	36
Tabela 4.1 – Principais métricas de avaliação para os modelos de análise de redundância e suas transformações.	41
Tabela 4.2 – Resultados da Análise de Redundância para os dados padronizados.	43
Tabela 4.3 – Resultados da Análise de Redundância para os dados composicionais transformados pelo método isometric log-ratio.	46
Tabela 4.4 – Principais métricas de avaliação para os modelos de análise de correspondência canônica e suas transformações.	49
Tabela 4.5 – Resultados da análise de correspondência canônica para os dados transformados pelo método qui-quadrado.	50
Tabela 4.6 – Resultados da análise de correspondência canônica para os dados composicionais transformados pelo método centered log-ratio.	53

LISTA DE QUADROS

Quadro 2.1 – Resumo dos dois tipos de representação de biplots para análise de redundâncias de acordo com Ter Braak (1994).	25
Quadro 2.2 – Resumo dos dois tipos de representação de biplots para análise de correspondência canônica de acordo com Ter Braak (1994).	26

SUMÁRIO

1	INTRODUÇÃO	12
2	REFERENCIAL TEÓRICO	15
2.1	ANÁLISE DE REDUNDÂNCIA (RDA)	15
2.2	ANÁLISE DE CORRESPONDÊNCIA CANÔNICA (CCA)	20
2.2.1	Regressão Ponderada	20
2.3	Biplot	21
2.3.1	Biplot na análise de redundância	25
2.3.2	Biplot na análise de correspondência canônica	26
2.4	TESTE DE PERMUTAÇÃO	26
2.5	ANÁLISE COMPOSICIONAL	28
2.5.1	Espaço amostral simplex	28
2.5.2	Operação de fechamento	29
2.5.3	Operação perturbação	29
2.5.4	Operação Potência	30
2.5.5	Produto interno, norma e distância	30
2.5.6	Transformações de razões logarítmicas	31
2.5.6.1	Razão Logarítmica Aditiva (alr)	31
2.5.6.2	Razão Logarítmica Centrada (clr)	32
2.5.6.3	Razão Logarítmica Isométrica (ilr)	33
3	MATERIAIS E MÉTODOS	34
3.1	Materiais	34
3.2	Algoritmo de Regressão Polinomial	37
3.2.1	Teste de permutação	40
3.2.2	Recursos Computacionais	40
4	RESULTADOS	41
4.1	Análise de Redundância	41
4.1.1	Dados Padronizados	43

4.1.2	Transformação Isometric Log-Ratio	46
4.2	Análise de Correspondência Canônica	49
4.2.1	Transformação Qui-quadrado	50
4.2.2	Transformação Centered Log-Ratio	53
5	DISCUSSÃO	56
6	CONCLUSÃO	59
	REFERÊNCIAS	60
	APENDICE A – Resultados Análise de Redundância	64
	APENDICE B – Resultados Análise de Correspondência Canônica	72

1 INTRODUÇÃO

O café figura como uma das principais *commodities* brasileiras, ou seja, é considerado uma matéria-prima pouco industrializada, produzida em larga escala que influencia o comportamento de determinados setores econômicos ou até da economia como um todo. Contudo, a forma de produção e consumo da bebida vêm sendo redefinidos conforme a "Terceira onda do café", descrita por Guimarães (2016) e caracterizada pelo mercado de cafés especiais voltada à autonomia dos cafeicultores e maior sustentabilidade social, ambiental e econômica da atividade, onde pequenas empresas (cafeterias independentes, também chamadas de coffee Bars e internet) fornecem serviços adotando uma abordagem científica e de experimentação acerca do produto e os consumidores (jovens em sua maioria) buscam novas experiências de consumo excepcional.

A prática de misturar diferentes variedades de café é de suma importância, pois agrega ao produto atributos sensoriais que complementam essa experiência de consumo. Do ponto de vista comercial, a elaboração de blends proporciona ao produto final uma expressiva capacidade de competição no mercado, tendo em vista o maior rendimento industrial e pelos menores preços médios em sua comercialização (RIBEIRO et al., 2014; IVOGLO et al., 2008).

Nesse contexto, diversas pesquisas foram realizadas para investigar a qualidade e aceitação dos consumidores em relação a formulação desses blends. Portanto, experimentos sensoriais e testes de aceitação são amplamente realizados para avaliar a qualidade, o grau de aceitação e a preferência por uma nova mistura específica (CIRILLO et al. 2019). Costa et al. (2018) sugerem que a qualidade do café depende de muitos fatores, que vão desde a escolha da espécie, até a variedade da cultura e método de preparo do café.

Variados métodos estatísticos já foram utilizados e formulados com esse propósito. No que se refere a modelos estatísticos, Ribeiro et al. (2014) utilizaram análise de variância e análise de regressão para avaliar blends com diferentes proporções das espécies *C. canephora* e *C. arabica* e observaram diferenças significativas nas variáveis de análises químicas e atributos sensoriais dos blends com maiores proporções de *C. canephora*. Ainda, Paulino et al. (2019) propuseram um modelo misto por mínimos quadrados para avaliar dados sensoriais de quatro

experimentos de blends com diferentes padrões de qualidade e concluíram que a inclusão dos parâmetros aleatórios no modelo, representados pelos experimentos permitiu comparar o efeito de cada componente, simultaneamente a todos os experimentos.

Métodos exploratórios multivariados são apropriados quando se deseja sintetizar a estrutura de variabilidade dos dados nos quais várias variáveis são estudadas simultaneamente. Como exemplo dessas técnicas, Costa, Brighenti e Cirillo (2018) utilizaram uma nova abordagem para o cálculo das coordenadas na análise de correspondência através da incorporação de resíduos aplicado em dados de grãos de café.

Da Costa et al. (2020) propuseram um índice de seleção métrica que permitiu agregar uma estatística que justifique a métrica mais adequada na análise de correspondência, evitando a subjetividade nas interpretações das similaridades entre os tipos de blends e classe de notas.

A análise de redundância (RDA) e a análise de correspondência canônica (CCA) são técnicas que combinam modelos de regressão linear e métodos de redução de dimensionalidade característicos das análises multivariadas. Diferente da análise de correspondência simples que estudam diversas variáveis em seus níveis categóricos, essas técnicas permitem avaliar um conjunto de variáveis respostas em detrimento de um conjunto de variáveis explicativas. Como vantagens, essas técnicas permitem avaliar variáveis expressas em valores contínuos (embora a análise de correspondência canônica seja geralmente recomendada em casos de variáveis de contagem e valores relativos), além de permitir testar formalmente hipóteses estatísticas sobre a significância das relações entre os conjuntos de dados.

No entanto, apesar das vantagens citadas em relação as técnicas anteriormente utilizadas nos experimentos sensoriais, tanto a análise de redundância como a análise de correspondência canônica, assumem linearidade nas relações entre os grupos de variáveis, o que pode não representar a realidade nas respostas sensoriais dos avaliadores em relação a prova dos blends. Como alternativa, Makarenkov e Legendre (2002) propuseram um método alternativo baseado na regressão polinomial para considerar relações não lineares entre variáveis de dois conjuntos de dados multivariados.

Dado o exposto, o presente trabalho teve como objetivo principal propor a utilização da análise de redundância e análise de correspondência canônica tradicionais e suas abordagens não lineares (MAKARENKOV; LEGENDRE, 2002) fundamentadas na regressão polinomial para descrever as relações entre as diferentes notas de atributos sensoriais e composições de blends de café. Como objetivos secundários; realizar um estudo de comparação entre os modelos utilizados, através do percentual de variância explicada e testes de significância; verificar diferentes transformações nos conjuntos de dados que forneçam melhores interpretações gráficas das relações entre as respostas sensoriais e as formulações dos blends de café.

Desse modo, os modelos de análise de redundância polinomial e análise de correspondência polinomial forneceram um aumento estatisticamente significativo na quantidade de variação das respostas sensoriais explicada pelas diferentes formulações de blends de café, se comparada aos modelos de RDA e CCA clássicos.

2 REFERENCIAL TEÓRICO

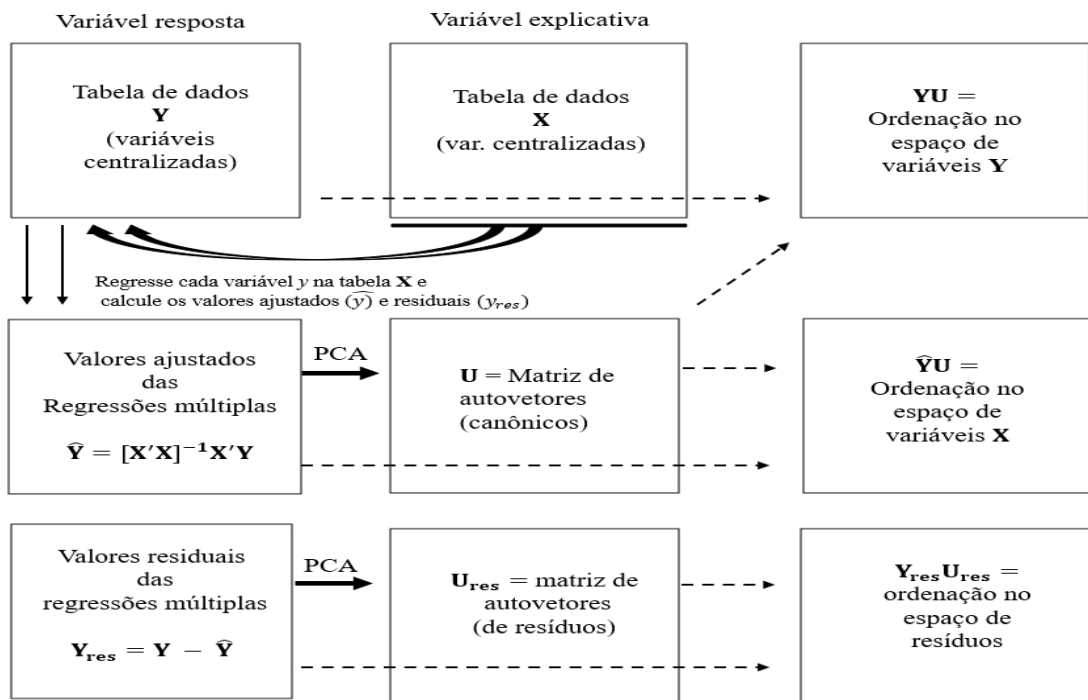
Este capítulo tem como objetivo apresentar as fundamentações teóricas das técnicas empregadas no estudo, abordando conceitos, pressupostos, álgebra relacionada e recomendações através da revisão de literatura.

2.1 ANÁLISE DE REDUNDÂNCIA (RDA)

Inicialmente, para estabelecer as relações entre dois conjuntos de dados, Stewart e Love (1968) propuseram o índice de redundância. Este índice busca determinar a variância total de um conjunto de dados resposta Y a partir de um conjunto de variáveis preditoras X , por meio de predição linear. Utilizando esta medida de redundância, Van den Wollenberg (1977) introduziu o método de análise de redundância, que consiste em extrair fatores do conjunto X que maximizam a redundância entre os dois conjuntos de variáveis.

Segundo Lazraq e Cléroux (2002) a análise de redundância pode ser vista como um associação entre a análise de correlação canônica e a regressão multivariada. Também é equivalente à análise de componentes principais em relação a um conjunto de variáveis instrumentais.

Figura 2.1 – Fluxograma do algoritmo de análise de redundância



Fonte: Traduzido de Legendre e Legendre (2012)

Conforme a Figura 2.1, a análise de redundância pode ser entendida como um processo de duas etapas: a primeira consiste na regressão das variáveis de Y através das variáveis regressoras em X para obtenção dos valores ajustados; a segunda etapa compreende a utilização de técnicas de redução de dimensionalidade na matriz de valores ajustados, como a decomposição em valores singulares para obtenção de autovalores e autovetores.

No modelo de regressão linear multivariado, pode-se expressar cada variável resposta y em uma amostra de n observações como uma função linear das variáveis predictoras x mais um erro aleatório, ε :

$$\begin{aligned}
y_1 &= b_0 + b_1x_{11} + b_2x_{12} + \cdots + b_mx_{1m} + \varepsilon_1 \\
y_2 &= b_0 + b_1x_{21} + b_2x_{22} + \cdots + b_mx_{2m} + \varepsilon_2 \\
&\vdots \\
y_n &= b_0 + b_1x_{n1} + b_2x_{n2} + \cdots + b_mx_{nm} + \varepsilon_n.
\end{aligned} \tag{2.1}$$

Usando a notação matricial, os modelos para as n observações em (2.1) podem ser escritos como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

em que \mathbf{X} representa a matriz de variáveis explicativas de ordem $(n \times m)$ e $\boldsymbol{\beta}$ é a matriz dos coeficientes de regressão de todas as n variáveis de resposta \mathbf{y} nas m variáveis explicativas \mathbf{X} .

Considerando as suposições para os termos de erro aleatório não correlacionados $E(\boldsymbol{\varepsilon}) = 0$ e $cov(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, assume-se as seguintes suposições para \mathbf{y} :

1. $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$,
2. $cov(\mathbf{y}) = \sigma^2\mathbf{I}$.

Conforme Rencher e Schaalje (2008), se a primeira suposição for válida pode-se estimar $E(\mathbf{y})$ através da estimação dos coeficientes $\boldsymbol{\beta}$ pelo método dos mínimos quadrados. As estimativas por mínimos quadrados de $\beta_0, \beta_1, \dots, \beta_m$ minimizam a soma dos quadrados dos desvios dos valores preditos de \hat{y}_i pelo modelo. Dessa forma busca-se $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$ que minimize

$$SSE = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1x_{i1} - \hat{\beta}_2x_{i2} - \cdots - \hat{\beta}_mx_{im} \right)^2 \tag{2.3}$$

Assim, tomando $\varepsilon = \mathbf{y} - \mathbf{X}\beta$ pode-se estimar $\hat{\beta}$ de modo que

$$\begin{aligned}
\sum_{i=1}^n \hat{\varepsilon}_i^2 &= \varepsilon \varepsilon' \\
\varepsilon \varepsilon' &= [\mathbf{y} - \mathbf{X}\beta]' [\mathbf{y} - \mathbf{X}\beta] \\
\varepsilon \varepsilon' &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta \\
\varepsilon \varepsilon' &= \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta \\
\frac{\partial \varepsilon \varepsilon'}{\partial \beta} &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0 \\
\mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{X}'\mathbf{y} \\
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}
\end{aligned} \tag{2.4}$$

Para Legendre e Legendre (2012), na análise de redundância a centralização das variáveis preditoras em suas respectivas médias é necessário para linearizar as relações e tornar as distribuições de frequência o mais simétricas possível, reduzindo assim o efeito de outliers. Assim, o modelo em (2.1) para cada y_i pode ser escrito em termos de variáveis de \mathbf{X} centradas como:

$$\begin{aligned}
y_i &= b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \\
&= \alpha + \beta_1 (x_{i1} - \bar{x}_1) + \beta_2 (x_{i2} - \bar{x}_2) + \dots + \beta_k (x_{ik} - \bar{x}_k) + \varepsilon_i
\end{aligned} \tag{2.5}$$

em que $\alpha = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \dots + \beta_k \bar{x}_k$ e $\bar{x}_j = \sum_{i=1}^n x_{ij}/n, j = 1, 2, \dots, k..$ Desse modo o modelo matricial equivalente à (2.2) é

$$\mathbf{y} = (\mathbf{j}, \mathbf{X}_c) \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} + \varepsilon \tag{2.6}$$

em que \mathbf{j} representa um vetor de 1's. A matriz de variáveis preditoras \mathbf{X}_c centralizada por suas médias é construída por

$$\mathbf{X}_c = \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{X} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{pmatrix} \quad (2.7)$$

Como consequência dessa adequação, os estimadores de mínimos quadrados serão $\hat{\alpha} = \bar{y}$ e $\beta_1 = (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{y}$.

O objetivo dos métodos de ordenação é representar os dados ao longo de um número reduzido de eixos ortogonais, construídos de forma que representem, em ordem decrescente, as principais tendências de variação dos dados, portanto, para este método, a análise é feita por meio da matriz de variâncias e covariâncias dos valores preditos. As inter-relações entre as variáveis envolvidas na análise canônica podem ser representadas pela matriz de covariância particionada, resultante da concatenação das variáveis.

$$\mathbf{S}_{\hat{\mathbf{Y}}|\hat{\mathbf{Y}}} = \mathbf{S}_{\mathbf{YX}} \mathbf{S}_{\mathbf{XX}}^{-1} \mathbf{S}'_{\mathbf{YX}} \quad (2.8)$$

em que $\mathbf{S}_{\mathbf{YX}}$ é a matriz de covariâncias ($p \times m$) entre as variáveis resposta e explicativas obtida por $\mathbf{X}'_c \mathbf{y} / n - 1$, $\mathbf{S}_{\mathbf{XX}}$ a matriz de covariâncias ($m \times m$) entre as variáveis explicativas $\mathbf{X}'_c \mathbf{X}_c / n - 1$.

Pode-se ainda calcular o coeficiente de determinação canônica, aqui chamado de estatística de redundância bivariada. Esta estatística avalia a intensidade da relação linear entre \mathbf{Y} e \mathbf{X} :

$$R^2_{\hat{\mathbf{Y}}|\mathbf{X}} = \frac{SS(\hat{\mathbf{Y}})}{SS(\mathbf{Y})} \quad (2.9)$$

em que $SS(\hat{\mathbf{Y}})$ é a soma total dos quadrados de $\hat{\mathbf{Y}}$ e $SS(\mathbf{Y})$ é a soma total dos quadrados de \mathbf{Y} . O R^2 canônico é construído da mesma maneira e tem o mesmo significado que a estatística R^2 na regressão múltipla.

2.2 ANÁLISE DE CORRESPONDÊNCIA CANÔNICA (CCA)

A análise de correspondência canônica é um método de ordenação assimétrica canônica desenvolvido por ter Braak (1986) e como o nome sugere, trata-se da forma canônica da análise de correspondência. Basicamente, é uma forma ponderada da RDA aplicada a uma matriz $\bar{\mathbf{Q}}$ de contribuições para a estatística χ^2 usada na análise de correspondência (Legendre e Legendre 2012).

Por meio de uma tabela de contigência a matriz \mathbf{Y} de variáveis resposta é transformada em uma matriz $\bar{\mathbf{Q}}$ ($n \times m$) por:

$$\bar{\mathbf{Q}} = [\bar{q}_{ij}] = \left[\frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right] \quad (2.10)$$

em que p_{ij} representa as frequências relativas da i -ésima linha e j -ésima coluna, p_{i+} corresponde às frequências relativas totais das linhas i e p_{+j} corresponde às frequências relativas totais das colunas j .

2.2.1 Regressão Ponderada

De acordo com teorema de Gauss-Markov o estimador de mínimos quadrados será eficiente se for não viesado e se os erros do modelo de regressão linear forem independentes e homocedásticos.

Conforme Cameron e Trivedi (2005), ao assumir a matriz de variância dos erros como $\mathbf{V} = \sigma^2 \mathbf{I}$. Se \mathbf{V} for conhecido e não singular, pode-se pré-multiplicar o modelo de regressão linear (2.2) por \mathbf{V}^{-1} , em que $\mathbf{V}^{1/2} \mathbf{V}^{1/2} = \mathbf{V}$, para obter

$$\mathbf{V}^{-1/2} \mathbf{y} = \mathbf{V}^{-1/2} \mathbf{X} \boldsymbol{\beta} + \mathbf{V}^{-1/2} \boldsymbol{\varepsilon} \quad (2.11)$$

Os erros neste modelo transformado possuem média zero, são não correlacionados e homocedásticos. Dessa forma o melhor estimador de mínimos quadrados generalizado não viesado $\hat{\beta}$ é obtido através de:

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (2.12)$$

Segundo Rencher e Schaalje (2008), caso a matriz \mathbf{V} seja conhecida e diagonal esse método é chamado de mínimos quadrados ponderados.

No caso da análise de correspondência canônica, a matrix \mathbf{X} será ponderada por uma matriz diagonal $\mathbf{D}^{1/2}(p_{i+})$ da raiz quadrada dos pesos das linhas da matriz de variáveis respostas \mathbf{Y} .

$$\mathbf{X}_w = \mathbf{D}(p_{i+})^{1/2}\mathbf{X} \quad (2.13)$$

Assim a regressão multivariada ponderada será dada por:

$$\hat{\mathbf{Q}} = \mathbf{X}_w\mathbf{B} = \mathbf{X}_w[\mathbf{X}'_w\mathbf{X}_w]^{-1}\mathbf{X}_w\bar{\mathbf{Q}} \quad (2.14)$$

A matriz de covariâncias para a CCA pode ser calculado por meio da expressão $\mathbf{S} = \hat{\mathbf{Q}}'\hat{\mathbf{Q}}$.

2.3 Biplot

A técnica biplot foi descrita minuciosamente por Gabriel (1971), e tem por objetivo a representação gráfica de linhas e colunas da matriz de dados como pontos em um espaço vetorial de dimensão reduzida.

Eles são muito usados no contexto da análise de componentes principais (Jolliffe 2002), mas também são ferramentas úteis para inspeções de dados no contexto da modelagem estatística. Como técnica de projeção, eles compartilham semelhanças com muitas outras técnicas de projeção, como escala multidimensional (Kruskal e Wish 1978), análise de coordenadas principais (Fenty 2004) e análise de correspondência (Blasius e Greenacre 1998).

Conforme Legendre e Legendre (2012) a RDA e a CCA são análogas a técnica de componentes principais, na qual um componente principal é semelhante a um eixo canônico, sendo o primeiro uma combinação linear das variáveis e a segunda uma combinação linear das variáveis ajustadas por uma regressão. Dessa maneira, esta seção abordará a técnica biplot no contexto de componentes principais e técnicas análogas.

De acordo com Johnson e Wichern (1998), a obtenção de um biplot originou-se da fatoração de uma matriz proveniente de uma amostra multivariada. Seja uma matriz $\mathbf{X}_{n \times p}$, sendo n o número de observações e p o número de variáveis, essa matriz pode ser decomposta em valores singulares de forma que:

$$\mathbf{X}_{n \times p} = \mathbf{U}_{n \times p} \mathbf{\Lambda}_{p \times p} \mathbf{V}'_{p \times p} \quad (2.15)$$

em que $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, uma matriz diagonal formada pelos p valores singulares, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$; \mathbf{U} , uma matriz ortogonal dos autovetores de $\mathbf{X}_{n \times p} \times \mathbf{X}'_{p \times n}$; \mathbf{V} , uma matriz ortogonal dos autovetores de $\mathbf{X}'_{p \times n} \times \mathbf{X}_{n \times p}$.

Com essa caracterização, é possível aproximar a matriz $\mathbf{X}_{n \times p}$ de uma matriz \mathbf{Y}_k considerando uma certa quantidade de autovalores, em que k representa o número de dimensões da aproximação, geralmente $k = 2$.

A matriz \mathbf{Y}_k pode ser definida pelo produto de uma matriz \mathbf{G} das coordenadas de linhas por uma matriz \mathbf{H} das coordenadas de colunas, de forma que:

$$\mathbf{G} = \mathbf{U}_{(k)} \mathbf{\Lambda}_{(k)}^\alpha \quad (2.16)$$

$$\mathbf{H} = \mathbf{V}_{(k)} \mathbf{\Lambda}_{(k)}^{1-\alpha} \quad (2.17)$$

Se tomarmos $k = 2$ o número de dimensões a serem plotadas, a matriz de aproximação \mathbf{Y} será dada por:

$$\hat{\mathbf{Y}}_2 = \mathbf{G}\mathbf{H}' = \mathbf{U}_{(2)} \mathbf{\Lambda}_{(2)}^\alpha \mathbf{\Lambda}_{(2)}^{1-\alpha} \mathbf{V}'_{(2)} = \mathbf{U}_{(2)} \mathbf{\Lambda}_{(2)} \mathbf{V}'_{(2)} \quad (2.18)$$

Na expressão (2.16), Gabriel (1971) propôs três representações distintas para os biplots baseado no valor do parâmetro $\alpha \in [0, 1]$.

1. Para $\alpha = 1/2$, tem-se o Biplot SQRT. Este biplot atribui a mesma importância para as coordenadas de linhas e colunas.

$$\mathbf{Y} = \begin{pmatrix} u_{11}\sqrt{\lambda_1} & u_{21}\sqrt{\lambda_2} \\ \vdots & \vdots \\ u_{n1}\sqrt{\lambda_1} & u_{n2}\sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} v_{11}\sqrt{\lambda_1} & \cdots & v_{1p}\sqrt{\lambda_2} \\ v_{21}\sqrt{\lambda_1} & \cdots & v_{2p}\sqrt{\lambda_2} \end{pmatrix} \quad (2.19)$$

2. Para $\alpha = 0$, tem-se o Biplot GH. Este biplot Prioriza coordenadas de coluna, obtém uma máxima qualidade de representação para as colunas e mínima para as linhas.

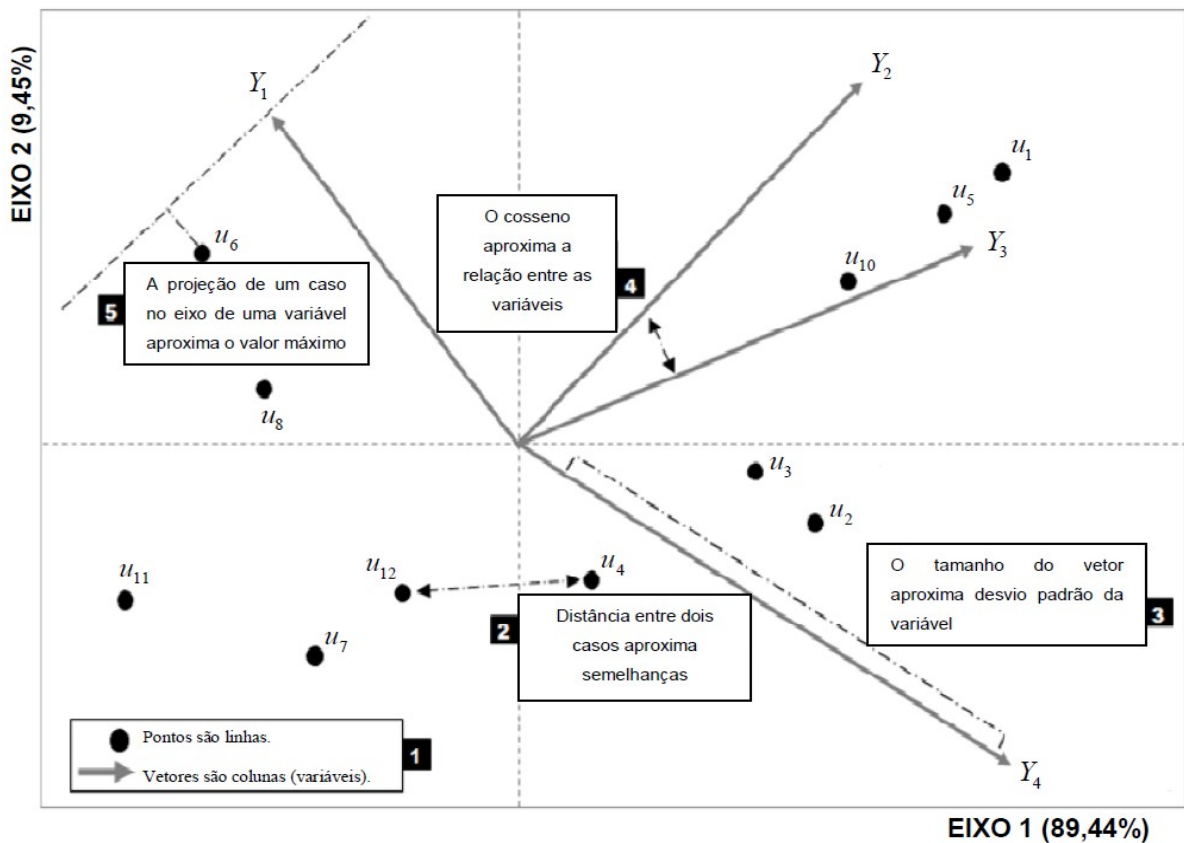
$$\mathbf{Y} = \begin{pmatrix} u_{11} & u_{21} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{pmatrix} \begin{pmatrix} v_{11}\lambda_1 & \cdots & v_{1p}\lambda_2 \\ v_{21}\lambda_1 & \cdots & v_{2p}\lambda_2 \end{pmatrix} \quad (2.20)$$

3. Para $\alpha = 1$, tem-se o Biplot JK. Este biplot prioriza coordenadas de linha, obtém uma máxima qualidade de representação para as linhas e mínima para as colunas.

$$\mathbf{Y} = \begin{pmatrix} u_{11}\lambda_1 & u_{21}\lambda_2 \\ \vdots & \vdots \\ u_{n1}\lambda_1 & u_{n2}\lambda_2 \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1p} \\ v_{21} & \cdots & v_{2p} \end{pmatrix} \quad (2.21)$$

Segundo Ferreira (2019), para qualquer tipo de biplots supracitados a sua interpretação básica pode ser dada conforme a Figura 2.2 a seguir.

Figura 2.2 – Interpretação básica de uma representação biplot.



Fonte: Traduzido de Torres-Salinas et al. (2013)

No biplot detalhado pela Figura 2.2, as variáveis hipotéticas Y_1, Y_2, Y_3, Y_4 são representadas por vetores cujos seus comprimentos refletem seus desvios-padrão. Os cossenos dos ângulos formados entre os vetores retratam a intensidade da correlação entre as variáveis de forma que:

- ângulos inferiores a 90° indicam correlação positiva entre variáveis;
- ângulos superiores a 90° indicam correlação negativa entre variáveis;
- ângulos iguais a 90° indicam que a correlação entre as variáveis é nula.

Cada ponto disposto no gráfico representa uma observação ou unidade amostral. A projeção de um ponto em relação a um vetor variável, aproxima o valor da variável para aquela

observação. Assim, a inspeção visual sugere que o objeto u_6 possui valores maiores da variável Y_1 em relação à maioria dos outros objetos. Ainda, a distância entre os objetos reflete suas semelhanças.

2.3.1 Biplot na análise de redundância

Sendo U uma matriz de autovetores normalizados de \mathbf{u}_k é possível calcular as coordenadas responsáveis pela representação gráfica no biplot. Segundo Ter Braak (1994) duas formas de representações gráficas podem ser obtidas na análise de redundância, de acordo com o propósito do estudo.

Quadro 2.1 – Resumo dos dois tipos de representação de biplots para análise de redundâncias de acordo com Ter Braak (1994).

Representação	Espaço de Y	Espaço de X	Recomendação
Tipo 1	$\mathbf{F} = \mathbf{YU}$	$\mathbf{Z} = \mathbf{XBU}$	Considera a distância Euclidiana entre os objetos
Tipo 2	$\mathbf{F}\Lambda^{-1/2}$	$\mathbf{Z}\Lambda^{-1/2}$	Considera a correlação entre as variáveis resposta

Fonte: Adaptado de Legendre e Legendre (2012)

Na representação do tipo 1 mostrada no Quadro 2.1, as similaridades entre os objetos (linhas) são calculadas pela distância euclidiana e os ângulos entre as variáveis respostas e explicativas retratam suas correlações, porém os ângulos entre as variáveis respostas não têm significado. Para a representação do tipo 2 os objetos não são discriminados por suas distâncias euclidianas, mas são plotados em relação aos centróides das variáveis, e as correlações entre as variáveis respostas são refletidas pelos seus ângulos.

Dessa forma, o tipo de representação depende da ênfase que se deseja dar ao biplot, ou seja, exibição de distâncias entre objetos ou de correlações entre variáveis.

2.3.2 Biplot na análise de correspondência canônica

De forma análoga à análise de redundância, na análise de correspondência canônica também pode-se obter dois tipos de representações gráficas. As expressões para se obter as coordenadas necessárias para a construção do biplot são mostradas no Quadro 2.2.

Quadro 2.2 – Resumo dos dois tipos de representação de biplots para análise de correspondência canônica de acordo com Ter Braak (1994).

Representação	Coordenadas		
	Objetos	Espaço de \mathbf{Y}	Espaço de \mathbf{X}
Tipo 1	$\mathbf{F} = \hat{\mathbf{V}}\Lambda^{1/2}$	$\mathbf{V} = \mathbf{D}(p_{+j})^{-1/2}\mathbf{U}$	$\mathbf{Z} = \mathbf{D}(p_{i+})^{-1/2}\hat{\mathbf{Q}}\mathbf{U}$
Tipo 2	$\hat{\mathbf{V}} = \mathbf{D}(p_{i+})^{-1/2}\hat{\mathbf{Q}}\mathbf{U}\Lambda^{-1/2}$	$\hat{\mathbf{f}} = \mathbf{V}\Lambda^{1/2}$	$\mathbf{Z} = \mathbf{D}(p_{i+})^{-1/2}\hat{\mathbf{Q}}\mathbf{U}\Lambda^{-1/2}$

Fonte: Adaptado de Legendre e Legendre (2012)

A representação do tipo 1 é apropriada quando pesquisador está interessado na ordenação dos objetos dispostos nas linhas, e são plotados em relação aos centróides das variáveis respostas. Quando o enfoque é na relação entre as variáveis respostas, a representação indicada é a do tipo 2. Um ponto a ser considerado é que, devido ao fato dessas análises gerarem coordenadas gráficas para as amostras, variáveis respostas e variáveis explicativas, os biplots passam a ser denominados como triplots.

2.4 TESTE DE PERMUTAÇÃO

Os testes de permutação são uma maneira eficaz de determinar a significância estatística através de uma reamostragem dos dados. Esses testes podem ser utilizados para testar qualquer estatística de teste e determinar a significância estatística quando não se tem conhecimento da distribuição dos dados ou quando se possui um tamanho amostral pequeno. Segundo Legendre e Legendre (2012), as primeiras referências conhecidas sobre testes de permutação podem ser encontradas em Fisher (1935).

Para avaliar a adequabilidade de uma regressão linear por meio de testes de hipóteses para os parâmetros do modelo ou para a construção de intervalos de confiança, o valor da estatís-

tica de teste calculado a partir dos dados é comparado com a distribuição esperada da estatística sob a hipótese nula testada e com base nessa comparação, estima-se a probabilidade de obter resultados diferentes daqueles esperados sob a hipótese nula. Para tal, A distribuição esperada da estatística de teste é derivada da suposição sobre a distribuição dos dados originais, nesse caso, assume-se a normalidade dos resíduos na regressão de mínimos quadrados. Porém, segundo Smilauer e Leps (2014) na análise de redundância e análise de correspondência canônica a distribuição da estatística de teste sob a hipótese nula de independência não é conhecida. Esta distribuição depende do número de variáveis respostas, sua estrutura de correlação, além da distribuição dos valores dos dados de resposta. Todavia, a distribuição pode ser simulada através da permutação de Monte Carlo.

Este método consiste na permutação dos dados originais em número suficientemente grande de amostras em que calcula-se alguma estatística a partir delas (média, mediana etc.), e o valor médio destas estatísticas indicará o valor real desse parâmetro na população. No caso da RDA e CCA a estatística de teste utilizada é chamada de valor pseudo-F , análogo ao valor F da regressão linear, dado por

$$pseudo - F = \frac{var\hat{Y}(\hat{Q})}{var.totY(\bar{Q}) - var\hat{Y}(\hat{Q})} \quad (2.22)$$

Denomina-se essa estatística de “pseudo-F” porque os graus de liberdade não estão incluídos no numerador e no denominador, uma vez que o número de parâmetros no modelo polinomial pode mudar em cada permutação realizada. Os graus de liberdade são constantes multiplicativas por meio das permutações, incluí-los ou não, não altera o resultado do teste. A hipótese nula considera que os dados explicativos não têm efeito sobre os dados de resposta.

2.5 ANÁLISE COMPOSICIONAL

De modo geral, dados composicionais consistem em vetores nos quais seus componentes (variáveis) representam proporções de um todo. Dados com esse tipo de estrutura estão presentes em diversas áreas de pesquisa, como a geologia, ecologia, economia e também na análise composicional de alimentos. Contudo, tais composições devem ser analisadas com bastante cautela, uma vez que cada componente estará relacionada com as restantes levando a interpretações equivocadas nos resultados.

2.5.1 Espaço amostral simplex

Conforme Aitchison (1986) a etapa primordial na análise estatística de experimentos consiste na elaboração de um espaço amostral adequado. Os dados composicionais compreendem um espaço amostral em que um vetor $x = [x_1, x_2, \dots, x_D]$ representa uma composição de D partes que satisfazem a restrição da soma constante de forma que :

$$x_1 \geq 0, \dots, x_D \geq 0; \quad (2.23)$$

$$\sum_{i=1}^D x_i = k. \quad (2.24)$$

Segundo Pawlowsky-Glahn et al. (2015), vetores com essas propriedades não pertencem a um espaço vetorial linear com uma estrutura métrica, portanto operações matemáticas com esses vetores, como multiplicação por um escalar e cálculo de distâncias entre pontos não são adequados para dados composicionais. Considerando essa restrição, Aitchison (1986) propôs que os vetores composicionais compreendem um espaço amostral denominado simplex, formalmente expresso como:

$$S^D = \{\mathbf{X} = [x_1, x_2, \dots, x_D] \mid x_i > 0, i = 1, 2, \dots, D; \sum_{l=1}^D x_l = k\}. \quad (2.25)$$

Apesar das operações comuns ao espaço real euclidiano não serem adequadas no espaço simplex, é possível realizar operações e estabelecer métricas análogas. Para isso é necessário compreender o conceito de fechamento (closure).

2.5.2 Operação de fechamento

Qualquer composição pode ser expressa em proporção usando um fator de escala apropriado, esse conceito é importante para que os vetores composicionais sejam equivalentes de forma a garantir que as somas de seus componentes sejam igual a uma constante k . A operação de atribuir um elemento de soma constante a uma composição é chamada de fechamento e é definida como:

$$C_{(z)} = \left[\frac{k \times z_1}{\sum_{i=1}^D z_i}, \frac{k \times z_2}{\sum_{i=1}^D z_i}, \dots, \frac{k \times z_D}{\sum_{i=1}^D z_i} \right] \quad (2.26)$$

para $\mathbf{Z} = [z_1, \dots, z_D] \in \mathbb{R}_+^D$, $z_i > 0$ para todo $i = 1, 2, \dots, D$.

O resultado do fechamento é um reescalonamento do vetor inicial de forma que a soma de seus componentes seja k . Assim, o fechamento nada mais é do que a projeção de um vetor com componentes positivos no simplex (PAWLOWSKY-GLAHN et al., 2015).

2.5.3 Operação perturbação

A perturbação é considerada a operação fundamental na análise de dados composicionais, tão importante quanto a adição de vetores no espaço vetorial real usual (VAN DEN BOOGAART E TOLOSANA-DELGADO, 2013). Assim, dada uma composição \mathbf{x} de D partes e \mathbf{y} um vetor com D elementos positivos. Então a operação

$$\mathbf{X} = \mathbf{x} \oplus \mathbf{y} = C[x_1 y_1, x_2 y_2, \dots, x_D y_D] \in S^D. \quad (2.27)$$

é denominado uma perturbação com a composição original \mathbf{x} sendo operado pelo vetor perturbador \mathbf{y} para formar uma composição perturbada \mathbf{X} .

2.5.4 Operação Potência

Segundo Pawlowsky-Glahn et al. (2015), a operação denominada potência no espaço simplex é análoga ao produto de um vetor por um escalar no espaço euclidiano. Essa operação pode ser definida da seguinte forma:

$$\alpha \odot \mathbf{x} = C[x_1^\alpha, x_2^\alpha, \dots, x_D^\alpha] \in S^D. \quad (2.28)$$

Como o simplex, S_D, \oplus, \odot é caracterizado como um espaço vetorial, as propriedades válidas no espaço euclidiano (comutativas, associativas, distributiva, elemento neutro e elemento inverso) são mantidas (PEREIRA, 2020).

2.5.5 Produto interno, norma e distância

O produto interno, norma e distância são conceitos importantes para se obter uma estrutura do espaço vetorial euclidiano, tais conceitos são possíveis também no espaço simplex conforme a geometria proposta por Aitchison (1986) representada pelo subíndice a .

Produto interno de $\mathbf{x}, \mathbf{y} \in S^D$,

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i y_i}{x_j y_j}. \quad (2.29)$$

Norma de $\mathbf{x} \in S^D$,

$$\| \mathbf{x} \|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2}. \quad (2.30)$$

Distância entre \mathbf{x} e $\mathbf{y} \in S^D$,

$$d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}. \quad (2.31)$$

2.5.6 Transformações de razões logarítmicas

A geometria desenvolvida por Aitchison para dados de natureza composicional adequou algumas das mais importantes operações e métricas do espaço real \mathbb{R}^D para o espaço simplex S^D , porém, métodos mais robustos de análise estatística não possuem essa equivalência, principalmente quando se baseiam em estruturas de variância e correlação.

Aitchison (1982) propôs a utilização de transformações como uma ferramenta descritiva para o estudo da variabilidade composicional. Tais transformações consideram a natureza relativa dos dados, expressos em termos de razões e a praticidade de se trabalhar em espaços amostrais complicados utilizando logaritmos.

2.5.6.1 Razão Logarítmica Aditiva (alr)

Proposta por Aitchison (1986), a transformação de razão logarítmica aditiva mapeia uma composição pertencente ao espaço simplex S^D para um vetor euclidiano dimensional $D - 1$ por meio do logaritmo da razão entre os componentes, de modo que o denominador seja uma parte comum entre todos os componentes. Assim, seja $\mathbf{x} = [x_1, x_2, \dots, x_D]$ uma composição e x_D uma parte (variável) de referência. A transformação *alr* é definida da seguinte maneira:

$$alr(\mathbf{x}) = \zeta = \left[\ln \frac{x_1}{x_D}, \ln \frac{x_2}{x_D}, \dots, \ln \frac{x_{D-1}}{x_D} \right] \in \mathbb{R}^{D-1} \quad (2.32)$$

A transformação inversa de $\zeta = [\zeta_1, \zeta_2, \dots, \zeta_{D-1}]$ para retornar a \mathbf{x} é obtida por meio de:

$$\mathbf{x} = alr^{-1}(\zeta) = C [\exp(\zeta_1), \exp(\zeta_2), \dots, \exp(\zeta_{D-1}), 1] \quad (2.33)$$

Como consequência desse procedimento, remove-se o efeito da restrição de soma constante, e os dados transformados podem se ajustar a uma distribuição normal, tornando os dados adequados para análises estatísticas clássicas, como a regressão linear múltipla. No entanto, para o cálculo de produtos internos e distâncias não há isometria entre S^D e \mathbb{R}^{D-1} , isto é, a *alr* não conserva os segmentos geométricos do espaço amostral original, ponto particularmente crítico para análises multivariadas.

2.5.6.2 Razão Logarítmica Centrada (*clr*)

A partir do conceito da transformação de razões logarítmicas, Aitchison (1986) formulou uma nova transformação com o propósito de contornar a ausência de isometria apresentado na *alr*. Denominada de razão logarítmica centrada (*clr*), esta nova transformação considera a média geométrica das componentes em vez de uma componente referência no denominador. Assim, seja $\mathbf{x} = [x_1, x_2, \dots, x_D]$, a transformação *clr* é definida como:

$$clr(\mathbf{x}) = \left(\ln \frac{x_1}{gm(x)}, \dots, \ln \frac{x_i}{gm(x)}, \dots, \ln \frac{x_D}{gm(x)} \right) \quad (2.34)$$

em que

$$gm(\mathbf{X}) = \left(\prod_{i=1}^D x_i \right)^{\frac{1}{D}} \quad (2.35)$$

A transformação inversa é obtida através da expressão:

$$\mathbf{x} = clr^{-1}(\xi) = C [exp(\xi_1), exp(\xi_2), \dots, exp(\xi_{D-1})] \quad (2.36)$$

Além de ser isométrica e permitir sua utilização em análises que envolvam métricas como distâncias, temos como outra vantagem o fato de que, para essa transformação não se faz necessário abrir mão de uma das componentes para o cálculo das razões.

Contudo, apesar dos ganhos em relação à *alr*, a amostra transformada pela *clr* resulta em uma nova restrição, na qual suas somas correspondem à zero. Isso significa que, para amostras aleatórias, a matriz de covariâncias de $clr(\mathbf{x}) = \xi$ será singular, ou seja, seu determinante será igual a zero.

2.5.6.3 Razão Logarítmica Isométrica (*ilr*)

Baseadas nas transformações pioneiras de Aitchison, novas técnicas matemáticas foram desenvolvidas para trabalhar de forma mais eficaz com dados composicionais, evitando as potenciais limitações descritas nas seções anteriores. Para solucionar tais limitações, Egozcue et al, (2003) propuseram a transformação razão logarítmica isométrica (*ilr*).

Essa nova transformação foi desenvolvida através da associação de uma base ortonormal com uma partição do vetor de composição. Desta forma, para qualquer composição $\mathbf{x} = [x_1, x_2, \dots, x_D]$, a transformação de razão logarítmica associada a uma base ortonormal de Aitchison e_i , $i = 1, 2, \dots, D - 1$ é dada por:

$$\mathbf{y} = ilr(\mathbf{x}) = [\langle x, e_1 \rangle_a, \langle x, e_2 \rangle_a, \dots, \langle x, e_{D-1} \rangle_a] \quad (2.37)$$

em que

$$y_i = \sqrt{\frac{i}{i+1}} \ln \left[\frac{g(x_1, \dots, x_i)}{x_{i+1}} \right], \quad i = 1, 2, \dots, D - 1, \quad (2.38)$$

A transformação inversa é obtida através da expressão:

$$\mathbf{x} = ilr^{-1}(\mathbf{y}) = \bigoplus_{i=1}^{D-1} (\langle \mathbf{y}, \vec{e}_i \rangle \otimes \mathbf{e}_i) \quad (2.39)$$

onde $\langle \mathbf{y}, \vec{e}_i \rangle = \langle x, e_i \rangle_a = y_i$.

Uma base ortonormal permite representar qualquer elemento do simplex, por suas coordenadas em eixos ortogonais como normalmente é feito em espaços reais. Com isso, a partir dessa transformação podem ser empregadas técnicas de redução de dimensionalidade como análise de componentes principais.

3 MATERIAIS E MÉTODOS

Este capítulo tem como objetivo apresentar o experimento científico responsável pela produção dos dados do estudo, assim como o método de adaptação das análises de redundância e correspondência canônica clássicas para as suas extensões polinomiais.

3.1 Materiais

Para avaliar os benefícios propostos pela abordagem multivariada não linear descrita neste trabalho, utilizou-se um conjunto de dados referentes a amostras de blends de café especiais e comerciais encontrados em Paulino et al. (2019).

Conforme descrevem os autores, os blends foram formulados a partir de proporções de variedades dos cafés especiais da espécie Arábica, Bourbon amarelo (CEB) e Acaiá (CEA), café Conilon (CC) e de uma marca comercial de café torrado (CT).

As amostras de cafés especiais foram produzidas no município de Carmo de Minas, localizado na região Mantiqueira de Minas, MG, Brasil, reconhecida mundialmente pela produção de cafés especiais (RIBEIRO et al., 2016). Já a amostra de café Conilon, representa uma mistura de lotes produzidos no estado do Espírito Santo, Brasil. O café comercial torrado representa uma marca comercializada no mercado nacional, comumente consumida na região sul do estado de Minas Gerais. Desta forma, quatro experimentos foram realizados, considerando os cafés especiais, além das concentrações da bebida definida nas porcentagens de 0,07 e 0,10 (m/v), caracterizando blends com bebidas com diferentes concentrações. Uma descrição resumida dos experimentos encontra-se na Tabela 3.1.

Dadas essas concentrações, o preparo das amostras foi feito utilizando água potável a 93 °C sem adição de açúcar. O tempo de extração foi de 4 minutos, empregando o 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.

Tabela 3.1 – Caracterização dos tratamentos utilizados na composição dos blends processados via seca.

Experimento	Variedade	Altitude	Concentração (m\v)
1	Bourbon amarelo	Acima de 1200m	0,07
2	Bourbon amarelo	Acima de 1200m	0,10
3	Acaiaá	Abaixo de 1100m	0,07
4	Acaiaá	Abaixo de 1100m	0,10

Fonte: Paulino et al. (2019).

Com o propósito de inferir o efeito da concentração da bebida definida nas porcentagens 0,07 e 0,10 m/v (35g/500 ml), em que 0,07 (m/v) significa 7 gramas de pó de café (m de massa) para cada 100 ml de água (v de volume), o mesmo a 0,10, os experimentos foram avaliados conjuntamente, considerando as composições descritas na (Tabela 3.2). Dessa forma, a identificação dos blends, na análise conjunta, segue a codificação na amostra ($k = 1, \dots, 36$), referindo-se aos blends analisados nos experimentos 1; 2; 3 e 4. Quanto ao contexto dos blends, trata-se de um produto a ser comercializado, o café da espécie *Canephora*, sendo tratado, doravante por Conilon, conforme sugere a descrição da Tabela 3.2, em função do tipo de processamento.

Tabela 3.2 – Composição dos blends formados pelos cafés Bourbon (CEB), Comercial torrado (CT), Conilon (CC) e Acaiaí (CEA)

Experimento 1					Experimento 3				
Amostra (K)	CEB	CT	CC	CEA	Amostra (K)	CEB	CT	CC	CEA
1	1,000	0,000	0,000	0,000	19	1,000	0,000	0,000	0,000
2	0,670	0,330	0,000	0,000	20	0,670	0,330	0,000	0,000
3	0,340	0,330	0,330	0,000	21	0,340	0,330	0,330	0,000
4	0,500	0,500	0,000	0,000	22	0,500	0,500	0,000	0,000
5	0,500	0,000	0,500	0,000	23	0,500	0,000	0,500	0,000
6	0,340	0,660	0,000	0,000	24	0,340	0,660	0,000	0,000
7	0,340	0,000	0,660	0,000	25	0,340	0,000	0,660	0,000
8	0,000	1,000	0,000	0,000	26	0,000	1,000	0,000	0,000
9	0,000	0,000	1,000	0,000	27	0,000	0,000	1,000	0,000
Experimento 2					Experimento 4				
Amostra (K)	CEB	CT	CC	CEA	Amostra (K)	CEB	CT	CC	CEA
10	0,000	0,000	0,000	1,000	28	0,000	0,000	0,000	1,000
11	0,000	0,330	0,000	0,670	29	0,000	0,330	0,000	0,670
12	0,000	0,330	0,330	0,340	30	0,000	0,330	0,330	0,340
13	0,000	0,500	0,000	0,500	31	0,000	0,500	0,000	0,500
14	0,000	0,000	0,500	0,500	32	0,000	0,000	0,500	0,500
15	0,000	0,660	0,000	0,340	33	0,000	0,660	0,000	0,340
16	0,000	0,000	0,660	0,340	34	0,000	0,000	0,660	0,340
17	0,000	1,000	0,000	0,000	35	0,000	1,000	0,000	0,000
18	0,000	0,000	1,000	0,000	36	0,000	0,000	1,000	0,000

Fonte: Paulino et al. (2019).

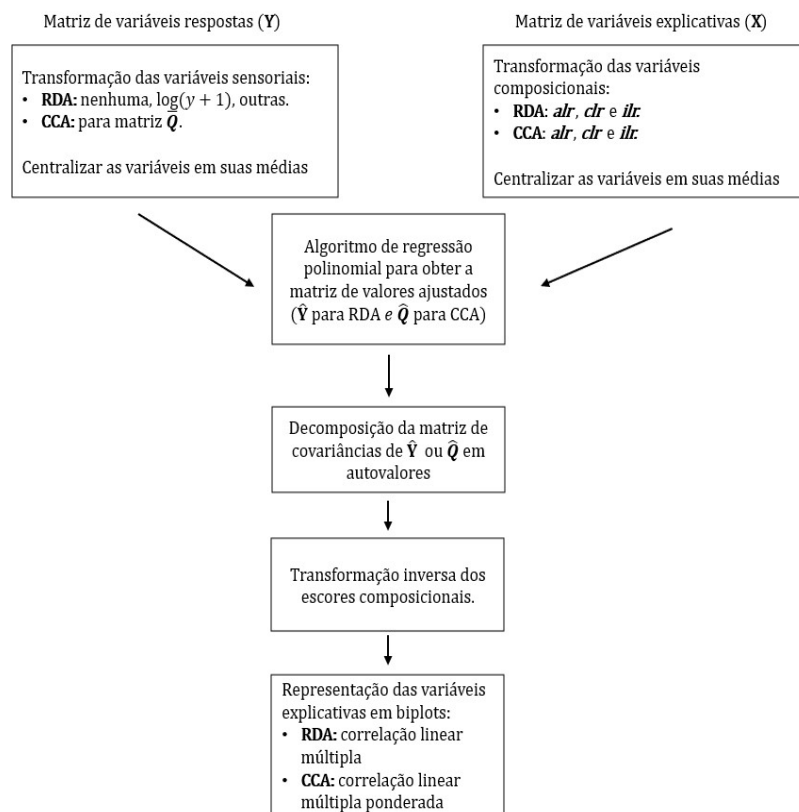
Cada experimento foi executado, em sessões distintas com intervalos de 24 horas, em virtude do número excessivo de avaliações. O grupo de cinco potenciais indivíduos avaliadores passou por uma seleção prévia através de experimento teste, para serem considerados aptos a diferenciar as amostras nos experimentos sensoriais. Após a seleção, o grupo final foi composto por cinco provadores qualificados. 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 em bancadas individuais para análise sensorial.

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: sabor, corpo, acidez, amargor e nota final, representando a impressão global da qualidade descrita pelos avaliadores.

3.2 Algoritmo de Regressão Polinomial

O modelo de algoritmo de regressão polinomial é uma extensão da análise de regressão linear multivariada para a análise de redundância e análise de correspondência canônica não lineares. Essa abordagem foi proposta por Makarenkov e Legendre (2002) com o objetivo de modelar a relação não linear entre comunidades de espécies e variáveis ambientais. A Figura 3.1, a seguir, apresenta o fluxo do algoritmo de regressão polinomial aplicado a análise de redundância e análise de correspondência canônica.

Figura 3.1 – Algoritmo de Regressão Polinomial



Fonte: Adaptado de Makarenkov e Legendre (2002)

Conforme a Figura 3.1, o algoritmo é composto de quatro etapas básicas, que são repetidas ($m - 1$) vezes à medida que a matriz das variáveis explicativas \mathbf{X} com m colunas é reduzida a um único vetor.

O procedimento inicial deste algoritmo corresponde a etapa inicial da RDA e CCA clássicas. É realizada uma regressão de cada variável resposta y em todas as variáveis do conjunto explicativo \mathbf{X} , seguindo um modelo de regressão linear múltipla para a RDA (Equação 2.2) e um modelo de regressão linear múltipla ponderada para o CCA (Equação 2.12).

Posteriormente, deve-se obter a matriz de valores residuais a partir da regressão múltipla. Para a obtenção desses valores, na RDA utiliza-se a expressão $y_{res} = y - \hat{y}$ e para a CCA $\bar{q}_{res} = \bar{q} - \hat{q}$.

Em seguida é necessário obter o par de variáveis de \mathbf{X} que proporcione a melhor aproximação quadrática dos resíduos. Isso pode ser feito criando-se uma matriz \mathbf{X}^{jk} (onde j e k são índices superiores) que representa um par de variáveis de \mathbf{X} . As colunas dessa nova matriz contém as variáveis $x_j, x_k, x_j x_k, x_j^2, x_k^2$ acrescida de uma coluna de constantes de valor 1. De forma hipotética, com $j = 1$ e $k = 2$ a matriz \mathbf{X}^{jk} é construída da seguinte forma:

$$\mathbf{X}^{12} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{11}\mathbf{x}_{12} & \mathbf{x}_{11}^2 & \mathbf{x}_{12}^2 & 1 \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{21}\mathbf{x}_{22} & \mathbf{x}_{21}^2 & \mathbf{x}_{22}^2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_{1n} & \mathbf{x}_{2n} & \mathbf{x}_{1n}\mathbf{x}_{2n} & \mathbf{x}_{1n}^2 & \mathbf{x}_{2n}^2 & 1 \end{bmatrix} \quad (3.1)$$

Assim é calculada uma regressão linear múltipla do vetor y_{res} com a nova matrix \mathbf{X}^{12} formada como preditora. Para a análise de redundância essa regressão será dada por:

$$\hat{\mathbf{y}}_{res}^{12} = \mathbf{X}^{12} \mathbf{c}^{12} \quad (3.2)$$

em que \mathbf{c}^{12} é o vetor de coeficientes de regressão para as variáveis explicativas $j = 1$ e $k = 2$, é calculado usando mínimos quadrados, como a matriz \mathbf{B} da Equação (2.8). Para levar em consideração os pesos na análise de correspondência canônica basta pré-multiplicar a matriz diagonal da raiz quadrada dos pesos das linhas da matriz de variáveis respostas de modo que $\hat{\mathbf{q}}_{res}^{12} = \mathbf{D}(p_{i+})^{1/2} \mathbf{X}^{12} \mathbf{c}^{12}$.

Este procedimento é repetido para cada par j, k de colunas de \mathbf{X} , e para cada par observado é calculado o coeficiente de determinação múltipla $R^2(j, k)$. O par de variáveis (j, k) que obtiver o maior coeficiente de determinação $R^2(j, k)$ é retido e será utilizado na próxima etapa.

As duas colunas j e k selecionadas na etapa anterior são combinadas para formar uma nova variável t em \mathbf{X} , que substitui j e k pelo restante da análise. A fórmula a seguir é usada para calcular a nova variável combinada t para cada observação $i (i = 1, \dots, n)$:

$$x_{it} = x_{ij}b_j x_{ik}b_k + \hat{y}_{res,i}^{jk} \quad (3.3)$$

em que os coeficientes b são os da equação (2.2). A fórmula (3.3) é válida para a RDA, para a adaptação à CCA é necessário apenas multiplicar o componente de resíduos por $p_i^{-1/2}$, resultando em:

$$x_{it} = x_{ij}b_j x_{ik}b_k + p_i^{-1/2} \hat{q}_{res,i}^{jk} \quad (3.4)$$

Os passos descritos anteriormente são repetidos $(m - 1)$ vezes até que a matriz \mathbf{X} ($n \times m$) seja transformada em uma matriz \mathbf{X} ($n \times 1$), que é um vetor simples. Para se obter o vetor final \hat{y} a ser usado na análise no lugar de y , realiza-se uma regressão linear simples de y em \mathbf{X} ($n \times 1$).

O algoritmo de regressão polinomial permite modelar as relações polinomiais entre as matrizes de resposta e as variáveis explicativas consideradas na RDA e na CCA, além de determinar quais termos devem ser mantidos no modelo. Desse modo, a matriz de valores ajustados $\hat{\mathbf{Y}}$ usados na análise não será mais uma combinação linear das variáveis explicativas em \mathbf{X} , mas uma combinação polinomial entre elas. Os polinômios gerados por esse algoritmo incluem subconjuntos de termos do seguinte modelo:

$$\begin{aligned} \hat{y} &= b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \\ &= + b_{m+1} x_1^2 + \dots + b_{2m} x_m^2 \\ &= + b_{2m+1} x_1 x_2 + \dots + b \prod_i x_i \prod_{j(j \neq i)} x_j^2 \end{aligned} \quad (3.5)$$

3.2.1 Teste de permutação

Caso os modelos linear e polinomial forem significativos, deve-se avaliar a diferença entre os dois modelos. Para isso, um procedimento de permutação derivado do teste pseudo-F foi utilizado.

$$pseudo - F = \frac{var_{polinomial} - var_{linear}}{var_{tot\mathbf{Y}(\bar{\mathbf{Q}})} - var_{polinomial}} \quad (3.6)$$

em que $var_{polinomial}$ corresponde a variância contabilizada de nos modelos polinomiais, var_{linear} a variância contabilizada pelos modelos lineares e $var_{tot\mathbf{Y}(\bar{\mathbf{Q}})}$ a variância total de \mathbf{Y} ou $\bar{\mathbf{Q}}$.

3.2.2 Recursos Computacionais

Os dados foram carregados e preparados por meio do software estatístico R (R CORE TEAM, 2021) versão 1.3.1073 de livre acesso. As transformações de log-razão indicadas para dados composicionais foram realizadas através do pacote *zcompositions* (PALAREA-ALBALADEJO e MARTÍN-FERNÁNDEZ, 2015).

Para a aplicação das técnicas de RDA e CCA clássicas e suas variantes polinomiais foi utilizado o software *RDACCA* (MAKARENKOV e LEGENDRE, 2001) de livre acesso.

4 RESULTADOS

Este capítulo tem como objetivo apresentar a análise dos resultados, onde, inicialmente serão apresentados os resultados referentes a análise de redundância clássica e sua extensão polinomial, seguido dos resultados referentes a análise de correspondência canônica clássica e polinomial.

4.1 Análise de Redundância

Os resultados registrados na Tabela 4.1 correspondem às principais métricas utilizadas para avaliar os modelos de análise de redundância. A média de R^2 é obtida através da média dos coeficientes de determinação de cada regressão efetuada durante o processamento de cada modelo, tanto para os lineares como para os polinomiais. A variância explicada é referente ao total da variância explicada por todos os eixos gerados em cada modelo. Por fim, os resultados dos testes de permutação de cada modelo e também o teste para diferença entre a modelos lineares e polinomiais.

Tabela 4.1 – Principais métricas de avaliação para os modelos de análise de redundância e suas transformações.

RDA Linear				
Transformação	Média R^2	Var. Explicada	Pseudo-F	Pseudo-F Diferença
Padronização	0,5567	55,6695	0,001	-
Additive Log Ratio	0,5228	58,6670	0,001	-
Centered Log Ratio	0,5228	58,6670	0,001	-
Isometric Log Ratio	0,5228	58,6670	0,001	-
RDA Polinomial				
Padronização	0,6221	62,2165	0,001	0,447
Additive Log Ratio	0,6371	70,7186	0,001	0,041
Centered Log Ratio	0,6372	70,4974	0,001	0,046
Isometric Log Ratio	0,6331	70,1243	0,001	0,045

Fonte: Autor (2022).

Observa-se que na Tabela 4.1, para os modelos lineares as médias dos coeficientes de determinação são iguais para as transformações baseadas em log-razão, com um pequeno ganho ao utilizar-se a padronização dos dados. Quanto a variância explicada esse cenário se inverte com as transformações de log-razão fornecendo um maior percentual de variância explicada das variáveis de notas sensoriais pelas diferentes proporções de blends de cafés. Ainda, pelo teste de permutação Pseudo-F todos os modelos de RDA linear foram significativos ao nível de significância $\alpha = 0,05$.

Nos modelos polinomiais as médias dos coeficientes de determinação apresentaram valores semelhantes, contudo os modelos de log-razão mostraram-se mais eficazes em explicar a variância total dos dados sensoriais. Essa eficácia é corroborada pelo fato de que, apenas esses modelos apresentaram significância estatística no teste Pseudo-F da diferença ao nível de $\alpha = 0,05$.

Considerando esses resultados e a grande quantidade de interpretações gráficas a serem feitas, optou-se por apresentar nesta seção apenas os resultados dos modelos que se referem aos modelos *ilr*, pelo fato de não possuírem limitações como as outras transformações. Além disso, para verificar as diferenças de se optar por uma abordagem que leve em consideração as restrições espaço amostral composicional, estes resultados serão comparados com os modelos com os dados padronizados, os quais não levam em consideração tais restrições. Os demais resultados se encontram na seção Anexo A.

4.1.1 Dados Padronizados

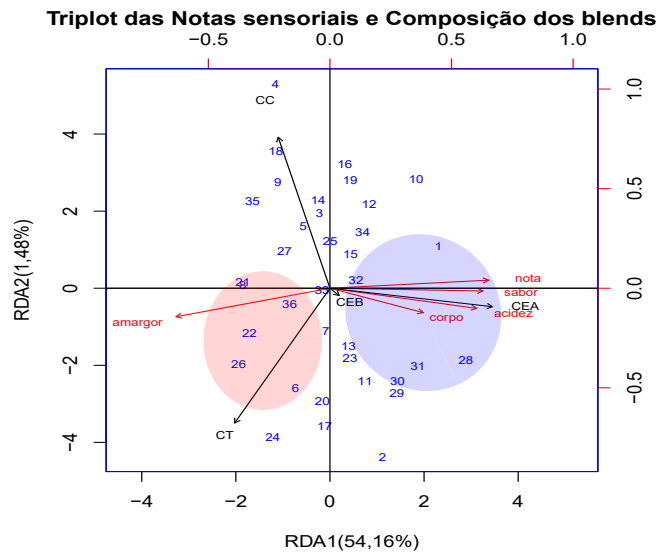
Tabela 4.2 – Resultados da Análise de Redundância para os dados padronizados.

CCA LINEAR					
Eixos canônicos	I	II	III	-	-
Autovalores em relação à variância total	2,7085	0,07427	0,0007	-	-
Fração da variância	54,1699	1,4854	0,0142	-	-
Fração da variância acumulada	54,1699	55,6553	55,6695	-	-
Média coeficiente de determinação múltipla	0,5567				
CCA POLINOMIAL					
Eixos canônicos	I	II	III	IV	V
Autovalores em relação à variância total	2,8748	0,1184	0,0704	0,0453	0,0018
Fração da variância	57,4959	2,3686	1,4083	0,9069	0,0367
Fração da variância acumulada	57,4959	59,8646	61,2729	62,1798	62,2165
Média coeficiente de determinação múltipla	0,6221				

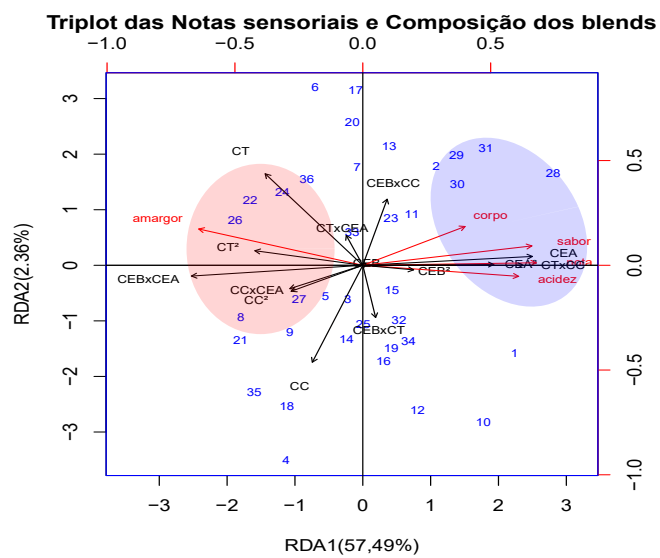
Fonte: Autor (2022).

Enquanto a RDA linear produziu três eixos canônicos (Tabela 4.2) explicando 55,67% da variação total dos dados de resposta, a abordagem polinomial gerou cinco eixos canônicos com 62,21% de variância explicada. Isso pode ser explicado pelo fato de que na Figura 4.1 os termos de interação e quadráticos relacionados as composições dos blends de cafés estão próximos das variáveis sensoriais *amargor*, *acidez* e *nota*.

Figura 4.1 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de análise de redundância para os dados padronizados.



(a)



(b)

Fonte: Autor (2022).

No triplot da Figura 4.1 (a) para a abordagem linear, é possível dividir as amostras em dois grandes grupos distintos conforme suas composições e notas atribuídas. O primeiro grupo identificado pela área em azul translúcido, corresponde às amostras que obtiveram melhores notas para os atributos *sabor*, *acidez*, *corpo* e *nota final*, uma vez que estão próximos aos vetores que os representam. Como esses vetores estão em ângulo bem próximo aos que representam as componentes de café Bourbon amarelo (CEB) e Acaia (CEA), as amostras pertencentes a essa área possuem maiores valores das mesmas, como as amostras puras 1 e 28 que possuem apenas um componente desses cafés especiais, ou amostras binárias (duas componentes) 31 e 32 que também possuem essas características.

O segundo grupo indicado pela área em vermelho representa as amostras de blends que estão mais associadas a maiores notas para o atributo *amargor* e também à componente de café comercial torrado (CT). Assim, as amostras de blends puros 26 e 8, blends binários 22 e 6 e ternário 21 que possuem em sua totalidade ou em partes, componentes de café Comercial Torrado (CT) receberam maiores notas para *amargor*.

O triplot do modelo polinomial da Figura 4.1 (b) apresenta uma configuração similar ao do modelo linear, a área em azul translúcido contém basicamente as mesmas amostras presentes na mesma área do modelo linear, com exceção da amostra binária 2 que desta vez apresentou maior similaridade com este grupo e as amostras 1 e 32 que mostraram maior dissimilaridade.

Em relação ao segundo grupo, houve maiores disparidades entre os modelos, a amostra 6 de café puro apresentou uma maior distância do vetor do atributo em questão, excluindo-a da área de relação. Pôde-se ainda, evidenciar a presença dos blends puros 9 e 27, dos blends binários 5 e 24 nesta área de agrupamento.

Resumidamente, em ambos os modelos, as amostras de café puros puderam ser discriminadas em relação aos blends, já que obtiveram coordenadas mais próximas as pontas dos vetores, que os aproximam do valor máximo das variáveis. Ainda, na área referente às amostras relacionadas ao atributo *amargor*, o modelo polinomial a incorporou mais amostras no agrupamento, explicado pelo acréscimo dos termos quadráticos e de interação das componentes de cafés.

4.1.2 Transformação Isometric Log-Ratio

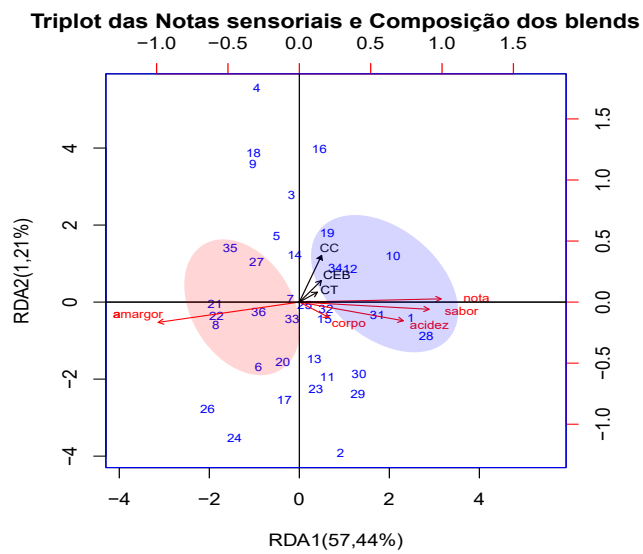
Tabela 4.3 – Resultados da Análise de Redundância para os dados composicionais transformados pelo método isometric log-ratio.

CCA LINEAR					
Eixos canônicos	I	II	III	-	-
Autovalores em relação à variância total	5,2586	0,1107	0,0010	-	-
Fração da variância	57,4451	1,2101	0,0117	-	-
Fração da variância acumulada	57,4451	58,6553	58,6670	-	-
Média coeficiente de determinação múltipla	0,5228				
CCA POLINOMIAL					
Eixos canônicos	I	II	III	IV	V
Autovalores em relação à variância total	6,2134	0,1226	0,0597	0,0217	0,0017
Fração da variância	67,8754	1,3398	0,6524	0,3273	0,0192
Fração da variância acumulada	67,8754	69,2152	69,8677	70,1051	70,1243
Média coeficiente de determinação múltipla	0,6331				

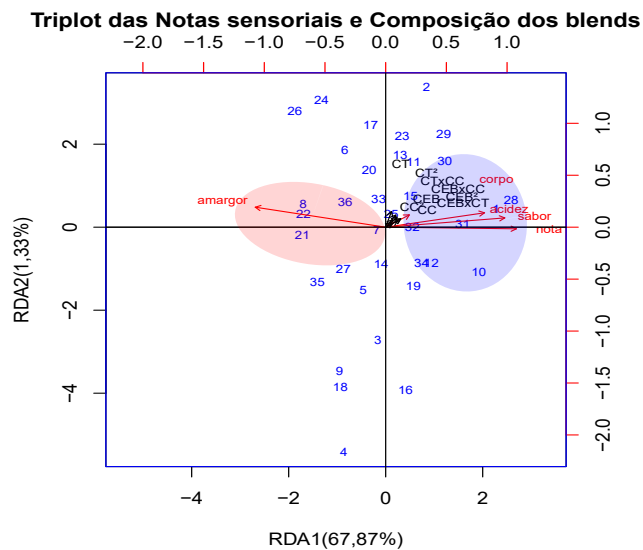
Fonte: Autor (2022).

No contexto dos dados que passaram pela transformação isométrica, o modelo linear produziu três eixos canônicos, os quais propiciaram uma explicação de 58,66% da variação total dos dados sensoriais. Em contrapartida, o modelo polinomial gerou cinco eixos canônicos reponsáveis por explicar 70,12% da variação total dos dados sensoriais, e além disso os dois primeiros eixos necessários para a formação dos biplots possuem uma explicação significamente maior (69,21%) do que em todos os eixos do modelo linear.

Figura 4.2 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de análise de redundância para os dados composicionais transformados pelo método isométrico log-ratio.



(a)



(b)

Fonte: Autor (2022).

Conforme o triplot da Figura 4.2 (a), pode-se assim como nos modelos anteriores, considerar dois grupos de amostras, o primeiro em azul translúcido com amostras associadas às notas sensoriais *corpo*, *acidez*, *sabor* e *nota final*. Nessa região, as amostras de cafés puros 1 e 28 se destacam das demais, se posicionando mais próximos dos valores máximos das notas sensoriais, especialmente dos atributos *acidez*, *sabor* e *nota final*, enquanto as amostras de blends binários 15 e 32 estão mais próximos do valor máximo do atributo *corpo*. No mesmo conglomerado, mas no primeiro quadrante a amostra de café puro 10 foi discriminada em relação aos outros blends do mesmo quadrante, esses últimos, em sua maioria, binários e ternários.

A segunda área agrupou 10 amostras de blends, definidas pela associação positiva com o atributo *amargor*, possui três amostras (8, 21 e 22) de café puro, blends binário e ternário respectivamente, fortemente associados à valores intermediários do atributo supracitado e outras três (7, 33 e 36) com as mesmas características fortemente associadas à valores mais baixos para *amargor*. À vista disso, concluí-se que para este grupo não foi possível observar diferenças significativas entre as amostras de café puro e os blends binários e ternários.

De modo similar ao modelo linear, na Figura 4.2 (b) do modelo polinomial, a área em azul translúcido mostra que as amostras de cafés puros 1 e 28 estão fortemente associadas à valores maiores dos atributos *acidez*, *sabor* e *nota final* e a amostra de café puro 10 moderadamente associada à valores maiores de *nota final*, a diferença é que para este modelo houve uma inversão de quadrante, onde as duas primeiras passaram a pertencer ao quadrante I e a terceira para o quadrante 4.

Na segunda área, observou-se o mesmo cenário, em que as amostras pertencentes ao grupo permaneceram as mesmas, porém as amostras que estavam no quadrante III passaram para o quadrante II deslocados pela direção do vetor do atributo *amargor*.

Os deslocamentos dos vetores dos atributos seguido das amostras, são ocasionados pela adição dos termos quadráticos e de interação presentes no quadrante I.

De modo geral, se comparado ao modelo polinomial com dados padronizados que apresentaram componentes mais dispersos, este modelo resultou em uma concentração das mesmas no primeiro quadrante do biplot. Nesse contexto, as componentes CC , CC^2 , $CEB \times CC$,

$CEB \times CT$, CEB^2 e CEB , apresentaram moderada correlação positiva com os atributos *sabor*, *acidez* e *nota final*, forte correlação positiva com o atributo *corpo*. Já as componentes CT e CT^2 mostraram uma forte correlação positiva com a variável sensorial *corpo*, fraca correlação com as variáveis *sabor*, *acidez* e *nota final* e fraca correlação positiva com a variável *amargor*.

4.2 Análise de Correspondência Canônica

De maneira similar a seção anterior, na Tabela 4.4 estão registradas as principais métricas utilizadas para avaliar os modelos de análise de correspondência canônica.

Tabela 4.4 – Principais métricas de avaliação para os modelos de análise de correspondência canônica e suas transformações.

CCA Linear				
Transformação	Média R^2	Var. Explicada	Pseudo-F	Pseudo-F Diferença
Qui-quadrado	0,5148	61,8665	0,001	
Additive Log Ratio	0,5162	59,4160	0,001	
Centered Log Ratio	0,5162	59,4160	0,001	
Isometric Log Ratio	0,5162	59,4160	0,001	
CCA Polinomial				
Qui-quadrado	0,6329	69,0054	0,001	0,300
Additive Log Ratio	0,6675	72,3490	0,001	0,025
Centered Log Ratio	0,6704	71,7732	0,001	0,036
Isometric Log Ratio	0,6602	70,5468	0,001	0,058

Fonte: Autor (2022).

Na Tabela 4.4 nota-se uma configuração de resultados similares aos ocorridos na análise de redundância, em que os resultados das médias dos coeficientes de determinação e variância explicada de cada abordagem considerando as diferentes transformações são bem próximos.

Tem-se que todos os modelos foram significativos, porém apenas os modelos que utilizaram as transformações additive log ratio e centered log ratio justificaram o emprego da abordagem polinomial, considerando o fato de que apenas estes foram significativos pelo teste pseudo-F da diferença ao nível de $\alpha = 0,05$.

Considerando estes resultados, esta seção considerou os modelos lineares e polinomiais cujo os dados foram submetidos às transformações *qui-quadrado* e *centered log ratio*, uma vez que, para essa última não foi necessário abrir mão de uma componente para o cálculo das razões. Outros resultados dessa seção encontram-se no Apêndice B.

4.2.1 Transformação Qui-quadrado

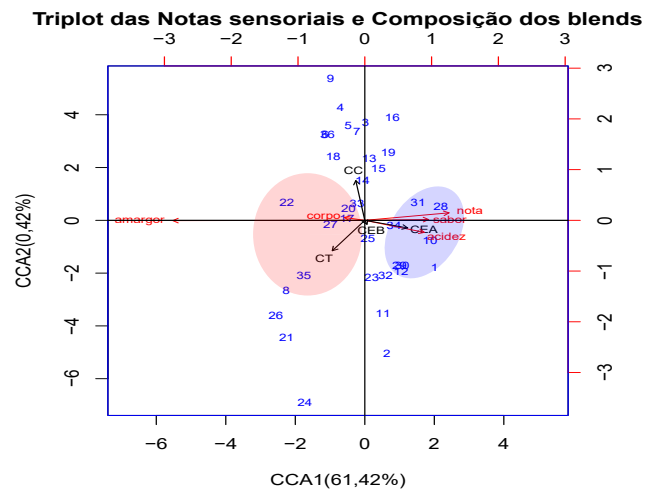
Tabela 4.5 – Resultados da análise de correspondência canônica para os dados transformados pelo método qui-quadrado.

CCA LINEAR					
Eixos canônicos	I	II	III	-	-
Autovalores em relação à variância total	0,0348	0,0002	0,0000	-	-
Fração da variância	61,4296	0,4251	0,0117	-	-
Fração da variância acumulada	61,4296	61,8548	61,8665	-	-
Média coeficiente de determinação múltipla	0,5148				
CCA POLINOMIAL					
Eixos canônicos	I	II	III	IV	V
Autovalores em relação à variância total	0,0380	0,0006	0,0003	0,0000	0,0000
Fração da variância	67,0860	1,1334	0,6273	0,1452	0,0133
Fração da variância acumulada	67,0860	68,2195	68,8468	68,9921	69,0054
Média coeficiente de determinação múltipla	0,6329				

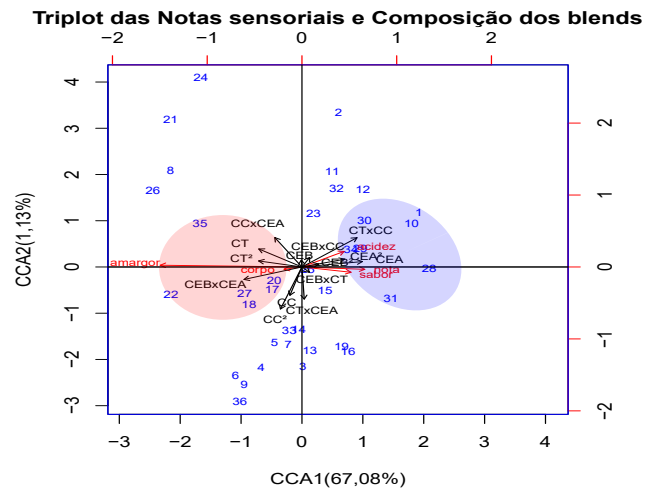
Fonte: Autor (2022).

Diante desses resultados observou-se que o modelo polinomial produziu cinco eixos canônicos, os quais corresponderam a 69% da variação total da matriz de atributos sensoriais. Os dois primeiros eixos explicam 68,21% da variação, valor superior aos 61,86% obtido pelos três eixos gerados pela abordagem linear.

Figura 4.3 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de de análise de correspondência canônica para os dados composicionais transformados pelo método qui-quadrado.



(a)



(b)

Fonte: Autor (2022).

Por meio da Figura 4.3 (a) referente ao modelo de correspondência canônica convencional, seguiu-se o padrão presente nos modelos de análise de redundância para seleção de grupos de amostras similares. Em um aspecto geral, a área em azul é caracterizada por amostras que estão correlacionadas positivamente aos atributos *nota final*, *sabor* e *acidez*, assim como as componentes de café Bourbon Amarelo e café Acaiá. A amostra de café puro Bourbon Amarelo 28 possui forte correlação com valores mais altos dos atributos *sabor* e *notas final* enquanto que a amostra de blend binário 31 está igualmente associada a esses atributos, porém para valores inferiores. Já, as amostras de café puro Acaiá 1 e 10 apresentaram forte correlação com altos valores para o atributo *acidez*, enquanto a amostra binária 34 composta de café Acaiá e Conilon mostrou-se associada à valores menores do atributo mencionado.

No segundo grupo, que na análise de redundância foi definida por amostras correlacionadas apenas ao atributo *amargor*, desta vez é caracterizada também pela presença do atributo sensorial de *corpo*. A amostra de café puro Conilon 27 e as amostras de blends binários apresentaram forte correlação com maiores valores para o atributo *corpo*, não mostrando dissimilaridades quanto a classificação das amostras. As demais amostras se mostraram correlacionadas à valores intermediários para *amargor*.

Para o modelo polinomial, a Figura 4.3 (b) percebe-se que na área em azul as posições das variáveis sensoriais permanecem aproximadamente as mesmas em comparação com o gráfico do modelo linear, com exceção da variável *acidez*. Neste aspecto, a diferença se deu na disposição das componentes, agora considerando os termos quadráticos e de interação e a aproximação dos blends binários para as pontas dos vetores de atributos caracterizando atribuição de notas sensoriais maiores. O mesmo comportamento foi observado na área vermelha em que as amostras se aproximaram dos pontos máximos dos vetores de atributos. Isso explica o fato de que não houve diferença significativa entre os dois modelos pelo teste pseudo-F.

4.2.2 Transformação Centered Log-Ratio

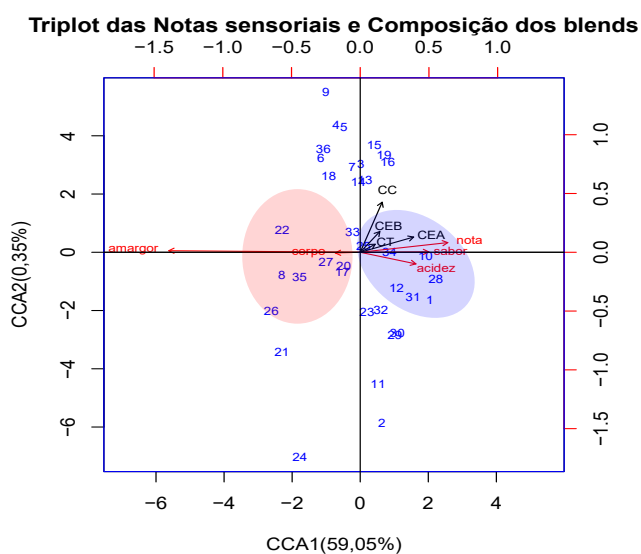
Tabela 4.6 – Resultados da análise de correspondência canônica para os dados composicionais transformados pelo método centered log-ratio.

CCA LINEAR					
Eixos canônicos	I	II	III	-	-
Autovalores em relação à variância total	0,0334	0,0002	0,0000	-	-
Fração da variância	59,0581	0,3509	0,0068	-	-
Fração da variância acumulada	59,0581	59,4091	59,4160	-	-
Média coeficiente de determinação múltipla	0,5162				
CCA POLINOMIAL					
Eixos canônicos	I	II	III	IV	V
Autovalores em relação à variância total	0,0396	0,0005	0,0003	0,0001	0,0000
Fração da variância	69,9326	0,9072	0,6597	0,2583	0,0152
Fração da variância acumulada	69,9326	70,8398	71,4996	71,7579	71,7732
Média coeficiente de determinação múltipla	0,6704				

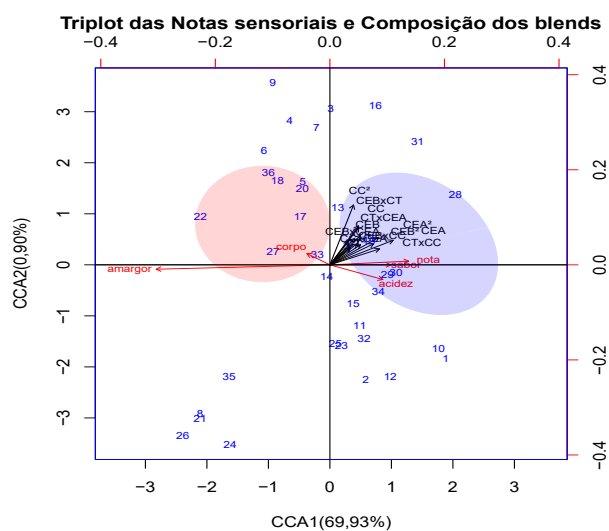
Fonte: Autor (2022).

Se tratando dos modelos resultantes da transformação *centered log ratio*, assim como nos modelos com os dados transformados pelo método qui-quadrado, a abordagem linear produziu três eixos canônicos enquanto que a polinomial produziu cinco eixos e o percentual de variância explicada foram de 51,62% e 71,77% respectivamente, para todos os eixos canônicos.

Figura 4.4 – Representação gráfica para o modelo linear (a) e modelo polinomial (b) de de análise de correspondência canônica para os dados composicionais transformados pelo método centered log-ratio.



(a)



(b)

Fonte: Autor (2022).

Na Figura 4.4 (a), a área em azul agregou 7 amostras, sendo que a amostra de café puro Acaiá 10 apresentou forte associação com valores máximos dos vetores dos atributos *sabor* e *acidez*, já a amostra 28 do mesmo café puro apresentou forte associação com maiores valores do atributo *acidez*, mas uma associação menor com o atributo *sabor*, assim como a amostra de café Boubon amarelo, porém com intensidade de associação um pouco mais moderada.

Na mesma figura, a área em vermelho agrupou 7 amostras, onde a amostra de café puro Conilon 27, a amostra de café puro comercial torrado 17 e amostra de blend binário 20 se mostraram fortemente correlacionadas à maiores notas para o atributo *corpo*. Ainda, as amostras de café puro comercial torrado 8 e 35 apresentaram forte associação com valores intermediários do atributo *amargor*, assim como a amostra de blend binário 22.

Se tratando do modelo polinomial da Figura 4.4 (b), a área azul dessa abordagem apresentou uma maior dissimilaridade entre os cafés puros as amostras de blends binários se comparados ao modelo linear clássico, visto o maior distanciamento entre as amostras. Dessa vez, as amostras de blends binário 29 e ternário 30 apresentaram forte correlação com maiores valores dos atributos *acidez* e *sabor*, enquanto a amostra binária 34, forte correlação com *acidez* e moderada com *sabor*.

Na comparação entre as áreas em vermelho das abordagens, a amostra binária 33 se aproximou mais de valores maiores para o atributo *corpo*, ao passo que a amostra de café puro conilon 27 foi distanciada das amostras de café puro comercial torrado 17 e da amostra binária 20.

5 DISCUSSÃO

Como a metodologia deste trabalho envolveu várias combinações de métodos empregados a um mesmo banco de dados, é pertinente comparar o desempenho de cada método considerando seus propósitos e suas particularidades.

A respeito das transformações, notou-se que as derivadas de razões logarítmicas apresentaram resultados similares, tanto na média do coeficiente de determinação quanto ao percentual de variância explicada, conforme resultados encontrados em Messias (2016) que utilizou essas transformações em dados composicionais para aplicação da análise de componentes principais.

Essas três transformações também produziram biplots equivalentes entre si, com os modelos lineares possuindo maior semelhança que os polinomiais. Isso indica que, a medida que novas variáveis são inseridas nos modelos, maior será a dissimilaridade nos resultados dessas transformações, corroborando mais uma vez os achados no trabalho de Messias (2016).

Ao comparar as transformações de razões logarítmicas com as demais, evidenciou-se um pequeno ganho no percentual de variância explicada quando se utilizou os primeiros métodos nos modelos de análise de redundância linear. Já nos modelos de análise de redundância polinomial, esse ganho se mostrou ainda maior, o que permitiu a significância através do teste Pseudo-F da diferença. Nos modelos de análise de correspondência canônicas, essas diferenças foram menos evidentes.

Os biplots dos modelos *alr*, *clr* e *ilr* apresentaram um agrupamento das componentes no primeiro quadrante, enquanto a padronização na RDA e a transformação qui-quadrado, essas componentes possuíram coordenadas mais dispersas em todo o gráfico.

De forma teórica e sucinta, segundo Legendre e Legendre (2012) a diferença entre a análise de redundância e a análise de correspondência canônica, é que na primeira utiliza-se a regressão linear para ajustar os valores de \mathbf{Y} e a segunda ajusta esses valores por meio da regressão linear ponderada. Dito isso, as soluções dadas pela análise de correspondência canônica diferem muito pouco da análise de redundância, tanto nas métricas de avaliação como no biplots obtidos, especialmente se considerarmos o uso das transformações de *log-razão*.

Em grande parte dos modelos aqui construídos, a abordagem polinomial forneceu uma maior porcentagem de variância explicada significativa em detrimento dos modelos lineares. Em termos gerais, nesses casos observa-se uma maior relação das notas sensoriais com termos de interação e quadráticos dos blends. O mesmo comportamento foi observado em Makarenkov e Legendre (2002), em que houve ganho significativo na variância explicada na distribuição de espécies de aranhas por termos quadráticos e de interação de variáveis ambientais. Nos triplots gerados por esses autores, a maioria das espécies de aranhas estavam mais associadas a termos quadráticos das variáveis *água*, *reflexo do solo* e a interação entre as duas.

No que diz respeito a análise sensorial dos blends, as abordagens convencionais das técnicas canônicas forneceram valores de variância explicada próximos a 60%. A integração de termos que levam a melhor aproximação quadrática dos resíduos por meio do algoritmo polinomial levou à um aumento significativo na proporção de variância explicada, com valores superiores a 70%. A modelagem polinomial resultou em percentuais condizentes com os encontrados em outros trabalhos, como Costa et.al (2018) que obtiveram resultados satisfatórios para a análise de correspondência utilizando diferentes métricas de distâncias.

Quanto aos triplots, vale destacar que suas interpretações são bastantes subjetivas cabendo ao pesquisador estabelecer critérios de similaridades em relação as distâncias entre amostras e variáveis. A partir dessa compreensão, segue-se que para todos os modelos de análise de redundância apresentam um padrão em que boa parte das amostras são subdivididas em dois grupos antagônicos no triplot. Em um primeiro padrão, as amostras estão correlacionadas positivamente com os atributos *sabor*, *acidez*, *corpo* e *nota final*, enquanto que, no segundo padrão as amostras estão correlacionadas positivamente com o atributo *amargor*.

Considerando o modelo polinomial isometric log-ratio, no primeiro grupo as amostras de cafés puros especiais Bourbon Amarelo e Acaiá não foram discriminadas em relação a *sabor* e *acidez*, uma vez que apresentaram coordenadas similares no gráfico. Utilizando modelos mistos para comparar os efeitos de cada componente dos blends, Paulino et. al (2019) verificou os mesmos resultados. Ainda, foi possível observar que os blends puderam ser discriminados em relação as amostras de cafés puros, com os blends binários e ternários em pontos mais

próximos do início e até metade das setas dos atributos. Essa diferenciação entre as categorias de blends também foi observada em Cirillo et.al (2019) que propuseram painel sensorial de blends de cafés através da distribuição de dados extremos para os scores.

No segundo grupo não houve diferenciação entre categorias de blends, porém foi possível observar que as amostras que tiveram maior correlação com maiores notas do atributo *amargor*, possuíam em sua composição componentes de café comercial torrado e ausência de café Acaiá. Há, portanto, evidências de que o café comercial torrado tenha menor influência no café acaiá em relação ao efeito do atributo *amargor* em detrimento do cafés Bourbon Amarelo e Conilon.

Para o modelo polinomial centered log-ratio de análise de correspondência canônica, as amostras de café puro Bourbon Amarelo e Acaiá foram discriminadas quanto a sua concentração, as amostras com nível de concentração de 0,07 m/v foram correlacionadas ao atributo *acidez*, enquanto que as amostras com concentração 0,10 m/v estabeleceram maior associação com os atributos *sabor* e *nota final*. A semelhança entre estes dois tipos de cafés especiais podem ser atestadas em Ossani et al. (2017) e justificadas pela condição genética e condições ambientais como altitude, e neste caso mostram que os cafés são semelhantes até em diferentes níveis de concentração.

A segunda área desse modelo, além do atributo *amargor* como na análise de redundância, também passou a ser relacionada com o atributo *corpo* e passou a aglomerar outras amostras de café Conilon que não estavam presentes na análise de redundância. De acordo com Lima Filho et al. (2015) a indústria utiliza certas variedades de café *Canephora* como o Robusta para a formulação de blends em conjunto com cafés especiais, dado que o Robusta possui a característica de ser mais amargo e encorpado, adicionando essas características aos cafés especiais que são mais frutados e ácidos.

6 CONCLUSÃO

Tendo em vista os objetivos deste trabalho e resultados obtidos, conclui-se que a análise de redundância e análise de correspondência polinomial apresentaram maiores percentuais de variância explicada dos dados sensoriais de amostras de blends de cafés. Ainda, forneceram interpretações gráficas com particularidades não compreendidas pelos modelos convencionais e que se mostraram presentes em outros estudos de análise sensorial de blends de cafés.

REFERÊNCIAS

- AITCHISON, J. The statistical analysis of compositional data. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 44, n. 2, p. 139-160, 1982.
- AITCHISON, J. **The statistical analysis of compositional data. Monographs on statistics and applied probability**. London: Chapman and Hall, 1986.
- BLASIUS, J.; GREENACRE, M. **Visualization of categorical data**. San Diego: Academic Press, 1998.
- CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications**. 1st. ed. New York: Cambridge university press, 2005.
- CIRILLO, M. A. et al. Statistical procedure for the composition of a sensory panel of blends of coffee with different qualities using the distribution of the extremes of the highest scores. **Acta Scientiarum. Agronomy**, v. 41, e 2019.
- COSTA, A. L. A.; BRIGHENTI, C. R. G.; CIRILLO, M. A. A new approach to simple correspondence analysis with emphasis on the violation of the independence assumption of the levels of categorical variables. **Acta Scientiarum. Technology**, v. 40, p. e34953-e34953, 2018.
- DA COSTA, A. S. et al. Proposal of a metric selection index for correspondence analysis: an application in the sensory evaluation of coffee blends. **Semina: Ciências Agrárias**, v. 41, n. 2, p. 479-492, 2020.
- EGOZCUE, J. J. et al. Isometric logratio transformations for compositional data analysis. **Mathematical geology**, v. 35, n. 3, p. 279-300, 2003.
- FENTY, J. Analyzing distances. **The Stata Journal**, v. 4, n. 1, p. 1-26, 2004.
- FERREIRA, H. A. **Componentes de efeitos de safras representados em biplots corrigidos por predições de modelos GEE na classificação granulométrica de cafés**. 2019. 55 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2019.
- FISHER, R. A. **The design of experiments**. 1st. ed. London: Oliver & Boyd, 1935.
- GABRIEL, K. R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v. 58, n. 3, p. 453-467, 1971.

GUIMARÃES, E. R. **Terceira Onda do Café: Base Conceitual e Aplicações**. 2016. Tese (Doutorado em Administração) - Universidade Federal de Lavras, Lavras, 2016.

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

JOLLIFFE, I. T. **Principal component analysis for special types of data**. 2nd. ed. New York: Springer, 2002.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th. ed. New Jersey: Prentice hall, 1998.

KRUSKAL, J. B; WISH, M. **Multidimensional scaling**. Newbury Park: Sage, 1978.

LAZRAQ, A.; CLÉROUX, R. Testing the significance of the successive components in redundancy analysis. **Psychometrika**, v. 67, n. 3, p. 411-419, 2002.

LEGENDRE, Pierre; LEGENDRE, Louis. **Numerical ecology**. 3rd. ed. Amsterdam: Elsevier, 2012.

LIMA FILHO, T. et al. Características físico-químicas de bebidas de café tipo expresso preparadas a partir de blends de café arábica e conilon. **Revista Ceres**, v. 62, n. 4, p. 333-339, 2015.

MAKARENKOV, V.; LEGENDRE, P. Program Polynomial RdaCca—User's notes. **Département de sciences biologiques, Université de Montréal**, Montréal, 2001.

MAKARENKOV, V.; LEGENDRE, P. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. **Ecology**, v. 83, n. 4, p. 1146-1161, 2002.

MESSIAS, Ricardo Matioli. **Transformações em dados composicionais para a aplicação da análise de componentes principais**. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, 2016.

OSSANI, P. C. et al. Qualidade de cafés especiais: uma avaliação sensorial feita com consumidores utilizando a técnica MFACT1. **Revista Ciência Agronômica**, v. 48, p. 92-100, 2017.

PALAREA-ALBALADEJO, J.; MARTÍN-FERNÁNDEZ, J. A. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. **Chemometrics and Intelligent Laboratory Systems**, v. 143, p. 85-96, 2015.

PAULINO, A. L. B. et al. A mixed model applied to joint analysis in experiments with coffee blends using the least squares method. **Revista Ciência Agronômica**, v. 50, n. 3, p. 345-352, 2019.

PAWLOWSKY-GLAHN, V.; EGOZCUE, J. J.; TOLOSANA-DELGADO, R. **Modeling and analysis of compositional data**. Chichester: John Wiley & Sons, 2015.

PEREIRA, A. B. **Avaliação de metodologias alternativas de estimativa para variáveis composicionais em depósitos multivariados**. 2020. 136 p. Dissertação (Mestrado em Engenharia Mineral) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 2020.

RENCHER, Alvin C.; SCHAALJE, G. Bruce. **Linear models in statistics**. 2nd. ed. Hoboken: John Wiley & Sons, 2008.

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. et al. Interaction of genotype, environment and processing in the chemical composition expression and sensorial quality of Arabica coffee. **African Journal of Agricultural Research**, v. 11, n. 27, p. 2412-2422, 2016.

ŠMILAUER, P.; LEPŠ, J. **Multivariate analysis of ecological data using CANOCO 5**. 2nd. ed. New York: Cambridge university press, 2014.

STEWART, D.; LOVE, W. A general canonical correlation index. **Psychological bulletin**, v. 70, n. 3p1, p. 160, 1968.

TER BRAAK, C. J. F. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. **Ecology**, v. 67, n. 5, p. 1167-1179, 1986.

TER BRAAK, C. J. F. Canonical community ordination. Part I: Basic theory and linear methods. **Ecoscience**, v. 1, n. 2, p. 127-140, 1994.

TORRES-SALINAS, D. et al. On the use of biplot analysis for multivariate bibliometric and scientific indicators. **Journal of the American Society for Information Science and Technology**, [S.l.] v. 64, n. 7, p. 1468-1479, 2013.

VAN DEN BOOGAART, K. G.; TOLOSANA-DELGADO, R. **Analyzing compositional data with R**. Berlin: Springer, 2013.

VAN DEN WOLLENBERG, A. L. Redundancy analysis an alternative for canonical correlation analysis. **Psychometrika**, v. 42, n. 2, p. 207-219, 1977.

APÊNDICE A – Resultados Análise de Redundância

ADDITIVE LOG-RATIO

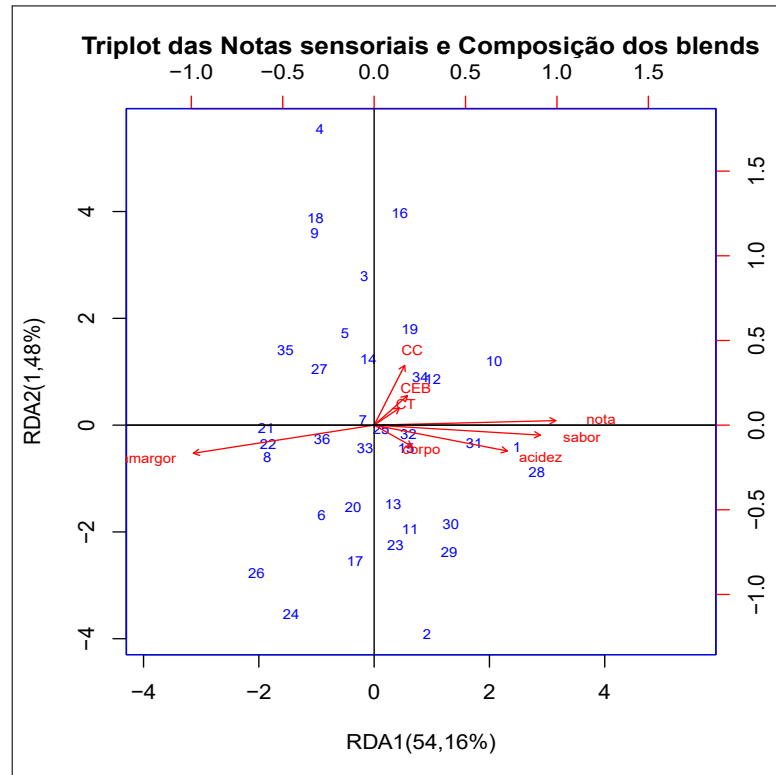
Modelo Linear

Tabela 1 – Resultados da Análise de Redundância linear para os dados composicionais transformados pelo método additive log-ratio.

Eixos canônicos		
I	II	III
Autovalores em relação à variância total de (Y = 5,37050)		
5,25865	0,11078	0,00107
Fração da variância total em Y		
57,44519	1,21017	0,01171
Fração acumulada da variância total de Y nos eixos I, II e III		
57,44519	58.65536	58.66707
Média coeficiente de determinação múltipla = 0,52284		

Fonte: Autor (2022).

Figura 1 – Representação gráfica para o modelo linear de análise de redundância para os dados composicionais transformados pelo método additive log-ratio.



Fonte: Autor (2022).

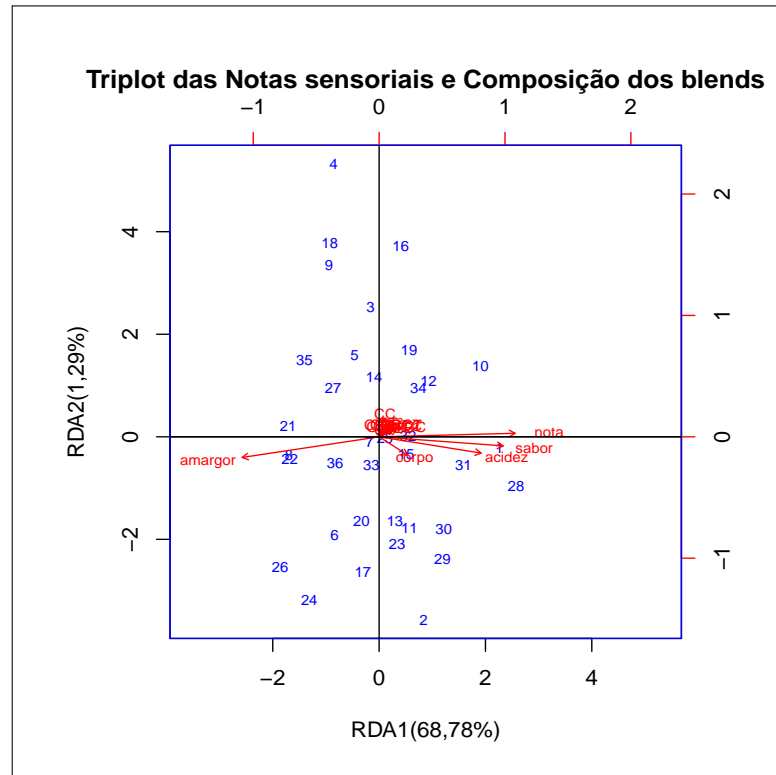
Modelo Polinomial

Tabela 2 – Resultados da Análise de Redundância polinomial para os dados composicionais transformados pelo método additive log-ratio.

Eixos canônicos				
I	II	III	IV	V
Autovalores em relação à variância total de (Y = 6,47372)				
6,29662	0,11858	0,04124	0,01605	0,00125
Fração da variância total em Y				
68,78393	1,29533	0,45047	0,17529	0,01363
Fração acumulada da variância total de Y nos eixos I, II e III				
68,78393	70,07926	70,52973	70,70502	70,71864
Média coeficiente de determinação múltipla = 0,63715				

Fonte: Autor (2022).

Figura 2 – Representação gráfica para o modelo polinomial de análise de redundância para os dados composicionais transformados pelo método additive log-ratio.



Fonte: Autor (2022).

CENTERED LOG-RATIO

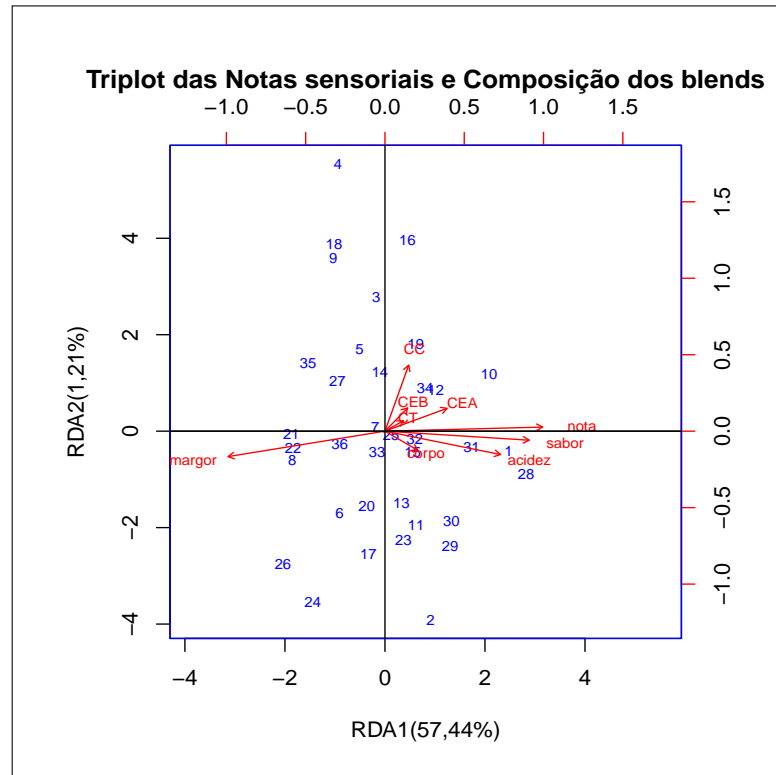
Modelo Linear

Tabela 3 – Resultados da Análise de Redundância linear para os dados composicionais transformados pelo método centered log-ratio.

Eixos canônicos		
I	II	III
Autovalores em relação à variância total de (Y = 5,37050)		
5,25865	0,11078	0,00107
Fração da variância total em Y		
57,44519	1,21017	0,01171
Fração acumulada da variância total de Y nos eixos I, II e III		
57,44519	58.65536	58.66707
Média coeficiente de determinação múltipla = 0,52284		

Fonte: Autor (2022).

Figura 3 – Representação gráfica para o modelo linear de análise de redundância para os dados composicionais transformados pelo método centered log-ratio.



Fonte: Autor (2022).

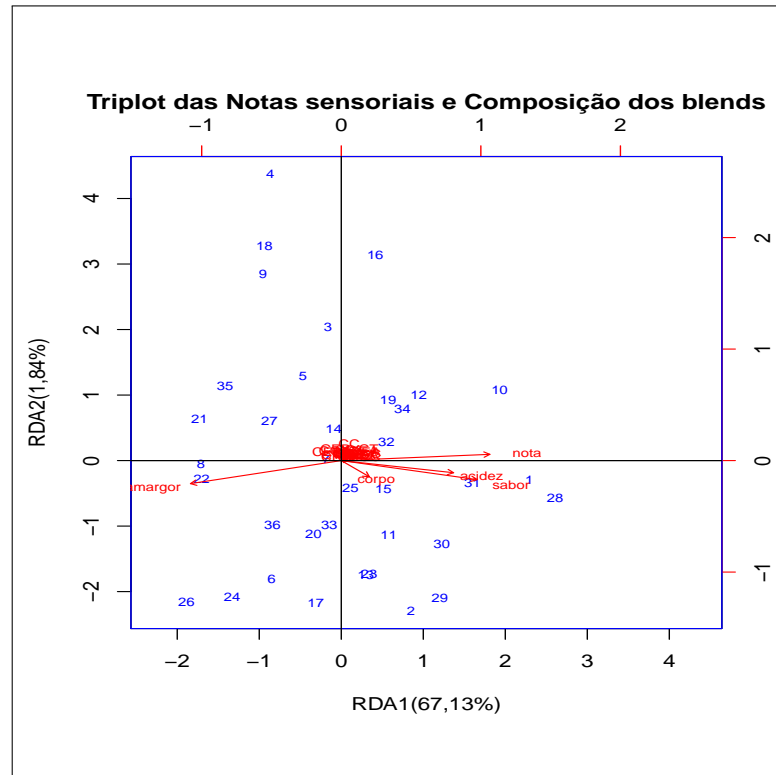
Modelo Polinomial

Tabela 4 – Resultados da Análise de Redundância polinomial para os dados composicionais transformados pelo método centered log-ratio.

Eixos canônicos				
I	II	III	IV	V
Autovalores em relação à variância total de (Y = 6,45348)				
6,14547	0,16845	0,09020	0,03971	0,00965
Fração da variância total em Y				
67,13286	1,84015	0,98532	0,43379	0,10536
Fração acumulada da variância total de Y nos eixos I, II e III				
67,13286	68,97301	70,52973	70,39212	70,49748
Média coeficiente de determinação múltipla = 0,63721				

Fonte: Autor (2022).

Figura 4 – Representação gráfica para o modelo polinomial de análise de redundância para os dados composicionais transformados pelo método centered log-ratio.



Fonte: Autor (2022).

APÊNDICE B – Resultados Análise de Correspondência Canônica

ADDITIVE LOG-RATIO

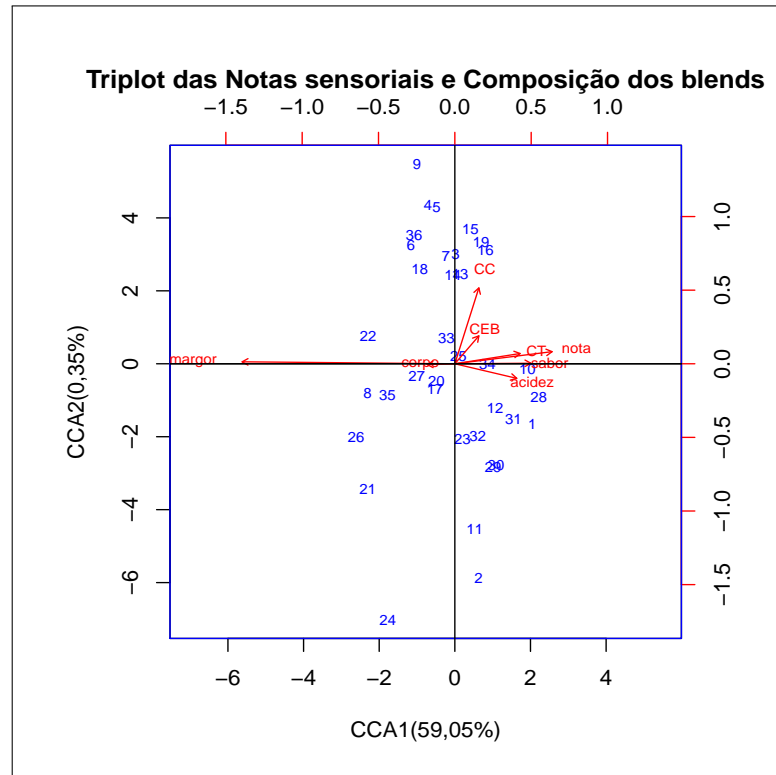
Modelo Linear

Tabela 5 – Resultados da análise de correspondência canônica linear para os dados composicionais transformados pelo método additive log-ratio.

Eixos canônicos		
I	II	III
Autovalores em relação à variância total de ($\bar{Q} = 0,03367$)		
0,03347	0,00020	0,00000
Fração da variância total em Y		
59,05818	0,35099	0,00688
Fração acumulada da variância total de Y nos eixos I, II e III		
59,05818	59,40917	59,41606
Média coeficiente de determinação múltipla = 0,51622		

Fonte: Autor (2022).

Figura 5 – Representação gráfica para o modelo linear de análise de correspondência canônica para os dados composicionais transformados pelo método additive log-ratio.



Fonte: Autor (2022).

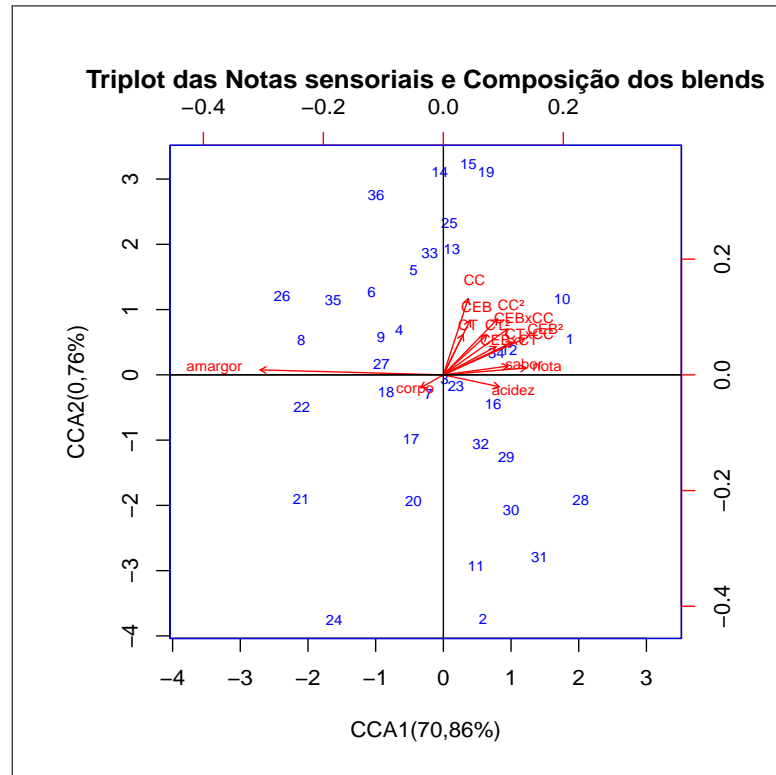
Modelo Polinomial

Tabela 6 – Resultados da análise de correspondência canônica polinomial para os dados composicionais transformados pelo método additive log-ratio.

Eixos canônicos				
I	II	III	IV	V
Autovalores em relação à variância total de ($\bar{Q} = 0,04100$)				
0,04016	0,00043	0,00031	0,00009	0,00002
Fração da variância total em Y				
70,86254	0,76722	0,53960	0,15079	0,02893
Fração acumulada da variância total de Y nos eixos I, II e III				
70,86254	71,62975	72,16936	72,32015	72,34908
Média coeficiente de determinação múltipla = 0,66751				

Fonte: Autor (2022).

Figura 6 – Representação gráfica para o modelo polinomial de análise correspondência canônica para dados composicionais transformados pelo método additive log-ratio..



Fonte: Autor (2022).

ISOMETRIC LOG-RATIO

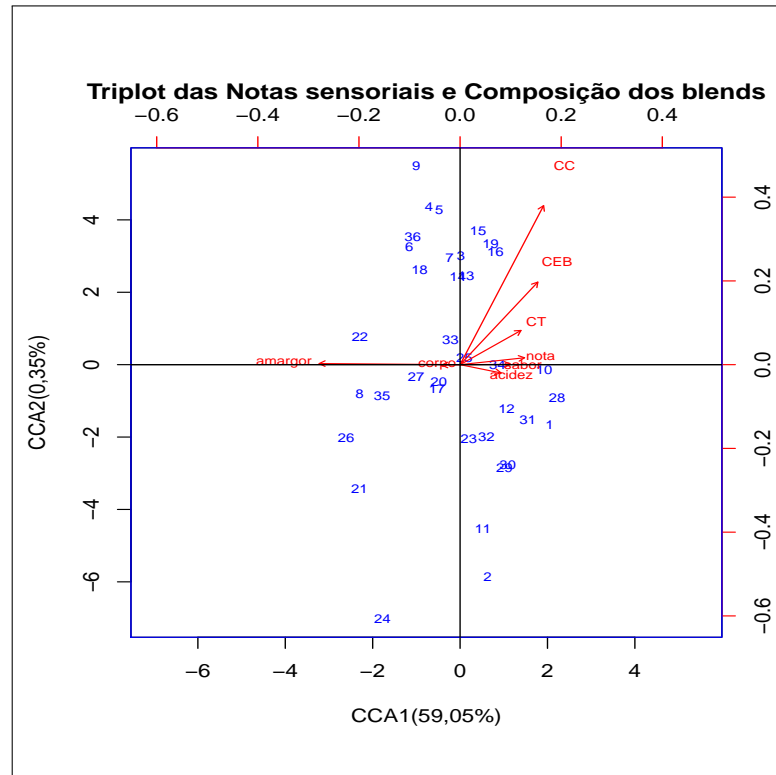
Modelo Linear

Tabela 7 – Resultados da análise de correspondência canônica linear para os dados composicionais transformados pelo método isometric log-ratio.

Eixos canônicos		
I	II	III
Autovalores em relação à variância total de ($\bar{Q} = 0.03367$)		
0,03347	0,00020	0,00000
Fração da variância total em Y		
59,05818	0,35099	0,00688
Fração acumulada da variância total de Y nos eixos I, II e III		
59,05818	59,40917	59,41606
Média coeficiente de determinação múltipla = 0.51622		

Fonte: Autor (2022).

Figura 7 – Representação gráfica para o modelo linear de análise de correspondência canônica para os dados composicionais transformados pelo método isometric log-ratio.



Fonte: Autor (2022).

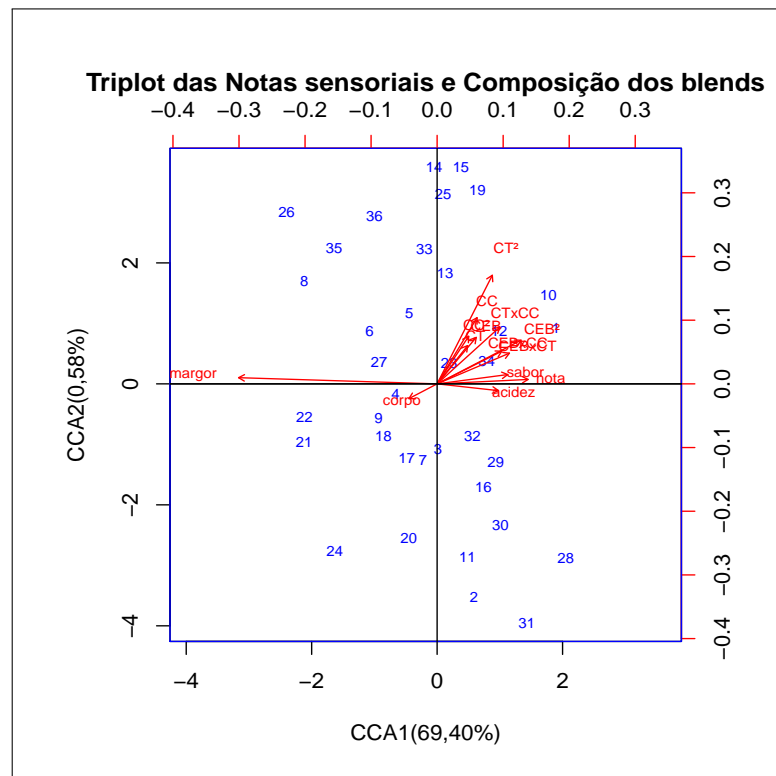
Modelo Polinomial

Tabela 8 – Resultados da análise de correspondência canônica polinomial para os dados composicionais transformados pelo método isometric log-ratio.

Eixos canônicos				
I	II	III	IV	V
Autovalores em relação à variância total de ($\bar{Q} = 0,03998$)				
0,03933	0,00033	0,00023	0,00007	0,00002
Fração da variância total em Y				
69,40327	0,58401	0,40403	0,11934	0,03622
Fração acumulada da variância total de Y nos eixos I, II e III				
69,40327	69,98728	70,39131	70,51065	70,54687
Média coeficiente de determinação múltipla = 0,66025				

Fonte: Autor (2022).

Figura 8 – Representação gráfica para o modelo polinomial de análise correspondência canônica para dados composicionais transformados pelo método isometric log-ratio.



Fonte: Autor (2022).