

**UNIVERSIDADE MUNICIPAL DE SÃO CAETANO DO SUL  
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO  
MESTRADO**

**Marcio Alexandre Freire**

**FATORES QUE CONTRIBUÍRAM PARA QUE AS IES PRIVADAS  
BRASILEIRAS SE DISTANCIASSEM DE SEU PONTO DE EQUILÍBRIO  
FINANCEIRO: um estudo baseado em dados do Censo da Educação  
Superior de 2018 a 2020**

**São Caetano do Sul**

**2023**

**MARCIO ALEXANDRE FREIRE**

**FATORES QUE CONTRIBUÍRAM PARA QUE AS IES PRIVADAS  
BRASILEIRAS SE DISTANCIASSEM DE SEU PONTO DE EQUILÍBRIO  
FINANCEIRO: um estudo baseado em dados do Censo da Educação  
Superior de 2018 a 2020**

Dissertação apresentada ao Programa de Pós-Graduação em Administração da Universidade Municipal de São Caetano do Sul como requisito para a obtenção do título de Mestre em Administração.  
Área de Concentração: Gestão e Regionalidade

Orientador: Professor Doutor Marco Antonio Pinheiro da Silveira

**São Caetano do Sul**

**2023**

## FICHA CATALOGRÁFICA

FREIRE, Marcio Alexandre

Fatores que contribuíram para que as IES privadas brasileiras se distanciassem de seu ponto de equilíbrio financeiro: um estudo baseado em dados do Censo da Educação Superior de 2018 a 2020 / Marcio Alexandre Freire. – São Caetano do Sul: USCS, 2023.

133p.

Orientador: Marco Antonio Pinheiro da Silveira

Dissertação (mestrado) – USCS, Universidade Municipal de São Caetano do Sul, Programa de Pós-Graduação em Administração, 2023.

1. Ensino superior. 2. Ponto de equilíbrio. 3. Indicadores financeiros. 4. Censo educacional. 5. Redes organizacionais e inovação I. Título. II. Universidade Municipal de São Caetano do Sul.

**Reitor da Universidade Municipal de São Caetano do Sul**  
Prof. Dr. Leandro Campi Prearo

**Pró-reitora de Pós-graduação e Pesquisa**  
Prof.<sup>a</sup> Dra. Maria do Carmo Romeiro

**Gestor do Programa de Pós-graduação em Administração**  
Prof. Dr. Eduardo de Camargo Oliva

Dissertação defendida e aprovada em 03/04/2023 pela Banca Examinadora constituída pelos professores:

Prof. Dr. Marco Antonio Pinheiro da Silveira - orientador  
Universidade Municipal de São Caetano do Sul

Profa. Dra. Luisa Veras de Sandes Guimarães  
Universidade Municipal de São Caetano do Sul

Prof. Dr. Emerson Antonio Maccari  
Associação Educacional Nove de Julho - UNINOVE

## **AGRADECIMENTOS**

Agradeço ao incentivo de minha família, esposa e aos colegas de trabalho que sempre me desafiaram a continuar. Aos professores do Programa de Pós-graduação da Universidade Municipal de São Caetano do Sul e a todos os funcionários que contribuíram para viabilizar a realização da pesquisa de Mestrado.

Em particular, gostaria de expressar minha mais sincera e profunda gratidão ao meu orientador, Prof. Dr. Marco Antonio Pinheiro da Silveira, pela paciência e orientações, que me ajudaram a conduzir a pesquisa.

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.” (Arthur Schopenhauer, 1851, p.172)

FREIRE, Marcio Alexandre. **Fatores que contribuíram para que as IES privadas brasileiras se distanciassem de seu ponto de equilíbrio financeiro:** Um estudo baseado em dados do Censo da Educação Superior de 2018 a 2020. Universidade Municipal de São Caetano do Sul. São Caetano do Sul, SP, 2023.

## RESUMO

No ensino superior privado brasileiro, manter o equilíbrio entre a receita adquirida pelas mensalidades e o custo de prestação do serviço educacional é necessário para alcance da sustentabilidade financeira. Este estudo avaliou, na população de mantenedoras de ensino superior privado brasileiro e em seus cursos de graduação, fatores que influenciaram o distanciamento, positivo ou negativo, do ponto de equilíbrio financeiro utilizando os valores informados de receitas, despesas e indicadores operacionais das mantenedoras. Também avaliou um comparativo dos impactos destes fatores identificados entre grandes grupos consolidadores e as instituições de menor porte tanto nas modalidades de ensino presencial, quanto a distância. Por meio de revisão bibliográfica, foi mostrado o desenvolvimento da rede privada de ensino superior, a concentração do mercado educacional, a definição dos custos de operação e os indicadores que compõem o ponto de equilíbrio financeiro, além dos métodos estatísticos utilizados. Como metodologia de análise, após extração, transformação e consolidação das informações utilizando dados secundários do Censo da Educação Superior de 2018 a 2020, foram utilizadas técnicas de estatística descritiva, regressão linear múltipla, análise de componentes principais e florestas aleatórias, nas variáveis transformadas e definidas como base de dados. Este trabalho avaliou, como pergunta-problema: Quais fatores contribuíram para que as IES se distanciassem do ponto de equilíbrio financeiro, no período dos censos 2018 a 2020? Foram identificados como fatores importantes para o distanciamento do ponto de equilíbrio o valor médio da mensalidade, o custo de folha docente e os custos de operação e, em menor grau, as razões entre alunos e docentes ou funcionários. Adicionalmente, foi verificada a relação em que o porte das instituições ou o pertencimento aos grandes grupos consolidadores contribuiu como fator no distanciamento do ponto de equilíbrio financeiro.

**Palavras-chave:** ensino superior; ponto de equilíbrio; indicadores financeiros; censo educacional; redes organizacionais e inovação.



FREIRE, Marcio Alexandre. **Factors Contributing to the Brazilian Private Higher Education Institutions' Deviation from their Financial Break-Even Point: A Study Based on the 2018-2020 Higher Education Census Data.** Universidade Municipal de São Caetano do Sul. São Caetano do Sul, SP, 2023.

## **ABSTRACT**

In Brazilian private higher education, maintaining a balance between the revenue acquired through tuition fees and the cost of providing educational services is necessary to achieve financial sustainability. This study evaluated factors that influenced positive or negative deviation from the financial equilibrium point in the population of private higher education maintainers and their undergraduate courses, using reported values of revenue, expenses, and operational indicators of maintainers. A comparison of the impacts of these identified factors between large consolidating groups and smaller institutions was also evaluated for both "face-to-face" and "distance" learning modalities. A literature review was conducted on the development of the private higher education network, the concentration of the educational market, the definition of operating costs, and the indicators that make up the financial equilibrium point, as well as the statistical methods used. As a method of analysis, after extraction, transformation, and consolidation of the information using secondary data from the higher education census from 2018 to 2020, descriptive statistics, multiple linear regression, principal component analysis, and random forests techniques were used in the transformed variables defined as the database. Therefore, this study evaluated, as a problem question: What factors contributed to IES deviating from the financial equilibrium point during the 2018 to 2020 census period? The average tuition fee, the cost of the teaching staff, and the operating costs were identified as important factors for deviating from the equilibrium point, and to a lesser extent, the ratios between students and teachers or employees. Additionally, a relationship was found in which the size of institutions or membership in large consolidating groups contributed as a factor in the deviation from the financial equilibrium point.

**Keywords:** higher education; break-even point; financial indicators; educational census; organizational networks; and innovation.

## LISTA DE FIGURAS

Figura 1 - Censo 2020: comparativo entre portes e matrículas.....	50
Figura 2 - Hiperplano de variáveis em regressão múltipla .....	56
Figura 3 - Exemplo de Floresta Randômica.....	61
Figura 4 - Correlograma das variáveis - 2018.....	69
Figura 5 - Correlograma das variáveis - 2019.....	71
Figura 6 - Correlograma das variáveis - 2020.....	73
Figura 7 - Gráfico dos componentes principais - 2018.....	75
Figura 8 - Gráfico dos componentes principais 2019.....	76
Figura 9 - Gráfico dos componentes principais 2020.....	77
Figura 10 - Análise de qualidade de regressão e resíduos .....	80
Figura 11 - Importância das variáveis independentes por ano censo - RF .....	83
Figura 12 - Árvore de Decisão.....	84
Figura 13 - Comparativo da Variável Ponto de Equilíbrio .....	87
Figura 14 - Comparativo de ponto de equilíbrio após remoção de outliers .....	88
Figura 15 - Comparativo EAD x Custo Docente/Aluno.....	93
Figura 16 - Comparativo EAD x Ticket Médio/Aluno .....	94
Figura 17 - Comparativo EAD x Ponto de Equilíbrio .....	95

## LISTA DE TABELAS

Tabela 1 - Evolução do ensino superior privado (1997-2021) .....	31
Tabela 2 - Grupos consolidadores no ensino superior em 2021.....	33
Tabela 3 - EAD nos grupos consolidadores .....	34
Tabela 4 - Registros totais da base de dados do Censo da Educação Superior .....	45
Tabela 5 - Validação dos registros importados.....	45
Tabela 6 - Total de mantenedoras privadas por porte .....	49
Tabela 7 - Total de matrículas privadas por porte .....	49
Tabela 8 - Estatística descritiva da base de dados .....	64
Tabela 9 - Redução amostral na análise de outliers.....	65
Tabela 10 - Estudo dos registros outliers .....	65
Tabela 11 - Bases de dados após tratamento de outliers.....	66
Tabela 12 - Intervalos de confiança de correlação - VDPEQ.....	67
Tabela 13 - Avaliação de multicolinearidade - VIF .....	68
Tabela 14 - Análise dos componentes principais - 2018 .....	74
Tabela 15 - Análise dos componentes principais - 2019 .....	76
Tabela 16 - Análise de componentes principais - 2020 .....	77
Tabela 17 - Resultado das regressões por variável e ano.....	78
Tabela 18 - Resultados da regressão .....	79
Tabela 19 - Resultado da análise por Florestas Aleatórias .....	82
Tabela 20 - Comparativo de quantidade de mantenedoras .....	86
Tabela 21 - Quantidade de mantenedoras pós-outliers.....	88
Tabela 22 - Resumo das descobertas, métodos e variáveis .....	89
Tabela 23 - Análise de regressão por estrato de porte.....	91
Tabela 24 – Comparativo: fatores e grupos educacionais.....	97
Tabela 25 - Comparativo Fatores e Ponto de Equilíbrio .....	101

## LISTA DE QUADROS

Quadro 1 - Formação das primeiras faculdades e cursos superiores no Brasil .....	26
Quadro 2 - Legislações do Ensino Superior Brasileiro – séculos XIX e XX.....	29
Quadro 3 - Margem contábil/financeira.....	38
Quadro 4 - Variáveis disponíveis na base do censo .....	39
Quadro 5 - Relação agrupador contábil com variáveis da base do Censo.....	40
Quadro 6 - Variáveis transformadas .....	46
Quadro 7 - Estratificação das mantenedoras.....	48
Quadro 8 - Variáveis obtidas no censo .....	51
Quadro 9 - Variáveis dependentes .....	52
Quadro 10- Variáveis independentes .....	52
Quadro 11 - Dicionário de Variáveis do Censo da Educação Superior – IES .....	114
Quadro 12 - Variáveis do Censo da Educação Superior - Financeiros (Sala Segura) .....	132

## LISTA DE ABREVIATURAS E SIGLAS

ABED	Associação Brasileira de Ensino a Distância
ABNT	Associação Brasileira de Normas Técnicas
CADE	Conselho Administrativo de Defesa Econômica
Cibec	Centro de Informação e Biblioteca em Educação
EAD	Ensino a Distância
IBGE	Instituto Brasileiro de Geografia e Estatística
IQR	<i>Interquartile Range</i> (Intervalo Interquartil)
IES	Instituição de Ensino Superior
MEC	Ministério da Educação e Cultura
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
RLM	Regressão Linear Multivariada
SEMESP	Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo
VIF	<i>Variance Inflation</i> (Fator de multicolinearidade)

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>17</b>
1.1 Problema da pesquisa .....	19
1.2 Objetivos da pesquisa.....	20
1.3 Delimitação do estudo .....	21
1.4 Justificativa e relevância do trabalho .....	21
1.5 Organização do relatório do trabalho .....	22
1.6 Contribuições da pesquisa.....	23
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>25</b>
2.1 O ensino superior brasileiro .....	25
2.2 Mercado educacional e sustentabilidade financeira .....	30
2.3 Aplicação educacional do ponto de equilíbrio financeiro .....	36
2.4 Ponto de equilíbrio contábil e indicadores fornecidos pelo Censo.....	38
<b>3 PROCEDIMENTOS METODOLÓGICOS</b> .....	<b>43</b>
3.1 Dados secundários e base de dados do Censo da Educação Superior .....	43
3.2 Bases de dados e definição das variáveis do estudo.....	44
3.3 Tratamento e agrupamento dos dados .....	47
3.4 Criação das variáveis de regressão.....	51
3.5 Referencial dos métodos estatísticos utilizados .....	54
3.5.1 Regressão Linear Multivariada .....	54
3.5.2 Análise de Componentes Principais .....	58
3.5.3 Análise por Floresta Aleatória ( <i>Random Forest</i> ) .....	59
<b>4 RESULTADOS OBTIDOS</b> .....	<b>63</b>
4.1 Bases de dados.....	63
4.2 Remoção de dados extremos ou inválidos - <i>Outliers</i> .....	64
4.3 Análise de correlações e covariâncias nas bases de dados .....	66
4.4 Análise de correlação e colinearidade .....	67
4.5 Análise de correlação – 2018 .....	69
4.6 Análise de correlação – 2019 .....	70
4.7 Análise de correlação – 2020 .....	72
4.8 Análise de Componentes Principais (PCA).....	74
4.9 Regressão Linear Múltipla da base de dados .....	78
4.10 Análise por Florestas Aleatórias ( <i>Random Forest</i> ).....	81

<b>5 DISCUSSÃO DOS RESULTADOS.....</b>	<b>85</b>
5.1 Visão geral do processo de geração das bases de dados.....	86
5.2 Resumo das descobertas dos métodos de análise utilizados.....	89
5.3 Segmentação por porte e análise das variáveis .....	91
5.4 Impacto do ensino a distância no custo e na receita .....	92
5.5 Os grandes grupos e o comportamento dos fatores .....	94
5.6 Comportamento dos fatores e o ponto de equilíbrio .....	100
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>103</b>
6.1 Lacunas e futuras pesquisas .....	105
<b>REFERÊNCIAS .....</b>	<b>113</b>
<b>ANEXO A – Dicionários de dados das variáveis do Censo da Educação</b>	
<b>Superior .....</b>	<b>114</b>
<b>ANEXO B – Fontes dos algoritmos utilizados - R.....</b>	<b>133</b>





## 1 INTRODUÇÃO

Com o aumento da competição no mercado educacional, na educação básica e superior e principalmente nas instituições privadas brasileiras, manter o equilíbrio entre a receita adquirida – que depende fundamentalmente das matrículas ativas e mensalidades recebidas – e o custo de prestação do serviço educacional é regra de sustentabilidade financeira. Ademais, cada vez mais relevante é o esforço de digitalização das metodologias educacionais e sua gestão, para medir resultados e adaptar-se ao ensino remoto (síncrono e assíncrono) e híbrido (parte presencial, parte remoto) na busca dessa sustentabilidade.

A importância da sustentabilidade, nas figuras da margem de contribuição e do ponto de equilíbrio, – entre outros indicadores –, foi avaliada por Lugoboni (2017) em uma pesquisa qualitativa que considerou uma amostra de 109 instituições de ensino, com 4,5% de taxa amostral em relação à população de 2.416 IES presentes no Censo da Educação Superior 2012 e teve como objetivo a qualificação dos principais indicadores presentes na gestão das IES. Nesse trabalho, o autor ressaltou que a situação financeira das instituições de ensino também pode ser afetada por diversos outros fatores, como a queda no número de alunos, a falta de investimentos em infraestrutura e tecnologia, a má gestão dos recursos financeiros, entre outros.

Gestores das IES responderam a essa pesquisa, em que avaliaram os indicadores utilizados na dimensão financeira e destacaram os indicadores de Receitas Financeiras, Margem ou Rentabilidade, além de considerarem como de grande importância itens como Custo Docente, Administrativo e Despesas de Operação.

Adicionalmente, a Sinopse Estatística do Ensino Superior, publicada em fevereiro de 2022 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (INEP, 2022), destaca pela primeira vez, tanto nas instituições públicas quanto no mercado privado e desde o início do acompanhamento dos indicadores de fluxo do Censo da Educação Superior, que o número de ingressantes em modalidade a distância superou o número de ingressos na modalidade presencial. Esse resultado é complementado pelas informações de que mais de 87% das IES são pertencentes à rede privada, sendo boa parte delas lucrativas.

Ainda em 2016, o Conselho Administrativo de Defesa Econômica – Cade (2016) reforçou as metodologias de análise de concentração de mercado, movimento iniciado em 2005, principalmente no âmbito da prestação do serviço na modalidade a distância. O reconhecimento das instituições na busca de escala para a redução dos custos é destacado na publicação e fato reconhecido pelas instituições.

Além do exposto anteriormente, com a divulgação de notícias por organismos do setor, conseguimos ter uma percepção de queda significativa nos indicadores principais de instituições de ensino, principalmente nas menores, como totalidade do corpo docente ativo, novos ingressos e dificuldades na busca da lucratividade financeira, principalmente devido à popularização do ensino a distância em grupos de maior poder econômico.

Expostos todos esses motivadores, tem-se como pergunta-problema: Quais fatores contribuíram para que IES privadas brasileiras se distanciassem de seu ponto de equilíbrio financeiro no período dos censos 2018 a 2020?

Com esse cenário, este trabalho avaliou, no período compreendido do ano referência dos censos do ensino superior de 2018 até 2020 – cuja data de referência é 31 de dezembro de 2020 –, fatores que afetaram o equilíbrio financeiro das mantenedoras e seus cursos de graduação. Analisaram-se, entre outros indicadores, o impacto do custo de prestação de serviços das IES privadas e a sua diluição em número cada vez maior de alunos, a gestão de despesas e a interiorização da oferta (busca de mercados externos à sede principal da IES, seja pelo Ensino a Distância, seja pela abertura de polos avançados) avaliados através de conceitos como ponto de equilíbrio financeiro e seus componentes – receitas e despesas –, buscando-os, para a composição das variáveis, na base de dados do censo, no mesmo período citado.

Para responder à pergunta-problema e atingir objetivo específico deste trabalho, foram utilizados os valores financeiros de receitas e despesas, calculando assim o ponto de equilíbrio. Quanto às matrículas, foram usados totalizadores de matrículas ativas, nas modalidades tanto curso a distância quanto presencial, pelos recortes de porte institucional. Também foi apresentada a avaliação da transferência de matrículas entre as IES privadas, proporcionalizando os portes das instituições – estudo da consolidação – e a migração entre as modalidades de ensino, presencial e a distância.

Para extração e definição das variáveis da base de dados, foram utilizados dados secundários, obtidos nos microdados do Censo Educacional Superior 2018, 2019 e 2020 (INEP, 2022), publicados pelos órgãos oficiais e utilizados como resultados de suas coletas e pesquisas.

Justifica-se o uso dessas bases por terem volume adequado de dados, atingindo toda a população de mantenedoras, e serem igualitárias, pois todas as IES selecionadas no procedimento metodológico devem utilizá-las e publicar seus resultados regularmente, passando por etapas de validação e conferência pelo órgão competente.

### 1.1 Problema da pesquisa

Com a divulgação de pesquisas empíricas e notícias divulgadas por organismos do setor, conseguimos ter uma percepção de queda significativa nos indicadores principais de instituições de ensino, como corpo docente ativo, ingressos no ensino presencial e financeiros como receitas com as mensalidades.

Na educação, tanto básica quanto superior e principalmente nas instituições privadas brasileiras, cuja receita depende fundamentalmente das matrículas ativas e mensalidades, em pouco tempo mudar o modelo de ensino, digitalizar a forma de ensinar, medir resultados e até mesmo mudar o relacionamento com o corpo docente, treinando-os para o ensino remoto (síncrono) e híbrido (parte presencial, parte remoto), fora as mudanças legais e de regime laboral, impostas ou recomendadas nestes cenários.

Ainda no cenário do período abarcado por este trabalho, em parte, temos as recomendações de isolamento social devido a declaração de situação pandêmica, que significaram o fechamento de escolas e faculdades em suas atividades presenciais e a recomendação de transformação das atividades presenciais em atividades remotas, conforme as portarias 343 e 345, publicadas em março de 2020. (BRASIL, MEC, 2020; BRASIL, MEC, 2020)

As ações governamentais em relação ao ensino e as instituições educacionais foram a permissão e substituição de aulas presenciais por aulas remotas principalmente no Ensino Superior, reconhecimento de atividades externas e remotas como carga horária integralizada do curso, integralização do estágio obrigatório antecipada em áreas de formação em Saúde, redução de carga tributária,

treinamentos virtuais para docentes e autorização para que defesas de teses e dissertações de mestrado e doutorado pudessem ser realizadas virtualmente. A política instada de emergência pelo governo federal, juntamente com as ações de estados e municípios, quando olhado para a Educação Básica, tinha como objetivo permitir que todos os estudantes, mesmo isolados e cumprindo a contingência pela emergência sanitária instaurada, continuassem os estudos e não perdessem o calendário escolar já comprometido.

Um dos efeitos que se observa no mercado educacional é a percepção de perda de receita pela redução de alunos no ensino presencial com a migração gradativa para o ensino a distância, redução do lucro pela evasão e mensalidades mais baixas além da queda no número de novos ingressantes nas instituições denominadas de pequeno porte. Os dados da pesquisa qualitativa organizada pelo SEMESP (Sindicato das Entidades Mantenedoras de Ensino Superior) indicam que 858.000 alunos deixaram o Ensino Superior, por evasão ou por não ingressarem no segundo semestre de 2020, o que representou 13,2% do total de matrículas privadas no país. Os cursos presenciais, fortemente afetados pela pandemia, representaram 40% da queda registrada. (PALHARES, 2020)

Portanto, avaliando estes vários fatores mercadológicos e estudo necessário dos impactos econômicos e financeiros nos indicadores das instituições de ensino superior, temos como pergunta-problema avaliada neste trabalho: Quais fatores contribuíram para que IES privadas brasileiras se distanciassem de seu ponto de equilíbrio financeiro, lucro ou prejuízo, no período dos censos 2018, 2019 e 2020?

## 1.2 Objetivos da pesquisa

Identificar fatores que levaram IES privadas brasileiras e suas mantenedoras a se distanciar de seu ponto de equilíbrio financeiro.

Especificamente, os objetivos são:

- a) fazer levantamento de variáveis independentes que tem relação com o desempenho das IES e suas mantenedoras no distanciamento do ponto de equilíbrio (variável dependente);
- b) avaliar o peso das variáveis independentes que tem relação com o desempenho das IES e suas mantenedoras no distanciamento do ponto

de equilíbrio (variável dependente), possibilitando apresentar orientações para bom desempenho das instituições;

- c) avaliar os indicadores de matrículas por porte institucional e modalidade de ensino.

### 1.3 Delimitação do estudo

Este trabalho avaliou a população de mantenedoras brasileiras de ensino superior privado brasileiro e seus cursos de graduação presentes nos dados secundários dos censos 2018, 2019 e 2020, representando o período delimitado.

A análise de fatores que podem influenciar o ponto de equilíbrio de uma instituição de ensino, neste estudo, foi limitada pelos valores de receitas e despesas fornecidos publicamente pelos microdados do Censo da Educação Superior.

Foram excluídos do estudo conjuntos que não possuíam dados ou que na análise de pontos extremos (*outliers*) foram considerados não aderentes ao modelo de dados, além das instituições de categoria administrativa pública, como federais, estaduais e municipais.

Para o caso da análise de fatores que envolvem os grandes grupos educacionais consolidadores – grupos empresariais privados, de capital aberto ou não, que detêm grande volume de matrículas – foi usada a lista das mantenedoras de cada grupo, fornecida pelos respectivos materiais disponíveis nos sítios de relações com investidores.

### 1.4 Justificativa e relevância do trabalho

A importância econômica do setor educacional privado é apresentada em números: em 2019, a contribuição do setor para o PIB foi de 126,5 bilhões de reais e o ensino privado correspondia a 1,65% do PIB. A educação particular contribuiu com quase 20 bilhões de reais para a Previdência Oficial e o FGTS sendo que o ensino privado tinha 1,25 milhões de vínculos empregatícios formais em 2020. Esse número cresce para aproximadamente 1,5 milhões em 2022. Isso representa 3,4% de todos os empregos formais do setor privado no Brasil (FENEP, 2022).

A percepção de queda significativa nos indicadores principais de instituições de ensino, como totalidade do corpo discente ativo, novos ingressos e financeiros

como receitas com as mensalidades e a rentabilidade bruta são alguns dos aspectos investigados neste trabalho. Entende-se uma oportunidade de realizar uma avaliação estatística destes impactos, utilizando os dados validados pelo INEP/MEC e informados pelas instituições de ensino superior privadas referentes a seus cursos de graduação.

Esta análise, a partir dos dados secundários, configura uma oportunidade de enriquecer os estudos do comportamento de indicadores financeiros em IES, contribui com a avaliação estatística não só do impacto de variáveis no distanciamento das IES do ponto de equilíbrio, de forma positiva ou negativa, mas também do impacto de grandes grupos educacionais consolidadores e os resultados da adoção da modalidade EAD na busca da rentabilidade e sustentabilidade financeira, considerando a consolidação mercadológica como fator importante na obtenção de resultados pela diluição dos custos de prestação de serviços

O uso dos dados validados pelo Inep/MEC e informados pelas instituições de ensino superior privadas referentes a seus cursos de graduação utilizados neste trabalho pretendem enriquecer indicadores e auxiliar as instituições de ensino a medir o impacto dos resultados financeiros, índices e compará-los com os mesmos estratos de mercado ao qual pertencem bem como com a concorrência.

### 1.5 Organização do trabalho

Este trabalho é organizado com uma seção introdutória, resumindo o desenvolvimento realizado. A fundamentação teórica concentra-se nas literaturas que fundamentam o ensino superior brasileiro e seus indicadores financeiros e contábeis, contabilização de custos e cálculos de margem e ponto de equilíbrio, finalizando com a relação entre os dados fornecidos nos microdados, a literatura de custos e margem e um resumo teórico dos métodos estatísticos utilizados.

Já no procedimento metodológico, é descrito o processo de análise e aplicação dos fundamentos estatísticos referenciados, que servem de base para a seção em que se descrevem os resultados obtidos.

Encerrando com a conclusão, discutimos os resultados encontrados e possíveis justificativas para eles.

## 1.6 Contribuições da pesquisa

Tanto econômica quanto socialmente, este trabalho procura apresentar contribuições efetivas resultantes do mapeamento e análise de indicadores e aspectos que os afetaram durante o período de estudo, permitindo assim que acadêmicos estudiosos do tema e as mantenedoras do ensino superior privado brasileiro possam utilizar o estudo de forma particular, na própria avaliação, evidenciando e criando planos de ação de acordo com os resultados indicados.

Adicionalmente, esta pesquisa pode gerar insumos aos órgãos governamentais para apoio na definição de políticas públicas de gestão do ensino superior.





## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, é feita a apresentação, através de revisão bibliográfica, da fundação e evolução das organizações que atuam no ensino superior brasileiro, buscando mostrar características deste mercado e a forma como se consolidou. São apresentados também indicadores financeiros e contábeis aplicáveis a instituições de ensino superior bem como sua relação com os indicadores de *performance* financeira e operacional informados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

Na subseção sobre o Ensino Superior Brasileiro, é apresentado um histórico da fundação do ensino superior no Brasil, desde as primeiras escolas superiores até os marcos regulatórios que determinam o modelo do mercado atual.

Já na subseção Mercado Educacional e Sustentabilidade Financeira, conceituam-se as naturezas administrativas, a consolidação do mercado e os marcos de sustentabilidade financeira, que deram suporte à busca de indicadores e variáveis utilizadas no procedimento metodológico.

Na subseção Ponto de Equilíbrio Financeiro, é apresentada a justificativa e os modelos de formulação suportados pela teoria contábil e visões, como custo/volume/lucro, utilizadas neste trabalho.

É feita uma relação, na subseção Ponto de Equilíbrio e Indicadores fornecidos pelo Censo, que cumpre o papel de justificativa documental da relação entre a fonte teórica conceituada na subseção anterior e seus vínculos com os dados fornecidos pelo Censo Educacional.

Encerra-se a conceituação teórica com uma breve descrição bibliográfica dos principais métodos estatísticos utilizados – Regressão Linear Multivariada, Componentes Principais e Florestas Aleatórias – importante para direcionar os resultados apresentados.

### 2.1 O ensino superior brasileiro

Conforme descrito por Barreyro (2008), a fundação da educação superior no Brasil começou nos primórdios da catequização jesuítica, com a criação dos cursos de Artes e Teologia no Colégio dos Jesuítas da Bahia, sendo o ensino superior não religioso iniciado com a transferência da sede do império português para o Brasil, criando-se cursos de Cirurgia, Medicina e Matemática, relacionados com as

atividades militares, e Agronomia, Desenho Técnico, Economia, Química e Arquitetura, destinados à burocracia estatal. Houve outros cursos ligados à Academia de Belas Artes, tais como os de Desenho, História, Arquitetura e Música, conforme Quadro 1, a seguir, que representa a criação dos primeiros cursos e faculdades no Brasil (MOROSINI, 2006)

**Quadro 1 - Formação das primeiras faculdades e cursos superiores no Brasil**  
(continua)

Ano	RJ	BA	PE	SP	MG	RS
1808	Academia de Marinha Cadeira de Anatomia	Escola de Cirurgia (Hospital Real da Bahia)				
1809	Cadeira de Medicina Teórica e Prática (Hospital Real Militar e da Marinha)					
1810	Academia Real Militar					
1812		Curso de Agricultura				
1813	Academia Médico- Sanitária					
1814	Curso de Agricultura					
1815		Academia Médico- Cirúrgica 1				
1817		Curso de Química (Industrial, Geologia e Minerologia)				
1820	Academia de Artes					
1826	Academia de Belas Artes					
1827			Curso de Ciências Jurídicas e Sociais de Olinda	Curso de Ciências Jurídicas e Sociais de São Paulo		
1832	Faculdade de Medicina	Faculdade de Medicina da Bahia				
1833	Academia Naval / Militar					
1839	Escola Militar				Faculdade de Farmácia	
1841	Escola Nacional de Música					
1854			Faculdade de Direito de Recife	Faculdade de Direito de São Paulo		

(conclusão)

Ano	RJ	BA	PE	SP	MG	RS
1855	Escola de Aplicação do Exército					
1858	Escola Central – Curso de Matemática e Ciências Físicas e Naturais					
1874	Escola Politécnica do Rio de Janeiro					
1875		Escola Superior de Agronomia (Cruz das Almas)			Escola de Minas e Metalurgia	
1878	Curso de Odontologia					
1883						Escola de Medicina Veterinária e Agricultura Prática (Pelotas)
1884	Escola de Farmácia					
1888		Escola Politécnica				
1889						

Fonte: Adaptado pelo autor com base em Morosini (2006, p. 18)

No Quadro 1, apresenta -se uma linha do tempo de formação de instituições educacionais e programas de ensino superior em diferentes estados brasileiros, incluindo Rio de Janeiro (RJ), Bahia (BA), Pernambuco (PE), São Paulo (SP), Minas Gerais (MG) e Rio Grande do Sul (RS) e suas respectivas datas de fundação. A lista relaciona academias, escolas e faculdades de várias disciplinas, como estudos marinhos, medicina, agricultura, direito, arte, música, engenharia e ciências. Algumas instituições também se concentravam em áreas específicas, como estudos militares e navais, e outras tinham enfoques específicos, como química industrial e mineração.

Essa formação inicia se em 1808, no Rio de Janeiro e Bahia, com cursos da área de saúde (Anatomia e Cirurgia) e no ano seguinte a cadeira de Medicina, ainda em academias militares. Entre 1810 e 1820, temos a inclusão de cursos de formação militar, sanitária e agricultura, já introduzindo na Bahia os cursos de química industrial, com foco em mineralogia.

Já entre 1820 e 1840, avança a fundação de cursos de humanidades (Belas Artes) e os primeiros cursos de ciências jurídicas em São Paulo e Pernambuco, além de Farmácia em Minas Gerais. Entre 1840 e 1888, criam-se as Faculdades de Direito de São Paulo e Recife, agronomia, ciências naturais e físicas no Rio de Janeiro, junto com odontologia e prática veterinária no Rio Grande do Sul. Nesse período, também se inicia a formação da Escola de Minas e Metalurgia em Minas Gerais e a Politécnica da Bahia.

O ensino superior privado aparece a partir da fundação da República (1889), com a Constituição de 1891, que descentralizou o que era exclusivo do poder vigente na época (BARREYRO, 2008), pois, até então, se tratava de experiências confessionais e profissionais, que respondiam às necessidades das elites detentoras dos poderes econômico, político e cultural (MOROSINI, 2006), ainda que, conforme descrito por Sguissardi (2015, p. 275), se levem em conta as precárias experiências de universidade das primeiras duas décadas do século XX: em Manaus (1909), São Paulo (1910) e Curitiba (1912),.

As primeiras universidades públicas – em Minas Gerais, Rio Grande do Sul, Rio de Janeiro e São Paulo – surgiram nas décadas de 1920 e 1930, sendo seguidas pela fundação de instituições católicas privadas, que originou as primeiras universidades católicas, modelo que perdura até hoje, quando temos as instituições superiores públicas (federais, estaduais e municipais) e as de natureza jurídica privada – confessionais ou sociedades privadas lucrativas com a reafirmação liberal do modelo de ensino superior através da Constituição de 1988 (BARREYRO, 2008; SAMPAIO, 2011)

No Quadro 2, é descrita a evolução ao longo dos séculos XIX e XX da universidade brasileira, a partir das legislações vigentes à época.

**Quadro 2 - Legislações do Ensino Superior Brasileiro – séculos XIX e XX**

<b>DATA</b>	<b>DENOMINAÇÃO (DECRETO)</b>	<b>SUBSTÂNCIA</b>
1808 – 18/2	Carta Régia	Autorização para a criação do 1º Curso Superior
1808 – 23/2	Carta Régia	1º Estatuto do Curso Superior
1813 – 1º/4	Reforma Álvares de Carvalho	Estatutos de Cirurgia
1832 – 3/10	Reforma Senador Vergueiro	Nova organização das Academias Médicas da Bahia e do Rio de Janeiro
1854 – 28/4	Reforma Luiz Pereira de Couto Ferraz – 1386	Novo Estatuto para os Cursos Jurídicos
1879 – 1884 19/4	Reforma Leôncio de Carvalho – 7274	Criação de Faculdades Livres
1891 – 2/1	Reforma Benjamin Constant (1232H)	Regulamento para a instituição de Ensino Jurídico
1901 – 1/1	Reforma Epiácio Pessoa (3890)	Código para o Ensino superior – equiparação
1911 – 5/4	Reforma Rivadávia Corrêa (8659)	Oficialização completa do ensino à sua total independência
1915 – 18/3	Reforma Carlos Maximiliano (11530)	Restabelecimento da Fiscalização
1925 – 13/1	Reforma Rocha Vaz (16782 <sup>a</sup> )	Criação de Universidades
1931	Estatuto das Universidades Brasileiras (19.850) (19.851) (19.852)	Criação do Conselho Nacional de Educação Organização das Universidades Organização da Universidade do Rio de Janeiro
1961	Lei de Diretrizes e Bases da Educação Nacional – LDB (4024)	Organização do Sistema de Educação
1968	Reforma Universitária (5.540)	Modernização da Universidade
1991	Mercosul Educativo	Integração dos sistemas educativos dos países do Mercosul
1996	Lei de Diretrizes e Bases da Educação Nacional – LDB (9394)	Flexibilização do Sistema de Educação Superior

Fonte: Adaptado pelo autor com base em Morosini (2006, p. 19)

Percebemos no Quadro 2, que a Reforma Benjamin Constant (1232H), em 1891, teve como objetivo modernizar e tornar mais eficiente o sistema educacional brasileiro, introduzindo mudanças significativas na estrutura e no conteúdo do ensino, incluindo a criação de escolas normais para treinar professores, a unificação dos cursos de ensino médio e superior e a instituição de exames de admissão e progressão, além de fundar um sistema nacional de ensino, com a criação de uma Secretaria de Educação e Ensino.

Nas duas décadas seguintes, a essa reforma sucederam-se várias outras que visavam modernizar e democratizar o sistema educacional brasileiro, ampliando a oferta de ensino para todas as camadas da sociedade e prevendo a criação de escolas primárias em todos os municípios do país, além da obrigatoriedade do ensino primário para todas as crianças entre 7 e 14 anos de idade. Também se

previu a criação de uma Secretaria Nacional de Educação, com o objetivo de centralizar e padronizar a administração e a supervisão do ensino no país.

A formação das primeiras universidades do país começa em 1925 e é organizada até final dos anos 1950, quando as reformas propostas pelos Estatutos das Universidades Brasileiras criaram o Conselho Nacional de Educação, organizaram as universidades e propuseram o estabelecimento da primeira universidade do Rio de Janeiro, etapa seguida por normalização e reformas até o ano de 1996, com a publicação da atual LDB (BRASIL, 1996) em vigor e que estabeleceu principalmente a autonomia no ensino superior, inclusão e priorização da pesquisa e inovação.

A adoção – e possibilidades de flexibilização mercadológica do ensino superior – permitiu a transformação acelerada verificada no mercado educacional e a inserção de grupos investidores, abrindo assim caminhos para aumento de instituições com naturezas lucrativas e a necessidade de busca da sustentabilidade financeira.

## 2.2 Mercado educacional e sustentabilidade financeira

A grande transformação no ensino superior brasileiro veio no ano de 1997 com a mudança legal nos estatutos das instituições por meio do decreto 2306/97 (BRASIL, 1997), que permitiu, entre outras transformações, a das regras para a criação, reconhecimento e credenciamento de cursos de educação a distância, assim como as responsabilidades das instituições que os oferecem. Esse decreto também regulamentou o reconhecimento de diplomas e graus estrangeiros obtidos por educação a distância e a permissão às IES para troca de natureza, podendo ser privadas com finalidade lucrativa ou sem fins lucrativos, cada uma contando com incentivos financeiros e econômicos governamentais (SAMPAIO, 2011).

Entende-se, ainda em Sampaio (2011), que instituições de natureza de fins lucrativos são aquelas instituições de ensino que visam obter lucro financeiro pela oferta de cursos e programas de estudo, geralmente por meio de mensalidades e taxas para seus alunos e podem ter investidores ou acionistas que buscam retorno financeiro. Já as de natureza sem fins lucrativos referem-se a instituições de ensino mantidas por organizações não governamentais e que não têm como objetivo principal obter lucro financeiro, diferenciando-se das instituições de natureza pública

pela origem do financiamento, suprido pelo Estado. Elas geralmente reinvestem o lucro obtido com mensalidades em incentivos à educação e podem depender de doações, subvenções ou financiamento governamental para sua operação, cujo objetivo principal é oferecer educação de qualidade e acessível para a comunidade.

Com isso, ao se analisarem os números do mercado de ensino superior brasileiro, apresentados na Tabela 1, percebe-se um salto comparativo: entre, 1997 e 2021, houve não só um aumento de 228% no número de IES privadas – superior aos 48% de crescimento de IES públicas no mesmo período –, como também da popularização do ensino superior, com um aumento de matrículas no ensino privado na ordem de 482%, em contrapartida aos 173% nas matrículas em IES públicas (INEP, 1999; 2022).

**Tabela 1 - Evolução do ensino superior privado (1997-2021)**

Ano Censo	1997			2021		
	#	Privadas	Públicas	Total	Privadas	Públicas
Instituições	689	211	900	2.261	313	2.574
Matrículas	1.186.433	759.182	1.945.615	6.907.893	2.078.661	8.986.554

Fonte: Elaborado pelo autor com base em Inep (1999; 2022)

Para efeito comparativo, segundo o IBGE (1999; 2002), nas faixas etárias jovens, alvo do ingresso no ensino superior (15-19 e 20-24 anos), houve um crescimento estimado de 9,4%.

Esse crescimento na quantidade de instituições e aumento do volume de matrículas no ensino superior privado é descrito por Barreyro (2008) como efeito da interiorização das ofertas, autonomia institucional universitária para criação de novos cursos e ofertas, expansão do EAD (ensino a distância) como modalidade, atingindo as classes mais carentes da população e encontra sustentação na forma mercantilizada (SGUISSARDI, 2015; DINIZ; GOERGEN, 2019) de busca do lucro na prestação do serviço educacional.

Em análise dos cenários da participação de mercado entre IES sem fins lucrativos e com fins lucrativos, Corrêa (2015) ressalta que houve variação numérica entre as matrículas nas instituições sem fins lucrativos, entre os anos de 2000 e 2012, respectivamente de 53,2% para 36,9%. O autor verificou também que, no mesmo período, as matrículas migraram para o cenário de natureza jurídica lucrativa

e subiram de 12% para 36,2%. O restante das matrículas concentrou-se, segundo o autor, em instituições administrativas públicas (respectivamente 34,8% e 26,9%).

Atualizando os números de Corrêa (2015), conforme Tabela 1, após tabulação de informações de matrículas ativas no censo de 2021 (INEP, 2022), temos a participação nas matrículas ativas de 19,7% em IES privadas sem fins lucrativos e 57,1% nas de natureza lucrativa, resultando em 23,1% em instituições na categoria administrativa pública.

A mercantilização do ensino superior privado no Brasil tem sido um tema crescente desde 1997, conforme descrito por Barreyro (2008). De acordo com Fritsch, Jacobus e Vitelli (2020), essa tendência foi impulsionada pela política de privatização do ensino superior implementada pelos governos, que incentivaram a entrada de investidores privados no setor e isso resultou em crescente competição entre as instituições de ensino, com as instituições privadas buscando atrair alunos oferecendo serviços e infraestrutura cada vez mais sofisticados e, ao mesmo tempo, buscando a redução de custos de operação para competir no mercado com mensalidades mais baixas, atingindo públicos de classes menos favorecidas.

Segundo Sguissardi (2015), a mercantilização do ensino – com a entrada de grupos financeiros (estrangeiros ou não), o fomento e o financiamento estudantil como política pública e a busca do lucro – levou a um cenário de concentração no ensino superior privado em que a maior fatia das matrículas está presente em poucos grupos consolidadores, fato corroborado pelo Cade (2016), citando a consolidação do setor.

Ainda em 2016, o Cade publicou documento afirmando que o cenário de consolidação do Ensino Superior, considerando que o aumento de solicitações – 62 atos de concentração até 2015 com as análises inicialmente tramitadas de forma sumária – se tornava preocupante, já que grande parcela da população, dados os cenários descritos anteriormente, estava tendo acesso ao ensino superior. Ao mesmo tempo, com a permissão de entrada de fundos de investimento de capitais nas instituições com fins lucrativos e a pulverização da oferta, consolidou-se em poucos grupos educacionais de natureza lucrativa grande volume de matrículas (CADE, 2016). A partir da publicação, foram desenvolvidos critérios mais sofisticados de análise visando impedir a concentração e a criação de monopólios, restringindo autorizações para consolidações, com análise profunda de impactos no mercado, concentração e aplicação de restrições a fusões e aquisições (CADE,



2016). Neste sentido, Colmán e Portes (2019, p. 144) assim se referem aos processos de fusões e aquisições:

Esse processo de privatização do ensino superior caminha ao lado de outro processo que é o de oligopolização. A realidade se expressa pela expansão da educação superior e pela formação de grandes oligopólios, constituída pela compra/fusão das Instituições de Educação Superior privadas e pela abertura de capitais dessas empresas nas bolsas de valores.

Segundo a consultoria Hoper Educação (HOPER, 2021), temos atualmente, derivados desse processo de consolidação mercadológico, oito grandes grupos consolidadores – organizações privadas com fins lucrativos, que têm como um de seus objetivos criar e compartilhar estruturas, gestão e recursos – seis dos quais com capital aberto, conforme mostrado na Tabela 2.

**Tabela 2 - Grupos consolidadores no ensino superior em 2021**

<b>Consolidador</b>	<b>Matrículas</b>	<b>Part. (%)</b>	<b>Mantenedoras</b>	<b>Mantidas</b>
(não consolidadas)	2.942.812	43%	1.468	1.961
COGNA	872.317	13%	10	87
VITRU	822.388	12%	3	17
YDUCS	754.826	11%	32	64
UNIP	421.036	6%	1	12
CRUZEIRO DO SUL	362.450	5%	12	13
ANIMA	320.269	5%	21	56
SER EDUCACIONAL S.A.	237.817	3%	9	44
UNINOVE	173.978	3%	2	7
<b>Total Geral</b>	<b>6.907.893</b>	<b>100%</b>	<b>1.551</b>	<b>2.261</b>

Fonte: adaptado pelo autor com base em Hoper (2021) e Inep (2022)

Ao se avaliarem os microdados de matrículas apresentadas no Censo da Educação Superior 2021, última publicação no tempo deste trabalho, e com base nas composições das estruturas das mantenedoras dos grupos consolidadores apresentados em relatórios com investidores (COGNA, 2023; VITRU, 2023; YDUCS, 2023; CRUZEIRO DO SUL EDUCACIONAL, 2023; ÂNIMA, 2023; SER EDUCACIONAL, 2023), percebe-se a concentração no mercado de ensino superior brasileiro. No ano do Censo 2021, 57% do total de matrículas ativas no mercado de ensino superior brasileiro pertenciam a esses oito grandes grupos consolidadores. Às demais instituições, representadas por 1.468 mantenedores, que detêm o registro de 1.961 mantidas, restaram 43% das matrículas ativas. Ressalta-se que o Censo

da Educação Superior traz resultados dos cursos de graduação autorizados pelo MEC, não incluindo, por exemplo, matrículas de pós-graduações, mestrados, doutorados e cursos livres, o que pode fazer com que este número de matrículas seja subestimado em relação ao quantitativo real.

Ademais, como estratégia de atuação no mercado, a modalidade EAD tornou-se um dos alvos das grandes *holdings* de educação para conquistar cada vez maiores fatias e aumentar a capilaridade, fazendo uso de grandes estruturas, investimentos de capital e aquisições estratégicas para chegar ao ponto de consolidação visto autalmente (JULIANI; SANTOS; FÁVERO, 2022).

O Censo ABED (2022, p. 30) mostra, com base nas pesquisas setoriais e nos dados do censo educacional, que a concentração do mercado de ensino a distância, conforme Tabela 3, ainda está nas mãos de oito grandes grupos consolidadores do mercado.

**Tabela 3 - EAD nos grupos consolidadores**

<b>Consolidador</b>	<b>Matrículas</b>	<b>Part. (%)</b>		<b>Mantenedoras</b>	<b>Mantidas</b>
(não consolidadas)	828.627	23%	23%	396	402
VITRU	802.196	23%		2	2
COGNA	668.768	19%		2	5
YDUCS	469.756	13%		7	9
UNIP	262.794	7%		1	1
CRUZEIRO DO SUL	233.451	7%	77%	7	7
SER EDUCACIONAL S.A.	118.727	3%		7	8
ANIMA	100.438	3%		15	15
UNINOVE	59.392	2%		1	1
<b>Total Geral</b>	<b>3.544.149</b>	<b>100%</b>		<b>438</b>	<b>450</b>

Fonte: adaptado pelo autor com base em Hoper (2021); Inep (2022); Abed (2022)

Na Tabela 3, observa-se que a modalidade EAD, no ano do censo 2021, tem um total de 3.544.149 matrículas, que representa 51,3% do total de alunos matriculados no Brasil, 6.907.893. Também se verifica que, das 3.544.149 matrículas em EAD, 77% delas pertencem aos oito grandes grupos consolidadores brasileiros, justificando um cenário de massificação e concentração dessa modalidade.

Ainda considerando o cenário de consolidação no ensino a distância, Gomes, Leher e Costa (2020, p. 14) apontam:

[...] as formas de ampliar a taxa de lucro, como o uso intensivo de tecnologias, materiais pedagógicos padronizados, disciplinas a

distância, ampliação das matrículas de cursos a distância, reorganização dos ciclos básicos de modo a torná-los comuns a diversas áreas conexas, reduzindo, assim, o número de disciplinas.

Corroborando essas observações, França Filho, Antunes e Couto (2020) e Novaes, Nascimento e Gonçalves (2022) pontuam que as instituições de ensino superior se conduzem à modalidade a distância em busca de redução de custos, principalmente com a reconfiguração do trabalho docente, aumentando a quantidade de alunos por docente, reduzindo disciplinas e aumentando as comuns, entre outras formas de aplicação de ferramentas de otimização do recurso humano.

É nesse cenário que as instituições, integrantes ou não dos grandes grupos educacionais, buscam otimizar o recebimento das receitas e a redução dos custos na prestação dos serviços educacionais. As instituições de ensino superior consideradas com fins lucrativos, conforme Corrêa (2015, p. 28):

[...] não possuem restrições em relação às aplicações dos seus excedentes financeiros, e, muito menos, em relação à diversificação do seu portfólio de produtos, cuja boa gestão é um dos indicativos da boa governança e, por consequência, da sustentabilidade e sucesso do negócio.

Justamente na questão da busca do lucro, rentabilidade e sustentação do negócio educacional, principalmente no ensino superior privado, que tem repousado a maior parte das preocupações dos gestores educacionais. Para Machado (2005, p.113-114), o produto da indústria educacional é o aluno formado, satisfeito e adequado ao mercado de trabalho e a melhor medida de eficácia de uma IES é o lucro/superávit nessa prestação de serviço.

Para Castro (2013, p. 179), a sustentabilidade financeira de uma instituição de ensino superior depende fundamentalmente da receita adquirida por meio das mensalidades pagas pelos alunos.

Para Xavier (2014, p. 18), “[...] faz-se mister o conhecimento desse momento orçamentário para uma IES privada, percebendo-se em qual momento a instituição começa a cobrir seus gastos fixos e variáveis, e a partir de qual momento ela passa a ter resultados financeiros positivos”.

Notoriamente, na linha da sustentabilidade financeira e busca do equilíbrio receitas e despesas, a modalidade EAD tem figurado de maneiras controversas, sempre dependendo da visão. Conforme descrito por Jesus (2020), percebe-se o

aumento do atendimento pelo EAD a camadas sociais não atingidas pelo modelo presencial, em razão não só da flexibilidade de acesso e horários, mas também por ter custos de mensalidade inferiores ao ensino presencial, permitindo o acesso de classes C, D e E (ABED, 2022). Por outro lado, Jesus (2020) alerta que as modalidades a distância tendem à retirada da figura e da autonomia do professor, restringindo o aprendizado do aluno a conteúdos apostilados.

Ainda em 2009, Moran (2009, p. 61) descreve: “Há instituições que veem a EAD mais como negócio, mercado e investimento e todos os esforços são direcionados para a ocupação rápida do mercado, para a rentabilidade máxima dos acionistas, para ter um baixo custo aluno”. Para o autor, essa reorganização do ensino o torna massificado, tecnológico dependente e, ao mesmo tempo, reduz a interação com o docente e com seus pares.

É baseado no esforço de redução de custos, rentabilidade e ajuste da operação vocacionando ao lucro que a aplicação dos conceitos financeiros de ponto de equilíbrio e análises de volume e lucro são importantes para as instituições mantenedoras alvo deste trabalho.

### 2.3 Aplicação educacional do ponto de equilíbrio financeiro

É na gestão das IES privadas que se verifica a importância da aplicação de conceitos de Margem e Ponto de Equilíbrio, definidas por Assaf Neto (2020, p. 310) como o ponto em que os custos e despesas totais igualam-se aos de receitas totais: acima desse ponto se tem lucro; abaixo, prejuízo. Quando se tem sobra (acima do ponto de equilíbrio), há a margem de contribuição, ou lucro operacional, ou resultado bruto.

Conforme descrito por Oyadomari *et al.* (2018), o resultado da diferença entre as receitas e as despesas em determinado período podem resultar em lucro (receita > despesa) ou em prejuízo (receita < despesa), obtido pela demonstração de resultados do exercício.

O acompanhamento dos indicadores de desempenho financeiro e operacionais é destacado na gestão educacional, de todos os portes e naturezas jurídicas das instituições. A margem de contribuição e o ponto de equilíbrio são medidos sempre buscando equilibrar a venda do produto (alunos matriculados) com

o custo de prestação do serviço (folha de pagamento docente e administrativo, despesas e investimentos na prestação do serviço).

Neste ponto Lugoboni (2010, p. 92), relaciona uma série de indicadores de desempenho financeiro e operacional validados por especialistas em gestão educacional, referentes a Receitas, Custos e Despesas Administrativas, quando na dimensão financeira, e Evolução no número de matrículas ou alunos, no âmbito operacional. Esses indicadores são também referendados pelo mesmo autor em sua tese de doutorado (LUGOBONI, 2017, p.149).

Para Machado (2005), dentro dos preceitos contábeis e de controladoria em Instituições de Ensino Superior, os relatórios de análise de custo do serviço prestado e ponto de equilíbrio, bem como a margem de contribuição, podem ser classificados como ótimos elementos para tomada de decisão, sendo o ponto de equilíbrio aquele em que as despesas (custos) igualam-se às receitas (mensalidades).

Esta análise de ponto de equilíbrio também é conhecida como análise de custo/volume/lucro, margem de contribuição, ruptura, entre outras denominações, e descreve o conceito analítico de avaliar um valor mínimo necessário para cobrir os custos e tem importância no processo de tomada de decisão, considerado como procedimento básico para estabelecimento dos valores de mensalidades de graduação e análise de sensibilidade das variáveis que afetam o custo da instituição de ensino (LUCHESA; MACHADO, 2011).

A aplicação da análise custo/volume/lucro, essência da análise de ponto de equilíbrio, é citada por Xavier (2014, p. 61), que a considera essencial para identificar, com a aplicação de gestão de custos em instituições de ensino superior, indicadores corroborados por Lugoboni (2010), como os lucros são afetados por fatores componentes tanto das receitas quanto das despesas.

O capital financeiro, reconhecido pelos processos contábeis, é alcançado, segundo Silva e Rezende (2017), pelo gerenciamento de itens que contribuem para a valorização patrimonial e, no caso de instituições de ensino pautadas principalmente na receita de mensalidades, para atingimento do ponto de equilíbrio financeiro.

Ainda na definição contábil, mas com uma visão operacional, Faria (2019) e Meinertz (2019) ressaltam a identificação do ponto de operação que cubra todos os custos e garanta sobras para reinvestimento na IES e isso só é possível através do monitoramento da relação custo/receita.

Com a fundamentação do uso da visão contábil do ponto de equilíbrio, por meio da análise de indicadores de receitas e despesas de instituições de ensino, a distância ou afastamento será relacionada diretamente com o lucro ou prejuízo observado no exercício, que, de acordo com Oyadomari *et al.* (2018), é a medição da eficiência operacional da prestação dos serviços e que, neste trabalho, será relacionada com as informações disponibilizadas pelo Censo da Educação Superior.

Portanto, o ponto de equilíbrio é o nível de vendas (recebimento de mensalidades) a partir do qual a instituição de ensino passa a ter lucro, que é calculado subtraindo os custos diretos (como salários dos professores, aluguel, materiais didáticos, entre outros) do valor das mensalidades. Isso significa que, para uma instituição de ensino atingir o ponto de equilíbrio, ela precisa vender determinado número de matrículas para cobrir todos os seus custos e assim gerar a margem de contribuição institucional, conceito financeiro que mede a diferença entre o preço de venda de um produto ou serviço e os custos diretos relacionados à sua produção ou prestação. Em relação ao ensino superior, a margem de contribuição é utilizada para calcular o lucro ou prejuízo de uma instituição de ensino.

Luchesa e Machado (2011, p.114) definem a margem de contribuição como uma medida da rentabilidade das operações de uma instituição de ensino superior, enquanto o ponto de equilíbrio é o nível de vendas a partir do qual a instituição passa a ter lucro.

Portanto, com as definições de ponto de equilíbrio contábil/financeiro e suas aplicações educacionais, podemos relacionar as variáveis definidas pela teoria com os indicadores fornecidos pelo censo e detalhados a seguir.

#### 2.4 Ponto de equilíbrio contábil e indicadores fornecidos pelo censo

Para Oyadomari *et al.* (2018), podemos definir os indicadores componentes do resultado contábil-financeiro como apresentados no Quadro 3.

**Quadro 3 - Margem contábil/financeira**

Indicador	Definição
Receita bruta	Receita bruta no período obtida com a prestação do serviço
Receita líquida	Receita bruta - Deduções da receita
Despesas da prestação dos serviços	Custo de prestação dos serviços
Resultado bruto (margem / ponto de equilíbrio)	Resultado bruto = Receita líquida – Custo dos serviços

Fonte: Adaptado pelo autor com base em Oyadomari *et al.* (2018, p. 52-55)

No Quadro 3, apresentam-se os indicadores necessários para o componente do resultado financeiro ponto de equilíbrio, sendo a receita bruta e a receita líquida correspondentes a mensalidades. Já as despesas com a prestação do serviço educacional são detalhadas como custos de folha de pagamento e outras despesas, de acordo com o dicionário de dados do Censo da Educação Superior.

No dicionário de dados e manual de preenchimento das informações no Censo da Educação Superior 2020 (INEP, 2021), apresentado no Quadro 4, apresentam-se os seguintes indicadores e as definições de lançamento dos valores, com definição resumida.

**Quadro 4 - Variáveis disponíveis na base do censo**

	Variável	Definição
Receitas	Receitas próprias	Valores obtidos com a operação e prestação dos serviços educacionais, mensalidades e financiamento estudantil
	Transferências	Recebimento de receitas de outros entes, tais como verbas governamentais
	Outras receitas	Receitas que não estão contempladas em receitas próprias nem em transferências
Despesas	Remuneração docentes	Despesas relativas ao pagamento de salários, gratificações e outros benefícios integrantes da remuneração pelo exercício do cargo de docente
	Remuneração administrativos	Remuneração do pessoal de apoio pedagógico, técnico-administrativo e suporte, registrando as despesas relativas ao pagamento de salários, gratificações e outros benefícios
	Benefícios e encargos	Despesas relativas ao pagamento de INSS, previdência oficial, abono de férias, 13º salário e outros encargos referentes à folha de pagamento
	Custeio	Despesas administrativas, tais como honorários da diretoria, aluguéis, materiais de consumo, seguros, depreciação, serviços de terceiros e despesas comerciais, tais como comissões, publicidade e propaganda e marketing
	Investimentos	Aplicação de capital em meios de produção, visando ao aumento da capacidade produtiva (instalações, máquinas, transporte, infraestrutura), exceto as despesas com pesquisa e desenvolvimento
	Pesquisa	Devem ser considerados: auxílios financeiros a recenseador e à manutenção de sua equipe, aquisição de material permanente, equipamentos, obras de infraestrutura e instalações para pesquisa
	Outras despesas	Despesas que não estão contempladas nas demais despesas conceituadas

Fonte: Adaptado pelo autor (2023) de Inep (2021)

No Quadro 4, apresentam-se, de acordo com o dicionário de dados e manual de preenchimento, os valores necessários para o cálculo do indicador de ponto de equilíbrio, delimitado pelas variáveis informadas pelo Inep (2021). Temos apresentados para o cálculo das receitas os valores recebidos com mensalidades

(receita própria), transferência de entes (receitas transferidas) e outras formas de receita, que podem ser prestação de serviços adicionais em educação, alocação de espaços, entre outros. Para efeito deste trabalho, será considerada somente a receita obtida com mensalidades (receita própria).

Como despesas, temos os valores descritos para Remuneração Docentes e Administrativos (despesas com salários, férias e outros benefícios), Encargos e Benefícios (impostos e contribuições obrigatórias). Complementam as despesas os valores de custeio, que incluem as despesas administrativas, aluguéis e materiais, investimentos, pesquisa e outras despesas não qualificadas. Para este trabalho, serão usadas as Despesas Docentes, Administrativas, Encargos, Custeio e outras. Investimento e Pesquisa não será alvo deste trabalho, porque o objetivo é o resultado do ponto de equilíbrio da prestação do serviço de educação.

Considerando as dimensões financeiras e operacionais acima descritas, com base em Oyadomari *et al.* (2018) e Inep (2021), podemos fazer uma relação de indicadores entre a literatura contábil e os dados preenchidos pelas instituições no Censo da Educação Superior, apresentado no Quadro 5:

**Quadro 5 - Relação agrupador contábil com variáveis da base do Censo**

Agrupador Contábil por Oyadomari <i>et al.</i> (2018)	Inep - Censo (INEP, 2020)
Receita bruta	Receitas próprias
	Transferências
	Outras receitas
Custo dos serviços prestados	Remuneração docentes
	Remuneração administrativos
	Benefícios e encargos
	Custeio
	Investimentos
	Pesquisa
	Outras despesas

Fonte: Adaptado pelo autor com base em Oyadomari *et al.* (2018); Inep (2021)

O ponto de equilíbrio, descrito anteriormente por Oyadomari *et al.* (2018), consiste em Receita Bruta - Custo dos Serviços Prestados, em sendo positivo o resultado, temos lucro; se negativo, prejuízo. Com a relação entre os dados fornecidos pelo Censo da Educação Superior, podemos calcular a estimativa do ponto de equilíbrio das IES, em que se formata a Equação (I), a seguir.



(I)

$$PE = \frac{(\text{Receitas Próprias} + \text{Transferências} + \text{Outras Receitas})}{(\text{Rem. Docentes e Administrativos} + \text{Encargos} + \text{Custeio} + \text{Outras Despesas} + \text{Pesquisa} + \text{Investimentos})}$$

Alguns fatores podem influenciar o alcance do ponto de equilíbrio definido por Soares (2016, p.427) como o “número de alunos necessários para que a instituição cubra os seus custos”. Já Silva (2017, p. 16) diz: “[...] para que uma IES privada se torne competitiva num mercado consolidado, onde prevalece maior a oferta que a demanda, elas devem procurar aperfeiçoar continuamente seus produtos-cursos, serviços e processos [...]”.

Ainda quanto ao equilíbrio financeiro e a oferta, Zamin (2019, p. 16-17), no entanto, argumenta:

[...] Na busca pelo crescimento, muitas instituições de ensino superior acabam ampliando significativamente a oferta de cursos com o intuito de abarcar novos extratos de alunos. Apesar de ser eficiente em um primeiro momento, esta estratégia pode fragilizar a instituição em um contexto de retração de demanda, aumento de mensalidades e inadimplência. Esta fragilização ocorre principalmente pela dificuldade de, com um número menor de alunos, equilibrar os altos custos desta estrutura de oferta ampliada.

Alguns desses fatores afetam o ponto de equilíbrio na estrutura de custos e podem ser resumidos como despesas com mão de obra e *ticket* médio praticado (ZAMIN, 2019). Esses pontos são corroborados por Silva (2017, p. 81), para quem não só fatores como quantidade de alunos e valores de mensalidades colaboram para o atingimento do equilíbrio financeiro, mas também a capacidade de obtenção de melhores resultados no controle de gastos operacionais, inovação de mercados, flexibilidade e agilidade.

Os fatores que serão investigados neste trabalho, de acordo Machado (2005), Lugoboni (2010), Xavier (2014), Silva e Rezende (2017), Faria (2019) e Meinertz (2019) são:

- *Ticket* Médio (razão entre os valores recebidos em receitas de mensalidades e a quantidade de alunos, no ano) juntamente com ações de fomento como bolsas e financiamentos;

- Custos de Folha de Pagamento de Docentes e Encargos relativos, Indicadores Operacionais de Qualidade administrativa e docente;

- Proporção Alunos por Professor e Funcionário, eficiência operacional na análise custo/volume/lucro, bem como indicadores de eficiência de ocupação das vagas;

- Exposição ao EAD e polos de ensino a distância, que se alinham com novos mercados e ferramentas de concorrência.

Portanto, esclarecidas as origens das informações nas dimensões financeira e operacional, a literatura permite estabelecer as bases para aplicação dos procedimentos metodológicos da estruturação dos indicadores de ponto de equilíbrio e matrículas, calculados pelos dados informados no Censo da Educação Superior, no período 2018 a 2020, apurando, de acordo com cada estrato de IES classificado, as evoluções de Receitas, Despesas e Rentabilidade Operacional (margem bruta e ponto de equilíbrio), e medir os efeitos nos volumes de matrículas, por estrato e modalidade.

### 3 PROCEDIMENTOS METODOLÓGICOS

Este estudo adota o tipo de pesquisa chamado de quantitativa, que, de acordo com Proetti (2017, p.18), permite remover a subjetividade da análise e testar as teorias necessárias para a quantificação dos dados, medir valores e importância, mensurando medidas e comprovando-as, atribuindo precisão e confiabilidade por análise estatística.

Segundo Sampieri, Collado e Lucio (2013), o enfoque quantitativo é sequencial e comprobatório, utilizando coleta de dados para testar hipóteses com base na medição numérica e análise estatística, Severino (2017, p. 140) corrobora essa definição e acrescenta ser a abordagem quantitativa a preferida para estabelecer relações matemáticas experimentais entre fenômenos, medindo assim causa e efeito.

O processo quantitativo passa por fases como a formulação da ideia e do problema, utilizando literaturas e passando pelos marcos teóricos com o alcance da visualização do estudo. Com isso Elaboramos as hipóteses e definimos as variáveis com o desenvolvimento do desenho das pesquisas. Após a seleção da amostra e coleta dos dados, realiza-se a análise das descobertas e resultados (SAMPIERI; COLLADO; LUCIO, 2013).

#### 3.1 Dados secundários e base de dados do censo da educação superior

A partir de dados secundários e arquivos oficiais, obtidos nos sítios oficiais do Inep/MEC, nesta pesquisa foram analisados dados e indicadores, utilizando ferramentas de análise estatística descritiva e regressão, principalmente análises de variação dos indicadores referenciados.

Dados secundários, segundo Sampieri, Collado e Lucio (2013), são aqueles coletados por outros pesquisadores e implicam revisão de documentos e registros públicos, em arquivos físicos ou eletrônicos. Metodologia de pesquisa que utiliza dados já existentes, coletados anteriormente, para responder a questões de pesquisa, essa abordagem é utilizada quando os dados primários (coletados especificamente para o projeto de pesquisa atual) incluem a análise de registros governamentais, dados de empresas ou dados de pesquisas anteriores.

Severino (2017) menciona como uma das fontes de dados secundários, figurando como agência de fomento à pesquisa, a fonte dos arquivos do Censo da Educação Superior, mantidos e publicados pelo Inep, autarquia federal vinculada ao Ministério da Educação (MEC) e mantida pelo Centro de Informação e Biblioteca em Educação (Cibec).

### 3.2 Bases de dados e definição das variáveis do estudo

Conforme descrito anteriormente, com base em Severino (2017), quando trata do uso de dados secundários e recomendações de fontes de pesquisa, este estudo utiliza as bases de dados do Censo da Educação Superior, de 2018 a 2020, disponibilizadas pelo Inep e mantidas pelo Cibec. Foi realizado *download* direto do sítio Microdados (INEP, 2022) complementado por acesso presencial e extração de dados financeiros das IES e seus polos diretamente na biblioteca do Cibec sediado em Brasília/DF, conforme solicitação e processo número 23036.004169/2022-41 (INEP, 2022).

O dicionário de dados das entidades utilizadas neste estudo está apresentado no Anexo A, bem como os sítios de acesso aos códigos fontes e *scripts* utilizados em R, no Anexo B, para importação de geração das bases de dados.

Utilizada como ferramenta para geração das bases de dados, a linguagem R (R CORE TEAM, 2022) mostra-se adequada para as tarefas de importação e tratamento de bases de dados com componentes de análise estatística, pois permite a realização das avaliações necessárias para o atingimento dos objetivos deste trabalho. Para algumas exibições gráficas, como tabelas, quadros e figuras gráficas, será usado o pacote MS Excel.

Na Tabela 4, apresentam-se os resultados diretos da importação e base de dados criada, sem filtros e sem delimitação do estudo, nas quantidades de registros evidenciadas.

**Tabela 4 - Registros totais da base de dados do Censo da Educação Superior**

Entidade	2018	2019	2020
IES	2.537	2.608	2.457
Cursos	38.256	40.709	41.978
Matrículas	8.451.748	8.604.526	8.680.945
Locais de oferta (polos)	767.724	1.254.044	*
Docentes	397.893	399.428	378.492

\* Informação omitida em 2020

Fonte: Elaborado pelo autor, com base nos microdados (INEP, 2022)

A validação, realizada sobre os números gerais de IES e matrículas ativas, é feita sobre comparação direta com as sinopses estatísticas publicadas pelo Inep, prévias à liberação dos microdados (INEP, 2022) e apresentados na Tabela 5, mostrando que as informações totais são iguais, comparando o resultado da importação dos dados analíticos e o resumo da publicação oficial do Inep.

**Tabela 5 - Validação dos registros importados**

Ano Censo	2018		2019		2020	
	Importação	Sinopse	Importação	Sinopse	Importação	Sinopse
Mantenedoras	2.537	2.537	2.608	2.608	2.457	2.457
Matrículas	8.451.748	8.451.748	8.604.526	8.604.526	8.680.945	8.680.945

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022) e Sinopses (INEP, 2022)

Conforme validado por comparação, na Tabela 5, a apresentam-se, para os anos do censo 2018, 2019 e 2020, respectivamente 2.537, 2.608 e 2.457 mantenedoras educacionais de ensino superior. Entende-se que a redução na quantidade de mantenedoras únicas, entre 2019 e 2020, pode ser devido a processos de encerramento de atividades com transferência de mantenças e unificação de mantenedoras de grupos educacionais. Já as quantidades absolutas de matrículas mantêm-se em elevação, no mercado privado: nossa base de dados começa em 2018 com 8.451.748 matrículas, prossegue 2019 com 8.604.526 matrículas na educação superior privada e encerra-se com um total de 8.680.945 matrículas em 2020.

A seguir, no processo metodológico, foi feita a transformação das variáveis categóricas numéricas (classificações) em seus textos significados, de acordo com o dicionário de dados, conforme Quadro 6.

Quadro 6 - Variáveis transformadas

Nome da variável original	Nome da variável transformada	Descrição	Descrição das categorias
<b>Dados da IES</b>			
TP_CATEGORIA_ADMINISTRATIVA	NO_CATEGORIA_ADMINISTRATIVA	Tipo da Categoria Administrativa	1= Pública Federal 2= Pública Estadual 3= Pública Municipal 4= Privada com fins lucrativos 5= Privada sem fins lucrativos 6= Privada - Particular em sentido estrito 7= Especial 8= Privada comunitária 9= Privada confessional
<b>Dados do curso</b>			
TP_GRAU_ACADEMICO	NO_GRAU_ACADEMICO	Tipo do grau acadêmico conferido ao aluno pela conclusão dos requisitos exigidos pelo curso	1= Bacharelado 2= Licenciatura 3= Tecnológico 4= Bacharelado e Licenciatura
TP_MODALIDADE_ENSINO	NO_MODALIDADE_ENSINO	Tipo da modalidade de ensino do curso	1= Presencial 2= Curso a distância
TP_NIVEL_ACADEMICO	NO_NIVEL_ACADEMICO	Tipo do nível acadêmico do curso	1= Graduação 2= Sequencial de Formação Específica

Fonte: Adaptado pelo autor (2023) com base no dicionário de dados dos microdados (INEP, 2020)

Para o objetivo deste estudo e para a importação das IES, de acordo com o Quadro , as variáveis transformadas para suas classificações categóricas são: Categoria Administrativa (TP\_CATEGORIA\_ADMINISTRATIVA em NO\_CATEGORIA\_ADMINISTRATIVA), Organização Acadêmica (TP\_ORGANIZACAO\_ACADEMICA em NO\_ORGANIZACAO\_ACADEMICA) e Região Geográfica da IES (CO\_REGIAO em NO\_REGIÃO). Para os Cursos, foram transformadas e utilizadas as variáveis Grau Acadêmico (TP\_GRAU\_ACADEMICO para NO\_GRAU ACADEMICO), Modalidade de Ensino (TP\_MODALIDADE\_ENSINO para NO\_MODALIDADE\_ENSINO) e Nível Acadêmico (TP\_NIVEL\_ACADEMICO para NO\_NIVEL\_ACADEMICO). Outras variáveis foram transformadas, mas não houve sua aplicação direta no tratamento, filtros e agrupamentos dos dados, assim, foram utilizadas em suas apresentações originais.

### 3.3 Tratamento e agrupamento dos dados

Deve-se considerar que, de acordo com as orientações do manual e dicionário de dados apresentado no Anexo A – Dicionários de Dados das Variáveis do (INEP, 2021) e em posse das informações obtidas na sala segura (INEP, 2022), as informações financeiras utilizadas podem ser relativas especificamente a IES em questão, TP\_REFERENTE = 2, ou referentes à mantenedora à qual a IES pertence, neste caso TP\_REFERENTE = 1, sendo os valores iguais para todas as mantidas (IES) vinculadas a esta mantenedora. Observando a variável de agregação informada, uma das etapas do processo de transformação de dados consistiu em realizar o agrupamento dos valores financeiros por mantenedora, opção que torna possível a comparação posterior.

Quanto às variáveis relativas ao corpo discente, quantidade de ingressantes e matrículas ativas, docentes e funcionários, bem como se as matrículas são na modalidade “a distância” ou “presencial”, foram feitas sumarizações das tabelas analíticas de matrículas por ano do censo, IES e mantenedora, observadas as situações de matrícula utilizadas (matrícula ativas e formados).

Para efeito de estratificação das mantenedoras educacionais e comparação pelo porte (número de alunos), foi aplicado o procedimento metodológico de acordo com a estratificação utilizada pelo Sindicato das Entidades Mantenedoras de

Estabelecimentos de Ensino Superior no Estado de São Paulo – Semesp (SEMESP, 2021, p. 12), que permitiu comparabilidade entre os diferentes grupos de mantenedoras, conforme Quadro 7, a seguir.

**Quadro 7- Estratificação das mantenedoras**

<b>Porte</b>	<b>Alunos</b>
Pequenas (P)	Até 3.000 alunos
Médias (M)	De 3.000 a 7.000 alunos
Grandes (G)	De 7.000 a 20.000 alunos
Gigantes (GG)	Acima de 20.000 alunos

Fonte: Adaptado pelo autor com base em Semesp (2021)

Até 3.000 alunos matriculados (ativos), conforme descrito no Quadro 7, são consideradas pequenas IES, que foram consolidadas em suas mantenedoras. De 3.000 até 7.000 alunos ativos matriculados, foram consideradas mantenedoras de médio porte. Acima de 7.000 alunos ativos matriculados até 20.000 alunos ativos matriculados foram consideradas grandes mantenedoras de ensino superior. Já acima de 20.000 alunos ativos matriculados foram consideradas as gigantes mantenedoras do mercado privado educacional.

O resultado da transformação e agrupamento, em termos numéricos, é apresentado na Tabela 6 e na Tabela 7, para cada ano censo, com os totalizadores considerando as situações apresentadas anteriormente e a estratificação utilizada. Ressalte-se que neste momento já são consideradas todas as mantenedoras privadas brasileiras, com fins lucrativos ou não, de acordo com a classificação apresentada nos microdados, excluindo-se assim as mantenedoras de categoria administrativa pública.



**Tabela 6 - Total de mantenedoras privadas por porte**

Ano Censo	Mantenedoras					
	2018		2019		2020	
	#	%	#	%	#	%
Pequenas (até 3 mil matrículas)	1.240	79,6%	1.248	79,9%	1.186	80,7%
Médias (de 3 a 7 mil matrículas)	176	11,3%	174	11,1%	150	10,2%
Grandes (de 7 a 20 mil matrículas)	94	6,0%	92	5,9%	89	6,1%
Gigantes (acima de 20 mil matrículas)	48	3,1%	48	3,1%	44	3,0%
<b>Total</b>	<b>1.558</b>	<b>100,0%</b>	<b>1.562</b>	<b>100,0%</b>	<b>1.469</b>	<b>100,0%</b>

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

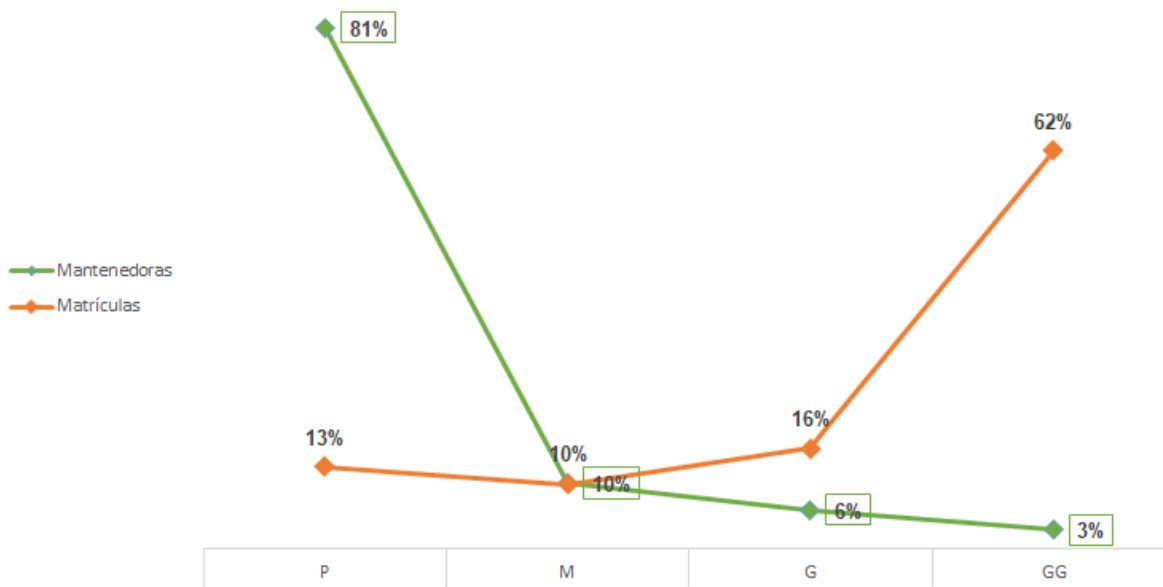
Os dados apresentados na Tabela 6 mostram que a distribuição de mantenedoras em portes tem, de forma consistente ao longo dos anos censitários, a maior quantidade alocada na categoria de pequeno porte (até 3 mil matrículas), com um total no ano censo de 2020 de 1.186 mantenedoras únicas, representando 80,7% do total de mantenedoras do país. Em contrapartida 3% das mantenedoras do país têm mais de 20.000 matrículas ativas, representadas em 44 mantenedoras.

**Tabela 7 - Total de matrículas privadas por porte**

Ano Censo	Matrículas					
	2018		2019		2020	
	#	%	#	%	#	%
Pequenas (até 3 mil matrículas)	881.132	13,8%	881.862	13,5%	852.930	12,7%
Médias (de 3 a 7 mil matrículas)	770.516	12,1%	762.113	11,7%	676.404	10,1%
Grandes (de 7 a 20 mil matrículas)	1.154.721	18,1%	1.098.018	16,8%	1.046.899	15,6%
Gigantes (acima de 20 mil matrículas)	3.566.905	56,0%	3.781.685	58,0%	4.147.769	61,7%
<b>Total</b>	<b>6.373.274</b>	<b>100,0%</b>	<b>6.523.678</b>	<b>100,0%</b>	<b>6.724.002</b>	<b>100,0%</b>

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Já na distribuição das matrículas pelos portes e anos de censo, na Tabela 7, temos que a maior quantidade de matrículas na rede privada de ensino superior está concentrada nas instituições classificadas como “gigantes”, acima de 20 mil alunos, estando 12,7% dos alunos matriculados em 2020 distribuídos nas mantenedoras de pequeno porte. Um gráfico comparativo, apresentado na Figura 1, ilustra a distribuição de mantenedoras e matrículas no ano censo 2020.

**Figura 1 - Censo 2020: comparativo entre portes e matrículas**

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Na Discussão dos Resultados, presente na seção 5.3 deste trabalho, será avaliado o efeito da consolidação das matrículas nas mantenedoras de grande porte, representado na Figura 1, em que 3% das mantenedoras detêm 62%, de forma arredondada, das matrículas do ensino superior no Censo 2020.

No Quadro 8, a seguir, adaptado do manual de preenchimento das variáveis do Censo da Educação Superior pelas IES (INEP, 2021), apresentam-se as variáveis obtidas e computadas nos microdados, relativas a cada mantenedora e ano do censo.

**Quadro 8 - Variáveis obtidas no censo**

<b>RECEITAS</b>	
Receitas próprias	Pagamentos de mensalidades, taxas e serviços cobrados dos alunos de graduação, de forma continuada ou não Financiamento público, bolsas, contratos
Transferências	Valores transferidos para a instituição por outros entes
Outras receitas	Renúncia de Receita fiscal, isenções e outras não enquadradas anteriormente
<b>DESPESAS</b>	
Docentes	Remuneração dos docentes, registrando as despesas relativas ao pagamento de salários, gratificações e outros benefícios
Técnicos Administrativos	Remuneração do pessoal de apoio pedagógico, técnico-administrativo e suporte
Encargos	Despesas relativas ao pagamento de INSS, previdência oficial, abono de férias, 13º salário e outros encargos
Custeio	No caso das instituições públicas, considerar: despesas totais com serviços terceirizados, pagamento de serviços públicos
Investimentos	Investimentos feitos em infraestrutura, aquisição de imóveis, concessão de empréstimos, participação em constituição ou aumento de capital de empresas ou entidades comerciais ou financeiras, aquisição de títulos representativos de capital de empresa em funcionamento e outras inversões financeiras, e demais despesas da IES que resultam no acréscimo do patrimônio
Pesquisa e Desenvolvimento	Auxílios financeiros a recenseador e à manutenção de sua equipe; aquisição de material permanente, equipamentos, obras de infraestrutura e instalações para pesquisa; entre outros
Outras despesas	Despesas que não estão contempladas nas demais despesas conceituadas

Fonte: Adaptado pelo autor (2023) de Inep (2021)

Após realização das extrações acima descritas para os anos 2018, 2019 e 2020, será realizado o cálculo de variação entre os indicadores relacionados ao ponto de equilíbrio e fluxo de matrículas.

### 3.4 Criação das variáveis de regressão

Com a importação das variáveis e a criação das bases de dados, relacionadas no Quadro 8, foi utilizada a análise de regressão linear múltipla para estabelecer a relação entre os indicadores calculados, para cada ano do censo de forma separada, contribuindo para a formação de expressões que possam representar o distanciamento do ponto de equilíbrio.

Uma questão que se coloca é comparar valores e quantitativos com escalas métricas variadas, considerando o porte das instituições. Para esta análise, foram calculadas e utilizadas variáveis relativas, sempre com base na quantidade de alunos apresentada pela mantenedora, quando o valor já não foi relativizado em sua origem.

Os dados absolutos são aqueles obtidos diretamente da fonte, sem manipulação, representando uma contagem ou medida. Já dados relativos, resultado de transformações dos dados absolutos, permitem igualar o denominador de comparação, obtendo-se índices como coeficientes ou taxas (ASSIS; SOUSA; DIAS, 2019).

Quanto ao objetivo de investigação, fatores que afetam o ponto de equilíbrio financeiro, dado o levantamento bibliográfico realizado e as variáveis obtidas com os microdados do censo, será construída a estrutura de dados, representados no Quadro 9 e no Quadro 10, a seguir.

**Quadro 9 - Variáveis dependentes**

Variável	Descrição	Definição
<b>VDEPQ</b>	Variável Distância do Ponto de Equilíbrio	Variável que será utilizada para regressão linear múltipla como variável dependente, representando o distanciamento do ponto de equilíbrio financeiro, em relação aos custos (Receitas – Despesas) / Despesas), podendo ser negativo ou positivo. Consideram-se somente receitas com mensalidades.

Fonte: Elaborado pelo autor, (2023)

**Quadro 10- Variáveis independentes**

Variável	Descrição	Definição
<b>VRAD</b>	Variável alunos/docentes	Variável que descreve a relação entre as quantidades absolutas de alunos e docentes. Representa eficiência e escala na alocação docente.
<b>VRAF</b>	Variável alunos/funcionários	Variável que descreve a relação entre as quantidades absolutas de alunos e funcionários administrativos. Representa eficiência e escala na operação administrativa.
<b>VCFPA</b>	Variável custos de pessoal / aluno	Variável que descreve o custo anual operacional por aluno, sendo representado por (Despesas Docentes + Administrativas + Encargos) / Quantidade de Alunos. É diretamente ligado com a linha contábil de custos de operação.
<b>VCODA</b>	Variável custeio + outras despesas) / aluno	Variável que descreve os custos operacionais adicionais (exceto investimento e pesquisa) por aluno, sendo representado por (Despesas Custeio + Outras) / Quantidade de Alunos
<b>VTMA</b>	Variável de <i>ticket</i> médio do aluno	Variável que representa a receita própria anual (mensalidades e serviços) / Quantidade de Alunos
<b>VOCIO</b>	Variável de ociosidade no ingresso.	Variável que representa a capacidade gerencial, comercial e de marketing de ocupação das vagas e ofertas. Quantidade de Ingressantes no ano / Vagas Ofertadas no ano.
<b>VEAD</b>	Variável de exposição à operação em EAD	Variável que representa a quantidade de alunos matriculados em EAD sobre a quantidade total da mantenedora.

<b>VPOL</b>	Variável de exposição à interiorização da oferta.	Variável que representa a quantidade de alunos matriculados em polos externos ao principal sobre a quantidade total da mantenedora.
<b>IQCTA</b>	Indicador de qualidade técnico-administrativa	Variável que representa o indicador do MEC (INEP, 2018) de gestão técnico-administrativa. Representa investimento e custo na equipe administrativa. É formulado por: $IQCTA = (5D + 3M + 2E + 1G + 0,75MD + 0,5B) / (D + M + E + G + MD + B)$ , em que: IQCTA = índice de qualificação do corpo técnico-administrativo; D = número de doutores; M = número de mestres; E = número de especialistas; G = número de graduados; MD = número de funcionários com ensino médio; B = número de funcionários com ensino fundamental.
<b>IQCDO</b>	Indicador de qualidade docente	Variável que representa indicador do MEC (INEP, 2018) representa investimento e custo na equipe docente. É formulado por: $IQCDO = (5D + 3M + 2E + 1G) / (D + M + E + G)$ , em que: IQCDO = índice de qualificação do corpo docente; D = número de doutores; M = número de mestres; E = número de especialistas; G = número de graduados.
<b>VFINE</b>	Indicador de exposição a financiamento estudantil	Representa a quantidade absoluta de alunos que estão sinalizados como recebedores de qualquer tipo de financiamento sobre a quantidade total de alunos da mantenedora, relacionado ao equilíbrio financeiro.
<b>VBOLS</b>	Indicador de aplicação de suporte, bolsas e apoio ao estudo	Representa a quantidade absoluta de alunos que estão sinalizados como recebedores de qualquer tipo de apoio social ou bolsas, sobre a quantidade total de alunos da mantenedora, relacionado ao equilíbrio financeiro.
<b>VEVAD</b>	Indicador de evasão em relação ao total absoluto da mantenedora	Indicador da evasão medida na mantenedora, relacionado ao equilíbrio financeiro.
<b>VRENC</b>	Relação entre as despesas com encargos trabalhistas e a remuneração da folha de pagamento	Variável que representa uma razão entre os gastos com encargos e a folha de pagamento, representando maior ou menor aproveitamento de recursos e planejamento tributário.

Fonte: Elaborado pelo autor, (2023)

A linguagem R (R CORE TEAM, 2022) é utilizada como ferramenta para geração das bases de variáveis dependentes e independentes e os cálculos relacionados.

### 3.5 Métodos estatísticos utilizados

Esta sessão apresenta um resumo que embasa e qualifica os métodos estatísticos utilizados nos procedimentos metodológicos, bem como as qualificações desses modelos e valores de confiança.

#### 3.5.1 Regressão Linear Multivariada

A Regressão Linear Multivariada (RLM) é uma das técnicas de regressão mais utilizadas e caracteriza-se por avaliar múltiplas medidas, dimensões ou variáveis sobre um ou mais indivíduos em investigação (HAIR *et al.*, 2009). Nela, aplicam-se os mesmos pressupostos de uma regressão linear simples (uma única variável dependente ( $y$ ) com uma única variável independente ( $x$ ) (OPENINTRO.ORG, s.d.).

Montgomery, Peck e Vining (2017) defendem que a Regressão Linear Multivariada é um método estatístico que permite modelar a relação entre uma variável dependente (ou resposta) e múltiplas variáveis independentes (ou explicativas). Trata-se de um caso especial da análise de regressão que permite lidar com mais de uma variável independente.

Segundo Chein (2019), o instrumento utilizado – regressão linear múltipla – permite analisar quantitativamente as relações entre variáveis, sendo uma dependente e outras independentes ou explicativas, traçando múltiplos modelos de retas ( $y = a+bx$ ) que tragam a inclinação da reta que mais explica o efeito da mudança em  $y$ . Similar definição é descrita por Assis, Sousa e Dias (2019, p.256) como “um modelo estatístico no qual a relação entre o preditor e a variável resposta pode ser ilustrada por uma linha reta”, enquanto uma equação linear simples pode ser representada conforme a Equação (II).

(II)

$$Y_i = \beta_0 + \beta_1 V_1$$

A equação de regressão múltipla adiciona cada coeficiente de regressão, de forma positiva ou negativa, à respectiva variável, obedecendo ao intercepto e erro ajustado, exemplificado a seguir na Equação (III)

$$y_i = \beta_0 + \beta_1 V_{1+} + \beta_2 V_{2+} + \beta_n V_n + e \quad (\text{III})$$

em que:

$y_i$  = Variável dependente ou preditora; neste estudo, única;

$\beta_0$  = intercepto ou constante da regressão;

$\beta_n$  = variação calculada ou coeficiente de regressão, para cada variável;

$V_n$  = valor registrado na variável correspondente;

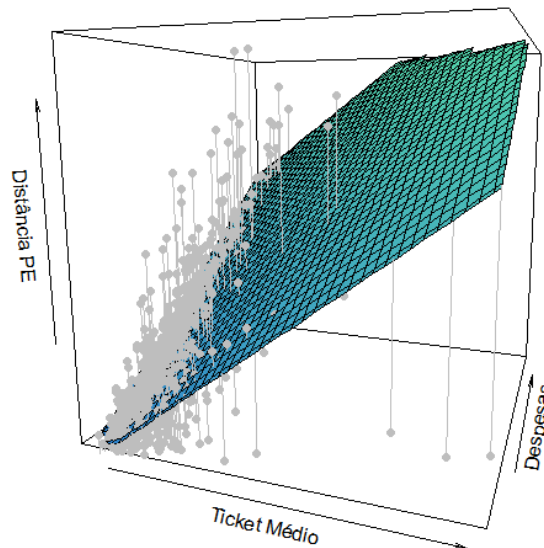
$e$  = erro ou desvio do modelo.

Conforme descrito por Burgel e Anzanello (2018, p. 158), o modelo procura ajustar uma reta (linearidade) reduzindo ao máximo os resíduos entre os dados calculados:

Na regressão linear múltipla procura-se minimizar a soma do desvio quadrado entre a resposta ajustada e a resposta verdadeira no espaço medido pelas variáveis explicativas. Neste processo, cada variável independente é ponderada para assegurar a máxima previsão a partir do conjunto de variáveis independentes.

Enquanto em uma regressão linear simples temos como representação gráfica uma linha cruzando os eixos  $x$  (independente ou preditora) e  $y$  (dependente ou predita), com os pontos ajustando-se o mais próximo da reta, na regressão linear múltipla, a representação visual é através de um hiperplano, representando  $n$  variáveis, a variável predita e os pontos com os valores, conforme Figura 2, que representa nosso objeto de estudo (ponto de equilíbrio) em relação a duas variáveis, *Ticket Médio* e *Despesas por Aluno*.

**Figura 2 - Hiperplano de variáveis em regressão múltipla**  
**Hiperplano Ticket Médio x Despesas**



Fonte: Elaborado pelo autor, (2023).

Percebe-se, na Figura 2, que, em uma análise multidimensional, a reta de regressão de uma análise linear simples torna-se um plano que passa o mais próximo dos pontos relacionados a cada variável regredida e valores obtidos, minimizando o erro médio.

A aplicação de RLM justifica-se pelo objetivo do estudo, cujo propósito é estabelecer uma relação entre uma variável contínua dependente e um conjunto de variáveis independentes, procurando estabelecer uma correlação entre elas, além de ser, segundo a Associação Brasileira de Normas Técnicas (ABNT), a mais utilizada na avaliação deste comportamento (ABNT, 2010). A descrição desta relação também pode ser entendida como modelagem probabilística de relação entre uma variável predita, ou dependente, e variáveis preditoras, ou independentes, com o intuito de fornecer uma saída relacional dentro de um intervalo de confiança do modelo (DEVORE, 2006; KREMER; DEINA; SIQUEIRA, 2019; OPENINTRO.ORG, 2022).

Segundo HAIR *et al.* (2009, p. 30), a técnica de RLM é escolhida para avaliar a dependência da variável predita e as preditoras, considerando somente uma predita (dependente) e esta ser métrica (contínua), estabelecendo assim um caminho de escolha adequado dos dados selecionados no modelo estatístico.



Para adequação do modelo de regressão, algumas regras sobre os dados devem ser observadas, como tamanho da amostra, tratamento dos *outliers* extremos e multicolinearidade (ABNT, 2010; FÁVERO, 2015).

O tamanho da amostra, conforme descrito por Hair *et al.* (2009, p. 167), tem impacto direto sobre a qualidade da regressão, assim amostras grandes (com mais de mil observações), como o caso deste estudo, tornam os dados de significância estatística muito sensíveis.

Para reduzir a sensibilidade, aumentar a linearidade e trazer estabilidade ao modelo, recomenda-se realizar um tratamento de *outliers* extremos e valores vazios. Para o caso de dados multivariados, um dos métodos de tratamento e exclusão dos *outliers* mais utilizado é o modelo por distância de Mahalanobis (TABACHNICK; FIDELL, 2019), que leva em consideração não somente a distância euclidiana (*z-score*), tal como usada em métodos univariados, mas também a correlação entre as variáveis.

Este modelo é baseado na teoria da estatística multivariada e utiliza a matriz de covariância dos dados para transformar as distâncias euclidianas tradicionais em distâncias mais precisas, que levam em conta a relação entre as variáveis. Esta medida de distância, ainda segundo Tabachnick e Fidell (2019, p. 74) estabelece que valores de  $p < 0,001$  são potenciais *outliers*.

Conforme Gujarati (2011, p. 74), a multicolinearidade, que representa o quanto as variáveis são redundantes entre si, pode ser verificada pelo teste VIF (*Variance Inflation*). Recomenda-se que no modelo o valor estimado do teste VIF, para cada variável, não seja superior a 10 (HAIR *et al.*, 2009; GUJARATI, 2011; SALVIAN, 2016). A fórmula do cálculo VIF pode ser representada pela Equação (IV), a seguir.

$$VIF_n = \frac{1}{1 - r_n^2} \quad (IV)$$

em que:

$VIF_n$  = valor calculado do VIF, para n variáveis preditoras;

$r_n^2$  = Coeficiente de correlação da regressão ajustada, para n variáveis.

O efeito das variáveis no modelo de regressão linear multivariada pode ser medido pelo coeficiente de determinação ( $R^2$ ), que varia de zero (mais fraco) até um (mais forte), descrito por Hair *et al.* (2009, p. 160) e Fávero (2015) como a medida de precisão da previsão do modelo, é representado pela Equação (V).

$$\text{Coeficiente de Determinação } (R^2) = \frac{\text{Soma dos quadrados da regressão}}{\text{Soma total dos quadrados}} \quad (V)$$

A interpretação do nível de precisão da previsão pode ser assim classificada quanto à força, de acordo com Fávero (2015) e Garson (2016):

- $0 \leq R^2 \leq 0,3$  representa fraca capacidade de predição (a variabilidade dos dados preditores explica em até 30% a variável predita);
- $0,3 \leq R^2 \leq 0,6$  representa uma moderada capacidade de predição (a variabilidade dos dados preditores explica entre 30% e 60% a variável predita);
- $0,6 \leq R^2 \leq 1$  representa forte capacidade de predição (a variabilidade dos dados preditores explica entre 60% e 100% a variável predita).

Ainda segundo Hair *et al.* (2009, p. 160) e Fávero (2015), outra medida para descrever a importância relativa de cada variável preditora na explicação da variabilidade da variável resposta é o Beta padronizado, que coloca todas as variáveis na mesma escala e nos permite comparar diretamente seus efeitos relativos na variável resposta.

### 3.5.2 Análise de Componentes Principais

Uma dentre várias técnicas aplicadas na Regressão Linear Multivariada, a Análise de Componentes Principais (*Principal Component Analysis – PCA*) permite a identificação da significância relativa de variáveis preditoras, em um modelo fatorial, agrupando as variáveis possibilitando obter a maior significância para o modelo com o menor número delas. (ASSIS; SOUSA; DIAS, 2019, p. 69).

A Análise de Componentes Principais (PCA) é uma técnica estatística utilizada para analisar e entender a variação de dados multidimensionais, que procura encontrar os "componentes principais", ou seja, as direções de maior variação nos dados, e os representa como novos eixos ortogonais (perpendiculares entre si), chamados de componentes principais. Esses componentes são ordenados de acordo com a quantidade de variação explicada, sendo o primeiro componente principal o que explica a maior parte da variação e assim por diante.

Esta técnica, amplamente utilizada em áreas como inteligência artificial, bioinformática e análise de sinais (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2013), permite identificar, em conjuntos de dados muito grandes, variáveis preditoras que em conjunto possam determinar, através de correlações entre elas, boa parte do Coeficiente de Predição (R<sup>2</sup>) obtido pela técnica de regressão linear multivariada, reduzindo assim a quantidade de variáveis preditoras necessárias para atingir boa parte da predição (TABACHNICK; FIDELL, 2019).

A técnica de Análise de Componentes Principais cria grande quantidade de combinações lineares entre as variáveis preditoras, combinadas em fatores – normalmente a quantidade de fatores relacionados com a quantidade de variáveis preditoras –, gerando assim uma grande matriz de sucesso e ordenando os fatores em ordem decrescente de sucesso (TABACHNICK; FIDELL, 2019).

Ainda em Weisberg (2014, p. 118), entende-se que a Análise de Componentes Principais pode ser resumida como uma técnica estatística que substitui muitas variáveis pela combinação linear entre elas, sendo estes novos regressores muito mais fáceis de manipular e explicar, nos termos do modelo criado, sendo a essência desta técnica também compartilhada por Hair *et al.* (2009, p. 112) e Gujarati (2011, p. 76).

### 3.5.3 Análise por Floresta Aleatória (*Random Forest*)

Outra técnica aplicada à análise multivariada é a análise por Floresta Aleatória (*Random Forest*), derivada de algoritmos utilizados em análises de inteligência artificial (IA). Esta técnica deriva dos algoritmos de decisão por árvore, que se resume na criação de agrupamentos (nós) que derivam em outros nós até que estes não tenham “nós filhos” (determinados como folhas).

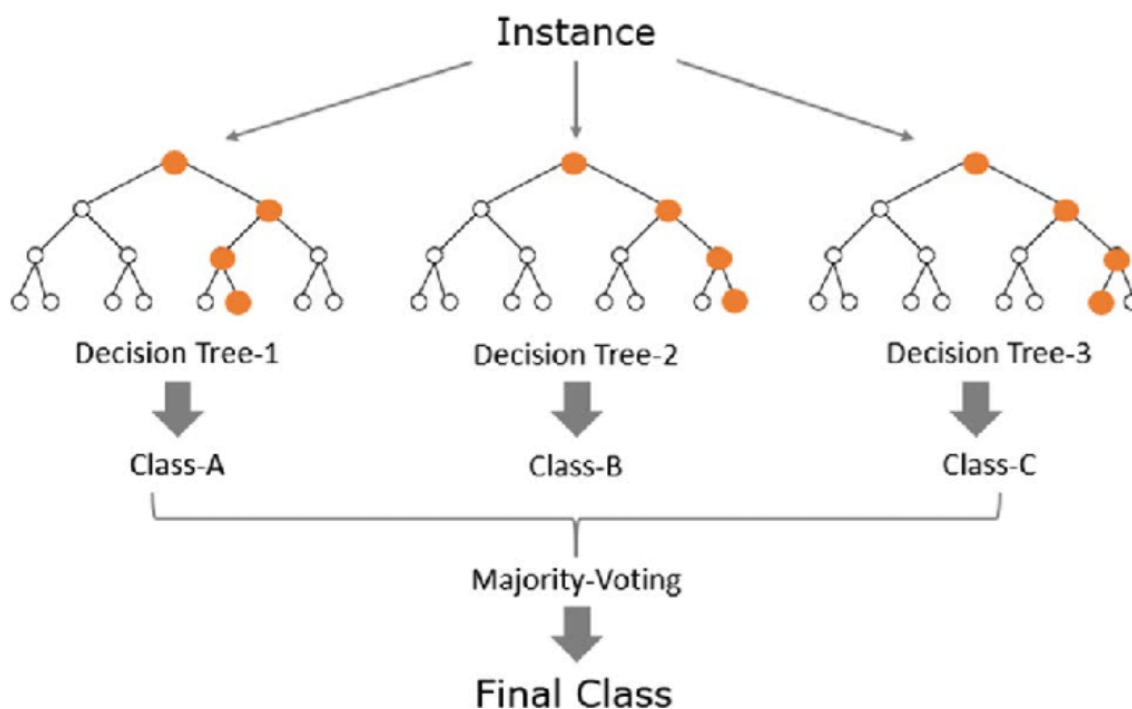
Também é definido como um algoritmo de aprendizado de máquina supervisionado que consiste em criar várias árvores de decisão, treinadas com um subconjunto aleatório dos dados de treinamento, e depois faz a classificação ou previsão tomando a maioria das classificações ou previsões dessas árvores. Isso ajuda a reduzir o super ajuste e melhora a precisão das previsões, sendo também usado para seleção de características e identificação de padrões importantes nos dados. (JAMES *et al.*, 2013)

A grande vantagem do uso desta técnica é que não necessitam de tratamento inicial em seus dados e variáveis e a assertividade não é determinada pelas premissas de normalidade (PORFIRIO, 2021).

Portanto, o conceito e árvores randômicas acaba por tratar, como o nome por sugerir, da criação de múltiplas árvores de decisão que são agrupadas em classes mais presentes, de acordo com os elementos encontrados. Além disso, conforme Avelar (2022, p. 49), é uma melhor aplicação dos algoritmos de árvore de decisão, onde cada árvore funciona como um classificador independente integrando-se em um único resultado (BREIMAN, 2001).

O funcionamento de uma floresta randômica é ilustrado na Figura 3, em que Tran (2019) defende a aplicação e conseqüentemente a precisão da classificação em situações multivariadas, em que a partir da instância principal de dados (base), para cada situação (classe) de nossa variável predita, busca-se criar árvores decisórias atribuindo pesos (distâncias e agrupamentos) criando assim as classificações. O algoritmo, por final, agrupa as classificações e entrega uma classificação final ao processo.

Figura 3 - Exemplo de Floresta Randômica



Fonte: Tran (2019, p.12)

Como pode ser observado na Figura 3, o número de árvores de decisão criadas é fator importante no algoritmo de classificação. Conforme discutido por Khan et al. (2020, p. 97), um número baixo de árvores definidas no algoritmo de classificação pode aumentar a não acurácia da predição, acarretando baixa diversidade de análise das variáveis, embora não represente esforço computacional em sua geração. Já a adoção de um número elevado de árvores de decisão no modelo impacta diretamente no esforço computacional, que, para grandes quantidades de variáveis e registros de teste, pode comprometer o sucesso da solução, embora traga uma acurácia maior ao modelo preditivo (KHAN et al., 2020).

Na próxima seção, serão apresentadas as descobertas de cada base de dados relativos aos anos censo e um resumo das variáveis e índices de forma agrupada para cada método estatístico utilizado.



## 4 RESULTADOS OBTIDOS

Nesta seção, será realizada a apresentação dos resultados obtidos da base de dados criada a partir dos dados secundários do censo e as variáveis derivadas e apresentadas na seção anterior, procedimentos metodológicos.

Inicialmente será feita a apresentação tabulada das bases de dados criadas e, em tópico seguinte, o tratamento de *outliers*, conforme método estatístico de Mahalanobis. Em seguida será apresentado, para cada ano censo as estatísticas descritivas e gerais e a preparação para as regressões, com estudo da correlação e multicolinearidade.

A análise de componentes principais e a regressão linear múltipla encerram esta seção, preparando para a discussão dos resultados encontrados.

### 4.1 Bases de dados

As análises apresentadas neste estudo foram geradas a partir dos dados apresentados nos microdados do Censo da Educação Superior, para os anos 2018, 2019 e 2020. Neste tópico serão apresentadas uma análise descritiva dos dados, bem como estatísticas que irão nortear as análises posteriores.

Após o processo de importação e tratamento dos dados brutos apresentados nos procedimentos metodológicos, temos o resumo da base de dados já apresentado nas Tabelas 6 e 7, em que temos as informações descritivas, para cada ano censo e sintetizamos o quantitativo de mantenedoras obtido bem como a média e o IQR calculado, de cada uma das variáveis a serem utilizadas na análise de regressão.

O IQR pode ser uma medida de dispersão estatística que representa a diferença entre o terceiro e o primeiro quartil de uma distribuição de dados. Ele é frequentemente usado para identificar *outliers* (valores atípicos) em um conjunto de dados, conforme descrito por (FIELD, 2009), esse IQR representando a faixa de distribuição em que 50% dos dados estão posicionados (ASSIS; SOUSA; DIAS, 2019).

Na Tabela 8, apresentam-se as médias calculadas e o IQR para cada variável.

Tabela 8 - Estatística descritiva da base de dados

	Ano Censo		
	2018	2019	2020
Mantenedoras	1558	1562	1469
Variável	Média (IQR)	Média (IQR)	Média (IQR)
<b>VDPEQ</b> (Dist. Ponto Eq.)	0.15 (-0.11,0.34)	0.17 (-0.14,0.34)	2.42 (-0.15,0.40)
<b>VRAD</b> (Alunos/Docentes)	18 (8,21)	18 (7,21)	19 (8,21)
<b>VRAF</b> (Alunos/Funcion.)	25 (10,31)	26 (10,32)	29 (11,36)
<b>VCFPA</b> (Folha Docente)	19,255 (2,934,8,974)	40,886 (2,868,9,127)	16,649 (2,772,9,582)
<b>VCODA</b> (Custo Operação)	9,909 (1,020,5,172)	12,773 (976,5,658)	9,452 (976,5,949)
<b>VTMA</b> (Ticket Médio)	26,586 (4,936,14,401)	32,434 (4,780,14,816)	38,351 (4,865,16,137)
<b>VOCIO</b> (Ocios. de Vagas)	0.32 (0.14,0.44)	0.32 (0.13,0.43)	0.26 (0.10,0.35)
<b>VEAD</b> (EAD)	0.03 (0.00,0.00)	0.05 (0.00,0.00)	0.07 (0.00,0.00)
<b>VPOL</b> (Polos)	0.86 (1.00,1.00)	0.86 (1.00,1.00)	0.15 (0.00,0.14)
<b>IQCTA</b> (Qual. Adm.)	1.22 (0.93,1.41)	1.25 (0.95,1.44)	1.28 (1.00,1.47)
<b>IQCDO</b> (Qual. Docente)	3.12 (2.80,3.41)	3.15 (2.84,3.44)	3.19 (2.88,3.49)
<b>VFINE</b> (Financiamento)	0.33 (0.01,0.55)	0.31 (0.00,0.52)	0.29 (0.00,0.47)
<b>VBOLS</b> (Bolsas)	0.04 (0.00,0.00)	0.04 (0.00,0.00)	0.04 (0.00,0.00)
<b>VEVAD</b> (Evasão)	0.25 (0.16,0.32)	0.26 (0.16,0.33)	0.28 (0.18,0.36)
<b>VRENC</b> (Encargos)	0.38 (0.22,0.48)	0.37 (0.22,0.48)	0.38 (0.19,0.46)

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Na Tabela 8, entre as variáveis criadas, destacam-se algumas considerações, ressaltadas por uma análise descritiva e comparativa entre as variáveis.

A variável dependente, VDPEQ, apresenta grande variação de média calculada entre os anos censo selecionados, 2018/2019 e comparados com 2020, embora mantenha um intervalo interquartil praticamente estável.

Já nas variáveis independentes, observa-se o crescimento gradual de exposição à modalidade de ensino a distância (VEAD), da relação alunos-funcionários e alunos-docentes (VRAF, VRAD) e de qualificação técnico-administrativa e docente (IQCTA e IQCDO).

#### 4.2 Remoção de dados extremos ou inválidos – *outliers*

Com as bases de dados anteriormente estabelecidas, no segundo passo antes da regressão, será executado o processo de avaliação de dados extremos ou dados inválidos, conhecidos como *outliers*. O objetivo deste trabalho é utilizar a base de dados multivariada desenvolvida, recomendando-se a análise por distância de



Mahalanobis (TABACHNICK; FIDELL, 2019) e filtrando-se, assim, os registros indicativos com valores de  $p < 0,001$ , considerados potenciais *outliers*. A aplicação desse processo permite aumentar a estabilidade dos modelos criados e é comumente usada para identificar e classificar *outliers* em conjuntos de dados multivariados.

Na Tabela 9, a seguir, apresenta-se, após a aplicação do algoritmo de Mahalanobis nos conjuntos de dados, ano a ano do censo, qual foi a redução amostral.

**Tabela 9 - Redução amostral na análise de outliers**

Número de Registros	Ano Censo		
	2018	2019	2020
Antes	1.558	1.562	1.469
Depois	1.450	1.482	1.376
Diferença	108	80	93
Redução (%)	-6,9%	-5,1%	-6,3%

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

A análise de Mahalanobis (TABACHNICK; FIDELL, 2019) e posterior remoção de *outliers* nas bases de dados permitiu que a redução amostral se estabelecesse em torno da média absoluta 6,1%, com 94 registros excluídos na média total, mantendo-se, assim, uma amostra da base de dados para análise e regressões adequada em quantidade de registros.

Em uma observação dos registros classificados como *outliers* ( $p > 0,001$ ) da amostra original, podemos estabelecer uma classificação dos motivos de exclusão ou posicionamento fora da distância calculada, mostrada na Tabela 10.

**Tabela 10 - Estudo dos registros outliers**

Classificação dos registros outliers	Ano Censo		
	2018	2019	2020
Variáveis necessárias vazias ou inválidas	26	17	44
Dados inválidos ou fora de distância	82	63	49
Total	108	80	93

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Pode-se verificar que a maioria dos registros excluídos das bases de dados é resultado do cálculo da distância de Mahalanobis e, em menor grau, dados vazios ou inválidos. Também pela baixa contribuição de *outliers* nas bases de dados (6,1%),

entende-se que esta amostra se percebe consistente, estabelecendo assim parâmetros adequados para a investigação seguinte, a análise de correlações e covariância.

#### 4.3 Análise de correlações e covariâncias nas bases de dados

Após a remoção dos registros *outliers*, foi estabelecida a base de dados de trabalho, para cada ano censo, cujo resumo descritivo é apresentado na Tabela 11.

**Tabela 11 - Bases de dados após tratamento de *outliers***

	Ano Censo		
	2018	2019	2020
Mantenedoras	1.450	1.481	1.469
Variável	Média (IQR)	Média (IQR)	Média (IQR)
<b>VDPEQ</b>	0.16 (-0.10,0.34)	0.18 (-0.13,0.34)	0.45 (-0.15,0.39)
<b>VRAD</b>	17 (8,21)	16 (8,20)	17 (8,21)
<b>VRAF</b>	24 (11,31)	24 (10,32)	27 (11,35)
<b>VCFPA</b>	7,845 (2,941,8,692)	9,221 (2,891,9,027)	9,625 (2,794,9,389)
<b>VCODA</b>	4,987 (1,024,4,978)	6,827 (982,5,581)	5,914 (979,5,843)
<b>VTMA</b>	13,536 (5,016,14,046)	16,829 (4,859,14,667)	18,885 (5,027,15,810)
<b>VOCIO</b>	0.32 (0.14,0.44)	0.32 (0.13,0.43)	0.26 (0.09,0.35)
<b>VEAD</b>	0.02 (0.00,0.00)	0.03 (0.00,0.00)	0.06 (0.00,0.00)
<b>VPOL</b>	0.86 (1.00,1.00)	0.85 (1.00,1.00)	0.15 (0.00,0.14)
<b>IQCTA</b>	1.21 (0.93,1.39)	1.24 (0.95,1.42)	1.28 (1.00,1.46)
<b>IQCDO</b>	3.11 (2.80,3.41)	3.14 (2.84,3.43)	3.18 (2.87,3.48)
<b>VFINE</b>	0.34 (0.02,0.56)	0.32 (0.00,0.53)	0.29 (0.00,0.47)
<b>VBOLS</b>	0.020 (0.000,0.000)	0.02 (0.00,0.00)	0.019 (0.000,0.000)
<b>VEVAD</b>	0.25 (0.16,0.32)	0.25 (0.16,0.32)	0.28 (0.18,0.36)
<b>VRENC</b>	0.37 (0.23,0.48)	0.37 (0.22,0.47)	0.36 (0.18,0.46)

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Em uma verificação dos números apresentados após o tratamento, percebe-se uma homogeneidade superior em relação aos dados da base de dados sem o tratamento dos dados extremos, principalmente nos valores médios calculados. Comparando ano a ano, as faixas interquartis também se apresentam com pouca ou nenhuma variação, mostrando assim que a maior parte dos dados variáveis é distribuída de forma parecida nos três anos do censo.

#### 4.4 Análise de correlação e colinearidade

Com a base de dados estabelecida e validada, fizemos o teste de correlação, que, conforme Assis, Sousa e Dias (2019, p. 72), mede o grau em que as variáveis estão associadas e o impacto que estabelecem uma sobre a outra. Esta análise será realizada e direcionada aplicando-se a variável dependente (VDPEQ) em relação às outras variáveis preditoras, comparando ano a ano. O índice de correlação varia de -1 (mais forte correlação negativa) até 1 (mais forte correlação positiva). A correlação entre a variável dependente (VDPEQ) e as independentes deste estudo está, a cada ano censo, tabuladas na Tabela 12, que resume a análise de todos os anos censo da base de dados.

**Tabela 12 - Intervalos de confiança de correlação - VDPEQ**

	Ano Censo		
	2018	2019	2020
VRAD	0,20	0,07	0,01
VRAF	0,22	0,07	0,04
VCFPA	-0,14	-0,04	-0,02
VCODA	-0,10	-0,03	-0,03
VTMA	0,13	0,21	0,66
VOCIO	0,06	0,00	0,01
VEAD	-0,01	-0,02	0,00
VPOL	-0,02	0,02	0,00
IQCTA	-0,08	-0,01	-0,03
IQCDO	-0,04	-0,04	-0,01
VFINE	0,10	0,02	-0,01
VBOLS	-0,04	-0,02	0,08
VEVAD	-0,13	-0,09	-0,03
VRENC	-0,06	-0,07	0,01

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Na Tabela 12, percebe-se, comparando ano a ano, que o maior fator de correlação individual entre a variável predita, VDPEQ, e as demais variáveis preditoras está, em 2020, com a variável VTMA (0,66). Aprofundando a discussão dos resultados neste trabalho, é forte indicativo que, para este ano censo e este grupo de variáveis, o *Ticket Médio do Aluno* foi importante, no sentido positivo, no distanciamento do ponto de equilíbrio financeiro.

Uma análise necessária para a avaliação do resultado das regressões é o estudo de multicolinearidade, recomendado para estabelecer a qualidade de um modelo gerado (Tabela 13). Conforme descrições referenciadas anteriormente, a multicolinearidade representa o quanto as variáveis são redundantes entre si e para isso o teste a ser utilizado é o *Variance Inflation (VIF)*, cujo resultado não pode ser superior a 10 (HAIR *et al.*, 2009; GUJARATI, 2011; SALVIAN, 2016).

**Tabela 13 - Avaliação de multicolinearidade - VIF**

Variável	Ano Censo		
	2018	2019	2020
VRAD	1,88	1,96	2,20
VRAF	1,57	1,58	1,69
VCFPA	3,43	3,19	1,76
VCODA	2,87	3,52	1,53
VTMA	4,97	6,87	1,53
VOCIO	1,11	1,09	1,08
VEAD	1,23	1,34	1,44
VPOL	1,12	1,12	1,01
IQCTA	1,04	1,06	1,03
IQCDO	1,14	1,18	1,09
VFINE	1,08	1,07	1,05
VBOLS	1,01	1,01	1,02
VEVAD	1,06	1,04	1,04
VRENC	1,08	1,07	1,04

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

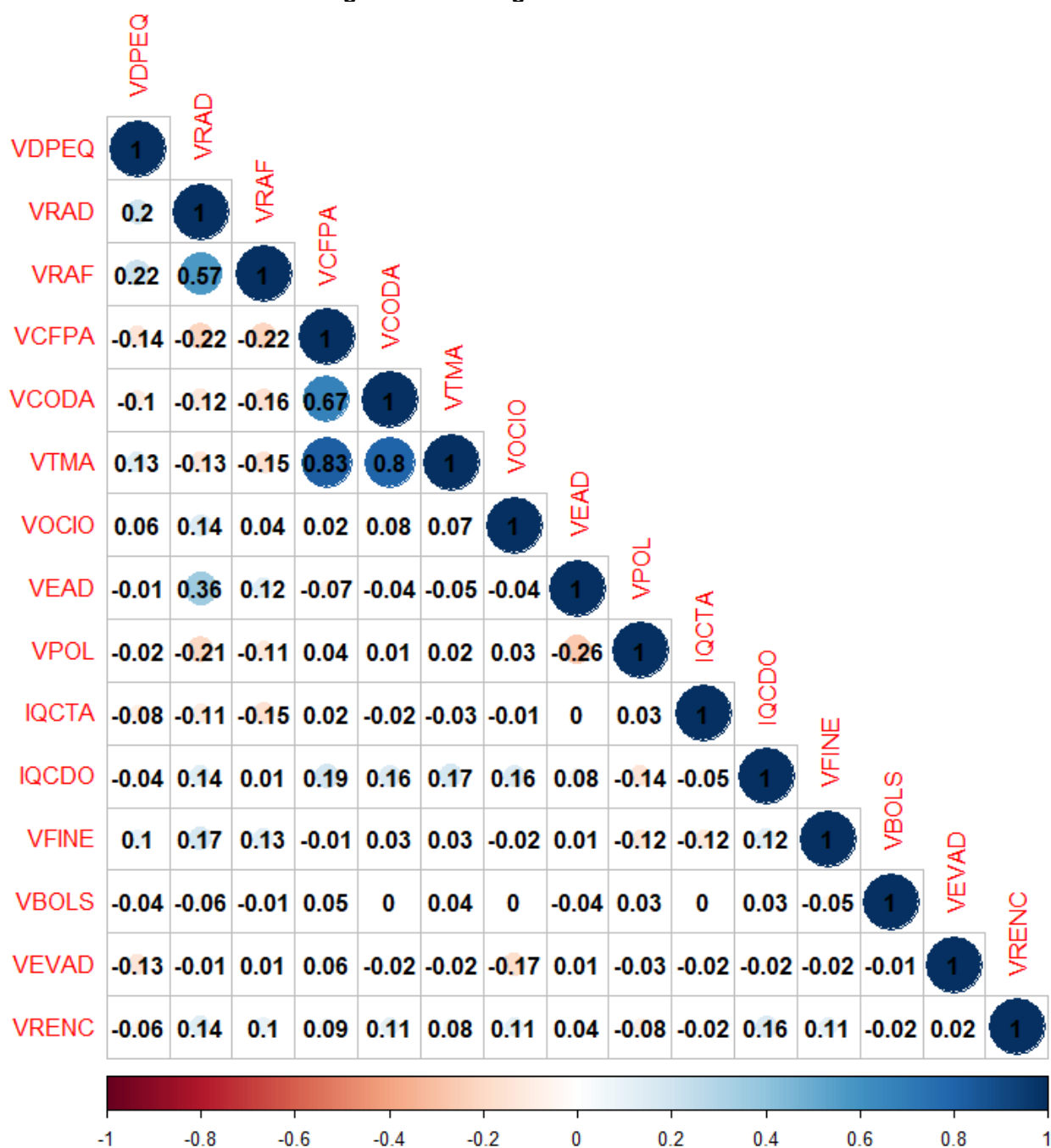
Seguindo as regras de avaliação propostas para detecção de multicolinearidade, percebe-se que os dados estão abaixo do limite de tolerância indicado,  $VIF < 10$ , embora tenham apresentado alguma significância para atribuímos maior atenção à variável VTMA nos anos 2018 e 2019. Todas as demais apresentaram valores abaixo de 4, demonstrando assim que a multicolinearidade não é significativa nos modelos gerados.

Nas subseções seguintes, iremos detalhar os achados nas correlações ano censo a ano censo, o que irá nos permitir verificar outros fatores influentes nas variáveis.

## 4.5 Análise de correlação – 2018

O estudo de correlação na base de dados para o ano censo 2018 é apresentado na Figura 4, na qual se ilustram as variáveis predita e predictoras, permitindo a interpretação da relação entre elas. Foi utilizada a escala de Garson (2016) para estabelecimento do grau de correlação.

Figura 4 - Correlograma das variáveis - 2018



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

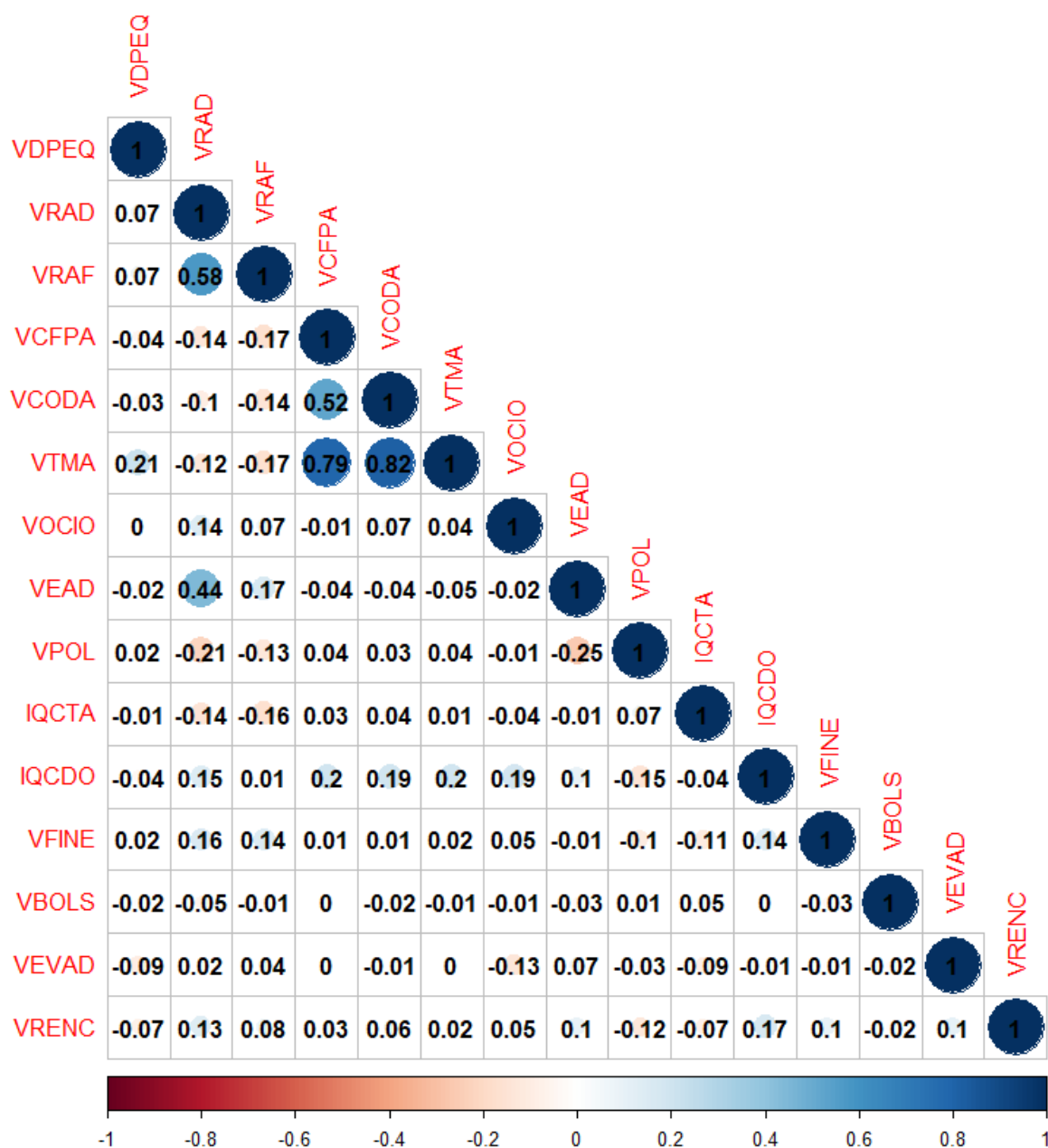
No correlograma exposto na Figura 4, podemos verificar que a variável predita, VDPEQ, possui correlações significativas, classificadas como fracas, com as variáveis relacionadas à distribuição de alunos por docentes e funcionários, custo de folha de pagamento docente e *ticket* médio (respectivamente VRAD, VRAF, VCFPA e VTMA). Já a relação entre as variáveis preditoras mostra algumas relações de importância, tais como a relação entre *ticket* médio do aluno e custos de folha de pagamento (relação correlacional forte entre a variável de receita VTMA e as variáveis de custos VCFPA e VCODA). Outra relação importante é entre a exposição da mantenedora à modalidade EAD e a quantidade de alunos na relação com os docentes, já classificada como de moderada força. Esta relação, moderada positiva, mostra que, quanto mais exposição ao EAD, maior a quantidade de alunos em relação aos docentes, o que será abordado na seção dedicada à discussão dos resultados.

Já correlações mais fracas são encontradas na variável que mede a interiorização da oferta (VPOL), principalmente em relação ao EAD (VEAD) e aos docentes e funcionários (VRAD e VRAF).

#### 4.6 Análise de correlação – 2019

O estudo de correlação na base de dados para o ano censo 2019, notadamente, apresenta variações importantes em relação às correlações encontradas comparativamente no ano censo 2018, conforme apresentado na Figura 5, na qual se apresenta a relação das variáveis preditas e preditoras, permitindo assim a interpretação da relação entre elas.

Figura 5 - Correlograma das variáveis - 2019



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Novamente, é possível verificar que a variável predita, VDPEQ, só possui correlações significativas, classificadas como fracas, com a variável relacionada ao *ticket* médio (VTMA). Já a relação entre as variáveis predictoras mostra algumas relações de importância, tais como aumento na força da relação entre *ticket* médio do aluno e custos de folha de pagamento (relação correlacional forte entre a variável de receita VTMA e as variáveis de custos VCFPA e VCODA). Outro aumento de força de relação verifica-se entre a exposição da mantenedora à modalidade EAD e a

quantidade de alunos na relação com os docentes, já classificada anteriormente como de moderada força, mas que apresenta ganho numérico, beirando o patamar de correlação forte. Essa relação, moderada positiva, mostra que, quanto mais exposição ao EAD, maior a quantidade de alunos em relação aos docentes, o que será discutido na seção dedicada à discussão dos resultados.

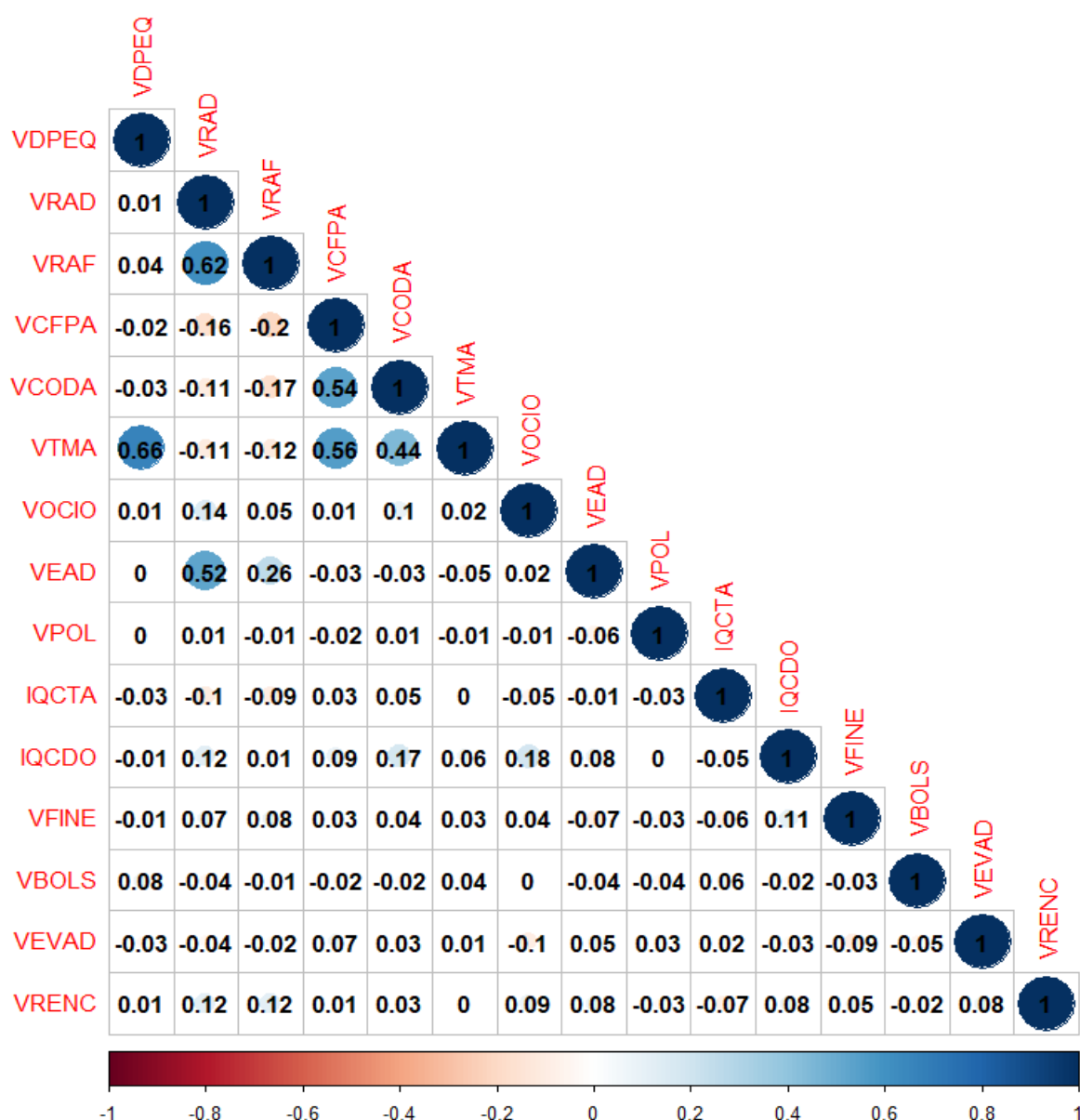
Destaca-se, ainda, como de fraca correlação, a qualificação do corpo docente (IQCDO) em relação às variáveis de custos de folha e *ticket* médio (VCFPA e VTMA), mostrando ainda correlações mais fracas encontradas na variável que mede a interiorização da oferta (VPOL), principalmente em relação ao EAD (VEAD), e relações docentes e funcionários (VRAD e VRAF).

#### 4.7 Análise de correlação – 2020

Encerrando a análise detalhada das correlações e utilizando o ano censo 2020, apresentamos o correlograma na Figura 6, que ilustra a relação não só da variável predita com as preditoras, como também entre todas as variáveis. Percebe-se o aumento da força – de fraca para forte correlação – entre a variável que mede o *ticket* médio do aluno com a variável predita (VDPEQ), algo numericamente três vezes maior em relação ao ano anterior.



Figura 6 - Correlograma das variáveis - 2020



Fonte: Elaborado pelo autor com base nos microdados (INEP, 2022)

Percebe-se no gráfico que a interação entre as variáveis tem significâncias, destacando a variável VTMA, já ressaltada, com VCFPA e VCODA, além do destaque da interação das variáveis VEAD (Proporção de alunos na modalidade EAD) com os indicadores administrativos VRAD (Número de Alunos / Docentes) e VRAF (Número de Alunos por Funcionários), mostrando um caminho a avaliar nos estudos de regressão e nos modelos gerados.

#### 4.8 Análise de Componentes Principais (PCA)

O recurso de Análise de Componentes Principais (*PCA*) será visto nesta subseção como análise adicional na classificação das variáveis no modelo de regressão, estudando-se individualmente a importância que cada uma tem no resultado gerado.

Conforme referenciado na subseção 3.5.2 Análise de Componentes Principais, a análise utiliza uma técnica em que agrupa as variáveis independentes e gera regressões entre elas (cada grupo de variáveis é um componente) e a variável dependente. Cada grupo recebe um percentual de explicação no modelo e, com a junção das variáveis dentre os grupos mais importantes, temos quais grupos de variáveis realmente impactam o modelo de regressão, permitindo assim avaliar e reduzir o modelo.

Primeiramente, apresentamos a análise para o ano censo 2018, utilizando a matriz de variáveis independentes sem padronização, cujo resultado é mostrado na Tabela 14.

**Tabela 4 - Análise dos componentes principais - 2018**

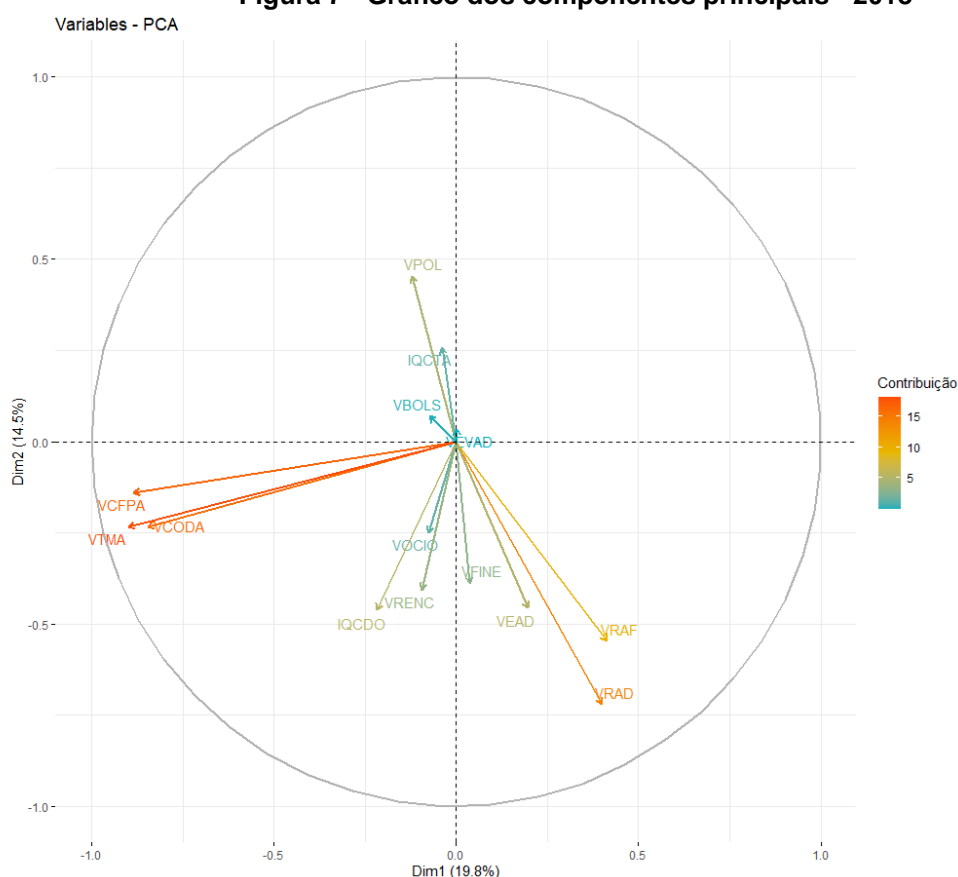
	Grupos de Componentes Avaliados						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7 a PC 14
Proporção de Variância Explicada	19,8%	14,5%	8,8%	8,0%	7,3%	7,1%	34,6%
Proporção Acumulada	19,8%	34,2%	43,0%	51,0%	58,3%	65,4%	100,0%

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Ao analisar a Tabela 14, verifica-se que os dois componentes principais (PC1 e PC2) representam sozinhos 34,2% da variação explicativa do modelo, em que cada componente representa sozinho a correlação entre múltiplas variáveis do modelo. Dos componentes PC3 até o PC14, temos o restante da variabilidade explicada, reduzindo a importância gradativamente.

Já em uma visão gráfica da distribuição dos componentes calculados, em termos de importância, temos o apresentado na Figura 7.

**Figura 7 - Gráfico dos componentes principais - 2018**



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

O gráfico bidimensional representado na Figura 7 mostra que para os dois primeiros grupos de componentes principais calculados, referentes a 34,2% da explicação da variabilidade, temos como variáveis de maior contribuição o VTMA (*Ticket Médio do Aluno*), VCFPA (*Custos de Folha Docente*) e VCODA (*Custos Operacionais*). Também importantes no destaque as relações alunos e docentes ou funcionários (VRAD e VRAF).

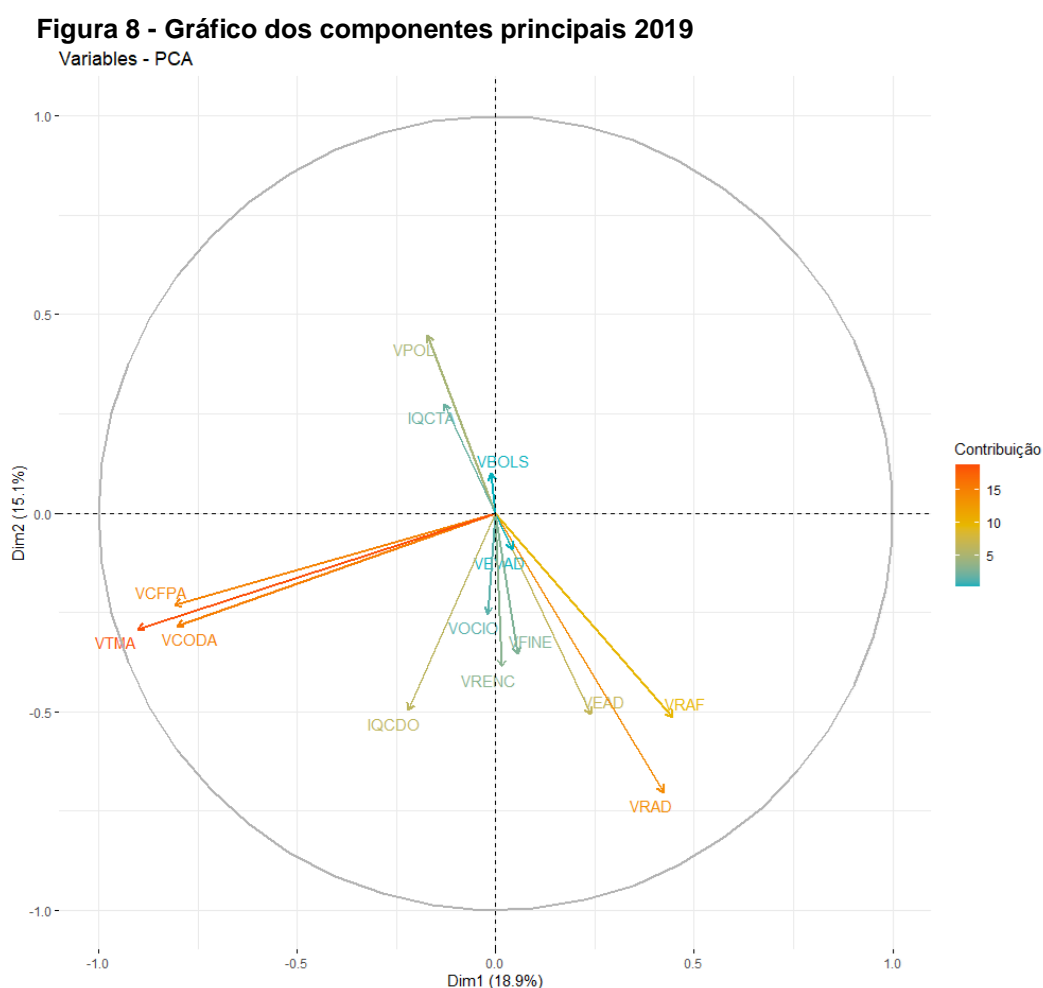
Em análise feita com a base de dados criada para o ano de 2019, temos, na Tabela 15, uma distribuição similar, em que os dois primeiros componentes principais representam 34% da explicação do modelo.

**Tabela 15 - Análise dos componentes principais - 2019**

	Grupos de Componentes Avaliados						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7 a PC 14
Proporção de Variância Explicada	18,9%	15,1%	8,6%	8,0%	7,5%	7,1%	34,8%
Proporção Acumulada	18,9%	34,0%	42,6%	50,6%	58,1%	65,2%	100,0%

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

No gráfico bidimensional (Figura 8), representando os componentes dos dois primeiros grupos, temos 34% da variabilidade, com a representação das variáveis de destaque.



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

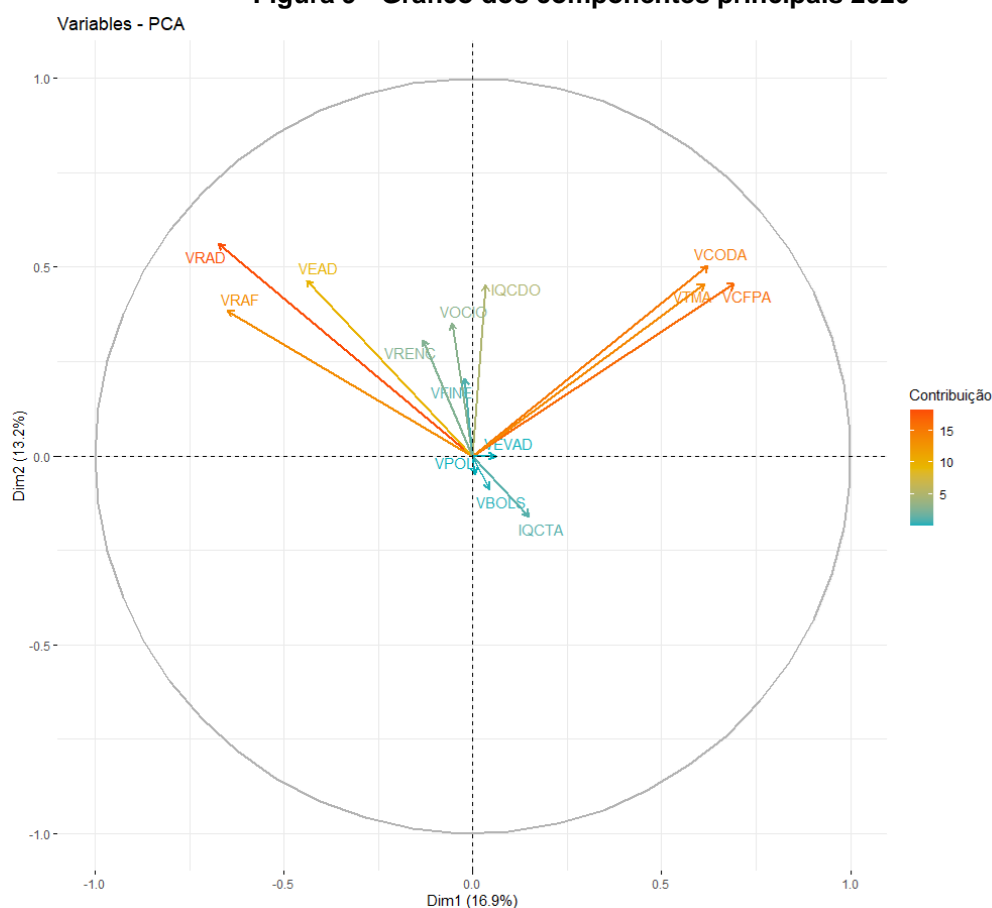
Já com os dados 2020 e as variáveis calculadas, temos uma concentração ainda menor de representatividade de componentes principais, sendo maior a distribuição representada na Tabela 16.

**Tabela 16 - Análise de componentes principais - 2020**

	Grupos de Componentes Avaliados						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7 a PC 14
Proporção de Variância Explicada	16,86%	13,20%	8,88%	7,95%	7,31%	7,25%	38,6%
Proporção Acumulada	16,9%	30,1%	38,9%	46,9%	54,2%	61,4%	100,0%

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

A distribuição da contribuição dos componentes principais no ano 2020, de acordo com a Tabela 16, é visivelmente maior, com menor agrupamento nos primeiros componentes principais. Já na visão bidimensional dos dois primeiros grupos de componentes, responsáveis por 30,1% da variabilidade correlacional explicada, temos sua representação na Figura 9.

**Figura 9 - Gráfico dos componentes principais 2020**

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

No gráfico bidimensional da Figura 10, apresenta-se a contínua importância das variáveis VTMA (*Ticket* Médio do Aluno), VCFPA (Custos de Folha Docente) e

VCODA (Custos Operacionais), mas com igual destaque prático das variáveis que tratam as relações alunos e docentes ou funcionários (VRAD e VRAF) e, na primeira aparição nos modelos, a exposição à modalidade de ensino a distância (VEAD).

#### 4.9 Regressão linear múltipla da base de dados

Entre os objetivos específicos deste trabalho está a busca da relação e peso entre a variável dependente – distância do ponto de equilíbrio (VDPEQ) – e as demais variáveis independentes modeladas e testadas anteriormente, tanto em distância quanto em correlação e multicolinearidade. Para atingimento desses objetivos específicos, será aplicado um modelo de Regressão Linear Múltipla, conforme escolha estabelecida anteriormente.

Para cada ano censo será criado um modelo de regressão e cada um dos parâmetros principais dos modelos será avaliado separadamente e depois conjuntamente, construindo assim um dos argumentos de análise e resposta aos objetivos específicos.

Na Tabela 17, é apresentado o resultado das regressões, variável a variável, para cada ano censo.

**Tabela 17 - Resultado das regressões por variável e ano**

Variável	Ano Censo					
	2018		2019		2020	
	Coeficiente	Beta P.	Coeficiente	Beta P.	Coeficiente	Beta P.
Intercepto	0,3318	NA	0,1881	NA	0,0973	NA
VRAD	0,0038	0,0869	0,0045	0,0535	-0,0025	-0,0082
VRAF	0,0031	<b>0,1132</b>	0,0038	0,0674	0,0039	0,0196
VCFPA	0,0000	<b>-0,6539</b>	0,0000	<b>-0,8857</b>	-0,0001	<b>-0,4563</b>
VCODA	0,0000	<b>-0,5394</b>	0,0000	<b>-0,9449</b>	-0,0001	<b>-0,2328</b>
VTMA	0,0000	<b>1,1375</b>	0,0000	<b>1,7069</b>	0,0001	<b>1,0211</b>
VOCIO	0,0442	0,0185	-0,1618	-0,0326	0,1235	0,0055
VEAD	-0,3773	-0,0513	-0,3471	-0,0330	0,8463	0,0286
VPOL	0,0033	0,0017	0,0695	0,0181	0,1511	0,0088
IQCTA	-0,0374	-0,0252	0,1084	0,0372	-0,0999	-0,0075
IQCDO	-0,0485	-0,0380	-0,0397	-0,0154	0,0177	0,0016
VFINE	0,0925	0,0495	-0,0336	-0,0088	-0,2375	-0,0146
VBOLS	-0,2942	-0,0464	-0,1724	-0,0153	1,1704	0,0228
VEVAD	-0,3089	-0,0728	-0,7613	-0,0868	-0,1136	-0,0033
VRENC	-0,1481	-0,0536	-0,1206	-0,0220	0,3772	0,0181

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Na Tabela 17, percebe -se, nos valores destacados em negrito, as variáveis que no ano respectivo tiveram maior importância no resultado da regressão, pelo *Beta* padronizado. Para o ano 2018, as variáveis VRAF, VCFPA, VCODA e VTMA são as mais destacadas. Com este indicativo, temos, portanto, que as variáveis que representam a relação alunos e docentes, custos de folha de pagamento docente e operacional e o *ticket* médio são identificadas como fatores relevantes para o modelo regredido no ano censo 2018. As demais, apesar de notadamente importantes para o modelo, ficaram em menor destaque na predição.

Similarmente, no ano 2019, as variáveis de custos de folha de pagamento (VCFPA e VCODA) e de *ticket* médio (VTMA) mantiveram o destaque em significância. Aparecem com destaque as relações entre docentes e alunos (VRAD) e aluno por funcionários (VRAF). Cabe destacar, no cenário de variáveis, a significância estatística do indicador de evasão estudantil (VEVAD).

Já a regressão do modelo para o ano censo 2020 trouxe como descoberta de significância estatística somente as variáveis relacionadas com os custos (VCFPA e VCODA) e com o *ticket* médio dos alunos (VTMA).

Com esse resultado, pode-se verificar que as variáveis com maior significância estatística e que explicam a variabilidade da predita são as que representam os custos de folha de pagamento docente, custo operacional (VCFPA e VCODA) e o *ticket* médio dos alunos (VTMA) em todos os cenários de bases de dados.

Na Tabela 18, é identificado o nível de predição do modelo obtido, sendo detalhada também a avaliação de resíduos.

**Tabela 18 - Resultados da regressão**

Medida	Ano Censo		
	2018	2019	2020
	Resultado	Resultado	Resultado
R <sup>2</sup>	0,3604	0,4490	0,6935
Resíduos	0,4648	0,8898	2,7627

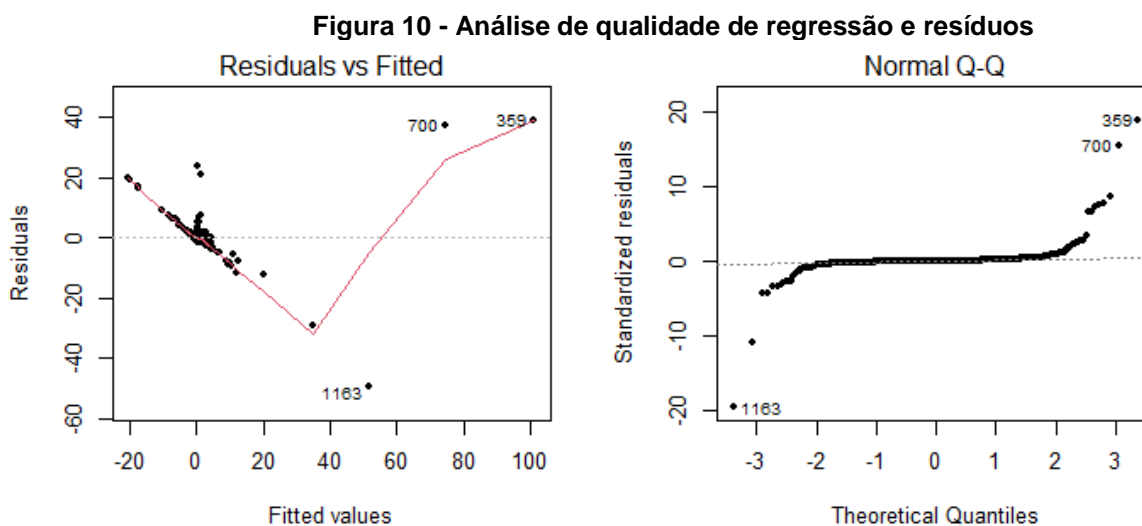
Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Conforme descrito por Fávero (2015) e Garson (2016), o fator de predição é de 36,04%, representando que este percentual dos registros da variável dependente

em 2018 pode ser explicado pelo modelo criado, resultando em uma classificação de predição moderada.

Já em 2019, temos 44,9% – ainda moderada – de adequação e em 2020 uma forte relação preditiva, com 69,35% das variações na predita sendo explicadas pelas variáveis predictoras.

Um gráfico de resíduos e normalidade é exibido na Figura 10, relativo ao ano censo 2020, mais recente.



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

O que se pode verificar nos gráficos da Figura 10 é que existe uma distribuição dos resíduos diante da curva da regressão gravitando em torno do valor zero. A distribuição Q-Q, para análise de normalidade dos resíduos, também mantém a maior parte dos registros sobre a linha pontilhada no eixo zero, embora os testes de normalidade de distribuição dos resíduos apontem para a não normalidade, com o teste de normalidade Shapiro-Wilk tendo resultado inferior a 0,05.



#### 4.10 Análise por florestas aleatórias (*random forest*)

Esta subseção é dedicada a avaliar, em nossa base de dados do censo, para cada ano, o comportamento das variáveis quando executamos os algoritmos de previsão por Florestas Aleatórias (*Random Forest*), que, conforme descrito por Porfirio (2021), nos permite alcançar precisão na predição dos comportamentos das variáveis dependentes em situações com múltiplas variáveis preditoras ou independentes.

Algumas considerações que devemos levar em conta na modelagem e configuração da floresta aleatória proposta são necessárias para completo entendimento do modelo. A primeira é a quantidade de árvores geradas para o processo preditivo.

Conforme Kahn *et al.* (2020), a quantidade de árvores deve ser suficientemente equilibrada para avaliar todas as variáveis e cenários, sem comprometer o custo computacional, apresentando assim um resultado preditivo eficiente. Para este processo de regressão, iremos considerar o valor padrão do algoritmo,  $n_{tree} = 500$ , suficiente para rodar as variações das bases de dados com custo computacional baixo.

A outra variável importante é a classificação da importância da variável e neste caso usaremos o método de menor perda por erro médio absoluto ( $loss = mae - mean\ absolute\ error$ ). Com este parâmetro, conseguimos reduzir a distância entre os agrupamentos e assim classificar as variáveis pela importância preditiva.

Portanto, para a geração da floresta, foram usados como parâmetros informados ao algoritmo de classificação:

- a quantidade de regressões e árvores criadas igual a 500, número suficiente para cada registro passar pelo classificador;
- o uso do algoritmo *K-means* (proximidade);
- análise de importância das variáveis pela menor perda, feita com o método *mae*, erro médio absoluto.

Com esses parâmetros aplicados para a base de dados de cada ano censo, temos o resultado geral mostrado na Tabela 19.

Tabela 14 - Resultado da análise por Florestas Aleatórias

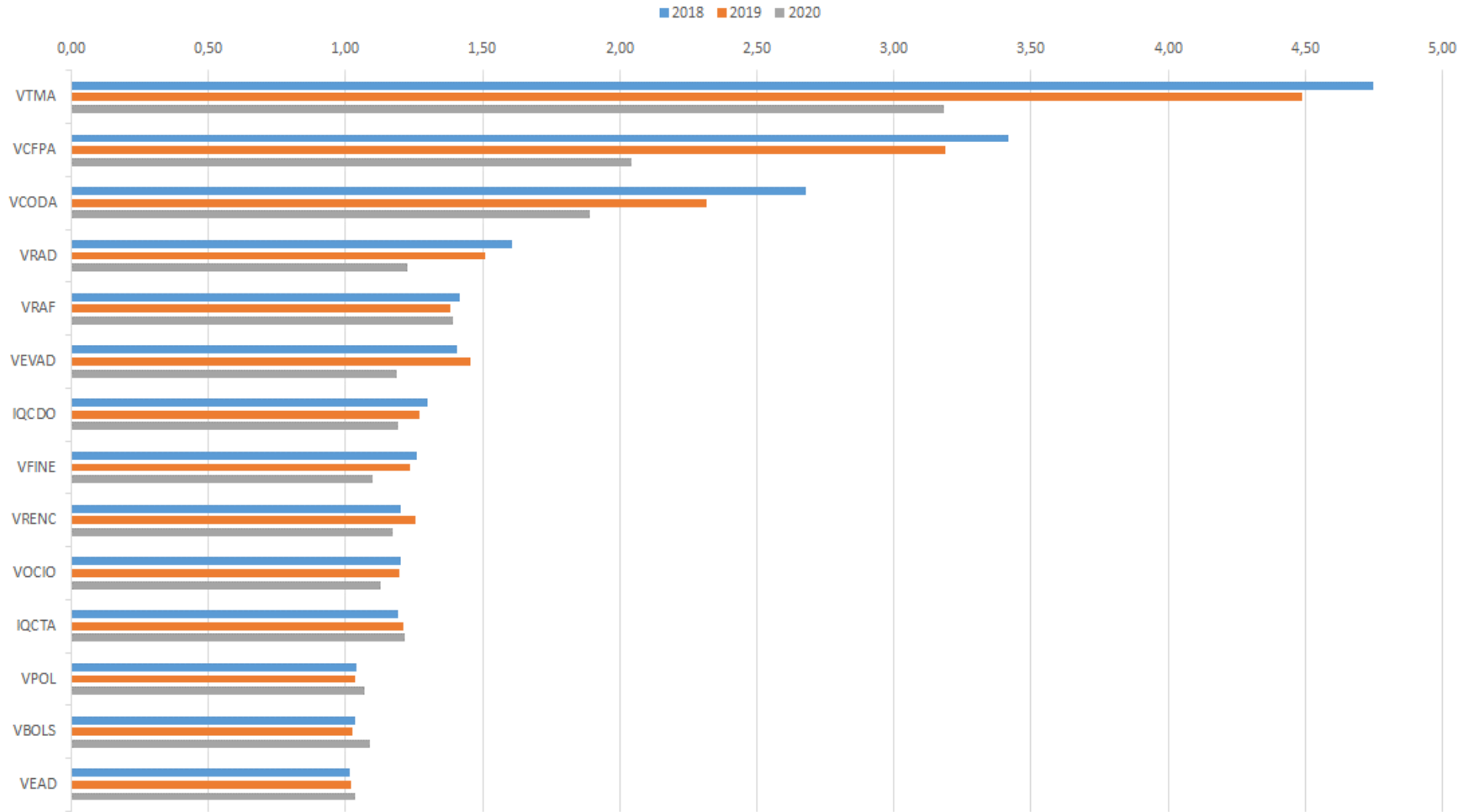
Variável	Ano Censo		
	2018	2019	2020
	Importância	Importância	Importância
VTMA	4,75	4,49	3,18
VCFPA	3,42	3,19	2,04
VCODA	2,68	2,32	1,89
VRAD	1,61	1,51	1,23
VRAF	1,42	1,38	1,39
VEVAD	1,41	1,46	1,19
IQCDO	1,30	1,27	1,19
VFINE	1,26	1,23	1,10
VRENC	1,20	1,26	1,17
VOCIO	1,20	1,20	1,13
IQCTA	1,19	1,21	1,22
VPOL	1,04	1,04	1,07
VBOLS	1,04	1,03	1,09
VEAD	1,02	1,02	1,03

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Podemos verificar que os valores de importância atribuídos pelos algoritmos de florestas aleatórias mostram, novamente, no topo as variáveis VTMA (*Ticket Médio do Aluno*), seguido por variáveis de custos docentes e operacionais (VCFPA e VCODA). Destaquem-se também as razões docente e funcionários por aluno (VRAD e VRAF).

Podemos verificar o desempenho detalhado da classificação, ano a ano, no gráfico apresentado na Figura 11.

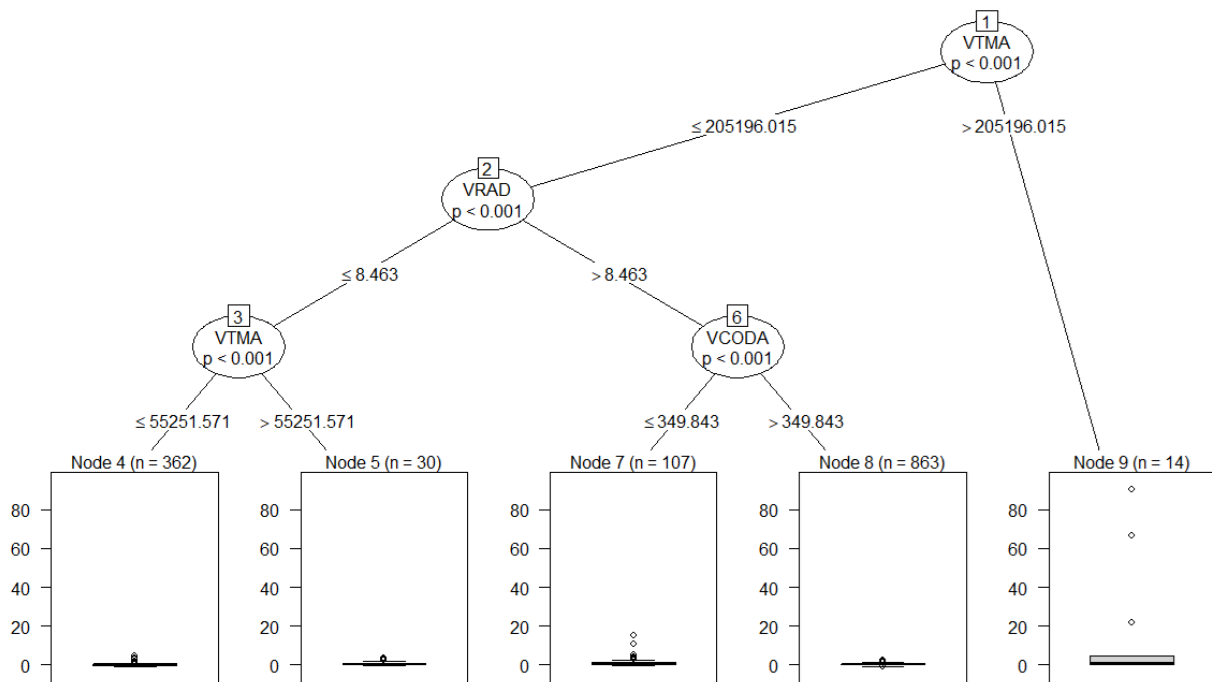
Figura 11 - Importância das variáveis independentes por ano censo - RF



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Já uma visão de uma das árvores de decisão, para o ano censo 2020, é mostrada na Figura 12, com descritivo posterior.

Figura 12 - Árvore de Decisão



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

A árvore apresentada na Figura 12 foi criada para apresentação visual dos três primeiros nós, o que facilita a visualização gráfica e mostra como a decisão do caminho preditivo para cada nó (variável) foi tomada.

Iniciou-se no topo (1) com a análise do *Ticket Médio do Aluno* (VTMA) realizando uma quebra que representa a primeira interação. Nesta decisão, já se caminhou para a relação aluno/docente (VRAD) no segundo nível (2) e seguiu-se criando vários nós conforme análise das variáveis e classificação de importância. Na base, temos a distribuição por nós, não exibidos, em que se pôde chegar até o detalhe da folha (registro), mostrando assim o caminho completo da predição.

## 5 DISCUSSÃO DOS RESULTADOS

Nesta seção, serão analisadas e discutidas as descobertas apresentadas na seção 4, detalhando a importância para o objetivo deste trabalho.

Após a análise do modelo de regressão obtido e a avaliação dos modelos, o objetivo, nesta seção, é aprofundar a análise das relações encontradas e discutir os resultados obtidos a partir dos referenciais teóricos abordados.

Um resumo das variáveis conseguidas, suas justificativas teóricas e a relação entre elas é apresentada na primeira subseção, permitindo, assim, o aprofundamento na análise e posterior discussão e comparação dos resultados entre os modelos de regressão utilizados.

Na segunda subseção, apresenta-se uma análise mais aprofundada das variáveis identificadas na base de dados, em diferentes estratos de portes das mantenedoras, para determinar se o resultado das variáveis consideradas principais varia de acordo com esta classificação.

Os resultados encontrados da distribuição quantitativa das matrículas entre as mantenedoras educacionais são apresentados na terceira subseção, o que, além de permitir a classificação em estratos de porte e modalidades de ensino, detalha a análise e especializa os resultados deste modelo de amostras.

O resultado do impacto da modalidade ensino a distância em três variáveis preditoras – Custo Docente, *Ticket* Médio e Ponto de Equilíbrio – é apresentado na subseção 5.3, na qual temos constatações importantes sobre a aplicação do modelo e a busca da redução de custos, aplicando esforços na tendência da interiorização da oferta.

Na subseção 5.4, encerramos a análise trazendo as informações de concentração de mercado e a relação com a bibliografia citada que indica a concentração mercadológica, influenciando o resultado dos fatores que impactam nos resultados gerais das mantenedoras, encerrando na subseção 5.5, com a análise das mantenedoras que ficaram negativamente distantes do ponto de equilíbrio e comparação com a média.

### 5.1 Visão geral do processo de geração das bases de dados

Foi realizada a importação e geração das bases de dados brutas, com as informações e dados sem nenhum tipo de tratamento, dos anos censo 2018, 2019 e 2020 utilizando a ferramenta R (R CORE TEAM, 2022) e pacotes de importação, com os quais chegamos às seguintes constatações sobre a base bruta, já delimitando no estudo as instituições mantenedoras de ensino classificadas na categoria privada, conforme Tabela 20. Foi adicionado o ano censo 2021, não objeto deste estudo, publicado após as extrações na sala segura do Inep (2022) para efeito comparativo e maior abrangência da análise.

**Tabela 15 - Comparativo de quantidade de mantenedoras**

<b>Ano</b>	<b>Mantenedoras</b>	<b>Variação</b>	
<b>2018</b>	1558		
<b>2019</b>	1562	0,3%	-11,7%
<b>2020</b>	1469	-6,0%	
<b>2021</b>	1376	-6,3%	

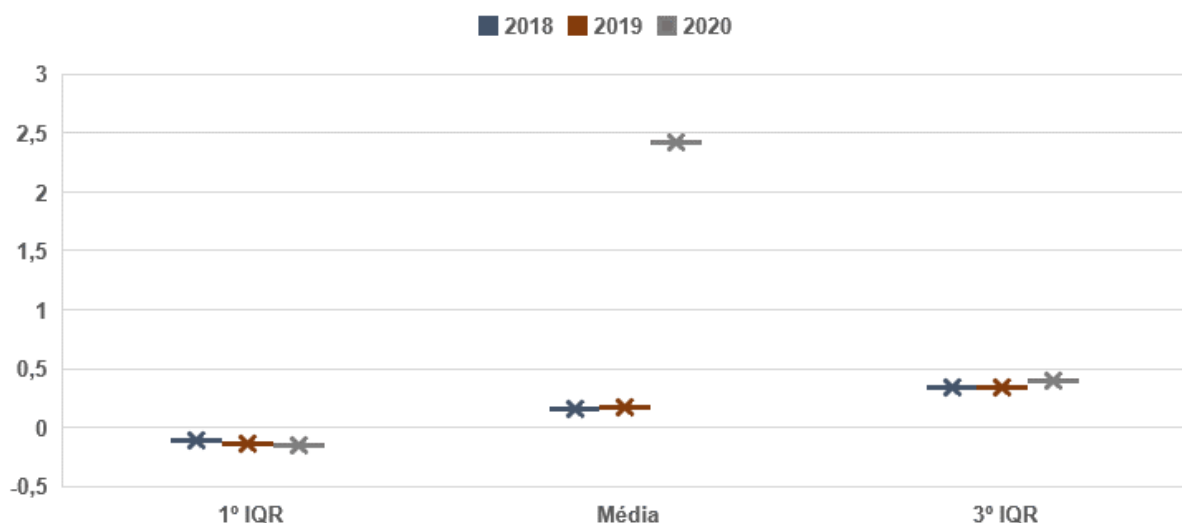
Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Uma das primeiras constatações é a redução quantitativa, a partir da base estabelecida em 2018, no número de mantenedoras privadas. Apesar de um ligeiro crescimento em 2019 (0,3%), houve reduções consecutivas no quantitativo de mantenedoras respectivamente de 6,0% e 6,3% entre os anos de 2019-2020 e 2020-2021, totalizando uma redução total na quantidade de mantenedoras privadas no mercado brasileiro de 11,7% entre os anos 2018 e 2021.

Pode-se, em estudo futuro, avaliar os motivos desta redução, comparando-a com movimentos de fusões e aquisições, encerramento e transferências de manutenções e/ou encerramento de atividades educacionais.

Quanto às variáveis calculadas, podemos ilustrar a distribuição no gráfico da Figura 13, em relação à variável calculada VDPEQ, que verifica a distância do ponto de equilíbrio das mantenedoras ao longo dos anos censo.

Figura 13 - Comparativo da Variável Ponto de Equilíbrio



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Percebe-se pela distribuição gráfica na Figura 13, que existe estabilidade ao longo dos anos censo dos valores calculados para a distância do ponto de equilíbrio, mesmo sem realizar o tratamento de *outliers*. Em relação ao primeiro intervalo interquartil (1º IQR), a variável VDPEQ mostra que, ao longo dos anos 2018, 2019 e 2020, manteve-se próxima, na faixa negativa em torno de 13%. Já a média traz uma variação grande, com potenciais *outliers* para o ano 2020, tendo, em 2018 e 2019, se mantido em torno de 16% na escala positiva. O último intervalo (3º IQR) também mostra estabilidade, em torno de 36% na escala positiva. Após a análise dos *outliers*, serão retomados estes comparativos de distribuição.

Na redução de dados extremos ou inválidos – *outliers* –, o uso do método de Mahalanobis (TABACHNICK; FIDELL, 2019) e do intervalo de confiança  $p < 0,001$  apresentou bons resultados, com a remoção de 6,1% em média dos registros das bases de dados, sem redução amostral significativa que comprometesse os resultados das regressões, preocupação para criação dos modelos. A amostra final gerada é apresentada na Tabela 21.

Tabela 16 - Quantidade de mantenedoras pós-outliers

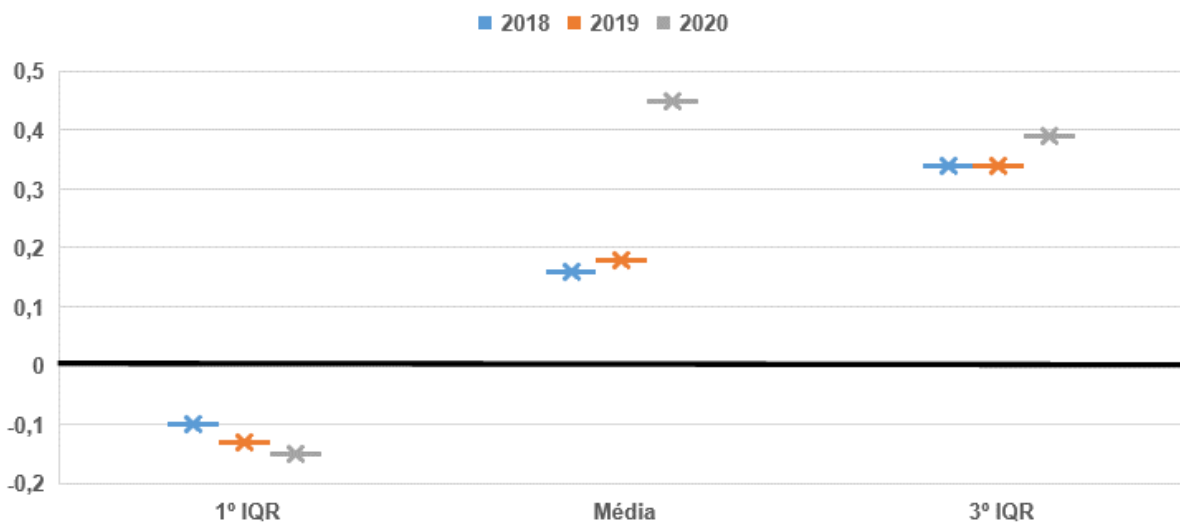
Ano	Mantenedoras	Varição da base total	Média da Redução
2018	1450	- 6,9%	
2019	1482	- 5,1%	-6,1%
2020	1376	- 6,3%	

Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Com esse tratamento, as bases de dados que foram trabalhadas para atingir os objetivos deste trabalho tiveram, para os anos censo 2018, 2019 e 2020 respectivamente 1.450, 1.482 e 1.376 registros, conforme mostrado na Tabela 21.

Na Figura 14, a seguir, ilustram-se os efeitos da redução de pontos extremos na variável calculada VDPEQ, distância do ponto de equilíbrio:

Figura 14 - Comparativo de ponto de equilíbrio após remoção de outliers



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

A remoção dos *outliers* modificou o cenário anteriormente apresentado. Enquanto os valores do primeiro e terceiro intervalo interquartil se mantiveram estáveis entre os anos censo, as posições médias do intervalo aumentaram, o que significa que o modelo de dados tem maior quantidade de registros dentro do intervalo interquartil, cobrindo assim mais de 50% dos dados.

Já na média, notou-se primeiro que o valor muito acima em 2020 foi relativamente reduzido, de aproximadamente 240% (2,4) para 45% (0,45), aproximando-o dos anos anteriores.



Com a operação de importação e tratamento da base de dados, temos, portanto, registros suficientes para a geração dos modelos de análise estatística e descobertas, detalhados na próxima subseção, com dados mais homogêneos, sem riscos de introduzir no modelo dados faltantes ou não coerentes em seu preenchimento original.

## 5.2 Resumo das descobertas dos métodos de análise utilizados

Na seção 4, conseguimos extrair, utilizando os três métodos estatísticos indicados – Regressão Linear Múltipla, Componentes Principais e Floresta Aleatória –, quais as variáveis mais importantes no relacionamento entre elas e a variação da distância de atingimento do ponto de equilíbrio.

Cabe agora fazermos uma comparação entre os três métodos e o resultado obtido na importância das variáveis calculadas, ano a ano, como apresentado na Tabela 22.

**Tabela 17 - Resumo das descobertas, métodos e variáveis**

Variável	Ano Censo					
	2018		2019		2020	
	RLM	RF	RLM	RF	RLM	RF
VRAD	***	**	***	**	*	**
VRAF	***	**	***	**	*	**
VCFPA	***	***	***	***	***	***
VCODA	***	***	***	***	***	**
VTMA	***	***	***	***	***	***
VOCIO	*	*	*	*	*	*
VEAD	***	*	*	*	*	*
VPOL	*	*	*	*	*	*
IQCTA	*	*	*	*	*	*
IQCDO	*	*	*	*	*	*
VFINE	***	*	*	*	*	*
VBOLS	***	*	*	*	*	*
VEVAD	***	*	***	*	*	*
VRENC	***	*	*	*	*	*

Nota: \*\*\* Alta relevância \*\* Média Relevância \* Baixa ou sem Relevância

Fonte: Elaborado pelo autor, (2023)

Na Tabela 22 consideramos, para cada ano censo, os dois métodos utilizados para avaliação de importância das variáveis (RLM – Regressão Linear Múltipla e RF - Floresta Aleatória). Em cada um dos métodos, temos uma escala de importância (\* = Sem ou baixa importância, \*\* = Média importância e \*\*\* = Alta importância) e, em cada variável, modelo de regressão e ano censo foi inserida a classificação correspondente.

Com essa avaliação, o cenário mostra que, independentemente do ano censo e do método de análise utilizado, as variáveis *Ticket* Médio do Aluno (VTMA) e Custo Folha Docente (VCFPA) são de alta importância nos modelos. Seguem-se em importância as variáveis Custo Operacional (VCODA), as razões docentes e alunos (VRAD) e funcionários administrativos e alunos (VRAF). Nas demais variáveis, existe variação ano a ano ou divergência entre os métodos de avaliação.

A avaliação por Componentes Principais (PCA) não demonstrou fatores de variáveis que tivessem dominância sobre a explicação da variável dependente, distribuindo-se ao longo dos fatores.

Esses achados, em relação ao referencial teórico, são corroborados principalmente por Soares (2016), Silva (2017), Zamin (2019), que identificam os fatores que influenciam o alcance do ponto de equilíbrio, como número de alunos adequado aos custos, receita com mensalidades e *ticket* médio além de despesas com a mão de obra. Silva (2017, p. 81) ainda ressalta que, além desses fatores, o controle de gestão dos resultados em gastos operacionais, inovação e flexibilidade também pode impactar os fatores da relação adequada de funcionários e docentes ao número de alunos.

Portanto, neste ponto, os cenários gerais de fatores que influenciam o distanciamento do ponto de equilíbrio estão corroborados tanto pelo suporte teórico quanto pelas descobertas na análise estatística feita neste trabalho.

Entende-se que nosso objetivo específico, fatores que influenciaram o distanciamento do ponto de equilíbrio bem como os pesos atribuídos a eles, foi satisfatoriamente atingido. Na próxima subseção, será apresentada uma especialização do estudo, comparando se os fatores variam de acordo com o estrato de porte ao qual as mantenedoras pertencem.

### 5.3 Segmentação por porte e análise das variáveis

Conforme adotado como padrão para classificação de mantenedoras educacionais pelo porte, nesta subseção utilizaremos a estrutura proposta pelo Semesp (2021, p. 12) e exibida anteriormente no Quadro 7 para avaliar o quanto a estratificação impacta nas variáveis determinantes do nosso modelo.

Buscando respaldo para esta análise na literatura já citada na fundamentação teórica, usaremos as afirmações de Sguissardi (2015) sobre a mercantilização e a concentração do mercado no ensino superior – já mostradas nos procedimentos metodológicos, na Figura 3, em que as instituições consideradas gigantes (GG), apesar de somarem 3% das mantenedoras brasileiras, detêm 68% do total de alunos matriculados. A pergunta que pretendemos responder agora, analisando as variáveis predictoras, é se o porte da mantenedora tem influência no comportamento da distância do Ponto de Equilíbrio (para cima ou para baixo).

Conforme a Tabela 23, somente para o ano censo 2020, temos os registros da regressão linear multivariada segmentados nos estratos propostos pelo Semesp (2021, p. 12).

**Tabela 18 - Análise de regressão por estrato de porte**

Ano Censo 2020				
Porte	P	M	G	GG
R <sup>2</sup>	0,716	0,210	0,675	0,940
Variável	Beta P.	Beta P.	Beta P.	Beta P.
VRAD	-0,0162	<b>0,2808</b>	0,2819	-0,0439
VRAF	0,0292	-0,0496	0,2977	0,0377
VCFPA	<b>-0,4708</b>	-0,0695	<b>-0,8720</b>	<b>-1,2178</b>
VCODA	-0,2394	<b>-0,2493</b>	<b>-0,7302</b>	<b>-1,1671</b>
VTMA	<b>1,0424</b>	<b>0,2121</b>	<b>1,0321</b>	<b>1,9878</b>
VOCIO	0,0094	-0,0711	-0,1214	-0,0124
VEAD	0,0020	<b>0,2549</b>	-0,2598	-0,0890
VPOL	0,0125	-0,0079	-0,1428	-0,0201
IQCTA	-0,0069	-0,0013	-0,0948	-0,0408
IQCDO	0,0103	-0,0095	0,0183	0,0052
VFINE	-0,0128	0,0586	0,1587	0,0430
VBOLS	0,0224	0,0055	0,3329	-0,0257
VEVAD	0,0013	-0,0701	-0,1369	0,0355
VRENC	0,0212	-0,0167	-0,1339	-0,0070

Nota: valores *Beta P.* em destaque são identificados como relevantes.

Fonte: Elaborado pelo autor, (2023)

Novamente, vemos que, independentemente dos estratos para os quais foram realizadas as análises, as variáveis correspondentes ao *Ticket* Médio do Aluno (VTMA) e Custo Folha Docente (VCFPA) são de alta importância em todos os estratos, exceto onde o porte é “M”, correspondendo na amostra a 145 mantenedoras, 10,5% do total. Para este estrato, a significância do resultado é baixa na comparação com os outros estratos, sem destaque para nenhuma das variáveis, tornando o resultado desta amostra inconclusivo.

Em linhas gerais, nos demais estratos, houve aumento significativo da relevância estatística da significância, redução nos resíduos e aparente aumento preditivo.

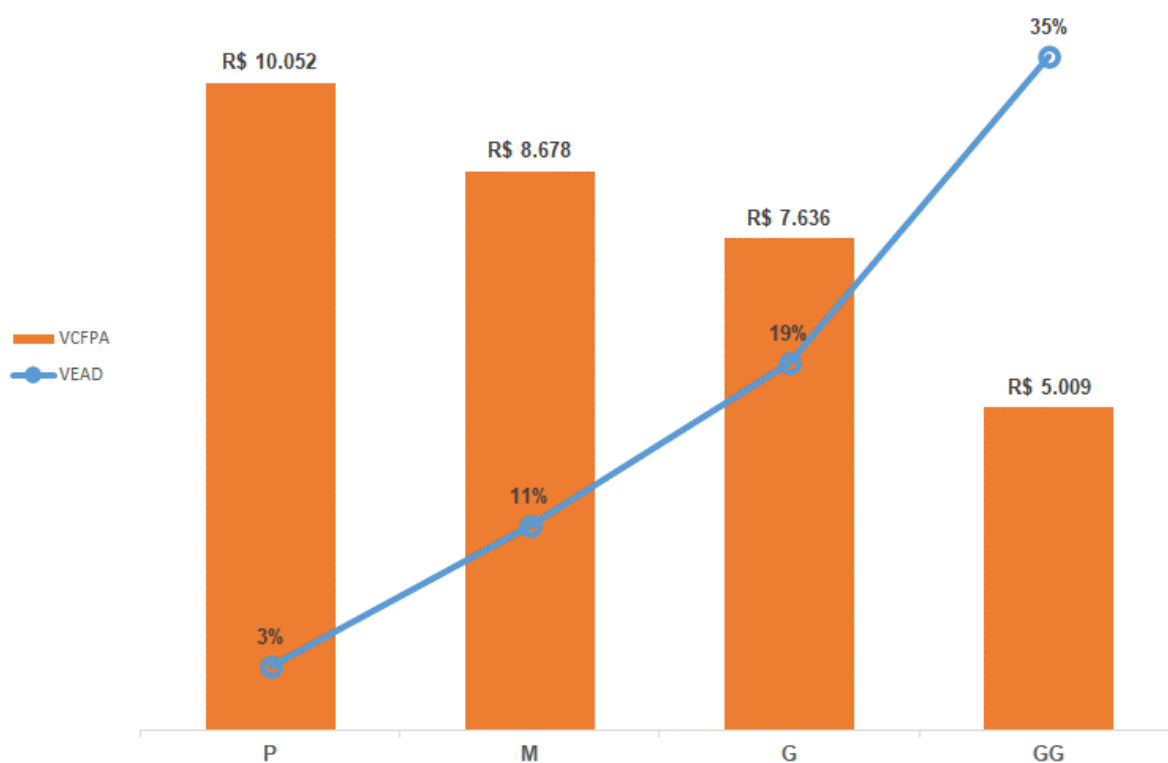
#### 5.4 Impacto do ensino a distância no custo e na receita

A variável VCFPA (Custo de Folha Docente por Aluno), em todos os modelos, sempre teve significância alta, mostrando que a busca pelo menor custo docente e operacional, vinculado ao *ticket* médio do aluno adequado, leva as instituições de ensino superior a manterem um distanciamento positivo do ponto de equilíbrio.

Embora não tenham aparecido como relevantes nas análises de variáveis importantes, temos registrado em Barreyro (2008), Sguissardi (2015) e Diniz e Goergen (2019) que a busca pelo lucro na prestação de serviço educacional passa pela criação de novas ofertas, interiorização e expansão do EAD.

Com o intuito de verificar de forma empírica se existe relação entre o custo docente e a aplicação da modalidade EAD como ferramenta de gestão e expansão, conseguimos avaliar, para o ano censo 2020, uma distribuição constante no gráfico da Figura 15.

**Figura 15 - Comparativo EAD x Custo Docente/Aluno**



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Percebe-se, na Figura 15, que, de acordo com o estrato ao qual pertence a mantenedora, a relação de Custo de Folha Docente por Aluno (VCFPA) e a exposição ao EAD (VEAD) é inversa em sua curva. A linha em azul representa, para cada estrato, o percentual de exposição ao EAD no grupo. Em barras de cor laranja, temos o custo docente médio para o estrato, em valores monetários anualizados.

Em instituições de porte “P”, entende-se que, em média, o custo da folha docente por aluno no ano (excluem-se os custos de encargos) é de R\$ 10.052,00; no mesmo estrato essas instituições têm exposição média ao EAD de 3% do alunado. Ao estender-se a análise para os outros estratos, percebe-se claramente que, quanto maior o porte, menor o custo de folha docente por aluno, aumentando o percentual de exposição do alunado ao EAD.

Esse achado alinha-se com o objetivo deste trabalho e com a fundamentação teórica, mostrando que a ferramenta para redução dos custos docentes em grandes instituições tem sido a aplicação da modalidade EAD massificada.

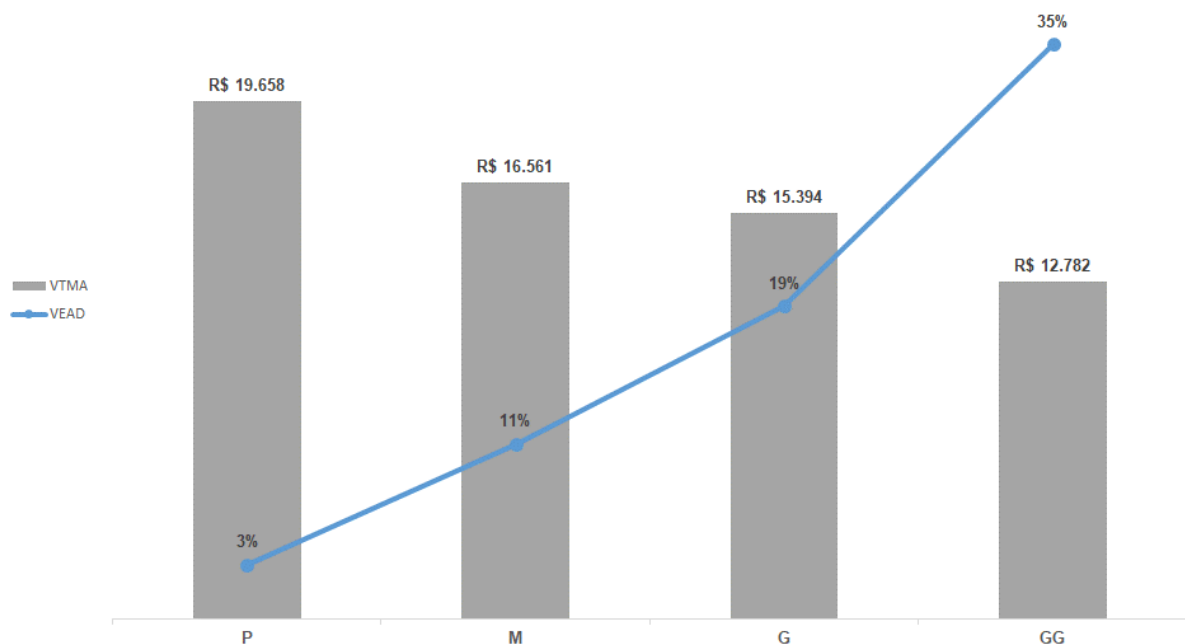
Na linha das receitas, a variável VTMA (*Ticket Médio por Aluno no ano*), em todos os anos censo, também, de forma similar, sempre teve significância alta, que,

relacionada ao *ticket* médio do aluno adequado aos custos, leva as instituições de ensino superior a manterem um distanciamento positivo do ponto de equilíbrio.

Para Castro (2013) e Xavier (2014), a sustentabilidade financeira de uma instituição de ensino superior depende da receita de mensalidades pagas pelos alunos e tem de apresentar um equilíbrio entre custo, quantidade de alunos pagantes e demais despesas

Para verificar, empiricamente, a relação entre o *Ticket* Médio do Aluno e a aplicação da modalidade EAD como ferramenta de gestão e expansão, adicionamos a variável no modelo anterior, de custos docentes, para o ano 2020, constante no gráfico da Figura 16.

**Figura 16 - Comparativo EAD x *Ticket* Médio/Aluno**



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

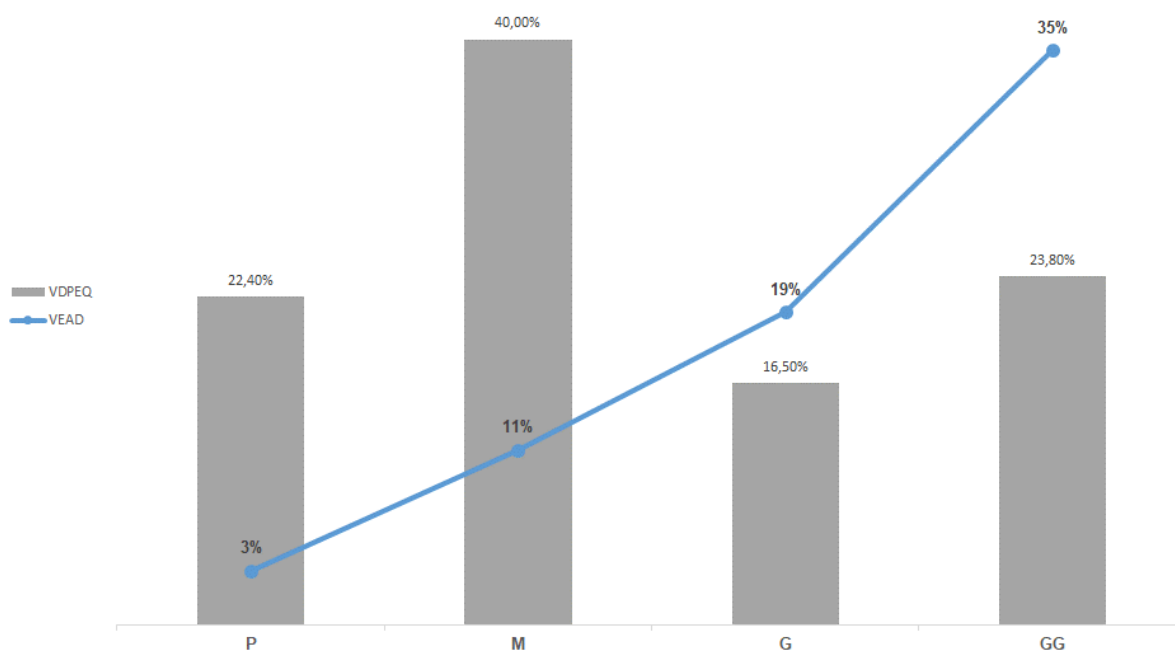
O que se percebe, na Figura 16, é uma curva parecida com a curva docente de acordo com o estrato ao qual pertence a mantenedora, com a mesma relação inversa de *Ticket* Médio e a exposição ao EAD (VEAD). A linha em azul representa, para cada estrato, o percentual de exposição ao EAD no grupo. Em barras de cor cinza, temos o *ticket* médio para o estrato, em valores monetários anualizados.

Em instituições de porte “P”, em média, o *Ticket Médio* por aluno no ano é de R\$ 19.658,00; no mesmo estrato essas instituições têm exposição média ao EAD de 3% do alunado. Ao estender-se a análise para os outros estratos, percebe-se claramente que, quanto maior o porte, menor o *Ticket Médio* por aluno conforme aumenta o percentual de exposição do alunado ao EAD.

Cabe uma comparação entre ambos os modelos, de custos e receitas em relação ao EAD: enquanto a diferença entre o custo docente relacionado aos estratos P e GG é de 100,1% (dobro), a diferença entre os *Tickets Médios* é de 53,8%, mostrando que em instituições menores o custo docente é maior do que nas grandes mantenedoras, implicando *ticket* médio maior e sugerindo um cenário desfavorável de competitividade.

Cabe uma análise similar para a variável que mede a distância do ponto de equilíbrio (VDPEQ), apresentado na Figura 7.

**Figura 17 - Comparativo EAD x Ponto de Equilíbrio**



Fonte: Elaborado pelo autor (2023) com base nos microdados (INEP, 2022)

Na Figura 17, percebe-se que pequenas mantenedoras possuem uma média de 22,4%, positiva, de distância do ponto de equilíbrio, com uma exposição ao EAD de 3%; já as de porte gigante, “GG”, apresentam uma distância, também positiva, de 23,8%, mas com exposição de 35% ao EAD.

Esses achados alinham-se com o objetivo deste trabalho e com a fundamentação teórica, sugerindo que a ferramenta para redução dos custos docentes em grandes instituições tem sido a aplicação da modalidade EAD massificada. Junta-se a esses fatores a concentração do número de matrículas anteriormente mostrado, de 68% do total neste grupo, podendo-se depreender disso uma alta rentabilidade financeira, objetivo dos grandes grupos consolidadores e mantenedoras.

### 5.5 Os grandes grupos e o comportamento dos fatores

Anteriormente adotamos a estratificação das mantenedoras educacionais pelo porte utilizando a estrutura proposta pelo Semesp (2021, p. 12) e exibida anteriormente no Quadro para avaliar o quanto a estratificação impacta nas variáveis determinantes do modelo de regressão.

Nesta seção, iremos verificar se existe impacto significativo no desvio dos fatores quando se consideram os grandes grupos educacionais do país, conforme descrito pela Hoper (2021). O intuito é verificar se, de acordo com os autores Colmán e Portes (2019) e Juliani, Santos e Fávero (2022), as grandes *holdings* de ensino conseguem aplicar seu poder econômico aos conceitos de custo/volume/lucro nos fatores identificados, apresentados na Tabela 24, a seguir.



Tabela 19 – Comparativo: fatores e grupos educacionais

Amostra	Total dos Registros	Grupos Consolidadores	Não Consolidadas
# Mantenedoras	1.376	81	1.295
# Matrículas	6.724.002	3.706.681	3.017.321
Variável	Média	Média	Média
VDPEQ	0,45	0,46	0,45
VRAD	17	39	16
VRAF	27	61	25
VCFPA	R\$ 9.624,69	R\$ 4.724,86	R\$ 9.931,17
VCODA	R\$ 5.914,48	R\$ 6.163,47	R\$ 5.898,91
VTMA	R\$ 18.884,53	R\$ 14.648,94	R\$ 19.149,46
VOCIO	0,26	0,34	0,25
VEAD	0,06	0,12	0,06
VPOL	0,15	0,19	0,15
IQCTA	1,28	1,26	1,28
IQCDO	3,18	3,42	3,17
VFINE	0,29	0,49	0,28
VBOLS	0,02	0,01	0,02
VEVAD	0,28	0,32	0,28
VRENC	0,36	0,46	0,35

Fonte: adaptado pelo autor com base em Hoper (2021) e Inep (2022)

Na Tabela 24, apresenta-se cada uma das variáveis calculadas e a média apresentada em cada amostra para análise, sendo primeiro para toda a amostra de 1.376 mantenedoras. A amostra seguinte considera somente as mantenedoras pertencentes aos oito grandes grupos consolidadores, 81 mantenedoras. Já as não consolidadas, 1.295 mantenedoras, são a última amostra desta análise.

Quanto à variável VDPEQ, que mede a rentabilidade ou distância do ponto de equilíbrio, entre as três amostras, é observado um equilíbrio, com valor de média em torno de 0,45. Essa variável nos indica que, na média, a rentabilidade das mantenedoras do ensino superior é positiva e com ligeira diferença positiva de 1% para os grandes grupos educacionais consolidadores.

Já analisando a média das distribuições Alunos/Docentes e Alunos/Funcionários Administrativos (respectivamente as variáveis VRAD e VRAF), identifica-se uma grande distância entre a média geral, instituições não pertencentes a grupos educacionais, e as que pertencem aos grandes grupos educacionais. Entre a média geral para a variável VRAD (17 alunos por docente) e a mesma variável medida em grandes grupos consolidadores (39 alunos por docente), temos uma

variação numérica de 129%. Já na variável que mede a relação alunos por funcionários administrativos (VRAF), temos respectivamente 27 alunos na média geral e 61 para as mantenedoras dos grandes grupos, com um incremento entre ambas de 122%. As mantenedoras não consolidadas têm respectivamente 8% (16) e 12% (25) abaixo da média geral, respectivamente, para as variáveis VRAD e VRAF.

Quando se analisa a variável VCFPA, que investiga a relação de custos de folha de pagamento totais da instituição, excluindo seus encargos, pela quantidade de alunos, percebe-se claramente a aplicação do conceito custo/volume/lucro, descrito anteriormente em Juliani, Santos e Fávero (2022).

A média do custo de folha de pagamento por aluno, considerando toda a amostra, é de R\$ 9.624,69/aluno-ano. Em instituição não consolidada, o valor medido é de R\$ 9.931,17/aluno-ano, praticando custos de folha de pagamento em torno de 3,2% acima da média geral. Quando se analisa a mesma variável – VCFPA – em grupos consolidadores, é apresentada uma redução de 51% no valor médio da mesma variável (R\$ 4.724,86/aluno-ano), o que significa praticamente que os grandes grupos consolidadores conseguem atender um número superior de alunos com a metade do custo de folha de pagamento docente quando comparados com outras instituições não consolidadas.

A análise de custos de operação (VCODA) – que inclui despesas gerais, de honorários, aluguéis, materiais de consumo, seguros, depreciação, serviços de terceiros e despesas relativas a vendas dos serviços, tais como comissões e marketing – mostra pouca variação entre as médias gerais, de consolidadas e não consolidadas. Temos como média geral R\$ 5.914,48/aluno-ano e em grandes grupos consolidadores um valor médio 4,2% acima – R\$ 6.163,47/aluno-ano, mostrando que este grupo de instituições investe mais em operações de marketing e suporte institucional. A variação é relativamente muito baixa em instituições não consolidadas, com redução de 0,3% em relação à média.

A análise da variável de receitas utilizadas neste estudo (VTMA), *ticket* médio do aluno no ano, mostra, em linha com a pesquisa de custo / volume / lucro de Juliani, Santos e Fávero (2022), que as instituições consolidadores podem, em decorrência de custos inferiores, praticar uma política de *ticket* médio mais agressiva. Na média geral, um aluno desembolsou em 2020 um valor de R\$ 18.884,53. Considerando-se uma instituição não consolidada, o valor médio de

anuidade foi de R\$ 19.149,46, o que representa 1,4% acima da média geral. Já em grupos consolidadores, o valor médio no ano foi de R\$ 14.648,94, com uma redução estimada de 22,4%.

Encerrando a análise de receitas e despesas, na comparação entre consolidadas e não consolidadas, para a avaliação de lucro ou prejuízo utilizando as variáveis VTMA (receitas) e VCFPA e VCODA (despesas), temos respectivamente lucro estimado de 35% e 21%, o que mostra que mesmo as consolidadoras, tendo reduzido o *ticket* médio e aumentado a competição mercadológica, devido ao custo reduzido em relação às não consolidadas, conseguiram aumentar a margem operacional.

A variável que mede a ociosidade, relação direta entre ingressantes e a quantidade de vagas abertas, definida na variável VOCIO, mostra que grupos consolidadores têm 33,1% maior eficiência na ocupação das vagas disponibilizadas em relação à média geral das mantenedoras. Temos uma média geral de 0,26 para a amostra total das mantenedoras diante de 0,34 em grupos consolidadores e 0,25 em não consolidadas (redução de 2,1%). Fatores de atração mercadológica, *ticket* médio agressivo e popularização da modalidade EAD podem explicar essa variação.

Na variável que mede a exposição à modalidade a distância (VEAD), temos uma variação medida de 110% acima da média geral (0,060) em relação aos grupos consolidadores (0,124) e abaixo da média em 6,9% para as não consolidadas (0,055), representando parte da explicação e estratégias desses grupos. Em linha com a exposição à modalidade a distância, a abertura de polos é relativamente maior em grupos consolidados (0,19) com 25,5% mais alunos matriculados fora da sede principal do que a média geral (0,15) e 27,1% em relação às instituições não pertencentes às consolidadoras (0,15).

Quanto à qualidade medida, na escala de 1 a 5 e preconizada pelo MEC, conforme fórmula descrita em Inep (2018), as notas técnico-administrativa (IQCTA) e docente (IQCDO) mostraram pouca variação entre a média geral, respectivamente 1,28 e 3,18 para 1,28 e 3,17 em instituições não consolidadas e 1,26 e 3,42 nas consolidadas. A variação de 7,4% na qualidade docente em grupos consolidadores é indicativa que rebate o argumento de Moran (2009) da redução da qualidade docente em favor a massificação. Embora Gomes, Leher e Costa (2020) argumentam que o uso de disciplinas comuns, com alocação de mestres e doutores – que têm maior peso no índice – atingindo um grande número de alunos e cursos,

pode ser uma estratégia dos grandes grupos consolidadores para a melhoria desse índice.

Os indicadores de exposição a financiamento estudantil e bolsas de estudo mostram que os grupos consolidadores possuem exposição em 66,4% acima da média geral (0,49 ante 0,29) sendo 4,2% menor em relação à média geral em instituições não consolidadas (0,28). Já em relação às bolsas, numericamente a relação captada entre bolsistas e matrículas é baixa, sendo em média 0,02, com menor exposição em grupos consolidados (0,01).

A evasão é um tema relevante para discussão, medido através da variável VEVAD. A média da evasão medida em 2020 foi de 0,28 (27,9%), em grandes grupos consolidadores foi de 0,32 (31,9%) e não consolidadas 0,28 (27,7%). Esses dados revelam que a preocupação dos grandes grupos é manter o aluno atraído e engajado no curso escolhido, devido a ter 14,2% a mais de evasão que a média do mercado.

Encerrando a análise da Tabela 24, temos a relação de encargos de folha de pagamento com os custos diretos de folha de pagamento administrativa e docente. Esta relação pode mostrar o quanto a instituição consegue administrar, reduzir ou aumentar os encargos em relação ao despendido em folha. A média da relação, em geral, é de 0,36 (a cada R\$ 100,00 gastos, R\$ 36,00 são encargos trabalhistas). Para os grandes grupos, esta relação esteve em 0,46 (a cada R\$ 100,00 gastos, R\$ 46,00 são encargos trabalhistas), 28% maior que a média geral; em não consolidadas, esta razão está em 0,35. Cabe recomendação de aprofundamento na investigação deste fator, principalmente devido a incentivos fiscais e trabalhistas apresentados nesse ano referência.

## 5.6 Comportamento dos fatores e o ponto de equilíbrio

Nesta subseção, será conduzida a avaliação e resposta a uma pergunta significativa para o objetivo deste trabalho: Qual o comportamento dos fatores entre as instituições mantenedoras que tiveram distância positiva do ponto de equilíbrio (lucro operacional), distância negativa (prejuízo operacional) e as que ficaram neutras (neste caso, potencial prejuízo após as avaliações contábeis).

Tabela 20 - Comparativo Fatores e Ponto de Equilíbrio

Porte	Total dos Registros	P.E. (positivo)	P.E. (negativo)
<b># Mantenedoras</b>	<b>1376</b>	<b>832</b>	<b>544</b>
Variável	Média	Média	Média
VDPEQ	0,45	0,94	-0,30
VRAD	17	19	14
VRAF	27	31	22
VCFPA	9.624,69	8.273,57	11.691,10
VCODA	5.914,48	5.243,09	6.941,32
VTMA	18.884,53	23.193,95	12.293,67
VOCIO	0,26	0,27	0,23
VEAD	0,059	0,062	0,055
VPOL	0,149	0,160	0,133
IQCTA	1,28	1,26	1,30
IQCDO	3,18	3,16	3,21
VFINE	0,29	0,32	0,26
VBOLS	0,02	0,02	0,02
VEVAD	0,279	0,262	0,306
VRENC	0,36	0,35	0,37

Fonte: adaptado pelo autor com base em Inep (2022)

Na Tabela 25, representa-se cada uma das variáveis calculadas e a média apresentada em cada amostra para análise, em primeiro lugar para toda a amostra de 1.376 mantenedoras. A amostra seguinte considera somente as mantenedoras que tiveram o ponto de equilíbrio calculado positivamente, sendo a amostra representada por 832 instituições, 60,4% do total. As que tiveram resultado negativo ou neutro (distância do ponto de equilíbrio = zero) representam 39,6% do total, com 544 instituições.

Considerando a média das distribuições Alunos/Docentes e Alunos/Funcionários Administrativos (respectivamente as variáveis VRAD e VRAF), percebe-se que as instituições com resultados positivos tiveram uma distribuição de recursos maior do que a média geral, sendo Alunos/Docentes uma relação de 17 na média geral, 19 nas mantenedoras com resultados positivos e 14 nas de resultado negativo. Isso representa que as instituições com resultado negativo tiveram uma relação Alunos/Docentes menos otimizada. A mesma interpretação se verifica na variável que descreve a relação Alunos/Funcionários Administrativos.

A média do custo de folha de pagamento por aluno, considerando toda a amostra, é de R\$ 9.624,69 / aluno-ano. Nas instituições positivas no ponto de

equilíbrio, o valor medido é R\$ 8.273,57/aluno-ano, inferior à média geral. Quando se analisa a mesma variável – VCFPA – em instituições com ponto de equilíbrio negativo ou neutro, é apresentado um aumento de 21,5% do valor médio na mesma variável (R\$ 11.691,10 / aluno-ano), significando praticamente que as instituições com ponto de equilíbrio negativo tendem a gastar mais do que a média em Folha Docente. A mesma análise se verifica com a variável de Custos de Operação (VCODA), com uma variação de 17,4%.

A análise da variável de receitas utilizadas neste estudo (VTMA), *ticket* médio do aluno no ano, mostra que as instituições lucrativas praticaram uma política de *ticket* médio 22,8% superior à da média geral e 47% maior do que as mantenedoras na linha negativa. Este achado pode representar que as instituições integrantes da amostra de ponto de equilíbrio negativo têm dificuldades em manter uma política de preços adequada e em linha com o custo de operação educacional.

A análise da evasão mostrou que as instituições de ponto de equilíbrio positivo têm praticamente a mesma média da amostra total, com 26,2%. Já nas negativas, temos maior evasão medida, sendo de 3,6 pontos percentuais em relação às instituições positivas (30,6%). Cabe recomendação de aprofundamento do impacto da evasão neste perfil de instituições.

## 6 CONSIDERAÇÕES FINAIS

O estudo apresentado neste trabalho buscou, por meio de referencial teórico sobre o ensino superior privado brasileiro e modelos de análise estatística, quando aplicados a dados secundários fornecidos pelo órgão oficial responsável pelas estatísticas da educação brasileira (INEP), contribuir na avaliação de possíveis fatores que influenciaram o resultado financeiro, distanciamento do ponto de equilíbrio, nas mantenedoras educacionais de ensino superior brasileiras, identificando desde fatores diretamente relacionados aos eventos financeiros a cenários mercadológicos.

O referencial teórico foi resumido e tratou sobre ensino o superior privado brasileiro, a concentração do mercado de ensino privado e a busca do lucro, em razão da qual grande parte das mantenedoras migrou para o modelo administrativo com fins lucrativos. Também foi evidenciado o embasamento das variáveis objeto deste estudo como o ponto de equilíbrio e as teorias de custo/volume/lucro, bem como suas relações com o mercado educacional e os dados fornecidos pelo Inep em seus dados disponibilizados. Cumpre parte do referencial teórico o estabelecimento dos métodos estatísticos e de análise de componentes para esclarecimento das escolhas, prévias à aplicação do procedimento metodológico.

O procedimento metodológico realizado, com importação dos microdados e geração das bases de dados, tratamento e remoção dos pontos extremos, resultou favorável ao prosseguimento das análises, sem redução significativa dos registros. A análise seguiu com dados de um total de 1.450 mantenedoras em 2018, 1.482 em 2019 e 1.376 em 2020, apresentando, assim, uma média de representatividade de 93,9% do total importado e disponibilizado pelo Inep para análises dos microdados.

O estudo começou com o modelo de análise de componentes principais (PCA), com as mesmas bases de dados. O intuito desta análise foi observar quais grupos de componentes (variáveis independentes) têm maior peso na variação observada na variável de estudo, dependente. Em todos os anos censo, quando usado o modelo em que as variáveis foram normalizadas, não houve um grupo de destaque (acima de 30% de contribuição na variação), sendo a contribuição da variação bem distribuída entre os componentes. Com uma média de contribuição de 33%, considerando-se somente os dois primeiros componentes, novamente temos

Custo Docente, Custo de Operação e Ticket Médio do Aluno como os fatores decisivos, embora não dominantes pelos fatores.

O primeiro modelo de análise utilizado foi o de regressão linear multivariada (RLM), aplicada separadamente a cada base de dados, separadas em anos censo. Os valores calculados de  $R^2$  de 36,04% para 2018 e 44,9% para 2019 mostraram que existe um fator calculado de intensidade moderada na explicação da variação da variável dependente (distância do ponto de equilíbrio) em relação às variáveis independentes (demais variáveis calculadas). Já no ano censo de 2020, tivemos uma explicação considerada forte na variação, com um  $R^2$  de 69,4%. Este modelo trouxe como fatores determinantes no resultado e importantes para a variação da distância do ponto de equilíbrio as variáveis de Custo Docente, Custo de Operação e *Ticket* Médio do Aluno, sendo seguidos pelas relações Aluno/Docente e Aluno/Funcionários Administrativos.

Adicionalmente foi realizado um estudo com a utilização do modelo estatístico de Florestas Aleatórias (*Random Forest*), introdutória a modelos de análise de inteligência artificial. A execução deste algoritmo trouxe algumas descobertas importantes, trazendo para primeiro lugar como fator de distanciamento positivo do ponto de equilíbrio o *Ticket* médio e depois o fator de Custo Docente. Os demais fatores foram sendo distribuídos ao longo dos nós decisores inferiores da árvore.

Após a análise da amostra total das bases de dados, cabia a análise por amostragem filtrada pelo porte das mantenedoras, respondendo à pergunta do fator mercadológico ser importante no distanciamento. Foi observado, usando a técnica de regressão linear multivariada (RLM) que o  $R^2$  em cada grupo aumentou de significância, onde nos casos o grau atribuído foi forte explicação nos estratos "P", "G" e "GG". O estrato "M" foi inconclusivo. Foi mostrada a participação dos fatores Custo Docente, Custo de Operação e *Ticket* Médio do Aluno como relevantes na análise.

Cabe destacar nesta análise, apesar de não ter aparecido com relevância nas regressões, o fator Ensino a Distância, que, conforme o referencial teórico apresentado por Gomes, Leher e Costa (2020), França Filho, Antunes e Couto (2020) e Novaes, Nascimento e Gonçalves (2022), é buscado pelas grandes instituições como ferramenta de redução e custos e *ticket* médio. Foi mostrado que existe uma relação entre Ponto de Equilíbrio, Custo Docente e *Ticket* Médio do aluno de acordo com o porte da instituição, podendo as grandes mantenedoras exercer



maior poder econômico ao praticar mensalidades menores que as instituições de menor porte, devido à redução de custos gerais e análise de custo/volume/lucro.

Cumprido ressaltar a confirmação do referencial teórico sobre a concentração do mercado educacional superior brasileiro, no qual oito grandes grupos educacionais detêm 57% do total de matrículas do mercado privado de ensino superior e, quando se fala de matrículas na modalidade a distância, a dominação é de 77% das matrículas, mostrando, assim, que o mercado de ensino superior brasileiro é muito consolidado.

Portanto, respondendo aos objetivos deste trabalho, foi feito o levantamento na base de dados do censo de variáveis independentes que podem ter relação com o desempenho das IES e suas mantenedoras no distanciamento do ponto de equilíbrio em consonância com a literatura apresentada. Foi avaliado o peso das variáveis independentes, que podem ter relação com o desempenho das IES e suas mantenedoras no distanciamento do ponto de equilíbrio (variável dependente), possibilitando concluir que os fatores Custo Docente, Custo de Operação e *Ticket Médio*, bem como fatores administrativos, como a relação Aluno/Docentes e Alunos/Funcionários, são importantes no distanciamento do ponto de equilíbrio.

Os indicadores de matrículas por porte institucional e modalidade de ensino mostraram a consolidação de mercado e a migração de alunado sugere fator mercadológico para o distanciamento do ponto de equilíbrio.

Com a definição dos fatores e o atingimento dos objetivos deste trabalho, cumpre-se e recomenda-se exploração de assuntos adicionais e complementares de estudo.

### 6.1 Lacunas e futuras pesquisas

A limitação de análise do impacto da evasão, ponto importante citado no referencial teórico, não teve descoberta como fator importante nos modelos de regressão, cabendo estudo exclusivo do impacto deste fator, principalmente na modalidade Ensino a Distância.

Ainda na modalidade de Ensino a Distância, recomenda-se o aprofundamento da análise sobre a interiorização da oferta – abertura de polos fora da sede principal – e o impacto que o conceito institucional atribuído pelo MEC tem neste indicador.

Outra recomendação de estudo é sobre o impacto fiscal e tributário na gestão de instituições de ensino quanto ao atingimento da distância do ponto de equilíbrio. Neste caso a pergunta que fica em lacuna é se instituições com maior porte têm maior poder de aplicação de benefícios tributário e de encargos, principalmente em folha de pagamento.

## REFERÊNCIAS

- ABED – ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA. **Censo EAD.BR**: relatório analítico da aprendizagem a distância no Brasil 2020. Curitiba: InterSaber, 2022. Disponível em: [https://abed.org.br/arquivos/CENSO\\_EAD\\_2020\\_PORTUGUES.pdf](https://abed.org.br/arquivos/CENSO_EAD_2020_PORTUGUES.pdf). Acesso em: 15 dez. 2022.
- ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14.653-2**. Rio de Janeiro: [S.n.], 2010.
- ASSAF NETO, Alexandre. **Finanças corporativas e valor**. 8a. ed. São Paulo: Atlas, 2020.
- ASSIS, Janilson Pinheiro de; SOUSA, Roberto Pequeno de; DIAS, Carlos Tadeu dos Santos. **Glossário de estatística**. Mossoró/RN: EdUFERSA - Universidade Federal do Semi-Árido, 2019. 901 p. Disponível em: <https://livraria.ufersa.edu.br/glossario-de-estatistica/>. Acesso em: 20 mar. 2022.
- AVELAR, Ewerton E. A. Algoritmo *random forest* para previsão de comportamento de preços de ativos. **Revista FSA**, Teresina/PI, v. 19, n. 10, p. 45-65, out. 2022.
- BARREYRO, Gladys B. **Mapa do ensino superior privado**. Brasília: Inep, 2008. Série Documental Relatos de Pesquisa.
- BRASIL. Lei Nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União –DOU**, Brasília, DF, 20 dez. 1996.
- BRASIL, MEC. Portaria nº 343, de 17 de Março de 2020. **Diário Oficial da União - DOU**, 17 mar. 2020. Brasília, DF, 17 mar. 2020.
- BRASIL, MEC. Portaria nº 345, de 17 de Março de 2020. **Diário Oficial da União - DOU**, 17 mar. 2020. Brasília, DF, 17 mar. 2020.
- BRASIL. Decreto Nº 2.306, de 19 de agosto de 1997. **Diário Oficial da União – DOU**, Brasília/DF, 20 ago. 1997.. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=20/08/1997&jornal=1&pagina=3&totalArquivos=168>. Acesso em: 25 jan. 2022.
- BREIMAN, Leo. Random forests. **Machine Learning** **45**, Berkeley, p. 5-32, 2001.
- BURGEL, Evandro; ANZANELLO, Michel J. Abordagem para seleção de variáveis preditivas no contexto de controle de inventários. **Revista Gestão Industrial – RGI**, Ponta Grossa, v. 14, n. 4, p. 154-195. out./dez. 2018. Disponível em: <https://periodicos.utfpr.edu.br/revistagi/article/download/7690/5438>. Acesso em: 27 out. 2022.
- CADE – CONSELHO ADMINISTRATIVO DE DEFESA ECONÔMICA. **Atos de Concentração no Mercado de Prestação de Serviços de Ensino Superior**.

Conselho Administrativo de Defesa Econômica – Departamento de Estudos Econômicos do Ministério da Justiça. Brasília - DF. 2016.

CASTRO, Alessandro de. **Gestão Financeira: a sustentabilidade financeira e a interação acadêmica nas instituições de ensino superior.** Porto Alegre: Penso - Grupo A, 2013. p. 179-195.

CHEIN, Flávia. **Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas.** Brasília/DF: Enap Fundação Escola Nacional de Administração Pública, 2019. 76 p. Disponível em: <http://repositorio.enap.gov.br/handle/1/4788>. Acesso em: 20 mar. 2022.

COLMÁN, Evaristo; PORTES, Lorena Ferreira. A orientação privatista-mercantil do ensino superior brasileiro e suas expressões nas propostas do governo federal – “Future-se” e da lei geral das universidades – “LGU” do governo do estado do Paraná. **Humanidades & Inovação**, v. 6, n. 17, 2019.

CORREIA, Eneida Cardoso de Britto. **Modelo regulatório híbrido da educação superior privada: possibilidades, limites e desafios.** 2015. Dissertação (Mestrado) – Escola Brasileira de Administração Pública e de Empresa, Centro de Formação Acadêmica e Pesquisa, 2015. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/14223>. Acesso em: 27 jan. 2022.

DEVORE, L. J. **Probabilidade e estatística: para engenharia e ciências.** 6a. ed. São Paulo: Cengage Learning, 2006.

DINIZ, Virgínia R.; GOERGEN, Pedro L. Educação Superior no Brasil: panorama da contemporaneidade. **Avaliação**, Campinas, v. 24, n. 03, p. 573-593, nov. 2019. Disponível em: <https://www.scielo.br/j/aval/a/KWJWLBPfjBkKbzSXw7TStb/?lang=pt>. Acesso em: 14 jan. 2022.

FARIA, Ricardo Rocha. **Gestão de riscos corporativos e monitoramento por meio de indicadores de desempenho: práticas em instituições comunitárias de educação superior.** 2019. 133p. Dissertação (Mestrado) – Universidade Metodista de São Paulo, São Bernardo do Campo, 2019.

FÁVERO, Luiz Paulo. **Análise de dados.** Rio de Janeiro: Elsevier, 2015.

FENEP – FEDERAÇÃO NACIONAL DAS ESCOLAS PARTICULARES. **Números da Educação Privada Brasileira.** São Paulo: Fenep, 2022

FIELD, A. **Discovering statistics using SPSS.** 3rd ed. London: Sage, 2009.

FRANÇA FILHO, A. L. de F.; ANTUNES, C. da F.; COUTO, M. A. C. Alguns apontamentos para uma crítica da educação a distância (EaD) na educação brasileira em tempos de pandemia. **Revista Tamoios**, São Gonçalo (RJ), ano 16, n. 1, p. 16-31, maio 2020. Especial Covid-19.

FRITSCH, Rosângela; JACOBUS, Artur Eugênio; VITELLI, Ricardo Ferreira. Diversificação, mercantilização e desempenho da educação superior brasileira. **Avaliação**, Campinas, v. 25, n. 1, jan.-abr. 2020.

GARSON, G. David. **Logistic regression: Binary & multinomial**. 2016. USA: Statistical Associates Publishing, 2016.

GOMES, T. A. M.; LEHER, R.; COSTA, H. B. Teaching work in higher education corporate groups and the stock market: A conflict-based study. **Archivos Analíticos de Políticas Educativas / Education Policy Analysis Archives**, v. 28, p. 1-29, 2020.

GUJARATI, Damodar. **Econometrics by example**. [S.l.]: Palgrave Macmillan, 2011.

HAIR, Joseph F. *et al.* **Análise multivariada de dados**. 6a. ed. Porto Alegre: Bookman, 2009.

HOPER EDUCAÇÃO. **Análise Setorial da Educação Superior Privada – Brasil**. 14a. ed. Foz do Iguaçu/PR: Hoper, 2021. Edição Eletrônica, 2021.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estudos e pesquisas. Informação demográfica e socioeconômica**. Rio de Janeiro: IBGE, 1999.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Projeção da população do Brasil e das unidades da Federação**. Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/apps/populacao/projecao/index.html>. Acesso em: 25 jan. 2022.

INEP. **Sinopse estatística do ensino superior: graduação 1997**. Brasília: MEC, 1999.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse estatística da educação superior 2017**. Brasília: Inep, 2018.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Nota Técnica Nº 2/2018/CGACGIES/DAES**. Brasília: SEI/Inep, 2018. Disponível em: [https://download.inep.gov.br/educacao\\_superior/avaliacao\\_cursos\\_graduacao/legislacao\\_normas/2017/nota\\_tecnica\\_sei\\_inep\\_0159763.pdf](https://download.inep.gov.br/educacao_superior/avaliacao_cursos_graduacao/legislacao_normas/2017/nota_tecnica_sei_inep_0159763.pdf). Acesso em: 10 abr. 2022.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados Censo da Educação Superior 2019. **Microdados Censo da Educação Superior**. Brasília: Inep, 2020. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>. Acesso em: 26 jan. 2022.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Manual de preenchimento do Censo da Educação Superior 2020: Módulo Instituição de Educação Superior (IES)**. Brasília/DF: Inep, 2021.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Censo da Educação Superior 2018, 2019 e 2020: notas estatísticas**. Brasília: Inep, 2022.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Comunicado de deferimento da 1ª extração de resultados**. Brasília: Inep, 2022.

INSTITUTO SEMESP. Estudo Efeitos da Pandemia na Educação Superior Brasileira. **SEMESP**, São Paulo, p. 7, 2020. Disponível em: <https://www.semesp.org.br/pesquisas/estudo-efeitos-da-pandemia-na-educacao-superior-brasileira/>. Acesso em: 15 out. 2021.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning: with applications**. New York: R. Springer, 2013.

JESUS, Nataly Batista de. O sucateamento do ensino superior através da educação a distância. **Arma da Crítica – Revista Eletrônica**, n. 13, maio 2020.

JULIANI, Marcio Pedroso; SANTOS, Antônio Pereira dos; FÁVERO, Altair Alberto. A mercantilização do Ensino Superior a partir da Educação a Distância: Aspectos gerais de um diagnóstico crítico. **Tear: Revista de Educação Ciência e Tecnologia**, Porto Alegre/RS, v. 11, n. 2, 2022.

KHAN, Zardad E. A., GUL, Asma; PERPEROGLU, Aris, MIFTAHUDDIN, Miftahuddin; MAHMOUD, Osama; ADLER, Werner; LAUSEN, Berthold Ensemble of optimal trees, random forest and random projection ensemble classification. **Advances in Data Analysis and Classification**, v. 14, n. 1, p. 97-116, 2020.

KREMER, Guilherme M.; DEINA, Carolina; SIQUEIRA, Hugo. Correlação e regressão linear de variáveis que interferem no Produto Interno Bruto do Brasil: uma análise estatística de dados. **Revista Gestão Industrial – RGI**, Ponta Grossa, v. 15, n. 2, p. 233-254, abr./jun. 2019. Disponível em: <https://periodicos.utfpr.edu.br/revistagi/article/view/9968>. Acesso em: 27 out. 2022.

LUCHESA, Cláudio José; MACHADO, Cristiane Ribas. A margem de contribuição na gestão de instituições de ensino superior – IES. **Teoria e Prática da Educação**, v.14, n. 2, p. 113-122, 2011.

LUGOBONI, Leonardo Fabris. **Modelos de avaliação de desempenho organizacional em instituições de ensino superior na Grande São Paulo**. 2010. 127 p. Dissertação (Mestrado) – Universidade Municipal de São Caetano do Sul, São Caetano do Sul, 2010. Disponível em: <https://www.uscs.edu.br/pos-stricto-sensu/arquivo/363>. Acesso em: 28 jan. 2022.

LUGOBONI, Leonardo Fabris. **A relação entre os sistemas de gestão por indicadores e o desempenho das instituições de ensino superior**. 2017. 300 p. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2017.

- MACHADO, Lucio de S. **A controladoria no processo de gestão de instituições privadas de ensino superior de Goiás: um estudo sob a ótica de contabilidade e finanças**. 2005. 227f. Dissertação (Mestrado) – Fecap - Faculdade Escola de Comércio Álvares Penteado, São Paulo, 2005. Disponível em: <http://tede.fecap.br:8080/handle/tede/564>. Acesso em: 13 jan. 2022
- MEINERTZ, Carmo André. **Gestão estratégica de indicadores aplicada a uma instituição de ensino superior confessional**. 2019. 142p. Dissertação (Mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Gestão Educacional, Porto Alegre/RS, 2019. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/8981>. Acesso em: 28 jan. 2022.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 6th. ed. [S.l.]: John Wiley & Sons., 2017.
- MORAN, José Manuel. Modelos e avaliação do ensino superior a distância no Brasil. **ETD – Educação Temática Digital**, Campinas, v. 10, n. 2, p. 54-70, 2009. Disponível em: [http://www2.eca.usp.br/moran/?page\\_id=33](http://www2.eca.usp.br/moran/?page_id=33). Acesso em: 05 dez. 2022.
- MOROSINI, Marilia C. **A universidade no Brasil: concepções e modelos**. 2a. ed. Brasília/DF: Inep, 2006. vol.1.
- NOVAES, M. A. B. de. ; NASCIMENTO, N. C. A. ; GONCALVES, R. M. P. . Educação a distância nos cursos de licenciaturas e a mercantilização da educação superior. **Revista Cocar** , v. 16, n. 34, p. 1-20, 2022.
- OPENINTRO.ORG. **Livro de probabilidade e estatística EAD – UFRGS**. S.d. Disponível em: <https://www.ufrgs.br/probabilidade-estatistica/livro/livrocompleto/index.html>. Acesso em: 27 out. 2022.
- OYADOMARI, José Carlos Tiomatsu *et al.* **Contabilidade gerencial: ferramentas para melhoria de desempenho empresarial**. São Paulo: Atlas, 2018.
- PALHARES, Isabela. Mais de 850 mil alunos deixam de cursar ensino superior após pandemia. Folha de São Paulo - Educação, 19 out. 2020. Disponível em: <https://www1.folha.uol.com.br/educacao/2020/10/mais-de-850-mil-alunos-deixam-de-cursar-ensino-superior-apos-pandemia.shtml>. Acesso em: 15 out. 2021.
- PORFIRIO, Karen Rayane C. **Análise de risco de crédito via técnicas de estatística multivariada**. 2021. Trabalho de Conclusão de Curso – Unifesp, São José dos Campos, São Paulo, 2021.
- PROETTI, Sidney. As pesquisas qualitativa e quantitativa como métodos de investigação científica: um estudo comparativo e objetivo. **Revista Lumen**, São Paulo, v. 2, n. 4, jul.-dez. 2017. Disponível em: <http://dx.doi.org/10.32459/revistalumen.v2i4>. Acesso em: 13 nov. 2022
- R CORE TEAM. **R: A language and environment for statistical. R Foundation for Statistical Computing**, Vienna, 2022. Disponível em: <https://www.R-project.org/>. Acesso em: 13 nov. 2022.

SALVIAN, Mayara. **Multicolinearidade**. Piracicaba/SP: Escola Superior de Agricultura Luiz de Queiroz, 2016. Disponível em: [https://edisciplinas.usp.br/pluginfile.php/2340848/mod\\_resource/content/0/Mayara\\_Multicolinearidade.pdf](https://edisciplinas.usp.br/pluginfile.php/2340848/mod_resource/content/0/Mayara_Multicolinearidade.pdf). Acesso em: 13 nov. 2022

SAMPAIO, Helena. O setor privado de ensino superior no Brasil: continuidades e transformações. **Ensino Superior Unicamp**, Campinas, n. 4, p. 28-43, out. 2011. Disponível em: [http://www.gr.unicamp.br/ceav/revistaensinosuperior/edicoes/ed04\\_outubro2011/05\\_ARTIGO\\_PRINCIPAL.pdf](http://www.gr.unicamp.br/ceav/revistaensinosuperior/edicoes/ed04_outubro2011/05_ARTIGO_PRINCIPAL.pdf). Acesso em: 25 jan. 2022.

SAMPIERI, Roberto Hernández; COLLADO, Carlos Fernández; LUCIO, María del Pilar Baptista. **Metodologia de pesquisa**. 5a. ed. Porto Alegre: Penso/Mc Graw Hill, 2013.

SCHOPENHAUER, Arthur. **Parerga y paralipómena**. Parerga y paralipómena, 1851.

SEMESP – SINDICATO DAS ENTIDADES MANTENEDORAS DE ESTABELECIMENTOS DE ENSINO SUPERIOR NO ESTADO DE SÃO PAULO. **Mapa do Ensino Superior no Brasil**. 11a ed. São Paulo: Instituto Semesp, 2021.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 24a. ed. São Paulo: Cortez, 2017.

SGUISSARDI, Valdemar. Educação superior no Brasil: democratização ou massificação mercantil? **Revista Educ. Soc.**, Campinas, v. 36, n. 133, p. 867-889, out./dez. 2015. Disponível em: [www.scielo.br/pdf/es/v36n133/1678-4626-es-36-133-00867.pdf](http://www.scielo.br/pdf/es/v36n133/1678-4626-es-36-133-00867.pdf). Acesso em: 26 jan. 2022.

SILVA, Claudio Marcos Maciel da ; REZENDE, José Francisco de Carvalho. Geração de valor no ensino superior privado: uma análise do Plano de Desenvolvimento Institucional (PDI). **Contextus: Revista Contemporânea de Economia e Gestão**, Fortaleza, p. 66-99, set./dez. 2017. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/29165>. Acesso em: 22 jun. 2022

SILVA, Renato. **Sustentabilidade financeira das instituições de ensino superior privadas**: um modelo de gestão estratégica com base no sistema nacional de avaliação superior (SINAES). 2017. 184 p. Tese (Doutorado) – Centro Universitário La Salle, Canoas/RS, 2017. Disponível em: <http://hdl.handle.net/11690/925>. Acesso em: 18 mar. 2022.:

SOARES, Thiago Coelho. **Avaliação de desempenho em instituições de ensino superior**: um estudo de fatores e competitividade. 2016. 424 p. Tese (Doutorado) – Universidade Federal de Santa Catarina, Florianópolis, 2016. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/172589>. Acesso em: 15 mar. 2022.

TABACHNICK, Barbara G.; FIDELL, Linda S. **Using multivariate statistics**. 7a. ed. New York: Pearson, 2019.

TRAN, Hieu. **A survey of machine learning and data mining techniques used in multimedia system**. 2019. USA: ResearchGate, 2019.



WEISBERG, Sanford. **Applied linear regression**. 4th. ed. Minneapolis: Wiley, 2014.

XAVIER, Aline da Rocha. **Análise de custos como ferramenta de gestão estratégica da educação superior**: um estudo dos conceitos do ponto de equilíbrio e margem de contribuição e sua aplicabilidade em IES privada. 2014. 118 p. Dissertação (Mestrado) – Universidade Federal do Ceará, Fortaleza, 2014.

ZAMIN, Marcia Regina. **Análise dos indicadores da gestão estratégica de custos dos cursos de graduação com estruturas diferenciadas**. 2019. 105 p. TCC – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí/RS, 2019. Disponível em: <http://bibliodigital.unijui.edu.br:8080/xmlui/handle/123456789/5581>. Acesso em: 19 mar. 2022.

## Anexo A – Dicionários de Dados das Variáveis do Censo da Educação Superior

Quadro 11 - Dicionário de Variáveis do Censo da Educação Superior – IES

#	Nome da Variável	Descrição da Variável	Categoria
1	NU_ANO_CENSO	Ano de referência do Censo da Educação Superior	
2	NO_REGIAO_IES	Nome da região geográfica da sede administrativa ou reitoria da IES	
3	CO_REGIAO_IES	Código da região geográfica	
4	NO_UF_IES	Nome da Unidade da Federação da sede administrativa ou reitoria da IES	
5	SG_UF_IES	Sigla da Unidade da Federação da sede administrativa ou reitoria da IES	
6	CO_UF_IES	Código da UF da sede administrativa ou reitoria da IES	
7	NO_MUNICIPIO_IES	Nome do Município da sede administrativa ou reitoria da IES	
8	CO_MUNICIPIO_IES	Código do Município da sede administrativa ou reitoria da IES	
9	IN_CAPITAL_IES	Sede administrativa ou reitoria da IES está localizada em uma capital da Unidade da Federação	0. Não 1. Sim
10	NO_MESORREGIAO_IES	Nome da Mesorregião da sede administrativa ou reitoria da IES	
11	CO_MESORREGIAO_IES	Código da Mesorregião da sede administrativa ou reitoria da IES	
12	NO_MICRORREGIAO_IES	Nome da Microrregião da sede administrativa ou reitoria da IES	
13	CO_MICRORREGIAO_IES	Código da Microrregião da sede administrativa ou reitoria da IES	
14	TP_ORGANIZACAO_ACADEMICA	Tipo de Organização Acadêmica da IES	1. Universidade 2. Centro Universitário 3. Faculdade 4. Instituto Federal de Educação, Ciência e Tecnologia 5. Centro Federal de Educação Tecnológica

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
15	TP_CATEGORIA_ADMINISTRATIVA	Tipo de Categoria Administrativa da IES	1. Pública Federal 2. Pública Estadual 3. Pública Municipal 4. Privada com fins lucrativos 5. Privada sem fins lucrativos 6. Privada – Particular em sentido estrito 7. Especial 8. Privada comunitária 9. Privada confessional
16	NO_MANTENEDORA	Nome da mantenedora da IES	
17	CO_MANTENEDORA	Código único de identificação da mantenedora da IES	
18	CO_IES	Código único de identificação da IES	
19	NO_IES	Nome da IES	
20	SG_IES	Sigla da IES	
21	DS_ENDERECO_IES	Endereço da sede administrativa/reitoria da IES	
22	DS_NUMERO_ENDERECO_IES	Número do endereço	
23	DS_COMPLEMENTO_ENDERECO_IES	Complemento	
24	NO_BAIRRO_IES	Bairro	
25	NU_CEP_IES	CEP	
26	QT_TEC_TOTAL	Quantidade de funcionários técnico-administrativos	
27	QT_TEC_FUNDAMENTAL_INCOMP_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com ensino fundamental incompleto	
28	QT_TEC_FUNDAMENTAL_INCOMP_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com ensino fundamental incompleto	
29	QT_TEC_FUNDAMENTAL_COMP_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com ensino fundamental completo	
30	QT_TEC_FUNDAMENTAL_COMP_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com ensino fundamental completo	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
3 2	QT_TEC_MEDIO_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com ensino médio	
3 3	QT_TEC_SUPERIOR_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com nível superior	
3 4	QT_TEC_SUPERIOR_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com nível superior	
3 5	QT_TEC_ESPECIALIZACAO_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com especialização	
3 6	QT_TEC_ESPECIALIZACAO_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com especialização	
3 7	QT_TEC_MESTRADO_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com mestrado	
3 8	QT_TEC_MESTRADO_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com mestrado	
3 9	QT_TEC_DOUTORADO_FEM	Quantidade de funcionários técnico-administrativos do sexo feminino com doutorado	
4 0	QT_TEC_DOUTORADO_MASC	Quantidade de funcionários técnico-administrativos do sexo masculino com doutorado	
4 1	IN_ACESSO_PORTAL_CAPESS	Informa se as bibliotecas da IES têm acesso ao portal Capes de periódicos	0. Não 1. Sim
4 2	IN_ACESSO_OUTRAS_BASES	Informa se as bibliotecas da IES têm acesso a outras bases de dados licenciadas ou compradas	0. Não 1. Sim
4 3	IN_ASSINA_OUTRA_BASE	Informa se as bibliotecas da IES assinam outras bases de dados licenciadas ou compradas	0. Não 1. Sim
4 4	IN_REPOSITORIO_INSTITUCIONAL	Informa se a IES possui base de dados <i>online</i> que reúne de maneira organizada a produção científica da instituição	0. Não 1. Sim
4 5	IN_BUSCA_INTEGRADA	Informa se a IES oferece ferramenta eletrônica que possibilita pesquisar simultaneamente diversas fontes de informação e obter resultado em uma única interface	0. Não 1. Sim
4 6	IN_SERVICO_INTERNET	Informa se as bibliotecas da IES oferecem serviços pela internet	0. Não 1. Sim
4 7	IN_PARTICIPA_REDE_SOCIAL	Informa se as bibliotecas participam de redes sociais	0. Não 1. Sim

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
48	IN_CATALOGO_ONLINE	Informa se a IES possui ferramenta de recuperação de informação que permite ao usuário consultar, de forma local ou remota, a existência e disponibilidade de itens do acervo das bibliotecas	0. Não 1. Sim
49	QT_PERIODICO_ELETRONICO	Quantidade de títulos de periódicos eletrônicos adquiridos pelas bibliotecas por meio de compra, doação ou permuta	
50	QT_LIVRO_ELETRONICO	Quantidade de títulos de livros eletrônicos disponibilizados pelas bibliotecas convertidos ao formato digital ou originalmente produzidos nesse formato para serem lidos em computador ou outros dispositivos	
51	QT_DOC_TOTAL	Quantidade total de docentes (em exercício e afastados)	
52	QT_DOC_EXE	Quantidade total de docentes em exercício	
53	QT_DOC_EX_FEMI	Quantidade de docentes em exercício do sexo feminino	
54	QT_DOC_EX_MASC	Quantidade de docentes em exercício do sexo masculino	
55	QT_DOC_EX_SEM_GRAD	Quantidade de docentes em exercício sem curso de graduação	
56	QT_DOC_EX_GRAD	Quantidade de docentes em exercício com curso de graduação	
57	QT_DOC_EX_ESP	Quantidade de docentes em exercício com especialização	
58	QT_DOC_EX_MEST	Quantidade de docentes em exercício com mestrado	
59	QT_DOC_EX_DOUT	Quantidade de docentes em exercício com doutorado	
60	QT_DOC_EX_INT	Quantidade de docentes em exercício em tempo integral	
61	QT_DOC_EX_INT_DE	Quantidade de docentes em exercício em tempo integral com dedicação exclusiva	
62	QT_DOC_EX_INT_SEM_DE	Quantidade de docentes em exercício em tempo integral sem dedicação exclusiva	
63	QT_DOC_EX_PARC	Quantidade de docentes em exercício em tempo parcial	
64	QT_DOC_EX_HOR	Quantidade de docentes em exercício horista	
65	QT_DOC_EX_0_29	Quantidade de docentes em exercício - até 29 anos	
66	QT_DOC_EX_30_34	Quantidade de docentes em exercício - de 30 a 34 anos	
67	QT_DOC_EX_35_39	Quantidade de docentes em exercício - de 35 a 39 anos	

(conclusão)

#	Nome da Variável	Descrição da Variável	Categoria
68	QT_DOC_EX_40_44	Quantidade de docentes em exercício - de 40 a 44 anos	
69	QT_DOC_EX_45_49	Quantidade de docentes em exercício - de 45 a 49 anos	
70	QT_DOC_EX_50_54	Quantidade de docentes em exercício - de 50 a 54 anos	
71	QT_DOC_EX_55_59	Quantidade de docentes em exercício - de 55 a 59 anos	
72	QT_DOC_EX_60 MAIS	Quantidade de docentes em exercício - de 60 anos ou mais	
73	QT_DOC_EX_BRANCA	Quantidade de docentes em exercício - Cor/Raça branca	
74	QT_DOC_EX_PRETA	Quantidade de docentes em exercício - Cor/Raça preta	
75	QT_DOC_EX_PARDA	Quantidade de docentes em exercício - Cor/Raça parda	
76	QT_DOC_EX_AMARELA	Quantidade de docentes em exercício - Cor/Raça amarela	
77	QT_DOC_EX_INDIGENA	Quantidade de docentes em exercício - Cor/Raça indígena	
78	QT_DOC_EX_COR_ND	Quantidade de docentes em exercício - Cor/Raça não dispõe da informação ou não declarada	
79	QT_DOC_EX_BRA	Quantidade de docentes em exercício - nacionalidade brasileira ou brasileiro nascido no exterior ou naturalizado	
80	QT_DOC_EX_EST	Quantidade de docentes em exercício - nacionalidade estrangeira	
81	QT_DOC_EX_COM_DEFICIENCIA	Quantidade de docentes em exercício - com deficiência	

Fonte: Reprodução direta de Inep (2022)

Quadro 12 - Variáveis do Censo da Educação Superior – Cursos

#	Nome da Variável	Descrição da Variável	Categoria
1	NU_ANO_CENSO	Ano de referência do Censo da Educação Superior	
2	NO_REGIAO	Nome da região geográfica do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
3	CO_REGIAO	Código da região geográfica	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
4	NO_UF	Nome da Unidade da Federação do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
5	SG_UF	Sigla da Unidade da Federação do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
6	CO_UF	Código do IBGE da Unidade da Federação do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
7	NO_MUNICIPIO	Nome do Município do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
8	CO_MUNICIPIO	Código do IBGE do Município do local de oferta do curso	(.) Dados de Cursos a Distância não calculados para esta dimensão geográfica e/ou Cursos a distância oferecidos por instituição de ensino brasileira no exterior
9	IN_CAPITAL	Informa se o local de oferta do curso está localizado em capital de estado	0. Não 1. Sim (.) Não aplicável (Curso a distância)
10	TP_ORGANIZACAO_ACADEMICA	Tipo da Organização Acadêmica da IES	1. Universidade 2. Centro Universitário 3. Faculdade 4. Instituto Federal de Educação, Ciência e Tecnologia 5. Centro Federal de Educação Tecnológica
11	TP_CATEGORIA_ADMINISTRATIVA	Tipo da Categoria Administrativa da IES	1. Pública Federal 2. Pública Estadual 3. Pública Municipal 4. Privada com fins lucrativos 5. Privada sem fins lucrativos 6. Privada - Particular em sentido estrito 7. Especial 8. Privada comunitária 9. Privada confessional
12	TP_REDE	Rede de Ensino	1. Pública 2. Privada



(continua)

#	Nome da Variável	Descrição da Variável	Categoria
13	CO_IES	Código da Instituição	
14	NO_CINE_ROTULO	Nome do curso, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
15	CO_CINE_ROTULO	Código de identificação do curso, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
16	CO_CINE_AREA_GERAL	Código de identificação da área geral, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
17	NO_CINE_AREA_GERAL	Nome da área geral, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
18	CO_CINE_AREA_ESPECIFICA	Código de identificação da área específica, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
19	NO_CINE_AREA_ESPECIFICA	Nome da área específica, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
20	CO_CINE_AREA_DETALHADA	Código de identificação da área detalhada, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
21	NO_CINE_AREA_DETALHADA	Nome da área detalhada, conforme adaptação da Classificação Internacional Normalizada da Educação Cine/Unesco	
22	TP_GRAU_ACADEMICO	Tipo do grau acadêmico conferido ao aluno pela conclusão dos requisitos exigidos pelo curso	1. Bacharelado 2. Licenciatura 3. Tecnológico 4. Bacharelado e Licenciatura (.) Não aplicável (cursos com nível acadêmico igual a sequencial de formação específica ou cursos de área básica de curso identificada pela variável TP_ATRIBUTO_INGRESSO)
23	TP_MODALIDADE_ENSINO	Tipo da modalidade de ensino do curso	1. Presencial 2. Curso a distância
24	TP_NIVEL_ACADEMICO	Tipo do nível acadêmico do curso	1. Graduação 2. Sequencial de Formação Específica

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
2 5	QT_CURSO	Número de cursos	
2 6	QT_VG_TOTAL	Quantidade total de vagas oferecidas nos cursos	
2 7	QT_VG_TOTAL_DIURNO	Quantidade total de vagas oferecidas nos cursos no turno diurno	(.) Não aplicável (Curso a distância)
2 8	QT_VG_TOTAL_NOTURNO	Quantidade total de vagas oferecidas nos cursos no turno noturno	(.) Não aplicável (Curso a distância)
2 9	QT_VG_TOTAL_EAD	Quantidade total de vagas oferecidas nos cursos EAD	
3 0	QT_VG_NOVA	Quantidade de vagas novas oferecidas nos cursos	
3 1	QT_VG_PROC_SELETIVO	Quantidade de vagas nos processos seletivos oferecidos nos cursos	
3 2	QT_VG_REMANESC	Quantidade de vagas remanescentes oferecidas nos cursos	
3 3	QT_VG_PROG_ESPECIAL	Quantidade de vagas oferecidas para programas especiais	
3 4	QT_INSCRITO_TOTAL	Quantidade total de inscritos nos cursos	
3 5	QT_INSCRITO_TOTAL_DIURNO	Quantidade total de inscritos nos cursos no turno diurno	(.) Não aplicável (Curso a distância)
3 6	QT_INSCRITO_TOTAL_NOTURNO	Quantidade total de inscritos nos cursos no turno noturno	(.) Não aplicável (Curso a distância)
3 7	QT_INSCRITO_TOTAL_EAD	Quantidade total de inscritos nos cursos EAD	
3 8	QT_INSC_VG_NOVA	Quantidade de inscritos no processo seletivo de vagas novas	
3 9	QT_INSC_PROC_SELETIVO	Quantidade de inscritos nos processos seletivos oferecidos nos cursos	
4 0	QT_INSC_VG_REMANESC	Quantidade de inscritos no processo seletivo de vagas remanescentes	
4 1	QT_INSC_VG_PROG_ESPECIAL	Quantidade de inscritos no processo seletivo de vagas oferecidas para programas especiais	
4 2	QT_ING	Quantidade de ingressantes nos cursos	Cálculo de ingressante (soma do número de alunos com data de ingresso de 01 de janeiro e 01 de julho do ano de referência do censo)

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
4 3	QT_ING_FEM	Quantidade de ingressantes do sexo feminino nos cursos	
4 4	QT_ING_MASC	Quantidade de ingressantes do sexo masculino nos cursos	
4 5	QT_ING_DIURNO	Quantidade de ingressantes nos cursos no turno diurno	(.) Não aplicável (Curso a distância)
4 6	QT_ING_NOTURNO	Quantidade de ingressantes nos cursos no turno noturno	(.) Não aplicável (Curso a distância)
4 7	QT_ING_VG_NOVA	Quantidade de ingressantes em vagas novas oferecidas nos cursos	
4 8	QT_ING_VESTIBULAR	Quantidade de ingressantes nos cursos por vestibular	
4 9	QT_ING_ENEM	Quantidade de ingressantes nos cursos pelo ENEM	
5 0	QT_ING_AVALIACAO_S ERIADA	Quantidade de ingressantes nos cursos por meio da avaliação seriada	
5 1	QT_ING_SELECAO_SIM PLIFICA	Quantidade de ingressantes nos cursos por meio de seleção simplificada	
5 2	QT_ING_EGR	Quantidade de alunos que se formaram em um curso de Bacharelado Interdisciplinar ou Licenciatura Interdisciplinar e que ingressaram em um curso de terminalidade	
5 3	QT_ING_OUTRO_TIPO_ SELECAO	Quantidade de ingressantes em outro tipo de seleção para os cursos	
5 4	QT_ING_PROC_SELETI VO	Quantidade de ingressantes nos processos seletivos oferecidos nos cursos	
5 5	QT_ING_VG_REMANES C	Quantidade de ingressantes nos cursos por meio de vagas remanescentes	
5 6	QT_ING_VG_PROG_ES PECIAL	Quantidade de ingressantes nos cursos por meio de vagas para programas especiais	
5 7	QT_ING_OUTRA_FORM A	Quantidade de ingressantes nos cursos por outras formas de ingresso	
5 8	QT_ING_0_17	Quantidade de ingressantes nos cursos - Até 17 anos de idade	
5 9	QT_ING_18_24	Quantidade de ingressantes nos cursos - De 18 a 24 anos de idade	
6 0	QT_ING_25_29	Quantidade de ingressantes nos cursos - De 25 a 29 anos de idade	
6 1	QT_ING_30_34	Quantidade de ingressantes nos cursos - De 30 a 34 anos de idade	
6 2	QT_ING_35_39	Quantidade de ingressantes nos cursos - De 35 a 39 anos de idade	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
6 3	QT_ING_40_49	Quantidade de ingressantes nos cursos - De 40 a 49 anos de idade	
6 4	QT_ING_50_59	Quantidade de ingressantes nos cursos - De 50 a 59 anos de idade	
6 5	QT_ING_60 MAIS	Quantidade de ingressantes nos cursos - De 60 ou mais anos de idade	
6 6	QT_ING_BRANCA	Quantidade de ingressantes nos cursos - Cor/Raça Branca	
6 7	QT_ING_PRETA	Quantidade de ingressantes nos cursos - Cor/Raça Preta	
6 8	QT_ING_PARDA	Quantidade de ingressantes nos cursos - Cor/Raça Parda	
6 9	QT_ING_AMARELA	Quantidade de ingressantes nos cursos - Cor/Raça Amarela	
7 0	QT_ING_INDIGENA	Quantidade de ingressantes nos cursos - Cor/Raça Indígena	
7 1	QT_ING_CORND	Quantidade de ingressantes nos cursos - Cor/Raça Não Declarada	
7 2	QT_MAT	Quantidade de matrículas	Cálculo de matrícula (soma do número de alunos com situação de vínculo ao curso igual a: 2. Cursando e 6. Formado)
7 3	QT_MAT_FEM	Quantidade de matrículas do sexo feminino	
7 4	QT_MAT_MASC	Quantidade de matrículas do sexo masculino	
7 5	QT_MAT_DIURNO	Quantidade de matrículas - turno diurno	(.) Não aplicável (Curso a distância)
7 6	QT_MAT_NOTURNO	Quantidade de matrículas - turno noturno	(.) Não aplicável (Curso a distância)
7 7	QT_MAT_0_17	Quantidade de matrículas - Até 17 anos de idade	
7 8	QT_MAT_18_24	Quantidade de matrículas - De 18 a 24 anos de idade	
7 9	QT_MAT_25_29	Quantidade de matrículas - De 25 a 29 anos de idade	
8 0	QT_MAT_30_34	Quantidade de matrículas - De 30 a 34 anos de idade	
8 1	QT_MAT_35_39	Quantidade de matrículas - De 35 a 39 anos de idade	
8 2	QT_MAT_40_49	Quantidade de matrículas - De 40 a 49 anos de idade	
8 3	QT_MAT_50_59	Quantidade de matrículas - De 50 a 59 anos de idade	
8 4	QT_MAT_60 MAIS	Quantidade de matrículas - De 60 ou mais anos de idade	
8 5	QT_MAT_BRANCA	Quantidade de matrículas - Cor/Raça Branca	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
86	QT_MAT_PRETA	Quantidade de matrículas - Cor/Raça Preta	
87	QT_MAT_PARDA	Quantidade de matrículas - Cor/Raça Parda	
88	QT_MAT_AMARELA	Quantidade de matrículas - Cor/Raça Amarela	
89	QT_MAT_INDIGENA	Quantidade de matrículas - Cor/Raça Indígena	
90	QT_MAT_CORND	Quantidade de matrículas - Cor/Raça Não Declarada	
91	QT_CONC	Quantidade de concluintes	
92	QT_CONC_FEM	Quantidade de concluintes do sexo feminino	
93	QT_CONC_MASC	Quantidade de concluintes do sexo masculino	
94	QT_CONC_DIURNO	Quantidade de concluintes - turno diurno	(.) Não aplicável (Curso a distância)
95	QT_CONC_NOTURNO	Quantidade de concluintes - turno noturno	(.) Não aplicável (Curso a distância)
96	QT_CONC_0_17	Quantidade de concluintes - Até 17 anos de idade	
97	QT_CONC_18_24	Quantidade de concluintes - De 18 a 24 anos de idade	
98	QT_CONC_25_29	Quantidade de concluintes - De 25 a 29 anos de idade	
99	QT_CONC_30_34	Quantidade de concluintes - De 30 a 34 anos de idade	
100	QT_CONC_35_39	Quantidade de concluintes - De 35 a 39 anos de idade	
101	QT_CONC_40_49	Quantidade de concluintes - De 40 a 49 anos de idade	
102	QT_CONC_50_59	Quantidade de concluintes - De 50 a 59 anos de idade	
103	QT_CONC_60 MAIS	Quantidade de concluintes - De 60 ou mais anos de idade	
104	QT_CONC_BRANCA	Quantidade de concluintes - Cor/Raça Branca	
105	QT_CONC_PRETA	Quantidade de concluintes - Cor/Raça Preta	
107	QT_CONC_AMARELA	Quantidade de concluintes - Cor/Raça Amarela	
109	QT_CONC_CORND	Quantidade de concluintes - Cor/Raça Não Declarada	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
1 1 0	QT_ING_NACBRAS	Quantidade de ingressantes nos cursos - nacionalidade brasileira ou brasileira - nascido no exterior ou naturalizado	
1 1 1	QT_ING_NACESTRANG	Quantidade de ingressantes nos cursos - nacionalidade estrangeira	
1 1 2	QT_MAT_NACBRAS	Quantidade de matrículas - nacionalidade brasileira ou brasileira - nascido no exterior ou naturalizado	
1 1 3	QT_MAT_NACESTRANG	Quantidade de matrículas - nacionalidade estrangeira	
1 1 4	QT_CONC_NACBRAS	Quantidade de concluintes - nacionalidade brasileira ou brasileira - nascido no exterior ou naturalizado	
1 1 5	QT_CONC_NACESTRANG	Quantidade de concluintes - nacionalidade estrangeira	
1 1 6	QT_ALUNO_DEFICIENTE	Quantidade de alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação	
1 1 7	QT_ING_DEFICIENTE	Quantidade de ingressantes nos cursos - alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação	
1 1 8	QT_MAT_DEFICIENTE	Quantidade de matrículas - alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação	
1 1 9	QT_CONC_DEFICIENTE	Quantidade de concluintes - alunos com deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação	
1 2 0	QT_ING_FINANC	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil	
1 2 1	QT_ING_FINANC_REEMB	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil reembolsável	
1 2 2	QT_ING_FIES	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil reembolsável - FIES	
1 2 3	QT_ING_RPFIES	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil reembolsável - administrado pela instituição	
1 2 4	QT_ING_FINANC_REEMB_OUTROS	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
1 2 6	QT_ING_PROUNII	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Integral	
1 2 7	QT_ING_PROUNIP	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Parcial	
1 2 8	QT_ING_NRPFIES	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil não reembolsável - administrado pela instituição	
1 2 9	QT_ING_FINANC_NREEMB_OUTROS	Quantidade de ingressantes nos cursos - alunos que utilizam financiamento estudantil não reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	
1 3 0	QT_MAT_FINANC	Quantidade de matrículas - alunos que utilizam financiamento estudantil	
1 3 1	QT_MAT_FINANC_REEMB	Quantidade de matrículas - alunos que utilizam financiamento estudantil reembolsável	
1 3 2	QT_MAT_FIES	Quantidade de matrículas - alunos que utilizam financiamento estudantil reembolsável - FIES	
1 3 3	QT_MAT_RPFIES	Quantidade de matrículas - alunos que utilizam financiamento estudantil reembolsável - administrado pela instituição	
1 3 4	QT_MAT_FINANC_REEMB_OUTROS	Quantidade de matrículas - alunos que utilizam financiamento estudantil reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	
1 3 5	QT_MAT_FINANC_NREEMB	Quantidade de matrículas - alunos que utilizam financiamento estudantil não reembolsável	
1 3 6	QT_MAT_PROUNII	Quantidade de matrículas - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Integral	
1 3 7	QT_MAT_PROUNIP	Quantidade de matrículas - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Parcial	
1 3 8	QT_MAT_NRPFIES	Quantidade de matrículas - alunos que utilizam financiamento estudantil não reembolsável - administrado pela instituição	
1 3 9	QT_MAT_FINANC_NREEMB_OUTROS	Quantidade de matrículas - alunos que utilizam financiamento estudantil não reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
1 4 0	QT_CONC_FINANC	Quantidade de concluintes - alunos que utilizam financiamento estudantil	
1 4 1	QT_CONC_FINANC_RE EMB	Quantidade de concluintes - alunos que utilizam financiamento estudantil reembolsável	
1 4 2	QT_CONC_FIES	Quantidade de concluintes - alunos que utilizam financiamento estudantil reembolsável - FIES	
1 4 3	QT_CONC_RPFIES	Quantidade de concluintes - alunos que utilizam financiamento estudantil reembolsável - administrado pela instituição	
1 4 4	QT_CONC_FINANC_RE EMB_OUTROS	Quantidade de concluintes - alunos que utilizam financiamento estudantil reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	
1 4 5	QT_CONC_FINANC_NR EEMB	Quantidade de concluintes - alunos que utilizam financiamento estudantil não reembolsável	
1 4 6	QT_CONC_PROUNII	Quantidade de concluintes - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Integral	
1 4 7	QT_CONC_PROUNIP	Quantidade de concluintes - alunos que utilizam financiamento estudantil não reembolsável - PROUNI Parcial	
1 4 8	QT_CONC_NRPFIES	Quantidade de concluintes - alunos que utilizam financiamento estudantil não reembolsável - administrado pela instituição	
1 4 9	QT_CONC_FINANC_NR EEMB_OUTROS	Quantidade de concluintes - alunos que utilizam financiamento estudantil não reembolsável - pelo governo estadual, municipal, entidades externas à IES e outras entidades	
1 5 0	QT_ING_RESERVA_VA GA	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas	
1 5 1	QT_ING_RVREDEPUBLI CA	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas - escola pública	
1 5 2	QT_ING_RVETNICO	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas - cunho étnico	
1 5 3	QT_ING_RVPDEF	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas - pessoas com deficiência	
1 5 4	QT_ING_RVSOCIAL_RF	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas - cunho social/renda familiar	



(continua)

#	Nome da Variável	Descrição da Variável	Categoria
1 5 5	QT_ING_RVOUTROS	Quantidade de ingressantes nos cursos - alunos que participam do programa de reserva de vagas - outros tipos	
1 5 6	QT_MAT_RESERVA_VAGA	Quantidade de matrículas - alunos que participam do programa de reserva de vagas	
1 5 7	QT_MAT_RVREDEPUBLICA	Quantidade de matrículas - alunos que participam do programa de reserva de vagas - escola pública	
1 5 8	QT_MAT_RVETNICO	Quantidade de matrículas - alunos que participam do programa de reserva de vagas - cunho étnico	
1 5 9	QT_MAT_RVPDEF	Quantidade de matrículas - alunos que participam do programa de reserva de vagas - pessoas com deficiência	
1 6 0	QT_MAT_RVSOCIAL_RF	Quantidade de matrículas - alunos que participam do programa de reserva de vagas - cunho social/renda familiar	
1 6 1	QT_MAT_RVOUTROS	Quantidade de matrículas - alunos que participam do programa de reserva de vagas - outros tipos	
1 6 2	QT_CONC_RESERVA_VAGA	Quantidade de concluintes - alunos que participam do programa de reserva de vagas	
1 6 3	QT_CONC_RVREDEPUBLICA	Quantidade de concluintes - alunos que participam do programa de reserva de vagas - escola pública	
1 6 4	QT_CONC_RVETNICO	Quantidade de concluintes - alunos que participam do programa de reserva de vagas - cunho étnico	
1 6 5	QT_CONC_RVPDEF	Quantidade de concluintes - alunos que participam do programa de reserva de vagas - pessoas com deficiência	
1 6 6	QT_CONC_RVSOCIAL_RF	Quantidade de concluintes - alunos que participam do programa de reserva de vagas - cunho social/renda familiar	
1 6 7	QT_CONC_RVOUTROS	Quantidade de concluintes - alunos que participam do programa de reserva de vagas - outros tipos	
1 6 8	QT_SIT_TRANCADA	Quantidade de alunos com situação de vínculo no curso - matrícula trancada	
1 6 9	QT_SIT_DESVINCULADO	Quantidade de alunos com situação de vínculo no curso - aluno desvinculado do curso	
1 7 0	QT_SIT_TRANSFERIDO	Quantidade de alunos com situação de vínculo no curso - aluno transferido para outro curso da mesma IES	

(continua)

#	Nome da Variável	Descrição da Variável	Categoria
171	QT_SIT_FALECIDO	Quantidade de alunos com situação de vínculo no curso - aluno falecido	
172	QT_ING_PROCESPUBLICA	Quantidade de ingressantes que terminaram o ensino médio em escolas públicas	
173	QT_ING_PROCESPRIVADA	Quantidade de ingressantes que terminaram o ensino médio em escolas privadas	
174	QT_ING_PROCNAINFORMADA	Quantidade de ingressantes que não informaram o tipo de escola que terminaram o ensino médio	
175	QT_MAT_PROCESPUBLICA	Quantidade de matrículas de alunos que terminaram o ensino médio em escolas públicas	
176	QT_MAT_PROCESPRIVADA	Quantidade de matrículas de alunos que terminaram o ensino médio em escolas privadas	
177	QT_MAT_PROCNAINFORMADA	Quantidade de matrículas de alunos que não informaram o tipo de escola que terminaram o ensino médio	
178	QT_CONC_PROCESPUBLICA	Quantidade de concluintes que terminaram o ensino médio em escolas públicas	
179	QT_CONC_PROCESPRIVADA	Quantidade de concluintes que terminaram o ensino médio em escolas privadas	
180	QT_CONC_PROCNAINFORMADA	Quantidade de concluintes que não informaram o tipo de escola que terminaram o ensino médio	
181	QT_PARFOR	Quantidade de alunos que participam do programa especial para a formação de professores em exercício na rede pública de educação básica (PARFOR)	
182	QT_ING_PARFOR	Quantidade de ingressantes que participam do programa especial para a formação de professores em exercício na rede pública de educação básica (PARFOR)	
183	QT_MAT_PARFOR	Quantidade de matrículas de alunos que participam do programa especial para a formação de professores em exercício na rede pública de educação básica (PARFOR)	
184	QT_CONC_PARFOR	Quantidade de concluintes que participam do programa especial para a formação de professores em exercício na rede pública de educação básica (PARFOR)	
185	QT_APOIO_SOCIAL	Quantidade de alunos que recebem algum tipo de apoio social	

(conclusão)

#	Nome da Variável	Descrição da Variável	Categoria
186	QT_ING_APOIO_SOCIAL	Quantidade de ingressantes que recebem algum tipo de apoio social	
187	QT_MAT_APOIO_SOCIAL	Quantidade de matrículas de alunos que recebem algum tipo de apoio social	
188	QT_CONC_APOIO_SOCIAL	Quantidade de concluintes que recebem algum tipo de apoio social	
189	QT_ATIV_EXTRACURRICULAR	Quantidade de alunos que participam de algum tipo de atividade extracurricular	
190	QT_ING_ATIV_EXTRACURRICULAR	Quantidade de ingressantes que participam de algum tipo de atividade extracurricular	
191	QT_MAT_ATIV_EXTRACURRICULAR	Quantidade de matrículas de alunos que participam de algum tipo de atividade extracurricular	
192	QT_CONC_ATIV_EXTRACURRICULAR	Quantidade de concluintes que participam de algum tipo de atividade extracurricular	
193	QT_MOB_ACADEMICA	Quantidade de alunos de curso de graduação que se vincularam temporariamente a outra instituição, sendo ela nacional ou internacional	
194	QT_ING_MOB_ACADEMICA	Quantidade de alunos ingressantes em curso de graduação que se vincularam temporariamente a outra instituição, sendo ela nacional ou internacional	
195	QT_MAT_MOB_ACADEMICA	Quantidade de alunos que estão regularmente matriculados em cursos de graduação, que se vincularam temporariamente a outra instituição, sendo ela nacional ou internacional	
196	QT_CONC_MOB_ACADEMICA	Quantidade de concluintes em cursos de graduação que se vincularam temporariamente a outra instituição, sendo ela nacional ou internacional	

Fonte: Reprodução direta de Inep (2022)

**Quadro 12 - Variáveis do Censo da Educação Superior - Financeiros (Sala Segura)**

#	NOME DA VARIÁVEL	DESCRIÇÃO DA VARIÁVEL	DESCRIÇÃO DAS CATEGORIAS
1	NU_ANO_CENSO	Ano de referência do Censo da Educação Superior	
2	CO_IES	Código único de identificação da IES	
3	NO_IES	Nome da IES	
4	SG_IES	Sigla da IES	
5	CO_MANTENEDORA	Código único de identificação da mantenedora	
6	NO_MANTENEDORA	Nome da mantenedora	
38	TP_REFERENTE	Tipo de entidade à qual é referente o dado financeiro	1. Mantenedora 2. Instituição
39	VL_RECEITA_PROPRIA	Valor das receitas próprias auferidas pela Mantenedora ou pela IES no ano de referência	
40	VL_RECEITA_TRANSFERENCIA	Valor de transferências auferidas pela Mantenedora ou pela IES no ano de referência	
41	VL_RECEITA_OUTRA	Valor de outras receitas auferidas pela Mantenedora ou pela IES no ano de referência	
42	VL_DESPESA_PESSOAL_DOCENTE	Valor das despesas com remuneração de pessoal - docentes ativos da IES ou da mantenedora	
43	VL_DESPESA_PESSOAL_TECNICO	Valor das despesas com remuneração de pessoal técnico-administrativo ativo/pedagógicos da IES ou da mantenedora	
44	VL_DESPESA_PESSOAL_ENCARGO	Valor das despesas com benefícios e encargos sociais de todo o pessoal da IES ou da mantenedora	
45	VL_DESPESA_CUSTEIO	Valor das despesas para a manutenção e custeio da IES ou da mantenedora (Não inclui despesas com pessoal.)	
46	VL_DESPESA_INVESTIMENTO	Valor das despesas de investimentos (despesas de capital) realizadas nas IES ou na mantenedora	
47	VL_DESPESA_PESQUISA	Valor das despesas com Pesquisa e Desenvolvimento da IES ou da mantenedora	
48	VL_DESPESA_OUTRA	Valor de dispêndios não contemplados nos demais campos relativos à despesa	

Fonte: Adaptação pelo autor a partir do dicionário dos microdados (INEP, 2022)

## **ANEXO B – Fontes dos Algoritmos Utilizados - R**

Com o intuito de permitir a reprodução destes resultados, o código fonte utilizado nas extrações, transformações e geração dos resultados deste trabalho está disponível em [http://rpubs.com/maf27br/Proj\\_Mestrado\\_IES](http://rpubs.com/maf27br/Proj_Mestrado_IES)