

O uso da análise de correspondências múltiplas (ACM) na pesquisa: o auxílio da tabela de proporções condicionais

João Wachelke
Raianne Silva Calixto
Jéssica Bruna Borges Pereira
Priscilla Martins Dornelas

RESUMO

A análise de correspondências múltiplas (ACM) é uma técnica pertinente para a psicologia, pois permite estudar associações entre diversas variáveis categóricas tais como características individuais e respostas a questionários, produzindo um mapa que permite visualizá-las. No entanto, é frequentemente mal interpretada, pois trabalha com um nível alto de dimensões, o que torna difícil verificar os dados originais. O trabalho apresenta uma técnica auxiliar para a ACM, a tabela de proporções condicionais, que apresenta as proporções das variáveis incluídas isoladamente e em relações bivariadas, possibilitando a identificação de associações e sua magnitude de modo concreto. O trabalho inclui a aplicação empírica de análise de atribuições de tarefas domésticas a homens e mulheres, proveniente de pesquisa com 958 adultos que forneceram opiniões em levantamento de dados realizado no interior do Brasil. A principal contribuição da tabela de proporções condicionais está no esforço de tornar mais acessível e precisa a utilização da ACM para pesquisadores de psicologia.

Palavras-chave: análise de correspondências múltiplas; análise de dados exploratória; estatística descritiva.

ABSTRACT

The use of multiple correspondence analysis (MCA) in research: the aid of conditional proportions table

Multiple correspondence analysis (MCA) is a relevant technique for psychology, since it allows the study of associations among various categorical variables such as individual characteristics or questionnaires responses, producing a map that makes it possible to visualize them. However, it is often misinterpreted as it deals with a high level of dimensions that makes it difficult to verify the original data. The paper presents an auxiliary technique for MCA, the conditional proportions table, which presents the proportions of the included variables in isolation and in bivariate relationships, enabling the identification of associations and their size in a concrete way. The paper includes an empirical application of the analysis of the attribution of domestic tasks to men and women, from a study with 958 adults who provided their opinions on a survey carried out in the Brazilian countryside. The main contribution of the conditional proportions table lies in the effort of making the use of MCA more accessible and precise for psychology researchers.

Keywords: multiple correspondence analysis; exploratory data analysis; descriptive statistics.

As análises multivariadas de dados são técnicas estatísticas que avaliam as relações simultâneas de mais de duas variáveis, isto é, múltiplas medidas sobre indivíduos ou objetos (Hair, Anderson, Tatham, & Black, 1998). Em psicologia e ciências sociais, essas análises permitem caracterizar relações sofisticadas entre os fenômenos de interesse, geralmente emoções, ações e pensamentos, seja em termos dos que ocorrem naturalmente, seja no que diz respeito a situações artificiais como em delineamentos experimentais.

Sobre os Autores

J. W.
orcid.org/0000-0003-4364-8598
Universidade Federal de
Uberlândia (UFU) - Uberlândia,
MG
joao.wachelke@ufu.br

R. S. C.
orcid.org/0000-0002-2605-5427
Universidade Federal de
Uberlândia (UFU) - Uberlândia,
MG
raianne.calixto@gmail.com

J. B. B. P.
orcid.org/0000-0002-3655-4406
Universidade Federal de
Uberlândia (UFU) - Uberlândia,
MG
jess.borges.63@gmail.com

P. M. D.
orcid.org/0000-0002-4657-2990
Universidade Federal de
Uberlândia (UFU) - Uberlândia,
MG
pridornelas@hotmail.com

Direitos Autorais

Este é um artigo de acesso aberto e pode ser reproduzido livremente, distribuído, transmitido ou modificado, por qualquer pessoa desde que usado sem fins comerciais. O trabalho é disponibilizado sob a licença Creative Commons CC-BY-NC.

Dentre as análises multivariadas, algumas têm finalidade descritiva, sintetizando dados sem recurso a modelos inferenciais de generalização, reduzindo a informação de diversas variáveis em um número menor de dimensões, também chamados fatores. Essas técnicas, chamadas de métodos fatoriais, podem lidar com variáveis de níveis de mensuração nominais, ordinais, intervalares ou de razão (Baccini, 2010).

Em alguns casos, contudo, as relações identificadas envolvendo diversas variáveis são complexas, de difícil interpretação ou comunicação. Nessas situações, os pesquisadores precisam de estratégias suplementares que lhes permitam demonstrar a importância dessas relações para que tenham maior segurança em termos do que afirmam, e também tornar esses resultados mais transparentes e inteligíveis para o público em geral, que não tenha os dados à disposição para verificar as conclusões.

Este trabalho apresenta uma proposta de cruzamentos bivariados de dados, a tabela de proporções condicionais, para auxiliar na interpretação de relações identificadas em uma técnica específica de análise multivariada de variáveis categóricas: a análise de correspondências múltiplas (ACM). A ACM é bastante popular em estudos psicossociais, pois permite avaliar associações entre características individuais, respostas a questionários, crenças, e assim por diante. Primeiramente faremos uma apresentação conceitual da ACM, indicando suas finalidades e como é interpretada, para em seguida ressaltar as dificuldades de sua interpretação prática. Posteriormente apresentamos a tabela de proporções condicionais e alternativas de visualização gráfica para atenuar essas dificuldades. A pesquisa apresentada é uma ilustração com dados empíricos, e finalizamos com uma discussão das características da técnica proposta.

ANÁLISE DE CORRESPONDÊNCIAS MÚLTIPLAS: INTRODUÇÃO CONCEITUAL

Primeiramente, cabe esclarecer que a ACM é uma técnica de análise estatística de dados categóricos. Por vezes, categorias são consideradas, erroneamente, características necessárias e exclusivas da pesquisa qualitativa (Bauer, Gaskell, & Allum, 2003), especialmente quando associadas a procedimentos interpretativos, como a análise de conteúdo. A pesquisa qualitativa é um enfoque amplo de pesquisa, geralmente baseado em textos e que produz textos, voltado para a compreensão aprofundada de relações sem isolar variáveis. Portanto, ela assume o caráter subjetivo de interpretação dos fenômenos culturais de interesse do pesquisador (Günther, 2006; Ragin & Amoroso, 2011). No contexto de pesquisas com estratégias estatísticas de análise de dados, as categorias dizem respeito a níveis de

mensuração nominais e ordinais. Essas categorias podem ser independentes, sem relação de magnitude entre si, chamadas também de variáveis nominais. É o caso da variável nacionalidade, que pode ser brasileira, argentina, francesa; ou sexo, masculino ou feminino. Além disso, há variáveis categóricas que podem ser ordenadas em termos de intensidade ou magnitude, como classe social baixa, média ou alta; ou grupo etário, jovem, adulto ou idoso.

A ACM permite estudar as correspondências, isto é, atrações e repulsões, entre variáveis categóricas e casos, que em pesquisas psicológicas tendem a ser indivíduos. Uma aplicação comum se dá em estudos de levantamentos de dados, em que pessoas respondem a questionários, seja com opiniões, preferências, registros de comportamentos ou outros. Torna-se possível verificar se quem responde a alguma questão ou item de algum modo tende a selecionar ou rejeitar respostas de outras questões. É possível tanto avaliar as associações de variáveis entre si – modalidades de variáveis associadas tendem a ser escolhidas pelas mesmas pessoas – quanto associações de indivíduos – indivíduos que são representados próximos no mapa tendem a escolher as mesmas opções de resposta nas questões ou itens.

Talvez a aplicação mais célebre da ACM seja a pesquisa sociológica de Bourdieu (1979), em que variáveis de posição social (ocupação, escolaridade, renda) são relacionadas num mesmo espaço de variáveis de consumo cultural, como leitura de livros, frequência a teatros e cinemas, apreciação de música e assim por diante.

Como muitas variáveis psicológicas e psicossociais são nominais, e mesmo variáveis numéricas podem ser recodificadas como variáveis nominais – um escore em algum teste ou escala pode ser convertido em categorias como baixo e alto, por exemplo –, a ACM pode ser útil na exploração de dados de muitos projetos de pesquisa psicológicos. Em psicologia, contudo, as pesquisas que se utilizam dessa técnica são menos numerosas que as análises fatoriais baseadas em variáveis quantitativas, comumente empregadas em investigações de validação de instrumentos psicométricos. Particularmente, o uso da ACM parece ser mais razoavelmente difundido em pesquisas sobre representações sociais, em tarefas de associação de palavras – consideradas como variáveis nominais textuais (mencionou a palavra ou categoria ou não a mencionou) –, ou em termos de variáveis que indiquem atitudes ou crenças. Estímulo importante para essa relativa popularização foi o da obra de Doise, Clemence, & Lorenzi-Cioldi (1992), que desenvolve uma perspectiva de estudo quantitativo das representações sociais.

Em termos de exemplos de pesquisas brasileiras, Lima e Vala (2005) utilizaram a ACM junto a uma amostra de

universitários brasileiros para estudar as relações de indicadores de infra humanização, como sentimentos, emoções, traços de natureza e cultura e atribuição de cor da pele a grupos raciais. Neves e Lima (2007) fizeram análises de correspondências múltiplas para desvendar as relações entre atitudes sobre cotas para negros, posição social e percepções sobre justiça social de estudantes brasileiros. Em Wachelke (2017), a ACM voltou-se para as correspondências entre indicadores de opinião sobre o tema do trabalho. Wachelke, Matos, Ferreira, e Costa (2015) categorizaram os objetos e grupos de pesquisas de representações sociais publicadas em periódicos científicos e incluíram essas categorias numa ACM.

Considerando suas características de técnica estatística, a ACM pode ser entendida como um caso mais complexo da análise de correspondências simples ou binária. A análise de correspondências binária representa graficamente as associações entre linhas e colunas de uma tabela de contingências bivariadas, isto é, um cruzamento simples de variáveis categoriais, como a associação entre preferências por candidatos numa eleição e o nível de escolaridade das pessoas. Quando as tabelas têm poucas células, a inspeção visual é simples, mas quando a tabela possui grandes dimensões, como digamos num caso com muitos candidatos e diversas categorias de escolaridade, a análise de correspondências é particularmente útil porque apresenta descrições parciais dos dados, os fatores, dimensões ou eixos, que caracterizam os principais desvios nas proporções de resposta em relação ao perfil médio dos dados. Com uma representação gráfica, geralmente bidimensional, um mapa num plano cartesiano, é possível visualizar as principais associações de uma tabela (Greenacre, 2007).

A ACM trabalha com uma tabela diferente, chamada matriz indicadora, com tantas linhas quanto casos – em estudos de questionários geralmente são os indivíduos respondentes – e tantas colunas quanto modalidades de variáveis categóricas – se temos duas variáveis com duas modalidades cada, como sexo alto ou baixo e desempenho num teste aprovado ou reprovado, resultam quatro colunas, referentes a cada modalidade. Posteriormente, gera uma matriz combinando todas as tabulações bivariadas entre pares de variáveis (Greenacre, 2007; Lebart, Morineau, & Piron, 1995).

A ACM permite avaliar associações entre os casos (indivíduos) e entre as modalidades das variáveis (colunas). Cada fator, dimensão ou eixo descreve contrastes específicos entre indivíduos, ou entre variáveis: é como se fosse um “ponto de vista” das diferenças dos dados. Essas variações são denominadas inércia, equivalente à noção de variância em outras análises. A importância dos fatores é decrescente

em termos de inércia explicada: o primeiro fator descreve os principais contrastes, o segundo fator descreve os principais contrastes, uma vez que os contrastes do primeiro tenham sido removidos, o terceiro diz respeito aos contrastes seguintes à remoção dos dois primeiros fatores, e assim por diante (Le Roux & Rouanet, 2010).

As modalidades das variáveis e indivíduos obtêm escores em cada um dos fatores, que se refletem em coordenadas. Os contrastes são identificados em termos de identificação de pontos referentes às modalidades ou indivíduos com coordenadas positivas, que se opõem aos que possuem coordenadas negativas. Escores próximos de zero (a origem) indicam respostas próximas ao perfil médio da amostra. Projetando os escores de modalidades e indivíduos com pontos num mapa formado por um plano cartesiano de dois fatores, é possível representar as diferenças e semelhanças entre indivíduos e entre variáveis. Isso é possível porque, devido a características da análise fatorial de correspondências, a distância entre pontos significa que indivíduos que tenham coordenadas próximas num mesmo fator provavelmente selecionaram as mesmas categorias. Modalidades de variáveis representadas próximas foram selecionadas pelos mesmos indivíduos (Lebart et al., 1995).

Como os fatores da ACM permitem sintetizar as informações totais da tabela, evidenciando as diferenças mais importantes, trata-se de uma técnica de redução de dados. Com um mapa dos dois primeiros fatores é possível dar conta dos principais contrastes. Há diversos critérios para selecionar a quantidade de fatores a serem analisados, de modo equivalente ao que ocorre em análise de correspondências simples: desde analisar todas as dimensões até chegar numa proporção desejada de cobertura – o que é pouco prático, haja vista que a tabela indicadora que dá origem à ACM pode ter muitas colunas. Isso implica numa quantidade muito grande de dimensões, até selecionar as dimensões que apresentem diferenciação clara na quantidade de inércia explicada (Husson, Lê, & Pagès, 2011; Le Roux & Rouanet, 2010).

Dentro de cada fator, as modalidades de variáveis (ou indivíduos) que devem ser interpretadas são as que contribuem de modo importante para a dimensão. Isso é dado por um índice denominado contribuição para o fator ou contribuição absoluta. As somas das contribuições absolutas das modalidades das variáveis entre si somam 100%, bem como as dos indivíduos/casos. É comum calcular a contribuição absoluta média (por exemplo, se há 20 modalidades de variáveis, 100 dividido por 20 modalidades resulta em 5) e interpretar as modalidades com contribuição absoluta superiores à média (no exemplo, > 5; Le Roux & Rouanet, 2010).

Finalmente, as modalidades de variáveis mencionadas até aqui constituem variáveis ativas, cujas relações são analisadas diretamente pela ACM. No entanto, é possível acrescentar outras variáveis à análise, que são posicionadas com escores nas dimensões calculadas previamente, sem interferir nos cálculos. Essas variáveis, ditas suplementares ou ilustrativas, quando têm suas modalidades posicionadas em contrastes dos fatores, podem ajudar a interpretar os resultados. Frequentemente, variáveis sociodemográficas como sexo, faixa etária ou renda desempenham esse papel para esclarecer relações entre questões (Husson et al., 2011; Le Roux & Rouanet, 2010).

À primeira vista, a interpretação da ACM parece ser bastante intuitiva: um mapa projeta um fator no eixo horizontal e outro no eixo vertical, divididos pela origem. No presente caso, trataremos somente das relações entre modalidades de variáveis, deixando de lado os indivíduos, até mesmo porque as modalidades representam grupos de indivíduos, então concentram geralmente o interesse das pesquisas (Husson et al., 2011). Tomando por referência o eixo horizontal, as variáveis que contribuem mais que a média para o fator e estão do mesmo lado do mapa (lado positivo ou negativo) apresentam desvios importantes do perfil da amostra, e estão associadas entre si: tendem a ser escolhidas pelos mesmos indivíduos. De modo geral, quanto mais distantes da origem, isto é, mais à direita ou mais à esquerda, com escores mais extremos positivos ou negativos, maior a importância das variáveis para o fator, representando uma variação maior. Por sua vez, variáveis de lados opostos do fator se repelem, isto é, quem seleciona uma modalidade de um lado tende a selecionar a modalidade do lado oposto em proporção inferior a do perfil médio. É possível destacar associações de modo mais nítido quanto mais há desvios em relação ao perfil geral da amostra.

Uma interpretação frequente, ainda que incorreta, entende que por se apresentarem próximas num fator ou plano do mapa, duas variáveis ocorrerão majoritariamente juntas. Mas isso não é verdade: a atração de duas variáveis significa que, para as pessoas que selecionaram uma delas, a seleção de uma modalidade associada tem proporção maior que a proporção observada para a amostra global. De modo semelhante, variáveis em eixos opostos do fator indicam que, para os que selecionam uma modalidade, a proporção da outra modalidade associada é menor que a proporção da amostra global.

Como exemplo, imaginemos duas variáveis, questões num questionário, B e C, e para cada uma delas seja possível responder “concordo” ou “discordo”. A amostra geral pode ter tido o seguinte padrão de resposta para C: 20% concordaram, 80% discordaram. Se a modalidade C_concordo (resposta “concordo” para a questão C) mostra-se associada à variável

B_concordo numa ACM, isso quer dizer que, dentre os participantes que concordam com B, a proporção de concordância com C será maior que a proporção geral da amostra, de 20%. Pode ser, para fins desta ilustração, 25%. É importante perceber que a interpretação de que, estando duas modalidades associadas, quem opta por uma também opta por outra, é incorreta: no exemplo, concordar com C ainda é resposta minoritária. Portanto, duas modalidades próximas no fator não indicam que ocorrem majoritariamente em conjunto. Os resultados sempre devem ser interpretados tendo por referência tão somente o contraste com o perfil geral da amostra.

Além disso, uma grande dificuldade da ACM está na verificação das relações nos dados originais. Como a tabela que baseia a análise geralmente envolve muitas variáveis, seriam necessários cruzamentos que superam em muito as tabelas bivariadas. Se pensamos em cinco variáveis, a verificação de uma relação entre elas para interpretação envolveria uma tabela de cinco entradas, multidimensional. Torna-se difícil pensar em proporções condicionais e computar percentuais, e a relação dos resultados da ACM com os cruzamentos multivariados mostra-se obscura e pouco útil.

Conforme Bock (2017), um dos princípios-chave para a interpretação correta e útil da análise de correspondências – que podemos certamente transpor para a ACM – está na interpretação das tabelas de cruzamentos originais entre as variáveis categóricas. Desse modo, a ACM pode ser muito útil para análise exploratória ao evidenciar os principais contrastes e remeter o pesquisador para relações específicas. Porém, como mencionado, é exatamente nesse ponto que a técnica de análise de dados apresenta obstáculos para sua correta interpretação. Cabe apontar, também, que as pesquisas de psicologia que fazem uso da técnica raramente apresentam análises desse tipo de relação, o que torna seus resultados pouco transparentes.

TABELA DE PROPORÇÕES CONDICIONAIS

A visualização de cruzamentos bivariados entre variáveis nominais presentes numa ACM não reproduz exatamente as relações multivariadas evidenciadas pela técnica, mas permite uma aproximação que facilita a interpretação parcial das relações multivariadas. Propomos a construção de tabelas de proporções condicionais para apresentar os cruzamentos condicionados de variáveis. Conforme Alencar Filho (2002), uma proposição condicional é uma relação de implicação em que uma condição antecedente p determina uma condição conseqüente q . Essa relação de implicação é expressa pelo símbolo \rightarrow . Para nossos fins, uma proporção condicional é a proporção de ocorrência de uma modalidade

categorica (consequente) dada a ocorrência de outra modalidade categorica (antecedente).

A tabela de proporções condicionais consiste na apresentação das proporções condicionais de modalidades categoricas incluídas na ACM, determinadas pela condição prévia de escolha de modalidades de outras variáveis dessa análise, juntamente com as proporções dessas modalidades obtidas junto à amostra global. Desse modo, torna-se mais fácil entender a importância de cada relação, tanto em termos de ocorrência ou não de uma variação grande ou pequena relativamente à proporção de resposta global quanto em termos de força maior ou menor de associação. Este trabalho tem por objetivo apresentar a construção dessa tabela e de uma alternativa gráfica para complementar a interpretação de uma ACM realizada com dados empíricos, esclarecendo e exemplificando seu uso e demonstrando sua pertinência para a interpretação dessa análise.

MÉTODO

Os dados apresentados fazem parte de um levantamento realizado em locais públicos junto a adultos de Uberlândia/MG. Trata-se de uma pesquisa com amostragem não probabilística sobre opiniões acerca da divisão de gênero de atividades domésticas.

PARTICIPANTES

Após exclusão de dados omissos, a amostra teve um total de 958 participantes, dos quais 504 (52,6%) eram mulheres. Em termos de escolaridade, 135 (14,1%) concluíram no máximo o ensino fundamental, 587 (61,3%) terminaram o ensino médio e 236 (24,6%) o ensino superior. A maior parte dos participantes tinha de 20 a 29 anos de idade ($n = 407$ ou 42,4%), enquanto 315 (32,9%) tinham de 30 a 39 anos de idade e 236 tinham 40 a 49 anos de idade (24,6%).

INSTRUMENTO

Utilizamos um roteiro de entrevista estruturada, muito semelhante a um questionário, em que os participantes avaliaram a divisão de gênero de tarefas domésticas, os traços pertinentes para os conceitos de masculino e feminino e as práticas de lazer atribuídas a cada gênero. Para o presente trabalho, interessam-nos seis indicadores acerca da divisão de gênero de tarefas domésticas, identificados por meio dos seguintes códigos: “ae”, acompanhar atividades escolares dos filhos; “li”, limpar a casa; “la”, lavar a louça; “ie”, interferir quando filhos fazem algo errado; “cz”, cozinhar no dia a dia; e “cs”, realizar consertos em casa. Para cada uma dessas atividades, o participante deveria indicar se, de acordo

com sua opinião, cabiam sempre ao homem, mais ao homem, mais à mulher, sempre à mulher, a ambos igualmente, ou se não sabia responder. Além disso, também havia no instrumento questões sociodemográficas, como as informações referentes ao sexo, escolaridade e faixa etária dos participantes.

PROCEDIMENTO

Conforme o Art. 1 da Resolução n. 510 de 7 de abril de 2016 do Conselho Nacional de Saúde, “Não serão registradas em avaliadas pelo sistema CEP/CONEP: I. pesquisa de opinião pública com participantes não identificados”. Como este é o caso da pesquisa aqui relatada, o projeto não foi submetido a comitê de ética. Os participantes foram convidados a responder à entrevista em locais públicos de bairros de diversas regiões da cidade de Uberlândia/MG. Três autoras deste trabalho foram as entrevistadoras encarregadas da aplicação.

Para fins de análise de dados, as respostas “sempre o homem” e “mais o homem” foram fundidas na modalidade “homem” (h), assim como as duas respostas referentes à atribuição das tarefas associadas às mulheres compuseram a modalidade “mulher (m). A resposta “ambos” também foi considerada (a). As não respostas e respostas “não sei” foram consideradas como dados omissos, excluídos da análise. Os 958 casos do exemplo dizem respeito aos dados válidos.

Para apresentar a construção da tabela de proporções condicionais e sua visualização gráfica, primeiramente realizamos uma ACM dos seis indicadores, juntamente com as variáveis de sexo, escolaridade e faixa etária como suplementares. Com apenas quatro variáveis ativas, é uma análise mais simples que as realizadas habitualmente com a ACM, que tende a sintetizar um volume de dados maior. No entanto, essa análise simplificada permitirá entender a interpretação da tabela de proporções condicionais. Realizamos as análises no ambiente de programação estatística R (R Core Team, 2018), com utilização dos pacotes *FactoMineR* (Lê, Josse, & Husson, 2008) e *factoextra* (Kassambara & Mundt, 2017). Posteriormente à identificação das modalidades e variáveis relacionadas, construímos uma tabela de proporções condicionais e visualização gráfica correspondente, a serem detalhados na seção Resultados.

RESULTADOS

A ACM apresentou um total de 12 fatores. O primeiro fator explicou 25,2% da inércia, enquanto o segundo foi responsável por 10,8% e o terceiro por 10%. Ainda que a ACM

tenha as proporções de inércia dos fatores diluída devido à usual maior quantidade de variáveis incluídas do que o que ocorre com a análise binária – não é tanto o caso aqui, já que a ACM da ilustração diz respeito a uma quantidade pequena de variáveis –, é perceptível uma diferença em potencial explicativo do primeiro fator para os demais, assinalando sua maior importância. Apresentaremos os dois primeiros fatores, responsáveis, portanto, por 35,9% da inércia.

Há 12 modalidades de variáveis referentes às atividades domésticas, então a contribuição média para cada fator foi de $100 / 12 = 8,3$. A seguir indicamos as modalidades com contribuições superiores à média no primeiro fator (valor da contribuição entre parênteses): li_a (11,1), li_m (8,9), la_a (10,4), la_m (10,8), cz_a (10,7), cs_a (8,4) e ae_m (9,2). A razão de correlações (η^2) é uma medida da contribuição de cada variável ou questão (aqui, um indicador considerando suas três modalidades). Em termos desse índice, as principais variáveis do primeiro fator foram lavar a louça, limpar a casa ($\eta^2 = 0,6$), cozinhar e acompanhar atividades dos filhos ($\eta^2 = 0,5$). O segundo fator só teve duas modalidades com contribuição maior que a média, com valores elevados, ie_h (36,7) e ae_h (31,8), e as variáveis

interferir quando filhos fazem algo errado ($\eta^2 = 0,6$) e acompanhar atividades escolares dos filhos ($\eta^2 = 0,4$) tiveram razões de correlação claramente maiores que as demais.

Le Roux e Rouanet (2010) recomendam interpretar modalidades de variáveis suplementares cuja diferença de suas coordenadas (escores) nos fatores seja de pelo menos 0,5. Nenhuma das variáveis suplementares incluídas (escolaridade, faixa etária, sexo) apresentou essas diferenças entre quaisquer de suas modalidades, indicando que seu poder explicativo para dar conta dos contrastes dos dois primeiros fatores é pequeno. Contudo, incluímos a variável faixa etária no mapa da ACM e na tabela de proporções condicionadas para fins de exemplo.

A Figura 1 apresenta o mapa da ACM, incluindo somente as modalidades com contribuições maiores que a média nos dois primeiros fatores e a variável suplementar faixa etária. A dispersão concentrada no eixo horizontal indica a maior importância do primeiro fator. A ACM indica que há atração entre as respostas de tarefas referentes à atribuição relacionada às mulheres nas atividades de limpar a casa,

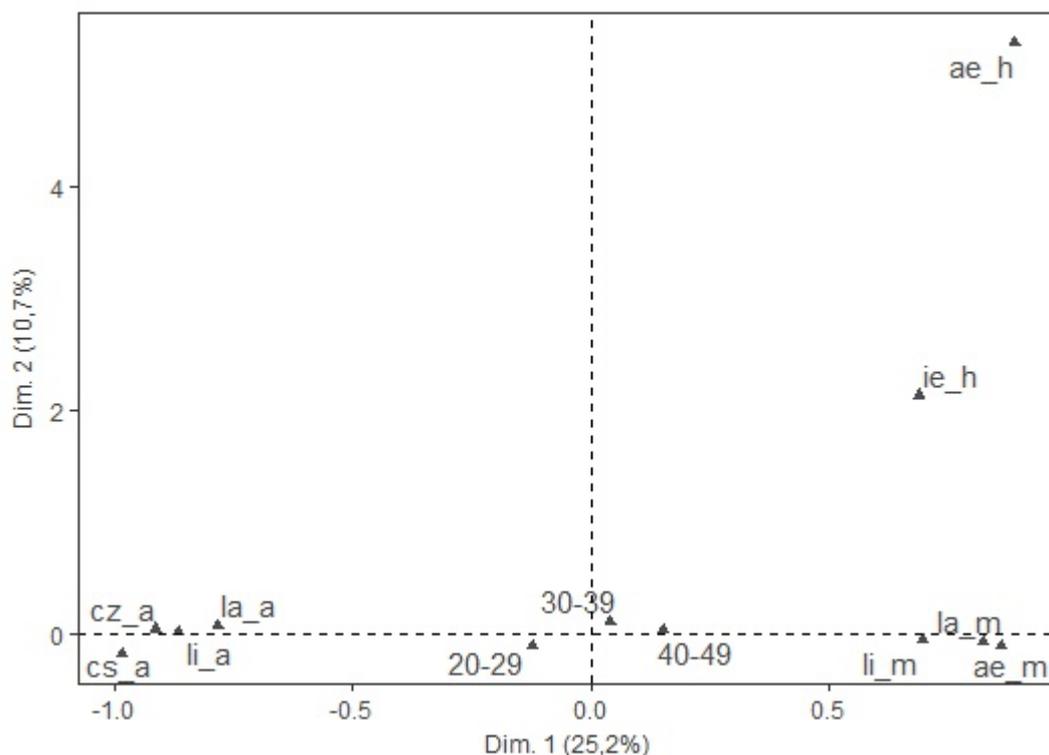


Figura 1. Mapa das duas primeiras dimensões da análise de correspondências múltiplas (ACM) referentes às respostas aos indicadores li (limpar a casa), la (lavar a louça), cs (realizar consertos em casa), cz (cozinhar no dia a dia), ae (acompanhar atividades escolares dos filhos), ie (interferir quando filhos fazem algo errado), com faixa etária como variável suplementar. Letras após os traços sublinhados indicam modalidade de resposta homem (h), ambos (a) e mulher (m).

lavar a louça e acompanhar atividades escolares dos filhos, que por sua vez repelem outro grupo de respostas relacionadas: respostas de atribuição igualitária (ambos) a cozinhar no dia a dia, lavar a louça, fazer consertos em casa e limpar a casa. No polo que atribui a responsabilidade à mulher, há associação com os participantes mais velhos da amostra (40 a 49 anos), enquanto os participantes mais jovens associam-se ao lado igualitário, e os de 30 a 39 anos estão mais próximos da origem. No segundo fator há indicação de associação mais forte entre as duas modalidades ligadas ao cuidado com os filhos atribuído ao homem.

Os resultados fazem sentido: pessoas mais velhas tendem a atribuir atividades ligadas ao cuidado da casa à mulher, e pessoas mais novas são mais igualitárias na divisão dessas atividades. Provavelmente a maior parte dos estudos que utilizam a ACM em psicologia teria feito essa interpretação. No entanto, essa descrição falha em permitir avaliar a magnitude dessas respostas, ou seja, se a associação tem importância prática.

Também é importante fazer a ressalva de que as duas modalidades ligadas ao segundo fator, apesar de sua projeção gráfica na dimensão horizontal, tiveram contribuições inferiores à média e, portanto, não foram consideradas na interpretação do primeiro. Nem sempre esse tipo de distinção é levado em conta em pesquisas que enfocam a interpretação somente do mapa gráfico da ACM.

Para qualificar melhor a importância das relações evidenciadas, inicialmente seria possível fazer um cruzamento das variáveis consideradas relacionadas entre si. Foi o que fizemos na Tabela 1, uma tabela de contingências considerando as variáveis mais importantes do primeiro fator, com todas as suas modalidades. É uma tabela de interpretação difícil: não há referências claras para trabalhar com proporções ou percentuais, e é possível observar as associações entre atribuições femininas a lavar a louça e limpar a casa, com efetivos mais altos, bem como entre as respostas igualitárias, também mais numerosas e concentradas no centro da tabela. As repulsões também se refletem em baixas frequências. No entanto, a tabela é pouco informativa, além de fornecer essas impressões gerais, e se mostra pouco útil para interpretar ou esclarecer a ACM. Além disso, cabe lembrar que traz apenas quatro variáveis das seis incluídas na análise, o que indica que não cobre todos os resultados. Uma ACM baseada em uma matriz maior, com mais variáveis, como é o caso da maior parte das aplicações da técnica, seria pouco viável para a representação em tabela de contingências.

A proposta da tabela de proporções condicionadas é apresentada na Tabela 2. Com exceção da segunda coluna,

que apresenta as frequências de cada resposta ou modalidade (no caso de faixa etária), indicada por n , todas as demais informações dizem respeito a proporções expressas em percentuais. As proporções em negrito, na diagonal, são as proporções de cada resposta na amostra global de 958 participantes. Na primeira coluna à esquerda, identificada pelo símbolo de implicação (\rightarrow), há a modalidade condicionante. Todas as proporções à direita de cada modalidade condicionante implicam a proporção de respostas à modalidade indicada em cada coluna, dada a modalidade condicionante.

Tabela 1. Tabela de contingências cruzando as frequências de respostas aos indicadores li (limpar a casa), la (lavar a louça), cz (cozinhar no dia a dia) e cs (realizar consertos em casa)*.

		li				TOTAL
la	cz	cs	h	a	m	
h	h	h	0	0	0	0
		a	0	0	0	0
		m	0	0	0	0
a	h	h	0	2	2	4
		a	0	0	0	0
		m	0	0	0	0
m	h	h	0	1	8	9
		a	0	0	0	0
		m	0	0	0	0
a	h	h	1	3	2	6
		a	0	2	0	0
		m	0	0	0	0
a	h	h	0	129	30	159
		a	0	151	9	160
		m	0	0	1	1
m	h	h	0	70	48	118
		a	0	19	18	37
		m	0	6	1	7
m	h	h	0	0	2	2
		a	0	1	0	1
		m	0	0	2	2
a	h	h	1	2	34	37
		a	0	2	4	6
		m	0	0	2	2
m	h	h	0	21	303	324
		a	1	8	35	44
		m	0	10	27	37
TOTAL			3	427	528	958

* Nota: h: atribuição ao homem; a: atribuição a ambos; m: atribuição à mulher.

Por exemplo, 38,5% dos participantes escolheram a modalidade cz_a, de que cabe a ambos, homens e mulheres, cozinhar no dia a dia. O mapa da Tabela 1 aponta forte associação entre cz_a e cs_a, indicando também que cabe a ambos fazer consertos em casa. Na Tabela 2, observa-se que, dentre as pessoas que consideram que as ambos devem

Tabela 2. Tabela de proporções condicionais envolvendo respostas de 958 participantes aos indicadores: *li* (limpar a casa), *la* (lavar a louça), *cs* (realizar consertos em casa), *cz* (cozinhar no dia a dia), *ae* (acompanhar atividades escolares dos filhos), *ie* (interferir quando filhos fazem algo errado) e faixa etária dos participantes*.

→	n	ae_h	ae_a	ae_m	ie_h	ie_a	ie_m	cz_h	cz_a	cz_m	cs_h	cs_a	cs_m
ae_h	14	1,4	-	-	71,4	7,1	21,4	7,1	14,3	78,6	85,7	7,1	7,1
ae_a	586	-	61,1	-	4,8	91,1	4,1	0,9	50,7	48,5	59,6	37,7	2,7
ae_m	358	-	-	37,4	17	35,5	47,5	2	19,6	78,5	83,2	7,8	8,9
ie_h	99	10,1	28,3	61,6	10,3	-	-	30,3	26,2	70,7	91,9	7,1	1
ie_a	662	0,2	80,7	19,2	-	69,1	-	0,9	47	52,1	63,1	34	2,9
ie_m	197	1,5	12,2	86,2	-	-	20,6	2	16,2	81,7	76,1	9,1	14,7
cz_h	13	7,7	38,5	53,8	23,1	46,2	30,8	1,4	-	-	61,5	23,1	15,4
cz_a	369	0,5	80,5	19	7	84,3	8,7	-	38,5	-	54,2	45	0,8
cz_m	576	1,9	49,3	48,8	12,2	59,9	28	-	-	60,1	78,3	14	7,6
cs_h	659	1,8	53	45,5	13,8	63,4	22,8	1,2	30,3	68,4	68,8	-	-
cs_a	250	0,4	88,4	11,2	2,8	90	7,2	1,2	66,4	32,4	-	26,1	-
cs_m	49	2	32,7	65,3	2	38,8	59,2	4,1	6,1	89,8	-	-	5,1
20-29a	407	0,7	61,4	37,8	5,9	74	20,1	1,2	45,9	52,8	68,5	26,7	4,7
30-39a	315	2,2	65,4	32,4	13,7	69,2	17,1	0,6	34,9	64,4	70,8	25,4	3,8
40-49a	236	1,7	55,1	43,2	13,6	60,6	25,8	2,5	30,5	66,9	66,5	25,8	7,7

*Nota: h: atribuição ao homem.; a: atribuição a ambos; m: atribuição à mulher.

fazer consertos em casa (isto é, *cs_a* como modalidade condicionante), que são em número de 250, mais de um quarto da amostra, quase dois terços (ou 66,4%) também julgam que ambos devem cozinhar. É uma proporção bem maior que os 38,5% globais e é isso indica essa proximidade no fator e no mapa.

Do mesmo modo, *cz_a* encontra-se oposto no fator à resposta de que as mulheres são as responsáveis por acompanhar as atividades escolares dos filhos (*ae_m*). Com esta última modalidade como condicionante, a proporção de *cz_a* é de 19,6%, próxima da metade da proporção global. Há, portanto, repulsão entre as duas modalidades. É possível ler a relação, evidentemente, também no outro sentido, com *cz_a* condicionando a resposta a *ae_m*; nesse caso, a proporção dos participantes que respondem *cz_a* que também respondem *ae_m* é de 19%, inferior à proporção global de 37,4%.

Por sua vez, as variáveis suplementares, próximas à

origem, apontam baixa importância no entendimento das diferenças indicadas no mapa. Basta observar que, para a modalidade *ae_h* (os homens são responsáveis por acompanhar atividades escolares), a amostra global teve proporção de 1,4%, e a variável faixa etária pouco explicou as variações: a proporção foi de 0,7% dentre os participantes mais jovens, a 2,2% para os participantes com 30 a 39 anos. A baixa importância da variável suplementar em questão é facilmente verificada com a tabela de proporções condicionadas.

De modo alternativo, é possível apresentar as informações da tabela de proporções condicionadas graficamente. A Figura 2 apresenta as proporções de respostas aos indicadores *ae* e *ie*, com as modalidades condicionantes no eixo das abscissas, e a proporção geral na primeira barra à esquerda. É fácil visualizar a atração entre as respostas de atribuição masculina para interferência nos erros dos filhos e o acompanhamento de atividades

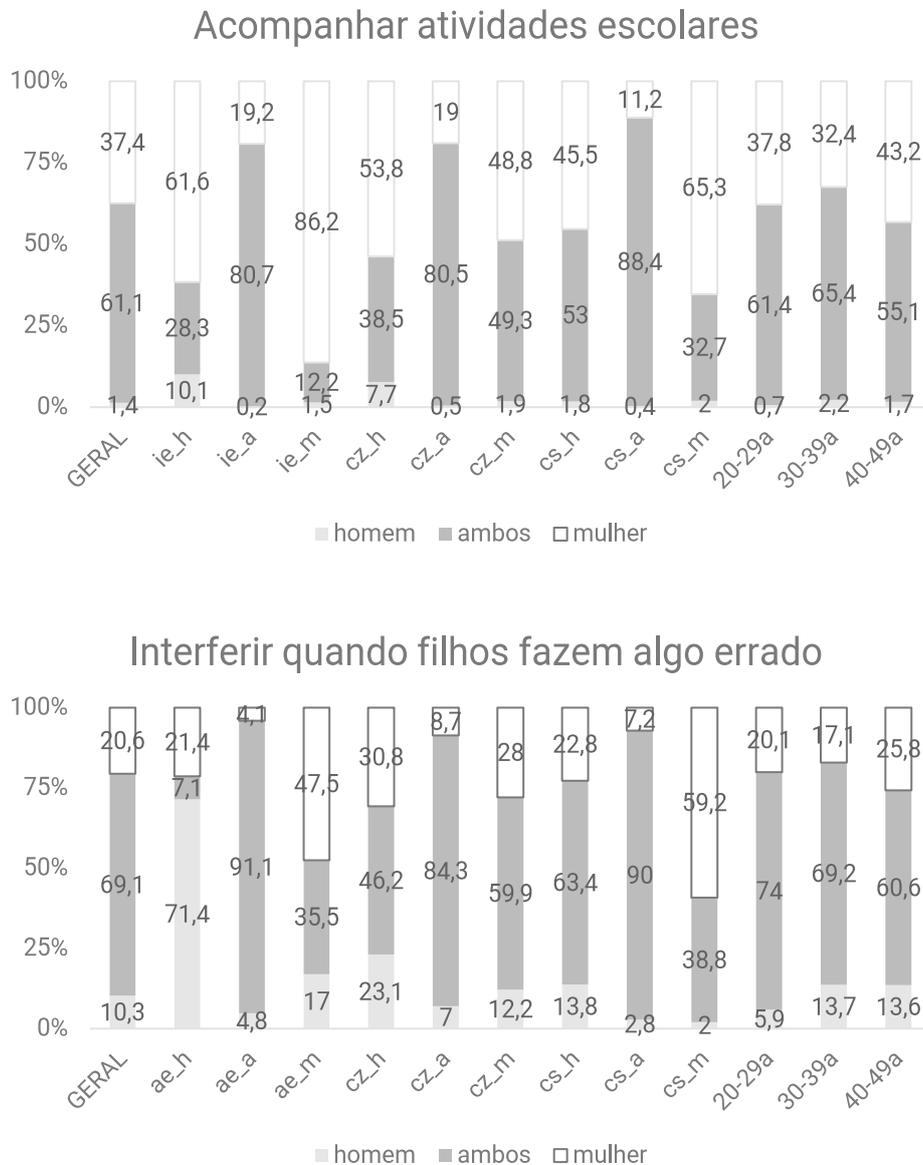


Figura 2: visualização gráfica da tabela de proporções condicionadas com gráficos de barras referentes às respostas aos indicadores ae (acompanhar atividades escolares dos filhos) e ie (interferir quando filhos fazem algo errado). Letras após os traços sublinhados indicam modalidade de resposta homem (h), ambos (a) e mulher (m).

escolares. Para os que consideram que os homens é que devem acompanhar essas atividades, 71,4% também pensam que são eles que interferem quando os filhos se comportam mal, frequência bem maior que os 10,3% da amostra global.

DISCUSSÃO

Face à constatação da utilidade da ACM para pesquisas

em psicologia, este trabalho indicou algumas dificuldades para que essa técnica seja corretamente aproveitada em termos de seu potencial informativo e para garantir a inteligibilidade dos resultados a ela associados. Para atenuar esses obstáculos, propusemos uma técnica auxiliar de apresentação dos resultados: a tabela de proporções condicionadas, que ao proporcionar diversos pontos de vista sobre algumas relações evidenciadas pela ACM, permite uma apreciação mais precisa das associações envolvidas e de

sua importância.

A construção de tabelas de proporções condicionais é simples, podendo ser realizada com cruzamentos de dados simples gerados por qualquer programa ou ambiente de estatística. Do mesmo modo, os gráficos podem ser gerados a partir de planilhas padrão, como Calc ou Excel. A ACM também é realizada por diversos programas comerciais ou de livre acesso. Particularmente, o ambiente estatístico R, gratuito, contém vários pacotes com alternativas de implementação, como *FactoMineR*, utilizado aqui, *ade4* (Chessel, Dufour, & Thioulouse, 2004) e *ca* (Nenadic & Greenacre, 2007). Portanto, a técnica de análise multivariada, a tabela de proporções condicionais e a opção de visualização gráfica com diagramas de barras são plenamente acessíveis para pesquisadores.

A ACM é um pouco mais complexa que a antevisão de seu mapa faz parecer intuitivamente. No entanto, é necessário aproximá-la de leitores instruídos, enfatizar ou tornar claras as informações pertinentes, para que seu potencial como técnica exploratória, geradora de ideias e hipóteses, seja aproveitado corretamente. Em psicologia, ocasionalmente encontram-se esforços para capacitar pesquisadores e estudantes em técnicas de análise de dados sem conhecimento de suas bases matemáticas e estatísticas, por meio de tutoriais ou instruções de uso de programas de computador e diretrizes para interpretação. Em que pesem as maiores facilidades e acessibilidade de técnicas complexas a partir da difusão de programas de estatística, alguns bastante intuitivos, a utilização adequada – seja para aplicação ou interpretação – dessas ferramentas demanda conhecimento ao menos básico da estatística, para não incorrer em erros que prejudicam a ciência e a população, na medida em que podem inspirar decisões equivocadas. Em outras palavras, para fazer uma analogia, a utilização de técnicas de análise de dados quantitativos sem o conhecimento necessário de estatística é tão problemática quanto o emprego de técnicas de análise do discurso, na pesquisa qualitativa, desconhecendo minimamente os conceitos da linguística e semiótica.

Se, conforme metáfora de Tukey (1977), a análise exploratória de dados é trabalho de detetive, acreditamos que a tabela de proporções condicionais, técnica complementar de visualização de relações bivariadas dentro do contexto multivariado, contribui para tornar mais preciso o entendimento da análise multivariada, dando confiabilidade às “pistas” e reduzindo a ambiguidade da visualização atraente da ACM.

Cabe ressaltar que, estritamente, a tabela de proporções condicionais é independente da ACM. A verificação de fato dos cruzamentos originais da ACM ocorreria por meio de

tabelas de contingência multivariadas, e sua alta dimensionalidade dificulta qualquer esforço interpretativo. No entanto, a ACM fornece indicações de quais variáveis estão associadas, e a tabela de proporções condicionais apresenta os cruzamentos bivariados, facilmente interpretáveis, de algumas dessas variáveis. Se, por um lado, não há índices ou parâmetros da ACM que se relacionam diretamente com os resultados da tabela de proporções condicionais, por outro lado os cruzamentos desta não refletem estimativas de qualquer tipo, mas ligações observadas reais, que permitem ao pesquisador realizar afirmações seguras e aproximar suas afirmações de seu público leitor, exigindo não mais que a capacidade de entender percentuais majoritários e minoritários. Com o auxílio dessa técnica simples de visualização de cruzamentos orientados, o leitor aproxima-se corretamente do entendimento multivariado e a ACM deixa de ser um recurso esotérico para direcionar de modo eficiente a ênfase nos dados.

Finalmente, em contraste com o que apresentamos no exemplo empírico deste texto, em vez de construir uma tabela de proporções condicionais parcial apenas de variáveis indicadas pela ACM, é certamente possível construir uma tabela de proporções condicionais de todas as variáveis ativas e ilustrativas envolvidas numa análise. A aplicação de rotinas de programação facilitaria essa tarefa, capaz de conter toda a informação desse tipo dos dados. No entanto, esse ganho provavelmente não compensaria a maior simplicidade e legibilidade de uma tabela menor e mais focada, que conduz o leitor ao entendimento das principais relações dos dados.

DECLARAÇÃO DE FINANCIAMENTO

A pesquisa relatada no manuscrito foi financiada parcialmente pela bolsa de mestrado da terceira autora (CAPES)

DECLARAÇÃO DA CONTRIBUIÇÃO DOS AUTORES

JW, RSC, JBBP e PMD elaboraram o projeto que deu origem aos dados empíricos, incluindo aspectos teóricos e construção dos instrumentos; JW realizou a análise formal dos dados, construção dos dispositivos de visualização, conceitualização principal e redação inicial do manuscrito; RSC, JBBP e PMD conduziram empiricamente a investigação em termos de coleta de dados e tabulação, contribuíram com a redação final e revisão do manuscrito.

DECLARAÇÃO DE CONFLITOS DE INTERESSE

Os autores declaram que não há conflitos de interesse no manuscrito submetido.

REFERÊNCIAS

- Alencar Filho, E. (2002). *Iniciação à lógica matemática*. São Paulo: Nobel.
- Baccini, A. (2010). *Statistique descriptive multidimensionnelle (pour les nuls)*. Toulouse: Institut de Mathématiques de Toulouse / Université Paul Sabatier. Recuperado de <https://www.math.univ-toulouse.fr/~baccini/zpedago/asdm.pdf>. Acesso em 28 de agosto de 2018.
- Bauer, M. W., Gaskell, G., & Allum, N. C. (2003). Qualidade, quantidade e interesses do conhecimento – evitando confusões. In M. W. Bauer, G. Gaskell (Eds.), *Pesquisa qualitativa com texto, imagem e som* (Pedrinho A. Guareschi, Trans., 2a ed., pp. 17-36). Petrópolis: Vozes.
- Bock, T. (2017). *How to interpret correspondence analysis plots (it probably isn't the way you think)* [Postagem de blog]. Recuperado de <https://www.displayr.com/interpret-correspondence-analysis-plots-probably-isnt-way-think/>. Acesso em 29 de agosto de 2018.
- Bourdieu, P. (1979). *La distinction. Critique sociale du jugement*. Paris: Les Éditions de Minuit.
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 package – I: one-table methods. *R News*, 4(1), 5-10.
- Doise, W., Clemence, A., & Lorenzi-Cioldi, F. (1992). *Représentations sociales et analyse des données*. Grenoble: PUG.
- Greenacre, M. (2007). *Correspondence analysis in practice* (2a ed.). Boca Raton: Chapman & Hall / CRC. doi: 10.1201/9781420011234
- Günther, H. (2006). Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão? *Psicologia Teoria e Pesquisa*, 22(2), 201-210. doi:10.1590/s0102-37722006000200010
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Upper Saddle River: Prentice Hall.
- Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory multivariate analysis by example using R*. Londres: Chapman & Hall/CRC. <https://doi.org/10.1201/b21874>
- Kassambara, A., & Mundt, F. (2017). *factoextra: Extract and visualize the results of multivariate data analysis. R package version 1.0.5*. Recuperado de <https://CRAN.R-project.org/package=factoextra>. Acesso em 29 de agosto de 2018.
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1-18. doi:10.18637/jss.v025.i01
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Le Roux, B., & Rouanet, H. (2010). *Multiple correspondence analysis*. Thousand Oaks: Sage. doi: 10.4135/9781412993906
- Lima, M. E. O., & Vala, J. (2005). A cor do sucesso: efeitos da performance social e econômica no branqueamento e na infra-humanização dos negros no Brasil. *Psicologia USP*, 16(3), 143-155. doi:10.1590/S0103-65642005000200008
- Nenadic, O., & Greenacre, M. (2007). Correspondence analysis in R with two-and three-dimensional graphics: the ca package. *Journal of Statistical Software*, 20(3), 1-13. doi: 10.18637/jss.v020.i03
- Neves, P. S. C., & Lima, M. E. O. (2007). Percepções de justiça social e atitudes de estudantes pré-vestibulandos e universitários sobre as cotas para negros e pardos nas universidades públicas. *Revista Brasileira de Educação*, 12(34), 17-38. doi:10.1590/S1413-24782007000100003
- Ragin, C. C., & Amoroso, L. M. (2011). *Constructing social research. The unity and diversity of method* (2a ed.). Thousand Oaks: Sage/Pine Forge.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.
- Wachelke, J. (2017). Ideologia nas opiniões de estudantes de ensino médio sobre sucesso no trabalho. *Psicologia: Ciência e Profissão*, 37(3), 652-668. doi:10.1590/1982-3703004062016
- Wachelke, J., Matos, F. R., Ferreira, G. C. S., & Costa, R. R. L. (2015). Um panorama da literatura relacionada às representações sociais publicada em periódicos científicos. *Temas em Psicologia*, 23(2), 309-325. doi: 10.9788/TP2015.2-06-Pt

Data de submissão: 03/09/2018
Primeira decisão editorial: 21/11/2018
Aceite em 03/12/2018