



**UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

DANIEL MIRANDA ALVES

**IMPACTO DO MESTRADO EM COMPUTAÇÃO NA REMUNERAÇÃO DOS
EGRESSOS: UMA ANÁLISE EXPLORATÓRIA E CAUSAL**

**RIO DE JANEIRO
2024**



DANIEL MIRANDA ALVES

IMPACTO DO MESTRADO EM COMPUTAÇÃO NA REMUNERAÇÃO DOS
EGRESSOS: UMA ANÁLISE EXPLORATÓRIA E CAUSAL

Dissertação apresentada como requisito parcial
para obtenção do título de mestre pelo Programa
de Pós-Graduação em Informática da
Universidade Federal do Estado do Rio de Janeiro
(UNIRIO).

Orientador: Carlos Eduardo Ribeiro de Mello

RIO DE JANEIRO
2024

Catálogo informatizada pelo(a) autor(a)

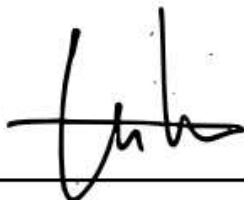
- A474 Alves, Daniel Miranda
Impacto do Mestrado em Computação na remuneração dos egressos: uma análise exploratória e causal / Daniel Miranda Alves. -- Rio de Janeiro, 2024.
73 f.
- Orientador: Carlos Eduardo Ribeiro de Mello.
Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Informática, 2024.
1. Análise Exploratória de Dados. 2. Análise Causal. 3. Mestrado em Computação. I. de Mello, Carlos Eduardo Ribeiro, orient. II. Título.

DANIEL MIRANDA ALVES

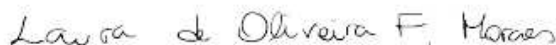
**IMPACTO DO MESTRADO EM COMPUTAÇÃO NA REMUNERAÇÃO DOS
EGRESSOS: UMA ANÁLISE EXPLORATÓRIA E CAUSAL**

Dissertação apresentada como requisito parcial
para obtenção do título de Mestre pelo Programa
de Pós-Graduação em Informática da
Universidade Federal do Estado do Rio de Janeiro
(UNIRIO)

Aprovada por:



Dr. Carlos Eduardo Ribeiro de Mello (Orientador)
Universidade Federal do Estado do Rio de Janeiro - UNIRIO



Dr.ª Laura de Oliveira Fernandes Moraes
Universidade Federal do Estado do Rio de Janeiro - UNIRIO



Dr.ª Fernanda De Negri
Instituto de Pesquisa Econômica Aplicada - IPEA

RIO DE JANEIRO
2024

AGRADECIMENTOS

Gostaria de agradecer a todos, incluindo você, que está lendo este trabalho agora. Todo o esforço que tive para completar o Mestrado, ao longo de tantas coisas que aconteceram em minha vida nesse meio tempo, valeu a pena. Eu faria essa caminhada mil vezes de novo, se tivesse a chance. A todos que me apoiaram em cada momento em que precisei de força para chegar ao final dessa trajetória: sem vocês, eu não conseguiria!

À minha esposa, Mariana, obrigado por nunca ter soltado minha mão e por acreditar tanto em mim. Me sinto feliz e orgulhoso por termos conquistado isso juntos! Com certeza, é mais uma vitória das muitas que vamos celebrar ao longo de nossas vidas. Meu combustível sempre foi te ver sorrindo e dizendo que sabia que eu conseguiria. Nunca vou esquecer tudo o que você fez para me apoiar.

Aos meus pais e minha irmã, Alcy, Maria Helena e Raphaella, obrigado por acreditarem em cada sonho que tive na minha vida. Vocês, mais do que ninguém, sabem o quanto desejei este momento. Vocês fizeram parte disso. Que sorte eu tenho de ter meus melhores amigos na minha família.

À minha avó, Edith, tão presente na minha criação e formação de vida. Você é a minha maior inspiração. Me lembro de você me levando para a escola desde pequeno, reforçando que eu até poderia jogar bola, mas que nunca deveria parar de estudar. Aqui estou eu, vó: estudando e me aperfeiçoando. Não existem palavras para expressar minha gratidão.

Ao Cadu, meu orientador, que em nenhum momento deixou de acreditar que eu chegaria ao final. Cadu, obrigado por me inspirar e por mostrar que o papel do professor vai muito além de passar conteúdo. Seu incentivo e paciência comigo foram simplesmente impressionantes. Vou lembrar dessa nossa caminhada para sempre, com muito carinho. Obrigado!

À UNIRIO, que me acolheu tanto na graduação quanto na pós-graduação. Quero usar este trecho para reconhecer que vocês, professores e servidores, construíram muito de mim. Quando entrei aqui pela primeira vez, apenas um jovem de 18 anos, jamais imaginaria que estaria vivendo um Mestrado tantos anos depois! Como pode o mesmo lugar me trazer tantas memórias boas?

Aos meus amigos, por cada ligação perguntando como estava indo o trabalho. Prometo uma comemoração merecida após a conclusão desta etapa acadêmica! Obrigado pela paciência!

Às pessoas que não podem mais estar comigo porque já se foram: eu lembro de vocês todos os dias, e sei que estão num lugar especial sorrindo e celebrando isso.

Aos membros da minha banca por disporem do seu tempo para avaliar e contribuir com o trabalho. É uma grande honra ter a chance de receber as críticas de vocês para acrescentar à dissertação.

Finalmente, quero encerrar os agradecimentos dividindo o meu sentimento agora, ao escrever este parágrafo. É muito bom poder ser grato a tantas pessoas. É maravilhoso chegar ao final deste texto sabendo que poderia escrever páginas e páginas sobre quantas pessoas importantes participaram disso. Só consigo me considerar alguém de muita sorte por ter tanta gente incrível na minha vida. Obrigado, Deus, pela vida que eu tenho. Que sorte!

ALVES, Daniel Miranda. **Impacto do Mestrado em Computação na remuneração dos egressos: uma análise exploratória e causal**. UNIRIO, 2024. 73 f. Dissertação (Mestrado em Informática) - Departamento de Informática Aplicada, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

RESUMO

A pós-graduação é amplamente reconhecida por proporcionar aos discentes habilidades avançadas e especializadas que são essenciais para a evolução da sociedade. No entanto, permanece a questão de até que ponto essas habilidades adquiridas se traduzem em um aumento significativo na remuneração dos egressos. Neste estudo, foi realizada uma análise exploratória detalhada utilizando dados coletados entre 2015 e 2019, abrangendo os egressos do Mestrado em Computação no Brasil. A análise exploratória utilizou dados individualizados da Relação Anual de Informações Sociais (RAIS) e dados públicos disponibilizados pela CAPES. Com isso, foram identificadas diferenças entre os discentes que concluíram a formação e aqueles que não conseguiram, explorando características como etnia, gênero, região, tipo de vínculo empregatício, a avaliação do programa de pós-graduação e outras. Os resultados da EDA geraram hipóteses, levantando discussões sobre as possíveis causas das disparidades observadas. Em seguida, foi realizada uma análise causal focada nas turmas de 2015 e 2016, utilizando o método de matching por *Coarsened Exact Matching* (CEM). Essa abordagem permitiu que fosse estimado o impacto do título de mestre na remuneração dos egressos, concluindo-se que, após cinco anos, houve um aumento médio de 18% na remuneração. Contudo, esse aumento foi principalmente impulsionado pelos funcionários públicos que concluíram o mestrado, o que levanta discussões sobre a integração da academia com o setor privado. As evidências apresentadas neste trabalho contribuem para o debate sobre o papel da pós-graduação na evolução remuneratória e sugerem uma reflexão sobre a relação entre a formação acadêmica e as demandas do mercado de trabalho, especificamente na indústria privada.

Palavras-chave: Análise Exploratória de Dados. Inferência Causal. Matching. Mestrado. Computação.

ALVES, Daniel Miranda. **Impact of the Master 's Degree in Computer Science on Graduate Salaries: An Exploratory and Causal Analysis**. UNIRIO, 2024. 73 f. Dissertação (Mestrado em Informática) - Departamento de Informática Aplicada, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

ABSTRACT

Postgraduate education is widely recognized for providing students with advanced and specialized skills essential for society. However, the extent to which these acquired skills translate into significant increases in graduates' compensation remains a question. This study conducted a detailed exploratory analysis using data collected between 2015 and 2019, focusing on graduates of the Master's in Computing programs in Brazil. The exploratory data analysis (EDA) utilized individualized data from the *Relação Anual de Informações Sociais* (RAIS) and public data provided by CAPES. The analysis identified differences between students who completed their degrees and those who did not, exploring characteristics such as ethnicity, gender, region, type of employment, program evaluation, and others. The EDA results generated hypotheses, sparking discussions on the possible causes of the observed disparities. Subsequently, a causal analysis was conducted, focusing on the 2015 and 2016 cohorts, using the *Coarsened Exact Matching* (CEM) method. This approach allowed for estimating the impact of obtaining a master's degree on graduates' compensation, concluding that after five years, there was an average compensation increase of 18%. However, this increase was primarily driven by public sector employees who completed the master's degree, raising discussions about the integration of academia with the private sector. The evidence presented in this work contributes to the debate on the role of postgraduate education in compensation progression and suggests a reflection on the relationship between academic training and the demands of the labor market, specifically in the private industry.

Keywords: Exploratory Data Analysis. Causal Inference. Matching. Master's Degree. Computing.

LISTA DE ILUSTRAÇÕES

Figura 1. Fluxo geral de dados da metodologia.....	25
Figura 2. Divisão da base acadêmica tratada de Pós-Graduação em Computação.....	27
Figura 3. Divisão da base acadêmica tratada de Pós-Graduação em Computação.....	28
Figura 4. Evolução absoluta de discentes de pós-graduação em computação de acordo com o grau acadêmico.....	34
Figura 5. Evolução de discentes de mestrado por região.....	35
Figura 6. Distribuição de programas de pós graduação em informática em 2019.....	36
Figura 7. Distribuição da idade no momento da matrícula.....	36
Figura 8. Correlação entre idade e salário na matrícula para discentes de 2019 empregados na área de Informática.....	37
Figura 9. Titulados por grupo de conceito de avaliação do programa.....	38
Figura 10. Titulados de conceitos avaliados com nota seis ou sete por região.....	38
Figura 11. Distribuição da idade de titulados de acordo com o grupo de conceito de avaliação do programa.....	39
Figura 12. Percentual de discentes com emprego no momento da matrícula por ano e situação.....	40
Figura 13. Remuneração média dos discentes empregados no momento da matrícula.....	40
Figura 14. Proporção de titulados com emprego no momento da matrícula por grupo de conceito avaliativo e ano da matrícula.....	41
Figura 15. Matriculados empregados de acordo com o tipo de organização no momento da matrícula por grupo de conceito.....	42
Figura 16. Proporção de funcionários públicos no momento da matrícula por grupo de conceito.....	43
Figura 17. Taxa de conclusão de acordo com a natureza jurídica do emprego.....	43
Figura 18. Remuneração média no momento da matrícula de acordo com perfil do empregador.....	44
Figura 19. Média de idade do funcionário da iniciativa pública e privada no momento da matrícula.....	45
Figura 20. Distribuição da etnia por ano de matrícula.....	45
Figura 21. Média remuneratória por etnia dos empregados no momento da matrícula.....	46
Figura 22. Discentes por gênero.....	47
Figura 23. Média remuneratória na matrícula por gênero.....	47
Figura 24. Quantidade de discentes por área de atuação.....	48
Figura 25. Distribuição da remuneração dos discentes empregados com ocupações de informática.....	49
Figura 26. Média remuneratória por natureza jurídica para discentes matriculados em 2015 e 2016 com remuneração na matrícula.....	49
Figura 27. Média remuneratória por natureza jurídica e situação final do discente.....	50
Figura 28. Filtros para a seleção do grupo da análise causal.....	51
Figura 29. Histograma da diferença média de remuneração no momento da matrícula após mil pareamentos e sem controle por variáveis confundidoras.....	52
Figura 30. Histograma da diferença média de remuneração no momento da matrícula após mil pareamentos e com controle por variáveis confundidoras através do coarsened exact matching.....	54
Figura 31. Histograma da diferença média de remuneração cinco anos depois da matrícula após mil pareamentos e com controle por variáveis confundidoras através do coarsened exact matching.....	54
Figura 32. QQ plot da distribuição de médias de remuneração após cinco anos da matrícula e com 1000 processos de pareamentos.....	55
Figura 33. Proporção de pares no estudo causal de acordo com a natureza jurídica por grupo de conceito de programa.....	57
Figura 34. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula.....	60
Figura 35. Histograma da diferença média de remuneração cinco anos depois da matrícula para	

discentes sem remuneração na matrícula que viraram funcionários públicos..... 60

Figura 36. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula que trabalharam para entidades empresariais..... 61

Figura 37. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula que trabalharam para entidades empresariais..... 61

LISTA DE TABELAS

Tabela 1. Detalhamento do fluxo de dados da metodologia.....	26
Tabela 2. Antes e depois do processo de pareamento com coarsened exact matching.....	53
Tabela 3. ATT por natureza jurídica do empregador.....	56
Tabela 4. ATT por grupo de conceito do programa.....	56
Tabela 5. ATT por faixa etária na matrícula.....	58
Tabela 6. ATT por grupo remuneratório no momento da matrícula.....	58
Tabela 7. ATT por região.....	59

LISTA DE ABREVIATURAS E SIGLAS

ATE - *Average Treatment Effect*

ATT - *Average Treatment Effect on Treated*

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

EDA - *Exploratory Data Analysis*

FIES - Fundo de Financiamento Estudantil

IPCA - Índice Nacional de Preços ao Consumidor Amplo

PROUNI - Programa Universidade para Todos

RAIS - Relação Anual de Informações Sociais

RCT - *Randomized controlled trial*

REUNI - Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais

TE - *Treatment Effect*

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 Motivação.....	14
1.2 Objetivos.....	15
1.3 Organização e estrutura do trabalho.....	15
2 TRABALHOS RELACIONADOS.....	17
3 FUNDAMENTAÇÃO METODOLÓGICA.....	20
3.1 Análise Exploratória Descritiva.....	20
3.2 Inferência causal.....	21
3.3 Matching.....	23
3.4 Coarsened exact matching (CEM).....	24
4 METODOLOGIA.....	25
4.1 Coleta e tratamento de dados.....	26
4.2 Abordagem na análise exploratória de dados.....	28
4.3 Abordagem na inferência causal.....	29
4.3.1 <i>Composição dos grupos de controle e tratamento</i>	30
4.3.2 <i>Matching</i>	30
4.3.3 <i>Desfecho</i>	32
5 ANÁLISE EXPLORATÓRIA DE DADOS.....	34
6 ANÁLISE CAUSAL.....	51
7 CONCLUSÕES.....	63
7.1 Contribuições.....	66
7.2 Limitações.....	66
7.3 Trabalhos futuros.....	67
REFERÊNCIAS.....	69

1 INTRODUÇÃO

A pós-graduação tem um papel fundamental no desenvolvimento de uma nação. Quando iniciada no Brasil, tinha entre seu escopo técnico as funções de formar professores para escalar a quantidade de profissionais altamente qualificados, estimular o desenvolvimento de pesquisa científica e treinar trabalhadores intelectuais para suprir as necessidades do país em todos os setores (CUNHA, 1974). A pós-graduação é reconhecida como um motor essencial para a inovação e a competitividade econômica (MOROSINI, 2009). No contexto brasileiro, vem se destacando também o papel da Pós-Graduação na internacionalização da ciência brasileira, que tem contribuído para a inserção do Brasil em redes globais de pesquisa e inovação, fortalecendo a produção científica e tecnológica do país (BALBACHEVSKY, 2005).

Por essa relevância, diversas políticas públicas foram implementadas ao longo dos anos para incentivar o ensino superior e a Pós-Graduação no Brasil. Iniciativas como o REUNI, PROUNI, e FIES ampliaram o acesso ao ensino superior, enquanto programas como o Ciência sem Fronteiras promoveram a internacionalização e a inovação acadêmica. Além disso, o financiamento de pesquisas por agências como CAPES e CNPq tem sido essencial para o desenvolvimento científico e tecnológico do país (FAVATO & RUIZ, 2018; PROLO et al., 2019).

Todo esse investimento busca suprir a necessidade urgente de formar mão de obra qualificada no Brasil. A lacuna entre a demanda por profissionais altamente capacitados e a oferta disponível reflete as dificuldades enfrentadas no processo de desenvolvimento do país, muitas vezes limitando o potencial de crescimento econômico e a inovação. Em uma economia globalizada, a qualificação profissional é essencial para a competitividade (POWELL & SNELLMAN, 2004) e a Pós-Graduação surge como uma peça fundamental para a superação dessas barreiras, contribuindo não apenas para a formação de especialistas, mas também para o avanço da pesquisa científica e tecnológica.

Essa discussão que mistura a alta necessidade de profissionais qualificados com os altos investimentos realizados no Brasil levanta uma questão crucial: será que essa formação no nível da Pós-Graduação tem realmente um efeito significativo na remuneração dos discentes? Especialmente na área de Computação, que tem

sido uma protagonista na inovação recente, é importante entender como essa formação impacta o mercado de trabalho. Estudos internacionais, como o de Bessen (2019), destacam a transformação do mercado de trabalho pela tecnologia, mas será que a formação a nível de Pós-Graduação dessa área está traduzindo em melhores remunerações no Brasil?

1.1 Motivação

Existe um movimento que critica fortemente a eficácia das universidades em formar profissionais adequados às demandas atuais do mercado, especialmente alegando que muitas instituições de ensino superior estão desconectadas da realidade da indústria (CHRISTENSEN & EYRING, 2011). Na área de tecnologia, essa crítica é amplificada pelo surgimento de inúmeros cursos de curta duração, como os bootcamps de programação, que têm ganhado popularidade por prometerem capacitação rápida e prática num formato alternativo ao tradicional (KAPLAN, 2021).

Em um setor que sofre com a escassez de profissionais qualificados, esses cursos muitas vezes são vistos como uma solução suficiente para preencher as lacunas de conhecimento.

No meio de tantos questionamentos, surge a dúvida sobre o papel do Mestrado: será que, após sua conclusão, os discentes já estão em uma condição melhor em relação àqueles que optaram por outros caminhos? Dado que o esforço investido em uma formação acadêmica costuma ser significativamente maior em comparação com esses cursos alternativos, seria de se esperar um "ganho" proporcional em termos de oportunidades e reconhecimento no mercado de trabalho. No entanto, a questão permanece: esse ganho realmente se materializa na prática?

1.2 Objetivos

Assim, introduzir esse debate já é o suficiente para direcionar a pergunta central da pesquisa:

- Como avaliar o impacto do Mestrado na área de Computação no mercado de trabalho através de dados analíticos?

Ao aprofundar essa investigação, o estudo revelou que existe um ganho significativo em uma subpopulação entre os egressos, apontando para a importância de uma análise segmentada. Para chegar nesse resultado, essa questão geral desmembrou o trabalho em dois objetivos:

- Conduzir uma análise exploratória de dados contendo detalhamento acadêmico e remuneratório para identificar relações, levantar novas hipóteses e debates sobre a evolução dos ganhos dos egressos do curso de Mestrado em áreas da Computação.
- Desenvolver um estudo causal para identificar o efeito do Mestrado na remuneração de seus egressos na área de Computação.

1.3 Organização e estrutura do trabalho

No capítulo 2, será apresentado um conjunto de trabalhos relacionados que buscaram entender o efeito da Pós-Graduação e suas principais características. Esses estudos fornecem uma base sobre a importância da educação avançada na trajetória profissional dos indivíduos, ao mesmo tempo em que revelam algumas lacunas e oportunidades para novas contribuições.

Já no capítulo 3, serão apresentados os fundamentos metodológicos necessários para conduzir tanto a análise exploratória quanto a análise causal da pesquisa. Esse capítulo irá detalhar o método escolhido, explicando suas principais características e justificando sua adequação para responder às questões levantadas no estudo.

O capítulo seguinte será dedicado à explicação da metodologia utilizada na pesquisa. Neste capítulo 4, a coleta dos dados é explicada, assim como eventuais

limitações geradas e como foram contornadas. O destaque fica para as decisões tomadas para a condução do estudo causal.

A análise exploratória dos dados é apresentada no capítulo 5, focando na seleção dos pontos mais relevantes encontrados ao longo do processo de descoberta. Os *insights* obtidos permitiram levantar hipóteses importantes, que não só enriqueceram o entendimento inicial dos dados, mas também alimentaram e prepararam o estudo para a análise causal.

O Capítulo 6 apresenta os resultados da análise causal realizada no estudo. Neste capítulo, são detalhados os resultados do pareamento, que permitiram isolar o efeito específico do Mestrado na área de Computação. O capítulo desdobra esses efeitos em recortes específicos, sugerindo uma subpopulação onde o efeito observado foi significativo, oferecendo assim uma visão mais granular dos impactos do Mestrado.

Finalmente, as conclusões, contribuições, limitações e oportunidades para pesquisas futuras, são apresentadas no capítulo 7. Este capítulo oferece uma oportunidade para entender de forma sucinta o trabalho realizado, destacando os principais achados e sua relevância no contexto do impacto do Mestrado na área de Computação.

2 TRABALHOS RELACIONADOS

Existe um interesse comum na área de Educação em compreender o comportamento do egresso em aspectos remuneratórios, pois esta discussão pode levar instituições a revisarem suas grades curriculares para um alinhamento maior com alguns requisitos da indústria (ALPAYDIN & KÜLTÜR, 2022). Em países vizinhos, como nos Estados Unidos, o acompanhamento da remuneração após universidade é considerado algo comum, como parte do processo de ajuste e melhoria dos cursos, visando o sucesso financeiro dos seus alunos (WITTEVEEN & ATTEWELL, 2021).

No Brasil, esse tipo de pesquisa conta com resultados em algumas áreas, como Enfermagem (SOUZA et. al, 2011), Turismo (SILVEIRA et. al, 2020) e Educação Física (SALLES et. al, 2015). Contudo, costuma-se enfrentar um problema comum: a escassez de dados sobre remunerações dos seus egressos.

Essa barreira traz sérias limitações, tornando-se um entrave para condução de estudos mais profundos. Nos Estados Unidos, onde as instituições têm um acesso facilitado a esses dados remuneratórios, diversos estudos apontam resultados interessantes (DONHARDT, 2004). Algumas pesquisas brasileiras limitaram-se à percepção do ganho salarial por parte do discente, como em Saliba et. al (2012), no ramo da odontologia. Ainda, grande parte dos estudos nacionais sofrem com forte viés de seleção na coleta de faixas salariais, o que nem sempre traz estimativas adequadas. As baixas taxas de respondentes e o viés associado nas respostas acabam sendo um grande obstáculo metodológico, havendo inclusive alguma discussão sobre mecanismos de incentivo para estimular um volume maior de respostas, o que poderiam eventualmente direcionar mais ainda os resultados das análises (ENGSTROM et. al, 2020).

Nesse contexto, os dados disponíveis da Relação Anual de Informações Sociais - RAIS conseguem preencher uma parte significativa desta lacuna, uma vez que, dentre outras informações, contém os dados das relações de trabalhadores de todo o Brasil, contratados sob regime da Consolidação das Leis Trabalhistas (CLT) e estatutários (servidores públicos de todas esferas). Portanto, esta base de dados é potencialmente útil para a condução de um estudo mais preciso e abrangente do acompanhamento da evolução dos rendimentos dos profissionais e egressos da universidade (GERARD et. al, 2021). Apesar disso, o acesso a esses dados requer

acordos específicos com os órgãos de Estado responsáveis, o que pode ser um elemento de dificuldade aos pesquisadores.

Dentre os trabalhos no Brasil com base na RAIS, existe uma série de estudos buscando caracterizar o egresso em relação às características da sua remuneração apenas após sua formação, como, por exemplo, Griboski et. al (2017) e Coelho (2017), no qual se discute os empregos dos egressos da Universidade de Brasília. Outros estudos, voltados especificamente a profissionais de Computação, como Alvares et. al (2020) e Ferreira et. al (2022), implementam essa abordagem de acompanhamento remuneratório para egressos do curso e investigam questões de gênero na área de Computação.

Existem trabalhos que investigam de maneira longitudinal a evolução da remuneração através da estruturação de um estudo em painel. Nessa linha, trabalhos, como Rocha et. al (2016), Oliveira et. al (2019) e Ferreira e De Almeida (2022), abordam informações de renda para estimar o efeito causal de discentes envolvidos em programas sociais após os estudos. Para separar o grupo controle e tratamento, boa parte dos estudos desconsideram fatores e características das instituições. Contudo, segundo Weiss (1995), a instituição carrega, implícita ou explicitamente, ao potencial empregador, informações sobre a produtividade e capacidade do trabalhador em questão. Neste sentido, não considerá-la incorre em algum potencial enviesamento.

Um dos métodos mais observados em trabalhos a partir da RAIS que abordam a intervenção de algum programa social para separar o grupo controle do teste é o método de *propensity score matching*. Apesar de ser a forma mais popular de se trabalhar o pareamento dos grupos, esta dissertação optou por usar o *coarsened exact matching* (CEM) por ser muitas vezes considerado superior às técnicas tradicionais, quando existem um volume de observações suficiente no grupo controle (KING, 2018).

Outros trabalhos já foram conduzidos com objetivo de estimar o efeito da Educação na remuneração em escopos mais amplos, como Barbosa Filho & Pessoa (2008). Esses trabalhos se diferenciam pela fonte da remuneração e pela abordagem centrada nas equações mincerianas – modelo econométrico amplamente usado – porém discussões quanto ao seu uso em contextos dinâmicos da sociedade ainda são necessárias (LEMIEUX, 2006).

Entre essa discussão metodológica e a escassez de alternativas de trabalhos estimando o efeito da Pós-Graduação no Brasil consumindo os dados da RAIS, Lima (2016) fez sua investigação nesse escopo. Seu trabalho apontou um ganho de 9% no Mestrado. Para o pareamento do grupo de controle e teste, o *coarsened exact matching* também foi usado como parte do processo. Por fim, o estudo de Lima (2016) interpreta os resultados de todas as áreas de conhecimento como um todo, deixando uma oportunidade para entender como cada campo de estudo se comporta em relação a esse efeito. É essa lacuna que este trabalho visa contribuir, especificamente na perspectiva do Mestrado em áreas de Computação, não só para o efeito em si, como também para sua caracterização.

3 FUNDAMENTAÇÃO METODOLÓGICA

3.1 Análise Exploratória Descritiva

A análise exploratória de dados, ou *exploratory data analysis* (EDA), é um conjunto de técnicas desenvolvidas para identificar padrões, anomalias, criar e testar hipóteses a partir de conjuntos de dados (HOAGLIN, MOSTELLER & TUKEY, 2000). Apesar dessa definição enfatizar o uso de ferramentas, é relevante ressaltar que a análise exploratória de dados depende em maior parte do conhecimento em relação à área de domínio por conta dos pesquisadores que estão conduzindo o estudo. Provost e Fawcett (2013) dizem que esse é um dos maiores desafios na condução da EDA, já que a intimidade com o problema que está sendo investigado é o que vai garantir que os *insights* sejam acionáveis e úteis. A ideia de permitir que os dados falem por si mesmo enquanto o pesquisador tem a capacidade de interpretar o que está sendo mostrado é o que gera o valor da análise exploratória (VIGNI, DURANTE & COCCHI, 2013).

Toda essa contextualização e discussão, apesar de ser uma das etapas que mais direciona valor para o estudo, é apenas a primeira. Com isso em mãos, a análise exploratória caminha para a coleta e preparação de dados. Na maior parte dos projetos, essa etapa costuma ser a que mais demanda tempo e energia (MYATT, 2007), e é importante garantir que a discussão prévia sobre o estudo seja apenas materializada por variáveis nesse momento, a fim de buscar explorar melhor o assunto. É esse balanceamento entre ter quantidades de observações suficientes e a quantidade de variáveis relevantes referentes aos pontos de investigação que vai garantir um bom fluxo da análise exploratória (CHATFIELD, 1986).

Atualmente, o processamento e armazenamento dos dados tem sofrido uma revolução por conta de sua disponibilidade excessiva em razão das evoluções ao redor do conceito de *big data*, que acaba gerando um diferencial competitivo para as organizações (ARDAGNA, CERAVOLO & DAMIANI, 2016). Ao mesmo tempo que isso permite análises que anteriormente não poderiam ser feitas, também ressalta a importância da definição do problema inicial investigado pela análise exploratória para garantir descobertas relevantes sobre o tema, sem se perder na imensidão de dados disponíveis.

Na investigação do dado em si, um dos recursos cruciais são as visualizações de dados. Esse recurso permite que os pesquisadores identifiquem

rapidamente padrões, relacionamentos e anomalias que muitas vezes não seriam tão óbvios (FEW, 2009), entretanto as visualizações precisam ser bem selecionadas para evitar que informações relevantes sejam omitidas ou levar a interpretações erradas (EVERGREEN, 2019).

Além das visualizações, a Análise Exploratória de Dados também envolve técnicas estatísticas que aprofundam a compreensão dos dados, que já foram mencionadas na definição desse conceito. Estatísticas descritivas, como médias, medianas, e testes de normalidade, são complementadas por métodos como a análise de correlação e a análise de componentes principais (PCA), que permitem examinar relações entre variáveis e reduzir a dimensionalidade dos dados (MARTINEZ & MARTINEZ, 2019; JAMES et al., 2021). Esses métodos estatísticos não apenas reforçam as visualizações, mas também fornecem uma base sólida para a compreensão do que está sendo investigado.

Embora a Análise Exploratória de Dados (EDA) seja poderosa para identificar padrões, relacionamentos e anomalias nos dados, ela tem suas limitações, especialmente quando se trata de responder a perguntas sobre causalidade (HERNÁN & ROBINZ, 2020). A EDA, por sua natureza, não é projetada para determinar relações de causa e efeito, o que é essencial em contextos onde se deseja entender o impacto de variáveis específicas sobre um resultado (PEARL et al., 2000). Essas questões, que envolvem a necessidade de estabelecer uma ligação causal robusta, exigem uma abordagem analítica diferente, que será discutida a seguir.

3.2 Inferência causal

Inferência causal é o processo de identificar e quantificar relações de causa e efeito entre variáveis, indo além de associações observadas nos dados. Essa abordagem busca responder a perguntas do tipo "o que aconteceria se?" ao estimar o impacto de uma intervenção ou alteração em uma variável sobre outra. As técnicas usadas permitem isolar o efeito causal e são fundamentais para a validação de políticas públicas, estudos clínicos, e outras áreas onde compreender a causalidade é crucial (KING, KEOHANE & VERBA, 2021).

A análise de efeito causal e a Análise Exploratória de Dados (EDA) se complementam, pois enquanto a EDA é eficaz para identificar padrões e sugerir

possíveis relacionamentos, ela não pode, por si só, estabelecer uma relação causal entre variáveis. A EDA ajuda a formular hipóteses iniciais e direciona a atenção para potenciais associações nos dados, enquanto a análise causal aprofunda essa investigação ao isolar e quantificar os efeitos diretos de uma variável sobre outra (GELMAN, 2007). Juntas, essas duas abordagens oferecem uma visão mais completa do fenômeno estudado: a EDA destaca o que pode estar acontecendo, enquanto a análise causal explica por que e como isso ocorre, permitindo uma compreensão mais robusta e fundamentada.

Um exemplo da aplicação da inferência causal no contexto da educação é o estudo realizado por IACUS, KING, e PORRO (2012), que investigou o impacto de políticas de diversidade em instituições de ensino superior nos Estados Unidos. Nesse estudo, os pesquisadores conseguiram criar grupos comparáveis de instituições com e sem políticas de diversidade, controlando rigorosamente por diversas características relevantes. Através dessa abordagem, eles foram capazes de isolar o efeito causal das políticas de diversidade sobre os resultados educacionais e a composição demográfica dos alunos, e concluíram que as políticas de diversidade tiveram um impacto positivo significativo em aumentar a representação de minorias sem comprometer os resultados acadêmicos.

Um dos maiores desafios da análise causal é justamente encontrar grupos comparáveis para isolar o efeito de uma variável sobre outra. O padrão-ouro para se obter esses grupos comparáveis são os testes aleatórios controlados (RCTs), onde os participantes são distribuídos aleatoriamente entre grupos de tratamento e controle, minimizando vieses e garantindo que as diferenças observadas nos resultados sejam devidas ao tratamento (HARITON & LOCASCIO, 2018). No entanto, quando a randomização não é possível ou ética, existem várias técnicas na análise causal que permitem aproximar essa condição ideal. Uma dessas abordagens é o *matching*, que busca emparelhar unidades de tratamento e controle com características semelhantes para reduzir diferenças entre os grupos, aumentando a comparabilidade (IMBENS & RUBIN, 2015). Essas estratégias são essenciais para inferir causalidade em contextos observacionais, e é isso que será discutido a seguir, explorando uma das diversas maneiras de construir comparabilidade entre grupos e extrair conclusões causais.

3.3 Matching

A divisão entre o grupo de controle e tratamento no estudo causal é uma das etapas cruciais para garantir que os efeitos observados sejam atribuídos ao tratamento e não a diferenças pré-existentes entre os grupos. Quando a randomização não é possível, uma técnica amplamente utilizada para criar grupos comparáveis é o *matching*. *Matching* é um método que visa emparelhar unidades de tratamento com unidades de controle que são similares em termos de suas características observadas, reduzindo assim o viés de seleção e permitindo uma comparação mais justa entre os grupos (ROSENBAUM & RUBIN, 1983).

De acordo com Stuart (2010), o objetivo do *matching* é replicar o que seria feito em um experimento randomizado, garantindo que as unidades comparadas sejam tão semelhantes quanto possível em todas as covariáveis relevantes, exceto pelo tratamento aplicado. Essa técnica é essencial para fortalecer a validade interna dos estudos observacionais e é amplamente utilizada em pesquisas que buscam inferir relações causais a partir de dados não experimentais.

Com os grupos de controle e tratamento definidos, é essencial validar que as condições iniciais sejam iguais entre os dois grupos no marco zero do estudo (SHADISH, COOK & CAMPBELL, 2002). A ausência de equilíbrio inicial entre os grupos pode introduzir vieses que comprometem a validade dos resultados (AUSTIN, 2011)

Existem formas diferentes de se alcançar esse pareamento entre os grupos, cada uma com suas próprias vantagens e limitações, dependendo do contexto do estudo. Alguns métodos buscam o ponto mais próximo, emparelhando unidades de tratamento e controle com base em características similares, enquanto outros estratificam os dados em camadas ou blocos para garantir que as comparações sejam feitas dentro de grupos homogêneos (HO et al., 2007). Embora todos esses métodos ajudem a minimizar vieses e a criar grupos comparáveis, cada um deles também possui suas desvantagens que devem ser consideradas no desenho do estudo. A seguir, um desses métodos será explorado em mais detalhes, destacando como ele pode ser aplicado para fortalecer a validade dos resultados em estudos observacionais.

3.4 Coarsened exact matching (CEM)

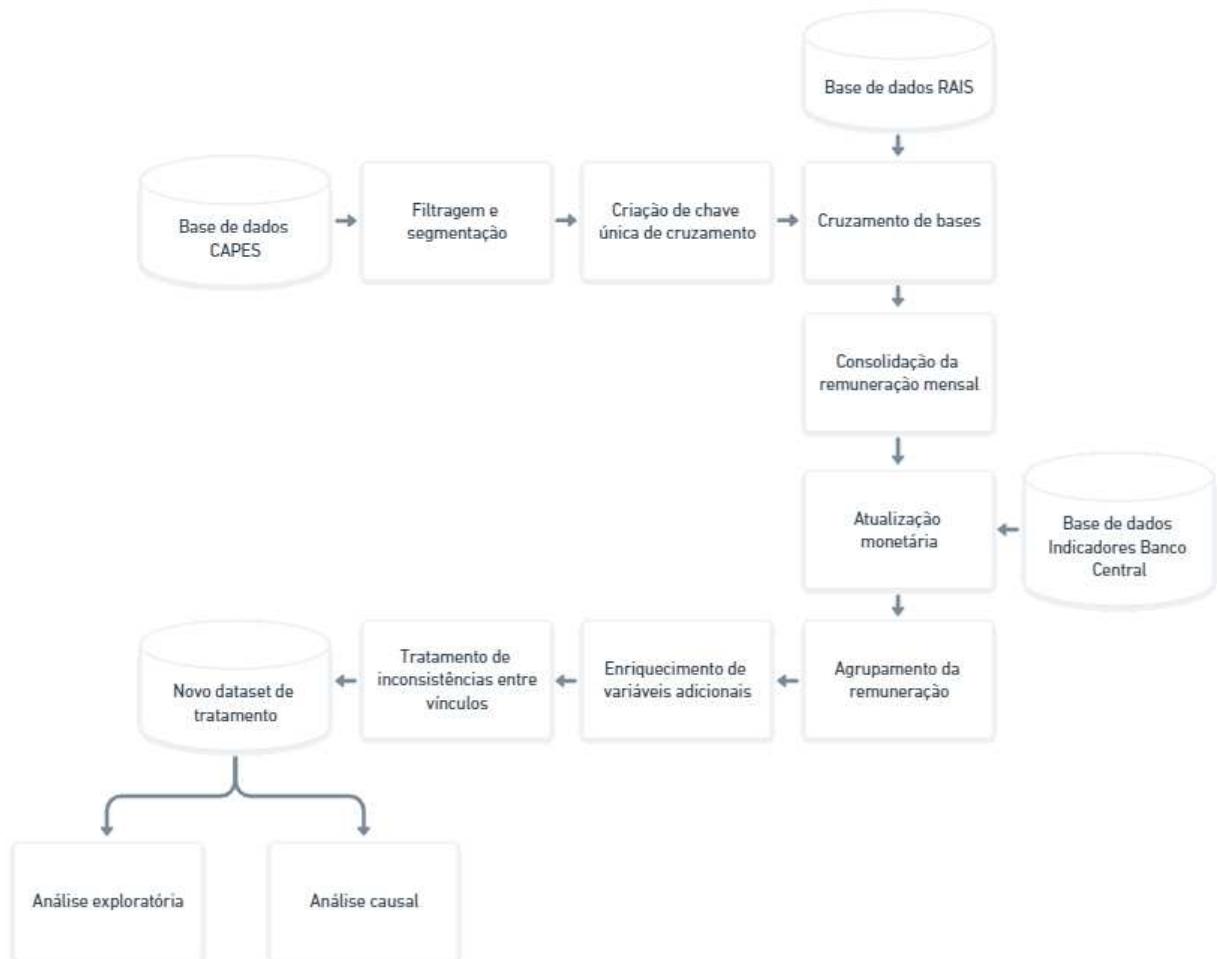
Um desses métodos de matching se chama *Coarsened Exact Matching* (CEM). O CEM é uma técnica de pareamento que agrupa as unidades de tratamento e controle em categorias discretas com base em variáveis observadas, garantindo que as unidades emparelhadas sejam exatamente iguais dentro dessas categorias (BLACKWELL et al., 2009). Esse método é particularmente eficaz para controlar variáveis confundidoras, que são aquelas variáveis que estão correlacionadas tanto com a variável de tratamento quanto com a variável de desfecho, potencialmente distorcendo a relação causal.

É importante identificar essas variáveis pois elas podem introduzir viés nas estimativas causais se não forem adequadamente controladas (MORGAN & WINSHIP, 2015). O processo de seleção dessas variáveis envolve a identificação de fatores que possam distorcer a relação causal desejada, garantindo que todas as variáveis relevantes sejam incluídas no modelo de *matching*. Esse cuidado é essencial para melhorar a validade das inferências causais, já que o sucesso do matching depende diretamente da correta identificação e controle dessas variáveis.

4 METODOLOGIA

A metodologia consistiu em duas grandes etapas: análise exploratória de dados e análise do efeito causal do Mestrado em Computação sobre os valores da remuneração. Para viabilizar essas etapas, será discutida primeiro a obtenção e tratamento dos dados nessa seção, para enfim detalhar a condução das demais etapas. A seguir, o fluxo geral de dados da metodologia geral é apresentado para ilustrar as etapas antes das análises.

Figura 1. Fluxo geral de dados da metodologia



Para sumarizar as etapas, será apresentada a seguir uma tabela que traz uma visão geral do objetivo de cada fase. Nessa tabela, será possível observar como cada etapa foi estruturada para viabilizar um conjunto de dados (*dataset*) que permita a condução da análise exploratória e a análise causal, facilitando a compreensão do processo e das escolhas feitas ao longo do trabalho.

Tabela 1. Detalhamento do fluxo de dados da metodologia

Etapa	Descrição
Filtragem e segmentação	Filtrar os dados para incluir o Mestrado em Computação e segmentar os discentes entre Titulados e Não Titulados.
Criação de chave única de cruzamento	Criar identificador único a partir da base CAPES (discentes) e base RAIS (remuneração).
Cruzamento de bases	Juntar as duas bases de dados utilizando a chave criada para cruzamento.
Consolidação da remuneração mensal	Agrupar as remunerações de diferentes vínculos empregatícios em uma única métrica.
Atualização monetária	Corrigir os valores de remuneração usando o IPCA mês a mês.
Agrupamento de remuneração	Organizar os dados para conter a remuneração ao longo do tempo, desde a matrícula até ano a ano.
Enriquecimento de variáveis adicionais	Adicionar variáveis como gênero, etnia, funcionalismo público e CBO.
Tratamento de inconsistências entre vínculos	Resolver as inconsistências nos dados, utilizando como critério a remuneração mais alta em caso de divergências em variáveis conflitantes.
Anonimização dos discentes	Anonimizar os dados pessoais dos discentes para garantir a privacidade.

4.1 Coleta e tratamento de dados

Para identificar os discentes que estavam matriculados na Pós-Graduação de Computação foram utilizados os dados da plataforma de dados abertos da CAPES trazendo discentes de pós-graduação entre os anos de 2013 e 2021 para os programas de mestrado, mestrado profissional e doutorado.

Essa tabela de discentes gerada inclui, entre outras variáveis, dados pessoais do estudante como o nome, seu CPF parcialmente ocultado, sua data de nascimento, nacionalidade, sua situação de matrícula (podendo ser matriculado, titulado, desligado, abandonado ou mudança de nível sem defesa), seu orientador principal e o título de seu trabalho final, quando titulado. Em relação à instituição, a base traz aspectos como a identificação da programa, o conceito de avaliação do programa e do curso (podendo ser avaliado entre 2 e 7, além do status A para os cursos recém habilitados), dependência administrativa (privada ou pública), status jurídico (particular, estadual, federal) e sua localização (estado e município).

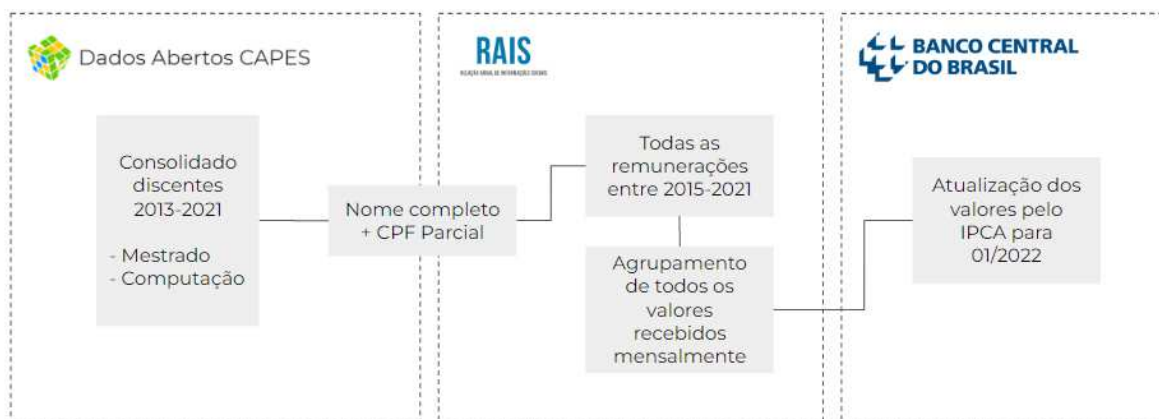
Por essa base vinda da CAPES trazer informações ano a ano ao longo da trajetória do estudante, filtrou-se apenas o estado final em relação a cada aluno, seja ele titulado, abandonado ou desligado, no qual as duas últimas foram agrupadas como uma classe chamada de “não titulados”. Finalmente, apenas a área de Computação da Pós-Graduação foi considerada nesse recorte pelo interesse do presente estudo.

A Figura 1 traz um panorama geral das divisões da base de dados da perspectiva acadêmica. Os discentes da base que são do grau acadêmico Mestrado e que tem um estado final (seja ele titulado ou não) correspondem a 72,6%. Desses estudantes, existem 70,6% que chegaram à conclusão com sucesso, enquanto o restante se divide entre os que abandonaram sua formação (6,8%) e os desligados (22,6%).

Figura 2. Divisão da base acadêmica tratada de Pós-Graduação em Computação



Com a chave necessária para a junção das bases, o enriquecimento dos dados por conta da remuneração se deu inicialmente com a renda no momento da matrícula. Isso fez com que os discentes matriculados em 2013 e 2014 fossem descartados, pois estão numa janela não contemplada pelos dados disponibilizados pela RAIS. Além disso, o interesse do estudo consiste no impacto da remuneração após a conclusão do Mestrado, portanto novas variáveis foram criadas com a remuneração do discente ano após ano em relação ao mês de sua matrícula.

Figura 3. Divisão da base acadêmica tratada de Pós-Graduação em Computação

Houveram enriquecimentos de variáveis relacionadas à característica do empregador ou por ele registradas, tal qual a Classificação Brasileira de Ocupações (CBO), tamanho da empresa, gênero, natureza jurídica do estabelecimento, entre outras. Em alguns casos, o discente pode ter tido mais de um empregador em um determinado mês com CBO diferente, por exemplo. Em todos os casos que houver alguma situação similar, será considerada a variável com maior remuneração associada.

Por fim, os dados, após o cruzamento, foram anonimizados. O objetivo dessa anonimização foi garantir que nenhum dado sensível individual fosse exposto. Todas as análises foram realizadas em grupo, focando apenas na busca pelas conclusões do estudo. Qualquer vestígio de identificação de um indivíduo foi então removido, assegurando que as informações pessoais não fossem comprometidas em nenhuma etapa da pesquisa.

4.2 Abordagem na análise exploratória de dados

A Análise Exploratória de Dados foi conduzida para conseguir compreender quem é o discente de Pós-Graduação em Computação, seja o titulado ou os discentes que não conseguiram a conclusão, seja por abandono ou por desligamento do programa. A maior parte dos trabalhos que investigam o comportamento do egresso em suas áreas de conhecimento investigam variáveis sócio-demográficas e a evolução da taxa de conclusão ao longo do tempo, por isso analisar esses pontos são relevantes também para eventuais comparações entre áreas de conhecimento distintas no futuro.

Por ter acesso aos microdados identificados da RAIS, a base disponível permite explorações que não foram possíveis em trabalhos relacionados em específico para a área de Computação. Essas variáveis permitem a investigação sobre características da relação com o empregador tal qual sua natureza jurídica e o Código Brasileiro de Ocupações (CBO). Esses aspectos, em conjunto com as variáveis tradicionalmente avaliadas em outros trabalhos, permitem uma contribuição nova para o recorte temporal selecionado e para o Mestrado em Computação.

As análises envolvendo intervalos de confiança na etapa de EDA foram realizadas utilizando a técnica de *bootstrap*, uma abordagem robusta para estimar a variabilidade dos dados (EFRON & TIBSHIRANI, 1993). Essa técnica permite gerar várias amostras simuladas a partir dos dados originais, proporcionando uma estimativa mais precisa dos intervalos de confiança. Em todas as análises, foi utilizado um nível de confiança de 95%, assegurando que os intervalos refletissem uma alta probabilidade de conter o verdadeiro valor dos parâmetros estimados.

4.3 Abordagem na inferência causal

Numa situação idealmente perfeita para compreender qual o impacto do Mestrado na área de Computação, seria necessário contar com um grupo grande de indivíduos e, em seguida, randomizar para cada um deles a decisão de fazer ou não o Mestrado. Tal aleatorização seria essencial para proteger o experimento de fatores externos que influenciam a decisão de realizar o Mestrado, assegurando que as diferenças observadas nas remunerações fossem devidas exclusivamente à titulação. No entanto, implementar esse método é inviável, tanto por motivos éticos, já que não seria justo impor ou negar a oportunidade de educação superior a indivíduos de maneira aleatória, quanto por questões práticas, como a dificuldade de gerenciamento e controle desse tipo de estudo.

Outra alternativa hipotética seria avaliar o efeito na remuneração para cada discente, considerando as duas condições: conclusão e não conclusão do mestrado. Se fosse possível expor a mesma observação a essas duas condições distintas, o efeito poderia ser calculado pela diferença observada nas remunerações. No entanto, não é viável aplicar essa abordagem ao mesmo discente, pois não é

possível observar simultaneamente os dois cenários para a mesma pessoa. Essa impossibilidade destaca a necessidade de realizar um estudo causal.

4.3.1 Composição dos grupos de controle e tratamento

Para a condução do estudo causal, é necessário separar os indivíduos em dois grupos: aqueles que receberam a intervenção (grupo de tratamento) e aqueles que não receberam (grupo de controle). Nesse contexto, os discentes que concluíram com sucesso um Mestrado da área de Computação representam o grupo de tratamento, enquanto os discentes que não se titularam por qualquer motivo (seja por uma decisão própria ou por decisão da instituição) compõem o grupo de controle.

A motivação da escolha do grupo de controle por indivíduos que não concluíram o Mestrado se deve por serem discentes que, apesar de não terem completado sua trajetória acadêmica por diferentes motivos, passaram pelo mesmo processo de seleção que os presentes no grupo de tratamento. Essa similaridade de aptidões no momento da matrícula faz com que haja uma homogeneidade das suas características iniciais, reforçando a variável de intervenção e seu possível impacto.

Outro fator que orientou a decisão do grupo controle foi a disponibilidade de dados remuneratórios, que estava restrita aos discentes que, pelo menos, se matricularam no curso de Mestrado em Computação. Essa delimitação não apenas foi necessária devido às limitações dos dados disponíveis, mas também viabilizou o processamento e armazenamento dos dados para o desenvolvimento do estudo.

Essa simples separação entre controle e tratamento pode não ser suficiente para garantir a validade do efeito causal, pois existem fatores ocultos que podem introduzir vieses nos resultados. Por exemplo, se homens mais velhos formarem uma maioria absoluta do grupo de controle, enquanto apenas jovens estiverem no grupo de tratamento, a análise pode ser distorcida. Esse cenário hipotético sugerindo que pessoas mais experientes, que tendem a ter uma remuneração maior, poderia causar uma distorção na avaliação do efeito causal do Mestrado sobre a remuneração. Assim, é necessário garantir que estamos sendo justos na construção do grupo de controle e tratamento.

4.3.2 Matching

Para garantir a correta separação entre os grupos de controle e tratamento, foi utilizado o *Coarsened Exact Matching* (CEM). Esse método agrupa indivíduos similares de acordo com variáveis relevantes para mitigar vieses, conhecidas como variáveis confundidoras. As variáveis confundidoras são definidas por impactar tanto a variável de intervenção (fazer ou não o Mestrado) quanto a de desfecho (a remuneração). Por exemplo, o gênero do discente é uma variável confundidora, pois o acesso à educação superior em áreas de tecnologia é menor para mulheres e há uma conhecida discrepância remuneratória entre homens e mulheres na sociedade. Dessa forma, ao usar um método para dividir os grupos como o CEM, assegura-se que os grupos comparados sejam semelhantes em relação a essa e outras variáveis, melhorando a precisão do estudo causal.

Além do gênero, o estudo considerou a faixa etária no momento da matrícula como uma variável confundidora. Da perspectiva do impacto na intervenção do estudo, que é a conclusão ou não do Mestrado, pessoas mais velhas podem ter outras responsabilidades que dificultam o retorno aos estudos. Por outro lado, do ponto de vista da remuneração, indivíduos mais velhos tendem a ter mais experiência profissional, o que geralmente se reflete em remunerações mais altas.

Da mesma forma, a característica geográfica da instituição de ensino foi compreendida como confundidora. A conclusão do Mestrado pode ser influenciada pela região devido às variações nos custos de vida, que podem impactar a escolha do discente em concluir sua formação, especialmente quando há outras responsabilidades financeiras. Ainda nessa linha, em algumas regiões, as oportunidades podem ser mais escassas, levando os discentes a tomarem decisões de curto prazo para gerar renda para suas famílias. No estudo, essa variável foi considerada em termos de regiões do Brasil. Para o impacto direto no desfecho, ou seja, na remuneração, é importante notar que as regiões possuem mercados de trabalho distintos, o que influencia significativamente na proporção de discentes com emprego e na renda.

Partindo para a perspectiva da instituição de ensino que o discente se matricula, a avaliação que a instituição de ensino recebeu também tem impacto direto na intervenção e no desfecho. Para a intervenção, o prestígio da instituição, frequentemente relacionado à sua avaliação, atrai um perfil de discente com foco na

qualidade e reputação do ensino. É esperado que essas instituições possuam um corpo docente melhor qualificado, o que incentiva alguns discentes a escolherem e concluírem esses programas de Mestrado. Além disso, o reconhecimento do mercado em áreas específicas pode motivar alguns discentes a se matricularem em determinadas instituições. Em relação à remuneração, o próprio reconhecimento da indústria, mesmo que em nichos, desempenha um papel crucial. É esperado que instituições bem avaliadas tendem a oferecer uma qualidade de ensino superior, resultando em um melhor preparo dos discentes. Isso, por sua vez, agregaria mais valor aos negócios e se refletiria positivamente em termos remuneratórios.

Por fim, duas características do empregador foram elencadas como confundidoras: a natureza jurídica e o patamar remuneratório no momento da matrícula. Para a primeira, sabe-se que os trabalhadores da administração pública têm uma motivação direta para conclusão da sua Pós-Graduação, o que pode afetar a decisão de completar sua formação e também a sua remuneração diretamente. Já o patamar remuneratório, como foi chamado no presente estudo um grupo de valores no qual a remuneração do discente se encaixa no momento da matrícula, é uma aproximação do seu momento de carreira através de quanto ganha no marco zero do experimento - ou seja, na matrícula.

Mesmo com todas essas etapas, é necessário garantir uma análise para verificar se os grupos de controle e tratamento estão em condições similares para seguir com o estudo causal. Como o *coarsened exact matching* garante que os pares terão as mesmas características, resta avaliar através da variável de desfecho, a remuneração, se isso está de fato acontecendo.

4.3.3 Desfecho

Após a compreensão de como controlar o estudo, é necessário discutir a análise da variável de desfecho, que é a remuneração. O grupo de controle, composto por discentes que não concluíram o Mestrado, não possui uma data de referência para quando titularam, pois não chegaram a esse marco. Contudo, há uma data de referência a partir da matrícula. Assim, assume-se que o tempo hipotético de conclusão do Mestrado seja entre dois e três anos após a matrícula. Dessa forma, a remuneração a ser considerada nesse estudo em termos de efeito

causal é de cinco anos após a matrícula, ou seja, três anos após a conclusão seja para titulados (tratamento) ou não titulados (controle).

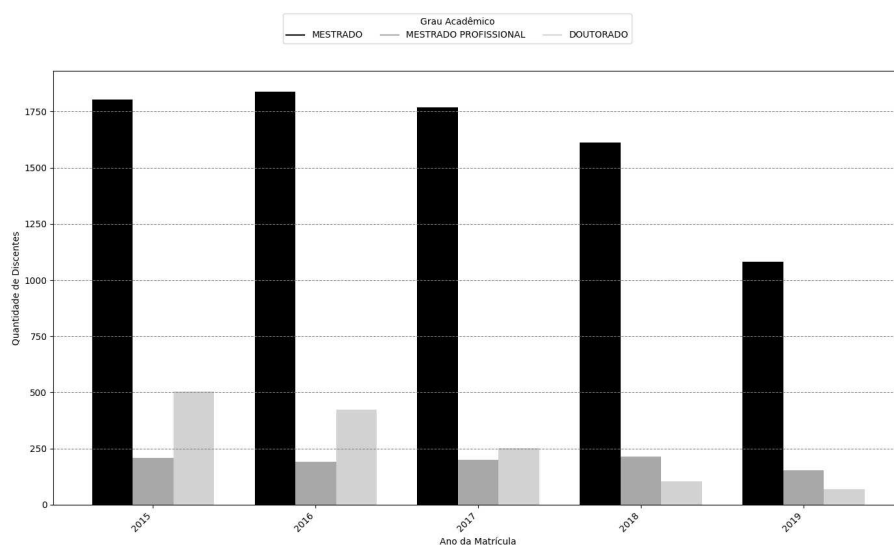
Por disponibilidade dos dados remuneratórios, a análise será centrada nos alunos que se matricularam em 2015 e 2016. A base tratada e apresentada anteriormente conterá os valores de suas remunerações dentro da janela de tempo estipulada. Os valores de remuneração foram corrigidos pelo IPCA para a mesma data de referência, conforme descrito nas etapas anteriores, para garantir uma comparação justa entre os diferentes períodos analisados.

5 ANÁLISE EXPLORATÓRIA DE DADOS

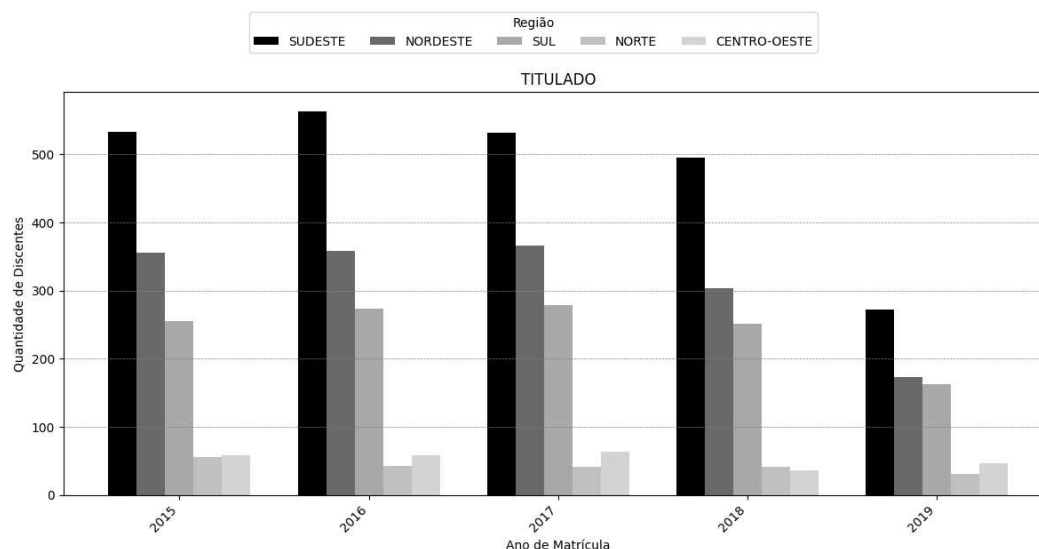
Essa investigação é fundamental para entender as variáveis que impactam os egressos, como gênero, qualidade da instituição de ensino, região e faixa etária. Foram incluídas apenas as análises que trouxeram os *insights* mais relevantes, proporcionando discussões aprofundadas e o levantamento de hipóteses essenciais antes de avançar para estudos de causalidade.

Na etapa inicial de nossa análise exploratória, a distribuição dos discentes entre os graus acadêmicos oferecidos pela pós-graduação em computação — Mestrado, Mestrado Profissional e Doutorado — é examinada. O objetivo é determinar qual grau apresenta maior relevância em termos de quantidade de alunos, influenciando a decisão sobre o foco de nossas investigações futuras.

Figura 4. Evolução absoluta de discentes de pós-graduação em computação de acordo com o grau acadêmico



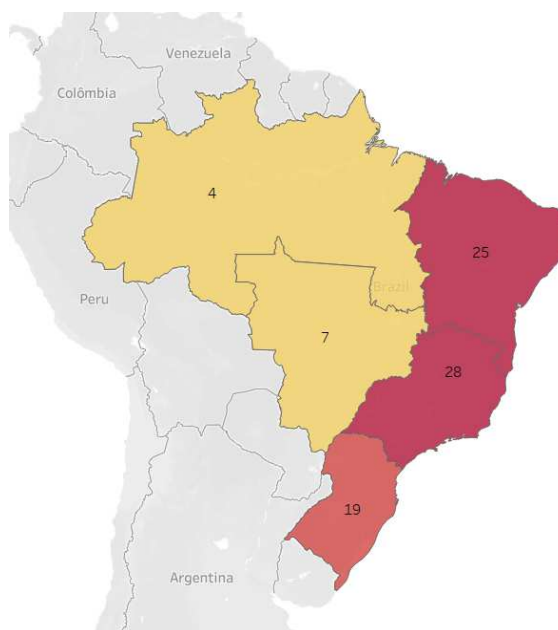
Ao longo dos anos, tem-se observado que o Mestrado atrai mais interessados do que o Mestrado Profissional e o Doutorado. Esse fato se justifica pela oferta de vagas, que é maior para esse grau. Mesmo com a queda em número absoluto, o mestrado é o grau que cresce proporcionalmente mais rápido na Pós-Graduação de Computação. Esse aumento chama atenção para um recorte específico deste grupo, devido à sua relevância em quantidade e impacto potencial.

Figura 5. Evolução de discentes de mestrado por região

Ao se analisar as regiões dos discentes que alcançaram a titulação, observa-se que o Sudeste predomina consistentemente ao longo da série histórica. Mesmo com a queda no número absoluto de alunos, proporcionalmente, essa região continua sendo a que mais forma discentes. Assim, analisar os dados a nível nacional sempre será influenciado significativamente pelo comportamento dessa região, tanto em termos de mercado de trabalho quanto no acesso às oportunidades de Pós-Graduação.

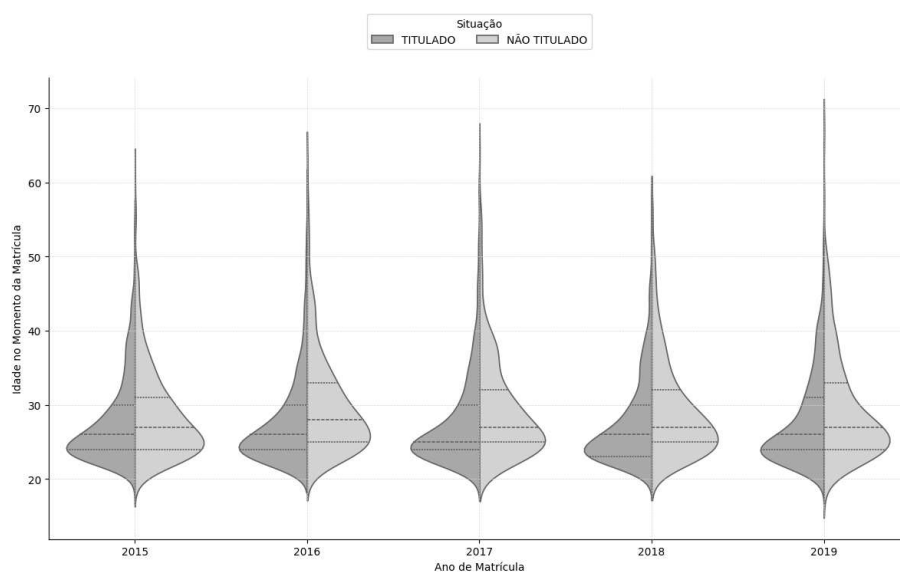
A concentração dos programas de Pós-Graduação em Informática está majoritariamente no Sudeste e no Nordeste, alinhando-se com a distribuição de discentes titulados nessas áreas. Esse cenário alerta para o problema de incentivo desses programas em regiões que carecem de maior atenção, como o Centro-Oeste e o Norte. Essa concentração de programas e discentes titulados no Sudeste e Nordeste resulta em uma densa concentração de mão de obra qualificada nessas áreas, o que pode desestimular o desenvolvimento da indústria de tecnologia nas outras partes do país.

Figura 6. Distribuição de programas de pós graduação em informática em 2019



Ao olhar a idade dos discentes no momento da matrícula, observa-se que os discentes que não conseguem se titular são ligeiramente mais velhos do que aqueles que chegam ao título. Essa tendência é consistente ao longo dos anos, sugerindo que a idade pode ser um fator com alguma influência na conclusão do Mestrado. Esse comportamento é também preservado ao olhar por regiões do Brasil.

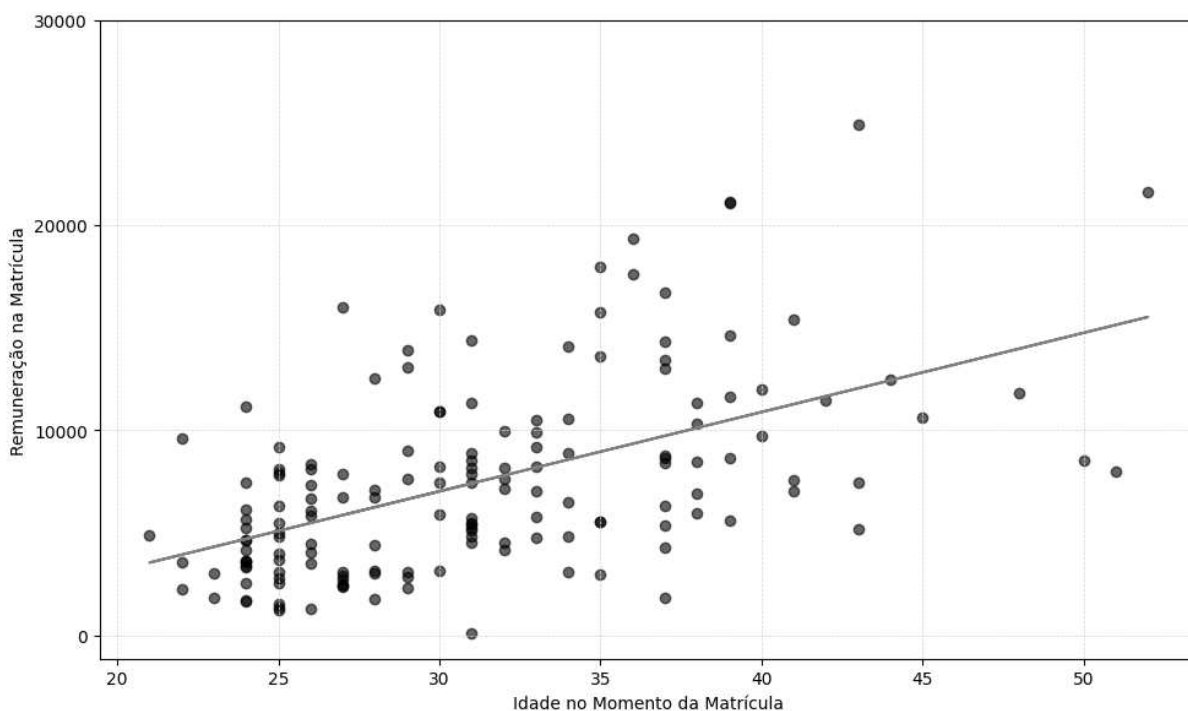
Figura 7. Distribuição da idade no momento da matrícula



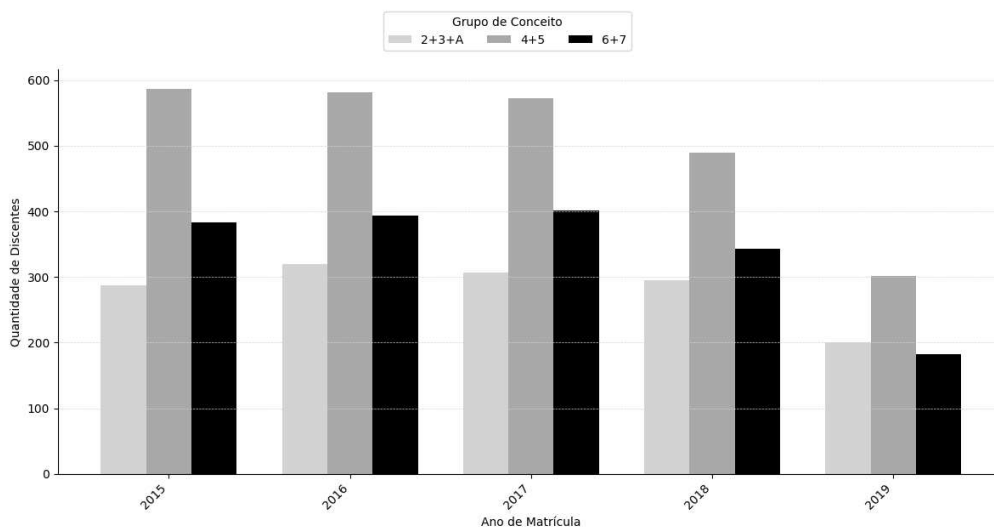
Essa diferença de idade previamente discutida entre os discentes que chegaram à titulação é relevante porque, frequentemente, a idade é associada à experiência profissional. Com isso, é esperado que isso se reflita na remuneração.

Ao analisar a idade na matrícula e a remuneração para os alunos de 2019 que atuam em profissões voltadas à informática, observou-se um coeficiente de Pearson de 0.53, indicando uma correlação positiva moderada. Esse ponto reforça também a necessidade de controle de faixa etária no estudo seguinte, o causal.

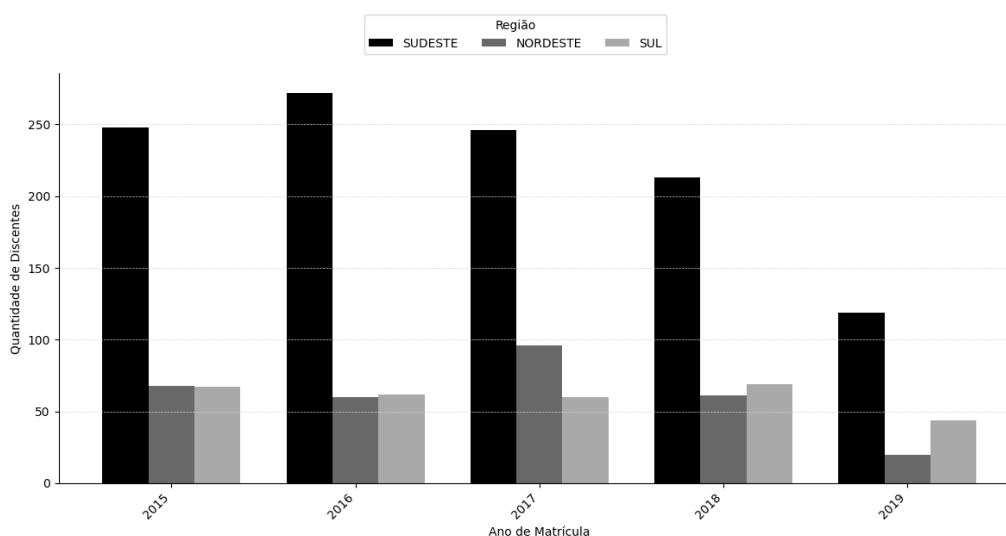
Figura 8. Correlação entre idade e salário na matrícula para discentes de 2019 empregados na área de Informática



Ao compreender como os discentes titulados se dividem nos grupos de conceitos de avaliação dos seus programas, percebe-se que o grupo de conceito intermediário (programas com avaliação 4 ou 5) sempre prevalece ao longo da série histórica, o que faz com que essa predominância influencie fortemente análises a nível nacional. As instituições com menor conceito têm uma participação menor, exceto no ano mais recente da série histórica. Além disso, mesmo com a queda absoluta no número de discentes que alcançaram a titulação por ano de matrícula, o comportamento proporcional entre grupos de conceito permanece semelhante. Esses padrões mostram que, independentemente das variações nos números absolutos, a distribuição dos conceitos das instituições se mantém estável ao longo dos anos, reforçando o impacto do grupo de conceito intermediário na análise consolidada.

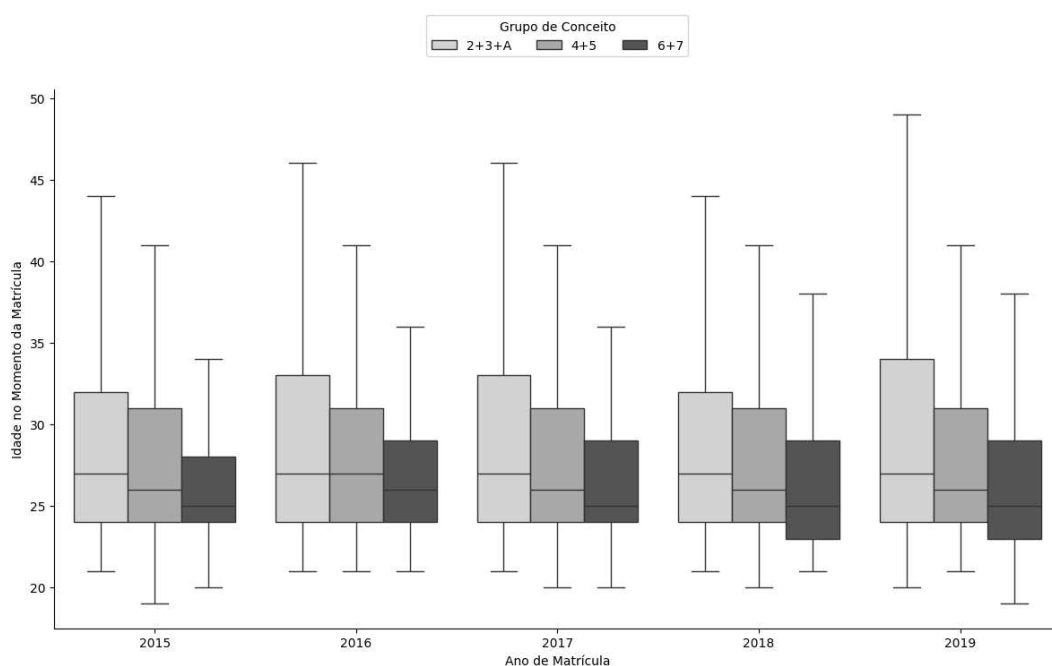
Figura 9. Titulados por grupo de conceito de avaliação do programa

Buscando compreender melhor onde estão os discentes titulados das instituições com melhor avaliação, observa-se que o Sudeste concentra esses titulados ao longo do tempo, sem que qualquer outra região se aproxime. Esse fato evidencia que o Sudeste não só impulsiona a quantidade de alunos, mas também a qualidade, medida através do conceito da avaliação dos programas. Os acessos e facilidades disponíveis nesta região, juntamente com o interesse das pessoas em residir nela, contribuem significativamente para essa concentração de mão de obra qualificada. Esse cenário ressalta a carência de desenvolvimento nas demais regiões do país, destacando a necessidade de estratégias que equilibrem essa distribuição e promovam o crescimento uniforme em todas as áreas geográficas.

Figura 10. Titulados de conceitos avaliados com nota seis ou sete por região

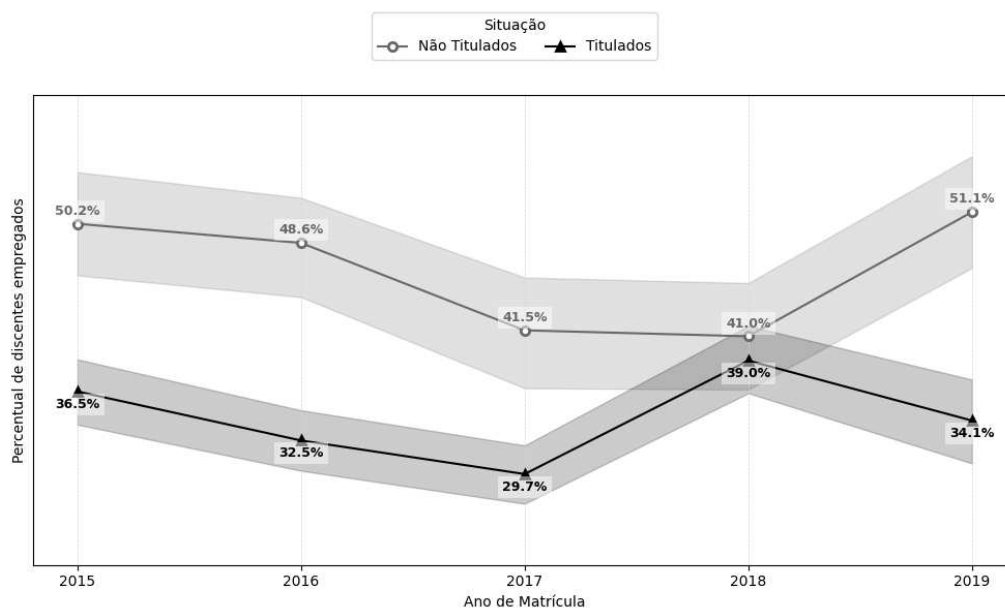
Olhando a distribuição por idade desses grupos de conceito de avaliação, observa-se que a distribuição dos discentes no grupo de maior conceito é composta por indivíduos ligeiramente mais jovens. Acredita-se que esses programas demandam mais tempo e dedicação, o que pode ser mais compatível com a rotina de pessoas mais novas, que tendem a possuir menos responsabilidades financeiras de uma forma geral em virtude do seu momento de vida.

Figura 11. Distribuição da idade de titulados de acordo com o grupo de conceito de avaliação do programa



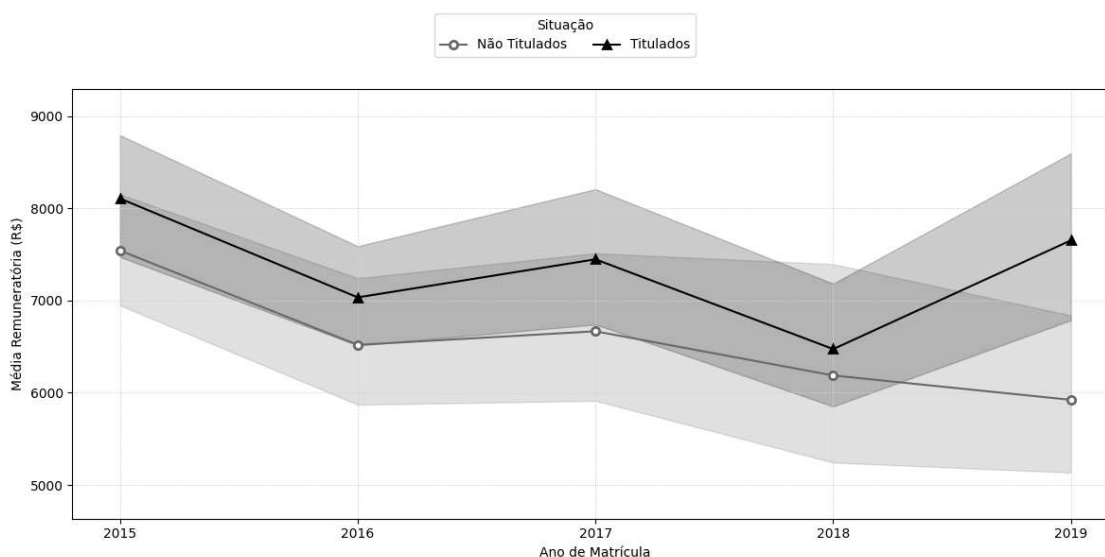
Embora a proporção de discentes com emprego não seja o desfecho principal do estudo, o fato de o aluno estar empregado no momento da matrícula pode introduzir um viés no estudo, o que torna relevante essa análise. Observou-se que os titulados, ao longo dos anos, apresentaram uma proporção de discentes com emprego menor do que os não titulados no momento da matrícula, com exceção de 2019, onde as taxas se aproximam. Esse comportamento sugere que a dificuldade de conciliar a presença no mercado de trabalho com uma Pós-Graduação em Computação pode estar influenciando a capacidade de conclusão do Mestrado de alguns discentes.

Figura 12. Percentual de discentes com emprego no momento da matrícula por ano e situação



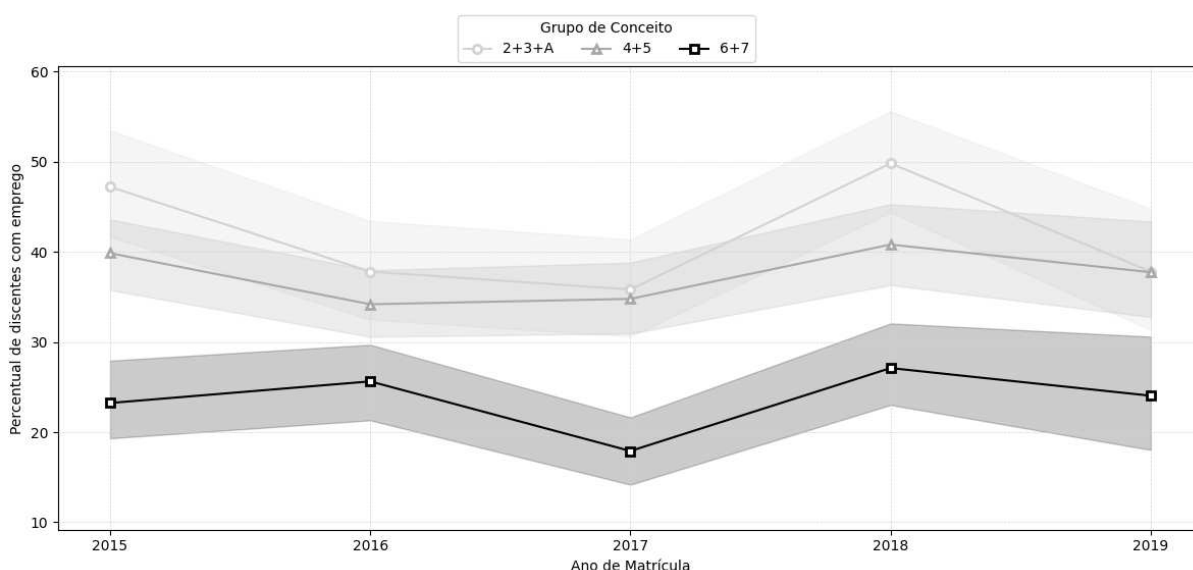
Olhando a remuneração média dos titulados e não titulados que estavam empregados (ou seja, tendo qualquer valor de remuneração) no momento da matrícula, percebe-se que não existe uma diferença significativa entre os grupos. O valor da remuneração foi impulsionado para cima, pois os discentes sem renda não foram incluídos na análise. Sabe-se que o percentual de discentes titulados com emprego no momento da matrícula é menor, portanto, se todos os discentes fossem considerados, seria esperado uma discrepância nos valores e uma diminuição na média remuneratória dos titulados.

Figura 13. Remuneração média dos discentes empregados no momento da matrícula



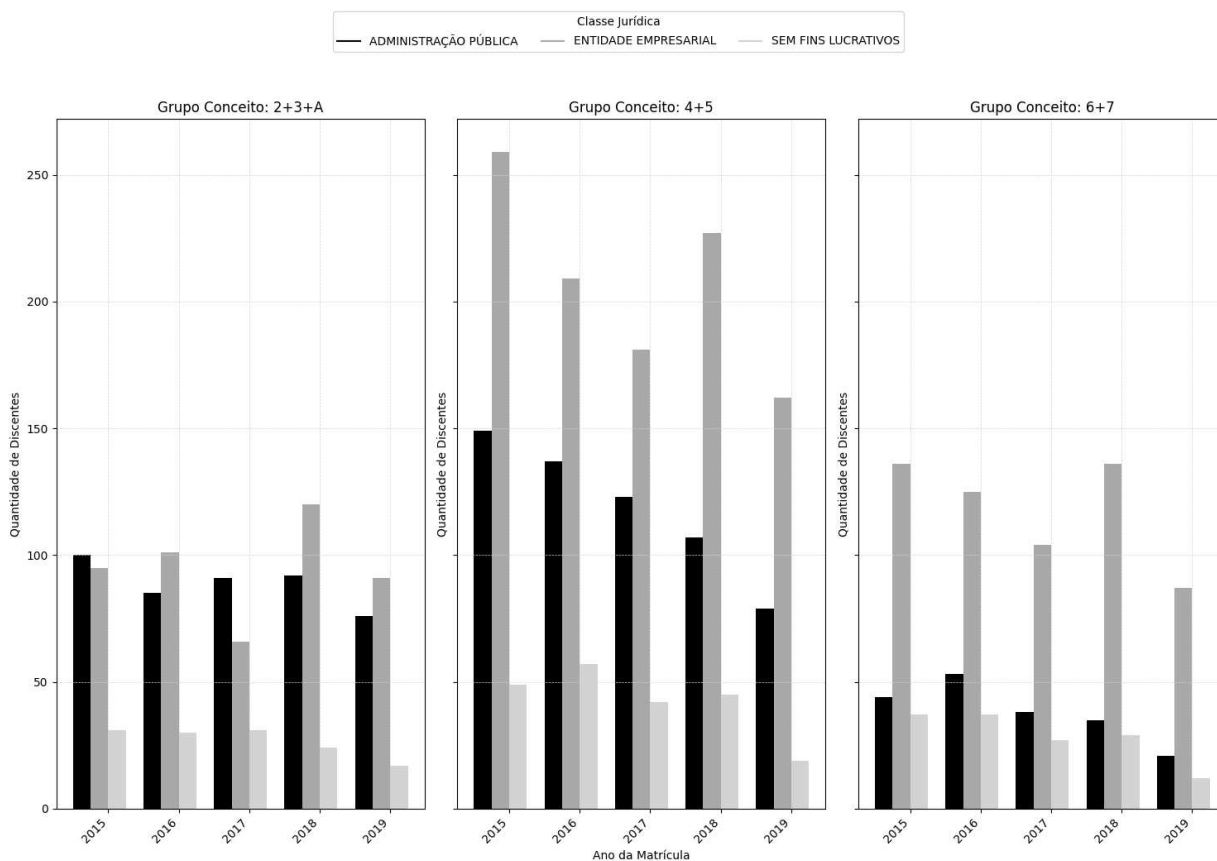
Na investigação da proporção de discentes com emprego por grupo de conceito de avaliação, constatou-se que os discentes titulados de instituições com programas melhores avaliados têm uma proporção menor do que os demais. Esse padrão foi observado de forma consistente em diversas regiões do país. Tal diferença pode indicar a necessidade de dedicação integral a esses programas. Essa exigência pode ser um fator limitante, uma vez que nem todos os indivíduos podem abrir mão do trabalho para se dedicar exclusivamente ao Mestrado. Além disso, a concentração desses programas no sudeste, uma região com alto custo de vida, agrava ainda mais essa questão, tornando a acessibilidade a esses programas um desafio maior para muitos discentes.

Figura 14. Proporção de titulados com emprego no momento da matrícula por grupo de conceito avaliativo e ano da matrícula

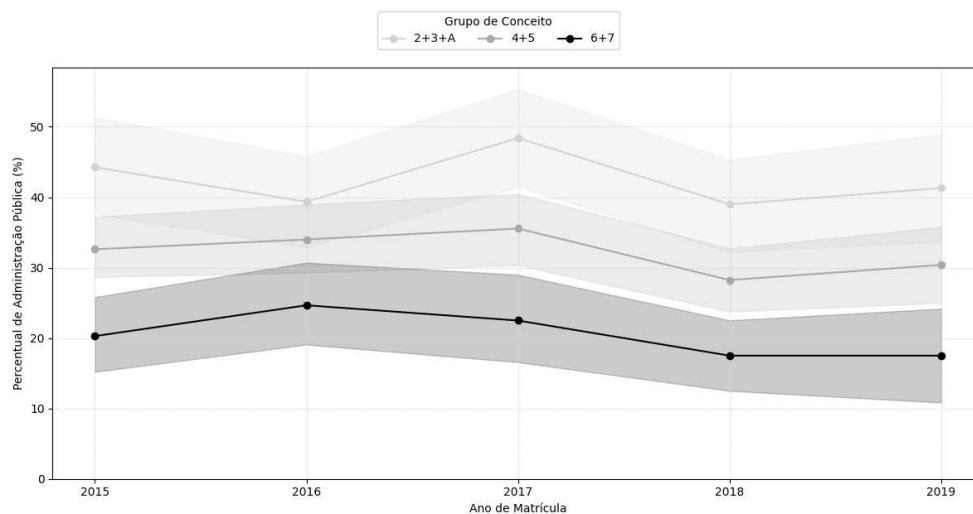


O panorama da proporção de discentes com emprego nos leva a investigar também a participação de funcionários públicos, privados e de instituições sem fins lucrativos. Nota-se que programas de conceito 4 e 5 constituem a maior parte dos discentes. Dessa forma, é natural que os funcionários públicos estejam majoritariamente concentrados nessas instituições. No entanto, ao olhar para cada grupo de conceitos, existe uma maior presença proporcional de funcionários públicos em instituições com programas de menor nível de avaliação. À medida que o nível de avaliação dos programas sobe, nota-se uma diminuição na proporção de funcionários públicos presentes nesses programas.

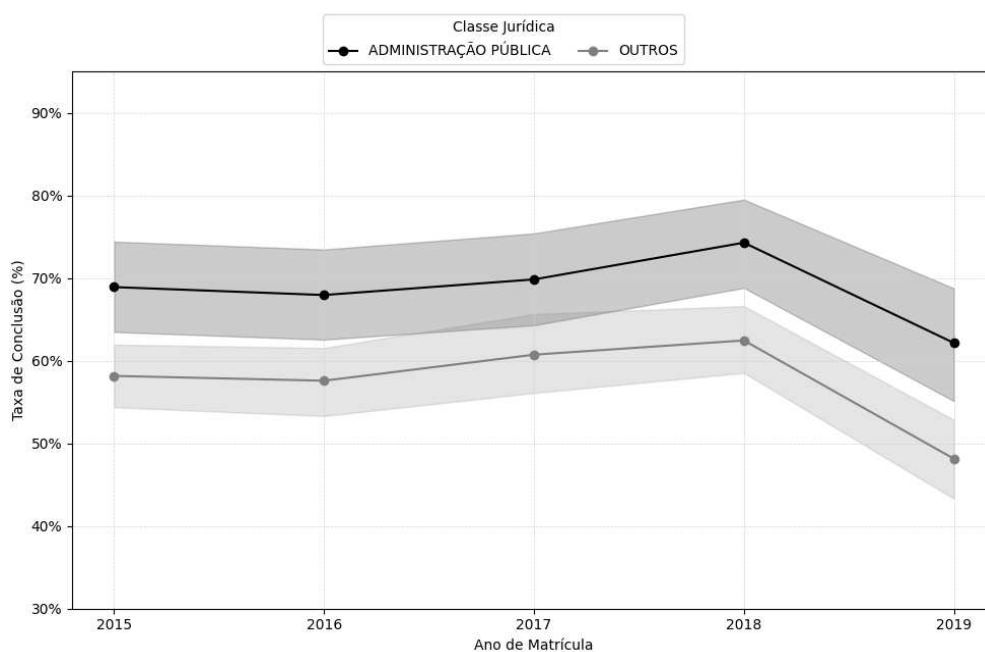
Figura 15. Matriculados empregados de acordo com o tipo de organização no momento da matrícula por grupo de conceito



Essa dinâmica sugere que, embora os programas de qualidade intermediária atraiam mais discentes que trabalham para a administração pública em números absolutos, a distribuição proporcional é invertida em relação aos programas de menor avaliação. Tal comportamento pode indicar que parte dos funcionários públicos opta por instituições com menor nível de avaliação devido à compatibilidade com suas agendas profissionais, além de proporcionar um retorno financeiro através da obtenção do título acadêmico independente do grau de avaliação do programa cursado.

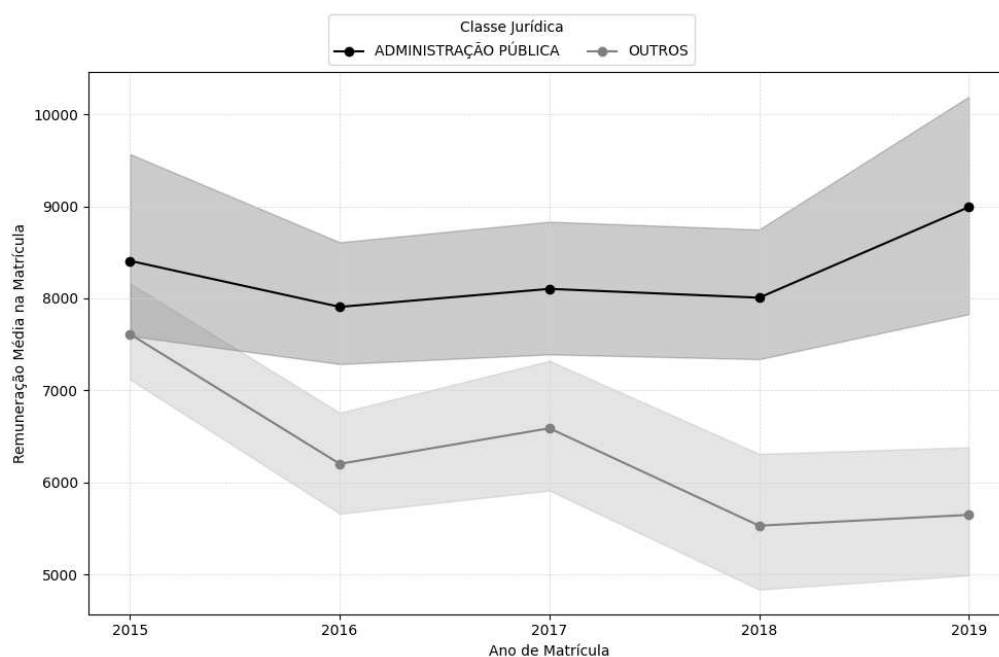
Figura 16. Proporção de funcionários públicos no momento da matrícula por grupo de conceito

Ainda investigando o funcionário público no momento da matrícula, observou-se que sua taxa de conclusão é maior do que a dos demais discentes. Isso pode ser atribuído, em grande parte, ao estímulo direto na remuneração recebido em muitos casos ao obter novos títulos acadêmicos. Ademais, pode-se considerar que a iniciativa privada muitas vezes não proporciona a flexibilidade necessária para que o discente concilie seus estudos com suas obrigações profissionais, algo que parece ser mais viável no setor público. Essa maior flexibilidade pode ser um fator determinante para a elevada taxa de conclusão observada entre esses profissionais.

Figura 17. Taxa de conclusão de acordo com a natureza jurídica do emprego

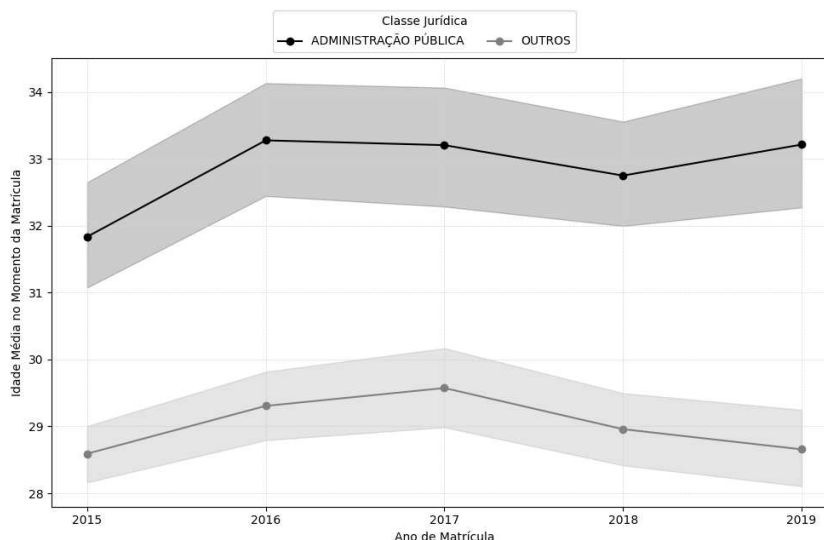
O funcionário público que ingressa no Mestrado também tende a ganhar mais do que os funcionários que não trabalham para a iniciativa pública. A remuneração mais elevada pode tanto sugerir que esses indivíduos estejam em um estágio diferente de suas carreiras, ou trazendo um panorama no qual a administração pública melhor reconhece financeiramente o valor do profissional em relação às empresas privadas.

Figura 18. Remuneração média no momento da matrícula de acordo com perfil do empregador



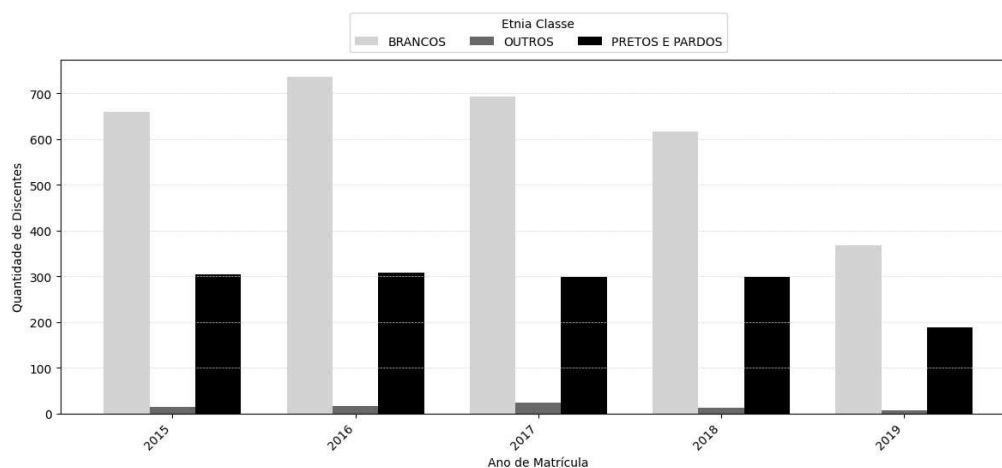
Em relação à idade, o funcionário público é consistentemente mais velho do que seus colegas de instituições privadas e sem fins lucrativos. A maior experiência profissional, frequentemente associada à idade, é um fator significativo que pode explicar parte da disparidade remuneratória entre os grupos. A idade mais avançada dos funcionários públicos indica uma trajetória profissional mais consolidada, refletindo em remunerações mais altas.

Figura 19. Média de idade do funcionário da iniciativa pública e privada no momento da matrícula



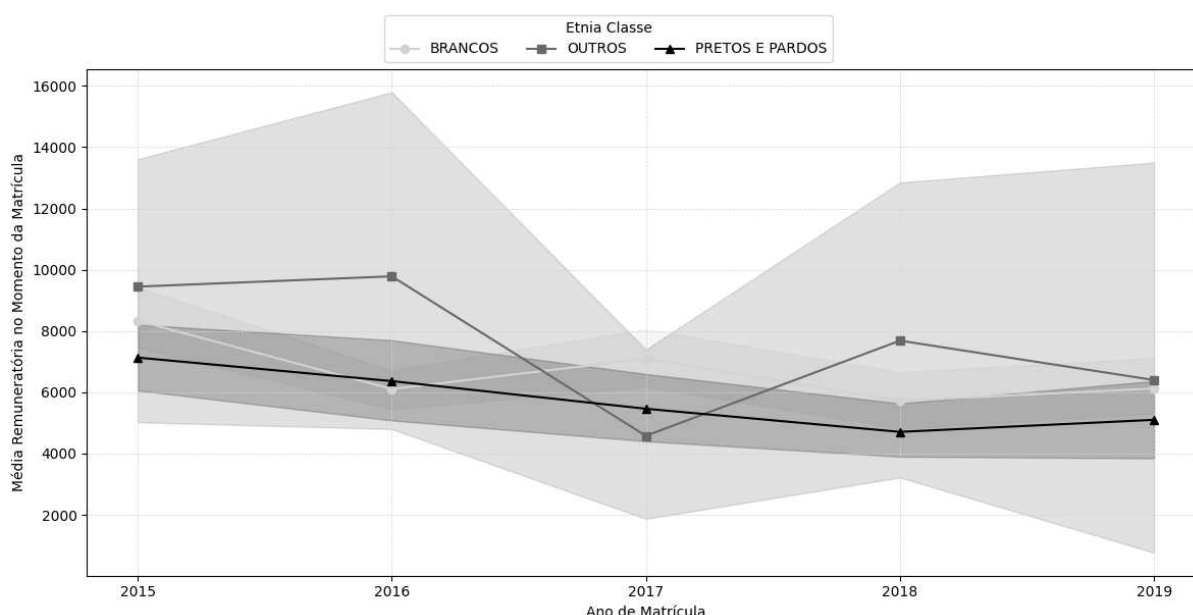
Ao analisar os discentes com alguma identificação de etnia na base da RAIS, verifica-se que a distribuição de discentes pretos e pardos não reflete a composição da população brasileira em geral. Nota-se um acesso limitado e contínuo desse grupo à pós-graduação em computação, tanto entre titulados quanto entre não titulados. Essa discrepância no acesso evidencia uma barreira persistente para esse segmento da população, que também se verifica no cenário de Pós-Graduação em Computação. A baixa representatividade de pretos e pardos entre os discentes é um indicativo das dificuldades enfrentadas por esses grupos para ingressar nos programas de Pós-Graduação, o que aponta para a necessidade de medidas que promovam uma maior inclusão e equidade no acesso a esses cursos.

Figura 20. Distribuição da etnia por ano de matrícula

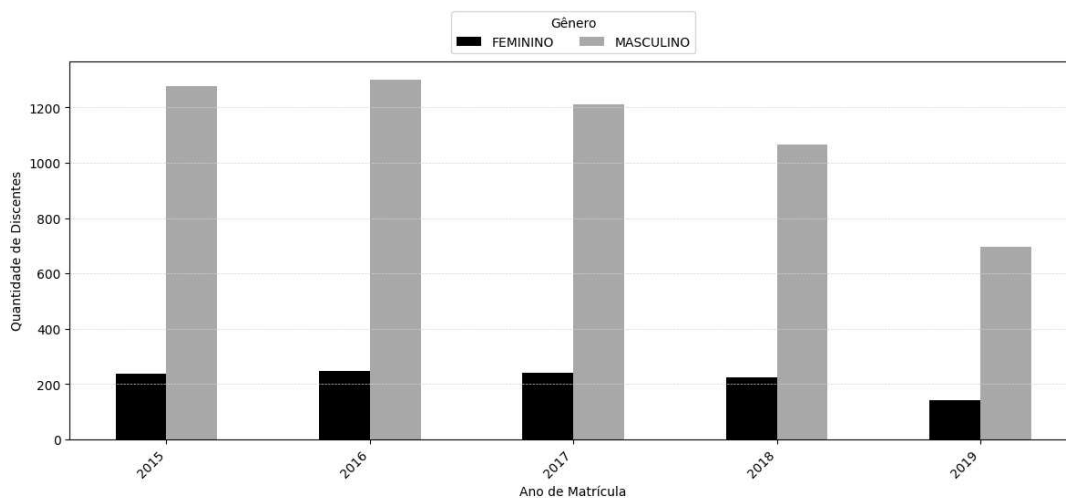


Apesar da dificuldade de acesso dos discentes identificados como pretos ou pardos discutida anteriormente, a média remuneratória entre as etnias que puderam ser identificadas no momento da matrícula não apresenta distinções significativas. Observa-se que os intervalos de confiança amplos indicam uma variedade considerável de perfis de alunos em relação às suas rendas. Esses intervalos amplos sugerem que, dentro de cada grupo étnico, há uma diversidade de situações econômicas, o que dificulta a identificação de padrões claros apenas com base na etnia.

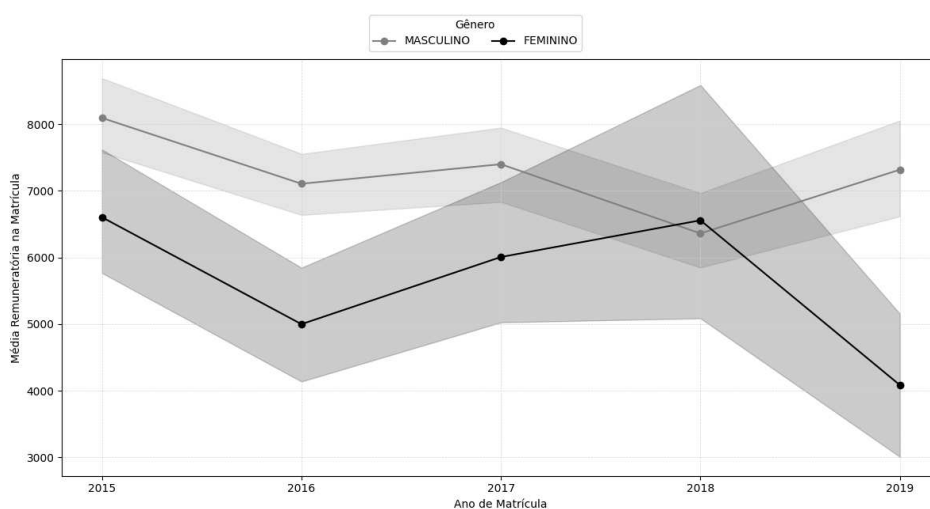
Figura 21. Média remuneratória por etnia dos empregados no momento da matrícula



Os programas de Pós-Graduação em Computação também mostraram ser majoritariamente masculinos, e essa característica não vem mudando ao longo do tempo. A diferença na proporção de gênero ocorre em todos os graus de avaliação dos programas, sendo, portanto, uma ocorrência generalizada. No entanto, é importante ressaltar que, mesmo com essa disparidade de acesso, as pessoas identificadas com o gênero feminino não apresentam qualquer diferença nas suas taxas de conclusão do Mestrado em comparação com seus colegas de gênero masculino. Isso indica que a questão é principalmente referente ao acesso inicial. Dessa forma, é nítido a necessidade de um esforço maior para incentivar a participação de mulheres na tecnologia, buscando um equilíbrio mais justo e representativo dentro desse campo.

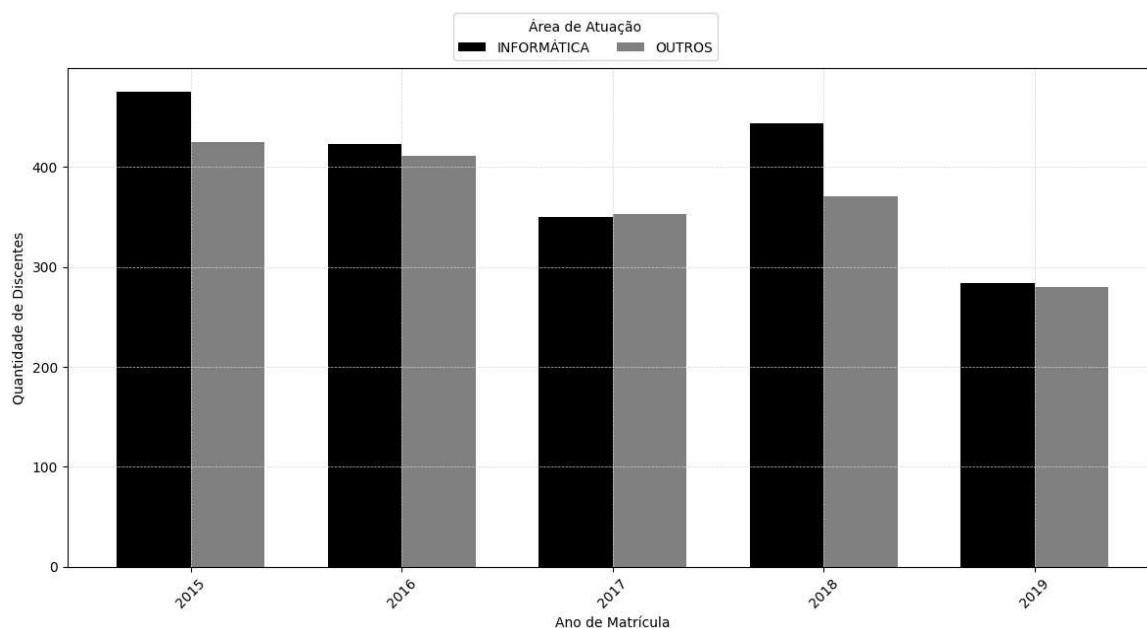
Figura 22. Discentes por gênero

Sabe-se que, estruturalmente, existe uma diferença da média salarial entre gêneros, e ao se analisar a remuneração na Pós-Graduação em Computação no momento da matrícula, essa disparidade também acontece. Com exceção do ano de 2017 e 2018, todos os demais anos apresentaram uma diferença significativa entre as remunerações do gênero masculino e feminino. Não há uma diferença de idade relevante entre os gêneros que justifique essa disparidade salarial, e é importante lembrar que está sendo considerado apenas os discentes empregados no momento da matrícula. Esse dado reforça que a diferença remuneratória entre gêneros se aplica também no cenário da Pós-Graduação em Computação, especificamente no Mestrado.

Figura 23. Média remuneratória na matrícula por gênero

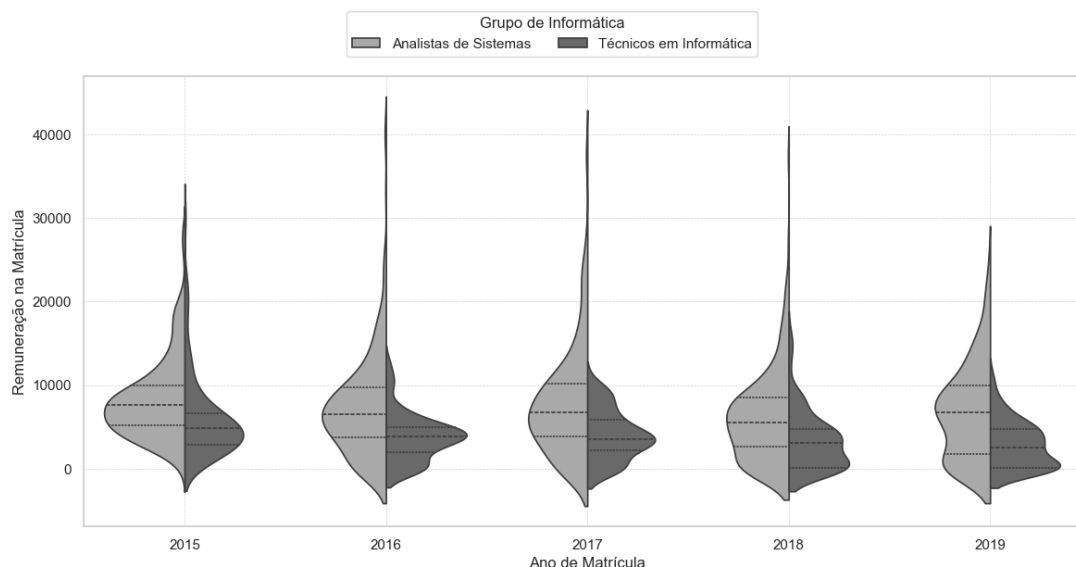
Os dados remuneratórios também permitem uma análise detalhada da área de atuação dos discentes no momento da matrícula. Foram investigados apenas aqueles que tinham remuneração registrada nesse período, utilizando as CBO cadastradas. Entre todas as ocupações, duas subfamílias se destacaram, representando metade de todas as profissões: analistas de sistemas e técnicos em informática.

Figura 24. Quantidade de discentes por área de atuação



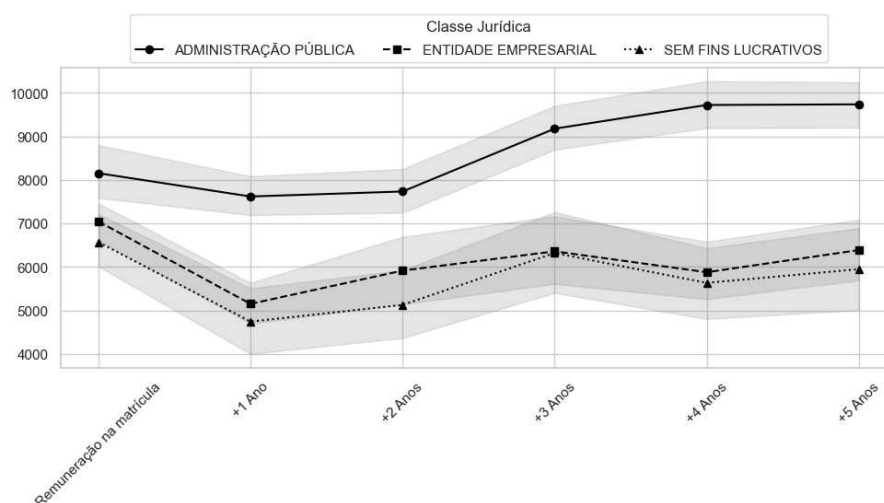
Essas duas ocupações apresentam uma diferença relevante na média remuneratória entre si, visível na distribuição a seguir. Os analistas de sistemas são, em média, mais velhos, o que lhes confere uma vantagem que justifica parcialmente a discrepância remuneratória observada. Apesar dessa diferença etária e de renda, a taxa de conclusão entre analistas de sistemas e técnicos em informática é similar.

Figura 25. Distribuição da remuneração dos discentes empregados com ocupações de informática



Ao analisar a evolução remuneratória dos discentes matriculados entre 2015 e 2016 que já possuíam alguma remuneração no momento da matrícula, observou-se um ganho significativo apenas entre funcionários públicos. Esse aumento foi mais evidente após três anos, período em que se esperava a conclusão do Mestrado. O mesmo padrão foi identificado entre os titulados, sugerindo que a alavanca de remuneração pode estar relacionada à natureza jurídica do emprego.

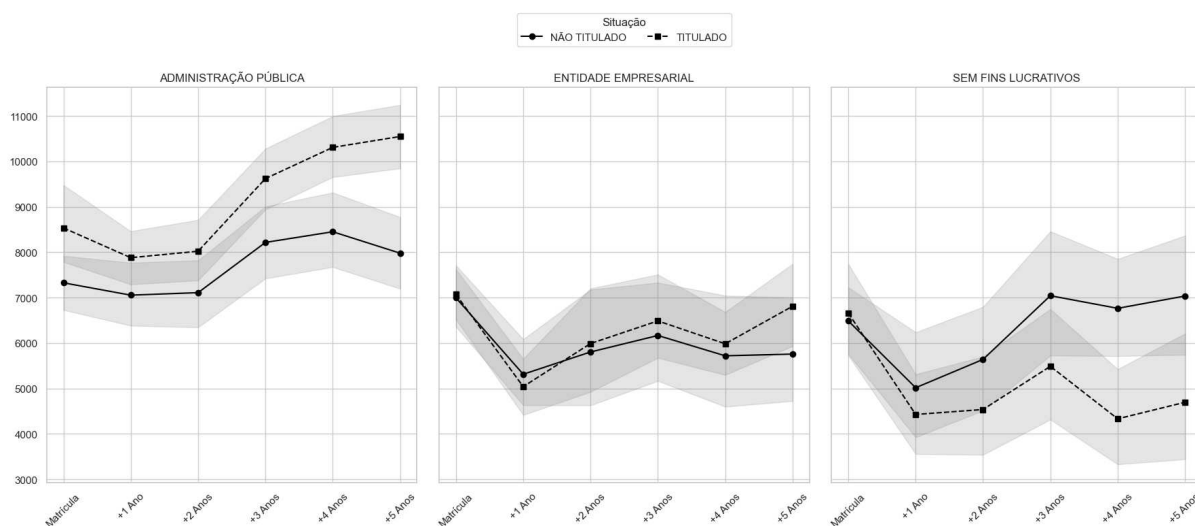
Figura 26. Média remuneratória por natureza jurídica para discentes matriculados em 2015 e 2016 com remuneração na matrícula



A análise anterior foi aprofundada, agora separando o grupo de discentes titulados dos não titulados. Assim, observa-se que os egressos de entidades empresariais e sem fins lucrativos apresentam trajetórias remuneratórias

semelhantes após três anos, sejam eles mestres ou não. No entanto, os titulados da administração pública se destacam dos não titulados, sugerindo novamente que os funcionários públicos podem estar sendo beneficiados com a formação em aspectos remuneratórios.

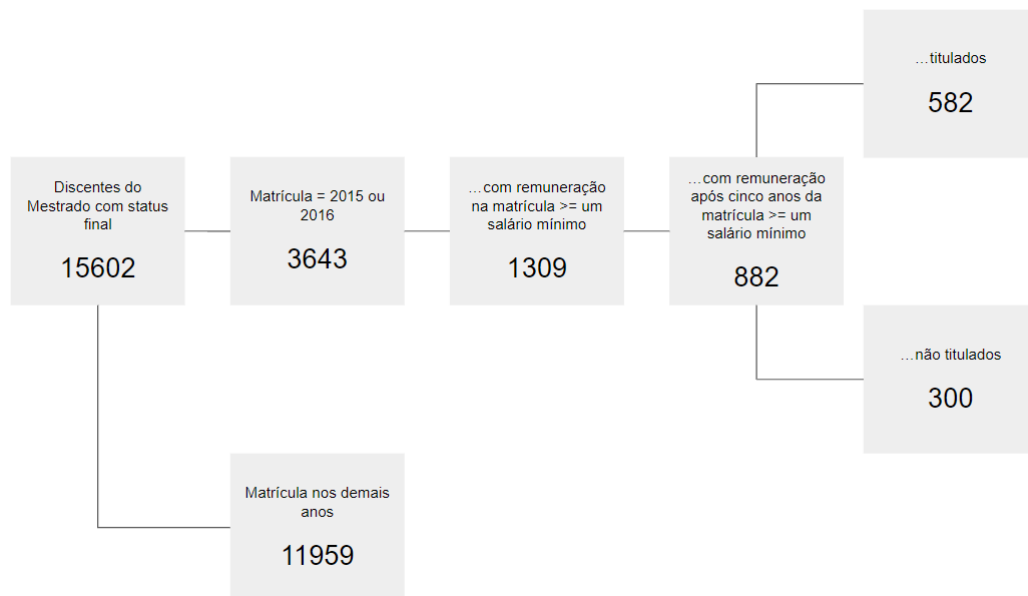
Figura 27. Média remuneratória por natureza jurídica e situação final do discente



6 ANÁLISE CAUSAL

Para conduzir a análise causal, só foram considerados os discentes com remuneração de pelo menos um salário mínimo tanto no momento da matrícula quanto após cinco anos. Esse critério resultou na remoção de alguns estudantes, conforme ilustrado na Figura 27. Além disso, apenas os matriculados em 2015 e 2016 foram incluídos na análise, pois eram os únicos para os quais havia dados disponíveis referentes aos cinco anos após a matrícula.

Figura 28. Filtros para a seleção do grupo da análise causal



A primeira etapa da análise, então, consiste em avaliar como o grupo de tratamento se comparou em relação ao grupo de controle no marco inicial do experimento com o objetivo de entender a necessidade de algum método de *matching*. Para isso, cada discente titulado teve um par aleatoriamente definido que abandonou ou foi desligado do Mestrado. Todos os indivíduos envolvidos tiveram a remuneração maior ou igual a um salário mínimo tanto no momento da matrícula quanto cinco anos após.

Em cada um dos pares formados (Ti, Cj) , onde Ti é o i -ésimo discente no grupo de tratamento e Cj é o j -ésimo discente no grupo de controle, o efeito do tratamento ou *treatment effect* (TE) pode ser calculado com base na remuneração Y , ou seja, a variável de desfecho, conforme a equação 6.1.

$$TE = Y_{Ti} - Y_{Cj} \quad (6.1)$$

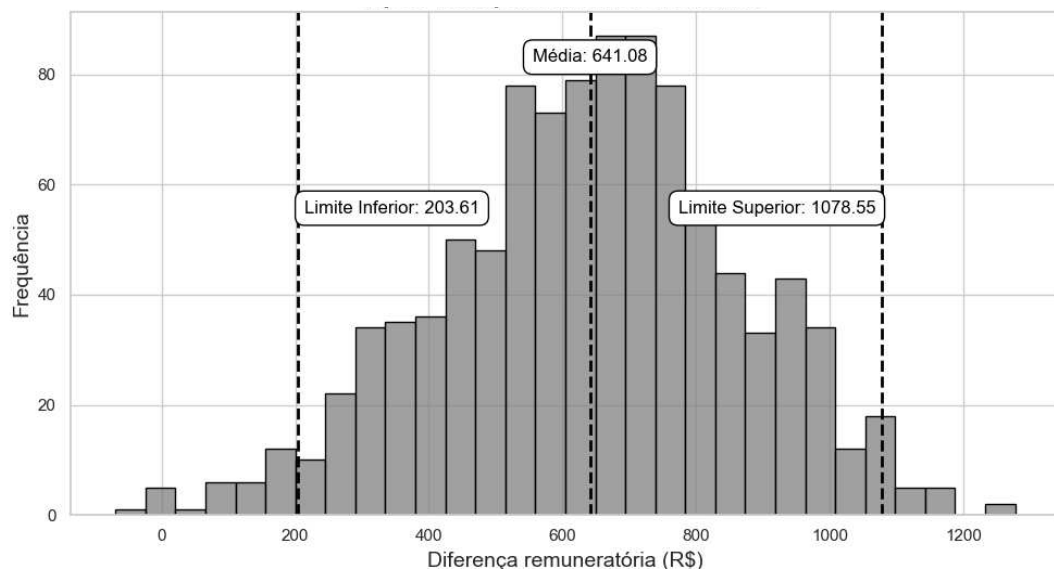
Com o efeito do tratamento calculado por par, foi possível calcular o *average treatment effect on treated* (ATT) através do seu valor absoluto e percentual, de acordo com as equações 6.2 e 6.3 a seguir, onde n é o número de pares formados.

$$ATT = \frac{1}{n} \sum_{i=1}^n (Y_{Ti} - Y_{Cj}) \quad (6.2)$$

$$ATT (\%) = \frac{\frac{1}{n} \sum_{i=1}^n (Y_{Ti} - Y_{Cj})}{\frac{1}{n} \sum_{i=1}^n Y_{Cj}} \quad (6.3)$$

Esse processo foi feito sem reposição das observações, entretanto, foi repetido por mil vezes visando variar os pares e permitir análises de confiança frente ao resultado observado. Cada uma das médias observadas ao longo dessas mil iterações está representada no histograma a seguir.

Figura 29. Histograma da diferença média de remuneração no momento da matrícula após mil pareamentos e sem controle por variáveis confundidoras



O histograma das médias das diferenças evidencia que, sem nenhum tipo de controle no pareamento dos indivíduos, houve uma diferença no qual os discentes titulados naturalmente ganham 8.2% a mais já no momento da matrícula. Esse resultado demonstra a necessidade de uma cautela maior na definição desses pares visando uma diminuição dessa diferença no momento inicial de comparação, tornando então justa qualquer análise ao longo do tempo do grupo de discentes que se titulou em relação aos demais.

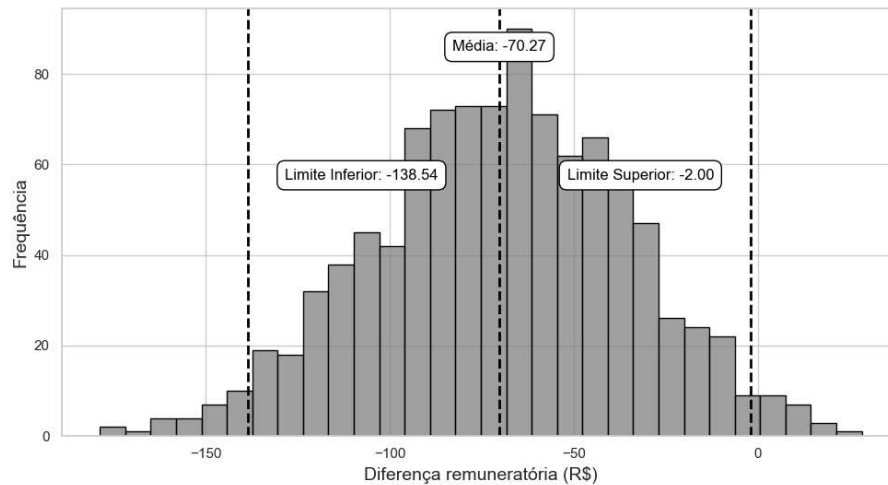
O mesmo processo foi conduzido, porém dessa vez aplicando o *coarsened exact matching* com as variáveis confundidoras previamente definidas e explicadas na seção 4. A aplicação do método fez com que os mesmos 882 discentes inicialmente fossem agrupados em 525 estratos. Cada um desses estratos conta com grupos de discentes contendo exatamente os mesmos valores das variáveis confundidoras, porém foram descartados os estratos que não existiam no grupo de tratamento e os que apenas existiam no mesmo, sem que fosse possível encontrar um par. O resultado do processo está exibido na Tabela 1.

Tabela 2. Antes e depois do processo de pareamento com coarsened exact matching

	Controle	Tratamento	Total	Estratos	Observações do tratamento descartadas	Observações do controle descartadas
Antes do descarte	300	582	882	525	-	-
Depois do descarte	126	126	252	98	456	174

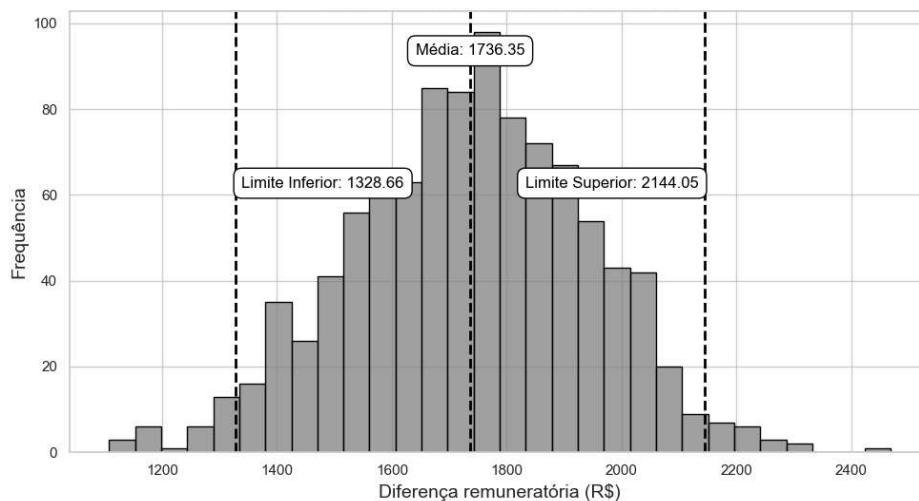
Com o pareamento feito, foi possível observar que a diferença média da remuneração dos discentes titulados em relação aos não titulados no momento da matrícula caiu para apenas 1%, mas agora em favor do grupo controle. Essa condição encontrada satisfaz a expectativa de igualdade entre os grupos do experimento para a análise causal, como demonstrado na Figura 29 a seguir.

Figura 30. Histograma da diferença média de remuneração no momento da matrícula após mil pareamentos e com controle por variáveis confundidoras através do *coarsened exact matching*.



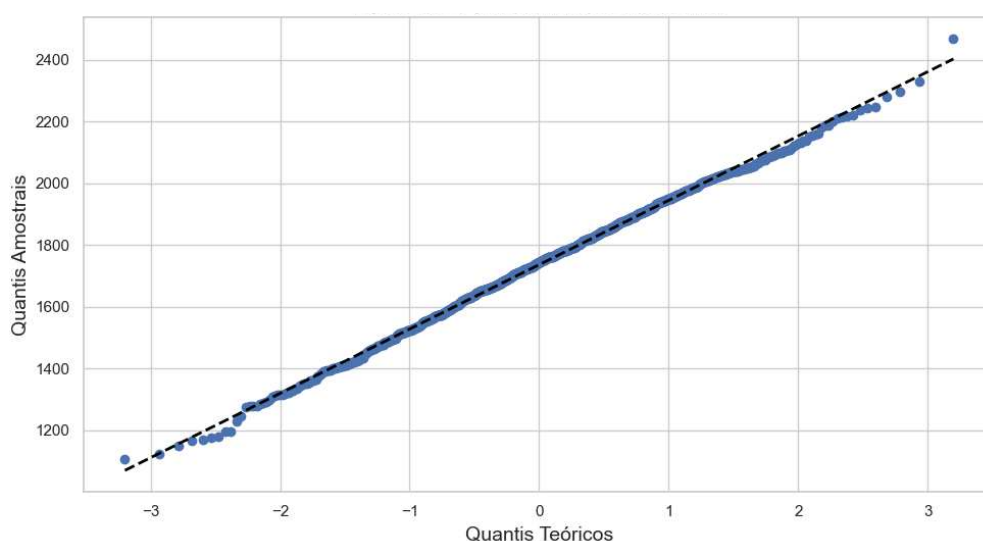
Com a condição de igualdade no marco inicial do experimento, que agora consiste em dois grupos com as mesmas características em relação ao gênero, faixa etária, região, natureza jurídica do empregador, grupo do conceito de avaliação do programa e mesmo patar remuneratório no momento da matrícula, foi possível observar que depois de cinco anos, os discentes que concluíram agora tem uma média de diferença de remuneração positiva. **O impacto do Mestrado na área de Computação é estimado em 18.2% após cinco anos da matrícula.** O histograma apresentado na Figura 30 demonstra as médias de diferenças de remuneração depois de mil processos de pareamentos realizados.

Figura 31. Histograma da diferença média de remuneração cinco anos depois da matrícula após mil pareamentos e com controle por variáveis confundidoras através do *coarsened exact matching*



Cabe ressaltar que os intervalos de confiança foram calculados considerando que a distribuição anteriormente apresentada na Figura 30 corresponde a uma distribuição normal, como exibido no “Quantile-Quantile plot” ou QQ plot na Figura 31. Essa visualização é comumente usada para comparar a distribuição de um conjunto de dados (no caso, a distribuição das médias remuneratórias) em relação à uma distribuição gaussiana.

Figura 32. QQ plot da distribuição de médias de remuneração após cinco anos da matrícula e com 1000 processos de pareamentos



Apesar do impacto como um todo ser relevante para o estudo, buscou-se entender como as características dos discentes contribuíram ou não para esse resultado. As análises desses desdobramentos do efeito causal foram realizadas filtrando as observações de cada variável confundidora e seus possíveis valores. Utilizando os resultados de mil pareamentos, foi examinado o *average treatment effect on treated*, visando entender detalhadamente os efeitos observados.

Um dos pontos levantados na análise exploratória de dados foi sobre a evolução remuneratória do funcionário público, e por isso foi a primeira característica escolhida para entender o efeito. Quando avaliando os pares de acordo com a natureza jurídica do seu trabalho no momento da matrícula, observou-se que, embora os ganhos dos discentes que trabalham para entidades privadas e sem fins lucrativos fossem similares, esses não foram significativos. Isso abre o debate sobre o alinhamento entre a indústria e a academia, que pode não estar gerando uma

percepção de valor na área de Computação a ponto de ser reconhecida em termos de remuneração por instituições fora da esfera pública.

Em contraste, os funcionários públicos apresentaram um ganho estatisticamente relevante. Esse impacto pode ser explicado pelo adicional remuneratório associado ao título para funcionários públicos, em muitos casos. Além disso, há a hipótese de que esses titulados que trabalham na administração pública buscaram outras alavancas adicionais de ganho salarial, demonstrando, através da conclusão do Mestrado, que teriam uma motivação para uma atitude assim.

Tabela 3. ATT por natureza jurídica do empregador

Natureza jurídica do empregador	Pares analisados	Estratos	Limite Inferior do ATT	ATT	Limite Superior do ATT	Houve efeito significativo?
Entidade empresarial	48	40	-R\$ 611,72	R\$ 219,38	R\$ 1.050,48	Não
Sem fins lucrativos	10	10	-R\$ 903,03	R\$ 303,48	R\$ 1.509,98	Não
Administração Pública	68	48	R\$ 2.568,15	R\$ 3.016,84	R\$ 3.464,52	Sim

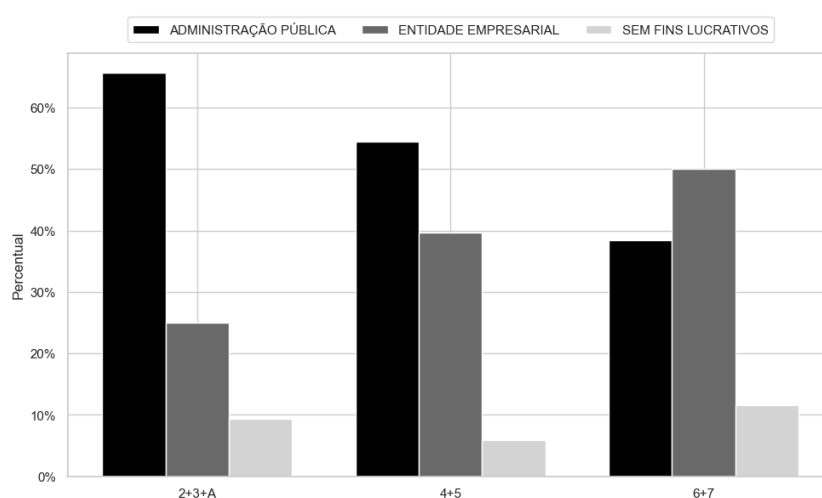
Ao olhar o impacto do Mestrado pelo grupo de conceito dos programas, observou-se que os grupos de conceito mais baixo e intermediário apresentaram um ganho significativo para o grupo de tratamento. No entanto, o grupo de conceito mais alto não apresentou um efeito positivo, uma vez que seu intervalo de confiança inferior foi menor do que zero. Um possível motivo para essa diferença é que esses discentes podem estar evoluindo em seus estudos, e cinco anos após a matrícula no Mestrado pode não ser tempo suficiente para a conclusão de um Doutorado, por exemplo, que habilitaria outros desafios profissionais com melhor remuneração.

Tabela 4. ATT por grupo de conceito do programa

Conceito do Programa	Pares analisados	Estratos	Limite Inferior do ATT	ATT	Limite Superior do ATT	Houve efeito significativo?
2+3+A	32	29	R\$ 1.737,73	R\$ 2.572,12	R\$ 3.406,51	Sim
4+5	68	48	R\$ 1.299,59	R\$ 1.839,03	R\$ 2.378,47	Sim
6+7	26	21	-R\$ 419,34	R\$ 436,46	R\$ 1.292,27	Não

Outro ponto que merece ser citado, é que foi visto na análise exploratória de dados que conforme aumenta o grupo de conceito de avaliação dos programas, menor era a proporção de funcionários públicos. Esse comportamento se repetiu nos pares da análise causal, e como os funcionários públicos influenciam positivamente o efeito, a análise por grupo de conceito avaliativo pode ter sido impactada por isso. A Figura 32 a seguir evidencia essa proporção de funcionários públicos nos pares da análise causal por grupo de conceito.

Figura 33. Proporção de pares no estudo causal de acordo com a natureza jurídica por grupo de conceito de programa



O recorte do experimento por faixa etária no momento da matrícula mostrou que, com exceção da faixa etária mais jovem, todos os demais grupos apresentaram um efeito positivo observado. Esse grupo mais jovem pode estar aguardando completar sua jornada acadêmica, incluindo o Doutorado, antes de buscar concursos públicos e outras oportunidades. Além disso, o grupo da faixa etária mais jovem foi o único que não tinha nenhum par de funcionários públicos, que é justamente o grupo que mais alavanca o resultado.

Tabela 5. ATT por faixa etária na matrícula

Faixa etária	Pares analisados	Estratos	Limite Inferior do ATT	ATT	Limite Superior do ATT	Houve efeito significativo?
19 a 24 anos	12	12	-R\$ 1.095,06	R\$ 1.233,82	R\$ 3.562,70	Não
25 a 29 anos	46	31	R\$ 1.323,60	R\$ 2.059,69	R\$ 2.795,78	Sim
30 a 34 anos	45	34	R\$ 779,62	R\$ 1.244,70	R\$ 1.709,92	Sim
35 a 39 anos	14	12	R\$ 1.541,72	R\$ 2.309,82	R\$ 3.077,93	Sim
40 anos ou mais	9	9	R\$ 1.122,00	R\$ 2.311,00	R\$ 3.501,11	Sim

Em relação ao grupo remuneratório do discente no momento da matrícula, observou-se que apenas os titulados com remuneração do menor grupo não apresentaram um ganho significativo. Um possível motivo para isso é que esses discentes estão em um momento mais júnior de suas carreiras, com tarefas mais operacionais. Dessa forma, podem não ter tido oportunidades de se destacar com o conhecimento adquirido e portanto traduzir isso em ganho de renda. Em relação a proporção de funcionários públicos que poderia explicar essa diferença, o grupo melhor remunerado no momento da matrícula tem uma proporção menor de funcionários públicos (aproximadamente 18%) do que o grupo com pior remuneração nesse momento, que conta com aproximadamente 30%. Mesmo assim, a presença dos funcionários públicos nessa faixa de remuneração não fez o efeito ser positivo. Isso fortalece a hipótese do escopo de trabalho de cada perfil remuneratório no momento inicial do estudo.

Tabela 6. ATT por grupo remuneratório no momento da matrícula

Grupo remuneratório na matrícula	Pares analisados	Estratos	Limite Inferior do ATT	ATT	Limite Superior do ATT	Houve efeito significativo?
Até R\$5.0000	36	28	-447,79	324,91	1097,6	Não
Entre R\$5.000 e R\$10.000	79	60	1872,05	2423,14	2974,23	Sim
Acima de R\$10.000	11	10	1278,86	1416,85	1554,85	Sim

Por fim, todas as regiões apresentaram um ganho significativo do tratamento em relação ao controle, destacando o impacto positivo do Mestrado em diferentes partes do país. O maior destaque ficou para a região centro-oeste, que teve um ATT mais expressivo por ter todos seus pares vindos da administração pública, no qual o maior efeito acontece.

A região Norte não foi representada nos resultados da Tabela 6 devido à avaliação de apenas um par, o que impede a obtenção de conclusões relevantes para essa região específica. Essa mesma situação ocorreu com o gênero feminino, onde os poucos pares não conseguiram variar ao longo dos mil pareamentos realizados por sua relação de sempre ter o mesmo tratamento para o único controle disponível, fazendo com que não tivesse representatividade para essa discussão de resultados.

Tabela 7. ATT por região

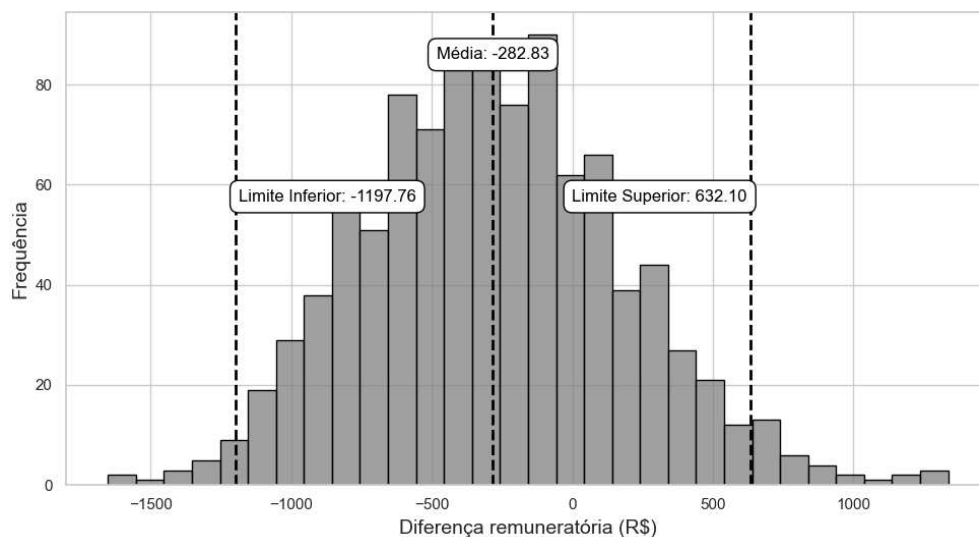
Região	Pares analisados	Estratos	Limite Inferior do ATT	ATT	Limite Superior do ATT	Houve efeito significativo?
Sudeste	51	39	R\$ 795,38	R\$ 1.504,97	R\$ 2.214,56	Sim
Nordeste	23	19	R\$ 1.059,28	R\$ 1.932,93	R\$ 2.806,59	Sim
Sul	41	32	R\$ 767,70	R\$ 1.369,22	R\$ 1.970,74	Sim
Centro-Oeste	10	7	R\$ 2.020,00	R\$ 3.437,36	R\$ 4.854,47	Sim

Apesar do escopo do estudo ter sido limitado aos discentes descritos anteriormente, o resultado motivou também uma investigação inicial para aqueles que não tinham renda no momento da sua matrícula, mas que tinham acima de um salário mínimo após cinco anos.

Essa análise foi conduzida controlando as mesmas variáveis confundidoras, porém como esses discentes não tiveram remuneração no momento da matrícula, não existe natureza jurídica do empregador no marco inicial. Da mesma forma, o grupo remuneratório na matrícula é o mesmo para todos: nenhuma remuneração, portanto não foi necessário avaliar se a variável de desfecho estava equilibrada no marco inicial do experimento. Sabendo que as remunerações mais altas da administração pública foram exploradas na etapa de EDA deste trabalho, optou-se então por controlar a natureza jurídica do empregador após cinco anos, seja ela pública, sem fins lucrativos ou uma entidade empresarial. Além disso, o pareamento foi feito sem reposição, porém repetido mil vezes, exatamente como o caso anterior.

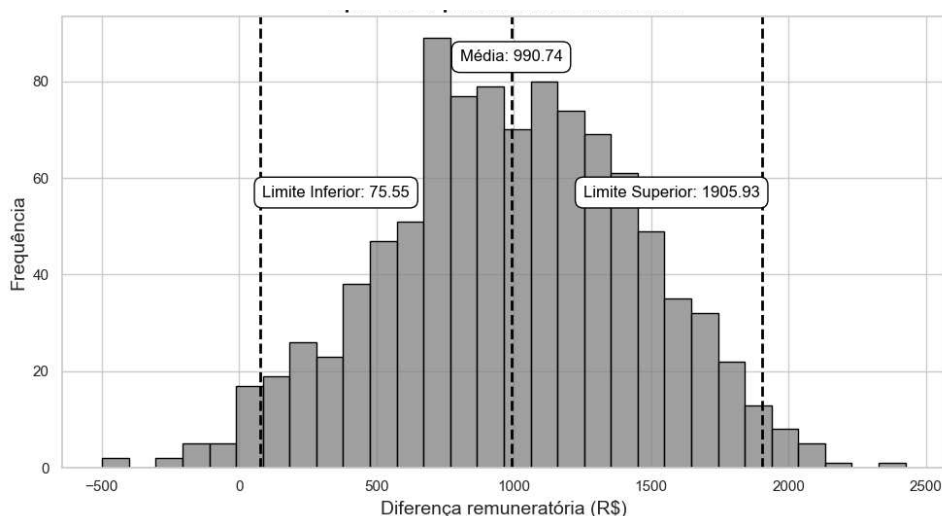
Diferente da primeira análise causal, desta vez o incremento da média remuneratória não foi significativo o suficiente para dizer que houve efeito, conforme Figura 33, que evidencia o intervalo de confiança contando a diferença média da remuneração entre tratamento e controle contendo o valor zero.

Figura 34. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula



Para compreender esse resultado, o foco inicial foi entender a distribuição de funcionários públicos, já partindo da compreensão que estes têm o potencial de alavancar positivamente o resultado. A Figura 34 mostra que o efeito positivo é de R\$1000,00 em média para o grupo da administração pública, com o intervalo de confiança abrangendo apenas valores acima de zero.

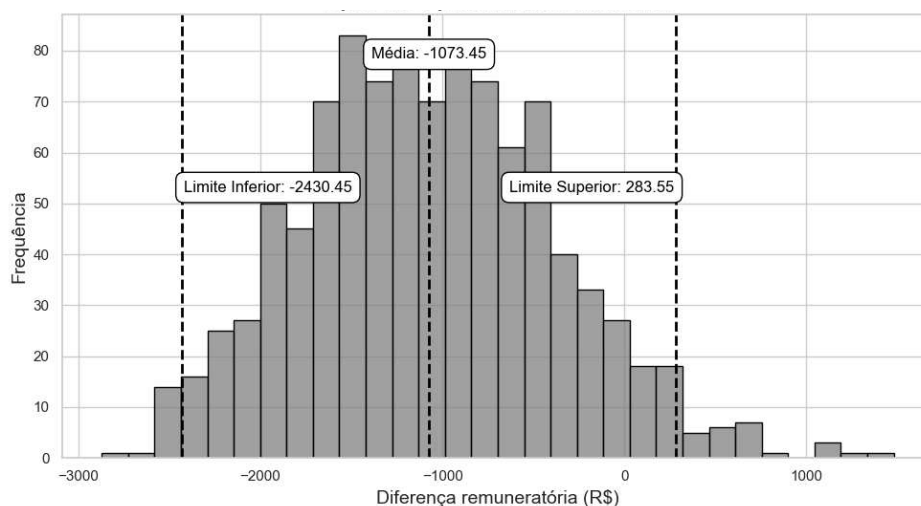
Figura 35. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula que viraram funcionários públicos



Já para os discentes que optaram por entidades empresariais após cinco anos, não houve evidência, conforme Figura 35 aponta. Apesar da média ser

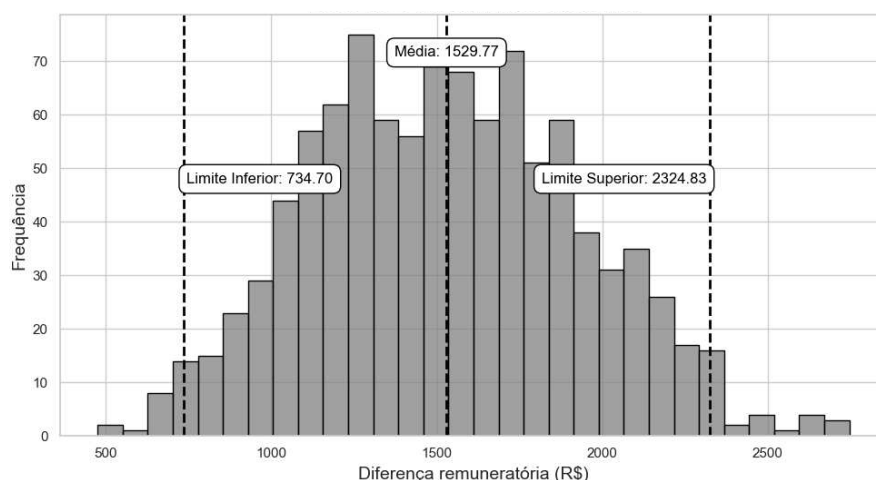
negativa, o intervalo de confiança abrange valores positivos, fazendo com que não se possa afirmar o efeito.

Figura 36. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula que trabalharam para entidades empresariais



Os discentes de instituições sem fins lucrativos apresentaram um ganho significativo, conforme aponta a Figura 36. Apesar disso, é importante ressaltar que esses pares correspondem a apenas 7.6% do total. A amplitude do intervalo de confiança mostra uma grande variabilidade neste grupo.

Figura 37. Histograma da diferença média de remuneração cinco anos depois da matrícula para discentes sem remuneração na matrícula que trabalharam para entidades empresariais



A composição dos pares deste experimento foi composta por aproximadamente 28.7% de funcionários públicos, enquanto o anterior teve 53.9%

de discentes da administração pública. Isso significa que o primeiro estudo, que apontou efeito positivo, tinha uma proporção maior do grupo de funcionários públicos do que este, movendo a média para cima. Essa diferença da proporção parece explicar os resultados do efeito causal, além de confirmar o impacto dos funcionários públicos no resultado agrupado.

7 CONCLUSÕES

Este estudo buscou investigar o retorno em termos remuneratórios para os egressos do Mestrado na área de Computação de programas oferecidos por instituições de ensino brasileiras. Para esse fim, inicialmente foi conduzido uma análise exploratória dos dados de origem pública vindos da CAPES e dados sigilosos do Relatório Anual de Informações Sociais - RAIS. O cruzamento dessas fontes permitiu que essa análise identificasse relações, levantasse novas hipóteses e caracterizasse o discente no momento da sua matrícula, distinguindo os que tiveram sucesso em sua titulação dos demais.

Posteriormente, esses mesmos dados foram usados para conduzir um estudo observacional para compreender se e o quanto a conclusão do Mestrado contribui para a remuneração do discente que completou sua pós-graduação. Os alunos escolhidos para o experimento foram os que recebiam ao menos um salário mínimo tanto no momento da sua matrícula, quanto cinco anos após. Além disso, esses discentes foram matriculados nos anos de 2015 e 2016 e foram comparados olhando essa remuneração no ano de sua matrícula e após cinco anos, com todos os valores corrigidos pelo Índice Nacional de Preço ao Consumidor Amplo (IPCA) para o mesmo momento no tempo - janeiro de 2022. Esse escopo foi limitado pela disponibilidade dos dados obtidos, já que a base remuneratória continha os valores entre 2015 e 2021, fazendo com que apenas alunos matriculados nos anos de 2015 e 2016 contemplassem a janela de tempo conduzida pelo estudo.

A análise exploratória evidenciou a concentração de discentes e de programas na região Sudeste frente às demais, tanto em quantidade quanto na qualidade dos seus programas. Isso ressalta a necessidade urgente de se olhar para as outras regiões, investindo na construção de mão de obra altamente qualificada em todo o país. Espalhar o desenvolvimento de programas de excelência pelo Brasil é essencial para evitar a centralização em um único pólo e promover um crescimento mais equilibrado e inclusivo.

A proporção de discentes com emprego também foi estudada, revelando que aqueles que se titulam estão proporcionalmente menos empregados no momento da matrícula do que os que não conseguem concluir seu Mestrado, possivelmente devido à dificuldade de conciliar estudo e trabalho. Observou-se também que os estudantes das instituições com maior conceito em seus programas

são mais jovens e estão proporcionalmente menos empregados do que os demais grupos de conceito de avaliação, sugerindo que essas instituições podem requerer um tempo de dedicação mais exclusivo ao estudo, algo que talvez não seja viável para pessoas mais velhas.

Analisando especificamente os funcionários públicos, percebe-se que eles estão mais presentes nos programas com avaliação intermediária. No entanto, quando observamos a proporção dentro de cada grupo de conceitos de avaliação, eles têm uma presença maior nos programas de conceito mais baixo. Uma hipótese para tal é que como o impacto da remuneração dos discentes da administração pública não depende da instituição, a busca desses alunos é por programas que permitam maior flexibilidade na sua rotina. Além disso, os funcionários públicos têm uma taxa de conclusão superior à dos discentes que trabalham para entidades privadas e sem fins lucrativos, possivelmente devido à maior facilidade de equilibrar a vida profissional e acadêmica, dispensas e/ou motivação pela relação direta da sua remuneração com a conclusão do curso em alguns casos.

Das pessoas com algum tipo de identificação em relação à sua etnia, pretos e pardos seguem sendo minoria, o que contrasta com a composição da sociedade brasileira. As mulheres também são minoria nos cursos de Mestrado em Computação e, nesse caso, recebem menos do que os homens no momento da matrícula, refletindo um contraste salarial já conhecido na sociedade como um todo. Essas evidências mostraram um acesso ainda limitado desses grupos na área de Computação, indicando a necessidade de políticas mais inclusivas e de desenvolvimento para ampliar essa participação.

Ao fim da análise exploratória, mostrou-se que seria necessária um estudo causal, pois as diferenças nos grupos de discentes que se titularam e não titularam precisavam ter um controle de um experimento para uma melhor compreensão do seu impacto de remuneração. Por isso, o critério de porta dos fundos foi utilizado para selecionar o gênero, faixa etária, grupo de conceito de avaliação do programa, natureza jurídica do empregador do discente, região do programa e grupo remuneratório do discente no momento da matrícula para controlar o estudo observacional que foi conduzido.

O experimento considerou como grupo de tratamento aqueles discentes que concluíram com sucesso seu Mestrado, enquanto o grupo de controle foram aqueles que abandonaram ou foram desligados do seu curso. Através do método de

pareamento *coarsened exact matching* (CEM), concluiu-se que existe um efeito de 18.2% na remuneração do discente que completa seu Mestrado e que ganha ao menos um salário mínimo no momento da sua matrícula e após cinco anos.

O resultado total foi desdobrado para compreender melhor quais variáveis estavam trazendo impacto em relação ao mestrado. Notou-se um aumento significativo no grupo de funcionários públicos em comparação aos demais, levantando a questão de um possível alinhamento entre as necessidades da indústria privada e da academia.

Os discentes que eram funcionários públicos no momento da sua matrícula impactaram positivamente o efeito, enquanto não houve evidências de que os funcionários de entidades privadas e de instituições sem fins lucrativos tenham tido algum efeito positivo com a conclusão do Mestrado após cinco anos da matrícula.

O grupo com remuneração mais baixa no momento da matrícula não apresentou ganhos significativos após o Mestrado, sugerindo que podem estar envolvidos em trabalhos mais operacionais, com poucas oportunidades para aplicar o conhecimento adquirido. De maneira semelhante, a menor faixa etária também não teve ganho significativo, possivelmente por estar em um momento mais júnior de suas carreiras.

Além disso, os discentes de programas com conceitos de avaliação mais altos não tiveram ganhos significativos que superassem os de programas com conceitos mais baixos, o que pode ser explicado pela expectativa de seguirem para o Doutorado, por exemplo. Assim, a janela escolhida para o estudo causal não foi tempo suficiente para observar um salto significativo em suas carreiras. Outro ponto é que, conforme visto na EDA, a proporção de funcionários públicos desse grupo é menor, o que não impulsiona o efeito da remuneração para esse recorte de instituições.

Finalmente, um estudo similar foi conduzido, porém considerando apenas discentes sem remuneração na matrícula, mas com remuneração de no mínimo um salário mínimo após cinco anos. Não foi possível ter evidências do efeito causal para esse grupo, porém os discentes que viraram funcionários públicos após cinco anos tiveram um efeito positivo, diferente dos que foram para o caminho de entidades empresariais. Esse comportamento reforçou o achado do primeiro estudo.

7.1 Contribuições

O estudo levantou hipóteses e discussões sobre o perfil do discente de Mestrado na área de Computação, assim como a sua caracterização através da análise exploratória de dados. Foram avaliados aspectos sobre a regionalização dos programas, a concentração de discentes por quantidade e qualidade, o fator da idade na conclusão do curso, a proporção de discentes com emprego no início da formação e seu possível impacto na conclusão, a relação do funcionário público com o Mestrado, além das presenças das minorias na área de Computação ao longo dos anos. Todos esses pontos debatidos visam colaborar para a contínua evolução da participação da Pós-Graduação na sociedade, em específico da área de Computação, tão relacionado ao desenvolvimento de um país.

Em adição a isso, a análise exploratória de dados indicou a necessidade de conduzir uma análise causal para estimar o efeito do Mestrado na área de Computação sobre a remuneração dos discentes. O resultado apontou um efeito de 18.2% na remuneração média após cinco anos em relação à matrícula para os alunos que possuíam remuneração de ao menos um salário mínimo no seu início de curso e também após cinco anos. O método consistiu em discutir as variáveis confundidoras para reduzir o viés de seleção do experimento e aplicar o coarsened exact matching (CEM) para a estimação do efeito.

Outra contribuição vem com o desdobramento desse efeito da remuneração para avaliar quais recortes de discentes foram mais impactados. Entre as descobertas, debates como o menor efeito de discentes que trabalham em entidades privadas frente aos funcionários públicos possibilita uma discussão sobre o alinhamento da indústria e a academia dentro da área de Computação, por exemplo.

7.2 Limitações

Apesar de explicitar com cuidado cada restrição e condição assumida tanto para a análise exploratória dos dados quanto para a análise causal ao longo do trabalho, é relevante ressaltar que todas as hipóteses levantadas, debates e relações discutidas se aplicam a um contexto específico. Mesmo assim, alguns pontos merecem ser citados nessa seção para garantir a compreensão.

Os dados sigilosos disponibilizados pela Relação Anual de Informações Sociais (RAIS) contam com as remunerações dos discentes com vínculos formais de trabalho, sejam CLT ou estatutários. Isso faz com que os alunos que optaram pela vida empreendedora sejam excluídos da análise exploratória e causal. Além disso, na área de tecnologia em geral, é comum que muitos profissionais atuem como Pessoa Jurídica (PJ), seja no Brasil ou fora do país. Esses casos também se enquadram na limitação do estudo.

Outra limitação é em relação à janela de dados analisada. Para a análise exploratória de dados, observou-se os dados no momento da matrícula entre 2015 e 2019 por questão de disponibilidade, o que significa que o comportamento nos anos após isso pode ter mudado. Já para a análise causal, optou-se por considerar o impacto após cinco anos a matrícula do discente, o que fez com que apenas os matriculados em 2015 e 2016 fizessem parte do experimento.

Por fim, a análise causal considerou apenas os discentes que estavam sendo remunerados com ao menos um salário mínimo em relação a janeiro de 2022 dado a correção via IPCA. Esse recorte não garante as mesmas discussões para alunos em outra situação, como por exemplo, os que não estão trabalhando no momento de sua matrícula e/ou após cinco anos.

7.3 Trabalhos futuros

O estudo permite uma continuidade em diferentes frentes. O mesmo conjunto de dados pode ser usado para explorar outras condições iniciais do estudante do Mestrado em Computação, tal como os que iniciaram suas trajetórias acadêmicas sem remuneração e obtiveram um emprego ao longo do tempo. Esse ponto poderia ser abordado tanto da perspectiva do Mestrado, Mestrado Profissional ou Doutorado.

A evolução do efeito remuneratório ao longo do tempo é uma outra oportunidade, entretanto seria relevante conseguir um conjunto de dados com uma janela temporal maior do que a obtida para o presente trabalho. Da mesma forma que anteriormente discutido, uma contribuição interessante seria a comparação desse ganho remuneratório de acordo com diferentes graus e regiões do Brasil.

Do ponto metodológico, novos estudos podem ser conduzidos em condição similar, porém aplicando outros métodos para comparar o resultado, em especial no

processo de *matching*. A avaliação desses resultados traria uma contribuição sobre a estabilidade das conclusões obtidas no que foi apresentado.

Finalmente, os dados vindos da RAIS limitam o acesso à remuneração de vínculos trabalhistas, porém seria relevante enriquecer o conjunto de dados com uma fonte alternativa que trouxesse a remuneração dos empreendedores. Assim, seria possível desdobrar o trabalho também comparando o sucesso financeiro das trajetórias CLT e PJ.

REFERÊNCIAS

- ALPAYDIN, Yusuf; KÜLTÜR, Kürşad. Improving the transition from higher education to employment: A review of current policies. **Education Policies in the 21st Century: Comparative Perspectives**, p. 45-60, 2022.
- ALVARES, Reinaldo V.; LOUTFI, Marcelo S.; CAMPOS, Nathielly de S. Onde estão meus egressos? Relato sobre um mapeamento automatizado da vida profissional dos formados em Sistemas de Informação da UNIRIO. In: **Anais do XXVIII Workshop sobre Educação em Computação**. SBC, 2020. p. 56-60.
- ARDAGNA, Claudio A.; CERAVOLO, Paolo; DAMIANI, Ernesto. Big data analytics as-a-service: Issues and challenges. In: **2016 IEEE international conference on big data (big data)**. IEEE, 2016. p. 3638-3644.
- BALBACHEVSKY, Elizabeth. A pós-graduação no Brasil: novos desafios para uma política bem-sucedida. **Os desafios da educação no Brasil**. Rio de Janeiro: Nova Fronteira, v. 1, p. 285-314, 2005.
- BARBOSA FILHO, Fernando de Holanda; PESSOA, Samuel de Abreu. Retorno da educação no Brasil. 2008.
- BLACKWELL, Matthew et al. cem: Coarsened exact matching in Stata. **The Stata Journal**, v. 9, n. 4, p. 524-546, 2009.
- BORMAN, Geoffrey D. Experiments for educational evaluation and improvement. In: **Reevaluating Evaluation**. Routledge, 2018. p. 7-27.
- CAPPELLI, Peter H. Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States. **ILR review**, v. 68, n. 2, p. 251-290, 2015.
- CHATFIELD, Chris. Exploratory data analysis. **European journal of operational research**, v. 23, n. 1, p. 5-13, 1986.
- COELHO, Vanilda de Oliveira. Retorno financeiro à educação superior da Universidade de Brasília: análise a partir dos egressos de graduação. 2017.
- CHRISTENSEN, Clayton M.; EYRING, Henry J. **The innovative university: Changing the DNA of higher education from the inside out**. John Wiley & Sons, 2011.

CUNHA, Luiz Antonio CR. A pós-graduação no Brasil: função técnica e função social. **Revista de Administração de Empresas**, v. 14, p. 66-70, 1974.

DONHARDT, Gary L. In search of the effects of academic achievement in post graduation earnings. **Research in Higher Education**, v. 45, p. 271-284, 2004.

DUFLO, Esther; GLENNERSTER, Rachel; KREMER, Michael. Using randomization in development economics research: A toolkit. **Handbook of development economics**, v. 4, p. 3895-3962, 2007.

ENGSTROM, Elyne Montenegro; HORTALE, Virginia Alonso; MOREIRA, Carlos Otávio Fiuza. Trajetória profissional de egressos de Curso de Mestrado Profissional em Atenção Primária à Saúde no Município de Rio de Janeiro, Brasil: estudo avaliativo. **Ciência & Saúde Coletiva**, v. 25, p. 1269-1280, 2020.

EVERGREEN, Stephanie DH. **Effective data visualization: The right chart for the right data**. SAGE publications, 2019.

FAVATO, Maria Nilse; RUIZ, Maria José Ferreira. REUNI: política para a democratização da educação superior. **Revista eletrônica de educação**, v. 12, n. 2, p. 448-463, 2018.

FERREIRA, Mércia et al. Panorama comparativo da atuação acadêmica e profissional de egressas e egressos de um curso de bacharelado em ciência da computação. In: **Anais do XXX Workshop sobre Educação em Computação**. SBC, 2022. p. 227-238.

FERREIRA, Francisco Danilo da Silva; DE ALMEIDA, Aléssio Tony Cavalcanti. Ação Afirmativa e Ganhos Iniciais Após a Graduação: Uma Análise Para Egressos da UFPB. **Planejamento e Políticas Públicas**, n. 64, 2022.

FEW, Stephen. **Now you see it: simple visualization techniques for quantitative analysis**. Analytics Press, 2009.

GELMAN, Andrew. **Data analysis using regression and multilevel/hierarchical models**. Cambridge University Press, 2007.

GERARD, François et al. Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil. **American Economic Review**, v. 111, n. 10, p. 3418-3457, 2021.

GRIBOSKI, CLAUDIA; BEDRITCHUK, AMANDA GUEDES ANDRADE; FERREIRA, GUILHERME VIANA. Autoavaliação institucional: Uma análise da formação e inserção profissional dos egressos da UNB. 2017.

HARITON, Eduardo; LOCASCIO, Joseph J. Randomised controlled trials—the gold standard for effectiveness research. **BJOG: an international journal of obstetrics and gynaecology**, v. 125, n. 13, p. 1716, 2018.

HERNÁN, Miguel A.; ROBINS, James M. Causal inference. 2010.

HO, Daniel E. et al. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. **Political analysis**, v. 15, n. 3, p. 199-236, 2007.

HOAGLIN, David C.; MOSTELLER, Frederick; TUKEY, John W. (Ed.). **Understanding robust and exploratory data analysis**. John Wiley & Sons, 2000.

IACUS, Stefano M.; KING, Gary; PORRO, Giuseppe. Causal inference without balance checking: Coarsened exact matching. **Political analysis**, v. 20, n. 1, p. 1-24, 2012.

IMBENS, Guido W.; RUBIN, Donald B. **Causal inference in statistics, social, and biomedical sciences**. Cambridge university press, 2015.

JAMES, G. An Introduction to Statistical Learning. 2013.

KAPLAN, Andreas. **Higher education at the crossroads of disruption: The university of the 21st century**. Emerald Publishing Limited, 2021.

KING, Gary. Matching Methods for Causal Inference. **Institute for Quantitative Social Science, Harvard University**, 2018.

KING, Gary; KEOHANE, Robert O.; VERBA, Sidney. **Designing social inquiry: Scientific inference in qualitative research**. Princeton university press, 2021.

KUBOTA, Luis Claudio; ROSA, Maurício Benedeti. Adoção de tecnologias da indústria 4.0 por empresas brasileiras. 2024.

LEMIEUX, Thomas. The “Mincer equation” thirty years after schooling, experience, and earnings. In: **Jacob Mincer a pioneer of modern labor economics**. Boston, MA: Springer US, 2006. p. 127-145.

LIMA, Diego de Faveri Pereira. **Quanto vale uma pós-graduação stricto-sensu no Brasil?: os efeitos do mestrado e doutorado na remuneração de seus egressos**. 2016. Tese de Doutorado.

MARTINEZ, Wendy L.; MARTINEZ, Angel R. **Computational statistics handbook with MATLAB**. Chapman and Hall/CRC, 2001.

MORGAN, Stephen L.; WINSHIP, Christopher. **Counterfactuals and causal inference**. Cambridge University Press, 2015.

MOROSINI, Marília Costa. A pós-graduação no Brasil: formação e desafios. **Revista Argentina de Educación Superior: RAES**, n. 1, p. 125-152, 2009.

MYATT, Glenn J. **Making sense of data: a practical guide to exploratory data analysis and data mining**. John Wiley & Sons, 2007.

OLIVEIRA, Guilherme Resende et al. Efeitos do programa de qualificação profissional Bolsa Futuro na empregabilidade e salário dos trabalhadores. **Pesquisa e Planejamento Econômico**, v. 49, n. 1, 2019.

PEARL, Judea et al. Models, reasoning and inference. **Cambridge, UK: Cambridge University Press**, v. 19, n. 2, p. 3, 2000.

POWELL, Walter W.; SNELLMAN, Kaisa. The knowledge economy. **Annu. Rev. Sociol.**, v. 30, n. 1, p. 199-220, 2004.

PROLO, Ivor et al. Internacionalização das universidades brasileiras-contribuições do programa ciência sem fronteiras. **Administração: Ensino e Pesquisa**, v. 20, n. 2, p. 319-361, 2019.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. " O'Reilly Media, Inc.", 2013.

ROCHA, Wilsimara M.; EHRL, Philipp; MONASTERIO, Leonardo M. Análise de impacto do FIES sobre o salário do trabalhador formal. **44º Encontro Nacional de Economia-Anpec**, 2016.

ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41-55, 1983.

SALIBA, Nemre Adas et al. Percepção do cirurgião-dentista sobre formação profissional e dificuldades de inserção no mercado de trabalho. **Revista de Odontologia da UNESP**, v. 41, p. 297-304, 2012.

SALLES, William das Neves; FARIAS, Gelcemar Oliveira; NASCIMENTO, Juarez Vieira do. Inserção profissional e formação continuada de egressos de cursos de graduação em Educação Física. **Revista Brasileira de Educação Física e Esporte**, v. 29, p. 475-486, 2015.

SILVEIRA, Carlos Eduardo; MEDAGLIA, Juliana; NAKATANI, Marcia Shizue Massukado. O mercado de trabalho dos egressos de cursos superiores em turismo: comparações dos dados de 2012-2018. **Revista Brasileira de Pesquisa em Turismo**, v. 14, p. 83-94, 2020.

SOUZA, Norma Valéria Dantas de Oliveira et al. O egresso de enfermagem da FENF/UERJ no mundo do trabalho. **Revista da Escola de Enfermagem da USP**, v. 45, p. 250-257, 2011.

STUART, Elizabeth A. Matching methods for causal inference: A review and a look forward. **Statistical science: a review journal of the Institute of Mathematical Statistics**, v. 25, n. 1, p. 1, 2010.

TUFTE, Edward R.; GRAVES-MORRIS, Peter R. **The visual display of quantitative information**. Cheshire, CT: Graphics press, 1983.

VIGNI, Mario Li; DURANTE, Caterina; COCCHI, Marina. Exploratory data analysis. In: Data handling in science and technology. Elsevier, 2013. p. 55-126.

WEISS, Andrew. Human capital vs. signalling explanations of wages. **Journal of Economic perspectives**, v. 9, n. 4, p. 133-154, 1995.

WITTEVEEN, Dirk; ATTEWELL, Paul. Delayed time-to-degree and post-college earnings. **Research in Higher Education**, v. 62, p. 230-257, 2021.