

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Faculdade de Filosofia e Ciências Humanas - FAFICH  
Programa de Pós-Graduação em Psicologia: Cognição e  
Comportamento

Heitor Blesa Farias

**UMA METODOLOGIA PRÁTICA PARA TESTAR O CRITÉRIO DO CUSTO DE  
COMPLEXIDADE AO PODAR ÁRVORES DE REGRESSÃO**

Belo Horizonte  
2022

Heitor Blesa Farias

**UMA METODOLOGIA PRÁTICA PARA TESTAR O CRITÉRIO DO CUSTO DE  
COMPLEXIDADE AO PODAR ÁRVORES DE REGRESSÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia: Cognição e Comportamento da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Psicologia.

Área de Concentração: Cognição e Comportamento

Linha de Pesquisa: Mensuração e Intervenção em Psicologia

Orientador: Prof. Dr. Cristiano Mauro Assis Gomes.

Belo Horizonte - MG

2022

153.4	Farias, Heitor Blesa.
F224m	Uma metodologia prática para testar o critério do custo de
2022	complexidade ao podar árvores de regressão [manuscrito] /
	Heitor Blesa Farias. - 2022.
	87 f.
	Orientador: Cristiano Mauro Assis Gomes.
	Dissertação (mestrado) - Universidade Federal de Minas
	Gerais, Faculdade de Filosofia e Ciências Humanas.
	Inclui bibliografia.
	1. Psicologia - Teses. 2. Análise de regressão – Teses.
	I. Gomes, Cristiano Mauro Assis. II. Universidade Federal de
	Minas Gerais. Faculdade de Filosofia e Ciências Humanas. III.
	Título.



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS DA UFMG - FAFICH  
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA: COGNIÇÃO E COMPORTAMENTO

## ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO

### HEITOR BLESIA FARIAS

Realizou-se, no dia 26 de agosto de 2022, às 09:00 horas, videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *UMA METODOLOGIA PRÁTICA PARA TESTAR O CRITÉRIO DO CUSTO DE COMPLEXIDADE AO PODAR ÁRVORES DE REGRESSÃO*, apresentada por HEITOR BLESIA FARIAS, número de registro 2020679331, graduado no curso de PSICOLOGIA, como requisito parcial para a obtenção do grau de Mestre em PSICOLOGIA: COGNIÇÃO E COMPORTAMENTO, à seguinte Comissão Examinadora: Prof(a). Cristiano Mauro Assis Gomes - Orientador (UFMG), Prof(a). Maicon Rodrigues Albuquerque (UFMG), Prof(a). Víthor Rosa Franco (Universidade São Francisco).

A Comissão considerou a dissertação:

( X ) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.



Documento assinado eletronicamente por **Víthor Rosa Franco, Usuário Externo**, em 26/08/2022, às 11:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiano Mauro Assis Gomes, Professor do Magistério Superior**, em 26/08/2022, às 13:49, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maicon Rodrigues Albuquerque, Professor do Magistério Superior**, em 03/09/2022, às 22:32, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1687043** e o código CRC **FEB54EBF**.

## AGRADECIMENTOS

Todo processo educacional e todo período formativo apresentam desafios a serem superados. Esses dois anos e meio de mestrado apresentaram para além dos desafios esperados para essa formação, desafios extras em função da pandemia e do processo de adaptação às novas demandas decorrentes dela. Se não fosse todas as pessoas que me auxiliaram e estiveram comigo durante este período, eu não conseguiria ter chegado até aqui. Quero agradecer a todos que de alguma forma se fizeram presentes em minha trajetória e contribuíram para minha formação enquanto pessoa e pesquisador. Não será possível nomear todos, esse é sempre um problema dos agradecimentos, mas citarei algumas pessoas que quero deixar registrado meus agradecimentos.

Agradeço aos meus pais, Adalberto e Dursilene, por todo amor, carinho, suporte, incentivo, compreensão e ensinamentos. Vocês sempre me incentivam a perseguir os meus sonhos e a ser uma pessoa melhor. Sem vocês eu não estaria aqui. Muito obrigado! Agradeço também a todos os meus familiares, em especial à minha irmã Beatriz, por todo apoio e atenção.

Agradeço aos meus amigos, Abraão, João, Lucas Lacerda, Lucas Matias e Paulo, por todo o companheirismo e por me ouvirem falar o tempo todo sobre medida, psicometria, estatística, R, ciência, psicologia, anime e a vida em geral. Nossos encontros virtuais semanais durante a pandemia foram muito importantes para mim e sou grato a vocês pela nossa amizade. Quero registrar também meu muito obrigado à Elder, por ser meu parceiro de mudança para Belo Horizonte, do xadrez e por me ensinar diversas coisas, até mesmo a cozinhar, isso foi muito importante!

Falando em Belo Horizonte... Quero agradecer à Nazaré por me recepcionar tão gentilmente em sua casa. Todo o seu suporte e acolhimento foram muito importantes para mim. Lhe sou muito grato! Quero agradecer também as minhas amigas Jalile e Fabíola por tudo o que fizeram por mim em Belo Horizonte. A companhia de vocês tornou minha estadia em BH mais fácil e agradável, mesmo durante a pandemia. Muito obrigado!

Quero registrar também o meu agradecimento ao meu orientador, o professor Cristiano Gomes. Com o professor Cristiano pude ter uma formação científica de alta qualidade, a qual me permite refletir sobre a construção de conhecimentos sobre a realidade e o fazer científico. Também pude aprender sobre o que é aprender de forma profunda e a ensinar de uma maneira didática. Para além dessa formação científica, também pude aprender com o professor Cristiano sobre como ser um ser humano melhor. A sua dedicação, perseverança e resiliência serão para

sempre exemplos para mim. Muito obrigado pelo privilégio e pela honra de ter sido orientado pelo senhor.

Quero agradecer também aos meus colegas do Laboratório de Investigação da Arquitetura Cognitiva (LaiCo), Diogo e Jhonys, por toda a parceria e companheirismo, e ao professor Enio Jelihovschi, por todos os ensinamentos e orientações.

Agradeço também aos meus alunos do curso de Psicologia da UFBA, pela confiança no meu trabalho e por todo incentivo. Vocês me inspiram a ser um profissional melhor.

Agradeço aos professores do programa de Pós-graduação em Psicologia: Cognição e Comportamento por toda a dedicação e empenho para realizarem uma formação de qualidade, enfrentando as diversas adversidades para o ensino durante a pandemia.

Agradeço também aos professores que compuseram a banca pelo aceite do convite e pelo olhar criterioso da dissertação.

Agradeço também à FAPEMIG por todo auxílio e financiamento concedido.

A todos, meu muito obrigado!

## Resumo

O método de árvore é uma abordagem bem estabelecida em aprendizado de máquina. A maioria de seus algoritmos constroem uma árvore inicial para a previsão de um determinado resultado e depois "podam" essa árvore como forma de minimizar o *overfitting*. O critério custo de complexidade é provavelmente o mais utilizado para poda, pois é um critério objetivo para definir o ponto na árvore onde a previsão é a melhor possível, levando em consideração a capacidade preditiva do modelo em outras amostras da população. Pesquisadores têm usado esse critério para realizar a poda da árvore de regressão com base nas recomendações da literatura de que esse é um bom critério. No entanto, não existe uma metodologia em que o pesquisador seja capaz de avaliar a eficácia desse critério para gerar uma árvore empírica com poda adequada, ou seja, uma árvore que não tenha *overfitting* e que encontre a melhor previsão possível para outras amostras da população. Considerando a relevância das técnicas de regressão de árvores para predição e a necessidade de poda dessas árvores para lidar com o problema de *overfitting*, é necessário desenvolver uma metodologia que permita ao pesquisador avaliar se o critério custo de complexidade é adequado, tomando como referência a sua própria árvore empírica podada. Esta dissertação tem como objetivo desenvolver uma metodologia prática para avaliar a adequação do critério custo de complexidade para poda de árvores de regressão. A dissertação é composta por dois artigos. O estudo um é uma simulação que apresenta evidências iniciais de que o critério de custo de complexidade é sensível ao tamanho da amostra e gera árvores podadas inadequadamente dependendo do tamanho dessas amostras. Devido a essa inadequação, é necessário testar se a poda pelo critério de custo de complexidade é adequada para um dado empírico. No estudo dois, apresenta-se de forma didática o problema do critério custo da complexidade, bem como a metodologia desenvolvida para verificar a adequação deste critério. Neste estudo, também é apresentado um exemplo de como implementar a metodologia e sua avaliação via simulação.

Palavras-chave: critério de custo de complexidade, método de árvore de regressão, metodologia, simulação, software R, algoritmo CART.

## **Abstract**

The tree method is a well-established approach in machine learning. Most of its algorithms build an initial tree for the prediction of a given outcome and then "prune" this tree as a way to minimize overfitting. The cost of complexity criterion is probably the most used for pruning, because it is an objective criterion to define the point in the tree where the prediction is the best possible, taking into account the predictive ability of the model in other samples of the population. Researchers have used this criterion to perform regression tree pruning based on literature recommendations that this is a good criterion. However, there is no methodology in which the researcher is able to assess the effectiveness of this criterion to generate an empirical tree with adequate pruning, that is, a tree that does not have overfitting and that finds the best possible prediction for other samples of the population. Considering the relevance of tree regression techniques for prediction and the need for pruning these trees to deal with the overfitting problem, it is necessary to develop a methodology that allows the researcher to assess whether the criterion cost of complexity is adequate, taking as reference your own pruned empirical tree. This dissertation aimed to develop a practical methodology to evaluate the adequacy of the cost of complexity criterion for pruning regression trees. The dissertation consists of two articles. Study one is a simulation that presents initial evidence that the complexity cost criterion is sensitive to sample size and generates inadequately pruned trees depending on the size of these samples. Due to this inadequacy, it is necessary to test whether pruning via the complexity cost criterion is adequate for a given empirical data. In study two, the problem of the cost of complexity criterion is presented in a didactic way, as well as the methodology developed to verify the adequacy of this criterion. In this study, an example of how to implement the methodology and its evaluation via simulation is also presented.

**Keywords:** complexity cost criterion, regression tree method, methodology, simulation, R software, CART algorithm.



## SUMÁRIO

<b>1 - Minha inserção no LAICO .....</b>	<b>9</b>
<b>2 - Introdução.....</b>	<b>15</b>
<b>3 - Objetivos .....</b>	<b>19</b>
3.1 - Objetivo Geral.....	19
3.2 - Objetivos específicos .....	19
<b>4 - Delineamento da dissertação .....</b>	<b>19</b>
<b>5 - ESTUDO 1: Pruning trees, complexity cost may not be a good approach: initial results.....</b>	<b>20</b>
5,1 - Background .....	21
5.2 - Question of Research .....	22
5.3 - Scopus of the Paper .....	23
5.4 - Results .....	23
5.5 - Conclusion.....	23
5.6 - References .....	24
<b>6 - ESTUDO 2: A Practical methodology to test the complexity cost criterion when pruning trees .....</b>	<b>37</b>
6.1 - Introduction .....	38
6.2 - The Data of Our Didactic Example.....	45
6.3 - Methodology in action: an example .....	47
6.4 - Testing the method in different true trees .....	56
6.5 - Conclusion.....	57
6.6 - References .....	58
<b>7 - Considerações finais.....</b>	<b>61</b>
<b>8 - Referências .....</b>	<b>62</b>
<b>9 - ANEXOS .....</b>	<b>80</b>
9.1 - ANEXO 1: Código do R para simular as árvores .....	81
9.2 - ANEXO 2: Função complexity_cost.test.....	83

## **1 - Minha inserção no LAICO**

Para iniciar a apresentação da minha dissertação, falarei um pouco sobre minha trajetória acadêmica e como ela me conduziu aos estudos que realizei nesta dissertação e no Laboratório de Investigação da Arquitetura Cognitiva (LAICO), que está vinculado ao Programa de Pós-Graduação em Psicologia: Cognição e Comportamento da Universidade Federal de Minas Gerais (UFMG).

Para falar sobre minha trajetória acadêmica, iniciarei com a minha escolha profissional. Acredito ser importante falar sobre essa escolha, pois ela influenciou muito meu perfil profissional. O meu interesse na área da psicologia surgiu durante o terceiro ano do ensino médio, no ano de 2012. Antes de me interessar pela psicologia, o meu interesse profissional era a ciência da computação. Em função desse meu interesse, realizei quatro cursos técnicos, a saber: técnico em manutenção de computadores, em redes, em programação e em sonoplastia. Foram cursos de curta duração, mas que estavam me preparando para o que eu queria profissionalmente. A mudança da ciência da computação para psicologia ocorreu durante algumas aulas de filosofia no ensino médio. Nestas aulas, a professora explicava as grandes mudanças que ocorreram na forma de pensar no ocidente em função das ideias introduzidas por Copérnico, Darwin e Freud. Busquei mais informações sobre a psicologia e descobri a analogia cérebro-computador e os estudos da ciência cognitiva. Deste ponto em diante, decidi estudar não mais os computadores, mas a cognição humana.

Na busca por uma maior compreensão da cognição humana, entrei no curso de Psicologia da Universidade Federal da Bahia, Instituto Multidisciplinar de Saúde, Campus Anísio Teixeira (UFBA-IMS-CAT), no ano de 2013. Acho importante mencionar o campus, pois ele faz parte do processo de interiorização das universidades federais e nesta época o curso de psicologia existia apenas há três anos. Em função disso, ainda eram poucas as oportunidades de pesquisa e extensão na área da psicologia, mas consegui me inserir no Núcleo de Estudos em Psicologia Humanista (NEPH). Neste grupo, tive a oportunidade de estudar sobre fundamentos históricos e epistemológicos da psicologia e ter meu primeiro contato com a pesquisa. Como resultado da minha participação neste grupo, publiquei meus primeiros artigos (Castelo Branco, Farias, Carpes et al, 2015; Castelo Branco, Farias et al, 2017; Castelo Branco & Farias, 2020) e realizei apresentações em congressos. Desse contato inicial com a pesquisa, comecei a desenvolver o hábito da leitura de artigos. Esses artigos possuíam análises estatísticas e a utilização de instrumentos psicológicos que eu não conseguia compreender. Conversando com colegas dos semestres mais avançados, descobri que tínhamos duas disciplinas que

possivelmente me ajudariam a entender esses artigos, a disciplina de bioestatística e a de psicometria.

Essas duas disciplinas eram ofertadas no quinto semestre e também eram consideradas as disciplinas mais difíceis do curso, sendo a disciplina de bioestatística a disciplina com maior taxa de reprovação do curso. Para mim, essas disciplinas foram reveladoras, pois realmente consegui entender melhor os artigos que lia e pude entender melhor sobre a construção do conhecimento psicológico e da ciência, de uma forma geral. Apesar da relevância dessas disciplinas, muitos assuntos importantes não foram ensinados. Por exemplo, a disciplina de bioestatística não ensinava sobre correlações, análise muito utilizada nos artigos de Psicologia. Para tentar entender sobre esse assunto, li muito sobre o tema e escrevi um programa na linguagem de programação Pascal, que aprendi em meu curso técnico de programação, para realizar análises de correlação de Pearson. Lembro-me de ter ficado muito empolgado com isso na época e mostrei para o monitor da disciplina de psicometria, meu amigo Lucas Matias. Para a minha surpresa, ele me mostrou uma maneira fácil de fazer aquilo no computador. O meu código na linguagem Pascal tinha mais de 200 linhas de programação, o código que ele me mostrou só usava uma linha para fazer a mesma análise. A maneira fácil que o Lucas me mostrou de fazer a análise de correlação foi com o uso da linguagem de programação R.

Com a descoberta dessa poderosa ferramenta, a linguagem R, pude não apenas fazer análises de correlação, mas comecei a fazer as análises que tinha aprendido e ouvido falar nas disciplinas de psicometria e bioestatística, como comparações de médias de grupos, análises fatoriais exploratórias, regressões lineares, etc. Todo esse contato com a bioestatística, psicometria e R ocorreu no ano de 2015. No final deste ano, foi oferecido um curso de extensão do R na UFBA-IMS-CAT e neste curso pude conhecer Fernanda David Vieira. Ela é a psicóloga técnica da UFBA-IMS-CAT e era, na época, uma doutoranda do curso de Psicologia da UFBA de Salvador e estava utilizando o R em suas análises. Ela me convidou para fazer do grupo de pesquisa que ela coordenava e tive a oportunidade de aprender a realizar diversas técnicas estatísticas mais avançadas no R, como regressões múltiplas, análises fatoriais confirmatórias, *path analysis* e equações estruturais. Dessa experiência, participei da publicação de um dos artigos da tese, intitulado de “A Memória de Trabalho como Preditora do Desempenho em Aritmética de Escolares Brasileiros” (Vieira et al., 2021). Este artigo e a pesquisa envolvida nele marcam o início do meu interesse por estudos de predição, desempenho acadêmico e variáveis cognitivas. Foi um período de muito aprendizado, onde pude entender mais dos desafios epistemológicos e metodológicos da ciência psicológica, assim como de diversas análises de dados. Em função do meu interesse por análises de dados, a professora Mariana

Teles, professora da disciplina de psicometria, conseguiu uma vaga para mim em um curso do R, em 2016, ofertado pelo professor Enio Jelihovschi. Foi a partir do contato com esses professores que tive acesso às pesquisas desenvolvidas no LAICO.

O Laboratório de Investigação da Arquitetura Cognitiva (LAICO) foi criado em 2008 pelo professor Cristiano Mauro Assis Gomes. Este laboratório possui como foco o estudo e análise da validade de construtos e teorias da psicologia da educação e das relações entre essas teorias. Para alcançar este fim, são realizadas pesquisas no LAICO que integram os campos psicologia da educação, psicometria e análises estatísticas estado-da-arte. Atualmente, a produção científica realizada no LAICO conta com mais de cem artigos publicados. Além desse aspecto da produção acadêmica, os testes e instrumentos que são construídos no LAICO são, em sua maioria, disponibilizados de forma gratuita para a comunidade acadêmica. Para que o leitor possa visualizar a produção do LAICO de forma mais detalhada, irei apresentar os trabalhos realizados no LAICO em quatro categoriais e, por fim, apresentarei os testes distribuídos de forma gratuita.

A primeira categoria de estudos será aqui denominada de “Construção e/ou validação de testes”. Essa categoria versa sobre a construção e/ou validação de instrumentos em quatro grandes áreas do saber: Psicologia, educação, saúde e musicoterapia.

Na psicologia, os estudos investigaram a construção de instrumentos para avaliar construtos de diferentes áreas da psicologia. Na psicologia cognitiva, foram construídos instrumentos para mensurar a inteligência (Golino & Gomes, 2014b; Gomes, 2010a, 2010b, 2012b, Gomes & Borges, 2007, 2008c, 2009b, 2009c; Muniz et al., 2016), o raciocínio indutivo (Golino & Gomes, 2012, 2019, Golino, Gomes, Commons et al, 2014; Gomes, Araujo, Lima et al., 2021), a memória de curto-prazo (Gomes, 2011b) e a metacognição (Diaz & Gomes, 2021a, 2021b; Golino & Gomes, 2011; Gomes, 2021a; Gomes, Araujo & Castillo-Diaz, 2021). Na psicologia da educação, os instrumentos construídos mensuram as abordagens de aprendizagem (Gomes, 2011a, 2013, 2021b, 2022b, 2022c; Gomes, Araujo et al, 2020; Gomes et al., 2011; Gomes & Linhares, 2018; Gomes, Linhares et al., 2021; Gomes & Nascimento, 2021h, 2021k; Gomes, Quadros et al., 2020; Gomes & Rodrigues, 2021), o interesse em pensar reflexivamente (Gomes, 2022a), as crenças de estudantes sobre ensino-aprendizagem (Gomes & Borges, 2008a), as aspirações escolares (Gomes & Gjikuria, 2018), os estilos de pensamento (Gomes & Marques, 2016), o risco dos estudantes abandonarem o ensino superior (Casanova et al., 2021), as cognições acadêmicas autorreferentes (Costa et al., 2017), as expectativas acadêmicas para o ensino superior (Fleith, Almeida et al., 2020), as práticas docentes para criatividade (Fleith & Gomes, 2019). Além dessas grandes áreas, também foram construídos instrumentos para avaliar

o funcionamento adaptativo (Mecca et al., 2015), a habilidade experiencial (Silveira et al., 2012; Silveira & Gomes, 2014), a procrastinação (Gomes & Rozenberg, 2021) e a personalidade (Gomes, 2012a, Pinheiro et al., 2009).

Na educação, os estudos investigaram a avaliação docente (Gomes & Borges, 2008b), o exame nacional do ensino médio (Gomes, 2005; Gomes & Borges, 2009a; Gomes, Golino & Peres, 2016, 2018, 2020, 2021), as concepções de avaliação de estudantes (Matos et al., 2019) e a criação de exames escolares (Pires & Gomes, 2017, 2018). Na saúde, os estudos investigaram o estresse e a recuperação de treinadores esportivos (Costa et al., 2012), conhecimento dos dentistas sobre HIV/AIDS (Moura et al., 2014), dependência de açúcar (Rosa et al., 2013) e o desenvolvimento motor grosso (Salami et al., 2021). Na musicoterapia, os estudos investigaram a comunicabilidade musical e relação criança-terapeuta na experiência musical (André et al., 2016, 2017, 2018, 2019, 2020a, 2020b, 2020c, 2020d, 2020e, 2021a, 2021b, 2021c), e a relação entre música e atenção (Rosário et al., 2019, 2020).

A segunda categoria de estudos será aqui denominada como “Estudos sobre intervenção cognitiva”. Nestes estudos, são realizadas investigações sobre programas de intervenções cognitivas e educacionais (Cardoso et al., 2019; Gomes, 2002; Gomes, 2007a; 2007b, 2020a, 2020b; Gomes, Carneiro et al., 2008; Gomes, Golino, Santos et al., 2014; Pereira et al., 2019; Ricci et al., 2020). A terceira categoria de estudos, “Preditores de resultados acadêmicos”, busca investigar importantes variáveis para predizer o desempenho acadêmico. Nesse sentido, as variáveis inteligência (Alves et al., 2016, 2018; Gomes & Golino, 2012a; Laros et al., 2014; Martins et al., 2018), metacognição (Castillo-Diaz & Gomes, 2022; Gomes, Golino & Menezes, 2014), motivação, estilos cognitivos, crenças e abordagens de aprendizagem (Alves et al., 2012; Gomes, 2010c, 2011a; Gomes & Golino, 2012c; Gomes, Marques et al., 2014; Rodrigues & Gomes, 2020; Gomes, Farias et al., 2022), nível socioeconômico e família (Alves et al., 2016, 2017, 2018; Martins et al., 2018; Pazeto et al., 2019, 2020) têm sido investigadas como potenciais preditoras do desempenho acadêmico.

A quarta categoria é a “Aplicações estado-da-arte da psicometria e *machine learning*”. Esta categoria versa sobre a aplicação de diversos métodos quantitativos avançados, ou pouco aplicados, para o estudo de construtos psicológicos. Os estudos estado-de-arte da psicometria envolvem análises de classes latente (Araújo et al. 2018; Monteiro et al., 2020), análise de invariância (Fleith et al., 2020; Matos, Brown, Gomes, 2019), modelo Rasch (Golino et al., 2015; Golino & Gomes, 2015a, 2015b, 2015c, 2015d, 2015e, Gomes & Golino, 2015a), medida intraindividual (Ferreira & Gomes, 2017; Golino & Gomes, 2014, Gomes, Araujo & Jelihovschi, 2022; Gomes, Araujo, Nascimento et al., 2018; Gomes, Araújo et al., 2014; Gomes

& Golino, 2015b; Gomes et al., 2013; Jelihovschi & Gomes, 2019), modelagem por equações estruturais confirmatória e exploratória (Gomes & Gjikuria, 2017, 2018; Gomes et al, 2017; Gomes & Golino, 2012a, 2012c, 2014), análise de caminhos (Gomes & Golino, 2012b), *direct acyclic graphics* (Gomes, Nascimento et al., 2019) e análise fatorial confirmatória para testar modelos bifatoriais (Gomes, Golino & Peres, 2016; Muniz, Pasian, & Gomes, 2016; Valentini et al., 2015). Muitas dessas técnicas e diversas outras foram utilizadas nas três categorias de estudos anteriores e tais estudos não foram citados novamente nesta parte. Além desse aspecto técnico da psicometria, o LAICO também produz estudos teóricos práticos da integração entre psicometria e neuropsicologia ((Dias et al., 2015; Gauer et al., 2010; Haase et al, 2010; Reppoldd et al., 2015). Os estudos estado-de-arte da área de machine learning utilizam diversas técnicas, como *learning trees*, *bagging*, *random forest*, *boosting* (Golino & Gomes, 2014a, 2014c, 2016). A área de *learning trees* possui diversos algoritmos. Dentre tais algoritmos, o algoritmo Classification and Regression Tree (CART) tem sido extensivamente utilizado nas pesquisas realizadas no LAICO (Gomes & Almeida, 2017; Gomes, Amantes et al., 2020; Gomes, Farias et al., 2021; Gomes, Farias et al., 2022; Gomes, Fleith et al., 2020; Gomes & Jelihovschi, 2019; Gomes, Lemos et al., 2020, 2021; Mansur-Alves et al., 2021; Teodoro et al., 2021). Além dessas análises, é importante mencionar outros métodos quantitativos também empregados no LAICO, como análise de modelo escondido de Markov (Gomes, Golino, & Costa, 2013), análise de ponto de mudança (Gomes & Jelihovschi, 2016), séries temporais (Gomes & Valentini, 2019) e tamanho de efeito (Gomes, 2020a).

Dos instrumentos produzidos no LAICO, vários instrumentos deles foram disponibilizados para o público de forma aberta e gratuita, a partir de 2021. Esses instrumentos são a Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) (Gomes & Nascimento, 2021a, 2021b, 2021c, 2021d, 2021d, 2021e, 2021f, 2021g, 2021i, 2021j, 2021l, 2021m, 2021n, 2021o, 2021p; Gomes, Nascimento & Araujo, 2021a, 2021b, 2021c, 2021d), a Escala de Interesse em Pensar Reflexivamente (Gomes, 2022a), o Teste Abordagem-em-Processo Versão 2 (Gomes, 2022b) e a Escala de Abordagens de Aprendizagem (EABAP) (Gomes, 2022c).

Das quatro categorias de produção acadêmica do LAICO, os meus interesses como pesquisador, atualmente, se concentram nas categoriais 1) Construção e/ou validação de testes; 3) Preditores de resultados acadêmicos; e 4) Aplicações estado-da-arte da psicometria e *machine learning*. Como resultados desses meus interesses e da minha inserção no LAICO, em 2020, tive a oportunidade de participar de pesquisas que envolvem a aplicação de técnicas estado-da-arte da psicometria para avaliar a invariância da Escala de Abordagens de Aprendizagem nas variáveis sexo, nível educacional e tipo de escola. Também tive a

oportunidade de pesquisar sobre a validade incremental das abordagens de aprendizagem, quando controlado o efeito da inteligência, para prever o desempenho acadêmico, utilizando técnicas estado-da-arte da psicometria, como a análise Rasch de modelos bifatoriais e multidimensionais, e *machine learning*, como o algoritmo CART (Gomes, Farias et al., 2022).

A presente dissertação faz parte dos estudos desenvolvidos por mim no LAICO. Os estudos que a compõem e serão apresentados a seguir, podem ser incluídos na categoria “Aplicações estado-da-arte da psicometria e *machine learning*”.

## 2 - Introdução

Na experiência de quase todo estudante, a preparação para a semana de avaliação é um fenômeno bastante comum. Nesta preparação, os estudantes realizam diversas atividades estratégicas, como ler o material passado pelo professor, responder listas de exercício, buscar informações complementares em outras fontes e analisar logicamente as informações adquiridas, pois acreditam que ao fazerem isso conseguirão obter notas satisfatórias e serem aprovados nas disciplinas. Essa prática recorrente de quase todo aluno, pode ser entendida como um modelo intuitivo de predição do desempenho escolar. Nesse modelo, os alunos acreditam que quanto mais atividades realizarem, maior será sua futura nota, ou seja, os alunos constroem uma maneira de prever/inferir qual será seu desempenho acadêmico. Esse modelo intuitivo dos estudantes é interessante e exemplifica a importância prática da predição. No entanto, a aparente relevância desse modelo intuitivo de previsão das notas precisa ser testada e melhor investigada. A utilização do método científico é uma forma de fazer essa investigação. O método científico busca criar novos modelos e aprimorar modelos como esses dos alunos, para responder, com a utilização de critérios, questões importantes sobre as variáveis no modelo. Por exemplo, quanto do desempenho acadêmico é possível prever utilizando essas atividades? Quais são as atividades mais importantes na predição do desempenho acadêmico? Quanto cada atividade contribui para prever o desempenho acadêmico? Como as atividades preditoras se relacionam entre si?

Essas perguntas não se aplicam apenas ao contexto da predição do desempenho acadêmico, mas podem ser extrapoladas para diversos fenômenos psicológicos e para a investigação científica de um modo geral. Investigar como prever um fenômeno é uma atividade tão importante do fazer científico e da psicologia, que Shaughnessy, Zechmeister e Zechmeister (2012) dizem que a predição é um dos quatro objetivos da ciência, junto com a descrição, a explicação e a aplicação. Segundo esses autores, a descrição diz respeito a forma como os pesquisadores definem, classificam e catalogam os fenômenos que estudam e as relações entre eles. No caso do modelo dos estudantes, precisaria ser feita uma descrição do que são cada uma das atividades estratégicas e do desempenho acadêmico. A predição, como dito anteriormente, seria a capacidade de prever um determinado resultado/desfecho (desempenho acadêmico), com base nas relações da variável desfecho com as variáveis preditoras (atividades estratégicas). Para Shaughnessy, Zechmeister e Zechmeister (2012), a explicação de um fenômeno é a identificação do que causa esse fenômeno. Por exemplo, suponha que analisar logicamente as informações adquiridas durante o estudo seja a principal



atividade estratégica para a predição do desempenho acadêmico. A explicação da relação entre esses fenômenos deve identificar o porquê isso acontece. Por fim, temos a aplicação. A aplicação e a pesquisa aplicada são formas de utilizar os conhecimentos adquiridos nos outros objetivos para propor maneiras de mudar a vida das pessoas para melhor (Shaughnessy, Zechmeister & Zechmeister, 2012). Por exemplo, conhecendo a relação entre analisar logicamente as informações adquiridas durante o estudo e o desempenho acadêmico, uma aplicação poderia ser a criação de um programa interventivo que ensina os estudantes a analisarem de forma lógica o material de estudo e assim melhorarem seu desempenho acadêmico.

Considerando a relevância da predição para a atividade científica e a construção de conhecimento, é muito importante saber como diferentes abordagens metodológicas operacionalizam este conceito. A literatura científica da psicologia e da educação utiliza, predominantemente, a abordagem estatística do modelo linear geral (ou seja, Modelagem de Equações Estruturais, Regressão Múltipla, Regressão Multinível, ANOVA, correlações) para gerar as evidências de predição (Gomes, Lemos, & Jelihovschi, 2021). Menos usual, mas que tem sido cada vez mais utilizada, é abordagem do método de árvore, que é uma abordagem bem estabelecida na área de *machine learning* (Dwyer et al, 2018; Franco, 2021; Gomes & Almeida, 2017; Gomes, Amantes et al., 2020; Gomes, Farias et al., 2021; Gomes, Farias et al., 2022; Gomes, Fleith et al., 2020; Gomes & Jelihovschi, 2019; Gomes, Lemos et al., 2020, 2021; Korkmaz, & Correia, 2019; Mansur-Alves et al., 2021; Shatte et al, 2019; Teodoro et al., 2021; Yarkoni, & Westfall, 2017). A abordagem do método de árvore apresenta vantagens em comparação às técnicas do modelo linear geral para a predição de desfechos, pois as técnicas do método de árvore não possuem pressupostos restritivos sobre a estrutura dos dados como o modelo linear geral (Gomes, Lemos, & Jelihovschi, 2021; Gomes & Jelihovschi, 2019). Por exemplo, para a adequada análise de dados e construção de evidências sobre a predição de um desfecho, o modelo linear geral requer que a relação entre as variáveis seja linear, que a distribuição dos erros da predição seja normal, que haja homogeneidade de variância (homocedasticidade) e que haja independência das observações (Gomes, Lemos, & Jelihovschi, 2021; Gomes & Jelihovschi, 2019). Caso alguma dessas condições não sejam satisfeitas ou especificamente modeladas, como o caso de uma relação não linear, a predição se torna inadequada. Outra vantagem das técnicas do método de árvore em relação ao modelo linear geral é o tratamento de dados que possuem variáveis nominais com muitas categorias e variáveis ordinais, como é o caso de muitas variáveis da psicologia e da educação (Gomes & Almeida, 2017; Gomes & Jelihovschi, 2019; Gomes, Lemos, & Jelihovschi, 2021). Em função

dessas vantagens, as predições que utilizam as técnicas de método de árvore conseguem prever uma variância muito maior da variável desfecho do que as técnicas do modelo linear geral. Como uma abordagem da área de *machine learning*, o método de árvore também apresenta vantagens em comparação com outras técnicas dessa área. Em muitas técnicas de *machine learning* não é fácil, ou mesmo possível, inspecionar como a predição foi feita. Essa falta de transparência e impossibilidade de compreensão humana das predições de muitos algoritmos de *machine learning* é denominada de problema da caixa preta (Rudin, 2019). Para evitar esse problema, a literatura tem apontando a importância da utilização de algoritmos de *machine learning* que são interpretáveis e transparentes para os seres humanos (Rudin, 2019). Dentre tais algoritmos interpretáveis e transparentes, os métodos de regressão em árvore se destacam dos demais e por isso são muito utilizados em diversas áreas (Loh, 2014).

Os métodos de regressão em árvore são muito utilizados, pois são flexíveis, simples computacionalmente e seus resultados são de fácil interpretação. A flexibilidade diz respeito à aplicação em problemas das mais diversas áreas do conhecimento científico. A simplicidade computacional, se refere a facilidade com que os computadores executam o método, sendo assim uma ferramenta acessível. Já a facilidade da interpretação, se refere a forma como o resultado do método é interpretado. O resultado da regressão em árvore é um conjunto de regras se-então. Por exemplo, suponhamos que queremos prever, via o método de regressão em árvore, o desempenho acadêmico utilizando as mesmas variáveis apresentadas no início desta introdução. Poderíamos encontrar o seguinte resultado: “Se um grupo de alunos lê o material passado pelo professor (variável preditora), e se esse grupo responde a lista de exercício (variável preditora), e se esse grupo analisa logicamente o material estudado (variável preditora), então esse grupo de alunos deve atingir um desempenho acadêmico de 95 pontos (variável desfecho)” (Golino & Gomes, 2014a; Gomes & Jelihovschi, 2019). Por produzir semelhante tipo de resultado, o método de regressão em árvore cria um mapa da interação das variáveis preditoras no processo de predição. Este tipo de resultado é muito importante para a expansão da compreensão teórica da variável desfecho, pois fornece além de uma predição, um modelo explicativo. Este tipo de resultado não seria encontrado com a utilização dos métodos do modelo linear geral que são de mais difícil interpretação (Gomes & Almeida, 2017; Gomes & Jelihovschi, 2019; Gomes, Lemos, & Jelihovschi, 2021).

Como o método de regressão em árvore mapeia as relações entre as variáveis preditoras e a variável desfecho para produzir a maior predição possível, este método pode gerar modelos preditivos dependentes da amostra. Isso significa que caso o modelo de predição encontrado seja utilizado para prever o desfecho em dados de uma outra amostra, o poder preditivo do

modelo será menor do que o modelo executado no primeiro dado. A literatura científica denomina esse problema como *overfitting*. Para evitar este problema, diversas alternativas tem sido propostas, como as técnicas de conjunto (bagging, random forest e boosting, Golino & Gomes, 2014a) e a utilização da poda da árvore via vários métodos, como a poda de erro reduzido, a remoção de erro mínimo, a poda pessimista, a remoção baseada em erros, a poda de valor crítico, a interpretabilidade e o critério do custo de complexidade (Rokach & Maimon, 2015, Gomes & Jelihovschi, 2019). Dentre tais métodos de poda das árvores de regressão, o critério do custo de complexidade é provavelmente o mais utilizado, pois fornece um indicador objetivo sobre o tamanho da árvore que apresenta a previsão que é a melhor possível para outras amostras da população. Por exemplo, suponhamos que as mesmas variáveis foram mensuradas em duas amostras e que aplicamos a regressão em árvore na primeira amostra, gerando uma árvore de regressão não podada com 50 folhas. Suponhamos também que essa árvore não podada com 50 folhas explique 90% da variância da variável desfecho da primeira amostra, mas apenas 20% da variância dessa variável desfecho na segunda amostra. Esse tipo de resultado indica que a árvore com 50 folhas não apresenta a previsão que é a melhor possível em outras amostras. Vamos supor agora que essa árvore de regressão com 50 folhas foi podada para 20 folhas. Suponhamos também que essa árvore podada com 20 folhas explique a variância da variável desfecho em 70% na primeira amostra e em 60% na segunda amostra. Com este tipo de resultado, podemos dizer que 20 folhas é o tamanho da árvore que apresenta a previsão que é a melhor possível para outras amostras da população. É justamente esse tamanho da árvore, esse ponto de corte da árvore não podada, que o critério do custo de complexidade se propõe a encontrar. O leitor interessado nos aspectos matemáticos de como esse ponto de corte é encontrado pelo critério do custo de complexidade pode consultar Gomes e Jelihovschi (2019).

Os pesquisadores têm utilizado o critério do custo de complexidade para realizar a poda de árvores de regressão com base em recomendações da literatura de que este é um bom critério. No entanto, uma simulação realizada no LAICO, a qual foi disponibilizada como pre-print na plataforma OSF (Gomes, Farias et al., 2021), indica que esse critério pode ser substancialmente inadequado e, por isso, necessita ser testado. Os detalhes dessa simulação e seus resultados são apresentados no estudo 1 desta dissertação.

Após uma busca na literatura da área, não foram encontradas maneiras e metodologias que permitam aos pesquisadores avaliarem a eficácia desse critério para gerar uma árvore empírica com a poda adequada, ou seja, uma árvore que não possui *overfitting* e que encontra a melhor previsão possível para outras amostras da população.

Considerando a relevância das técnicas de regressão em árvore para a predição e a necessidade da poda dessas árvores para tratar o problema do *overfitting*, se torna necessário o desenvolvimento de uma metodologia que permita ao pesquisador avaliar se o critério do custo de complexidade é adequado, tomando como referência sua própria árvore empírica podada. É objetivo desta dissertação o desenvolvimento desta metodologia.

### **3 - Objetivos**

#### **3.1 - Objetivo Geral**

Desenvolver uma metodologia prática para avaliar a adequação do critério custo de complexidade para podar árvores de regressão.

#### **3.2 - Objetivos específicos**

- 1) Avaliar a adequação do critério do custo de complexidade para podar árvores.
- 2) Descrever a racionalidade e desenvolvimento da metodologia.
- 3) Operacionalizar a metodologia via uma função na linguagem R.
- 4) Analisar a metodologia e a função via simulação computacional.

### **4 - Delineamento da dissertação**

A dissertação é composta por dois artigos. O estudo 1, conforme dito, é uma simulação que apresenta evidências iniciais de que o critério do custo de complexidade gera árvores podadas inadequadas a depender do tamanho da amostra e, por isso, necessita ser testado. Devido a sua relevância para a composição do problema da dissertação, ele é apresentado na íntegra. Por sua vez, o estudo 2 trata o problema da dissertação, apresentando uma metodologia para testar o critério do custo de complexidade.

## 5 - ESTUDO 1: Pruning trees, complexity cost may not be a good approach: initial results

Cristiano Mauro Assis Gomes (LAICO, Universidade Federal de Minas Gerais)

Heitor Blesa Farias (LAICO, Universidade Federal de Minas Gerais)

Jhonys de Araujo (LAICO, Universidade Federal de Minas Gerais)

Enio Galinkin Jelihovschi (LAICO, Universidade Estadual de Santa Cruz)

## 5,1 - Background

The Laboratory for Cognitive Architecture Mapping [Laboratório de Investigação da Arquitetura Cognitiva (LAICO)] was founded in 2009 in the department of psychology of the Federal University of Minas Gerais, Brazil. Its mission consists in articulating the psychometrics fields in educational issues and investigating the validity of psychological constructs, particularly related to the educational psychology field. In other words, LAICO aims to contribute to three research lines. The first involves the construction and validation of psychological tests, particularly related to educational outcomes (Costa, Gomes & Fleith, 2017; Gomes, & Golino, 2012; Golino & Gomes, 2012, 2014b; Golino, Gomes, & Andrade, 2014; Gomes, 2010b, 2011b, 2021a, 2012b; Gomes & Borges, 2007, 2008a, 2008c, 2009b, 2009c; Gomes & Gjikuria, 2017, 2018; Gomes & Golino, 2012a, 2012b; Golino & Gomes, 2011; Gomes & Golino, 2014; Gomes, Golino & Menezes, 2014; Gomes, 2010c, 2011a, 2013; Gomes, Araujo, & Jelihovschi, 2020; Gomes & Golino, 2012c; Gomes et al., 2019; Gomes, Golino, Pinheiro, Miranda, & Soares, 2011; Gomes & Linhares, 2018; Gomes et al., 2021; Gomes & Marques, 2016; Gomes, Marques, & Golino, 2014; Pinheiro, Gomes & Braga, 2009; Rodrigues & Gomes, 2020; Silveira & Gomes, 2014; Silveira, Gomes & Golino, 2012). Many of these tests are available for research and teaching in researchgate (Gomes & Nascimento, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2021g, 2021h, 2021i, 2021j, 2021k, 2021l, 2021m, 2021n, 2021o, Gomes, Nascimento & Araujo, 2021a, 2021b, 2021c, 2021d). The second contribution of LAICO involves cooperation with many research centers, concerning the validation of theoretical models or instruments of measurement (Alves, Gomes, Martins, & Almeida, 2016, 2017, 2018; Araujo, Gomes, Almeida, & Núñez, 2018; Casanova, Gomes, Bernardo, Núñez, & Almeida, 2021; Fleith et. al 2020; Fleith & Gomes, 2019; Fleith, Gomes, Marinho-Araujo, & Almeida, 2020; Matos, Brown, Gomes, 2019; Monteiro, Almeida, Gomes, & Sinval, 2020; Golino, Gomes, Commons, Miller, 2014; Golino & Gomes, 2015, 2019; Martins, Gomes, Alves, & Almeida, 2018; Andre, Gomes, & Loureiro, 2016, 2017, 2018, 2020a, 2020b, 2020c; Rosário, Gomes, & Loureiro, 2019; Sampaio, Loureiro & Gomes, 2015; Costa, Gomes, Andrade, & Samulski, 2012; Moura, Gomes, Blanc, Mesquita, & Ferreira, 2014; Rosa et al., 2013; Cardoso, Seabra, Gomes, & Fonseca, 2019; Dias et al., 2015; Mecca et al., 2015; Pazeto, Dias, Gomes, & Seabra, 2019, 2020; Reppold et al., 2015; Gomes & Borges, 2007, 2008c; Laros, Valentini, Gomes, & Andrade, 2014; Muniz, Pasian, & Gomes, 2016; Valentini et al., 2015). The third contribution of LAICO involves both the applications of innovative statistics and algorithms of machine learning, as well as, the elaboration of original

methodologies (Cardoso, Seabra, Gomes, & Fonseca, 2019; Ferreira & Gomes, 2017; Fleith & Gomes, 2019; Gauer, Gomes, & Haase, 2010; Golino, Andrade, & Gomes, 2014; Golino & Gomes, 2012, 2014a, 2014c, 2016, 2019; Golino, Gomes, Amantes, & Coelho, 2015; Gomes, 2007a, 2007b, 2010a, 2020a, 2020b; Gomes & Almeida, 2017; Gomes, Almeida, & Núñez, 2017; Gomes, Amantes, & Jelihovschi, 2020; Gomes, Araujo, Ferreira, & Golino, 2014; Gomes, Araujo, Nascimento, & Jelihovschi, 2018; Gomes & Borges, 2008b, 2009a; Gomes, Fleith, Marinho-Araujo, & Rabelo, 2020; Gomes & Gjikuria, 2017; Gomes & Golino, 2015; Gomes, Golino, & Costa, 2013; Gomes, Golino & Peres, 2016, 2018, 2020; Gomes, Golino, Santos, & Ferreira, 2014; Gomes & Jelihovschi, 2016, 2019; Gomes, Lemos, & Jelihovschi, 2020; Gomes, Linhares, Jelihovschi, & Rodrigues, 2021; Gomes, Nascimento, & Peres, 2019; Gomes & Valentini, 2019; Jelihovschi, & Gomes, 2019; Matos, Brown, Gomes, 2019; Pereira, Golino & Gomes, 2019; Pires & Gomes, 2017, 2018; Ricci, Gomes, Nico, & Seabra, 2020; Rodrigues & Gomes, 2020).

## **5.2 - Question of Research**

The researchers of LAICO have been using the CART algorithm for tree analysis in many educational and psychological data and many papers resulted from this work, as Gomes and Almeida (2017) and Gomes and Jelihovschi (2019), which discussed the use of decision and regression tree methods in educational datasets. The CART algorithm along with other algorithms for tree analysis require pruning the tree, in order to eliminate the leaves resulting from noise and do not contribute to the prediction of the outcome. The mainstream approach consists in using the cross-validation K-folder and the identification of the cutoff point value which minimizes the error of prediction to do the pruning. This approach is called complexity cost criterion. Another criterion also largely used is the interpretability, which involves pruning the tree by selecting only a number of leaves which are small enough to allow the researcher to interpret them and get a clear view of their meaning. The complexity cost criterion is considered an objective criterion, because it uses the cut-point where the tree is pruned to the smallest error of prediction. Compared to it, the interpretability criterion seems to be an arbitrary cutoff point chosen by convenience. However, this apparent adequacy of the complexity cost criterion is a mistake. Two simulation analyses conducted by the researchers of LAICO show that this criterion is not suitable and the interpretability criterion is much more appropriate to analyse a tree.

### 5.3 - Scopus of the Paper

We created a true tree with 20 leaves and ran two simulation analyses (the code is in appendix), the first generating 100 samples of  $N = 239$  and the second generating 100 samples of  $N = 10000$ . We used those simulations to test the adequacy of the complexity cost criterion in pruning the trees, that is, whether by using this criterion the trees were pruned to the correct number of leaves.

### 5.4 - Results

The simulation of  $N = 239$  produced pruned trees with an average of 12.21 (SD = 3.86) leaves, and a range of 2 to 20 leaves. The simulation of  $N = 10000$  produced an average of 53.62 (SD = 12.11) leaves for the pruned trees, and a range of 36 to 85 leaves. The complexity cost criterion generated pruned trees which clearly underestimated the true number of leaves in the small samples ( $N = 239$ ) and largely overestimated the correct number of leaves in the large samples ( $N = 10000$ ). In short, these results show that the complexity cost criterion does not lead to the correct number of leaves.

### 5.5 - Conclusion

Our study brings initial evidence that the complexity cost criterion may be ineffective. In small samples, there is a tendency to underestimate the correct number of leaves, while in large samples there is a tendency to overestimate the correct number of leaves. Although the interpretability criterion appears to be guided by convenience, it nonetheless may be better than the complexity cost criterion to estimate the correct number of leaves. The lack of objectivity of the pruning guided by this criterion, can be evaluated through an old approach used in the scientific field, that is, the replicability in different samples. The number of leaves and the adequacy of the tree produced by the interpretability criterion can be evaluated by observing the number of times that these leaves occur in different samples in a similar context of study. If the same variables are used in a similar predictive model in one hundred samples, for example, then the leaves that are continually present and show consistent and similar values may be seen as valid. Instead of searching a true tree through a criterion to cut the tree, our study suggests that it may be more effective to evaluate the validity of the tree through this old and powerful strategy of the scientific field.



## 5.6 - References

- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2018). The structure of intelligence in childhood: age and socio-familiar impact on cognitive differentiation. *Psychological Reports, 121*(1), 79-92. doi: 10.1177/0033294117723019
- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2016). Social and cultural contexts change but intelligence persists as incisive to explain children's academic achievement. *PONTE: International Scientific Researches Journal, 72*(9), 70-89. doi: 10.21506/j.ponte.2016.9.6
- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2017). Cognitive performance and academic achievement: How do family and school converge? *European Journal of Education and Psychology, 10*(2), 49-56. doi: 10.1016/j.ejeps.2017.07.001
- Alves, F. A., Flores, R. P., Gomes, C. M. A., Golino, H. F. (2012). Preditores do rendimento escolar: inteligência geral e crenças sobre ensino-aprendizagem. *Revista E-PSI, 1*, 97-117. <https://revistaepsi.com/artigo/2012-ano2-volume1-artigo5/>
- André, A. M., Gomes, C. M. A., Loureiro, C. M. V. (2016). Escalas Nordoff Robbins: uma revisão bibliográfica. *Percepta, 3*(2), 117-131, 2016. doi: 10.34018/2318-891X.3(2)117-131
- André, A. M., Gomes, C. M. A., Loureiro, C. M. V. (2017). Equivalência de itens, semântica e operacional da versão brasileira da Escala Nordoff Robbins de Comunicabilidade Musical. *OPUS, 23*(2), 153. doi:10.20504/opus2017b2309.
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2018). Reliability Inter-examiners of The Nordoff Robbins Musical Communicativeness Scale Brazilian Version. *11th International Conference of Students of Systematic Musicology, 101–105*. <http://musica.ufmg.br/sysmus2018/wp-content/uploads/2018/07/Reliability-Interexaminers-of-the-Nordoff-Robbins-Musical-Communicativeness-Scale-BrazilianVersion.pdf>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020a). Confiabilidade Interexaminadores da Escala de Relação Criança-Terapeuta na Experiência Musical Coativa para validação no contexto brasileiro. *Hodie, 20*, e64243, 1–18. doi:10.5216/mh.v20.64243
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020b). Confiabilidade Interexaminadores da versão brasileira da Escala Nordoff Robbins de Comunicabilidade

- Musical. In *Estudos Latino-americanos em Música*. vol.2 (pp. 152–163). Artemis. doi:10.37572/EdArt\_13210092015
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020c). Equivalência de itens, semântica e operacional da “Escala de Musicabilidade: Formas de Atividade, Estágios e Qualidades de Engajamento.” *Orfeu*, 5(2), 1–22. doi:10.5965/2525530405022020e0010
- Araújo, A. M., Gomes, C. M. A., Almeida, L. S., & Núñez, J. C. (2018). A latent profile analysis of first-year university students’ academic expectations. *Anales De Psicología / Annals of Psychology*, 35(1), 58-67. doi:10.6018/analesps.35.1.299351
- Cardoso, C. O., Seabra, A. G., Gomes, C. M. A., & Fonseca, R. P. (2019). Program for the neuropsychological stimulation of cognition in students: impact, effectiveness, and transfer effect on student cognitive performance. *Frontiers in Psychology*, 10, 1-16. doi: 10.3389/fpsyg.2019.01784
- Casanova, J. R., Gomes, C. M. A., Bernardo, A. B., Núñez, J. C., & Almeida, L. S. (2021). Dimensionality and reliability of a screening instrument for students at-risk of dropping out from higher education. *Studies in Educational Evaluation*, 68, 100957. doi: 10.1016/j.stueduc.2020.100957
- Costa, V. T., Gomes, C. M. A., Andrade, A. G. P., & Samulski, Di. M. (2012). Validação das propriedades psicométricas do RESTQ-Coach na versão brasileira. *Motriz: Revista de Educação Física*, 18(2), 218-232. doi:10.1590/S1980-65742012000200002
- Costa, B. C. G., Gomes, C. M. A., & Fleith, D. S. (2017). Validade da Escala de Cognitiones Acadêmicas Autorreferentes: autoconceito, autoeficácia, autoestima e valor. *Avaliação Psicológica*, 16(1), 87-97. doi: 10.15689/ap.2017.1601.10
- Dias, N. M., Gomes, C. M. A., Reppold, C. T., Fioravanti-Bastos, A., C., M., Pires, E. U., Carreiro, L. R. R., & Seabra, A. G. (2015). Investigação da estrutura e composição das funções executivas: análise de modelos teóricos. *Psicologia: teoria e prática*, 17(2), 140-152. doi: 10.15348/1980-6906/psicologia.v17n2p140-152
- Ferreira, M. G., & Gomes, C. M. A. (2017). Intraindividual analysis of the Zarit Burden Interview: a Brazilian case study. *Alzheimers & Dementia*, 13, P1163-P1164. doi: 0.1016/j.jalz.2017.06.1710
- Fleith, D. S., Almeida, L. S., Marinho-Araujo, C. M., Gomes, C. M. A., Bisinoto, C., & Rabelo, M. L. (2020). Validity evidence of a scale on academic expectations for higher education. *Paidéia*, 30, e3010. doi: 10.1590/1982-4327e3010

- Fleith, D. S., & Gomes, C. M. A. (2019). Students' assessment of teaching practices for creativity in graduate programs. *Avaliação Psicológica*, 18(3), 306-315. doi: 10.15689/ap.2019.1803.15579.10
- Fleith, D. S., Gomes, C. M. A., Marinho-Araujo, C. M., & Almeida, L. S. (2020). Expectativas de sucesso profissional de ingressantes na educação superior: estudo comparativo. *Avaliação Psicológica*, 19(3), 223-231. doi: 10.15689/ap.2020.1903.17412.01
- Gauer, G., Gomes, C. M. A., & Haase V. G. (2010). Neuropsicometria: Modelo clássico e análise de Rasch. In *Avaliação Neuropsicológica*, (pp. 22-30). Porto Alegre: Artmed, 2010.
- Golino, H. F., & Gomes, C. M. A. (2011). Preliminary internal validity evidences of two Brazilian Metacognitive Tests. *International Journal of Testing*, 26, 11-12. <https://www.intestcom.org/files/ti26.pdf>
- Golino, H. F., & Gomes, C. M. A. (2012). The Structural validity of the Inductive Reasoning Developmental Test for the measurement of developmental stages. *International Journal of Testing*, 27, 10-11. [https://www.researchgate.net/publication/269985854\\_The\\_Structural\\_VValidity\\_of\\_the\\_Inductive\\_Reasoning\\_Developmental\\_Test\\_for\\_the\\_Measurement\\_of\\_Developmental\\_Stages](https://www.researchgate.net/publication/269985854_The_Structural_VValidity_of_the_Inductive_Reasoning_Developmental_Test_for_the_Measurement_of_Developmental_Stages)
- Golino, H. F., & Gomes, C. M. A. (2014a). Four Machine Learning methods to predict academic achievement of college students: a comparison study. *Revista E-Psi*, 1, 68- 101. Retirado de <https://revistaepsi.com/artigo/2014-ano4-volume1-artigo4/>
- Golino, H.F., & Gomes, C.M.A. (2014b). Psychology data from the “BAFACALO project: The Brazilian Intelligence Battery based on two state-of-the-art models – Carroll’s Model and the CHC model”. *Journal of Open Psychology Data*, 2(1), p.e6. doi: 10.5334/jopd.af
- Golino, H. F., & Gomes, C. M. A. (2014c). Visualizing random forest’s prediction results. *Psychology*, 5, 2084-2098. doi: 10.4236/psych.2014.519211
- Golino, H. F., Gomes, C. M. A., & Andrade, D. (2014). Predicting academic achievement of high-school students using machine learning. *Psychology*, 5, 2046-2057. doi: 10.4236/psych.2014.518207
- Golino, H. F., & Gomes, C. M. A. (2015). Investigando estágios de desenvolvimento do raciocínio indutivo usando a análise fatorial confirmatória, o Modelo Logístico Simples de Rasch e o modelo de teste logístico linear (Rasch Estendido). In H. F. Golino, C. M. Gomes, A. Amantes, & G. Coelho, *Psicometria Contemporânea: Compreendendo os Modelos Rasch* (pp. 283-331). São Paulo: Casa do Psicólogo/Pearson.

- Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Method in Education*, 39(4), 401-421. doi: 10.1080/1743727X.2016.1168798
- Golino, H. F. & Gomes, C. M. A. (2019) *TDRI: Teste de Desenvolvimento do Raciocínio Indutivo*. São Paulo: Hogrefe. Golino, H. F., Gomes, C. M. A., Amantes, A., & Coelho, G. (2015). *Psicometria contemporânea: compreendendo os Modelos Rasch* (1<sup>st</sup> ed., p. 416). São Paulo: Casa do Psicólogo.
- Golino, H. F., Gomes, C. M. A., Commons, M. L., & Miller, P. M. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin*, 19(3), 37-54. doi: 10.1037/h0100589
- Gomes, C. M. A. (2005). Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio (ENEM). *Tese de Doutorado*. Programa de Pós-Graduação em Educação da Universidade Federal de Minas Gerais. <http://hdl.handle.net/1843/FAEC-85RJNN>
- Gomes, C. M. A. (2007a). *Apostando no desenvolvimento da inteligência; em busca de um novo currículo educacional para o desenvolvimento do pensamento humano*. Rio de Janeiro: Lamparina.
- Gomes, C.M.A. (2007b). Softwares educacionais podem ser instrumentos psicológicos. *Psicologia Escolar e Educacional*, 11(2), 391-401. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-85572007000200016&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-85572007000200016&lng=pt&tlng=pt).
- Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: notas escolares e inteligência fluida. *Psicologia em Estudo*, 15(4), 841-849. doi: 10.1590/S1413-73722010000400020
- Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo). *Avaliação Psicológica*, 9(3), 449-459. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712010000300011&lng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300011&lng=pt).
- Gomes, C. M. A. (2010c). Perfis de estudantes e a relação entre abordagens de aprendizagem e rendimento Escolar. *Psico (PUCRS. Online)*, 41(4), 503-509. <http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/6336>
- Gomes, C. M. A. (2011a). Abordagem profunda e abordagem superficial à aprendizagem: diferentes perspectivas do rendimento escolar. *Psicologia: Reflexão e Crítica*, 24(3), 438-447. doi: 10.1590/S0102-79722011000300004

- Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC). *Estudos de Psicologia (Natal)*, 16(3), 235-242. doi:10.1590/S1413-294X2011000300005
- Gomes, C. M. A. (2012a). A estrutura fatorial do inventário de características da personalidade. *Estudos de Psicologia (Campinas)*, 29(2), 209-220. doi:10.1590/S0103-166X2012000200007
- Gomes, C. M. A. (2012b). Validade de construto do conjunto de testes de inteligência cristalizada (CTIC) da bateria de fatores cognitivos de alta-ordem (BaFaCAIO). *Gerais: Revista Interinstitucional de Psicologia*, 5(2), 294-316. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt).
- Gomes, C. M. A. (2013). A construção de uma medida em abordagens de aprendizagem. *Psico (PUCRS. Online)*, 44(2), 193-203. <http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11371>
- Gomes, C. M. A. (2020a). Análises estatísticas para estudos de intervenção. In M. MansurAlves & J. B. Lopes-Silva, *Intervenção cognitiva: dos conceitos às práticas baseadas em evidências para diferentes aplicações* (pp. 93-107). Belo Horizonte: T.Ser.
- Gomes, C. M. A. (2020b). Programa de Enriquecimento Instrumental: evidências de eficácia para intervenção cognitiva. In M. Mansur-Alves & J. B. Lopes-Silva, *Intervenção cognitiva: dos conceitos às práticas baseadas em evidências para diferentes aplicações* (pp. 621-639). Belo Horizonte: T.Ser.
- Gomes, C. M. A., & Almeida, L. S. (2017). Advocating the broad use of the decision tree method in education. *Practical Assessment, Research & Evaluation*, 22(10), 1-10. <https://pareonline.net/getvn.asp?v=22&n=10>
- Gomes, C. M. A., Almeida, L. S., & Núñez, J. C. (2017). Rationale and applicability of exploratory structural equation modeling (ESEM) in psychoeducational contexts. *Psicothema*, 29(3), 396-401. doi: 10.7334/psicothema2016.369
- Gomes, C.M.A., Amantes, A., & Jelihovschi, E.G. (2020). Applying the regression tree method to predict students' science achievement. *Trends in Psychology*, 28, 99-117. doi: 10.9788/s43076-019-00002-5
- Gomes, C. M. A., Araujo, J., Nascimento, E., & Jelihovisch, E. (2018). Routine Psychological Testing of the Individual Is Not Valid. *Psychological Reports*, 122(4), 1576-1593. doi: 10.1177/0033294118785636

- Gomes, C. M. A., Araujo, J., & Jelihovschi, E. G. (2020). Approaches to learning in the nonacademic context: construct validity of Learning Approaches Test in Video Game (LAT-Video Game). *International Journal of Development Research*, 10(11), 41842-41849. doi: 10.37118/ijdr.20350.11.2020
- Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, 6(2), 167-179. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt).
- Gomes, C. M. A., & Borges, O. N. (2008a). Avaliação da validade e fidedignidade do instrumento crenças de estudantes sobre ensino-aprendizagem (CrEA). *Ciências & Cognição (UFRJ)*, 13(3), 37-50. <http://www.cienciasecognicao.org/revista/index.php/cec/article/view/60>
- Gomes, C. M. A., & Borges, O. (2008b). Limite da validade de um instrumento de avaliação docente. *Avaliação Psicológica*, 7(3), 391-401. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712008000300011&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712008000300011&lng=pt&tlng=pt).
- Gomes, C. M. A., & Borges, O. (2008c). Qualidades psicométricas de um conjunto de 45 testes cognitivos. *Fractal: Revista de Psicologia*, 20(1), 195-207. doi:10.1590/S1984-02922008000100019
- Gomes, C. M. A. & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, 20(42), 73-88. doi: 10.18222/eae204220092060
- Gomes, C. M. A.s, & Borges, O. N. (2009b). Propriedades psicométricas do conjunto de testes da habilidade visuo-espacial. *PsicoUSF*, 14(1), 19-34. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt).
- Gomes, C. M. A., & Borges, O. (2009c). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica*, 8(1), 17-32. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712009000100003&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000100003&lng=pt&tlng=pt).
- Gomes, C. M. A., de Araújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the Cattell-Horn-Carroll model on the intraindividual approach. *Behavioral Development Bulletin*, 19(4), 22-30. doi: 10.1037/h0101078

- Gomes, C. M. A., Fleith, D. S., Marinho-Araujo, C. M., & Rabelo, M. L. (2020). Predictors of students' mathematics achievement in secondary education. *Psicologia: Teoria e Pesquisa*, 36, e3638. doi:10.1590/0102.3772e3638
- Gomes, C. M. A., & Gjikuria, J. (2017). Comparing the ESEM and CFA approaches to analyze the Big Five factors. *Avaliação Psicológica*, 16(3), 261-267. doi:10.15689/ap.2017.1603.12118
- Gomes, C. M. A., & Gjikuria, E. (2018). Structural Validity of the School Aspirations Questionnaire (SAQ). *Psicologia: Teoria e Pesquisa*, 34, e3438. doi:10.1590/0102.3772e3438
- Gomes, C. M. A., & Golino, H. F. (2012a). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? *Psicologia: teoria e prática*, 14(1), 126-139. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt).
- Gomes, C. M. A., & Golino, H. F. (2012b). Relações hierárquicas entre os traços amplos do Big Five. *Psicologia: Reflexão e Crítica*, 25(3), 445-456. doi:10.1590/S0102-79722012000300004
- Gomes, C. M. A., & Golino, H. F. (2012c). Validade incremental da Escala de Abordagens de Aprendizagem (EABAP). *Psicologia: Reflexão e Crítica*, 25(4), 400-410. doi:10.1590/S0102-79722012000400001
- Gomes, C. M. A., & Golino, H. F. (2014). Self-reports on students' learning processes are academic metacognitive knowledge. *Psicologia: Reflexão e Crítica*, 27(3), 472-480. doi: 10.1590/1678-7153.201427307
- Gomes, C. M. A., & Golino, H. (2015). Factor retention in the intra-individual approach: Proposition of a triangulation strategy. *Avaliação Psicológica*, 14(2), 273-279. doi: 10.15689/ap.2015.1402.12
- Gomes, C. M. A., Golino, H. F., & Costa, B. C. G. (2013). Dynamic system approach in psychology: proposition and application in the study of emotion, appraisal and cognitive achievement. *Problems of Psychology in the 21st Century*, 6, 15-28. <http://www.journals.indexcopernicus.com/abstracted.php?level=5&icid=1059487>
- Gomes, C. M. A., Golino, H. F., & Menezes, I. G. (2014). Predicting School Achievement Rather than Intelligence: Does Metacognition Matter? *Psychology*, 5, 1095-1110. doi: 10.4236/psych.2014.59122
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2016). Investigando a validade estrutural das competências do ENEM: quatro domínios correlacionados ou um modelo bifatorial.

- Boletim na Medida (INEP-Ministério da Educação)*, 5(10), 33-30.  
<http://portal.inep.gov.br/documents/186968/494037/BOLETIM+NA+MEDIDA+-+N%C2%BA+10/4b8e3d73-d95d-4815-866c-ac2298dff0bd?version=1.1>
- Gomes, C. M. A. Golino, H. F., & Peres, A. J. S. (2018). Análise da fidedignidade composta dos escores do enem por meio da análise fatorial de itens. *European Journal of Education Studies*, 5(8), 331-344. doi:10.5281/zenodo.2527904
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2020). Fidedignidade dos escores do Exame Nacional do Ensino Médio (Enem). *Psico (RS)*, 54(2), 1-10. doi: 10.15448/1980-8623.2020.2.31145.
- Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. M. T. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra Brasileira. *Psicologia: Reflexão e Crítica*, 24(1), 19-27. doi: 10.1590/S0102-79722011000100004
- Gomes, C. M. A., Golino, H. F., Santos, M. T., & Ferreira, M. G. (2014). Formal-Logic Development Program: Effects on Fluid Intelligence and on Inductive Reasoning Stages. *British Journal of Education, Society & Behavioural Science*, 4(9), 1234-1248.  
<http://www.sciencedomain.org/review-history.php?iid=488&id=21&aid=4724>
- Gomes, C. M. A., & Jelihovisch, E. (2016). Proposing a new approach and a rigorous cut-off value for identifying precognition. *Measurement*, 93, 117-125. doi: 10.1016/j.measurement.2016.06.066
- Gomes, C. M. A., & Jelihovschi, E. (2019). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education*, 43(2), 201–221. doi: 10.1080/1743727X.2019.1654992
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020). Comparing the predictive power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1), 87-96. doi: 10.15689/ap.2020.1901.17737.10
- Gomes, C. M. A., & Linhares. (2018). Investigação da validade de conteúdo do TAPPensamento. Pôster. *I Encontro Anual da Rede Nacional de Ciência para Educação (CPE)*. doi: 10.13140/RG.2.2.31110.40006
- Gomes, C. M. A., Linhares, I. S., Jelihovschi, E. G., & Rodrigues, M. N. S. (2021). Introducing rationality and content validity of SLAT-Thinking. *International Journal of Development Research*, 11(1), 43264-43272, doi: 10.37118/ijdr.20586.01.2021



- Gomes, C. M. A., & Marques, E. L. L. (2016). Evidências de validade dos estilos de pensamento executivo, legislativo e judiciário. *Avaliação Psicológica*, 15(3), 327-336. doi: 10.15689/ap.2016.1503.05
- Gomes, C. M. A., Marques, E. L. L., & Golino, H. F. (2014). Validade Incremental dos Estilos Legislativo, Executivo e Judiciário em Relação ao Rendimento Escolar. *Revista E-Psi*, 2, 31-46. <https://revistaepsi.com/artigo/2013-2014-ano3-volume2-artigo3/>
- Gomes, C. M. A., Nascimento, E., & Peres, A. J. S. (2019). Investigating causal relations in personality by combining path analysis and Search algoritms. Poster. *3rd World Conference on Personality*, World Association for Personality Psychology (WAPP), Hanoi, Vietnam.
- Gomes, C. M. A., & Nascimento, D. F. (2021a). A medida da habilidade de fluência do modelo CHC: apresentando o Teste de Fluência Ideativa 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.35726.28481/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021b). Acesso aberto ao Teste de Fluência Figural da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) como medida da habilidade ampla de fluência do modelo CHC de inteligência. Preprint. doi: 10.13140/RG.2.2.15593.62564/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021c). Acesso aberto e gratuito ao Teste de Fluência Ideativa 1 da BAFACALO. Preprint. doi: 10.13140/RG.2.2.24821.09442/3.
- Gomes, C. M. A., & Nascimento, D. F. (2021d). Acesso aberto e gratuito ao Conjunto de Testes de Inteligência Fluida: Teste de Raciocínio Geral da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.30509.61921/1
- Gomes, C. M. A., & Nascimento, D. F. (2021e). Apresentando o Teste de Flexibilidade de Fechamento da BAFACALO. Preprint. doi: 10.13140/RG.2.2.31920.28164.
- Gomes, C. M. A., & Nascimento, D. F. (2021f). Disponibilizando de forma gratuita e aberta o Teste de Memória Associativa 1 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.29964.03201/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021g). Disponibilizando de forma gratuita e aberta o Teste de Velocidade Numérica da BAFACALO. Preprint. doi: 10.13140/RG.2.2.24114.94407/1
- Gomes, C. M. A., & Nascimento, D. F. (2021h). Medidas de inteligência cristalizada: disponibilizando o Teste de Compreensão Verbal 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.36085.09447/1.

- Gomes, C. M. A., & Nascimento, D. F. (2021i). Medindo a habilidade de rapidez cognitiva do modelo CHC: apresentando o Teste de Velocidade Perceptiva 1 da BAFACALO. Preprint. doi: 10.13140/RG.2.2.28564.83848/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021j). Presenting SLAT-Thinking Second Version and its content validity. *International Journal of Development Research*, 11(3), 45590-45596. doi: 10.37118/ijdr.21368.03.2021
- Gomes, C. M. A., & Nascimento, D. F. (2021k). Projeto de acesso aberto e gratuito aos testes do LAICO: Teste de Raciocínio Lógico da Bateria de Fatores Cognitivos de AltaOrdem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.25476.45445/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021l). Projeto de acesso aberto e gratuito à Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): o Teste de Compreensão Verbal 1 do Conjunto de Testes de Inteligência Cristalizada. Preprint. doi: 10.13140/RG.2.2.22663.32165/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021m). Projeto de acesso aos testes de inteligência da BAFACALO: Teste de Compreensão Verbal 3. Preprint. doi: 10.13140/RG.2.2.10499.84001/2.
- Gomes, C. M. A., & Nascimento, D. F. (2021n). Projeto de acesso da BAFACALO: Teste de Memória Associativa 2. Preprint. doi: 10.13140/RG.2.2.23253.14565/1.
- Gomes, C. M. A., & Nascimento, D. F. (2021o). Teste de Memória Visual da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.33319.47529.
- Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021a). Medindo a inteligência fluida: o Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Preprint. doi: 10.13140/RG.2.2.17087.84641/3.
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021b). Acesso aberto ao Teste de Dobraduras (VZ) da BAFACALO. Preprint. doi: 10.13140/RG.2.2.21853.95201/2.
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021c). Projeto de testes gratuitos e abertos do LAICO: Teste de Velocidade Perceptiva 3 da BAFACALO. Preprint. doi: 10.13140/RG.2.2.36278.42563/2.
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021d). Teste de Velocidade Perceptiva 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): disponibilização aberta e gratuita aos testes de medida de rapidez cognitiva do LAICO. Preprint. doi: 10.13140/RG.2.2.29567.53928/1.

- Gomes, C. M. A., & Valentini, F. (2019). Time series in educational psychology: application in the study of cognitive achievement. *European Journal of Education Studies*, 6(8), 214-229. doi: 10.5281/zenodo.3551953
- Jelihovschi, E. G., & Gomes, C. M. A. (2019). Proposing an achievement simulation methodology to allow the estimation of individual in clinical testing context. *Revista Brasileira de Biometria*, 37(4), 1-10. doi: 10.28951/rbb.v37i4.423
- Laros, J. A., Valentini, F., Gomes, C. M. A., & Andrade, J. M. (2014). Modelos de inteligência. In A. G. Seabra, J. A. Laros, E. C. Macedo & N. Abreu (Eds.), *Inteligência e funções executivas: Avanços e desafios para a avaliação neuropsicológica* (pp. 17-38). São Paulo: Editora Memnon
- Martins, A. A., Gomes, C. M. A., Alves, A. F., Almeida, L. S. (2018). The structure of intelligence in childhood: age and socio-familiar impact on cognitive differentiation. *Psychological Reports*, 121(1), 79-92. doi: 10.1177/0033294117723019
- Matos, D. A. S., Brown, G. T. L., & Gomes, C. M. A. (2019). Bifactor invariance analysis of student conceptions of assessment inventory. *Psico-USF*, 24(4), 737-750. doi: 10.1590/1413-82712019240411
- Mecca, T. P., Dias, N. M., Reppold, C. R., Muniz, M., Gomes, C. M. A., Fioravanti-Bastos, A. C. M., Yates, D. B., Carreiro, L. R. R., & Macedo, E. C. (2015). Funcionamento adaptativo: panorama nacional e avaliação com o adaptive behavior assessment system. *Psicologia: Teoria e Prática*, 17(2), 107-122, doi:0.15348/1980-6906/psicologia.v17n2p107-122
- Monteiro, S., Almeida, L. S., Gomes, C. M. A., & Sinval, J. (2020). Employability profiles of higher education graduates: a person-oriented approach. *Studies in Higher Education*, 1-14. doi: 10.1080/03075079.2020.1761785
- Moura, M. D. G., Gomes, C. M. A., Blanc, S. L., Mesquita, R. A., & Ferreira, E. F. (2014). Development of questionnaire on dentists' knowledge of HIV/AIDS. *Arquivos em Odontologia (UFMG. Online)*, 50(1), 6-12, 2014. doi:10.7308/aodontol/2014.50.1.01
- Muniz, M., Gomes, C. M. A., & Pasian, S. R. (2016). Factor structure of Raven's Coloured Progressive Matrices. *Psico-USF*, 21(2), 259-272. doi: 10.1590/1413-82712016210204
- Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2019). Prediction of arithmetic competence: role of cognitive abilities, socioeconomic variables and the perception of the teacher in early childhood education. *Estudos de Psicologia*, 24(3), 225-236. doi: 10.22491/1678-4669.20190024

- Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2020). Prediction of reading and writing in elementary education through early childhood education. *Psicologia: Ciência e Profissão*, 40, e205497, 1-14. doi: 10.1590/1982-3703003205497
- Pereira, B. L. S., Golino, M. T. S., & Gomes, C. M. A. (2019). Investigando os efeitos do programa de enriquecimento instrumental básico em um estudo de caso único. *European Journal of Education Studies*, 6(9), 35-52. doi:10.46827/ejes.v0i0.2669
- Pinheiro, C. A. R., Gomes, C. M. A., & Braga, A. G. (2009). Construção e validação do Inventário dos Adjetivos de Personalidade - 50 (IAP-50) [Resumo]. In *Anais do IV Congresso Brasileiro de Avaliação Psicológica e XIV Conferência Internacional de Avaliação Psicológica: Formas e Contextos* (p. 182). Campinas, SP: Instituto Brasileiro de Avaliação Psicológica.
- Pires, A. A. M., & Gomes, C. M. A. (2017). Three mistaken procedures in the elaboration of school exams: explicitness and discussion. *PONTE International Scientific Researches Journal*, 73(3), 1-14. doi: 10.21506/j.ponte.2017.3.1
- Pires, A. A. M., & Gomes, C. M. A. (2018). Proposing a method to create metacognitive school exams. *European Journal of Education Studies*, 5(8), 119-142. doi:10.5281/zenodo.2313538
- Reppold, C. T., Gomes, C. M. A., Seabra, A. G., Muniz, M., Valentini, F., & Laros, J.A. (2015). Contribuições da psicometria para os estudos em neuropsicologia cognitiva. *Psicologia: teoria e prática*, 17(2), 94-106. doi: 10.15348/1980-6906/psicologia.v17n2p94-106
- Ricci, K., Gomes, C. M. A., Nico, M. A. N., Seabra, A. G. (2020). Programa de Enriquecimento Instrumental (PEI) Básico em crianças com TDAH e dislexia. *Psicologia desde el Caribe*, 37(3), 259-282. <http://rcientificas.uninorte.edu.co/index.php/psicologia/article/view/12397>
- Rodrigues, M. N. S., & Gomes, C. M. A. (2020). Testing the hypothesis that the deep approach generates better academic performance. *International Journal of Development Research*, 10(12), 42925-42935. doi:10.37118/ijdr.20579.12.2020
- Rosa, M. A. C., Gomes, C. M. A., Rocha, N. S., Kessler, F. H. P., Slavutzky, S. M. B., Ferreira, E. F., & Pechansky, F. (2013). Dependence module of the MINI plus adapted for sugar dependence: psychometric properties. *Psicologia: Reflexão e Crítica*, 26(1), 77-86. doi:10.1590/S0102-79722013000100009
- Rosário, V. M., Gomes, C. M. A., & Loureiro, C. M. V. (2019). Systematic review of attention testing in allegedly "untestable" populations. *International Journal of Psychological Research and Reviews*, 2(19), 1-21. doi: 10.28933/ijpr-2019-07-1905

- Sampaio, R. T., Loureiro, C. M. V., & Gomes, C. M. A. (2015). A Musicoterapia e o Transtorno do Espectro do Autismo: uma abordagem informada pelas neurociências para a prática clínica. *Per Musi*, 32, 137-170. doi:10.1590/permusi2015b3205
- Silveira, M. S., & Gomes, C. M. A. (2014). Avaliação do desenvolvimento experiencial de pacientes com prótese ocular: a focalização no atendimento clínico. *Psicologia Clínica*, 26(1), 181-196. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S0103-56652014000100012&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0103-56652014000100012&lng=pt&tlng=pt).
- Silveira, M. B., Gomes, C. M. A., Golino, H. F., & Dias, F. S. (2012). Construção do Teste de Habilidade Experiencial (THE): Evidências iniciais de validade e confiabilidade. *Revista E-psi*, 1, 77-96. <https://revistaepsi.com/artigo/2012-ano2-volume1-artigo4/>
- Therneau, T. & Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. <https://CRAN.R-project.org/package=rpart>
- Valentini, F., Gomes, C. M. A., Muniz, M., Mecca, T. P., Laros, J. A., & Andrade, J. M. (2015). Confiabilidade dos índices fatoriais da Wais-III adaptada para a população brasileira. *Psicologia: teoria e prática*, 17(2), 123-139. doi: 10.15348/1980-6906/psicologia.v17n2p123-139

6 - ESTUDO 2: A Practical methodology to test the complexity cost criterion when pruning trees

Cristiano Mauro Assis Gomes – Universidade Federal de Minas Gerais

Heitor Blesa Farias – Universidade Federal de Minas Gerais

Enio Galinkin Jelihovschi – Universidade Estadual de Santa Cruz

Manuscrito encaminhado à revista *Practical Assessment, Research, and Evaluation*. O artigo já foi avaliado e novamente submetido com as modificações necessárias para possível publicação.

## 6.1 - Introduction

The tree method is a well-established data-driven data analysis approach in the machine learning field that predicts both continuous outcomes by the regression tree and categorical outcomes by the classification tree. The method is relatively old, its first algorithm began in the 1960s, with a wide diversity of techniques, about a hundred algorithms nowadays (Loh, 2014).

There are strong arguments in favor of the tree method over the general linear model techniques with regard to predictive ability. The tree method does not require the data to have any specific characteristics, which makes it suitable for treating any type of data. This is not the case with general linear model techniques which requires the data to have implausible characteristics to work properly, making predictive analysis very difficult. Gomes and Jelihovschi (2019) and Gomes et al. (2021) show that general linear model techniques are very inefficient for handling databases with categorical variables having many categories, ordinal variables, and nonlinear relationships, which are abundant in educational data.

In addition of being a flexible approach, appropriate for the most varied types of data, the tree method generates information about the relationships between the predictors and the outcome (Gomes & Almeida, 2017). This is very advantageous, since other important machine learning methods are very effective in prediction, such as random forest and deep learning, but show almost no contribution in explaining the relationships between variables. The tree method is unique since it puts together, in the same approach, predictive and explanatory analysis (Gomes, Amantes et al., 2020; Gomes, Fleith et al., 2020).

The purpose of any predictive method is to decrease the prediction error. The tree method uses a very interesting algorithm to do this; it splits the data into parts. When splitting the data, the method checks whether the sum of the errors of the split parts generates a smaller predictive error than the predictive error prior to the break. It analyzes the possible splits in the data and evaluates which of them will produce the largest drop in the prediction error. This break is selected. The algorithm continues this procedure until some criterion determines that it stop the splitting (Gomes & Jelihovschi, 2019; Gomes, Lemos et al., 2020).

Following, we explain the rationality of the tree method. First, we create a fictitious example. Suppose we want to predict the performance of 10 students on a mathematics test with scores ranging from 0 to 7 points and that we use two categorical predictors with two categories each. These predictors are sex (male and female) and student motivation to study mathematics (low or high). Suppose again that the 10 students in our example had an average performance of 4.5 points on the mathematics test. Table 1 shows the performance of each of the students

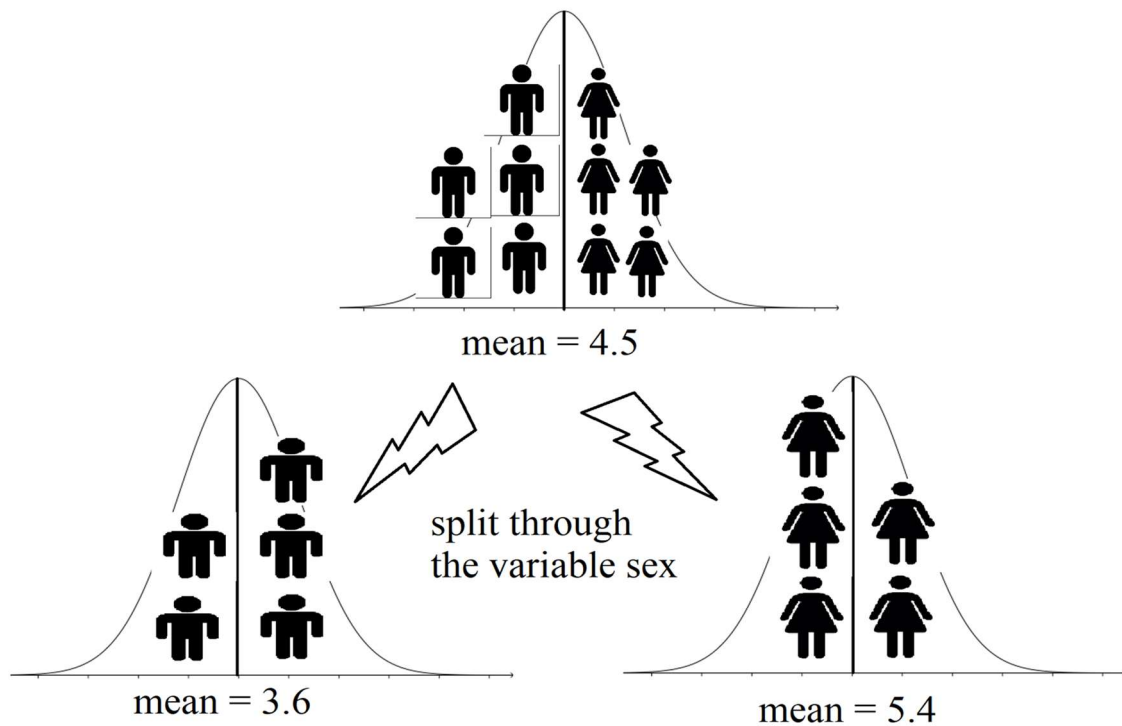
and the average performance, the difference of each student's performance from the average and the squared value of this difference. This difference represents the prediction error about student performance when the mean is understood as the expected value for student performance. Table 1 shows that student 1 has prediction error of 2.25 points, because his performance is not perfectly predicted by the value of the mean, which is 4.5 points. The total error of prediction is 10.5 points.

**Table 1**

*Student's Score, Mean Performance, Differences from Mean, Squared Differences, and Sum of Squared Differences*

Student	Score	Mean	Difference	Square
1	3	4.5	-1.5	2.25
2	3	4.5	-1.5	2.25
3	4	4.5	-0.5	0.25
4	4	4.5	-0.5	0.25
5	4	4.5	-0.5	0.25
6	5	4.5	0.5	0.25
7	5	4.5	0.5	0.25
8	5	4.5	0.5	0.25
9	6	4.5	1.5	2.25
10	6	4.5	1.5	2.25
Mean =	4.5		Sum =	10.5





**Figure 1.** Example of how the tree method algorithm splits the data.

Let the first 5 students, represented in Table 1 be male and the other 5 students female. The tree method will then evaluate which of the two predictors, sex and motivation, gives the best break. Let's assume that the algorithm of tree method first evaluates the squared errors by taking the predictor sex and its two categories as reference. The algorithm splits the data in two parts, the first part being composed of men and the second of women. Figure 1 shows a schematic diagram of this breakdown. The first part of the split data contains the students with grades 3, 3, 4, 4, and 4. The average performance of these 5 students is 3.6 points on the math test and the sum of squared differences in this group is only 1.2 points. The second part of the data contains the students with grades 5, 5, 5, 6, and 6. The average score of this group is 5.4 points and the sum of squared differences is also only 1.2 points. Note that the sum of the squared errors of the two parts is only 2.4 points, much smaller than the 10.5 points value of the entire data set.

We note that the algorithm does not consider only one predictor, but inspects all predictors before judging which one is the best split. In our example, the method will also inspect the predictor motivation in studying mathematics before deciding which of the two predictors it should use to perform the first split on the data. The algorithm will then analyze the split on the data using the motivation predictor. Let's say that, in doing so, it finds that the prediction error is 7.5 points, after adding up the prediction errors from the part of students with

low motivation and from the part of students with high motivation. The algorithm will then compare the squared errors between the two splits and therefore it finds that splitting the data by using sex is much more effective than splitting through motivation to study mathematics. Therefore, the algorithm will choose sex to perform the first real split on the data. The algorithm will continue this process and do as many splits as possible on the new groups being formed until a certain rule prevents it from continuing this process.

Different algorithms carry out the tree method, each one with its own peculiarities to do the splits as well as to decide when to stop and terminate the tree. Our example provides an adequate scenario of the general rationale of the method. The output of the method is a result of the splitting process. This result is called a "tree". The initial data is called the root of the tree. The parts resulting from the various data splits are called nodes and the final nodes are called the leaves of the tree.

Most algorithms of the tree method create an initial tree and then reduce the number of leaves of the tree. This process is called pruning the tree. This is done because the tree method tends to generate overfitting which is a prediction that is too optimal for the sample being analyzed what causes a loss of its predictive value for other samples in the population. One of the mainstream criteria for pruning is the complexity cost (Rokach & Maimon, 2015). It is probably the most widely used criterion since it provides an objective indicator about the size of the tree bearing the prediction which is the best possible for other samples in the population.

Now we present a brief example to explain the cost of complexity criterion. First, suppose we use a tree algorithm and another one that prunes the tree. Second, suppose that the tree algorithm produced an unpruned tree with a total of 50 leaves which explain 90% of the variance of an outcome in the sample. The prediction looks very good. However, suppose that the 50 leaves of the unpruned tree explain only 20% of the variance of that outcome in other samples of the population. This clearly shows a loss in predictive power when other samples are considered. Third, suppose that the first 20 leaves of the tree explain 70% of the variance of the outcome in the sample, which is a worse performance than the previous 50 leaves of the tree. However, suppose again that the first 20 leaves of the tree are able to explain 60% of the variance of the outcome in other samples. This predictive performance is much better than that of the unpruned tree. In short, if we use only the initial 20 leaves, we will have a much better prediction because its predictive power is much better for other samples in the population, although it is worse in relation to the sample being analyzed. Those initial 20 leaves indicate the size of the tree that best predicts the outcome in other samples which is smaller than the initial one. The cost of complexity criterion prunes the tree by searching for this point,

indicating the size of the tree where the prediction is the best possible considering the predictive power relative to other samples.

The complexity cost criterion is carried out by making use of the K-fold cross-validation method (Lantz, 2015). The following example explains it. Suppose we use the 10-fold cross-validation in our example of the 50-leaf unpruned tree. Applying 10-fold means that we randomly split the data into 10 parts. The first nine parts of the data serve as a training sample of the predictive model while the tenth part is used as a test to ascertain the predictive ability of the model. The analysis is done until all parts of the data are used as a test sample. This process generates 10 different predictive analyses. The result of these analyses is assumed to be the performance of the tree method on "other samples" from the population to allow the cost of complexity criterion to identify the tree size where the prediction of the outcome is the best possible on these "other samples".

As far as we know, researchers have used the complexity cost criterion to perform tree pruning based on the literature recommendation that this is a good criterion. In fact, many methodological choices made by researchers come from literature recommendations. Nevertheless, many of them have a weak or even non-existent empirical basis. In the case of pruning criteria, we are unaware of the existence of any methodology in which the researcher is able to evaluate the effectiveness of these criteria to generate an adequate empirical pruned tree. Considering the above argument, we believe that an important methodological advance for the area would be the development of a methodology that would allow the researcher to evaluate whether the cost of complexity criterion is effective, taking his/her own empirical pruned tree as a reference. Despite the fact that in no empirical data does the researcher know the true structure of the data, by explaining our methodology we will show that it is possible to evaluate the effectiveness of the cost of complexity criterion by taking the empirical pruned tree as reference.

We now make use of another example to explain the approach of our methodology. Suppose we have a predictive model and we apply to our data the most famous and recognized tree algorithm, the CART algorithm (Breiman et al., 1984). We use this algorithm to predict a continuous outcome variable, namely the students' annual school grade ranging from 0 to 100 points. Imagine that our sample contains 10,000 students and that we use two predictors, one categorical (boys and girls) and one continuous (previous year's annual student school grade) to predict our outcome. Suppose that by running the CART algorithm and the 10-fold cross-validation method, we obtain in this sample an unpruned tree of 100 leaves, and the information by the cost of complexity criterion that the pruned tree should have 30 leaves. In theory, if the

complexity cost criterion generates an empirical pruned tree that eliminates the leaves from the unpruned tree that are mere noise, then it is correct to assume that the 30-leaf tree is the true structure. Our methodology accepts this assumption of the cost of complexity criterion and fairly evaluates the criterion by taking the 30-leaf tree as the true one.

It is important to explain why our method evaluates the complexity cost criterion by taking as reference the empirical pruned tree generated by a given k-fold. As we noted above, the complexity cost criterion indicated a 30-leaf pruned tree from a specific k-fold. The k-fold cross-validation is a random partitioning of the data, so that every time it is run, it tends to generate variations in the outcome of the cost of complexity criterion and so, different empirical pruned trees. Thus, if we ran the k-fold cross-validation on the 100-leaf unpruned tree again, we might see a different empirical pruned tree. For instance, this new empirical pruned tree could have 37 leaves. Therefore, the evaluation of the complexity cost criterion is always conditional on the empirical pruned tree generated by a specific k-fold.

The rationale of our methodology is shown below:

- 1 Our method runs the CART algorithm using the empirical data and obtain the unpruned tree.
- 2 After step 1, the method runs the k-fold cross-validation and the complexity cost criterion to generate the *empirical pruned tree* which is now assumed to be the true tree.
- 3 The method creates 100 *bootstrapped pruned trees* which simulate the empirical pruned tree structure:
  - 3.a The first step creates a bootstrapped simulation of the empirical data, but only of the predictors from the empirical pruned tree prediction model.
  - 3.b The empirical pruned tree prediction model is used on the data simulated in step 3.a to predict the values of the outcome variable. These values are inserted into the simulated data from step 3.a, so that the simulated data now has the predictors and the outcome.
  - 3.c Then CART, k-fold cross-validation and the complexity cost criterion are run on the data generated in step 3.b.
  - 3.d This process is carried out 100 times generating 100 bootstrapped pruned trees.
- 4 **FIRST EVALUATION OF THE METHOD:** The method evaluates the quality of the 100 bootstrapped pruned trees as follows. The  $R^2$  of the empirical pruned tree and the  $R^2$  of the 100 bootstrapped pruned trees are extracted. Then the method

calculates the coefficient of variation of the  $R^2$  of the bootstrapped pruned trees with respect to the  $R^2$  of the empirical pruned tree. If the coefficient of variation is equal to or less than 10%, then an output informs the user that the method was adequate to generate bootstrapped pruned trees similar to the empirical pruned tree. In short, the statement needs to be favorable to the method for the results of the following steps to be valid.

- 5 EVALUATION OF THE COMPLEXITY COST CRITERION: In this step, the method evaluates whether the complexity cost criterion is adequate. It does this as follows: first, it brings forth the mean, standard deviation, maximum and minimum of the number of leaves of the bootstrapped pruned trees; after, it calculates the coefficient of variation of the leaves of the bootstrapped pruned trees relative to the number of leaves of the empirical pruned tree. If the coefficient of variation is 10% or less, then the method says that the cost of complexity criterion is adequate to perform pruning of the empirical data otherwise, the method says that the cost of complexity criterion is inadequate.
- 6 SECOND EVALUATION OF THE METHOD: In this step, the ability of the method to evaluate the cost of complexity criterion, is itself evaluated. It does this by assessing the probability of the statement produced by the method in step 5 to be wrong:
  - 6.a The method finds the number of leaves in each bootstrapped pruned tree and classifies each of them as suitable if their number of leaves has the same number of leaves (+ or - 10 %) as the empirical pruned tree; otherwise, it classifies them as unsuitable.
  - 6.b The method generates a table showing the total number of adequate and inadequate bootstrapped pruned trees, taking the classification from step 6.a as reference.
  - 6.c In this part we test the adequacy of the method in obtaining good results, that is, results which reflect the truth about the number of leaves of a tree. The test tries to answer the question on whether the values of step 6.a are real or mere random outcomes. The binomial test is used on the values resulting in step 6.b. The null hypothesis  $H_0$  is:  $\pi = 0.5$  that is no difference between suitable and unsuitable bootstrapped pruned trees and the alternative hypothesis  $H_1$  is:  $\pi \neq 0.5$  that is, there a difference between suitable and unsuitable bootstrapped pruned trees. It is important to highlight that the

alternative hypothesis  $H_1$  means that the method is appropriate to evaluate the complexity cost criterion, because it shows that the method is able to discriminate between suitable and unsuitable bootstrapped pruned trees and the results of the method are not random, whereas the null hypothesis  $H_0$  shows no discrimination at all. The p-value will determine the result of the test and its cut value is 0.05, which is the most widely used. In other words, if the p-value is less than 0.05 we should decide for  $H_1$ , otherwise, decide for  $H_0$ .

Our methodology has two limitations. It is restricted to outcomes of continuous variables and it works only via the CART algorithm. However, any researcher can customize it for other algorithms and contexts.

Next, we show how our methodology can be used by giving an example so as to help in the understanding of the rationality of the method and its application. The methodology is carried out using R software (R Core Team, 2020) and our example can be reproduced using the syntaxes of the code.

After presenting the example, we will test our method on different true trees (20, 50, 100 and 500 leaves). To generate the different true trees, we will use the same parameters as the empirical pruned tree from our example. If our method generates adequate bootstrapped pruned trees as well as statements on the complexity cost criterion with a p-value less than .05 on all the true trees, this will be evidence in favor of the adequacy of our method.

## 6.2 - The Data of Our Didactic Example

The National High School Exam (ENEM) is a large-scale standardized educational test in Brazil organized by the National Institute of Educational Studies and Research Anísio Teixeira (INEP), an agency within the Brazilian Ministry of Education (MEC / INEP, 2012). ENEM evaluates four areas of knowledge: Languages, Codes and their Technologies (LC); Mathematics and its Technologies (MT); Nature Sciences and its Technologies (CN); and Human Sciences and its Technologies (CH) (MEC / INEP, 2012). Since 2009 it is used both for the diagnosis of the knowledge acquired by students who finish high school and to select students who wish to enter Brazilian federal universities (MEC / INEP, 2019).

ENEM collects many information about the student and this information is stored by INEP through the so-called microdata. The data in our example are the microdata from the 2011 edition of ENEM. The 2011 microdata are organized in 8 blocks: (1) student data; (2) student school data; (3) data of the municipality where students took the exam; (4) multiple choice

exam data; (5) argumentative essay data; (6) elementary school census data; (7) socioeconomic questionnaire; and (8) data of participants who will apply for high school certification (MEC / INEP, 2012).

Gomes and Jelihovschi (2019) argue that ENEM microdata are an example of data that (1) are not drawn from a specific theory; (2) have large amounts and types of information; (3) contain a large number of nominal variables with many categories; and (4) possibly contain many nonlinear relationships among variables. The tree regression method is well suited for dealing with ENEM-type data because the method does not require any assumptions about the data and handles variables with many categories as well as nonlinear relationships between variables (Gomes & Almeida, 2017; Gomes & Jelihovschi, 2019; Gomes et al., 2021).

The choice for the 2011 microdata is due to the fact that this edition has been extensively studied in terms of the structural validity and reliability of its measures (Gomes et al., 2016, 2018; Gomes, Golino et al., 2020). In our example, we selected only 6 predictor variables from the first and seventh blocks of the ENEM 2011 microdata. These variables were selected based on results of previous studies on the prediction of ENEM scores based on microdata (Gomes, Amantes et al., 2020; Gomes, Fleith et al., 2020; Gomes & Jelihovschi, 2019). The variables selected as predictors were sex, monthly family income, motivation to obtain a high school leaving certificate, motivation to obtain a scholarship, the type of school the student attended in high school, and the student's state of residence. Our outcome variable is represented by the students' scores in the mathematics and its technologies (MT) domain, which are stored in the fourth block of the ENEM 2011 microdata.

Every year INEP makes the microdata of ENEM available to the general public, safeguarding the anonymity of the respondents, as recommended by the code of ethics in research. All microdata, from all editions, are available and can be accessed and free downloaded from the site: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. In our analysis of the microdata from the 2011 version of ENEM, we will use only the data from those who were present in the two days of the exam, answered the socioeconomic questionnaire and wrote the argumentative essay. These conditions reduced the sample from 5,380,856 registrants to a sample of 3,670,089 participants, with participants here being those who met the stipulated conditions. We have made the sample of participants available on the osf.io platform, which can be downloaded via the following link <https://osf.io/b6pd7/download>.

### 6.3 - Methodology in action: an example

Each step in our example is accompanied by a syntax, highlighted in gray, which was built in version 4.0.3 of the statistical software R (R Core Team, 2020). The reader can use the syntaxes in order to reproduce the entire example. An important note: when reproducing the example, the user may find a different result than the one presented in this example, because both the cost complexity criterion by the k-fold cross-validation and bootstrapping produce random variations. Each command in the syntaxes has been numbered to make it easier to explain their functionality. However, in order to properly reproduce the syntaxes, you must remove the numbers that precede the commands. All syntaxes can be run directly in the R console or in an IDE (Integrated Development Environment), such as RStudio (RStudio Team, 2021) or Tinn-R (Tinn-R Team, 2021). We also present the output of the syntaxes, highlighting them with a black border, and comment on their most important aspects. To understand the syntaxes at all, a minimal knowledge of the R language is required.

#### Step 1: Importing the data

The first step in our example consists in importing the empirical data into the R software environment, given that our methodology is applied in this environment. We have devised a data import syntax so that you have access to the data we have deposited at the OSF platform link <https://osf.io/b6pd7/download>.

```

                                SYNTAX 1
1  dir <- getwd()
2  path_file <- paste(dir, "/", "microdata_enem_2011.zip", sep = "")
3  download.file("https://osf.io/b6pd7/download", path_file, mode='wb')
4  microdata_path <- unzip(path_file, exdir = dir)
5  DB_Enem <- read.csv(microdata_path, sep = ",")
```

If you are a user of the R software, then you already know that every time you open R, either from your console or some IDE, such as RStudio or Tinn-R, it starts operating from a certain folder on your computer. The first line of syntax 1 has exactly the function of identifying by the `getwd()` command what this folder is and save this identification in the `dir` object. The second line defines that the data should be downloaded to the folder where R is running. It also defines that the name of the data to be downloaded is "microdata\_enem\_2011.zip". Line 3 extracts the data from the OSF platform and directs it to the folder where R is running on your computer. Line 4 decompresses the data, since it is zipped. Finally line 5 opens the unzipped data in R and saves it as an object with the name `DB_Enem`.

#### Step 2: Handling the data



The data in our example has many variables. Since our interest is only to show as an example, we will restrict our data to the variables we are interested in. Syntax 2 reduces the DB\_Enem object to the only seven variables of interest. All other variables from the original data are removed from this object at syntax 2, line 6. To facilitate manipulation of the DB\_Enem object, the variables of interest are renamed on line 7.

```

                                SYNTAX 2
6      DB_Enem <- subset(DB_Enem, select = c("Sex",
"Student.s.home.location..State.in.Brazil",
"Score.in.Mathematics", "Q04.Student.s.family.monthly.income", "Q26.Motivation.to.tak
e.the.Exam..To.obtain.a.Secondary.Education.certificate.", "Q27.Motivation.to.take.t
he.Exam..To.obtain.a.scholarship.", "Q33.Type.of.school.attended.in.Secondary.Educ
ation..Private..public.or.other.options."))
7  colnames(DB_Enem) <- c("Sex", "UF", "Math", "Q4", "Q26", "Q27", "Q33")

```

The DB\_Enem object has more than 3 million observations, in our case, it is not necessary to use all cases. Furthermore, to analyze such a large data set requires a computer with higher performance configurations (e.g., RAM above 8GB). Since our goal is just the reproduction of the example, we will reduce the sample size to 100,000 observations via syntax 3. This reduction is done by randomly sampling the observations without replacing the cases. If you want to reproduce the same sampling we did, you should run line 8, because the set.seed command allows the result to be reproduced. If you don't do this, your sample of 100,000 observations will be different from ours, due to its random composition. Line 9 determines that the DB\_Enem object has only 100,000 observations.

```

                                SYNTAX 3
8  set.seed(1234)
9  DB_Enem <- DB_Enem[sample(1:3670089, 100000), ]

```

The function of syntax 4 is to check whether the data has been saved correctly in the R object. It must have been saved as a data frame, and the variables must have been correctly transformed into numeric, nominal and ordinal variables.

```

                                SINTAXE 4
10 str(DB_Enem)

```

Output 1 shows two important pieces of information. The first, shown in line 1, indicates that the Enem data has been transformed into a data frame (DB\_Enem) and that the 100,000 observations have been incorporated. The second information indicates how each variable was transformed into the data frame. For example, line 2 shows that the variable sex was transformed into characters (chr). Line 4 shows that the variable math grade was transformed into numeric (num) and line 6 shows that the variable Q26, about the motivation to take ENEM,

was transformed into integer (int). This information is relevant, because it is important to modify variables that have not been correctly transformed. Nominal and categorical variables should be entered into the DB\_Enem object as factors, not characters, so that we will change them.

```

                                OUTPUT 1
1 'data.frame':      100000 obs. of  7 variables:
2  $ Sex: chr  "Male" "Female" "Male" "Female" ...
3  $ UF : chr  "RN" "MS" "MG" "MG" ...
4  $ Math: num  425 458 355 629 431 ...
5  $ Q4 : chr  "Until 1 Minimum Wage" "Until 1 Minimum Wage" "Until 1 Minimum
Wage" "1.5-2 Minimum Wages" ...
6  $ Q26 : int   5 0 0 1 5 5 5 5 3 5 ...
7  $ Q27 : int   5 5 5 5 5 5 5 5 4 5 ...
8  $ Q33 : chr  "Only Public" "Only Public" "Only Public" "Only Private" ...

```

Syntax 5 shows how variables sex, UF, Q4 and Q33, initially transformed as characters, are changed into factors.

```

                                SYNTAX 5
11 DB_Enem$Sex <- factor(DB_Enem$Sex)
12 DB_Enem$UF <- factor(DB_Enem$UF)
13 DB_Enem$Q4 <- factor(DB_Enem$Q4, levels = c("No income",
                                                "Until 1 Minimum Wage",
                                                "1-1.5 Minimum Wage",
                                                "1.5-2 Minimum Wages",
                                                "2-5 Minimum Wages",
                                                "5-7 Minimum Wages",
                                                "7-10 Minimum Wages",
                                                "10-12 Minimum Wages",
                                                "12-15 Minimum Wages",
                                                "15-30 Minimum Wages",
                                                "Above 30 Minimum Wages"))
14 DB_Enem$Q33 <- factor(DB_Enem$Q33)

```

We want to ensure that the categories of the transformed variables are being read properly. Syntax 6 tabulates most of our variables, showing their categories and frequencies.

```

                                SYNTAX 6
15 lapply(DB_Enem[,c(1,2,4,5,6,7)], table)

```

Output 2 shows the results of syntax 6. Here we can see that the categories of the variables have been correctly identified. For example, lines 1 to 3 shows us that our sample consists of 59,524 females and 40,476 males.

```

                                OUTPUT 2
1  $Sex
2    Female      Male
3    59524      40476

4  $UF
5    AC      AL      AM      AP      BA      CE      DF
6    727      1462     2260     423     7188     5536     1215
7    ES      GO      MA      MG      MS      MT      PA
8    2341     2764     4006     11587     2314     2306     4504
9    PB      PE      PI      PR      RJ      RN      RO

```

10	2544	4932	2338	5009	6502	2215	1121
11	RR	RS	SC	SE	SP	TO	
12	284	6193	1781	1059	16642	747	
13	\$Q4						
14			No income			Until 1 Minimum Wage	
15			1691			19267	
16		1-1.5 Minimum Wage				1.5-2 Minimum Wages	
17			25202			28650	
18		2-5 Minimum Wages				5-7 Minimum Wages	
19			11862			5145	
20		7-10 Minimum Wages				10-12 Minimum Wages	
21			3378			1491	
22		12-15 Minimum Wages				15-30 Minimum Wages	
23			1244			1449	
24		Above 30 Minimum Wages					
25			621				
26	\$Q26						
27	0	1	2	3	4	5	
28	34985	4582	4580	7493	9366	38994	
29	\$Q27						
30	0	1	2	3	4	5	
31	5512	1556	1719	3261	5111	82841	
32	\$Q33						
33			Majority Indigenous			Majority in Public	
34			5			3153	
35			Majority in Private			Not Attended	
36			2063			932	
37			Only Public			Only Private	
38			77327			16491	
39		Majority Quilombola Community				Only Quilombola Community	
40			6			2	
41			Only Indigenous				
42			21				

Syntax 7 inspects the mean, standard deviation, quartile ranges, and minimum and maximum values of the outcome variable.

```

16 summary(DB_Enem$Math)
17 sd(DB_Enem$Math)

```

**SYNTAX 7**

Output 3 presents the results of Syntax 7. We can see that our sample had a mean math score of 521.2 and a standard deviation of 116.94, with a minimum of 321.6 and a maximum of 953.0.

OUTPUT 3						
1	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2	321.6	423.5	509.2	521.2	605.3	953.0
3	[1]	116.94				

### Step 3: Package installation

Syntax 8 execute the installation of the packages that will be used by the `complexity_cost.test()` function, prepared by the second author of this paper to operationalize our methodology in the R environment. The `rpart` package (Therneau & Atkinson, 2019) is

installed for the implementation of the CART algorithm. The packages `doParallel` (Corporation & Weston, 2020a) and `foreach` (Corporation & Weston, 2020b) are installed to optimize computational performance during the application of the methodology. The `progress` package (Csárdi & FitzJohn, 2019) is installed to allow the user to visualize the progress of the results coming from the analysis involved in the methodology and to inform the time remaining for the conclusion of its processing. The `psych` package (Revelle, 2021) is installed for the summarized results of the methodology to be presented and the `effectsize` package (Ben-Shachar, Lüdtke, & Makowski, 2020) to calculate the effect size of the hypothesis test that assesses the adequacy of our method. The package `caret` (Kuhn, 2021) has been used to predict the values of the outcome variable.

```

                                SYNTAX 8
18 install.packages("rpart", dependencies = TRUE)
19 install.packages("doParallel", dependencies = TRUE)
20 install.packages("foreach", dependencies = TRUE)
21 install.packages("progress", dependencies = TRUE)
22 install.packages("psych", dependencies = TRUE)
23 install.packages("effectsize", dependencies = TRUE)
24 install.packages("caret", dependencies = TRUE)

```

#### **Step 4: Running the `complexity_cost.test()` function.**

Step 4 of our example is the most important part, because it introduces our methodology. All steps of the method rationality are executed in step 4 via the `complexity_cost.test()` function. The `complexity_cost.test()` function has five arguments: (1) formula, (2) data, (3) `nsplit`, (4) `n`, (5) `kfolds`. In the formula argument, the user must inform the predictive model, defining the outcome variable and its predictors. The model to be informed follows the default structure of any predictive model in the R environment. In the data argument, the user inserts his empirical data to be analyzed by our methodology. The argument `nsplit` indicates the number of splits in order to define the structure of the true tree. By default, they are the same as the splits of the empirical pruned tree, but you can change them if you wish. The argument `n` defines the size of the empirical pruned tree and the bootstrapped pruned trees. By default, the value of `n` is the same as the size of the object in the data parameter, but you can change its value. When the parameter `n` is user-defined with a size different from the empirical data, our function uses R's `sample` command to obtain the `n` demanded by the user. If the user demands an `n` greater than the empirical data, the `sample` command will create new data to increase the sample size until it reaches the `n` desired by the user through bootstrapping. If the user demands an `n` smaller than the empirical data, then the `sample` command will select a random set of values from the empirical data. We suggest changing the size of the `n` only in situations where you want to check whether a larger or smaller empirical and bootstrapped pruned trees might impact the adequacy

of the cost of complexity criterion. Finally, the `kfolds` argument sets the number of folders for K-fold cross-validation. By default, the value is 10, but this value can be changed by the user. We recommend the user to not change the default of this parameter when evaluating the cost of complexity criterion, since the number of folds interferes with the quality of the result of the cost of complexity criterion. The literature defaults to a value of 10, so this is advisable. Very large samples are an exception, the literature recommends 3 folds in that case (Gomes, Amantes et al., 2020; Gomes, Fleith et al., 2020; Gomes & Jelihovschi, 2019).

In short, to run the function, the user only needs to define the predictive model in the `formula` parameter and enter the empirical data in data frame format in the `data` parameter. The script for the `complexity_cost.test()` function can be found in the appendix of this article. To load the function into the R environment you need to copy all the lines from this attachment, paste them into the R console or some IDE, and run the function commands.

Syntax 9 executes the `complexity_cost.test()` function. The first step of the function involves running the predictive model on the empirical data, applying the CART algorithm and the complexity cost criterion via 10-fold cross-validation. Line 25 of syntax 9, shows that the outcome is the variable `Math` and the predictors are `Sex`, `UF`, `Q4`, `Q26`, `Q27` and `Q33`. The empirical data to be analyzed is the data `DB_Enem`. The results of the analysis appear in the summarized formats which are shown in output 4.

```

                                SYNTAX 9
## To load the function complexity_cost.test() into the R environment you need to
## copy all the lines from this attachment, paste them into the R console or some IDE,
## and run the function commands.

25 cct_Enem <- complexity_cost.test(formula = "Math ~ Sex + UF + Q4 + Q26 + Q27 +
Q33", data = DB_Enem)

```

Line 4 of output 4 tells us that the total number of leaves of unpruned tree performed on the empirical data by the CART algorithm was 3346 (the number of leaves is determined in CART by the number of splits plus one). Line 5 shows that the cost of complexity criterion determined that the unpruned tree should be divided 200 times, so that the empirical pruned tree should have 201 leaves. Line 7 shows that the empirical pruned tree has 201 leaves, that is, the user chose to create a real tree with the number of leaves defined by the complexity cost criterion.

```

                                OUTPUT 4
1  =====
2  ===== Unpruned tree and pruned tree =====
3  =====
4  Number of leaves of the unpruned tree:  3346
5  Number of splits according to the complexity criterion:  200
6  Number of splits according to the user:  200

```

```

7 Number of leaves according to the user: 201
8 =====

9 = Evaluating the bootstrapped pruned trees =
10 =====
11 100 bootstrapped pruned trees, each sample containing 1e+05 observations.
12 Pruned tree R2: 0.33
13 Descriptive statistics of the bootstrapped pruned trees' R2.
14 mean sd min max
15 0.34 0 0.33 0.35

16 The Coefficient of Variation for the R2 was 2.57 %
17 This is a suitable method to generate bootstrapped pruned trees.
18 =====

19 = Evaluating the complexity cost criterion =
20 =====
21 Descriptive statistics of the bootstrapped pruned trees' leaves.
22 mean sd min max
23 leaves 185.39 17.61 134 224

24 The Coefficient of Variation of the leaves was 11.41 %
25 The pruned tree of the Complexity Cost Criterion is NOT adequate.
26 =====

27 ===== Affirmative Test =====
28 Adequate Inadequate p-value Cohen's h IC-low IC-high Cohen's h
Interpretation
29 64 36 0.006637121 0.57 0.29 0.84 Medium

30 The affirmative test for the Practical Methodology to Test the Complexity
Cost Criterion confirms it is adequate for evaluating the Complexity Cost
Criterion.
31 =====

```

The second part of output 4, which runs from line 9 to 18, evaluates the ability of the method to generate bootstrapped pruned trees similar to the empirical pruned tree, in terms of their R2. Row 12 shows that the R2 of the empirical pruned tree was .33 indicating that this tree explained 33% of the variance of the outcome variable. Row 15 shows the mean, standard deviation, minimum and maximum R2 of the bootstrapped pruned trees. The mean R2 was .34 with a standard deviation of 0, a minimum R2 of .33 and a maximum R2 of .35. Except for the p-value, the results of the method and hence of output 4 consider 2 decimals. In this sense, the reader should understand that the result of 0 standard deviation actually means a standard deviation of 0.00, so there may be some value in the third decimal place, but this value is negligible. Row 16 shows that the coefficient of variation of the R2 was 2.57%. The interpretation of this value is shown in row 17, indicating that the method was adequate for generating bootstrapped pruned trees. The result of the second stage of output 4 is central because if the method produces inadequate bootstrapped pruned trees, then all the rest of the output is invalid.

The third part of output 4 concerns the evaluation of the cost of complexity criterion. Line 23 shows that the average number of leaves das bootstrapped pruned trees is 185 leaves

( $M = 185.39$ ), which is below the true value of 201 leaves. The coefficient of variation of the bootstrapped pruned trees relative to the empirical pruned tree was 11.41% (See line 24). This value of the coefficient of variation indicates that the cost of complexity criterion was not adequate to generate an appropriate empirical pruned tree (See line 25).

The fourth part of output 4 evaluates the ability of the method itself to evaluate the cost of complexity criterion. Row 29 shows that of the 100 bootstrapped pruned trees, 64 generated adequate results and 36 inadequate results. The p-value of the binomial hypothesis test indicates that there is a 0.66 % probability that the statement generated by the method in the third part of output 4 is wrong. This p-value is less than 5%, so we reject the null hypothesis that the adequate and inadequate values are equal and conclude that the method is adequate to produce the statement evaluating the cost of complexity criterion. The effect size between the adequate and inadequate values is calculated using Cohen's  $h$ . Its value was 0.57, indicating a medium effect size (see row 29).

Now, it is important to make some final remarks. Our example used only an empirical pruned tree, but the function `complexity_cost.test()` allows the user to create true trees of different sizes. Syntax 10 provides an example of how to proceed to generate distinct true trees. The `nsplit` parameter is used to define the number of splits that will generate the number of true leaves. For example, the vector `c(0, 19, 49, 99, 499)` in syntax 10 shows that the user is asking the function to create five different true trees. The value 19 in this vector indicates that a true tree be created with 19 splits, generating a true tree of 20 leaves, since the number of true leaves is determined by the number of splits plus one. In syntax 10 true trees of 20, 50, 100 and 500 leaves are generated. The exception is the number zero. The number zero indicates the generation of a true tree pruned by the complexity cost criterion.

#### SYNTAX 10

```
26 cct_Enem_2 <- complexity_cost.test(formula = "Math ~ Sex + UF + Q4 + Q26 + Q27 + Q33", data = DB_Enem, nsplit = c(0,19,49,99,499), n = 100000, kfolds = 10)
```

It is possible to generate true trees with a different size. For instance, in syntax 11 the parameter `n` was changed to 1000. This implies that the user is determining that the empirical pruned tree and bootstrapped pruned tree has 1000 cases. Furthermore, it is possible for the user to change the number of K-folds of the complexity cost criterion that will be used to generate both the empirical pruned tree and the bootstrapped pruned trees. In syntax 11 the user is determining 20 folds. Only when determining true trees can the user create a vector, as in the example in syntax 10. For the `n`- and `k`-fold parameters, if you want to run different values, you have to run the function again with a new value.

## SYNTAX 11

```
27 cct_Enem_3 <- complexity_cost.test(formula = "Math ~ Sex + UF + Q4 + Q26 + Q27 +
Q33", data = DB_Enem, n = 1000, kfolds = 20)
```

Suppose that the method was adequate, according to the evaluation criteria of step 2 and 4 of its output, and also suppose that the criterion of the cost of complexity of the empirical pruned tree was adequate. It would not be useful for the user to get this information and not have access to the output of the empirical pruned tree. Line 28 of syntax 12 uses the same predictors and data as in the example. This line also defines that two true trees may be produced, one with 20 leaves and one with 25, according to the settings in the `nsplit` parameter. It also stipulates that the empirical pruned tree and the bootstrapped pruned trees have 10,000 observations and are generated through a 10-fold cross-validation. All results generated in line 28 were saved in the object called "cct\_Enem\_4".

## SYNTAX 12

```
28 cct_Enem_4 <- complexity_cost.test(formula = "Math ~ Sex + UF + Q4 + Q26 + Q27 +
Q33", data = DB_Enem, nsplit = c(19, 24), n = 10000, kfolds = 10)
29 cct_Enem_4$results
30 library(rpart.plot)
31 dpi=1200
32 png("empiricalprunedtree_20l.png", height = 15, width = 30, units = 'cm', res=dpi)
33 split.fun <- function(x, labs, digits, varlen, faclen)
34 {
35   # replace commas with spaces (needed for strwrap)
36   labs <- gsub(",", " ", labs)
37   for(i in 1:length(labs)) {
38     # split labs[i] into multiple lines
39     labs[i] <- paste(strwrap(labs[i], width=25), collapse="\n")
40   }
41   labs}
42 rpart.plot(cct_Enem_4$results[[1]]$empirical_pruned_tree,
43           box.palette="Grays", split.fun=split.fun,
44           cex=0.6, nn=TRUE, fallen.leaves=F, space=1, nn.cex=0.6,
45           branch=0.9, type=0, yesno=TRUE, branch.lwd=10)
46 dev.off()
47 dpi=1200
48 png("empiricalprunedtree_25l.png", height = 15, width = 30, units = 'cm', res=dpi)
49 rpart.plot(cct_Enem_4$results[[2]]$empirical_pruned_tree,
50           box.palette="Grays", split.fun=split.fun,
51           cex=0.6, nn=TRUE, fallen.leaves=F, space=1, nn.cex=0.6,
52           branch=0.9, type=0, yesno=TRUE, branch.lwd=10)
53 dev.off()
```

The object "cct\_Enem\_4" gives access to the two empirical pruned trees generated. Line 29 shows the command that allows access to a lot of important information. We will not show the output generated by line 29, but only highlight the sequence of information in this output. Line 28 generates two empirical pruned trees, the output of the command in line 29 first shows information concerning the first empirical pruned tree generated, that is, the 20-leaf tree. Next,



the same about the second empirical pruned tree, the 25 leaves tree. For each empirical pruned tree the output in row 29 first presents descriptive statistics about each bootstrapped pruned tree produced in row 28. It then reports the parameters of the empirical pruned tree. These parameters allow the user to have direct access to this tree. To get these parameters the user simply enters the command on line 29 and adds information to the command that specifically indicates the object in which the empirical pruned tree is stored. For the first empirical pruned tree, the command is `cct_Enem_4$results[[1]]$empirical_pruned_tree`, while for the second it is `cct_Enem_4$results[[2]]$empirical_pruned_tree`. Lines 30 to 53 allow the user to get a picture for each of the two empirical pruned trees generated in line 28. This figure shows the nodes of the tree, their leaves, the predictors and the respective cut-off values used for the tree splits. Syntax 12 defines that the figures will be saved in the directory where R is running, with the names `empiricalprunedtree_20l.png` and `empiricalprunedtree_25l.png`. We will not show the figures generated by syntax 12, because of space limitations.

#### 6.4 - Testing the method in different true trees

We now create a 20, 50, 100 and 500-leaf true tree using the parameters of the true tree from our example. These real trees have 100,000 cases and 10 folds for cross-validation.

**Table 2**

*Results of the Practical Methodology to Test the Complexity Cost Criterion in 20, 50, 100 and 500-leaf true tree*

True Leaves	R2 [Mean (SD)]	CV R2	Leaves [Mean (SD)]	CV Leaves	Adequated/ Inadequated	p-value	Cohen's h	Cohen's h interpretation
20	0.30 (0.00)	3.23%	20.17 (0.40)	2.18%	100/0	1.577722e-30	3.14	Large
50	0.32 (0.00)	1.85%	51.45 (2.24)	5.31%	94/6	2.00596e-21	2.15	Large
100	0.33 (0.00)	1.77%	104.67 (5.12)	6.91%	91/9	3.322049e-18	1.92	Large
500	0.34 (0.00)	6.24%	316.98 (46)	37.61%	1/99	1.593499e-28	-2.74	Large

Table 2 shows that the method was adequate to generate bootstrapped pruned trees on all true trees, because the coefficients of variation of the R2 (CV R2) were less than 10%. If the method is not able to produce bootstrapped pruned trees with R2 similar to the R2 of the empirical pruned tree, then all subsequent results of the method are invalid. The method was also adequate to evaluate the complexity cost criterion on all true trees, as the smallest p-value on these trees closest to .05 was 3.322049e-18, indicating that there is an extremely low

probability that the statement generated by the method about the adequacy of the complexity cost criterion is wrong.

Since the method was appropriate for testing the complexity cost criterion on all true trees in the simulation, we will present the results of the method with respect to the complexity cost criterion. The method evaluated the cost of complexity criterion as appropriate for the 20, 50, and 100 leaf true trees because the coefficient of variation (CV Leaves in Table 2) of the bootstrapped pruned trees relative to the empirical pruned tree was less than 10%. The method evaluated the complexity cost criterion as inadequate for the 500-leaf true tree.

It is important to note that the evaluation of the method on itself is independent of the evaluation it produces on the cost of complexity criterion. This can be seen when we compare the results of the example and the simulation themselves. Notice that on the 20, 50 and 100 leaf true trees the method proved adequate to generate appropriate bootstrapped pruned trees and to evaluate the cost of complexity criterion, with a very low p-value. For these trees, the cost of complexity criterion was adequate. On the 500-leaf true tree, on the other hand, the method proved adequate, also with a very low p-value, but the cost of complexity criterion proved inadequate. Finally, the example presented a 200-leaf true tree where the method proved adequate. In this example, the cost of complexity criterion proved inadequate. In short, the method being evaluated as adequate does not imply that the cost of complexity criterion will be evaluated as adequate or inadequate.

## **6.5 - Conclusion**

In this paper, we have presented an original and novel methodology that allows the researcher to verify whether the cost of complexity criterion is suitable to be used in his empirical data, taking as reference the empirical pruned tree generated by that criterion with the specific k-fold. To the best of our knowledge, our methodology is the first to allow such type of analysis.

In addition to the methodological proposition, we have made our methodology fully accessible to the academic community by making it operational in the statistical software R. A user with minimal knowledge in R is able to apply our methodology. Finally, throughout this text we presented our methodology in a didactic way, using a simple language to reach everyone interested in this theme. We also evaluated the adequacy of our method by using simulation.

We hope that our article will encourage researchers to always test the complexity cost criterion when using the tree method. Pruning is a key step in the tree method, and our methodology allows the evidence of this method to have a good scientific robustness.

## 6.6 - References

- Ben-Shachar M, Lüdtke, D. & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall/CRC.
- Corporation, M., & Weston, S. (2020a). doParallel: Foreach parallel adaptor for the 'parallel' package (R package version 1.0.16) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=doParallel>
- Corporation, M., & Weston, S. (2020b). foreach: Provides foreach looping construct (R package version 1.5.1) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=foreach>
- Csárdi, G., & FitzJohn, R. (2019). progress: Terminal progress bars (R package version 1.2.2) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=progress>
- Gomes, C. M. A., & Almeida, L. S. (2017). Advocating the Broad Use of the Decision Tree Method in Education. *Practical Assessment, Research & Evaluation*, 22(10), 1-10. <https://doi.org/10.7275/y36w-hg55>
- Gomes, C. M. A., & Jelihovschi, E. (2019). Presenting the Regression Tree Method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education*, 43(2), 201-221. <https://doi.org/10.1080/1743727X.2019.1654992>
- Gomes, C. M. A., Amantes, A., & Jelihovschi, E. G. (2020). Applying the regression tree method to predict students' science achievement. *Trends in Psychology*, 28, 99–117. <https://doi.org/10.9788/s43076-019-00002-5>
- Gomes, C. M. A., Fleith, D. de S., Marinho-Araujo, C. M., & Rabelo, M. L. (2020). Preditores do Desempenho em Matemática de Estudantes do Ensino Médio. *Psicologia: Teoria e Pesquisa*, 36, e3638. <https://dx.doi.org/10.1590/0102.3772e3638>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2016). Investigando a validade estrutural das competências do ENEM: quatro domínios correlacionados ou um modelo bifatorial. *Boletim na Medida* (INEP-Ministério da Educação), 5(10), 33-38. Retrived from <http://portal.inep.gov.br/documents/186968/494037/BOLETIM+NA+MEDIDA+-+N%C2%BA+10/4b8e3d73-d95d-4815-866cac2298dff0bd?version=1.1>

- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2018). Análise da fidedignidade composta dos escores do Enem por meio da análise fatorial de itens. *European Journal of Education Studies*, 5(8), 331-344. <https://doi.org/10.5281/zenodo.2527904>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2020). Fidedignidade dos escores do Exame Nacional do Ensino Médio (ENEM). *Psico*, 51(2), e31145. <https://doi.org/10.15448/1980-8623.2020.2.31145>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020). Comparing the predictive power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1), 87-96. <https://doi.org/10.15689/ap.2020.1901.17737.10>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2021). The reasons why the Regression Tree Method is more suitable than General Linear Model to analyze complex educational datasets. *Revista Portuguesa De Educação*, 34(2), 42–63. <https://doi.org/10.21814/rpe.18044>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2012). *Microdados Enem 2011*. Retrieved from <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019). *Microdados do Enem 2019*. Retrieved from <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>
- Kuhn, M. (2021). caret: Classification and Regression Training (R package version 6.0-90) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=caret>
- Lantz, B. (2015). *Machine learning with R*. Birmingham: Packt Publishing.
- Loh, W-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3), 329-348. <https://doi.org/10.1111/insr.12016>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Revelle, W. (2021). psych: Procedures for psychological, psychometric, and personality research (R package version 2.1.6) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=psych>
- Rokach, L., & Maimon, O. (2015). *Data mining with decision trees: Theory and applications*. Singapore: World Scientific Publishing.
- RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>

- Therneau, T., & Atkinson, B. (2019). rpart: Recursive partitioning and regression trees (R package version 4.1-15) [Computer software]. The Comprehensive R Archive Network. Available from <https://CRAN.R-project.org/package=rpart>
- Tinn-R Team (2021). Tinn-R Editor - GUI for R Language and Environment. URL <https://tinn-r.org/en/>

## 7 - Considerações finais

Esta dissertação objetivou desenvolver uma metodologia prática para a avaliação da adequação do critério do custo de complexidade para podar árvores de regressão. Dois estudos foram desenvolvidos para mostrar a relevância dessa metodologia. No primeiro estudo, foi realizada uma simulação e os resultados evidenciam que o critério do custo de complexidade é sensível ao tamanho da amostra. Em amostras pequenas, o critério subestima o número de folhas na árvore. Já em amostras grandes, o método superestima o número de folhas. Este resultado é muito importante para estudos de predição que utilizam o método de árvore, pois alerta à comunidade acadêmica que as árvores podadas pelo critério do custo de complexidade podem não ser adequadas para tratar o *overfitting* e encontrar a melhor previsão possível para outras amostras da população. Estes resultados foram obtidos via simulação, onde o número verdadeiro de folhas nas árvores simuladas era conhecido. Nos estudos empíricos, o número verdadeiro de folhas da árvore não é conhecido. Desta forma, a avaliação do critério do custo de complexidade em dados empíricos se torna difícil, pois não são conhecidos parâmetros de comparação para avaliar se o critério subestimou ou superestimou o número de folhas.

Para tratar essa problemática, no segundo estudo foi criada uma metodologia que permite os pesquisadores testarem a adequação do critério do custo de complexidade para podar árvores em dados empíricos (objetivo específico um). Esta metodologia é inovadora e apresenta uma grande contribuição para os estudos que utilizam o algoritmo CART para fazer estudos de predição, pois com esta metodologia é possível testar a adequação do critério do custo de complexidade em podar árvores de regressão. Considerando a relevância desta metodologia, no estudo dois sua racionalidade foi apresentada (objetivo específico dois), foi criada uma função na linguagem R para operacionalizá-la (objetivo específico três) e foram feitas simulações em diferentes tamanhos de árvores para testá-la (objetivo específico quatro). Os resultados encontrados nas simulações são indicativos da adequação da metodologia para avaliar o critério do custo de complexidade. Com estes resultados, esta dissertação contribui para os diversos campos do saber que utilizam o algoritmo CART para fazer predições e explicações teóricas (Dwyer et al, 2018; Franco, 2021; Gomes & Almeida, 2017; Korkmaz, & Correia, 2019; Loh, 2014; Shatte et al, 2019; Yarkoni, & Westfall, 2017). Novos estudos podem transpor e testar a adequação da metodologia para a técnica de decisão em árvore em que os desfechos são categóricos.

## 8 - Referências

Obs. Nesta seção são incluídas unicamente as referências citadas na apresentação, introdução e considerações finais da dissertação. As referências dos estudos 1 e 2 são incluídas no final de cada estudo nas suas respectivas seções.

- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2017). Cognitive performance and academic achievement: How do family and school converge? *European Journal of Education and Psychology*, 10(2), 49-56. <https://doi.org/10.1016/j.ejeps.2017.07.001>
- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2016). Social and cultural contexts change but intelligence persists as incisive to explain children's academic achievement. *PONTE: International Scientific Researches Journal*, 72(9), 70-89. <https://doi.org/10.21506/j.ponte.2016.9.6>
- Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. S. (2018). The structure of intelligence in childhood: age and socio-familiar impact on cognitive differentiation. *Psychological Reports*, 121(1), 79-92. <https://doi.org/10.1177/0033294117723019>
- Alves, F. A., Flores, R. P., Gomes, C. M. A., Golino, H. F. (2012). Preditores do rendimento escolar: inteligência geral e crenças sobre ensino-aprendizagem. *Revista E-PSI*, 1, 97-117. Portuguese. <https://revistaepsi.com/artigo/2012-ano2-volume1-artigo5/>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2018). Reliability inter-examiners of the Nordoff Robbins Musical Communicativeness Scale brazilian version. In Davi Alves Mota & Tairone Nunes Magalhães, *Proceedings of the 11th International Conference of Students of Systematic Musicology*. (pp. 101-105). Belo Horizonte, Minas Gerais. <https://doi.org/10.5281/zenodo.1345176>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2019). Tradução e validação das Escalas Nordoff Robbins: “Relação criança terapeuta na experiência musical coativa” e “Musicabilidade, formas de atividades, estágios e qualidades de engajamento. In Regina Antunes Teixeira dos Santos & Marcos Nogueira (eds.), *Anais Completos do XIV Simpósio Internacional de Cognição e Artes Musicais*. (pp. 486-493). Campo Grande/MS. ISSN: 2236-4366
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020a). Análise de confiabilidade da Escala de Comunicabilidade Musical. *Per Musi*, 40, 1-12. e204016. Portuguese. <https://doi.org/10.35699/2317-6377.2020.12459>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020b). Confiabilidade inter-examinadores da Escala de Relação Criança-Terapeuta na Experiência Musical Coativa

- para validação no contexto brasileiro. *Revista Música Hodie*, 20, e64243. Portuguese. <https://doi.org/10.5216/mh.v20.64243>.
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020c). Confiabilidade inter-examinadores da versão brasileira da Escala Nordoff Robbins de Comunicabilidade Musical. In Javier Albornoz (org.), *Estudos Latino-Americanos sobre Música*: vol II. (pp. 152-163). Curitiba: Artemis. [https://doi.org/10.37572/EdArt\\_13210092015](https://doi.org/10.37572/EdArt_13210092015)
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020d). *Equivalência de itens, semântica e operacional da “Escala de Musicabilidade: Formas de Atividade, Estágios e Qualidades de Engajamento”*. Orfeu, 5(2), 1-22. Portuguese. <https://doi.org/10.5965/2525530405022020e0010>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2020e). Equivalência de itens, semântica e operacional da Escala Nordoff Robbins de Relação Criança-Terapeuta na Experiência Musical Coativa. *Percepta*, 8(1), 125-144. Portuguese. <https://doi.org/1.34018/2318-891X>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2021a). Estudo de revisão da utilização das escalas Nordoff Robbins: “Relação Criança-Terapeuta na Experiência Musical Coativa” e “Musicabilidade: Formas de Atividade, Estágios e Qualidades de Engajamento”. *Revista Música*, 21(1), 443-468. Portuguese. <https://doi.org/10.11606/rm.v21i1.173943>
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2021b). Measuring the structural validity of two Nordoff-Robbins scales for a patient with autism. In Silvia Inés Del Valle Navarro & Gustavo Adolfo Juarez, *Ciências humanas: estudos para uma visão holística da sociedade: vol I*. (pp.51-66). Curitiba: Artemis. [https://doi.org/10.37572/EdArt\\_2706213786](https://doi.org/10.37572/EdArt_2706213786)
- André, A. M. B., Gomes, C. M. A., & Loureiro, C. M. V. (2021c). Measuring the structural validity of two Nordoff-Robbins scales for a patient with tuberous sclerosis. In Francisca de Fátima dos Santos Freire (org.), *Serviços e cuidados em saúde 3*. (pp. 195-212). Ponta Grossa: Atena. <https://doi.org/10.22533/at.ed.00221180619>
- André, A. M., Gomes, C. M. A., Loureiro, C. M. V. (2016). Escalas Nordoff Robbins: uma revisão bibliográfica. *Percepta*, 3(2), 117-131, 2016. Portuguese. [https://doi.org/10.34018/2318-891X.3\(2\)117-131](https://doi.org/10.34018/2318-891X.3(2)117-131)
- André, A. M., Gomes, C. M. A., Loureiro, C. M. V. (2017). Equivalência de itens, semântica e operacional da versão brasileira da Escala Nordoff Robbins de Comunicabilidade Musical. *OPUS (BELO HORIZONTE. ONLINE)*, 23(2), 197-215. Portuguese.



- <http://www.anppom.com.br/revista/index.php/opus/article/view/459>
- Araújo, A. M., Gomes, C. M. A., Almeida, L. S., & Núñez, J. C. (2018). A latent profile analysis of first-year university students' academic expectations. *Anales De Psicología / Annals of Psychology*, 35(1), 58-67. <https://doi.org/10.6018/analesps.35.1.299351>
- Branco, P. C. C., & Farias, H. B. (2020). Cientometria e bibliometria do campo da abordagem centrada na pessoa e Gestalt-terapia no Brasil: Análise das redes de autoria e produção. *Phenomenology, Humanities and Sciences*, 1(1), 18-43. Retrived from <https://phenomenology.com.br/index.php/phe/article/view/6>
- Branco, P. C. C., Farias, H. B., & Leite, L. S. (2017). Influências humanistas na produção de conhecimento psicológico da Pontifícia Universidade Católica-Campinas. *Revista do NUFEN*, 9(3), 1-15. <http://dx.doi.org/10.26823/RevistadoNUFEN.vol09.n03artigo11>
- Branco, P. C. C., Farias, H. B., Carpes, C. D. O., & Leite, L. S. (2015). Producción de artículos en Psicodrama en Brasil: revisión sistemática (1996-2014). *Revista Brasileira de Psicodrama*, 23(2), 16-23. <http://dx.doi.org/10.15329/2318-0498.20150003>
- Cardoso, C. O., Seabra, A. G., Gomes, C. M. A., & Fonseca, R. P. (2019). Program for the neuropsychological stimulation of cognition in students: impact, effectiveness, and transfer effect on student cognitive performance. *Frontiers in Psychology*, 10, 1-16. <https://doi.org/10.3389/fpsyg.2019.01784>
- Casanova, J. R., Gomes, C. M. A., Bernardo, A. B., Núñez, J. C., & Almeida, L. S. (2021). Dimensionality and reliability of a screening instrument for students at-risk of dropping out from higher education. *Studies in Educational Evaluation*, 68, 100957. <https://doi.org/10.1016/j.stueduc.2020.100957>
- Castillo-Diaz, M.A., Gomes, C.M.A. (2022). Monitoring and Intelligence as Predictors of a Standardized Measure of General and Specific Higher Education Achievement. *Trends in Psychology*, 1, 1-21. <https://doi.org/10.1007/s43076-022-00160-z>
- Costa, B. C. G., Gomes, C. M. A., & Fleith, D. S. (2017). Validade da Escala de Cognições Acadêmicas Autorreferentes: autoconceito, autoeficácia, autoestima e valor. *Avaliação Psicológica*, 16(1), 87-97. Portuguese. <https://doi.org/10.15689/ap.2017.1601.10>
- Costa, V. T., Gomes, C. M. A., Andrade, A. G. P., & Samulski, Di. M. (2012). Validação das propriedades psicométricas do RESTQ-Coach na versão brasileira. *Motriz: Revista de Educação Física*, 18(2), 218-232. Portuguese. <https://doi.org/10.1590/S1980-65742012000200002>
- Dias, N. M., Gomes, C. M. A., Reppold, C. T., Fioravanti-Bastos, A., C., M., Pires, E. U., Carreiro, L. R. R., & Seabra, A. G. (2015). Investigação da estrutura e composição das

- funções executivas: análise de modelos teóricos. *Psicologia: teoria e prática*, 17(2), 140-152. Portuguese. <https://doi.org/10.15348/1980-6906/psicologia.v17n2p140-152>
- Diaz, M. A. C., & Gomes, C. M. A. (2021a). Presenting the Meta-Performance Test, a metacognitive battery based on performance. *International Journal of Educational Methodology*, 7(2), 289-303. <https://doi.org/10.12973/ijem.7.2.289>
- Diaz, M. A. C., & Gomes, C. M. A. (2021b). Validade estrutural do Meta-Texto: evidências que permitem repensar os componentes metacognitivos. Conference. *10 Congresso Brasileiro de Avaliação Psicológica*. <https://doi.org/10.13140/RG.2.2.32944.48642>
- Dwyer, D., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91-118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Ferreira, M. G., & Gomes, C. M. A. (2017). Intraindividual analysis of the Zarit Burden Interview: a Brazilian case study. *Alzheimers & Dementia*, 13, 1163-1164. <https://doi.org/0.1016/j.jalz.2017.06.1710>
- Fleith, D. S., Gomes, C. M. A., Marinho-Araujo, C. M., & Almeida, L. S. (2020). Expectativas de sucesso profissional de ingressantes na educação superior: estudo comparativo. *Avaliação Psicológica*, 19(3), 223-231. Portuguese. <https://doi.org/10.15689/ap.2020.1903.17412.01>
- Fleith, D. S., & Gomes, C. M. A. (2019). Students' assessment of teaching practices for creativity in graduate programs. *Avaliação Psicológica*, 18(3), 306-315. <https://doi.org/10.15689/ap.2019.1803.15579.10>
- Fleith, D. S., Almeida, L. S., Marinho-Araujo, C. M., Gomes, C. M. A., Bisinoto, C., & Rabelo, M. L. (2020). Validity evidence of a scale on academic expectations for higher education. *Paidéia*, 30, e3010. <https://doi.org/10.1590/1982-4327e3010>
- Franco, V. R. (2021). Aprendizado de Máquina e Psicometria: Inovações Analíticas na Avaliação Psicológica. *Avaliação Psicológica*, 20(3), a-c. <https://dx.doi.org/10.15689/ap.2021.2003.ed>
- Gauer, G., Gomes, C. M. A., & Haase V. G. (2010). Neuropsicometria: Modelo clássico e análise de Rasch. In *Avaliação Neuropsicológica*, (pp. 22-30). Porto Alegre: Artmed, 2010. ISBN-10: 8536322101
- Golino, H. F. & Gomes, C. M. A. (2019) *TDRI: Teste de Desenvolvimento do Raciocínio Indutivo*. São Paulo: Hogrefe.
- Golino, H. F., & Gomes, C. M. A. (2011). Preliminary internal validity evidences of two Brazilian Metacognitive Tests. *International Journal of Testing*, 26, 11-12.

<https://www.intestcom.org/files/ti26.pdf>

- Golino, H. F., & Gomes, C. M. A. (2012). The Structural validity of the Inductive Reasoning Developmental Test for the measurement of developmental stages. *International Journal of Testing*, 27, 10-11. [https://www.researchgate.net/publication/269985854\\_The\\_Structural\\_Validity\\_of\\_the\\_Inductive\\_Reasoning\\_Developmental\\_Test\\_for\\_the\\_Measurement\\_of\\_Developmental\\_Stages](https://www.researchgate.net/publication/269985854_The_Structural_Validity_of_the_Inductive_Reasoning_Developmental_Test_for_the_Measurement_of_Developmental_Stages)
- Golino, H. F., & Gomes, C. M. A. (2014a). Four Machine Learning methods to predict academic achievement of college students: a comparison study. *Revista E-Psi*, 1, 68-101. <https://revistaepsi.com/artigo/2014-ano4-volume1-artigo4/>
- Golino, H. F., & Gomes, C. M. A. (2014b). Psychology data from the “BAFACALO project: The Brazilian Intelligence Battery based on two state-of-the-art models – Carroll’s Model and the CHC model”. *Journal of Open Psychology Data*, 2(1), p.e6. <https://doi.org/10.5334/jopd.af>
- Golino, H. F., & Gomes, C. M. A. (2014c). Visualizing random forest’s prediction results. *Psychology*, 5, 2084-2098. <https://doi.org/10.4236/psych.2014.519211>
- Golino, H. F., & Gomes, C. M. A. (2015a). Aprendendo a ler a expressão de invariância dos parâmetros. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 47-80). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Golino, H. F., & Gomes, C. M. A. (2015b). Avançando na compreensão da invariância dos parâmetros: probabilidade e suas características. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 81-107). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Golino, H. F., & Gomes, C. M. A. (2015c). Investigando estágios de desenvolvimento do raciocínio indutivo usando a análise fatorial confirmatória, o modelo logístico simples de Rasch e o modelo de teste logístico linear (Rasch estendido). In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 283-338). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Golino, H. F., & Gomes, C. M. A. (2015d). O modelo logístico simples de Rasch para dados dicotômicos. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 111-154). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Golino, H. F., & Gomes, C. M. A. (2015e). Teoria da medida e o modelo Rasch. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 13-46). São Paulo: Casa do Psicólogo. ISBN: 97885845989

- Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Method in Education*, 39(4), 401-421. <https://doi.org/10.1080/1743727X.2016.1168798>
- Golino, H. F., Gomes, C. M. A., & Andrade, D. (2014). Predicting academic achievement of high-school students using machine learning. *Psychology*, 5, 2046-2057. <https://doi.org/10.4236/psych.2014.518207>
- Golino, H. F., Gomes, C. M. A., & Peres, A. J. S. (2021). Creating an objective measurement for the ENEM: an analysis using the Rasch model. *Psicologia: Teoria e Prática*, 23(1), 1-21. <https://doi.org/10.5935/1980-6906/ePTPPA12625>
- Golino, H. F., Gomes, C. M. A., Amantes, A., & Coelho, G. (2015). *Psicometria contemporânea: compreendendo os Modelos Rasch* (1<sup>st</sup> ed., p. 416). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Golino, H. F., Gomes, C. M. A., Commons, M. L., & Miller, P. M. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin*, 19(3), 37-54. <https://doi.org/10.1037/h0100589>
- Gomes, C. M. A., & Linhares. (2018). Investigação da validade de conteúdo do TAP-Pensamento. *Pôster. I Encontro Anual da Rede Nacional de Ciência para Educação (CPE)*. <https://doi.org/10.13140/RG.2.2.31110.40006>
- Gomes, C. M. A., Araujo, J., & Castillo-Díaz, M.A. (2021). Testing the Invariance of the Metacognitive Monitoring Test. *Psico-USF*, 26(4), 685-696. <https://doi.org/10.1590/1413-82712021260407>
- Gomes, C. M. A. & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, 20(42), 73-88. Portuguese. <https://doi.org/10.18222/ae204220092060>
- Gomes, C. M. A. & Nascimento, D. F. (2021h). Evidências de validade do Teste de Abordagens de Aprendizagem: Identificação do Pensamento contido em Textos 2. *Anais do XVI Congresso Internacional Galego-Português de Psicopedagogia*, 1 a 3 de Setembro de 2021, UMinho, Braga, Portugal (pp. 2426-2438).
- Gomes, C. M. A. (2002). *Feuerstein e a construção mediada do conhecimento*. Porto Alegre: Artmed Editora.
- Gomes, C. M. A. (2005). *Uma análise dos fatores cognitivos mensurados pelo Exame Nacional do Ensino Médio (ENEM)*. [Tese de Doutorado, Universidade Federal de Minas Gerais]. <http://hdl.handle.net/1843/FAEC-85RJNN>

- Gomes, C. M. A. (2007a). *Apostando no desenvolvimento da inteligência; em busca de um novo currículo educacional para o desenvolvimento do pensamento humano*. Rio de Janeiro: Lamparina.
- Gomes, C. M. A. (2007b). Softwares educacionais podem ser instrumentos psicológicos. *Psicologia Escolar e Educacional*, 11(2), 391-401. Portuguese. <https://doi.org/10.1590/S1413-85572007000200016>
- Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: notas escolares e inteligência fluida. *Psicologia em Estudo*, 15(4), 841-849. Portuguese. <https://doi.org/10.1590/S1413-73722010000400020>
- Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo). *Avaliação Psicológica*, 9(3), 449-459. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712010000300011&lng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300011&lng=pt).
- Gomes, C. M. A. (2010c). Perfis de estudantes e a relação entre abordagens de aprendizagem e rendimento Escolar. *Psico (PUCRS. Online)*, 41(4), 503-509. Portuguese. <http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/6336>
- Gomes, C. M. A. (2011a). Abordagem profunda e abordagem superficial à aprendizagem: diferentes perspectivas do rendimento escolar. *Psicologia: Reflexão e Crítica*, 24(3), 438-447. Portuguese. <https://doi.org/10.1590/S0102-79722011000300004>
- Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC). *Estudos de Psicologia (Natal)*, 16(3), 235-242. Portuguese. <https://doi.org/10.1590/S1413-294X2011000300005>
- Gomes, C. M. A. (2012a). A estrutura fatorial do inventário de características da personalidade. *Estudos de Psicologia (Campinas)*, 29(2), 209-220. <https://doi.org/10.1590/S0103-166X2012000200007>
- Gomes, C. M. A. (2012b). Validade de construto do conjunto de testes de inteligência cristalizada (CTIC) da bateria de fatores cognitivos de alta-ordem (BaFaCAIO). *Gerais: Revista Interinstitucional de Psicologia*, 5(2), 294-316. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt).
- Gomes, C. M. A. (2013). A construção de uma medida em abordagens de aprendizagem. *Psico (PUCRS. Online)*, 44(2), 193-203. Portuguese. <http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11371>
- Gomes, C. M. A. (2020a). Análises estatísticas para estudos de intervenção. In M. Mansur-

- Alves & J. B. Lopes-Silva, *Intervenção cognitiva: dos conceitos às práticas baseadas em evidências para diferentes aplicações* (pp. 93-107). Belo Horizonte: T.Ser.
- Gomes, C. M. A. (2020b). Programa de Enriquecimento Instrumental: evidências de eficácia para intervenção cognitiva. In M. Mansur-Alves & J. B. Lopes-Silva, *Intervenção cognitiva: dos conceitos às práticas baseadas em evidências para diferentes aplicações* (pp. 621-639). Belo Horizonte: T.Ser.
- Gomes, C. M. A. (2021a). Apresentação de uma metodologia para criação de provas metacognitivas. *Conference. XVI Congresso Internacional Galego-Português de Psicopedagogia*. <https://doi.org/10.13140/RG.2.2.33129.62569>
- Gomes, C. M. A. (2021b). Avaliação educacional focada no processo: apresentando o teste SLAT-Thinking 2. *Conference. XVI Congresso Internacional Galego-Português de Psicopedagogia*. <https://doi.org/10.13140/RG.2.2.24903.42408>
- Gomes, C. M. A. (2022a). *Apresentação da Escala de Interesse em Pensar Reflexivamente*. Preprint. <https://doi.org/10.13140/RG.2.2.35149.51684>
- Gomes, C. M. A. (2022b). *Apresentação do Teste Abordagem-em-Processo Versão 2*. Preprint. <https://doi.org/10.13140/RG.2.2.29156.24962>
- Gomes, C. M. A. (2022c). *Projeto de Acesso Aberto: Escala de Abordagens de Aprendizagem (EABAP)*. Preprint. <https://doi.org/10.31219/osf.io/4edy7>
- Gomes, C. M. A. Golino, H. F., & Peres, A. J. S. (2018). Análise da fidedignidade composta dos escores do enem por meio da análise fatorial de itens. *European Journal of Education Studies*, 5(8), 331-344. Portuguese. <https://doi.org/10.5281/zenodo.2527904>
- Gomes, C. M. A., & Almeida, L. S. (2017). Advocating the broad use of the decision tree method in education. *Practical Assessment, Research & Evaluation*, 22(10), 1-10. <https://pareonline.net/getvn.asp?v=22&n=10>
- Gomes, C. M. A., & Borges, O. (2008b). Limite da validade de um instrumento de avaliação docente. *Avaliação Psicológica*, 7(3), 391-401. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712008000300011&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712008000300011&lng=pt&tlng=pt).
- Gomes, C. M. A., & Borges, O. (2008c). Qualidades psicométricas de um conjunto de 45 testes cognitivos. *Fractal: Revista de Psicologia*, 20(1), 195-207. Portuguese. <https://doi.org/10.1590/S1984-02922008000100019>
- Gomes, C. M. A., & Borges, O. (2009c). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica*, 8(1), 17-32. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-)

- 04712009000100003&lng=pt&tlng=pt.
- Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, 6(2), 167-179. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt).
- Gomes, C. M. A., & Borges, O. N. (2008a). Avaliação da validade e fidedignidade do instrumento crenças de estudantes sobre ensino-aprendizagem (CrEA). *Ciências & Cognição (UFRJ)*, 13(3), 37-50. Portuguese. <http://www.cienciasecognicao.org/revista/index.php/cec/article/view/60>
- Gomes, C. M. A., & Borges, O. N. (2009b). Propriedades psicométricas do conjunto de testes da habilidade visuo espacial. *PsicoUSF*, 14(1), 19-34. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt).
- Gomes, C. M. A., & Gjikuria, E. (2018). Structural Validity of the School Aspirations Questionnaire (SAQ). *Psicologia: Teoria e Pesquisa*, 34, e3438. <https://doi.org/10.1590/0102.3772e3438>
- Gomes, C. M. A., & Gjikuria, J. (2017). Comparing the ESEM and CFA approaches to analyze the Big Five factors. *Avaliação Psicológica*, 16(3), 261-267. <https://doi.org/10.15689/ap.2017.1603.12118>
- Gomes, C. M. A., & Golino, H. (2015b). Factor retention in the intra-individual approach: Proposition of a triangulation strategy. *Avaliação Psicológica*, 14(2), 273-279. <https://doi.org/10.15689/ap.2015.1402.12>
- Gomes, C. M. A., & Golino, H. F. (2012a). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? *Psicologia: teoria e prática*, 14(1), 126-139. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt).
- Gomes, C. M. A., & Golino, H. F. (2012b). Relações hierárquicas entre os traços amplos do Big Five. *Psicologia: Reflexão e Crítica*, 25(3), 445-456. Portuguese. <https://doi.org/10.1590/S0102-79722012000300004>
- Gomes, C. M. A., & Golino, H. F. (2012c). Validade incremental da Escala de Abordagens de Aprendizagem (EABAP). *Psicologia: Reflexão e Crítica*, 25(4), 400-410. Portuguese. <https://doi.org/10.1590/S0102-79722012000400001>
- Gomes, C. M. A., & Golino, H. F. (2014). Self-reports on students' learning processes are academic metacognitive knowledge. *Psicologia: Reflexão e Crítica*, 27(3), 472-480.

<https://doi.org/10.1590/1678-7153.201427307>

- Gomes, C. M. A., & Golino, H. F., (2015a). A medida de habilidades cognitivas amplas da Bateria de Fatores Cognitivos de Alta Ordem (BAFACALO): empregando o modelo Rasch bifatorial. In Hudson F. Golino et al., *Psicometria contemporânea: compreendendo os Modelos Rasch* (pp. 361-385). São Paulo: Casa do Psicólogo. ISBN: 97885845989
- Gomes, C. M. A., & Jelihovschi, E. (2016). Proposing a new approach and a rigorous cut-off value for identifying precognition. *Measurement*, 93, 117-125. <https://doi.org/10.1016/j.measurement.2016.06.066>
- Gomes, C. M. A., & Jelihovschi, E. (2019). Presenting the regression tree method and its application in a large-scale educational dataset. *International Journal of Research & Method in Education*, 43(2), 201-221. <https://doi.org/10.1080/1743727X.2019.1654992>
- Gomes, C. M. A., & Marques, E. L. L. (2016). Evidências de validade dos estilos de pensamento executivo, legislativo e judiciário. *Avaliação Psicológica*, 15(3), 327-336. Portuguese. <https://doi.org/10.15689/ap.2016.1503.05>
- Gomes, C. M. A., & Nascimento, D. F. (2021a). *A medida da habilidade de fluência do modelo CHC: apresentando o Teste de Fluência Ideativa 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.35726.28481/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021b). *Acesso aberto ao Teste de Fluência Figural da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) como medida da habilidade ampla de fluência do modelo CHC de inteligência*. Preprint. <https://doi.org/10.13140/RG.2.2.15593.62564/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021c). *Acesso aberto e gratuito ao Conjunto de Testes de Inteligência Fluida: Teste de Raciocínio Geral da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.30509.61921/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021d). *Acesso aberto e gratuito ao Teste de Fluência Ideativa 1 da BAFACALO*. Preprint. <https://doi.org/10.13140/RG.2.2.24821.09442/3>
- Gomes, C. M. A., & Nascimento, D. F. (2021e). *Apresentando o Teste de Flexibilidade de Fechamento da BAFACALO*. Preprint. <https://doi.org/10.13140/RG.2.2.31920.28164>
- Gomes, C. M. A., & Nascimento, D. F. (2021f). *Disponibilizando de forma gratuita e aberta o Teste de Memória Associativa 1 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.29964.03201/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021g). *Disponibilizando de forma gratuita e aberta*



- o Teste de Velocidade Numérica da BAFACALO*. Preprint. <https://doi.org/10.13140/RG.2.2.24114.94407/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021i). *Medidas de inteligência cristalizada: disponibilizando o Teste de Compreensão Verbal 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.36085.09447/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021j). *Medindo a habilidade de rapidez cognitiva do modelo CHC: apresentando o Teste de Velocidade Perceptiva 1 da BAFACALO*. Preprint. <https://doi.org/10.13140/RG.2.2.28564.83848/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021k). *Presenting SLAT-Thinking Second Version and its content validity*. *International Journal of Development Research*, 11(3), 45590-45596. <https://doi.org/10.37118/ijdr.21368.03.2021>
- Gomes, C. M. A., & Nascimento, D. F. (2021l). *Projeto de acesso aberto e gratuito à Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): o Teste de Compreensão Verbal 1 do Conjunto de Testes de Inteligência Cristalizada*. Preprint. <https://doi.org/10.13140/RG.2.2.22663.32165/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021m). *Projeto de acesso aberto e gratuito aos testes do LAICO: Teste de Raciocínio Lógico da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.25476.45445/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021n). *Projeto de acesso aos testes de inteligência da BAFACALO: Teste de Compreensão Verbal 3*. Preprint. <https://doi.org/10.13140/RG.2.2.10499.84001/2>
- Gomes, C. M. A., & Nascimento, D. F. (2021o). *Projeto de acesso da BAFACALO: Teste de Memória Associativa 2*. Preprint. <https://doi.org/10.13140/RG.2.2.23253.14565/1>
- Gomes, C. M. A., & Nascimento, D. F. (2021p). *Teste de Memória Visual da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.33319.47529>
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021a). *Acesso aberto ao Teste de Dobraduras (VZ) da BAFACALO*. Preprint. <https://doi.org/10.13140/RG.2.2.21853.95201/2>
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021c). *Projeto de testes gratuitos e abertos do LAICO: Teste de Velocidade Perceptiva 3 da BAFACALO*. Preprint. doi: 10.13140/RG.2.2.36278.42563/2
- Gomes, C. M. A., & Nascimento, D. F., & Araujo, J. (2021d). *Teste de Velocidade Perceptiva*

- 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): disponibilização aberta e gratuita aos testes de medida de rapidez cognitiva do LAICO*. Preprint. <https://doi.org/10.13140/RG.2.2.29567.53928/1>
- Gomes, C. M. A., & Rodrigues, M. N. S. (2021). *Teste Abordagem-em-Processo*. Preprint. <https://doi.org/10.13140/RG.2.2.17602.71363/2>.
- Gomes, C. M. A., & Rozenberg, M. P. (2021) Bi-factor hierarchical model of procrastination: presentation and initial evidence of validity. In Ezequiel Martins Ferreira (org.), *A pesquisa em psicologia: contribuições para o debate metodológico 2*. (pp. 137-156). Ponta Grossa: Atena. <https://doi.org/10.22533/at.ed.30321151215>
- Gomes, C. M. A., & Valentini, F. (2019). Time series in educational psychology: application in the study of cognitive achievement. *European Journal of Education Studies*, 6(8), 214-229. <https://doi.org/10.5281/zenodo.3551953>
- Gomes, C. M. A., Almeida, L. S., & Núñez, J. C. (2017). *Rationale and applicability of exploratory structural equation modeling (ESEM) in psychoeducational contexts*. *Psicothema*, 29(3), 396-401. <https://doi.org/10.7334/psicothema2016.369>
- Gomes, C. M. A., Araujo, J., & Jelihovschi, E. G. (2020). Approaches to learning in the non-academic context: construct validity of Learning Approaches Test in Video Game (LAT-Video Game). *International Journal of Development Research*, 10(11), 41842-41849. <https://doi.org/10.37118/ijdr.20350.11.2020>
- Gomes, C. M. A., Araujo, J., & Jelihovschi, E. G. (2022). The current assessment of the student's academic achievement is a big mistake. *International Journal of Development Research*, 12(03), 54795–54798. <https://doi.org/10.37118/ijdr.24160.03.2022>
- Gomes, C. M. A., Araujo, J., Lima, I. P. C., Chaves, V. N. B., & Golino, H. F. (2021). Inductive Reasoning Developmental Test – Second Revision (TDRI-SR): content validity. In Ezequiel Martins Ferreira (org.), *A pesquisa em psicologia: contribuições para o debate metodológico*. (pp. 36-49). Ponta Grossa: Atena. <https://doi.org/10.22533/at.ed.1692115124>
- Gomes, C. M. A., Araujo, J., Nascimento, E., & Jelihovschi, E. (2018). Routine Psychological Testing of the Individual Is Not Valid. *Psychological Reports*, 122(4), 1576-1593. <https://doi.org/10.1177/0033294118785636>
- Gomes, C. M. A., Carneiro, I. J. P., & Soares, J. M. T. (2008). Aspectos processuais de programas de educação baseados no ensino do pensamento. In *Anais Completos do IV Colóquio Franco Brasileiro de Filosofia da Educação*, UERJ (pp. 1-19).
- Gomes, C. M. A., de Araújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the

- Cattell-Horn-Carroll model on the intraindividual approach. *Behavioral Development Bulletin*, 19(4), 22-30. <https://doi.org/10.1037/h0101078>
- Gomes, C. M. A., Farias, H. B., & Jelihovschi, E. G. (2022). Approaches to learning does matter to predict academic achievement. *Revista de Psicologia*, 40(2), 905-933. <https://doi.org/10.18800/psico.202202.010>
- Gomes, C. M. A., Farias, H. B., Araujo, J., & Jelihovschi, E. G. (2021). *Pruning trees, complexity cost may not be a good approach: initial evidence*. Preprint OSF. <http://doi.org/10.31219/osf.io/4xcwm>
- Gomes, C. M. A., Fleith, D. S., Marinho-Araujo, C. M., & Rabelo, M. L. (2020). Predictors of students' mathematics achievement in secondary education. *Psicologia: Teoria e Pesquisa*, 36, e3638. <https://doi.org/10.1590/0102.3772e3638>
- Gomes, C. M. A., Golino, H. F., & Costa, B. C. G. (2013). Dynamic system approach in psychology: proposition and application in the study of emotion, appraisal and cognitive achievement. *Problems of Psychology in the 21st Century*, 6, 15-28. <http://www.journals.indexcopernicus.com/abstracted.php?level=5&icid=1059487>
- Gomes, C. M. A., Golino, H. F., & Menezes, I. G. (2014). Predicting School Achievement Rather than Intelligence: Does Metacognition Matter? *Psychology*, 5, 1095-1110. <https://doi.org/10.4236/psych.2014.59122>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2016). Investigando a validade estrutural das competências do ENEM: quatro domínios correlacionados ou um modelo bifatorial. *Boletim na Medida (INEP-Ministério da Educação)*, 5(10), 33-30. Portuguese. <http://portal.inep.gov.br/documents/186968/494037/BOLETIM+NA+MEDIDA+-+N%C2%BA+10/4b8e3d73-d95d-4815-866c-ac2298dff0bd?version=1.1>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2020). Fidedignidade dos escores do Exame Nacional do Ensino Médio (Enem). *Psico (RS)*, 54(2), 1-10. Portuguese. <https://doi.org/10.15448/1980-8623.2020.2.31145>
- Gomes, C. M. A., Golino, H. F., & Peres, A. J. S. (2021). Evidências desfavoráveis ao postulado de cargas fatoriais simples do Exame Nacional do Ensino Médio (ENEM). *Avaliação Psicológica*, 20(3). Portuguese. <http://dx.doi.org/10.15689/ap.2021.2003.15777.05>
- Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. M. T. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra Brasileira. *Psicologia: Reflexão e Crítica*, 24(1), 19-27. Portuguese. <https://doi.org/10.1590/S0102-79722011000100004>
- Gomes, C. M. A., Golino, H. F., Santos, M. T., & Ferreira, M. G. (2014). Formal-Logic

- Development Program: Effects on Fluid Intelligence and on Inductive Reasoning Stages. *British Journal of Education, Society & Behavioural Science*, 4(9), 1234-1248.  
<http://www.sciencedomain.org/review-history.php?iid=488&id=21&aid=4724>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2020). Comparing the predictive power of the CART and CTREE algorithms. *Avaliação Psicológica*, 19(1), 87-96.  
<https://doi.org/10.15689/ap.2020.1901.17737.10>
- Gomes, C. M. A., Lemos, G. C., & Jelihovschi, E. G. (2021). The reasons why the regression tree method is more suitable than general linear model to analyze complex educational datasets. *Revista Portuguesa de Educação*, 34(2), 42-63.  
<http://doi.org/10.21814/rpe.18044>
- Gomes, C. M. A., Linhares, I. S., Jelihovschi, E. G., & Rodrigues, M. N. S. (2021). Introducing rationality and content validity of SLAT-Thinking. *International Journal of Development Research*, 11(1), 43264-43272.  
<https://doi.org/10.37118/ijdr.20586.01.2021>
- Gomes, C. M. A., Marques, E. L. L., & Golino, H. F. (2014). Validade Incremental dos Estilos Legislativo, Executivo e Judiciário em Relação ao Rendimento Escolar. *Revista E-Psi*, 2, 31-46. Portuguese. <https://revistaepsi.com/artigo/2013-2014-ano3-volume2-artigo3/>
- Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021b). *Medindo a inteligência fluida: o Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO)*. Preprint. <https://doi.org/10.13140/RG.2.2.17087.84641/3>
- Gomes, C. M. A., Nascimento, E., & Peres, A. J. S. (2019). Investigating causal relations in personality by combining path analysis and Search algorithms. Poster. *3rd World Conference on Personality, World Association for Personality Psychology (WAPP)*, Hanoi, Vietnam.
- Gomes, C. M. A., Quadros, J. S., Araujo, J., & Jelihovschi, E. G. (2020). Measuring students' learning approaches through achievement: structural validity of SLAT-Thinking. *Estudos de Psicologia*, 25(1), 33-43. <https://doi.org/10.22491/1678-4669.20200004>
- Gomes, C.M.A., Amantes, A., & Jelihovschi, E.G. (2020). Applying the regression tree method to predict students' science achievement. *Trends in Psychology*, 28, 99-117.  
<https://doi.org/10.9788/s43076-019-00002-5>
- Haase, V. G., Gauer, G., & Gomes, C. M. A. (2010). Neuropsicometria: modelos nomotético e ideográfico. In Leandro Malloy-Diniz et al., *Avaliação Neuropsicológica*, (pp. 31-37). Porto Alegre: Artmed. ISBN-10: 8536322101
- Jelihovschi, E. G., & Gomes, C. M. A. (2019). Proposing an achievement simulation

- methodology to allow the estimation of individual in clinical testing context. *Revista Brasileira de Biometria*, 37(4), 1-10. <https://doi.org/10.28951/rbb.v37i4.423>
- Laros, J. A., Valentini, F., Gomes, C. M. A., & Andrade, J. M. (2014). Modelos de inteligência. In A. G. Seabra, J. A. Laros, E. C. Macedo & N. Abreu (Eds.), *Inteligência e funções executivas: Avanços e desafios para a avaliação neuropsicológica* (pp. 17-38). São Paulo: Editora Memnon
- Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3), 329-348. <https://doi.org/10.1111/insr.12016>
- Mansur-Alves, M., Gomes, C. M. A., Peixoto, C. B., Bocardi, M. B., Diniz, M. L. N., Freitas, S. K. P., Pereira, E. G., Alvares-Teodoro, J., Ribeiro, P. C. C., Teodoro, M. L. M. (2021). A longitudinal model for psychological distress in the COVID-19 crisis among brazilian graduate students. *PSICO(RS)*, 52(3), 1-15. <http://dx.doi.org/10.15448/1980-8623.2021.3.41332>
- Martins, A. A., Gomes, C. M. A., Alves, A. F., Almeida, L. S. (2018). The structure of intelligence in childhood: age and socio-familiar impact on cognitive differentiation. *Psychological Reports*, 121(1), 79-92. <https://doi.org/10.1177/0033294117723019>
- Matos, D. A. S., Brown, G. T. L., & Gomes, C. M. A. (2019). Bifactor invariance analysis of student conceptions of assessment inventory. *Psico-USF*, 24(4), 737-750. <https://doi.org/10.1590/1413-82712019240411>
- Mecca, T. P., Dias, N. M., Reppold, C. T., Muniz, M., Gomes, C. M. A., Fioravanti-Bastos, A., C., M., Yates, D. B., Carreiro, L. R. R., & Macedo, E. C. (2015). Funcionamento adaptativo: panorama nacional e avaliação com o adaptive behavior assessment system. *Psicologia: Teoria e Prática*, 17(2), 107-122. <https://doi.org/10.15348/1980-6906/psicologia.v17n2p107-122>
- Monteiro, S., Almeida, L. S., Gomes, C. M. A., & Sinval, J. (2020). Employability profiles of higher education graduates: a person-oriented approach. *Studies in Higher Education*, 47(3), 499-512. <https://doi.org/10.1080/03075079.2020.1761785>
- Moura, M. D. G., Gomes, C. M. A., Blanc, S. L., Mesquita, R. A., & Ferreira, E. F. (2014). Development of questionnaire on dentists? knowledge of HIV/AIDS. *Arquivos em Odontologia (UFMG. Online)*, 50(1), 6-12, 2014. <https://doi.org/10.7308/aodontol/2014.50.1.01>
- Muniz, M., Gomes, C. M. A., & Pasian, S. R. (2016). Factor structure of Raven's Coloured Progressive Matrices. *Psico-USF*, 21(2), 259-272. <https://doi.org/10.1590/1413-82712016210204>

- Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2019). Prediction of arithmetic competence: role of cognitive abilities, socioeconomic variables and the perception of the teacher in early childhood education. *Estudos de Psicologia*, 24(3), 225-236. <https://doi.org/10.22491/1678-4669.20190024>
- Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A., & Seabra, A. G. (2020). Prediction of reading and writing in elementary education through early childhood education. *Psicologia: Ciência e Profissão*, 40, e205497, 1-14. <https://doi.org/10.1590/1982-3703003205497>
- Pereira, B. L. S., Golino, M. T. S., & Gomes, C. M. A. (2019). Investigando os efeitos do Programa de Enriquecimento Instrumental Básico em um estudo de caso único. *European Journal of Education Studies*, 6(7), 35-52. Portuguese. <https://doi.org/10.5281/zenodo.3477577>
- Pinheiro, C. A. R., Gomes, C. M. A., & Braga, A. G. (2009). Construção e validação do Inventário dos Adjetivos de Personalidade - 50 (IAP-50) [Resumo]. In *Anais do IV Congresso Brasileiro de Avaliação Psicológica e XIV Conferência Internacional de Avaliação Psicológica: Formas e Contextos* (p. 182). Campinas, SP: Instituto Brasileiro de Avaliação Psicológica.
- Pires, A. A. M., & Gomes, C. M. A. (2017). Three mistaken procedures in the elaboration of school exams: explicitness and discussion. *PONTE International Scientific Researches Journal*, 73(3), 1-14. <https://doi.org/10.21506/j.ponte.2017.3.1>
- Pires, A. A. M., & Gomes, C. M. A. (2018). Proposing a method to create metacognitive school exams. *European Journal of Education Studies*, 5(8), 119-142. <https://doi.org/10.5281/zenodo.2313538>
- Reis, C. P., Morales, J. C. P., Gomes, C. M. A., Pereira, F. A. A., & Ibáñez, S. J. (2021). Construct validation of a new instrument to measure declarative tactical knowledge in basketball. *Perceptual and Motor Skills*, 128(4), 1712-1729. <https://doi.org/10.1177/00315125211016247>
- Reppold, C. T., Gomes, C. M. A., Seabra, A. G., Muniz, M., Valentini, F., & Laros, J. A. (2015). Contribuições da psicometria para os estudos em neuropsicologia cognitiva. *Psicologia: Teoria e Prática*, 17(2), 94-106. Portuguese. <https://doi.org/10.15348/1980-6906/psicologia.v17n2p94-106>
- Ricci, K., Gomes, C. M. A., Nico, M. A. N., & Seabra, A. G. (2020). Programa de Enriquecimento Instrumental (PEI) básico em crianças com TDAH e Dislexia. *Psicologia desde el Caribe*, 37(3), 1-29. Portuguese. <http://rcientificas.uninorte.edu.co/index.php/psicologia/article/viewFile/12397/214421>

444770.

- Rodrigues, M. N. S., & Gomes, C. M. A. (2020). Testing the hypothesis that the deep approach generates better academic performance. *International Journal of Development Research*, 10(12), 42925-42935. <https://doi.org/10.37118/ijdr.20579.12.2020>
- Rokach, L., & Maimon, O. (2015). Data mining with decision trees: Theory and applications. Singapore: World Scientific Publishing.
- Rosa, M. A. C., Gomes, C. M. A., Rocha, N. S., Kessler, F. H. P., Slavutzky, S. M. B., Ferreira, E. F., & Pechansky, F. (2013). Dependence module of the MINI plus adapted for sugar dependence: psychometric properties. *Psicologia: Reflexão e Crítica*, 26(1), 77-86. <https://doi.org/10.1590/S0102-79722013000100009>
- Rosário, V. M., Gomes, C. M. A., & Loureiro, C. M. V. (2019). Systematic review of attention testing in allegedly "untestable" populations. *International Journal of Psychological Research and Reviews*, 2(19), 1-21. <https://doi.org/10.28933/ijpr-2019-07-1905>
- Rosário, V. M., Loureiro, C. M. V., & Gomes, C. M. A. (2020). A relação entre música e atenção: fundamentos, evidências neurocientíficas e reabilitação. *Per Musi*, 40, 1-18. E204015. Portuguese. <https://doi.org/10.35699/2317-6377.2020.14912>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Salami, S., Bandeira, P. F. R., Gomes, C. M. A., & Dehkordi, P. S. (2021). The Test of Gross Motor Development—Third Edition: A Bifactor Model, Dimensionality, and Measurement Invariance, *Journal of Motor Learning and Development*, 10(1), 116-131. <https://doi.org/10.1123/jmld.2020-0069>
- Sampaio, R. T., Loureiro, C. M. V., & Gomes, C. M. A. (2015). A Musicoterapia e o Transtorno do Espectro do Autismo: uma abordagem informada pelas neurociências para a prática clínica. *Per Musi*, 32, 137-170. Portuguese. <https://doi.org/10.1590/permusi2015b3205>
- Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, 1–23. <https://doi.org/10.1017/S0033291719000151>
- Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). Metodologia de pesquisa em psicologia. Porto Alegre: AMGH Editora.
- Silveira, M. B., Gomes, C. M. A., Golino, H. F., & Dias, F. S. (2012). Construção do Teste de Habilidade Experiencial (THE): Evidências iniciais de validade e confiabilidade. *Revista E-psi*, 1, 77-96. Portuguese. <https://revistaepsi.com/artigo/2012-ano2-volume1->

artigo4/

- Silveira, M. S., & Gomes, C. M. A. (2014). Avaliação do desenvolvimento experiencial de pacientes com prótese ocular: a focalização no atendimento clínico. *Psicologia Clínica*, 26(1), 181-196. Portuguese. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S0103-56652014000100012&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0103-56652014000100012&lng=pt&tlng=pt).
- Teodoro, M. L. M., Alvares-Teodoro, J., Peixoto C. B., Pereira E.G., Diniz, M. L. N., Freitas, S. K. P., Ribeiro, P. C. C., Gomes, C. M. A., & Mansur-Alves, M. (2021). Mental health in college students during covid-19 pandemic. *REFACS*, 9(2), 372-82. <https://doi.org/10.18554/refacs.v9i2.5409>
- Valentini, F., Gomes, C. M. A., Muniz, M., Mecca, T. P., Laros, J. A., & Andrade, J. M. (2015). Confiabilidade dos índices fatoriais da Wais-III adaptada para a população brasileira. *Psicologia: teoria e prática*, 17(2), 123-139. Portuguese. <https://doi.org/10.15348/1980-6906/psicologia.v17n2p123-139>
- Vieira, F. D., Ribeiro, D. O., Farias, H. B., & Freitas, P. M. (2021). The Working Memory as Predictor of Performance in Arithmetic of Brazilian Students. *Paidéia* (Ribeirão Preto), 31, e3119. <https://doi.org/10.1590/1982-4327e3119>
- Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/17456916176933>



## **9 - ANEXOS**

Nesta seção são apresentados os algoritmos e as funções utilizadas nos estudos. No anexo 1, consta o algoritmo utilizado no estudo 1 que permite simular árvores de regressão. No anexo 2, a função que operacionaliza o método para testar o custo de complexidade, utilizada no estudo 2, está disponibilizada.

## 9.1 - ANEXO 1: Código do R para simular as árvores

```

set.seed(2004)
leafs <- NULL
n=239
library(rpart)

for(i in 1:100){
  TDRI <- rnorm(n, mean = 3.78, sd = 0.96)
  EABAP_AS <- rnorm(n, mean = 2.62, sd = 0.88)
  EABAP_AP <- rnorm(n, mean = 3.81, sd = 0.86)
  TAP_AP <- rnorm(n, mean = 0.11, sd = 0.1)
  TAP_AS <- rnorm(n, mean = 0.79, sd = 0.24)
  TAP_INT_P <- rnorm(n, mean = 0.32, sd = 0.1)
  e<-rnorm(n)

  enem
  510*I (TDRI<2.7)*I (TAP_AS<0.92)+492*I (TDRI<4.4)*I (EABAP_AP<4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<0.7
5)*I (TDRI>=2.7)*I (TAP_AS<0.92)+586*I (TDRI>=4.4)*I (EABAP_AP<4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<0.
75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+558*I (TDRI>=4.8)*I (EABAP_AP>=4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<
0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+628*I (TDRI<4.8)*I (EABAP_AP>=4.2)*I (EABAP_AP>=3.1)*I (TAP_AS
<0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+637*I (EABAP_AP<3.1)*I (TAP_AS<0.75)*I (TDRI>=2.7)*I (TAP_AS<
0.92)+543*I (EABAP_AP<4.3)*I (EABAP_AS<2.7)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.7
5)*I (TDRI>=2.7)*I (TAP_AS<0.92)+621*I (EABAP_AP>=4.3)*I (EABAP_AS<2.7)*I (EABAP_AS>=2.1)*I (TAP_I
NT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+573*I (TAP_AP<0.085)*I (EABAP_AS>=2.7
)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+692*I (TAP
_AP>=0.085)*I (EABAP_AS>=2.7)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.
7)*I (TAP_AS<0.92)+656*I (EABAP_AS<2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP
_AS<0.92)+688*I (TAP_INT_P<0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+596*I (EABAP_AP<
3.4)*I (TAP_AS>=0.92)+609*I (EABAP_AP>=4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+6
09*I (TDRI>=3.9)*I (TAP_INT_P>=0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS
>=0.92)+660*I (TDRI<3.9)*I (TAP_INT_P>=0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*
I (TAP_AS>=0.92)+621*I (EABAP_AS>=3)*I (TAP_INT_P<0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP
_AP>=3.4)*I (TAP_AS>=0.92)+668*I (TDRI<3.9)*I (EABAP_AS<3)*I (TAP_INT_P<0.44)*I (EABAP_AP<4.5)*I (
EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+705*I (TDRI>=3.9)*I (EABAP_AS<3)*I (TAP_INT_P<0.
44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+711*I (EABAP_AS>=3.5)*I (
EABAP_AP>=3.4)*I (TAP_AS>=0.92) + e

  datal <-as.data.frame(cbind(enem,TDRI,EABAP_AS,EABAP_AP,TAP_AS,TAP_AP,TAP_INT_P))
  cfit <- rpart (enem ~ TAP_AP + TAP_INT_P + TAP_AS + EABAP_AS + EABAP_AP + TDRI,na.action
= na.rpart,
               data = datal,
               method = 'anova',
               control = rpart.control(xval = 10, cp = 0))
  cpvalue <- cfit$cpstable[which.min(cfit$cpstable[, "xerror"]), "CP"]
  cfit <- rpart (enem ~ TAP_AP + TAP_INT_P + TAP_AS + EABAP_AS + EABAP_AP + TDRI,
               na.action = na.rpart,
               data = datal,
               method = 'anova',
               control = rpart.control(xval = 10, cp = cpvalue))
  x1 <- cfit$cpstable[which.max(cfit$cpstable[, "nsplit"]), "nsplit"]
  leafs <- append(leafs, x1)
  leafs + c(1)}
summary(leafs)

set.seed(2004)
leafs <- NULL
n=10000

for(i in 1:100){
  TDRI <- rnorm(n, mean = 3.78, sd = 0.96)
  EABAP_AS <- rnorm(n, mean = 2.62, sd = 0.88)
  EABAP_AP <- rnorm(n, mean = 3.81, sd = 0.86)
  TAP_AP <- rnorm(n, mean = 0.11, sd = 0.1)
  TAP_AS <- rnorm(n, mean = 0.79, sd = 0.24)
  TAP_INT_P <- rnorm(n, mean = 0.32, sd = 0.1)
  e<-rnorm(n)

  enem
  510*I (TDRI<2.7)*I (TAP_AS<0.92)+492*I (TDRI<4.4)*I (EABAP_AP<4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<0.7
5)*I (TDRI>=2.7)*I (TAP_AS<0.92)+586*I (TDRI>=4.4)*I (EABAP_AP<4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<0.
75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+558*I (TDRI>=4.8)*I (EABAP_AP>=4.2)*I (EABAP_AP>=3.1)*I (TAP_AS<
0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+628*I (TDRI<4.8)*I (EABAP_AP>=4.2)*I (EABAP_AP>=3.1)*I (TAP_AS
<0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+637*I (EABAP_AP<3.1)*I (TAP_AS<0.75)*I (TDRI>=2.7)*I (TAP_AS<
0.92)+543*I (EABAP_AP<4.3)*I (EABAP_AS<2.7)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.7
5)*I (TDRI>=2.7)*I (TAP_AS<0.92)+621*I (EABAP_AP>=4.3)*I (EABAP_AS<2.7)*I (EABAP_AS>=2.1)*I (TAP_I
NT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+573*I (TAP_AP<0.085)*I (EABAP_AS>=2.7
)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+692*I (TAP
_AP>=0.085)*I (EABAP_AS>=2.7)*I (EABAP_AS>=2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.
7)*I (TAP_AS<0.92)+656*I (EABAP_AS<2.1)*I (TAP_INT_P>=0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP
_AS<0.92)+688*I (TAP_INT_P<0.065)*I (TAP_AS>=0.75)*I (TDRI>=2.7)*I (TAP_AS<0.92)+596*I (EABAP_AP<
3.4)*I (TAP_AS>=0.92)+609*I (EABAP_AP>=4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+6
09*I (TDRI>=3.9)*I (TAP_INT_P>=0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS
>=0.92)+660*I (TDRI<3.9)*I (TAP_INT_P>=0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*
I (TAP_AS>=0.92)+621*I (EABAP_AS>=3)*I (TAP_INT_P<0.44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP
_AP>=3.4)*I (TAP_AS>=0.92)+668*I (TDRI<3.9)*I (EABAP_AS<3)*I (TAP_INT_P<0.44)*I (EABAP_AP<4.5)*I (
EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+705*I (TDRI>=3.9)*I (EABAP_AS<3)*I (TAP_INT_P<0.
44)*I (EABAP_AP<4.5)*I (EABAP_AS<3.5)*I (EABAP_AP>=3.4)*I (TAP_AS>=0.92)+711*I (EABAP_AS>=3.5)*I (
EABAP_AP>=3.4)*I (TAP_AS>=0.92) + e

```

```

5)*I(TDRI>=2.7)*I(TAP_AS<0.92)+621*I(EABAP_AP>=4.3)*I(EABAP_AS<2.7)*I(EABAP_AS>=2.1)*I(TAP_INT_P>=0.065)*I(TAP_AS>=0.75)*I(TDRI>=2.7)*I(TAP_AS<0.92)+573*I(TAP_AP<0.085)*I(EABAP_AS>=2.7)*I(EABAP_AS>=2.1)*I(TAP_INT_P>=0.065)*I(TAP_AS>=0.75)*I(TDRI>=2.7)*I(TAP_AS<0.92)+692*I(TAP_AP>=0.085)*I(EABAP_AS>=2.7)*I(EABAP_AS>=2.1)*I(TAP_INT_P>=0.065)*I(TAP_AS>=0.75)*I(TDRI>=2.7)*I(TAP_AS<0.92)+656*I(EABAP_AS<2.1)*I(TAP_INT_P>=0.065)*I(TAP_AS>=0.75)*I(TDRI>=2.7)*I(TAP_AS<0.92)+688*I(TAP_INT_P<0.065)*I(TAP_AS>=0.75)*I(TDRI>=2.7)*I(TAP_AS<0.92)+596*I(EABAP_AP<3.4)*I(TAP_AS>=0.92)+609*I(EABAP_AP>=4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+609*I(TDRI>=3.9)*I(TAP_INT_P>=0.44)*I(EABAP_AP<4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+660*I(TDRI<3.9)*I(TAP_INT_P>=0.44)*I(EABAP_AP<4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+621*I(EABAP_AS>=3)*I(TAP_INT_P<0.44)*I(EABAP_AP<4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+668*I(TDRI<3.9)*I(EABAP_AS<3)*I(TAP_INT_P<0.44)*I(EABAP_AP<4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+705*I(TDRI>=3.9)*I(EABAP_AS<3)*I(TAP_INT_P<0.44)*I(EABAP_AP<4.5)*I(EABAP_AS<3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92)+711*I(EABAP_AS>=3.5)*I(EABAP_AP>=3.4)*I(TAP_AS>=0.92) + e

data1 <- as.data.frame(cbind(enem, TDRI, EABAP_AS, EABAP_AP, TAP_AS, TAP_AP, TAP_INT_P))
cfrit <- rpart (enem ~ TAP_AP + TAP_INT_P + TAP_AS + EABAP_AS + EABAP_AP + TDRI, na.action = na.rpart,
               data = data1,
               method = 'anova',
               control = rpart.control(xval = 10, cp = 0))
cpvalue <- cfrit$cptable[which.min(cfrit$cptable[, "xerror"]), "CP"]
cfrit <- rpart (enem ~ TAP_AP + TAP_INT_P + TAP_AS + EABAP_AS + EABAP_AP + TDRI,
               na.action = na.rpart,
               data = data1,
               method = 'anova',
               control = rpart.control(xval = 10, cp = cpvalue))
x1 <- cfrit$cptable[which.max(cfrit$cptable[, "nsplit"]), "nsplit"]
leafs <- append(leafs, x1)
leafs + c(1)}
summary(leafs)

```

## 9.2 - ANEXO 2: Função complexity\_cost.test

```
# =====
complexity_cost.test <- function(formula, data, nsplit = 0, n = 0, kfold = 10){

# Step 01 - The function checks if the 'data' object is a data.frame().
# -----
# If it is not a data.frame(), the function stops,
# otherwise, parts 2, 3, 4, and 5 of the function are executed.
if(class(data)[1] != "data.frame"){
  stop("The 'data' must be of the data.frame() class")
}else{
# -----

# Step 02 - Loading necessary packages
# -----
require(rpart)
require(doParallel)
require(foreach)
require(progress)
require(effectsize)
require(caret)
# -----

# Step 03 - Selecting 'data' variables from the 'formula'
#
# Suppose the file 'data' has 5 variables: VD, VI1, VI2, VI3 and VI4
# If we need only 3 variables use this : formula = "VD ~ VI1 + VI2".
# In this way, only these 3 variables will be selected to be
# in the 'data' object.
# In case we need to include all variables, there are two ways to do this:
# either by writing all of them or by using a period after the tilde.
# For example: formula = "VD ~ VI1 + VI2 + VI3 + VI4" or formula = "VD ~."
# Unselected variables are excluded from 'data'.
# -----
if(any(unlist(strsplit(formula, "")) == ".") == TRUE){
  data <- data
  x <- gsub("[~+]", " ", formula)
  x <- unlist(strsplit(x, " "))
  ifelse(any(x == ""), x <- x[-which(x == "")], x <- x)
  x <- unlist(lapply(x, function(x){grep(x,colnames(data))}))
  x <- x[!duplicated(x)]
  data <- data[,x]
}else{
  x <- gsub("[~+]", " ", formula)
  x <- unlist(strsplit(x, " "))
  ifelse(any(x == ""), x <- x[-which(x == "")], x <- x)
  x <- unlist(lapply(x, function(x){grep(x,colnames(data))}))
  data <- data[,x]
# -----

# Step 04 - Inspect the categorical variables to check whether the values use comma
# (brazilian style) and replaces those with a period (R style).

# For example, the categorical variable family income has the categories
# 'Between 1 and 1,5 Minimum Wages' and 'Between 1,5 and 2 Minimum Wages',
# these commas will be replaced by periods.
# From : 'Between 1 and 1,5 Minimum Wages' to 'Between 1 and 1.5 Minimum Wages'
# This will be important for step 08, where a series of results will be separated
# by commas.
# -----
vc <- sapply(data, is.factor)
data[vc] <- lapply(data[vc], as.character)
vcharacter <- sapply(data, is.character)
vp <- unlist(lapply(data[vcharacter], function(x) any(unlist(strsplit(x, "")) == ",")))
```

```

vp <- which(colnames(data) == names(which(vp == TRUE)))
data[vp] <- lapply(data[vp], function(x) gsub(",", ".", x))
rm(vc, vp)
# -----

# Step 05 - Running the CART Algorithm
# First the function checks whether the outcome variable is of type numeric or integer,
# if it is not, the function stops.
# If the outcome variable is of type numeric or integer, the function rpart will be executed
# on the object 'data'.
#
# The unpruned tree is made after the execution of the function rpart.
# -----
if(class(data[,1]) != "integer" & class(data[,1]) != "numeric"){
  stop("The target (dependent) variable should be integer or numeric.")}else{

  n <- nrow(data)
  ifelse(n == 0, n <- nrow(data), n<- n)
  if(n < nrow(data)){data <- data[sample(1:nrow(data),n, replace = FALSE),]}
  if(n > nrow(data)){data <- data[sample(1:nrow(data),n, replace = TRUE),]}

  kfold <- kfold
  model <- rpart(formula, data = data,
                 na.action = na.rpart, method = 'anova',
                 control = rpart.control(xval = kfold, cp = 0))
# -----

# Step 06 - Create the function for the 'Practical Methodology to Test the Complexity Cost
Criterion'

sim_bootstrapping <- function(model_CART, formula, data, nsplit, n, kfold){
  if(class(data)[1] != "data.frame"){
    stop("The 'data' must be of the data.frame() class")
  }else{

# Step 06.a - Identifies the cut point of the unpruned tree
# by the complexity cost criterion and the pruned tree's R2.
# -----
model <- model_CART
nsplit_model <- min(model$cptable[which.min(model$cptable[,4]),2])
min_div <- min(model$cptable[,2])
max_div <- max(model$cptable[,2])
nsplit <- nsplit
ifelse(nsplit == 0, nsplit <- nsplit_model, nsplit <- nsplit)
ifelse(any(model$cptable[,2] == nsplit), nsplit <- nsplit, nsplit <-
model$cptable[which.min(abs(model$cptable[,2]-nsplit)),2])
R2_original <- round(1 - model$cptable[which.min(abs(model$cptable[,2]-nsplit)),4],2)
# -----

# Step 06.b - Create an empty data.frame for bootstrapped pruned trees.
# This data.frame has either the number of rows of the 'data' object or the number
# entered in the 'n' parameter of the function
# -----
n<- n
ifelse(n == 0, n <- nrow(data), n<- n)
new_data <- as.data.frame(matrix(NA,n,ncol(data)))
colnames(new_data) <- colnames(data)
# -----

# Step 06.c - Pruning the unpruned tree generated in step 05.
# -----
parameter_complexity <- data.frame(model$cptable)
modelp <- prune(model, cp= parameter_complexity$nsplit == nsplit,1)
# -----

# Step 06.d - Create the bootstrapped pruned trees.

```

```

# -----
sd_data <- sd(data[,1], na.rm = T)
nsim <- 100; leafs <- list(); j<-1
pb <- progress_bar$new(
  format = "Calculating [:bar] :percent || Time to completion :eta",
  total = nsim, clear = FALSE, width= 60)
cl <- makeCluster(detectCores())
registerDoParallel(cl)
while(j <= nsim){

# Step 06.d.1 - Simulate the predictors of 'data' through bootstrapping.
# The values are placed in the empty data.frame ('new_data') from step 6.b.
# -----
  new_data[,2:ncol(new_data)] <- data[sample(1:nrow(data), n, replace =
TRUE),2:ncol(data)]
# -----

# Step 06.d.2 - Create the values of the outcome variable using the predict function.
# The values are placed in the 'new_data'.
# -----
  new_data[,1] <- predict(modelp, new_data[,2:ncol(new_data)])
  sd_error <- sqrt(sd_data^2 - sd(new_data[,1], na.rm = T)^2)
  e <- rnorm(n,0,sd_error)
  new_data[,1] <- new_data[,1] + e
# -----

# Step 06.d.3 - Run the CART and produce the bootstrapped unpruned tree
# and the bootstrapped pruned tree.
# -----
  cfit <- rpart(formula, na.action = na.rpart,
    data = new_data ,
    method = 'anova',
    control = rpart.control(xval = kfolds,cp = 0))

  cpvalue <- cfit$cpstable[which.min(cfit$cpstable[, "xerror"]), "CP"]
  cfit2 <- prune(cfit, cp= cpvalue)
# -----

# Step 06.d.4 - Descriptive statistics of the outcome variable of 'new_data'.
# Descriptive statistics of the bootstrapped pruned tree.
# -----
  leafs[[j]] <- data.frame(meansample = mean(new_data[,1] , na.rm = T),sdsample =
sd(new_data[,1] , na.rm = T),minsample = min(new_data[,1] , na.rm = T),maxsample =
max(new_data[,1] , na.rm = T),t(cfit2$cpstable[which.max(cfit2$cpstable[, "nsplit"]),]))
# -----

  pb$tick()
  Sys.sleep(1 / 100)
  j<- j+1}
stopCluster(cl)
# -----

# Step 06.e - Output of the 'Practical Methodology to Test the Complexity Cost Criterion'
# -----
  leafs <- do.call(rbind.data.frame, leafs)
  leafs <- cbind(leafs, leaves = leafs$nsplit + 1)
  s <- length(leafs$leaves)
  true_leaf <- nsplit+1
  cv <- round(sqrt(sum(((leafs$leaves) - true_leaf)^2) / (s))/true_leaf * 100,2)
  yleafs <- leafs$leaves
  zleafs <- yleafs >= true_leaf*.9 & yleafs <= true_leaf*1.1
  cbind(yleafs,zleafs)
  zleafs <- as.factor(zleafs)

  ifelse(levels(zleafs)[1] == "FALSE", levels(zleafs) <- c("FALSE","TRUE"), levels(zleafs)
<- levels(zleafs))

```

```

ifelse(levels(zleafs)[1] == "TRUE", levels(zleafs) <- c("TRUE","FALSE"), levels(zleafs)
<- levels(zleafs))

tab_results <- data.frame(matrix(NA,1,7))
colnames(tab_results) <- c("Adequate","Inadequate","p-value","Cohen's h", "IC-low","IC-
high","Cohen's h Interpretation")
rownames(tab_results) <- c("")
result_table <- table(zleafs)
tab_results$Inadequate <- result_table[names(result_table) == "FALSE"]
tab_results$Adequate <- result_table[names(result_table) == "TRUE"]
tab_results$p-value <- binom.test(tab_results$Adequate, n = 100, p = .5, alternative =
"two.sided")$p.value
cont <- matrix(NA,2,2)
cont[1,] <- as.numeric(tab_results[,1:2])
cont[2,] <- 100 - cont[1,]
x<- cohens_h(cont)
tab_results$`Cohen's h` <- round(x$Cohens_h,2)
tab_results$`IC-low` <- round(x$CI_low,2)
tab_results$`IC-high` <- round(x$CI_high,2)
#Using Cohen (1988, p. 198):
if (abs(tab_results$`Cohen's h`) < 0.2){
  tab_results$`Cohen's h Interpretation` <- "Negligible"
} else if (abs(tab_results$`Cohen's h`) < 0.5){
  tab_results$`Cohen's h Interpretation` <- "Small"
} else if (abs(tab_results$`Cohen's h`) < 0.8){
  tab_results$`Cohen's h Interpretation` <- "Medium"
} else {
  tab_results$`Cohen's h Interpretation` <- "Large"}
R2boot <- data.frame(psych::describe(round(1 - leafs$error,2))[,c(3,4,8,9)]);
row.names(R2boot)<-""
cvR2 <- round(sqrt(sum(((round(1 - leafs$error,2)) - R2_original)^2) / (s))/R2_original *
100,2)

cat(" =====")
cat("\n ===== Unpruned tree and pruned tree =====")
cat("\n =====")
cat("\n Number of leaves of the unpruned tree: ", max_div+1)
cat("\n Number of splits according to the complexity criterion: ", nsplit_model)
cat("\n Number of splits according to the user: ", nsplit)
cat("\n Number of leaves according to the user: ", nsplit+1)
cat("\n =====")
cat("\n")
cat("\n = Evaluating the bootstrapped pruned trees =")
cat("\n =====")
cat("\n",nsim, "bootstrapped pruned trees, each sample containing", n, "observations.")
cat("\n Pruned tree R2: ",R2_original)
cat("\n Descriptive statistics of the bootstrapped pruned trees' R2.")
cat("\n")
print(round(R2boot,2))
if(cvR2 > 10){
  cat("\n The Coefficient of Variation for the R2 was",cvR2, "%")
  cat("\n This is a NOT suitable method to generate bootstrapped pruned trees.")
} else {
  cat("\n The Coefficient of Variation for the R2 was",cvR2, "%")
  cat("\n This is a suitable method to generate bootstrapped pruned trees.")
}
cat("\n =====")
cat("\n")
cat("\n = Evaluating the complexity cost criterion =")
cat("\n =====")
cat("\n Descriptive statistics of the bootstrapped pruned trees' leaves.")
cat("\n")
print(round(data.frame(psych::describe(leafs))[10,c(3,4,8,9)],2))
if(cv > 10){
  cat("\n The Coefficient of Variation of the leaves was ",cv, "%")
  cat("\n The pruned tree of the Complexity Cost Criterion is NOT adequated.")
} else {
  cat("\n The Coefficient of Variation of the leaves was ",cv, "%")

```

```

    cat("\n The pruned tree of the Complexity Cost Criterion is adequated.")}
cat("\n =====")
cat("\n")
cat("\n ===== Affirmative Test =====")
cat("\n")
print(tab_results)
if(tab_results$`p-value` < .05) {
  cat("\n The affirmative test for the Practical Methodology to Test the Complexity Cost
Criterion confirms it is adequate for evaluating the Complexity Cost Criterion.")
} else{
  cat("\n The affirmative test for the Practical Methodology to Test the Complexity Cost
Criterion confirms it is inadequate for evaluating the Complexity Cost Criterion.")
cat("\n =====")
cat("\n")
return(list(leafs = leafs, empirical_pruned_tree = modelp))}}}}
# -----

# Step 7 - The sim_bootstrapping function is executed according to the
#          nsplit parameter vector and the results of each execution are
#          saved in a list that will return such values to the users.
# -----
results <- list()
for(p in 1:length(nsplit)){
  results[[p]] <- sim_bootstrapping(model, formula, data, nsplit[p], n=n, kfolds=kfolds)
return(list(results = results, regression = model))}
# -----
# =====

```