

Variáveis Psicológicas e Desempenho Acadêmico: Uma Análise Da Existência de Correlação Canônica

Psychological Variables and Academic Performance: an Analysis of Canonical Correlation Existence

¹Anderson Cristiano Neisse; ²Kuang Hongyu

¹Graduando em Estatística pela Universidade Federal de Mato Grosso

²Professor Doutor; Departamento de Estatística – Universidade Federal de Mato Grosso

RESUMO: A análise multivariada é um conjunto de técnicas estatísticas que permite a análise e interpretação de conjuntos de dados de natureza quantitativa com grande número de variáveis de forma simplificada. Dentre o rol de técnicas encontra-se a análise de correlação canônica (ACC) que visa estabelecer estrutura de relação entre dois grupos de variáveis por meio de combinações lineares que maximizem a correlação entre ambos. O objetivo deste artigo foi utilizar a técnica de ACC para analisar a estrutura de correlação entre três variáveis psicológicas e quatro variáveis de desempenho acadêmico. As variáveis psicológicas são locus de controle, autoconceito e motivação enquanto que as variáveis acadêmicas são leitura, escrita, matemática e ciências. A análise de correlação canônica se apresentou efetiva no estabelecimento de relação entre as variáveis psicológicas e de desempenho acadêmico apresentadas nos dados obtidos. Uma das três funções canônicas ajustadas apresentou correlação canônica significativa de 44,64% e explicou 19,93% da variabilidade conjunta dos dados. Apesar da efetividade, outras variáveis psicológicas podem ser medidas para, possivelmente, explicar proporção maior da variabilidade do desempenho acadêmico.

Palavras-chave: Análise multivariada. Análise de correlação canônica. Psicologia. Desempenho acadêmico.

ABSTRACT: Multivariate analysis is a set of statistical techniques that allows the analysis and interpretation of quantitative data sets with large variable sets in a simplified form. Among the list of techniques, there is the canonical-correlation analysis (CCA), technique that aims to establish relation structure between two sets of variables by means of linear combinations that maximizes the correlation in between both sets. This article's objective was to apply the CCA method to analyze the correlation structure in between three psychological variables and four academic performance variables. The psychological variables are locus of control, self-concept and motivation while the academic performance variables are reading, writing, math and science. The canonical-correlation analysis showed to be an effective method in establishing relation between psychological and academic performance variables as presented in the data. One out of three adjusted canonical functions turned out to be significant, reaching 44,64% canonical correlation and explaining 19,93% of the shared variability. Although effective, more psychological variables could be measured for the canonical correlation to explain greater proportions of shared variability.

Keywords: Multivariate analysis. Canonical-correlation analysis. Psychology. Academic performance.

INTRODUÇÃO

A análise multivariada é um conjunto de técnicas estatísticas que permite a análise e interpretação de conjuntos de dados de natureza quantitativa com grande número de variáveis de forma simplificada. O advento das técnicas multivariadas se deu pelo avanço tecnológico e o uso de computadores para análise estatística, sem os quais não seria possível analisar grande número de variáveis de forma conjunta (XING *et al.*, 2016).

Por conta da facilidade de análise e interpretação de grande número de variáveis, as técnicas multivariadas têm se tornado muito populares em várias áreas do conhecimento apesar de terem sido desenvolvidas para problemas específicos, dentre as áreas de grande aplicação atualmente estão: agronomia, zootecnia, ecologia, florestal, psicologia, etc.

Dentre as várias possíveis aplicações das técnicas multivariadas está o estudo da relação de fatores psicológicos com o desempenho em diferentes tarefas de cunho acadêmico. A Análise de Correlação Canônica (ACC) é uma técnica multivariada de modelagem e estabelecimento de estrutura da correlação entre dois grupos de variáveis por meio da transformação linear em pares de variáveis canônicas de forma que a correlação entre essas combinações seja máxima (TRUGILHO *et al.*, 2003).

A ACC está relacionada à redução de massa de dados através da combinação de variáveis de forma a maximizar a correlação e facilitar a interpretação. Apesar de as técnicas multivariadas terem sido desenvolvidas para soluções de problemas muito específicos, a possibilidade da redução de massa de dados tornou tais técnicas populares em várias áreas do conhecimento (ROCHA & PELOGIO, 2014; KAN *et al.*, 2016).

O presente estudo apresentou a técnica multivariada de ACC como instrumento de redução de massa de dados que possibilitam a análise de relação entre conjuntos com grande quantidade de variáveis. As análises foram feitas considerando as variáveis psicológicas *locus* de controle (*control*), autoconceito (*concept*) e motivação (*motivation*) de 600 estudantes e seus respectivos desempenhos acadêmicos em escrita (*writing*), leitura (*reading*), matemática (*math*) e ciência (*science*). Locus de controle é definido como a percepção que um indivíduo tem sobre o quanto os eventos de sua vida tem influência própria ou externa, se um indivíduo possui *locus* de controle interno, acredita ter maior influência sobre os desfechos de sua vida. Autoconceito é definido como um conjunto de crenças que um indivíduo tem sobre suas mais variadas capacidades, quanto maior o autoconceito de um indivíduo mais fortes são suas crenças nas próprias habilidades. Já a motivação pode ser definida como a condição de um indivíduo que o impulsiona a realizar alguma tarefa.

O objetivo deste trabalho foi utilizar a ACC para analisar a correlação de variáveis psicológicas com o desempenho acadêmico. A proposta é, por meio da ACC, detalhar a variação conjunta das variáveis psicológicas e acadêmicas de forma a possibilitar interpretação com base em poucas variáveis canônicas.

MATERIAL E MÉTODOS

Banco de Dados

O conjunto de dados utilizado para as análises apresentadas neste trabalho são de domínio público, disponibilizados pela UCLA: *Institute for Digital Research and Education*, que incentiva pesquisas baseadas em análise de dados utilizando os mais variados softwares estatísticos. Tais dados são uma amostra de 600 estudantes de um estudo longitudinal realizado nos EUA pelo projeto *High School and Beyond*. O estudo original obteve informações psicológicas e de desempenho acadêmico de 58.270 estudantes do ensino médio. O intuito da pesquisa é evidenciar como variáveis de natureza psicológica estão relacionadas ao desempenho acadêmico de estudantes. A amostra é composta por 273 (45,5%) de estudantes do sexo masculino e 327 (54,5%) estudantes do sexo feminino.

As três variáveis psicológicas são: *locus* de controle (*control*); autoconceito (*concept*) e motivação (*motivation*). As variáveis psicológicas são quantitativas, baseadas em testes padronizados. A variável *locus* de controle apresenta valores abaixo de zero para indivíduos com *locus* de controle interno e acima de zero caso contrário, sendo que aqueles com grau mais elevado terão valor mais distantes de zero.

Analogamente, variável autoconceito apresenta valores acima de zero para autoconceito avaliado como positivo, e valores abaixo de zero caso contrário. Os valores se distanciam de zero conforme o grau de autoconceito aumenta. Já a variável de motivação apresenta valores de 0 a 1 de acordo com o grau avaliado de motivação do indivíduo.

As variáveis de desempenho acadêmico são testes padronizados de: leitura (*reading*); escrita (*writing*); matemática (*math*); e ciências (*science*). Estas variáveis foram utilizadas para demonstrar a aplicação da técnica multivariada ACC para relacionar características psicológicas com o desempenho acadêmico. Toda a análise foi realizada por meio do software R 3.0.1 (R DEVELOPMENT CORE TEAM, 2014).

Análise de Correlação Canônica

A análise de correlação canônica pode ser considerada uma generalização de uma regressão linear múltipla para mais de uma variável dependente, apesar de a ACC não diferenciar variáveis dependentes de independentes simplesmente analisando dois grupos de variáveis. A ACC resume as informações contidas em ambos os conjuntos de variáveis em duas combinações lineares que maximizem a correlação entre os dois grupos, chamadas de variáveis estatísticas canônicas (MINGOTI, 2005; ROCHA & PELOGIO, 2014). As correlações canônicas em si são os coeficientes de correlação entre ambas variáveis estatísticas canônicas.

Dados dois conjuntos de dados com p e q observações respectivamente, a quantidade de correlações canônicas que podem ser obtidas é de $\min(p, q)$ sendo que a sua medida diminui conforme a ordem de estimação. O primeiro coeficiente de correlação canônica é sempre maior ou igual, em módulo, a qualquer valor de correlação simples que possa ser obtida entre os dois grupos originais de variáveis (PROTÁSIO *et al.*, 2012; ROCHA & PELOGIO, 2014).

A multicolinearidade pode aumentar estimativas de variância de coeficientes de forma que nenhuma variável se apresente significativa mesmo com alto R^2 ou produzir estimativas de coeficientes de magnitude errônea (KAN *et al.*, 2016). Para verificar a existência de multicolinearidade em um conjunto de variáveis é utilizado o Fator de Inflação da Variância (VIF) que, para a i -ésima variável, é definido por $1/(1-R^2)$. É comum o VIF que assuma valor acima de 10, ser considerado indicador de existência de multicolinearidade excessiva (O'BRIEN, 2007). Portanto, é indicado realizar verificação de presença de multicolinearidade antes da aplicação de ACC aos dados.

Sejam $\mathbf{X}' = [x_1, x_2, \dots, x_p]$ e $\mathbf{Y}' = [y_1, y_2, \dots, y_q]$ os vetores de características dos grupos de variáveis I e II respectivamente, o problema está em obter estimações dos coeficientes de combinações lineares de ambos os grupos de forma a estimar a correlação máxima entre as combinações. Considerando X_1 e Y_1 as primeiras combinações dos grupos de variáveis I e II respectivamente, tem-se que

$$X_1 = a_1x_1 + a_2x_2 + \dots + a_px_p$$

e

$$Y_1 = b_1y_1 + b_2y_2 + \dots + b_qy_q$$

em que $\mathbf{a}' = [a_1, a_2, \dots, a_p]$ é o vetor $1 \times p$ de pesos das características do grupo I e $\mathbf{b}' = [b_1, b_2, \dots, b_q]$ é o vetor $1 \times q$ de pesos das características do grupo II (JOHNSON & WICHERN, 2007; KAN *et al.*, 2016).

Sendo \mathbf{R}_{11} a matriz de correlação $p \times p$ das variáveis do grupo I, \mathbf{R}_{22} a matriz de correlação $q \times q$ das variáveis do grupo II e \mathbf{R}_{12} a matriz de correlação $p \times q$ entre os grupos de variáveis I e II, define-se (JOHNSON & WICHERN, 2007):

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}'_{12} & \mathbf{R}_{22} \end{bmatrix}$$

sendo a matriz de correlação da união dos grupos de variáveis I e II. Desta forma, a estimação dos vetores \mathbf{a} e \mathbf{b} é realizada por meio da maximização da função r^2 que fica sujeita à restrição $\mathbf{a}'\mathbf{R}_{11}\mathbf{a} = \mathbf{b}'\mathbf{R}_{22}\mathbf{b}$ de forma a possibilitar a obtenção de estimadores únicos de \mathbf{a} e \mathbf{b} , também significando que a variância de cada combinação linear é igual a 1 (JOHNSON & WICHERN, 2007; KAN *et al.*, 2016).

A obtenção dos autovalores é o primeiro passo na ACC por meio das equações características:

$$|\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}'_{12} - \lambda\mathbf{I}| = 0$$

e

$$|\mathbf{R}_{22}^{-1}\mathbf{R}'_{12}\mathbf{R}_{11}^{-1}\mathbf{R}_{12} - \lambda\mathbf{I}| = 0$$

seguida do cálculo dos seus respectivos autovetores associados. Desta forma obtém-se a primeira correlação canônica r_1 entre as combinações lineares dos grupos de variáveis I e II obtida por:

$$r_1 = \sqrt{\lambda_1},$$

em que λ_1 é o maior autovalor da matriz $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}'_{12}$, que em geral, não-simétrica de ordem p (JONSON & WICHERN, 2007).

O mais importante fator canônico é obtido por $X_1 = \mathbf{a}'\mathbf{X}$ e $Y_1 = \mathbf{b}'\mathbf{Y}$, em que \mathbf{a} é o autovetor associado ao primeiro autovalor $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}'_{12}$ e \mathbf{b} é o autovetor associado ao primeiro autovalor de $\mathbf{R}_{22}^{-1}\mathbf{R}'_{12}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$. As demais correlações e fatores canônicos são obtidas com base nos autovalores e autovetores das expressões conforme a ordem decrescente (TRUGILHO *et al* 2003; XING *et al.*, 2016).

Os autovalores λ_i são as correlações canônicas das respectivas funções canônicas ajustadas aos conjuntos de dados originalmente observados. Os autovetores \mathbf{a} e \mathbf{b} são chamados de pesos canônicos ou coeficientes canônicos, provenientes das duas funções lineares que maximizaram a correlação entre ambos os conjuntos de dados. Os pesos canônicos são utilizados para obtenção dos escores canônicos, que são os valores resultantes da aplicação das cargas às variáveis observadas originalmente.

Ao se obter as funções canônicas, a primeira preocupação é avaliar sua significância através de testes multivariados. Existem vários testes passíveis de aplicação ao modelo completo, ou seja, eles avaliam as variâncias compartilhadas entre variável dependente e independente em todas as funções canônicas. Tais testes podem ter suas estatísticas convertidas à estatística mais familiar F , que pode ser utilizada para avaliar a significância do modelo. O fato de cada teste ser baseado em diferentes estruturas teóricas resulta na possibilidade de diferentes resultados. O método de Lambda de Wilks (λ) é o mais comumente utilizado pelo fato de possuir aplicação mais geral (SHERRY & HENSON 2005).

As cargas canônicas (*canonical loadings*) ou correlações estruturais se tratam das correlações entre as variáveis observadas originalmente e os respectivos escores da variável canônica sintetizada na ACC. Como estes coeficientes se tratam de simples correlações de Pearson devem variar no intervalo fechado de -1 a +1.

Os *loadings* fornecem interpretação pela definição da estrutura da variável canônica, isto é, evidenciando quais variáveis observadas possuem importância na obtenção da variável canônica e, por isso, podem ser úteis no modelo em si. Já os *loadings* quadrados, que são os *loadings* elevados ao quadrado como o nome sugere, indicam a proporção de variação que uma variável observada original compartilha linearmente com a variável canônica do respectivo grupo (COURVILLE & THOMPSON 2001; SHERRY & HENSON 2005).

Uma vez que os coeficientes das funções canônicas e cargas canônicas (*loadings*) são voltados para a maximização da correlação entre os grupos de variáveis, não são ideais para a interpretação das relações entre os grupos. Desta forma com intuito de melhor interpretar as funções canônicas obtidas através da análise de correlação canônica, as cargas canônicas cruzadas (*cross-loadings*) são mais utilizadas. Tratam-se das correlações entre as variáveis originais e a variável canônica do grupo oposto (HAIR JR *et al.*, 2007).

O índice de redundância foi apresentado por Stewart & Love (1968) como adição à ACC, que é nada mais do que a quantidade média de variância das variáveis de um conjunto que é explicada por uma variável canônica do outro conjunto, representada pela notação:

$$\mathcal{R}_{y_i} = \mu_i^2 \frac{1}{m_y} f'_{y\hat{y}_i} f_{y\hat{y}_i} \quad (1)$$

e

$$\mathcal{R}_y = \sum_i \mathcal{R}_{y_i} \quad (2)$$

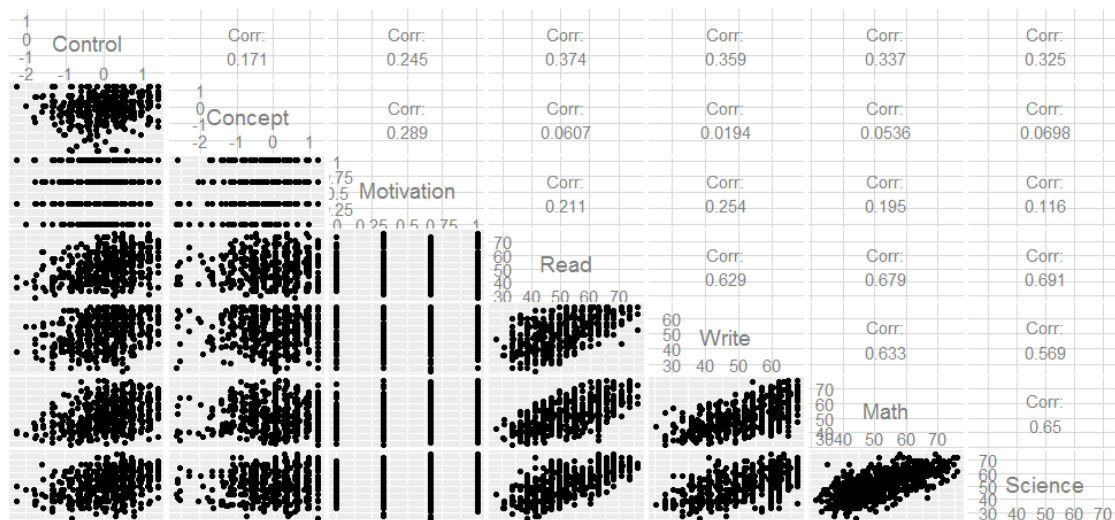
em que (1) é a redundância da i-ésima variável canônica do grupo de variáveis independentes e (2) é a redundância geral. O símbolo m_y diz respeito ao número de variáveis, μ_i é a i-ésima correlação canônica e $f'_{y\hat{y}_i} = v' R_{yy}$ é o vetor de *loadings* das variáveis y para a i-ésima função canônica (VAN DEN WOLLENBERG 1977).

RESULTADOS E DISCUSSÕES

Na Figura 1, foi apresentada a matriz de gráficos de dispersão e respectivas correlações entre variáveis. A parte inferior da matriz de correlações, que é simétrica, foi preenchida com os gráficos de dispersão de forma a expressar visualmente as correlações apresentadas. A análise foi realizada tanto dentro de cada grupo como entre os grupos, o que possibilita a verificação de indícios da existência de multicolinearidade nos dados.

Dentro do grupo de variáveis psicológicas a maior correlação observada foi de **28,9%** entre as variáveis *concept* e *motivation*, considerada fraca. Já dentre as variáveis do grupo acadêmico as correlações são consideradas moderadas, a maior foi de **69,1%** entre as variáveis *science* e *reading* com as demais correlações estando acima de **56,9%**. Em análise de correlação simples entre os dois conjuntos a maior foi de **37,4%** entre as variáveis *control* e *reading* sendo que as demais estão acima de **5,36%** e são todas consideradas correlações fracas (Figura 1).

Figura 1: Matriz de gráficos de dispersão e correlações entre as variáveis psicológicas e acadêmicas.



Fonte: Autoria Própria.

Adicionalmente ao cálculo das correlações lineares simples foram calculados os VIFs para todas as variáveis com o objetivo de diagnosticar a possível existência de multicolinearidade. O maior valor VIF encontrado foi de **2,54** para a variável *reading* que por ser menor do que **10** indica a não existência de multicolinearidade significativa nos dados psicológicos e acadêmicos.

Na Tabela 1, apresentou as correlações canônicas obtidas, o R^2 canônico e o teste de significância Lambda de Wilks para cada uma das funções. Para decidir qual função canônica analisar são necessários três critérios. De acordo com Hair Jr. et al. (2007), os três critérios são (1º) nível de significância estatística da função, (2º) magnitude da correlação canônica e (3º) medida de redundância para o percentual de variância explicada a partir dos dois conjuntos de dados. Foi possível observar que as funções canônicas 1 e 2 foram consideradas significativas ao nível de 5%, possibilitando o descarte da função canônica 3.

A correlação canônica da função 1 foi considerada moderada com 44,64%, enquanto que a da função 2 com 16,75%. O R^2 canônico presente na tabela é uma estimativa da variância compartilhada entre as variáveis canônicas psicológica e acadêmica, ou seja, a quantidade de variância explicada pelo par de variáveis canônicas. A variância compartilhada na função foi de **19,93%**, sendo baixa e resultante da correlação canônica moderada da mesma. A função canônica 2 apresenta variância compartilhada entre as variáveis muito baixa (**2,35%**) por consequência da correlação canônica fraca (Tabela 1).

Como a correlação canônica e a proporção de variância explicada foram indicativos da influência das características psicológicas no desempenho acadêmico, uma vez que as funções canônicas sejam significativas podem ser feitas análises das relações entre os grupos, entretanto é consenso entre pesquisadores de que os p-valores em análises de correlação canônica podem apresentar significância mesmo para efeitos ínfimos quando a amostra é suficientemente grande (SHERRY & HENSON 2005). Desta forma foi escolhido interpretar somente a função canônica 1 considerando que a segunda função canônica, apesar de significativa, apresentou variância compartilhada de somente 2,35% possibilitando o seu descarte (Tabela 1).

Tabela 1: Teste multivariado Lambda de Wilks (aproximação da distribuição F) para as três funções canônicas.

Função Canônica	Correl. Canônica	R^2 canônico	Fa	GL ₁	GL ₂	p-valor
1	0,4464	0,1993	12,7735	12	1.569,22	0,0000
2	0,1675	0,0235	2,4210	6	1.188,00	0,0249
3	0,0225	0,0005	0,1507	2	595,00	0,8601

Fonte: Autoria Própria.

A função canônica 1 foi representada pelo par de variáveis em (3) e (4) abaixo, que correspondem aos grupos de variáveis psicológicas (independentes) e acadêmicas (dependentes) respectivamente. Levando em consideração os coeficientes de (3), a mesma pode ser interpretada como contraste de *control* e *motivation* com *concept*, entretanto deve-se observar a menor influência da variável *concept* na variável canônica indicando menor importância dela na função canônica ajustada.

A variável de (4) pode ser interpretada como contraste de *math* e *reading* com *writing* e *science*, sendo que *science* possui a menor influência na variável canônica em questão, desta forma tendo menor importância na função canônica ajustada na análise.

$$U_1 = -1,2501 * control + 0,2367 * concept - 1,2491 * motivation \quad (3)$$

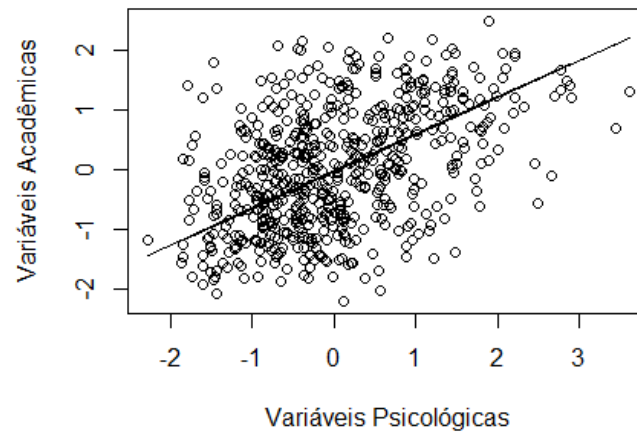
$$V_1 = -0,0440 * reading + 0,0551 * writing - 0,0194 * math + 0,0038 * science \quad (4)$$

Na Figura 2 encontram-se os escores obtidos por meio da combinação linear das variáveis originais considerando os coeficientes obtidos na função canônica 1, e também é mostrada a linha descrita pelo modelo linear ajustado entre ambas as variáveis canônicas. O modelo ajustado ilustra a correlação canônica obtida de 44,64% na função 1.

As cargas canônicas e cargas canônicas cruzadas da função canônica 1 para os dois grupos de variáveis foram apresentadas na Tabela 2. Observou-se que a variável psicológica *concept* apresentou cargas baixas, o que indica sua baixa importância na derivação da variável estatística

canônica. Já no grupo de variáveis acadêmicas, todas as cargas foram parecidas indicando a importância equânime de todas as variáveis na obtenção da variável estatística canônica.

Figura 2: Diagrama de dispersão e linha ajustada para a função canônica 1 entre as variáveis psicológicas e acadêmicas.



Fonte: Autoria Própria.

Tanto cargas canônicas quanto cargas canônicas cruzadas possuem o mesmo sinal (negativo) em ambos os grupos de variáveis, o que indica correlação positiva entre as variáveis na função canônica 1, principalmente entre as variáveis *control* com *reading* e *writing* (Tabela 2).

Os coeficientes de correlação entre as variáveis originalmente observadas e sua variável canônica são chamados como cargas canônicas. É perceptível que a variável *control* foi a variável primária, *motivation* teve contribuição secundária e *concept* teve contribuição considerada ínfima em relação às demais. No grupo de variáveis acadêmicas, todas tiveram contribuição significativa para a obtenção da variável canônica sendo que *science* e *math* possuem menores contribuições enquanto *writing* e *reading* possuem as maiores influências.

As cargas canônicas cruzadas se tratam dos coeficientes de correlação das variáveis originais com a variável canônica do grupo oposto e são mais utilizadas para interpretação de funções canônicas, uma vez que os coeficientes e cargas são otimizados para correlação e não para interpretação. Observou-se que as cargas cruzadas são mais fracas e moderadas para ambos os grupos de variáveis do que as cargas comuns.

Tabela 2: Cargas canônicas e cargas canônicas cruzadas para a função canônica 1.

Variáveis	Cargas canônicas	Cargas canônicas cruzadas
Psicológicas		
<i>Control</i>	-0,9143	-0,4082
<i>Concept</i>	-0,0997	-0,0446
<i>Motivation</i>	-0,5853	-0,2613
Acadêmicas		
<i>Reading</i>	-0,8804	-0,3931
<i>Writing</i>	-0,9101	-0,4063
<i>Math</i>	-0,7999	-0,3571
<i>Science</i>	-0,6941	-0,3099

Fonte: Autoria Própria.

No grupo psicológico a variável *control* apresentou maior correlação negativa (-0,9141), seguida da variável *motivation* (-0,5853) e *concept* apresentou menor correlação (-0,0997) na função canônica 1. No grupo acadêmico as variáveis *writing* e *reading* apresentaram maiores correlações negativas (-0,9101 e -0,8804) enquanto as demais apresentaram correlação fraca com a variável canônica oposta (Tabela 2). Apesar de reduzir os coeficientes em relação às cargas canônicas, o comportamento permaneceu similar quanto às cargas canônicas cruzadas.

A variável *concept* não possui cargas significativamente grandes na função canônica 1, logo, pode-se afirmar que o desempenho acadêmico tende a ser mais elevado quando as pontuações das características psicológicas também são elevadas e vice-versa. Entretanto, deve-se considerar que a variância explicada pela função canônica 1 é somente de 19,93% apesar de significativa segundo o teste de Lambda Wilks, pode ser conferida na Tabela 1.

Na Tabela 3 foram apresentadas as cargas canônicas quadradas, cargas canônicas cruzadas quadradas e índices de redundância da função canônica 1 para os dois grupos de variáveis. As cargas canônicas quadradas representam a proporção de variância compartilhada entre as variáveis originalmente observadas e a respectiva variável canônica. No grupo de variáveis psicológicas foi observada alta variância compartilhada na variável *control* (83,59%), variância compartilhada baixa para a variável *motivation* (34,26%) e proporção insignificante (0,99%) para a variável *concept*. Já no grupo de variáveis acadêmicas, *writing* (82,83%) e *reading* (77,51%) apresentaram alta proporção de variância compartilhada com a respectiva variável canônica e as variáveis *math* (63,98%) e *science* (48,18%) apresentaram proporções moderadas.

As cargas canônicas cruzadas quadradas representam a proporção de variância compartilhada entre as variáveis originais e a variável canônica oposta. No grupo psicológico, as variáveis *control* (16,66%) e *motivation* (6,82%) apresentaram proporção pequena de variação compartilhada com a variável canônica oposta enquanto que a proporção da variável *concept* (0,19%) foi insignificante. Quanto ao grupo acadêmico, todas as variáveis apresentaram proporções pequenas de variação compartilhada com a variável canônica oposta obtida na análise. As cargas canônicas cruzadas quadradas apresentaram comportamento similar às cargas canônicas cruzadas apesar de menores por se tratarem de relação entre dois grupos diferentes. Indicador que possui interpretação similar às cargas canônicas cruzadas quadradas, mas que avalia todo o grupo, é o índice de redundância também mostrado na Tabela 3.

Tabela 3: Cargas canônicas quadradas, cargas canônicas cruzadas quadradas e índices de redundância da função canônica 1 para os dois grupos de variáveis.

Variáveis	Cargas canônicas quadradas	Cargas canônicas cruzadas quadradas	Índice de redundância (%)
Psicológicas			
<i>Control</i>	0,8359	0,1666	8,46
<i>Concept</i>	0,0099	0,0019	
<i>Motivation</i>	0,3425	0,0682	
Acadêmicas			
<i>Reading</i>	0,7751	0,1545	13,94
<i>Writing</i>	0,8282	0,1650	
<i>Math</i>	0,6398	0,1275	
<i>Science</i>	0,4817	0,0960	

Fonte: Autoria Própria.

Quanto aos índices de redundância encontrados, pode-se observar que apenas 8,46% da variância presente na variável estatística canônica independente é explicada pela variável

estatística canônica dependente. Já a variável estatística canônica dependente possui 13,94% da variância explicada pela variável estatística canônica independente.

Não existe padrão de classificação para os índices de redundância, entretanto pode-se observar que os índices para a correlação canônica dos grupos de variáveis psicológicas e acadêmicas se apresentaram relativamente baixos.

Günther e Günther (1998) utilizaram análise de correlação canônica para avaliar perspectivas de futuro entre adolescentes. Como variáveis sócio-pessoais, idade, gênero, série, trabalho, escola que frequentam, foram relacionadas com perspectivas de futuro. Resultou em duas correlações, apontando duas relações significativas: adolescentes, frequentando uma escola privada e cursando uma série mais avançada, revelaram maiores expectativas de concluir o segundo grau e entrar para a universidade.

Rutter (1983) chamou atenção para características de escolas e turmas/classes que funcionam como significativas fontes de variação na busca da competência acadêmica de estudantes. Estas características contribuem, por meio de processos interpessoais, para estabelecer as diferenças de desempenho e realização entre grupos econômicos. Os jovens que frequentam uma escola privada e cursam uma série mais avançada referem maior expectativa de concluir o segundo grau e de se pensar como estudante universitário. Tais resultados são congruentes com as conclusões de Oliveira e Costa (1997).

CONCLUSÃO

De acordo com os resultados, a Análise de Correlação Canônica aplicada foi efetiva pelo fato de ter possibilitado detalhamento da estrutura de relação entre os grupos de variáveis psicológicas e acadêmicas em suas 600 observações. Tal detalhamento é possível por meio do estabelecimento de duplas combinações lineares ortogonais que maximizem a correlação entre os grupos de variáveis.

Quanto aos resultados da Análise de Correlação Canônica, apesar de as duas primeiras funções canônicas terem se apresentado significativas, somente a primeira apresentou proporção de variância explicada satisfatória. A análise de tal função evidenciou a relação entre as variáveis psicológicas de locus de controle (*control*) e motivação (*motivation*) com as variáveis de desempenho acadêmico. Também indicou que a variável psicológica autoconceito (*concept*) possui pouca relação com variáveis acadêmicas, de acordo com os dados estudados. Entretanto, deve-se observar que, apesar das relações discriminadas terem sua significância afirmada, a variância explicada pela função canônica ajustada foi de 19,93% com uma correlação canônica de 44,64%, desta forma as interpretações representam apenas parte da variância contida nos dados.

Por fim, a análise de correlação canônica se apresentou efetiva no estabelecimento de estrutura de relação e na análise de correlação entre as variáveis psicológicas e de desempenho acadêmico apresentadas nos dados obtidos. Entretanto, de forma a possibilitar a explicação de maior parte da variabilidade podem ser necessárias variáveis psicológicas adicionais.

REFERÊNCIAS

- COURVILLE T, THOMPSON B (2001) Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, 61(2): 229-248.
- GÜNTHER, I.A.; GÜNTHER, H. Brasília pobres, Brasília ricas: perspectivas de futuro entre adolescentes. *Psicol. Reflex. Crit.* vol.11 n.2 Porto Alegre, 1998.

HAIR JR., J. F.; ANDERSON, R. E.; TATHAN, R. L.; BLACK, W. C. Análise multivariada de dados. 5.ed. São Paulo: Bookman, 2007.

JOHNSON, R.A; WICHERN, D,W. Multivariate statistics, a practical approach. 6th Edition. P.600. 2007.

KAN, M.; SHAN, S; MEMBER, S.; IEEE; ZHANG, H.; LAO, S.; CHEN, X.; Member, S. Multi-View Discriminant Analysis. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, v. 38, n. 1, 2016.

MINGOTI, S.A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. 1 ed. Editora UFMG, 2005.

O'BRIEN RM (2007) A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673-690.

Oliveira, C. A. A. & Costa, A. E. B. Categorias de conflitos no cotidiano de adolescentes mineiros. *Psicologia: Reflexão e Crítica*, 10, 87-104. 1997.

PROTÁSIO, T.P.; TRUGILHO, P.F.; NEVES, T.A.; VIEIRA, C.M.M. Análise de correlação canônica entre características da madeira e do carvão vegetal de Eucalyptus. *Sci. For.*, Piracicaba, v. 40, n. 95, p. 317-326, set. 2012.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, Vienna, 2014.

ROCHA, L.C.S.; PELOGIO, E.A. Relação entre cultura e clima organizacionais: um estudo empírico em um campus do instituto federal de ensino. *Holos*, v.5. 2014.

RUTTER, M.; School effects on pupil progress: Research findings and policy implications. *Child Development*, 54, 1-29. 1983.

SHERRY, Alissa; HENSON, Robin K. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of personality assessment*, 84(1): 37-48.

STEWART D, LOVE W (1968) A general canonical correlation index. *Psychological bulletin*, 70(3): 1-160.

TRUGILHO PF, LIMA JT, MORI FA (2003) Correlação canônica das características químicas e físicas da madeira de clones de Eucalyptus grandis e Eucalyptus saligna. *Cerne*, 9(1): 066-080.

UCLA: *Institute for Digital Research and Education*. R Data Analysis Examples: Canonical Correlation Analysis. Disponível em: <<http://www.ats.ucla.edu/stat/data/mmreg.csv>>. Acesso em 14 de Novembro de 2016.

VAN DEN WOLLENBERG, AL. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, 42(2): 207-219. 1977.

XING, X.; WANG, K.; YAN, T.; ZHUOWEN, L. Complete canonical correlation analysis with application to multi-view gait recognition. *Pattern Recognition*. V. 50, 107–117. 2016.