

Guilherme Souza

**Variáveis determinantes para formação do  
conceito preliminar de curso nas avaliações do  
Enade**

UNIVERSIDADE  
FEDERAL  
FLUMINENSE

Niterói

2018

Guilherme Souza

## **Variáveis determinantes para formação do conceito preliminar de curso nas avaliações do Enade**

Monografia apresentada ao Curso de Administração da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Bacharel em Administração.

Universidade Federal Fluminense – UFF  
Faculdade de Administração e Ciências Contábeis – EST  
Departamento de Administração – STA  
Curso de Graduação em Administração

Orientador: Prof. Dr. Ariel Levy

Niterói  
2018

Ficha catalográfica automática - SDC/BAC

S719v Souza, Guilherme  
Variáveis determinantes para formação do conceito preliminar de curso nas avaliações do Enade / Guilherme Souza ; Ariel Levy, orientador. Niterói, 2018.  
59 f.

Trabalho de Conclusão de Curso (Graduação em Administração)-Universidade Federal Fluminense, Faculdade de Administração e Ciências Contábeis, Niterói, 2018.

1. Enade. 2. Conceito Preliminar de Curso. 3. Qualidade do ensino superior. 4. Avaliação de curso de ensino superior. 5. Produção intelectual. I. Título II. Levy, Ariel, orientador. III. Universidade Federal Fluminense. Faculdade de Administração e Ciências Contábeis. Departamento de Administração.

CDD -

Guilherme Souza

## **Variáveis determinantes para formação do conceito preliminar de curso nas avaliações do Enade**

Monografia apresentada ao Curso de Administração da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Bacharel em Administração.

Banca Examinadora. Niterói, Julho de 2018:

---

**Prof. Dr. Ariel Levy**

Orientador

Universidade Federal Fluminense

---

**Prof. Dr. Eduardo Camilo da Silva**

Universidade Federal Fluminense

---

**Prof. Dr. Joel de Lima Pereira Castro  
Júnior**

Universidade Federal Fluminense

Niterói

2018

# Agradecimentos

Agradeço a todas as pessoas que contribuíram, de alguma forma, para minha formação e para que eu esteja nesse momento redigindo este trabalho. Agradeço a todos os membros da minha família e a todos os integrantes da Universidade Federal Fluminense e colegas que fiz nesta instituição.

Em especial, agradeço:

- i. A Deus;
- ii. A minha Mãe, Maria, por ter me ensinado valores e me oferecido todo suporte para educação;
- iii. Ao meu orientador e grande amigo, Ariel Levy, cujo conhecimento é inspirador e por ter sido referência às minhas escolhas de aprendizagem.

## Lista de ilustrações

Figura 1 – Percentual de cursos por organização acadêmica . . . . .	30
Figura 2 – Percentual de cursos por categoria administrativa . . . . .	31
Figura 3 – Cursos com conceito . . . . .	32
Figura 4 – Matriz correlograma: correlação entre os pares de variáveis segundo uma escala de cor. . . . .	37
Figura 5 – Matriz correlograma com as duas bandas . . . . .	38
Figura 6 – Matriz de pares de variáveis . . . . .	39
Figura 7 – Critério de seleção dos componentes: a partir do 5º componente, observa- se redução no decaimento da variabilidade, o que diminui a importância do componente para uma possível seleção. . . . .	43
Figura 8 – Matriz correlograma . . . . .	47
Figura 9 – Critério de seleção dos fatores: a partir do 5º fator, observa-se uma tendência de horizontalização da curva de variabilidade explicada pelos fatores, o que torna os fatores a partir desta contagem menos suscetíveis em participar de uma seleção. . . . .	50

## Lista de tabelas

Tabela 1 – Cursos por região . . . . .	29
Tabela 2 – Cursos com conceito . . . . .	32
Tabela 3 – Sumarização da base de dados após uso do método $k$ -NN: medidas estatísticas básicas . . . . .	34
Tabela 4 – Estatísticas sobre as variáveis padronizadas pelo INEP . . . . .	36
Tabela 5 – Características dos Componentes . . . . .	44
Tabela 6 – Componentes com magnitudes $\pm .50$ . . . . .	45

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	Motivação	12
1.2	Justificativa	13
1.3	Estrutura do Trabalho	15
<b>2</b>	<b>AVALIAÇÃO DO ENSINO SUPERIOR NO</b>	
	<b>BRASIL</b>	<b>17</b>
2.1	Breve panorama evolutivo das metodologias de avaliação	17
2.2	Dificuldades e necessidades em avaliar IES	19
<b>3</b>	<b>AVALIAÇÃO DO ENSINO SUPERIOR NO</b>	
	<b>MUNDO</b>	<b>22</b>
<b>4</b>	<b>METODOLOGIA</b>	<b>23</b>
4.1	Classificação da Pesquisa	23
4.2	Ferramenta e Técnicas de análise de dados	24
4.3	Análise Exploratória de dados	24
4.4	Análise do Componente Principal	35
4.4.1	Padronização das variáveis	35
4.4.2	Matriz de correlação das variáveis preditoras	36
4.4.3	Aplicação da Análise do Componente Principal	39
4.4.4	Componentes Principais	40
4.4.5	Seleção dos Componentes	42
4.4.6	Características dos Componentes	43
4.5	Análise do Fator	46
4.5.1	Rotação dos Fatores	50
<b>5</b>	<b>RESULTADOS</b>	<b>53</b>
<b>6</b>	<b>CONCLUSÃO</b>	<b>55</b>
	<b>REFERÊNCIAS</b>	<b>57</b>



## Resumo

O desenvolvimento de modelos de mensuração da qualidade de cursos superiores e instituições ainda é assunto bastante discutido na literatura internacional dada a diversidade de dimensões e interesses de avaliação. Organismos multilaterais têm estimulado a avaliação como instrumento de aprimoramento e controle social. No que se refere às políticas públicas do Brasil, houve um considerável aumento de gastos em ensino, sobretudo na última década. Nacionalmente, uma das métricas de mensuração da qualidade do ensino superior e das instituições é o Enade, com provas realizadas em ciclos trienais de avaliação. Este trabalho tem como principal objetivo investigar os dados de avaliação de 2015 disponibilizados pelo INEP e verificar a composição do Conceito Preliminar de Cursos (CPC) como indicador de qualidade. Com este intuito foram utilizadas três técnicas estatísticas: Análise Exploratória de Dados, Análise do Componente Principal e Análise do Fator. Para diferentes configurações de amostragem, os resultados indicam que a Organização Didático-pedagógica, Infraestrutura e Instalações e Oportunidades de Ampliação da Formação são as variáveis que mais exercem influência na composição do CPC.

**Palavras-chave:** Enade, Conceito Preliminar de Cursos, Qualidade do ensino superior.

## Abstract

The development of models for measuring the quality of higher education and institutions is still very much discussed in the international literature given the multiplicity of evaluation optics and interests. Multilateral organizations have stimulated evaluation as an instrument of improvement and social control. With regard to public policies in Brazil, there has been a considerable increase in education spending, especially in the last decade. Nationally, one of the higher education quality measurement metrics is Enade with its trienal exams. The main objective of this study is to investigate the evaluation data of 2015 made available by INEP and verify if the Preliminary Course Concept is an indicator of quality. In this regard, three statistical methodologies has been used: Exploratory Data Analysis, Principal Component Analysis and Factor Analysis. Through different sampling arrangements, results indicate that Didactic pedagogical Organization, Infrastructure and Facilities and Opportu-

nities for Expansion of Training are variables that influence the most in the composition of the Preliminary Course Concept.

**Keywords:** Enade, Preliminar Course Concept, Higher education quality.

## Lista de abreviaturas e siglas

**CI:** Concluintes Inscritos;

**CP:** Concluintes Participantes

**NBFG:** Nota Bruta - FG (Formação Geral)

**NBCE:** Nota Bruta - CE (Conhecimentos Específicos)

**NBG:** Nota Bruta - Geral

**NCE:** Nota Contínua do Enade

**NBODP:** Nota Bruta - Organização Didático-Pedagógica

**NPODP:** Nota Padronizada - Organização Didático-Pedagógica

**NBIIF:** Nota Bruta - Infraestrutura e Instalações Físicas

**NPIIF:** Nota Padronizada - Infraestrutura e Instalações Físicas

**NBOAF:** Nota Bruta - Oportunidades de Ampliação da Formação

**NPOAF:** Nota Padronizada - Oportunidades de Ampliação da Formação

**CPNE:** Concluintes Participantes com nota no Enem

**PCPNE:** Percentual de Concluintes participantes com nota no Enem

**NBIDD:** Nota Bruta - IDD

**NPIDD:** Nota Padronizada - IDD

**ND:** Número de Docentes

**NBM:** Nota Bruta - Mestres

**NPM:** Nota Padronizada - Mestres

**NBD:** Nota Bruta - Doutores

**NPD:** Nota Padronizada - Doutores

**NBRT:** Nota Bruta - Regime de Trabalho

**NPRT:** Nota Padronizada - Regime de Trabalho

**CPCC:** CPC Contínuo

**CPCF:** CPC Faixa

**INEP:** Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

**AF:** Análise do Fator

**AED:** Análise Exploratória de Dados

**ACP:** Análise do Componente Principal

# 1 INTRODUÇÃO

## 1.1 Motivação

É notável que educação se tornou um tema recorrente nos meios de comunicação, sobretudo recentemente com o aparente aumento de demanda por qualificação. Essa crescente demanda por educação é impulsionada pelos constantes avanços de base tecnológica. De fato, este progresso vem acompanhado por um lado da ameaça à mão de obra tradicional e por outro, cada vez mais, das exigências por profissionais mais capacitados. Neste sentido, para aumentar as condições de acesso aos cargos do futuro, é essencial que seja dada a devida atenção para as desigualdades no acesso ao aprendizado ([OCDE, 2017](#), p. 45).

Educação não gera apenas intelecto, bons empregos e remuneração associada. Sabe-se que quanto mais o indivíduo se qualifica, maior é o seu engajamento social e melhor é a sua saúde. É fundamental, portanto que os governos tenham conhecimento desta associação e promovam ações que possibilitem acesso facilitado às fontes de qualificação que deem suporte a essa expansão ([OCDE, 2017](#), p. 45).

Como importantes fontes de mão de obra para o mercado de trabalho, as universidades tem importância significativa para o desenvolvimento de uma nação. É por meio delas que são forjadas as competências profissionais e ocorre a transferência de tecnologia e o surgimento de novas. Além das competências específicas para atuação no mercado de trabalho, as universidades desempenham papel fundamental na formação dos valores individuais. Nesse sentido, a prosperidade de uma sociedade e o aprimoramento do ensino superior são temas indissociáveis.

Percebe-se que, especificamente no caso brasileiro, o tema educação vem recebendo cada vez mais atenção governamental em termos de alocação de recursos. De acordo com [Mendes \(2015, p. 1\)](#), esta área passou a absorver parcelas significativas de destinações orçamentárias, sobretudo na última década. De acordo com o autor, estas receitas passaram de 4% em 2004 para 9,3% em 2014, um acréscimo de 130% no período. Isso revela uma importante transição da política educacional no país.

Entende-se que o expressivo aumento na destinação de recursos está associado à influência do governo federal na criação de vagas. De acordo com [Zoghbi, Rocha e Mattos \(2013, p. 2\)](#)

isso ocorreu por intermédio da expansão de organizações de ensino privadas e pela criação do programa Universidade para Todos (ProUni) em 2004 pelo Ministério da Educação (MEC).

Com a expansão da oferta, foram aprimorados mecanismos de mensuração da qualidade dos cursos prestados por instituições de ensino superior (IES) públicas e privadas. Com isso, é natural que a análise do desempenho e da qualidade do ensino superior tenha se tornado uma importante área de pesquisas com a utilização diversas metodologias quantitativas e qualitativas. Isso decorre da necessidade de mensurar a eficiência do gasto público em educação, principalmente por intermédio de indicadores, como a proporção de alunos por professor e o custo por estudante, por exemplo (COSTA et al., 2012, p. 417).

O grande obstáculo em mensurar a qualidade de cursos ou instituições de ensino superior reside na dificuldade em determinar as características e dimensões importantes para gerar resultados (ou produtos). Neste raciocínio, Costa et al. (2012, 417) classificam características inerentes ao setor de produção educacional. São elas:

- i) a natureza múltipla e intangível do produto: produtos podem ser conhecimento e habilidades, valores, atitudes, entre outros;
- ii) a participação do cliente no processo produtivo: o aluno tem participação ativa no processo produtivo e não apenas demanda a “mercadoria”;
- iii) a heterogeneidade dos serviços: uma vez que o estudante atua no processo produtivo, as unidades produtivas se diferenciam umas das outras;
- iv) a dimensão temporal: é necessário observar a trajetória completa da vida dos estudantes para mensurar a eficiência do processo produtivo;
- v) o caráter acumulativo do ensino;
- vi) a incidência de fatores exógenos: como experiências adquiridas fora do setor educacional.

## 1.2 Justificativa

A complexidade na avaliação da qualidade dos cursos superiores reside na diversidade de dimensões subjetivas dos participantes, inerente ao processo de educação. Apesar disso, a

avaliação é um componente fundamental pois é capaz de determinar, por exemplo, a sensibilidade dos investimentos em relação aos retornos para a sociedade. Com esse intuito, a formulação de indicadores é central para a tomada de decisão em diversos níveis. Com base nos indicadores, políticas de alocação de recursos podem ser desenvolvidas. Da mesma forma, gestores com atuação direta em instituições de ensino podem fundamentar suas decisões com base em dados concretos.

No Brasil, um dos principais instrumentos de avaliação dos cursos superiores é o Exame Nacional de Desempenhos dos Estudantes de ensino Superior (Enade). Este exame é um dos componentes do Sistema Nacional de Avaliação da Educação Superior SINAES estabelecido na Lei 1086/04 de 14 de Abril de 2004 cujas principais diretrizes, de acordo com Brito e Regina (2008, p. 842), são:

- i) Avaliação Institucional: auto-avaliação (pelas CPAs<sup>1</sup> e avaliação externa in loco, desenvolvida pelos avaliadores institucionais capacitados pelo INEP nos moldes do SINAES.
- ii) Avaliação de Curso: pelos pares na avaliação in loco, pelos estudantes, através do ADES (questionário de Avaliação Discente da Educação Superior que é enviado aos estudantes da amostra do ENADE), pelos coordenadores de curso, mediante questionário dos coordenadores e avaliações realizadas pelos professores dos cursos e a CPA.
- iii) Avaliação do Desempenho dos estudantes ingressantes e concluintes: através de um exame em larga escala aplicado aos estudantes que preenchem os critérios estabelecidos pela legislação vigente.

Ainda de acordo com a autora, o SINAES se estabelece como elemento norteador das políticas de educação superior no Brasil cujo principal foco é avaliar a situação dos cursos de graduação. A avaliação pelo SINAES tem como intuito:

- i) identificar as condições de ensino oferecidas aos estudantes, em especial as relativas à organização didático pedagógica, corpo social e instalações físicas;

---

<sup>1</sup> Comissão Própria de Avaliação. Trata-se de uma comissão responsável pelo processo de avaliação interno da instituição. Prestam as informações diretamente ao Inep. Visa garantir a participação da sociedade e da comunidade universitária. (INEP, 2018, p. 26)

- ii) verificar a articulação entre PDI (Plano de Desenvolvimento Institucional), Projeto Pedagógico de Curso - PPC, currículo, vocação institucional e inserção regional;
- iii) analisar a aderência às Diretrizes Curriculares Nacionais – DCN's.

A prática de ranqueamento de instituições com base em pontuação no Enade parece ser uma prática comum, principalmente nos mecanismos de marketing de instituições de ensino particulares. Percebe-se que isso é algo que se encontra presente desde que o INEP iniciou os ciclos de avaliação ([VERHINE; DANTAS; SOARES, 2006](#), p. 294). De fato, existe um grande gosto popular por tais comparações com base em disputa por posições, embora de nada sirvam para mensurar a capacidade de agregação de habilidades e competências dos estudantes nos cursos. ([BRITO; REGINA, 2008](#), p. 847).

A amplitude de variáveis analisadas pelo Enade, à primeira vista, parece ser capaz de aferir com precisão aspectos evolutivos de componentes individuais dos cursos de graduação. Contudo, uma interpretação mais atenta das diretrizes estabelecidas pelo SINAES leva a entender que evoluções em itens específicos estão aquém dos seus reais objetivos. Há que se investigar a relação entre a diversidade destas variáveis com o vetor qualidade. Para isso é necessário que se racionalize o universo de variáveis em um conjunto delimitado de dimensões.

Diante do exposto, este trabalho é resultado da tentativa de análise das variáveis que constituem o Enade por meio da redução das dimensionalidades. Acredita-se que seja possível extrair perfis de variáveis representativos dos cursos superiores capazes de descrevê-los em arranjos significativos de fácil compreensão.

O propósito deste trabalho consiste em identificar quais variáveis são determinantes na avaliação dos cursos superiores pelo Enade e com isso desenvolver uma melhor compreensão sobre as avaliações do INEP e dos cursos. Procurou-se alcançar esta finalidade por meio da aplicação de técnicas estatísticas de descrição de dados (Análise Exploratória de Dados), determinação de componentes (Análise dos Componentes Principais) e fatores (Análise de Fator).

### **1.3 Estrutura do Trabalho**

Esta pesquisa encontra-se organizada em seções. A presente seção apresenta a introdução. Na segunda (Avaliação do Ensino Superior no Brasil), será feito um panorama histórico dos



principais instrumentos de avaliação do nível superior no Brasil. A terceira seção (Avaliação do ensino superior no mundo) é destinada a apresentar as práticas internacionais de avaliação bem como a situação dos outros países perante o Brasil. A quarta seção é destinada ao desenvolvimento da Metodologia. Por fim, as seções 5 e 6 apresentam, respectivamente, os resultados e conclusão provenientes desta pesquisa.

## 2 AVALIAÇÃO DO ENSINO SUPERIOR NO BRASIL

Na literatura em torno da avaliação do ensino superior no Brasil é comum constatar que diversos autores destacam a complexidade deste sistema e os diversos mecanismos que orbitam ao seu redor. Não obstante, o Ministério da Educação por meio do INEP publica com relativa frequência as notas técnicas e portarias onde são descritas de forma esmiuçada novas metodologias e mudanças aplicadas na avaliação do ensino superior. O que se observa é uma permanente tentativa do governo em aprimorar os critérios e métricas em vias de racionalizar a mensuração da qualidade do ensino superior no país. Nessa perspectiva, o objetivo desta sessão é apresentar uma visão dos principais aspectos da sistemática de avaliação das instituições e cursos superiores pela abordagem das principais dificuldades na formulação de indicadores e métricas, bem como um panorama evolutivo dos processos de avaliação.

### 2.1 Breve panorama evolutivo das metodologias de avaliação

[Barbosa, Freire e Crisóstomo \(2011, p. 318\)](#) classificam a avaliação das instituições de ensino superior (IES) em externa e interna e evidenciam que estas têm estado no centro dos debates tanto nacional quanto internacionalmente. Tal fato tem relação com a percebida necessidade de otimização de recursos humanos e materiais das universidades. Evidentemente, os processos de avaliação fazem uso extensivo de indicadores. De acordo com esses autores, apesar de serem muito criticados no Brasil, são instrumentos de firme aceitação pelos governos que parecem não abandoná-los. *Ibid.*, p. 319.

[Zandavalli \(2009, p. 385\)](#) aponta a influência do banco mundial no desenvolvimento de políticas públicas que subsidiam o entendimento do sistema de avaliação da educação superior. De forma geral, é possível constatar pela análise da literatura disponível, os instrumentos de avaliação em sua concepção e finalidade, têm certa influência estrangeira. Como esse autores de forma bem íntegra analisam, as movimentações iniciais pela reestruturação do ensino superior no Brasil coincidem com as mobilizações de universitários contrários ao regime militar de 1960. *Ibid.*, p. 386.

Neste mesmo ano tem grande destaque o plano Atcon<sup>2</sup> como um dos primeiros processos voltados a avaliar a estrutura das universidades brasileiras com a proposta de associar investimentos e resultados (BARBOSA; FREIRE; CRISÓSTOMO, 2011, p. 320). Esse plano se estabeleceu como um marco de primeira prática de avaliação do ensino durante o regime militar pela adequação da educação superior ao modelo econômico capitalista (ZANDAVALLI, 2009, 389). As instituições brasileiras foram satisfatoriamente beneficiadas pelos estudos e propostas do Plano Atcon. Uma de suas recomendações que mais repercutiram a sua tentativa de gerar uma correspondência entre investimentos e resultados parece ser o quadro-organograma e os *decision-makers* da “universidade empresa”. Ibid., p. 390.

A década de 1970 foi marcada em sua maior parte pela operacionalização progressiva dos planos traçados pelo Plano Atcon em 60. Em vista disso, esse período se mostrou relativamente apagado em termos de novas propostas de reformulação. O Plano Atcon, embora seja um marco pela sua importância na concretização de um considerável avanço em termos de adequação da estrutura do ensino no Brasil, carecia do envolvimento e diálogo com os professores e a sociedade civil. Ibid., p. 401. Em decorrência disso, em 1980 houve a implantação do Programa de Avaliação da Reforma Universitária (PARU), onde passou-se a considerar as propostas da sociedade e dos discentes.

De acordo com Zandavalli (2009, p. 402), o PARU contou com uma participação mais significativa de pesquisadores e professores universitários quando de sua criação e aplicação – condições fundamentais para reformulação da educação superior. O autor destaca que as principais demandas situavam-se em torno da democratização e acesso ao ensino pelas diferentes classes da sociedade e nos princípios básicos das universidades: formação de profissionais, produção e disseminação de conhecimentos. (JUNIOR, 2004 apud ZANDAVALLI, 2009, p. 403).

A avaliação do ensino superior volta ao centro do cenário político na década de 80, no governo Sarney, momento em que a grande escassez de recursos públicos impulsionava os debates na Comissão Nacional para Reformulação do Ensino Superior (CNRES) (Decreto

---

<sup>2</sup> Rudolph Atcon foi um consultor norte americano responsável pelo estudo para reformulação das estruturas das universidades brasileiras. Seu trabalho se deu com base em visita a 12 universidades e a procura por fatores que refletissem a perspectiva de modernização. Seu estudo se deu com base nos pressupostos americanos de racionalidade, eficiência e eficácia das instituições. Deu origem ao documento “Rumo à Reformulação Estrutural da Universidade Brasileira” com as propostas para reformulação do ensino superior no país. (ZANDAVALLI, 2009, 387)

no 91.117/1985) e do Grupo Executivo para a Reformulação do Ensino Superior (GERES) (Portaria no 100/1986). As propostas levantadas por esses dois grupos de trabalho apresentam pontos em comum em relação às instituições de ensino superior (IES): implantação de sistemas de avaliação e valorização do desempenho de universidades e necessidade de melhoria no processo de financiamento pelo aprimoramento da gestão do uso de recursos públicos. (BARBOSA; FREIRE; CRISÓSTOMO, 2011, p. 321).

A partir de então, o estabelecimento de estruturas de controle baseadas em indicadores tornou-se uma necessidade da gestão pública. Contudo, apenas em 1993 é criado o Programa de Avaliação das Universidades Brasileiras (PAIUB) que introduzia, dentre outras, as propostas de indicadores de avaliação das IES; verificação em nível nacional; carência de eficiência no financiamento e de prestação de contas aos financiadores. O programa também aponta a necessidade de verificação de dimensões quantitativas e qualitativas em consonância com a frequência de processos de verificação de desempenho. É reforçada a necessidade de criação de um sistema de controle de avaliação de desempenho institucional e a qualidade dos cursos coordenado pela Secretaria de Educação Superior (SESu) do Ministério da Educação – MEC e envolvimento da comunidade acadêmica. Ibid., p. 322.

De acordo com Barbosa, Freire e Crisóstomo (2011), neste mesmo período, o estabelecimento dos indicadores das IES se deu paralelamente a caracterização das dimensões das universidades, quais sejam avaliação de cursos, avaliação de alunos, avaliação de professores, avaliação didático-pedagógica do ensino, avaliação de servidores técnico-administrativos e avaliação das carreiras. O critério de organização da avaliação com base em dimensões é um dos fundamentos do Sistema Nacional de Avaliação da Educação Superior (SINAES – Lei 10.861/2004) como será visto mais à adiante.

## 2.2 Dificuldades e necessidades em avaliar IES

A avaliação da qualidade de cursos superiores ou instituições de ensino é um processo inerentemente difícil pela diversidade de perspectivas e interesses. De forma ilustrativa, o gestor de recursos públicos poderia voltar sua atenção ao custo por aluno. O estudante, por outro lado, poderá valorizar mais a empregabilidade após formado – dimensão esta quase sempre negligenciada pelo primeiro observador. O professor universitário, por sua vez, pode-

ria voltar sua preocupação a quantidade de alunos em sala de aula ou mesmo às condições de estabilidade.

Como é possível observar, a aplicação de mecanismos de avaliação capazes de aferir de forma eficiente a qualidade das IES constitui tarefa trabalhosa. De acordo com [Barbosa, Freire e Crisóstomo \(2011, p. 322\)](#) o tema está em constante discussão no Brasil. Para estes autores, outro fator que contribui para a dificuldade do processo de avaliação é a distinção entre a avaliação do ensino superior e avaliação de instituições públicas. A primeira é aplicável tanto às instituições públicas quanto privadas. A avaliação de instituições públicas, por sua vez, deve se preocupar, concomitantemente à qualidade do ensino, a eficiência na alocação dos recursos públicos, uma vez que necessita realizar a prestação de contas para com a sociedade.

Historicamente no Brasil há uma concentração maior do ensino superior nas instituições de ensino públicas ([BARBOSA; FREIRE; CRISÓSTOMO, 2011](#)), realidade que pode estar relacionada à inexistência de pagamento de mensalidades pelos alunos e a proliferação de instituições particulares com cursos de baixa qualidade. Isso gera necessidade ao governo de prestar atenção especial às políticas de ampliação de acesso ao ensino superior bem como criar mecanismos de avaliação, auditoria e gestão das IES de maneira geral.

A subjetividade envolvida no processo de avaliação de IES torna necessária a utilização de sistemas de avaliação de desempenho aplicados nacionalmente. O desempenho alcançado em tais avaliações passou a figurar como parâmetro de balizamento da qualidade de cursos e instituições. Este é um dos fatores que contribuem tanto para o desenvolvimento de um mercado da educação, quanto para necessidade de se determinar a conformidade dos sistemas baseados em avaliação para emitir juízo de valor acerca da qualidade. ([BERTOLIN; MARCON, 2015, p. 108](#)).

De acordo com [Barbosa, Freire e Crisóstomo \(2011, p. 318\)](#) a avaliação desempenho baseada em exames é motivada tanto no Brasil quanto internacionalmente pela responsabilidade institucional para com os *stakeholders* (ou financiadores), principalmente no caso de IES públicas. Por outro lado [Zandavalli \(2009, p. 389\)](#) aponta que na realidade, o que se busca com os dados estatísticos confiáveis é um melhor domínio das IES a fim de acabar com a gratuidade no ensino superior e criar uma lógica de mercado ao sistema universitário.

Sob uma dimensão funcional, a dificuldade em avaliar as IES decorre do fato de estas serem entidades muito distintas das empresas. As IES apresentam uma complexidade muito mais

acentuada pela existência de peculiaridades distintas das organizações empresariais uma vez que têm como objetivo formar profissionais capazes de exercer atividades de ensino, pesquisa e extensão. Deste modo, as dificuldades estão concentradas na mensuração da qualidade destas atividades. (BARBOSA; FREIRE; CRISÓSTOMO, 2011, p. 318).

A dificuldade em avaliar IES gira em torno da diversidade de grupos de interesse com diferentes graus de subjetividade e visões distintas sobre qualidade em educação. Em vista disso as avaliações tem sido concebidas e implementadas de formas diversas tanto no Brasil quanto em outros países. Apesar disso, organismos multilaterais têm estimulado em nível mundial a criação de mecanismos de avaliação como forma de maximização de benefícios e garantias sociais. (BERTOLIN; MARCON, 2015, p. 106)

Fatores importantes na avaliação de IES são métricas e indicadores adequados. Segundo Barbosa, Freire e Crisóstomo (2011, p. 319) as métricas e indicadores em uso no Brasil têm sido alvo de muitas críticas e questionamentos. Neste sentido, para o autor é necessário que se façam constantes avaliações no sentido de aperfeiçoá-los. Todavia, é visível que há um certo esforço dos governos em não abandoná-los.

A visualização do desempenho de cursos e IES a partir dos indicadores parecer ser a forma mais coerente de se trabalhar com métricas para realizar a mensuração e acompanhamento da qualidade do ensino. Apesar disso, existem dimensões que encontram-se não abrangidas por tais métricas. Uma dessas dimensões é abordada por Bertolin e Marcon (2015, p. 114) e faz referência ao *background* dos alunos. Este elemento pode ser entendido como condições relacionadas ao contexto familiar e situação sócio-econômica dos mesmos. Neste contexto, o autor faz referência ao capital cultural proposto por Bordieu e aborda os elementos que impactam diretamente nas condições de aprendizagem dos alunos e conseqüentemente seu desempenho em exames de grande abrangência.

O entendimento deste “pano de fundo” dos estudantes como fator determinante no desempenho em exames faz surgir controvérsias acerca da validade dos exames de mensuração de desempenho de ensino superior. Alunos com melhores panos de fundo são aqueles nascidos em famílias com melhores condições financeiras. Podem estudar sem se preocupar com provisões e tem melhores condições de aprendizagem pelo fato de os pais terem tido maior tempo de estudo e, portanto serem capazes de transmitir como mais facilidade o conhecimento aos seus filhos.

### 3 AVALIAÇÃO DO ENSINO SUPERIOR NO MUNDO

[Costa et al. \(2012, p. 417\)](#) destacam algumas características inerentes ao setor de produção educacional. São elas: i) a natureza múltipla e intangível do produto – os produtos educacionais podem ser classificados como: conhecimento e habilidades, valores, atitudes, entre outras características; ii) a participação do cliente no processo produtivo – o cliente (aluno) não é meramente um demandante da mercadoria, mas atua de forma decisiva no processo produtivo; iii) a heterogeneidade dos serviços – devido à participação do estudante no processo produtivo, as unidades produtivas se diferenciam umas das outras; iv) a dimensão temporal – os resultados obtidos no processo produtivo podem não ser suficientes para uma mensuração completa da produção do setor educativo, visto que é necessário observar uma trajetória completa da vida dos estudantes; v) o caráter acumulativo do ensino; vi) a incidência de fatores exógenos – essa característica tem como embasamento a denominada educação informal, que não é obtida pelos anos de estudos, mas sim por experiências fora do setor educacional.

O tema sobre como devem ser alocados os recursos públicos no setor de educação superior vem direcionando a grande maioria dos estudos para a mensuração da eficiência das IES públicas. Ao longo dos anos, muitos estudos têm como objetivo mensurar a eficiência e ranquear as IES públicas através de seu grau de eficiência. Ademais, cada país tem sua estrutura de financiamento e alocação de recursos que serve como base para a estimação da eficiência do setor educacional superior. *Ibid.*, p. 418

Os métodos mais utilizados para medir a eficiência dentro do contexto do setor educacional são os paramétricos e os não paramétricos. As técnicas estatísticas empregadas são baseadas nos Mínimos Quadrados Ordinários (MQO) de regressão para análise de fronteira estocástica. ([COSTA et al., 2012, p. 418](#)).

Por outro lado, como [Bernhardt \(1998, p. 1\)](#) afirma, a mensuração das unidades escolares com base em indicadores múltiplos pode não ser suficiente para o aprimoramento da unidade de ensino. De acordo com a autora, isso chegar a ser uma forma enganosa de se compreender o desempenho.

Deste modo, qualquer avaliação de desempenho que envolva mais de uma variável deve incluir em meio a estas, outras variáveis relacionadas a aspectos adicionais. São elas: dimensão demográfica, percepções, aprendizado do estudante e processos escolares. São esses fatores que fornecem as respostas que necessitam para melhorar os resultados. Ibid..

## 4 METODOLOGIA

Explicar a qualidade de cursos superiores compreende uma tarefa complexa que pode ser objeto de várias abordagens e modelagens matemáticas distintas. Verificou-se na literatura relacionada ao tema trabalhado – e em especial no trabalho de [Coelho \(2016\)](#), que a análise sobre a qualidade do Enade como indicador se dá pela aplicação da técnica de Análise Envoltória de dados (*Data Envelopment Analysis* - DEA), que se destina a encontrar as melhores práticas em uma amostra de organizações pertencentes a um conjunto de unidades comparáveis (*Decision Making Units* - DMU).

Em resumo, esta técnica baseia-se na identificação de variáveis de entrada (*input*) e saída (*output*) que implicam na eficiência das unidades de decisão. As inferências são realizadas de modo a se determinar a proporção de diminuição (ou aumento) de uma determinada variável de entrada (ou saída) a fim de tornar uma DMU eficiente.

### 4.1 Classificação da Pesquisa

A presente pesquisa é qualitativa uma vez que almeja fornecer um entendimento objetivo sobre a metodologia de avaliação de cursos pelo INEP por intermédio da interpretação das narrativas dos autores utilizados como fundamentação. Por outro lado, este trabalho é também quantitativo, uma vez que a motivação principal é entender os indicadores por meio de informações numéricas. Portanto a investigação qualitativa é complementada com a técnica de *mineração dos dados* do censo dos cursos superiores, aliada ao uso de metodologia estatística sobre os mesmos.



## 4.2 Ferramenta e Técnicas de análise de dados

Para manipulação dos dados foi utilizado R. Trata-se de uma linguagem e um ambiente para computação estatística e criação de gráficos similar a linguagem *S* desenvolvida pelo *Bell Labs*. É uma aplicação livre e de código aberto sob a licença GNU/GPL (*General Public License*) composta por uma extensa variedade de pacotes utilizados para as mais diferentes finalidades. O R é mantido por uma comunidade global de desenvolvedores e possui uma expressiva aceitação em pesquisas.

A crescente utilização deste ambiente de programação em substituição a outras soluções comerciais como *Microsoft Excel*, *IBM SPSS*, *Sata* e *Minitab* se deve a sua facilidade de uso e a possibilidade de aplicação em diversas áreas relacionadas à análise de dados. A opção por este ambiente se justifica pelo fato de o mesmo ser uma ferramenta gratuita e conter todos os recursos necessários às fases que compõem a aplicação da metodologia deste trabalho.

Esta pesquisa utilizou um conjunto de três técnicas estatísticas destinadas à preparação e análise de dados. São elas: Análise Exploratória de Dados (AED), Análise do Componente Principal (ACP) e Análise Fatorial. Uma breve descrição relativa a cada uma delas será apresentada nas subseções a seguir.

## 4.3 Análise Exploratória de dados

A análise exploratória neste trabalho foi possível pela versatilidade do pacote **dplyr**<sup>3</sup>. De acordo com Wickham et al. (2017) o pacote disponibiliza um conjunto extenso de instruções baseadas em “verbos” para manipulação do conjunto de dados. A ferramenta é uma iteração do pacote **plyr** idealizada para lidar com **dataframes**, por isso o prefixo *d*.

Os dados utilizados estão disponíveis publicamente em formato *Microsoft Excel* (.xlsx) e podem ser obtidos diretamente do site do INEP. A presente análise utiliza dados referentes ao senso de 2015. A escolha deste ano se deu pelo fato de o ciclo de avaliação dos cursos superiores pelo INEP ser trienal. Desta forma, os dados mais atuais para utilização são

---

<sup>3</sup> De acordo com os autores, o **plyr** tem três principais objetivos: identificar os verbos mais importantes para manipulação de dados e torná-los fáceis de utilizar com o R; promover rapidez para o tratamento dos dados em memória pela utilização do C++ como linguagem de desenvolvimento de componentes chaves do pacote; e utilizar a mesma interface para manipulação dos dados independentemente da forma como os mesmos estejam armazenados (seja como **dataframe**, *data tabel* ou *database*).

relativos ao último ciclo avaliativo de 2015 no qual os cursos da Área de Ciências Sociais Aplicadas foram avaliados.

Para elaboração da Análise Exploratória de Dados, procurou-se primeiramente descrever as variáveis presentes na base de dados do CPC, especificando nome, classe, além de uma breve definição. A seguir é apresentada a enumeração das mesmas juntamente aos seus atributos.

1. **Ano:** Categórica. Ano de realização do censo pelo INEP. Neste censo, foram avaliados os cursos em 2015.
2. **Código da IES:** Categórica. Código único de identificação das Instituições de ensino.
3. **Nome da IES:** Categórica. Nome da IES.
4. **Sigla da IES:** Categórica. Sigla da instituição de ensino.
5. **Código do Curso:** Categórica. Código de identificação composto por números inteiros positivos, únicos para cada curso nacionalmente. Varia de 1 a 5001295.
6. **Código da Área:** Categórica. Faz referência ao curso. Cada curso tem um código de área específico. Mesmos cursos em IES's distintas têm o mesmo código. Varia de 1 a 804.
7. **Área de Enquadramento:** Categórica. Nome do curso superior.
8. **Código do Município:** Categórica. Código da cidade onde o curso é ofertado. Varia de 1100023 a 5300108. 756 níveis.
9. **Município do Curso:** Categórica. Nome do município em que se localiza a IES que oferta o curso. 753 níveis.
10. **Sigla da UF:** Categórica. Sigla do estado. 27 níveis.
11. **Organização Acadêmica:** Categórica. As IES podem ter os possíveis 5 (cinco) níveis de enquadramento:
  - a) Centro Federal de Educação Tecnológica

- b) Centro Universitário
- c) Faculdade
- d) Instituto Federal de Educação, Ciência e Tecnologia
- e) Universidade

12. **Categoria Administrativa:** enquadramento organizacional da IES em 10 (dez) níveis:

- a) Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Civil
- b) Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Mercantil ou Comercial
- c) Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Associação de Utilidade Pública
- d) Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Fundação
- e) Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Sociedade
- f) Pessoa Jurídica de Direito Público - Estadual
- g) Pessoa Jurídica de Direito Público - Federal
- h) Pessoa Jurídica de Direito Público - Municipal
- i) Privada com fins lucrativos
- j) Privada sem fins lucrativos

13. **Concluintes Inscritos:** Inteira, discreta. Número de estudantes inscritos para realizar a avaliação do Enade.

14. **Concluintes Participantes:** Inteira, discreta. Quantidade de estudantes que efetivamente realizaram a avaliação.

15. **Nota Bruta FG:** Real, contínua. Nota dos alunos na parte de formação geral em escala de 0 a 100.

16. **Nota Bruta - CE:** Real, contínua. Nota dos alunos na parte de conhecimento específico em escala de 0 a 100.

17. **Nota Bruta - Geral:** Real, contínua. Média ponderada envolvendo as notas nas provas de conhecimentos específicos e gerais.
18. **Nota Contínua do Enade:** Real, contínua. Nota atribuída com base no desempenho do estudante na prova do Enade, calculada para cada curso de graduação. Varia de 0 a 5.
19. **Nota Bruta - Organização Didático-Pedagógica:** Real, contínua. Nota média dos alunos nas questões do questionário do estudante referentes à Organização Didático-Pedagógica em escala de 0 a 6.
20. **Nota Padronizada - Organização Didático-Pedagógica:** Real, contínua. nota padronizada dos alunos nas questões do questionário do estudante referentes à Organização Didático-Pedagógica em escala de 0 a 6.
21. **Nota Bruta - Infraestrutura e Instalações Físicas:** Real, contínua. Nota média dos alunos nas questões do questionário do estudante referentes à infraestrutura em escala de 0 a 6.
22. **Nota Padronizada - Infraestrutura e Instalações Físicas:** Real, contínua. Nota padronizada dos alunos nas questões do questionário do estudante referentes à infraestrutura em escala de 0 a 6.
23. **Nota Bruta - Oportunidades de Ampliação da Formação:** Real, contínua. Nota média dos alunos nas questões do questionário do estudante referentes à Oportunidades de ampliação em escala de 0 a 6.
24. **Nota Padronizada - Oportunidades de Ampliação da Formação:** Real, contínua. nota padronizada dos alunos nas questões do questionário do estudante referentes à oportunidades de ampliação em escala de 0 a 6.
25. **Concluintes Participantes com nota no Enem:** Inteira, discreta. Quantidade de estudantes que realizaram o Enem como forma de ingresso nas IES's.

26. **Percentual de Concluintes participantes com nota no Enem:** Real, contínua. Relação entre os concluintes que participaram da avaliação do Enade e os que tiveram nota no Enem.
27. **Nota Bruta - IDD:** Real, contínua. Indicador de diferença de desempenho. Mede o valor agregado pelo processo formativo ao desenvolvimento dos estudantes concluintes oferecido pelo curso. Mensura o valor agregado pelo curso considerando resultado do Enade e as características dos estudantes ao ingressarem no curso avaliada pela nota do estudante no Enem. Nesta base de dados, variam de -23.6500 a 24.4700.
28. **Nota Padronizada - IDD:** Real, contínua. Indicador de diferença de desempenho após padronização em escala de 0 a 5.
29. **Nr. de Docentes:** Inteira, discreta. Quantidade de professores no curso.
30. **Nota Bruta - Mestres:** Real, contínua. Percentual de mestres do curso de 0 a 100%.
31. **Nota Padronizada - Mestres:** Real, contínua. Nota padronizada atribuída ao percentual de mestres após padronização em escala de 0 a 5.
32. **Nota Bruta - Doutores:** Real, contínua. Percentual de doutores do curso de 0 a 100%.
33. **Nota Padronizada - Doutores:** Real, contínua. Nota padronizada atribuída ao percentual de doutores após padronização em escala de 0 a 5.
34. **Nota Bruta - Regime de Trabalho:** Real, contínua. Percentual de professores com regime integral + parcial do curso de 0 a 100%.
35. **Nota Padronizada - Regime de Trabalho:** Nota padronizada atribuída ao percentual de professores com regime integral + parcial após padronização.
36. **CPC Contínuo:** Real, contínua. Conceito preliminar de curso é a nota atribuída ao curso superior, em escala de 0 a 5.
37. **CPC Faixa:** Inteira, discreta. Variável resposta. Faixa do CPC em escala de 0 a 5.

Quanto a estrutura da base de dados, e informações pertinentes ao censo algumas informações podem ser destacadas:

1. **Dimensões da base de dados:** 8121 linhas  $\times$  38 colunas

2. **Dados sobre o censo:**

- a) Cursos avaliados: 8121
- b) IES's avaliadas: 1758
- c) Municípios com cursos avaliados: 753
- d) Estados com cursos avaliados: 27
- e) Percentual de participação nas provas: 81.4%

3. **Cursos por regiões:**

Área de Enquadramento	Centro Oeste	Nordeste	Norte	Sudeste	Sul	Total
ADMINISTRAÇÃO	189	312	97	841	367	1806
DIREITO	117	208	70	455	216	1066
CIÊNCIAS CONTÁBEIS	115	190	71	444	224	1044
TECNOLOGIA EM GESTÃO DE RECURSOS HUMANOS	41	69	17	314	82	523
PSICOLOGIA	39	85	31	200	111	466
PUBLICIDADE E PROPAGANDA	30	49	16	200	60	355
TECNOLOGIA EM LOGÍSTICA	15	44	12	210	56	337
JORNALISMO	23	49	22	126	55	275
TECNOLOGIA EM MARKETING	10	37	7	167	50	271
TECNOLOGIA EM GESTÃO FINANCEIRA	11	23	5	139	45	223
TECNOLOGIA EM PROCESSOS GERENCIAIS	9	24	7	113	63	216
CIÊNCIAS ECONÔMICAS	15	36	9	83	48	191
DESIGN	8	20	5	70	75	178
TECNOLOGIA EM GESTÃO COMERCIAL	9	32	6	77	46	170
TURISMO	16	35	10	61	27	149
TECNOLOGIA EM GASTRONOMIA	10	24	2	49	22	107
RELAÇÕES INTERNACIONAIS	9	8	5	58	21	101
TEOLOGIA	8	14	5	37	30	94
TECNOLOGIA EM DESIGN DE INTERIORES	8	14	2	38	20	82
TECNOLOGIA EM GESTÃO DA QUALIDADE	6	6	48	15	6	81
TECNOLOGIA EM COMÉRCIO EXTERIOR	2	3	2	57	16	80
TECNOLOGIA EM DESIGN GRÁFICO	6	12	4	43	12	77
TECNOLOGIA EM GESTÃO PÚBLICA	15	11	4	19	16	65
SECRETARIADO EXECUTIVO	7	10	4	22	18	61
TECNOLOGIA EM DESIGN DE MODA	2	13	2	22	19	58
ADMINISTRAÇÃO PÚBLICA	6	16	2	19	8	51

Tabela 1 – Cursos por região

#### 4. Percentual de cursos por Organização Acadêmica:

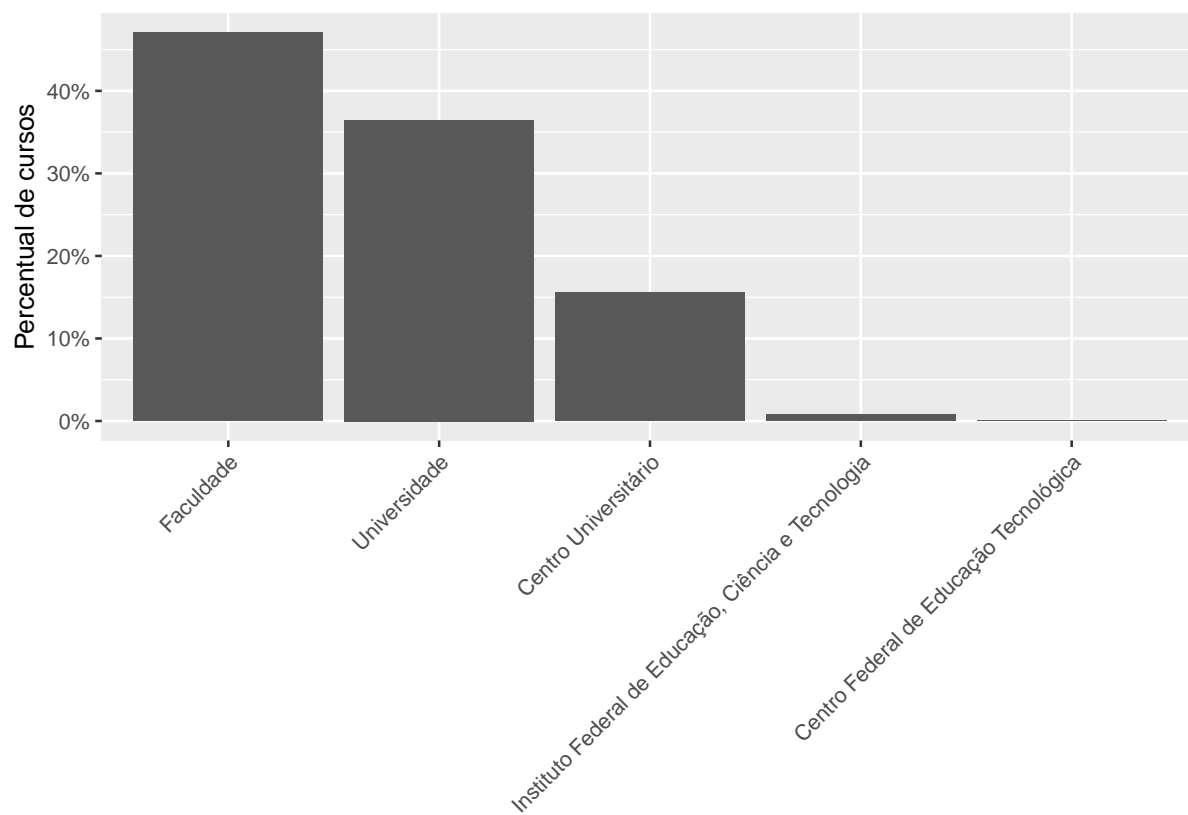


Figura 1 – Percentual de cursos por organização acadêmica

### 5. Percentual de cursos por Categoria Administrativa:

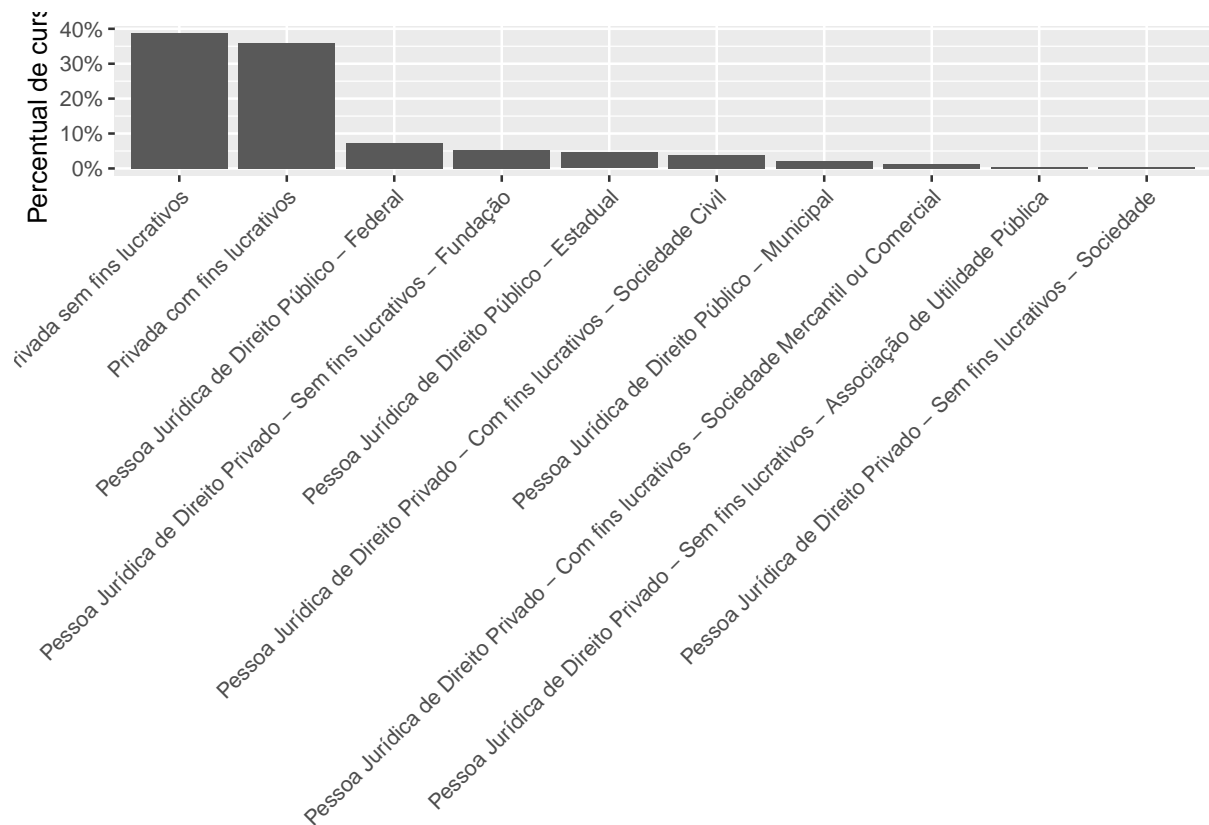


Figura 2 – Percentual de cursos por categoria administrativa

### 6. IES por categoria administrativa:

Categoria Administrativa	Total
Privada com fins lucrativos	681
Privada sem fins lucrativos	671
Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Fundação	94
Pessoa Jurídica de Direito Público - Federal	91
Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Civil	70
Pessoa Jurídica de Direito Público - Estadual	68
Pessoa Jurídica de Direito Público - Municipal	43
Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Mercantil ou Comercial	25
Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Associação de Utilidade Pública	9
Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Sociedade	6

7. **Variável resposta:** Conceito Preliminar de Curso. Sobre a variável resposta *CPC Faixa* foi observado que 12,97% dos cursos não obtiveram notas<sup>4</sup>. Os demais cursos



estão relacionados na figura 3 e na tabela 2:

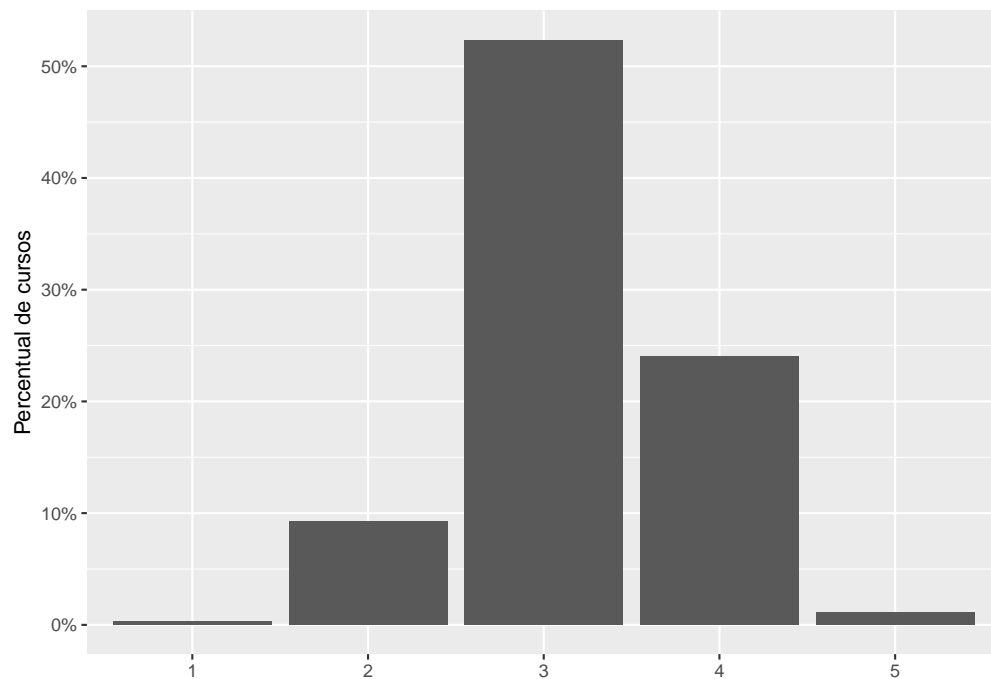


Figura 3 – Cursos com conceito

Conceito	Frequência	Percentual
1	22	0.3%
2	753	9.3%
3	4252	52.4%
4	1949	24.0%
5	91	1.1%

Tabela 2 – Cursos com conceito

8. **Análise das variáveis numéricas:** Sabe-se que alguns cursos não obtiveram conceito pelo fato de estarem na condição de não reconhecidos, sem conceito ou sub judice. Foi observado, contudo que tais cursos apesar de estarem com o campo CPC Faixa vazios (e portanto sem a principal variável, a de resposta), tiveram notas nas variáveis preditoras. A partir desta constatação foi possível concluir que a utilização dessas observações nas modelagens estatísticas traria prejuízos significativos em termos de precisão.

<sup>4</sup> Cursos não reconhecidos: 847 (10.4%); Cursos sem conceito: 205 (2.52%); Cursos sub judice: 2 (0.0246%)

Com a finalidade de minimizar os efeitos da influência de observações vazias, duas abordagens foram inicialmente estudadas. A primeira consiste em utilizar o método de imputação de dados denominado  $k$ -NN<sup>5</sup> (*k-Nearest Neighbors algorithm*) procura estimar o valor da observação vazia a partir de um parâmetro  $k$  que representa a distância deste aos seus  $k$  vizinhos mais próximos.

Na segunda abordagem foram omitidas as observações vazias sob a hipótese de que as mesmas estão suficientemente representadas pelas demais observações. Desta forma a criação de um modelo não perderia eficiência nem estaria enviesado. A tabela 3 contém a sumarização estatística das variáveis preditoras. Nela foi utilizada a técnica de estimação das observações vazias pelo método  $k$ -NN. Não foram verificadas diferenças significativas nas medidas estatísticas que implicassem em impossibilidade de estimação pelo método proposto.

Como Josse, Husson et al. (2016, p. 2) de forma bem elucidativa aborda, manipular bases com dados faltantes utilizando PCA pode ser uma tarefa complicada pela limitação que alguns softwares têm em gerar resultados a partir de campos vazios. Segundo esse autor a aplicação do método  $k$ -NN pressupõe que as variáveis estejam sob distribuição Gaussiana e é seguramente um método de imputação de variáveis contínuas bastante popular Ibid., p. 24. A sumarização estatística utilizando o método pode ser vista na tabela 3.

---

<sup>5</sup> Este método encontra-se implementado no pacote **DMwR** sob autoria de Torgo (2010).

Variável	Média	Desvio padrão	Mediana	Moda	Mínimo	Máximo	n
Concluintes Inscritos	67.6625	204.194	39	25	1	11155	8121
Concluintes Participantes	55.0494	162.6907	32	20	0	10059	8121
Nota Bruta - FG	54.1955	6.2626	53.8867	55.8	4.7127	80.65	8121
Nota Bruta - CE	42.5299	8.228	41.46	44.6	4.1746	80.8091	8121
Nota Bruta - Geral	45.4587	7.0596	44.7091	47.6	4.3141	79.75	8121
Nota Contínua do Enade	2.3915	0.8577	2.324	5	0	5	8121
Nota Bruta - Organização Didático-Pedagógica	5.2105	0.481	5.221	6	1.7121	6	8121
Nota Padronizada - Organização Didático-Pedagógica	3.0204	1.1832	3.0213	5	0	5	8121
Nota Bruta - Infraestrutura e Instalações Físicas	4.9685	0.6723	5.0136	6	2	6	8121
Nota Padronizada - Infraestrutura e Instalações Físicas	3.1313	1.2076	3.1954	5	0	5	8121
Nota Bruta - Oportunidades de Ampliação da Formação	4.5659	0.8142	4.5565	6	1.1905	6	8121
Nota Padronizada - Oportunidades de Ampliação da Formação	2.9541	1.1725	2.9522	5	0	5	8121
Concluintes Participantes com nota no Enem	28.7791	67.3562	17	7	1	3842	8121
Percentual de Concluintes participantes com nota no Enem	0.5344	0.206	0.5385	0.5	0.0143	1	8121
Nota Bruta - IDD	-0.1043	2.3789	-0.0948	-1.5685	-23.6519	24.4723	8121
Nota Padronizada - IDD	2.4782	0.8428	2.4758	0	0	5	8121
Nr. de Docentes	26.1061	21.6867	20	16	0	298	8121
Nota Bruta - Mestres	0.7633	0.2083	0.8077	1	0	1	8121
Nota Padronizada - Mestres	3.5552	1.2234	3.7796	5	0	5	8121
Nota Bruta - Doutores	0.3124	0.2219	0.2766	0	0	1	8121
Nota Padronizada - Doutores	1.7291	1.2051	1.5541	0	0	5	8121
Nota Bruta - Regime de Trabalho	0.7592	0.238	0.8125	1	0	1	8121
Nota Padronizada - Regime de Trabalho	3.6919	1.2766	3.9688	5	0	5	8121
CPC Contínuo	2.6111	0.5807	2.609	1.0491	0.4257	4.6885	8121
CPC Faixa	3.1711	0.6371	3	3	1	5	8121

Tabela 3 – Sumarização da base de dados após uso do método  $k$ -NN: medidas estatísticas básicas

## 4.4 Análise do Componente Principal

### 4.4.1 Padronização das variáveis

De acordo com [Larose e Larose \(2015, p. 98\)](#) a padronização dos dados precede a aplicação da Análise do Componente Principal - ACP (ou *Principal Component Analysis*). Isso fica evidente com base na observação da sumarização estatística apresentada na tabela 3. Nesta tabela é possível observar a desproporcionalidade dos desvios-padrão entre as variáveis. Concluintes Inscritos, por exemplo, tem um desvio 896 vezes maior que Nota Bruta - Regime de Trabalho. Ainda de acordo com o mesmo autor, a padronização tem como principal objetivo evitar que a influência de uma variável domine a da outra no espectro de variabilidade.

Tanto a ACP como a AF (Análise do Fator, como será visto adiante) fizeram amplo uso do pacote `psych`<sup>6</sup>. Dada a sua importância, este pacote constitui elemento chave para a aplicação das metodologias estatísticas desta pesquisa. De acordo com [Revelle \(2017\)](#), abarca um conjunto de funções para análise multivariada permitindo a criação de estatísticas descritivas. Dentre estas funções podem ser destacadas `fa` e `principal` que possibilitaram a computação dos carregamentos para a AF e ACP, respectivamente.

A padronização dos dados foi feita com a utilização da função `scales` contida na instalação básica do R. Esta função posiciona as variáveis contínuas em uma unidade de escala pela subtração de sua média e divisão pelo desvio padrão (procedimento denominado *z-scoring*). Isso torna o desvio padrão das mesmas 1 e a média 0. Em termos matemáticos a padronização pode ser expressa pela seguinte equação:

$$Z_i = \frac{X_i - \mu_i}{\sigma_{ii}} \quad (1)$$

Foi observado que o *dataset* do CPC contém, para cada nota bruta, uma nota padronizada. Além disso, segundo [INEP \(2017\)](#) a metodologia utilizada para produção destas notas é expressa pela seguinte equação:

$$Z_{FGj} = \frac{FG_{kj} - \overline{FG_k}}{S_{FGk}} \quad (2)$$

Onde,

---

<sup>6</sup> Possui ferramentas estatísticas de propósito geral para aplicações em psicometria, personalidade e psicologia experimental.

$Z_{FGj}$  é o afastamento padronizado em  $FG$  da unidade de observação  $j$ ;

$FG_{kj}$  é a nota bruta em  $FG$  da  $j$ -ésima unidade de observação da área de avaliação  $k$ ;

$\overline{FG_k}$  é a média em  $FG$  da área de avaliação  $k$ ; e

$S_{FGk}$  é o desvio-padrão em  $FG$  da área de avaliação  $k$ .

Nesta pesquisa optou-se por manter tais variáveis no conjunto de preditoras. Esta decisão foi tomada com base na verificação de que tal metodologia é aplicada para a composição da respectiva variável. Além do mais, a mesma não necessariamente implica em desvio padrão unitário e média 0 quando a análise é feita sobre a totalidade de observações, como pode ser observado na tabela 4.

Variáveis padronizadas pelo INEP	Média	Desvio padrão	Mediana	Moda	Mínimo	Máximo	n
Nota Padronizada - Organização Didático-Pedagógica	3.0202	1.1832	3.0212	5	0	5	8121
Nota Padronizada - Infraestrutura e Instalações Físicas	3.1312	1.2076	3.1951	5	0	5	8121
Nota Padronizada - Oportunidades de Ampliação da Formação	2.9541	1.1724	2.9519	5	0	5	8121
Nota Padronizada - IDD	2.4801	0.8406	2.4773	0	0	5	8121
Nota Padronizada - Mestres	3.5556	1.2232	3.7834	5	0	5	8121
Nota Padronizada - Doutores	1.7295	1.2049	1.5556	0	0	5	8121
Nota Padronizada - Regime de Trabalho	3.6922	1.2768	3.9688	5	0	5	8121

Tabela 4 – Estatísticas sobre as variáveis padronizadas pelo INEP

#### 4.4.2 Matriz de correlação das variáveis preditoras

De acordo com [Larose e Larose \(2015, p. 98\)](#) a próxima etapa na Análise do Componente Principal consiste em verificar a existência de correlações entre as 23 variáveis preditoras e, portanto, a ocorrência de multicolinearidade. Duas variáveis do dataset são de resposta: CPCC refere-se à nota contínua do CPC e CPCF é a nota escalonada. Desta forma temos um total de 25 variáveis (23 preditoras e 2 de resposta).

A dificuldade em fazer uma matriz de correlação com as variáveis do CPC está em determinar quais das 25 variáveis apresentam correlação umas com as outras. A quantidade de variáveis preditoras é considerável e não é possível afirmar que as mesmas não apresentam correlação mútua.

Segundo [Larose e Larose \(2015, p. 92\)](#), modelar uma relação utilizando um número de variáveis muito grande pode complicar mais do que explicar e viola dois princípios básicos. O primeiro é o da parcimônia segundo o qual é preciso manter a quantidade de variáveis

preditoras em uma quantidade mínima de forma que haja facilidade na interpretação. O segundo é a tendência ao *overfittig* que diz que a generalidade dos resultados pode dificultar a formação de novos resultados com a mudança dos dados de entrada.

Para análise de correlação foi utilizada a matriz correlograma<sup>7</sup> da figura 4. Nesta matriz as variáveis são dispostas de forma que é possível determinar para cada uma o seu nível de correlação com os seus pares. Isso possibilita a verificação de possíveis ocorrências de multicolinearidade.

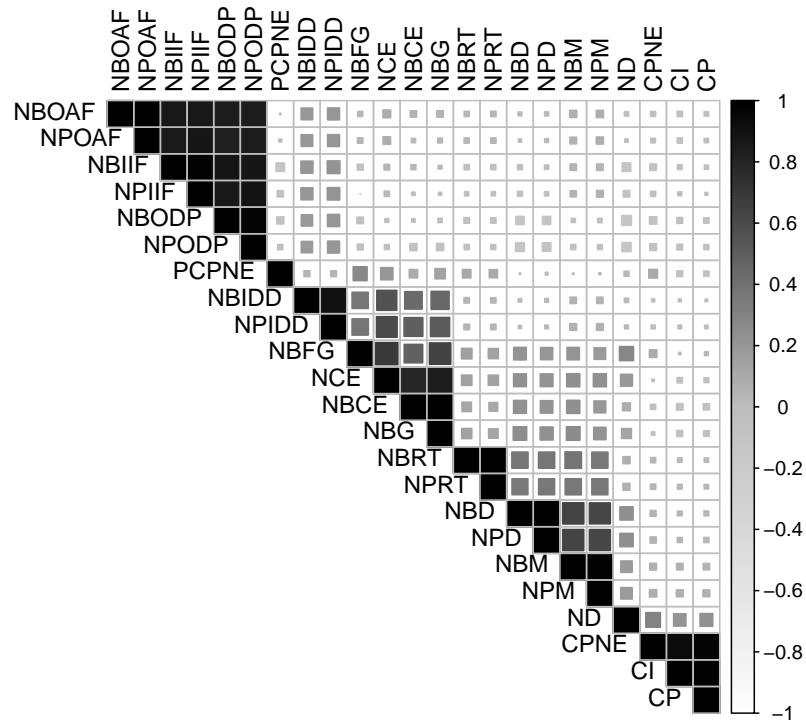


Figura 4 – Matriz correlograma: correlação entre os pares de variáveis segundo uma escala de cor.

Nesta matriz, a ordenação das variáveis foi feita por hierarquia de correlação, ou seja, da esquerda para direita, iniciando pelas correlações mais fortes. Com base na observação nas intensidades das cores, podemos verificar alguns trechos de correlações consideráveis especialmente nas variáveis relacionadas às oportunidades de ampliação da formação e infraestrutura e instalações físicas.

Na figura 5 vemos a representação das duas bandas da matriz correlograma. As variáveis assumem ordenamento de acordo com o índice de correlação e são posteriormente agrupadas

<sup>7</sup> Esta matriz é criada por intermédio da função `corrplot` do pacote de mesmo nome. De acordo com seus criadores, [Wei e Simko \(2017\)](#), o pacote permite visualizar matrizes de correlação e intervalos de confiança. Também contém algoritmos para reordenação de matrizes.

em *clusters*.

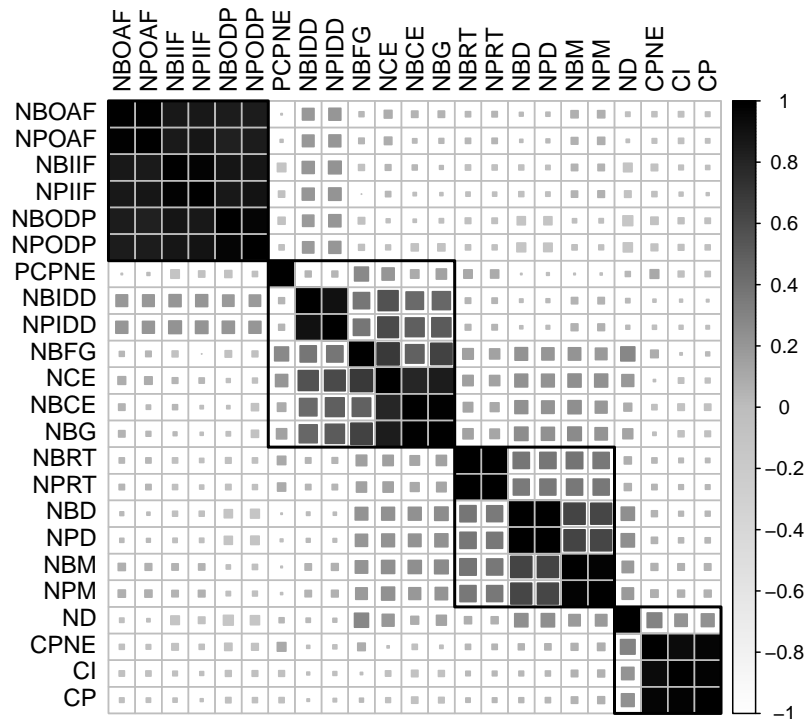


Figura 5 – Matriz correlograma com as duas bandas

Uma visualização alternativa das correlações pode ser obtida por meio da função `shinypairs` do pacote `pairsD3`. Este pacote permite produzir uma matriz interativa de gráficos de dispersão (TARR, 2015). É possível selecionar interativamente, de forma gradual dentre as variáveis preditoras, as que formarão a matriz de correlações. Com base nessa seleção de variáveis, copia-se da tela do `shiny`<sup>8</sup> o código que permite replicar a matriz com a função `pairs`.

Na figura 6 podemos ver a formação do *scatter-plot* com combinações<sup>9</sup> de variáveis obtidas pela interface do `pairsD3`.

<sup>8</sup> Facilitador de visualização interativa de dados. Para maiores informações visitar <https://shiny.rstudio.com>

<sup>9</sup> É importante observar que, dada a alta quantidade de variáveis contidas na base de dados, não é possível afirmar que a configuração de variáveis mostrada na matriz da figura 6 contém combinação de pares com maior correlação. Experimentações foram feitas com variações de pares tendo como referência a matriz correlograma da figura 5. Para representação com todas as 23 variáveis preditoras seria necessário espaço suficiente para alocar 529 relações, muito além das dimensões das páginas deste trabalho.

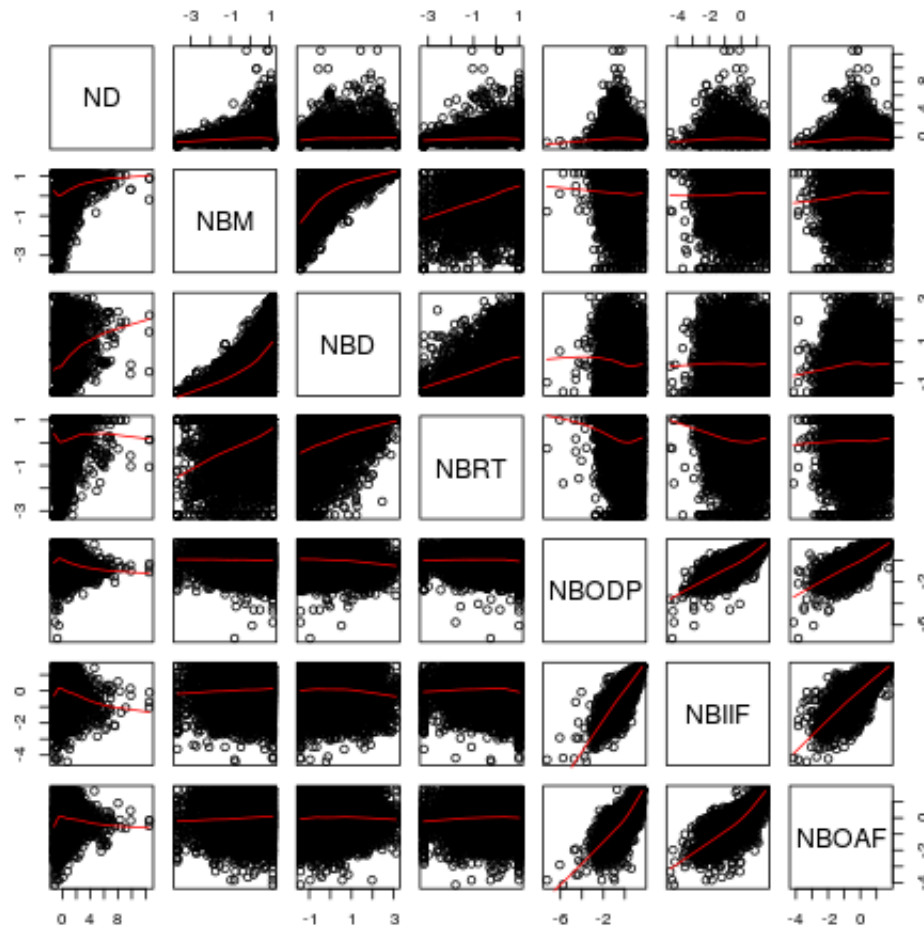


Figura 6 – Matriz de pares de variáveis

Devido ao elevado número de preditoras, esta forma de visualização, apesar de precisa, torna-se prejudicada. Contudo, o diagrama de matriz e a matriz de correlação são duas formas de olhar para a mesma coisa. (LAROSE; LAROSE, 2015, p. 98)

#### 4.4.3 Aplicação da Análise do Componente Principal

De acordo com Larose e Larose (2015, p. 98) a ACP (Análise do Componente Principal) é a ferramenta mais adequada quando são analisadas variáveis em multicolinearidade. Isto é, tenta-se explicar a estrutura de correlações com base em combinações lineares, também chamadas de componentes. A variabilidade total de todas as variáveis pode ser contabilizada por um conjunto de componentes, uma vez que existe a mesma quantidade de informações nos mesmos em relação ao conjunto original.



Em outras palavras, a análise do componente principal procura reduzir a dimensão de um determinado conjunto de características pela criação de um novo conjunto de propriedades representativas devendo, para tanto, contabilizar a maior variância possível.

#### 4.4.4 Componentes Principais

De acordo com [Larose e Larose \(2015, p. 95–96\)](#) a ACP é uma técnica de análise que se fundamenta em três corolários:

1. Assumindo que cada variável preditora possa ser representada em forma de vetor; que o conjunto dessas variáveis tenha sido padronizado e que a quantidade total de variáveis seja  $m$ , a variabilidade total do grupo de variáveis é igual a soma de cada vetor individual, que é igual à soma dos *eigenvectors*<sup>10</sup>, que é o total de variáveis predictoras. Desta forma, em termos matemáticos temos que:

$$\sum_{i=1}^m Var(Y_i) = \sum_{i=1}^m Var(Z_i) = \sum_{i=1}^m \lambda_i = m \quad (3)$$

Onde,

$Y_i$  é o  $i$ -ésimo componente principal;

$Z_i$  é o  $i$ -ésimo vetor  $Z$  (variável preditora pós padronização);

$\lambda_i$  é o  $i$ -ésimo *eigenvector* da variável preditora.

2. A proporção de variabilidade total que explica o  $i$ -ésimo componente principal é a razão entre o  $i$ -ésimo *eigenvector* pelo total de variáveis, isto é,  $\frac{\lambda_i}{m}$ .

A listagem 1 é o resultado da computação e visualização dos componentes principais das 23 predictoras sendo cada célula o peso de um componente:

---

<sup>10</sup> Em algebra linear, um *eigenvector* ou vetor característico de uma transformação linear é um vetor não nulo que se altera somente por um fator escalar quando uma transformação linear é aplicada ao mesmo.

---

Loadings:

	PC1	PC2	PC3	PC4	PC5
CI		0.103	0.639	0.722	
CP		0.108	0.642	0.734	
NBFG	0.175	0.624	-0.271	0.234	
NBCE	0.236	0.679	-0.440	0.195	
NBG	0.245	0.732	-0.445	0.223	
NCE	0.285	0.725	-0.430	0.255	
NBODP	0.891	-0.308			
NPODP	0.891	-0.327	0.112		
NBIIF	0.919	-0.237	0.121		
NPIIF	0.918	-0.237	0.145		
NBOAF	0.918	-0.164	0.140		
NPOAF	0.912	-0.172	0.148		
CPNE		0.152	0.621	0.736	
PCPNE		0.179	-0.124		0.306
NBIDD	0.413	0.379	-0.419	0.331	
NPIDD	0.439	0.399	-0.443	0.334	
ND		0.343	0.232	0.208	-0.197
NBM	0.157	0.653	0.392	-0.343	-0.242
NPM	0.154	0.631	0.410	-0.340	-0.245
NBD		0.670	0.376	-0.373	-0.284
NPD		0.663	0.386	-0.373	-0.284
NBRT		0.497	0.313	-0.334	0.712
NPRT		0.483	0.318	-0.332	0.720

	PC1	PC2	PC3	PC4	PC5
SS loadings	5.633	4.974	3.207	2.820	1.472
Proportion Var	0.245	0.216	0.139	0.123	0.064
Cumulative Var	0.245	0.461	0.601	0.723	0.787

---

Por definição o primeiro componente é o mais representativo de todos, ou seja, é a dimensão que contabiliza a maior variabilidade possível de todas as variáveis preditoras. Em outros termos é a dimensão que maximiza a variância.

Com base na listagem 1, observa-se no primeiro componente (PC1) que as variáveis NBODP, NBIIF e NBOAF tem os maiores pesos. Isso significa que a organização didático pedagógica, infra-estrutura e instalações e oportunidades de ampliação da formação dos estudantes nos cursos são as variáveis mais representativas deste componente. Além disso o primeiro componente contabiliza aproximadamente 1/4 da variância de todo conjunto de componentes.

#### 4.4.5 Seleção dos Componentes

A seleção dos componentes pode ser feita com base na seleção de um dentre vários critérios. Um deles é o da proporção da variância explicada que relaciona a área de pesquisa e a proporção de variabilidade que se deseja trabalhar. Em áreas como Ciências Sociais por exemplo, 60% da variabilidade é considerada satisfatória dada a imprevisibilidade da natureza do comportamento humano (LAROSE; LAROSE, 2015, p. 103). Sob essa ótica escolheu-se os quatro primeiros componentes, uma vez que os mesmos contabilizam 72,3% de toda a variabilidade.

A natureza exploratória da pesquisa fornece relativa liberdade em relação a seleção da quantidade de componentes. Pelo critério *Scree Plot*, a partir do quinto componente é possível verificar uma tendência de estabilização dos pesos, como pode ser observado no gráfico 7. Este é o ponto em que se recomenda que seja feita a extração dos componentes.

Sob essa ótica escolheu-se os três primeiros componentes, uma vez que os mesmos contabilizam 60,1% de toda a variabilidade. Ainda no gráfico 7, uma observação pertinente é que pelo critério de *tendência de horizontalização da curva de eigenvalues* sugere-se a possibilidade de selecionar os cinco primeiros componentes. Observa-se que a partir do quinto componente é possível verificar uma tendência de horizontalização da curva de *eigenvalues*. Desta, caso fosse utilizado tal critério seria computada 78,7% da variabilidade.

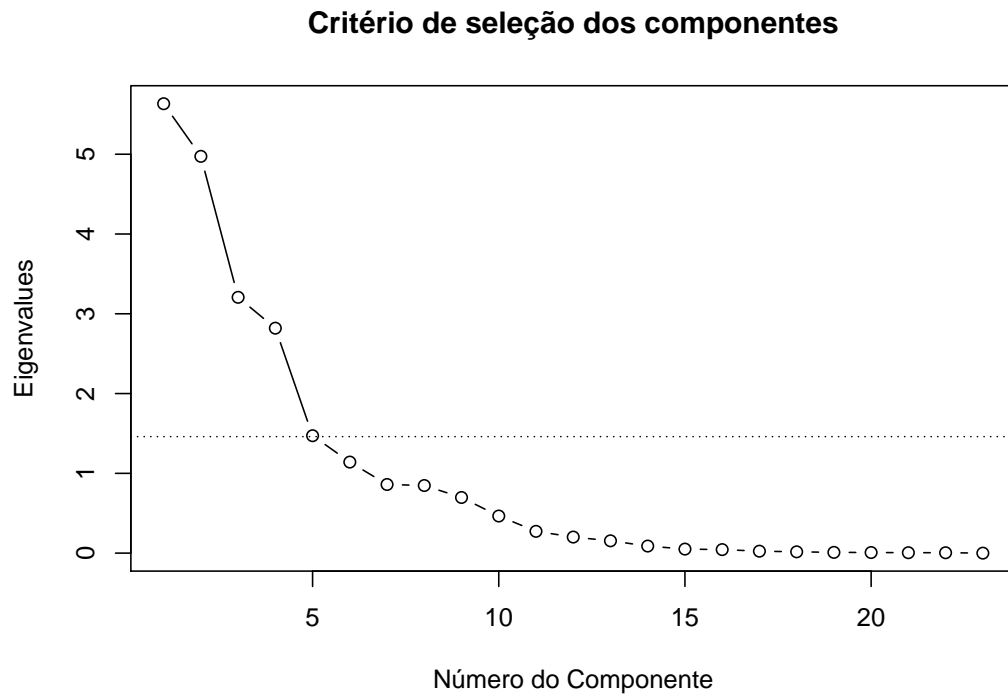


Figura 7 – Critério de seleção dos componentes: a partir do 5º componente, observa-se redução no decaimento da variabilidade, o que diminui a importância do componente para uma possível seleção.

#### 4.4.6 Características dos Componentes

A tabela 5 apresenta com mais clareza os pesos atribuídos aos 4 primeiros componentes pela aplicação da ACP.

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>
CI	-0.0849	0.1032	0.6388	0.7223
CP	-0.0748	0.1076	0.6423	0.7344
NBFG	0.1748	0.6244	-0.2712	0.2344
NBCE	0.2362	0.6789	-0.4401	0.1951
NBG	0.2452	0.732	-0.4448	0.2226
NCE	0.2849	0.7246	-0.4295	0.2552
NBODP	0.8914	-0.3083	0.0949	-0.0112
NPODP	0.8906	-0.327	0.1124	-0.0083
NBIIF	0.9189	-0.2369	0.1209	-0.017
NPIIF	0.9185	-0.2368	0.1453	-0.0077
NBOAF	0.9175	-0.1637	0.1402	-0.033
NPOAF	0.9124	-0.1723	0.1478	-0.03
CPNE	-0.0899	0.1519	0.6207	0.7365
PCPNE	-0.0035	0.1788	-0.124	0.0935
NBIDD	0.4134	0.3794	-0.4194	0.3308
NPIDD	0.439	0.3988	-0.4431	0.3344
ND	-0.0428	0.3428	0.2322	0.2084
NBM	0.1568	0.6534	0.3918	-0.3426
NPM	0.1542	0.6309	0.4099	-0.3402
NBD	0.0692	0.6702	0.3762	-0.3727
NPD	0.0741	0.6629	0.3857	-0.3727
NBRT	0.0739	0.4969	0.3134	-0.334
NPRT	0.0727	0.4829	0.3175	-0.332

Tabela 5 – Características dos Componentes

De acordo com [Larose e Larose \(2015, p. 107\)](#) o peso dos componentes representa a relação do mesmo com a variável. Para que o peso de um componente tenha significância prática este deve exceder  $\pm 0.50$  em magnitude. Desta forma, a [tabela 5](#) apresenta a mesma informação, sendo omitidos os pesos menores que 0.5.

	PC1	PC2	PC3	PC4
CI			0.6388	0.7223
CP			0.6423	0.7344
NBFG		0.6244		
NBCE		0.6789		
NBG		0.732		
NCE		0.7246		
NBODP	0.8914			
NPODP	0.8906			
NBIIF	0.9189			
NPIIF	0.9185			
NBOAF	0.9175			
NPOAF	0.9124			
CPNE			0.6207	0.7365
PCPNE				
NBIDD				
NPIDD				
ND				
NBM		0.6534		
NPM		0.6309		
NBD		0.6702		
NPD		0.6629		
NBRT				
NPRT				

Tabela 6 – Componentes com magnitudes  $\pm .50$ 

Determinados os pesos, é possível agora classificar os componentes principais de acordo com sua estrutura e pesos nas variáveis:

1. O primeiro componente (PC1) apresenta pesos significativos nas variáveis *Organização Didático-pedagógica, Infra-estrutura e Instalações* e *Oportunidades de Ampliação da Formação*. Com base nisto é possível concluir que esse componente apresenta um perfil do que o curso é capaz de entregar ao aluno, que é refletida nas notas que o mesmo atribui no momento da avaliação.
2. Em PC2 observamos uma concentração dos pesos nas variáveis associadas a duas dimensões: alunos e professores. No que se refere aos alunos, pesam mais as notas avaliadas de acordo com o desempenho dos mesmos nas provas nas áreas de *Formação*

*Geral e Conhecimentos Específicos*. Já em relação aos professores tem destaque as variáveis *Quantidade de Mestres* e *Quantidade de Doutores* nos cursos.

3. Em PC3 têm destaque as variáveis *Concluínates Inscritos* e *Percentual de Concluínates Participantes com nota no Enem*. Isso indica que este componente tem um perfil mais voltado a determinar a situação dos alunos no censo bem como a contabilização da participação dos mesmos.
4. PC4 apresenta perfil de composição similar à PC3. Os pesos, por sua vez, apresentam magnitudes levemente superiores. Isso se deve ao fato de esse componente ter tido uma proporção de variância individual menor em relação a PC3 (13,9% contra 21,6%) como pode ser observado na listagem 1. Desta forma, apesar de este componente ter apresentado pesos relativamente maiores, sua variância total menor faz com que seja posicionado em 4º na ordem dos componentes.

## 4.5 Análise do Fator

A análise do componente principal apresentou quais variáveis são passíveis de melhor explicarem o que influencia no desempenho do curso no CPC. Foi apresentado o conjunto de variáveis e o perfil de cada componente. Contudo, é oportuno incluir mais uma camada de robustez à pesquisa pela aplicação da Análise do Fator – AF ou (*Factor Analysis* – FA). Larose e Larose (2015, p. 110–111) destaca que, apesar de a ACP e AF serem técnicas de análise multivariada bem similares, a ACP deve ser em essência, um meio para a consecução da AF.

De acordo com Hair et al. (1998, p. 93) a AF desempenha um papel importante na identificação das interações entre as variáveis por intermédio do agrupamento (em fatores) das mesmas de acordo com seu grau de relação. Esse procedimento permite a redução da dimensionalidade. Isso quer dizer que procura-se descrever um evento utilizando um conjunto menor de variáveis relativamente ao conjunto original estudado com o mínimo de perda de informação.

Vale ressaltar que as conclusões desta pesquisa tem perspectiva exploratória. Procura-se por meio do AF estabelecer uma estrutura nos dados do Enade utilizando este método como

técnica de redução. Essa abordagem é diferente da confirmatória pois não são utilizados testes para se chegar a uma estrutura esperada com base em hipóteses. Ibid..

De acordo com [Larose e Larose \(2015, p. 111\)](#), precedem a aplicação da AF os seguintes procedimentos: (i) padronização das variáveis; (ii) determinação dos vetores-Z (ou *Z-vectors*) e (iii) a matriz de correlação.

Na figura 8 vemos a reapresentação da matriz de correlação das variáveis preditoras agrupadas em *clusters* de acordo com a intensidade de correlação.

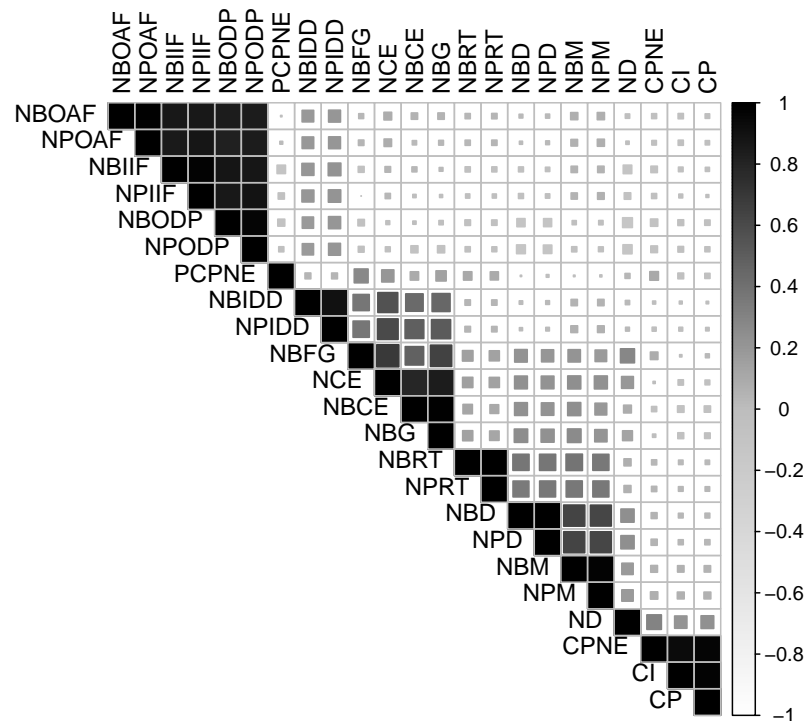


Figura 8 – Matriz correlograma

Para que seja empregada de forma correta, a AF requer que as variáveis estejam minimamente correlacionadas. Neste sentido, [Larose e Larose \(2015, p. 112\)](#) recomendam a realização do teste de *Esfericidade de Bartlett* para determinar se existe um nível mínimo de correlação como fase preliminar à aplicação da AF. Seguindo este autor, o referido teste foi realizado nos dados padronizados das variáveis do *dataset* em questão. O resultado é apresentado na listagem 2.



---

```
$chisq
[1] 469552.9

$p.value
[1] 0

$df
[1] 300
```

---

#### Listagem 2 – Teste de Esfericidade de Bartlett

O teste de Bartlett verifica a *hipótese nula* de que a matriz de correlação é uma matriz identidade, sendo as variáveis portanto não correlacionadas. A principal informação deste teste é o valor de  $p$  (ou *p-value*). Valores muito acima de 0.10 indicam que a aplicação do AF pode não ser adequada. Por outro lado valores pequenos indicam que há evidências contra a hipótese nula, consequentemente validando o método. Como observado na saída acima, o *p-value* é 0 o que indica que o AF pode ser aplicada.

Na listagem 3 são apresentados os dois resultados da aplicação da função `fa` sobre os dados padronizados. O primeiro resultado contém os *eigenvalues* de cada variável preditora computada. O segundo contém os carregamentos (ou *loadings*) juntamente à proporção de variância e variância acumulada.

---

```
[1] 6.82829446 5.06102146 3.02305954 2.57594785 0.89229036 0.73263127
    0.44322564 0.32873650 0.12874448
[10] 0.09205939 0.02032373 -0.01542126 -0.02141875 -0.04602155 -0.05979722
    -0.08471119 -0.09794112 -0.11387848
[19] -0.13637542 -0.23237097 -0.27069873 -0.29403878 -0.29933457 -0.31091515
    -0.65475511
```

Loadings:

	PA1	PA2	PA3	PA4
CI		0.129	0.844	0.451
CP		0.128	0.866	0.470
NBFG	0.453	0.382	-0.135	0.247
NBCE	0.552	0.415	-0.294	0.297

NBG	0.596	0.461	-0.299	0.332
NCE	0.652	0.446	-0.285	0.365
NBDP	0.551	-0.753		
NPODP	0.545	-0.772	0.100	
NBIIF	0.613	-0.713	0.110	
NPIIF	0.614	-0.715	0.135	
NBOAF	0.641	-0.650	0.122	
NPOAF	0.633	-0.654	0.129	
CPNE		0.172	0.835	0.469
PCPNE		0.123		
NBIDD	0.583	0.103	-0.241	0.406
NPIDD	0.628	0.108	-0.267	0.431
ND	0.125	0.258	0.222	
NBM	0.462	0.423	0.264	-0.463
NPM	0.452	0.407	0.278	-0.463
NBD	0.420	0.493	0.242	-0.495
NPD	0.422	0.486	0.250	-0.498
NBRT	0.307	0.322	0.167	-0.342
NPRT	0.300	0.312	0.170	-0.340
CPCC	0.958	0.248		
CPCF	0.862	0.219		

	PA1	PA2	PA3	PA4
SS loadings	6.828	5.061	3.023	2.576
Proportion Var	0.273	0.202	0.121	0.103
Cumulative Var	0.273	0.476	0.596	0.700

---

Listagem 3 – Carregamentos da AF: o primeiro fator demonstra similaridade com o primeiro componente da ACP o que confirma a solidez e consistência das constatações até aqui verificadas.

Observa-se que os quatro primeiros fatores computam uma variância de 68,4% – ligeiramente menor quando comparado à variância acumulada da mesma quantidade de componentes principais na fase de aplicação da ACP que foi de 72,3%. É verificado também que, assim

como na ACP, o primeiro fator é o que possui a maior variância.

Quanto a seleção de fatores, o mesmo critério foi utilizado em relação à escolha dos componentes na ACP: *Scree plot* (HAIR et al., 1998, p. 109). O gráfico da figura 9 demonstra que a partir do 5º fator há uma tendência à estabilização em relação à variância explicada pelos fatores individuais.

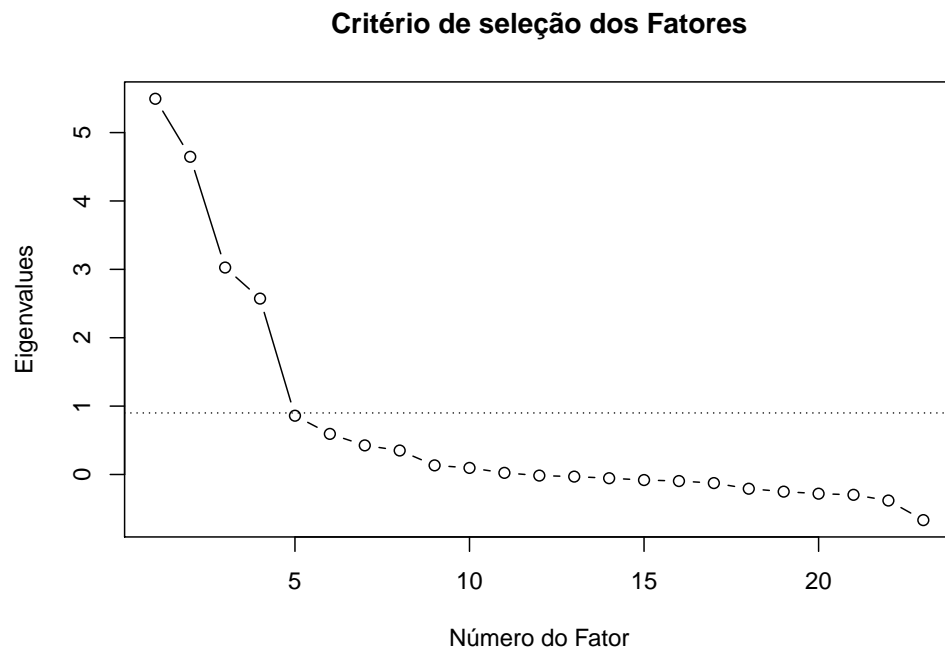


Figura 9 – Critério de seleção dos fatores: a partir do 5º fator, observa-se uma tendência de horizontalização da curva de variabilidade explicada pelos fatores, o que torna os fatores a partir desta contagem menos suscetíveis em participar de uma seleção.

#### 4.5.1 Rotação dos Fatores

A rotação dos fatores consiste em um elemento de validação da Análise do Fator. De acordo com Hair et al. (1998, p. 112), a rotação serve para redistribuir as proporções de variância residuais dos fatores mais afastados para os iniciais. Larose e Larose (2015, p. 114) faz a analogia de um cientista tentando obter maior resolução em uma imagem ao ajustar o foco de seu microscópio.

---

Loadings:

PA2	PA1	PA4	PA3
CI			0.965

CP			0.993	
NBFG	0.627	0.176		
NBCE	0.791	0.151		
NBG	0.857	0.168		
NCE	0.900	0.168		
NBODP	0.934			
NPODP	0.947			
NBIIF	0.947			
NPIIF	0.953			
NBOAF	0.919			
NPOAF	0.918			
CPNE			0.972	
PCPNE	0.169			
NBIDD	0.215	0.725		
NPIDD	0.233	0.779		
ND	0.133	0.212	0.255	
NBM	0.122	0.811		
NPM	0.104	0.804		
NBD	0.114	0.841		
NPD	0.108	0.843		
NBRT		0.580		
NPRT		0.571		
CPCC	0.395	0.741	0.524	
CPCF	0.358	0.667	0.468	

	PA2	PA1	PA4	PA3
SS loadings	5.672	4.814	4.056	2.946
Proportion Var	0.227	0.193	0.162	0.118
Cumulative Var	0.227	0.419	0.582	0.700

---

Listagem 4 – Carregamentos da AF após rotação

A rotação de fatores pode ser entendida como o esforço para obtenção da melhor confi-

guração de variâncias no plano cartesiano. A listagem 4 demonstra a rotação pelo método *Varimax* que maximiza a variabilidade para cada coluna. Pela comparação com a listagem 3, percebe-se que variabilidades maiores são obtidas após a rotação.

## 5 RESULTADOS

Por meio da aplicação da análise do componente principal foi possível identificar características relevantes sobre as variáveis preditoras do CPC. A referida análise também possibilitou o agrupamento destas variáveis em componentes principais. Verificou-se que o primeiro (principal) componente tem afinidade com a estrutura pedagógica do curso (Organização Didático-pedagógica), sua estrutura (Infra-estrutura e Instalações) e perspectiva de futuro do ponto de vista do aluno (Oportunidades de Ampliação da Formação). O desempenho nestes quesitos tem origem em avaliação subjetiva do aluno participante do censo. A classificação destas variáveis como constantes no primeiro componente permite afirmar que um maior desempenho nestes quesitos implicará em maior resposta na variável dependente (CPC Contínuo).

Foi possível verificar notável semelhança nos resultados das análises multivariadas AF e ACP no que se refere a composição do conjunto de variáveis que representam a redução de dimensionalidade. Nesta pesquisa, constatou-se o caráter corroborativo da AF em relação ao ACP pelo fato de os perfis e fatores terem apresentado magnitudes significativas em variáveis semelhantes.

A ACP e AF permitiram identificar as variáveis com maior participação na formação do CPC. A aplicação dessas duas metodologias possibilitou verificar que as variáveis relativas às percepções dos alunos são as que mais influenciam na formação do CPC. As variáveis *Organização Didático-pedagógica*, *Infra-estrutura e Instalações* e *Oportunidades de Ampliação da Formação* constituem o componente principal na ACP (pelo acúmulo de 1/4 do espectro de variabilidade) e o fator na AF (pelo acúmulo de 22,7% da variabilidade).

No segundo componente da ACP, em relação aos alunos, tem maiores pesos as variáveis *Formação Geral* e *Conhecimentos Específicos*. Já em relação aos professores tem destaque as variáveis *Quantidade de Mestres* e *Quantidade de Doutores* nos cursos. Este componente contabiliza 21,6% da variabilidade dos dados da ACP e 20,2% na AF.

O terceiro componente (ou fator) é composto, tanto na ACP quanto na AF, pelas variáveis *Concluíntes Inscritos*, *Concluíntes Participantes* e *Concluíntes Participantes com Nota no Enem*. Na ACP, o componente explica 13,9% da variabilidade e 12,1% AF.

O quarto fator (ou componente) é constituído na AF e na ACP pelas variáveis *Concluíntes Inscritos*, *Concluíntes Participantes* e *Concluíntes Perticipantes com Nota no Enem* e nota no *Indicador de Diferença de Desempenho*. Na AF o fator correspondente explica 10,3% da variabilidade. Já na ACP, 12,3%.

## 6 CONCLUSÃO

Apesar de a educação ser um fator essencial para o desenvolvimento de uma nação, o Brasil ainda desponta como uma das nações com maior proporção de adultos que não tiveram a educação primária (OCDE, 2017, p. 44). Por esse motivo, acredita-se que as contribuições ao tema, por menores que sejam, são valiosas. Ainda que se reconheça a relevância das fases de base da educação, bem como a variedade de instrumentos de avaliação das mesmas, buscou-se neste trabalho analisar a sistemática de avaliação do ensino superior.

Procurou-se, fornecer um instrumento informativo, possivelmente fundamentador de tomada de decisão aos atores responsáveis, direta e indiretamente pela gestão das unidades de ensino superior e cursos de graduação. À comunidade acadêmica das IES, foram apresentados os principais elementos que devem ser estimulados, desenvolvidos, e mantidos de forma a lograr qualidade na avaliação do Enade.

As aplicações de metodologias estatísticas deram sustentação ao desenvolvimento de respostas às perguntas e hipóteses previamente formuladas para a elaboração desta pesquisa. Neste sentido, a primeira das principais apurações é a de que a dimensão do Enade que incorpora a opinião do estudante incide fortemente na formação do CPC. Nessa perspectiva, acredita-se que existem dados de dimensões e variáveis latentes que ainda podem ser extraídos.

Verificou-se que a qualidade do ensino dos cursos de graduação resulta, principalmente:

- i) do adequado ajustamento da estrutura curricular levando em consideração a utilidade dos conteúdos ao exercício da profissão;
- ii) do ambiente onde se desenvolvem as atividades de ensino;
- iii) dos estímulos a continuidade dos estudos após a graduação;
- iv) da quantidade de mestres e doutores presentes dos cursos.

Em relação à primeira dimensão, é possível afirmar que fatores como coerência da estrutura curricular com os objetivos dos cursos; adequação e atualização das ementas e das disciplinas; uma seleção de conteúdos satisfatória; adequação, atualização e relevância da bibliografia são favoráveis à melhora do indicador de qualidade do curso de graduação.



No que se refere à Infra-estrutura e Instalações podem ser citados elementos como: ambientes projetados para atender a todos os requisitos necessários para a realização das atividades de ensino; que levem em consideração o conforto não apenas dos discentes, mas de toda comunidade acadêmica e que sejam atentos para questões relacionadas à acessibilidade. Aliado a estes, sistema de segurança, iluminação, ventilação, equipamentos e mobiliários adequados, podem impactar de forma significativa no CPC.

A promoção de estímulos à ampliação da formação dos discentes são fatores igualmente importantes na composição do CPC. Devem ser criados meios para que, durante o período da graduação, sejam apresentados os caminhos e possibilidades de prosseguimento dos estudos aos alunos. É seguro afirmar que os esforços seriam irrisórios quando comparados aos resultados favoráveis oriundos de avanços e contribuições em relação ao desenvolvimento de pesquisas.

Finalmente, aos desenvolvedores de políticas educacionais, busca-se por meio desta material viabilizar conteúdo informacional para que seja promovida a adequada ampliação do ensino superior pelas diversas classes da sociedade sem, contudo, negligenciar sua qualidade.

## Referências

- BARBOSA, G. d. C.; FREIRE, F. d. S.; CRISÓSTOMO, V. L. Análise dos indicadores de gestão das ifes e o desempenho discente no enade. Jul 2011. Disponível em: <http://repositorio.unb.br/handle/10482/14518>.
- BERNHARDT, V. L. *Data Analysis for Comprehensive Schoolwide Improvement*. [S.l.]: Eye On Education, 6 Depot Way West, Larchmont, NY 10538 (\$29, 1998. ISBN 978-1-883001-57-5).
- BERTOLIN, J. C. G.; MARCON, T. O (des)entendimento de qualidade na educação superior brasileira – das quimeras do provão e do enade à realidade do capital cultural dos estudantes. *Avaliação: Revista da Avaliação da Educação Superior*, v. 20, n. 1, 2015. ISSN 1414-4077. Disponível em: <http://www.redalyc.org/resumen.oa?id=219138341007>.
- BRITO, F. d.; REGINA, M. O sinais e o enade: da concepção à implantação. *Avaliação: Revista da Avaliação da Educação Superior*, v. 13, n. 3, 2008. ISSN 1414-4077. Disponível em: <http://www.redalyc.org/resumen.oa?id=219114874014>.
- COELHO, É. d. O. P. O papel das variáveis na eficiência da nota enade dos cursos de engenharia de produção nas instituições públicas de ensino superior. 2016.
- COSTA, E. M. et al. Efficiency and performance in higher education: a frontier analysis of the educational productivity of the brazilian federal institutes of higher education. *Revista de Economia Contemporânea*, v. 16, n. 3, p. 415–440, Dec 2012. ISSN 1415-9848.
- HAIR, J. F. et al. *Multivariate data analysis*. [S.l.]: Prentice hall Upper Saddle River, NJ, 1998. v. 5.
- INEP. Nota técnica nº 38/2017/cgcqes/daes. 2017. Disponível em: [http://download.inep.gov.br/educacao\\_superior/enade/notas\\_tecnicas/2016/nota\\_tecnica\\_n38.2017\\_cgcqes\\_daes\\_calculo\\_cpc.pdf](http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2016/nota_tecnica_n38.2017_cgcqes_daes_calculo_cpc.pdf).
- INEP. Avaliação in loco - glossário dos instrumentos de avaliação externa. Jul 2018. Disponível em: [http://download.inep.gov.br/educacao\\_superior/avaliacao\\_institucional/apresentacao/glossario\\_2edicao\\_25072018.pdf](http://download.inep.gov.br/educacao_superior/avaliacao_institucional/apresentacao/glossario_2edicao_25072018.pdf).
- JOSSE, J.; HUSSON, F. et al. missmda: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, Foundation for Open Access Statistics, v. 70, n. 1, p. 1–31, 2016.
- JUNIOR, V. d. P. A. *O processo de formação das políticas de avaliação da educação superior no Brasil (1983-1996)*. Tese (Doutorado) — Universidade Estadual de Campinas, 2004.

- LAROSE, D. T.; LAROSE, C. D. *Data mining and predictive analytics*. [S.l.]: John Wiley & Sons, 2015.
- MENDES, M. *Boletim nº 26, de 2015 - A Despesa Federal em Educação: 2004-2014 — Publicações Portal*. Senado Federal, 2015. Disponível em: <https://www12.senado.leg.br/publicacoes/estudos-legislativos/tipos-de-estudos/boletins-legislativos/bol26/view>.
- OCDE. *Education at a Glance 2017: OECD Indicators*. [s.n.], 2017. Disponível em: <http://dx.doi.org/10.1787/eag-2017-en>.
- REVELLE, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois, 2017. R package version 1.7.8. Disponível em: <https://CRAN.R-project.org/package=psych>.
- TARR, G. *pairsD3: D3 Scatterplot Matrices*. [S.l.], 2015. R package version 0.1.0. Disponível em: <https://CRAN.R-project.org/package=pairsD3>.
- TORGO, L. *Data Mining with R, learning with case studies*. Chapman and Hall/CRC, 2010. Disponível em: <http://www.dcc.fc.up.pt/~ltorgo/DataMiningWithR>.
- VERHINE, R. E.; DANTAS, L. M. V.; SOARES, J. F. Do provão ao enade: uma análise comparativa dos exames nacionais utilizados no ensino superior brasileiro. 2006.
- WEI, T.; SIMKO, V. *R package "corrplot": Visualization of a Correlation Matrix*. [S.l.], 2017. (Version 0.84). Disponível em: <https://github.com/taiyun/corrplot>.
- WICKHAM, H. et al. *dplyr: A Grammar of Data Manipulation*. [S.l.], 2017. R package version 0.7.4. Disponível em: <https://CRAN.R-project.org/package=dplyr>.
- ZANDAVALLI, C. B. Avaliação da educação superior no brasil: os antecedentes históricos do sinaes evaluation of higher education in brazil: the historical precedents of sinaes. 2009. Disponível em: <http://www.ingentaconnect.com/content/doi/14144077/2009/00000014/00000002/art00008>.
- ZOGHBI, A. C.; ROCHA, F.; MATTOS, E. Education production efficiency: Evidence from brazilian universities. *Economic Modelling*, v. 31, p. 94–103, Mar 2013. ISSN 0264-9993.