



MARIANA RESENDE

**ÍNDICES DE QUALIDADE DE AJUSTE DE MODELOS DE
EQUAÇÕES ESTRUTURAIS CONSIDERANDO REPETIÇÕES
E TRATAMENTOS COM APLICAÇÃO EM ANÁLISE
SENSORIAL DE CAFÉS**

**LAVRAS - MG
2022**

MARIANA RESENDE

**ÍNDICES DE QUALIDADE DE AJUSTE DE MODELOS DE EQUAÇÕES
ESTRUTURAIS CONSIDERANDO REPETIÇÕES E TRATAMENTOS COM
APLICAÇÃO EM ANÁLISE SENSORIAL DE CAFÉS**

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

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(a) autor(a).

Resende, Mariana.

Índices de qualidade de ajuste de modelos de equações estruturais considerando repetições e tratamentos com aplicação em análise sensorial de cafês / Mariana Resende. - 2022.

62 p. : il.

Orientador: Marcelo Ângelo Cirillo.

Tese (Doutorado) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Simulação. 2. *Outliers*. 3. GFI. I. Cirillo, Marcelo Ângelo.
II. Título.

MARIANA RESENDE

**ÍNDICES DE QUALIDADE DE AJUSTE DE MODELOS DE EQUAÇÕES
ESTRUTURAIS CONSIDERANDO REPETIÇÕES E TRATAMENTOS COM
APLICAÇÃO EM ANÁLISE SENSORIAL DE CAFÉS**

**QUALITY INDICES OF STRUCTURAL EQUATION MODEL ADJUSTMENT
CONSIDERING REPLICATES AND TREATMENTS WITH APPLICATION IN THE
SENSORY ANALYSIS OF COFFEE**

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

APROVADA em 23 de maio de 2022.

Dra. Evelise Roman Corbalan Góis Freire UFPA

Dr. Gilberto Rodrigues Liska UFSCAR

Dra. Lúcia Pereira Barroso USP

Dr. Renato Ribeiro de Lima UFPA

Prof. DSc. Marcelo Ângelo Cirillo
Orientador

**LAVRAS - MG
2022**

*Dedico esse trabalho à minha mãe, Bete (in memoriam),
meu maior exemplo de amor e força, que me ensinou que
devemos lutar até o fim sempre, independente das
circunstâncias. Sua lembrança me inspira e faz persistir.*

AGRADECIMENTOS

A Deus por ter me oferecido o dom da vida, por ter me regido, me iluminado e dado forças para a concretização deste trabalho.

À minha mãe, Bete, pelo amor, paciência, incentivo e apoio incondicional, em todas as minhas decisões nas diferentes etapas da minha vida. Que me ensinou ter valores e ser a pessoa que sou hoje. Sei que, mesmo não estando presente de forma física, continuou participando de cada segundo nesta jornada. Amo você mais que tudo nesta vida.

Ao meu irmão João Marcelo, por sempre me encorajar, nunca me deixando desistir e por estar sempre segurando minha mão.

Às minhas primas-irmãs, Sabrina e Camila, por todo o amor, carinho e por estarem sempre ao meu lado me ajudando a levantar nos tropeços da vida e vibrando com cada uma de minhas vitórias.

Ao meu marido, Diones, por todo o amor e apoio, por entender minhas ausências muitas vezes.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pelos ensinamentos. Em particular, ao meu orientador, professor Marcelo Ângelo Cirillo, pelas orientações, ensinamentos, apoio e compreensão constantes no momento mais difícil da minha vida. Serei eternamente grata. Fica minha admiração pela pessoa e profissional.

À Universidade Federal de Lavras e ao Departamento de Estatística pela oportunidade concedida para a realização deste doutorado.

Ao CNPq, pela bolsa de estudos concedida, fundamental para esta conquista.

RESUMO

A Modelagem de Equações Estruturais (MEE) é uma técnica multivariada que permite o estudo de diversas relações simultaneamente, inclusive envolvendo variáveis não observáveis. Em geral, a qualidade de ajuste de um modelo de equações estruturais é verificada por diversos índices, nos quais, em síntese, avaliam a correspondência entre a matriz de covariância amostral e a covariância implícita pelo modelo que represente apropriadamente as inter-relações entre os indicadores envolvidos no estudo, porém, em nenhum deles, a repetição entre as variáveis observadas são consideradas. Com essa motivação, este trabalho tem por objetivo propor a modificação em índices GFI e AGFI considerando repetições e tratamentos nas variáveis observadas e aplicação à análise sensorial de cafés. A validação foi feita por simulação Monte Carlo, em diversos cenários, representados por diferentes números de repetições, graus de heterogeneidade e quantidades de *outliers*, geradas por distribuições com desvios de simetria e excesso de curtose. O estudo de simulação mostrou que o aprimoramento dos índices de validade de ajuste GFI e AGFI foi promissor, uma vez que apresentou propriedades de robustez, em relação à presença de *outliers* e mostrou evidências de diagnóstico de um modelo com bom e ruim ajuste, mediante os graus de heterogeneidade impostos entre as matrizes de covariâncias amostral e hipotetizada pelo modelo. Em relação ao estudo de cafés especiais, dado o modelo proposto, as informações representadas, nas variáveis latentes, produção e variações ambientais, de fato, influenciam na percepção sensorial para os cafés produzidos na região da Serra da Mantiqueira e o aprimoramento dos índices apresentou melhorias, em validação do modelo, dada a inclusão das repetições das variáveis observadas.

Palavras-chave: Simulação. *Outliers*. GFI. AGFI. Sabor. Acidez.

ABSTRACT

Structural Equation Modeling (SEM) is a multivariate technique that allows studying several relationships simultaneously, including those involving unobservable variables. In general, the structural equation model goodness-of-fit is verified through several indices, which, in short, evaluate the correspondence between the sample covariance matrix and that implied by the model that properly represents the interrelationships among indexes involved in the study. However, none of them considers the repetition between the observed variables. In this sense, this study aims to propose the correction of GFI and AGFI indices considering observed variable repetitions and their application to coffee sensory analysis. Monte Carlo simulation validation was performed under different scenarios, which are represented by different numbers of repetitions, degrees of heterogeneity, and amounts of outliers generated by distributions with symmetry deviations and excess kurtosis. The simulation study showed that improving GFI and AGFI fit-validity indexes were promising since it showed robustness concerning outliers and diagnosed a model as good and bad fitted through degrees of heterogeneity imposed between the sample and hypothesized covariance matrices per model. Given the proposed model for the study of specialty coffees, the information represented in the latent variables production and environmental changes influence the sensory perception of coffees produced in the Serra da Mantiqueira region, and the improvement of indices enhanced model validation due to the inclusion of observed variable repetitions.

Keywords: Simulation. Outliers. GFI. AGFI. Flavor. Acidity.

LISTA DE FIGURAS

PRIMEIRA PARTE

Figura 1 - Diagrama de caminhos do modelo completo..... 12

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Figura 1 - Modelo de equações estruturais teórico utilizado na simulação Monte Carlo 31

Figura 2 - Desempenho dos índices modificados, em função do grau de heterogeneidade, considerando *outliers* gerados pela t-Student multivariada, com taxa de mistura α e r número de repetições das variáveis observadas: (A) $\alpha=5\%$ e $r=2$; (B) $\alpha=5\%$ e $r=4$; (C) $\alpha=5\%$ e $r=10$; (D) $\alpha=5\%$ e $r=20$; (E) $\alpha=15\%$ e $r=2$; (F) $\alpha=15\%$ e $r=4$; (G) $\alpha=15\%$ e $r=10$; (H) $\alpha=15\%$ e $r=20$ 40

Figura 3 - Desempenho dos índices modificados, em função do grau de heterogeneidade, considerando *outliers* gerados pela Lognormal multivariada, com taxa de mistura α e r número de repetições das variáveis observadas: (A) $\alpha=5\%$ e $r=2$; (B) $\alpha=5\%$ e $r=4$; (C) $\alpha=5\%$ e $r=10$; (D) $\alpha=5\%$ e $r=20$; (E) $\alpha=15\%$ e $r=2$; (F) $\alpha=15\%$ e $r=4$; (G) $\alpha=15\%$ e $r=10$; (H) $\alpha=15\%$ e $r=20$ 42

ARTIGO 2

Figura 1 - Diagrama de caminhos para a análise sensorial de cafés especiais 52

Figura 2 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação usual. 59

Figura 3 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação R_{θ} e com 2 repetições para os tratamentos. 59

Figura 4 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação R_{θ} e com 20 repetições para os tratamentos. 59

LISTA DE TABELAS

PRIMEIRA PARTE

Tabela 1 - Elementos básicos utilizados na construção de um diagrama de caminhos..... 13

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Tabela 1 - Índice de qualidade de ajuste (GFI) e índice de qualidade de ajuste corrigido (AGFI) pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição normal multivariada com grau de heterogeneidade $GH = 2$ 36

Tabela 2 - Índice de qualidade de ajuste (GFI) e índice de qualidade de ajuste corrigido (AGFI) pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição normal multivariada com grau de heterogeneidade $GH = 8$ 38

Tabela 3 - GFI e AGFI pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição t-Student e lognormal multivariadas..... 44

ARTIGO 2

Tabela 1 - Tratamentos utilizados no delineamento experimental com base na altitude. 54

Tabela 2 - Estimativas dos índices de qualidade de ajuste GFI AGFI para validar o modelo de equações estruturais com a inclusão de delineamento inteiramente casualizado com faixas de altitudes definidas como tratamento em estimativas de diferentes números de repetições. 57

SUMÁRIO

PRIMEIRA PARTE	11
1 INTRODUÇÃO	11
2 REFERENCIAL TEÓRICO.....	12
2.1 Modelagem de Equações Estruturais	12
2.2 Notação Matemática	14
2.2.1 Modelo de Mensuração.....	14
2.2.2 Modelo Estrutural.....	16
2.3 Estimação dos parâmetros do Modelo	18
2.3.1 Método dos Mínimos Quadrados Ordinários (MQO)	19
2.3.2 Método de Máxima Verossimilhança (MV).....	20
2.4 Qualidade de Ajuste.....	21
2.4.1 Índice de Qualidade do Ajuste (GFI) e Índice da Qualidade do Ajuste Corrigido (AGFI)	21
3 Principais aspectos de metodologias estatísticas aplicada a análise sensorial de café especiais e aplicabilidade em modelos de equações estruturais	23
REFERÊNCIAS	25
SEGUNDA PARTE - ARTIGOS.....	27
ARTIGO 1 - APRIMORAMENTO DOS ÍNDICES DE AJUSTES GFI E AGFI EM MODELOS DE EQUAÇÕES ESTRUTURAIS CORRIGIDOS PELA COVARIÂNCIA ENTRE AS REPETIÇÕES NAS VARIÁVEIS OBSERVADAS	27
ARTIGO 2 - UM ESTUDO DE CASO DA MODELAGEM DA COVARIÂNCIA ENTRE FATORES EXTERNOS RELACIONADO A PERCEPÇÃO SENSORIAL DE CAFÉS.....	47

PRIMEIRA PARTE

1 INTRODUÇÃO

A modelagem de equações estruturais (MEE) é uma técnica de análise multivariada, que permite formalizar e validar uma série de relações hipotéticas de causa-efeito entre grupos de variáveis sejam elas, observadas ou latentes, representadas pelos construtos (HAIR *et al.*, 2017). Ela é particularmente útil para testar teorias que contêm múltiplas equações envolvendo relações de dependência.

Uma variável latente não pode ser medida diretamente, mas pode ser representada por um conjunto de variáveis observadas (PILATIL; LAROS, 2007). Por exemplo, a preferência de um consumidor, em relação a um tipo de café, jamais pode ser medida precisamente, decorrente da existência de incertezas nas avaliações. Entretanto, relacionando com outros atributos sensoriais, torna-se possível inferir as tendências de preferências em relação a determinados tipos de cafés. A relação que explique essas tendências é dada por um conjunto de equações, que determinam a causalidade entre os atributos (variáveis observadas) e preferências e atitudes comportamentais (variáveis não observadas). Neste âmbito, surgem os modelos de equações estruturais, nos quais a forma de estabelecer relações entre as variáveis, em geral, é dada pela combinação de elementos teóricos e de evidências empíricas de pesquisas anteriores.

Uma maneira de verificar se um modelo de equações estruturais está bem ajustado é por avaliação de diversos indicadores ou índices de qualidade de ajuste. Todos esses índices apresentam um objetivo em comum de avaliar a correspondência entre a matriz de covariância amostral e a covariância imposta por um modelo estrutural que represente apropriadamente as inter-relações entre os indicadores envolvidos no estudo (CIRILLO; BARROSO, 2017). Porém nenhum desses índices leva em consideração a estrutura experimental com as repetições e os tratamentos entre as variáveis observadas.

Mediante o exposto, torna-se compreensível um estudo com abordagem de modelos de equações estruturais, considerando as repetições e tratamentos nas variáveis observadas, portanto o objetivo deste trabalho consiste em propor a modificação dos índices de qualidade de ajuste em modelos de equações estruturais corrigidos por uma estrutura de covariância considerando os tratamentos e diferentes repetições nas variáveis observadas. E utilizar os índices de qualidade de ajuste modificados na análise sensorial de cafés especiais no sentido de inferir com maior riqueza informações sobre a dispersão das respostas sensoriais.

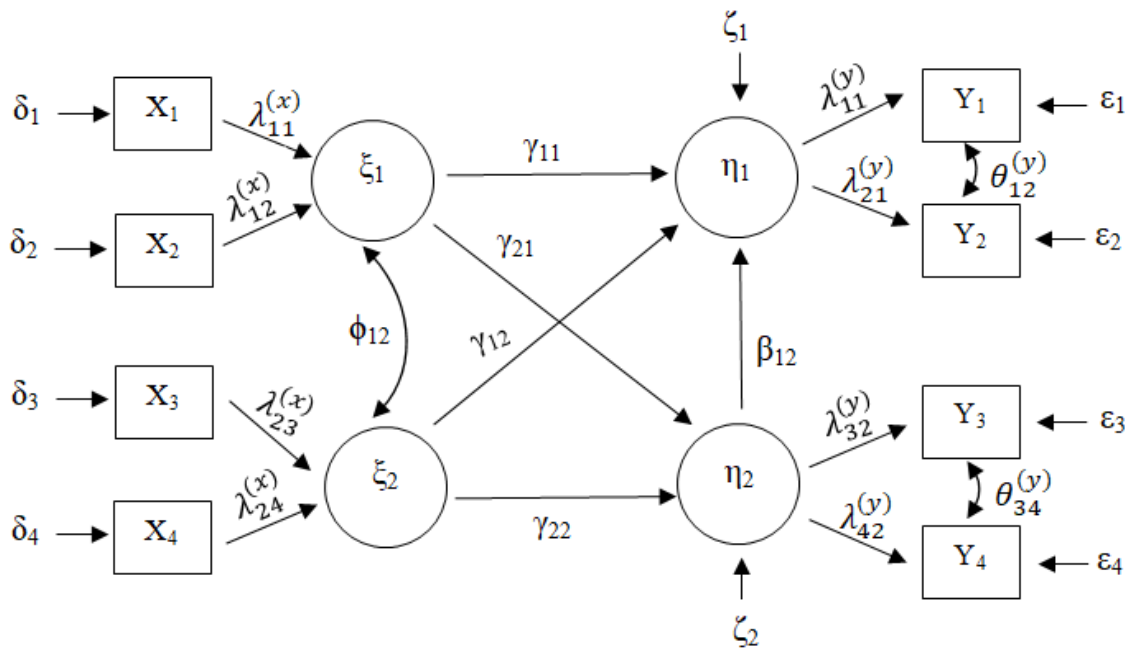
2 REFERENCIAL TEÓRICO

2.1 Modelagem de Equações Estruturais

A modelagem de equações estruturais (MEE) é um conjunto de modelos estatísticos que visa explicar as relações entre múltiplas variáveis. A MEE examina a estrutura de inter-relações expressas em uma série de equações. Tais equações descrevem todas as relações entre as variáveis dependentes e independentes envolvidas na análise (HAIR *et al.*, 2009).

A construção de um modelo de equações estruturais se inicia de um modelo teórico previamente definido o qual determinará as diversas relações existentes entre as variáveis do modelo, podendo ser relações causais ou de dependência. Um modelo teórico consiste de relações que fornecem explicações consistentes e abrangentes dos fenômenos, conforme é exemplificado na Figura 1.










Figura 1 - Diagrama de caminhos do modelo completo.



Fonte: Da autora (2022).

A Tabela 1 ilustra os elementos básicos utilizados na construção de um diagrama de caminhos.

Tabela 1 - Elementos básicos utilizados na construção de um diagrama de caminhos

Descrição	Elemento Básico
Variável latente ou construto	
Variável observada ou indicadora	
Relação causal direta ou direcional entre duas variáveis	
Relação não recursiva ou não direcional entre duas variáveis	
Correlação entre duas variáveis	
Relação entre duas variáveis latentes	
Relação entre uma variável observada e uma variável latente	
Erro de mensuração na variável observada	
Erro na predição da variável latente	

Fonte: Amorim *et al.* (2012).

Usualmente os efeitos são separados em conjuntos e são utilizadas letras distintas para representá-los, ou seja, efeitos de uma variável latente sobre variáveis observadas podem ser representados por uma letra adicionada de índices, enquanto, no mesmo diagrama, para representar efeitos de uma outra variável observada sobre determinado conjunto de variáveis latentes, pode ser usada outra letra e, para as correlações entre as variáveis observadas, uma terceira.

Todas as relações supostas pelo modelo devem ser representadas no diagrama, mesmo que a estimativa seja igual a zero no parâmetro correspondente. Relações não especificadas supõe-se que valham zero, como, por exemplo, duas variáveis não serem correlacionadas. Deste modo, o modelo de equações estruturais fica completamente representado pelo diagrama de caminhos, a partir do qual é possível explicitar matematicamente o modelo de forma completa (BRUNELLI, 2012).

Assim, na Figura 1, foi construído um diagrama de caminho, para exemplificar seus elementos básicos: ξ_1 , ξ_2 , η_1 e η_2 as variáveis latentes, $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3$ e Y_4 as variáveis observáveis, $\delta_1, \dots, \delta_4, \varepsilon_1, \dots, \varepsilon_4, \zeta_1$ e ζ_2 as variáveis de erro e flechas representando as relações entre essas variáveis.

Pela Figura 1 observa-se que as variáveis ξ_1 e ξ_2 são causadas por X_1, X_2 e X_3, X_4 , respectivamente. As variáveis ξ_1 e ξ_2 causam η_1 e η_2 ; as variáveis η_1 e η_2 causam Y_1, Y_2 e Y_3, Y_4 respectivamente. E as variáveis Y_1 e Y_2 e Y_3 e Y_4 são correlacionadas.

2.2 Notação Matemática

Um modelo de equações estruturais completo é composto por dois submodelos que podem ser observados no diagrama de caminhos: o modelo de mensuração e o modelo estrutural.

O modelo de mensuração especifica como as variáveis latentes afetam ou são afetadas pelas variáveis observáveis. Essa parte do modelo está relacionada ao uso de análise fatorial confirmatória. As equações de mensuração oferecem ainda uma descrição das propriedades de mensuração como validade e confiabilidade dessas variáveis (AMORIM *et al.*, 2012; CODES, 2005). O modelo de mensuração permite avaliar se cada construto tem confiabilidade satisfatória para se estimarem as relações causais em etapas posteriores de teste do modelo (BREI; LIBERALI NETO, 2006).

A parte estrutural mostra como os construtos estão associados uns com os outros. Seu desenvolvimento fundamenta-se no cálculo de sistemas de equações simultâneas. Nessa etapa da MEE, estão os procedimentos de especificação e estimação das associações das variáveis latentes entre si ou com outras variáveis observáveis, descrevendo seus efeitos e suas magnitudes. Incluem também as informações sobre a variância explicada e a não explicada de cada termo endógeno presente no modelo (AMORIM, 2012; KAPLAN, 2000).

Assim, o sistema de relações de causa-efeito do modelo de equações estruturais é decomposto nessas duas partes, e cada uma possui uma notação matemática específica, que é descrita a seguir.

2.2.1 Modelo de Mensuração

Considere $\mathbf{x}_{(q \times 1)}$ o vetor de q variáveis observadas que são indicadores das s variáveis latentes exógenas e $\mathbf{y}_{(p \times 1)}$ o vetor de p variáveis observadas que são indicadores das r variáveis latentes endógenas; além disso, sejam os vetores $\boldsymbol{\eta}_{(r \times 1)}$ e $\boldsymbol{\xi}_{(s \times 1)}$ compostos pelas variáveis latentes endógenas e exógenas, respectivamente.

Segundo Bollen (1989), as equações (2.1) e (2.2) compõem o que se conhece por modelo de mensuração e especificam as relações entre as variáveis observadas e as variáveis latentes.

$$\mathbf{y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (2.1)$$

$$\mathbf{x} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \quad (2.2)$$

em que:

- (2.1) é a equação de mensuração das variáveis latentes endógenas e (2.2), a das variáveis latentes exógenas;
- $\mathbf{\Lambda}_y(p \times r)$ e $\mathbf{\Lambda}_x(q \times s)$ são matrizes de parâmetros, que representam os efeitos das variáveis latentes (respectivamente, endógenas e exógenas) sobre seus respectivos indicadores;
- $\boldsymbol{\varepsilon}_{(p \times 1)}$ e $\boldsymbol{\delta}_{(q \times 1)}$ são os vetores de erros de mensuração para \mathbf{y} e \mathbf{x} , respectivamente.

Com as seguintes pressuposições:

- $\boldsymbol{\varepsilon}$ é não correlacionado com $\boldsymbol{\delta}$, $\boldsymbol{\xi}$ e $\boldsymbol{\eta}$;
- $\boldsymbol{\delta}$ é não correlacionado com $\boldsymbol{\varepsilon}$, $\boldsymbol{\xi}$ e $\boldsymbol{\eta}$;
- $E(\boldsymbol{\eta}) = 0$;
- $E(\boldsymbol{\varepsilon}) = 0$;
- $E(\boldsymbol{\xi}) = 0$;
- $E(\boldsymbol{\delta}) = 0$

A equação do modelo de mensuração especificado, em termos das variáveis exógenas, apresenta a seguinte forma matricial:

$$\begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_q \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} & \dots & \lambda_{1s}^{(x)} \\ \vdots & \vdots & \vdots \\ \lambda_{q1}^{(x)} & \dots & \lambda_{qs}^{(x)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_s \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_q \end{pmatrix}$$

A equação do modelo de mensuração especificado, em termos de variáveis endógenas, apresenta a seguinte forma matricial:

$$\begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_p \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} & \dots & \lambda_{1r}^{(y)} \\ \vdots & \vdots & \vdots \\ \lambda_{p1}^{(y)} & \dots & \lambda_{pr}^{(y)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_r \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_p \end{pmatrix}$$

2.2.2 Modelo Estrutural

As equações estruturais especificam as relações entre os vetores de variáveis latentes $\boldsymbol{\eta}_{(r \times 1)}$ e $\boldsymbol{\xi}_{(s \times 1)}$, os mesmos definidos no Item 2.5.1, então, a equação estrutural do modelo de equações estruturais pode ser definido como (BOLLEN, 1989):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (2.3)$$

A equação do modelo estrutural dada em (2.3) pode ser simplificada:

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \Rightarrow (\mathbf{I} - \mathbf{B})\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \Rightarrow \boldsymbol{\eta} &= (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}) \end{aligned} \quad (2.4)$$

em que:

- a) $\mathbf{B}_{(r \times r)}$ é uma matriz de parâmetros que representa os efeitos das variáveis latentes endógenas sobre os outros construtos. Verifica-se que a diagonal principal dessa matriz será sempre identicamente nula, pois um fator nunca será escrito como função dele mesmo.
- b) $\boldsymbol{\Gamma}_{(r \times s)}$ é uma matriz de parâmetros que representa o efeito das variáveis latentes exógenas sobre os outros construtos;
- c) $\boldsymbol{\zeta}_{(r \times 1)}$ é um vetor aleatório com erros das equações estruturais.

Com as seguintes pressuposições:

- a) $\boldsymbol{\zeta}$ é não correlacionado com $\boldsymbol{\xi}$;
- b) $E(\boldsymbol{\eta}) = 0$;
- c) $E(\boldsymbol{\zeta}) = 0$;
- d) $E(\boldsymbol{\xi}) = 0$;
- e) $(\mathbf{I} - \mathbf{B})$ é uma matriz não singular, de forma que $(\mathbf{I} - \mathbf{B})^{-1}$ existe.

As matrizes de covariância dos parâmetros do modelo são as seguintes:

- a) $\boldsymbol{\Phi}_{(s \times s)} = E(\boldsymbol{\xi}\boldsymbol{\xi}')$;
- b) $\boldsymbol{\Psi}_{(s \times r)} = E(\boldsymbol{\zeta}\boldsymbol{\zeta}')$;
- c) $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}(p \times p)} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$;
- d) $\boldsymbol{\Theta}_{\boldsymbol{\delta}(q \times q)} = E(\boldsymbol{\delta}\boldsymbol{\delta}')$;

A equação do modelo estrutural (2.3) apresenta a seguinte forma matricial:

$$\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \end{pmatrix} = \begin{pmatrix} 0 & \cdots & \beta_{1r}^{(y)} \\ \vdots & \vdots & \vdots \\ \beta_{r1}^{(y)} & \cdots & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \end{pmatrix} + \begin{pmatrix} \gamma_{11}^{(y)} & \cdots & \gamma_{1s}^{(y)} \\ \vdots & \vdots & \vdots \\ \gamma_{r1}^{(y)} & \cdots & \gamma_{rs}^{(y)} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_s \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_r \end{pmatrix}$$

Da Figura 1 podem-se escrever as equações (2.1), (2.2) e (2.3).

O modelo estrutural, que representa as relações entre as variáveis latentes, possui as seguintes equações:

$$\begin{aligned} \eta_1 &= \beta_{12}\eta_2 + \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \\ \eta_2 &= \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2 \end{aligned} \quad (2.5)$$

ou em forma matricial:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \quad (2.6)$$

Com as seguintes matrizes de covariância:

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \quad (2.7)$$

$$\Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} \quad (2.8)$$

As equações do modelo de mensuração em y são dadas por:

$$\begin{aligned} y_1 &= \lambda_{11}^{(y)}\eta_1 + \varepsilon_1 \\ y_2 &= \lambda_{21}^{(y)}\eta_1 + \varepsilon_2 \\ y_3 &= \lambda_{32}^{(y)}\eta_2 + \varepsilon_3 \\ y_4 &= \lambda_{42}^{(y)}\eta_2 + \varepsilon_4 \end{aligned} \quad (2.9)$$

ou em forma matricial:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} & 0 \\ \lambda_{21}^{(y)} & 0 \\ 0 & \lambda_{32}^{(y)} \\ 0 & \lambda_{42}^{(y)} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \quad (2.10)$$

As equações do modelo de mensuração em x são dadas por:

$$\begin{aligned} x_1 &= \lambda_{11}^{(x)} \xi_1 + \delta_1 \\ x_2 &= \lambda_{12}^{(x)} \xi_1 + \delta_2 \\ x_3 &= \lambda_{23}^{(x)} \xi_2 + \delta_3 \\ x_4 &= \lambda_{24}^{(x)} \xi_2 + \delta_4 \end{aligned} \quad (2.11)$$

ou em forma matricial:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} & 0 \\ \lambda_{12}^{(x)} & 0 \\ 0 & \lambda_{23}^{(x)} \\ 0 & \lambda_{24}^{(x)} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix} \quad (2.12)$$

2.3 Estimação dos parâmetros do Modelo

Na estimação dos parâmetros, assume-se que a distribuição das variáveis observadas pode ser descrita por meio do seu vetor de médias e da sua matriz de covariância, ignorando-se os momentos de ordem superior. Uma vez que não existem restrições para o vetor de médias a estimação dos parâmetros pode ser feita utilizando a matriz de covariância amostral (\mathbf{S}) ao invés da matriz de covariância imposta pelo modelo ($\mathbf{\Sigma}$).

Para estimar os parâmetros de um modelo de equações estruturais, parte-se da igualdade $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$, em que $\mathbf{\Sigma}$ é a matriz de covariância populacional das variáveis observadas e $\mathbf{\Sigma}(\boldsymbol{\theta})$ é uma função do vetor $\boldsymbol{\theta}$ de parâmetros livres do modelo. Como $\mathbf{\Sigma}$ não é conhecida, utiliza-se, neste caso, a matriz \mathbf{S} de covariâncias amostrais, para estimar o modelo, chegando ao sistema de equações determinado pela igualdade $\mathbf{S} = \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$.

A matriz $\mathbf{\Sigma}$ pode ser particionada em blocos que separam a variância de \mathbf{x} , a variância de \mathbf{y} e a covariância entre \mathbf{x} e \mathbf{y} , reescrevendo-as como transformações não lineares dos parâmetros do modelo (BOLLEN, 1989):

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{\Sigma}_{yy}(\boldsymbol{\theta}) & \mathbf{\Sigma}_{yx}(\boldsymbol{\theta}) \\ \mathbf{\Sigma}_{xy}(\boldsymbol{\theta}) & \mathbf{\Sigma}_{xx}(\boldsymbol{\theta}) \end{pmatrix} \quad (2.13)$$

A matriz covariância implícita de \mathbf{x} e \mathbf{y} pode ser escrita como:

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{\Lambda}_y (\mathbf{I} - \mathbf{B})^{-1} (\mathbf{\Gamma} \Phi \mathbf{\Gamma}' + \mathbf{\Psi}) (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda}_y' + \mathbf{\Theta}_\varepsilon & \mathbf{\Lambda}_y (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Gamma} \mathbf{\Psi} \mathbf{\Lambda}_x' \\ \mathbf{\Lambda}_x \mathbf{\Psi} \mathbf{\Gamma}' (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda}_y' & \mathbf{\Lambda}_x \mathbf{\Psi} \mathbf{\Lambda}_x' + \mathbf{\Theta}_\delta \end{pmatrix} \quad (2.14)$$

Para estimar os parâmetros do modelo de equações estruturais, são utilizadas as funções de discrepância. Funções de discrepância F são relações entre as matrizes \mathbf{S} e $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$, é uma função da diferença da matriz de covariâncias amostrais \mathbf{S} e da matriz de covariâncias gerada pelo modelo proposto com parâmetros $\boldsymbol{\theta}(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}))$.

As funções de discrepância são minimizadas com o objetivo de estimar o vetor de parâmetros $\boldsymbol{\theta}$ e denotadas por $F(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta}))$. Conforme Mulaik (2009), elas são duas vezes diferenciáveis, contínuas em \mathbf{S} e $\mathbf{\Sigma}(\boldsymbol{\theta})$, não negativas e satisfazem a condição $F(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta})) = 0$ se e somente se, $\mathbf{S} = \mathbf{\Sigma}(\boldsymbol{\theta})$. Fixando-se $\boldsymbol{\theta}$, $F(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta}))$ é um escalar. De acordo com Timm (2002), a minimização da função de discrepância é complexa, uma vez que envolve um sistema de equações não lineares restritas. Assim, são utilizados algoritmos iterativos, tais como Newton-Raphson e Gauss-Newton (JÖRESKOG, 1969).

Segundo Browne (1984), a função de discrepância definida, conforme os critérios acima, proporcionarão estimadores consistentes para os parâmetros do modelo. Logo, são essas as funções utilizadas pelos principais métodos de estimação, conforme será comentado adiante.

Existem várias funções F que podem ser utilizadas para indicar a magnitude das diferenças entre os elementos de \mathbf{S} e $\mathbf{\Sigma}(\boldsymbol{\theta})$. Nesta seção, serão contempladas as funções de discrepâncias construídas pelos métodos de estimação de Máxima Verossimilhança (MV) e Mínimos Quadrados Ordinários (MQO), por serem considerados os mais utilizados em modelos equações estruturais (BOLLEN, 1996).

2.3.1 Método dos Mínimos Quadrados Ordinários (MQO)

O método dos mínimos quadrados ordinários tem por princípio minimizar a soma de quadrados de cada elemento da matriz residual $(\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta}))$ (BOLLEN, 1989).

A função de discrepância a ser minimizada pelo algoritmo iterativo é:

$$F_{MQO}(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta})) = \frac{1}{2} tr \{ [\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})]' [\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})] \} = \frac{1}{2} tr \{ [\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})]^2 \} \quad (2.15)$$

em que $tr(\cdot)$ designa o operador traço, ou seja, é a soma dos elementos da diagonal principal da matriz e $[\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})]' = \mathbf{S}' - \mathbf{\Sigma}'(\boldsymbol{\theta}) = \mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})$.

Apesar da função (2.15), entre todas as funções de discrepância, ser considerada a mais simples delas e propiciar a obtenção de estimadores consistentes e não exigir a suposição de

que as variáveis observadas seguem alguma distribuição, ela apresenta, como desvantagem, a não determinação de estimadores assintoticamente mais eficientes (BOLLEN, 1989).

2.3.2 Método de Máxima Verossimilhança (MV)

O método de máxima verossimilhança (MV) baseia-se no princípio de máxima verossimilhança, isto é, de encontrar as estimativas dos parâmetros que maximizariam a probabilidade de observar \mathbf{S} . Esse método assume a suposição de que os vetores x e y seguem distribuição normal multivariada e, portanto parte da verossimilhança dessa distribuição (BOLLEN, 1989).

O estimador de máxima verossimilhança é obtido a partir da seguinte função de discrepância:

$$F_{MV}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \text{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] - \log[\mathbf{S}] - (p+q) \quad (2.16)$$

em que $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ é a matriz de covariância populacional do modelo que está sendo proposto, $|\mathbf{X}|$ é o operador determinante de uma matriz \mathbf{X} ; e $\text{tr}(\cdot)$ é o operador traço; p a quantidade de variáveis observadas endógenas (Y) e q a quantidade de variáveis observadas exógenas (X).

A função (2.16) deve ser minimizada com relação a \mathbf{B} , $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ e $\boldsymbol{\Psi}$, encontrando-se então as estimativas de máxima verossimilhança desses parâmetros. Segundo Bollen (1989), esses estimadores são consistentes, invariantes, assintoticamente não viciados, assintoticamente eficientes e com distribuição assintoticamente normal. Além disso, sob a hipótese de que $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, ou seja, que as estimativas obtidas realmente representam os verdadeiros valores populacionais das correlações e que, portanto o modelo está correto, vale que a função F_{MV} descrita acima, multiplicada por $N - 1$, converge para uma distribuição qui-quadrado:

$$(N-1)F_{MV}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \sim \chi^2\left(\frac{1}{2}(p+q)(p+q+1) - w\right), \quad (2.17)$$

em que w é o número de parâmetros livres no vetor $\boldsymbol{\theta}$.

2.4 Qualidade de Ajuste

Uma maneira de verificar se um modelo de equações estruturais está bem ajustado é por meio da análise do desempenho de alguns indicadores ou índices. Embora eles não possam ser analisados individualmente, para concluir sobre a qualidade do ajuste, uma vez que cada um possui suas vantagens e desvantagens, o estudo de várias dessas medidas em conjunto pode ser informativo (BOLLEN, 1989).

Para verificar a qualidade do ajuste do modelo, é necessário realizar duas verificações. Primeiro, se os valores estimados para os parâmetros fazem sentido (com relação ao sinal, se é necessário fixar limites para o valor). E, posteriormente, é necessário avaliar o ajuste geral do modelo: o quão boa está a aderência dos valores estimados aos dados realizados. Para isso, é avaliada a seguinte hipótese de interesse: $H_0 : \Sigma - \Sigma(\theta) = \mathbf{0}$, observando-se \mathbf{S} (matriz de covariâncias amostrais, que é conhecida) em vez de Σ e $\Sigma(\hat{\theta})$, ou seja, a matriz de covariância $\Sigma(\theta)$ estimada no ponto que minimiza a função de discrepância utilizada.

Sintetizando, os índices têm por objetivo comparar dois modelos distintos: o modelo definido pelo pesquisador e o modelo que não possui restrições sobre os momentos populacionais, dito saturado. Em geral, eles são construídos com base na função de discrepância ou na matriz de resíduos (BULHÕES, 2013).

2.4.1 Índice de Qualidade do Ajuste (GFI) e Índice da Qualidade do Ajuste Corrigido (AGFI)

O índice da qualidade do ajuste (GFI, *Goodness of Fit Index*) mede quanto a quantidade relativa de covariância em \mathbf{S} é predita por $\hat{\Sigma}$; e o índice de qualidade de ajuste corrigido (AGFI, *Adjusted Goodness of Fit Index*) foi proposto com a finalidade de melhor adequar o GFI aos graus de liberdade e à quantidade de variáveis observadas. Geralmente os dois são maiores que 0 e valem no máximo 1. Ambos apresentam distintas formas, variando conforme o método de estimação.

Jöreskog e Sorbom (1986) propuseram medidas GFI e AGFI, para modelos ajustados por MV, que são calculadas por:

$$GFI_{MV} = 1 - \frac{\text{tr} \left[\left(\Sigma^{-1}(\hat{\theta}) \mathbf{S} - \mathbf{I} \right)^2 \right]}{\text{tr} \left[\left(\Sigma^{-1}(\hat{\theta}) \mathbf{S} \right)^2 \right]} \quad (2.18)$$

$$AGFI_{MV} = 1 - \left[\frac{k(k+1)}{2gl_p} \right] [1 - GFI_{MV}] \quad (2.19)$$

Tanaka e Huba (1985) propuseram essas medidas, para modelos ajustados por MQO, que são calculadas por:

$$GFI_{MQO} = 1 - \frac{\text{tr} \left[\left(\mathbf{S} - \Sigma(\hat{\theta}) \right)^2 \right]}{\text{tr} \left[\mathbf{S}^2 \right]} \quad (2.20)$$

$$AGFI_{MQO} = 1 - \left[\frac{k(k+1)}{2gl_p} \right] [1 - GFI_{MQO}] \quad (2.21)$$

em que $k = p+q$, $\hat{\theta}$ é um vetor de estimativas do vetor de parâmetros θ , $\Sigma(\hat{\theta})$ é a matriz de covariâncias $\Sigma(\theta)$, avaliada no vetor $\hat{\theta}$, que minimiza a função de discrepância F considerada, \mathbf{S} é a matriz de covariâncias de $\mathbf{z} = [\mathbf{y}' \quad \mathbf{x}']$; $\text{tr}(\cdot)$ designa o operador traço; \mathbf{I} é a matriz identidade; p é o número de componentes de \mathbf{y} e q é o número de componentes de \mathbf{x} e, por fim, gl_p é o número de graus de liberdade do modelo do pesquisador.

Independentemente da função de discrepância, os indicadores GFI buscam medir a quantidade relativa de covariâncias \mathbf{S} que são previstas por $\Sigma(\hat{\theta})$. Eles podem, eventualmente, assumir valores negativos e possuem a seguinte característica: $GFI = 1$ se e só se, $\mathbf{S} = \Sigma(\hat{\theta})$. A partir de revisões da literatura e experiências práticas, Marôco (2010) sugere aos valores de GFI a seguinte interpretação: se $GFI < 0,90$, modelos com ajustamento ruim aos dados; se $0,90 \leq GFI < 0,95$, indicam um bom ajustamento; e se $GFI \geq 0,95$, indicadores de um ajustamento muito bom. $GFI = 1$ indica um ajustamento perfeito. Tanaka e Huba (1985) citam que o GFI tem tendência a aumentar à medida que o número de unidades amostrais aumenta.

O indicador AGFI também pode assumir valores negativos, assim como o GFI. Além disso, verifica-se facilmente que $GFI = 1$ se e somente se, $AGFI = 1$. Ambos possuem a mesma interpretação e compartilham o defeito de serem inflacionados à medida que o número de unidades amostrais aumenta.

3 PRINCIPAIS ASPECTOS DE METODOLOGIAS ESTATÍSTICAS APLICADA A ANÁLISE SENSORIAL DE CAFÉS ESPECIAIS E APLICABILIDADE EM MODELOS DE EQUAÇÕES ESTRUTURAIS

O café é uma bebida muito apreciada em diversos países no mundo, por ser um produto natural, com diversos aromas e sabores agradáveis, que o fazem distinto. Entre as espécies mais cultivadas, destaca-se o café arábica (*Coffea arabica L.*), essas cultivares são potenciais cafés dignos de serem classificados como especiais (RAMOS *et al.*, 2016).

Os cafés especiais têm grande importância no cenário econômico e sua qualidade sensorial é apreciada pelo setor produtivo e pelo mercado. Pesquisas têm sido constantemente realizadas na busca por melhores misturas, a fim de agregar valor e diferenciar preços, de acordo com a qualidade do produto (OSSANI *et al.*, 2021). Os três pilares no que se refere à qualidade do café são o manejo/processamento, a microbiologia e a físico-química, refletindo diretamente sobre a qualidade final do café, que é validada e certificada, seja no âmbito industrial ou acadêmico, necessariamente pela análise de um painel sensorial.

No tocante às projeções de consumo, conforme relatório da pesquisa, de tendências do mercado de cafés, em 2017, apresentada pela Associação Brasileira da Indústria de Café (ABIC, 2017), demonstra que o consumo de café aumentou 3,5% em 2017, comparado ao ano anterior (SOCIEDADE NACIONAL DE AGRICULTURA - SNA, 2018). Tal fato reflete que os apreciadores de café continuam cada vez mais consumindo o produto sob as mais diferentes formas, em função da melhoria da qualidade dos cafés do Brasil. Em síntese, segundo projeção do relatório Tendências do Mercado de Cafés, de 2017 a 2021, as taxas de crescimento do consumo no Brasil serão de 3,3% para o café em pó, de 4,3% para o café torrado e 9% de café em cápsulas.

Os resultados obtidos, por meio do conhecimento de diferentes perfis e apreciação por cafés de diferentes qualidades, agregaram informações que podem nortear uma expansão do consumo de cafés especiais, no crescimento da competitividade inter-regional da atividade cafeeira e na distribuição de rendas, com a geração de novas frentes de trabalhos em diversos setores da economia. Como exemplo, citamos marketing direcionado à segmentação de mercado, com buscas ao mercado internacional, no tocante à exportação de cafés e seus derivados com elevada qualidade.

Na perspectiva quimiométrica, a análise de atributos sensoriais, como, por exemplo, o sabor, é pautada na olfatométrica, centrada no indivíduo e preferências do consumidor

(YERETZIAN *et al.*, 2004, 2010), haja vista a influência de reflexos e percepções psicológicas que atuam no discernimento da qualidade de um produto.

Em se tratando da modelagem sobre essa perspectiva, torna-se complexa, uma vez que essas percepções não são variáveis observadas, além de serem intrínsecas a cada indivíduo. Diante disso, a modelagem de equações estruturais vem sendo utilizada na área cafeeira no qual pode-se mencionar os trabalhos de Pereira *et al.* (2021) e Santos *et al.* (2021), no primeiro os autores utilizam estimadores ridge generalizados adaptados à modelagem de equações estruturais na descrição do perfil de um grupo de consumidores de café. Através desses estimadores foi possível observar que na presença de variáveis observadas, independente do tamanho amostral ou do grau de multicolinearidade o mesmo desempenho em termos de acurácia, precisão ao considerar os cenários representados por modelo sem erro de especificação e modelo com erro de especificação. Mesmo apresentando a vantagem de resultarem EQM relativamente menores que o estimador de mínimos quadrados ordinários, os estimadores ridge generalizado foram na maioria dos casos analisados, responsáveis pelos maiores vieses. Entretanto, observou-se que quanto maior o tamanho amostral e o grau de multicolinearidade, mais os vieses dos estimadores ridge generalizados se aproximavam dos vieses empíricos do estimados de mínimos quadrados ordinários.

Santos *et al.* (2021) elaboraram construtos cujas variáveis possibilitassem a caracterização e distinção de indivíduos entre esses diferentes grupos de consumidores de cafés. Os construtos foram validados pelo índice adaptativo de variância média extraída (AVEADP). A transição entre grupos de consumidores é explicada e incentivada pelos avanços em seus conhecimentos conceituais e perceptivos.

Assim, os consumidores regulares devem ser endereçados com mensagens voltadas principalmente para o aspecto social do consumo; os entusiastas, dando ênfase as aspectos simples a moderados comumente utilizados como critérios de compra de produtos e especialistas, atraídos pela ênfase em critérios complexos relacionados ao conhecimento conceitual e perceptual dos cafés especiais, destacando sua influência no perfil sensorial da bebida.

Esses resultados possibilitaram uma melhor compreensão desses consumidores e podem orientar as estratégias de marketing de diferentes atores desse mercado

REFERÊNCIAS

- AMORIM, L. D. A. F. *et al.* **Modelagem com equações estruturais: princípios básicos e aplicações.** Salvador: [Ed. UFBA], 2012.
- ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **Demanda interna por cafés especiais deve ter expansão de até 7% em 2017.** 2017. Disponível em: http://www.redepeabirus.com.br/redes/form/post?topico_id=68170. Acesso em: 3 abr. 2022.
- BOLLEN, K. A. **Structural equations with latent variables.** New York: J. Wiley, 1989. 528 p.
- BREI, V. A.; LIBERALI NETO, G. O uso da técnica de modelagem em equações estruturais na área de marketing: um estudo comparativo entre publicações no Brasil e no exterior. **Revista de Administração Contemporânea**, Curitiba, v. 10, n. 4, p. 131-151, 2006.
- BROWNE, M. W. Asymptotic distribution free methods in analysis of covariance structures. **British Journal of Mathematical and Statistical Psychology**, London, v. 37, p. 62-83, 1984.
- BRUNELLI, R. T. **Análise do impacto de perturbações sobre medidas de qualidade de ajuste para modelos de equações estruturais.** 2012. 140 p. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2012.
- BULHÕES, R. D. S. **Contribuições à análise de outliers em modelos de equações estruturais.** 2013. 97 p. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.
- CIRILLO, M. A.; BARROSO, L. P. Effect of *outliers* on the GFI quality adjustment index in structural equation model and proposal of alternative indices. **Communications in Statistics Simulation and Computation**, Ontario, v. 46, n. 3, p. 1895-1905, 2017.
- CODES, A. L. M. Modelagem de equações estruturais: um método para a análise de fenômenos complexos. **Caderno CRH**, Salvador, v. 18, n. 45, p. 471-484, 2005.
- HAIR, J. F. *et al.* **Análise multivariada de dados.** 6. ed. Porto Alegre: Bookman, 2009. 688 p.
- HAIR, J. F. *et al.* **A Primer on partial least squares structural equation modeling (PLS-SEM).** 2nd ed. Thousand Oaks: Sage, 2017.
- JÖRESKOG, K. G. A general approach to confirmatory maximum like hood factor analysis. **Psychometrika**, Williamsburg, v. 34, n. 2, p. 183-202, 1969.
- JÖRESKOG, K. G.; SORBOM, D. **Lisrel VI: analysis of linear structural relationships by maximum likelihood and least square methods.** Michigan: Scientific Software, 1986.
- KAPLAN, D. **Structural equation modeling: foundations and extensions.** Thousand Oaks: Sage, 2000. 240 p.

MARÔCO, J. **Análise de equações estruturais: fundamentos teóricos, software & aplicações**. Pêro Pinheiro: Report Number, 2010. 384 p.

MULAIK, S. A. **Linear casual modeling with structural equations**. Boca Raton: Chapman & Hall; CRC, 2009. 468 p.

OSSANI, P. C. *et al.* Classification of specialty coffees using machine learning techniques. **Research, Society and Development**, Vargem Grande Paulista, v. 10, p. 1-15, 2021.

PEREIRA, G. A.; RESENDE, M.; CIRILLO, M. A. Generalized Ridge Estimators Adapted in Structural Equations Models. **Acta Scientiarum-Technology**, Maringá, v. 43, p. 1-10, 2021.

PILATIL, R.; LARO, J. A. Modelos de equações estruturais em psicologia: conceitos e aplicações. **Psicologia: teoria e pesquisa**, Brasília, DF, v. 23, n. 2, p. 205-216, 2007.

RAMOS, M. F. *et al.* Discrimination of the sensory quality of the *Coffea arabica* L. (cv. *Yellow Bourbon*) produced in different altitudes using decision trees obtained by the CHAID method. **Journal of the Science of Food and Agriculture**, London, v. 96, p. 3543-3551, 2016.

SANTOS, P. M. DOS; CIRILLO, M. Â.; GUIMARÃES, E. R. Specialty coffee in Brazil: transition among consumers' constructs using structural equation modeling. **British Food Journal**, Bradford, v. ahead-of-print, p. 1-15, 2021.

SOCIEDADE NACIONAL DE AGRICULTURA. **Consumo brasileiro de café cresce 3,5%, revela pesquisa da ABIC**. 2018. Disponível em: <http://www.sna.agr.br/consumo-brasileiro-de-cafe-cresce-35-revela-pesquisa-da-abic/>. Acesso em: 5 abr. 2022.

TANAKA, J. S.; HUBA, G. J. A Fit index for covariance structure models under arbitrary GLS estimation. **British Journal of Mathematical and Statistical Psychology**, London, v. 38, n. 2, p. 197-201, 1985.

TIMM, N. **Applied multivariate analysis**. New York: Springer, 2002. 675 p.

YERETZIAN, C. *et al.* Individualization of flavor preferences: toward a consumer-centric and individualized aroma science. **Comprehensive Reviews in Food Science and Food Safety**, Chicago, v. 3, n. 4, p. 152-159, 2004.

YERETZIAN, C. *et al.* The smell of coffee: an analytical perspective. **New Food**, Brasted, v. 13, n. 14, p. 1-17, 2010.

SEGUNDA PARTE - ARTIGOS**ARTIGO 1 - APRIMORAMENTO DOS ÍNDICES DE AJUSTES GFI E AGFI EM
MODELOS DE EQUAÇÕES ESTRUTURAIS CORRIGIDOS PELA COVARIÂNCIA
ENTRE AS REPETIÇÕES NAS VARIÁVEIS OBSERVADAS****(VERSÃO PRELIMINAR)****ARTIGO FORMATADO DE ACORDO COM AS NORMAS DO PERIÓDICO***Communications in Statistics - Simulation and Computation.*

Resumo

Embora existam metodologias que relacionem uma estrutura experimental ao ajuste de modelos de equações estruturais, notam-se algumas deficiências, nos índices de qualidade de ajuste, propostos na literatura, por não contemplarem características de natureza experimental. Com essa argumentação, este trabalho tem por objetivo propor um aprimoramento nos índices GFI e AGFI, incorporando uma correção que considere repetições nas variáveis observadas. Para validação desta proposta, diversos delineamentos são considerados, em função de critérios que determinam a qualidade de ajuste, bem como presença de *outliers*. Neste contexto, o aprimoramento é dado aos índices construídos nas abordagens de mínimos quadrados e máxima verossimilhança, sendo comparadas suas vantagens e desvantagens, em relação aos índices convencionais, utilizando simulação Monte Carlo. Concluiu-se que o aprimoramento dos índices GFI e AGFI, para incorporar o efeito das repetições das variáveis observadas, foi promissor, uma vez que apresentou propriedades de robustez quanto à presença de *outliers* e mostrou evidências de diagnóstico de um modelo com bom e ruim ajuste. Recomenda-se o uso de até quatro repetições para que não ocorram problemas de natureza numéricas, ao computar as estimativas dos índices.

Palavras-chave: Delineamentos, grau de heterogeneidade, *outliers*, robustez, falta de ajuste.

1 INTRODUÇÃO

Em geral, a qualidade de ajuste de um modelo de equações estruturais é verificada por diversos índices, nos quais, em síntese, avaliam a correspondência entre a matriz de covariância amostral e a covariância implícita pelo modelo que represente apropriadamente as inter-relações entre os indicadores envolvidos no estudo (CIRILLO e BARROSO, 2017). Entretanto os valores de referência, que justificam um bom ajuste são diferentes, uma vez que diferentes métodos de ajustes e metodologias são aplicados em suas formulações (CHEUNG e RENSVOLD, 2002).

Um fator motivador à realização de um aprimoramento aos índices existentes, bem como à pesquisa de novas metodologias se depara com um problema notório a índices resultantes da distribuição Qui-Quadrado, a exemplificar, Normed Fit Index (NFI); Non-Normed Fit Index (NNFI) e Comparative Fit Index (CFI). Sun (2005) mencionou que, na maioria dos modelos, os resultados são significativos, portanto ocorre uma tendência a rejeitar seus modelos, além de depender explicitamente do tamanho amostral.

Outra categoria de índices utilizados na validação dos modelos mencionada por Ullman (2007) refere-se aos índices fundamentados, na proporção da variância explicada, por meio da estimação do modelo, sendo os mais usuais o índice da qualidade do ajuste (GFI, *Goodness of Fit Index*) e o índice de qualidade de ajuste corrigido (AGFI, *Adjusted Goodness of Fit Index*),

o qual foi projetado para compensar o aumento na qualidade de ajuste de um modelo menos restrito obtido pela estimativa de mais parâmetros livres (MULAİK et al, 1989).

Convém ressaltar que, em termos práticos, a estimação desses índices requer apenas o conhecimento da matriz de covariância amostral e a matriz hipotetizada pelo modelo, além do conhecimento do número de parâmetros a ser adicionado no índice AGFI. Com essa perspectiva, torna-se fácil o aprimoramento desses índices, utilizando estimadores plug-in, submetidos a um procedimento de simulação computacional para generalizar sua eficiência em um conjunto de cenários a serem simulados.

Neste sentido, Cirillo e Barroso (2017) propuseram uma modificação, no índice GFI, substituindo a matriz de covariância amostral por matrizes de covariância robustas, descritas por Jackson e Chen (2004), sendo estimadas pelos procedimentos Minimum Covariance Determinant Estimator (MCD) e Minimum Volume Ellipsoid (MVE). Os autores observaram que com essa modificação, para os tamanhos amostrais $n \geq 100$, com *outliers* gerados por distribuições simétricas, o índice GFI computado pelas matrizes de covariâncias robustas apresentaram resultados similares.

Em se tratando de um modelo estrutural que contemple uma estrutura experimental podemos mencionar a multilevel structural equation modeling (MVSEM), em que é possível posicionar qualquer estrutura hierárquica que inclua variáveis latentes (Zitzmann, 2018).

Nos (MVSEM), as unidades são aninhadas em clusters, fazendo com que a suposição de independência entre as unidades seja violada, tornando-se plausível assumir estruturas de correlações. Logo, diferentes abordagens para estender os modelos de equações estruturais, para essas configurações multiníveis, seguindo um modelo estrutural, são encontradas na literatura, visto que a mais comum consiste na formulação de modelos separados dentro e entre os clusters (Rabe-Hesketh et al, 2007).

Convém ressaltar que, em várias abordagens da modelagem de equações estruturais, os índices de qualidade de ajuste, utilizados na validação de modelos de equações estruturais, em sua construção, considera que as variáveis observadas não apresentam repetições, portanto pode-se afirmar que, para tal situação, não há relatos na literatura de índices que contemplam o efeito de repetições na validação de um modelo.

Com essa motivação, este trabalho tem por objetivo propor uma modificação na construção dos índices GFI e AGFI, de tal forma que o efeito das repetições de um conjunto de variáveis observadas poderá ser contemplada na validação de um modelo estrutural.

2 METODOLOGIA

Em concordância com os objetivos propostos, a metodologia estatística proposta, neste trabalho, é dada nas seguintes seções:

2.1 Definição do modelo de equações estruturais e dos parâmetros utilizados no processo de simulação

O modelo de equações estruturais utilizados, neste trabalho, segue as equações abaixo:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (2.1)$$

$$\mathbf{Y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (2.2)$$

$$\mathbf{X} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}. \quad (2.3)$$

Com esperança $\Theta_\delta = E\{\delta\delta'\} = 0$, $\Theta_\varepsilon = E\{\varepsilon\varepsilon'\} = 0$, $\Theta = E\{\xi\xi'\} = 0$, $\Psi = E\{\zeta\zeta'\} = 0$ (McArdle e McDonald, 1984) e ε é não correlacionado com δ , ξ e η e δ é não correlacionado com ε , ξ e η , em que $\boldsymbol{\eta}_{(m \times 1)}$ representou o vetor de variáveis latentes endógenas; $\boldsymbol{\xi}_{(r \times 1)}$ o vetor de variáveis latentes exógenas; $\boldsymbol{\zeta}_{(m \times 1)}$ o vetor dos erros estruturais; $\mathbf{B}_{m \times m}$ a matriz dos coeficientes que relaciona as variáveis latentes endógenas e, por fim, $\boldsymbol{\Gamma}_{(m \times r)}$ a matriz dos coeficientes que relaciona as variáveis latentes exógenas com as variáveis latentes endógenas.

A relação das variáveis latentes endógenas com as variáveis observadas y_j ($j=1, \dots, p$) foi determinada pela matriz dos coeficientes $\boldsymbol{\Lambda}_{y(p \times m)}$, dado o erro de medida representado por $\boldsymbol{\varepsilon}_{(p \times 1)}$.

A Equação (2.3) referiu-se ao modelo de mensuração \mathbf{X} que relaciona às variáveis observadas x_k ($k=1, \dots, q$) à variável latente. A relação entre a variável latente exógena com as variáveis observadas foi especificada, por meio da matriz dos coeficientes $\boldsymbol{\Lambda}_{x(q \times r)}$ e, por fim, o erro de mensuração representado pelo vetor $\boldsymbol{\delta}_{(q \times 1)}$.

A estimação dos parâmetros do modelo estrutural (Figura 1) foi realizada, considerando, apenas, os coeficientes de regressão e as covariâncias entre as variáveis independentes, uma vez que Chou and Bentler (1995) mencionam que os erros $\boldsymbol{\delta}_{(q \times 1)}$, $\boldsymbol{\varepsilon}_{(p \times 1)}$ e $\boldsymbol{\xi}_{(r \times 1)}$ são especificados como variáveis exógenas, ou seja, não recebem setas caracterizando a ausência de influência de outras variáveis. As estimativas dos parâmetros referentes às suas correlações poderão ser suprimidas.

Mantendo as suposições do modelo estrutural, em que as esperanças dos vetores de erros e das variáveis latentes são iguais a zero, $\zeta_{(m \times 1)}$ e $\xi_{(r \times 1)}$ são não correlacionados, $\varepsilon_{(p \times 1)}$ é não correlacionado com $\eta_{(m \times 1)}$, $\xi_{(r \times 1)}$ e $\delta_{(q \times 1)}$, sendo esse não correlacionado com $\xi_{(r \times 1)}$, $\eta_{(m \times 1)}$ e $\varepsilon_{(p \times 1)}$.

Com essas especificações, a descrição do modelo teórico é ilustrada na Figura 1.

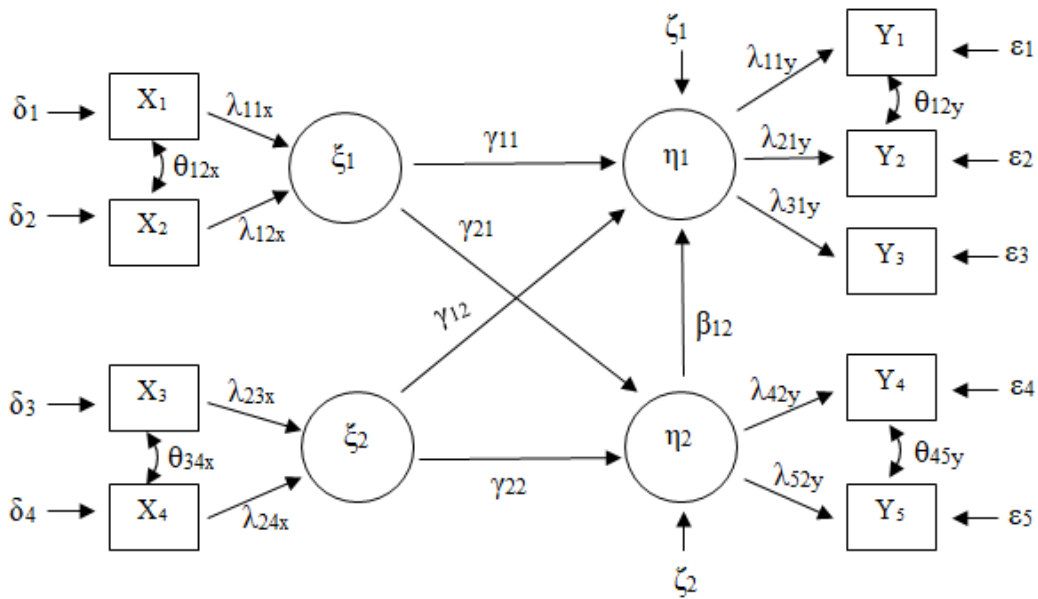


Figura 1 - Modelo de equações estruturais teórico utilizado na simulação Monte Carlo

Fonte: Da autora, 2022.

Para simular a matriz de covariância sobre a hipótese do modelo, utilizou-se o procedimento dado por McDonald e Hartmann (1992), respeitando a relação $\mathbf{v} = \mathbf{A}\mathbf{v} + \mathbf{u}$, matricialmente, dada com os valores paramétricos atribuídos, arbitrariamente, em (2.4).

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ \eta_1 \\ \eta_2 \\ \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0,2 & 0,6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,7 & 0,9 \\ 0 & 0 & 0 & 0 & 0 & -0,3 & 0,3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,4 & 0,4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ \eta_1 \\ \eta_2 \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \zeta_1 \\ \zeta_2 \\ \xi_1 \\ \xi_2 \end{bmatrix} \quad (2.4)$$

em que \mathbf{v} correspondeu ao vetor formado por variáveis observadas e latentes. A matriz dos coeficientes de regressão, com os valores paramétricos assumidos, na simulação Monte Carlo, foi definida em A.

Assumindo a como o número de variáveis definidas em \mathbf{v} e $k = p+q$, o número de variáveis observadas, para que a matriz de covariâncias imposta pelo modelo estrutural (Figura 1), fosse completamente especificada, necessariamente as seguintes matrizes foram definidas:

$$\mathbf{P}_{(a \times a)} = \begin{bmatrix} 1 & 0,3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0,5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -0,7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0,7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0,4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \mathbf{J}_{(k \times a)} = [\mathbf{I}_k : \mathbf{0}], \quad (2.5)$$

em que \mathbf{P} representou a matriz de covariâncias dos elementos do vetor \mathbf{u} , assumindo os valores paramétricos utilizados na simulação Monte Carlo, \mathbf{J} indicou uma matriz cujas primeiras k colunas formam a matriz identidade \mathbf{I}_k .

Com essas especificações, o parâmetro relativo à matriz de covariâncias, imposta pelo modelo estrutural, foi construído por meio da equação (2.6).

$$\Sigma_{\theta} = \mathbf{J}(\mathbf{I}_a - \mathbf{A})^{-1} \mathbf{P} \left[(\mathbf{I}_a - \mathbf{A})^{-1} \right]^t \mathbf{J}^t \quad (2.6)$$

2.2 Obtenção da matriz de covariância amostral considerando repetições nas variáveis observadas

Dada a matriz de covariância, postulada pelo modelo, geraram-se amostras multivariadas com $k = p+q$ variáveis observadas de tamanho $n = 5.000$, representadas por $\mathbf{G} \sim N_{(k)}(0, \Sigma_{\theta})$ sendo a matriz de covariância amostral \mathbf{S} , não considerando repetições entre as variáveis.

As repetições foram incluídas, atribuindo os valores para $r = 2; 4; 8; 10; 20; 40$ e 200 , calculando a razão $t = k/r$, subconjuntos (r,t) foram gerados, em que k ($k=p+q$) representa a soma das p variáveis endógenas com q variáveis exógenas, das r repetições. Desta forma, temos a caracterização de sete delineamentos experimentais multivariados inteiramente casualizados com as atribuições descritas em (2.7), logo o conjunto de variáveis observadas repetidas combinadas é dada abaixo pela sequência (r,t) :

$$(r,t) = (2, 2500); (4,1250); (8,625); (10,500); (20,250); (40,125); (200,25) \quad (2.7)$$

As variáveis respostas simuladas em G (2.8), para as p variáveis observadas exógenas e q endógenas, contemplaram as 5.000 parcelas, associadas, sequencialmente, a cada delineamento (r,t) .

$$\mathbf{G} = \left(\begin{array}{ccc|ccc} \mathbf{g}_{11} & \cdots & \mathbf{g}_{1p} & \mathbf{g}_{11} & \cdots & \mathbf{g}_{1q} \\ \mathbf{g}_{21} & \cdots & \mathbf{g}_{2p} & \mathbf{g}_{21} & \cdots & \mathbf{g}_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{g}_{50001} & \cdots & \mathbf{g}_{5000p} & \mathbf{g}_{5000q} & \cdots & \mathbf{g}_{5000q} \end{array} \right)_{5000 \times (k=p+q)} \quad (2.8)$$

em que $g_{.k}$ é o valor de cada observação, com $k = 1, \dots, 5.000$.

Fixado o número de variáveis observadas repetidas (t), a média, para cada variável (\bar{t}) foi obtida em função do número de repetições r . Dessa forma, computou-se a matriz T , denominada como matriz das médias das variáveis observadas repetidas, conforme encontra-se escrita em (2.9).

$$\mathbf{T} = \begin{pmatrix} \bar{t}_{11} & \cdots & \bar{t}_{p1} & | & \bar{t}_{11} & \cdots & \bar{t}_{q1} \\ \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\ \bar{t}_{1t} & \cdots & \bar{t}_{pt} & | & \bar{t}_{1t} & \cdots & \bar{t}_{qt} \end{pmatrix}_{t \times (k=p+q)} \quad (2.9)$$

Tendo como referência a matriz \mathbf{T} , efetuou-se a correção da média de cada tratamento em relação à média geral de cada variável, conforme recomendação dada por Konishi (2015), originando a matriz \mathbf{M} (2.10).

$$\mathbf{M} = \begin{pmatrix} \bar{t}_{11} - \bar{g}_1 & \cdots & \bar{t}_{p1} - \bar{g}_p & | & \bar{t}_{11} - \bar{g}_1 & \cdots & \bar{t}_{q1} - \bar{g}_q \\ \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\ \bar{t}_{1t} - \bar{g}_1 & \cdots & \bar{t}_{pt} - \bar{g}_p & | & \bar{t}_{1t} - \bar{g}_1 & \cdots & \bar{t}_{qt} - \bar{g}_q \end{pmatrix}_{t \times (k=p+q)} \quad (2.10)$$

Aplicando a decomposição em valores singulares, de modo que $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, sendo \mathbf{U} e \mathbf{V} matrizes ortogonais, cujas colunas são formadas, respectivamente, pelos autovetores de $\mathbf{M}\mathbf{M}^T$ e $\mathbf{M}^T\mathbf{M}$. A matriz de covariância, considerando-se as repetições nas variáveis observadas, foi determinada por $\Sigma_U = \mathbf{U}\mathbf{U}^T$.

2.3 Proposta de aprimoramento dos GFI e AGFI e cenários para avaliação

A proposta de aprimoramento dos índices GFI e AGFI é fundamentada na incorporação da matriz de covariância que representa a informação relativa às repetições das variáveis observadas, dadas por Σ_U . Com esse propósito, os índices GFI e AGFI, considerando-se as formulações obtidas pelos mínimos quadrados ordinários (MQO) (2.11-2.12) e máxima verossimilhança (MV) métodos (2.13-2.14), foram redefinidos, respectivamente por $GFI_{(MQO)}$; $AGFI_{(MQO)}$ e $GFI_{(MV)}$ e $AGFI_{(MV)}$.

$$GFI_{(MQO)} = 1 - \frac{tr \left[\left((S + \Sigma_U) - \Sigma_\theta \right)^2 \right]}{tr \left[(S + \Sigma_U)^2 \right]} \quad (2.11)$$

$$AGFI_{(MQO)} = 1 - \left[\frac{k(k+1)}{2gl_p} \right] \left[1 - GFI_{(MQO)} \right] \quad (2.12)$$

$$GFI_{(MV)} = 1 - \frac{tr \left[\left((\Sigma_\theta)^{-1} (S + \Sigma_U) - I \right)^2 \right]}{tr \left[\left((\Sigma_\theta)^{-1} (S + \Sigma_U) \right)^2 \right]} \quad (2.13)$$

$$AGFI_{(MV)} = 1 - \left[\frac{k(k+1)}{2gl_p} \right] [1 - GFI_{(MV)}] \quad (2.14)$$

Sendo gl_p o grau de liberdade do modelo adotado dado por: $\frac{1}{2} k(k+1) - w$, k é o número de variáveis observadas e w o número de parâmetros livres do modelo.

Entre os cenários utilizados no procedimento de simulação Monte Carlo, considerou-se como critério a eficiência relativa entre os estimadores das covariâncias em função do número

de variáveis observadas, uma vez que $\left(\frac{\det|S|}{\det|\Sigma_\theta|} \right)^{\frac{1}{k}} = GH$.

Assumindo baixos valores para GH, podemos observar uma similaridade na dispersão das matrizes de covariância de S e Σ_θ . Nesse caso, os valores amostrais serão mais concentrados, em relação ao vetor das médias, logo, os índices de qualidade de ajuste terão maior evidência em não rejeitar o modelo proposto. Adotando-se maiores valores para GH, normalmente, aumenta-se a heterogeneidade entre as matrizes de covariâncias S e Σ_θ e, como consequência, maior evidência de que os índices de qualidade de ajuste irão proporcionar resultados favoráveis à rejeição do modelo.

Atribuindo um valor numérico para GH tornou-se possível quantificar a eficiência do estimador da matriz de covariâncias, imposta pelo modelo, em relação à matriz de covariância amostral. Dessa forma, a relação que estabeleceu o grau de heterogeneidade entre essas matrizes, foi explicitada por $S = \sqrt[k]{GH} \Sigma_\theta$ atribuindo os valores para $GH = 2, 4, 6$ e 8 , ressaltando que, dada esta ordem de grandeza, a diferenciação entre as matrizes de covariâncias é dada em pequenas perturbações, por se considerar a raiz k -ésima do grau de heterogeneidade.

Em se tratando da violação da suposição de normalidade multivariada, procedeu-se à simulação de normais multivariadas, contaminadas com *outliers* gerados por distribuições simétricas e com excesso de curtose, neste contexto, consideraram-se as distribuições t-Student e Lognormal, com as probabilidades de mistura fixadas em $\alpha = 0,05$ e $0,15$.

Com especificação desses cenários, resultantes da combinação dos delineamentos (r,t), graus de heterogeneidade GH e probabilidades de misturas, considerando as distribuições retromencionadas, desenvolveu-se um script, no software R versão 4.1.2 (R Development Core Team, 2021), para a validação dos índices propostos, assumindo 1.000 simulações Monte Carlo, nas quais se computou a média das estimativas dos índices propostos.

3 RESULTADOS E DISCUSSÕES

Preliminarmente à discussão dos resultados, convém ressaltar alguns critérios para a interpretação dos resultados dos índices GFI e AGFI que classificam a qualidade de ajuste do modelo. Marsh e Hau, (1996) e Marôco (2010) mencionam que valores de $GFI > 0,95$ indicam um bom ajuste e $0,90$ é um valor de corte aceitável. Sun (2005) afirma que não há critério de corte claro, para esse índice, ele pode variar com a experiência do pesquisador.

3.1 Desempenho dos índices modificados, em relação à qualidade de ajuste do modelo, avaliada pelo grau de heterogeneidade entre as matrizes de covariâncias envolvidas nas formulações dos índices

Com base nas classificações mencionadas por diferentes autores, procedeu-se à discussão dos resultados, descritos inicialmente na Tabela 1.

Tabela 1 - Índice de qualidade de ajuste (GFI) e índice de qualidade de ajuste corrigido (AGFI) pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição normal multivariada com grau de heterogeneidade $GH = 2$.

	Sem repetição	Com repetição	
		Nº Repetições	
GFI _(MV)	0,9937	2	-0,1803 ± 0,361
		4	-0,1026 ± 0,381
		10	0,9577 ± 0,007
		20	0,9625 ± 0,011
		40	0,9620 ± 0,018
		100	0,9620 ± 0,015
AGFI _(MV)	0,9931	2	-0,2955 ± 0,395
		4	-0,1980 ± 0,412
		10	0,9535 ± 0,008
		20	0,9640 ± 0,012
		40	0,9618 ± 0,020
		100	0,9620 ± 0,011
GFI _(MQO)	0,9933	2	0,3128 ± 0,029
		4	0,3117 ± 0,031
		10	0,8786 ± 0,007
		20	0,8882 ± 0,010
		40	0,8893 ± 0,012
		100	0,8802 ± 0,013
AGFI _(MQO)	0,9926	2	0,2446 ± 0,034
		4	0,2457 ± 0,031
		10	0,8626 ± 0,008
		20	0,8773 ± 0,011
		40	0,8727 ± 0,013
		100	0,8700 ± 0,009

Fonte: Da autora, 2022.

Pelos resultados descritos, na Tabela 1, pode-se observar que, para o método de máxima verossimilhança, os valores, para os índices modificados para o método da máxima verossimilhança encontrados, a partir da décima repetição, são considerados de um modelo com bom ajuste, concordando com os resultados encontrados em sua forma convencional. Observa-se que, a partir da vigésima repetição, os índices apresentam uma tendência de estabilidade, o que sugere que não há necessidade de um número grande de repetições de um experimento. Já, os índices obtidos, por meio do método dos mínimos quadrados, segundo os critérios de Marsh e Hau (1996) e Marôco (2010) são indicativos de um modelo mal ajustado.

Sugawara e MacCallum (1993) e Fan et. al (1999) encontram resultados semelhantes; isto é, os índices de ajustes foram, consideravelmente, afetados pelo método de estimativa. Segundo Sugawara e MacCallum (1993), os valores do índice de ajuste, baseados nas estimativas de máxima verossimilhança sempre foram maiores que os valores correspondentes do índice de ajuste baseados no método dos mínimos quadrados generalizados.

Para pequenos números de repetições (duas e quatro) valores negativos foram encontrados, independentemente do método de estimação utilizado e do grau de heterogeneidade como pode ser observado nas tabelas 1 e 2. Yadama e Pandey (1995) e Mulaik et al. (1990) mencionam que é possível encontrar valores negativos, para esses índices, indicando que esses resultados podem ser um diagnóstico de que o modelo está mal ajustado, uma vez que valores negativos são incoerentes mediante a formulação do AGFI. É justificável encontrar valores negativos nessa situação em especial uma vez que os dados simulação são provenientes de uma normal multivariada e essa distribuição apresentam em seu domínio valores negativos.

Aumentando o grau de heterogeneidade ($\text{GH} = 8$) entre as matrizes de covariância, envolvidas nas formulações dos índices, para que se contemple uma situação maior de desajuste, os resultados descritos na Tabela 2 apresentam de um modo geral mesmo comportamento que a Tabela 1 ($\text{GH} = 2$), ou seja, apresentam estabilidade no valor dos índices após a vigésima repetição e os índices obtidos pelo método da máxima verossimilhança foram melhores que os obtidos pelo método dos mínimos quadrados.

Os índices obtidos para maior grau de heterogeneidade ($\text{GH} = 8$) indicam que não houve um bom ajuste de acordo com os critérios adotados, resultados esses esperados de serem encontrado uma vez que essa situação retrata maior distância das matrizes de covariância envolvidas nos índices.

Tabela 2 - Índice de qualidade de ajuste (GFI) e índice de qualidade de ajuste corrigido (AGFI) pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição normal multivariada com grau de heterogeneidade $GH = 8$.

	Sem repetição	Com repetição	
		Nº Repetições	
GFI _(MV)	0,9338	2	-0,6467 ± 0,478
		4	-0,3478 ± 0,522
		10	0,8458 ± 0,016
		20	0,8762 ± 0,013
		40	0,8761 ± 0,014
		100	0,8758 ± 0,014
AGFI _(MV)	0,9274	2	-0,8074 ± 0,524
		4	-0,4086 ± 0,573
		10	0,8307 ± 0,017
		20	0,8641 ± 0,012
		40	0,8640 ± 0,015
		100	0,8637 ± 0,016
GFI _(MQO)	0,9327	2	0,0165 ± 0,039
		4	0,1898 ± 0,044
		10	0,8173 ± 0,013
		20	0,8501 ± 0,014
		40	0,8499 ± 0,012
		100	0,8492 ± 0,012
AGFI _(MQO)	0,9262	2	0,1157 ± 0,043
		4	0,1883 ± 0,048
		10	0,7995 ± 0,015
		20	0,8355 ± 0,013
		40	0,8353 ± 0,013
		100	0,8345 ± 0,013

Fonte: Da autora, 2022.

Comparando os resultados sobre o grau de heterogeneidade (Tabelas 1 e 2), em geral, notou-se que as estimativas dos índices GFI e AGFI, de forma convencional, permaneceram próximas da unidade, indicando um bom ajuste do modelo. O mesmo ocorreu com os índices modificados a partir da décima repetição, quando se utilizou o método da máxima verossimilhança para estimar as matrizes de covariância. É possível destacar que, para o método de máxima verossimilhança para baixo grau crescente de heterogeneidade ($GH = 2$), os índices modificados apresentaram melhor ajuste que os índices em sua forma convencional, implicando robustez dos índices modificados propostos.

Em conformidade com Hoyle e Panter (1995) e Shah e Goldstein (2006), uma amostra que apresente normalidade multivariada para $n \geq 650$ não apresenta efeitos nos índices de qualidade de ajuste normalizados e índices comparativos. Nas situações simuladas neste trabalho, o tamanho amostral considerado foi de 5.000, portanto, mesmo aumentando o grau de heterogeneidade ($GH = 8$), os resultados computados, para ambos os índices neste trabalho, são

plausíveis de serem utilizados, principalmente os com base no método da máxima verossimilhança sendo menos sensível à redução das estimativas.

3.2 Desempenho dos índices modificados em relação ao grau de heterogeneidade e violação da suposição de normalidade multivariada

As Figuras 2(A)-(H) ilustram o efeito de observações *outliers*, nos índices de ajuste propostos, em função do grau de heterogeneidade (GH) das matrizes de covariância amostral e estrutural, do número de repetições e da quantidade de *outliers* (α) geradas pela distribuição t-Student multivariada. Com base nesses resultados, de um modo geral, observou-se que o aumento no número de repetições resultou em um decréscimo, nas estimativas dos índices, com uma estabilização, a partir da décima repetição, independentemente do método de estimação, grau de heterogeneidade e quantidade e porcentagens de *outliers* contidas na amostra original

Outro resultado que merece destaque é verificado, por meio das Figuras 2(A) e (E), em que os valores dos índices GFI e AGFI com baixas e altas contaminações de *outliers* ($\alpha = 5\%$ e 15%), dado um baixo grau de heterogeneidade ($GH = 2$), independentemente do método de estimação, apresentaram valores médios que evidenciam um ajuste aceitável ($GFI > 0,95$).

Para os demais casos (Figuras 2(B) - (D) e 2(F) - (H)), constatou-se uma mínima redução, ao verificar o incremento do grau de heterogeneidade e aumento do número de repetições. Esse fato permite compreender que o aprimoramento dos índices, para incorporar o efeito das repetições, evidencia uma certa robustez, visto que, mesmo com as perturbações violando a pressuposição de normalidade do modelo, causadas pelos fatores avaliados neste trabalho, as estimativas dos índices foram coerentes e sugestivas à interpretação da qualidade de ajuste do modelo estrutural.

Os resultados obtidos na proposta do aprimoramento, em relação à robustez, são concordantes com afirmações dadas por Yuan e Zhong (2013), em relação aos *outliers* terem uma tendência em proporcionar um bom ajuste ao modelo, embora muitos índices baseados na estimação de máxima verossimilhança acusam que o modelo seja ruim.

Yuan e Bentler (2001) apontam uma deficiência que, mesmo se o modelo estiver corretamente especificado, uma pequena proporção de *outliers* poderá distorcer as estimativas da matriz de covariância imposta pelo modelo estrutural, o que supostamente comprometerá fazer qualquer afirmação sobre a qualidade de ajuste do modelo. Diante dessas afirmações, percebe-se que essas deficiências foram corrigidas com incorporação do efeito de repetições nos índices GFI e AGFI modificados neste trabalho.

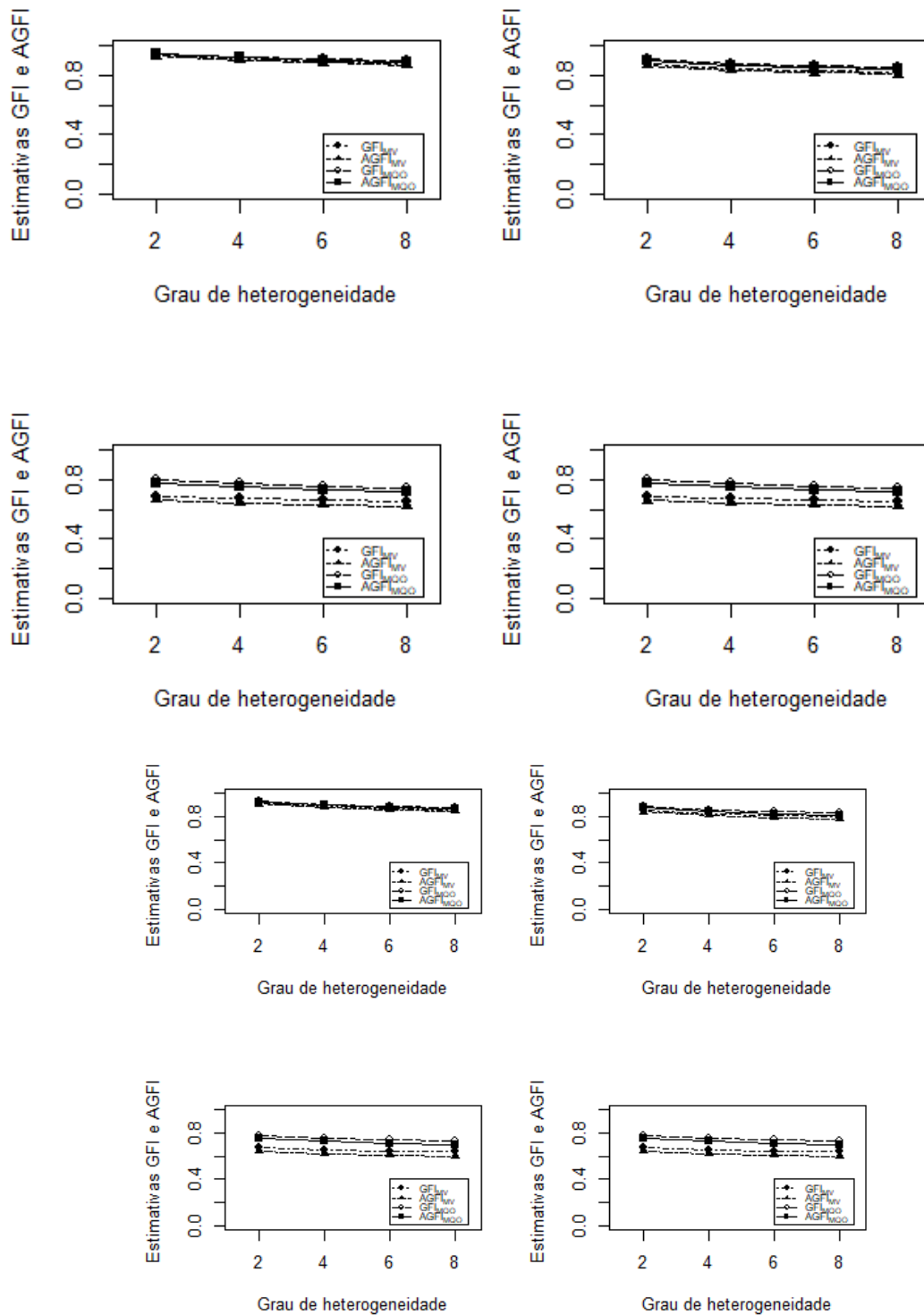


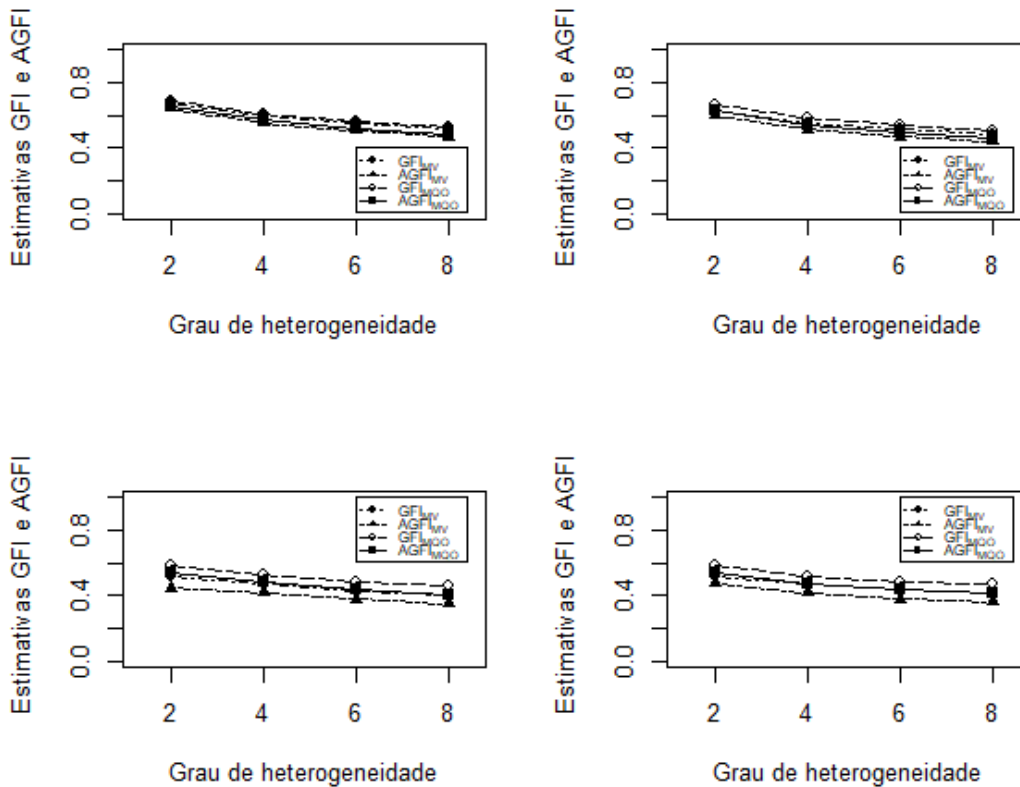
Figura 2 - Desempenho dos índices modificados, em função do grau de heterogeneidade, considerando *outliers* gerados pela t-Student multivariada, com taxa de mistura α e r número de repetições das variáveis observadas: (A) $\alpha=5\%$ e $r=2$; (B) $\alpha=5\%$ e $r=4$; (C) $\alpha=5\%$ e $r=10$; (D) $\alpha=5\%$ e $r=20$; (E) $\alpha=15\%$ e $r=2$; (F) $\alpha=15\%$ e $r=4$; (G) $\alpha=15\%$ e $r=10$; (H) $\alpha=15\%$ e $r=20$.

Fonte: Da autora, 2022.

Considerando os *outliers* gerados pela distribuição Lognormal, na qual a principal característica é assimetria e excesso de curtose, observou-se, por meio das figuras 3(A) - 3(H), que as estimativas dos índices não foram promissoras ao afirmar que o modelo está bem ajustado.

Em baixas concentrações de *outliers* ($\alpha = 5\%$), houve uma redução nas estimativas, conforme o grau de heterogeneidade e do número de repetições foi aumentado, porém alguns resultados relacionados à propriedade de robustez ainda foram preservados (Figuras 3(A)-(D)).

Para uma alta concentração de *outliers* ($\alpha = 15\%$), ocorreu redução expressiva, nos valores dos índices, ao comparar o número de repetições entre dois e quatro da segunda para a quarta repetição (Figuras 3(E) - (H)).



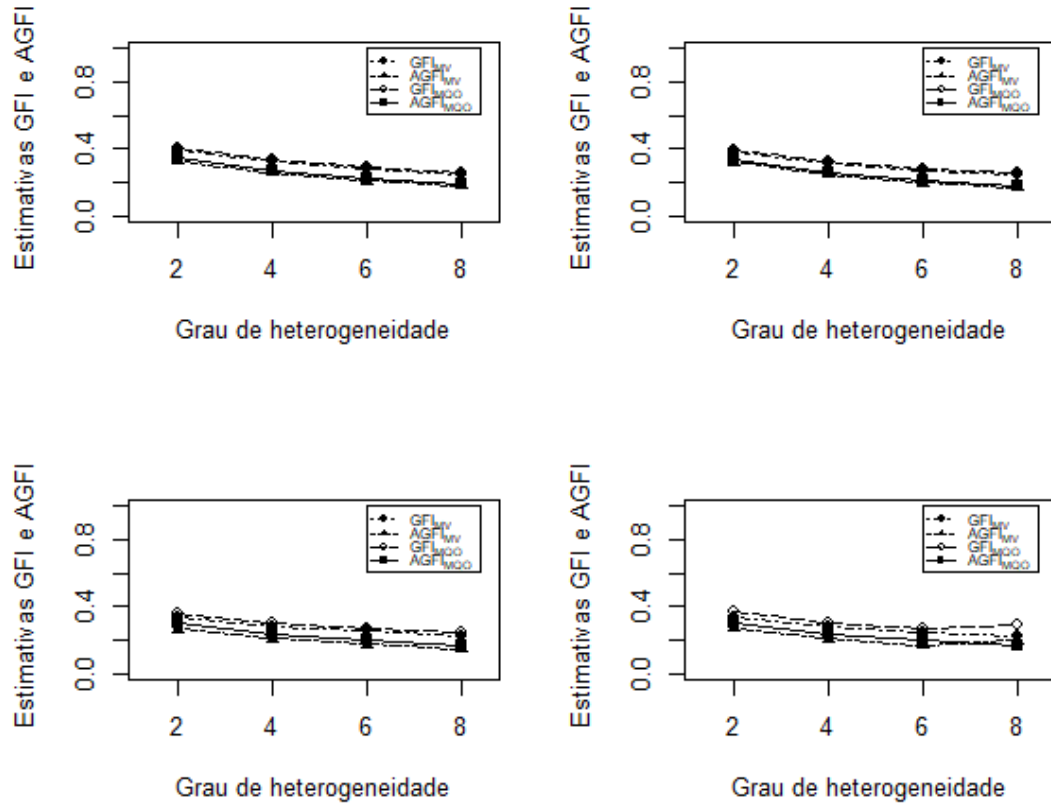


Figura 3 - Desempenho dos índices modificados, em função do grau de heterogeneidade, considerando *outliers* gerados pela Lognormal multivariada, com taxa de mistura α e r número de repetições das variáveis observadas: (A) $\alpha=5\%$ e $r=2$; (B) $\alpha=5\%$ e $r=4$; (C) $\alpha=5\%$ e $r=10$; (D) $\alpha=5\%$ e $r=20$; (E) $\alpha=15\%$ e $r=2$; (F) $\alpha=15\%$ e $r=4$; (G) $\alpha=15\%$ e $r=10$; (H) $\alpha=15\%$ e $r=20$.

Fonte: Da autora, 2022.

3.3 Considerações finais dos índices GFI e AGFI modificados, em relação ao método de estimação, em função do grau de heterogeneidade e violação da suposição de normalidade multivariada

Em função dos cenários contemplados na simulação Monte Carlo, neste estudo observaram-se algumas discrepâncias entre os índices modificados, em relação às abordagens dadas ao método de estimação, mínimos quadrados ordinários e máxima verossimilhança.

A ressaltar, o fato de os *outliers* sejam provenientes de distribuições com baixa e elevada curtose, a exemplificar, t-Student e Lognormal, avaliadas neste trabalho, observou-se, para o caso de baixo grau de heterogeneidade ($\text{GH} = 2$) e baixa concentração de *outliers* ($\alpha = 5\%$) considerando quatro repetições, concluiu-se, que para GFI modificado, os valores médios das estimativas, baseados em nos métodos MV e MQO, foram 0,86 e 0,91, respectivamente, ao passo que, para os mesmos métodos, os valores médios, para o AGFI modificado, foram 0,85 e 0,90. Logo, de acordo com critério de corte de ajuste de modelo, os

valores para a $AGFI_{(MV)}$, em ambos os pares, seriam julgados como indicando ajuste ruim, enquanto, para o $AGFI_{(MQO)}$, os últimos valores seriam interpretados como indicando ajuste aceitável (Sun, 2005).

Para o modelo simulado neste estudo, no qual essas discrepâncias ocorreram, os valores do índice de ajuste, com base no MQO, pareciam invariavelmente indicar um ajuste melhor que aqueles baseados no MV. Esse fato ocorreu para altas e baixas concentrações de *outliers* e para todos os graus de heterogeneidade da matriz de covariância, portanto há evidências de que, dependendo do método de estimação, os índices GFI e AGFI, utilizados em modelos com repetições de variáveis respostas, poderão proporcionar informações relativas à qualidade de ajuste do modelo.

Sugawara e MacCallum (1993) e Fan et. al (1999) encontraram resultados semelhantes, embora esses autores não tenham considerado o número de repetições das variáveis observáveis, no tocante, ao efeito do método de estimação em relação à validação do modelo.

Contudo há controvérsias, se compararmos estudos obtidos por Sugawara e MacCallum (1993), em que respaldam que os valores do índice de ajuste, baseados nas estimativas de MV, sempre foram maiores que os valores correspondentes do índice de ajuste, baseados no Método dos mínimos quadrados generalizados (GLS). Fan et. al. (1999) afirmam que os valores do índice de ajuste com base no GLS indicaram melhor ajuste que aqueles baseados no MV.

Particularizando as amostras contaminadas com *outliers* gerados por distribuição com baixa curtose, verificou-se que as estimativas dos índices GFI e AGFI, na forma convencional, para todos os graus de heterogeneidade e métodos de estimação, mantiveram-se próximos à unidade independente da taxa de contaminação de *outliers* (Tabela 3).

Tabela 3 - GFI e AGFI pelos métodos MV e MQO, obtidos no estudo de simulação da distribuição t-Student e lognormal multivariadas

	t- Student				Lognormal			
	$\alpha = 5\%$							
	$G_H = 2$	$G_H = 4$	$G_H = 6$	$G_H = 8$	$G_H = 2$	$G_H = 4$	$G_H = 6$	$G_H = 8$
$GFI_{(MV)}$	0,9886	0,9702	0,9564	0,9457	0,7192	0,6320	0,5827	0,5464
$AGFI_{(MV)}$	0,9875	0,9673	0,9521	0,9404	0,6918	0,5961	0,5420	0,5021
$GFI_{(MQO)}$	0,9890	0,9707	0,9570	0,9462	0,7224	0,6379	0,5895	0,5560
$AGFI_{(MQO)}$	0,9879	0,9678	0,9528	0,9410	0,6953	0,6026	0,5495	0,5127
	$\alpha = 15\%$							
	$G_H = 2$	$G_H = 4$	$G_H = 6$	$G_H = 8$	$G_H = 2$	$G_H = 4$	$G_H = 6$	$G_H = 8$
$GFI_{(MV)}$	0,9741	0,9504	0,9342	0,9220	0,4076	0,3340	0,2875	0,2568
$AGFI_{(MV)}$	0,9716	0,9457	0,9277	0,9144	0,3498	0,2691	0,2180	0,1843
$GFI_{(MQO)}$	0,9745	0,9513	0,9347	0,9227	0,4140	0,3427	0,2960	0,2659
$AGFI_{(MQO)}$	0,9720	0,9465	0,9283	0,9152	0,3568	0,2785	0,2273	0,1943

Fonte: Da autora, 2022.

Em se tratando das amostras contaminadas com *outliers*, pertencentes a uma distribuição com excesso de curtose, no caso a Lognormal, verificou-se que os resultados foram mais discrepantes, em relação aumento do grau de heterogeneidade das matrizes de covariâncias amostrais e estruturais com uma redução nas estimativas médias dos índices GFI e AGFI modificados.

Gao et al. (2008) ressaltam que, considerando um tamanho amostral moderado, embora os resultados apresentem-se desejáveis, a ausência de normalidade em uma amostra conduz a uma superestimação da estatística qui-quadrado, bem como nos demais índices de qualidade de ajuste que avaliam o grau de discrepância entre as matrizes de covariância amostral e estrutural.

Mediante o exposto, percebe-se, por meio dos resultados encontrados (Tabela 3), que essa deficiência foi corrigida, uma vez que as estimativas dos índices GFI e AGFI convencionais e modificados encontrados foram semelhantes em todos os casos simulados.

4 CONCLUSÕES

Conclui-se que o aprimoramento dos índices GFI e AGFI para incorporar o efeito das repetições das variáveis observadas foi promissor, pois apresentou propriedades de robustez, em relação à presença de *outliers* e mostrou evidências de diagnóstico de um modelo com bom e ruim ajuste, mediante os graus de heterogeneidade impostos entre as matrizes de covariâncias amostral e hipotetizada pelo modelo.

Quando utilizados dados contaminados com *outliers* ressalta-se maior performance o método de mínimos quadrados. Recomenda-se utilizar até 20 repetições uma vez os índices se estabilizam a partir desse valor.

REFERÊNCIAS

- Cirillo M A, Barroso L P (2017) "Effect of outliers on the GFI quality adjustment index in structural equation model and proposal of alternative index." *Communications in Statistics - Simulation and Computation* 46(3):1895-1905.
- Cheung G W, Rensvold R B (2002) "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 9(2):233-255.
- Chou C P, Bentler P M (1995) "Estimates and tests in Structural Equation Modeling." *Structural Equation Modeling: Concepts, Issues and Applications* 37-55.
- Fan X, Thompson B, Wang L (1999) "Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes." *Structural Equation Modeling: A Multidisciplinary Journal* 6(1):56-83.
- Gao S, Mokhtarian P L, Johnston R A (2008) "Nonnormality of Data in Structural Equation Models." *Transportation Research Record* 2082(2082):116-124.
- Hoyle R H, Panter A T (1995) "Writing about Structural Equation Models." *Structural Equation Modeling: Concepts, Issues and Applications* 1:159-176.
- Jackson D A, Chen Y (2004) "Robust principal component analysis and outlier detection with ecological data." *Environmetrics* 15(2):129-139.
- Konishi, T (2015) "Principal component analysis for designed experiments." *BMC Bioinformatics* 16(18):9-11.
- Marsh H W, Hau K T (1996) "Assessing goodness of fit: Is parsimony always desirable?" *The Journal of Experimental Education* 64(4):364-390.
- McArdle J J, McDonald R P (1984) "Some algebraic properties of the Reticular Action Model for moment structures." *British Journal of Mathematical and Statistical Psychology* 37:234-251.
- McDonald R P, Hartmann W M (1992) "A procedure for obtaining initial values of parameters in the RAM model." *Multivariate Behavioral Research* 27(1):57-76.
- Mulaik S A, James L R, Van Alstine J, Bonnett N, Lind S, Stillwell C D (1989) "Evaluation of goodness of fit indices for structural equation models." *Psychological Bulletin* 105(3):430-445.
- R Core Team (2021) "R: A language and environment for statistical computing." *R Foundation for Statistical Computing*, Vienna, Austria. Software available at <https://www.Rproject.org/>

- Rabe-Hesketh S, Skrondal A, Zheng X (2007) "Multilevel Structural Equation Modeling in Handbook of Latent Variable and Related Models", L. Sik-Yum, eds, North-Holland, Amsterdam, 209-227.
- Shah R, Goldstein S M (2006) "Use of Structural Equation Modeling in Operations Management Research: Looking Back and Forward." *Journal of Operations Management* 24(2):148--169.
- Sugawara H M, MacCallum R C (1993) "Effect of estimation method on incremental fit indexes for covariance structure models." *Applied Psychological Measurement* 17:365-377.
- Sun J (2005) "Assessing Goodness of Fit in Confirmatory Factor Analysis." *Measurement and Evaluation in Counseling and Development*, 37(4):240-256.
- Yadama G N, Pandey S (1995) "Effect of Sample Size on Goodness-Fit of-Fit Indices in Structural Equation Models." *Journal of Social Service Research* 20(3-4):49-70.
- Yuan K H, Bentler P M (2001)"Effects of "Outliers" on Estimators and Tests in Covariance Structure Analysis." *British Journal of Mathematical and Statistical Psychology* 54(1):161-175.
- Yuan K H, Zhong X (2013) "Robustness of fit indices to outliers and leverage observations in structural equation modeling." *Psychological Methods* 18(2):121-136.
- Ullman J B (2007)"Structural equation modeling", in *Using multivariate statistics*, Boston, MA: Pearson Education.
- Zitzmann S (2018) "A Computationally More Efficient and More Accurate Stepwise Approach for Correcting for Sampling Error and Measurement Error." *Multivariate Behavioral Research* 53(5):612-632.

**ARTIGO 2 - UM ESTUDO DE CASO DA MODELAGEM DA COVARIÂNCIA
ENTRE FATORES EXTERNOS RELACIONADO A PERCEPÇÃO SENSORIAL DE
CAFÉS**

**Artigo redigido conforme a NBR 6022 (ABNT, 2018) e formatado de acordo com o
Manual da UFLA de apresentação de teses e dissertações.**

RESUMO

A análise e inferência das percepções sensoriais em bebidas de cafés é complexa, justamente por inúmeras causas aleatórias intrínsecas à produtividade, preparo e, principalmente, subjetividade dos consumidores e/ou provadores. Diante do exposto, variáveis latentes, normalmente, compostas de combinação de outras variáveis observadas, são descartadas nas análises convencionais. Seguindo essa argumentação, objetivou-se, neste trabalho, pautar-se na proposição de um modelo de equações estruturais aplicado em banco de dados, indicação geográfica dos cafés da Serra da Mantiqueira, com uma contribuição metodológica caracterizada pela inclusão de efeito de tratamento, contemplado por diferentes altitudes, nas quais os cafés foram produzidos. Conclui-se que, dado o modelo proposto, as informações representadas nas variáveis latentes, produção e variações ambientais, de fato, influenciam na percepção sensorial, para os cafés produzidos na região da Serra da Mantiqueira e o aprimoramento dos índices de validade de ajuste, apresentou melhorias em validação do modelo, dada a inclusão das repetições das variáveis observadas.

Palavras-chave: Altitude. Sabor. Pluvisiodade. Acidez. GFI. AGFI.

1 INTRODUÇÃO

Preliminarmente, à definição da qualidade da bebida de café, existe uma cadeia de processos que, em síntese, envolve questões relacionadas ao plantio e variações ambientais, que supostamente, são factíveis de serem considerados como fatores externos que contribuem na qualidade da bebida, mormente, avaliada por diversos atributos sensoriais, com destaque, ao sabor e acidez, sendo estes, os mais apreciados pelo consumidor que apresente um grau inferior ou superior em relação a sua percepção sensorial.

No tocante à discriminação da percepção sensorial que desperte interesse ao consumidor em ter conhecimento de informações mais técnicas sobre a aquisição de um determinado café, Santos, Cirillo e Guimarães (2021) destacam uma pesquisa realizada, nos Estados Unidos e no Canadá, corroborada por Quintão e Brito (2016) e Quintão, Brito e Belk (2017a, 2017b), que identificaram a existência de comunidades heterogêneas relacionadas ao consumo de cafés especiais, ou seja, cafés de elevada pontuação em concursos internacionais. Assim, os autores subdividiram tais consumidores em três categorias, associadas em seu nível de expertise relacionado ao conhecimento conceitual e perceptual do produto.

A contextualização de diferentes segmentos de consumo pode ser verificada no interesse crescente dos consumidores, a ser observado, principalmente, nos jovens, que, cada vez mais, frequentam cafeterias que proporcionem cafés com sabores distintos e aromas intensos, resultando em bebidas agradáveis.

Nota-se também que o consumidor, apreciador de bebidas de qualidade, tem uma curiosidade em saber questões técnicas, como: grau de torração, preparo, origem e certificações, uma vez que fatores externos que contribuem para a qualidade do café, a citar formas de processamento, granulometria, microbiologia, ambientais e físico-química, repercutem diretamente na qualidade final do café e, certamente, com influências nas análises sensoriais para a sua validação.

No tocante às projeções de consumo, conforme relatório da pesquisa de tendências do mercado de cafés em 2017, apresentadas pela Associação Brasileira da Indústria de Café (AOUN, 2017), demonstram que o consumo de café aumentou 3,5%, em 2017, comparado ao ano anterior (SOCIEDADE NACIONAL DE AGRICULTURA - SNA, 2018). Tal fato reflete no fato de que os apreciadores de café continuam, cada vez mais, consumindo o produto sob as mais diferentes formas, em função da melhoria da qualidade dos cafés do Brasil.

Com todas essas argumentações, com destaque à subjetividade de cada consumidor, a aferição da qualidade e aptidões são limitadas, no uso de metodologias estatísticas uni ou

multivariadas, uma vez que, valorizam apenas respostas médias e/ou individuais, obtidas em variáveis observadas no decorrer do experimento.

Especificamente, ao estudo da dispersão, por meio da modelagem de uma estrutura de covariância, que contemple a inclusão de variáveis latentes definidas, por meio de construtos que envolvam características relacionadas a fatores externos, contextualizados, neste trabalho pela variável latente manejo e descrição ambiental, e especificação da percepção sensorial.

A composição desses construtos, por meio de algumas variáveis observadas, bem como, a formalização do modelo hipotético, encontra-se descrita em um diagrama ilustrado na metodologia.

Assim, como um estudo de caso definido como objetivo, deste trabalho, considerou uma amostra representativa da produção de cafés especiais em diversas lavouras, identificadas por meio, das coordenadas geográficas e variáveis referentes à área de plantio e controle de praga, neste trabalho, objetivou-se propor um modelo de equações estruturais para modelar a covariância que relacione a percepção sensorial da qualidade de cafés avaliada pelos atributos sabor e acidez, obtidos em diferentes altitudes.

2 MATERIAL E MÉTODOS

2.1 Caracterização do experimento

A base de dados utilizadas, para a formulação do modelo de equações estruturais, correspondeu a uma parte do banco de dados do projeto Identidade e Rastreabilidade com indicação geográfica dos cafés da Região da serra da Mantiqueira (BORÉM *et al.*, 2020), no qual a coleta das amostras de café (*Coffea arabica* L.) foi realizada, ao longo de três safras consecutivas (2010/11, 2011/12 e 2012/13), em lavouras comerciais de propriedades localizadas no município de Carmo de Minas, Minas Gerais, Brasil.

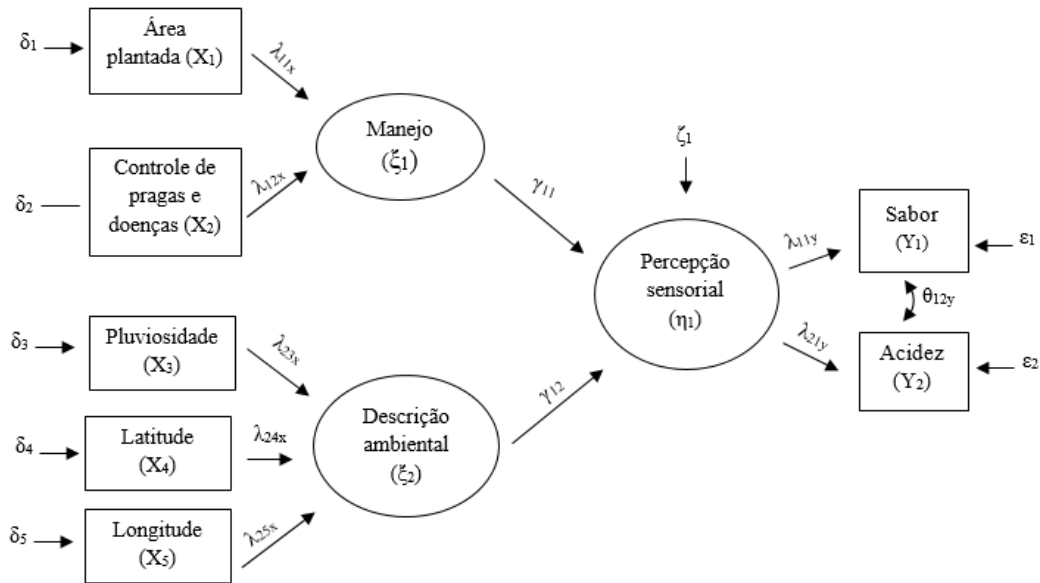
A área de estudo selecionada possui localização delimitada pelas coordenadas geográficas 22°07'21" de latitude sul e 45°07'45" de longitude oeste, com a extensão territorial de 32.332 ha.

O ambiente de cultivo do café foi estratificado em diferentes classes de altitude. Todos os pontos de coleta foram georreferenciados. As variáveis utilizadas, neste trabalho, encontram-se descritas na Figura 1.

2.2 Especificação do modelo estrutural

Em função dos dados obtidos, a formulação do modelo de equações estruturais encontra-se ilustrada, na Figura 1, com a especificação dos fatores representados pelos construtos que definem as variáveis latentes: produção, variações ambientais e percepções sensoriais.

Figura 1 - Diagrama de caminhos para a análise sensorial de cafés especiais



Nota: Os retângulos de x e y referem-se à variável indicadora; setas unilaterais representam efeitos causais; setas bilaterais representam correlação entre as variáveis indicadoras; as elipses η_1 , ξ_1 e ξ_2 representam as variáveis latentes.

Fonte: Da autora (2022).

A relação que explique essas tendências é dada por um conjunto de equações, que determinam a causalidade entre os atributos (variáveis observadas) e as variáveis latentes, conforme descrevem as equações a seguir (BOLLEN; NOBLE, 2011).

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (1)$$

$$\mathbf{X} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \quad (2)$$

$$\mathbf{Y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (3)$$

em que $\boldsymbol{\eta}_{(m \times 1)}$ representou o vetor de variáveis latentes endógenas; $\boldsymbol{\xi}_{(n \times 1)}$ muda o índice o vetor de variáveis latentes exógenas; $\boldsymbol{\zeta}_{(m \times 1)}$ o vetor dos erros estruturais; $\mathbf{B}_{m \times m}$ a matriz dos coeficientes que relaciona as variáveis latentes endógenas e, por fim, $\boldsymbol{\Gamma}_{(m \times r)}$ a matriz dos coeficientes que relaciona as variáveis latentes exógenas com as variáveis latentes endógenas. A relação das variáveis latentes endógenas com as variáveis observadas y_j ($j=1,2$) foi determinada pela matriz dos coeficientes $\boldsymbol{\Lambda}_{y(p \times m)}$, dado o erro de medida representado por $\boldsymbol{\varepsilon}_{(p \times 1)}$.

Na Equação (3), refere-se ao modelo de mensuração \mathbf{X} que relaciona as variáveis observadas x_k ($k=1,\dots,5$) à variável latente. A relação entre a variável latente exógena com as

variáveis observadas foi especificada, por meio da matriz dos coeficientes $\Lambda_{x(q \times r)}$ e, por fim, o erro de mensuração representado pelo vetor $\delta_{(q \times 1)}$.

Matematicamente, o diagrama (Figura 1) é representado pelas equações descritas matricialmente por Hampton (2015). Desta forma, para a Equação (1) temos

$$[\eta_1] = [\gamma_{11} \quad \gamma_{12}] \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\zeta_1] \quad \text{sendo} \quad (4)$$

os construtos definidos pelas variáveis latentes endógenas (η_1) e exógenas (ξ), hipoteticamente, representando os construtos percepções sensoriais, variações ambientais e produção.

As relações lineares entre essas variáveis encontram-se descritas em (2) e (3), respectivamente, matricialmente são definidas por (5) e (6).

$$\Lambda_x = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & \lambda_{23} \\ 0 & \lambda_{24} \\ 0 & \lambda_{25} \end{bmatrix}; \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & \lambda_{23} \\ 0 & \lambda_{24} \\ 0 & \lambda_{25} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} \quad (5)$$

$$\Lambda_y = \begin{bmatrix} \lambda_{12} \\ \lambda_{21} \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_{12} \\ \lambda_{21} \end{bmatrix} [\eta_1] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (6)$$

Cabe justificar que, em determinadas situações, torna-se previsível variação ambiental pelo genótipo da planta, pela procedência e pela forma de cultivo e processamento, contudo, podem existir oscilações de temperatura em uma mesma planta de acordo com a face exposta ao sol, pluviosidade. A análise conjunta de todos esses fatores, de forma precisa, dificulta a determinação de uma causa, nesse sentido pode ser considerado como uma variável latente.

A estimação da matriz de correlação correspondente ao modelo hipotético (Figura 1) foi feita, seguindo o procedimento definido por McDonald e Hartmann (1992), respeitando a relação $\mathbf{v} = \mathbf{A}\mathbf{v} + \mathbf{u}$, sendo expressa pelo conjunto de regressões lineares, representadas no diagrama por setas. Assim, \mathbf{v} refere-se ao vetor formado por variáveis observadas e latentes; \mathbf{A} matriz dos coeficientes de regressão que foram especificadas pelas cargas fatoriais obtidas no ajuste do modelo e, por fim, \mathbf{u} o vetor de erros estimado pelo método dos mínimos quadrados. Logo a matriz de covariância sobre o modelo hipotético (Diagrama 1) é obtida por (7)

$$\hat{\Sigma}_0 = \mathbf{J}(\mathbf{I}_a - \mathbf{A})^{-1} \mathbf{P} \left[(\mathbf{I}_a - \mathbf{A})^{-1} \right]^t \mathbf{J}^t \quad (7)$$

2.3 Incorporação do delineamento com efeito de tratamentos dado por diferentes categorias de altitude e adaptação dos índices de qualidade de ajuste

Em razão da matriz de covariância postulada pelo modelo (7) e a matriz de covariância amostral S, sem considerar as repetições entre as variáveis, a inclusão das repetições foi feita definindo a razão $t = k/r$ em que k ($k=p+q$) representa a soma das p variáveis endógenas, com q variáveis exógenas, das r repetições. Assim, fixado o número de variáveis observadas repetidas (t), a média para cada variável será obtida em função do número de repetições r . Seguindo esse procedimento, mediante as classes de altitudes, cinco tratamentos foram determinados, descritos na Tabela 1.

Tabela 1 - Tratamentos utilizados no delineamento experimental com base na altitude.

Tratamento	Altitude
T1	900 m - 1.000 m
T2	1.000 m - 1.100 m
T3	1.100 m - 1.200 m
T4	1.200 m - 1.300 m
T5	1.300 m - 1.400 m

Fonte: Da autora (2022).

Determinado os tratamentos, gerou-se a matriz T (8), denominada como matriz das médias das variáveis observadas repetidas, conforme encontra-se escrita abaixo:

$$\mathbf{T} = \begin{pmatrix} \bar{t}_{11} & \bar{t}_{21} & | & \bar{t}_{11} & \bar{t}_{21} \\ \vdots & \vdots & | & \vdots & \vdots \\ \bar{t}_{15} & \bar{t}_{25} & | & \bar{t}_{15} & \bar{t}_{25} \end{pmatrix}_{5 \times 4} \quad (8)$$

Tendo como referência a matriz T, efetuou-se a correção da média de cada tratamento, em relação à média geral de cada variável, conforme recomendação dada por Konishi (2015), originando a matriz M (9).

$$M = \begin{pmatrix} \bar{t}_{11} - \bar{x}_1 & \bar{t}_{21} - \bar{x}_2 & | & \bar{t}_{11} - \bar{y}_1 & \bar{t}_{21} - \bar{y}_2 \\ \vdots & \vdots & | & \vdots & \vdots \\ \bar{t}_{15} - \bar{x}_1 & \bar{t}_{25} - \bar{x}_2 & | & \bar{t}_{15} - \bar{y}_1 & \bar{t}_{25} - \bar{y}_2 \end{pmatrix}_{5 \times 4} \quad (9)$$

Posteriormente, foi aplicada a decomposição do valor singular, de modo que $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ sendo \mathbf{U} e \mathbf{V} matrizes ortogonais, cujas colunas são formadas, respectivamente, pelos autovetores de $\mathbf{M}\mathbf{M}^T$ e $\mathbf{M}^T\mathbf{M}$. A matriz de covariância, considerando-se as repetições nas variáveis observadas, será determinado por $\Sigma\mathbf{U} = \mathbf{U}\mathbf{U}^T$.

O índice da qualidade do ajuste (GFI, *Goodness of Fit Index*) mede quanto a quantidade relativa de covariância na amostra (\mathbf{S}) é predita pela matriz de covariância do modelo ($\hat{\Sigma}_\theta$) e o índice de qualidade de ajuste corrigido (AGFI, *Adjusted Goodness of Fit Index*) foi proposto com a finalidade de melhor adequar o GFI aos graus de liberdade e à quantidade de variáveis observadas. Geralmente, os dois são maiores que 0 e valem no máximo 1.

Os índices de qualidade de ajuste GFI e AGFI, aprimorados com a incorporação da matriz de covariância, que representa a informação relativa às repetições das variáveis observadas, considerando as formulações obtidas pelos mínimos quadrados ordinários (MQO), são dados por (10) e (11).

$$\text{GFI}_{(\text{MQO})} = 1 - \frac{\text{tr} \left[\left((\mathbf{S} + \Sigma_U) - \hat{\Sigma}_\theta \right)^2 \right]}{\text{tr} \left[(\mathbf{S} + \Sigma_U)^2 \right]} \quad (10)$$

$$\text{AGFI}_{(\text{MQO})} = 1 - \left[\frac{k(k+1)}{2gl_p} \right] \left[1 - \text{GFI}_{(\text{MQO})} \right] \quad (11)$$

em que $\hat{\Sigma}_\theta$ é a matriz de covariância populacional do modelo que está sendo proposto; \mathbf{S} é a matriz de covariância amostral; $\mathbf{S} + \hat{\Sigma}_\theta$ é a matriz de covariância com efeito de repetições e tratamentos (experimental); $\text{tr}(\cdot)$ é o operador traço; p a quantidade de variáveis observadas endógenas (Y) e q a quantidade de variáveis observadas exógenas (X) e, por fim, gl_p é o número de graus de liberdade do modelo do pesquisador.

Para a análise dos dados, segundo a metodologia proposta, foi utilizado o pacote lavaan (ROSSEEL, 2012) do software R versão 4.1.2 (R DEVELOPMENT CORE TEAM, 2021), para se obter a matriz de covariância do modelo e foi desenvolvida uma rotina, no mesmo software, para a obtenção dos índices de qualidade de ajuste modificados.

3 RESULTADOS E DISCUSSÕES

Preliminarmente, a modelagem da estrutura de covariância (7), imposta pelo modelo (Figura 1), as justificativas para seleção das variáveis que formam cada construto, representada pelas variáveis latentes são dadas a seguir.

A precipitação é um dos atributos climatológicos que mais contribui nas fases fenológicas do cafeeiro e é utilizada para ordenar, geograficamente, áreas aptas, restritas e inaptas ao cultivo. Atualmente, regiões restritas à disponibilidade hídrica, mas que adotaram práticas de irrigação suplementar, têm apresentado destaque, tanto na produtividade quanto na qualidade da bebida do café (BARBOSA *et al.*, 2012). WeldeMichael e Teferi (2020) relatam que, dos fatores ambientais mais frequentemente mencionados, a precipitação é considerada um fator negativo para a qualidade do café.

A acidez pode ser um atributo desejável ou não, dependendo do ácido predominante na bebida. Uma acidez agradável contribui para a vivacidade do café, aumentando a percepção da doçura e confere característica de fruta fresca. Em geral, cafés que apresentam baixa acidez não apresentam nota sensorial elevada (SPECIALTY COFFEE ASSOCIATION OF AMERICA - SCAA, 2008). Geralmente, os cafés despulpados apresentam atributos sensoriais que os distinguem dos cafés processados por via seca, como acidez mais acentuada.

O sabor representa a principal característica do café. Essa avaliação reflete a combinação de todas as percepções observadas na degustação. A pontuação dada ao sabor relata a intensidade, qualidade e complexidade da combinação entre o gosto e aroma sentido quando o café é provado (SCAA, 2008).

Isso posto, em concordância com a metodologia proposta, os resultados referentes à validação do modelo estrutural proposto, que justifique a inclusão de repetições e tratamentos, é dada de forma comparativa ao ajuste convencional. Dessa forma, os resultados encontram-se descritos na Tabela 2.

Tabela 2 - Estimativas dos índices de qualidade de ajuste GFI AGFI para validar o modelo de equações estruturais com a inclusão de delineamento inteiramente casualizado com faixas de altitudes definidas como tratamento em estimativas de diferentes números de repetições.

	Qtde para estimar o número de repetições		Estimativas dos índices	
	Número de repetições		Com repetições	Sem repetições
GFI _(MQO)	2		0,6087	0,1988
	5		0,6129	0,2125
	10		0,7769	0,6852
	20		0,9417	0,8977
AGFI _(MQO)	2		0,5616	0,1027
	5		0,5665	0,1779
	10		0,7501	0,5848
	20		0,9023	0,8329

Fonte: Da autora (2022).

Por meio dos resultados descritos na Tabela 2, nota-se que, em todas as situações, a inclusão das repetições proporcionou uma melhora nos índices de qualidade de ajuste avaliados. Tal fato sugere que, não considerar a informação dos tratamentos, representados pelas diferentes faixas de altitudes, poderá ocasionar uma falsa interpretação, bem como uma subestimativa entre a proximidade das matrizes de covariância entre os dados amostrais e o modelo, uma vez que a formulação dos índices segue esta fundamentação.

Pode-se observar que, de um modo geral, o aumento do número de repetições, tem-se evidências estatísticas da melhoria das estimativas dos índices. Assim, reportando-se às considerações de uma melhora no ajuste, cujo critério estabelecido por Joreskog (1993), considera-se um ajuste excelente, estimativas dos índices GFI acima de 0,90. Logo, na aplicação proposta, neste trabalho, pode-se recomendar 20 repetições das variáveis observadas para cada faixa de altitude, dado o modelo hipotético (Figura 1) que justifica a modelagem da covariância.

Contudo, ainda sobre os resultados, na Tabela 2, pode-se observar que ao considerar apenas 2 repetições, tem-se o ajuste pode ser considerado moderado, portanto, se o objetivo em consiste em modelar a covariância exigindo um número mínimo de repetições, ainda sem perda de generalidade, pode-se considerar o ajuste do modelo adequado.

Com a obtenção da matriz de covariância ajustada pelo modelo ($\hat{\Sigma}_\theta$), incluindo as informações das repetições dos tratamentos (altitudes), o usuário poderá utilizá-la em diversas metodologias de análise de dados, que envolvam redução de dimensão, assim exemplificam-

se: análises de componentes principais, biplots, escalonamento multidimensional, análise discriminante, enfim, técnicas que pressupõem, como entrada de dados, a matriz de covariância.

A título de ilustração, segue uma aplicação da técnica de componentes principais, com a descrição do principal resultado, a ser verificado na porcentagem da explicação da variação amostral dada pelos componentes principais (Figuras 2-4). Para tal aplicação, considerou-se a matrizes de correlação, geradas a partir de $(\hat{\Sigma}_\theta)$, para 2 e 20 repetições, definidas respectivamente.

$$R_\theta = \begin{bmatrix} 1,000 & & & & & & & \\ -0,347 & 1,000 & & & & & & \\ -0,106 & -0,041 & 1,000 & & & & & \\ 0,142 & 0,179 & -0,176 & 1,000 & & & & \\ -0,092 & -0,026 & 0,127 & -0,079 & 1,000 & & & \\ -0,142 & -0,025 & 0,151 & -0,103 & 0,919 & 1,000 & & \\ -0,144 & 0,126 & -0,002 & 0,382 & -0,016 & 0,004 & 1,000 & \end{bmatrix} \quad (12)$$

$$R_\theta = \begin{bmatrix} 1,000 & & & & & & & \\ -0,327 & 1,000 & & & & & & \\ -0,193 & -0,012 & 1,000 & & & & & \\ 0,222 & 0,105 & -0,199 & 1,000 & & & & \\ -0,082 & -0,121 & 0,058 & -0,091 & 1,000 & & & \\ -0,117 & -0,104 & 0,113 & -0,110 & 0,928 & 1,000 & & \\ -0,115 & 0,287 & -0,061 & 0,558 & -0,083 & -0,067 & 1,000 & \end{bmatrix} \quad (13)$$

Figura 2 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação usual.

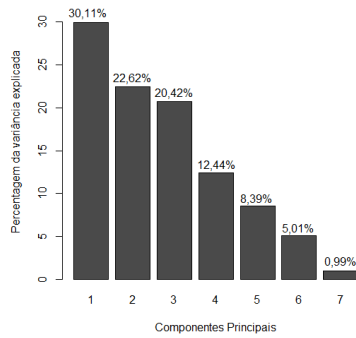


Figura 3 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação R_{θ} e com 2 repetições para os tratamentos.

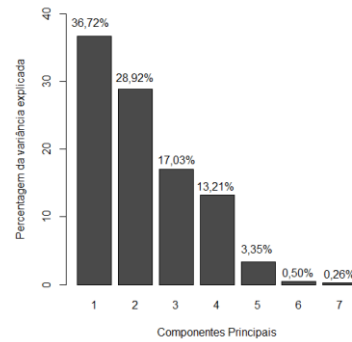
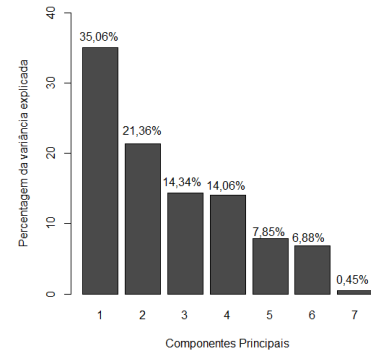


Figura 4 - Porcentagem da variação amostral explicada pelos componentes com matriz de correlação R_{θ} e com 20 repetições para os tratamentos.



Fonte: Da autora (2022).

Considerando a restituição da variabilidade amostral nos dois primeiros componentes, por meio das Figuras 2-4, nota-se que, com a inclusão das variáveis latentes e a estrutura do delineamento na estimativa da matriz de correlação (R_{θ}), tem-se um acúmulo de 65,64%. Aumentando o número de repetições, o acúmulo foi 56,42%. Em ambas as situações, percebe-se uma vantagem na aplicação da análise de componentes principais, na qual, considerando apenas a matriz de correlação amostral (R), o acúmulo foi de 52,73%.

4 CONCLUSÕES

A Modelagem da matriz de covariância proposta, neste trabalho, como estudo de caso, mostrou-se vantajosa por permitir a análise da relação de variáveis latentes que não são consideradas em uma análise sensorial.

A metodologia proposta permitiu incorporar variáveis ambientais, como, por exemplo, altitudes a serem consideradas repetições das variáveis observadas.

A estrutura de correlação gerada, a partir da matriz de covariância ajustada pelo modelo, resultou em estimativas mais promissoras ao aplicar a técnica de análise de componentes principais, sendo esta, um exemplo ilustrativo como instrumento de recomendação a ser empregada em outras técnicas de redução de dimensionalidade.

REFERÊNCIAS

- AOUN, L. A. **Demanda interna por cafés especiais deve ter expansão de até 7% em 2017**. 2017. Disponível em: http://www.redepeavirus.com.br/redes/form/post?topico_id=68170. Acesso em: 3 abr. 2022.
- BARBOSA, J. N. *et al.* Coffee quality and its interactions with environmental factors in Minas Gerais, Brazil. **Journal of Agricultural Science**, Cambridge, v. 4, p. 181-190, 2012.
- BOLLEN, K. A.; NOBLE, M. D. Structural equation models and the quantification of behavior. **PNAS**, Washington, DC, v. 108, n. 3, p. 15639-15646, 2011.
- BORÉM, F. M. *et al.* Coffee sensory quality study based on spatial distribution in the mantiqueira mountain region of Brazil. **Journal of Sensory Studies**, Westport, v. 35, n. 2, p. e12552, 2020.
- HAMPTON, C. Estimating and reporting structural equation models with behavioral accounting data. **Behavioral Research in Accounting**, Jacksonville, v. 27, n. 2, p. 1-34, 2015.
- JORESKOG, K. G. Testing structural equation models. *In*: BOLLEN, K. A.; LONG, J. S. (ed.). **Testing structural equation models**. Newbury Park: Sage, 1993. p. 294-316.
- KONISHI, T. Principal component analysis for designed experiments. **BMC Bioinformatics**, Tokyo, v. 16, n. 18, p. 9-11, 2015.
- MCDONALD, R. P.; HARTMANN, W. M. A procedure for obtaining initial values of parameters in the RAM model. **Multivariate Behavioral Research**, Fort Worth, v. 27, n. 1, p. 57-76, 1992.
- QUINTÃO, R. T.; BRITO, E. P. Z. Connoisseurship consumption and market evolution: an institutional theory perspective on the growth of specialty coffee consumption in the USA. **Brazilian Journal of Marketing**, São Paulo, v. 15, n. 1, p. 1-15, 2016.
- QUINTÃO, R. T.; BRITO, E. P. Z.; BELK, R. W. Connoisseurship consumption community and its dynamics. **Review of Business Management**, São Paulo, v. 19, n. 63, p. 1-17, 2017a.
- QUINTÃO, R. T.; BRITO, E. P. Z.; BELK, R. W. The taste transformation ritual in the specialty coffee market. **Revista de Administração de Empresas**, São Paulo, v. 57, n. 5, p. 483-494, 2017b.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: Foundation for Statistical Computing, 2021. 409 p.
- ROSSEEL Y. Lavaan: an R package for structural equation modeling. **Journal of Statistical Software**, Innsbruck, v. 48, n. 2, p. 1-36, 2012.

SANTOS, P. M. dos; CIRILLO, M. Â.; GUIMARÃES, E. R. Specialty coffee in Brazil: transition among consumers' constructs using structural equation modeling. **British Food Journal**, Bradford, v. 123, n. 5, p. 1-15, 2021.

SOCIEDADE NACIONAL DE AGRICULTURA. **Consumo brasileiro de café cresce 3,5%, revela pesquisa da ABIC**. 2018. Disponível em: <http://www.sna.agr.br/consumo-brasileiro-de-cafe-cresce-35-revela-pesquisa-da-abic/>. Acesso em: 5 abr. 2022.

SPECIALTY COFFEE ASSOCIATION OF AMERICA. **Protocolo para análise sensorial de café: metodologia SCAA**. 2008. Disponível em: http://coffeetraveler.net/wp-content/files/901-SCAA_CuppingProtocols_TSC_DocV_RevDec08_Portuguese.pdf. Acesso em: 2 abr. 2022.

WELDEMICHAEL, G.; TEFERI, D. The impact of climate change on coffee (*Coffea Arabica* L.) production and genetic resources. **International Journal of Research Studies in Agricultural Sciences**, Piplani, v. 5, p. 26-34, 2020.