
UMA ANÁLISE MULTIVARIADA DO PERFIL DE ALUNOS DE ENSINO SUPERIOR DE IES PÚBLICAS E PRIVADAS

A multivariate analysis of the profile of higher education students

Ana Maria de Paula Morais

Economista. Doutoranda em Economia Aplicada na Universidade Federal de Juiz de Fora.
Universidade Federal de Juiz de Fora, Faculdade de Economia. Rua José Lourenço Kelmer,
São Pedro. 36036330 - Juiz de Fora, MG – Brasil. ana.morais@economia.ufjf.br

Resumo: Nos últimos anos, o mercado de Ensino Superior no Brasil ganhou forte importância. Por um lado, houve um crescimento da oferta de cursos superiores, mas, por outro lado, a demanda também aumentou, principalmente em função da maior exigência de especialização da mão de obra. Diante disto, o presente trabalho utilizou algumas técnicas multivariadas em três períodos (2005, 2010 e 2017) para entender as principais características dos alunos de graduação, e os fatores que diferenciam alunos de instituições públicas e privadas de ensino. Foi realizado, primeiramente, uma análise de correspondência. Em seguida, utilizou-se de uma análise comparativa qualitativa e análise discriminante para verificar diferenças entre algumas variáveis sociais dos alunos, quando comparamos instituições públicas e privadas de ensino. Com essas análises, verificou-se que alunos de famílias mais carentes tiveram mais acesso ao Ensino Superior nos anos mais recentes, e que alunos de instituições privadas possuem preferência por cursos noturnos, talvez, por trabalharem durante o dia.

Palavras-chave: Análise multivariada; Instituições de Ensino Superior; Análise Discriminante.

Abstract: In recent years, the graduation courses market in Brazil has gained strong importance. There has been an increase in the supply of higher education courses, but on the other hand, there has also been an increase in the demand of this, mainly due to the greater need for specialized labor. So, the present study used some techniques to understand the main characteristics of graduation students. At first, a correspondence analysis was performed. Then, a qualitative comparative analysis and a discriminant analysis was used, to verify differences of some social variables of the students, when comparing public and private teaching institutes. With these analyzes, was found that students from poorer families had more access to higher education in recent years, and that students from private institutions may prefer night courses, perhaps because they work during the day.

Keywords: Systematic review; Multidimensional analysis; Words cluster; Development strategies; Iramuteq.

Código JEL: C38; C39.

1 INTRODUÇÃO

A crescente necessidade de especialização da mão de obra nas economias modernas tem instigado a população a obter níveis mais altos de escolaridade. Até um passado não muito distante, as pessoas conseguiam se manter empregadas sem a formação em um curso superior. Porém, isso não é mais realidade, o que implica em uma demanda crescente por Ensino Superior (FLANNERY, O'DONOGHUE, 2013; HEINESEN, 2018; LAM, SHIU, 2010; MALLIDIS *et al.*, 2012; OPPE-DISANO, 2011).

Fato é que a educação superior no Brasil só começou a obter alguma importância em meados dos anos 30, ainda com um desenvolvimento modesto, embora nos últimos anos tenha ocorrido um forte crescimento no número de alunos matriculados em cursos de Ensino Superior.

Ocorre que, especialmente a partir dos anos 2000, as políticas de modernização e de concursos públicos para ampliar as operações das Instituições de Ensino Superior (IES), de construção de novas universidades públicas, bem como programas de financiamento estudantil – como o “Programa de Financiamento Estudantil” (FIES) e o “Ensino Universitário para Todos” (ProUni), que concedeu bolsas de estudos a alunos de baixa renda –, provocaram um forte crescimento no número de matrículas em cursos de Ensino Superior (DEARDEN *et al.*, 2014; FLANNERY, O'DONOGHUE, 2013; LONG, 2019; PINTO, 2004; ROJAS *et al.*, 2016; SENHORAS *et al.*, 2006). Este fato pode ser observado tanto pela expansão das universidades públicas, como pelo crescente número de faculdades particulares espalhadas pelo Brasil (ZOGHBI *et al.*, 2013).

Assim, com a expansão do número de pessoas matriculadas em cursos de Nível Superior, é de se esperar que as características sociais e demográficas dos alunos, como renda familiar, estado civil, idade etc., tenham se tornaram mais heterogêneas. Isso significa que a tentativa de traçar um perfil geral dos alunos de Ensino Superior deve levar em consideração diversas variáveis.

As técnicas de Análise Multivariada têm sido amplamente utilizadas na literatura que trata desde a Educação Básica até o Ensino Superior. Farrell *et al.* (1994) e Webster (2001), por exemplo, utilizaram a Análise de Componentes Principais (ACP) a partir de dados coletados por questionários no estudo de variáveis relacionadas à educação. Farrell *et al.* (1994), por sua vez, buscaram determinar componentes baseados na relação entre a performance dos alunos, bem como os planos para estudo e dados de raça e gênero, com as chances de ingressarem no Ensino Superior, enquanto Webster (2001) analisou a continuação de onze critérios de qualidade acadêmica na determinação do ranqueamento de universidades dos Estados Unidos.

Já Marques (2010) e Marchesan, assim como Souza e Menezes (2011), aplicaram técnicas de Análise Multivariada para verificar a qualidade do processo de ensino e o desempenho dos alunos, a partir de questionários aplicados a alunos e professores. Marques (2010) utilizou também a Análise de Componentes Principais para reduzir o número de variáveis e analisar quais conjuntos de variáveis explicariam a maior parte da variabilidade do desempenho e infraestrutura de escolas públicas estaduais de Ensino Fundamental e Médio no Brasil. Com os componentes selecionados, aplicou a Análise Fatorial por componentes principais, rotacionada pelo Método Varimax, na intenção de chegar a indicadores com maior poder de explicação. Obteve, então, uma classificação das escolas analisadas em relação ao desempenho e infraestrutura, e concluiu que nenhuma das escolas obteve um aproveitamento superior a 70%.

Marchesan, Souza e Menezes (2011) utilizaram ainda a técnica de Análise de Agrupamento de acordo com a classificação hierárquica pelo Método de Ligação Simples, utilizando como métrica a distância euclidiana. O objetivo foi separar grupos similares e, portanto, com elevada correlação interna. Após a definição de tais grupos, utilizaram-se da análise de componentes principais para identificar as variáveis mais relevantes. E, finalmente, realizaram uma análise fatorial dentro dos grupos, com bons coeficientes de correlação de Pearson, definindo o número de fatores pelo critério gráfico Scree Plot. Dessa forma, obtiveram como principal resultado que os alunos valo-

rizam o incentivo por parte dos professores, de leitura, pesquisa, participação em discussões dos conteúdos e em eventos.

Assim, quando o assunto é educação, é comum se trabalhar com variáveis categóricas de avaliação. Portanto, outra técnica de análise multivariada amplamente utilizada na literatura trata-se da Análise de Correspondência, justamente por ser capaz de lidar com estes tipos de dados. Soutar e Turner (2002), Mazzarol e Soutar (2008), Shanka, Quintal e Taylor (2006) e Yavas e Shemwell (1996), se basearam nesta técnica na análise de questionários aplicados em alunos e universidades.

Soutar e Turner (2002) e Shanka, Quintal e Taylor (2006) buscaram entender as preferências de alunos que estavam concluindo o Ensino Médio no processo de escolha da universidade que desejariam se inserir. Através da análise de correspondência conjunta, Soutar e Turner (2002) verificaram que os principais determinantes nas preferências por universidades são: cursos adequados, perspectivas de emprego futuro e qualidade de ensino.

Mazzarol e Soutar (2008) utilizaram a análise de correspondência em uma amostra de 225 universidades australianas, que buscavam recrutar estudantes internacionais. Com as informações obtidas por essa análise, realizaram posteriormente, uma análise de agrupamento hierárquico para verificar a relação entre os países de origens dos estudantes e as instituições de ensino australianas.

Yeung e Lee (1999), por sua vez, realizaram um estudo do autoconceito de alunos do Ensino Médio, quando foram perguntados a respeito de seus conhecimentos em disciplinas verbais, matemáticas, acadêmicas e gerais. As respostas foram coletadas em dois períodos e analisadas através de uma análise fatorial confirmatória. Obtiveram como principais resultados que os autoconceitos dos alunos em disciplinas verbais e matemáticas possuem alta relação com autoconceitos acadêmicos, enquanto, por outro lado, existe uma correlação negativa entre autoconceitos de matemática e estudos verbais.

Portanto, esses trabalhos apresentam a relevância de se considerar técnicas de Análise Multivariada para analisar a relação entre diversas variáveis, que podem afetar as escolhas dos alunos ao decidirem se irão se inserir em um de curso Ensino Superior e em qual Instituições de Ensino Superior (IES). Ou seja, a utilização dessas técnicas pode contribuir com análises sobre o sistema educacional, especialmente em um país como o Brasil, que conta com IES com características administrativas e de ensino bastante diferentes.

Um exemplo a mencionar alude às IES privadas, que, em geral, possuem maior foco em preparar seus alunos para o mercado de trabalho, enquanto as públicas investem relativamente mais em pesquisas e produções acadêmicas. Além disso, no Brasil as IES são gratuitas, embora continuem não sendo acessíveis para grande parte da população, uma vez que contam com exames de seleção altamente competitivos e que acabam por aprovar apenas os alunos mais preparados (REY; DEL; ESTEVAN, 2020). Por outro lado, os alunos de IES privadas precisam pagar mensalidades, embora exista a possibilidade de conseguirem bolsas totais ou parciais a depender de alguns fatores, como renda familiar e nota nos vestibulares (SCHWARTZMAN, 2015). No entanto, analisar as principais características de tais alunos, e como elas se relacionam em um sistema educacional superior, com as peculiaridades encontradas no Brasil, é um assunto que muitas vezes tem sido tratado através de análises descritivas, que ignoram as relações simultâneas entre as variáveis em questão.

Portanto, diante de todas essas possibilidades de análises das variáveis referentes aos alunos, este trabalho tem como objetivo verificar o perfil geral dos alunos de Ensino Superior e entender as principais diferenças entre alunos de instituições públicas versus privadas. Essa perspectiva fomentará uma maior compreensão sobre os principais fatores que influenciam as características observadas em diferentes tipos de cursos superiores, que deve ser usada como base para a elaboração e o aprimoramento de políticas educacionais que possam ser pensadas para a evolução do Ensino Superior no Brasil.

Para a construção deste trabalho foram aplicadas quatro técnicas de Análise Multivariada, que serão apresentadas e discutidas na seção seguinte: análise de correspondência, análise de componentes principais, análise comparativa qualitativa e análise discriminantes. Mesmo que as duas primeiras técnicas já tenham sido aplicadas por outros trabalhos e apresentem resultados mais diretos, elas servem para descrever bem os dados e as relações entre eles. As técnicas de análise comparativa qualitativa e análise discriminante, por sua vez, trazem uma abordagem diferente do que já realizado pela literatura para estudos sobre a educação. Essas últimas, embora de simples aplicação, são capazes de proporcionar análises completas e importantes para o entendimento do setor. Desse modo, as quatro análises utilizadas, de certa forma, se completam e permitem uma boa compreensão do público envolvido no Ensino Superior no Brasil.

Além desta introdução e da segunda seção com a apresentação dos métodos de análise, na terceira seção apresenta-se a base de dados e as variáveis utilizadas neste estudo. Já na quarta seção têm-se os principais resultados obtidos. Por fim, o trabalho encerra-se com uma discussão das principais conclusões obtidas e destaca possibilidades de estudos futuros.

2 ESTRATÉGIA EMPÍRICA

A caracterização do perfil de alunos de Ensino Superior depende da análise de diversas variáveis. Por isso, este trabalho pretende aplicar algumas técnicas de Análise Multivariada na intenção de encontrar padrões confiáveis, que descrevam e diferenciem o público de alunos de instituições de ensino com diferentes características. Logo, nesta seção serão brevemente apresentados os métodos utilizados ao longo deste estudo. Em seguida, têm-se a descrição das variáveis utilizadas nas análises aqui descritas.

2.1 Análise de Correspondência

A Análise de Correspondência (AC) permite uma primeira e importante forma de interpretação dos dados utilizados nesta pesquisa. Trata-se de uma técnica multivariada útil para identificar associações existentes entre variáveis categóricas, como será o caso das variáveis utilizadas neste estudo, conforme descrito na próxima seção. O principal objetivo é transformar uma tabela de informação numérica em uma tabela de contingência (ou frequência), que possa ser representada graficamente de forma a facilitar a interpretação (GREENACRE, 2000).

A principal vantagem da Análise de Correspondência é a facilidade de representação de dados categóricos de uma tabela de contingência em um espaço dimensional reduzido, de modo que as relações de linha e coluna sejam comparáveis em termos de distâncias. Isso torna possível a percepção de semelhanças e diferenças nas categorias das colunas em relação a determinada categoria de linha (SHANKA; QUINTAL; TAYLOR, 2006). Além disso, a única restrição em termos de dados, é que não existam entradas negativas na tabela de contingência (YAVAS; SHEMWELL, 1996).

A AC também busca reduzir a estrutura dos dados sem a necessidade de assumir uma distribuição de probabilidade. Ou seja, pode ser considerada como um caso especial da Análise de Componentes Principais (ACP), descrita a seguir, todavia, para dados categóricos ao invés de contínuos. Além disso, a AC pode ser aplicada para analisar várias variáveis categóricas ao mesmo tempo. Neste caso, deve-se considerar uma matriz indicadora e Burt, que explora a relação entre o conjunto de variáveis categóricas. As informações contidas em uma matriz indicadora, X , com as variáveis e suas categorias, permitem gerar a matriz Burt, B , que representa uma tabulação cruzada de todas as categorias e variáveis, isto é, $B=X'X$.

Desse modo, neste trabalho é utilizada uma análise de correspondência conjunta (ACJ), proposta por Greenacre e Blasius (2006), que se baseia em um método iterativo de mínimos qua-

drados iniciado a partir de uma matriz $B_0=B$. Em cada interação os elementos fora da diagonal são substituídos pelos elementos de uma matriz de aproximação, até que ocorra uma convergência da matriz B.

A análise é repetida para dois conjuntos de variáveis: o primeiro referente à situação em que os alunos se encontram no período de graduação e, o segundo, referente à algumas situações pre-determinadas na vida do aluno. Além disso, essa mesma análise foi repetida e aplicada em três períodos, separados por alguns anos entre eles.

2.2 Análise de Componentes Principais

O segundo passo da pesquisa se deu através da aplicação da técnica de Análise de Componentes Principais (ACP). Este método foi utilizado na intenção de identificar as variáveis mais importantes e criar um conjunto menor de novas variáveis, capaz de manter a maior parte das informações dos dados. Além disso, esse método não requer que seja assumido um modelo *a priori*, pois não há uma hipótese a ser testada, não é necessário postular causalidades e não se requer a adoção de hipótese de distribuição normal multivariada.

A variabilidade inicial do sistema determinada por p variáveis, em geral, podem ser reproduzidas por p componentes principais. No entanto, dependendo da correlação entre as variáveis originais, menos componentes, $k < p$, pode reproduzir grande parte da variabilidade original. Isto é, essa técnica expressa a estrutura de variância dos dados através de poucas combinações lineares entre as variáveis originais. Essas combinações são conhecidas como componentes principais:

$$\begin{aligned} Y_1 &= a'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= a'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ Y_p &= a'_p X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned}$$

Ou seja, $Y_i = \alpha_i X$, em que Y_i são os componentes; α_i são os coeficientes que representam os pesos das variáveis, logo, a importância de cada variável para o valor de cada componente, e X é um vetor das variáveis originais.

A qualidade dos componentes obtidos depende da estrutura de dados. Para a aplicação desta técnica espera-se que as variáveis consideradas possuam forte correlação entre si¹. Portanto, como muitas das variáveis deste estudo não apresentaram correlação alta, a aplicação desta técnica foi realizada separadamente, considerando, em cada caso, apenas variáveis com correlações significativas entre si, conforme será apresentado nos resultados.

2.3 Análise Comparativa Qualitativa

Na próxima etapa é realizada uma Análise Comparativa Qualitativa (do inglês, *Qualitative Comparative Analysis*, QCA). Esse método requer que se considere um número limitado de variáveis. Porém, isso é possível ao utilizar os índices obtidos pela técnica ACP, que reduzem uma grande quantidade de variáveis em poucos índices.

Esta técnica baseia-se na Teoria dos Conjuntos (relações lógicas) e na Álgebra Booleana para avaliar múltiplas combinações de variáveis ("condições"), que são presentes ou ausentes quando um fenômeno de interesse ocorre ou não. Esse método apresenta também um padrão de associações entre os conjuntos, que, embora não prove as relações causais, fornece suporte para a existência de tais relações. Ou seja: o QCA possibilita encontrar caminhos teóricos diferentes de acordo com certos resultados, o fenômeno de "causalidade complexa".

As associações entre fatores produzem uma configuração específica, evidenciando também as combinações não associadas ao fenômeno de interesse. Os resultados revelam padrões de associações, que são avaliados em termos de suficiência e necessidade (BERGSCHLOSSER *et al.*, 2009).

1 Para a ACP, é recomendável que a correlação entre as variáveis seja superior a 0,3.

A QCA combina uma análise detalhada de dentro do caso, com as comparações sistemáticas entre os casos. Portanto, o processo de pesquisa é iterativo, envolvendo várias rodadas de comparação de análise dentro de cada caso e entre os casos. Nesse processo, os primeiros resultados obtidos devem induzir à redefinição dos conjuntos, que descrevem as condições e resultados, que, certamente, fornecerão maiores conhecimentos sobre os casos (LEGEWIE, 2013).

No caso do QCA, com conjuntos fuzzy (fsQCA), é possível se obter associações parciais ou completas, permitindo valores de condições entre 0 e 1, que são capazes de descrever o grau e o tipo de casos pertencentes a cada conjunto. Normalmente, se considera três âncoras, que definem um conjunto difuso: adesão plena (quando há uma associação igual a 1), nenhuma adesão (quando a associação é zero) e um ponto de corte (escore 0,5).

As combinações são organizadas em uma tabela verdade, que serve para identificar padrões causais de suficiência e necessidade de cada configuração. No entanto, são raras as condições em que todos os casos atendem à relação de suficiência e necessidade. Portanto, é oportuno avaliar o quão bem os dados se encaixam em termos de suficiência e necessidade. Para isso, existem duas medidas para avaliar o bom ajuste dos resultados: consistência e cobertura.

A consistência demonstra a proporção de casos consistentes com o resultado e assemelha-se a noção de significância em modelos estatísticos. Ou seja, avalia a participação do número de casos presente simultaneamente entre uma combinação x e o resultado y sobre o total de casos em x . Quanto mais próximo de 1 for o valor dessa medida (I_{xy}), maior é a consistência dos dados²:

$$I_{XY} = \frac{\sum \min(x_i, y_i)}{\sum x_i}$$

em que, X é a configuração, Y é o resultado, x_i é a adesão de cada caso na configuração X ; e y_i a adesão de cada caso em Y .

Já cobertura indica o número de casos empiricamente observados para cada configuração, isto é, avalia a parcela dos casos presentes simultaneamente em x e y em relação ao total de casos em y . Essa medida é análoga ao R^2 em modelos estatísticos e é definida por:

$$C_{XY} = \frac{\sum \min(x_i, y_i)}{\sum y_i}$$

Com isso, uma cobertura igual a 1 indica uma sobreposição entre x e y ; e a condição X abrange todos os casos pertencentes ao resultado y . Porém, quando há vários caminhos para um mesmo resultado, a cobertura pode ser pequena.

2.4 Análise Discriminante

A última etapa deste trabalho consiste na utilização da técnica multivariada de Análise Discriminante (AD). Aqui, a intenção é a separação de grupos distintos de objetos, permitindo se descrever as diferenças entre os grupos para, posteriormente, classificar novos objetos como membros em um dos grupos. Portanto, esta técnica requer um conhecimento prévio das características gerais dos dados.

Para a distribuição dos grupos, deve-se definir uma variável categórica como dependente e múltiplas variáveis métricas como independentes. Tais variáveis discriminantes medem as características nas quais os grupos devem diferir. Dessa forma, torna-se possível identificar a importância relativa de cada uma das variáveis independentes na previsão de participação em um dos grupos. O objetivo matemático é combinar linearmente as variáveis discriminantes, de modo que os grupos sejam os mais distintos estatisticamente quanto possível (PARKER, 1999).

² Espera-se que a consistência seja pelo menos igual a 0,8 para todas as configurações.

Portanto, deve-se construir uma regra de classificação, que permita classificar o novo objeto a partir da probabilidade de ele pertencer a cada um dos grupos. Nesse processo, deve-se minimizar o custo de classificação incorreta, ou seja, o erro em afirmar que um objeto pertence a um grupo, quando na verdade ele pertence a outro (JOHNSON; WICHERN *et al.*, 2002). Para cada observação das populações, π_1 e π_2 , é calculado o seu escore na função discriminante. A análise dos escores permite avaliar a qualidade da função em termos de erro de classificação e capacidade de discriminação. Se essa qualidade é boa, os escores de π_1 devem ser diferenciados dos de π_2 .

Será utilizada a função discriminante logística para que seja possível a análise com variáveis qualitativas. De acordo com Albert e Lesaffre (1986), em vez de fazer suposições sobre a distribuição dos dados dentro de cada grupo, a análise discriminante logística baseia-se na suposição de que as razões de verossimilhança dos grupos têm uma forma exponencial. Como a regressão logística multinomial é capaz de lidar com regressores binários e contínuos ao mesmo tempo, a análise discriminante logística também é apropriada para variáveis discriminantes binárias e contínuas. Além disso, este modelo é flexível em termos de permitir que as variáveis discriminantes não possuam distribuição normal multivariada.

A regra de classificação da função discriminante logística funciona da seguinte forma: a variável de resposta Y assume o valor "1" se o elemento amostral com o vetor das variáveis independentes, X , pertence à população π_1 ; e assume o valor "0" caso a observação pertença à π_2 . Ou seja, a observação com o seu respectivo vetor de valores, X , será alocada à π_1 se a condição da razão de chance (RC) abaixo for satisfeita, caso contrário será alocada à π_2 :

$$RC = \frac{\hat{p}(X)}{1 - \hat{p}(X)} = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) > 1$$

sendo o vetor β estimado por máxima verossimilhança.

3 BASE DE DADOS

Na intenção de examinar o perfil dos alunos de Instituições de Ensino Superior, foram utilizadas uma série de variáveis referentes aos alunos que prestaram o Exame Nacional de Desempenho de Estudantes (ENADE), disponibilizadas pelo banco de dados do ENADE, através do site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Os dados foram coletados para os anos de 2005³, 2010 e 2017, o que permite uma comparação entre diferentes períodos e, principalmente, possibilita uma análise crítica a respeito dos impactos de políticas educacionais na sociedade.

O ENADE avalia o rendimento e desempenho dos estudantes concluintes de Ensino Superior de alguns cursos selecionados, que variam de acordo com o ano, além de aplicar um "Questionário do Estudante", que coleta informações socioeconômicas e sobre o processo de formação dos estudantes. Portanto, os dados utilizados neste trabalho são basicamente de carácter categórico, tendo em vista que são baseados nas respostas de um questionário. Isso justifica a escolha dos métodos descritos na seção anterior, já que são capazes de proporcionar análises interessantes a partir de dados categóricos e binários. A descrição mais detalhada de cada variável considerada é apresentada a seguir.

3 O questionário que foi aplicado em 2005 é diferente do aplicado nos outros anos. Por isso, foi necessário se fazer algumas adaptações em algumas variáveis.

3.1 Análise Descritiva dos dados

As variáveis utilizadas neste estudo estão descritas nas Tabelas 1 e 2⁴. Em 2015, 200.100 alunos prestaram o ENADE; em 2010, foram 238.537 alunos; e, em 2017, 442.496 alunos foram observados.

A Tabela 1 apresenta a média e desvio padrão das variáveis contínuas deste estudo, isto é, a média das notas gerais obtidas pelos alunos no ENADE e a média da idade dos alunos que prestaram o ENADE. Nela pode-se observar que a média das notas dos alunos que prestaram o exame aumentou de 2005 para 2010, e de 2010 para 2017. Já a idade média dos alunos não apresentou grandes alterações ao longo desses anos.

Tabela 1 – Análise descritiva das variáveis contínuas

Variáveis	2005	2010	2017
Média das notas	37,998	41,182	44,353
	14,665	14,443	13,939
Média de idades	26,533	24,162	27,295
	7,833	6,256	6,965

Fonte: Elaboração Própria a partir de dados do INEP(2019).

Na Tabela 2 são apresentadas as variáveis que foram retiradas do formulário respondido pelos alunos no site do ENADE. Portanto, tais variáveis são de caráter qualitativo, seguindo as categorias conforme explicado na tabela.

Pode-se perceber que, em 2017, a quantidade de alunos que prestaram o ENADE – matriculados em instituições públicas de ensino – aumentou comparativamente às matrículas em instituições privadas, mesmo que a quantidade de bolsas de financiamento oferecidas tenha aumentado. Talvez, esses fatos tenham relação com o que se observa na variável de horas trabalhadas, pois, no ano de 2005, grande parte dos estudantes estavam trabalhando, o que mudou nos anos seguintes. Além disso, a quantidade de horas que os alunos se dedicam aos estudos parece ter aumentado nos anos mais recentes.

É possível perceber, também, que pessoas de famílias com renda mais baixa, filhos de pais com menor escolaridade e vindos de escolas públicas de Ensino Médio tiveram mais acesso ao Ensino Superior no ano de 2017.

4 RESULTADOS

Primeiramente foi realizada uma análise de correspondência, assim como feito por Soutar e Turner (2002), Mazzarol e Soutar (2008), Shanka, Quintal e Taylor (2006). Nesta etapa, a análise foi repetida para dois conjuntos de variáveis: o primeiro referente à situação em que os alunos se encontravam no período de graduação; já o segundo, referente à algumas situações pré-determinadas na vida do aluno. Essa divisão foi necessária, porque, como se está trabalhando com muitas variáveis, a correlação entre algumas delas é menor que 0,30, o que melhora ao considerar esta divisão⁵.

4 As informações faltantes no ano de 2005 se devem ao fato de não conter as mesmas opções no questionário deste ano em relação ao que foi aplicado nos demais anos. Sendo assim, para a aplicação das técnicas neste ano, as categorias foram ajustadas.

5 Foi realizada uma Análise de Correspondência para todas as variáveis, mas esta se mostrou inviável, pois seria necessário se considerar 7 dimensões para alcançar uma boa representatividade dos dados.

Tabela 2 – Análise descritiva das variáveis categóricas

Variáveis	Categoria	2005		2010		2017	
		Frequência	Porcentagem	Frequência	Porcentagem	Frequência	Porcentagem
Categoria Administrativa							
Pública	0	69218	34,59	55752	24,80	148701	42,05
Privada	1	130882	65,41	169080	75,20	204968	57,95
Turno							
Diurno	0	60612	30,29	89263	39,70	134271	37,97
Noturno	1	139488	69,71	135569	60,30	219398	62,03
Horas Trabalhadas							
Nenhuma	0	40401	20,19	141989	63,15	150417	42,53
Trabalha eventualmente até 20 horas semanais	1	11178	5,59	12935	5,75	28774	8,14
21 a 39 horas semanais	2	23497	11,74	9939	4,42	28533	8,07
40 horas ou mais	3	37092	18,54	23869	10,62	38619	10,92
	4	87932	43,94	36100	16,06	107326	30,35
Bolsa Financiamento							
Não possui	0	132032	65,98	151975	67,59	211475	59,79
Possui	1	68068	34,02	72857	32,41	142194	40,21
Estado Civil							
Casado	0	56952	28,46	37310	16,59	82212	23,25
Solteiro	1	143148	71,54	187522	83,41	271457	76,75
Horas de Estudo							
Nenhuma	0	14006	7,00	10012	4,45	14084	3,98
De uma a três	1	76199	38,08	108247	48,15	141707	40,07
De quatro a sete	2	65140	32,55	60342	26,84	10619	30,03
De oito a doze	3	24086	12,04	25965	11,55	46534	13,16
Mais de doze	4	20669	10,33	20266	9,01	45154	12,77
Bolsa Acadêmica							
Não possui						246354	69,66
Possui						107315	30,34
Escolaridade do Pai							
Nenhuma	0	13535	6,76	8761	3,90	22997	6,50
Ensino Fund. (1º ao 5º ano)	1	68553	34,26	53460	23,78	93148	26,34
Ensino Fund. (6º ao 9º ano)	2	33171	16,58	31563	14,04	52895	14,96
Ensino Médio	3	50692	25,33	75204	33,45	112952	31,94
Ensino Superior	4	34149	17,07	40736	18,12	52764	14,92
Pós-graduação	5			15108	6,72	18913	5,35
Escolaridade da Mãe							
Nenhuma	0	11846	5,92	6292	2,80	14641	4,14
Ensino Fund. (1º ao 5º ano)	1	61778	30,87	42564	18,93	77420	21,89
Ensino Fund. (6º ao 9º ano)	2	34971	17,48	28954	12,88	51296	14,50
Ensino Médio	3	55229	27,6	77093	34,29	118526	33,51
Ensino Superior	4	36276	18,13	46160	20,53	59473	16,82
Pós-graduação	5			23769	10,57	32313	9,14
Renda							
Nenhuma				12750	5,67		
Até 1,5 salários-mínimos	0	59887	29,93	19898	8,85	77034	21,78
De 1,5 a 5 salários-mínimos	1			58061	25,82	97727	27,63
De 3 a 4,5 salários-mínimos	2	104222	52,08	45413	20,20	72045	20,37
De 4,5 a 6 salários-mínimos	3			28971	12,89	38732	10,95
De 6 a 10 salários-mínimos	4	24974	12,48	30809	13,70	40221	11,37
De 10 a 30 salários-mínimos	5	6797	3,40	23359	10,39	24220	6,85
Mais de 30 salários-mínimos	6	4220	2,11	5571	2,48	3690	1,04
Escola do ensino médio							
Privada	0	66174	33,07	98570	43,84	109961	31,09
Pública	1	133926	66,93	126262	56,16	243708	68,91

Fonte: Elaboração Própria a partir de dados do INEP (2019).

A Tabela 3, na sequência, apresenta a Análise de Correspondência Conjunta para o primeiro conjunto de variáveis para os anos de 2005, 2010 e 2017⁶. Em todos os anos, a primeira dimensão foi capaz de atingir, sozinha, mais de 80% da estrutura de dados originais. A contribuição relativa de cada dimensão é apresentada na Tabela 3 como uma porcentagem da variância total. Considerando as duas primeiras dimensões, a variância desconsiderada é, em geral, menor do que 1%.

Tabela 3 – Análise de correspondência para situação dos alunos no período da graduação

	2005		2010		2017	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
Categoria Administrativa						
Pública	0,526	0,181	0,641	0,290	0,586	0,111
Privada	-0,278	-0,096	-0,211	-0,096	-0,425	-0,081
Turno						
Diurno	0,330	-0,036	0,188	-0,028	0,364	-0,072
Noturno	-0,143	0,015	-0,286	0,043	-0,223	0,044
Horas Trabalhadas						
Nenhuma	0,412	-0,326	0,240	-0,109	0,176	-0,124
Trabalha eventualmente até 20 horas semanais	0,183	-0,140	-0,067	0,002	0,059	-0,037
21 a 39 horas semanais	0,108	0,020	-0,196	0,080	0,088	0,005
40 horas ou mais	-0,047	0,061	-0,365	0,120	-0,037	0,040
Bolsa Financiamento						
Não possui	0,081	0,026	0,104	0,039	0,233	0,040
Possui	-0,156	-0,051	-0,218	-0,081	-0,346	-0,059
Estado Civil						
Casado	-0,141	0,111	-0,301	0,080	-0,184	0,129
Solteiro	0,056	-0,044	0,060	-0,016	0,056	-0,039
Horas de Estudo						
Nenhuma	-0,134	-0,002	-0,210	0,106	-0,140	0,106
De uma a três	-0,072	0,015	-0,094	0,005	-0,089	0,051
De quatro a sete	0,022	-0,002	0,048	-0,013	0,025	-0,013
De oito a doze	0,091	-0,015	0,157	-0,019	0,111	-0,057
Mais de doze	0,181	-0,028	0,258	-0,014	0,150	-0,104
Bolsa Acadêmica						
Não possui					-0,156	0,025
Possui					0,359	-0,058
Proporção	82,37	16,35	85,11	14,26	92,43	7,02
Proporção cumulativa	82,37	98,72	85,11	99,36	92,43	99,45

Fonte: Elaboração própria (2019).

No ano de 2005, as categorias que mais contribuíram para a definição da dimensão 1 foram: faculdades públicas, aulas do período da manhã e/ou tarde, alunos que não trabalham e que não possuem bolsa de financiamento⁷ e alunos que estudam mais do que doze horas de estudo (todos com sinal positivo). Já nos anos de 2010 e 2017, as categorias de cursos noturnos – alunos que estão inseridos no mercado de trabalho, com carga-horária de 40 horas e bolsas de financiamento – ganharam importância (ambas as variáveis com valor negativo). É provável que essas alterações tenham ocorrido por conta de mudanças sociais e culturais vivenciadas no país. As principais

⁶ A informação sobre bolsas acadêmicas só está disponível para o ano de 2017. Portanto, foi realizado neste ano a análise com e sem esta variável. Como os resultados foram bastante parecidos, optou-se por apresentar a versão mais completa.

⁷ Deve-se lembrar que a bolsa de financiamento é válida apenas para faculdades privadas.

políticas de financiamento e inclusão no Ensino Superior surgiram em 2005 e, junto com elas, se expandiu a oferta de cursos de graduação. Com essa situação e com a crescente necessidade de especialização da mão de obra, é natural que as pessoas que trabalhavam em funções mais básicas (sem requerer o Ensino Superior) aproveitassem as oportunidades de financiamento e o aumento da oferta de cursos para se especializarem, o que evidencia a maior importância de cursos noturnos nos últimos anos.

A representação gráfica da Análise de Correspondência encontra-se na Figura 1. Nela pode-se observar que as IES privadas estão graficamente mais próximas de alunos, que trabalham regularmente (40 horas por semana), que escolhem cursos noturnos e que utilizam de bolsas de financiamento. Por outro lado, alunos que não trabalham e que dedicam mais horas de seus dias ao estudo estão mais próximos de instituições públicas. Os pontos referentes a cursos diurnos e bolsa acadêmicas estão praticamente sobrepostos no ano de 2017.

Na Tabela 4, por sua vez, é apresentada a análise de correspondências para o segundo conjunto de variáveis e, também, para os anos 2005⁸, 2010 e 2017. Neste caso, embora a variância desconsiderada seja um pouco maior, ainda assim as duas primeiras dimensões estão cobrindo mais do que 90% da estrutura de dados originais (sendo a primeira dimensão responsável por mais de 70%).

Tabela 4 – Análise de correspondência de situações predeterminadas na vida dos alunos

	2005		2010		2017	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
Escolaridade do Pai						
Nenhuma	-0,965	1,110	-1,128	1,531	-0,908	1,007
Ensino Fundamental (1º ao 5º ano)	-0,478	-0,040	-0,619	0,091	-0,532	0,051
Ensino Fundamental (6º ao 9º ano)	-0,058	-0,348	-0,227	-0,286	-0,163	-0,271
Ensino Médio	0,331	-0,232	0,140	-0,283	0,193	-0,263
Ensino Superior	0,908	0,324	0,636	0,087	0,709	0,119
Pós-graduação			0,907	0,562	1,047	0,522
Escolaridade da Mãe						
Nenhuma	-0,846	0,835	-1,009	1,093	-0,914	0,909
Ensino Fundamental (1º ao 5º ano)	-0,474	-0,023	-0,651	0,075	-0,574	0,077
Ensino Fundamental (6º ao 9º ano)	-0,110	-0,258	-0,328	-0,190	-0,271	-0,188
Ensino Médio	0,279	-0,132	0,059	-0,167	0,087	-0,181
Ensino Superior	0,765	0,216	0,522	0,064	0,632	0,082
Pós-graduação			0,627	0,225	0,741	0,211
Renda						
Até 1,5 salários-mínimos	-0,398	0,026	-0,414	0,026	-0,446	0,067
De 1,5 a 5 salários-mínimos			-0,337	-0,063	-0,253	-0,077
De 3 a 4,5 salários-mínimos	0,007	-0,078	-0,152	-0,084	0,004	-0,112
De 4,5 a 6 salários-mínimos			0,060	-0,056	0,265	-0,052
De 6 a 10 salários-mínimos	0,532	0,109	0,302	0,018	0,522	0,064
De 10 a 30 salários-mínimos	0,819	0,293	0,655	0,197	0,949	0,324
Acima de 30 salários-mínimos	1,007	0,432	0,955	0,436	1,254	0,551
Tipo de escola do ensino médio						
Privada	0,488	0,101	0,456	0,062	0,622	0,086
Pública	-0,241	-0,050	-0,356	-0,048	-0,281	-0,039
Proporção	73,27	20,14	71,72	71,72	76,88	17,10
Proporção cumulativa	73,27	93,41	20,13	91,84	76,88	93,98

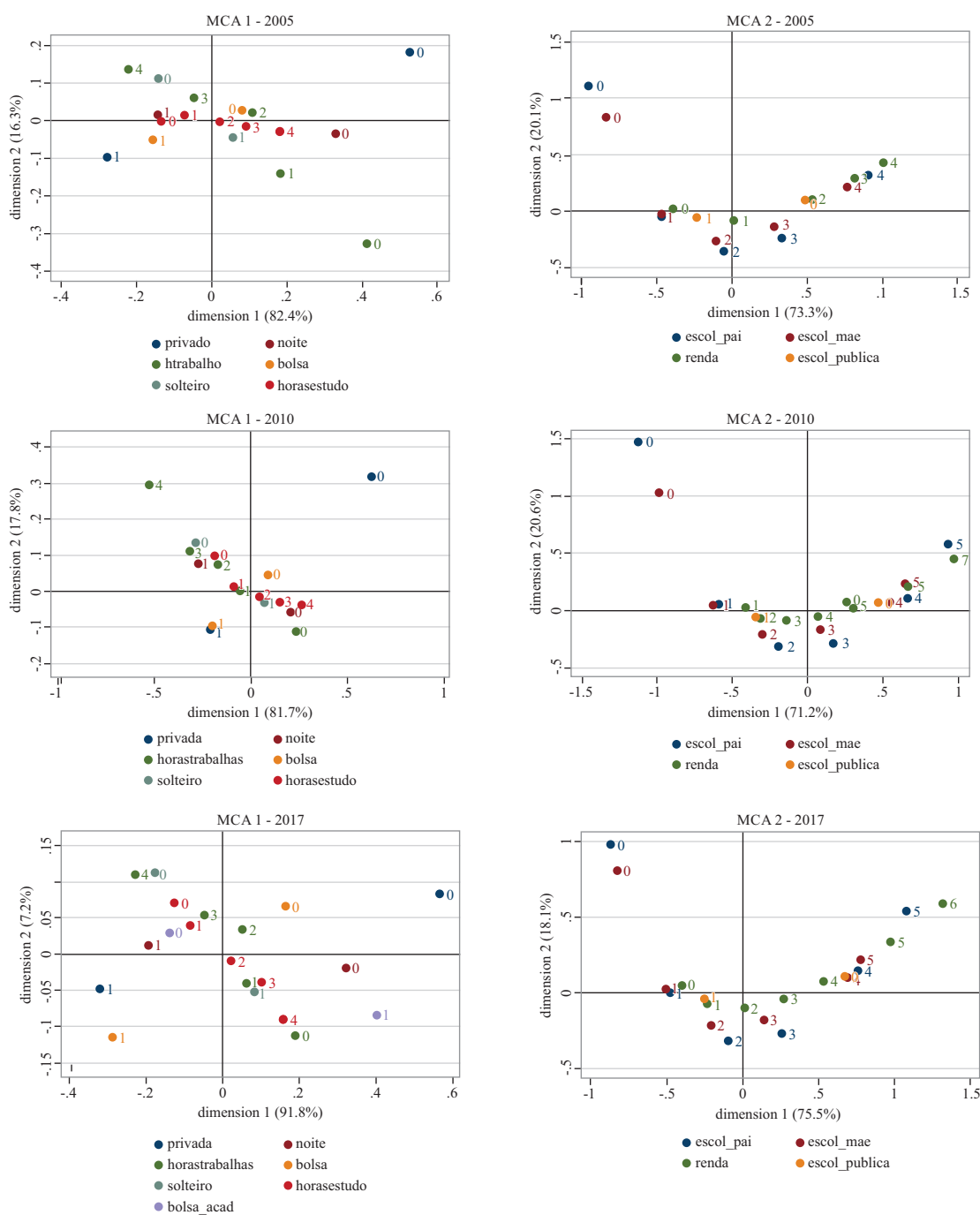
Fonte: Elaboração própria (2019).

⁸ Em 2005 não havia a opção de pais com pós-graduação no questionário, e a divisão das faixas salariais é diferente dos outros anos.

O que se observa nesta tabela é que as categorias que mais contribuíram para a dimensão 1 foram justamente os extremos de cada variável, um com valor negativo, e o outro com valor positivo. As categorias referentes à baixa escolaridade dos pais, baixa renda e estudo em escolas públicas no Ensino Médio, entraram com sinal negativo, enquanto o oposto (alta escolaridade dos pais, alta renda e estudo em escolas privadas) entrou com sinal positivo.

A representação gráfica desta segunda análise de correspondência está disposta na Figura 1. É interessante observar como o ponto de alta escolaridade do pai encontra-se próximo ao ponto de alta escolaridade da mãe, assim como os pontos de baixa escolaridade estão também próximos em todos os anos. Além disso, alunos que realizaram o Ensino Médio em escolas privadas estão mais próximos a situações de alta renda e alta escolaridade dos pais, como é de se esperar.

Figura 1 – Análise de correspondência



Fonte: elaboração própria (2019).

Como foi verificado que, no ano de 2017, a primeira dimensão da Análise de Correspondência foi capaz de representar mais de 90% da estrutura de dados, segue-se que a QCA será realizada para este mesmo ano.

Dentre as variáveis referentes aos alunos, a que teve maior importância na Análise de Correspondência foi a categoria administrativa das instituições de ensino. Portanto, na Análise Comparativa Qualitativa, que se segue, essa variável será tratada como o fenômeno de interesse, e as configurações das demais variáveis devem ser analisadas como presentes ou ausentes a esse fenômeno. O problema é que essa técnica requer que não se considere uma grande quantidade de variáveis. Portanto, foi necessário aplicar primeiro a Análise de Componentes Principais na tentativa de reduzir a quantidade de variáveis.

A maioria das variáveis consideradas não apresentou correlações altas entre elas, talvez por causa da grande quantidade de alunos sendo considerados, o que acaba trazendo uma heterogeneidade significativa entre eles. Por conta disto, a Análise de Componentes Principais não gerou índices capazes de expressar bem a estrutura de variância dos dados. Na tentativa de melhorar este problema, foi realizado, também nesta técnica, uma divisão entre um conjunto de variáveis associadas às características dos alunos no momento da graduação, e outro conjunto com as características pré-determinadas a esses alunos (assim como foi feito na Análise de Correspondência).

Para o primeiro conjunto de variáveis, ainda não foi possível reduzir o número de variáveis pelo ACP, pois seria necessário se considerar 7 componentes para se atingir pelo menos 70% da variância total dos dados, o que é inviável. Desse modo, ao fazer a análise para o segundo conjunto de variáveis, foi possível se chegar em um índice que será tratado como "índice família", que resume bem a variância das variáveis: escolaridade do pai, escolaridade da mãe e renda familiar.

Para o ano de 2017, o primeiro componente⁹ da ACP, que considera estas três variáveis, pode ser representado pela seguinte equação¹⁰:

$$f_{familia} = 0,59 \text{ escol}_{pai} + 0,60 \text{ escol}_{mae} + 0,53 \text{ renda}$$

Com isso, temos que as três variáveis possuem praticamente a mesma importância na definição do índice família. Agora, é possível proceder com a Análise Comparativa Qualitativa (QCA), que considera a variável privada (P) como o fenômeno de interesse. Como "condições" foram consideradas as variáveis idade (I), nota (N), noite (U), família (F) e horas trabalhadas (H).

Nas fórmulas de solução da fsQCA¹¹, o resultado e as condições causais são representados por letras. As maiúsculas significam pertencimento e as minúsculas significam ausência ou baixo pertencimento. Então, para facilitar a interpretação, letras minúsculas serão lidas como alunos jovens (i), notas baixas (n), cursos diurnos (u), piores condições familiares (f) e poucas horas trabalhadas (h).

A matriz de suficiência e necessidade, que apresenta a proporção de sobreposição entre pares de conjuntos, é apresentada nas tabelas 5, 6, 7 e 8¹². Nestas é possível se observar que o fato de um curso ser noturno é o "mais suficiente" em termos de instituições privadas de ensino¹³.

9 De acordo com o critério *Screeplot*, apenas o primeiro componente deve ser considerado. Além disso, o segundo componente já apresenta um autovalor menor do que 1.

10 Este componente atingiu um autovalor de 1,96, que explica 65% da variância total dessas informações. Embora ainda haja 35% da variância dos dados não sendo considerada, como se trata de uma amostra grande, esse resultado é satisfatório.

11 Os conjuntos *crisp* (conjuntos binários) foram transformados em conjuntos *fuzzy*, pois este segundo fornece meios mais adequados de acomodar complementariedades, de tal forma que a adesão de um conjunto pode assumir qualquer valor entre 0 e 1.

12 O triângulo superior das matrizes representa a consistência; já o triângulo inferior, a cobertura.

13 De acordo com Kent e Olsen (2008), é desejável que se tenha alta cobertura (representada pela necessidade), e alta consistência (representada pela suficiência), mas essas duas medidas frequentemente são forças opostas; então, deve-se dar preferência para uma consistência mais elevada.

Tabela 5 – Matriz de Suficiência e Necessidade (2017)

	P	I	N	U	F	H
H	0,659	0,548	0,461	0,747	0,478	1,000
F	0,572	0,337	0,564	0,545	1,000	0,480
U	0,746	0,502	0,440	1,000	0,431	0,594
N	0,512	0,379	1,000	0,549	0,556	0,457
I	0,613	1,000	0,424	0,702	0,373	0,609
P	1,000	0,470	0,440	0,799	0,484	0,561

Fonte: Elaboração própria (2019).

Tabela 6 – Matriz de Suficiência e Necessidade, 18 a 30 anos (2017)

	P	I	N	U	F	H
H	0,661	0,470	0,457	0,737	0,481	1,000
F	0,546	0,357	0,557	0,501	1,000	0,445
U	0,744	0,427	0,430	1,000	0,412	0,561
N	0,485	0,349	1,000	0,507	0,541	0,410
I	0,557	1,000	0,442	0,638	0,439	0,534
P	1,000	0,391	0,431	0,780	0,472	0,527

Fonte: Elaboração própria (2019).

Tabela 7 – Matriz de Suficiência e Necessidade, 30 a 50 anos (2017)

	P	I	N	U	F	H
H	0,713	0,450	0,496	0,828	0,542	1,000
F	0,690	0,370	0,550	0,750	1,000	0,517
U	0,756	0,445	0,489	1,000	0,492	0,518
N	0,624	0,427	1,000	0,724	0,534	0,460
I	0,651	1,000	0,478	0,736	0,402	0,466
P	1,000	0,447	0,479	0,860	0,515	0,508

Fonte: Elaboração própria (2019).

Tabela 8 – Matriz de Suficiência e Necessidade, mais de 50 anos (2017)

	P	I	N	U	F	H
H	0,582	0,452	0,529	0,745	0,506	1,000
F	0,588	0,508	0,592	0,692	1,000	0,527
U	0,688	0,485	0,503	1,000	0,477	0,534
N	0,551	0,469	1,000	0,68	0,552	0,514
I	0,565	1,000	0,474	0,664	0,479	0,443
P	1,000	0,49	0,482	0,816	0,48	0,495

Fonte: Elaboração própria (2019).

Na tabela 9, a seguir, são apresentadas apenas as combinações lógicas, que exibem relações suficientes – e estatisticamente significativas –, com a determinação de uma faculdade ser privada ou pública, cujas consistências são superiores à 0,8¹⁴.

14 Como se está lidando com 6 variáveis, existem $2^6=64$ combinações lógicas possíveis, mas algumas delas não são observadas empiricamente e não possuem importância. Portanto, considera-se apenas combinações com relações de suficiência.

Tabela 9 – Combinações Lógicas Suficientes

Combinações lógicas	Consistência				Casos e Frequência		
	P	1-P	F	P-value	Frequência	Relativa	Acumulada
inUFh	0,821	0,179	36,350	0	12170	3,440	3,440
inUFH	0,850	0,150	313,090	0	15670	4,430	7,870
InUFH	0,847	0,153	253,670	0	15049	4,260	12,130
Cobertura	0,176						
Consistência	0,841						

Fonte: Elaboração própria (2019).

O fsQCA usa a "minimização booleana" para reduzir expressões primitivas e identificar as combinações mais gerais de condições suficientes para que o resultado permaneça logicamente verdadeiro. Com essa redução, tem-se duas configurações finais: $i*n*U*F$ e $n*U*F*H$. A primeira configuração diz que a combinação de estudantes jovens, notas baixas, curso noturno e melhores condições familiares possui uma relação com graduação em instituições privadas. A segunda configuração representa notas baixas, cursos noturnos, boas condições familiares e muitas horas trabalhadas como associados a cursos de instituições privadas. Notas baixas, boas condições familiares e cursos noturnos são condições necessárias para o resultado faculdades privadas.

A consistência dessas configurações lógicas é de 84%, e a cobertura de apenas 18%. Isso significa que a maior parte das faculdades privadas são independentes dessa configuração. Porém, a consistência é elevada, que é o principal critério de validação dos resultados¹⁵.

A grande quantidade de observações consideradas pode estar dificultando e comprometendo os resultados obtidos. Por isso, a análise foi repetida separando os alunos em 3 faixas etárias: 18 a 30 anos, 30 a 50 anos e acima de 50 anos. Com isso, além da redução do número de casos, sendo considerados em cada análise, essa nova situação permite que se verifique se existem diferenças de configurações entre essas faixas etárias. A seguir, são apresentados os índices família obtidos pelo ACP para as três faixas etárias:

$$\begin{aligned}
 18 \text{ a } 30 \text{ anos: } & \textit{família} = 0,59 \textit{ escol}_{\textit{pai}} + 0,60 \textit{ escol}_{\textit{mae}} + 0,55 \textit{ renda} \\
 30 \text{ a } 50 \text{ anos: } & \textit{família} = 0,62 \textit{ escol}_{\textit{pai}} + 0,63 \textit{ escol}_{\textit{mae}} + 0,47 \textit{ renda} \\
 \textit{mais de } 50: & \textit{família} = 0,64 \textit{ escol}_{\textit{pai}} + 0,64 \textit{ escol}_{\textit{mae}} + 0,44 \textit{ renda}
 \end{aligned}$$

Em cada faixa etária, o primeiro componente não está reproduzindo muito bem a variância total dos dados (principalmente para grupos de pessoas mais velhas). Portanto, a análise por fsQCA foi realizada de duas formas: uma considerando o índice família obtido pela ACP; outra considerando a variável escolaridade mãe¹⁶ no lugar do índice família. Porém, os resultados obtidos por essas duas formas foram bastante parecidos, tanto em termos de cobertura e consistência como da configuração final. Portanto, optou-se por apresentar os resultados, utilizando o índice família, para facilitar a comparação com o resultado obtido para a amostra total.

Na Tabela 10, são apresentadas as combinações lógicas para as três faixas etárias mencionadas. Ao considerar apenas alunos de 18 a 30 anos, os resultados obtidos foram muito parecidos com o que se observou, utilizando a amostra completa. Inclusive, com a "minimização booleana", chega-se à mesma configuração final: $i*n*U*F$ e $n*U*F*H$. Isso ocorre porque, de fato, essa faixa etária representa quase 78% da amostra total e, portanto, é a principal responsável pelos resultados obtidos nas etapas anteriores.

15 Além disso, foi feito o teste de relaxar a restrição de consistência de 0,8 para 0,7 e, neste caso, se teria uma cobertura de 56%.

16 Esta variável foi escolhida por ser mais comum na literatura se considerar variáveis da mãe quando se está interessado nas características dos alunos. Além disso, essa variável apresentou correlação mais elevada com a variável renda.

Tabela 10 – Combinações lógicas suficientes (dividido por faixas etárias)

Combinações lógicas	Consistência				Casos e Frequência		
	P	1-P	F	P-value	Frequência	Relativa	Acumulada
18 a 30 anos							
inUFh	0,833	0,167	71,86	0	9240	3,36	3,36
inUFH	0,857	0,143	302,92	0	11522	4,19	7,55
InUFH	0,838	0,162	107,47	0	10010	3,64	11,19
Cobertura	0,176						
Consistência	0,841						
30 a 50 anos							
inUFH	0,871	0,129	235,6	0	5255	6,27	6,27
InUFH	0,862	0,138	103,09	0	3207	3,83	10,1
Cobertura	0,135						
Consistência	0,868						
Mais de 50 anos							
inUFh	0,786	0,214	6,11	0,013	140	2,29	2,29
inUFH	0,788	0,212	11,28	0,001	241	3,94	6,23
Cobertura	0,096						
Consistência	0,787						

Fonte: Elaboração própria (2019).

A configuração final – minimizada para alunos de 30 a 50 anos – é $n*U*F*H$ (nota baixa, cursos noturnos, boas condições familiares e carga horária de trabalho elevada). Já para alunos acima de 50 anos é $i*n*U*F$ (mais novos – dentro da faixa etária –, notas baixas, cursos noturnos e boas condições familiares). Porém, para esses casos, a cobertura da QCA foi muito baixa, especialmente, quando se considera apenas alunos com mais de 50 anos. Soma-se o fato de que, para este último grupo, foi necessário relaxar a restrição de consistência de 0,80 para 0,70, pois, caso contrário, não seria possível chegar a nenhuma configuração final. O que acontece é que, como os alunos com essas idades são minorias, é de se esperar que existam situações e características muito peculiares à vida e preferência de cada um, o que dificulta a definição de um padrão consistente.

Seguindo a mesma ideia de análise, com foco na variável de ensino que determina se uma IES é pública ou privada, a última etapa deste estudo consistiu em realizar uma análise discriminante considerando a variável binária privada como dependente. Ou seja: considera-se que existe um grupo de faculdades públicas e um grupo de faculdades privadas que são diferentes um do outro. Então, busca-se verificar como os alunos seriam separados entre esses dois grupos.

Na Tabela 11 são apresentadas as médias de cada variável considerada na função discriminante logística, para cada um dos dois grupos. Observa-se que em ambos os períodos considerados (2005, 2010 e 2017), as escolas públicas apresentaram, na média, alunos mais novos, solteiros, que não trabalham e que estudaram em escolas particulares durante o Ensino Médio. Além disso, os alunos de instituições públicas apresentaram notas mais elevadas no exame ENADE e dedicaram mais tempo de seus dias ao estudo. A média de escolaridade dos pais é mais alta para os alunos de universidades públicas, mas, por outro lado, a renda média das famílias é mais elevada para estudantes de faculdades privadas. É interessante observar que o mesmo padrão de informações foi obtido nos três períodos.

Tabela 11 – Médias de cada grupo da Análise Discriminante

Média	2005			2010			2017		
	Pública	Privada	Total	Pública	Privada	Total	Pública	Privada	Total
Idade	25,818	26,966	26,569 7,856	22,325	24,768	24,162 6,256	26,878	27,597	27,295 6,965
Nota	39,659	37,064	37,963 14,698	45,849	39,642	41,182 14,443	46,982	42,446	44,353 13,939
Noite	0,479	0,813	0,697 0,460	0,155	0,477	0,397 0,489	0,374	0,799	0,620 0,485
Solteiro	0,758	0,691	0,714 0,452	0,928	0,803	0,834 0,372	0,810	0,737	0,768 0,422
Escolaridade do pai	2,188	2,075	2,114 1,240	2,884	2,483	2,583 1,311	2,420	2,359	2,385 1,331
Escolaridade da mãe	2,331	2,143	2,208 1,227	3,171	2,712	2,825 1,317	2,718	2,590	2,644 1,342
Renda	0,923	0,975	0,957 0,866	3,143	3,230	3,208 1,757	1,823	1,908	1,872 1,574
Escola pública	0,623	0,695	0,670 0,470	0,429	0,605	0,562 0,496	0,649	0,718	0,689 0,463
Horas de estudo	2,008	1,698	1,806 1,076	2,026	1,626	1,725 1,030	2,023	1,822	1,907 1,092
Horas trabalhadas	2,196	2,826	2,607 1,561	0,438	1,327	1,107 1,592	1,393	2,068	1,784 1,749
Bolsa de financiamento	0,170	0,430	0,340 0,474	0,062	0,410	0,324 0,468	0,070	0,643	0,402 0,490
Bolsa acadêmica							0,524	0,143	0,303 0,460
Observações	71122	134212	205334	55752	169080	224832	148701	204968	353669

Fonte: Elaboração própria (2019).

A Tabela 12 apresenta como as observações foram classificadas entre o grupo de universidades públicas *versus* privada a partir de uma função discriminante logística¹⁷. Para o ano de 2005, houve uma quantidade considerável de observações, que foram erradamente classificadas (32% das informações foram classificadas no grupo de privadas, sendo na verdade de instituições públicas, e 27% foram classificadas no grupo das públicas, sendo de privadas). Mas, ainda assim, a função discriminante classificou corretamente a maior parte das observações. Já para os anos de 2010 e 2017, a diferenciação entre os dois grupos em questão é mais significativa, pois a proporção de observações classificadas erradamente é quase sempre menor do que 20% (principalmente para o ano de 2017).

Até aqui, analisou-se diferenças entre instituições públicas e privadas de ensino. Porém, com o crescimento da oferta de bolsas de financiamento, a partir de 2005, tem-se uma situação que pode gerar um terceiro grupo. Se for considerado que a principal diferença entre universidades públicas e privadas é que uma é gratuita e a outra não, a possibilidade de obter uma bolsa de financiamento gera um grupo diferente, em que é possível se estudar em universidades privadas, mas sem o custo financeiro tradicional. Portanto, na Tabela 13 é apresentado uma nova classificação das observações entre os grupos, também pela função discriminante logística.

17 Foi considerada a opção de probabilidades iguais, *a priori* de pertencimento a cada grupo, por ter apresentado uma menor taxa de erro de classificação na maioria dos casos.

Tabela 12 – Classificação entre os grupos pela análise discriminante

Grupo original	2005			2010			2017		
	Pública	Privada	Total	Pública	Privada	Total	Pública	Privada	Total
Pública	47.815 67,23%	23.307 32,78%	71.122 100%	45.217 81,10%	10.535 18,90%	55.752 100%	127.467 85,72%	21.234 14,28%	148.701 100%
Privada	36.842 27,45%	97.370 72,55%	134.212 100%	48.411 28,63%	120.669 71,37%	169.080 100%	38.052 18,56%	166.916 81,44%	204.968 100%
Total	84.657 41,23%	120.677 58,77%	205.334 100%	93.628 21,64%	131.204 58,36%	224.832 100%	165.519 46,80%	18.815 53,20%	353.669 100%

Fonte: Elaboração própria (2019).

Tabela 13 – Análise Discriminante

Grupo original	2005				2010				2017			
	Pública	Privada s/ finas.	Privada c/ finan.	Total	Pública	Privada s/ finas.	Privada c/ finan.	Total	Pública	Privada s/ finas.	Privada c/ finan.	Total
Pública	31.551 53,12%	1.068 17,98%	17.164 28,90%	59.395 100%	36.894 70,58%	8.586 16,42%	6.794 13,00%	52.274 100%	95.979 69,39%	20.194 14,60%	22.152 16,01%	138.325 100%
Privada s/ finas.	15.547 20,13%	31.476 40,75%	30.225 39,13%	77.248 100%	28.909 29%	40.154 40,27%	30.735 30,735	99.701 100%	13.421 18,35%	37.113 50,73%	22.621 30,92%	73.155 100%
Privada c/ finan.	15.547 20,13%	31.476 40,75%	30.225 39,13%	77.248 100%	21.595 29,64%	16.756 23%	34.506 47,36%	72.857 100%	33.478 25,40%	30.408 23,07%	67.932 51,53%	131.818 100%
Total	58.364 30,04%	58.545 30,13%	77.398 39,83%	194.307 100%	87.398 38,87%	65.496 29,13%	71.938 32%	224.832 100%	142.878 41,62%	87.715 25,55%	112.705 32,83%	343.298 100%

Fonte: Elaboração própria (2019).

Nesta nova situação, a média das notas foi mais alta para universidades públicas, e mais baixa entre os alunos de universidades privadas sem financiamento. Já os cursos noturnos continuam sendo mais presentes, em média, nas instituições privadas (principalmente entre os alunos com financiamento). Além disso, os alunos que recebem financiamento para estudarem em universidades privadas, em média, são os que possuem renda familiar mais baixa e pais com menor nível de escolaridade, assim como discutido por Rossetto e Gonçalves (2015). Talvez, por isso, esses são os alunos que trabalham mais horas. Em contrapartida, alunos de faculdades privadas sem financiamento possuem, em média, uma renda familiar mais elevada.

Ao considerar três grupos diferentes na análise discriminante, as taxas de erro de classificação foram bastante elevadas, principalmente, quando se trata dos dois grupos referentes às instituições privadas (especialmente no ano de 2005). Entretanto, a quantidade de classificações erradas foi menor para os anos mais recentes. Tanto em 2010 como em 2017, aproximadamente 70% dos alunos de faculdades públicas foram corretamente classificados neste grupo.

5 CONSIDERAÇÕES FINAIS

A partir de diferentes técnicas de análise multivariada, foi possível perceber alguns padrões observados no perfil dos alunos matriculados em cursos de Ensino Superior nos anos de 2005, 2010 e 2017. As técnicas utilizadas se mostraram úteis na caracterização e interpretação dos resultados, embora existam alguns aspectos que necessitam de uma investigação mais cuidadosa e detalhada.

O principal resultado obtido por ambas as técnicas realizadas, foi a constatação de que, no ano mais recente desta pesquisa (2017), houve um maior acesso à cursos de Ensino Superior por parte de alunos nascidos em famílias mais pobres e com pais menos escolarizados. Essa situação é de

extrema importância, pois mostra que o país pode estar caminhando em direção a uma situação de maior igualdade social nos próximos anos. Inclusive, isso pode ser resultado das políticas educacionais adotadas, principalmente a partir de 2010, de maior acesso e expansão do Ensino Superior (a exemplo do ProUni e FIES).

Tabela 14 – Análise Discriminante

Média	2005				2010				2017			
	Classificação				Classificação				Classificação			
	Pública	Privada s/ finan.	Privada c/ finan.	Total	Pública	Privada s/ finan.	Privada c/ finan.	Total	Pública	Privada s/ finan.	Privada c/ finan.	Total
Idade	25,550	27,095	26,818	26,541 7,860	22,268	24,667	24,831	24,162 6,256	26,892	27,406	27,705	27,314 6,976
Nota	39,671	36,626	37,634	37,856 14,511	46,077	38,256	41,672	41,182 14,443	47,284	42,417	42,462	44,395 1,394
Noite	0,487	0,804	0,824	0,713 0,452	0,142	0,461	0,493	0,397 0,489	0,362	0,765	0,818	0,623 0,485
Solteiro	0,766	0,689	0,692	0,714 0,452	0,932	0,808	0,800	0,834 0,372	0,811	0,765	0,721	0,767 0,423
Esc. Pai	2,229	2,178	1,932	2,121 1,236	2,918	2,632	2,275	2,583 1,311	2,446	2,663	2,190	2,394 1,331
Esc. Mãe	2,373	2,228	2,024	2,212 1,224	3,205	2,833	2,543	2,825 1,317	2,744	2,863	2,438	2,652 1,341
Renda	0,949	1,103	0,800	0,966 0,865	3,172	3,529	2,795	3,208 1,757	1,871	2,565	1,543	1,893 1,578
Esc.pública	0,615	0,644	0,764	0,671 0,470	0,412	0,523	0,722	0,562 0,496	0,638	0,566	0,803	0,686 0,464
h estudo	1,989	1,664	1,742	1,787 1,069	2,042	1,585	1,690	1,725 1,030	2,025	1,787	1,841	1,904 1,092
h trabalhadas	2,190	2,806	2,849	2,630 1,561	0,400	1,268	1,393	1,107 1,592	1,394	2,105	2,047	1,797 1,752
bolsa acad.									0,516	0,104	0,165	0,293 0,455
Observações	59.395	77.248	57.664	194.307	52.274	99.701	72.857	224.832	138.325	73.155	131.818	343.298

Fonte: Elaboração própria (2019).

Ainda sobre a questão da renda, foi possível verificar pela fsQCA que o estudo em instituições privadas parece que ainda não é acessível a todas as classes sociais. Desse modo, em todas as situações analisadas, o aluno ser de famílias mais ricas foi uma característica importante para se verificar uma matrícula em curso de faculdade privada. Além disso, as tentativas de comparação entre as diferentes características dos alunos e cursos de universidades públicas e privadas, constatou-se que instituições privadas de ensino são mais presentes nos cursos noturnos e, em média, possuem alunos mais velhos e que trabalham ao mesmo tempo em que estudam.

Já em relação ao desempenho dos alunos, ficou evidente em todos os anos que, em média, alunos de IES públicas apresentaram um melhor rendimento no exame e informaram que se dedicam mais horas ao estudo.

O sistema educacional de um país é fundamental para seu desenvolvimento. Este trabalho teve o papel de contribuir com discussões importantes dentro deste tema, porém, existe ainda uma série de situações, que valem ser estudadas mais detalhadamente em pesquisas futuras.

Este trabalho apresentou uma ideia inicial de que políticas de apoio a alunos que desejam obter um curso superior, como bolsas de financiamento e aumento na oferta de vagas de universidades públicas, foram importantes e capazes de proporcionar boas oportunidades às pessoas de famílias

mais carentes. Portanto, é fundamental que se avance em pesquisas mais minuciosas sobre esse tipo de política.

Uma análise nesse sentido pode ser realizada também para a questão das bolsas acadêmicas, como bolsas de iniciação científica e monitoria. Por ser uma variável mais nova no banco de dados, essa observação só pôde ser observada para o ano de 2017. Todavia, o que se verificou foi que alunos que possuem esse tipo de bolsa apresentaram um rendimento significativamente maior no exame prestado. Esse tipo de relação pode ser melhor analisado, considerando apenas anos mais recentes.

REFERÊNCIAS

ALBERT, A.; LESAFFRE, E. Multiple group logistic discrimination. In: **Statistical Methods of Discrimination and Classification**. [S.l.]: Elsevier. p. 209-224, 1986.

BERG-SCHLOSSER, D. *et al.* Qualitative comparative analysis (qca) as an approach. In: **Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques**, Sage Thousand Oaks, CA, v. 1, p. 18, 2009.

DEARDEN, L.; FITZSIMONS, E.; WYNESS, G. Money for nothing: Estimating the impact of student aid on participation in higher education. In: **Economics of Education Review**, 43, 66-78, 2014. Disponível em: <https://doi.org/10.1016/j.econedurev.2014.09.005> .

DEL REY, E.; ESTEVAN, F. Assessing Higher Education Policy in Brazil: A Mixed Oligopoly Approach. B.E. In: **Journal of Economic Analysis and Policy**, v. 20, n. 1, p. 1-16, 2020. Disponível em: <https://doi.org/10.1515/bejeap-2019-0240>.

FARRELL, W. C. *et al.* Assessing college aspirations among at-risk high school students: A principal component analysis. In: **The High School Journal**, JSTOR, v. 77, n. 4, p. 294-303, 1994.

FLANNERY, D.; O'DONOGHUE, C. The demand for higher education: A static structural approach accounting for individual heterogeneity and nesting patterns. In: **Economics of Education Review**. 34, 243-257, 2013. Disponível em: <https://doi.org/10.1016/j.econedurev.2012.12.001>

GREENACRE, M. Correspondence analysis of square asymmetric matrices. In: **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 49, n. 3, p. 297-310, 2000.

GREENACRE, M.; BLASIUS, J. **Multiple correspondence analysis and related methods**. [S.l.]: Chapman and Hall/CRC, 2006.

HEINESEN, E. Admission to higher education programmes and student educational outcomes and earnings—Evidence from Denmark. In: **Economics of Education Review**, v. n. 63, p. 1-19, 2018. Disponível em: <https://doi.org/10.1016/j.econedurev.2018.01.002> .

JOHNSON, R. A.; WICHERN, D. W. *et al.* **Applied multivariate statistical analysis**. [S.l.]: Prentice hall Upper Saddle River, NJ, 2002.

LAM, P.L., SHIU, A. Economic growth, telecommunications development and productivity growth of the telecommunications sector: Evidence around the world. In: **Telecommunications Policy**. 34, 185-199, 2010. Disponível em: <https://doi.org/10.1016/j.telpol.2009.12.001>

LEGEWIE, N. An introduction to applied data analysis with qualitative comparative analysis. In: **Forum Qualitative Sozialforschung/Forum: Qualitative Social Research**. 2013.

- LONG, N. VAN. Financing higher education in an imperfect world. In: **Economics of Education Review**. 71, 23-31, 2019. Disponível em: <https://doi.org/10.1016/j.econedurev.2018.06.004>
- KENT, R.; OLSEN, W. **Using fsqca a brief guide and workshop for fuzzy-set qualitative comparative analysis**. Citeseer, 2008.
- MALLIDIS, I.; DEKKER, R.; VLACHOS, D. The impact of greening on supply chain design and cost: A case for a developing region. In: **Journal of Transport Geography**, 22, 118-128, 2012. Disponível em: <https://doi.org/10.1016/j.jtrangeo.2011.12.007>.
- MARCHESAN, T. M.; SOUZA, A. M.; MENEZES, R. Avaliação do processo de ensino: uma abordagem multivariada. In: **Produção, São Paulo**, SciELO Brasil, v. 21, n. 2, 2011.
- MARQUES, A. F. Aplicação da análise multivariada na infraestrutura e no desempenho das escolas públicas do Ensino Fundamental e Médio pertencentes ao núcleo regional de educação de Paranavaí. In: **Acta Scientiarum. Technology, Universidade Estadual de Maringá**, v. 32, n. 1, 2010.
- MAZZAROL, T. W.; SOUTAR, G. N. Australian educational institutions' international markets: a correspondence analysis. In: **International Journal of Educational Management**, Emerald Group Publishing Limited, v. 22, n. 3, p. 229-238, 2008.
- OPPEDISANO, V. The (adverse) effects of expanding higher education: Evidence from Italy. In: **Economics of Education Review**, v. 30, n. 5, p. 997-1008, 2011. Disponível em: <https://doi.org/10.1016/j.econedurev.2011.04.010> .
- PARKER, A. A study of variables that predict dropout from distance education. In: **International journal of educational technology**, v. 1, n. 2, p. 1-10, 1999.
- PINTO, J. O acesso à Educação Superior no Brasil. In: **Educação & Sociedade**. 25, 727-756, 2004. Disponível em: <https://doi.org/10.1590/s0101-73302004000300005>.
- ROJAS, E.; SÁNCHEZ, R.; VILLENNA, M. Credit constraints in higher education in a context of unobserved heterogeneity. In: **Economics of Education Review**, v. 52, p. 225-250, 2016. Disponível em: <https://doi.org/10.1016/j.econedurev.2016.03.004> .
- SCHWARTZMAN, S. Perspectivas Para a Educação Superior No Brasil. In: **Desafios da Nação: artigos de apoio**, p. 333-353, 2015. Disponível em: http://ipea.gov.br/agencia/images/stories/PDFs/livros/livros/180413_desafios_da_nacao_artigos_vol2_cap25.pdf .
- SENHORAS, E.M.; TAKEUCHI, K. P.; TAKEUCHI, K. P. A análise estrutural do Ensino Superior privado sob perspectiva. In: **Simpósio de Excelência em Gestão e Tecnologia, III SEGeT**, 2006.
- SHANKA, T.; QUINTAL, V.; TAYLOR, R. Factors influencing international students' choice of an education destination-a correspondence analysis. In: **Journal of Marketing for Higher Education**, Taylor & Francis, v. 15, n. 2, p. 31-46, 2006.
- SOUTAR, G. N.; TURNER, J. P. Students' preferences for university: a conjoint analysis. In: **International journal of educational management**, MCB UP Ltd, v. 16, n. 1, p. 40-45, 2002.
- WEBSTER, T. J. A principal component analysis of the us news & world report tier rankings of colleges and universities. In: **Economics of Education Review**, Elsevier, v. 20, n. 3, p. 235-244, 2001.

YAVAS, U.; SHEMWELL, D. J. Graphical representation of university image: A correspondence analysis journal of marketing for higher education. In: **Journal of marketing for Higher Education**, Taylor & Francis Group, v. 7, n. 2, p. 75-84, 1996.

YEUNG, A. S.; LEE, F. L. Self-concept of high school students in china: Confirmatory factor analysis of longitudinal data. In: **Educational and Psychological Measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 59, n. 3, p. 431-450, 1999.

ZOGHBI, A. C.; ROCHA, F.; MATTOS, E. Education production efficiency: Evidence from Brazilian universities. In: **Economic Modelling**, v. 31, n. 1, 94-103, 2013. Disponível em: <https://doi.org/10.1016/j.econmod.2012.11.018>.