

UNIVERSIDADE FEDERAL DE ALFENAS
UNIFAL-MG

MARCELA COSTA ROCHA

**TESTES BASEADOS EM CORRELAÇÃO CANÔNICA PARA AVALIAR A
CONSONÂNCIA DE PAINÉIS SENSORIAIS**

Alfenas/MG

2015

MARCELA COSTA ROCHA

**TESTES BASEADOS EM CORRELAÇÃO CANÔNICA PARA AVALIAR A
CONSONÂNCIA DE PAINÉIS SENSORIAIS**

Dissertação apresentada à Universidade Federal de Alfenas, como parte dos requisitos para obtenção do título de Mestre em Estatística Aplicada e Biometria.

Orientador: Prof. Dr. Eric Batista Ferreira

Coorientador: Prof. Dr. Daniel Furtado Ferreira

Alfenas/MG

2015

Rocha, Marcela Costa

Testes baseados em correlação canônica para avaliar a consonância de painéis sensoriais / Marcela Costa Rocha. - Alfenas, 2015.

71 f. -

Orientador: Eric Batista Ferreira.

Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2015.

Bibliografia.

1. Avaliação sensorial. 2. Análise multivariada. 3. Testes de hipóteses estatísticas. I. Ferreira, Eric Batista. II. Título.

CDD: 519.535



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas / UNIFAL-MG
Programa de Pós-graduação em Estatística Aplicada e Biometria

Rua Gabriel Monteiro da Silva, 700. Alfenas - MG CEP 37130-000
Fone: (35) 3299-1392 (Secretaria) (35) 3299-1121 (Coordenação)
<https://www.unifal-mg.edu.br/ppgeab/>



MARCELA COSTA ROCHA

“TESTES BASEADOS EM CORRELAÇÃO CANÔNICA PARA AVALIAR A
CONSONÂNCIA DE PAINÉIS SENSORIAIS”.

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

Aprovado em: 20 de março de 2015.

Prof. Dr. Eric Batista Ferreira
Instituição: UNIFAL-MG

Assinatura: Eric Batista Ferreira

Prof. Dr. Daniel Furtado Ferreira
Instituição: UFLA

Assinatura: Daniel Furtado Ferreira

Prof.^a Dr.^a Adriana Dias
Instituição: UNIFAL-MG

Assinatura: Adriana Dias

Prof. Dr. Denismar Alves Nogueira
Instituição: UNIFAL-MG

Assinatura: Denismar Alves Nogueira

Ao meu pai, que foi meu maior incentivador enquanto estava presente e minha maior inspiração após partir.

AGRADECIMENTOS

A Deus, pela força concedida.

À minha mãe, pelo amor e compreensão em todos os momentos.

Ao meu irmão, Luís Eduardo, pela amizade incondicional e presença constante.

Ao meu namorado Marcos, pelo seu amor e companheirismo em todos os momentos.

Ao meu orientador, professor Dr. Eric Batista Ferreira, por seus ensinamentos, dedicação e, principalmente, por não me deixar esquecer de que não seria uma tarefa fácil, mas que eu conseguiria.

Ao meu coorientador, professor Dr. Daniel Furtado Ferreira, pelas contribuições fundamentais para o desenvolvimento deste trabalho.

Aos membros das bancas do exame de qualificação e da banca de defesa de mestrado, professores Dr. Denismar Alves Nogueira, Dra. Adriana Dias e professora Dra. Flávia Flávia Della Lucia, pelas orientações e pelo incentivo.

À Universidade Federal de Alfenas e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, juntamente com seus docentes, pela oportunidade concedida para a realização do mestrado.

Aos professores do Instituto de Ciências Exatas, pela atenção e contribuições para a minha formação.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, em especial ao Câmpus Machado, pelo apoio.

A todos que, de alguma forma, contribuíram para a realização desse trabalho.

RESUMO

Um painel sensorial é considerado consonante quando todos os julgadores pontuam um produto de maneira semelhante. Dessa forma, a consonância entre os julgadores que compõem um painel sensorial é uma das características necessárias para a confiabilidade da análise sensorial e pode ser mensurada pelo seu nível de unidimensionalidade. Na literatura existem testes para a unidimensionalidade de um painel sensorial, mas sua aplicação é restrita à análise para um atributo sensorial por vez. Diante disso, o objetivo do presente trabalho foi a generalização de cinco testes propostos para avaliar a consonância de painéis sensoriais, a saber: teste de Fujikoshi, teste Monte Carlo para unidimensionalidade, dois testes *bootstrap* paramétricos e teste sobre autovalores Monte Carlo. Tal generalização consistiu em utilizar a matriz de correlação canônica, de forma que fosse possível inferir sobre a consonância de painéis levando em consideração todos os atributos simultaneamente. A avaliação do desempenho das generalizações propostas, em termos de poder e taxa de erro tipo I, foi feita via simulação Monte Carlo. A generalização do teste sobre autovalores Monte Carlo foi aplicada para avaliar a consonância do painel de um experimento realizado por Pereira (2005). O teste sobre autovalores Monte Carlo apresentou desempenho igual ou superior aos demais testes e, por esse motivo, é recomendado para a análise da unidimensionalidade multivariada de painéis sensoriais.

Palavras-chave: Unidimensionalidade. Sensometria. Análise Multivariada.

ABSTRACT

A sensory panel is considered to be consonant when all referees point out a product in a similar way. Thus, the consonance between the referees that compose a sensory panel is one of the necessary characteristics for the reliability of sensory analysis, and can be measured by its level of unidimensionality. In literature there are tests for unidimensionality of a sensory panel, but its application is restricted to the analysis for one sensory attribute at a time. Thus, the objective of this study was the generalization of five tests proposed to evaluate the line of sensory panels, namely: the Fujikoshi test, the Monte Carlo test for unidimensionality, two parametric *bootstrap* tests and Monte Carlo test of eigenvalues. Such generalization is to use the canonical correlation matrix, so that it was possible to infer about the consonance of panels considering all attributes simultaneously. The performance evaluation of the proposed generalizations, in terms of power and type I error rate, was done through Monte Carlo simulation. The generalization of Monte Carlo test of eigenvalues test was applied to assess the panel's consonance from an experiment conducted by Pereira (2005). The Monte Carlo test of eigenvalues performed equally to or higher than the other tests and, therefore, it is recommended for the analysis of multivariate unidimensionality of sensory panels.

Keywords: Unidimensionality. Sensometrics. Multivariate Analysis

SUMÁRIO

1	INTRODUÇÃO	8
2	REVISÃO DE LITERATURA	10
2.1	Sensometria	10
2.2	Testes para a retenção de componentes principais	15
2.2.1	Teste de Fujikoshi	16
2.2.2	Testes <i>bootstrap</i> paramétricos	17
2.2.3	Teste Monte Carlo para unidimensionalidade	20
2.2.4	Teste sobre autovalores Monte Carlo	21
2.3	Análise de componentes principais	23
2.4	Análise de correlação canônica	24
2.5	Taxas de erro tipo I e poder de um teste	26
2.6	Simulação Monte Carlo	27
3	METODOLOGIA	28
3.1	Experimento de simulação	28
3.2	Experimento com dados reais	31
4	RESULTADOS E DISCUSSÃO	33
4.1	Generalização dos testes de unidimensionalidade	33
4.2	Validação e comparação dos testes	36
4.2.1	Discussão dos resultados de simulação	44
4.3	Experimento com dados reais	45
5	CONCLUSÕES	52
	REFERÊNCIAS	53
	GLOSSÁRIO	56
	APÊNDICES	57

1 INTRODUÇÃO

A busca pela melhoria da qualidade dos produtos e sua adequação ao mercado consumidor é constante e, com essa finalidade, a análise sensorial tem sido cada vez mais utilizada em vários segmentos da indústria. A análise sensorial é composta por um conjunto de técnicas para analisar e interpretar os atributos sensoriais, obtidos com base nas respostas aos estímulos percebidos pelos órgãos sensoriais humanos. Para esse fim, se utiliza da sensometria para quantificar tais respostas.

A análise sensorial é um método científico e, dessa forma, as medições devem ser precisas, ou seja, as respostas devem ser baseadas em algum critério ou padrão preestabelecido, sem erro sistemático ou ideia preconcebida (DUTCOSKY, 2011; PIGGOTT, 1995; PINTO, 2014). Por esse motivo, a avaliação sensorial é realizada por um painel sensorial, composto por um grupo de pessoas treinadas para avaliar os atributos em questão, denominados julgadores ou avaliadores.

A confiabilidade dos resultados da análise sensorial está diretamente ligada ao desempenho do painel, ou seja; quanto melhor o desempenho do painel, mais confiáveis os resultados da avaliação sensorial. Esse desempenho, em geral, está relacionado à capacidade dos julgadores em identificar pequenas diferenças entre produtos, à repetibilidade de suas avaliações para um mesmo produto e à consonância do painel (AMORIM et al., 2010).

A consonância do painel pode ser entendida como a concordância entre os julgadores, dessa forma, é refletida pelo grau de treinamento dos mesmos. Em um painel bem treinado, os julgadores pontuam o mesmo produto de maneira semelhante. Dessa forma, o desenvolvimento de técnicas que avaliem o nível de concordância entre os julgadores é fundamental para o aperfeiçoamento dos resultados da avaliação sensorial.

O objetivo do presente trabalho foi generalizar cinco testes existentes na literatura, a saber: teste sobre autovalores Monte Carlo (AV_{mc}) proposto por Fernandes (2012), teste Monte Carlo para unidimensionalidade (TMCU) proposto por Amorim et al. (2010), dois testes *bootstrap* paramétricos propostos por Gebert (2010) e teste de Fujikoshi proposto por Ferreira (2011). Além disso, objetivou-se avaliar o desempenho dos testes generalizados, em termos de poder e taxa de erro tipo I, via simulação Monte Carlo, e recomendar o teste com melhor desempenho.

O teste sobre autovalores Monte Carlo e o teste Monte Carlo para unidimensionalidade

foram propostos com a finalidade de avaliar a unidimensionalidade de painéis sensoriais para um determinado atributo. Já, os testes *bootstrap* paramétricos e teste de Fujikoshi são, essencialmente, testes para a retenção de componentes principais e foram inseridos no contexto sensorial por Fernandes (2012), com a mesma finalidade.

Levando-se em consideração que a utilização dos testes citados anteriormente é restrita à avaliação da concordância entre os julgadores para um determinado atributo, a generalização proposta consiste em utilizar a matriz de correlação canônica para a realização dos mesmos, de tal forma que seja possível inferir sobre a consonância do painel de uma maneira global, ou seja, considerando todos os atributos simultaneamente. Esse texto foi organizado em 5 seções. Na segunda seção é apresentada a revisão de literatura, que busca esclarecer o contexto que deu origem ao presente estudo, bem como a teoria que o fundamenta. Além disso, são descritos os critérios utilizados para avaliar o desempenho dos testes e os fundamentos da simulação Monte Carlo.

A terceira seção descreve o experimento com dados reais e a metodologia empregada para a validação e comparação das generalizações propostas sob diversas situações, resultantes da combinação entre o número de atributos sensoriais, produtos, avaliadores e diferentes graus de treinamento.

Na quarta seção são descritos os resultados metodológicos, ou seja, a proposição da generalização dos testes, apresentados os resultados da simulação Monte Carlo e o estudo com dados reais e a discussão dos resultados.

Finalmente, as conclusões são exibidas na quinta seção.

2 REVISÃO DE LITERATURA

A revisão de literatura buscou esclarecer o contexto que deu origem ao presente estudo e a teoria que o fundamenta. Para tal, são apresentados alguns trabalhos que propõem e/ou comparam métodos de avaliação de painéis e evidenciados os testes para a retenção de componentes principais que foram generalizados no estudo.

Além disso, foram detalhados a análise de componentes principais e a análise de correlação canônica, necessários à generalização proposta.

Finalmente, são descritos os critérios adotados para a avaliação dos testes (poder e taxa de erro tipo I), bem como os fundamentos da simulação Monte Carlo.

2.1 Sensometria

A análise sensorial é uma ciência que tem sido largamente utilizada na indústria, tanto no desenvolvimento de produtos quanto no controle de qualidade. Sua finalidade é interpretar, medir e compreender as respostas humanas a propriedades de um determinado produto percebidas pelos sentidos: paladar, olfato, visão, audição e tato (HUMMER, 1998; MARTENS, 1999).

De acordo com Minim (2006), a análise sensorial consiste em três etapas básicas, que seguem:

- i) Identificação dos atributos sensoriais;
- ii) escolha e aplicação do método para quantificar tais atributos; e
- iii) a seleção dos métodos estatísticos para análise e interpretação dos dados.

Os métodos estatísticos supracitados compõem a Sensometria, que é uma área da estatística que objetiva mensurar os estímulos percebidos pelos sentidos humanos básicos e, dessa forma, lidar com o caráter subjetivo das sensações (DAEMS; DELVAUX, 1997; SINESIO; MONETA; SABA, 1992). Além disso, a Sensometria fornece métodos que permitem inferir sobre o treinamento de painéis sensoriais, pois a qualidade desse treinamento é um fator imprescindível para a confiabilidade dos resultados da análise sensorial (CASTURA; FINDLAY; LESSCHAEVE, 2005).

Segundo Martins (2002), o ser humano pode ser considerado o instrumento de medida na análise sensorial e, dessa forma, deve se preparado para fornecer medidas tão precisas quanto possível. Dessa forma, o treinamento do painel objetiva familiarizar o julgador com os procedimentos dos testes, aperfeiçoar sua habilidade de identificar os atributos sensoriais, melhorar sua sensibilidade e memória e padronizar as medidas sensoriais com a finalidade de que os resultados possam ser reproduzidos.

A escolha dos julgadores e os procedimentos de treinamento podem variar de acordo com os objetivos da análise sensorial, mas, para que os resultados da análise sejam confiáveis, o painel deve apresentar discriminação entre os produtos, repetibilidade das avaliações e consonância entre os julgadores (FERREIRA; OLIVEIRA, 2007).

Diante da importância do treinamento dos julgadores para a credibilidade dos resultados da análise sensorial, é crescente o número de estudos que buscam quantificar o nível desse treinamento.

Isso posto, a seguir são evidenciados alguns estudos que propõem e/ou comparam métodos de avaliação da consonância de painéis sensoriais. Vale ressaltar que todos os estudos expostos nessa seção avaliam a concordância dos julgadores para um atributo fixo e, mesmo que não tenham relação direta com o presente trabalho, sua revisão tem por finalidade expor algumas investigações realizadas sobre o tema.

Dijksterhuis (1995) propôs um método baseado em análise de componentes principais (ACP) para avaliar a consonância de um painel, em relação a um determinado atributo. Segundo o autor, se os julgadores pontuarem esse atributo da mesma maneira, a maior parte da variância será explicada pelo primeiro componente principal e o painel será considerado unidimensional.

Considerando-se um experimento com n produtos, p julgadores e q atributos, os dados podem ser dispostos em uma hipermatriz \mathbf{X} de dimensão $n \times q \times p$, com elementos x_{ijk} , com $i = 1, 2, \dots, n$, $j = 1, 2, \dots, q$ e $k = 1, 2, \dots, p$.

A hipermatriz \mathbf{X} é formada por matrizes \mathbf{X}_j de ordem $n \times p$ e, tomando uma matriz \mathbf{X}_j , com $j \in \{1, 2, \dots, q\}$, ou seja, fixando-se um atributo, cada uma de suas p colunas é referente às notas de um determinado julgador para o atributo j , como é apresentado na Figura 1.

Dijksterhuis (1995) afirma que, quando a aplicação de ACP na matriz \mathbf{X}_j resulta em um primeiro autovalor grande ou alto percentual de explicação da variância, o painel deve ser considerado unidimensional, ou seja, existe alto grau de concordância entre os julgadores para o atributo fixado.

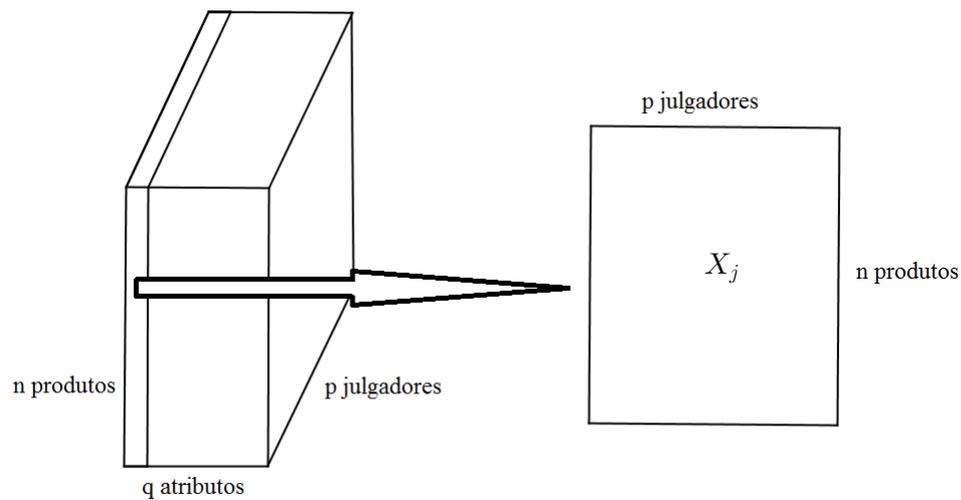


Figura 1 – À esquerda, a hipermatriz \mathbf{X} e, à direita, uma fatia da hipermatriz, para o atributo j .
Fonte: Adaptado de Dijksterhuis (1995).

Vale ressaltar que um resultado completamente unidimensional será raro e, dessa forma, a comparação entre a primeira dimensão com as demais, para um atributo j , pode ser feita por:

$$C_j = \frac{\lambda_1^2}{\sum_2^p \lambda_i^2}$$

em que λ_i^2 é o i -ésimo autovalor da matriz $\mathbf{X}_j^\top \mathbf{X}_j$.

Bi (2003) revisita o índice proposto por Finn (1970) para avaliar a concordância de um painel para um determinado atributo, que é dado por:

$$r_a = 1 - \frac{\hat{\sigma}_e^2}{\sigma_u^2}$$

em que $\hat{\sigma}_e^2$ é a variância do erro observada na amostra e σ_u^2 é a variância esperada para uma distribuição uniforme. Segundo o autor, σ_u^2 varia de acordo com a escala adotada e é dado por $\sigma_u^2 = \frac{(b-a)^2}{12}$, em que a e b são os extremos da escala, quando a escala é contínua, ou $\sigma_u^2 = \frac{(b-a+1)^2 - 1}{12}$, quando a escala é discreta.

Dessa forma, o coeficiente de concordância (r_a), é a proporção da variância que não se refere ao erro e varia de zero (situação ideal) a um.

Finalmente, considerando-se n produtos e p julgadores, o autor propõe um intervalo de confiança para o índice de concordância, como segue:

$$P \left[1 - \frac{\hat{\sigma}_e^2 \nu}{\sigma_u^2 \chi_{\alpha/2}^2} \leq r_a \leq 1 - \frac{\hat{\sigma}_e^2 \nu}{\sigma_u^2 \chi_{1-\alpha/2}^2} \right] = 1 - \alpha$$

em que χ_{α}^2 é o quantil superior α de uma distribuição qui-quadrado com $\nu = n(p - 1)$ graus de liberdade.

Derndorfer et al. (2005) propuseram um método para avaliação de painéis e o implementaram no *software* R Core Team. Os autores investigaram as seguintes características: discriminação entre os produtos, repetibilidade das avaliações, utilização da escala e concordância entre os julgadores. Para avaliar a concordância entre os julgadores, fixando-se um atributo, inicialmente foram determinados intervalos com 95% de confiança para as médias de cada um dos produtos. Os erros-padrão para a construção dos intervalos de confiança foram calculados a partir dos resíduos das análises de variância dos produtos, considerando as interações entre *atributo* \times *produto* e *julgador* \times *produto* incluídas no modelo, divididos pelo número de repetições. A seguir, foram calculadas as médias dadas por cada julgador para o produto, em todas as repetições, e a porcentagem de produtos que se encontravam dentro do intervalo de confiança construído foi considerada o grau de concordância do julgador com o painel.

Latreille et al. (2006) propuseram um método baseado em modelos lineares mistos, para avaliar vários critérios de confiabilidade de um painel. O método desenvolvido considera tanto o desempenho global do painel quanto o desempenho individual dos julgadores e, para isso, foram testadas características como discriminação, repetibilidade, severidade e concordância com o painel. O modelo proposto pelos autores foi aplicado em um estudo realizado com cosméticos e levava em consideração os efeitos julgador, sessão e produto, bem como as possíveis interações entre esses efeitos, como descrito a seguir:

$$Y_{ijs} = \mu + \alpha_i + \beta_j + \gamma_s + a_{ij} + b_{is} + c_{js} + \varepsilon_{ijs}$$

em que Y_{ijs} é o valor dado pelo julgador j para o produto i na sessão s , μ é o intercepto, α_i é o efeito fixo do produto, β_j é o efeito aleatório de julgador, γ_s é o efeito aleatório de sessão, a_{ij} , b_{is} e c_{js} são as interações *produto* \times *julgador*, *produto* \times *sessão* e *julgador* \times *sessão*,

respectivamente, ε_{ijs} é o erro aleatório e $\varepsilon_{ijs} \sim N(\mu, \sigma^2)$.

Para testar a concordância de um julgador com o painel foram utilizados contrastes da interação *produto* \times *julgador*. De acordo com os autores, o método proposto se mostrou adequado para uma pequena quantidade de produtos e sessões e sugeriram a utilização de modelos simples ou de métodos de análise multivariada em outros casos.

Pinto et al. (2014) propuseram a análise da concordância de um painel utilizando o coeficiente α de Cronbach (CA) e a compararam com o método baseado em ACP proposto por Dijksterhuis (1995) para avaliar a consonância de um painel.

Os autores afirmam que o coeficiente α de Cronbach (1951), proposto para medir a consistência de respostas em questionários, pode ser utilizado para mensurar a similaridade entre as avaliações dos julgadores, indicando aqueles que discordam do restante do painel.

Considere n produtos e s sessões. Então tem-se $n \times s$ notas para cada julgador, por atributo. A covariância entre as notas de dois julgadores reflete a concordância entre os mesmos para o atributo em análise.

Isso posto, o coeficiente α de Cronbach (α_c) pode ser calculado pela expressão:

$$CA = \frac{p \bar{c}}{\nu + (p - 1)\bar{c}} \quad (2.1)$$

em que p é o número de julgadores, \bar{c} é a covariância média entre todos os possíveis pares de julgadores e ν é a variância média de todos os julgadores.

Para avaliar o desempenho geral do painel, os dados devem ser dispostos em uma hipermatriz \mathbf{X} ($n \times q \times s$) em que n é o número de produtos, q é o número de atributos e s é o número de sessões. A partir da hipermatriz \mathbf{X} , calcula-se a média das sessões e, em seguida, o coeficiente α de Cronbach apresentado na expressão (2.1) considerando-se todos os julgadores, obtendo-se o coeficiente α de Cronbach geral para o painel (CA_G).

Para avaliar o desempenho do j -ésimo julgador (CA_j), calcula-se a expressão (2.1) da matriz \mathbf{X} , excluindo-se a j -ésima coluna, ou seja, excluindo-se as notas atribuídas por esse julgador. A partir daí, deve-se calcular o coeficiente α gradiente ($CA_{\Delta j}$), obtido por meio da expressão:

$$CA_{\Delta j} = \frac{(CA_G - CA_j)}{CA_G} \times 100 \quad (2.2)$$

A expressão (2.2) é usada para calcular o crescimento (ou decréscimo) do coeficiente α geral resultante da exclusão do j -ésimo julgador da matriz \mathbf{X} . Um valor positivo para $CA_{\Delta j}$ indica um julgador consonante com o restante do painel, pois a exclusão das suas notas ocasiona uma queda no valor do coeficiente α que indica a concordância do painel.

Os autores aplicaram o índice α de Cronbach em um conjunto de dados no qual 8 produtos foram pontuados por 9 julgadores, em relação a 26 atributos, comparando-o ao método proposto por Dijksterhuis (1995). De acordo com os autores, o método proposto por eles mostrou-se mais eficiente, por permitir classificar os julgadores em relação à sua concordância com o restante do painel.

2.2 Testes para a retenção de componentes principais

Considerando uma amostra aleatória p -variada, o objetivo da retenção de componentes principais é resumir a informação contida nas variáveis originais em k componentes principais, de tal forma que se tenha $k < p$.

A quantidade de componentes principais retidos (k), em geral, depende da proporção da variação da variância da amostra original que deve ser explicada pelos k primeiros componentes principais. Diante disso, o valor de k pode ser determinado por um teste de hipóteses de igualdade dos últimos autovalores.

Para a realização do teste deve-se considerar a hipótese nula $H_0 : \lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_p = 0$. Se a hipótese testada for rejeitada, deve-se retirar um componente principal da hipótese nula (λ_{k+1}) e realizar o teste novamente até que seja obtido o menor valor de k para o qual H_0 não seja mais rejeitada.

A seguir são apresentados os testes para a retenção de componentes principais generalizados no presente trabalho.

Vale notar que os testes são apresentados da mesma maneira que foram propostos. Assim, com exceção do teste Monte Carlo para unidimensionalidade e teste sobre autovalores Monte Carlo, os testes inferem sobre a explicação da variação pelos k primeiros componen-

tes principais (ρ_k^2). No entanto, para as generalizações propostas, é considerada a proporção da variação explicada pelo primeiro componente principal, ou seja, o caso particular em que $k = 1$.

2.2.1 Teste de Fujikoshi

Baseado no trabalho de Fujikoshi (1980), Ferreira (2011) propôs um teste para avaliar a proporção da variância explicada pelos k primeiros componentes principais amostrais, que pode ser calculada pela expressão:

$$R_k^2 = \frac{\sum_{i=1}^k \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i} \quad (2.3)$$

em que $R_k^2 \in [0, 1]$ e $\hat{\lambda}_i$ são os autovalores do estimador da matriz de covariâncias (Σ), que pode ser obtido pela expressão

$$\mathbf{S} = \frac{1}{n-1} \left[\sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top - \frac{\left(\sum_{i=1}^n \mathbf{Y}_i \right) \left(\sum_{i=1}^n \mathbf{Y}_i \right)^\top}{n} \right]. \quad (2.4)$$

De acordo com o autor, a distribuição de R_k^2 pode ser conhecida, utilizando o resultado encontrado por Fujikoshi (1980) que afirma que

$$\sqrt{n-1}(R_k^2 - \rho_k^2) \sim N(0, \omega^2)$$

em que ρ_k^2 é a proporção da variância explicada pelos k primeiros componentes principais populacionais e ω^2 é estimado pela expressão

$$\hat{\omega}^2 = \sqrt{\frac{2tr(\mathbf{S}^2)}{[tr(\mathbf{S})]^2} [(\rho_k^2)^2 - 2\hat{\beta}\rho_k^2 + \hat{\beta}]}$$

e

$$\hat{\beta} = \frac{\sum_{i=1}^k \hat{\lambda}_i^2}{\sum_{i=1}^p \hat{\lambda}_i^2}.$$

Fundamentado nesse resultado, o teste proposto por Ferreira (2011) para ρ_k^2 tem o seguinte par de hipóteses:

$$\begin{cases} H_0 : \rho_k^2 \geq \rho_{k0}^2 \\ H_1 : \rho_k^2 < \rho_{k0}^2. \end{cases}$$

A estatística do teste é dada pela expressão:

$$z_c = \frac{R_k^2 - \rho_{k0}^2}{\frac{\omega_0}{\sqrt{n-1}}}$$

em que $z_c \sim N(0, 1)$,

$$\omega_0 = \sqrt{\frac{2tr(\mathbf{S}^2)}{[tr(\mathbf{S})]^2} [(\rho_{k0}^2)^2 - 2\hat{\beta}\rho_{k0}^2 + \hat{\beta}]}$$

2.2.2 Testes *bootstrap* paramétricos

A construção dos testes *bootstrap* paramétricos propostos por Gebert (2010) dependem de uma distribuição normal p -variada com as seguintes características:

- i) Média igual ao vetor nulo, ou seja, $\boldsymbol{\mu}_b = [0 \cdots 0]$.
- ii) A matriz de covariâncias ($\boldsymbol{\Sigma}_b$) construída de tal forma que atenda à condição $\rho_k^2 = \rho_{k0}^2$.

Considerando uma amostra aleatória em \mathbb{R}^p e sua matriz de covariâncias $\boldsymbol{\Sigma}$, a proporção da variação explicada pelos k primeiros componentes principais (ρ_k^2) tem estimador (R_k^2) obtido pela expressão (2.3).

A matriz Σ_b pode ser construída a partir da seguinte decomposição espectral:

$$\Sigma_b = \hat{\mathbf{P}}\Lambda_b\hat{\mathbf{P}}^\top$$

em que $\hat{\mathbf{P}}^\top$ é a matriz dos autovetores de \mathbf{S} , que é a matriz de covariância amostral e pode ser obtida pela expressão (2.4), e Λ_b é uma matriz diagonal definida para que a proporção da variação explicada pelos k primeiros componentes principais seja igual a ρ_{k0}^2 (GEBERT, 2010; GEBERT; FERREIRA, 2013).

Os autovetores de \mathbf{S} são conhecidos e a construção da matriz Λ_b é explicada a seguir.

Considerando $\hat{V} = \sum_{i=1}^p \hat{\lambda}_i$ a variação total contida em \mathbf{S} , a proporção da variação que deve ser explicada pelos k primeiros componentes principais é representada por $\rho_{k0}^2 \hat{V}$.

Isso posto, como $\sum_{i=1}^k \hat{\lambda}_i < \rho_{k0}^2 \hat{V}$, a matriz Λ_b pode ser construída retirando a quantidade $\hat{\Delta} = \rho_{k0}^2 \hat{V} - \sum_{i=1}^k \hat{\lambda}_i$, de maneira proporcional, dos últimos $p - k$ autovalores de \mathbf{S} , ou seja, o i -ésimo elemento da diagonal de Λ_b pode ser calculado por

$$\lambda_{ib} = \hat{\lambda}_i - \frac{\hat{\lambda}_i}{\sum_{i=k+1}^p \hat{\lambda}_i} \hat{\Delta}$$

para $i = k + 1, k + 2, \dots, p$.

Os k primeiros elementos da diagonal de Λ_b são obtidos adicionando $\hat{\Delta}$, também de maneira proporcional, aos k primeiros autovalores de \mathbf{S} , ou seja

$$\lambda_{ib} = \hat{\lambda}_i + \frac{\hat{\lambda}_i}{\sum_{i=1}^k \hat{\lambda}_i} \hat{\Delta}$$

para $i = 1, 2, \dots, k$.

Definidos os parâmetros μ_b e Σ_b são geradas B amostras aleatórias de tamanho n de uma distribuição normal p -variada com esses parâmetros.

Assim, considerando o seguinte par de hipóteses

$$\begin{cases} H_0 : \rho_k^2 \geq \rho_{k0}^2 \\ H_1 : \rho_k^2 < \rho_{k0}^2. \end{cases} \quad (2.5)$$

cada uma das amostras *bootstrap* geradas estará sob a hipótese nula. Assim, considerando a ℓ -ésima amostragem *bootstrap* $\mathbf{Y}_{1\ell^*}, \mathbf{Y}_{2\ell^*}, \dots, \mathbf{Y}_{n\ell^*}$ com $\ell = 1, 2, \dots, B$ foram definidas as quantidades que tornam possível a construção de dois testes.

O primeiro teste, chamado Teste *bootstrap* paramétrico baseado na proporção da variação explicada pelos k primeiros componentes principais (baseado em $R_{k\ell}^2$), quantidade definida por

$$R_{k\ell}^2 = \frac{\sum_{i=1}^k \lambda_i^*}{\sum_{i=1}^p \lambda_i^*} \quad (2.6)$$

em que λ_i^* são os autovalores da matriz de covariância \mathbf{S}_ℓ^* , calculada de acordo com a expressão (2.4), a partir da ℓ -ésima amostra *bootstrap*.

Dessa forma o valor- p desse teste é definido por

$$\text{valor-}p = \frac{\sum_{\ell=1}^B I(R_{k\ell}^2 \leq R_k^2)}{B},$$

em que I é a função indicadora que retorna 1 se $R_{k\ell}^2 \leq R_k^2$ e 0, caso contrário. Dessa forma, o valor- p é a proporção dos casos em que a explicação dos k primeiros autovalores das amostras *bootstrap* foram inferiores ao valor obtido na amostra original.

O segundo teste, nomeado Teste *bootstrap* paramétrico baseado em $z_{c\ell}$, tem estatística de teste definida por

$$z_{c\ell} = \frac{R_{k\ell}^2 - \rho_{k0}^2}{\frac{\omega_{0\ell}}{\sqrt{n}}}$$

em que $z_{c\ell} \sim N(0,1)$, $R_{k\ell}^2$ é calculado conforme expressão (2.6) e $\omega_{0\ell}$ é calculado por

$$\omega_{0\ell} = \sqrt{\frac{2tr(\mathbf{S}_\ell^{*2})}{[tr(\mathbf{S}_\ell^*)]^2} [(\rho_{k0}^2)^2 - 2\hat{\beta}_\ell \rho_{k0}^2 + \hat{\beta}_\ell]}$$

e

$$\hat{\beta}_\ell = \frac{\sum_{i=1}^k \hat{\lambda}_i^{*2}}{\sum_{i=1}^p \hat{\lambda}_i^{*2}}.$$

O valor de z_c da amostra original é calculado por meio da expressão

$$z_c = \frac{R_k^2 - \rho_{k0}^2}{\frac{\omega_0}{\sqrt{n}}}$$

Assim, o valor- p do teste é dado pela proporção de casos em que os valores calculados para $z_{c\ell}$ são inferiores ao valor de z_c na amostra original, ou seja,

$$\text{valor-}p = \frac{\sum_{\ell=1}^B I(z_{c\ell} \leq z_c)}{B}$$

em que I é a função indicadora que retorna 1 se $z_{c\ell} \leq z_c$ e 0, caso contrário.

2.2.3 Teste Monte Carlo para unidimensionalidade

O teste Monte Carlo para unidimensionalidade (TMCU) foi proposto por Amorim et al. (2010) para decidir a respeito da unidimensionalidade de um painel.

Admitindo que $\Sigma = \sigma^2 \rho$; para o caso particular em que $\sigma^2 = 1$ tem-se a igualdade entre a matriz de covariância (Σ) e a matriz de correlação (ρ). A matriz ρ tem estrutura equicorrelacionada, ou seja, todas as correlações entre pares de variáveis diferentes são iguais a ρ .

O primeiro autovalor de Σ pode ser obtido pela expressão

$$\lambda_1 = \sigma^2[1 + (p - 1)\rho]$$

em que σ^2 é a variância das notas dadas pelos p julgadores, p é o total de julgadores que compõem o painel e ρ é a correlação entre os p julgadores (MORRISON, 1990 apud AMORIM et al., 2010).

Isso posto, a autora afirma que a porcentagem da explicação fornecida pelo primeiro

componente principal ($\rho_{1\rho}^2$) pode ser estimada por

$$R_{1\rho}^2 = \frac{1 + (p - 1)\rho}{p} \times 100\%. \quad (2.7)$$

Para a proposição do teste, a autora considerou o seguinte par de hipóteses:

$$\begin{cases} H_0 : \rho_{1\rho}^2 \geq \rho_{1\rho 0}^2 \\ H_1 : \rho_{1\rho}^2 < \rho_{1\rho 0}^2. \end{cases}$$

em que $\rho_{1\rho}^2$ é estimado pela expressão (2.7) e $\rho_{1\rho 0}^2$ é a proporção da variação explicada pelo primeiro componente principal estabelecida previamente.

Definido o par de hipóteses, foi gerada uma distribuição empírica sob H_0 para a porcentagem da variação explicada pelo primeiro autovalor (λ_1), utilizando simulação Monte Carlo. Foram geradas 1000 amostras, sob H_0 , de uma variável aleatória $\mathbf{X}_{p \times 1}$ tal que $\mathbf{X}_{p \times 1} \sim N(\mathbf{0}, \Sigma)$ e

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \ddots & \rho \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{bmatrix} = \boldsymbol{\rho}, \text{ com } \sigma^2 = 1,$$

em que ρ é a correlação entre os julgadores.

A porcentagem da variação explicada pelo primeiro componente principal deve ser calculada para cada uma das 1000 amostras Monte Carlo e para a amostra de interesse, conforme expressão (2.3), para $k = 1$. Além disso, deve-se determinar o valor crítico da distribuição empírica ($R_{1\alpha}^2$), em que α é nível nominal de significância adotado para o teste.

A regra de decisão do teste é que H_0 deverá ser rejeitada sempre que $R_{1\rho}^2 < R_{1\alpha}^2$.

2.2.4 Teste sobre autovalores Monte Carlo

Fernandes (2012) propôs uma versão de teste para autovalores assintótico (AVa), denominado teste sobre autovalores Monte Carlo (AVmc). Segundo a autora, o objetivo do teste é verificar a unidimensionalidade de painéis sensoriais e sua proposição justifica-se pela necessi-

dade de se obter um bom desempenho, em termos de poder e taxa de erro tipo I, para amostras pequenas.

Para tal, é estabelecido o seguinte par de hipóteses:

$$\begin{cases} H_0 : \lambda_2 = \lambda_3 = \dots = \lambda_p = 0 \\ H_1 : \text{ao menos um } \lambda_j \text{ difere de } 0, j > 1. \end{cases}$$

Sob H_0 , toda variação deve estar contida no primeiro componente principal, ou seja, no primeiro autovalor. Assim, a soma dos p autovalores deve ser igual ao primeiro autovalor e, dessa forma, a hipótese nula pode ser reescrita como $H_0 : \lambda_1 = \sum_{i=1}^p \lambda_i$ e, dessa forma, a estatística de teste é dada por:

$$z_c = \frac{\hat{\lambda}_1 - \sum_{i=1}^p \lambda_i}{\sum_{i=1}^p \lambda_i \sqrt{\frac{2}{(n-1)}}}, \quad (2.8)$$

em que $z_c \sim N(0,1)$.

Levando-se em consideração $\sum_{i=1}^p \lambda_i = \lambda_1$ e que λ_1 pode ser estimado por $\sum_{i=1}^p \hat{\lambda}_i$, a expressão 2.8 pode ser reescrita da seguinte maneira:

$$z_c = \frac{-\sum_{i=2}^p \hat{\lambda}_i}{\hat{\lambda}_1 \sqrt{\frac{2}{(n-1)}}} \quad (2.9)$$

em que $z_c \sim N(0,1)$.

Para a realização do teste, deve-se fixar um valor de significância (α). Além disso, são geradas B amostras de uma variável aleatória $\mathbf{X}_{p \times 1}$, que são vetores com as notas dos p julgadores para n produtos, tal que $\mathbf{X}_{p \times 1} \sim N(\mathbf{0}, \Sigma)$ e tenha estrutura equicorrelacionada ρ_{10}^2 .

Tanto para os valores observados quanto para cada uma das B amostras simuladas deve-se realizar a análise de componentes principais, determinar seu primeiro autovalor ($\hat{\lambda}_1$ e $\hat{\lambda}_{1l}$, respectivamente) e calcular a estatística de teste apresentada em (2.9) - z_c e z_{cl} , respectivamente.

A partir das estatísticas de teste é calculado o valor- $p = \frac{\sum_{l=1}^B I(z_{cl} \leq z_c)}{B}$ e H_0 é rejeitada

caso valor- $p < \alpha$.

2.3 Análise de componentes principais

A análise de componentes principais (ACP), introduzida por Pearson (1901), é uma metodologia estatística multivariada na qual são criadas q variáveis latentes (Y_1, Y_2, \dots, Y_q), denominadas componentes principais, que são combinações lineares de um conjunto original de p variáveis correlacionadas, encontrados de maneira que $q < p$ e cada combinação linear seja não correlacionada com a outra.

Para estabelecer os componentes principais, considere um vetor \mathbf{X} com p variáveis correlacionadas, digamos $\mathbf{X} = [X_1, X_2, \dots, X_p]^\top$, com vetor de médias $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^\top$ e matriz de covariâncias $\boldsymbol{\Sigma}$ ($p \times p$). O i -ésimo componente (ou variável latente) principal é definido por

$$Y_i = \mathbf{e}_i^\top \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad (2.10)$$

em que o vetor \mathbf{e}_i determina a i -ésima combinação linear para $i \in \{1, 2, \dots, p\}$.

A variância de Y_i é dada pela expressão

$$\text{Var}(Y_i) = \text{Var}(\mathbf{e}_i^\top \mathbf{X}) = \mathbf{e}_i^\top \text{Var}(\mathbf{X}) \mathbf{e}_i = \mathbf{e}_i^\top \boldsymbol{\Sigma} \mathbf{e}_i \quad (2.11)$$

e a covariância entre dois componentes principais Y_i e Y_j ($i \neq j$) é dado por

$$\text{Cov}(Y_i, Y_j) = \text{Cov}(\mathbf{e}_i^\top \mathbf{X}, \mathbf{e}_j^\top \mathbf{X}) = \mathbf{e}_i^\top \boldsymbol{\Sigma} \mathbf{e}_j. \quad (2.12)$$

Para determinar um componente principal Y_i o vetor \mathbf{e}_i deve ser obtido de maneira que $\text{Var}(Y_i)$ seja maximizada. No entanto, o máximo da variância do componente principal não existe, pois cresce para infinito na medida em que os componentes do vetor \mathbf{e}_i crescem e, por esse motivo, é conveniente maximizar a variância impondo a restrição $\mathbf{e}_i^\top \mathbf{e}_i = 1$, ou seja, a

soma dos coeficientes do vetor e_i deve ser igual à unidade (FERREIRA, 2011).

Para tal, deve-se maximizar a seguinte expressão em relação a e_i

$$\max[e_i^\top \Sigma e_i - \lambda_i(e_i^\top e_i - 1)] \quad (2.13)$$

em que λ_i é o multiplicador de Lagrange.

Derivando a expressão (2.13) em relação a e_i , igualando sua derivada a zero e simplificando obtém-se a seguinte expressão

$$(\Sigma - \lambda_i \mathbf{I})e_i = \mathbf{0}, \quad (2.14)$$

em que \mathbf{I} é a matriz identidade $p \times p$.

Da expressão (2.14) segue que

$$\Sigma e_i = \lambda_i e_i. \quad (2.15)$$

Além disso, é possível verificar que

$$\begin{aligned} \text{Var}(Y_i) &= e_i^\top \Sigma e_i = e_i^\top \lambda_i e_i = \lambda_i e_i^\top e_i = \lambda_i && \text{e que} \\ \text{Cov}(Y_i, Y_j) &= e_i^\top \Sigma e_j = e_i^\top \lambda_j e_j = \lambda_j e_i^\top e_j = 0 \end{aligned}$$

se $i \neq j$, já que e_i e e_j são ortogonais.

Observando (2.15) é possível concluir que os vetores e_i e os multiplicadores de Lagrange λ_i ($i \in \{1, 2, \dots, p\}$) são, respectivamente, os autovetores e autovalores da matriz Σ (FERREIRA, 2011). Isso justifica o uso termos “componentes principais” e “autovalores da matriz de covariâncias” como sinônimos neste trabalho.

2.4 Análise de correlação canônica

A análise de correlação canônica, introduzida por Hotelling (1935), objetiva o estudo das relações lineares existentes entre dois vetores ou conjuntos de variáveis. Para tal, deve-se gerar combinações lineares dos vetores de tal forma que a correlação entre esses vetores seja

maximizada.

Considere os vetores \mathbf{X} e \mathbf{Y} de dimensões p e q , respectivamente, e as covariâncias entre os pares de variáveis de \mathbf{X} e \mathbf{Y} , representados na matriz de covariâncias Σ_{XY} de dimensão $p \times q$. Nas situações em que p e q são grandes, existe maior dificuldade em analisar os elementos da matriz de covariâncias conjuntamente. Nesse caso, a análise de correlações canônicas é uma alternativa viável por resumir a informação contida nos pq parâmetros em um conjunto menor de coeficientes de correlação entre \mathbf{X} e \mathbf{Y} .

A análise de correlações canônicas consiste em criar pares de variáveis latentes, ou seja, combinações lineares dos vetores \mathbf{X} e \mathbf{Y} , de tal forma que sejam fortemente correlacionadas e que concentrem a maior parte informação contida na matriz de covariâncias. A cada etapa do procedimento é possível criar um novo par de variáveis, não correlacionado com o par anterior e que contenha a maior parte da informação restante, até que toda a informação de covariâncias tenha sido explicada pelos pares de variáveis latentes criados. As variáveis que são criadas a partir das combinações lineares do conjunto de variáveis são chamadas de variáveis canônicas e o k -ésimo par de variáveis canônicas é definido por

$$U_k = \mathbf{a}_k^\top \mathbf{X} \quad \text{e} \quad V_k = \mathbf{b}_k^\top \mathbf{Y}$$

em que \mathbf{a}_k e \mathbf{b}_k são vetores constantes de dimensões $p \times 1$ e $q \times 1$, respectivamente, escolhidos para maximizar a correlação entre as variáveis U_k e V_k , dada pela expressão:

$$\rho_{U_k, V_k} = Cor(U_k, V_k) = \frac{\mathbf{a}_k^\top \Sigma_{XY} \mathbf{b}_k}{\sqrt{\mathbf{a}_k^\top \Sigma_{XX} \mathbf{a}_k} \sqrt{\mathbf{b}_k^\top \Sigma_{YY} \mathbf{b}_k}} \quad (2.16)$$

de modo que a variância de U_k e V_k seja igual a 1 (MINGOTI, 2005).

Para maximizar a expressão dada em (2.16), deve-se derivá-la em relação aos vetores \mathbf{a}_k e \mathbf{b}_k obtendo, respectivamente, as seguintes equações normais:

$$\Sigma_{XY} \mathbf{b}_k - \rho_{U_k, V_k} \Sigma_{XX} \mathbf{a}_k = \mathbf{0} \quad (2.17)$$

$$\Sigma_{YX} \mathbf{a}_k - \rho_{U_k, V_k} \Sigma_{YY} \mathbf{b} = \mathbf{0} \quad (2.18)$$

Desenvolvendo as expressões (2.17) e (2.18), é possível encontrar os vetores \mathbf{a}_k e \mathbf{b}_k como segue:

$$\mathbf{a}_k = \Sigma_{XX}^{-1/2} \mathbf{c} \quad (2.19)$$

em que \mathbf{c} é o maior autovetor da matriz $\Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2}$

$$\mathbf{b}_k = \Sigma_{YY}^{-1/2} \mathbf{d} \quad (2.20)$$

em que \mathbf{d} é o maior autovetor da matriz $\Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2}$

Finalmente, para reduzir a dimensionalidade dos grupos originais de variáveis, o vetor \mathbf{a}_k pode ser estimado, de modo que seja relacionado ao vetor \mathbf{b}_k , pela seguinte transformação linear:

$$\mathbf{a}_k = \frac{1}{\rho_{U_k, V_k}} \Sigma_{XX}^{-1} \Sigma_{XY} \mathbf{b}_k.$$

2.5 Taxas de erro tipo I e poder de um teste

Ao se realizar qualquer procedimento ligado à inferência estatística, deve-se observar que as conclusões estão sujeitas a erros. Em particular, ao realizar testes de hipóteses, destacam-se o erro tipo I e o erro tipo II, que estão relacionados ao desempenho dos testes.

O erro tipo I refere-se à rejeição da hipótese de nulidade (H_0) quando esta hipótese é verdadeira. Neste trabalho o erro tipo I significa admitir que há, pelo menos, um julgador mal treinado quando, de fato, todos os julgadores possuem um bom treinamento. O valor de α (nível de significância do teste) é a probabilidade de ocorrência deste erro.

Um teste é dito conservador quando pratica taxas de erro tipo I menores que o nível de significância adotado (α) e, caso contrário, ou seja, quando pratica taxas significativamente maiores que α , é dito liberal. Quando pratica taxas iguais a α diz-se que ele controla ou respeita o nível nominal de significância (MACHADO et al., 2005).

O erro tipo II, cuja probabilidade de ocorrência é representada por β , é caracterizado

pela não rejeição da hipótese de nulidade (H_0) nos casos em que esta hipótese é falsa. O poder de teste é a probabilidade de que H_0 seja rejeitada quando é realmente falsa, ou seja, não rejeitar a hipótese alternativa (H_1) quando esta é verdadeira (MOOD; GRAYBILL; BOES, 1974).

Estes erros são inversamente proporcionais, assim, quanto menor for a significância do teste, maior será a taxa de erro tipo II e, por consequência, menor será o poder do teste. Em geral, o erro tipo I é previamente fixado (significância do teste) e o erro tipo II pode ser controlado aumentando-se o tamanho da amostra (FERREIRA, 2011).

2.6 Simulação Monte Carlo

A simulação é a elaboração de um modelo teórico para um sistema real, e são conduzidos experimentos com este modelo, para compreender o comportamento do sistema (SHANNON, 1975). Assim, a simulação serve como uma primeira avaliação de um sistema para gerar novas estratégias de análise e regras de decisões antes de se correr o risco de experimentá-las no sistema real (NAYLOR et al., 1971).

Diante disso, o emprego de simulações é essencial para a avaliação de desempenho dos métodos devido às complexas operações matriciais envolvidas, dificultando o julgamento analítico comparativo desses métodos. O método de Monte Carlo é muito utilizado para se determinar propriedades de estimadores. Todo processo simulado que envolve um componente aleatório de qualquer distribuição é considerado como pertencente ao método Monte Carlo (SANTOS, 2001; FERNANDES, 2012).

A simulação Monte Carlo é utilizada com os recursos e as técnicas computacionais em que amostras são geradas de acordo com determinadas distribuições teóricas conhecidas, visando estudar o comportamento de diferentes técnicas estatísticas que poderiam ser empregadas num dado problema (DACHS, 1988; RAMOS, 2009).

3 METODOLOGIA

O presente trabalho se constitui da generalização e avaliação de cinco testes multivariados de unidimensionalidade de painéis sensoriais. Além disso, um dos testes generalizados foi aplicado em um experimento com dados reais. A generalização dos testes em si será explicitada como resultado metodológico na seção (4.1). Dessa forma, nessa seção é apresentada a metodologia empregada para a validação e comparação dos testes propostos e detalhado o experimento com dados reais no qual o teste com melhor desempenho foi aplicado.

3.1 Experimento de simulação

Para validar e comparar os testes foram realizadas simulações computacionais para avaliar seu desempenho, em termos de poder e taxa de erro tipo I. Para isso, foram construídas rotinas em código R (R CORE TEAM, 2014), tanto para a implementação dos testes quanto para sua validação e comparação, que foram realizadas via simulação Monte Carlo.

Para a simulação computacional, foram considerados 216 cenários, gerados a partir da combinação entre o número de atributos ($q= 2, 5, 10$ e 20), número de produtos ($n= 5, 10, 15$ e 20), número de julgadores ($p=2, 5, 10$ e 15) e correlações entre os julgadores ($\rho=0,1; 0,3; 0,5; 0,7; 0,9$ e $0,99$). Os cenários foram estabelecidos de forma que $n > q$ e $n \geq p$.

O grau de treinamento ou proporção da variação explicada pelo primeiro componente principal (ρ_1^2), utilizado para a construção dos gráficos, indica a consonância do painel sensorial e varia de acordo com a configuração do cenário, pois é calculado em função do número de julgadores (p) e da correlação entre eles (ρ). Conforme resultados apresentados por Fernandes (2012), pode ser obtido por meio da expressão:

$$\rho_1^2 = \frac{\rho(p - 1) + 1}{p}$$

Para cada um dos cenários foram simuladas $N = 1000$ amostras de Monte Carlo. Além disso, em todos os experimentos, foram simuladas amostras de populações multivariadas com parâmetro Σ_q de modo que o grau de treinamento dos avaliadores estivesse confinado no inter-

valo entre 0 e 1.

Para a geração das amostras multivariadas deste estudo, deve-se observar que existe correlação entre os atributos sensoriais (variáveis) e entre os julgadores. As covariâncias entre atributos e entre julgadores foram impostas (fixadas) pelo procedimento de composição dos dados, que segue o modelo linear:

$$Y_{ijm} = \mu + \tau_j + \beta_m + e_{ijm},$$

em que Y_{ijm} é a observação da j -ésima variável do produto i , para o julgador m , μ é uma constante comum a todas as observações (zero, sem perda de generalidade), τ_j é o efeito da variável (atributo) j , β_m é o efeito aleatório do julgador m , e e_{ijm} é o erro aleatório associado a Y_{ijm} . A composição de cada observação é ilustrada na Figura (2) e explicada a seguir.

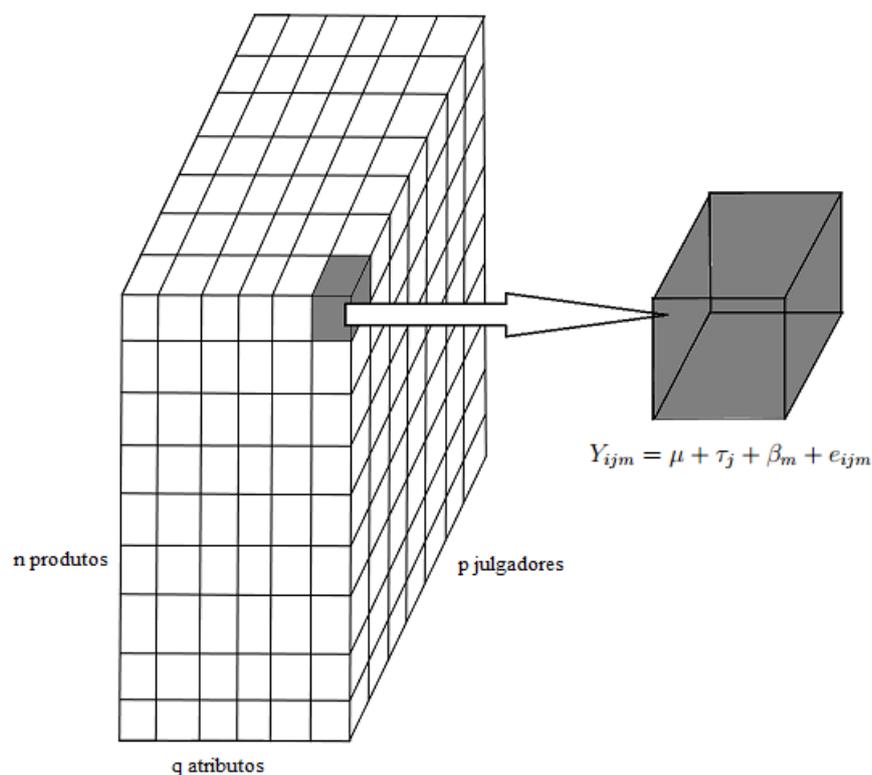


Figura 2 – À esquerda, a amostra, à direita, uma observação Y_{ijm} .
Fonte: Da autora.

Inicialmente foi gerada uma hipermatriz nula de dimensão $n \times q \times p$. Assim, a constante $\mu = 0$ foi atribuída a todas as observações.

Em seguida, foi somado o efeito do atributo (τ_j) a cada uma das fatias $n \times p$ da hi-

permatriz. Para isso, foi sorteada uma observação q variada, proveniente de uma distribuição $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}_q)$, sendo $\boldsymbol{\mu} = \mathbf{0}$ o vetor de médias dos atributos e $\boldsymbol{\Sigma}_q$ a matriz de covariâncias entre os atributos. No presente trabalho, a correlação entre os atributos sensoriais foi fixada em zero.

Analogamente, para cada uma das fatias $n \times q$ da hipermatriz, o efeito do julgador m (β_m) foi somado a todas as notas atribuídas pelo mesmo. Assim, para o sorteio dos efeitos de julgador (fixando-se a qualidade do treinamento desejada, ρ), foi utilizada uma distribuição $N_p(\mathbf{0}, \boldsymbol{\Sigma}_p)$, em que

$$\boldsymbol{\Sigma}_p = \sigma^2 \mathbf{R} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \ddots & \rho \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{bmatrix}.$$

Note que a matriz de correlação \mathbf{R} é equicorrelacionada, e a correlação (que indica a qualidade do treinamento) assumiu os valores mencionados anteriormente ($\rho=0,1; 0,3; 0,5; 0,7; 0,9$ e $0,99$). Foi fixado, sem perda de generalidade, $\sigma^2 = 1$.

Por fim, os erros aleatórios foram sorteados de uma distribuição normal padrão e somados a cada uma das observações.

Em cada amostra gerada foram realizados todos os testes construídos conforme proposta que apresentada na seção (4.1), fixando-se a correlação ideal em $\rho_0 = 0,75$. Foram computadas as porcentagens de rejeição da hipótese nula ao longo das $N = 1000$ amostras para o nível nominal de significância $\alpha = 5\%$ e, dessa forma, o estimador da função poder (FP), ou seja, a proporção de rejeições de H_0 é dada por

$$\text{FP} = \frac{\sum_{\ell=1}^N I(\text{valor-}p_{\ell} \leq \alpha)}{N} \quad (3.1)$$

em que $\text{valor-}p_{\ell}$ é o valor- p da ℓ -ésima amostra Monte Carlo simulada e I é a função indicadora que retorna 1 se ($\text{valor} - p_{\ell} \leq \alpha$) e 0, caso contrário.

Nos casos em que amostra gerada era proveniente de uma população em que $\rho < \rho_0$, a proporção de rejeição de H_0 em N simulações Monte Carlo foi considerada uma estimativa de poder. Analogamente, quando a amostra foi gerada de uma população com $\rho \geq \rho_0$, a proporção de rejeições de H_0 em N simulações será a estimativa da taxa de erro tipo I, com o valor nominal α fixado em 5% .

A partir da proporção de rejeição de H_0 em N simulações Monte Carlo os testes foram comparados.

3.2 Experimento com dados reais

Na seção (4.3) será proposta uma aplicação da generalização dos testes no experimento realizado por Pereira (2005), que objetivou produzir queijo minas padrão adicionado de probióticos - bactérias acidificantes (*Lactococcus lactis ssp.lactis* e *Lactococcus lactis ssp. cremoris*, bactérias probióticas (*Lactobacillus.acidophilus* e *Lactobacillus casei*) - e um prebiótico (inulina) em três concentrações (0%, 2%, 4%).

Os probióticos são aditivos alimentares que contém bactérias que propiciam certo benefício ao hospedeiro e que agem no equilíbrio da microbiota intestinal. Já os aditivos prebióticos são aditivos alimentares não digeríveis e que também trazem benefícios ao hospedeiro por estimularem a proliferação de bactérias (probióticos) no intestino. Os prebióticos podem ser combinados com probióticos para a produção de um alimento simbióticos, que são considerados alimentos funcionais, ou seja, que além de suprir as suas necessidades nutricionais, contribuam com benefícios à saúde e ao bem estar, proporcionando melhor qualidade de vida e prevenindo o aparecimento de determinadas doenças (PEREIRA, 2005).

A inulina é um importante prebiótico que pode tornar o queijo minas padrão um alimento funcional, mas sua adição não deve afetar a qualidade sensorial do queijo. Dessa forma, foi realizada a análise sensorial verificar se a adição de inulina é prejudicial às características sensoriais do queijo minas padrão.

Seiscentos litros de leite pasteurizado com 3,5 % de gordura, acidez entre 15 e 16 °D (Dornic), foram utilizados na fabricação dos queijos. Os três tratamentos continham as mesmas proporções dos ingredientes, sendo diferenciada na adição do prebiótico inulina. Amostras de queijo foram coletadas nos tempos 0, 15, 30 e 45 dias de maturação e submetidas às análises sensoriais.

Foi adicionado cloreto de cálcio ao leite na proporção de 250 mg/L de leite, culturas lácticas mesofílicas mista (*Lactococcus lactis ssp.lactis* e *Lactococcus lactis ssp. cremoris*, ma - 016 fermentec[®]), na proporção de 1,5%, coalho bovino líquido comercial na proporção indicação pelo fabricante e culturas lácticas probióticas (na proporção de 1,5 % (*Lactobacillus*

acidophilus - la14 Fermentec[®], *Lactobacillus paracasei ssp.paracasei* - lbc81 fermentec[®]). A salga dos queijos foi feita em solução de cloreto de sódio a 20 % (m/v) em água a 10°C a 12°C por 24 horas, foi utilizado a inulina em pó (Raftiline st, clariant[®]) com 92 % de grau de pureza.

Para a análise sensorial, foram consideradas as avaliações de doze tecnologias de fabricação de queijos, resultantes da combinação entre três concentrações de inulina (0%, 2% e 4%) e os quatro tempos de maturação dos queijos (0, 15, 30 e 45 dias) e foram pontuados quatro atributos, a saber: aparência, cor, sabor e textura.

As análises sensoriais foram realizadas no Laboratório de análise sensorial do Departamento de Ciência dos Alimentos (DCA) da Universidade Federal de Lavras (UFLA-MG).

Foram selecionados julgadores (alunos da pós-graduação e graduação, professores ou funcionários do departamento), os quais receberam um treinamento onde foi observado a disponibilidade de tempo, atenção, aptidão e responsabilidades. Dentro desses foi recrutado um grupo de 20 pessoas para receber o treinamento. Em seguida, foram selecionadas 15 pessoas que foram submetidas a uma série de testes de diferenças, que tinham por objetivo avaliar a habilidade sensorial dos julgadores.

As amostras foram apresentadas em quantidade suficiente e à temperatura ambiente, codificadas com três dígitos aleatórios e com duas repetições para os atributos sabor e textura, e uma repetição para os atributos cor e aparência. Além disso, as amostras foram avaliadas em dois horários, de manhã e a tarde.

Os queijos elaborados foram submetidos à análise sensorial para avaliação dos atributos, utilizando Teste Duplo de Ordenação-Preferência (DELLA-MODESTA, 1994). Também foram submetidos ao Teste de Aceitação, mediante o uso de escala hedônica de 9 pontos (STONE; SIDEL, 1985).

Vale notar que as análises foram realizadas em várias sessões e, por esse motivo, dez julgadores não compareceram a todas as sessões. Diante disso, na seção (4.3), foram considerados apenas os 5 julgadores que pontuaram os produtos em todas as sessões para a aplicação da generalização proposta no presente trabalho.

4 RESULTADOS E DISCUSSÃO

Na presente seção são descritos a proposta e validação dos testes, a comparação entre os mesmos, além da aplicação em um experimento com dados reais. A parte metodológica da generalização dos testes é detalhada na seção (4.1) e a validação e comparação dos testes é apresentada e discutida na seção (4.2).

4.1 Generalização dos testes de unidimensionalidade

Considere a avaliação de n produtos, por p avaliadores que utilizam q atributos (variáveis).

Considere que $\mathbf{X}_{ij}^\top = [X_{ij1}, X_{ij2}, \dots, X_{ijq}]$ seja o vetor q -dimensional avaliado no i -ésimo produto pelo j -ésimo julgador, sendo $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$. Assim, pode-se conceber a matriz \mathbf{X} ($n \times pq$) do painel como

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11}^\top & \mathbf{X}_{12}^\top & \cdots & \mathbf{X}_{1j}^\top & \cdots & \mathbf{X}_{1p}^\top \\ \mathbf{X}_{21}^\top & \mathbf{X}_{22}^\top & \cdots & \mathbf{X}_{2j}^\top & \cdots & \mathbf{X}_{2p}^\top \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{X}_{i1}^\top & \mathbf{X}_{i2}^\top & \cdots & \mathbf{X}_{ij}^\top & \cdots & \mathbf{X}_{ip}^\top \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{X}_{n1}^\top & \mathbf{X}_{n2}^\top & \cdots & \mathbf{X}_{nj}^\top & \cdots & \mathbf{X}_{np}^\top \end{bmatrix}. \quad (4.1)$$

A partir da matriz 4.1 tem-se o seguinte par de hipóteses:

$$\begin{cases} H_0 : & \text{O painel é unidimensional.} \\ H_1 : & \text{O painel não é unidimensional para, pelo menos, uma variável.} \end{cases} \quad (4.2)$$

Isso posto, são apresentadas cinco formas de se testar a hipótese nula.

Com a finalidade reduzir a dimensão de \mathbf{X} de modo que seja possível aplicar os testes descritos na seção (2.2), no algoritmo dos testes, substituiu-se a matriz de covariância amostral (\mathbf{S}) pela matriz de correlação canônica (\mathbf{R}). Dessa forma, a etapa inicial, que consiste em construir a matriz \mathbf{R} e é descrita a seguir, é comum a todos os testes.

As generalizações são propostas para os casos em que o número de atributos avaliados é superior a um ($q > 1$) e consistem, inicialmente, em reduzir a dimensão para o caso unidimensional, ou seja, a dimensão coluna da matriz de dados será reduzida de pq para p . A adaptação proposta deve-se ao fato de que a aplicação dos testes generalizados nesse trabalho só ser possível para um atributo fixo. Diante disso, a vantagem oferecida pela generalização dos testes é permitir analisar, simultaneamente, todos os atributos sensoriais.

Considerando as matrizes \mathbf{X}_j e $\mathbf{X}_{j'}$, que contém as notas de dois julgadores distintos (j e j'), são buscadas as combinações lineares $Y_j = \mathbf{a}_j^\top \mathbf{X}_j$ e $Y_{j'} = \mathbf{b}_{j'}^\top \mathbf{X}_{j'}$, tais que a correlação amostral $r_{Y_j, Y_{j'}}$ seja maximizada. Esse máximo é facilmente obtido utilizando a teoria das correlações canônicas. Se esse processo for repetido para todos os pares de julgadores, a matriz de correlação amostral \mathbf{R} pode ser construída por

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}.$$

Em termos paramétricos, a matriz de correlações canônicas $p \times p$ pode ser denotada por

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}.$$

Uma vez definida a matriz \mathbf{R} , cada um dos testes propostos é finalizado por um dos testes para a retenção de componentes principais descritos na seção (2.2).

Vale notar que os testes *bootstrap* paramétrico baseado em R_{kl}^2 (Bp), *bootstrap* paramétrico baseado em z_{cl} (Bpz) e do teste baseado nos resultados de Fujikoshi (Fuji) inferem a respeito da proporção da variação explicada pelo primeiro componente principal. Dessa forma, generalização desses testes implicam no seguinte par de hipóteses:

$$\begin{cases} H_0 : \rho_1 \geq \rho_0 \\ H_1 : \rho_1 < \rho_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \rho_1^2 \geq \rho_{10}^2 \\ H_1 : \rho_1^2 < \rho_{10}^2 \end{cases} \quad (4.3)$$

O teste sobre autovalores Monte Carlo (AVmc) infere sobre a igualdade dos últimos autovalores. Sendo assim, considerando $\lambda_1, \lambda_2, \dots, \lambda_p$ autovalores de $\boldsymbol{\rho}$, estimados pelos respectivos autovalores de \boldsymbol{R} suas hipóteses nula e alternativa podem ser expressas por por:

$$\begin{cases} H_0 : \lambda_{p+1} = \lambda_{p+2} = \dots = \lambda_{pq} = 0 \\ H_1 : \text{ao menos um } \lambda_j \text{ difere de } 0, j > p. \end{cases} \quad (4.4)$$

As etapas descritas anteriormente nesta seção são representadas na Figura 3.

Vale ressaltar que, embora haja diferenças entre os pares de hipóteses apresentados em (4.3) e (4.4), a não rejeição de H_0 em ambos os casos implica na conclusão que o painel pode ser considerado unidimensional.

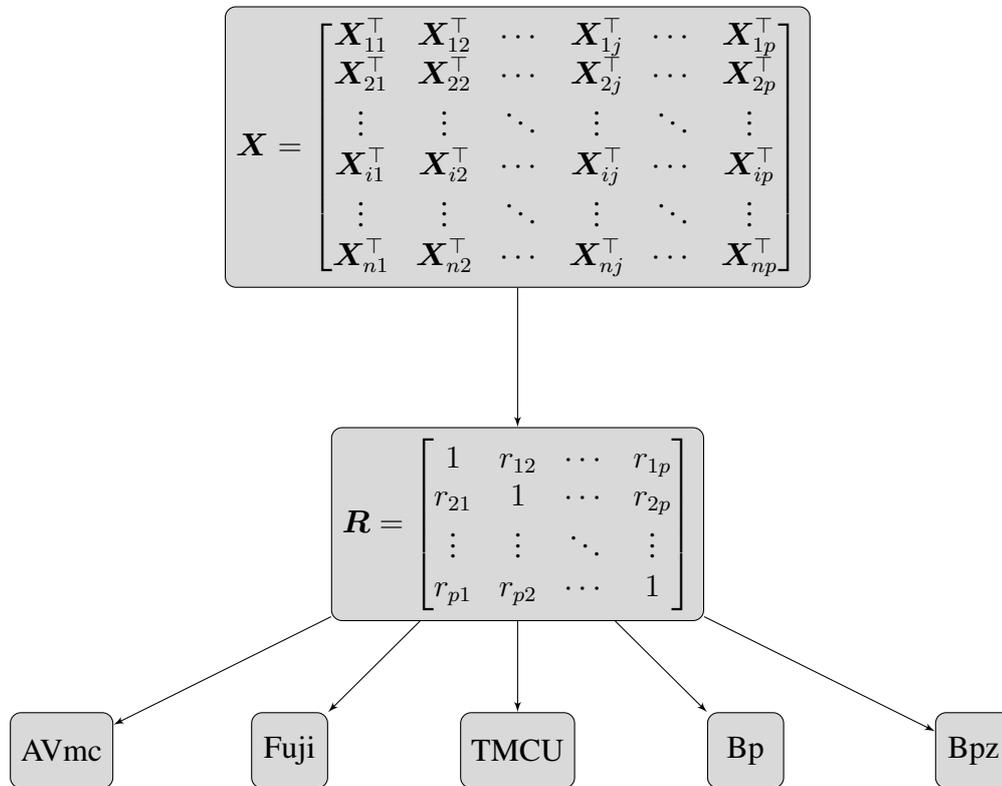


Figura 3 – Representação das etapas da generalização dos testes proposta na seção (4.1).
Fonte: Da autora.

Em qualquer um dos casos, a ideia é que, uma vez rejeitada a hipótese nula de unidimensionalidade multivariada, sejam aplicados testes de unidimensionalidade para cada atributo.

Enxerga-se, portanto, uma analogia entre o procedimento proposto e o teste F da análise de variância seguido de um teste de comparações múltiplas.

4.2 Validação e comparação dos testes

A seguir são descritos os resultados da simulação Monte Carlo e apresentados os gráficos referentes ao desempenho dos testes sobre autovalores Monte Carlo, Fujikoshi, *bootstrap* paramétrico baseado em R_{kl}^2 , *bootstrap* paramétrico baseado em z_{cl} , e teste Monte Carlo para unidimensionalidade, generalizados nesse trabalho.

Nas figuras, a linha pontilhada paralela à ordenada no ponto em que o grau de treinamento corresponde à correlação entre os julgadores igual a 0,75 representa a separação entre as regiões sob H_1 e sob H_0 e a linha pontilhada horizontal determina o nível de significância adotado ($\alpha = 0,05$). Para verificar a existência de diferenças entre o nível nominal de significância adotado (α) e as taxas de erro tipo I foi calculado o intervalo de confiança exato para proporção, com 99% de confiança, cujos extremos são representados pela linha tracejada paralela à abscissa.

A descrição detalhada do comportamento dos testes é exibida a seguir.

Avaliando $q = 2$ atributos e $p = 2$ julgadores, os testes Fujikoshi não rejeita a hipótese nula para todos os graus de treinamento estabelecidos, enquanto o teste Monte Carlo para unidimensionalidade os testes *bootstrap* paramétricos (Bp e Bpz) a rejeitam sempre. Como é possível notar pelas Figuras 4 e 5, esse resultado é recorrente, mesmo com o aumento do número de produtos. Nesses cenários, o teste sobre autovalores Monte Carlo apresenta poder baixo (37,8%) ao analisar $n = 5$ produtos, além de não controlar as taxas de erro tipo I na região próxima ao grau de treinamento tomado como referência, mas seu poder cresce com o aumento do número de produtos (n), superando 85% para $n = 20$. Além disso, o teste sobre autovalores Monte Carlo passa a controlar as taxas de erro tipo I quando foram considerados dez ou mais produtos.

Assim, analisando os cenários com dois atributos, o poder do teste sobre autovalores Monte Carlo melhora com o aumento do número de produtos (n). Ainda é possível notar que, conforme evidenciado pela comparação entre as Figuras 5 e 6, o controle das taxas de erro tipo I do teste sobre autovalores Monte Carlo diminui com o aumento do número de atributos (q), enquanto o desempenho dos demais testes permanece sem alterações, ver Apêndice A.

Pelas Figuras 7 e 8 é possível notar que, comparando-se com os casos em que $p = 2$, os testes Fujikoshi, teste Monte Carlo para unidimensionalidade e os testes *bootstrap* paramétricos apresentaram melhora no seu desempenho.

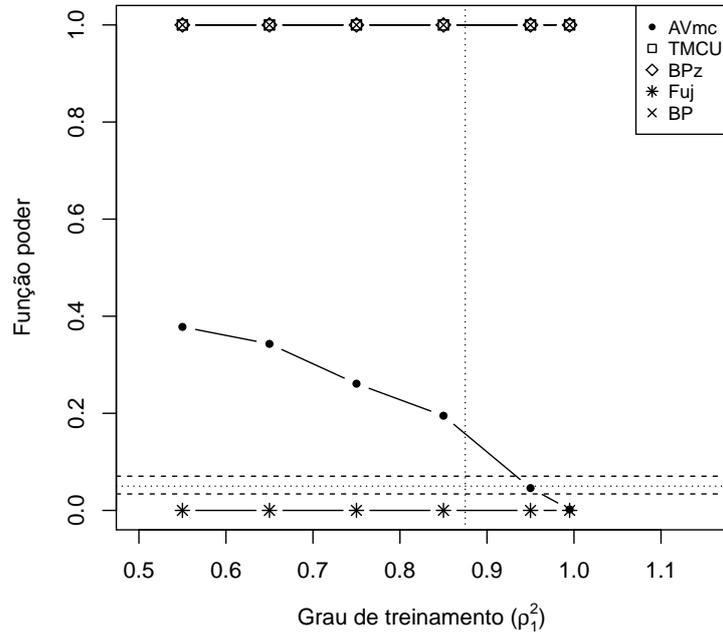


Figura 4 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 5$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

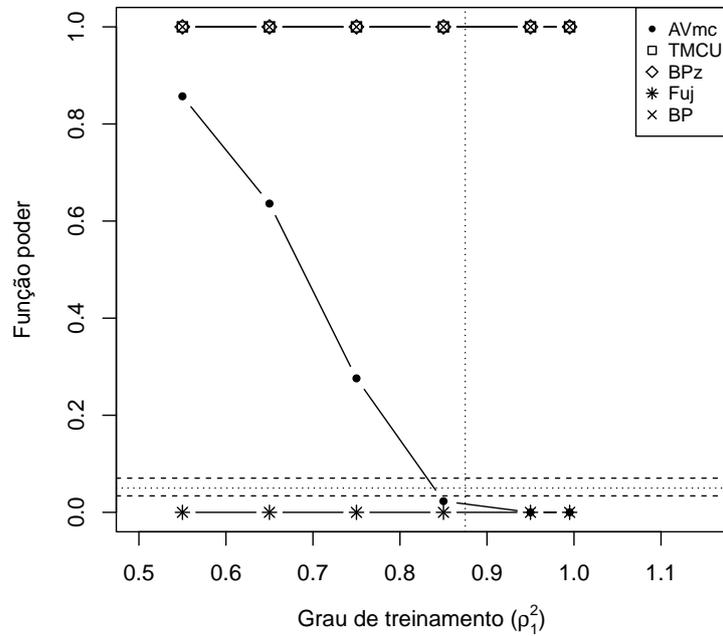


Figura 5 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 20$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

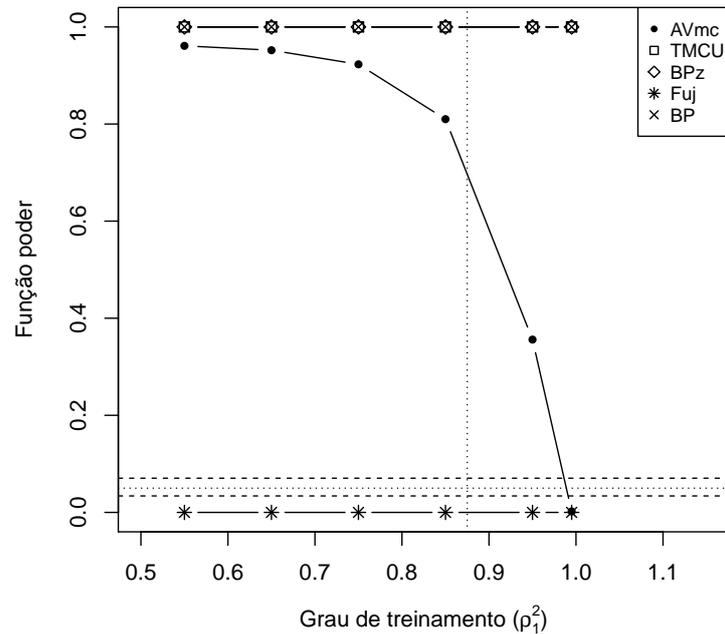


Figura 6 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 15$ atributos, $n = 20$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

De acordo com as Figuras 7 e 8, considerando ainda um número pequeno de atributos ($q = 2$), o poder apresentado pelos testes *bootstrap* paramétrico baseado em $R_{k\ell}^2$ e *bootstrap* paramétrico baseado em z_{cl} (89,27% e 95,04%, respectivamente) superou o poder dos testes sobre autovalores Monte Carlo (78,10%) e Monte Carlo para unidimensionalidade (65,80%). Considerando-se 5 julgadores e o número mínimo de atributos ($q = 2$), embora tenha apresentado um poder baixo quando $n = 5$ (abaixo de 35%), o teste de Fujikoshi mostrou uma melhora em relação ao poder com o aumento do número de produtos e, nesses casos, controlou melhor as taxas de erro tipo I em todas as situações, diferente do que ocorreu com os demais testes.

Além disso, nota-se pela Figuras do Apêndice B que, com o aumento do número de atributos (q), os testes sobre autovalores Monte Carlo, teste Monte Carlo para unidimensionalidade, Fujikoshi e os dois testes *bootstrap* paramétricos passam a ter curvas de poder cada vez mais semelhantes entre si, enquanto o teste Fujikoshi continua um pouco inferior aos demais testes em relação ao poder, mas aparenta controlar melhor a taxa de erro tipo I.

Nas Figuras 9, 10 e 11 são apresentados os resultados para $q = 2$ e $p = 10$. Nesses casos, as curvas de poder dos testes se aproximam. Em geral, os testes têm poder alto, mas não controlam a taxa de erro tipo I na região próxima ao ponto crítico.

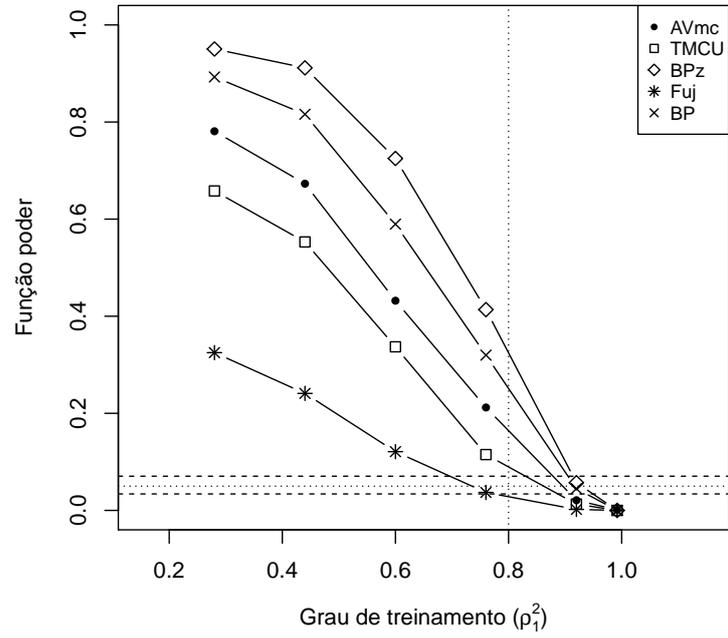


Figura 7 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 5$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

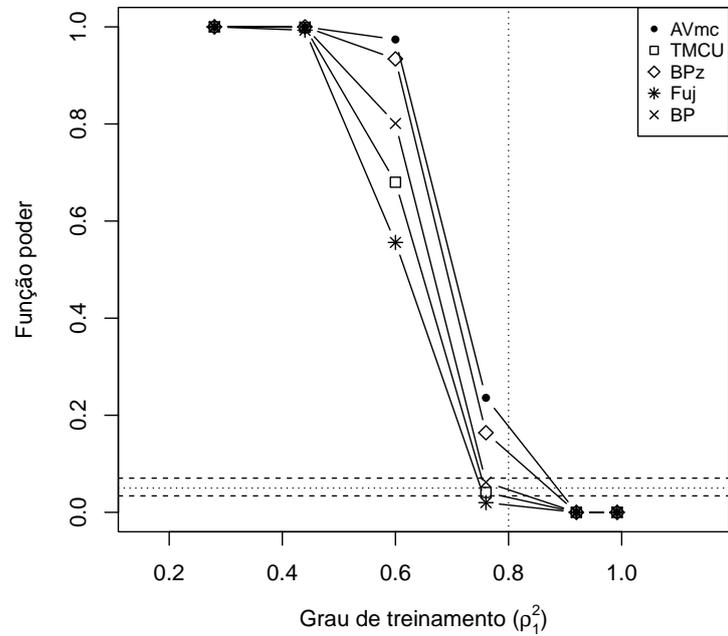


Figura 8 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 20$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

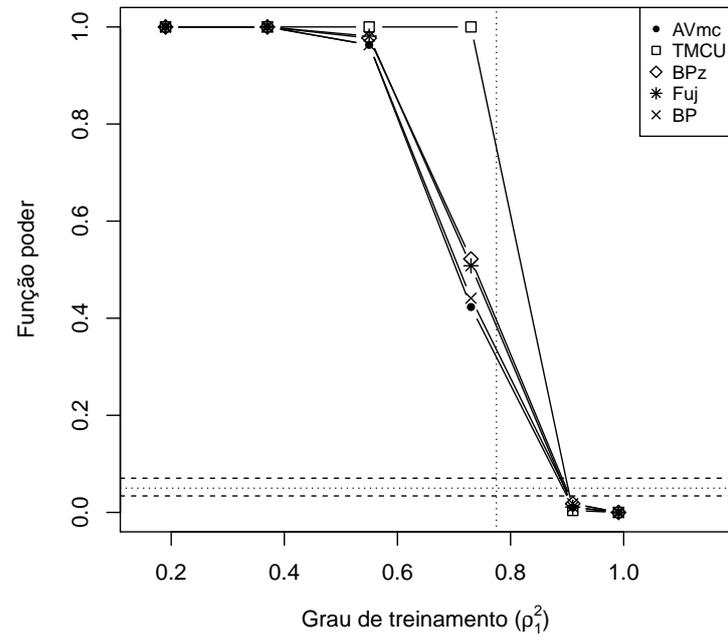


Figura 9 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 10$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

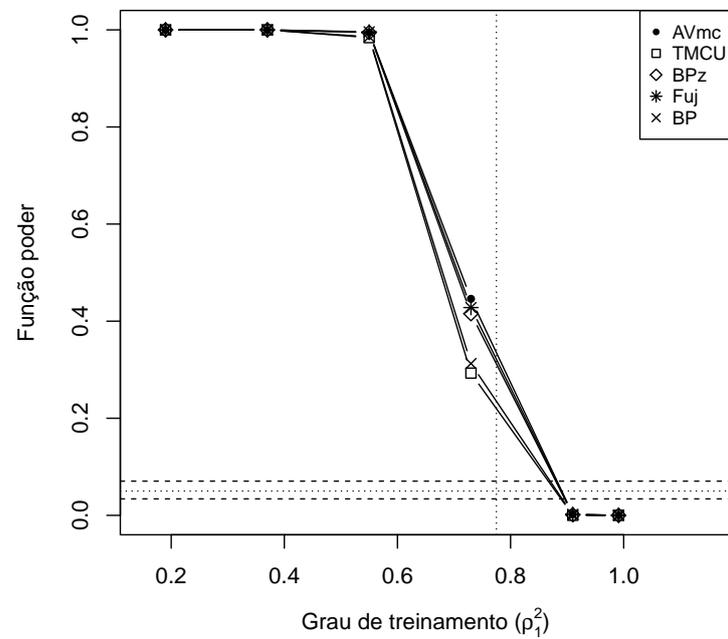


Figura 10 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 15$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

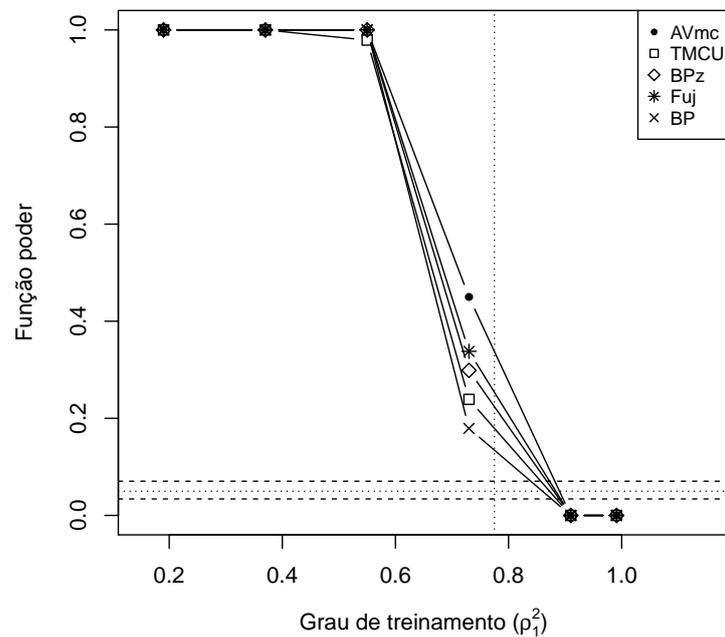


Figura 11 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 20$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

Os testes Fujikoshi e teste sobre autovalores Monte Carlo têm suas curvas de poder quase coincidentes quando se mantém o número de avaliadores em $p = 10$ e mostram-se ainda mais semelhantes ao aumentar o número de atributos (q) e o número de produtos (n) e, nessa situação, os dois testes *bootstrap* paramétricos e teste Monte Carlo para unidimensionalidade se mostram mais liberais na medida em que aumenta-se o número de atributos (conforme mostrado na Figura 12). Além disso, nota-se pelas Figuras do Apêndice C que todos os testes apresentam uma ligeira melhora no controle das taxas de erro tipo I com o aumento do valor de n .

Nas Figuras 13 e 14 são mostrados os resultados da simulação computacional nos cenários compostos por dois atributos e quinze avaliadores. Nota-se que testes têm alto poder (100% na maioria dos cenários simulados sob H_1) e apresentam comportamentos similares entre si. Além disso, a proximidade de suas curvas de poder é proporcional ao número de produtos (n).

Considerando-se ainda $p = 15$ e $q \geq 10$, nota-se pelas Figuras 15, 16 e Figuras do Apêndice D que, com o aumento do número de atributos, os testes *bootstrap* não controlaram as de erro tipo I em nenhum cenário e o teste Monte Carlo para unidimensionalidade é liberal sempre que $n = p$.

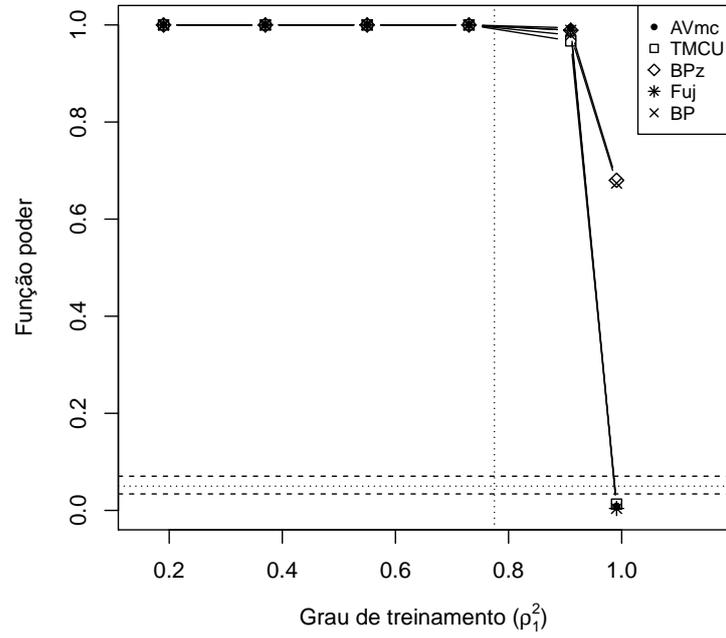


Figura 12 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

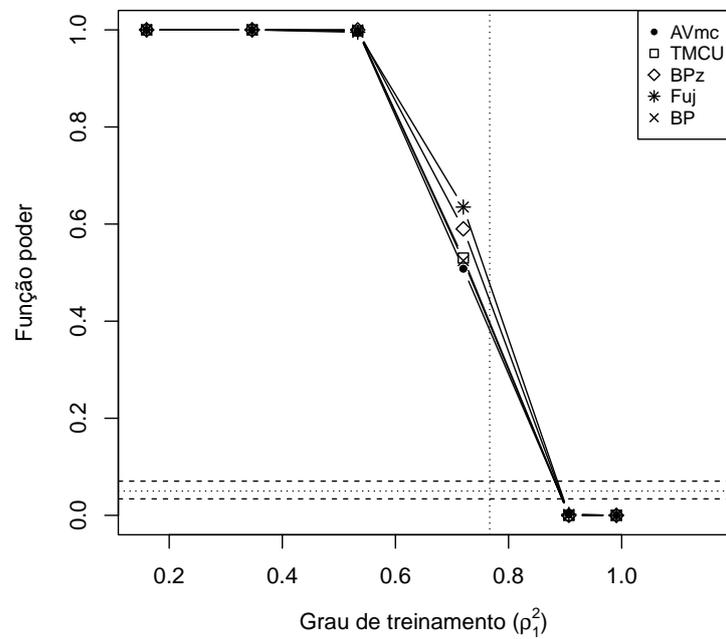


Figura 13 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 15$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

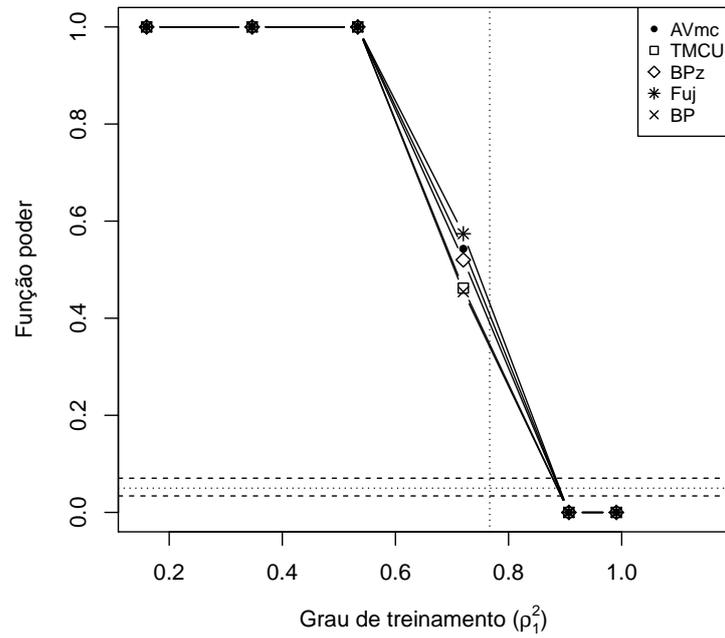


Figura 14 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 20$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

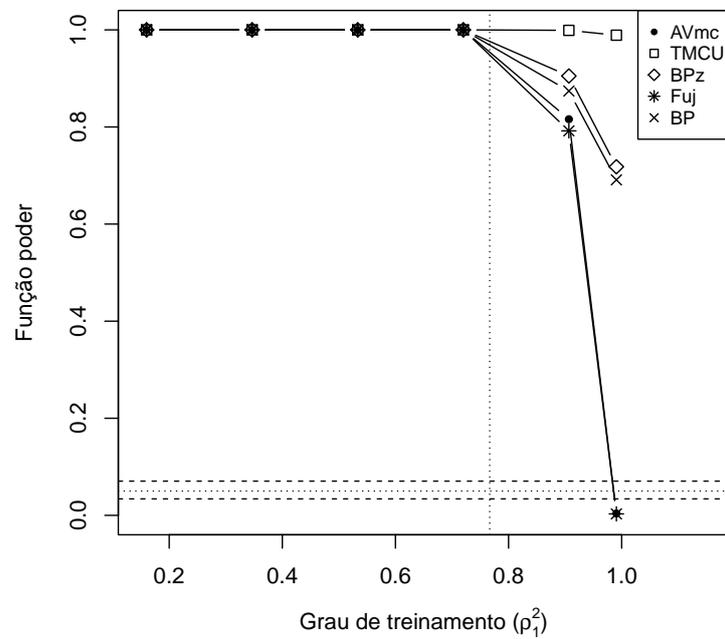


Figura 15 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 15$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

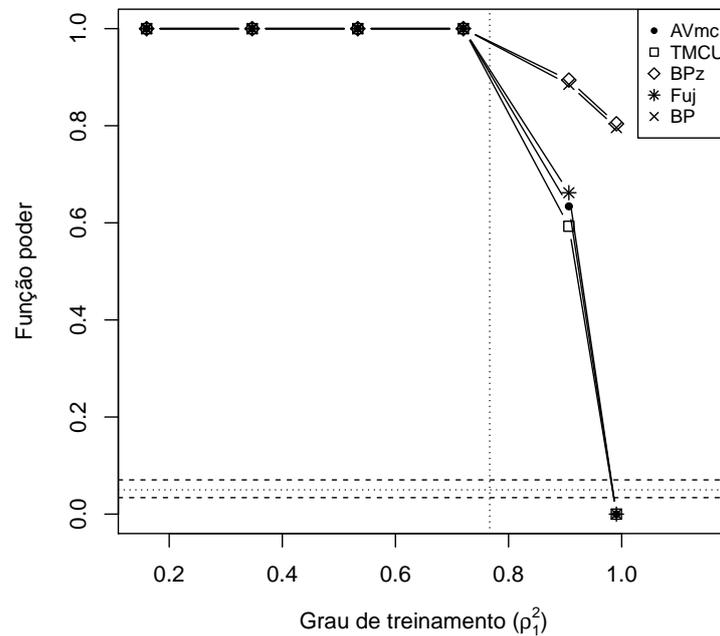


Figura 16 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

4.2.1 Discussão dos resultados de simulação

Para o número mínimo de avaliadores ($p = 2$), apenas o teste sobre autovalores Monte Carlo diferenciou os cenários simulados sob H_0 dos cenários sob H_1 . Ao avaliar o desempenho dos testes para um atributo, Gebert (2010) obteve resultados diferentes, pois considerou satisfatório o desempenho dos testes *bootstrap* paramétricos quando considerados $p = 2$ produtos. Amorim et. al. (2010), ao propor e avaliar o desempenho do teste Monte Carlo para unidimensionalidade para um atributo sensorial, afirmou que o teste mostrou-se mais liberal para um número pequeno de julgadores, mas, em geral, teve um bom desempenho.

Para $p \geq 5$, as curvas de poder dos dois testes *bootstrap* paramétricos, teste sobre autovalores Monte Carlo, testes Monte Carlo e teste de Fujikoshi se aproximam com o aumento do número de avaliadores. Resultado semelhante foi encontrado por Fernandes (2012), que comparou o desempenho dos testes para $q = 1$ atributo.

Em geral, o teste de Fujikoshi foi ligeiramente menos liberal do que os dois testes *bootstrap* paramétricos, teste sobre autovalores Monte Carlo e teste Monte Carlo para uni-

dimensionalidade em alguns dos cenários simulados. Nos testes realizados para um atributo, segundo Fernandes (2012), o teste Fujikoshi foi mais liberal que os demais. O desempenho do teste Fujikoshi contrasta com os resultados de Gebert (2010) que, ao comparar os testes para um atributo, afirmou que o teste de Fujikoshi teve desempenho inferior aos testes *bootstrap*.

Nos cenários em que $p \geq 5$ o aumento do número de produtos (n) implicou em uma melhora no comportamento dos testes sobre autovalores Monte Carlo, teste Monte Carlo para unidimensionalidade, Fujikoshi, testes *bootstrap* paramétricos (baseado em $R_{k\ell}^2$ e baseado em $z_{c\ell}$), com o aumento do poder e maior controle na taxa de erro tipo I, resultado que também foi obtido por Fernandes (2012) para $q = 1$.

O aumento do número de atributos (q) implicou no crescimento das taxas de erro tipo I, sendo que esse crescimento foi mais prejudicial ao desempenho dos testes *bootstrap* paramétricos. No entanto, o aumento do número de produtos (n) melhora o controle das taxas de erro tipo I, assim como nas simulações realizadas por Gebert (2010), Amorim et al. (2010) e Fernandes (2012) ao avaliarem o desempenho dos testes para um atributo.

Vale notar que o desempenho dos testes é melhor sempre que $n \geq pq$, pois apresentam poder mais alto e controlam melhor as taxas de erro tipo I. No entanto, essa situação é pouco utilizada na prática pois, em análises sensoriais, é bastante comum que o número de atributos e/ou julgadores supere o número de produtos.

4.3 Experimento com dados reais

O experimento analisado nessa sessão, conforme detalhado na seção 3.2, é referente à análise sensorial de 12 tecnologias de fabricação de queijo minas padrão, cujas notas foram atribuídas por 5 provadores treinados em 4 atributos: aparência, cor, sabor e textura (PEREIRA, 2005).

A partir das avaliações feitas pelos cinco provadores foi realizado o teste sobre autovalores Monte Carlo, generalizado nesse trabalho. A escolha do teste foi feita pelo seu desempenho ter sido melhor para um pequeno número de atributos no experimento de simulação computacional. Além disso, foi realizada a simulação, considerando-se o mesmo número de produtos, atributos e julgadores do experimento realizado por Pereira (2005) e, como é possível notar pela Figura 17, o teste sobre autovalores Monte Carlo apresentou bom desempenho nesse cenário.

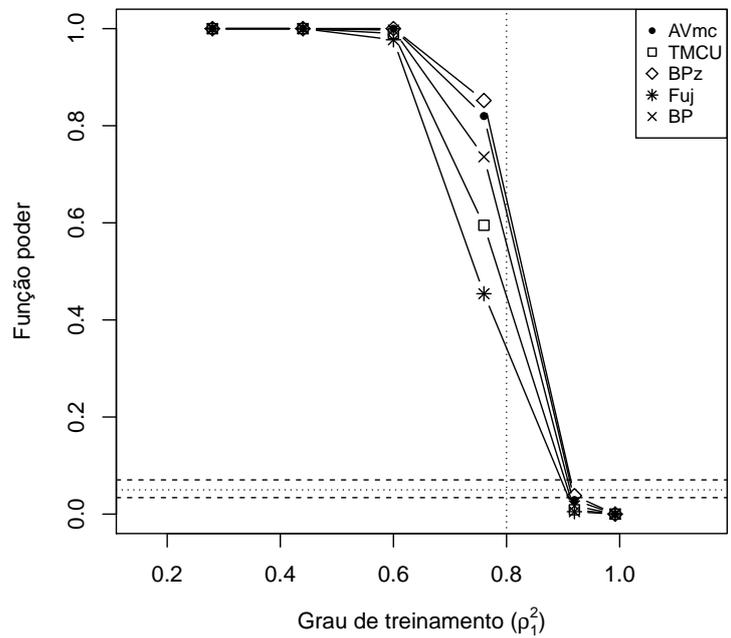


Figura 17 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 4$ atributos, $n = 12$ produtos e $p = 5$ julgadores.
 Fonte: Da autora.

A hipótese de unidimensionalidade do painel foi rejeitada ($\text{valor-}p < 0,001$), conforme ilustrado na Figura 18, ou seja, não há concordância entre os provadores, considerando-se todos os atributos simultaneamente.

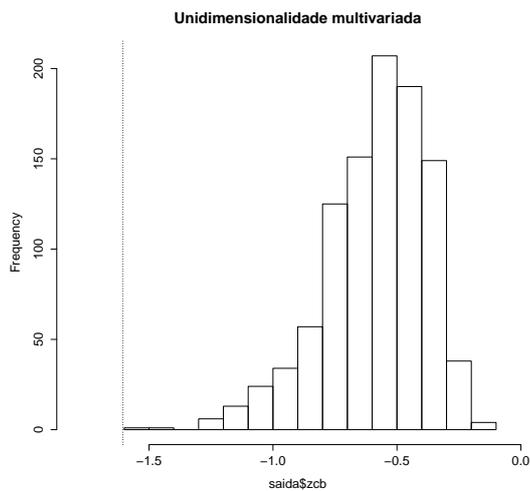


Figura 18 – Distribuição de frequências das estatísticas de teste sob H_0 (via Monte Carlo), ou seja, quando a correlação entre os provadores é 75%. A linha pontilhada indica o posicionamento da estatística de teste proveniente da amostra.
 Fonte: Da autora.

A avaliação da unidimensionalidade multivariada do painel, proposta neste trabalho, é encerrada neste ponto. A seguir são expostas duas possíveis abordagens levando em conta a rejeição da hipótese nula.

A primeira abordagem permite reconhecer os atributos para os quais os provadores necessitam ser retreinados através da análise de componentes principais. Conforme é detalhado a seguir, o procedimento consiste em verificar a unidimensionalidade do painel para cada um dos atributos. Dessa forma, em cada um dos atributos, foi aplicado o teste sobre autovalores Monte Carlo proposto e recomendado por Fernandes (2012), para verificar a unidimensionalidade do painel para um atributo.

De acordo com a Figura 19, considerando-se a aparência, ao nível nominal de 5% de significância, o painel pode ser considerado unidimensional para essa variável (valor- $p = 0,089$), ou seja, o painel é consonante em relação ao atributo aparência.

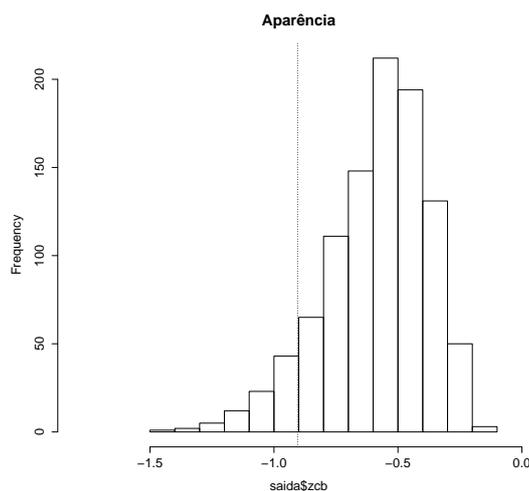


Figura 19 – Distribuição de frequências das estatísticas de teste sob H_0 para o atributo aparência. A linha pontilhada indica o posicionamento da estatística de teste proveniente da amostra.

Fonte: Da autora.

Ao avaliar o atributo cor, rejeita-se a hipótese de unidimensionalidade do painel (valor- $p = 0,002$), conforme ilustrado pela Figura 20, ou seja, pelo menos um dos provadores discorda dos demais em relação a esse atributo.

Pela Figura 21, é possível notar que o atributo cor foi mal compreendido pelos provadores, pois os provadores 1 e 2 pontuam o atributo de maneira oposta ao provador 3 e, além disso, os provadores 4 e 5 avaliam o atributo de maneira independente dos demais. A porcentagem da variação explicada pelos dois primeiros componentes principais foi 74,01%, sendo que dessa porcentagem apenas 40,06% da variabilidade é explicada pelo primeiro componente principal.

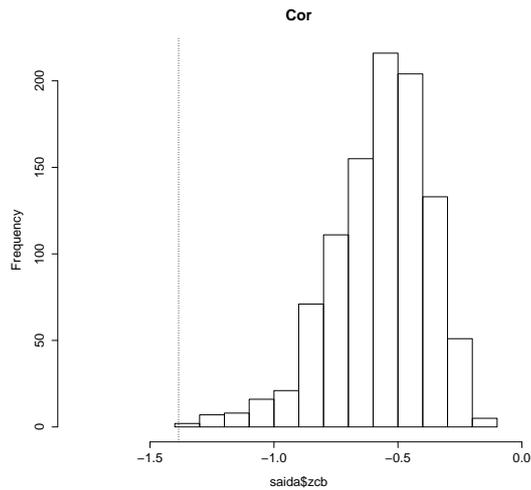


Figura 20 – Distribuição de frequências das estatísticas de teste sob H_0 para o atributo cor. A linha pontilhada indica o posicionamento da estatística de teste proveniente da amostra.

Fonte: Da autora

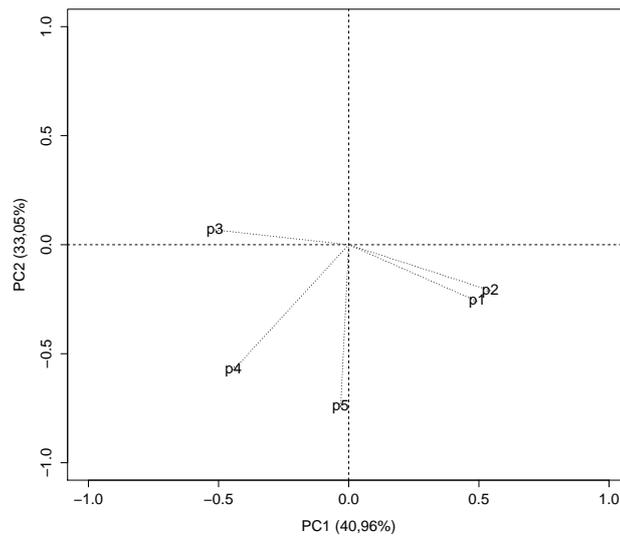


Figura 21 – Cargas da análise de componentes principais da matriz de correlações amostrais para o atributo cor. Fonte: Da autora.

Para o atributo sabor, a hipótese de consonância do painel foi rejeitada (valor- $p = 0,004$), conforme ilustrado na Figura 22. Dessa forma, pode-se concluir que há discordância entre os julgadores, considerando-se esse atributo.

A necessidade de retreinamento dos provadores para o atributo sabor é evidenciada pela Figura 23. Nota-se que os provadores 1 e 3 discordam dos demais provadores e tendem a pontuar o atributo de maneira oposta entre si. Além disso, a porcentagem da variabilidade explicada pelos dois primeiros componentes principais é baixa (42,53% e 22,11%, respectivamente).

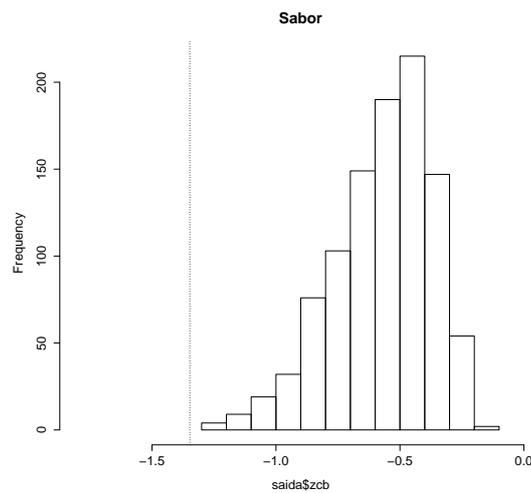


Figura 22 – Distribuição de frequências das estatísticas de teste sob H_0 para o atributo sabor. A linha pontilhada indica o posicionamento da estatística de teste proveniente da amostra.

Fonte: Da autora.

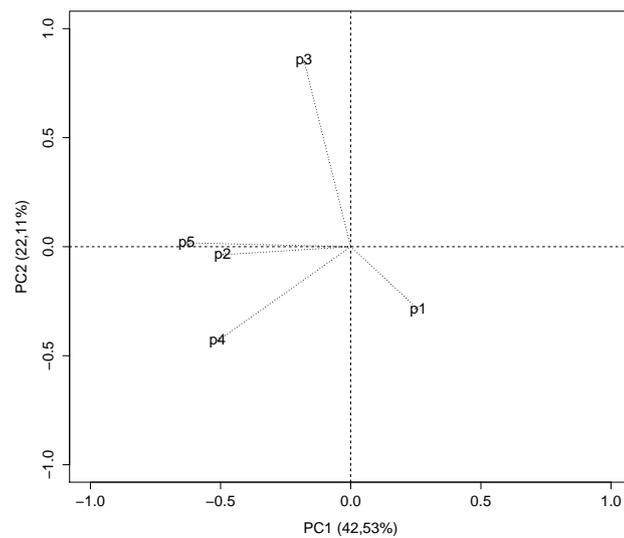


Figura 23 – Cargas da análise de componentes principais da matriz de correlações amostrais para o atributo sabor. Fonte: Da autora.

Conforme ilustrado na Figura 25, ao considerar o atributo textura, rejeita-se a hipótese de unidimensionalidade do painel (valor- $p = 0,046$), ou seja, o painel não pontua o atributo da mesma forma.

A ACP para o atributo revela que o atributo textura foi mal compreendido pelos provadores. É possível identificar na Figura 25 a avaliação feita pelos provadores 1 e 3, assim como os provadores 2 e 4. Além disso, a maior parte da variabilidade total das notas está concentrada nos dois primeiros componentes principais (58,08% e 19,81%, respectivamente).

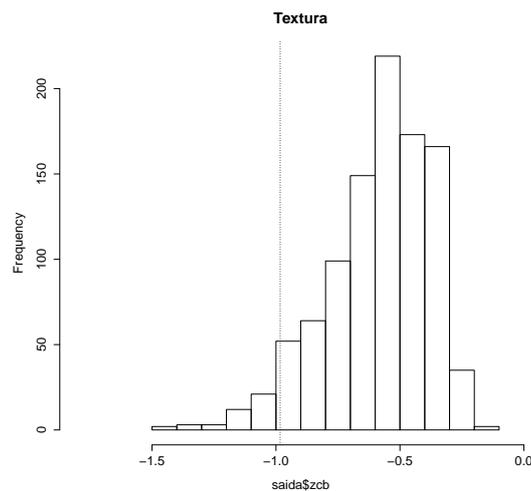


Figura 24 – Distribuição de frequências das estatísticas de teste sob H_0 para o atributo textura. A linha pontilhada indica o posicionamento da estatística de teste proveniente da amostra.

Fonte: Da autora.

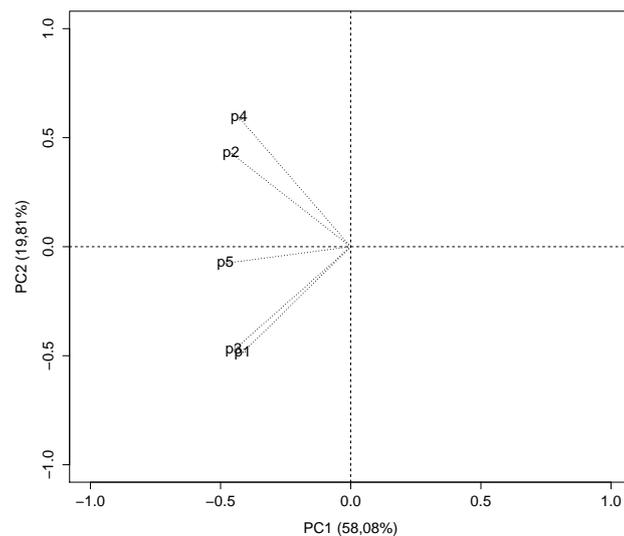


Figura 25 – Cargas da análise de componentes principais da matriz de correlações amostrais para o atributo textura. Fonte: Da autora.

Outra abordagem possível, diante da rejeição da hipótese de unidimensionalidade multivariada do painel sensorial é a análise de componentes principais na matriz de correlação canônica com a intenção de identificar o provador (ou provadores) discrepante dos demais e, dessa forma, retreiná-lo ou excluí-lo da análise sensorial. Assim, tal procedimento deve ser repetido até que seja obtida a unidimensionalidade multivariada do painel.

Vale ressaltar que a análise de componentes principais é um procedimento necessário para a execução do teste sobre autovalores Monte Carlo. Dessa forma, sem que haja necessi-

dade de mais análises, além de inferir sobre a consonância do painel, o procedimento proposto permite discutir sobre a possibilidade de retreinamento de um ou mais julgadores não tenham assimilado corretamente o treinamento para um ou mais atributos ou, caso o retreinamento não seja viável, fornece respaldo para decidir pela exclusão desse(s) julgadore(s) da análise sensorial.

5 CONCLUSÕES

Diante da necessidade de avaliar a unidimensionalidade de painéis sensoriais considerando todos os atributos sensoriais simultaneamente, foram propostas as generalizações de cinco testes para retenção de componentes principais existentes na literatura.

O teste sobre autovalores Monte Carlo foi superior aos demais testes quando o número de julgadores era mínimo ($p = 2$), pois foi o único que diferenciou as situações simuladas sob H_0 daquelas simuladas sob H_1 . Dessa forma, os demais testes não são recomendados para um número pequeno de julgadores.

Para amostras grandes, sobretudo com o aumento do número de atributos (q), os testes sobre autovalores Monte Carlo e Fujikoshi são mais recomendados do que os testes *bootstrap* paramétricos e o teste Monte Carlo para unidimensionalidade, pois apresentam alto poder e controlam melhor a taxa de erro tipo I.

Em geral, as adaptações propostas são liberais na vizinhança do ponto crítico, principalmente quando o número de atributos é grande. No entanto, passam a controlar melhor as taxas de erro tipo I na medida que aumenta-se o número de produtos.

A aplicação generalização do teste sobre autovalores Monte Carlo no experimento realizado por Pereira (2005) mostrou-se adequada e seus resultados foram corroborados pelos resultados obtidos ao avaliar a consonância do painel para cada um dos atributos.

As generalizações propostas podem ser utilizadas para avaliar a consonância de painéis sensoriais formados por julgadores que estejam avaliando qualquer produto. O número de atributos, produtos, julgadores e grau de treinamento adequado variam de acordo com o produto avaliado e os objetivos da análise sensorial, assim, para a avaliação e comparação dos testes, buscou-se estabelecer essas quantidades de modo que abrangessem grande parte das situações reais no contexto da análise sensorial.

Isso posto, para trabalhos futuros, sugere-se:

- a) Avaliar os testes em situações reais, por exemplo, considerando correlação entre os atributos e/ou produtos;
- b) particularizar o número de produtos, atributos, julgadores e grau de treinamento à uma determinada área de interesse; e
- c) analisar a influência da unidimensionalidade de um ou mais atributos, avaliados individualmente, na unidimensionalidade multivariada.

REFERÊNCIAS

- AMORIM et al. Monte Carlo based test for inferring about the unidimensionality of a Brazilian coffee sensory panel. **Food Quality and Preference**, v. 21, p. 319-323, 2010.
- BI, J. Agreement and reliability assessments for performance of sensory descriptive panel. **Journal of Sensory Studies**, n. 18, p. 61-76, 2003.
- CASTURA, J. C.; FINDLAY, C. J.; LESSCHAEVE, I. Monitoring calibration for descriptive sensory panels using distance from target measurements. **Food Quality and Preference**, Barking, v. 16, p. 682-690, 2005.
- CRONBACH, L.J. Coefficient alpha and the internal structure of tests. **Psychometrika**, n. 16, p. 297-334, 1951.
- DACHS, J. N. W. **Estatística computacional: uma introdução em Turbo Pascal**. Rio de Janeiro, 1988. 236p.
- DAEMS, V.; DELVAUX, F. Multivariate analysis of descriptive sensory data on 40 commercial beers. **Food Quality and Preference**, Barking, v. 8, p. 373-380, 1997.
- DELLA MODESTA, R. C. **Manual de análise sensorial de alimentos e bebidas**. Rio de Janeiro: EMBRAPA-CTTA, 1994.
- DERNDORFER, E. et al. A panel performance procedure implemented in R. **Journal of Sensory Studies**, n. 20, p. 217-227, 2005.
- DIJKSTERHUIS, G. Assessing panel consonance. **Food Quality and Preference**, Barking, v.6, p. 7-14, 1995.
- DUTCOSKY, S.D. **Análise sensorial de alimentos..** Curitiba: Champagnat, 2011. 426p.
- FERNANDES, F. M. O. **Proposta de um Teste Monte Carlo para unidimensionalidade de painéis sensoriais**. 2012. 80 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2012.
- FERREIRA, D. F. **Estatística Multivariada**. Lavras: Editora UFLA, 2011. 676 p.
- FERREIRA, E. B.; OLIVEIRA, M. S. **Sensometria: uma abordagem com ênfase em Procrustes**. Santa Maria: UFSM, 2007. 71 p. Apostila.
- FINN, R. H. A note on estimating the reliability of categorical data. **Educational and Psychological Measurement**, v.30, p. 71-76, 1970.
- FUJIKOSHI, Y. Asymptotic expansions for the distributions of the sample roots under nonnormality. **Biometrika**, London, v. 67, n. 1, p. 45-51, 1980.

- GEBERT, D.M.P. **Proposta de testes *bootstrap* para inferir sobre o número de componentes principais retidos**. 2010. 75 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)- Universidade Federal de Lavras, Lavras, 2010.
- GEBERT, D.M.P.; FERREIRA, D.F. Parametric *bootstrap* Tests for Determining the Number of Principal Components. **Journal of Statistical Theory and Practice**, v. 8, p. 674-691, 2013.
- HOTELLING, H. The most predictable criterion. **Journal of Educational Psychology**, v. 26, p. 139-142, 1935.
- HUMMER, S. Application of multivariate analysis of variance. **Food Quality and Preference**, Barking, v. 9, n. 3, p. 83-85, 1998.
- LATREILLE, J. et al. Measurement of the reliability of sensory panel performances. **Food Quality and Preference**, Barking, n. 17, p. 369-365, 2006.
- MACHADO, A. A. et. al. **Estatística experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais**. In: Anais da 50^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, Londrina, 2005. p. 290.
- MARTINS, C. M. R. **Proposta metodológica para otimização experimental de formulações: um estudo de caso no setor alimentício**. 2002. 144 p. Dissertação (Mestrado profissional em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- MARTENS, M. A. A philosophy for sensory science. **Food Quality and Preference**, Barking, v. 10, p. 233-234, 1999.
- MINGOTI, S. A.; **Análise de dados através de métodos estatística multivariada: Uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.
- MINIM, V. P. R. **Análise sensorial: estudos com consumidores**. Viçosa: Editora UFV, 2006. 225p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. Singapore: McGraw-Hill, 1974. 564 p.
- NAYLOR, T. H. et. al. **Técnicas de simulação em computadores**. Petrópolis: Vozes, 1971. 402p.
- PEARSON, K. Principal components analysis. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, n. 2, v. 6, p. 559, 1901.
- PEREIRA, C. **Elaboração de queijo minas padrão com adição de probióticos e prebióticos**. 2005. 123 p. Dissertação (Mestrado em Ciências dos Alimentos) - Universidade Federal de Lavras, Lavras, 2005.
- PIGGOTT, J. R. Design questions in sensory and consumer science. **Food Quality and Preference**, n.4, v.6, p. 217-220, 1995.

PINTO, F. S. T., et al. A method for panelists consistency assessment in sensory evaluations based on the Cronbachs alpha coefficient. **Food Quality and Preference**, Barking, v. 32,p. 41-47, 2014.

R CORE TEAM. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Software.

RAMOS, P.S. **Proposta e avaliação de uma solução Bayesiana para o problema de Behrens-Fisher multivariado**. 2009. 76 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2009.

SANTOS, A. C. dos. **Definição do tamanho amostral usando simulação Monte Carlo para os testes de normalidade univariado e multivariado baseados em assimetria e curtose**. 2001. 71p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2001.

SHANNON, R. E. **Systems simulation**. New Jersey: Prentice-Hall, 1975. 387 p.

SINESIO, F.; MONETA, E.; SABA, A. Comparison of multivariate methods of analysis to evaluate panellists performance. **Food Quality and Preference**, Barking, v. 3, n. 4, p. 201-208, 1992.

STONE, H.; SIDEL, J. L. **Sensory evaluation practices**. Orlando, Florida: Academic Press, 1985.

GLOSSÁRIO

Atributo: característica avaliada/pontuada em uma análise sensorial.

Consonância: a consonância de um painel é a concordância entre os julgadores.

Discriminação: capacidade de um julgador em detectar pequenas diferenças entre produtos.

Julgador ou avaliador: pessoa devidamente treinada para avaliar os atributos de um determinado produto.

Painel sensorial: conjunto de julgadores designados para pontuar os atributos de um determinado produto.

Provador: julgador especializado na análise sensorial de alimentos.

Repetibilidade: capacidade de um julgador em reproduzir suas notas para um mesmo produto e atributo.

Unidimensionalidade: relacionada à explicação da variação das notas atribuídas pelos julgadores. Um painel sensorial é dito unidimensional a variabilidade dos dados explicada pelo primeiro componente principal é suficientemente grande.

APÊNDICES

Apêndice A: Função poder empírica dos testes avaliados ($p = 2$).

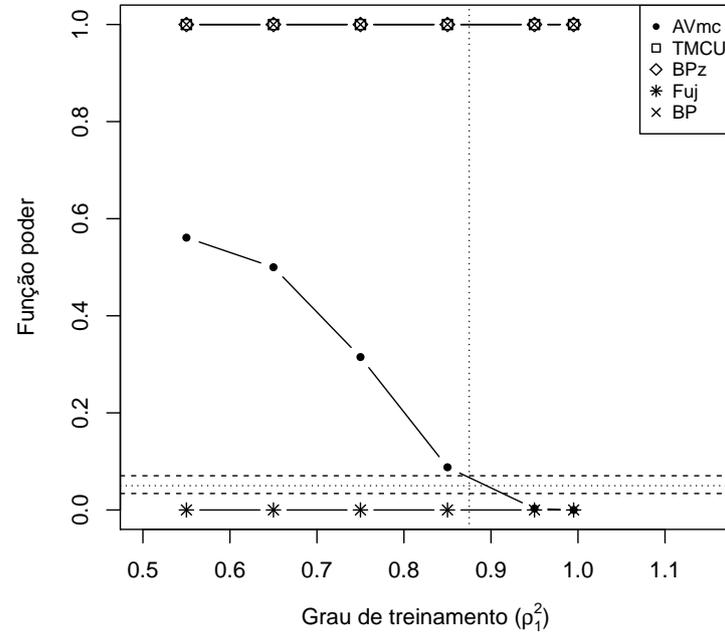


Figura 26 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 10$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

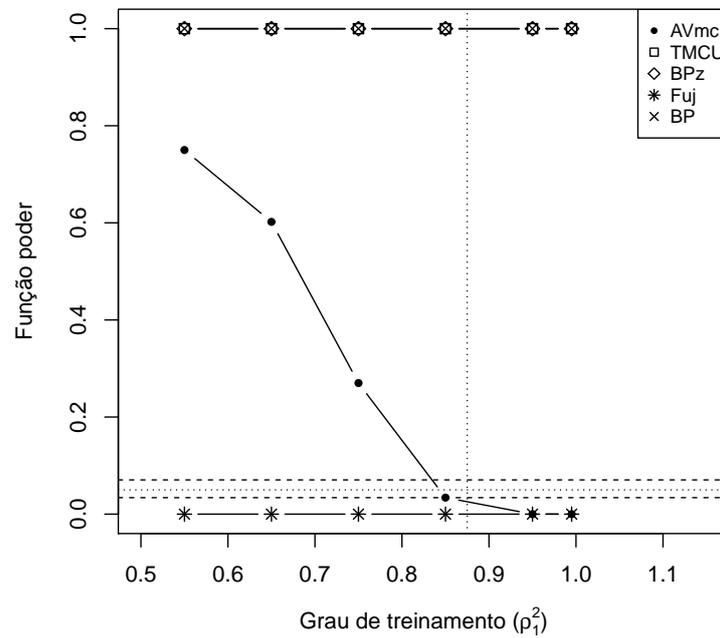


Figura 27 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 15$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

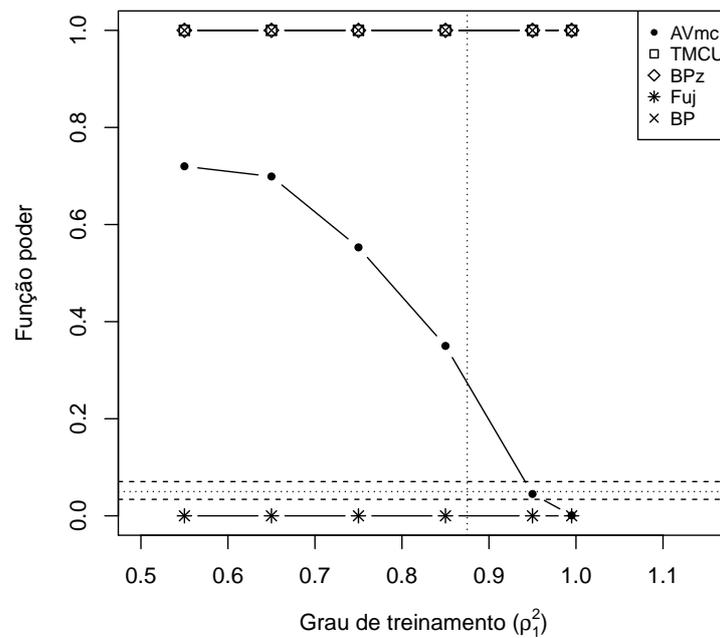


Figura 28 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 10$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

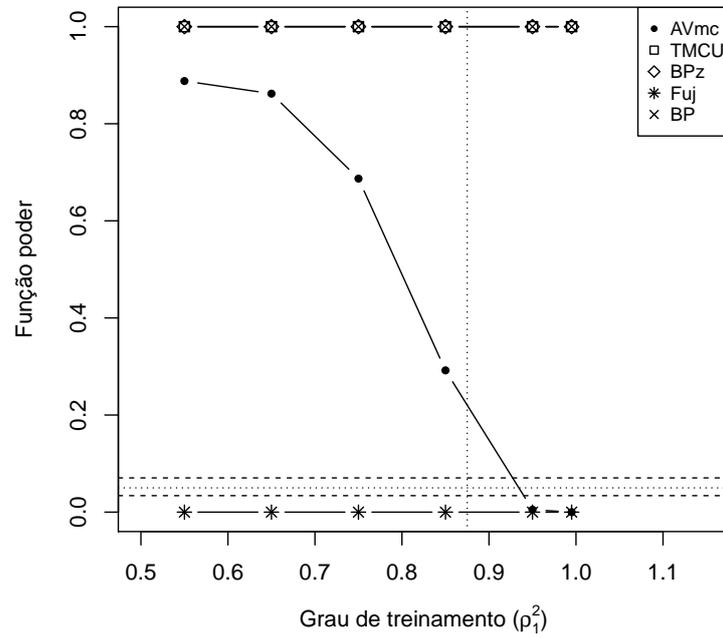


Figura 29 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 15$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

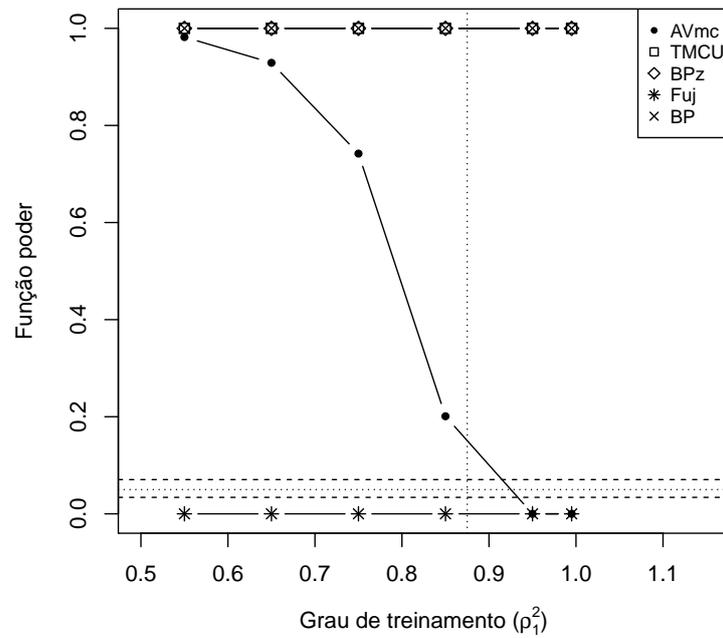


Figura 30 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 20$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

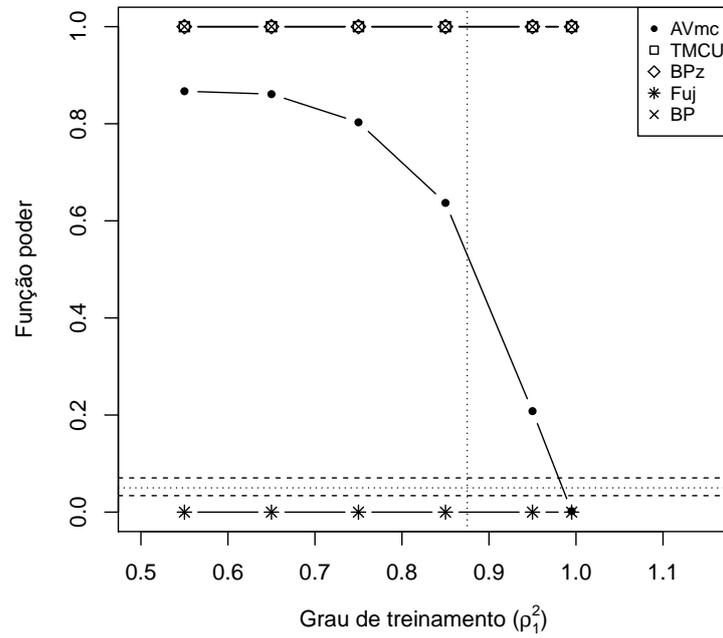


Figura 31 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 15$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

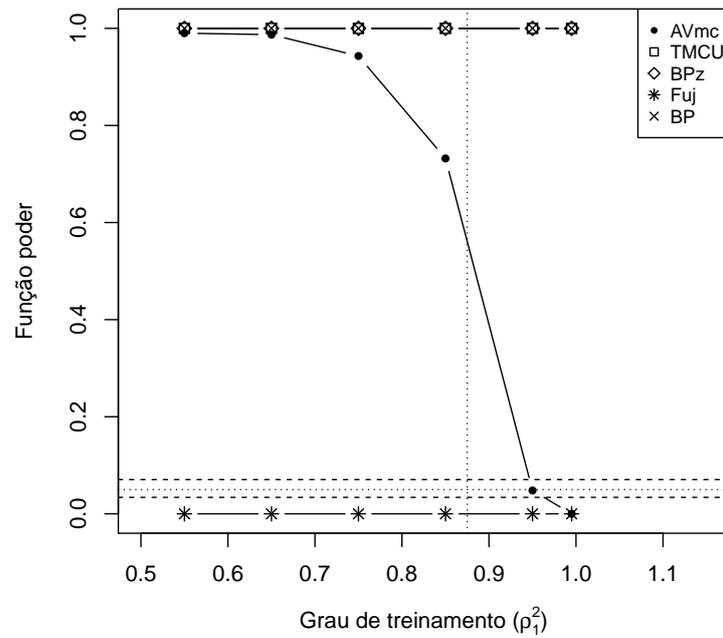


Figura 32 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 2$ julgadores.

Fonte: Da autora.

Apêndice B: Função poder empírica dos testes avaliados ($p = 5$)

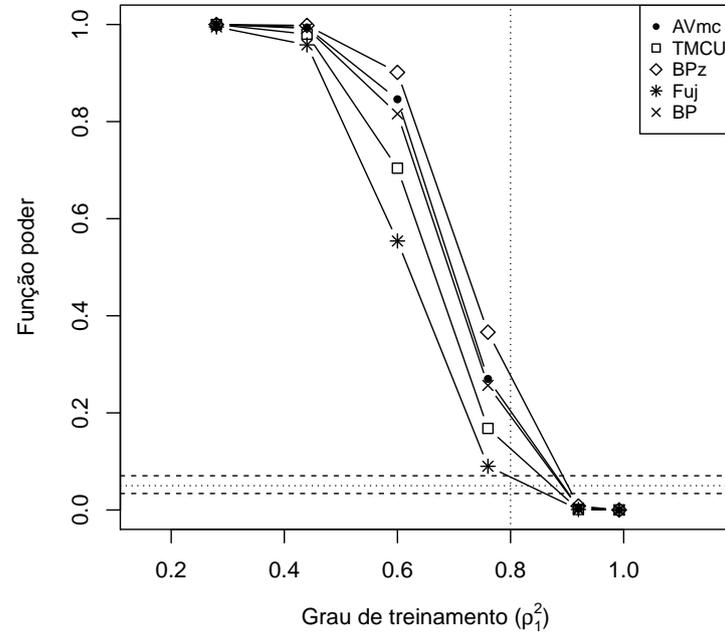


Figura 33 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 10$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

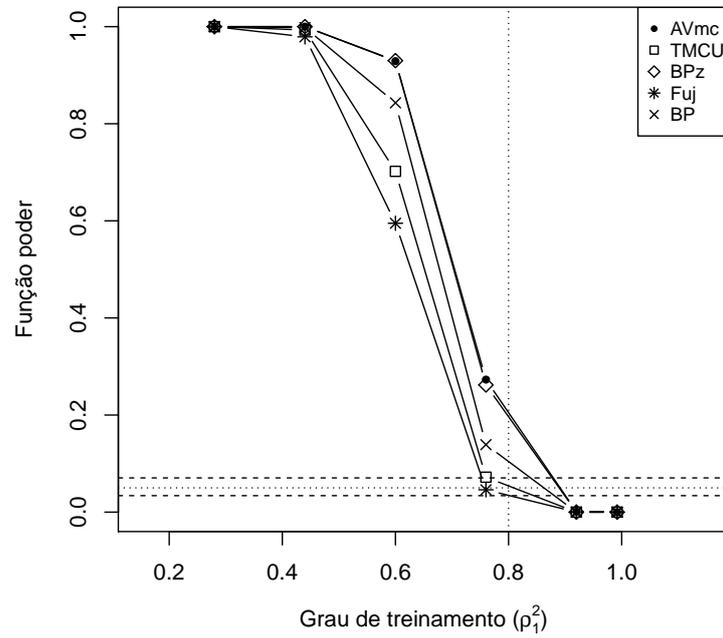


Figura 34 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 2$ atributos, $n = 15$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

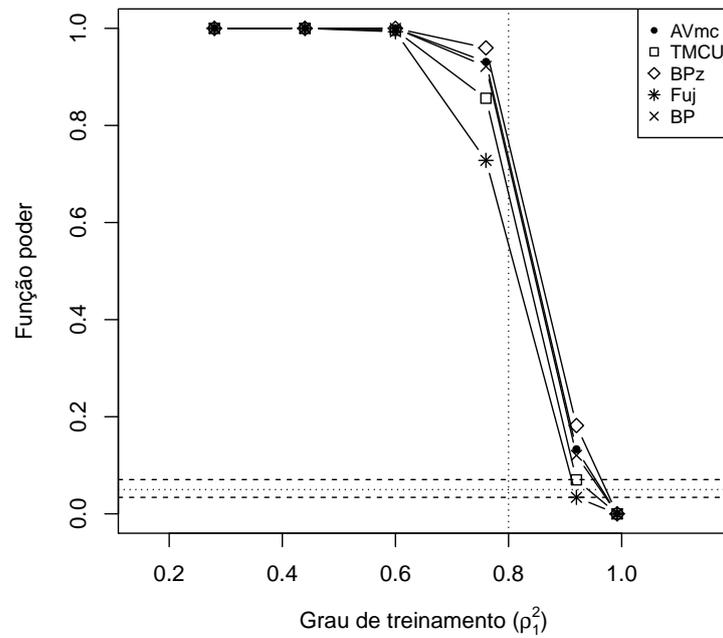


Figura 35 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 10$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

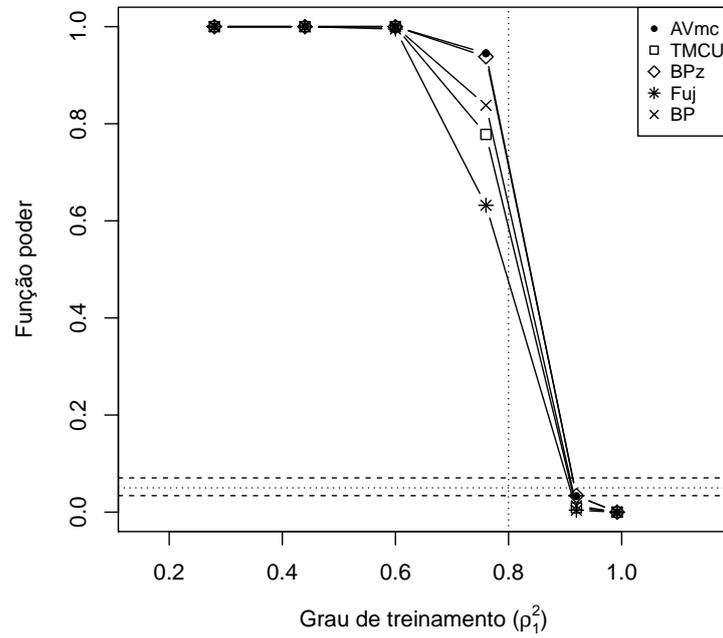


Figura 36 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 15$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

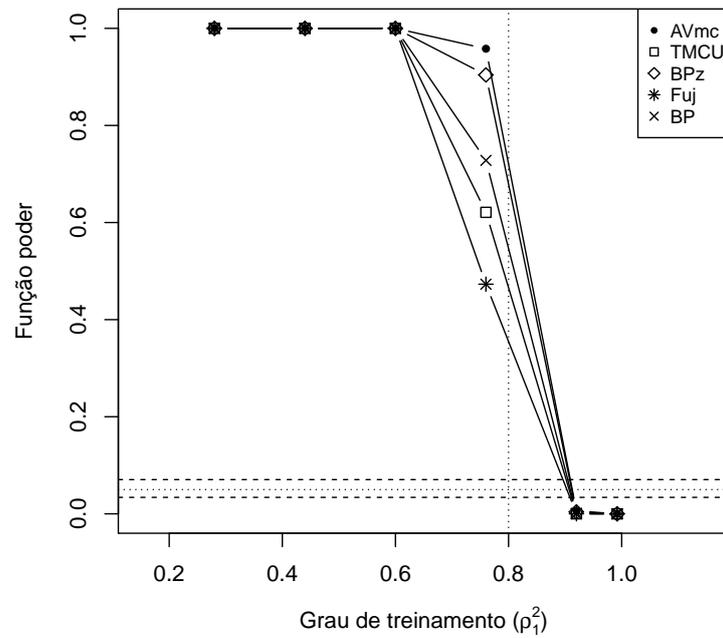


Figura 37 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 20$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

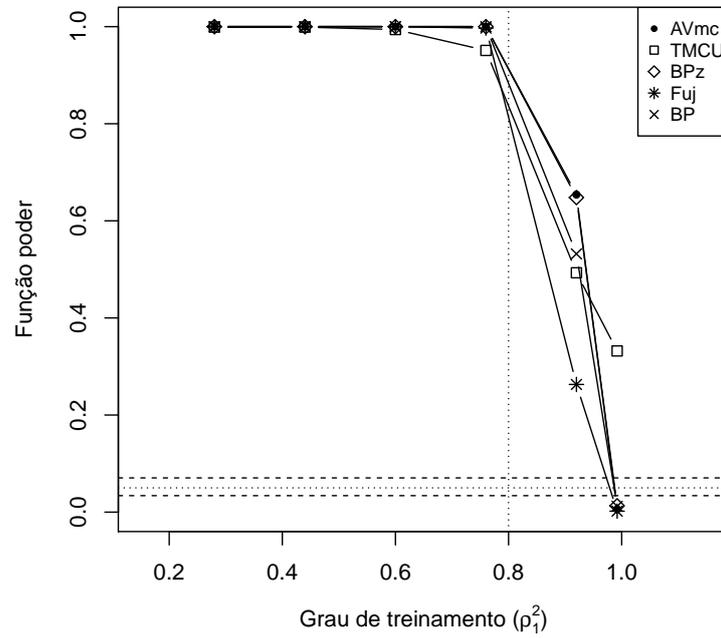


Figura 38 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 15$ produtos e $p = 5$ julgadores.

Fonte: Da autora.

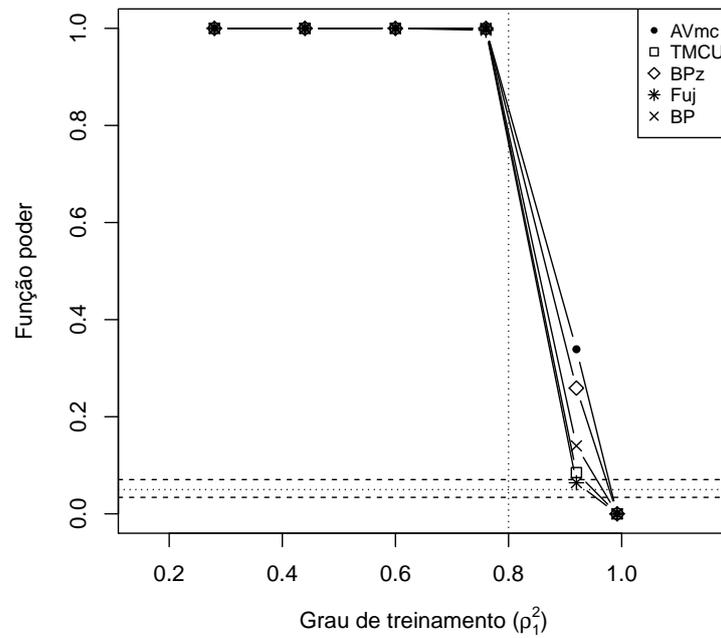


Figura 39 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

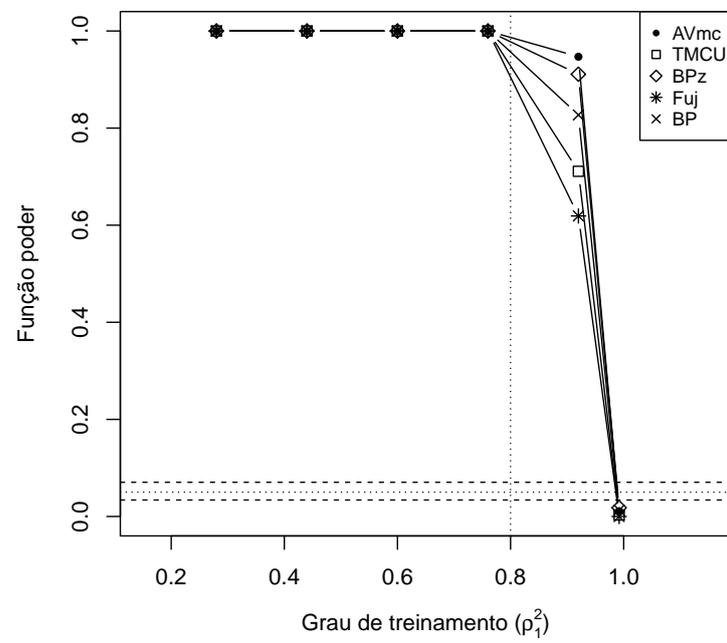


Figura 40 – Poder dos testes AVmc, Fuji, TMCU, Bp e Bpz, considerando $q = 15$ atributos, $n = 20$ produtos e produtos e $p = 15$ julgadores.

Fonte: Da autora.

Apêndice C: Função poder empírica dos testes avaliados ($p = 10$)

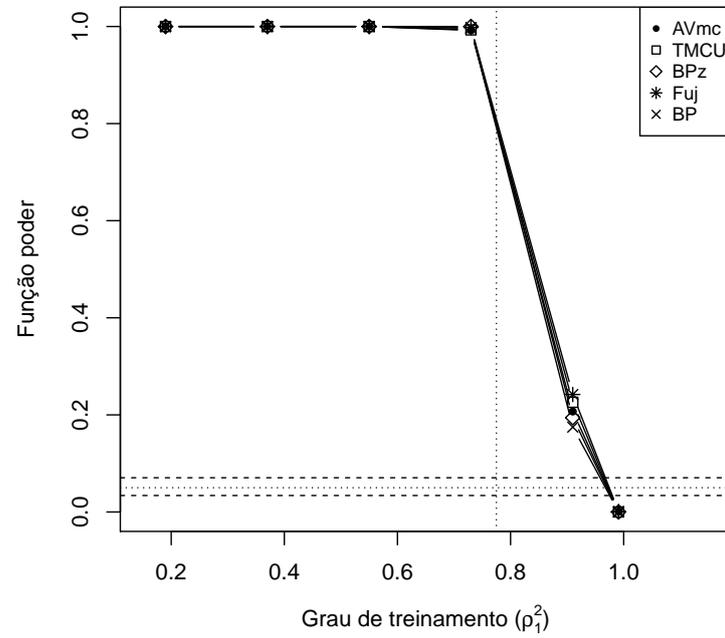


Figura 41 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 10$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

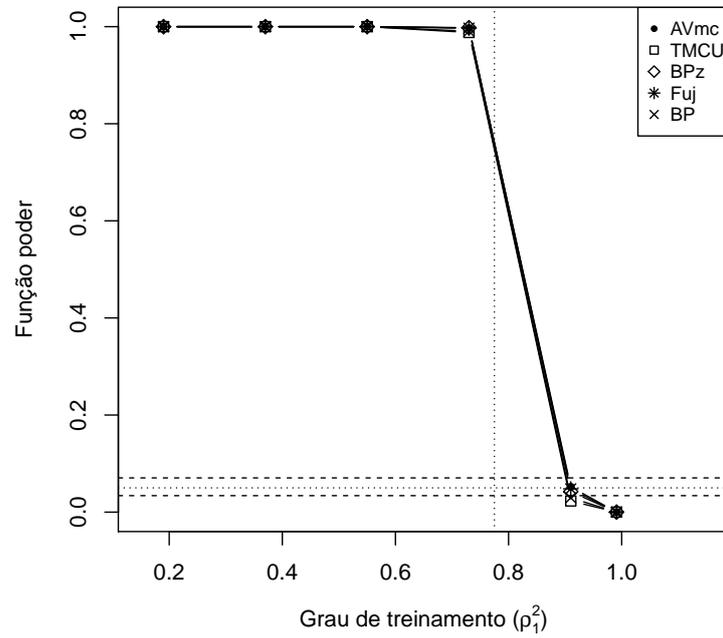


Figura 42 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 15$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

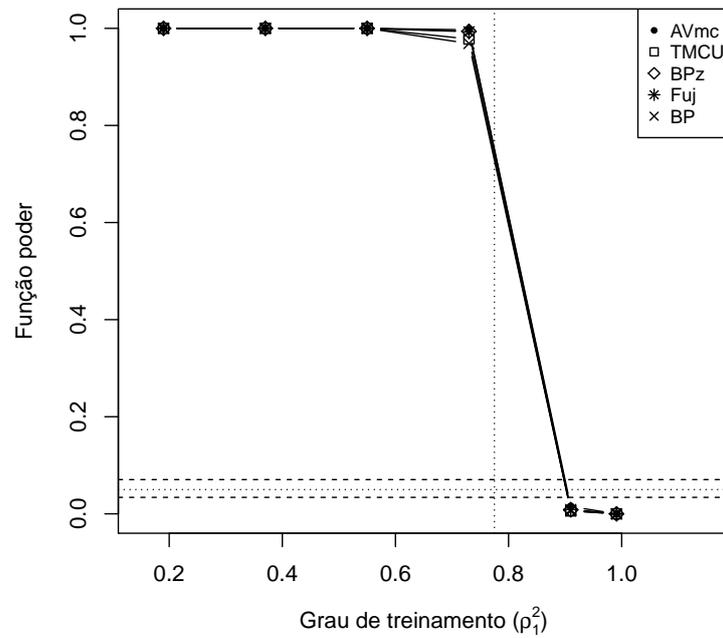


Figura 43 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 20$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

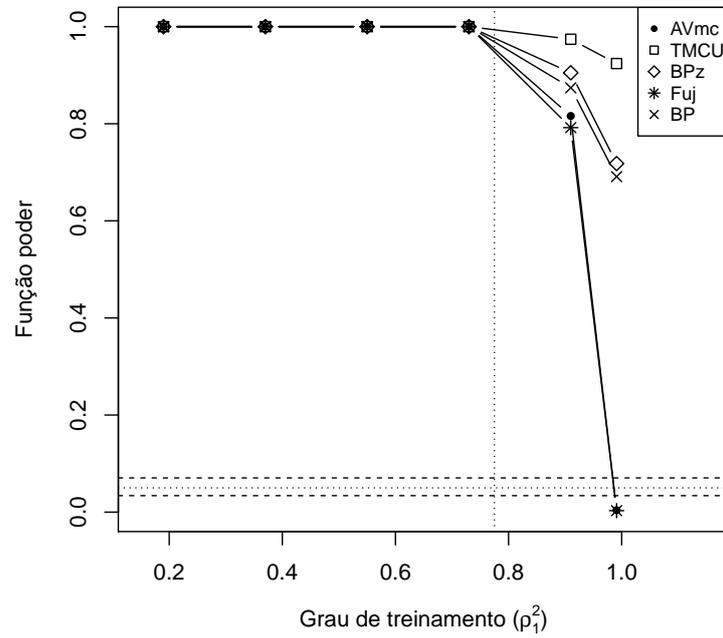


Figura 44 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 15$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

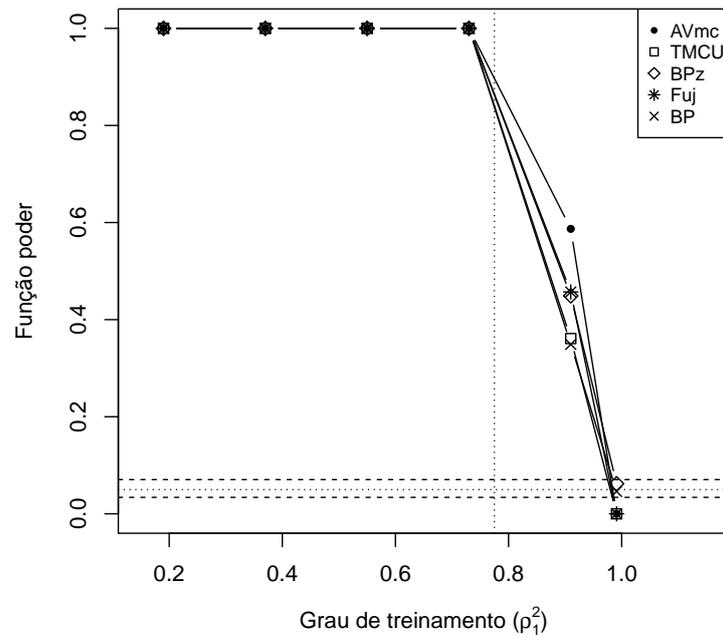


Figura 45 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 10$ julgadores.

Fonte: Da autora.

Apêndice D: Função poder empírica dos testes avaliados ($p = 15$)

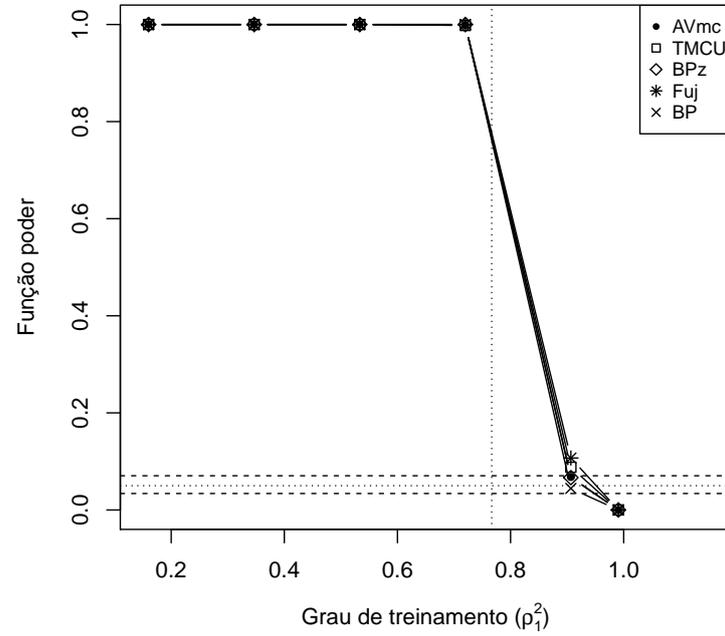


Figura 46 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 15$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

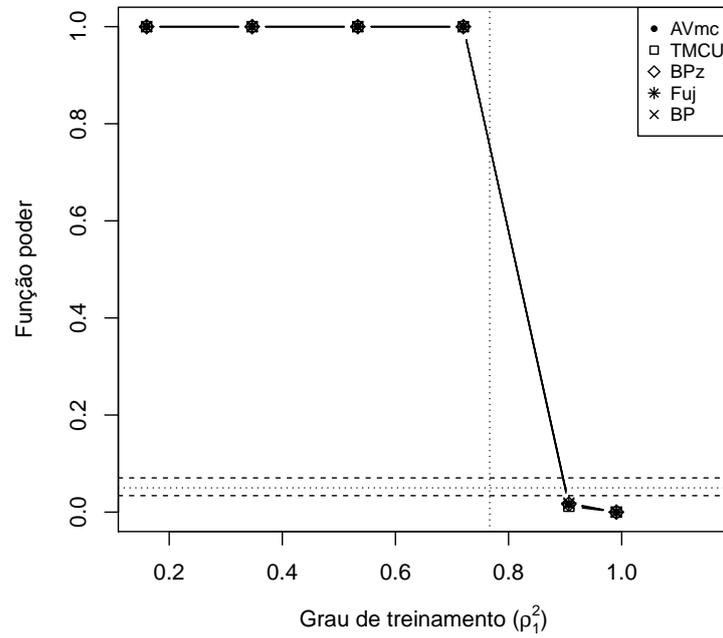


Figura 47 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 5$ atributos, $n = 20$ produtos e $p = 15$ julgadores.

Fonte: Da autora.

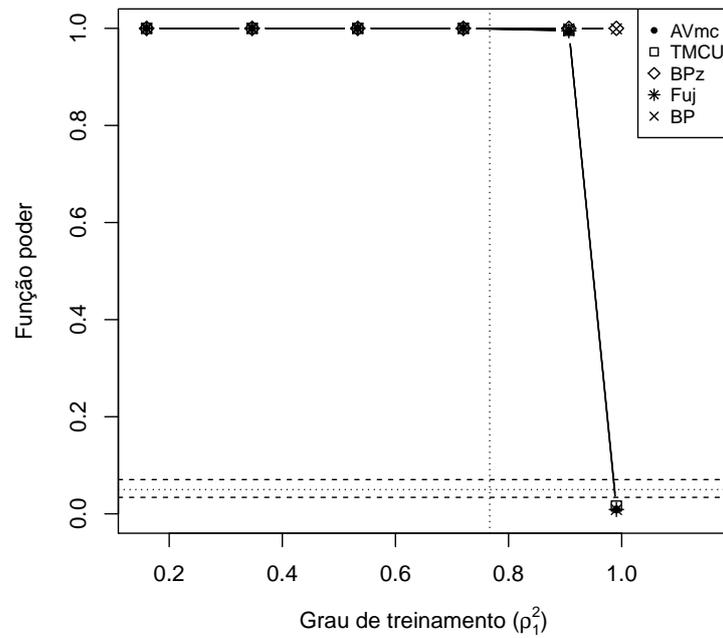


Figura 48 – Poder dos testes AVmc, Fuj, TMCU, Bp e Bpz, considerando $q = 10$ atributos, $n = 20$ produtos e $p = 15$ julgadores.

Fonte: Da autora.