

**UNIVERSIDADE FEDERAL DE ALFENAS**

**Tatiane Gomes de Araújo**

**Testes Monte Carlo para autovalores na verificação de  
consenso multivariado de painéis sensoriais**

ALFENAS, MG

2019

**TATIANE GOMES DE ARAÚJO**

**Testes Monte Carlo para autovalores na verificação de consenso  
multivariado de painéis sensoriais**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, Área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre.

Linha de Pesquisa: Estatística Aplicada e Biometria.

Orientador: Prof. Dr. Eric Batista Ferreira.

ALFENAS, MG

2019

Dados Internacionais de Catalogação-na-Publicação (CIP)  
Sistema de Bibliotecas da Universidade Federal de Alfenas

Araújo, Tatiane Gomes de.  
A663t Testes Monte Carlo para autovalores na verificação de consenso  
multivariado de painéis sensoriais / Tatiane Gomes de Araújo – Alfenas/MG,  
2019.  
86 f.: il. --  
  
Orientador: Eric Batista Ferreira .  
Dissertação (Mestrado em Estatística Aplicada e Biometria) -  
Universidade Federal de Alfenas, 2019.  
Bibliografia.  
  
1. Sensometria. 2. Concordância de painéis. 3. Teste Monte Carlo. I.  
Ferreira, Eric Batista. II. Título.

CDD-519



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 37130-001  
Fone: (35) 3701-9600  
<https://www.unifal-mg.edu.br/ppgeab/>



TATIANE GOMES DE ARAÚJO

**“TESTES MONTE CARLO PARA AUTOVALORES NA VERIFICAÇÃO DE CONSENSO  
MULTIVARIADO 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.  
Área de Concentração: Estatística Aplicada e  
Biometria

Aprovado em: 27 de fevereiro de 2019.

Prof. Dr. Eric Batista Ferreira

Instituição: UNIFAL-MG

Assinatura: Eric Batista Ferreira

Profa. Dra. Flávia Della Lúcia

Instituição: UNIFAL-MG

Assinatura: Flávia Della Lúcia

Profa. Dra. Adriana Dias

Instituição: UNIFAL-MG

Assinatura: Adriana Dias

*Dedico a Deus, aos meus  
pais e irmãos.*

## AGRADECIMENTOS

Agradeço à Deus que me proporcionou esta oportunidade de dar continuidade aos meus estudos e por ter sido minha força em todo tempo.

Aos meus pais, Samuel e Eunice, por estarem sempre ao meu lado me incentivando e sendo meu porto seguro. Aos meus irmãos, Gracielle e Guilherme, por dividirem minhas angústias e alegrias e por sempre me motivarem.

Às amigas que conviveram comigo durante este período, Camila, Érika, Ingrid, Loren, Stephany, Suellen e Sula, por fazerem deste tempo um marco na minha vida, por deixarem tudo mais alegre e divertido, pela amizade e companheirismo.

À Cássia e Luiz Otávio, sem os quais não teria concluído este trabalho. Obrigada por todo suporte, apoio, reflexões, risadas, parceria e conselhos.

À minha turma que deixou o curso mais leve, Luiz Henrique, Maria Caruline, Michele, Renan e Taís, obrigada pela troca de conhecimento, parceria, ajuda, apoio, incentivo, carinho e amizade.

Ao meu orientador, Prof. Dr. Eric Batista Ferreira, sem o qual não concluiria este trabalho, por toda paciência, conselhos, orientações, dedicação, apoio, incentivo, aprendizado, por sempre acreditar no meu potencial e pela parceria para desenvolver esta pesquisa.

Aos membros da banca dos exames de qualificação e defesa de mestrado, Profa. Dra. Adriana Dias, Prof. Dr. Davi Butturi Gomes e Profa. Dra. Flávia Della Lúcia, pelas sugestões e contribuições para o desenvolvimento deste trabalho.

À Universidade Federal de Alfenas e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, juntamente com seus docentes e servidores, por todas as oportunidades de aprendizado oferecidas. Em especial, Prof. Dr. Denismar Alves Nogueira, Prof. Dr. Fabricio Goecking Avelar, Prof. Dr. Luiz Alberto Beijo e Profa. Dra. Patrícia De Siqueira Ramos, pelos conhecimentos adquiridos e à Martha por toda paciência e prestatividade.

Por fim, agradeço a todos que contribuíram, direta ou indiretamente, para realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## RESUMO

Em análise sensorial a concordância entre os membros de um painel é fundamental para verificar a eficiência do treinamento por ele recebido. Para essa finalidade, encontram-se na literatura alguns métodos e testes, dentre os quais destacam-se os testes assintóticos de Ferreira (2017), que avaliam a consonância de um painel de forma multivariada. Por serem testes assintóticos, eles necessitam de um grande número de observações amostrais, o que não é comum no contexto sensorial. Diante disso, o objetivo deste trabalho foi propor versões Monte Carlo dos testes de Ferreira (2017) para obter melhores características de poder e erro tipo I em qualquer tamanho amostral, bem como compará-los com os testes assintóticos via simulação computacional. Também foi objetivo, adaptar o protocolo de verificação de desempenho de painéis proposto por Tomic et al. (2009), inserindo uma etapa de teste de consenso multivariado. Os quatro testes Monte Carlo propostos foram considerados exatos e apresentaram maior poder em 95,24% dos cenários analisados, quando comparados com as versões assintóticas. Recomenda-se o uso do teste  $InvH_2mc$ . A eficiência deste teste pôde ser comprovada ao verificar o desempenho de um painel sensorial ao avaliar duas marcas de refrigerante sabor guaraná em suas versões tradicionais e zero açúcar. Verificou-se que o painel não apresentou consenso multivariado, ou seja, os provadores não pontuaram os atributos de forma semelhante. A inserção do teste Monte Carlo no protocolo permitiu analisar de forma inferencial o comportamento do painel, corroborando com as fases exploratórias.

**Palavras-Chave:** Sensometria. Concordância de painéis. Teste Monte Carlo.

## ABSTRACT

In sensory analysis, the concordance between the members of a panel is fundamental to verify the efficiency of the training received by it. To this purpose, some methods and tests are found in the literature, among which are the asymptotic tests of Ferreira (2017), which evaluate the consonance of a panel in a multivariate form. Because they are asymptotic tests, they require a large number of sample observations, which is not common in the sensory context. Therefore, the objective of this work was to propose Monte Carlo versions of the Ferreira tests (2017) to obtain better characteristics of power and type I error in any sample size, as well as to compare them with the asymptotic tests via computerized simulation. It was also an objective to adapt the panel performance verification protocol proposed by Tomic et al. (2009), inserting a multivariate consensus test step. The four Monte Carlo tests proposed were considered accurate and presented higher power in 95,24% of the analyzed scenarios, when compared with the asymptotic versions. The use of the  $InvH_2mc$  test is recommended. The efficiency of this test could be proven by verifying the performance of a sensory panel when evaluating two brands of guarana soda flavor in both their traditional and zero sugar versions. It was found that the panel did not present a multivariate consensus, that is, the tasters did not score the attributes similarly. The insertion of the Monte Carlo test in the protocol allowed inferential analysis of the panel's behavior, corroborating the exploratory phases.

**Keywords:** Sensory analysis. Panel concordance. Monte Carlo Method.

## LISTA DE FIGURAS

Figura 1 –	Esquema para a proposição dos testes. . . . .	18
Figura 2 –	Diagrama: proposição dos testes Monte Carlo. . . . .	30
Figura 3 –	Poder dos testes $IAH_1$ , $IAH_2$ , $InvH_1$ , $InvH_2$ , $IAH_1mc$ , $IAH_2mc$ , $InvH_1mc$ , $InvH_2mc$ para o caso $p = 5$ provedores, $q = 5$ atributos, $n = 10$ observações. . . . .	33
Figura 4 –	Poder dos testes $IAH_1$ , $IAH_2$ , $InvH_1$ , $InvH_2$ , $IAH_1mc$ , $IAH_2mc$ , $InvH_1mc$ , $InvH_2mc$ para o caso $p = 5$ provedores, $q = 5$ atributos, $n = 50$ observações. . . . .	34
Figura 5 –	Poder dos testes $IAH_1$ , $IAH_2$ , $InvH_1$ , $InvH_2$ , $IAH_1mc$ , $IAH_2mc$ , $InvH_1mc$ , $InvH_2mc$ para o caso $p = 10$ provedores, $q = 5$ atributos, $n = 10$ observações. . . . .	35
Figura 6 –	Poder dos testes $IAH_1$ , $IAH_2$ , $InvH_1$ , $InvH_2$ , $IAH_1mc$ , $IAH_2mc$ , $InvH_1mc$ , $InvH_2mc$ para o caso $p = 10$ provedores, $q = 5$ atributos, $n = 50$ observações. . . . .	35
Figura 7 –	Poder dos testes $IAH_1$ , $IAH_2$ , $InvH_1$ , $InvH_2$ , $IAH_1mc$ , $IAH_2mc$ , $InvH_1mc$ , $InvH_2mc$ para o caso $p = 16$ provedores, $q = 4$ atributos, $n = 12$ observações. . . . .	48
Figura 8 –	Fluxo de trabalho proposto para a analisar o desempenho do provedor e do painel. . . . .	49
Figura 9 –	Resultado da ANAVA para verificação de seleção de atributos. . . . .	50
Figura 10 –	Gráficos Tucker-1 para cada atributo analisado. . . . .	51
Figura 11 –	Gráficos Manhattam para cada atributo analisado. . . . .	51
Figura 12 –	Gráfico MSE (a) e gráfico F (b) do painel para verificar repetibilidade e discriminação. . . . .	53
Figura 13 –	$P = 2$ provedores, $q = 2$ atributos, $n = 10$ observações. . . . .	57
Figura 14 –	$P = 2$ provedores, $q = 5$ atributos, $n = 10$ observações. . . . .	57
Figura 15 –	$P = 2$ provedores, $q = 2$ atributos, $n = 20$ observações. . . . .	58
Figura 16 –	$P = 2$ provedores, $q = 5$ atributos, $n = 20$ observações. . . . .	58
Figura 17 –	$P = 2$ provedores, $q = 10$ atributos, $n = 20$ observações. . . . .	59
Figura 18 –	$P = 2$ provedores, $q = 2$ atributos, $n = 30$ observações. . . . .	59

Figura 19 – P = 2 provedores, q = 5 atributos, n = 30 observações. . . . .	60
Figura 20 – P = 2 provedores, q = 10 atributos, n = 30 observações. . . . .	60
Figura 21 – P = 2 provedores, q = 2 atributos, n = 40 observações. . . . .	61
Figura 22 – P = 2 provedores, q = 5 atributos, n = 40 observações. . . . .	61
Figura 23 – P = 2 provedores, q = 10 atributos, n = 40 observações. . . . .	62
Figura 24 – P = 2 provedores, q = 20 atributos, n = 40 observações. . . . .	62
Figura 25 – P = 2 provedores, q = 2 atributos, n = 50 observações. . . . .	63
Figura 26 – P = 2 provedores, q = 5 atributos, n = 50 observações. . . . .	63
Figura 27 – P = 2 provedores, q = 10 atributos, n = 50 observações. . . . .	64
Figura 28 – P = 2 provedores, q = 20 atributos, n = 50 observações. . . . .	64
Figura 29 – P = 2 provedores, q = 2 atributos, n = 100 observações. . . . .	65
Figura 30 – P = 2 provedores, q = 5 atributos, n = 100 observações. . . . .	65
Figura 31 – P = 2 provedores, q = 10 atributos, n = 100 observações. . . . .	66
Figura 32 – P = 2 provedores, q = 20 atributos, n = 100 observações. . . . .	66
Figura 33 – P = 2 provedores, q = 2 atributos, n = 200 observações. . . . .	67
Figura 34 – P = 2 provedores, q = 5 atributos, n = 200 observações. . . . .	67
Figura 35 – P = 2 provedores, q = 10 atributos, n = 200 observações. . . . .	68
Figura 36 – P = 2 provedores, q = 20 atributos, n = 200 observações. . . . .	68
Figura 37 – P = 5 provedores, q = 2 atributos, n = 10 observações. . . . .	69
Figura 38 – P = 5 provedores, q = 2 atributos, n = 20 observações. . . . .	69
Figura 39 – P = 5 provedores, q = 2 atributos, n = 30 observações. . . . .	70
Figura 40 – P = 5 provedores, q = 5 atributos, n = 30 observações. . . . .	70
Figura 41 – P = 5 provedores, q = 2 atributos, n = 40 observações. . . . .	71
Figura 42 – P = 5 provedores, q = 5 atributos, n = 40 observações. . . . .	71
Figura 43 – P = 5 provedores, q = 2 atributos, n = 50 observações. . . . .	72
Figura 44 – P = 5 provedores, q = 10 atributos, n = 50 observações. . . . .	72
Figura 45 – P = 5 provedores, q = 2 atributos, n = 100 observações. . . . .	73
Figura 46 – P = 5 provedores, q = 5 atributos, n = 100 observações. . . . .	73
Figura 47 – P = 5 provedores, q = 10 atributos, n = 100 observações. . . . .	74
Figura 48 – P = 5 provedores, q = 20 atributos, n = 100 observações. . . . .	74
Figura 49 – P = 5 provedores, q = 2 atributos, n = 200 observações. . . . .	75
Figura 50 – P = 5 provedores, q = 5 atributos, n = 200 observações. . . . .	75

Figura 51 – P = 5 provadores, q = 10 atributos, n = 200 observações. . . . .	76
Figura 52 – P = 5 provadores, q = 20 atributos, n = 200 observações. . . . .	76
Figura 53 – P = 10 provadores, q = 2 atributos, n = 20 observações. . . . .	77
Figura 54 – P = 10 provadores, q = 2 atributos, n = 30 observações. . . . .	77
Figura 55 – P = 10 provadores, q = 2 atributos, n = 40 observações. . . . .	78
Figura 56 – P = 10 provadores, q = 2 atributos, n = 50 observações. . . . .	78
Figura 57 – P = 10 provadores, q = 2 atributos, n = 100 observações. . . . .	79
Figura 58 – P = 10 provadores, q = 5 atributos, n = 100 observações. . . . .	79
Figura 59 – P = 10 provadores, q = 10 atributos, n = 100 observações. . . . .	80
Figura 60 – P = 10 provadores, q = 2 atributos, n = 200 observações. . . . .	80
Figura 61 – P = 10 provadores, q = 5 atributos, n = 200 observações. . . . .	81
Figura 62 – P = 10 provadores, q = 10 atributos, n = 200 observações. . . . .	81
Figura 63 – P = 10 provadores, q = 20 atributos, n = 200 observações. . . . .	82
Figura 64 – P = 15 provadores, q = 2 atributos, n = 30 observações. . . . .	82
Figura 65 – P = 15 provadores, q = 2 atributos, n = 40 observações. . . . .	83
Figura 66 – P = 15 provadores, q = 2 atributos, n = 50 observações. . . . .	83
Figura 67 – P = 15 provadores, q = 2 atributos, n = 100 observações. . . . .	84
Figura 68 – P = 15 provadores, q = 5 atributos, n = 100 observações. . . . .	84
Figura 69 – P = 15 provadores, q = 2 atributos, n = 200 observações. . . . .	85
Figura 70 – P = 15 provadores, q = 5 atributos, n = 200 observações. . . . .	85
Figura 71 – P = 15 provadores, q = 10 atributos, n = 200 observações. . . . .	86

## SUMÁRIO

	<b>CAPÍTULO 1 - INTRODUÇÃO</b>	11
1	<b>REFERENCIAL TEÓRICO</b>	12
1.1	ANÁLISE SENSORIAL E SENSOMETRIA	12
1.2	AVALIAÇÃO COMPUTACIONAL DE TESTES	21
	<b>REFERÊNCIAS</b>	23
	<b>CAPÍTULO 2 - PROPOSIÇÃO E AVALIAÇÃO DE TESTES MONTE CARLO PARA CONSENSO MULTIVARIADO</b>	25
	<b>RESUMO</b>	25
1	<b>INTRODUÇÃO</b>	26
2	<b>METODOLOGIA</b>	28
2.1	PROPOSIÇÃO DOS TESTES	28
2.2	AVALIAÇÃO E COMPARAÇÃO DOS TESTES	32
3	<b>RESULTADOS E DISCUSSÃO</b>	33
4	<b>CONCLUSÃO</b>	37
	<b>REFERÊNCIAS</b>	39
	<b>CAPÍTULO 3 - TESTE MONTE CARLO NA VERIFICAÇÃO DO CONSENSO DO PAINEL DE BEBIDAS SABOR GUARANÁ</b>	41
	<b>RESUMO</b>	41
1	<b>INTRODUÇÃO</b>	42
2	<b>MATERIAL E MÉTODOS</b>	43
2.1	DADOS	43
2.2	MÉTODOS	44
2.2.1	Análise de variância com Modelo Misto	44
2.2.2	Tucker-1	45
2.2.3	Gráficos de Manhattam	46
2.2.4	Gráficos F	46
2.2.5	Gráficos MSE	46
2.2.6	Testes para verificar consenso multivariado de painéis sensoriais	47
3	<b>RESULTADOS</b>	49
4	<b>CONCLUSÃO</b>	54
	<b>REFERÊNCIAS</b>	55
	<b>CAPÍTULO 4 - CONSIDERAÇÕES FINAIS</b>	56
	<b>APÊNDICE</b>	57

## INTRODUÇÃO

Visando atender às exigências do mercado consumidor há uma constante preocupação quanto à melhoria e o desenvolvimento de produtos. Para isso, a análise sensorial tem sido amplamente utilizada em indústrias de alimentos, bebidas, celulares, cosméticos, automóveis, etc, possibilitando a mensuração e análise da qualidade sensorial e de sua aceitação de produtos, medidas essas feitas por meio dos sentidos humanos (tato, olfato, audição, paladar e visão).

Para quantificar as respostas dadas aos estímulos sensoriais percebidos, a análise sensorial utiliza-se da Sensometria, que é a área da estatística responsável por avaliar métodos e testes, tais como os testes descritivos e discriminativos. Os testes descritivos descrevem produtos por meio de um conjunto de atributos sensoriais percebidos pelo provador e os discriminativos elencam quais são os atributos sensoriais e quais as diferenças entre eles. Ambos são aplicados a provadores e são de interesse deste trabalho.

Para realizar a análise sensorial, um painel formado por provadores é constituído, treinado e seu desempenho é medido avaliando-se a capacidade deles em identificar pequenas diferenças entre produtos, replicar as suas notas para um mesmo produto e quantificando o grau de concordância entre eles. Se o painel foi bem treinado, os provadores pontuam um mesmo produto de forma semelhante, ou seja, concordam em sua avaliação e avaliam em conformidade com o treinamento recebido.

Neste sentido, avaliar a concordância do painel é de fundamental importância, pois por meio dela pode-se inferir sobre a eficiência do treinamento recebido. Para isso, alguns métodos foram propostos na literatura, destacando-se os testes assintóticos apresentados por Ferreira (2017), com a finalidade de avaliar a consonância do painel de forma multivariada.

Por serem testes assintóticos, eles necessitam de um tamanho amostral grande (muitos produtos diferentes ou muitas repetições de um mesmo produto), o que não é comum no contexto sensorial. Dessa forma, este trabalho tem por objetivo propor uma versão Monte Carlo destes testes com melhores características de poder e erro tipo I e compará-los com os testes assintóticos via simulação computacional. Além disso, adaptar o protocolo de verificação de desempenho de painéis proposto por Tomic et al. (2009), utilizando os dados de um painel sensorial semi-treinado ao avaliar a bebida sabor guaraná.

## 1 REFERENCIAL TEÓRICO

### 1.1 ANÁLISE SENSORIAL E SENSOMETRIA

A análise sensorial é "utilizada para evocar, medir, analisar e interpretar reações às características de alimentos e outros materiais da forma como são percebidos pelos sentidos da visão, olfato, gosto, tato e audição"(MINIM, 2013). Ela pode ser utilizada em diversos segmentos da indústria, principalmente os de alimentos e bebidas, como um suporte técnico para pesquisa, industrialização, marketing e controle de qualidade dos itens produzidos, qualificando decisões técnicas e administrativas (ROSSINI; ANZANELLO; FOGLIATTO, 2011).

A análise sensorial caracteriza-se por identificar os atributos ou propriedades de interesse na qualidade sensorial, selecionar o método mais adequado para quantificar os estímulos recebidos pelos sentidos humanos e aplicar os métodos estatísticos mais adequados para interpretar os resultados (MINIM, 2013).

Para analisar estes estímulos, a Sensometria é a área da estatística responsável por lidar com o caráter subjetivo das sensações. Ela ainda fornece métodos que permitem inferir sobre o grau de treinamento dos painéis sensoriais, que está diretamente relacionado com a confiabilidade dos resultados das análises (ROCHA, 2015).

Neste contexto, denomina-se o ser humano como provador ou julgador, sendo o conjunto de provadores um painel. O desempenho de um painel e de provadores individuais compreende a capacidade de detectar, identificar, medir e caracterizar os atributos de maneira semelhante a outros painéis e entre provadores. Além de discriminar os estímulos usando uma escala adequada, repetir seus próprios resultados e reproduzir os resultados de outros painéis e outros provadores (ROGERS, 2018).

Para a credibilidade dos resultados da análise sensorial, o treinamento dos provadores é fundamental, já que a concordância dos provadores influencia na credibilidade dos resultados da análise. Na literatura encontram-se diversos estudos que buscam quantificar a qualidade do treinamento verificando a concordância de painéis sensoriais.

Dijksterhuis (1995) propôs um método univariado baseado na análise de componentes principais (ACP) para avaliar a concordância de um painel em relação a um determinado atributo. Se os provadores foram bem treinados eles pontuarão da mesma forma o atributo avali-

ado, e, então, espera-se que a maior parte da variância seja explicada pelo primeiro componente principal. Quando isso ocorre o painel é considerado unidimensional, ou seja, os provadores concordam em relação a este atributo.

Já Latreille et al. (2006) propuseram um método baseado em modelos lineares mistos para medir a confiabilidade do desempenho do painel ao longo de várias sessões. Também verificou-se o desempenho global do painel e o desempenho individual de cada provador em termos de discriminação, repetibilidade e concordância com o painel. O modelo proposto é dado por:

$$Y_{ijl} = \delta + \alpha_j + s_i + b_l + a_{ji} + c_{jl} + d_{il} + \varepsilon_{ijl},$$

em que  $Y_{ijl}$  é a nota dada a um atributo pelo provador  $i$  para o produto  $j$  na sessão  $l$ ;  $\delta$  é o intercepto;  $\alpha_j$  é o efeito fixo do produto;  $s_i$  é o efeito aleatório do provador;  $b_l$  é o efeito aleatório da sessão;  $a_{ji}$  é o efeito aleatório de produto  $\times$  provador;  $c_{jl}$  é o efeito aleatório de produto  $\times$  sessão;  $d_{il}$  é o efeito aleatório de provador  $\times$  sessão e  $\varepsilon_{ijl}$  é o erro aleatório, onde  $\varepsilon_{ijl} \sim N(0, \sigma^2)$ .

Neste modelo, a concordância do provador com o painel pode ser testada utilizando contrastes da interação produto  $\times$  provador. Este método é indicado quando o número de produtos ou sessões é pequeno. Aumentado-se o número de produtos ou sessões, os autores indicaram o uso de modelos mais simples ou o uso de métodos multivariados.

Amorim et al. (2009) desenvolveram o teste Monte Carlo para unidimensionalidade (TMCU) para medir o desempenho do painel sensorial quanto à sua unidimensionalidade, ou seja, verificando o quanto da variabilidade é explicada pelo primeiro componente principal para um atributo ( $q = 1$ ). Para realizá-lo é necessário impor a hipótese nula ( $H_0$ ) considerando que a proporção da variância explicada pelo primeiro componente principal seja muito próxima de 100%, mas que nunca atinja 100%. Tal imposição é necessária, pois se os painéis fossem compostos por provadores idênticos, não haveria variância, o que não faz sentido para estes testes.

Assim, considerando que o painel é unidimensional,  $C_{\rho_n} \geq C_{\rho_0}$ , onde  $C_{\rho_n}$  é a proporção da variância explicada pelo primeiro componente principal e  $C_{\rho_0}$  é um valor arbitrário entre 0% e 100% escolhido previamente pelo pesquisador. A relação entre  $C_{\rho_n}$  e o coeficiente de correlação ( $\rho_n$ ), pode ser dada por:

$$C_{\rho_n} = \frac{100[1 + (p - 1)\rho_n]}{p}\%,$$

em que  $\rho_n$  é a correlação entre os  $p$  provadores e  $p$  é o número total de provadores. Portanto,  $\rho_n \leq C_{\rho_n} \leq 1$ , e quando  $N \rightarrow \infty$  então  $\rho_n = C_{\rho_n}$ .

Para compor os painéis virtuais, foram geradas sob a hipótese nula,  $N = 1000$  amostras Monte Carlo obtidas de uma distribuição normal  $p$ -variada equicorrelacionada com parâmetros  $\boldsymbol{\mu}$  e  $\boldsymbol{\Sigma}$ , ou seja,  $X \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ , em que  $X$  é um vetor  $p \times 1$  que contém as notas dos  $p$  provadores para um produto e  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$ , sendo

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_n & \cdots & \rho_n \\ \rho_n & 1 & \cdots & \rho_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_n & \rho_n & \cdots & 1 \end{bmatrix},$$

com  $\sigma^2 = 1$ , sem perda de generalidade e  $\rho_n$  é o coeficiente de correlação entre as notas dos  $n$  produtos atribuídas por dois provadores.

O nível de significância  $\alpha$  é previamente fixado pelo pesquisador e a porcentagem da variância explicada pelo primeiro componente principal de cada uma das  $N$  amostras simuladas e para a amostra de interesse são computadas e comparadas com os quantis da distribuição empírica sob a hipótese nula ( $C_\alpha$ ). A hipótese nula será rejeitada se  $\hat{C}_{\rho_n} < C_\alpha$ .

Os autores observaram que o teste Monte Carlo para unidimensionalidade (TMCU) é eficiente e poderoso, sendo mais indicado quando o número de provadores e produtos são grandes (Amorim et al., 2009).

Outra abordagem possível é a análise de variância (ANAVA), que permite investigar diferenças individuais e comparar, para cada provador, as diferenças entre as médias do produto com uma medida do erro aleatório com base nas diferenças entre as replicações sensoriais. O modelo a ser considerado é dado por: (NAES; BROCKHOFF; TOMIC, 2010)

$$Y_{ijk r} = \mu_{ik} + \beta_{ijk} + \varepsilon_{ijk r},$$

em que para cada combinação de provador  $i$  e atributo  $k$ ,  $\beta$  representa os efeitos do produto,  $j$  é o número de produtos avaliados,  $r$  é o número de replicações dos produtos e  $\varepsilon$  representa o erro aleatório. A análise deste modelo consiste no cálculo da Razão F, que é uma razão da soma dos quadrados das diferenças entre os valores médios dos produtos e a variância das replicações. Os valores  $F$  são computados para cada provador e atributo e são comparados com os valores

críticos tabelados (NAES; BROCKHOFF; TOMIC, 2010).

Em 2008, Ferreira propôs o teste de autovalores assintótico (AVa) para inferir sobre autovalores populacionais, baseado nas propriedades assintóticas demonstradas por Girshick (1939), Lawley (1956) e Anderson (1963).

Para compreendê-lo considere  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  uma amostra aleatória oriunda de uma população normal multivariada com vetor de média  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ , estimadas respectivamente, por  $\bar{\mathbf{X}}$  e  $\mathbf{S}$ . Seja ainda os autovalores  $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_p$  e autovetores  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ , da matriz de covariâncias  $\mathbf{S}$ . Considere também as decomposições espectrais:  $\mathbf{S} = \hat{\mathbf{P}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}'$  e  $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ , em que  $\hat{\mathbf{P}}$  é uma matriz composta pelos autovetores de  $\mathbf{S}$  em suas colunas e  $\hat{\boldsymbol{\Lambda}}$  é uma matriz diagonal que contém os autovalores de  $\mathbf{S}$ .  $\mathbf{P}$  é uma matriz cujas colunas contém os autovetores de  $\boldsymbol{\Sigma}$  e  $\boldsymbol{\Lambda}$  é uma matriz diagonal composta pelos autovalores de  $\boldsymbol{\Sigma}$ .

De acordo com Ferreira (2008), considerando que os componentes principais possuem as seguintes propriedades assintóticas demonstradas por Girshick (1939), Lawley (1956) e Anderson (1963), tem-se:

- a)  $\sqrt{n-1}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \sim N_p(\mathbf{0}, 2\boldsymbol{\Lambda}^2)$ , em que  $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p]'$  e  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_p]'$  são os vetores diagonais de  $\hat{\boldsymbol{\Lambda}}$  e  $\boldsymbol{\Lambda}$ , respectivamente;
- b)  $\sqrt{n-1}(\hat{e}_i - e_i) \sim N_{p-1}(\mathbf{0}, \boldsymbol{\Upsilon}_i)$ , em que

$$\begin{aligned} \boldsymbol{\Upsilon}_i &= \lambda_i \sum_{k=1}^p \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \mathbf{e}_k \mathbf{e}_k' \\ &= \lambda_i \mathbf{P}_i \boldsymbol{\Gamma}_i \mathbf{P}_i', \end{aligned}$$

onde  $\mathbf{P}_i = [\mathbf{e}_i, \dots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \dots, \mathbf{e}_p]$  é uma matriz de dimensão  $p \times (p-1)$  cujas colunas são compostas pelos autovetores de  $\boldsymbol{\Sigma}$ , exceto pelo  $i$ -ésimo autovetor. E a matriz diagonal  $\boldsymbol{\Gamma}_i$ , de dimensão  $(p-1) \times (p-1)$ , é dada por:

$$\boldsymbol{\Gamma}_i = \begin{bmatrix} \frac{\lambda_1}{(\lambda_1 - \lambda_i)^2} & 0 & \dots & 0 \\ 0 & \frac{\lambda_2}{(\lambda_2 - \lambda_i)^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\lambda_p}{(\lambda_p - \lambda_i)^2} \end{bmatrix},$$

desconsiderando-se a  $i$ -ésima linha e a  $i$ -ésima coluna correspondente ao autovalor  $\lambda_i$  da matriz

de covariâncias  $\Sigma$ .

c) A covariância entre dois autovetores amostrais distintos é dada pela matriz:

$$Cov(\hat{e}_i, \hat{e}_j) = \Upsilon_{ij} = \begin{bmatrix} -\frac{\lambda_i \lambda_i e_{i1} e_{j1}}{(n-1)(\lambda_i - \lambda_j)^2} & \cdots & -\frac{\lambda_i \lambda_i e_{i1} e_{jp}}{(n-1)(\lambda_i - \lambda_j)^2} \\ -\frac{\lambda_i \lambda_i e_{i2} e_{j1}}{(n-1)(\lambda_i - \lambda_j)^2} & \cdots & -\frac{\lambda_i \lambda_i e_{i2} e_{jp}}{(n-1)(\lambda_i - \lambda_j)^2} \\ \vdots & \ddots & \vdots \\ -\frac{\lambda_i \lambda_i e_{ip} e_{j1}}{(n-1)(\lambda_i - \lambda_j)^2} & \cdots & -\frac{\lambda_i \lambda_i e_{ip} e_{jp}}{(n-1)(\lambda_i - \lambda_j)^2} \end{bmatrix}.$$

d) Os elementos de  $\hat{\lambda}$  são assintoticamente independentes dos elementos de  $\hat{P}$ .

Então, pela propriedade (a), conclui-se que  $\hat{\lambda}_i \sim N\left(\lambda_i, \lambda_i^2 \left(\frac{2}{n-1}\right)\right)$ .

Assim, o par de hipóteses do teste é dado por

$$\begin{cases} H_0 : \lambda_i = \lambda_{i0} \\ H_1 : \lambda_i \neq \lambda_{i0} \end{cases}. \quad (1.1)$$

E a estatística de teste é

$$Z = \frac{\hat{\lambda}_i - \lambda_{i0}}{\lambda_{i0} \sqrt{\frac{2}{n-1}}} \sim N(0,1),$$

em que  $\lambda_{i0}$  é o valor sob hipótese para o parâmetro  $\lambda_i$ .

Fernandes (2012) propôs um teste Monte Carlo para o primeiro autovalor, modificando o teste proposto por Ferreira (2008). Para isso, estabeleceu-se o seguinte par de hipóteses:

$$\begin{cases} H_0 : \lambda_2 = \lambda_3 = \cdots = \lambda_p = 0 \\ H_1 : \text{ao menos um } \lambda_i, i > 1, \text{ é diferente de } 0 \end{cases}.$$

Pode-se reescrever a hipótese nula considerando sob  $H_0$ , que toda a variação está contida no primeiro componente principal (primeiro autovalor). Assim,  $H_0 : \lambda_1 = \lambda_{1,0}$ , onde  $\lambda_{1,0} = \sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i = \lambda_1$ . Logo, a estatística de teste é dada por

$$Z_c = \frac{\hat{\lambda}_1 - \lambda_{1,0}}{\lambda_{1,0} \sqrt{\frac{2}{n-1}}}.$$

Como os parâmetros são desconhecidos,  $\lambda_1$  pode ser estimado por  $\hat{\lambda}_1 = \sum_{i=1}^p \hat{\lambda}_i$ . Então,

$$Z_c = \frac{\hat{\lambda}_1 - \sum_{i=1}^p \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i \sqrt{\frac{2}{n-1}}} = \frac{-\sum_{i=2}^p \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i \sqrt{\frac{2}{n-1}}}. \quad (1.2)$$

Para realizar o teste sob  $H_0$ , são geradas  $B = 1000$  amostras de uma variável aleatória  $X_{p \times 1} \sim N(\mathbf{0}, \Sigma)$ , em que  $X_{p \times 1}$  são vetores com as notas dos  $n$  produtos avaliados por  $p$  provadores. Para cada uma das  $B$  amostras simuladas e para os dados observados, efetua-se a análise de componentes principais, determina-se o seu primeiro autovalor ( $\hat{\lambda}_{1l}$  e  $\hat{\lambda}_1$ ) e calcula-se a estatística de teste 1.2.

Fixando-se um valor nominal de significância  $\alpha$ , a hipótese  $H_0$  de unidimensionalidade será rejeitada se o valor- $p < \alpha$ , sendo

$$\text{valor-}p = \frac{\sum_{l=1}^B I(Z_{cl} \leq Z_c)}{B}.$$

Rocha (2015) propôs generalizações dos seguintes testes: o teste Monte Carlo para unidimensionalidade proposto por Amorim et al. (2009), dois testes bootstrap paramétricos propostos por Gebert (2010), o teste de Fujikoshi proposto por Ferreira (2008) e o teste sobre autovalores Monte Carlo proposto por Fernandes (2012).

Tais generalizações foram propostas com o intuito de inferir sobre a consonância do painel de forma multivariada, ou seja, avaliando todos os atributos simultaneamente. Para isso, no algoritmo dos testes, a matriz de covariância amostral foi substituída pela matriz de correlação canônica. Para testar o consenso multivariado de painéis sensoriais, a autora recomendou o teste sobre autovalores Monte Carlo, pois este apresentou desempenho igual ou superior aos demais testes.

Com o intuito de verificar a concordância de provadores treinados para vários atributos avaliados simultaneamente, Ferreira (2017) propôs generalizações do teste de autovalores de Ferreira (2008). O processo de obtenção dos quatro testes é descrito a seguir.

Seja  $\Omega$  um subconjunto do conjunto de índices  $I = \{1, 2, \dots, pq\}$  ( $\Omega \subseteq I$ ). Seja ainda a variável aleatória definida por  $Y_\Omega = \sum_{\Omega} \hat{\lambda}_i$ , tem-se que  $\hat{\lambda}_i \sim N\left(\lambda_i, \lambda_i^2 \left(\frac{2}{n-1}\right)\right)$ .

A partir da definição do par de hipóteses geral e da padronização de  $Y$  a seguir, foi possível realizar a generalização do teste assintótico de autovalores sob o par de hipóteses

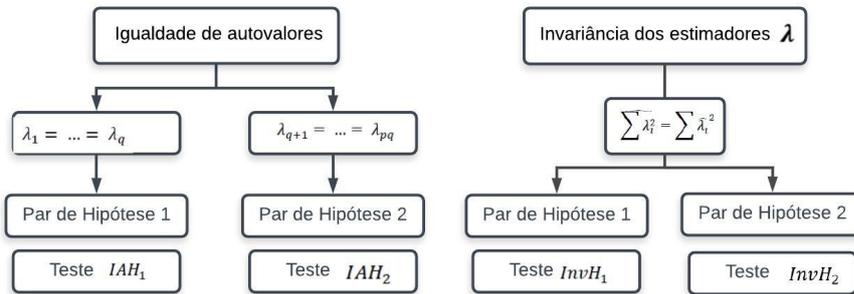
$$\begin{cases} H_0 : \text{ o painel apresenta consenso multivariado} \\ H_1 : \text{ o painel não apresenta consenso multivariado} \end{cases} \quad (1.3)$$

Com a estatística de teste dada por:

$$Z = \frac{Y - E[Y]}{\sigma_y} = \frac{Y - \sum \lambda_i}{\Omega} = \frac{Y - \sum \lambda_i}{\sqrt{\frac{2}{n-1} \sum \lambda_i^2}} \sim N(0,1). \quad (1.4)$$

Os quatro testes obtidos partem de duas premissas: a igualdade dos autovalores  $\lambda$  e a propriedade de invariância dos estimadores de  $\lambda$ . O esquema de proposição dos testes está ilustrado na Figura 1:

Figura 1 – Esquema para a proposição dos testes.



Fonte: Adaptado de Ferreira (2017, p.35).

Cada teste foi avaliado sob os seguintes pares de hipóteses:

$$\begin{cases} H_0 : \sum_{i=1}^q \lambda_i \geq \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 tr(\Sigma) \\ H_1 : \sum_{i=1}^q \lambda_i < \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 tr(\Sigma) \end{cases}$$

e

$$\begin{cases} H_0 : \sum_{i=q+1}^{pq} \lambda_i \leq (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) tr(\Sigma) \\ H_1 : \sum_{i=q+1}^{pq} \lambda_i > (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) tr(\Sigma) \end{cases}$$

sendo  $p$  o número de provedores,  $q$  o número de atributos,  $\rho_0$  é a proporção fixada pelo pesqui-

sador e  $\Sigma$  é a matriz de covariância da combinação entre provedores e atributos. Assumindo que  $\Sigma = R$ , então  $tr(\Sigma) = tr(R) = pq$ . Então, os pares de hipóteses serão dados por:

$$\left\{ \begin{array}{l} H_0 : \sum_{i=1}^q \lambda_i \geq \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 pq \\ H_1 : \sum_{i=1}^q \lambda_i < \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 pq \end{array} \right. \quad (1.5)$$

e

$$\left\{ \begin{array}{l} H_0 : \sum_{i=q+1}^{pq} \lambda_i \leq (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) pq \\ H_1 : \sum_{i=q+1}^{pq} \lambda_i > (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) pq \end{array} \right. \quad (1.6)$$

Tem-se ainda que se  $\lambda_r = \dots = \lambda_s$ , em que os índices  $r$  e  $s \in \Omega$ , então  $\sum_{\Omega} \lambda_i^2 = \frac{\left(\sum_{\Omega} \lambda_i\right)^2}{l}$ , em que  $l$  é a cardinalidade do conjunto  $\Omega$  ( $|\Omega| = l$ ). A partir desse resultado e do que foi disposto acima, foi possível obter a estatística de teste geral, que é dada por:

$$Z = \frac{\sum_{\Omega} \hat{\lambda}_i - \sum_{\Omega} \lambda_i}{\sum_{\Omega} \lambda_i \sqrt{\frac{2}{l(n-1)}}} \sim N(0,1). \quad (1.7)$$

Observe que, se  $\Omega = \{1, \dots, q\}$  então é obtida a estatística de teste para  $IAH_1$ , pois este teste tem como premissa a igualdade dos  $q$  primeiros autovalores. Se  $\Omega = \{q+1, q+2, \dots, pq\}$ , obtém-se a estatística para  $IAH_2$ , por assumir como premissa a igualdade dos últimos autovalores. O teste de hipótese  $IAH_1$  verifica se a variabilidade dos dados é explicada pelos  $q$  primeiros componentes principais. O par de hipótese é dado por 1.5 e a estatística de teste geral sob  $H_0$  é dada por:

$$Z_1 = \frac{\sum_{i=1}^q \hat{\lambda}_i - \rho_0 pq}{\rho_0 pq \sqrt{\frac{2}{q(n-1)}}} \sim N(0,1). \quad (1.8)$$

A premissa assumida pelo teste  $IAH_2$  é a de que a variabilidade dos dados é explicada

pelos  $pq$  últimos autovalores. Seu par de hipótese é dado por 1.6 e sua estatística de teste é dada por:

$$Z_2 = \frac{\sum_{i=q+1}^{pq} \hat{\lambda}_i - (1 - \rho_0)pq}{(1 - \rho_0)pq \sqrt{\frac{2}{q(p-1)(n-1)}}} \sim N(0,1). \quad (1.9)$$

Os testes  $InvH_1$  e  $InvH_2$  tem, respectivamente, os mesmos pares de hipóteses de  $IAH_1$  e  $IAH_2$  e consistem na seguinte relação:  $\sum \lambda_i^2 = \sum \hat{\lambda}_i^2$ . Assim, a estatística do teste  $InvH_1$  é definida por:

$$Z_3 = \frac{\sum_{i=1}^q \hat{\lambda}_i - \rho_0 pq}{\sqrt{\frac{2}{(n-1)}} \sum_{i=1}^q \hat{\lambda}_i^2} \sim N(0,1). \quad (1.10)$$

E para o teste  $InvH_2$  tem-se a estatística de teste em (3.9).

$$Z_4 = \frac{\sum_{i=q+1}^{pq} \hat{\lambda}_i - (1 - \rho_0)pq}{\sqrt{\frac{2}{(n-1)}} \sum_{i=q+1}^{pq} \hat{\lambda}_i^2} \sim N(0,1). \quad (1.11)$$

De acordo com Ferreira (2017), os testes apresentaram comportamento satisfatório quando o número de observações é grande, sendo  $InvH_2$  o mais indicado. Quando o número de provedores é pequeno ( $p = 2$ ) combinado com diferentes valores de atributos ( $q$ ) e observações ( $n$ ), a autora observou que não houve rejeição da hipótese nula mesmo quando a hipótese era falsa e o poder dos testes foi zero. Tais resultados foram obtidos pelo fato de os testes serem assintóticos e necessitarem de um grande número de observações para apresentar um comportamento satisfatório. Sendo esta uma limitação em experimentos sensoriais.

## 1.2 AVALIAÇÃO COMPUTACIONAL DE TESTES

A avaliação computacional de testes prevê o estabelecimento de cenários formados por combinações dos parâmetros de interesse. No contexto deste trabalho são combinados número de provedores ( $p$ ), atributos ( $q$ ), produtos ( $n$ ) e grau de treinamento ( $\rho^2$ ). Em cada cenário formado os testes foram avaliados de forma exaustiva, via processo Monte Carlo.

O método Monte Carlo caracteriza-se por realizar inferências utilizando-se de um componente aleatório de alguma distribuição probabilística associado ao uso de mecanismos numéricos computacionais para gerar valores desta distribuição. Ele é útil para avaliar as propriedades de um estimador, solucionar problemas estatísticos computacionalmente intensivos e avaliar as propriedades de um teste estatístico recém-proposto. Definições mais aprofundadas podem ser vistas em Ferreira (2013), que descreve computacionalmente tal método.

Nestas avaliações é comum analisar as taxas de erro tipo I e poder do teste simultaneamente em uma curva de poder empírica, qualificar o teste, verificando seu funcionamento em diversos pontos dos cenários estabelecidos como também comparar e avaliar testes candidatos. Neste trabalho, o erro tipo I significa concluir que há pelo menos um provedor mal treinado, quando na verdade o painel está bem treinado. Avaliações sobre a robustez dos testes também podem ser realizadas, porém isso não foi abordado neste trabalho. Mais detalhes sobre erro tipo I, poder e robustez podem ser verificadas em Mood, Graybill e Boes (1974).

## REFERÊNCIAS

- AMORIM, I. S. 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, 2009.
- ANDERSON, T. W. Asymptotic theory for principal components. **Annals of Mathematical Statistics**, Ann Arbor, v. 34, p. 122-148, 1963.
- DIJKSTERHUIS, G. Assessing panel consonance. **Food Quality and Preference**, v. 6, p. 7-14, 1995. Disponível em: <[https://doi.org/10.1016/0950-3293\(94\)P4207-M](https://doi.org/10.1016/0950-3293(94)P4207-M)>. Acesso em: 30 ago. 2018.
- FERNANDES, F. M. O. **Proposta de um teste Monte Carlo para unidimensionalidade de painéis sensoriais**. 2012. 80 f. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2012.
- FERREIRA, D. F. **Estatística multivariada**. Lavras: Editora UFLA, 2008. 662 p.
- FERREIRA, D. F. **Estatística computacional em Java**. Lavras: Editora UFLA, 2013. 695 p.
- FERREIRA, L. B. S. **Testes assintóticos para detectar consenso multivariado em painéis sensoriais**. 80 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, 2017.
- GEBERT, D. M. P. **Proposta de testes bootstrap para inferir sobre o número de componentes principais retidos**. 2010. 75 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)- Universidade Federal de Lavras, Lavras, 2010.
- GIRSHICK, M. A. On the sampling theory of roots of detrimental equations. **The Annals of Mathematical Statistics**, v. 10, p. 203-224, 1939.
- LATREILLE, J. et al. Measurement of the reliability of sensory panel performances. **Food Quality and Preference**, Barking, n. 17, p. 369-365, 2006. Disponível em: <<https://doi.org/10.1016/j.foodqual.2005.04.010>>. Acesso em: 12 Set. 2018.
- LAWLEY, D. N. Tests of significance of the latent roots and correlation matrices. **Biometrika**, London, v. 43, p. 128-136, 1956.
- MINIM, V. P. R. **Análise sensorial: estudos com consumidores**. 3. ed. Viçosa: UFV, 2013. 332 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3.ed. Tokyo: McGraw-Hill, 1974, 564p.
- NAES, T.; BROCKHOFF PER B.; TOMIC, O. **Statistics for sensory and consumer science**. West Sussex, Reino Unido, 2010. 282 p.
- ROCHA, M. C. **Testes baseados em correlação canônica para avaliar consonância de**

**painéis sensoriais.** 71 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, 2015.

ROSSINI, K.; ANZANELLO, M. J.; FOGLIATTO, F. S. Seleção de atributos em avaliações sensoriais descritivas. **Produção**, v. 22, n. 3, p. 380-390. 2011. Disponível em: <<http://dx.doi.org/10.1590/S1413-24782012000200011>>. Acesso em: 08 ago. 2018.

ROGERS, L. **Sensory panel management: a practical handbook for recruitment, training and performance.** Duxford: Elsevier Ltd., 2018. 376 p.

TOMIC, O. et al. Analysing sensory panel performance in a proficiency test using the PanelCheck software. **Eur Food Res Technol**, Dinamarca, v. 230, p. 497-511. 2009. Disponível em: <<https://doi.org/10.1007/s00217-009-1185-y>>. Acesso em: 05 dez 2018.

## CAPÍTULO 2 - PROPOSIÇÃO E AVALIAÇÃO DE TESTES MONTE CARLO PARA CONSENSO MULTIVARIADO

Tatiane Gomes de Araújo; Eric Batista Ferreira

**RESUMO:** Para avaliar atributos específicos de produtos, em geral alimentos, a formação de painéis sensoriais é útil, pois através das notas por eles atribuídas pode-se medir o desempenho do painel, que está diretamente relacionado com a confiabilidade dos resultados da análise sensorial. Para medir o desempenho de painéis pode-se utilizar diversos testes propostos na literatura. No entanto, a maioria destes testes avaliam apenas um atributo por vez ou necessitam de um grande número de observações para apresentarem um resultado satisfatório. Para contornar estas limitações, foram propostos quatro testes Monte Carlo para autovalores na verificação de consenso multivariado, os quais foram avaliados em diversos cenários via simulação Monte Carlo e foram computadas as taxas de erro tipo I e poder do teste. Os testes apresentaram comportamento semelhante, foram considerados exatos e controlaram as taxas de erro tipo I perto da região de rejeição da hipótese nula e são indicados independente do tamanho da amostra. E, por serem equivalentes entre si, indica-se o uso do teste *InvH<sub>2</sub>mc*.

**Palavras-Chave:** Sensometria. Computação intensiva. Análise sensorial. Checagem de painel.

# PROPOSAL AND EVALUATION OF MONTE CARLO TEST FOR MULTIVARIATE CONSENSUS

Tatiane Gomes de Araújo; Eric Batista Ferreira

**ABSTRACT:** To evaluate specific attributes of a product, in general food, sensory panels are usually formed. By means of the scores they give, one can not only evaluate foods but also measure the performance of the panel itself, which is directly related to the reliability of the sensory analysis results. An important dimension of panel performance is consensus or agreement, but most available tests evaluate only one attribute at a time or require a large number of observations. To circumvent these limitations, four Monte Carlo eigenvalues tests were proposed for the verification of multivariate consensus of sensory panels. They were evaluated in several scenarios, via Monte Carlo simulation, through their type I error rates and power. The tests showed similar behaviour, controlled the type I error rates and presented high power and could be indicated regardless of the sample size. Among them, the  $InvH_2mc$  test is recommended.

**Keywords:** Sensometry. Intensive computing. Sensory analysis. Panel check.

## 1 INTRODUÇÃO

Para avaliar os atributos específicos de produtos, em geral alimentos, a análise sensorial utiliza-se de um painel sensorial, cujas notas tem que ser críveis e os provadores tem que ter a capacidade de identificar pequenas diferenças entre produtos (discriminabilidade), replicar as mesmas notas para um mesmo produto em diferentes ocasiões (repetibilidade) e concordar com os demais provadores sobre a sensação sensorial dos atributos avaliados (consonância). Estes critérios são utilizados também para medir o desempenho do painel, que se bem treinado, fornece informações precisas e semelhantes sobre os produtos avaliados (VERHOEF; HUIJBERTS; VAESSEN, 2014).

O desempenho do painel está diretamente relacionado com a confiabilidade dos resultados da análise sensorial. Assim, é de fundamental importância o desenvolvimento de técnicas que permitem avaliar o grau de concordância entre provadores para validar os resultados da avaliação sensorial. Para isto, a Sensometria fornece métodos estatísticos úteis para verificar a qualidade do treinamento de painéis sensoriais (ROCHA, 2015).

A confiabilidade de um painel é outro aspecto importante a ser avaliado. Bi (2003) propõe analisá-la usando dois indicadores baseados na teoria da confiabilidade clássica: um indicador de concordância que descreve o consenso de classificações para um mesmo produto e um indicador de confiabilidade que estuda a capacidade de discriminação do painel. Já Latreille et al. (2006) propuseram um método baseado em modelos lineares mistos para medir a confiabilidade do desempenho global do painel e o desempenho individual de cada provador.

Para avaliar de forma univariada e descritiva a concordância de um painel em relação a um determinado atributo, Dijkstra (1995) propôs uma estatística baseada na análise de componentes principais (ACP). De acordo com o autor, se o painel for bem treinado seus provadores pontuarão os atributos avaliados de forma semelhante e, assim, a maior parte da variância será explicada pelo primeiro componente principal e o painel será considerado unidimensional, ou seja, os provadores concordam em relação a este atributo. Inspirado nesta ideia, Amorim et al. (2009) propôs o teste Monte Carlo para unidimensionalidade (TMCU) para verificar a unidimensionalidade de um painel sensorial ao avaliar qualquer produto, avaliando somente um atributo ou variável de cada vez.

Outro teste para verificar a unidimensionalidade de painéis sensoriais foi proposto por Fernandes (2012). A autora traz para o contexto sensorial o teste assintótico sobre autovalores

(Ava) para inferir sobre autovalores populacionais de Ferreira (2008), baseado nas propriedades assintóticas dos componentes principais demonstradas por Girshick (1939), Lawley (1956) e Anderson (1963). Fernandes (2012) modificou o teste Ava sugerindo uma versão Monte Carlo para o primeiro autovalor (teste AVmc), permitindo que o teste seja aplicado em amostras pequenas. Há ainda, testes que permitem inferir sobre a concordância do painel de forma multivariada, ou seja, avaliando todos os atributos simultaneamente. Rocha (2015) propôs a generalização de cinco testes. Tal generalização foi feita substituindo no algoritmo dos testes a matriz de covariância amostral pela matriz de correlação canônica. Os testes modificados foram: o teste Monte Carlo para unidimensionalidade proposto por Amorim et al. (2009), dois testes bootstrap paramétricos propostos por Gebert (2010), o teste de Fujikoshi proposto por Ferreira (2008) e o teste sobre autovalores Monte Carlo proposto por Fernandes (2012).

Com este mesmo intuito, Ferreira (2017) propôs quatro testes que são generalizações do teste de autovalores assintóticos (Ava): teste  $IAH_1$  que considera a igualdade dos  $q$  primeiros autovalores; teste  $IAH_2$  que considera a igualdade dos  $pq$  últimos autovalores; teste  $InvH_1$  e teste  $InvH_2$  baseados na propriedade de invariância dos estimadores dos autovalores. Estes testes apresentaram comportamento satisfatório quando o número de observações é grande, sendo o teste  $InvH_2$  o mais indicado. Tais resultados foram obtidos pelo fato de os testes serem assintóticos e necessitarem de um grande número de observações (muitos produtos diferentes ou repetições de um mesmo produto) para apresentar um comportamento satisfatório, sendo esta uma limitação em experimentos sensoriais.

Portanto, com o intuito de melhorar as características dos testes assintóticos para detectar o consenso multivariado em painéis sensoriais, este trabalho tem como objetivo propor uma versão Monte Carlo destes testes, a fim de melhorar seu desempenho em amostras pequenas. Além disso, analisar as características de poder e erro tipo I, comparando-as com os testes assintóticos via simulação computacional.

## 2 METODOLOGIA

### 2.1 PROPOSIÇÃO DOS TESTES

Os testes propostos neste trabalho são versões Monte Carlo dos testes assintóticos para detectar consenso multivariado em painéis sensoriais de Ferreira (2017). A hipótese nula ( $H_0$ ) é de que o painel apresenta consenso multivariado. As versões Monte Carlo possuem o mesmo par de hipóteses e as mesmas estatísticas de teste dos testes assintóticos e serão definidas a seguir.

Assuma que os testes partem de duas premissas: a igualdade dos autovalores de  $\lambda$  e a propriedade de invariância dos estimadores de  $\lambda$ . Além disso, eles podem ser abordados de duas maneiras: ou considerando os  $q$  primeiros autovalores de  $\lambda$  (par de hipóteses 1) ou considerando os  $pq - q$  últimos autovalores de  $\lambda$  (par de hipóteses 2). Cada par de hipóteses é dado, respectivamente, por:

$$\left\{ \begin{array}{l} H_0 : \sum_{i=1}^q \lambda_i \geq \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 pq \\ H_1 : \sum_{i=1}^q \lambda_i < \rho_0 \sum_{i=1}^{pq} \lambda_i = \rho_0 pq \end{array} \right. \quad (2.1)$$

e

$$\left\{ \begin{array}{l} H_0 : \sum_{i=q+1}^{pq} \lambda_i \leq (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) pq \\ H_1 : \sum_{i=q+1}^{pq} \lambda_i > (1 - \rho_0) \sum_{i=1}^{pq} \lambda_i = (1 - \rho_0) pq \end{array} \right. , \quad (2.2)$$

sendo  $p$  o número de provedores,  $q$  o número de atributos,  $\rho_0$  a proporção fixada pelo pesquisador e  $pq$  o traço da matriz de correlação da combinação entre provedores e atributos, assumindo que  $\Sigma = \mathbf{R}$ , então  $tr(\Sigma) = tr(\mathbf{R}) = pq$ .

O par de hipóteses 1 baseado na igualdade dos  $q$  primeiros autovalores, gera o teste assintótico  $IAH_1$  e sua versão Monte Carlo  $IAH_{1mc}$ . Estes testes verificam se a variabilidade dos dados é explicada pelos  $q$  primeiros componentes principais. O par de hipóteses é dado por

2.1 e sua estatística de teste sob  $H_0$  é dada por:

$$Z_1 = \frac{\sum_{i=1}^q \hat{\lambda}_i - \rho_0 pq}{\rho_0 pq \sqrt{\frac{2}{q(n-1)}}} \sim N(0,1). \quad (2.3)$$

Considerando-se ainda a igualdade dos autovalores e assumindo o par de hipóteses 2 ( $\lambda_{q+1} = \dots = \lambda_{pq}$ ), obtém-se o teste assintótico  $IAH_2$  e o teste Monte Carlo  $IAH_2mc$ . Eles assumem como premissa que a variabilidade dos dados é explicada pelos  $pq$  últimos autovalores. Seu par de hipóteses é dado por 2.2 e a estatística de teste é dada por:

$$Z_2 = \frac{\sum_{i=q+1}^{pq} \hat{\lambda}_i - (1 - \rho_0) pq}{(1 - \rho_0) pq \sqrt{\frac{2}{q(p-1)(n-1)}}} \sim N(0,1). \quad (2.4)$$

Tem-se ainda que o par de hipóteses 1 considerando a propriedade de invariância dos autovalores, gera o teste assintótico  $InvH_1$  e o teste Monte Carlo  $InvH_1mc$ , cujo par de hipóteses é dado por 2.1 e sua estatística de teste é definida por:

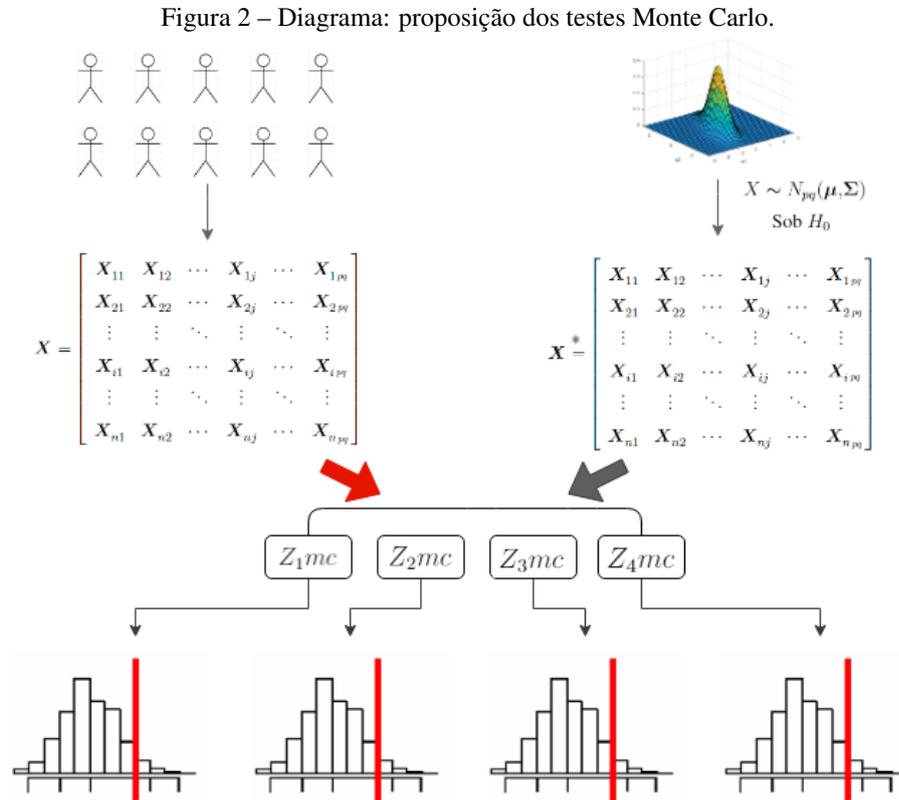
$$Z_3 = \frac{\sum_{i=1}^q \hat{\lambda}_i - \rho_0 pq}{\sqrt{\frac{2}{(n-1)} \sum_{i=1}^q \hat{\lambda}_i^2}} \sim N(0,1). \quad (2.5)$$

Por fim, ao considerar o par de hipóteses 2 e a propriedade de invariância dos autovalores, obtém-se o teste assintótico  $InvH_2$  e sua versão Monte Carlo  $InvH_2mcc$ . O par de hipóteses destes testes é definido por 2.2 e sua estatística de teste é dada por:

$$Z_4 = \frac{\sum_{i=q+1}^{pq} \hat{\lambda}_i - (1 - \rho_0) pq}{\sqrt{\frac{2}{(n-1)} \sum_{i=q+1}^{pq} \hat{\lambda}_i^2}} \sim N(0,1). \quad (2.6)$$

As versões Monte Carlo foram propostas com o intuito de eliminar o caráter assintótico dos testes de Ferreira (2017), a ideia principal é tomar decisão na distribuição empírica das estatísticas. Para isso são geradas  $M = 1000$  amostras de uma distribuição normal sob  $H_0$  e

calculadas as estatísticas de teste tanto para os valores observados quanto para cada uma das  $M$  amostras simuladas. A um nível  $\alpha$  de significância, calcula-se o valor- $p$  e  $H_0$  é rejeitada caso valor- $p < \alpha$ . O diagrama da Figura 2 ilustra como os testes Monte Carlo foram propostos.



Fonte: da autora.

Para uma melhor compreensão dos testes Monte Carlo propostos, construiu-se o seguinte algoritmo:

- i. Estabelecer as hipóteses e as estatísticas de teste.
- ii. Em uma matriz  $X_{n \times pq}$ , dispor as notas obtidas pelo painel sensorial, sendo que nas linhas estão as observações de cada produto e nas colunas as notas dos atributos dadas por todos os provedores.
- iii. Sob  $H_0$ , gerar  $M = 1000$  amostras aleatórias de uma distribuição normal  $pq$ -variada com parâmetros  $\mu = 0$  e  $\Sigma_{pq} = \sigma^2 R = \sigma^2(I \otimes P)$ :

$$\Sigma_{pq} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

em que é fixado  $\sigma^2 = 1$ , sem perda de generalidade, e  $\rho$  é o grau de treinamento previamente definido ( $\rho^2$ ).

iv. Computar as estatísticas de cada teste,  $Z_{1mc}$ ,  $Z_{2mc}$ ,  $Z_{3mc}$  e  $Z_{4mc}$ , tanto para os dados provenientes do painel, quanto para as amostras simuladas, e para estas, ordenar os valores da estatística de teste obtidos. v. Fixar um valor nominal de significância  $\alpha$ . Neste trabalho assumiu-se  $\alpha = 5\%$ . vi. Calcular o valor- $p$  como:

$$\text{valor} - p = \frac{\sum_{l=1}^M I(Z_{li} \leq Z_i)}{M}, \quad (2.7)$$

em que  $i = 1, \dots, 4$ , ou seja, se  $i = 1$  a estatística de teste corresponde ao teste Monte Carlo  $IAH_1$ , se  $i = 2$  a estatística de teste corresponde ao teste Monte Carlo  $IAH_2$ , se  $i = 3$  a estatística de teste se refere ao teste Monte Carlo  $InvH_1$  e se  $i = 4$  a estatística de teste se refere ao teste Monte Carlo  $InvH_2$ . Tem-se ainda que,  $Z_i$  é a estatística de teste amostral e  $Z_{li}$  é a estatística de teste amostral computada para cada uma das  $M = 1000$  amostras Monte Carlo.

vii. Rejeitar a hipótese  $H_0$  de que o painel apresenta consenso multivariado se  $\text{valor-}p < \alpha$ .

A matriz de covariâncias bloco diagonal foi estabelecida de maneira que se pudessem fixar as correlações entre atributos, entre provedores e de atributos e provedores de forma precisa. Foram fixadas em zero a correlação entre atributos e de atributos  $\times$  provedores para que somente a correlação entre provedores fosse avaliada (essa foi fixada em  $\rho$ ). Fixou-se a correlação hipotética em  $\rho_0 = 0,70$ , ou seja, entende-se que, se após o treinamento os provedores possuírem no mínimo 70% de similaridade ou concordância em suas notas, o treinamento pode ser considerado bem sucedido.

## 2.2 AVALIAÇÃO E COMPARAÇÃO DOS TESTES

A avaliação do desempenho e comparação dos testes originais e propostos foram feitas via simulação Monte Carlo, por meio da taxa de erro tipo I e poder. Foram considerados 63 cenários, resultantes da combinação de diferentes números de provadores  $p \in \{2, 5, 10, 15\}$ , número de atributos  $q \in \{2, 5, 10, 20\}$ , número de observações  $n \in \{10, 20, 30, 40, 50, 100, 200\}$  e da proporção da variação total explicada pelos primeiros  $q$  componentes principais ou o grau de treinamento  $\rho^2 \in \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 0,99\}$ , sendo  $n \geq pq$ , para que existam graus de liberdade suficientes para estimar os  $pq$  autovalores dos testes assintóticos.

Cada amostra aleatória foi gerada a partir de uma distribuição normal  $pq$ -variada com parâmetros  $\mu = 0$  e  $\Sigma_{pq}$ , definida anteriormente. A correlação hipotética foi fixada em  $\rho_0 = 0,70$  e as proporções de rejeição da hipótese nula ao longo das  $N = 1000$  simulações de Monte Carlo foram computadas para o nível nominal de significância  $\alpha = 5\%$ . Essa proporção sob  $H_0$  representa a taxa de erro tipo I e sob  $H_1$  representa o poder do teste. A função poder é estimada por:

$$FP = \frac{\sum_{l=1}^N I(\text{valor-}p_l \leq \alpha)}{N}, \quad (2.8)$$

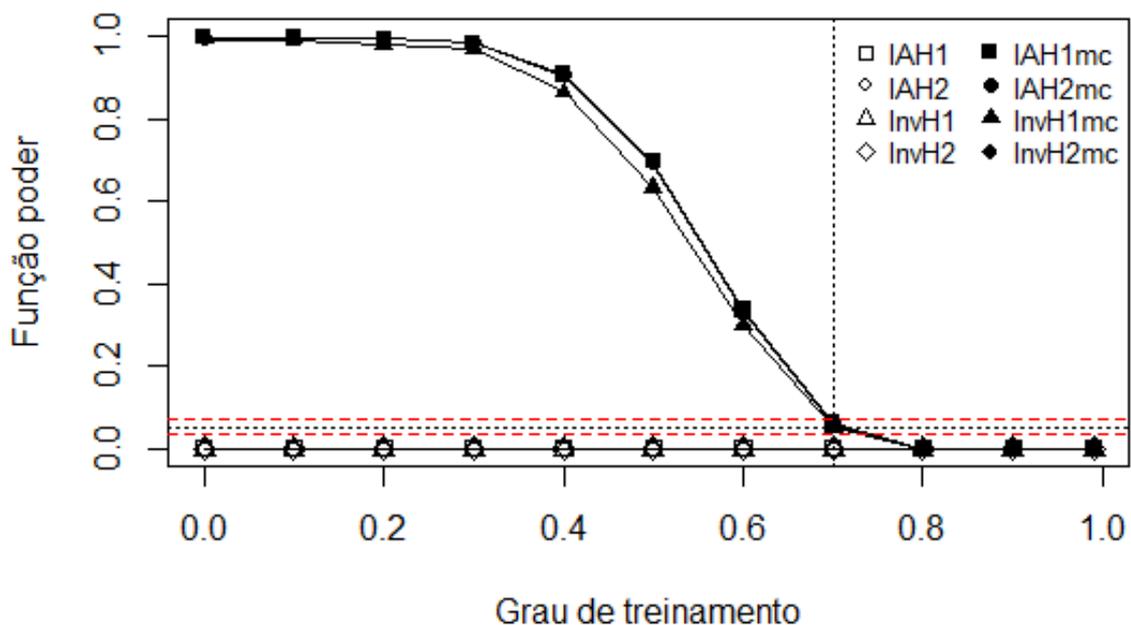
em que  $\text{valor-}p_l$  é o valor- $p$  da  $l$ -ésima amostra Monte Carlo simulada e  $I$  é a função indicadora que retorna 1 se a condição for verdadeira ou 0 se for falsa na  $l$ -ésima simulação. Se a amostra gerada for proveniente de uma população na qual  $\rho < \rho_0$ , então a FP é considerada uma estimativa de poder e se  $\rho \geq \rho_0$  FP será considerada uma estimativa do erro tipo I. Por fim, determinou-se um intervalo binomial exato com 99% de confiança para o nível de significância  $\alpha$ ,  $[0,033927; 0,070504]$ , para verificar se a taxa de erro tipo I praticada era igual ao nível nominal de significância adotado, conforme FERREIRA (2008). A programação dos testes, bem como o estudo de simulação foram desenvolvidos no software R (R CORE TEAM, 2019).

### 3 RESULTADOS E DISCUSSÃO

As Figuras 3 a 6 apresentam a proporção de rejeição de  $H_0$  dos oito testes comparados (quatro assintóticos e suas versões Monte Carlo). Essas proporções de rejeição sob  $H_0$  representam o poder na região sob  $H_1$  e representam o erro tipo I no *sup* de  $FP$  na região sob  $H_0$ . Estas regiões são separadas pela linha pontilhada paralela ao eixo ordenado no ponto em que o grau de treinamento corresponde à correlação entre provedores do painel sensorial. O nível de significância adotado,  $\alpha = 5\%$ , é representado pela linha paralela à abscissa e seu intervalo de confiança é representado pelas duas linhas vermelhas horizontais.

Avaliando o caso em que  $p = 5$  provedores,  $q = 5$  atributos e  $n = 10$  observações, tem-se que os testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$  e  $InvH_2$  não rejeitam a hipótese nula, mesmo quando  $H_0$  é nitidamente falsa e o poder dos testes é zero. Já os testes  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  são considerados exatos, apresentam poder semelhante, 69,5%, e são os mais indicados neste cenário (Figura 3).

Figura 3 – Poder dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  para o caso  $p = 5$  provedores,  $q = 5$  atributos,  $n = 10$  observações.



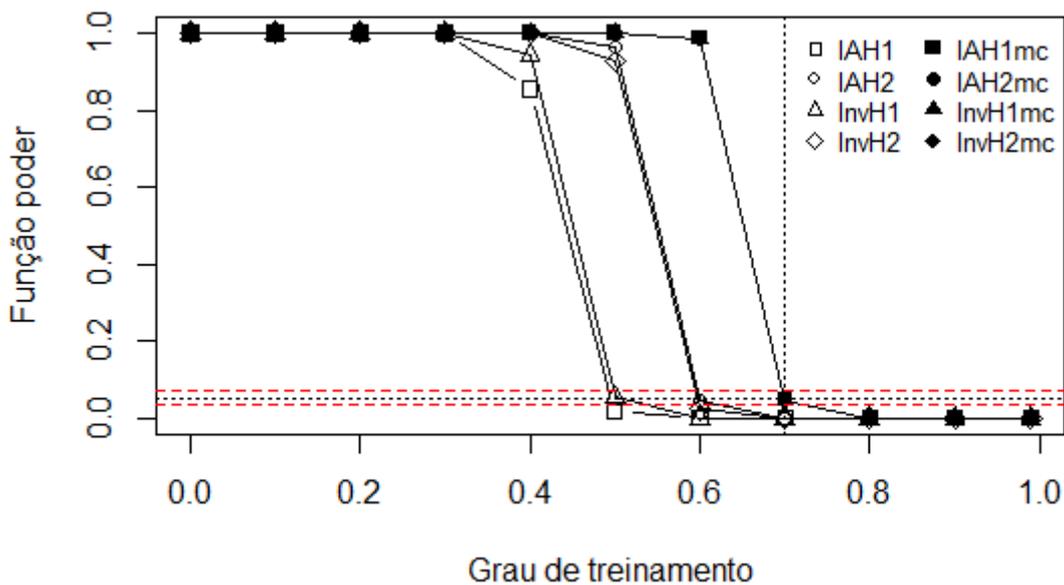
Fonte: da autora.

Mantendo fixo o número de atributos e provedores e aumentando o número de observações para  $n = 50$ , pode-se observar na Figura 4 que há uma queda na curva de poder dos testes assintóticos  $IAH_1$  e  $InvH_1$  no grau de treinamento  $\rho_0 = 0,50$ . Os testes  $IAH_2$  e  $InvH_2$  apre-

sentam alto poder (96,4% e 96,9%, respectivamente), porém eles são conservadores. O poder dos quatro testes Monte Carlo aumenta (98,5%), eles continuam exatos e são os mais indicados neste cenário.

Considerando  $p = 10$  provadores,  $q = 5$  atributos e  $n = 10$  observações, verifica-se na Figura 5, que os testes assintóticos apresentam poder nulo e os testes  $IAH_1mc$ ,  $IAH_2mc$  e  $InvH_2mc$  apresentam poder igual a 89% e são os mais indicados neste cenário.

Figura 4 – Poder dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  para o caso  $p = 5$  provadores,  $q = 5$  atributos,  $n = 50$  observações.



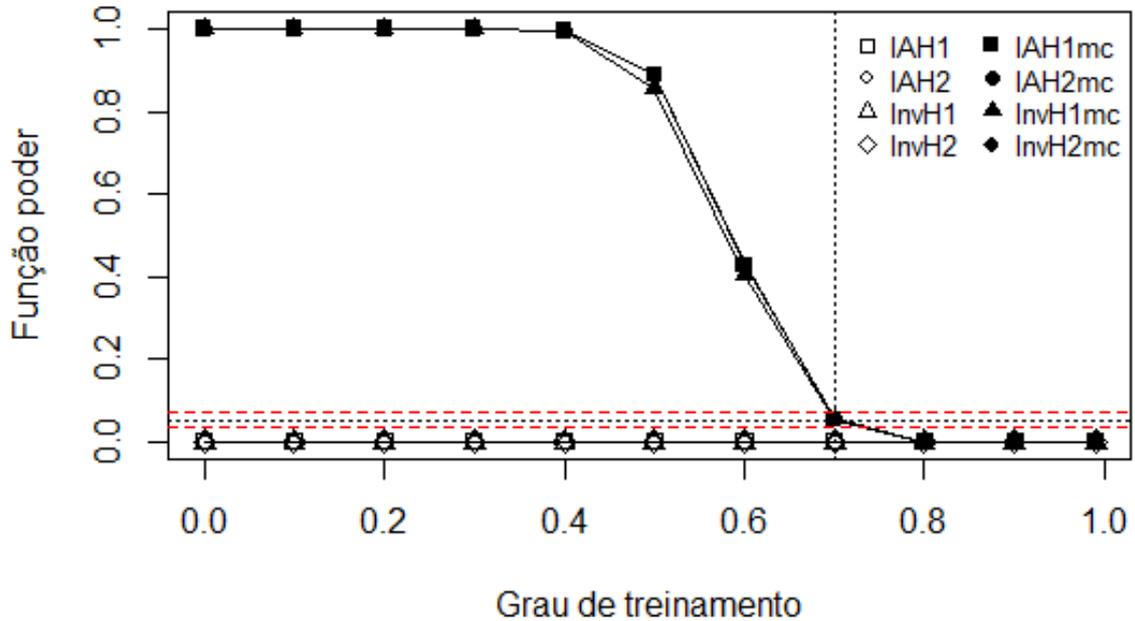
Fonte: da autora.

Para o caso em que aumenta-se o número de provadores para 10 ( $p = 10$ ) e mantém-se constante os valores de  $q = 5$  atributos e de  $n = 50$  observações (Figura 6), verifica-se que os testes assintóticos  $IAH_1$  e  $IAH_2$  apresentam menor poder que os demais testes e há uma queda na curva de poder no grau de treinamento  $\rho_0 = 0,60$ . O teste  $IAH_2$  apresenta alto poder (78,6%), porém ele é superado por todos os testes Monte Carlo (99,4%), que são os mais indicados também neste cenário.

Para as situações em que o número de provadores é  $p = 5$  e o número de observações é pequeno, Fernandes (2012) indicou o uso do teste Monte Carlo para unidimensionalidade (TMCU) que apresenta alto poder, porém não controla a taxa de erro tipo I próximo ao valor nominal de significância. Rocha (2015) recomendou o uso do teste Fujikoshi e Ferreira (2017) indica o teste  $IAH_2$ , mesmo ele sendo considerado liberal. Aumentando-se o número de atributos e mantendo fixos os números de observações e provadores, Ferreira (2017) indicou o teste

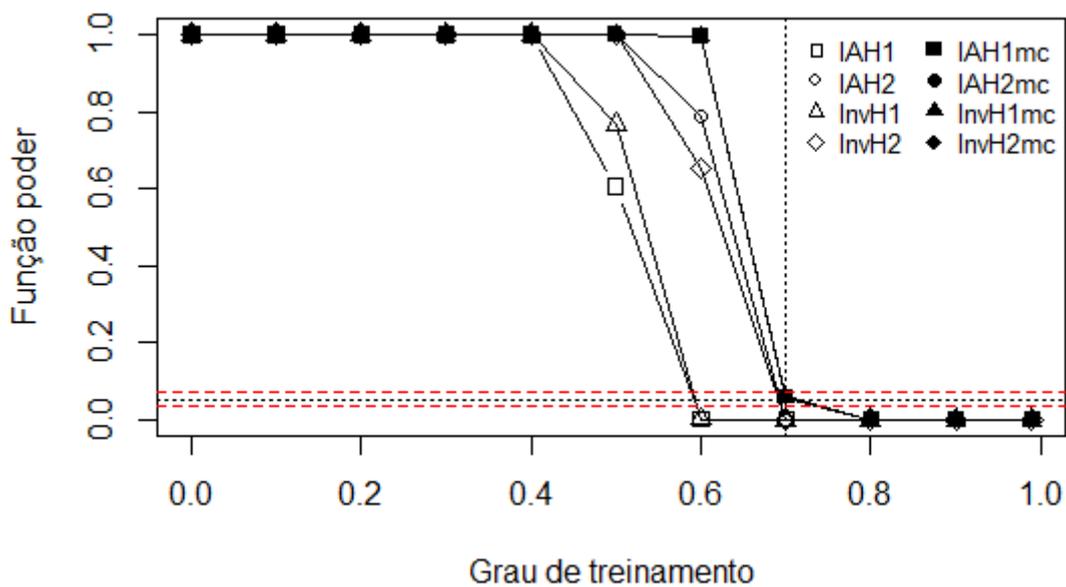
$InvH_2$  quando o grau de treinamento estipulado pelo pesquisador é maior que 0,70.

Figura 5 – Poder dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  para o caso  $p = 10$  provadores,  $q = 5$  atributos,  $n = 10$  observações.



Fonte: da autora.

Figura 6 – Poder dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  para o caso  $p = 10$  provadores,  $q = 5$  atributos,  $n = 50$  observações.



Fonte: da autora.

Considerando os cenários em que o número de provadores era  $p = 10$ , Ferreira (2017) obteve resultado semelhante a este trabalho e indicou o uso do teste assintótico  $InvH_2$ , por ser

considerado exato nos casos em que o número de observações é pequeno e menos liberal nos casos em que o número de observações é grande. Além disso, são mais indicados que os testes generalizados por Rocha (2015).

Pode-se observar ainda que, nos cenários em que o número de provadores era  $p = 15$  e o número de atributos  $q = 2$ , combinados com  $n = 30, 40$  e  $50$ , o teste assintótico  $IAH_2$  apresentou maior poder que os demais testes. No entanto, isso representa apenas 4,76% dos cenários analisados. Assim, de forma geral, verificou-se que os testes Monte Carlo propostos neste trabalho são equivalentes entre si e apresentam bom controle de taxas de erro tipo I, alto poder e exatidão.

Para estes cenários, Ferreira (2017) sugeriu o uso do teste  $InvH_2$ , pois além de ser poderoso, ele é menos liberal. Para amostras grandes, os testes sobre autovalores Monte Carlo e Fujikoshi generalizados por Rocha (2015) são mais recomendados do que os testes bootstrap paramétricos e o teste Monte Carlo para unidimensionalidade.

## 4 CONCLUSÃO

Os quatro testes Monte Carlo propostos apresentaram por meio do estudo de simulação um comportamento semelhante e foram considerados exatos por estarem dentro do intervalo de confiança para o nível de significância  $\alpha$ . Eles controlaram as taxas de erro tipo I perto da região de rejeição da hipótese nula e são indicados independente do tamanho da amostra. Além disso, apresentaram maior poder em 95,24% dos cenários analisados quando comparadas às versões assintóticas.

Assim, como os testes propostos neste trabalho são equivalentes entre si, indica-se o teste  $InvH_2 mc$ , pois ele apresentou alto poder e exatidão em todos os cenários analisados e, também, por Ferreira (2017) ter indicado a sua versão assintótica ( $InvH_2$ ).

## REFERÊNCIAS

- AMORIM, I. S. 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.
- ANDERSON, T. W. Asymptotic theory for principal components. **Annals of Mathematical Statistics**, Ann Arbor, v. 34, p. 122-148, 1963.
- BI, J. Agreement and reliability assessment for performance of sensory descriptive panel. **Journal of Sensory Studies**, Richmond, v.18, p. 61-67, 2003.
- DIJKSTERHUIS, G. Assessing panel consonance. **Food Quality and Preference**, v. 6, p. 7-14, 1995. Disponível em: <[https://doi.org/10.1016/0950-3293\(94\)P4207-M](https://doi.org/10.1016/0950-3293(94)P4207-M)>. Acesso em: 30 ago. 2018.
- FERNANDES, F. M. O. **Proposta de um teste Monte Carlo para unidimensionalidade de painéis sensoriais**. 2012. 80 f. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2012.
- FERREIRA, D. F. **Estatística multivariada**. Lavras: Editora UFLA, 2008. 662 p.
- FERREIRA, L. B. S. **Testes assintóticos para detectar consenso multivariado em painéis sensoriais**. 80 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, 2017.
- GEBERT, D. M. P. **Proposta de testes bootstrap para inferir sobre o número de componentes principais retidos**. 2010. 75 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)- Universidade Federal de Lavras, Lavras, 2010.
- GIRSHICK, M. A. On the sampling theory of roots of detrimental equations. **The Annals of Mathematical Statistics**, v. 10, p. 203-224, 1939.
- LATREILLE, J. et al. Measurement of the reliability of sensory panel performances. **Food Quality and Preference**, Barking, n. 17, p. 369-365, 2006. Disponível em: <<https://doi.org/10.1016/j.foodqual.2005.04.010>>. Acesso em: 12 Set. 2018.
- LAWLEY, D. N. Tests of significance of the latent roots and correlation matrices. **Biometrika**, London, v. 43, p. 128-136, 1956.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>.
- ROCHA, M. C. **Testes baseados em correlação canônica para avaliar consonância de painéis sensoriais**. 71 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, 2015.
- ROGERS, L. **Sensory panel management: a practical handbook for recruitment, training and performance**. Duxford: Elsevier Ltd., 2018. 376 p.

VERHOEF, A.; HUIJBERTS, G.; VAESSEN, W. Introduction of a quality index, based on Generalizability theory, as a measure of reliability for univariate and multivariate sensory descriptive data. **Food Quality and Preference**, Netherlands, v.40, p.296-303, 2014.

### **CAPÍTULO 3 - TESTE MONTE CARLO NA VERIFICAÇÃO DO CONSENSO DO PAINEL DE BEBIDAS SABOR GUARANÁ**

Tatiane Gomes de Araújo; Eric Batista Ferreira

**RESUMO:** Para avaliar a qualidade sensorial de bebidas, painéis sensoriais treinados são uma importante ferramenta. Contudo, faz-se necessário monitorar seu desempenho para verificar se as informações por ele fornecidas são confiáveis. Com esse intuito, Tomic et al. (2009) propuseram um protocolo para verificar de forma uni e multivariada as diferenças e similaridades na performance do painel. No entanto, este protocolo pode ser melhorado ao acrescentar uma etapa de teste que permita inferir sobre o comportamento de painéis. Dessa forma, foi proposta a inserção de um teste Monte Carlo para autovalores para testar o consenso multivariado. Verificou-se a qualidade do treinamento de um painel de refrigerante sabor guaraná de duas marcas usualmente encontradas no mercado brasileiro em suas versões tradicional e zero açúcar. Pode-se observar de forma gráfica e, também, no teste  $InvH_2mc$ , que o painel não apresenta consenso multivariado e necessita de um novo treinamento.

**Palavras-Chave:** Sensometria. Concordância de painéis. Bebida sabor guaraná.

# MONTE CARLO TEST ON VERIFICATION CONSENSUS IN A GUARANA FLAVOR DRINK PANEL

Tatiane Gomes de Araújo; Eric Batista Ferreira

**ABSTRACT:** In order to evaluate the sensory quality of beverages, trained sensory panels are an important tool. However, it is necessary to monitor their performance to verify that the information provided by it is reliable. To this purpose, Tomic et al. (2009) proposed a protocol to verify the differences and similarities in the panel performance in a uni and multivariate manner. However, this protocol can be improved by adding a test step that allows one to infer about the panels behavior. Thus, the insertion of a Monte Carlo test for self-values was proposed to test the multivariate consensus. The quality of the guarana soda flavor panel training from two brands usually found in the Brazilian market was verified in both their traditional and zero sugar versions. It can be graphically observed, and also in the  $InvH_2mc$  test, that the panel does not present a multivariate consensus and requires a new training.

**Keywords:** Sensory analysis. Panel concordance. Guarana Flavor Drink.

## 1 INTRODUÇÃO

Para avaliar a qualidade sensorial de bebidas, painéis sensoriais treinados são uma importante ferramenta. Contudo, faz-se necessário monitorar o desempenho de painéis, pois existem diversos problemas relacionados ao treinamento, estabilidade e qualidade das informações por ele transmitidas. Para verificar a performance dos painéis, vários métodos para detectar a sua discriminabilidade, repetibilidade e consonância foram desenvolvidos. Estas informações são úteis para validar o desempenho do painel e para que em sessões futuras haja um treinamento melhor e mais direcionado (TOMIC et al., 2009).

Na literatura existem diversos métodos para avaliar de forma descritiva o desempenho do provador e de todo painel. Tomic et al. (2009) propuseram um protocolo que visa verificar de forma multivariada as diferenças em um nível mais geral, considerando as correlações entre atributos e de forma univariada as diferenças para cada atributo. Todas as técnicas são graficamente apresentadas e fáceis de serem interpretadas e geradas no software PanelCheck, permitindo uma análise rápida e eficiente dos dados.

No entanto, este protocolo pode ser melhorado ao acrescentar em seu processo de verificação testes que permitem inferir sobre o comportamento de painéis. Os testes propostos na seção 2.1 do capítulo 2, permitem verificar de forma multivariada o consenso de painéis, tornando-se uma importante ferramenta de triagem e conferência do grau de treinamento recebido por provadores ao analisar produtos, inclusive bebidas.

Em análise sensorial é comum a avaliação de bebidas como vinhos, cafés, cervejas e cachaças (CASTILHOS; BIANCHI, 2011; AMORIM et al., 2009; VIEJO et al., 2018; ODELLO et al., 2009). Outra bebida muito consumida no Brasil, o refrigerante sabor guaraná foi avaliado por Fagundes (2018). Foram analisadas duas marcas tradicionalmente encontradas no mercado brasileiro em suas versões tradicional e zero açúcar. Para isso, um painel foi formado por meio da técnica de perfil descritivo otimizado (PDO) para pontuar atributos relacionados à visão, olfato e paladar.

Assim, o presente trabalho tem por objetivo propor uma etapa inferencial no protocolo sugerido por Tomic et al. (2009) e verificar a qualidade do treinamento do painel ao avaliar o refrigerante sabor guaraná.

## 2 MATERIAL E MÉTODOS

### 2.1 DADOS

Os dados utilizados neste trabalho são oriundos de um experimento realizado por Fagundes (2018), que avaliou refrigerantes sabor guaraná de duas marcas usualmente encontradas no mercado brasileiro nas versões tradicional e zero açúcar. Os testes sensoriais foram realizados no laboratório de Análise Sensorial de Alimentos e Bebidas da Faculdade de Nutrição da Universidade Federal de Alfenas (FANUT/Unifal-MG), sendo aprovada pelo Comitê de Ética em pesquisa com humanos da Unifal-MG sob parecer número 2.137.975.

O painel foi composto por 16 provadores ( $p = 16$ ), selecionados após um processo de triagem e treinados utilizando-se a técnica de Perfil Descritivo Otimizado (PDO), que garante um painel semi-treinado com tempo de treinamento reduzido (SILVA et al., 2014). Para selecionar os candidatos a compor o painel, aplicou-se quatro testes triangulares aleatorizados, com o objetivo de verificar se os candidatos conseguiam discernir o gosto doce. Na etapa seguinte, o método rede foi aplicado a fim de que os provadores selecionados levantassem os atributos a serem avaliados, indicando as similaridades e diferenças entre as amostras em relação à aparência, ao aroma e ao sabor.

Posteriormente, realizou-se uma discussão em grupo para ordenar os termos descritivos semelhantes levantados no método rede e gerar amostras de referências. A Tabela 1 apresenta a definição de cada termo descritivo e as referências para cada atributo desenvolvidos pela equipe sensorial de forma consensual. A intensidade de cada descritor foi avaliada três vezes em cada amostra por uma escala não estruturada de 10 centímetros (devido a um redimensionamento ao imprimir as fichas). Além disso, foram dispostos na cabine os termos de intensidade ancorados em seus extremos, sendo o mínimo à esquerda e o máximo à direita, para que o provador alocasse a intensidade da amostra em relação ao estímulo sensorial.

Tabela 1 – Definição dos termos descritivos e materiais de referências para refrigerante de guaraná.

Atributo	Definição	Referências
<i>Aparência</i>		
Caramelo	Intensidade da cor caramelo	Claro: 100ml refrigerante* + 45ml água Escuro: 100ml refrigerante* + 10ml de xarope de guaraná
Gás	Quantidade de bolhas de gás	Nenhuma: Água pura Muitas: Água com gás
<i>Aroma</i>		
Doce	Intensidade do cheiro doce	Fraco: 20g caramelo <sup>†</sup> + 300g de água Forte: 40g caramelo <sup>†</sup> + 300g de água
Guaraná	Intensidade do aroma característico de guaraná	Fraco: 1g xarope de guaraná + 54g água Forte: 10g xarope de guaraná + 54g água
<i>Sabor</i>		
Guaraná	Intensidade do sabor característico de guaraná	Fraco: 1g xarope de guaraná + 54g água Forte: 10g xarope de guaraná + 54g água
Amargo	Intensidade do gosto básico amargo provocado pelo adoçante	Fraco: 2g de Pó de Guaraná + 53,7g de água Forte: 5g Pó de Guaraná + 50g de água
Doce	Sensação do gosto básico provocado pela sacarose	Pouco: 0,6g açúcar refinado + 50g água Muito: 6g açúcar refinado + 50g água

Fonte: Adaptado de Fagundes (2018, p. 37).

## 2.2 MÉTODOS

Para avaliar o desempenho do painel sensorial seguiu-se o protocolo sugerido por Tomic et al. (2009) (Figura 8). Cada método gráfico recomendado será apresentado na ordem em que os autores sugerem a sua aplicação e eles podem ser obtidos no software gratuito PanelCheck.

### 2.2.1 Análise de variância com Modelo Misto

A análise de variância (ANAVA) com modelos mistos é utilizada para avaliar quais atributos são importantes para detectar diferenças significativas entre as amostras. O método consiste em modelar as amostras, provadores e interações no modelo ANAVA de duas vias ou modelar amostras, provadores, repetições e suas interações no modelo ANAVA de três vias e, em seguida, testar o efeito da amostra pelo teste F. Nos dois casos o provador e os efeitos da interação são considerados aleatórios. Somente os atributos significativos no nível de significância escolhido (neste trabalho definiu-se  $\alpha = 5\%$ ) para efeito do produto são considerados

\*Refrigerante de guaraná Antártica.

<sup>†</sup>Calda de açúcar.

para análise posterior (LOSÓ et al., 2012).

### 2.2.2 Tucker-1

O método de análise multivariada Tucker-1 é utilizado para se ter uma visão geral sobre o desempenho do provador e do painel avaliando múltiplos atributos. O método é essencialmente uma análise de componentes principais (ACP) aplicada em uma matriz desdobrada horizontalmente de dimensão  $n \times pq$  que consiste em todas as matrizes individuais ( $\mathbf{X}_i$ ) alinhadas horizontalmente, sendo  $n$  o número de observações,  $p$  o número de provadores e  $q$  o número de atributos. Cada matriz  $\mathbf{X}_i$  representa a matriz das notas de um provador e possui dimensão  $n \times q$ , na qual a pontuação da amostra é baseada na média das repetições (NAES et al., 2010)

O método fornece dois gráficos: um gráfico de pontuação comum e um gráfico de cargas de correlação. O gráfico de pontuação comum ilustra como as  $n$  amostras testadas se relacionam entre si, ou seja, apresenta quais as suas semelhanças e dissimilaridades ao longo dos componentes principais encontrados. Baseando-se na alta variância explicada nos primeiros componentes principais, pode-se considerar que o painel apresenta a capacidade de discriminação entre produtos (LOSÓ et al., 2012).

O gráfico de cargas de correlação fornece informações sobre o desempenho individual do provador e do painel em geral. O gráfico contém  $p \times q$  pontos, com cada ponto representando uma combinação provador-atributo. A posição dos pontos dentro do gráfico informa o desempenho do provador ou de todo o painel. Quanto mais ruído o atributo de um determinado provador contiver, mais próximo o ponto aparecerá no meio do gráfico. Quanto mais informação sistemática um atributo de um provador contiver, mais próximo ele aparecerá da elipse externa (variância explicada em 100% para esse atributo). A elipse interna representa 50% da variância explicada e pode ser considerada como um limite inferior de regra geral de quanto a variância explicada por pelo menos um atributo deve ser considerada boa o suficiente (TOMIC et al., 2009).

### 2.2.3 Gráficos de Manhattam

Os gráficos de Manhattam permitem identificar de forma rápida se algum provador apresenta desempenho diferente de outros provadores do painel. Estes gráficos podem ser obtidos utilizando-se diferentes métodos estatísticos, dentre eles a ACP, a qual é aplicada na matriz de dados individuais  $X_i$  e a variância explicada para cada atributo-provador é visualizada no gráfico (TOMIC et al., 2009).

Os gráficos de Manhattan apresentam, em tons de cinza, o quanto da variabilidade de cada atributo e de cada avaliador pode ser explicado por determinado número de componentes principais (eixo vertical). Um tom de cor escura indica que apenas uma pequena proporção da variação foi explicada, enquanto um tom de cor clara indica o oposto. Os pontos extremos são "pretos"(0% de variância explicada) e "brancos"(100% de variância explicada). No presente trabalho, interessa-se verificar o desempenho entre provadores. Para isto investiga-se um número total de  $q$  gráficos, compostos por  $p$  colunas, sendo que cada gráfico representa um atributo e cada coluna dentro dos gráficos representa um provador (TOMIC et al., 2009).

### 2.2.4 Gráficos F

Os gráficos F são baseados em valores F e são utilizados para verificar a capacidade de cada provador detectar diferenças entre as amostras para um determinado atributo. Os valores F são apresentados em um diagrama de barras, contendo duas linhas horizontais que representam o nível de significância adotado, geralmente de 1% e 5%. Pode-se dizer que quanto maior o valor F de um provador, maior é a sua capacidade em discriminar as amostras. Além disso, se houver diferenças entre as amostras espera-se que os provadores obtenham altos valores de F (LOSÓ et al., 2012).

### 2.2.5 Gráficos MSE

Os gráficos MSE são usados como uma medida de repetibilidade para cada provador. Os valores MSE são os erros quadráticos médios (estimativas aleatórias de variação de erro)

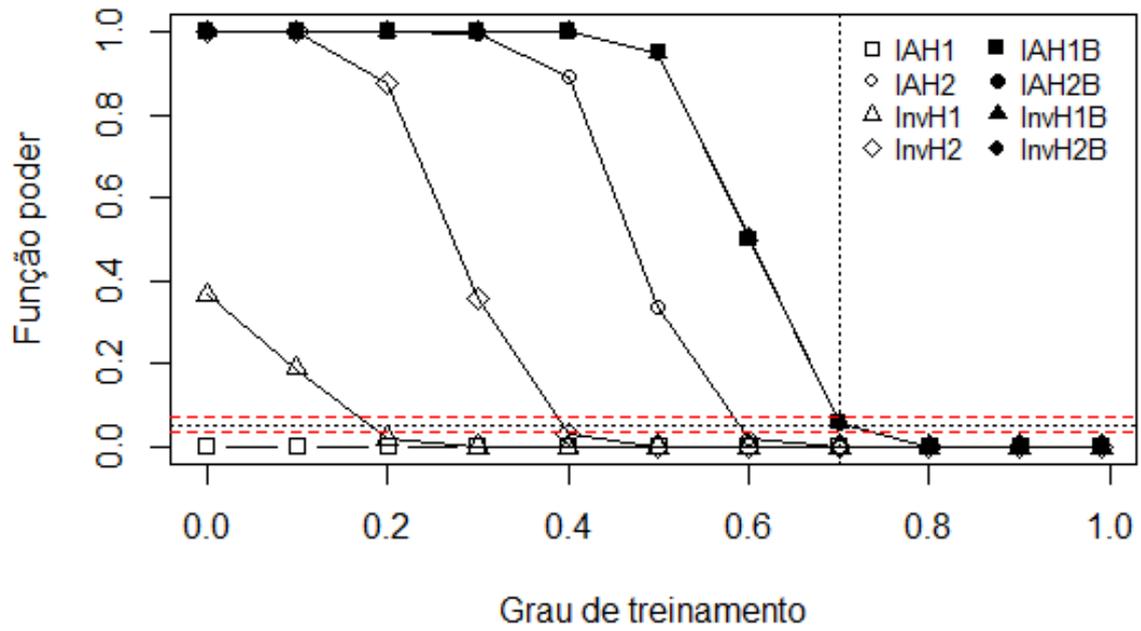
do modelo ANAVA, os quais são plotados em um gráfico de barras. Se o provador repetesse quase perfeitamente, esse valor deve ser próximo de zero. Este gráfico deve ser avaliado juntamente com o gráfico F para obter uma visão mais realista do desempenho do painel. Se existirem diferenças entre as amostras, um provador deve idealmente ter altos valores de F e baixos valores de MSE (LOSÓ et al., 2012).

### 2.2.6 Testes para verificar consenso multivariado de painéis sensoriais

Para verificar o consenso de painéis sensoriais de forma inferencial, pode-se utilizar os testes propostos em 2.1. Para decidir qual teste será aplicado, realizou-se um estudo de simulação no software R (R CORE TEAM, 2019), considerando o cenário  $p = 16$  provadores,  $q = 4$  atributos e  $n = 12$  observações, conforme indicado em 2.2.

Pode-se observar na Figura 7 que o teste assintótico  $IAH_1$  apresenta poder nulo e não rejeita a hipótese de nulidade, mesmo quando ela é nitidamente falsa. O teste  $IAH_2$  tem alto poder (89,3%), porém há uma queda de poder no grau de treinamento  $\rho_0 = 0,60$ . Os quatro testes Monte Carlo são equivalentes entre si, são considerados exatos e mais poderosos (94,9%), por isso são os mais indicados neste cenário. Assim, de acordo com o que foi obtido nesta simulação e seguindo a recomendação do Capítulo 2, optou-se por utilizar o teste  $InvH_{2mc}$  para verificar o consenso deste painel.

Figura 7 – Poder dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$ ,  $InvH_2mc$  para o caso  $p = 16$  provedores,  $q = 4$  atributos,  $n = 12$  observações.

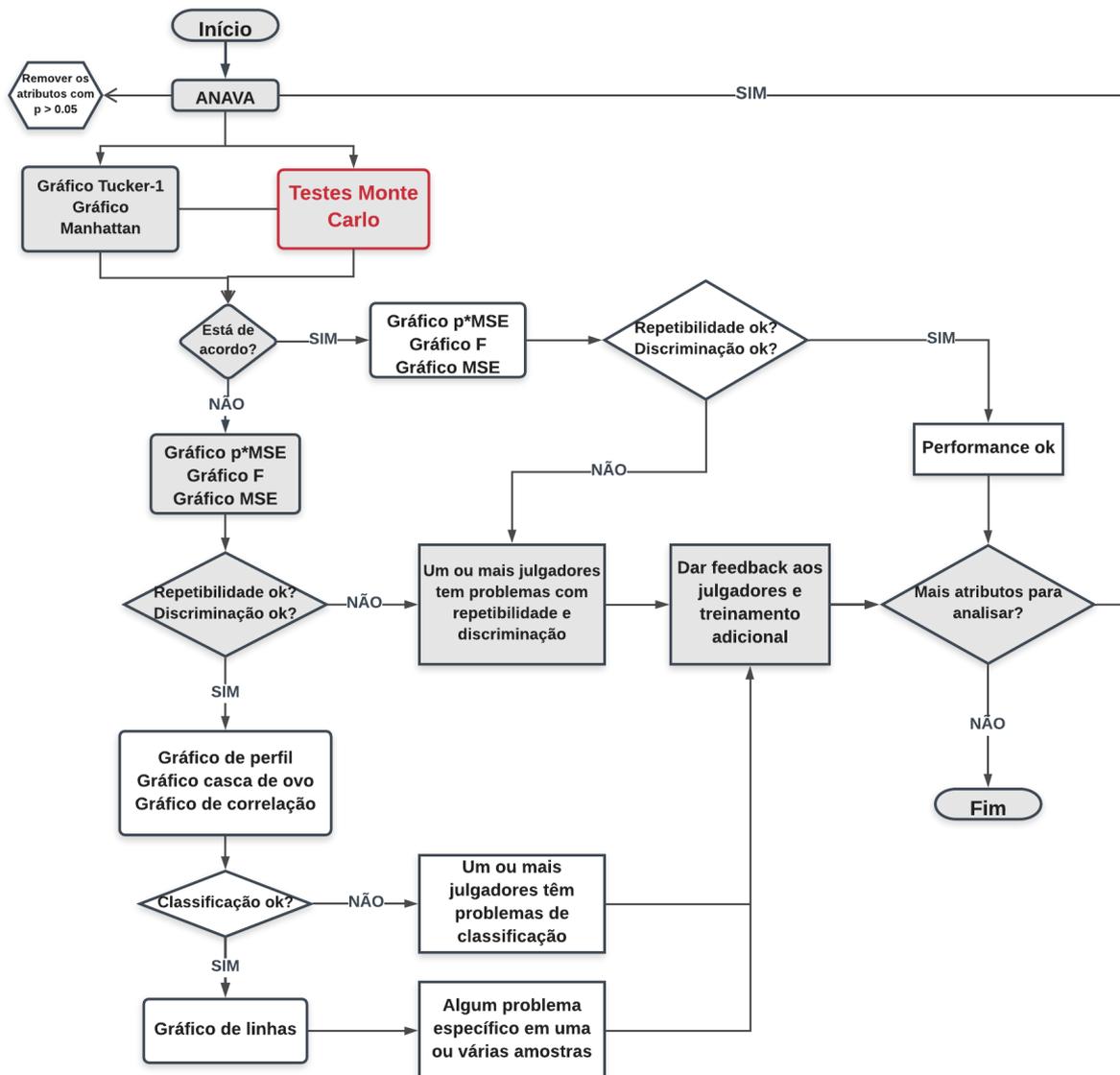


Fonte: da autora.

### 3 RESULTADOS

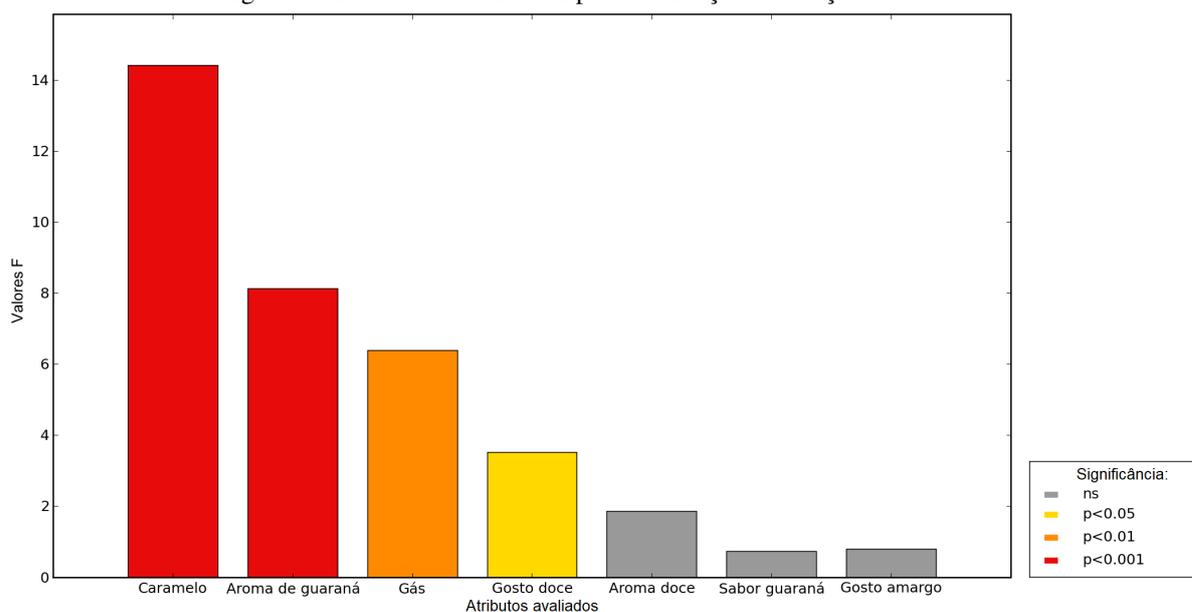
Para analisar o desempenho do painel sensorial ao avaliar os refrigerantes sabor guaraná, o fluxograma da Figura 8 indica primeiramente o uso da ANOVA com modelos mistos. Pode-se verificar na Figura 9 que os atributos intensidade do aroma doce (Doce), intensidade do sabor característico de guaraná (Gosto guaraná) e a intensidade do gosto básico amargo provocado pelo adoçante (Amargo) foram não significativos ( $\text{valor-}p > \alpha$ ). Assim, como estes atributos não contribuem de forma significativa para detectar diferenças entre as amostras eles foram retirados da análise.

Figura 8 – Fluxo de trabalho proposto para a analisar o desempenho do provador e do painel.



Fonte: Adaptado de Tomic et al. (2009).

Figura 9 – Resultado da ANOVA para verificação de seleção de atributos.

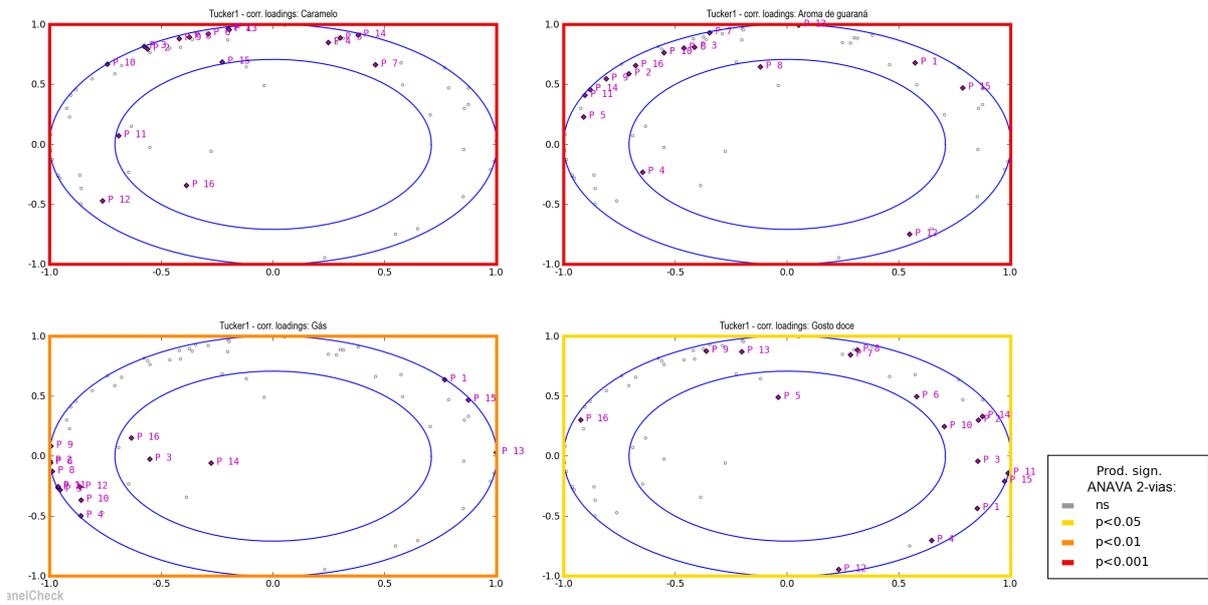


Fonte: da autora.

O próximo passo indicado é a aplicação do gráfico de Tucker-1 para avaliar de forma geral o desempenho de cada provador e do painel de forma geral. Pode-se observar na Figura 10, que o painel não apresenta um desempenho satisfatório. Para os atributos cor de caramelo (Caramelo) e aroma característico de guaraná (Guaraná), o painel apresenta um maior consenso, pois as notas atribuídas pela maioria dos provadores estão mais próximas. Em nenhum dos atributos todos os pontos estão concentrados entre as elipses interna e externa, indicando que são necessários mais de dois componentes principais para explicar mais de 50% da variabilidade dos dados.

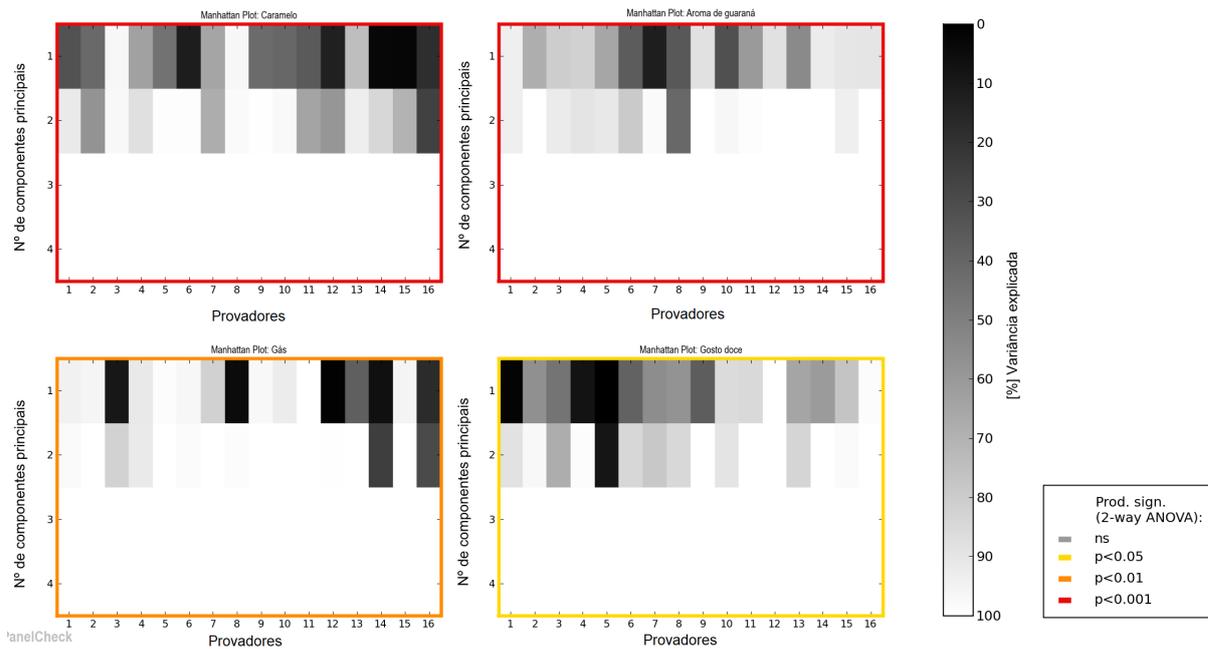
Os gráficos de Manhattan confirmam o que foi observado nos gráficos de Tucker-1 (Figura 11) quanto a variação sistemática para cada atributo provador. De forma geral, os provadores precisaram de mais de duas componentes principais para alcançar um alto nível de variância explicada. Pode-se observar ainda que alguns provadores apresentam mais dificuldade em detectar diferenças para um ou mais atributos e que não há consenso entre eles. Os provadores tiveram menos dificuldade ao avaliar o aroma característico de guaraná, com exceção do provador 8 que precisa de mais de dois componentes principais para alcançar um alto nível de variância explicada. Por isso, esse provador precisa receber novo treinamento para avaliar este atributo.

Figura 10 – Gráficos Tucker-1 para cada atributo analisado.



Fonte: da autora.

Figura 11 – Gráficos Manhattan para cada atributo analisado.



Fonte: da autora.

Verifica-se ainda que o provador 16 identifica melhor as diferenças entre produtos ao avaliar os atributos intensidade do aroma característico de guaraná (Guaraná) e a sensação do gosto básico provocado pela sacarose (Gosto doce), enquanto tem dificuldade para identificar diferenças ao avaliar a intensidade da cor caramelo (Caramelo) e a quantidade de bolhas de gás (Gás). Os provadores 3 e 8 têm mais facilidade para detectar diferenças entre a intensidade da cor caramelo (Caramelo) entre as amostras e o provador 7 teve mais dificuldade que os demais

membros do painel ao avaliar a intensidade do aroma característico de guaraná (Guaraná).

Ao aplicar o teste *InvH<sub>2</sub>mc* às notas obtidas pelo painel sensorial obteve-se valor-*p* = 0,0001. Assim, confirmando de forma inferencial o que havia sido constatado nos gráficos de Tucker-1 (Figura 10) e Manhattam (Figura 11), o painel não apresenta consenso multivariado. Logo, faz-se necessário investigar para quais atributos e/ou quais provadores deve ocorrer um novo treinamento.

Para isto, Tomic et al. (2009) indicam o gráfico MSE e o gráfico F para verificar, respectivamente, a capacidade de repetibilidade e discriminação do painel. Pode-se observar na Figura 12(a) que os valores MSE são altos, indicando que os provadores apresentaram pouca capacidade de replicar suas notas. Observa-se ainda na Figura 12(b) que não há altos valores de F, ou seja, o painel não apresenta boa capacidade de discriminação.

Além disso, apresentaram dificuldade em replicar suas notas ao avaliar: a intensidade da cor caramelo (Caramelo) os provadores 7, 12 e 13; a quantidade de bolhas de gás (Gás) os provadores 1, 2, 5, 6, 7, 8, 12, 13, 14 e 15; a intensidade do aroma característico de guaraná (Guaraná) os provadores 1, 2, 4, 7, 9, 12, 13 e 16; e a sensação do gosto básico provocado pela sacarose (Gosto doce) os provadores 1, 3, 4, 12, 13, 14 e 16 (Figura 12(a)).

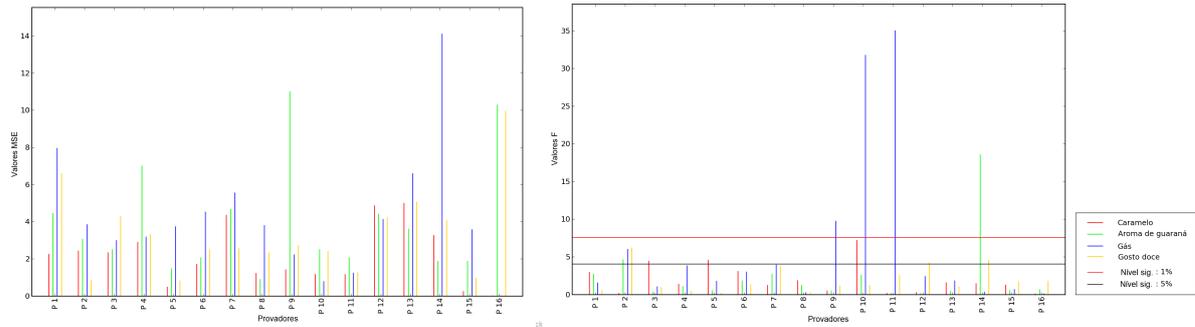
Portanto, como os provadores não apresentam capacidade de identificar pequenas diferenças entre produtos, replicar as mesmas notas para um mesmo produto e concordar entre si, faz-se necessário a aplicação de um novo treinamento para que o painel possa apresentar resultados confiáveis.

Carneiro et al. (2003) também avaliaram marcas de refrigerantes encontradas no comércio de Viçosa (MG), sendo que seis delas foram produzidas por pequenas empresas e uma era da empresa marca líder de mercado. Para isto, nove provadores foram treinados através da Análise Descritiva Quantitativa (ADQ) e selecionados após verificar a capacidade deles com relação a discriminação e repetibilidade. Eles avaliaram os seguintes atributos: cor de caramelo, aroma e sabor característicos de refrigerante sabor guaraná, aroma e sabor de tutti-fruti (associados a mistura de frutas), efervescência, gosto doce, sabor associado ao xarope infantil (medicinal), sabor associado a solução de ácido cítrico (frutas cítricas) e sabor residual associado a solução de sacarina (adoçante).

Através da aplicação da análise de variância eles observaram que os atributos aroma e sabor característico, aroma e sabor de tutti-fruti, cor caramelo e sabor medicinal contribuíram para identificar diferenças entre as amostras. Além disso, o refrigerante da marca líder

apresentou os atributos aroma e sabor característico em maior intensidade, enquanto que os refrigerantes produzidos por pequenas empresas apresentaram maior intensidade para os atributos aroma e sabor de tutti-fruti, sabor caramelo e sabor medicinal. No entanto, nenhum teste sobre o desempenho do painel foi aplicado.

Figura 12 – Gráfico MSE (a) e gráfico F (b) do painel para verificar repetibilidade e discriminação.



Fonte: da autora.

## 4 CONCLUSÃO

O protocolo proposto por Tomic et al. (2009) para verificar o desempenho de provadores e de todo o painel mostrou-se útil para verificar o consenso do painel ao avaliar refrigerantes sabor guaraná de duas marcas usualmente encontradas no mercado brasileiro em suas versões tradicional e zero açúcar. A inserção do teste Monte Carlo permitiu analisar de forma inferencial o comportamento do painel, corroborando com o que já havia sido observado de forma descritiva nos gráficos.

Pode-se observar que o painel necessita de um novo treinamento, pois ele não conseguiu identificar pequenas diferenças entre produtos, replicar as mesmas notas para um mesmo produto e concordar entre si. Alguns provadores tiveram dificuldades em identificar diferenças ao analisar os atributos relacionados a visão (intensidade de cor caramelo e quantidade de gás), mas tiveram bom desempenho ao avaliar os atributos relacionados ao olfato (intensidade do cheiro doce) e paladar (intensidade do sabor característico de guaraná), enquanto outros obtiveram bom desempenho ao analisar apenas um atributo.

## REFERÊNCIAS

- AMORIM, I. S. 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.
- CARNEIRO, J. D. D. S. et al. Avaliação sensorial e mapa de preferência interno de marcas comerciais de refrigerantes sabor guaraná. **Boletim do Centro de Pesquisa de Processamento de Alimentos**, Curitiba, v. 21, n. 2, p. 279-292, 2003. Disponível em: <DOI: <http://dx.doi.org/10.5380/cep.v21i2.1165>>. Acesso em: 8 jan. 2019.
- CASTILHOS, M. B. M.; Del Bianchi, V. L. Caracterização físico-química e sensorial de vinhos brancos da região noroeste de São Paulo. **Holos**, v. 4, p. 148-158, 2011. Disponível em: <<http://hdl.handle.net/11449/122449>>. Acesso em: 01 fev. 2019.
- FAGUNDES, C. S. S. **Análise sequencial e o ajuste de mapas de preferência**. 2018. 71 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, 2018.
- LOSÓ, V. et al. Comparison of the performance of a trained and an untrained sensory panel on sweet corn varieties with the PanelCheck software. **APSTRACT** Budapeste, v. 1-2. p. 77-85. 2012. Disponível em: <DOI: [10.19041/Apstract/2012/1-2/10](https://doi.org/10.19041/Apstract/2012/1-2/10)>. Acesso em: 10 Jan 2019.
- NAES, T.; BROCKHOFF PER B.; TOMIC, O. **Statistics for sensory and consumer science**. London: Wiley, 2010. 282 p.
- ODELLO, Luigi et al. Avaliação sensorial de cachaça. **Quím. Nova**, São Paulo, v. 32, n. 7, p. 1839-1844. 2009. Disponível em: <<http://dx.doi.org/10.1590/S0100-40422009000700029>>. Acesso em: 02 Fev. 2019.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>.
- SILVA, R. C. S. N. et al. Validation of optimized descriptive profile (ODP) technique: Accuracy, precision and robustness. **Food Research International**, v. 66, p. 445-453, 2014. Disponível em: <<https://doi.org/10.1016/j.foodres.2014.10.015>>. Acesso em: 30 mar. 2018.
- TOMIC, O. et al. Analysing sensory panel performance in a proficiency test using the PanelCheck software. **Eur Food Res Technol**, Dinamarca, v. 230, p. 497-511. 2009. Disponível em: <<https://doi.org/10.1007/s00217-009-1185-y>>. Acesso em: 05 dez 2018.
- VIEJO, C. G. et al. Integration of non-invasive biometrics with sensory analysis techniques to assess acceptability of beer by consumers. **Physiology and Behavior**, Austrália. v. 200, p.139-147. 2018. Disponível em <<https://doi.org/10.1016/j.physbeh.2018.02.051>>. Acesso em: 01 fev. 2019.

## CAPÍTULO 4 - CONSIDERAÇÕES FINAIS

As quatro versões Monte Carlo dos testes de Ferreira (2017) por meio do estudo de simulação apresentaram maior poder que suas versões assintóticas em 95,24% dos cenários analisados, como pode ser verificado nas Figuras do apêndice. Além disso, os quatro testes propostos apresentaram comportamento semelhante e foram considerados exatos por estarem dentro do intervalo de confiança para o nível de significância  $\alpha$ . Eles controlaram as taxas de erro tipo I perto da região de rejeição da hipótese nula e são indicados independente do tamanho da amostra.

Assim, como os testes propostos neste trabalho são equivalentes entre si, indica-se o teste  $InvH_2mc$ , pois ele apresentou alto poder e exatidão em todos os cenários analisados e, também, por Ferreira (2017) ter indicado a sua versão assintótica ( $InvH_2$ ).

A eficiência dos testes Monte Carlo pôde ser comprovada ao avaliar o desempenho de um painel sensorial para detectar diferenças entre duas marcas de refrigerante sabor guaraná e suas versões tradicionais e zero açúcar. Para verificar se o painel apresenta consenso multivariado o teste  $InvH_2mc$  foi o mais indicado para ser utilizado neste cenário. Ao aplicá-lo verificou-se que o painel não apresentou consenso multivariado, ou seja, os provadores não pontuaram os atributos de forma semelhante.

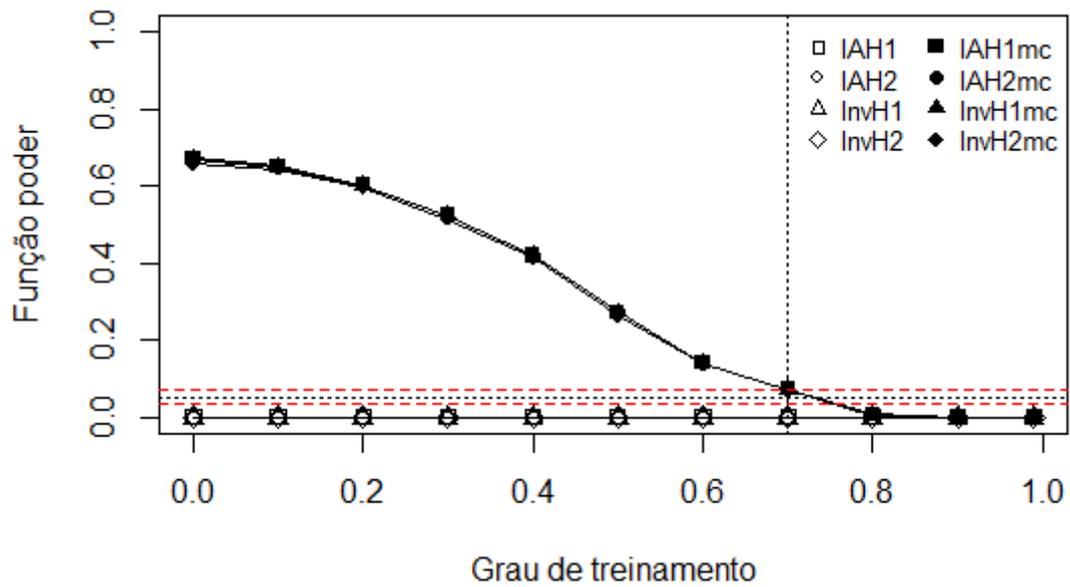
Tal fato também foi confirmado ao aplicar o protocolo proposto por Tomic et al. (2009) ao analisar os gráficos de Tucker-1 e Manhattam que indicaram que os provadores não concordaram entre si. Além disso, verificou-se que os provadores não conseguiram replicar suas notas para um mesmo produto e identificar pequenas diferenças entre os produtos.

Para trabalhos futuros, sugere-se verificar a robustez dos testes e considerar as correlações entre atributos e provadores  $\times$  atributos.

## APÊNDICE

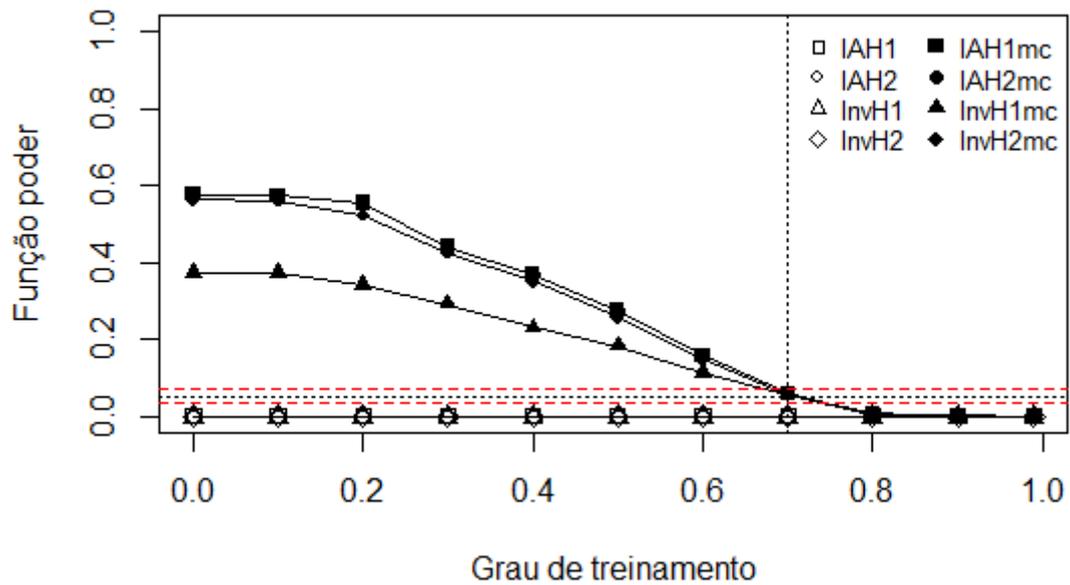
APÊNDICE - Gráficos das curvas dos poderes dos testes  $IAH_1$ ,  $IAH_2$ ,  $InvH_1$ ,  $InvH_2$ ,  $IAH_1mc$ ,  $IAH_2mc$ ,  $InvH_1mc$  e  $InvH_2mc$ .

Figura 13 – P = 2 provadores, q = 2 atributos, n = 10 observações.



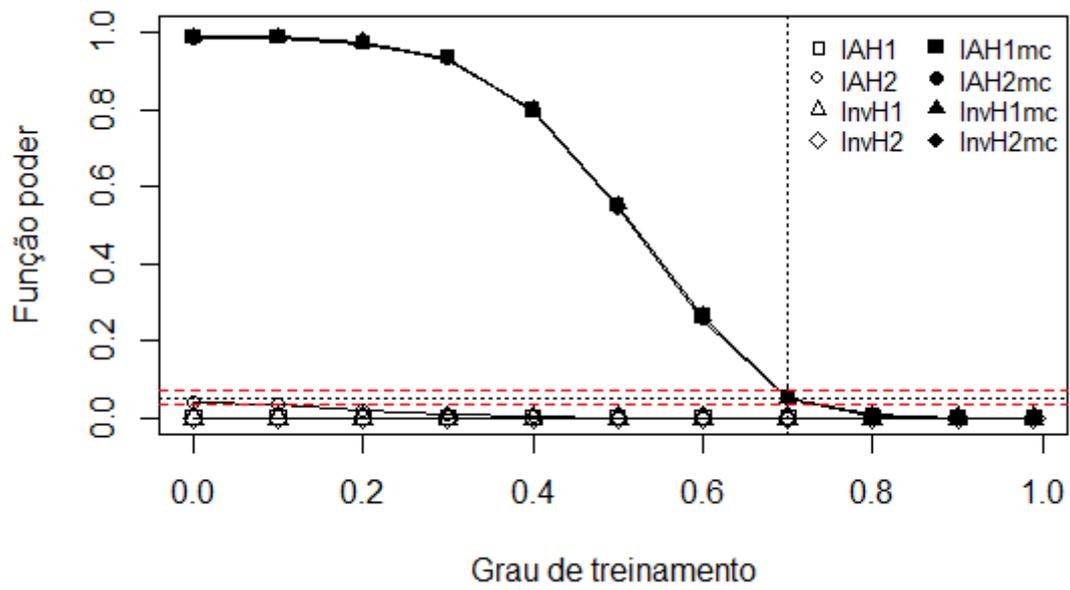
Fonte: da autora.

Figura 14 – P = 2 provadores, q = 5 atributos, n = 10 observações.



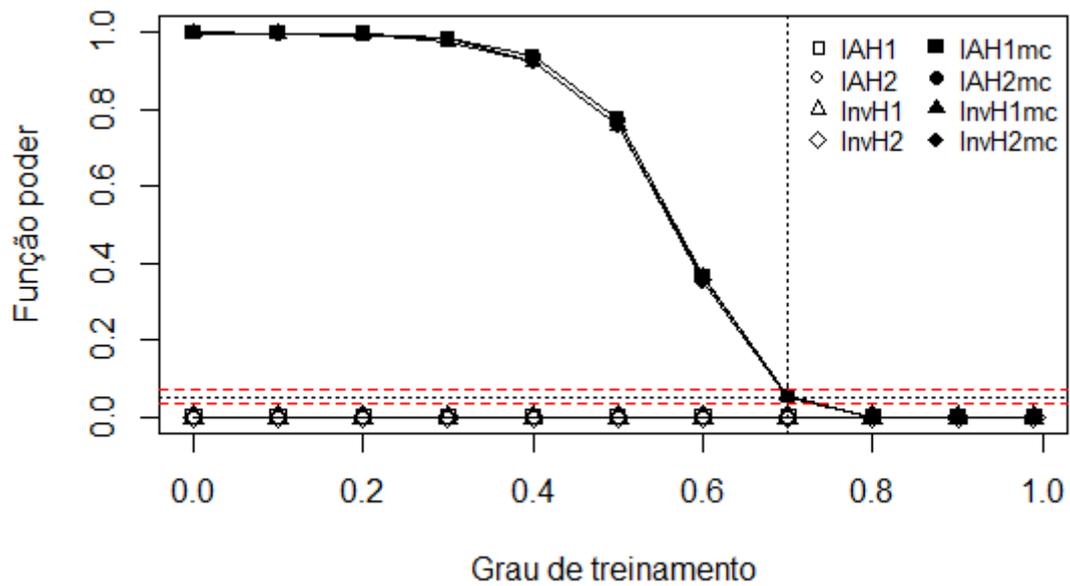
Fonte: da autora.

Figura 15 –  $P = 2$  provedores,  $q = 2$  atributos,  $n = 20$  observações.



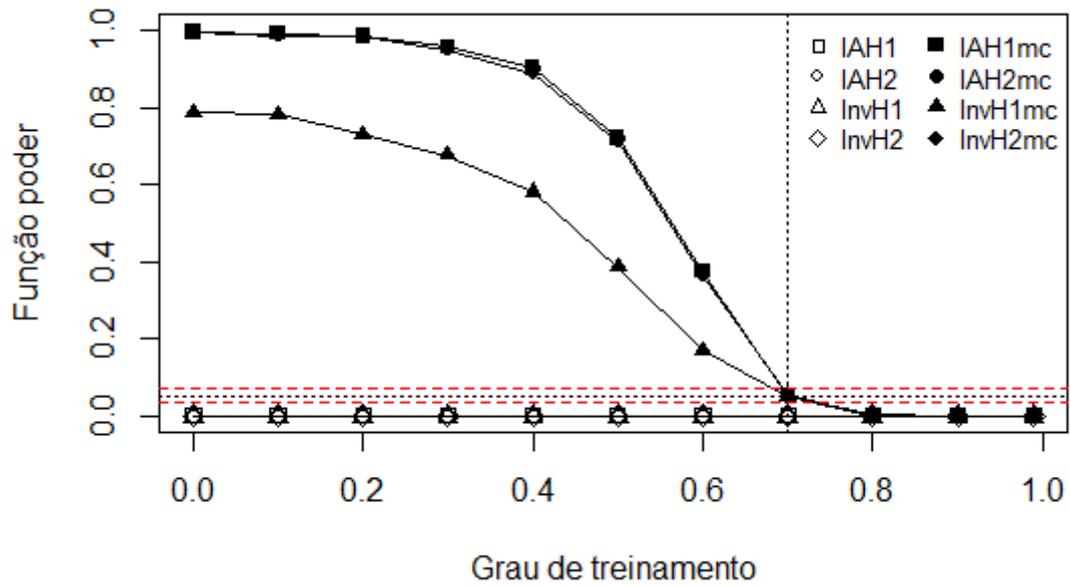
Fonte: da autora.

Figura 16 –  $P = 2$  provedores,  $q = 5$  atributos,  $n = 20$  observações.



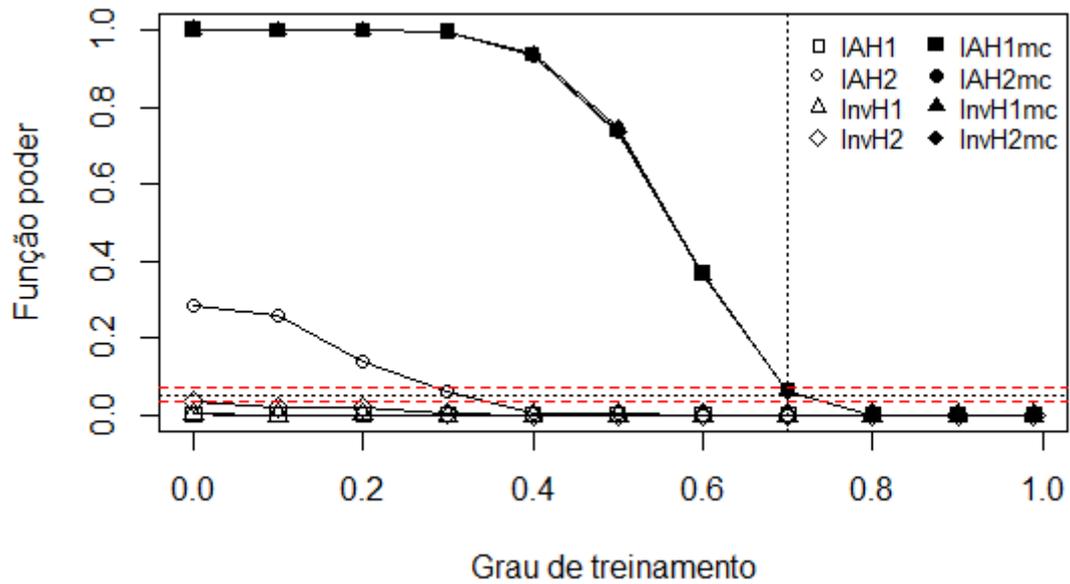
Fonte: da autora.

Figura 17 –  $P = 2$  provadores,  $q = 10$  atributos,  $n = 20$  observações.



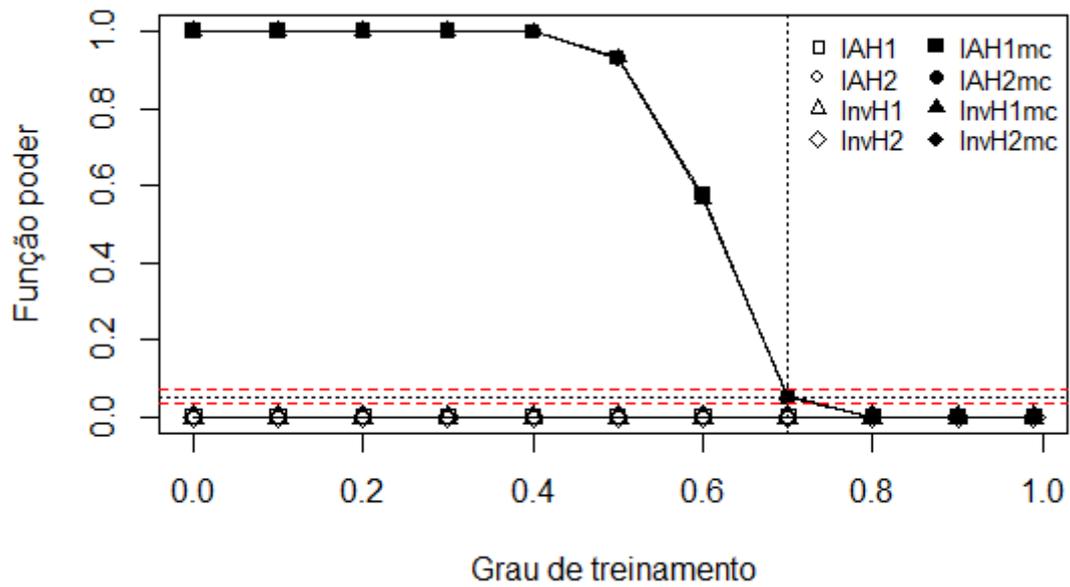
Fonte: da autora.

Figura 18 –  $P = 2$  provadores,  $q = 2$  atributos,  $n = 30$  observações.



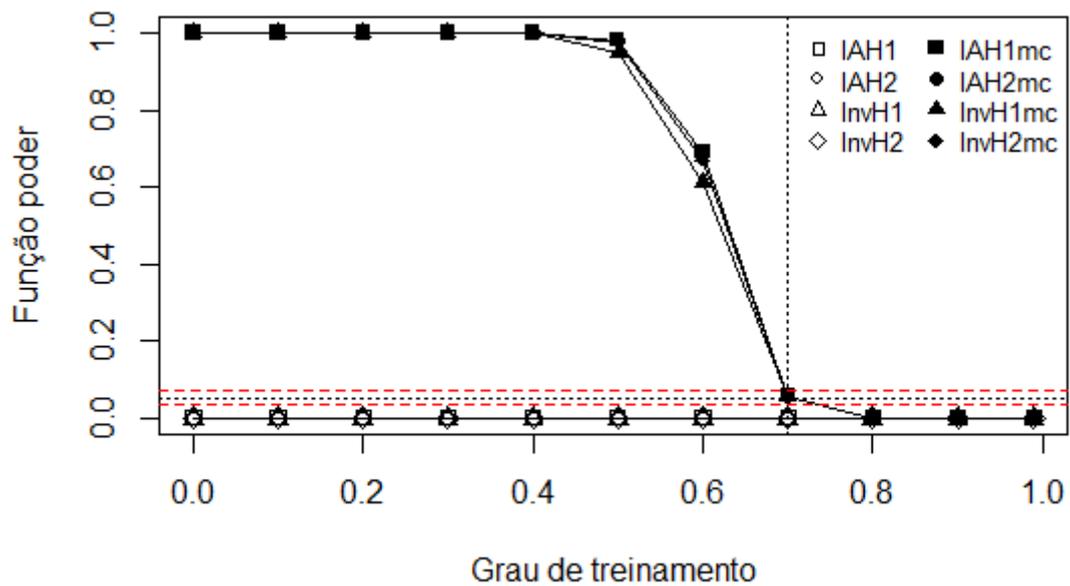
Fonte: da autora.

Figura 19 – P = 2 provadores, q = 5 atributos, n = 30 observações.



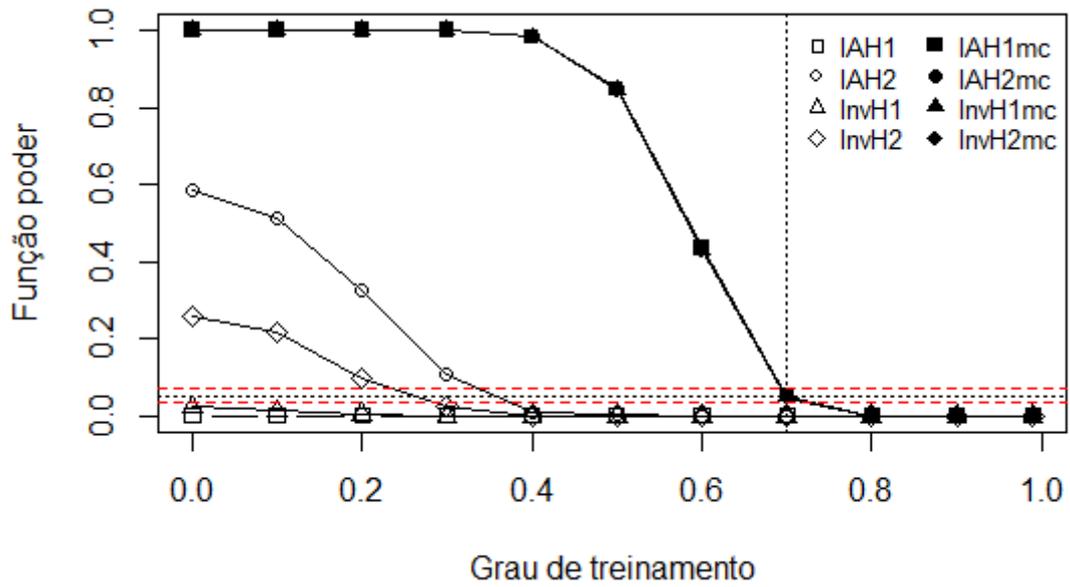
Fonte: da autora.

Figura 20 – P = 2 provadores, q = 10 atributos, n = 30 observações.



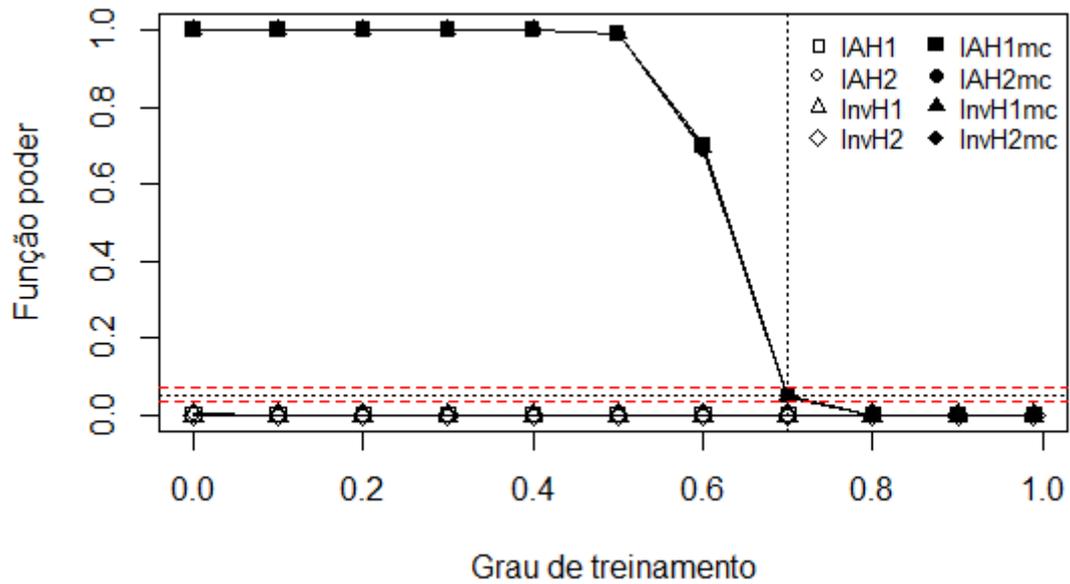
Fonte: da autora.

Figura 21 – P = 2 provedores, q = 2 atributos, n = 40 observações.



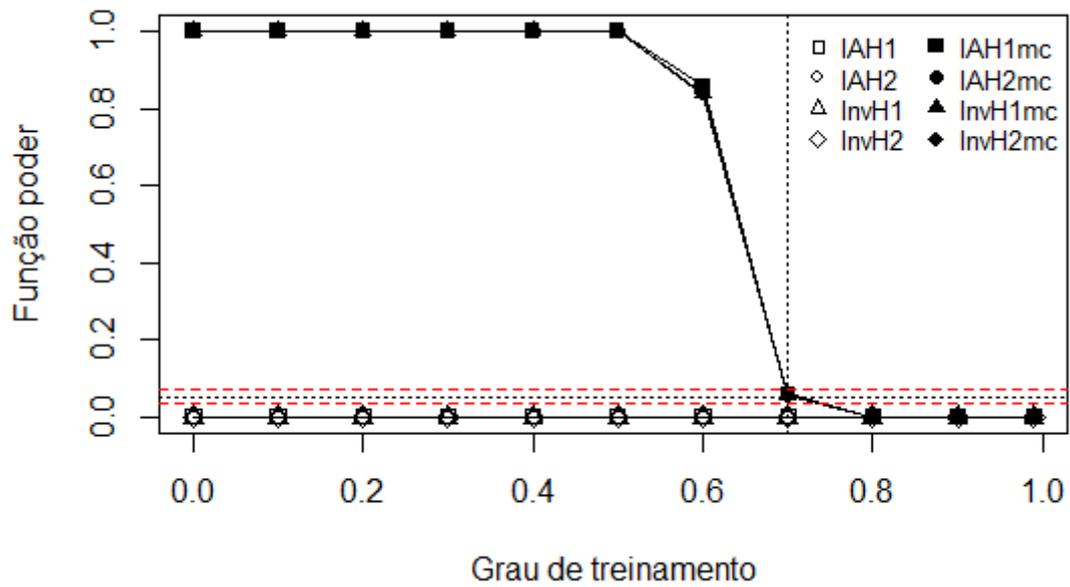
Fonte: da autora.

Figura 22 – P = 2 provedores, q = 5 atributos, n = 40 observações.



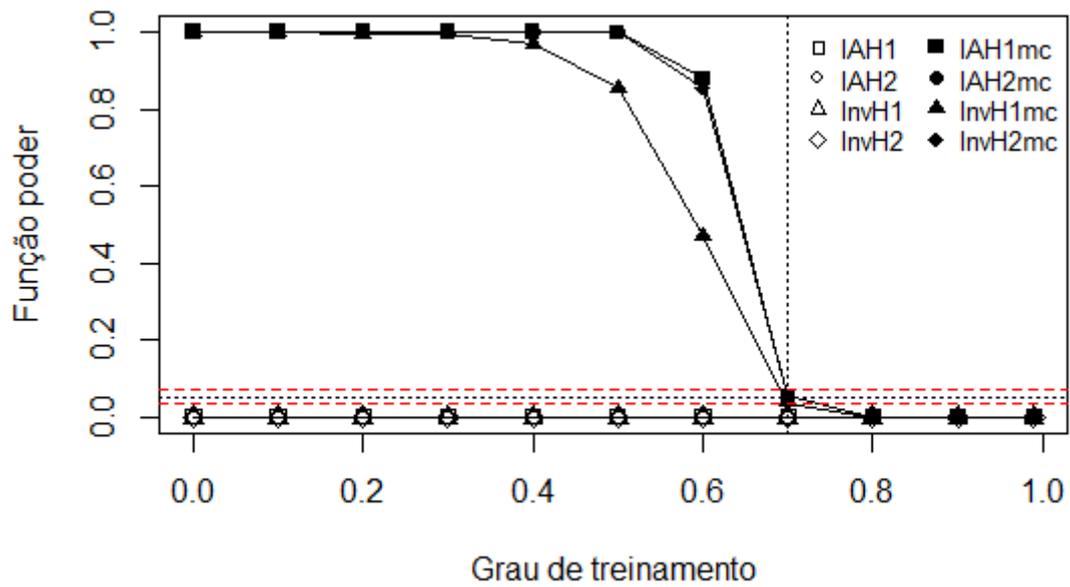
Fonte: da autora.

Figura 23 – P = 2 provadores, q = 10 atributos, n = 40 observações.



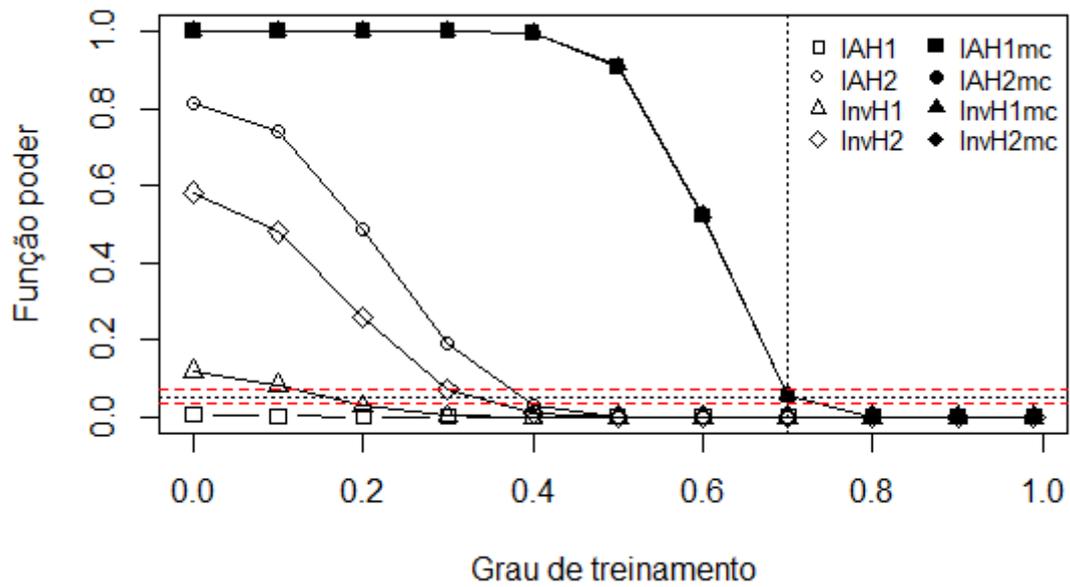
Fonte: da autora.

Figura 24 – P = 2 provadores, q = 20 atributos, n = 40 observações.



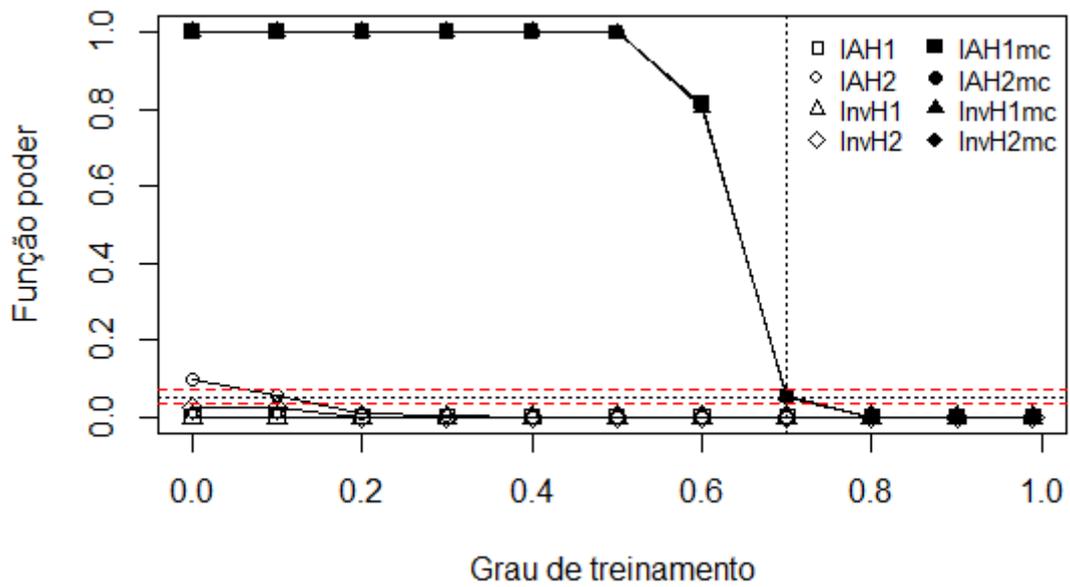
Fonte: da autora.

Figura 25 –  $P = 2$  provadores,  $q = 2$  atributos,  $n = 50$  observações.



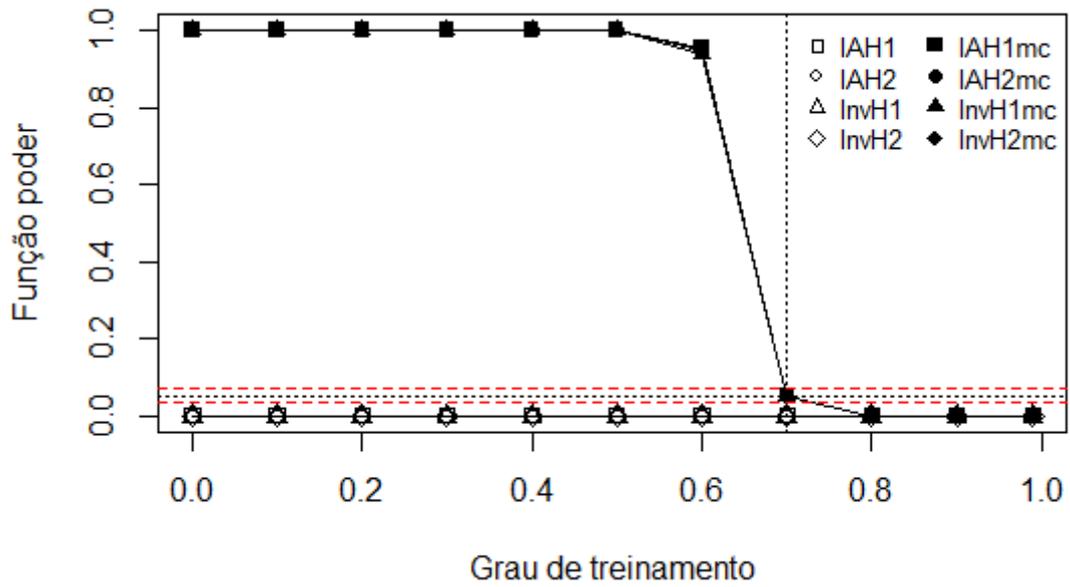
Fonte: da autora.

Figura 26 –  $P = 2$  provadores,  $q = 5$  atributos,  $n = 50$  observações.



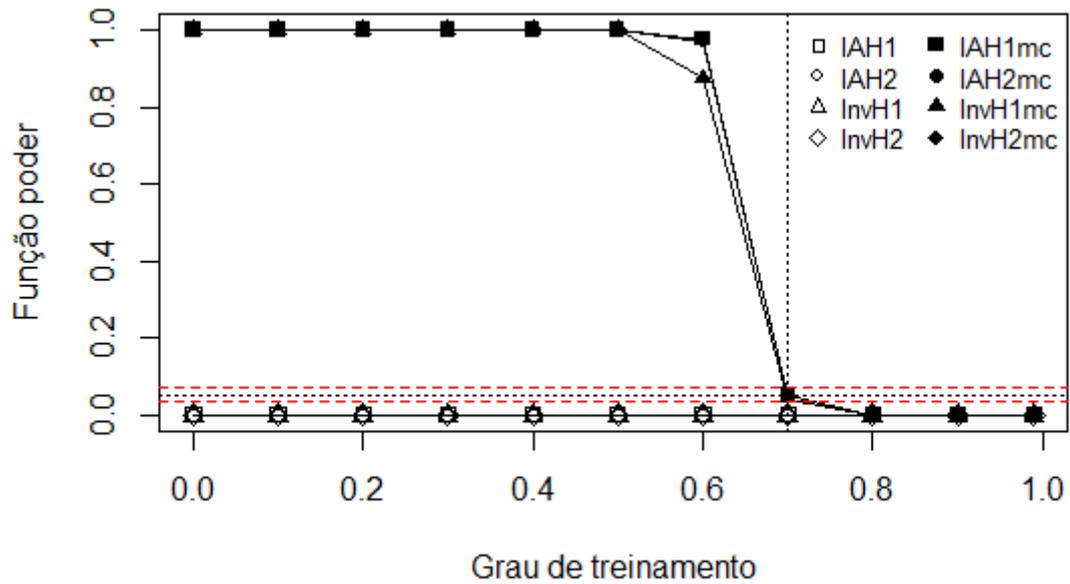
Fonte: da autora.

Figura 27 – P = 2 provadores, q = 10 atributos, n = 50 observações.



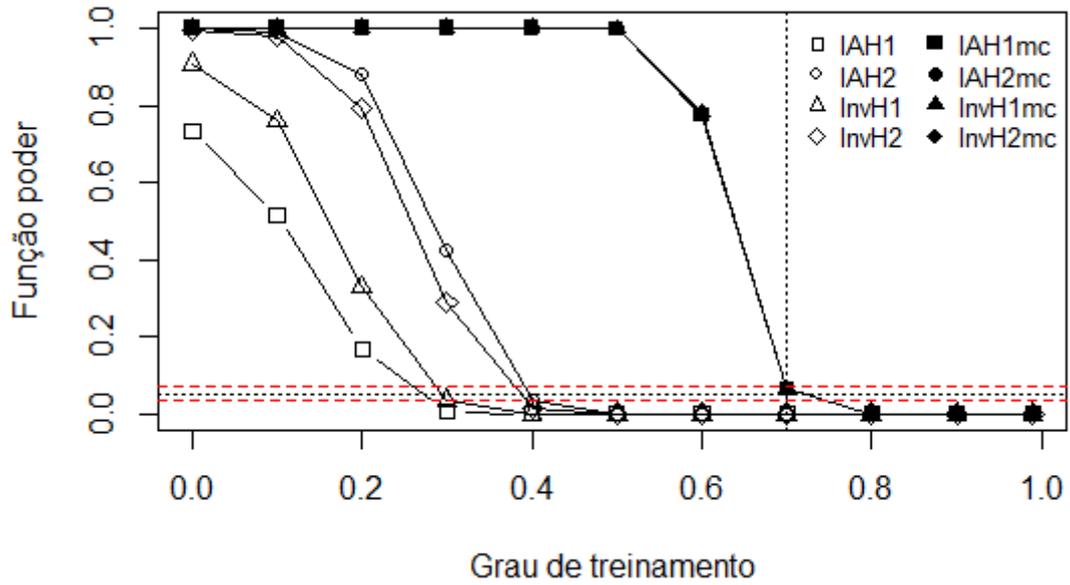
Fonte: da autora.

Figura 28 – P = 2 provadores, q = 20 atributos, n = 50 observações.



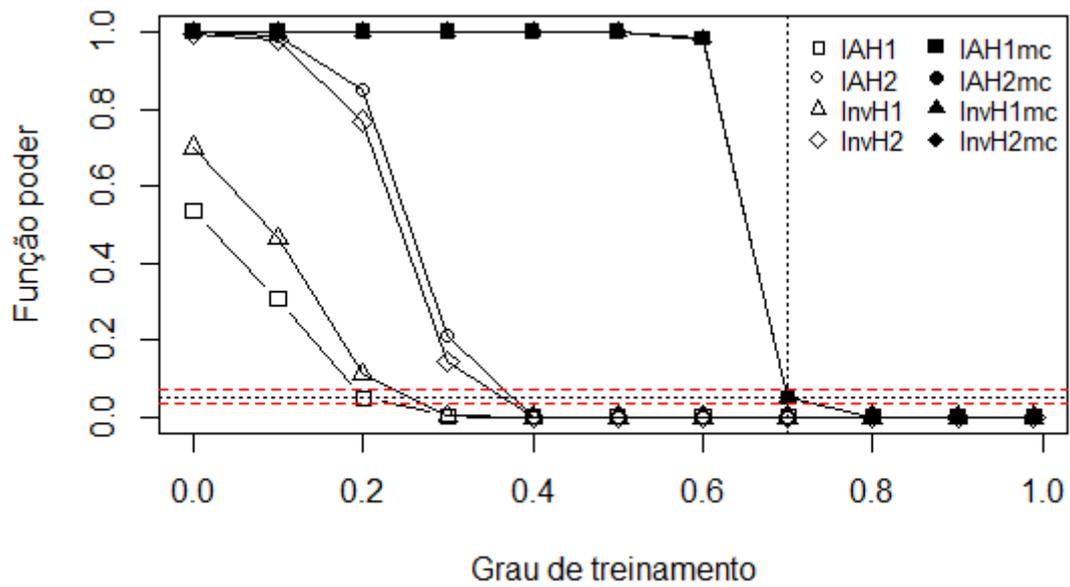
Fonte: da autora.

Figura 29 –  $P = 2$  provedores,  $q = 2$  atributos,  $n = 100$  observações.



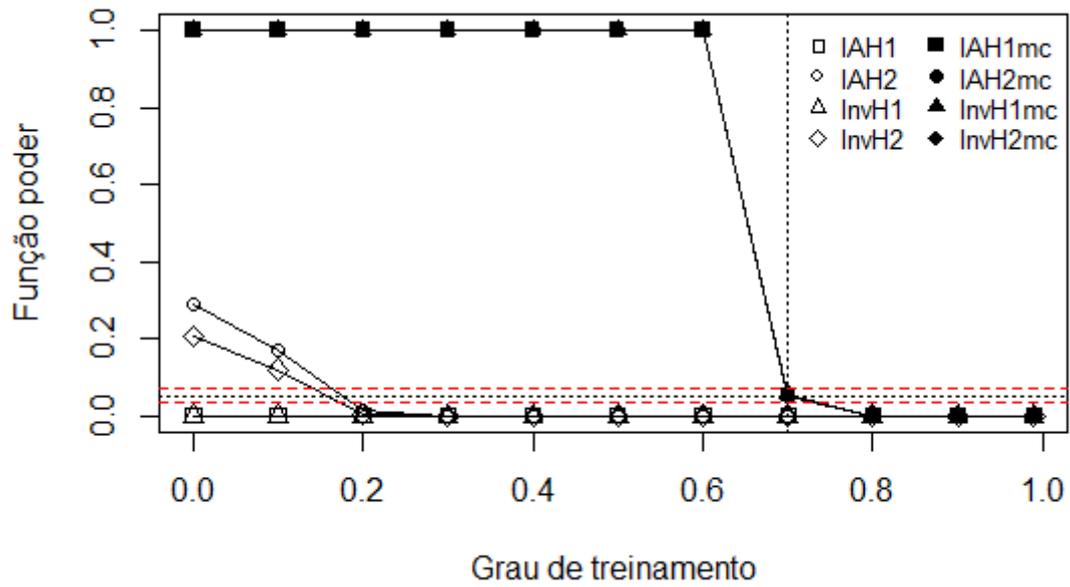
Fonte: da autora.

Figura 30 –  $P = 2$  provedores,  $q = 5$  atributos,  $n = 100$  observações.



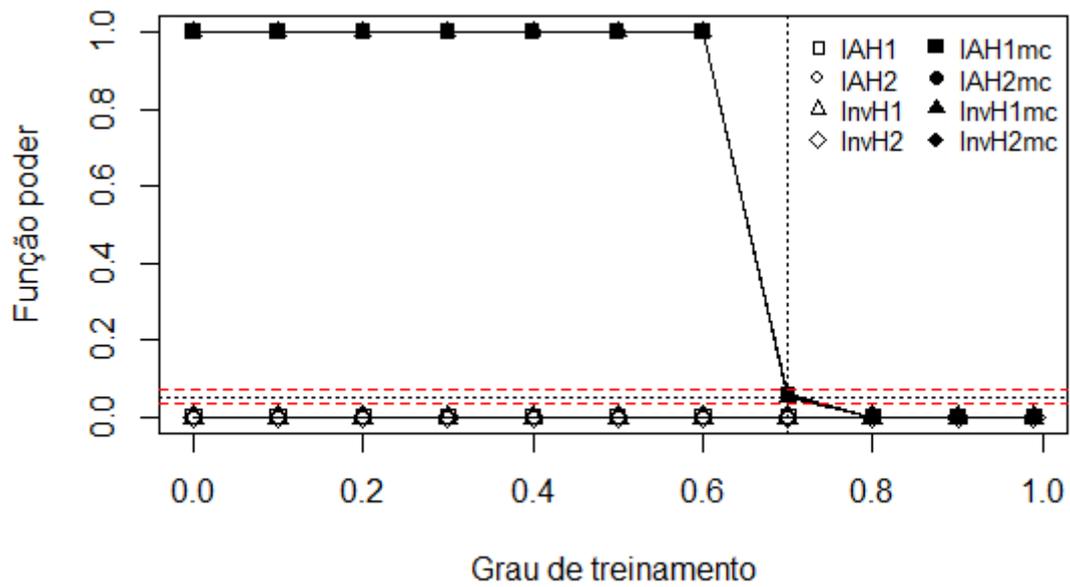
Fonte: da autora.

Figura 31 –  $P = 2$  provadores,  $q = 10$  atributos,  $n = 100$  observações.



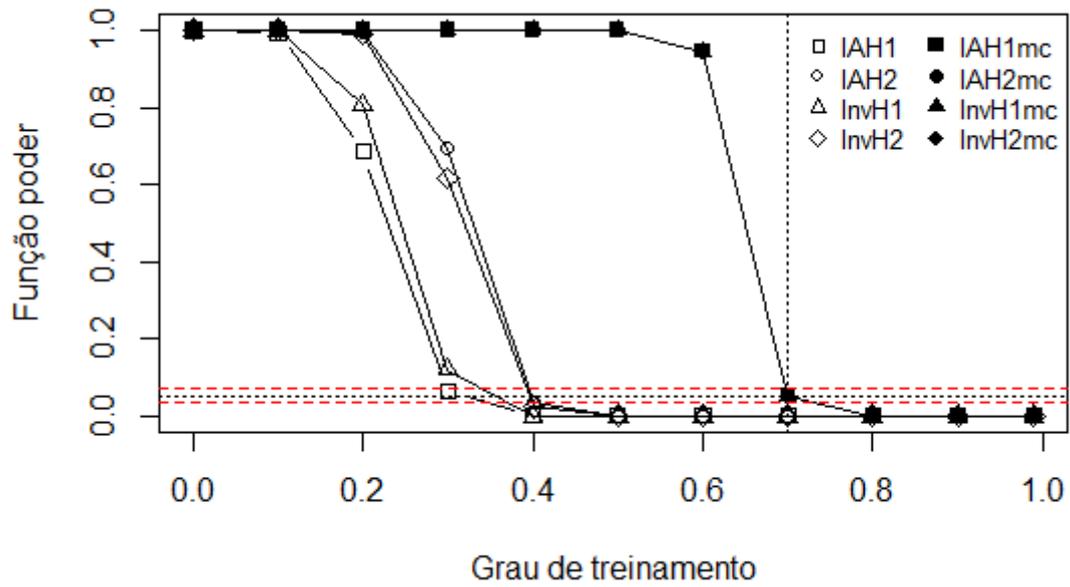
Fonte: da autora.

Figura 32 –  $P = 2$  provadores,  $q = 20$  atributos,  $n = 100$  observações.



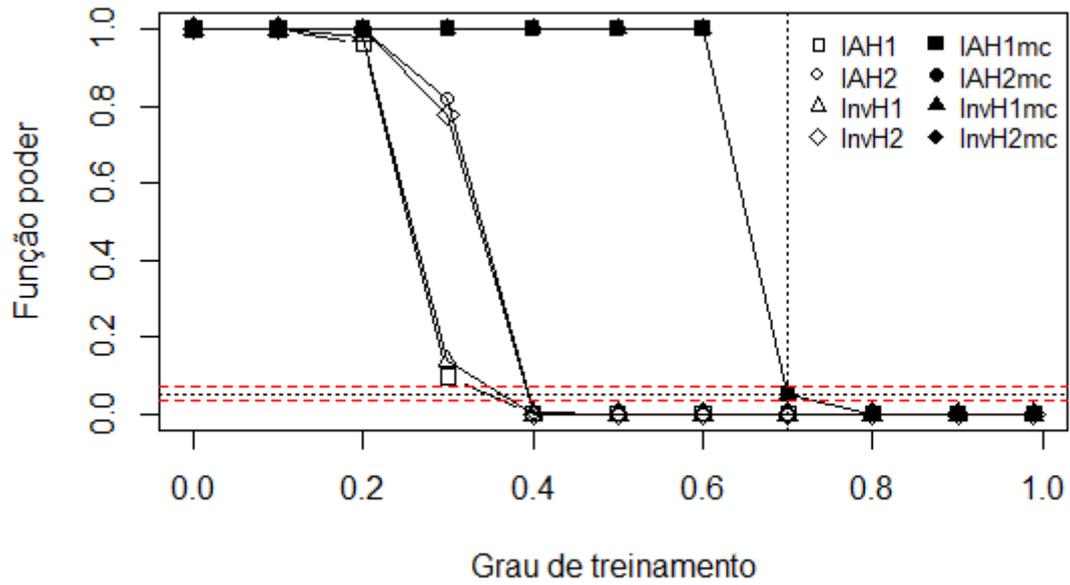
Fonte: da autora.

Figura 33 –  $P = 2$  provadores,  $q = 2$  atributos,  $n = 200$  observações.



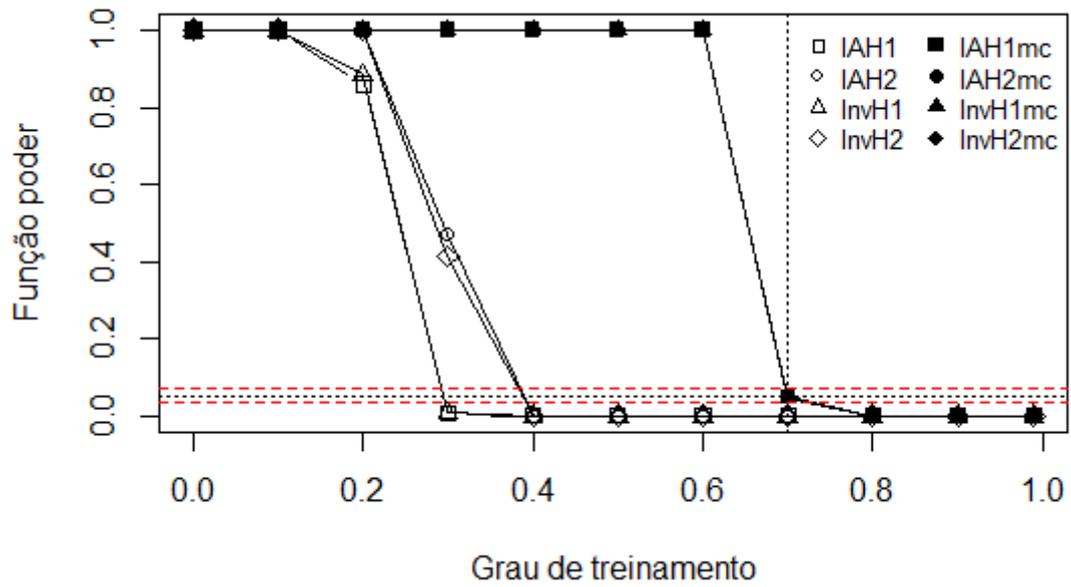
Fonte: da autora.

Figura 34 –  $P = 2$  provadores,  $q = 5$  atributos,  $n = 200$  observações.



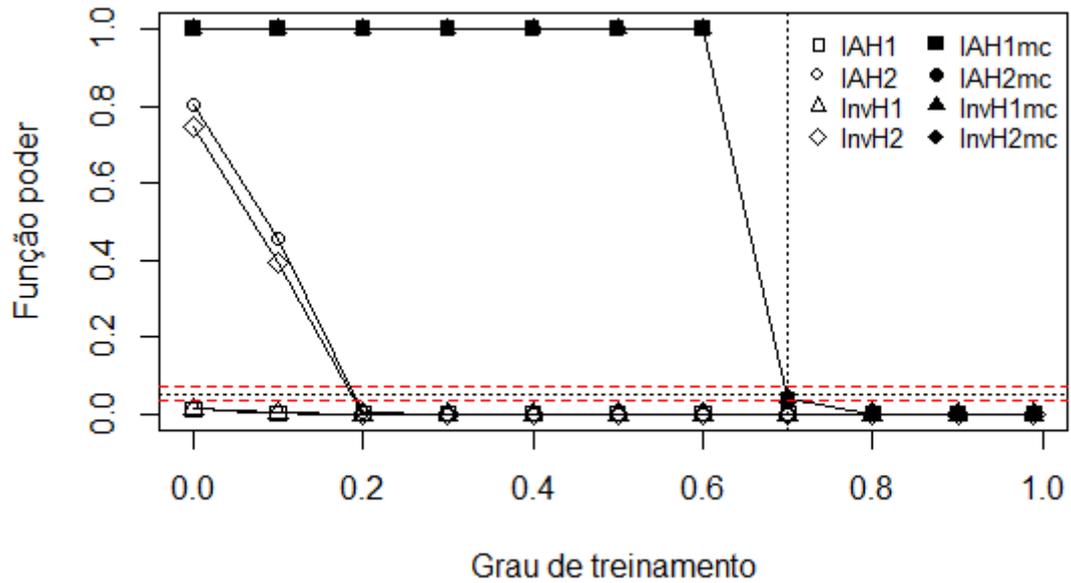
Fonte: da autora.

Figura 35 –  $P = 2$  provadores,  $q = 10$  atributos,  $n = 200$  observações.



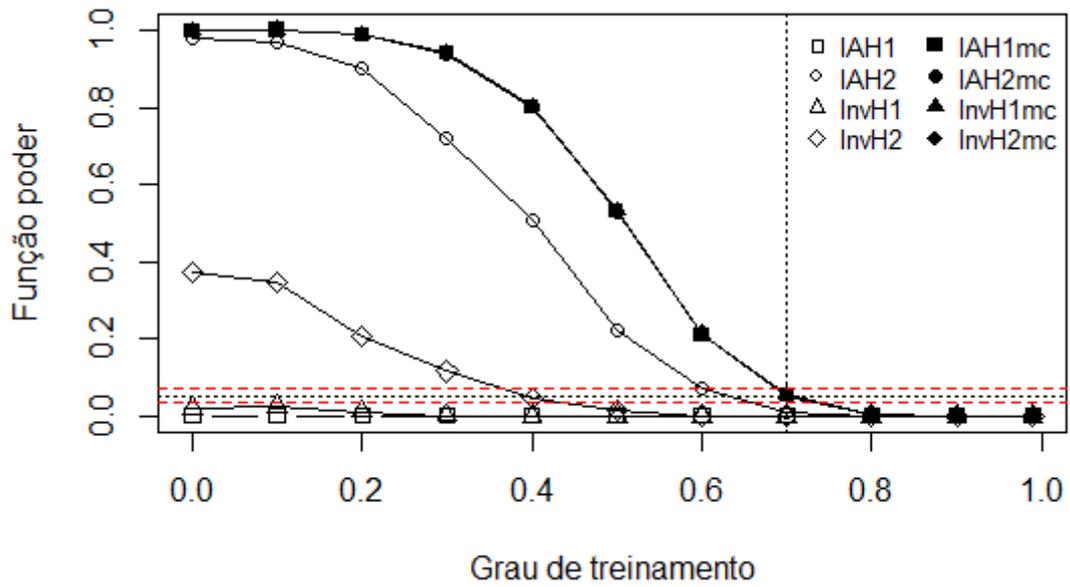
Fonte: da autora.

Figura 36 –  $P = 2$  provadores,  $q = 20$  atributos,  $n = 200$  observações.



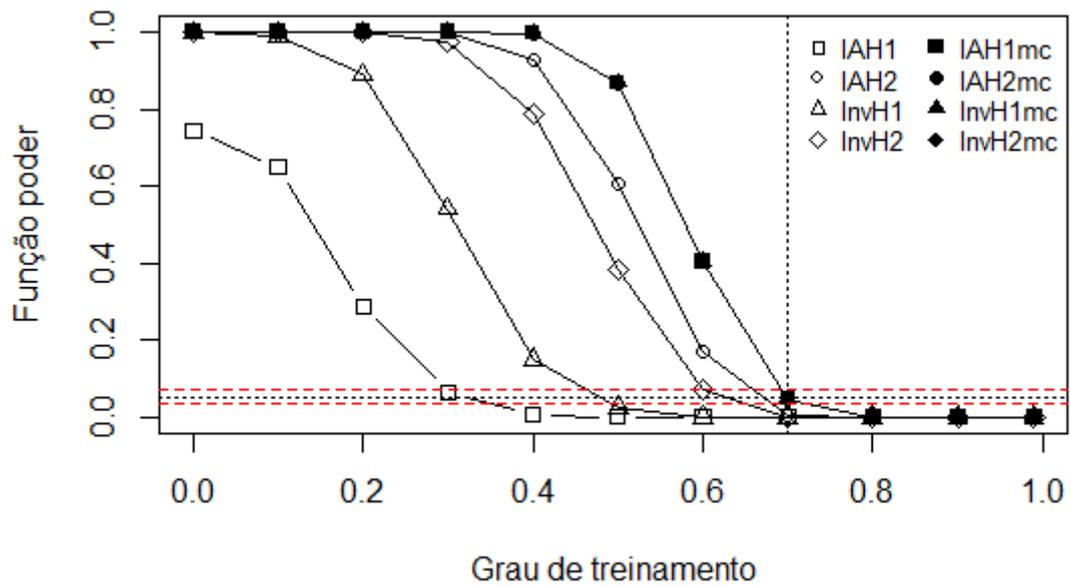
Fonte: da autora.

Figura 37 – P = 5 provedores, q = 2 atributos, n = 10 observações.



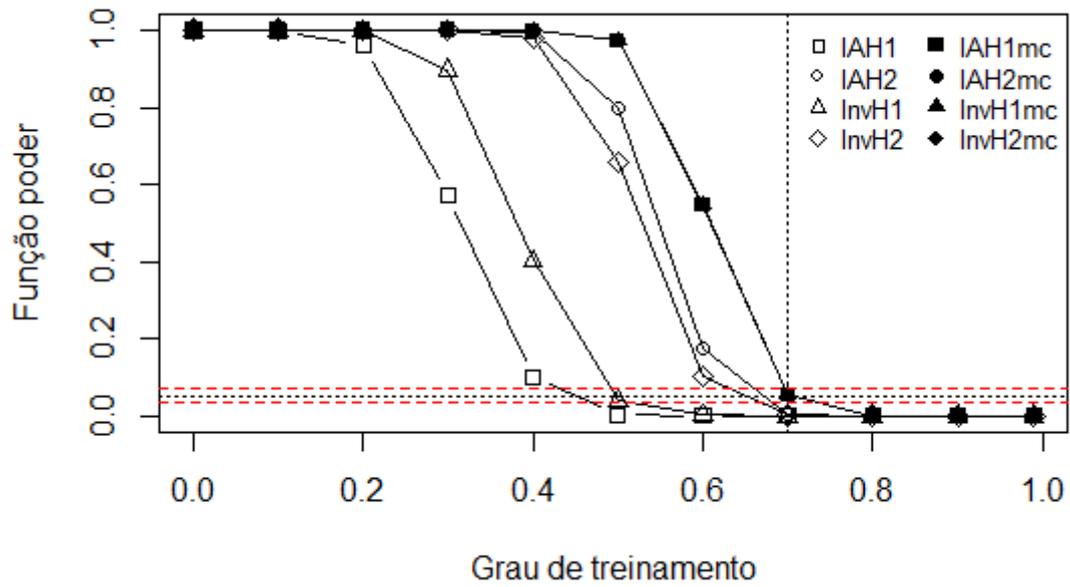
Fonte: da autora.

Figura 38 – P = 5 provedores, q = 2 atributos, n = 20 observações.



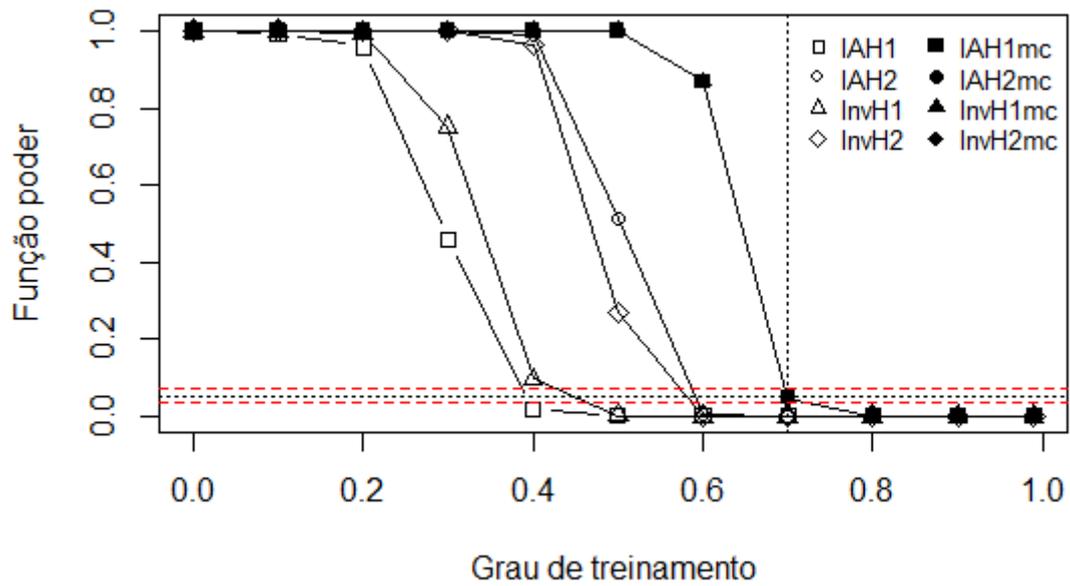
Fonte: da autora.

Figura 39 –  $P = 5$  provedores,  $q = 2$  atributos,  $n = 30$  observações.



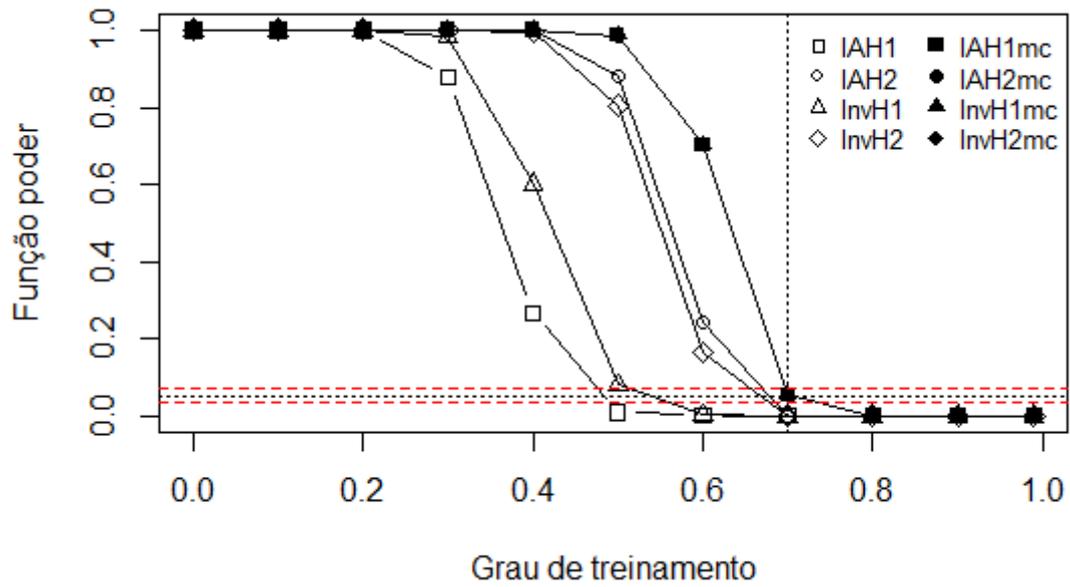
Fonte: da autora.

Figura 40 –  $P = 5$  provedores,  $q = 5$  atributos,  $n = 30$  observações.



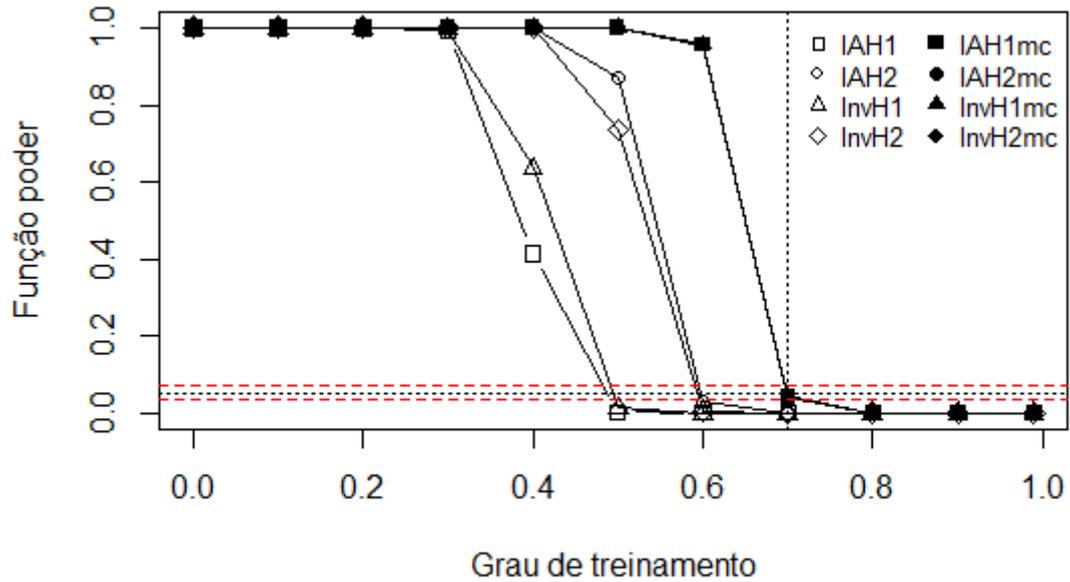
Fonte: da autora.

Figura 41 –  $P = 5$  provadores,  $q = 2$  atributos,  $n = 40$  observações.



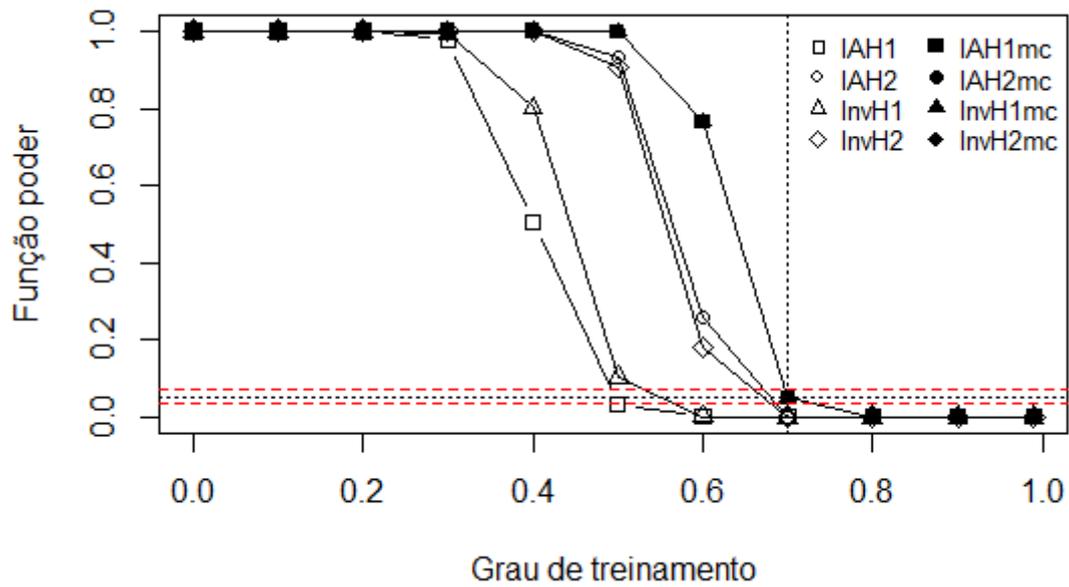
Fonte: da autora.

Figura 42 –  $P = 5$  provadores,  $q = 5$  atributos,  $n = 40$  observações.



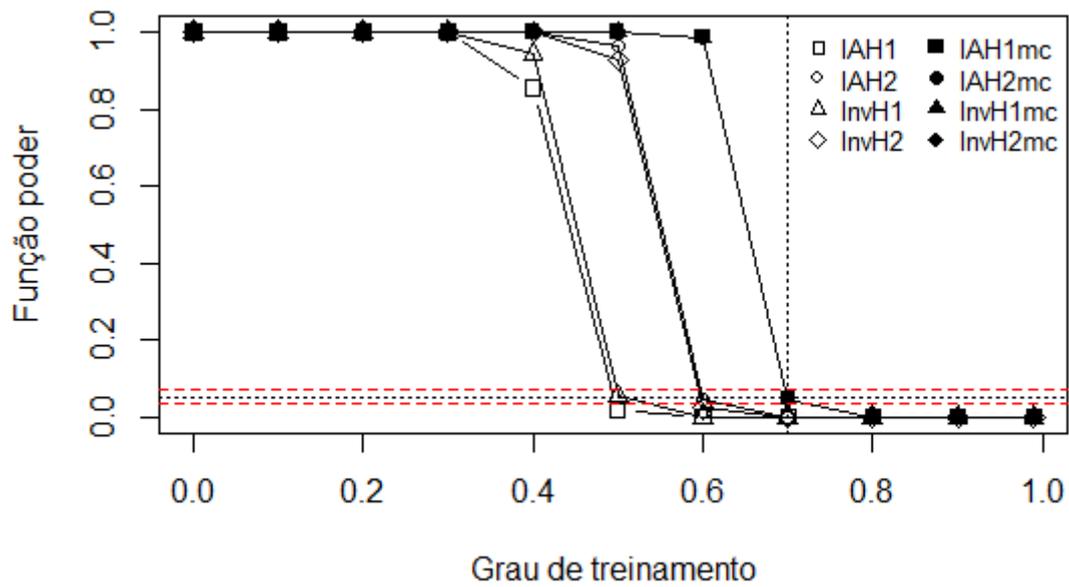
Fonte: da autora.

Figura 43 – P = 5 provedores, q = 2 atributos, n = 50 observações.



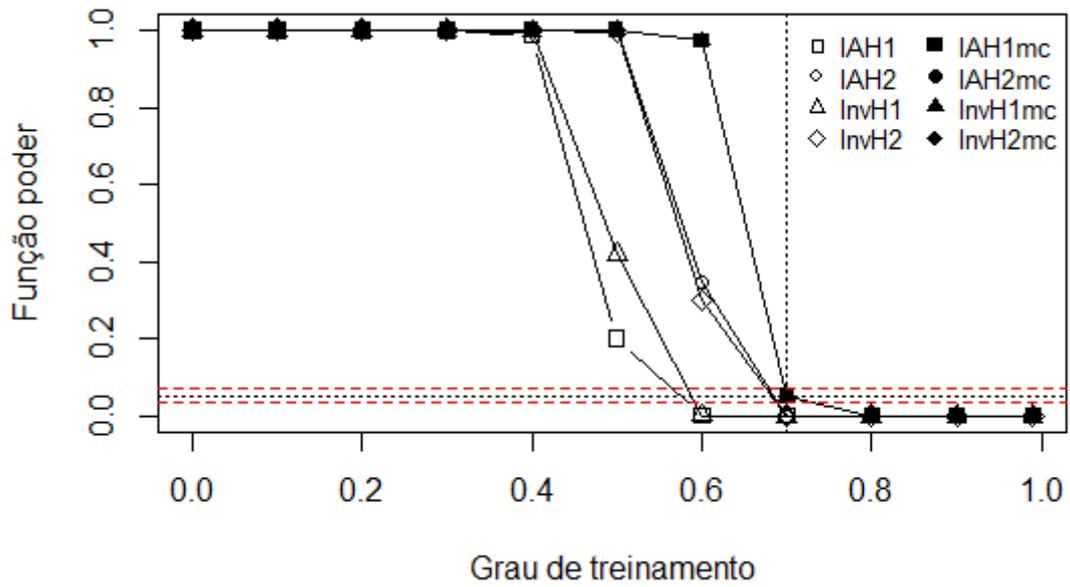
Fonte: da autora.

Figura 44 – P = 5 provedores, q = 10 atributos, n = 50 observações.



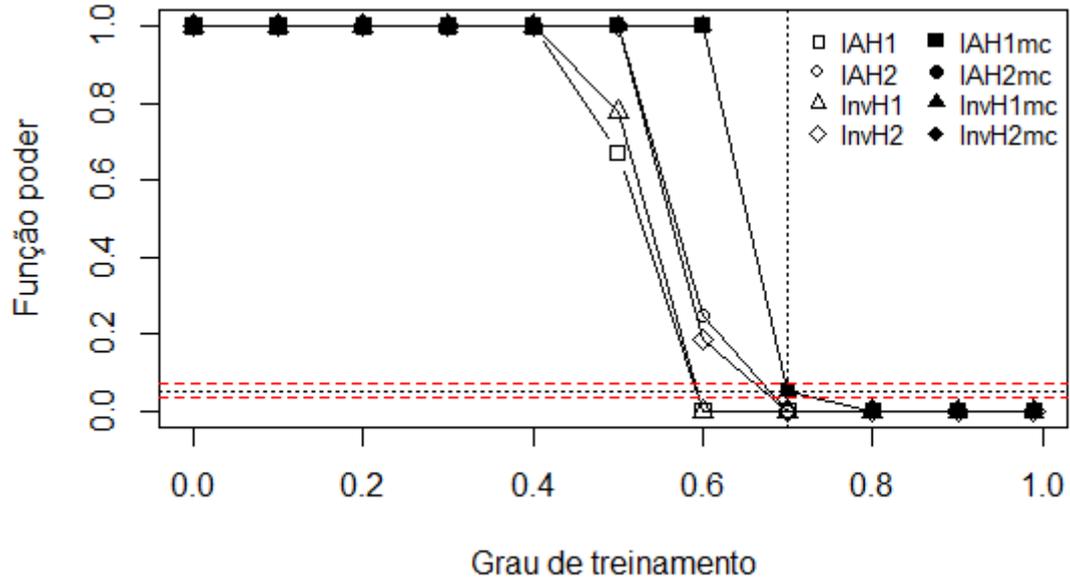
Fonte: da autora.

Figura 45 –  $P = 5$  provedores,  $q = 2$  atributos,  $n = 100$  observações.



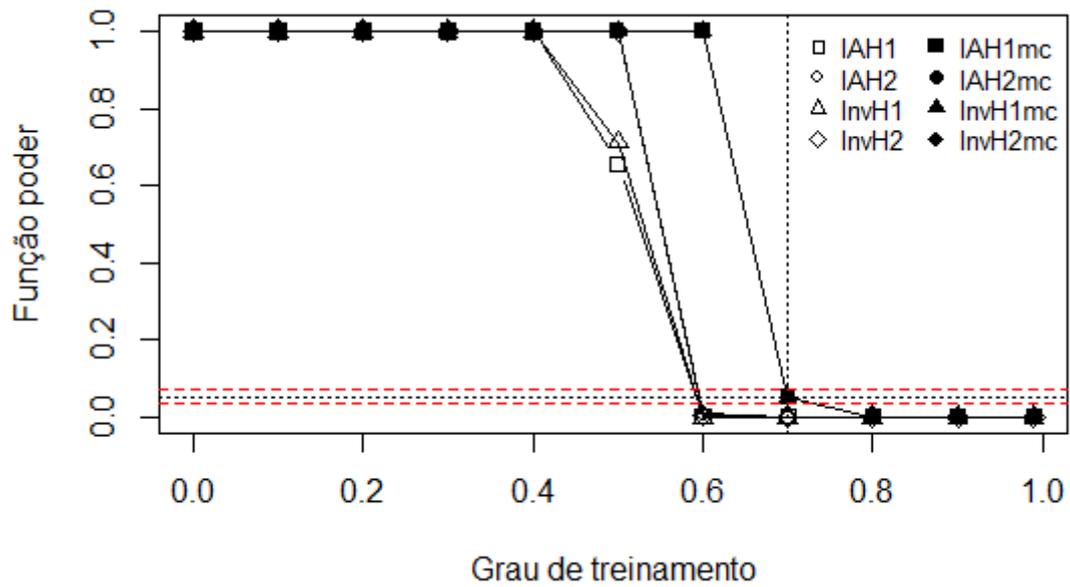
Fonte: da autora.

Figura 46 –  $P = 5$  provedores,  $q = 5$  atributos,  $n = 100$  observações.



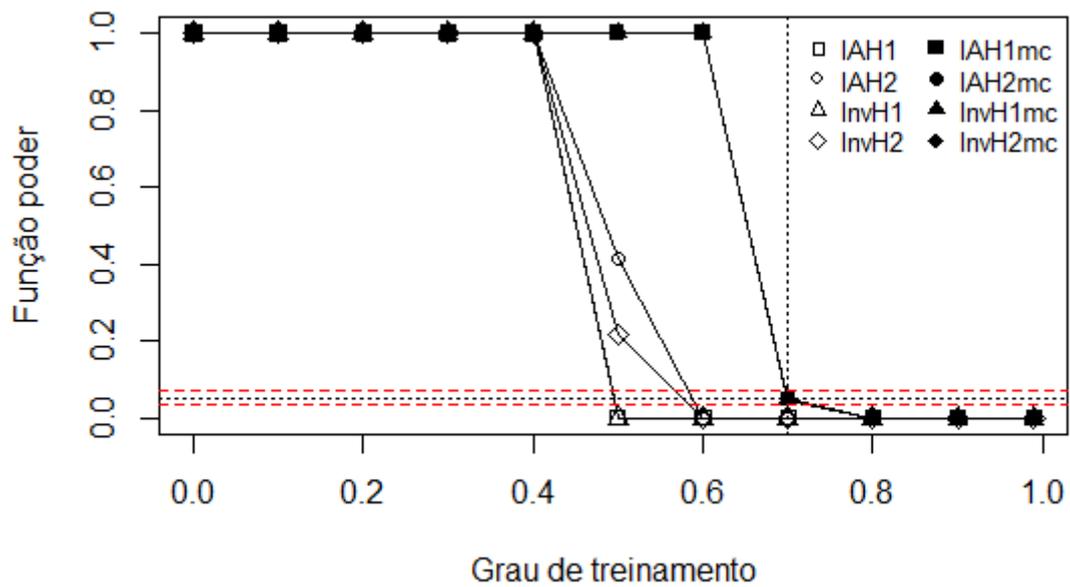
Fonte: da autora.

Figura 47 –  $P = 5$  provadores,  $q = 10$  atributos,  $n = 100$  observações.



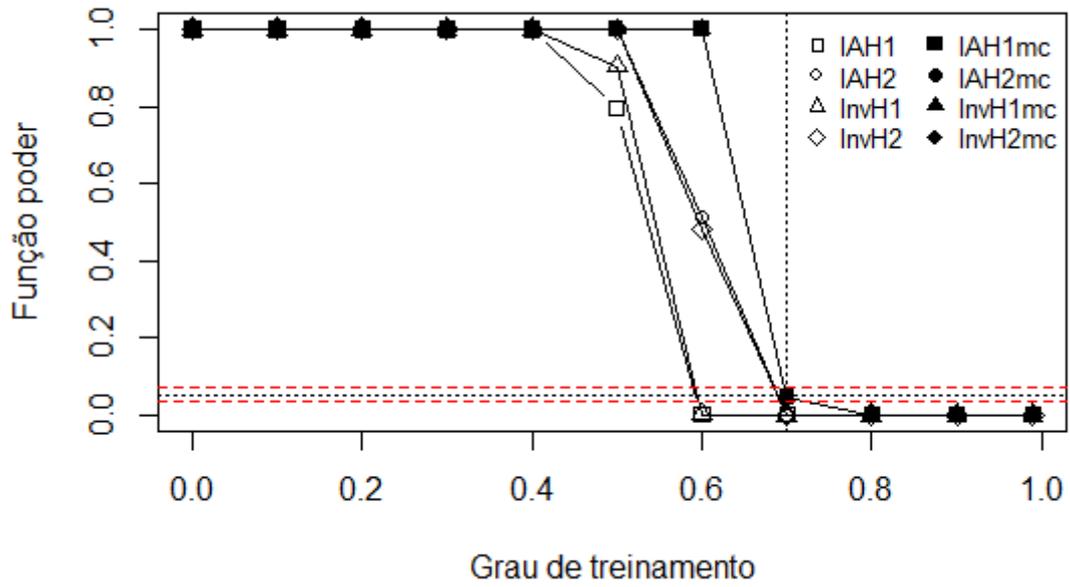
Fonte: da autora.

Figura 48 –  $P = 5$  provadores,  $q = 20$  atributos,  $n = 100$  observações.



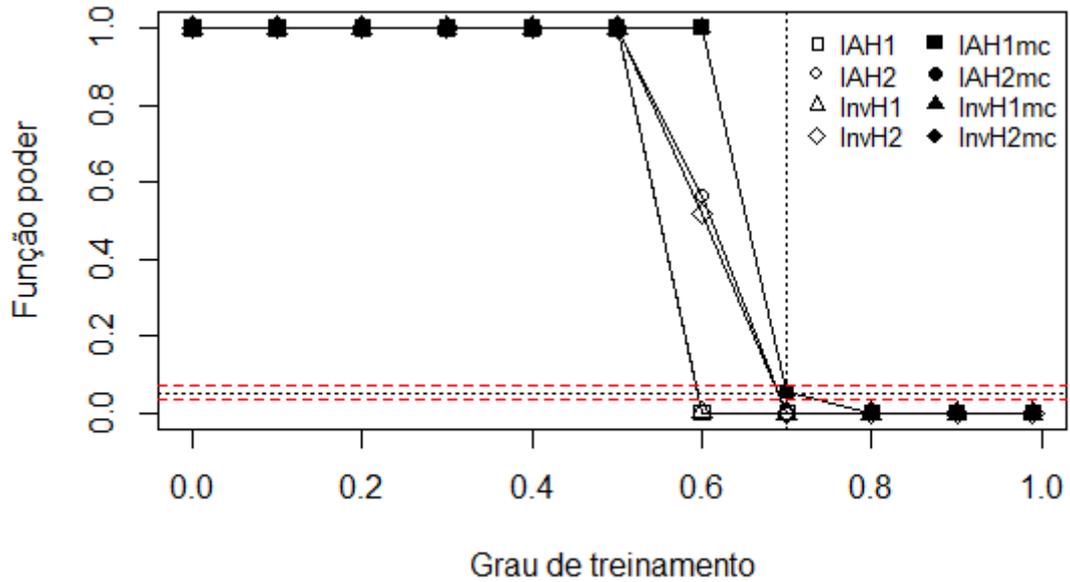
Fonte: da autora.

Figura 49 – P = 5 provadores, q = 2 atributos, n = 200 observações.



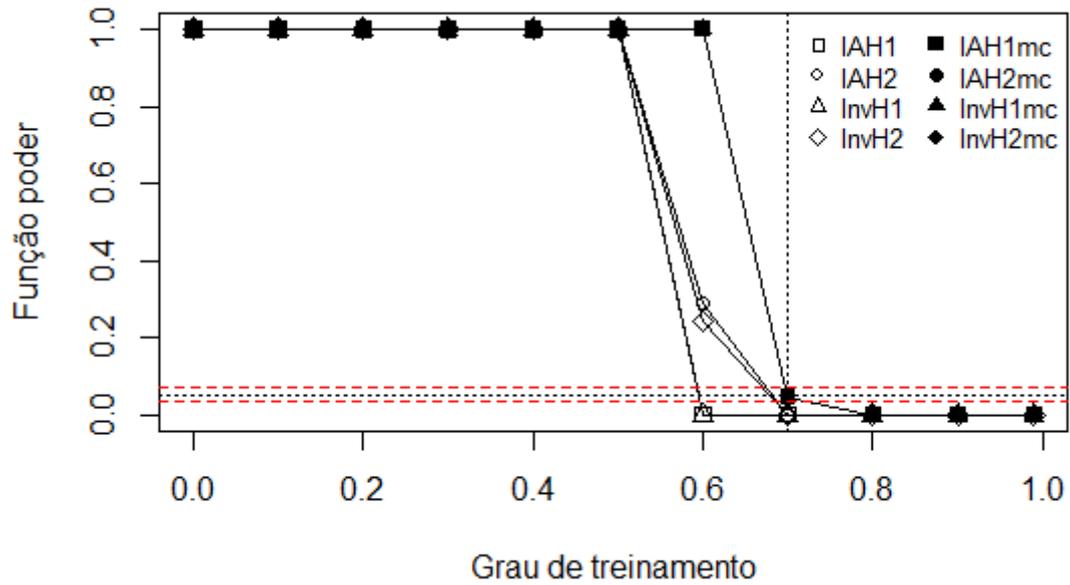
Fonte: da autora.

Figura 50 – P = 5 provadores, q = 5 atributos, n = 200 observações.



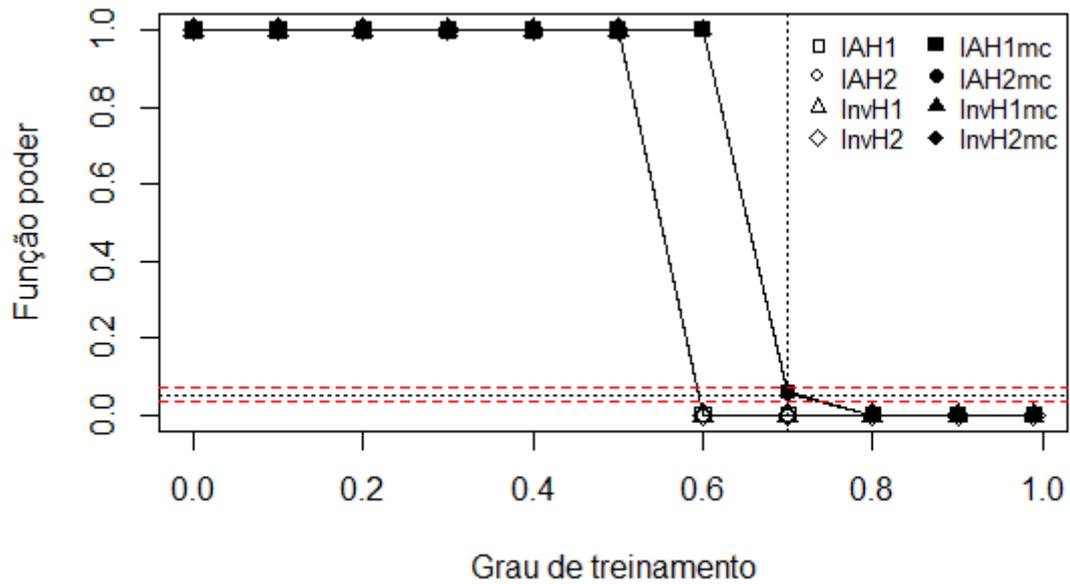
Fonte: da autora.

Figura 51 –  $P = 5$  provedores,  $q = 10$  atributos,  $n = 200$  observações.



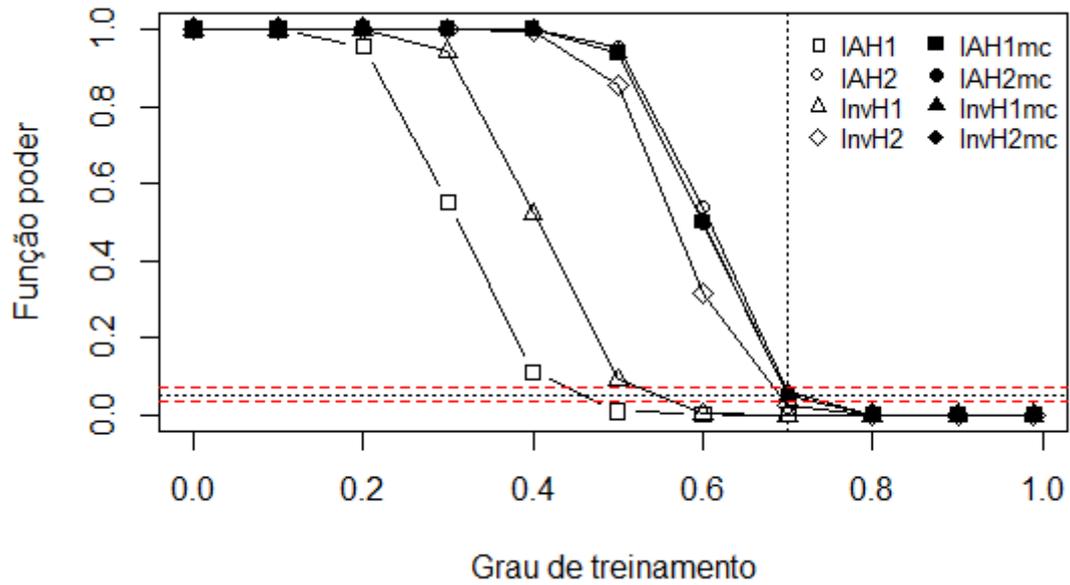
Fonte: da autora.

Figura 52 –  $P = 5$  provedores,  $q = 20$  atributos,  $n = 200$  observações.



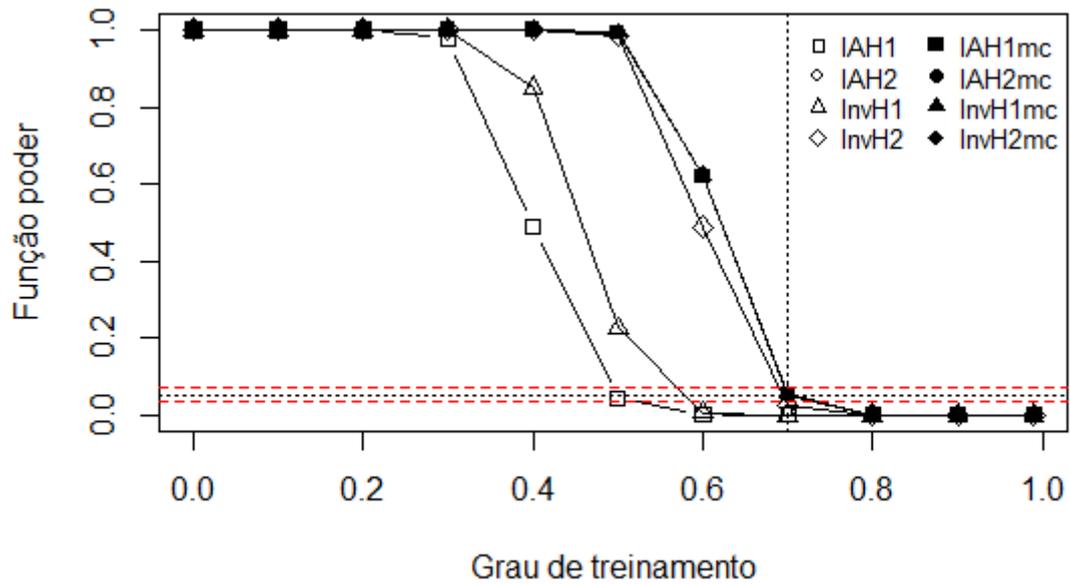
Fonte: da autora.

Figura 53 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 20$  observações.



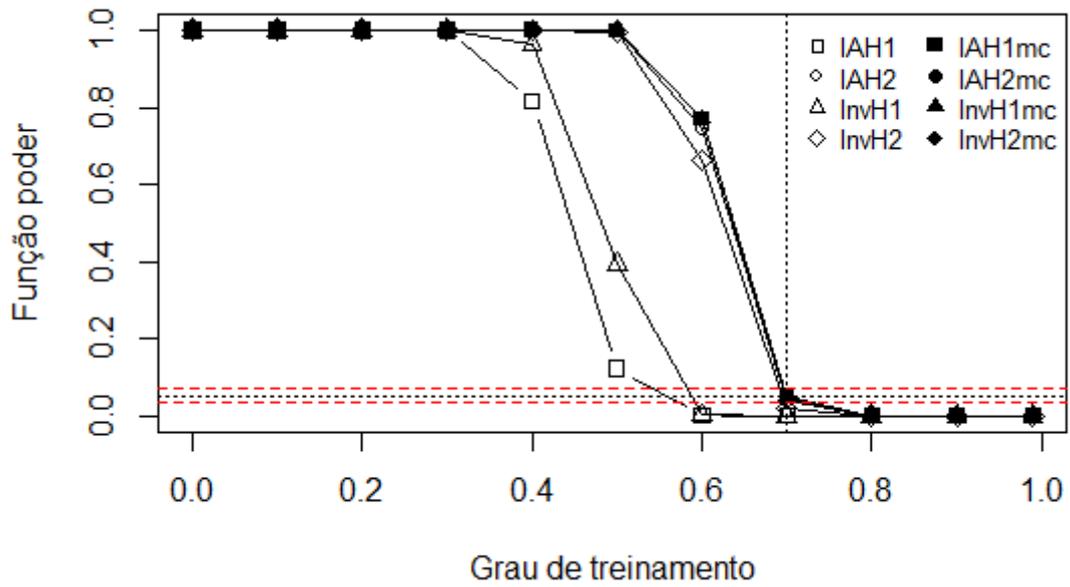
Fonte: da autora.

Figura 54 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 30$  observações.



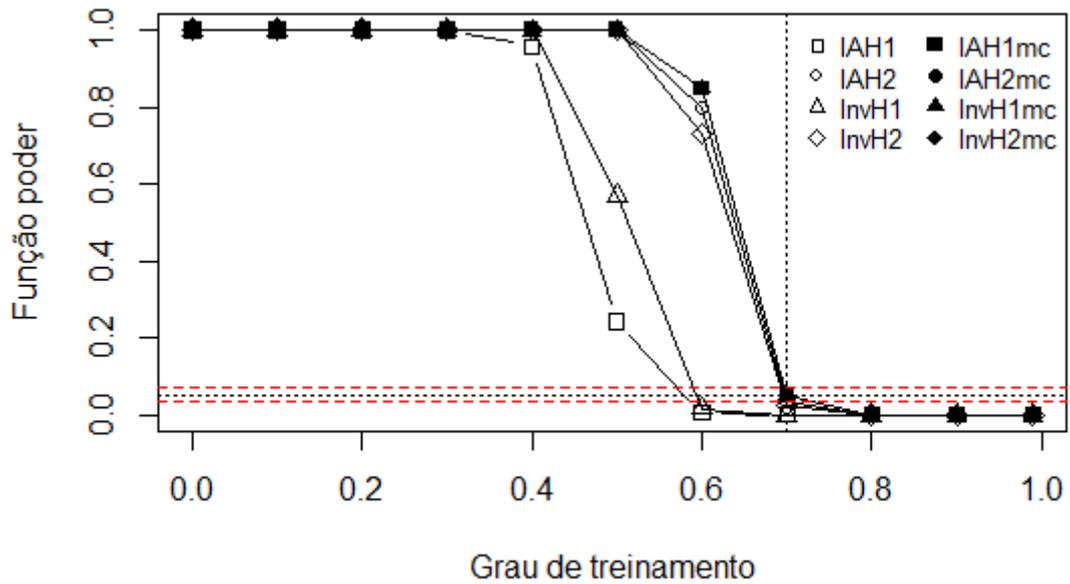
Fonte: da autora.

Figura 55 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 40$  observações.



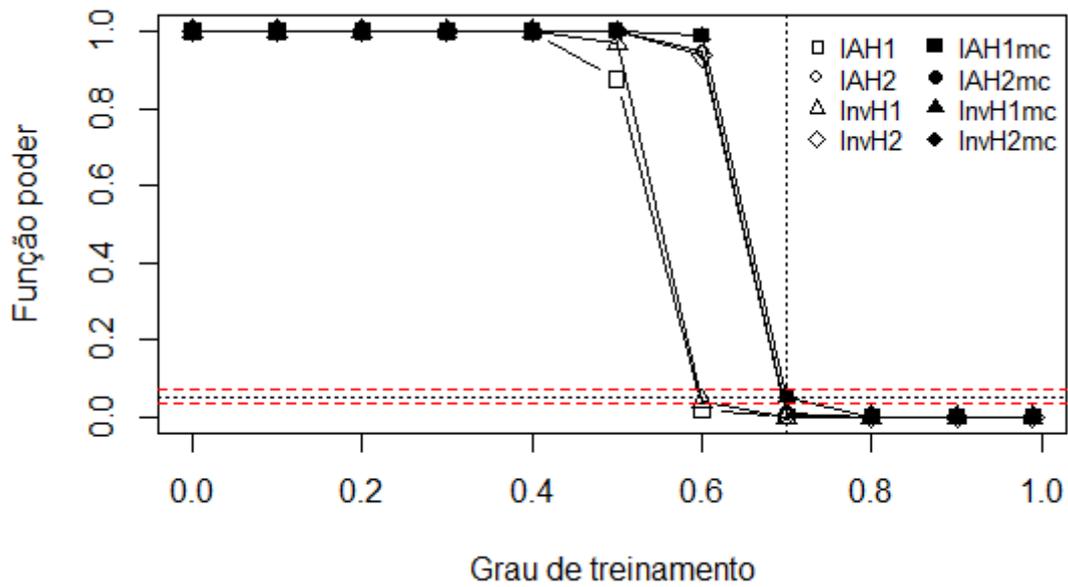
Fonte: da autora.

Figura 56 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 50$  observações.



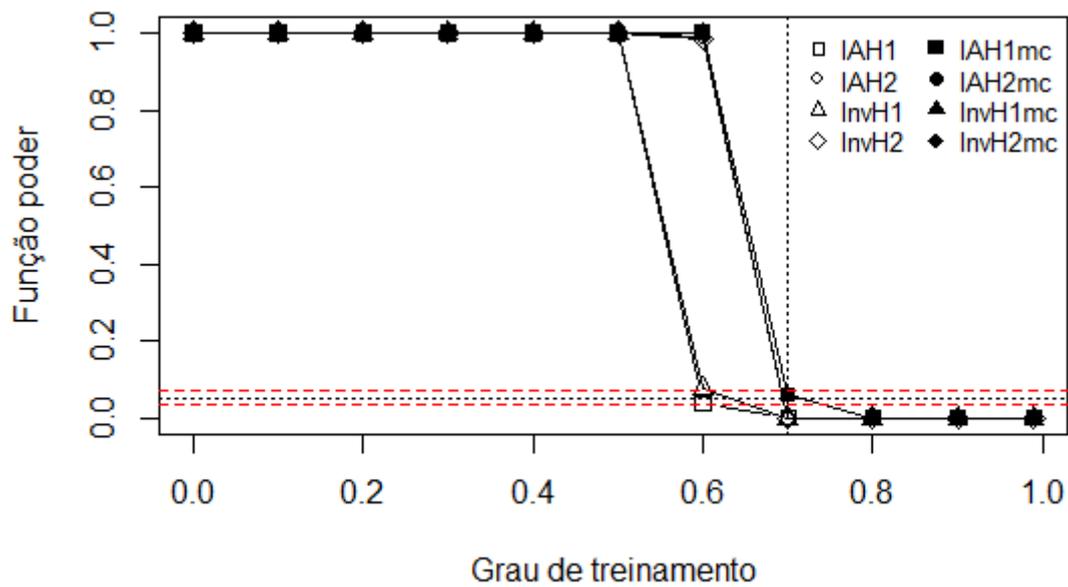
Fonte: da autora.

Figura 57 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 100$  observações.



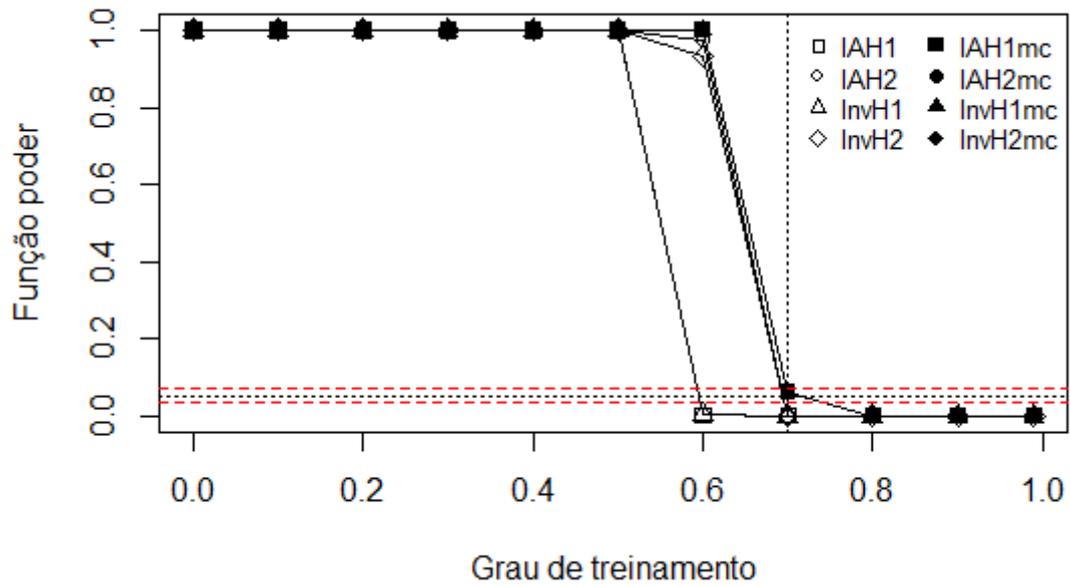
Fonte: da autora.

Figura 58 –  $P = 10$  provedores,  $q = 5$  atributos,  $n = 100$  observações.



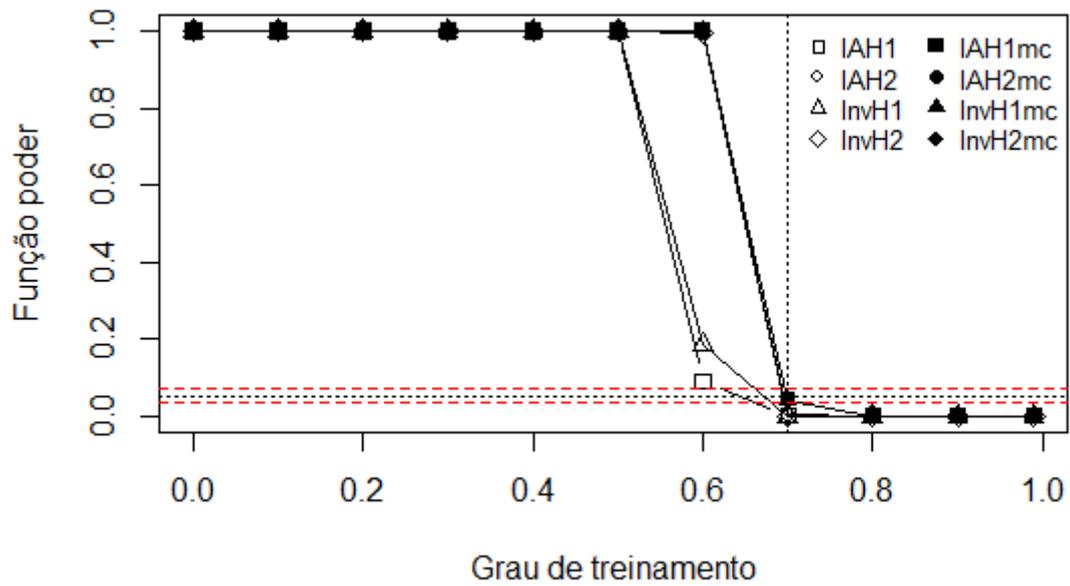
Fonte: da autora.

Figura 59 –  $P = 10$  provedores,  $q = 10$  atributos,  $n = 100$  observações.



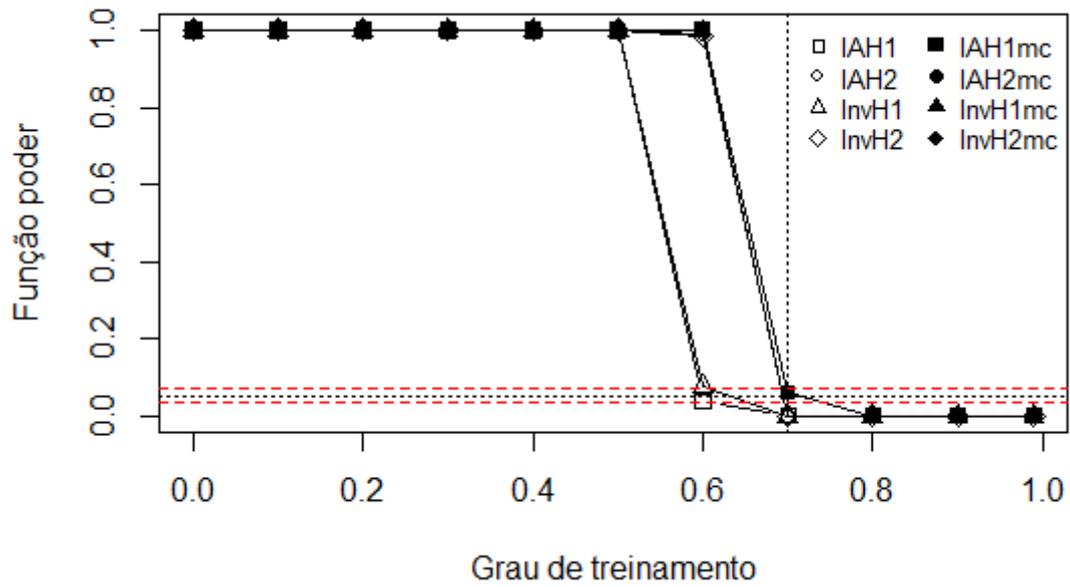
Fonte: da autora.

Figura 60 –  $P = 10$  provedores,  $q = 2$  atributos,  $n = 200$  observações.



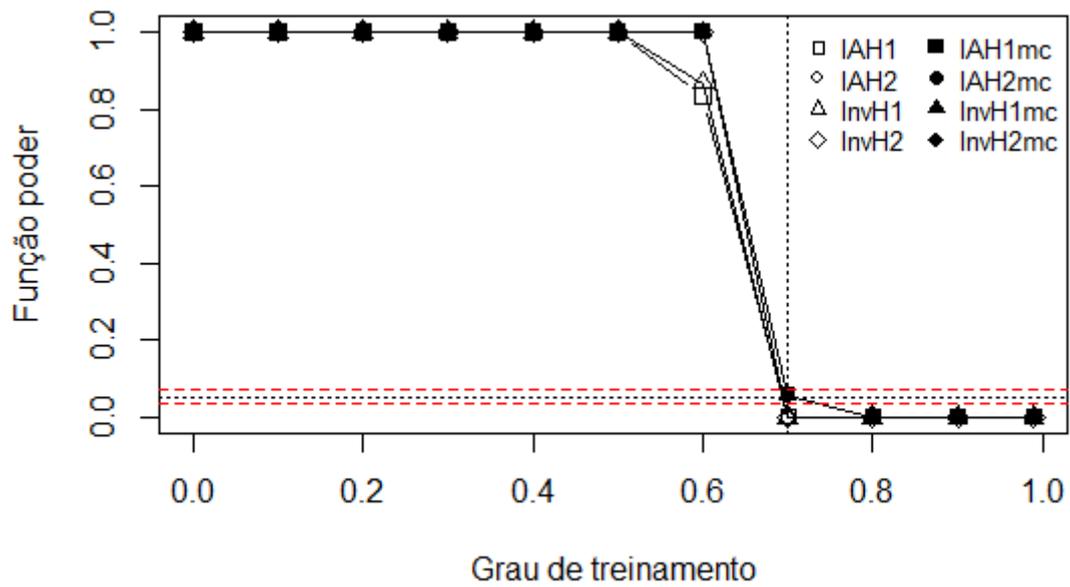
Fonte: da autora.

Figura 61 –  $P = 10$  provedores,  $q = 5$  atributos,  $n = 200$  observações.



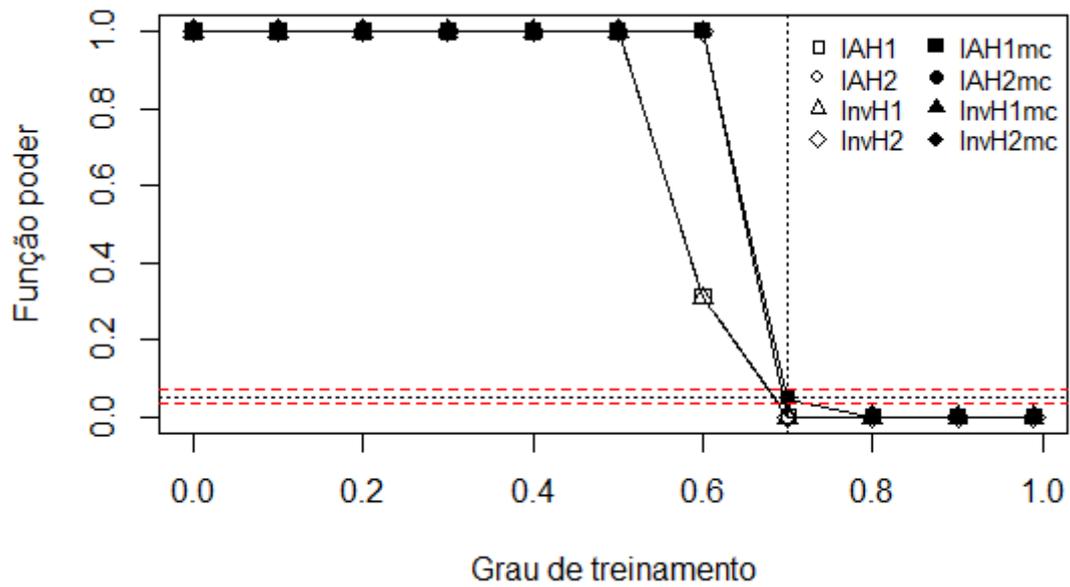
Fonte: da autora.

Figura 62 –  $P = 10$  provedores,  $q = 10$  atributos,  $n = 200$  observações.



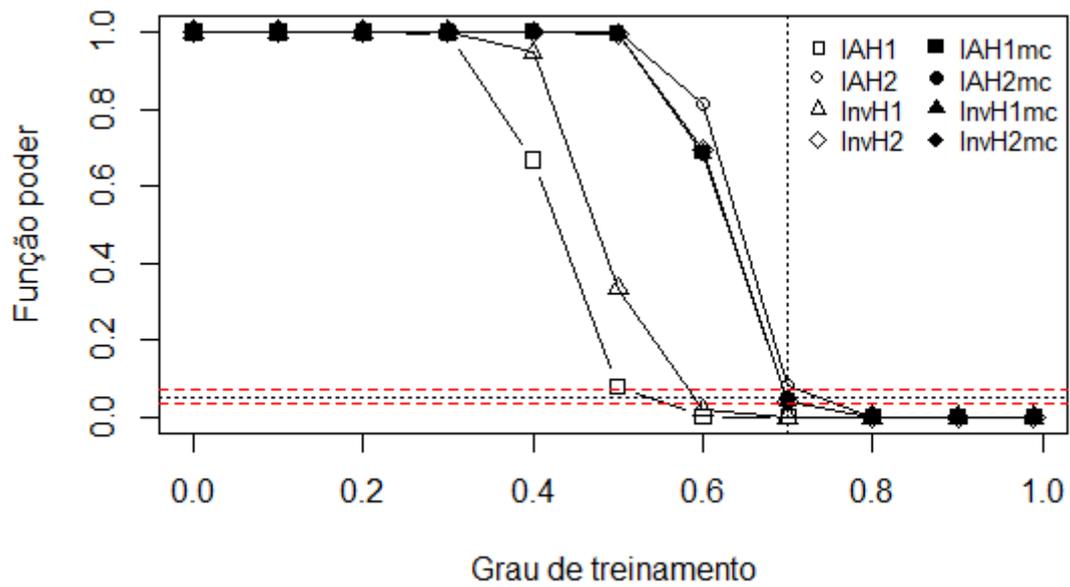
Fonte: da autora.

Figura 63 – P = 10 provadores, q = 20 atributos, n = 200 observações.



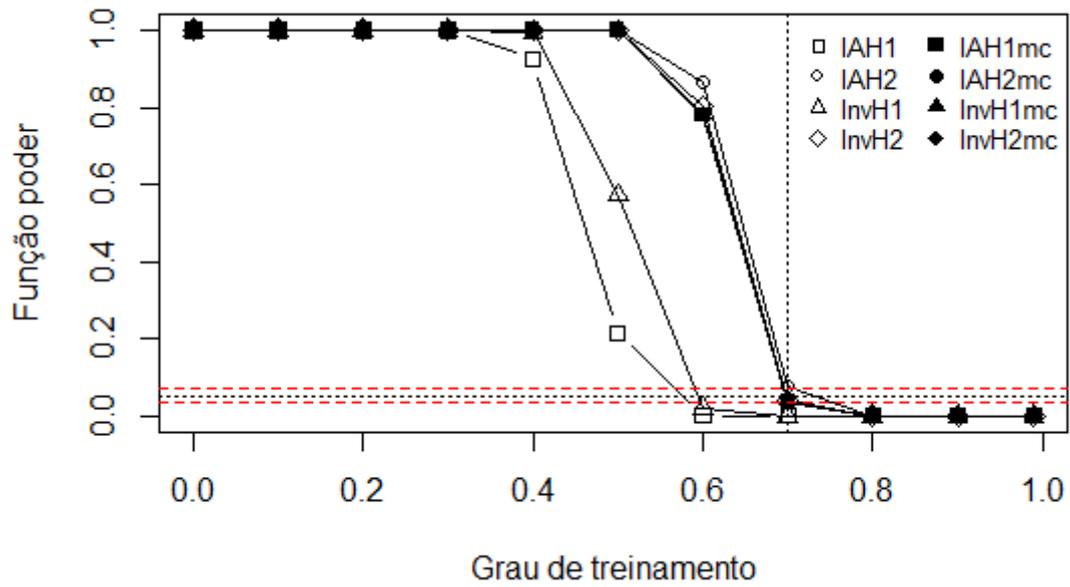
Fonte: da autora.

Figura 64 – P = 15 provadores, q = 2 atributos, n = 30 observações.



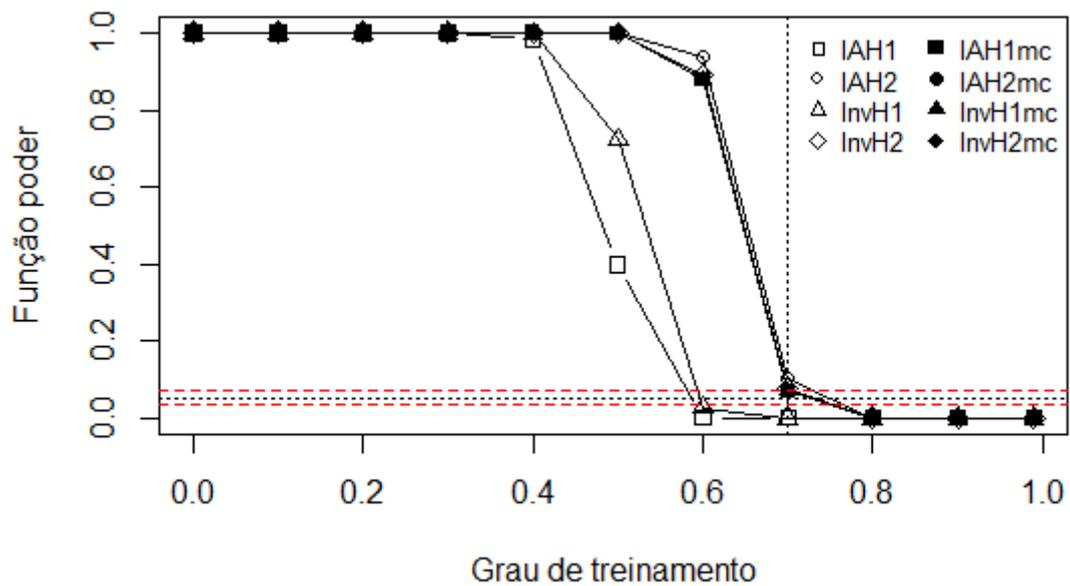
Fonte: da autora.

Figura 65 – P = 15 provadores, q = 2 atributos, n = 40 observações.



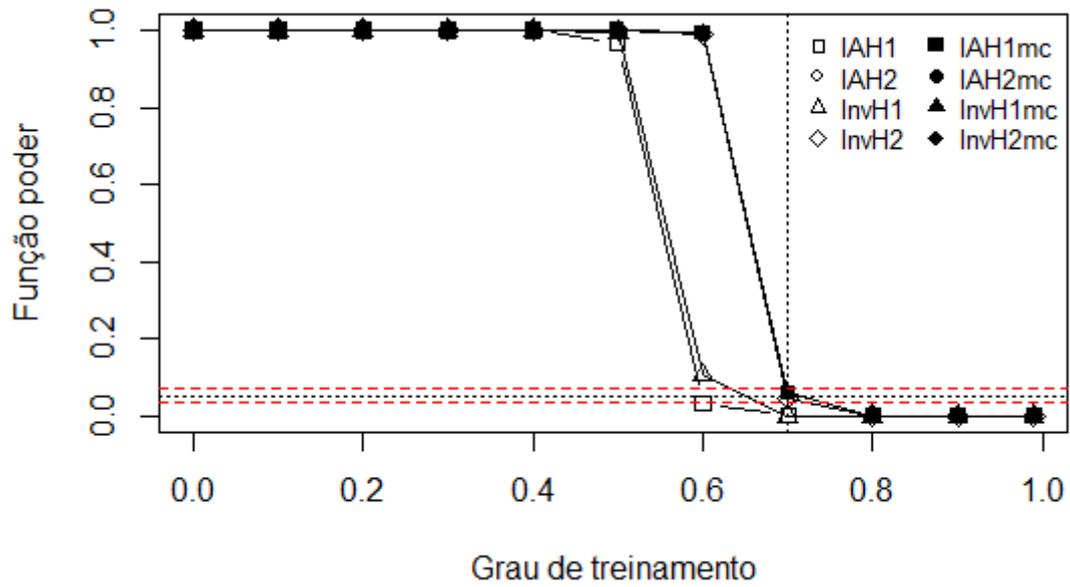
Fonte: da autora.

Figura 66 – P = 15 provadores, q = 2 atributos, n = 50 observações.



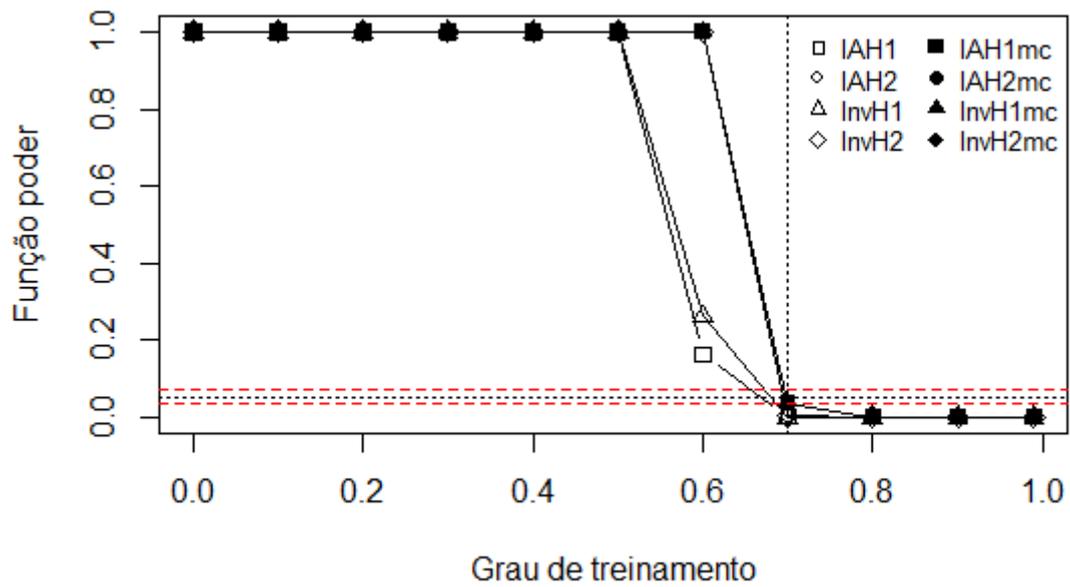
Fonte: da autora.

Figura 67 – P = 15 provedores, q = 2 atributos, n = 100 observações.



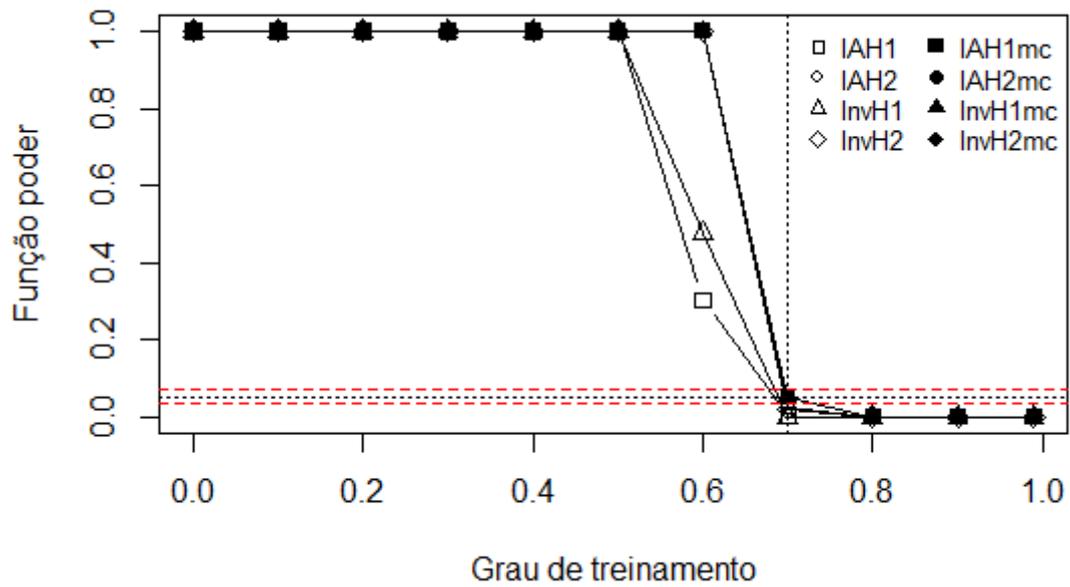
Fonte: da autora.

Figura 68 – P = 15 provedores, q = 5 atributos, n = 100 observações.



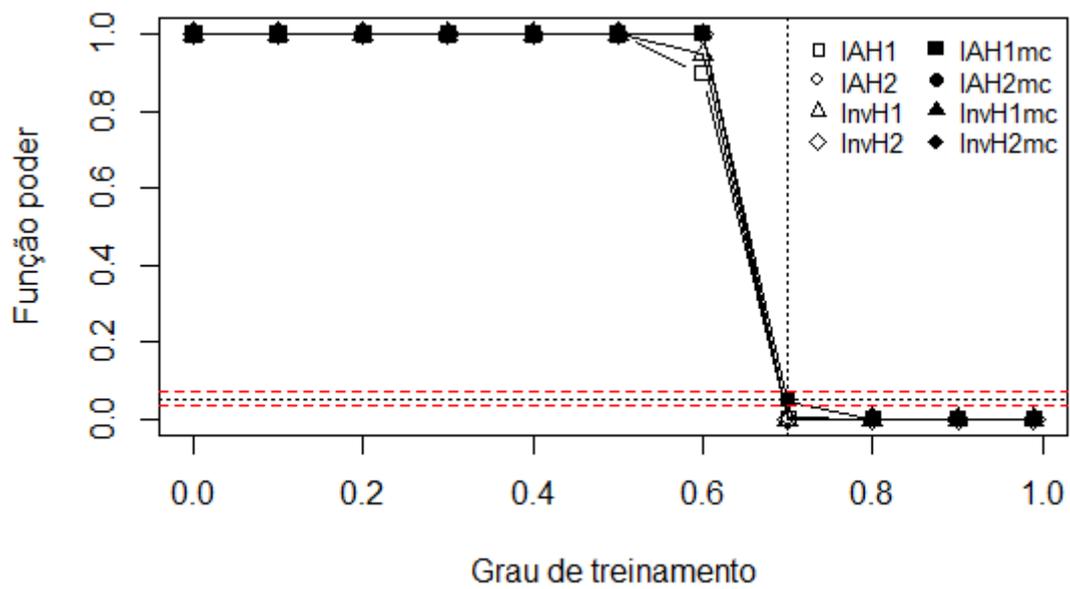
Fonte: da autora.

Figura 69 – P = 15 provedores, q = 2 atributos, n = 200 observações.



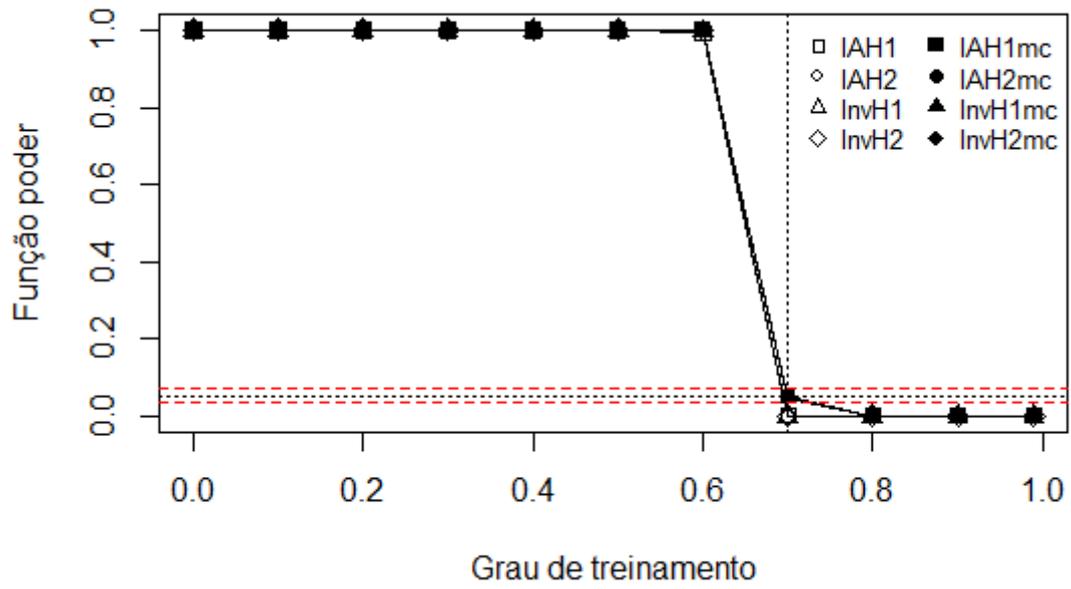
Fonte: da autora.

Figura 70 – P = 15 provedores, q = 5 atributos, n = 200 observações.



Fonte: da autora.

Figura 71 –  $P = 15$  provedores,  $q = 10$  atributos,  $n = 200$  observações.



Fonte: da autora.