

UNIVERSIDADE FEDERAL DO PARANÁ

CIBELE DE BIASI DA SILVA

ENSAIOS SOBRE DESIGUALDADE DE GÊNERO NO MERCADO DE TRABALHO

CURITIBA

2024

CIBELE DE BIASI DA SILVA

ENSAIOS SOBRE DEGIGUALDADE DE GÊNERO NO MERCADO DE TRABALHO

Tese de doutorado apresentada ao curso de Pós-Graduação em Desenvolvimento Econômico, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutora em Desenvolvimento Econômico.

Orientadora: Prof<sup>ª</sup>. Dra. Kênia Barreiro de Souza

CURITIBA

2024

DADOS INTERNACIONAIS DE CATALOÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Silva, Cibele De Biasi da  
Ensaio sobre a desigualdade de gênero no mercado de trabalho / Cibele  
De Biasi da Silva. – 2024.  
1 recurso on-line: PDF.

Tese (doutorado) - Universidade Federal do Paraná, Setor de  
Ciências Sociais Aplicadas, Programa de Pós-Graduação em  
Desenvolvimento Econômico.  
Orientadora: Kênia Berreiro de Souza.

1. Desenvolvimento Econômico. 2. Desigualdade de gênero. 3. Academia.  
4. Desempenho. 5. Mercado de trabalho. I. Souza, Kênia Berreiro de. II.  
Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas.  
Programa de Pós-Graduação em Desenvolvimento Econômico. III. Título.

Bibliotecário Eduardo Silveira – CRB – 9/1921

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação DESENVOLVIMENTO ECONÔMICO da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **CIBELE DE BIASI DA SILVA** intitulada: **ENSAIOS SOBRE DESIGUALDADE DE GÊNERO NO MERCADO DE TRABALHO**, sob orientação da Profa. Dra. KÊNIA BARREIRO DE SOUZA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua **APROVAÇÃO** no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 01 de Março de 2024.

Assinatura Eletrônica  
01/03/2024 18:21:27.0  
KÊNIA BARREIRO DE SOUZA  
Presidente da Banca Examinadora

Assinatura Eletrônica  
01/03/2024 20:20:07.0  
DANIELA VERZOLA VAZ  
Avaliador Externo (UNIVERSIDADE FEDERAL DE SÃO PAULO)

Assinatura Eletrônica  
01/03/2024 18:50:22.0  
DANIELLE CARUSI MACHADO  
Avaliador Externo (UNIVERSIDADE FEDERAL FLUMINENSE)

Assinatura Eletrônica  
01/03/2024 19:41:22.0  
VICTOR RODRIGUES DE OLIVEIRA  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

## RESUMO

Um dos fatores necessários para o desenvolvimento econômico de um país é a garantia de igualdade nas oportunidades de acesso ao mercado de trabalho entre mulheres e homens. Entretanto, mesmo quando há condições iguais para a participação, não necessariamente se observa igualdade na remuneração, com as mulheres invariavelmente ganhando menos do que os homens. Diante desse contexto, o objetivo geral dos três ensaios que constituem esta tese de doutorado é analisar a desigualdade de gênero no mercado de trabalho, tendo como foco as questões relacionadas à segregação ocupacional em carreiras científicas, à relação entre desempenho acadêmico e resultados no mercado de trabalho e ao *mismatch* educacional. Mais especificamente, o primeiro ensaio investiga se o prêmio salarial, medido pelo salário recebido por hora, é diferente entre mulheres e homens que declararam estar empregados em alguma função do grande grupo de “Profissionais das Ciências e Intelectuais” da Pesquisa Nacional por Amostra de Domicílios Contínua entre 2016 e 2022 a partir do método de *Propensity Score Matching*. Os resultados encontrados sugerem a existência de um prêmio salarial e esse valor é diferente entre os gêneros, uma vez que as mulheres recebem um prêmio aproximadamente 55% inferior quando comparadas aos homens. Na tentativa de analisar a robustez dessas respostas, sete cenários foram investigados e os resultados encontrados corroboram, entre outras coisas, a segregação ocupacional das mulheres em ocupações que pagam menores salários e que são menos valorizadas no mercado de trabalho, em especial àquelas que se relacionam aos cuidados. O segundo ensaio, por sua vez, analisa a relação entre o desempenho acadêmico e os resultados no mercado de trabalho. A partir da integração de dados administrativos de discentes egressos da Universidade Federal do Paraná e de informações da Relação Anual de Informações Sociais (RAIS) identificada de 2003 a 2021, busca-se averiguar se o desempenho de estudantes durante seus cursos de graduação possui um viés de gênero e se esse viés afeta os seus salários no mercado de trabalho. Os resultados encontrados apontam evidências estatísticas de que os indivíduos, tanto mulheres quanto homens, que apresentaram melhores performances acadêmicas no ensino superior também apresentaram melhores retornos no mercado de trabalho em termos salariais. Entretanto, constataram-se diferenças entre os resultados quando as áreas de formação são levadas em consideração: Saúde, Engenharia e Sociais Aplicadas apresentaram valores estatisticamente significativos e positivos para o retorno salarial enquanto a área de Biológicas apresentou uma relação negativa entre notas e salários. Por fim, o terceiro ensaio investiga o *mismatch* educacional entre homens e mulheres no mercado de trabalho formal brasileiro. A partir de um painel de dados que integrou informações de egressos da UFPR, como área de formação, às suas informações no mercado de trabalho formal da base de dados da Relação Anual de Informações Sociais (RAIS) identificada entre 2003 e 2021, como ocupação e remuneração, os resultados encontrados apontam que o *mismatch* educacional influencia negativamente os resultados no mercado de trabalho para todos os indivíduos, mas que as mulheres são ligeiramente mais penalizadas do que os homens. Além disso, quando as análises são feitas separadamente para cada área, os resultados indicam que áreas caracterizadas por habilidades mais específicas, como a Saúde e Engenharias, apresentam efeitos negativos significativamente mais acentuados no que diz respeito ao *mismatch* horizontal em comparação às demais áreas de habilidades menos específicos, como Humanas, Linguística; Letras e Artes e Sociais Aplicadas.

**Palavras-chave:** Segregação ocupacional. Desempenho acadêmico. *Mismatch* educacional. Desigualdade de gênero.

## ABSTRACT

One of the necessary factors for the economic development of a country is ensuring equal opportunities for access to the labor market between women and men. However, even when there are equal conditions for participation, equality in compensation is not necessarily observed, with women invariably earning less than men. In this context, the general objective of the three essays that constitute this doctoral dissertation is to analyze gender inequality in the labor market, focusing on issues related to occupational segregation in scientific careers, the relationship between academic performance and outcomes in the labor market, and educational mismatch. Specifically, the first essay investigates whether the wage premium, measured by the hourly wage received, differs between women and men who declared being employed in some function within the broad group of "Professionals in Sciences and Intellectuals" from the Continuous National Household Sample Survey between 2016 and 2022 using the Propensity Score Matching method. The findings suggest the existence of a wage premium, and this value differs between genders, as women receive a premium approximately 55% lower when compared to men. In an attempt to analyze the robustness of these responses, seven scenarios were investigated, and the findings corroborate, among other things, the occupational segregation of women in occupations that pay lower wages and are less valued in the labor market, especially those related to care. The second essay analyzes the relationship between academic performance and outcomes in the labor market. By integrating administrative data from graduates of the Federal University of Paraná and information from the identified Annual Social Information Report (RAIS) from 2003 to 2021, it seeks to ascertain whether the performance of students during their undergraduate courses has a gender bias and whether this bias affects their salaries in the labor market. The findings point to statistical evidence that individuals, both women and men, who presented better academic performances in higher education also had better returns in the labor market in terms of salaries. However, differences were observed in the results when considering the fields of study: Health, Engineering, and Applied Social Sciences showed statistically significant and positive values for salary returns, while the Biological Sciences field showed a negative relationship between grades and salaries. Finally, the third essay investigates educational mismatch between men and women in the formal Brazilian labor market. Using a panel of data that integrated information from UFPR graduates, such as field of study, with their information in the formal labor market from the database of the identified Annual Social Information Report (RAIS) from 2003 to 2021, such as occupation and remuneration, the findings indicate that educational mismatch negatively influences outcomes in the labor market for all individuals, but women are slightly more penalized than men. Additionally, when analyses are conducted separately for each field, the results indicate that areas characterized by more specific skills, such as Health and Engineering, present significantly more pronounced negative effects regarding horizontal mismatch compared to other areas with less specific skills, such as Humanities, Linguistics, Letters and Arts, and Applied Social Sciences.

**Keywords:** Occupational Segregation. Academic performance. Educational mismatch. Gender inequality.

## LISTA DE FIGURAS

Gráfico 1 – Densidade da distribuição dos valores dos propensity scores para os grupos de controle e tratamento antes do pareamento – Mulheres.....	36
Gráfico 2 – Densidade da distribuição dos valores dos propensity scores para os grupos de controle e tratamento antes do pareamento - Homens .....	36
Gráfico 3 – Densidade da distribuição dos valores dos propensity scores para os grupos de controle e tratamento antes do pareamento - Todos .....	37
Gráfico 4 – Densidade da distribuição dos valores dos propensity scores para cada grupo antes e depois do pareamento – Mulheres.....	44
Gráfico 5 – Densidade da distribuição dos valores dos propensity scores para cada grupo antes e depois do pareamento – Homens .....	44
Gráfico 6 – Densidade da distribuição dos valores dos propensity scores para cada grupo antes e depois do pareamento - Todos .....	44
Gráfico 7 – Densidade da média das notas dos alunos por gênero.....	73
Gráfico 8 – Densidade da média das notas normalizadas por curso para mulheres e homens	73
Gráfico 9 – Posição da média normalizada de cada gênero por área mais abrangente .....	74
Gráfico 10 – Posição da média normalizada de cada gênero por curso.....	75

## LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Mulheres .....	29
Tabela 2 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Homens.....	30
Tabela 3 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Todos .....	31
Tabela 4 – Resultados antes do pareamento – Mulheres .....	34
Tabela 5 – Resultados antes do pareamento – Homens .....	34
Tabela 6 – Resultados antes do pareamento – Todos .....	35
Tabela 7 – Resultados após o pareamento – Mulheres .....	38
Tabela 8 – Resultados após o pareamento – Homens.....	39
Tabela 9 – Resultados após o pareamento – Todos.....	40
Tabela 10 – Balanceamento das covariáveis após o pareamento – Mulheres.....	42
Tabela 11 – Balanceamento das covariáveis após o pareamento – Homens.....	42
Tabela 12 – Balanceamento das covariáveis após o pareamento – Todos.....	43
Tabela 13 – Efeito médio do tratamento sobre os tratados (ATT) .....	45
Tabela 14 – Efeito médio do tratamento sobre os tratados para o décimo decil de renda.....	47
Tabela 15 – Efeito médio do tratamento sobre os tratados considerando a base de dados apenas com indivíduos que declararam ter formação superior .....	48
Tabela 16 – Estatísticas descritivas para o grande grupo de “Profissionais das Ciências e Intelectuais” .....	49
Tabela 17 – Estatísticas descritivas para o subgrupo principal “Profissionais do Ensino” .....	51
Tabela 18 – Estatísticas descritivas para o subgrupo principal “Profissionais da Saúde” .....	53
Tabela 19 – Efeito médio do tratamento sobre os tratados considerando a base de dados sem os subgrupos do Ensino que auferem menores níveis salariais .....	54
Tabela 20 – Efeito médio do tratamento sobre os tratados considerando a base de dados sem os subgrupos da Saúde que auferem menores níveis salariais .....	55
Tabela 21 – Efeito médio do tratamento sobre os tratados para os subgrupos individualmente .....	56
Tabela 22 – Efeito médio do tratamento sobre os tratados separadamente por jornada de trabalho.....	58

Tabela 23 – Efeito médio do tratamento sobre os tratados separadamente por esfera de trabalho.....	59
Tabela 24 – Efeito médio do tratamento sobre os tratados considerando a base de dados entre os anos de 2016 e 2019 .....	59
Tabela 25 – Estatísticas descritivas e Teste t – base de dados UFPR.....	70
Tabela 26 – Estatísticas descritivas e Teste t - base de dados UFPR e RAIS .....	76
Tabela 27 – Resultados das estimações considerando todos os indivíduos .....	86
Tabela 28 – Resultados das estimações para as mulheres.....	88
Tabela 29 – Resultados das estimações para os homens .....	89
Tabela 30 – Resultados das estimações considerando todos os indivíduos formados.....	91
Tabela 31 – Resultados das estimações individuais para as mulheres formadas .....	92
Tabela 32 – Resultados das estimações individuais para os homens formados .....	92
Tabela 33 – Resultados das estimações para todos os indivíduos com outras variáveis de desempenho .....	93
Tabela 34 – Resultados das estimações para todos os indivíduos para cada área individualmente .....	95
Tabela 35 - Estatísticas descritivas e teste t .....	110
Tabela 36 - Resultados das estimações do painel de dados com efeitos fixos.....	112
Tabela 37 - Resultados das estimações por gênero e área do conhecimento .....	114

## LISTA DE QUADROS

Quadro 1 – Descrição das variáveis.....	24
Quadro 2 – Descrição das variáveis.....	80
Quadro 3 – Descrição das variáveis.....	107

## LISTA DE ABREVIATURAS E DE SIGAS

AGTIC	Agência de Tecnologia da Informação e Comunicação
ATE	<i>Population average treatment effect</i>
ATT	<i>Average treatment effect on the treated</i>
CBO	Classificação Brasileira de Ocupações
NN	<i>Nearest Neighbor</i>
PNADC	Pesquisa Nacional por Amostra de Domicílios Contínua
PSM	<i>Propensity Score Matching</i>
RAIS	Relação Anual de Informações Sociais
STEM	<i>Science, Technology, Engineering, Math</i>
UFPR	Universidade Federal do Paraná

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO GERAL .....</b>	<b>14</b>
<b>2</b>	<b>ENSAIO 1 – A SEGREGAÇÃO OCUPACIONAL ENTRE MULHERES E HOMENS EM CARREIRAS CIENTÍFICAS NO BRASIL (2016-2022).....</b>	<b>16</b>
2.1	INTRODUÇÃO.....	16
2.2	REVISÃO DE LITERATURA.....	18
2.2.1	As mulheres nas ciências e a segregação ocupacional .....	18
2.3	METODOLOGIA E TRATAMENTO DE DADOS.....	23
2.3.1	Base de dados e recorte amostral.....	24
2.3.2	Propensity score matching .....	26
2.4	RESULTADOS E DISCUSSÕES.....	29
2.4.1	Estatísticas descritivas .....	29
2.4.2	Propensity Score Matching .....	33
2.4.3	Análise de robustez.....	45
2.5	CONSIDERAÇÕES FINAIS.....	60
<b>3</b>	<b>ENSAIO 2 – A DEDICAÇÃO VALEU A PENA? UMA ANÁLISE DA RELAÇÃO ENTRE GÊNERO, DESEMPENHO ACADÊMICO E MERCADO DE TRABALHO FORMAL .....</b>	<b>62</b>
3.1	INTRODUÇÃO.....	62
3.2	REVISÃO DE LITERATURA.....	64
3.2.1	A teoria do capital humano e da sinalização e o desempenho acadêmico .....	65
3.3	METODOLOGIA E TRATAMENTO DE DADOS.....	69
3.3.1	Bases de dados e recorte amostral .....	69
3.3.2	Descrição das variáveis e das hipóteses.....	80
3.3.3	Correção de Heckman.....	81
3.4	RESULTADOS E DISCUSSÕES.....	84
3.4.1	Análise de robustez e heterogeneidade .....	90

3.4.2	Limitações metodológicas.....	96
<b>3.5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>97</b>
<b>4</b>	<b>ENSAIO 3: AS IMPLICAÇÕES DO MISMATCH EDUCACIONAL ENTRE MULHERES E HOMENS NO MERCADO DE TRABALHO FORMAL BRASILEIRO</b>	<b>99</b>
<b>4.1</b>	<b>INTRODUÇÃO.....</b>	<b>99</b>
<b>4.2</b>	<b>REVISÃO DE LITERATURA.....</b>	<b>101</b>
4.2.1	Literatura internacional.....	101
4.2.2	Literatura nacional.....	103
<b>4.3</b>	<b>METODOLOGIA E TRATAMENTO DE DADOS.....</b>	<b>105</b>
4.3.1	Base de dados e recorte amostral.....	105
4.3.2	Descrição das variáveis e das hipóteses.....	107
4.3.3	Mismatches vertical e horizontal.....	107
4.3.4	Estratégia de identificação e modelo empírico.....	108
<b>4.4</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>109</b>
4.4.1	Estatísticas descritivas.....	110
4.4.2	Resultados das estimações do painel de dados.....	111
4.4.3	Análise de heterogeneidade.....	113
<b>4.5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>115</b>
<b>5</b>	<b>CONCLUSÃO GERAL.....</b>	<b>117</b>
	<b>REFERÊNCIAS.....</b>	<b>120</b>

## 1 INTRODUÇÃO GERAL

O estudo sobre as mulheres e o trabalho, seja ele remunerado ou não, é de fundamental importância e ganhou mais visibilidade recentemente. Em outubro de 2023, por suas contribuições acerca da dinâmica das mulheres no mercado de trabalho, Claudia Goldin, professora e pesquisadora da Universidade de Harvard, foi laureada com a maior condecoração da área de Economia, o Prêmio do Banco da Suécia para as Ciências Econômicas em Memória de Alfred Nobel. Ela foi a terceira mulher a receber o prêmio desde 1969, quando ele passou a ser concedido para a área de Economia, e a primeira a recebe-lo sozinha. Em novembro do mesmo ano, o tema da redação do Exame Nacional do Ensino Médio (ENEM) no Brasil trouxe à tona o debate sobre o trabalho feminino invisível do cuidado.

Apesar da visibilidade do tema, da elevação da participação feminina na força de trabalho e da diminuição nas diferenças salariais entre os gêneros, ainda são inúmeros os fatores que continuam contribuindo para que as mulheres permaneçam em situação de desvantagem no que diz respeito às suas escolhas – ou imposições – profissionais.

Primeiramente, pode-se citar a dupla jornada de trabalho. Essencial para a sustentação e reprodução da vida humana e ainda assim pouco reconhecido, o cuidado da casa, dos filhos e da família é majoritariamente concentrado nas mulheres. Conforme destacado pela filósofa italiana Silvia Federici (2019), o trabalho doméstico não é amor, é trabalho não remunerado. De acordo com dados do IBGE (2023) para o Brasil, em 2022, as mulheres dedicavam cerca de 9,6 horas semanais a mais do que os homens aos afazeres domésticos e/ou cuidado de pessoas e essa diferença ainda permanecia mesmo quando apenas os indivíduos ocupados são considerados: em média, a parcela feminina ocupada dedicava 6,8 horas a mais por semana do que a masculinas com o trabalho não remunerado.

As normas sociais que atribuem esse trabalho invisível às mulheres muitas vezes também agem para que no mercado de trabalho elas se concentrem em profissões que são entendidas como extensão da casa, frequentemente relacionadas ao cuidado. Esse é um dos fatores que fomentam o problema da segregação ocupacional entre os gêneros, com as mulheres concentradas em funções que remuneram menos e que oferecem carreiras profissionais menos atrativas (BORROWMAN; KLASSEN, 2019).

Além disso, na tentativa de conciliar a dupla jornada de trabalho, muitas mulheres acabam optando por empregos que oferecem jornadas de trabalho mais flexíveis. A importância da flexibilidade da jornada de trabalho é uma das contribuições centrais do conjunto da obra de

Goldin. Quando a flexibilidade de horários em empregos de período integral não é uma opção, muitas mulheres acabam optando por trabalhos de meio período. Isso faz com que elas, frequentemente, acabem ocupando cargos que estão abaixo de sua qualificação (*mismatch* educacional) e que remuneram menos, mas que oferecem a flexibilidade de que necessitam. (GOLDIN, KATZ, 2011; GOLDIN, 2014).

Diante desse breve contexto, o objetivo geral dos três ensaios que constituem essa tese de doutorado é analisar a desigualdade de gênero entre indivíduos com grau superior<sup>1</sup>, tendo como norte importantes questões levantadas pela literatura como fatores que ainda contribuem para que as mulheres permaneçam em posição de desvantagem no mercado de trabalho, como a segregação ocupacional em carreiras científicas, a relação entre desempenho acadêmico e resultados no mercado de trabalho e o *mismatch* educacional.

Além dessa introdução geral, a presente tese de doutorado está organizada em outros três capítulos. O segundo capítulo diz respeito ao primeiro ensaio, intitulado “*A segregação ocupacional entre mulheres e homens em carreiras científicas no Brasil (2016-2021)*”, e se debruça sobre as questões relacionadas à segregação ocupacional.

O terceiro capítulo, por sua vez, cujo título é “*A dedicação valeu a pena? Uma análise da relação entre gênero, desempenho acadêmico e mercado de trabalho formal*”, apresenta o segundo ensaio, que objetiva investigar se o desempenho acadêmico no ensino superior possui um viés de gênero, e se esse viés se traduz, ou não, em melhores resultados no mercado de trabalho.

Por fim, o quarto capítulo traz o terceiro ensaio, intitulado “*As implicações do mismatch educacional entre mulheres e homens no mercado de trabalho formal brasileiro*”, que investiga acerca dos efeitos do *mismatch* educacional entre homens e mulheres.

---

<sup>1</sup> Muito embora os termos gênero e sexo possam ser entendidos como sinônimos, optou-se por priorizar a utilização do primeiro em concordância com Perez (2022, p.13): “Por “sexo”, entendo as características biológicas que determinam se uma pessoa nasce homem ou mulher. XX e XY. Por “gênero”, entendo os significados sociais que são impostos a esses fatores biológicos – o modo como as mulheres são tratadas por serem percebidas como mulheres. Um desses conceitos é obra humana, mas ambos são reais.”

## 2 ENSAIO 1 – A SEGREGAÇÃO OCUPACIONAL ENTRE MULHERES E HOMENS EM CARREIRAS CIENTÍFICAS NO BRASIL (2016-2022)

### 2.1 INTRODUÇÃO

Em 2017, a Elsevier realizou um estudo com o objetivo de analisar a (des)igualdade entre os gêneros no âmbito acadêmico ao longo de quase duas décadas (1996-2015) para doze regiões e encontrou que, entre 2011 e 2015, as mulheres representavam mais de 40% dos pesquisadores em regiões como Estados Unidos, União Europeia, Reino Unido, Canadá, Austrália, França e Dinamarca e 38% no Chile e no México. Portugal e Brasil apresentaram os melhores números, com 49% de participação feminina, enquanto o Japão era o país que apresentava o menor nível, com 33%. Os dados evidenciavam um aumento da proporção das mulheres no meio acadêmico em todas as regiões analisadas em comparação aos números do período anterior, 1996-2000 (ELSEVIER GENDER REPORT, 2017).

Entretanto, conforme Valentova *et al.* (2017), as mulheres cientistas produzem menos, recebem menos financiamentos de pesquisa e são mais prováveis de abandonar a carreira vis-à-vis homens cientistas com grau de qualificação similar. Isso porque, frequentemente, enfrentam barreiras visíveis e invisíveis no mercado de trabalho (VERDUGO-CASTRO *et al.*, 2021). Em relação às barreiras visíveis, as mulheres são muitas vezes obrigadas a postergar seus objetivos profissionais a fim de responder às demandas familiares e domésticas. As barreiras invisíveis, por sua vez, podem remeter aos estereótipos de gênero, frequentemente adquiridos em estágios iniciais da vida das crianças, que agem de forma a criar a imagem de que as mulheres são mais aptas a desenvolver certos tipos de atividades (BUCKLEY; FARRELL; TYNDALL, 2021), e a ciência não é uma delas.

Mesmo aquelas mulheres que superam essas barreiras e conseguem trabalhar no campo das ciências, ainda precisam enfrenta-las para se manter atuantes na área. A partir de dados coletados entre 2006 e 2007 nos Estados Unidos, Schiebinger e Gilmartin (2010) analisaram o desequilíbrio entre a quantidade de trabalho realizado por cientistas mulheres e homens e encontraram que, apesar dos ganhos consideráveis que as mulheres cientistas tiveram nos últimos anos, elas ainda realizam quase o dobro do trabalho doméstico quando comparadas aos seus colegas cientistas do gênero masculino, o que acrescenta mais de dez horas de trabalho não remunerado em suas jornadas semanais.

Diante desse cenário, o presente ensaio realiza uma investigação acerca da situação de indivíduos atuantes em carreiras científicas no Brasil entre 2016 e 2022 a partir de dados do grande grupo de “Profissionais das Ciências e Intelectuais” da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), com foco nas disparidades entre os gêneros. Em outras palavras, em um primeiro momento é investigado se existe um prêmio salarial por desempenhar alguma função desse grupo e, em caso afirmativo, posteriormente é analisado se esse prêmio é diferente entre mulheres e homens.

Em primeiro lugar, é importante ressaltar quem a PNAD Contínua classifica como intelectuais e cientistas. De acordo com a Classificação de Ocupações para Pesquisas Domiciliares, o grande grupo de “Profissionais das Ciências e Intelectuais” é composto por seis subgrupos principais que englobam áreas do conhecimento bastante heterogêneas como as exatas, a saúde, o ensino, a administração, as ciências sociais e a tecnologia da informação e comunicações. Contudo, apesar da heterogeneidade presente nas categorias que formam o grande grupo de intelectuais, a escolha dele para conduzir as análises no presente ensaio foi motivada, principalmente, pela disponibilidade de dados que podem ser utilizados na investigação de indivíduos atuantes em carreiras científicas. Dessa forma, as análises apresentam um aspecto mais amplo, não sendo restritas somente àqueles cientistas que atuam especificamente nas pesquisas acadêmicas.

Devido à natureza não aleatória da amostra e também na tentativa de mitigar o problema do viés de seleção, a análise empírica utilizou o método *Propensity Score Matching* (PSM). Com ele, foi possível realizar a comparação entre indivíduos presentes no grupo de tratamento, ou seja, aqueles que declararam ocupar uma função listada no grande grupo “Profissionais das Ciências e Intelectuais” que possuem características observáveis que mais se assemelham aos indivíduos do grupo de controle, ou seja, todos os outros indivíduos que não ocupam uma função desse grupo.

Dessa forma, o presente estudo busca contribuir para a agenda de pesquisa sobre gênero a partir da produção de resultados que lançam luz sobre a situação atual das mulheres no mercado de trabalho brasileiro, com foco nas carreiras intelectuais e científicas – ainda que heterogêneas –, e tendo como norte importantes questões apontadas pela literatura nacional e internacional, como o problema acerca da segregação ocupacional.

O presente ensaio está organizado em outras quatro seções, além dessa introdução. A segunda seção apresenta a revisão de literatura que norteia as análises. A seção três expõe a metodologia e o tratamento de dados. A seção quatro, por sua vez, contém os resultados, bem

como a análise de robustez. À guisa de conclusão, a seção cinco apresenta as considerações finais do ensaio.

## 2.2 REVISÃO DE LITERATURA

A revisão de literatura a seguir traz uma síntese de alguns dos principais trabalhos que analisam a questão do gênero nas ciências em períodos mais recentes. Os principais pontos de convergência dos estudos dizem respeito aos inúmeros motivos que dificultam a atuação das mulheres nessa esfera, como, por exemplo, os estereótipos de gênero, a escassez de *female role models*, o trabalho doméstico não remunerado, a maternidade e o cuidado de filhos ou parentes e a segregação ocupacional, com as mulheres concentradas em funções que remuneram menos e que possuem menor prestígio social – que será o foco do presente ensaio.

### 2.2.1 *As mulheres nas ciências e a segregação ocupacional*

No âmbito das ciências, Reuben, Sapienza e Zingales (2014) é um exemplo de estudo que analisa os estereótipos de gênero. Conforme os autores, apesar de as mulheres serem maioria nos cursos de graduação, é menos provável que elas se formem ou trabalhem em áreas como matemática e ciências. A fim de entender se a discriminação contribui para essa questão, os autores desenham um experimento no qual candidatos são selecionados para resolver uma prova de aritmética com o intuito de serem contratados para uma vaga de emprego. Os resultados encontrados apontam que, sem o fornecimento de informações sobre os candidatos com exceção das suas aparências, os homens possuem duas vezes mais chances de serem contratados para uma tarefa na área de matemática do que as mulheres. Quando informações adicionais sobre o desempenho dos candidatos em provas anteriores são fornecidas, a discriminação é reduzida, mas não eliminada. Os autores concluem que esse fato contribui para decisões sub ótimas de contratações viesadas em favor de candidatos do gênero masculino<sup>2</sup>.

Por sua vez, Porter e Serra (2020), sugerem que a baixa participação das mulheres em áreas como a Economia e outros campos da ciência conhecidamente dominados pelos homens,

---

<sup>2</sup> Na mesma linha, o estudo de Goldin e Rouse (2000) analisam como as audições às cegas promovem imparcialidade na contratação e aumento na proporção de mulheres em orquestras sinfônicas.

pode ser devida à escassez de *female role models*<sup>3</sup>. Sendo a economia uma dessas áreas de destaque na participação masculina e que, diferentemente das áreas como física e a matemática, tem progredido pouco ao longo dos anos na tentativa de atrair as mulheres, as autoras conduzem um experimento que tem como objetivo elevar o número de mulheres que se formam em economia na *Southern Methodist University* nos Estados Unidos. Na condução das análises, estudantes de aulas introdutórias nos anos de 2015 e 2016 tiveram contato com mulheres bem-sucedidas e carismáticas que haviam se formado na mesma universidade. A partir da metodologia econométrica de diferenças em diferenças, as autoras encontram que a intervenção impactou significativamente a matrícula das estudantes do sexo feminino no curso de economia, aumentando em oito pontos percentuais a probabilidade de concluírem seus estudos na área. Para Buckley, Farrell e Tyndall (2021), a falta de encorajamento para que mais mulheres entrem e permaneçam nessas áreas faz com que muitos países careçam de profissionais qualificadas e fracassem em se tornar competitivos. Na mesma linha, o estudo de Griffith e Main (2021) aponta que a promoção de mais oportunidades de interação entre mulheres – estudantes e professoras – pode ser uma forma de aumentar a representação feminina nessas áreas, uma vez que estas podem ser vistas como *role models* na medida em que refletem sucesso e êxito.

Os inúmeros fatores que dificultam a atuação e a permanência das mulheres em ocupações do campo das ciências contribuem também para a existência da segregação ocupacional e setorial, ou seja, a tendência de concentração específica de algum gênero em determinada ocupação ou setor da economia (BLAU; BRUMMUND; LIU, 2013). Esse problema é entendido, segundo Blau *et al.* (2013), como um dos principais geradores da desigualdade salarial entre os gêneros.

A teoria econômica convencional aponta potenciais fatores explicativos da segregação ocupacional tanto pelo lado da oferta quanto pelo lado da demanda por trabalho. Pelo lado da oferta, a explicação padrão sugere que as mulheres geralmente antecipam carreiras profissionais menos contínuas e são forçadas a escolher empregos compatíveis com suas tarefas familiares e domésticas, devido ao entendimento socialmente discriminatório de que a distribuição dessas responsabilidades deve sempre recair em maior grau para elas. Dessa forma, seria do interesse da parcela feminina ocupar funções que exigem menores níveis de investimento em capital

---

<sup>3</sup> “Usually, an individual rather than a group... the role model demonstrates for the individual how something is done in the technical sense... [The role model] is concerned with the “how” question. The essential quality of the role model is that he possesses skills and display techniques which the actor lacks (or he thinks he lacks) and from whom, by observation and comparison with own performance, the actor can learn” (KEMPER, 1968, p. 33).

humano e que impõem penalidades menos severas em caso de pausas em suas carreiras (DOLADO; FELGUEROSO; JIMENO, 2003).

Pelo lado da demanda, pode-se ressaltar o papel desempenhado pelos estereótipos de gênero, que agem de forma a criar a imagem de que as mulheres são mais aptas a desenvolver certos tipos de atividades percebidas como “femininas” (BUCKLEY; FARRELL; TYNDALL, 2021), e as questões referentes à maternidade, na medida em que essa incute menores níveis de produtividade às mulheres mães ou em idade reprodutiva. Ademais, as mulheres podem se defrontar com situações de discriminação de gênero, que se dá a partir da percepção por parte dos empregadores de que as mulheres são em média menos qualificadas do que os homens (DOLADO; FELGUEROSO; JIMENO, 2003; FERNANDEZ-MATEO; FERNANDEZ, 2016).

Esses e outros fatores contribuem para que, não obstante o aumento da participação feminina no mercado de trabalho e a redução do diferencial salarial e educacional entre mulheres e homens serem observados nas últimas décadas, as mulheres ainda sejam maioria entre os trabalhadores de empregos que oferecem piores remunerações, condições trabalhistas e benefícios (BORROWMAN; KLASSEN, 2019), geralmente de setores nos quais a alta competitividade e a baixa flexibilidade limita o seu poder de barganha (ARORA; BRAUNSTEIN; SEGUINO, 2023).

Uma das possíveis consequências dessa realidade são os salários mais baixos das mulheres, mesmo que em muitas vezes elas possam se equiparar aos seus pares masculinos em termos de qualificações (SOARES, 2000; SANTOS, 2008). Uma vez que, frequentemente, a disparidade salarial se mantém em um cenário no qual as diferenças educacionais entre mulheres e homens se tornam cada vez menores, fica evidente que o nível educacional entre os gêneros pode não ser mais uma variável com alto poder de explicação acerca da disparidade salarial (BÜTIKOFER; JENSEN; SALVANES, 2018).

Nesse sentido, existem estudos que sugerem que a segregação ocupacional e a segregação educacional andam de mãos dadas (SAHOO; KLASSEN, 2018; FRANCESCONI; PAREY, 2018; SCHNEEWEIS; ZWEIMÜLLER, 2012; SOOKRAM; STROBL, 2009; AGRAWAL, 2021). Isso porque, para os indivíduos que possuem ensino superior, a escolha acerca de qual área cursar direciona, em grande medida, o seu emprego no mercado de trabalho futuramente.

Barone e Assirelli (2018), por exemplo, apontam que a segregação educacional é reconhecida como um fator chave para explicar a persistência da desigualdade de gênero no mercado de trabalho, apesar de mulheres e homens serem cada vez mais similares em termos

de escolaridade. Os autores sugerem também que as mulheres normalmente são maioria em áreas como ciências sociais, humanidades e áreas relacionadas aos cuidados e ao ensino, como professores. Essas áreas oferecem, relativamente, perspectivas menos favoráveis ao mercado de trabalho em termos de rendimentos, taxas de desemprego, e aos riscos do *overeducation* (quando o grau de educação de um indivíduo vai além do nível requerido pelo seu trabalho), e do *mismatch* educacional horizontal (quando a área de formação é diferente da área de trabalho), quando comparadas às áreas de exatas, como ciências, tecnologia, engenharia e matemática (STEM – *Science, Technology, Engineering, Math*) que são relativamente mais predominadas por homens.

Já o trabalho de Borrowman e Klasen (2019) analisa a segregação setorial e ocupacional a partir de duas medidas agregadas de segregação aplicadas a dados de sessenta e nove países entre 1980 e 2011 e constata aumento dos dois tipos de segregação. Além disso, os autores apontam, entre outras coisas, que o aumento da participação feminina no mercado de trabalho está associado à queda da segregação setorial e ao aumento da segregação ocupacional e que o aumento do nível educacional tende a elevar o nível de segregação no geral.

Francesconi e Parey (2018) também analisam a influência das escolhas acadêmicas no *gap* salarial entre os gêneros no mercado de trabalho. Os autores apontam que, na Alemanha, mulheres e homens entram no ensino superior em número similar, mas mais mulheres do que homens concluem os cursos. Ademais, as mulheres entram na universidade com notas ligeiramente mais altas do ensino médio, mas deixam o curso com notas ligeiramente mais baixas do que os homens. Imediatamente após a conclusão da universidade, a carga horária semanal de trabalho entre os gêneros é bastante semelhante, mas os homens ganham mais do que as mulheres em toda a distribuição salarial. A variável que mais explica essa diferença nos salários é a área de estudo que o indivíduo cursou na universidade, apontando para a importância das escolhas ocupacionais feitas antes de entrar no mercado de trabalho. De acordo com os autores, tais escolhas podem ser, em parte, motivadas por diferenças nas preferências entre os gêneros, como, por exemplo, aversão ao risco, autoconfiança, competitividade, expectativas de ganhos e a valorização de benefícios não monetários.

Por sua vez, Sahoo e Klasen (2018), investigam a segregação de gênero em diferentes campos de estudo e como isso afeta o mercado de trabalho na Índia. Os autores encontram que as mulheres são menos prováveis de estudar em áreas das ciências, tecnologia, engenharia e matemática (STEM) e comércio vis-à-vis humanidades e artes. Essa disparidade se mantém significativa mesmo após controlar notas em testes passados dos indivíduos, indicando que ela não se dá por motivos de diferença em suas habilidades matemáticas. Por fim, os autores

concluem que as escolhas educacionais afetam as disparidades entre os gêneros no mercado de trabalho, tanto em termos de participação quanto na natureza e na remuneração dos empregos.

Em relação ao Brasil, Vaz e Hoffman (2011) analisam a segregação ocupacional no setor público brasileiro a partir de dados da RAIS e cálculos de medidas sumárias de segregação. Os autores sugerem que as mulheres estão mais propensas a trabalhar nas esferas estadual e municipal na medida em que ofertam proporcionalmente um maior número de postos de trabalhos relacionados às atividades socialmente entendidas como “femininas”.

Por sua vez, Botassio e Vaz (2020), analisam a segregação ocupacional no mercado de trabalho brasileiro entre 2004 e 2015. A partir do cálculo de indicadores como o índice de dissimilaridade, o índice de Gini e a classe de medidas aditivamente decomponíveis proposta por Hutchens (2004), os autores encontram que a segregação aumentou entre 2004 e 2015 e que esse crescimento foi impulsionado pela segregação entre os grandes grupos ocupacionais que contêm ocupações mais heterogêneas entre si.

O estudo de Silva e Vaz (2022) analisa por que as ocupações femininas pagam salários menores a partir da avaliação do impacto da transição entre ocupações distintas quanto à composição por gênero no rendimento dos trabalhadores. Utilizando dados da PNAD Contínua entre 2012 e 2019 aplicados a um painel de dados com efeitos fixos e a tipologia de integração ocupacional proposta por Oliveira (2001) para a classificação das ocupações em predominantemente femininas, predominantemente masculinas ou integradas, as autoras encontram evidências que suportam a hipótese da desvalorização do trabalho das mulheres, na medida em que o trabalhador experimenta uma queda no rendimento quando transita para uma ocupação classificada como feminina. Além disso, as autoras observam que a relação entre a composição ocupacional por gênero e os rendimentos não é linear uma vez que os maiores rendimentos são recebidos nas ocupações integradas.

A questão acerca da concentração de mulheres em ocupações que pagam menores salários é analisada também por Vaz e Botassio (2022) a partir de modelos dinâmicos de efeitos fixos com dados do Censo de 1980, 1991 e 2000 para o Brasil. Os autores não encontram evidências de que a concentração de mulheres em ocupações específicas afeta negativamente os salários dessas ocupações e sugerem que para o período analisado fatores não monetários, como a flexibilização das horas de trabalho, que são avaliados diferentemente entre mulheres e homens, são mais determinantes para explicar os diferenciais salariais entre os gêneros.

Em suma, os trabalhos indicam que fatores como os estereótipos de gênero, que incutem às mulheres tarefas socialmente entendidas como femininas, contribuem para que elas se concentrem em áreas que possuem menor prestígio social, tanto na escolha do curso de

formação superior quanto, posteriormente, na área de atuação no mercado de trabalho. Contudo, é importante ressaltar, como aponta Levanon, England e Allison (2009), que não são as mulheres que procuram piores empregos, e sim que as áreas perdem valor quando concentram um grande número da parcela feminina, indicando que é o trabalho mal pago que procura as mulheres e não o contrário. A relação entre esses fatores, portanto, será o foco de análise do presente ensaio.

### 2.3 METODOLOGIA E TRATAMENTO DE DADOS

Conforme a Classificação de Ocupações para Pesquisas Domiciliares da PNAD Contínua, a categoria de “Profissionais das Ciências e Intelectuais” contém indivíduos atuantes em seis subgrupos principais: “Profissionais das Ciências e Engenharia”; “Profissionais da Saúde”; “Profissionais do Ensino”; “Especialistas em Organização da Administração Pública e de Empresas”; “Profissionais de Tecnologia da Informação e Comunicações” e; “Profissionais em Direito, em Ciências Sociais e Culturais”<sup>4</sup>.

Devido à natureza não aleatória da amostra e também na tentativa de mitigar o problema do viés de seleção<sup>5</sup>, a análise empírica utilizou o método *Propensity Score Matching* (PSM). Com ele, foi possível realizar a comparação entre indivíduos presentes no grupo de tratamento, ou seja, aqueles que declararam ocupar um cargo listado no grande grupo “Profissionais das Ciências e Intelectuais” que possuem características observáveis que mais se assemelham aos indivíduos do grupo de controle, ou seja, todos os outros indivíduos que não ocupam uma função desse grupo.

Para tal, o primeiro passo foi estimar a probabilidade de ocupar uma função do grande grupo “Profissionais das Ciências e Intelectuais”, dado um vetor de covariáveis observáveis. Posteriormente, o pareamento entre os grupos foi realizado a partir do método do PSM. Por fim, após a avaliação e validação da qualidade do pareamento, o próximo e último passo consistiu em estimar o efeito do tratamento e seus erros-padrão. Para isso, foi ajustado um modelo de regressão linear com o salário/hora sendo a variável dependente e as covariáveis e

---

<sup>4</sup> Para mais detalhes sobre subgrupos e grupos de base, ver Quadro A1 no Apêndice.

<sup>5</sup> Avaliações microeconômicas normalmente precisam levar em consideração questões relacionadas ao viés de seleção. Esse problema acontece porque não é possível observar dois resultados para o mesmo indivíduo ao mesmo tempo. Além disso, as observações que recebem o tratamento podem ter diferentes características daquelas que não recebem. Portanto, não é aconselhável considerar o valor médio do grupo de controle como uma aproximação (CALIENDO; KOPEINIG, 2008).

os valores do *propensity score* estimados como as variáveis independentes. As covariáveis foram escolhidas conforme o modelo minceriano<sup>6</sup> de determinação salarial e outras variáveis de controle como número de habitantes do domicílio, número de filhos(as), ser chefe de família, ter cônjuge ou companheiro(a) habitando o mesmo domicílio, trabalhar no setor público ou privado, e trabalhar período integral ou meio-período também foram adicionadas ao modelo.

### 2.3.1 Base de dados e recorte amostral

O presente trabalho utilizou dados longitudinais da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) da quinta visita entre os anos de 2016 e 2022 por serem a sequência de anos mais recentes disponível, que inclui informações completas sobre os rendimentos e as horas trabalhadas. Implantada efetivamente em 2012, a PNAD Contínua tem como objetivo realizar o acompanhamento de flutuações do mercado de trabalho e de outros fatores que influenciam o desenvolvimento socioeconômico do país a partir da produção de indicadores conjunturais, estruturais e especiais que podem ter periodicidades mensais, trimestrais e anuais. A pesquisa é realizada por uma amostra probabilística de domicílios e a população alvo são residentes de domicílios particulares (permanentes ou improvisados) e em unidades de habitação em domicílios coletivos nas áreas urbana e rural que respondem às entrevistas respeitando um esquema de rotação<sup>7</sup> (IBGE, 2023).

O Quadro 1 a seguir descreve as variáveis utilizadas.

Quadro 1 – Descrição das variáveis

Variável	Descrição <sup>8</sup>
Intelectual	<i>Dummy</i> igual a um se o indivíduo ocupa um trabalho da categoria "Profissionais das ciências e intelectuais" e igual a zero caso contrário.

(continua)

<sup>6</sup> A equação de determinação salarial de Mincer (1974) é um método frequentemente utilizado para estimar diferenciais de salários entre grupos. De acordo com o modelo, o salário de um indivíduo é explicado por fatores como escolaridade (anos de estudo), experiência (idade), experiência ao quadrado (idade ao quadrado), além de um vetor de características observáveis como raça, gênero e região de domicílio. Para mais informações sobre a equação minceriana ver Heckman, Lochner e Todd (2006).

<sup>7</sup> A rotação ocorre em meses, no esquema 1-2(5), ou seja, um domicílio que responde a entrevista uma vez no mês, fica fora dois meses, volta a ser entrevistado no 4º mês, fica fora outros 2 meses, repetindo esse esquema 5 vezes. Já em trimestres, a notação desse esquema de rotação é 1 por 5: o domicílio é entrevistado uma vez no trimestre e permanece na pesquisa por 5 trimestres consecutivos, com 2 meses de intervalo entre as entrevistas. Para mais informações: ([https://www.ibge.gov.br/arquivo/projetos/sipd/terceiro\\_forum/terceiro\\_amostra.pdf](https://www.ibge.gov.br/arquivo/projetos/sipd/terceiro_forum/terceiro_amostra.pdf)).

<sup>8</sup> Todas as informações são autodeclaradas pelos indivíduos entrevistados.

(continuação)

Variável	Descrição <sup>9</sup>
Salário/hora	Remuneração em reais do trabalho principal recebido por hora aos preços de 2022, ajustados de acordo com o deflator da PNAD Contínua.
Idade	Idade do indivíduo.
Idade ao quadrado	Idade ao quadrado do indivíduo.
Número de pessoas no domicílio	Número de pessoas que residem no mesmo domicílio.
Filhos	<i>Dummy</i> igual a um se o indivíduo tem filho(a) ou enteado(a) no domicílio e igual a zero caso contrário.
Anos de estudo	Número de anos de estudo
Gênero	<i>Dummy</i> igual a um para mulheres e igual a zero para homens.
Raça Branca/Amarela	<i>Dummy</i> igual a um para brancos(as) ou amarelos(as) e igual a zero para pretos(as), pardos(as) ou indígenas.
Chefe de família	<i>Dummy</i> igual a um para chefe da família e igual a zero caso contrário.
Região metropolitana	<i>Dummy</i> igual a um se o indivíduo mora na capital ou na região metropolitana e igual a zero caso contrário.
Setor privado	<i>Dummy</i> igual a um se o indivíduo trabalha no setor privado e igual a zero caso contrário.
Setor público	<i>Dummy</i> igual a um se o indivíduo trabalha no setor público e igual a zero caso contrário.
Cônjuge	<i>Dummy</i> igual a um se o indivíduo tem cônjuge ou companheiro(a) morando no mesmo domicílio e igual a zero caso contrário.
Período integral	<i>Dummy</i> igual a um se o indivíduo trabalha 40 horas ou mais por semana e igual a zero caso contrário.
Regiões Sul, Sudeste, Nordeste, Norte e Centro-Oeste*	<i>Dummies</i> de regiões.
Anos 2016, 2017, 2018, 2019, 2020, 2021 e 2022*	<i>Dummies</i> de anos.

Fonte: Elaboração própria.

\* Categorias de base.

<sup>9</sup> Todas as informações são autodeclaradas pelos indivíduos entrevistados.

### 2.3.2 Propensity score matching

O método do *propensity score* é uma técnica estatística útil para estimar os efeitos de um tratamento quando não há aleatoriedade na amostra (GUO; FRASER, 2015). Em amostras aleatórias, o grupo de tratamento é o nome dado ao grupo que recebe o tratamento em um estudo. O tratamento, por sua vez, é a variável de interesse e pode ter diversas naturezas a depender do campo da pesquisa em questão. Já o grupo de controle é aquele que não recebe o tratamento ou que recebe um tratamento placebo. Quando a amostra não foi desenhada com o propósito de analisar os efeitos causais e não há um grupo de controle, o método do *propensity score matching* permite criar, de forma sintética, um grupo de controle, que possua características semelhantes ao grupo de tratamento, dadas as características observadas.

Formalmente, Rosenbaum e Rubin (1983) definem o método do *propensity score matching* como a probabilidade condicional de ser atribuído a um tratamento, dado um vetor de covariáveis observáveis. Guo e Fraser (2015) destacam que a modelagem do *propensity score matching* deve seguir três passos.

O primeiro passo consiste na escolha das covariáveis que podem causar um desbalanceamento entre os grupos de controle e de tratamento. A modelagem inicia sempre com a estimação da probabilidade de receber o tratamento, ou seja, com a estimação do *propensity score*<sup>10</sup> a partir de uma regressão binária logística conforme a equação (1):

$$P(W | X_i = x_i) = E(W_i) = \frac{e^{x\beta_i}}{1 + e^{x\beta_i}} = \frac{1}{e^{-x\beta_i} + 1} \quad (1)$$

Em que  $W_i$  ( $W_i = 1$ , quando a observação está no grupo de tratamento e  $W_i = 0$  quando está no grupo de controle) para o  $i$ -ésimo caso ( $i = 1, \dots, N$ ) é a condição binária de tratamento,  $X_i$  é o vetor de variáveis explicativas e  $\beta_i$  é o vetor dos parâmetros da regressão (GUO; FRASER, 2015).

Posteriormente, após a estimação dos *propensity scores*, o segundo passo consiste em parear as observações do grupo de controle com as observações do grupo de tratamento, com o objetivo de criar uma nova amostra na qual as observações possuem probabilidades similares

---

<sup>10</sup> Tanto as observações do grupo de controle quanto as observações que fazem parte do grupo de tratamento podem apresentar um valor de *propensity score* diferente de zero. Um par de participantes, sendo um do grupo de controle e o outro do grupo de tratamento que possuem um *propensity score* semelhante, são considerados como comparáveis, apesar de poderem apresentar valores diferentes para determinadas covariáveis (GUO; FRASER, 2015).

de serem atribuídas ao tratamento. Nessa etapa, existem alguns tipos de pareamento que podem ser escolhidos, como o método do vizinho mais próximo (*nearest neighbor* – NN) com ou sem reposição, a distância de Mahalanobis com ou sem *propensity scores* e o *optimal matching*<sup>11</sup>.

O presente trabalho utilizou o pareamento do *nearest neighbor* (NN) com reposição, que pode ser descrito conforme a equação (2). A escolha do *nearest neighbor* se deu por ser o método mais usual utilizado na literatura e a escolha da reposição foi devido ao fato de a base de dados apresentar valores de *propensity scores* bastante distintos entre os grupos de controle e tratamento<sup>12</sup> (CALIENDO; KOPEINIG, 2008). Sendo assim, o pareamento entre as observações ocorre da seguinte forma:

$$C(P_i) = \min_j ||P_i - P_j||, j \in I_0 \quad (2)$$

Em que  $P_i$  e  $P_j$  são os *propensity scores* para o grupo de tratamento e de controle, respectivamente,  $I_1$  é o conjunto de observações tratadas e  $I_0$  o conjunto de observações de controle. A vizinhança  $C(P_i)$  contém a observação de controle  $j$  (i.e.,  $j \in I_0$ ) como um pareamento para a observação tratada  $i$  (i.e.,  $i \in I_1$ ) se a diferença absoluta entre os *propensity scores* é a menor entre todos os possíveis pares de *propensity scores* entre  $i$  e  $j$ .

Se para cada  $i$  existe apenas um  $j$  na vizinhança, então tem-se um pareamento 1:1. Se para cada  $i$  é definido um número  $n$  de observações que podem ser pareadas na vizinhança, então o pareamento é 1:n (GUO; FRASER, 2015). Nessa etapa, a qualidade do pareamento precisa ser investigada e os efeitos do tratamento, bem como seus erros-padrão, precisam ser estimados (CALIENDO; KOPEINIG, 2008).

O terceiro passo, por sua vez, consiste na análise pós-pareamento. A nova amostra pareada, já com a correção do viés de seleção, pode ser utilizada para performar análises multivariadas como é normalmente feito com amostras aleatórias (GUO; FRASER, 2015). O método de *propensity score matching* possibilita a estimação de diversos parâmetros, sendo o efeito médio do tratamento sobre a população (*population average treatment effect* – ATE) e o

<sup>11</sup> Para mais informações sobre os métodos de pareamento ver Caliendo e Kopeinig (2008).

<sup>12</sup> Existe um *trade-off* entre viés e variância quando se utiliza o pareamento com reposição. Quando essa opção é utilizada, a qualidade média do pareamento irá melhorar e o viés irá diminuir. Isso é particularmente útil quando se tem uma base de dados na qual os valores de *propensity scores* entre os grupos de controle e tratamento bastante distintos. Por exemplo, se existem muitas observações no grupo de tratamento com valores altos de *propensity scores*, mas poucas observações com *propensity score* altos no grupo de controle, algumas observações com alto *propensity score* serão pareadas com observações de baixo *propensity score*. Dessa forma, a reposição pode ser útil na medida em que irá reduzir o número de observações com *propensity scores* muito diferentes na construção do contrafactual. A contrapartida será o aumento da variância das estimações (CALIENDO; KOPEINIG, 2008, p. 42).

efeito médio do tratamento sobre os tratados (*average treatment effect on the treated – ATT*) os mais frequentemente analisados na literatura. O ATE é a diferença esperada nos resultados entre aqueles que receberam o tratamento e aqueles que não receberam, conforme a equação (3):

$$\tau_{ATE} = E(\tau) = E[Y(1) - Y(0)] \quad (3)$$

Ou seja, o parâmetro busca entender qual seria o efeito sobre o resultado se as observações da população fossem designadas aleatoriamente para o tratamento<sup>13</sup>.

O ATT, por sua vez, foca explicitamente no efeito do tratamento para os tratados, conforme a equação (4):

$$\tau_{ATT} = E(\tau|W = 1) = E[Y(1)|W = 1] - E[Y(0)|W = 1] \quad (4)$$

Isto é, o valor esperado do ATT é definido como a diferença entre o resultado esperado dos valores com e sem o tratamento para aquelas observações que de fato receberam o tratamento (CALIENDO; KOPEINIG, 2008).

Ademais, Greifer (2022) adiciona uma quarta etapa às três anteriormente citadas na condução das análises do *propensity score matching* que diz respeito ao planejamento pré-pareamento<sup>14</sup>. Na fase do planejamento, três fatores precisam ser levados em consideração. O primeiro diz respeito à seleção do efeito a ser estimado, o segundo refere-se à população na qual o efeito do tratamento será generalizado e o terceiro corresponde à escolha das covariáveis.

Inicialmente, nas análises que antecedem o pareamento, Greifer (2022) aponta também para a importância da investigação do desbalanceamento das covariáveis antes do *matching*, para elucidar o que o método buscará eliminar, e depois do *matching*, com o objetivo de constatar a adequação do modelo e a validade das inferências causais, conforme as Tabelas 4, 5 e 6 apresentadas na seção 4.2.

O presente ensaio focou no efeito médio do tratamento sobre os tratados (ATT), ou seja, no efeito médio que ocupar uma função da categoria de “Profissionais das Ciências e Intelectuais” tem sobre mulheres e homens inclusos na base de dados da PNADC entre 2016 e 2022 com o objetivo de investigar se existe um prêmio salarial por ocupar uma função dessa

<sup>13</sup> Heckman (1997) enfatiza que a informação contida na estimação do ATE pode ser limitada, principalmente para formuladores de políticas públicas, uma vez que o parâmetro inclui o efeito nas observações para as quais o tratamento nunca foi intencionado.

<sup>14</sup> O *matching* foi estimado utilizando o pacote Matchit (Ho *et al.*, 2011) e as análises de balanceamento foram realizadas com o pacote cobalt (Greifer, 2022), ambos no software R.

categoria e, em caso afirmativo, se esse valor é diferente entre os gêneros. Vale ressaltar que o pareamento foi realizado com a amostra apenas de homens, apenas de mulheres, com a amostra completa, e por fim, separadamente, para cada uma das amostras na análise de robustez.

## 2.4 RESULTADOS E DISCUSSÕES

A seção de resultados e discussões a seguir traz as estatísticas descritivas da base de dados, os resultados do PSM e a análise de robustez, na qual são investigados sete cenários.

### 2.4.1 Estatísticas descritivas

As Tabelas 1, 2 e 3 a seguir apresentam as estatísticas descritivas para o grupo de tratamento, ou seja, aqueles indivíduos que declararam estar empregados em alguma função da categoria “Profissionais das Ciências e Intelectuais” e para o grupo de controle, composto por aqueles que não declararam ocupar uma função dessa categoria. A Tabela 1 contém os dados da amostra apenas para aqueles indivíduos que se autodeclararam mulheres, a Tabela 2 contém para aqueles que se autodeclararam homens e a Tabela 3 para toda a amostra.

É possível observar que, quando é considerada a amostra contendo apenas mulheres, antes do pareamento, o grupo de tratamento é composto por mais mulheres que declararam ser brancas ou amarelas em relação ao grupo de controle. Este, por sua vez, é composto relativamente por mais mulheres que declararam ser chefes de família e que trabalham 40 horas ou mais no setor privado. A média de anos de estudo do grupo de controle é de 10,58 anos e para o grupo de tratamento é de 15,64 anos.

Tabela 1 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Mulheres

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
<b>Características do indivíduo</b>					
Idade	40,7603	12,2830	41,5235	10,9550	-15,253***
Idade ao quadrado	1812,2790	1053,2579	1844,2140	9624,1810	-7,296***
Raça Branca/Amarela	0,4244	0,4942	0,5817	0,4933	-71,137***
Cônjuge	0,5634	0,4960	0,5801	0,4935	-7,5547***
Chefe de família	0,4150	0,4927	0,3844	0,4864	14,014***
Setor privado	0,8752	0,3305	0,4661	0,4989	191,13***

(continua)

(continuação)

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
Anos de estudo	10,5881	3,7051	15,6471	1,1466	-634,310***
Período integral	0,6266	0,4837	0,5508	0,4974	34,114***
Salário/hora	13,4372	32,8378	37,0744	68,9028	-81,367***
<b>Características do domicílio</b>					
Habitantes no domicílio	3,3841	1,4827	3,1799	1,2787	34,745***
Cônjuge	0,3800	0,4854	0,3928	0,4884	-5,8048***
Filhos(as)	1,2694	1,0935	1,1332	0,9438	31,4***
Metropolitana	0,5782	0,4938	0,4982	0,5000	35,766***
<b>Regiões</b>					
Norte	0,1154	0,3195	0,1185	0,3231	-2,109**
Nordeste	0,2691	0,4435	0,2448	0,4300	12,536***
Sudeste	0,3030	0,4596	0,3266	0,4690	-11,254***
Sul	0,1982	0,3986	0,2007	0,4005	-1,3779
Centro-Oeste	0,1143	0,3181	0,1094	0,3122	3,4436***
<b>Anos</b>					
2016	0,1600	0,3666	0,1487	0,3558	7,0356***
2017	0,1604	0,3670	0,1456	0,3527	9,3103***
2018	0,1584	0,3652	0,1499	0,3570	5,298***
2019	0,1582	0,3650	0,1565	0,3633	1,083
2020	0,1114	0,3146	0,1245	0,3301	-8,9118***
2021	0,1128	0,3163	0,1289	0,3351	-10,838***
2022	0,1388	0,3457	0,1459	0,3530	-4,5078***
<b>Observações</b>	<b>333.715</b>		<b>58.497</b>		

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Tabela 2 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Homens

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
<b>Características do indivíduo</b>					
Idade	41,1196	13,2751	42,9710	12,8203	-26,915***
Idade ao quadrado	1867,0460	1161,9240	2010,8590	1196,6820	-22,492***
Raça Branca/Amarela	0,3999	0,4899	0,6261	0,4838	-87,273***
Cônjuge	0,6616	0,4732	0,6417	0,4795	7,7818***
Chefe de família	0,5701	0,4951	0,6017	0,4896	-12,043***
Setor privado	0,9194	0,2723	0,6913	0,4620	94,23***
Período integral	0,7713	0,4200	0,6736	0,4689	39,194***
Anos de estudo	9,2893	3,9802	15,3245	1,8412	-555,11***
Salário/hora	15,6550	36,4556	54,2637	107,7683	
<b>Características do domicílio</b>					
Habitantes no domicílio	3,4522	1,5262	3,1066	1,2872	-68,904***
Cônjuge	0,1975	0,3981	0,1741	0,3792	11,507***

(continua)

(continuação)

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
Filhos(as)	1,2888	1,1652	1,0632	0,9800	49,518***
Metropolitana	0,6501	0,4769	0,4236	0,4941	42,448***
<b>Regiões</b>					
Norte	0,1345	0,3412	0,1179	0,3225	85,844***
Nordeste	0,2813	0,4497	0,1975	0,3981	9,5796***
Sudeste	0,2821	0,4500	0,3609	0,4803	39,025***
Sul	0,1905	0,3927	0,2093	0,4068	-30,764***
Centro-Oeste	0,1116	0,3149	0,1145	0,3184	-8,707***
<b>Anos</b>					
2016	0,1626	0,3690	0,1490	0,3561	-1,6859*
2017	0,1592	0,3659	0,1458	0,3530	7,0593***
2018	0,1564	0,3632	0,1488	0,3559	3,9429***
2019	0,1550	0,3619	0,1582	0,3649	-1,6554*
2020	0,1157	0,3198	0,1274	0,3335	-6,6097***
2021	0,1155	0,3196	0,1242	0,3299	-4,9798***
2022	0,1357	0,3425	0,1464	0,3535	-5,6434***
<b>Observações</b>	<b>546.663</b>		<b>37.280</b>		

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Tabela 3 – Estatísticas descritivas e Teste t para os grupos de tratamento e de controle antes do pareamento – Todos

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
<b>Características do indivíduo</b>					
Idade	40,9833	12,9091	42,0869	11,7377	-2,7352***
Idade ao quadrado	1846,2780	1122,2710	1909,0790	1062,8800	-17,268***
Gênero	0,3792	0,4852	0,6108	0,4876	-139,65***
Raça Branca/Amarela	0,4092	0,4917	0,5990	0,4901	-113,78**
Cônjuge	0,6244	0,4843	0,6041	0,4891	12,216***
Chefe de família	0,5113	0,4999	0,4690	0,4990	24,924***
Setor privado	0,9026	0,2965	0,5537	0,4971	213,12***
Período integral	0,7165	0,4507	0,5986	0,4902	71,191***
Anos de estudo	9,7818	3,9291	15,5215	1,4653	-908***
Salário/hora	14,8140	3,5144	43,7651	8,6547	-102,61***
<b>Características do domicílio</b>					
Habitantes no domicílio	3,4264	1,5102	3,1514	1,2825	61,858***
Cônjuge	0,2667	0,4225	0,3077	0,4615	-26,162***
Filhos(as)	1,2814	1,1386	1,1060	0,9586	52,742***
Metropolitana	0,6229	0,4847	0,4692	0,4991	90,774***

(continua)

(continuação)

Variável	Média	Desvio-padrão	Média	Desvio-padrão	Teste t
	Grupo de controle		Grupo de tratamento		
<b>Regiões</b>					
Norte	0,1273	0,3333	0,1182	0,3229	0,1896***
Nordeste	0,2767	0,4474	0,2264	0,4185	35,091***
Sudeste	0,2900	0,4538	0,3399	0,4737	-31,098***
Sul	0,1934	0,3950	0,2040	0,4030	-7,7711***
Centro-Oeste	0,1126	0,3161	0,1114	0,3146	1,1315***
<b>Anos</b>					
2016	0,1616	0,3681	0,1488	0,3559	10,511***
2017	0,1597	0,3663	0,1457	0,3528	11,588***
2018	0,1572	0,3639	0,1495	0,3566	6,2862***
2019	0,1562	0,3631	0,1571	0,3639	-0,7551***
2020	0,1140	0,3179	0,1256	0,3314	-10,309***
2021	0,1145	0,3184	0,1271	0,3331	-11,214***
2022	0,1369	0,3437	0,1461	0,3532	-7,6845***
<b>Observações</b>	<b>984.017</b>		<b>110.455</b>		

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Quando somente a parcela da amostra contendo indivíduos que declararam ser homens é considerada, antes do pareamento, o grupo de tratamento é composto por mais homens que declararam ser brancos ou amarelos e chefes de família relativamente ao grupo de controle. Este, por sua vez, é composto por mais homens que trabalham 40 horas ou mais por semana no setor privado. A média de anos de estudo do grupo de controle é de 9,28 anos enquanto o grupo de tratamento possui em média 15,32 anos de estudo.

Os valores para toda a amostra seguem a tendência dos resultados das amostras separadas. A variável que mais destoa entre os dois grupos é a referente aos anos de estudo formal: enquanto a média do grupo de controle é de 9,78 anos, a do grupo de tratamento é de 15,52 anos<sup>15</sup>.

Em relação ao salário recebido por hora, pode-se observar que os indivíduos que não declararam ocupar uma função na categoria “Profissionais das Ciências e Intelectuais” possuem a média de salário/hora no valor de R\$ 14,81 enquanto os indivíduos que declararam ocupar

<sup>15</sup> No Brasil, os níveis de ensino são categorizados da seguinte forma: a Educação Básica compreende a Educação Infantil (de 0 a 5 anos) o Ensino Fundamental (de 6 a 14 anos) e o Ensino Médio (15 a 17 anos). A Educação Superior pode ser dividida em: Graduação (tecnologia, licenciatura ou bacharelado); Pós-Graduação (*lato sensu*: especializações; *stricto sensu*: mestrado e doutorado) e pós-doutorado. Podemos considerar que os indivíduos do grupo de controle, que possuem uma média de 9,72 anos de estudo, são aqueles que possuem o ensino médio incompleto e os indivíduos do grupo de tratamento, que possui uma média de 15,51 anos de estudo, são aqueles que possuem o ensino superior completo. Para mais informações sobre grau de instrução: <<https://www.ipea.gov.br/atlasestado/arquivos/rmd/4874-conjunto4v10.html>>.

uma função dessa categoria recebem em média R\$ 43,76 reais/hora, ou seja, aproximadamente três vezes mais do que os indivíduos não intelectuais.

Quando analisamos os gêneros separadamente, observamos que os homens não intelectuais recebem em média R\$ 15,65 reais/hora enquanto as mulheres não intelectuais recebem em média R\$ 13,43 reais/hora, cerca de 86% do salário dos homens. Em relação aos intelectuais, os resultados apontam para uma diferença ainda mais acentuada: enquanto os homens intelectuais recebem em média R\$ 54,26 reais/hora, as mulheres intelectuais recebem em média R\$ 37,07 reais/hora, aproximadamente 68% do salário dos homens.

Os resultados obtidos com o Teste t indicam significância para a maior parte das variáveis analisadas, com exceção das variáveis do ano de 2019 e Região Sul que não são estatisticamente significativas na base de dados que considera somente as mulheres. Ou seja, na maior parte dos casos, existem evidências estatísticas de que se pode rejeitar a hipótese nula de igualdade entre as médias do grupo de tratamento e do grupo de controle.

Ainda que as diferenças sejam estatisticamente significativas, parte delas pode ser explicada por que as mulheres e homens possuem características produtivas distintas que influenciam os salários. Nesse caso, apenas comparar as médias entre os grupos não é o ideal, uma vez que a comparação estaria sendo feita para indivíduos que podem possuir características muito diferentes. Portanto, para contornar essa questão, o presente ensaio utiliza o método do *propensity score matching* – PSM no exercício empírico.

#### 2.4.2 *Propensity Score Matching*

A Tabela 4 contém as informações das covariáveis antes do pareamento considerando a base de dados somente com indivíduos que declararam ser mulheres, a Tabela 5 considera apenas indivíduos que declararam ser homens e a Tabela 6 considera todos os indivíduos. Conforme os resultados reportados, pode-se verificar o desbalanceamento da base a partir dos valores das diferenças das médias padronizadas e das densidades acumuladas médias e máximas que quanto mais distantes de zero mais apontam para a existência do desbalanceamento. Ou seja, indicam que a base de controle não pode ser diretamente comparada ao tratamento. Considerando toda a amostra, a variável mais desbalanceada na base é a Anos de Estudo, uma vez que o grupo de controle tem em média 9,78 anos de estudo enquanto o grupo de tratamento estudou em média 15,52 anos. Esse padrão pode ser observado também quando são analisadas as amostras separadas por gênero. O valor médio da variável para as mulheres do grupo de tratamento é de 15,64 anos de estudo enquanto para as mulheres que estão no grupo de controle

é de 10,58 anos. Para os homens, aqueles que estão no grupo de tratamento possuem em média 15,32 anos de estudo, e aqueles que estão no grupo de controle 9,29 anos de estudo.

Após a análise dos dados antes do pareamento, a próxima etapa diz respeito à estimação dos *propensity scores*, ou seja, das probabilidades de um indivíduo ser intelectual (tratado) considerando todas as covariáveis incluídas no modelo.

Tabela 4 – Resultados antes do pareamento – Mulheres

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima
Distância	0,5274	0,0828	2,2092	1363,4	0,4191	0,7758
Idade	41,5235	40,7603	0,0697	0,7954	0,0142	0,0656
Idade ao quadrado	1844,2141	1812,2794	0,0332	0,8349	0,0142	0,0656
Habitantes no domicílio	3,1799	3,3841	-0,1597	0,7438	0,0093	0,0587
Filhos(as)	1,1332	1,2694	-0,1443	0,7449	0,0113	0,0475
Anos de estudo	15,6471	10,5881	4,4122	0,0958	0,2976	0,7710
Raça Branca/Amarela	0,5817	0,4244	0,3190	.	0,1573	0,1573
Chefe de família	0,3844	0,4150	-0,0629	.	0,0306	0,0306
Região metropolitana	0,4982	0,5782	-0,1600	.	0,0800	0,0800
Setor privado	0,4661	0,8752	-0,8201	.	0,4091	0,4091
Cônjuge	0,5801	0,5634	0,0339	.	0,0167	0,0167
Período integral	0,5508	0,6266	-0,1523	.	0,0758	0,0758
Norte	0,1185	0,1154	0,0094	.	0,0030	0,0030
Nordeste	0,2448	0,2691	-0,0565	.	0,0243	0,0243
Sudeste	0,3266	0,3030	0,0503	.	0,0236	0,0236
Sul	0,2007	0,1982	0,0062	.	0,0025	0,0025
Centro-Oeste	0,1094	0,1143	-0,0155	.	0,0048	0,0048
Ano 2016	0,1487	0,1600	-0,0317	.	0,0113	0,0113
Ano 2017	0,1456	0,1604	-0,0420	.	0,0148	0,0148
Ano 2018	0,1499	0,1584	-0,0238	.	0,0085	0,0085
Ano 2019	0,1565	0,1582	-0,0049	.	0,0018	0,0018
Ano 2020	0,1245	0,1114	0,0397	.	0,0131	0,0131
Ano 2021	0,1289	0,1128	0,0482	.	0,0161	0,0161
Ano 2022	0,1459	0,1388	0,0201	.	0,0071	0,0071
<b>Observações</b>	<b>58.497</b>	<b>333.715</b>				

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela 5 – Resultados antes do pareamento – Homens

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima
Distância	0,3673	0,0432	1,9570	2,3805	0,4131	0,7824
Idade	42,9709	41,1196	0,1444	0,9327	0,0218	0,0779
Idade ao quadrado	2010,8593	1867,0458	0,1202	1,0607	0,0218	0,0779
Habitantes no domicílio	3,1066	3,4522	-0,2685	0,7114	0,0150	0,0833
Filhos(as)	1,0632	1,2888	-0,2302	0,7073	0,0174	0,0621
Anos de estudo	15,3245	9,2893	3,2780	0,2140	0,3550	0,7810

(continua)

(continuação)

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima
Raça Branca/Amarela	0,6261	0,3999	0,4676	.	0,2262	0,2262
Chefe de família	0,6017	0,5701	0,0645	.	0,0316	0,0316
Região metropolitana	0,4236	0,6501	-0,4585	.	0,2266	0,2266
Setor privado	0,6913	0,9194	-0,4938	.	0,2281	0,2281
Cônjuge	0,6417	0,6616	-0,0416	.	0,0200	0,0200
Período integral	0,6736	0,7713	-0,2085	.	0,0978	0,0978
Norte	0,1179	0,1345	-0,0515	.	0,0166	0,0166
Nordeste	0,1975	0,2813	-0,2107	.	0,0839	0,0839
Sudeste	0,3609	0,2821	0,1640	.	0,0788	0,0788
Sul	0,2093	0,1905	0,0463	.	0,0188	0,0188
Centro-Oeste	0,1145	0,1116	0,0090	.	0,0029	0,0029
Ano 2016	0,1490	0,1626	-0,0381	.	0,0136	0,0136
Ano 2017	0,1458	0,1592	-0,0379	.	0,0134	0,0134
Ano 2018	0,1488	0,1564	-0,0211	.	0,0075	0,0075
Ano 2019	0,1582	0,1550	0,0089	.	0,0032	0,0032
Ano 2020	0,1274	0,1157	0,0353	.	0,0118	0,0118
Ano 2021	0,1242	0,1155	0,0266	.	0,0088	0,0088
Ano 2022	0,1464	0,1357	0,0302	.	0,0107	0,0107
<b>Observações</b>	<b>37.280</b>	<b>546.336</b>				

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela 6 – Resultados antes do pareamento – Todos

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima
Distância	0,4590	0,0589	2,0917	1,925,2	0,4229	0,7910
Idade	42,0869	40,9833	0,0940	0,8267	0,0133	0,0740
Idade ao quadrado	1909,0787	1846,2784	0,0591	0,8970	0,0133	0,0740
Habitantes no domicílio	3,1514	3,4264	-0,2144	0,7212	0,0120	0,0717
Filhos(as)	1,1060	1,2814	-0,1830	0,7089	0,0135	0,0562
Anos de estudo	15,5215	9,7818	3,9171	0,1391	0,3376	0,7879
Gênero	0,6108	0,3792	0,4749	.	0,2316	0,2316
Raça Branca/Amarela	0,5990	0,4092	0,3873	.	0,1898	0,1898
Chefe de família	0,4690	0,5113	-0,0848	.	0,0423	0,0423
Região metropolitana	0,4692	0,6229	-0,3080	.	0,1537	0,1537
Setor privado	0,5537	0,9026	-0,7019	.	0,3489	0,3489
Cônjuge	0,6041	0,6244	-0,0415	.	0,0203	0,0203
Período integral	0,5986	0,7165	-0,2404	.	0,1178	0,1178
Norte	0,1182	0,1273	-0,0280	.	0,0090	0,0090
Nordeste	0,2264	0,2767	-0,1202	.	0,0503	0,0503
Sudeste	0,3399	0,2900	0,1054	.	0,0499	0,0499
Sul	0,2040	0,1934	0,0264	.	0,0106	0,0106
Centro-Oeste	0,1114	0,1126	-0,0039	.	0,0012	0,0012
Ano 2016	0,1488	0,1616	-0,0359	.	0,0128	0,0128
Ano 2017	0,1457	0,1597	-0,0396	.	0,0140	0,0140
Ano 2018	0,1495	0,1572	-0,0214	.	0,0076	0,0076

(continua)

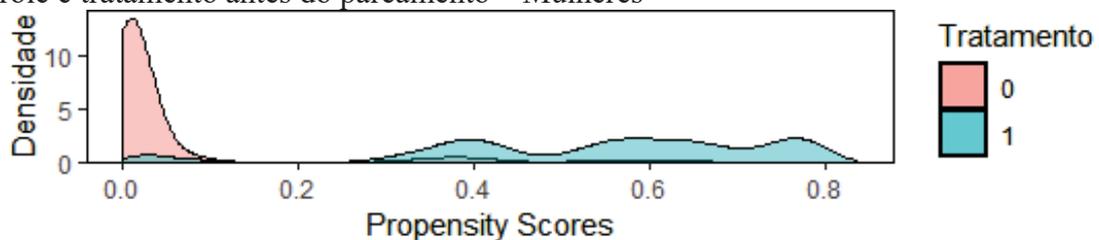
(continuação)

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima
Ano 2019	0,1571	0,1562	0,0026	.	0,0009	0,0009
Ano 2020	0,1256	0,1140	0,0349	.	0,0116	0,0116
Ano 2021	0,1271	0,1145	0,0380	.	0,0127	0,0127
Ano 2022	0,1461	0,1369	0,0261	.	0,0092	0,0092
<b>Observações</b>	<b>95.777</b>	<b>880.051</b>				

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

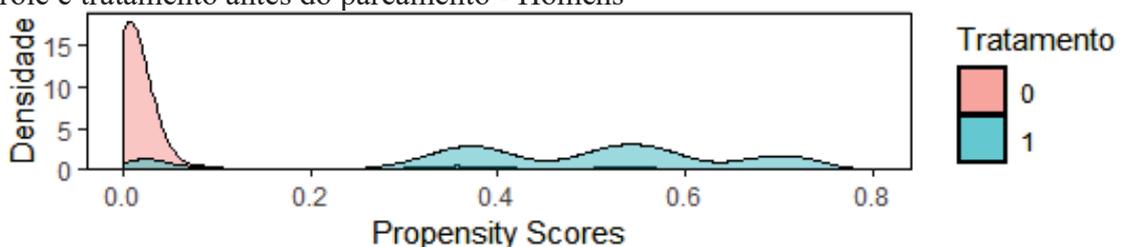
Depois da estimação dos *propensity scores* e antes do pareamento, a região de suporte comum (*common support*) precisa ser avaliada. Isso porque, a partir dela, será possível saber se o pareamento poderá ser realizado, ou seja, vai evidenciar se existe entre os grupos de controle e tratamento uma região de sobreposição, de modo que cada observação do grupo de tratamento possa ser pareada com uma observação do grupo de controle. Uma das formas mais diretas de analisar a sobreposição é visualmente a partir de gráficos de densidade da distribuição dos valores dos *propensity scores* para cada grupo (CALIENDO; KOPEINIG, 2008) conforme os Gráficos 1, 2 e 3 a seguir:

Gráfico 1 – Densidade da distribuição dos valores dos *propensity scores* para os grupos de controle e tratamento antes do pareamento – Mulheres



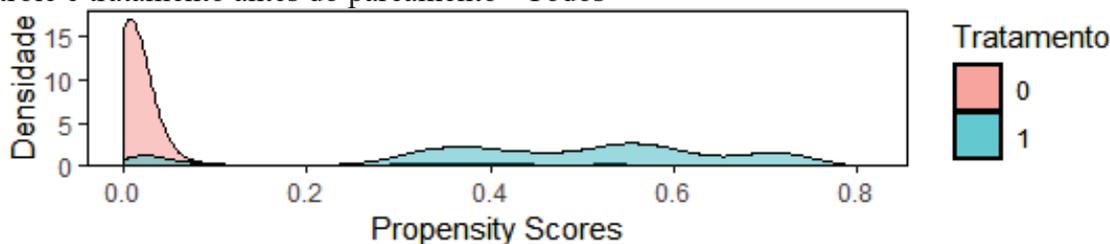
Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Gráfico 2 – Densidade da distribuição dos valores dos *propensity scores* para os grupos de controle e tratamento antes do pareamento - Homens



Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Gráfico 3 – Densidade da distribuição dos valores dos *propensity scores* para os grupos de controle e tratamento antes do pareamento - Todos



Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

A partir dos Gráficos 1, 2 e 3 podemos observar que, ainda que sejam bastante distintas, existem regiões de sobreposição entre os grupos de controle e tratamento em todos os casos. Assim, o próximo passo consiste em parear as observações do grupo de controle com as observações do grupo de tratamento, com o objetivo de criar uma amostra na qual as observações possuem probabilidades similares de serem atribuídas ao tratamento. Nessa etapa, o presente trabalho utilizou o pareamento do *nearest neighbor* (NN) com reposição. As Tabelas 7, 8 e 9 contêm os resultados após o pareamento.

Conforme os resultados reportados nas Tabelas 7, 8 e 9 pode-se verificar que houve uma melhora no balanceamento das covariáveis a partir dos valores das diferenças das médias padronizadas e das densidades acumuladas<sup>16</sup> médias e máximas mais próximas a zero. A variável Anos de Estudo, que antes apresentava a maior diferença entre os grupos, agora é bastante semelhante, evidenciando para todos os casos um valor médio de aproximadamente 15,5 anos.

<sup>16</sup> As estatísticas relacionadas às diferenças nas densidades empíricas acumuladas de cada covariável entre os grupos permitem avaliar o desbalanceamento em toda a distribuição das covariáveis, não apenas em sua média ou variância. O valor da densidade acumulada máxima, também conhecido como estatística Kolmogorov-Smirnov, é por vezes recomendada como uma forma de avaliação adicional do balanceamento (AUSTIN; STUART, 2015).

Tabela 7 – Resultados após o pareamento – Mulheres

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima	Dist. Padrão dos Pares
Distância	0,5274	0,5274	0,0000	1,0000	0,0000	0,0004	0,0000
Idade	41,5235	41,4511	0,0066	1,0045	0,0019	0,0111	0,9000
Idade ao quadrado	1844,2141	1837,6720	0,0068	1,0080	0,0019	0,0111	0,8800
Habitantes no domicílio	3,1799	3,1327	0,0369	1,0643	0,0022	0,0158	0,8700
Filhos(as)	1,1332	1,1096	0,0250	1,0417	0,0020	0,0117	0,8700
Anos de estudo	15,6471	15,6605	-0,0117	1,0409	0,0008	0,0051	0,0700
Raça Branca/Amarela	0,5817	0,6003	-0,0377	.	0,0186	0,0186	0,7180
Chefe de família	0,3844	0,3756	0,0179	.	0,0087	0,0087	0,7700
Região metropolitana	0,4982	0,4839	0,0286	.	0,0143	0,0143	0,6900
Sector privado	0,4661	0,4788	-0,0255	.	0,0127	0,0127	0,2500
Cônjuge	0,5801	0,5828	-0,0055	.	0,0027	0,0027	0,7400
Período integral	0,5508	0,5592	-0,0168	.	0,0084	0,0084	0,3200
Norte	0,1185	0,1160	0,0075	.	0,0024	0,0024	0,5100
Nordeste	0,2448	0,2406	0,0099	.	0,0042	0,0042	0,6500
Sudeste	0,3266	0,3298	-0,0069	.	0,0032	0,0032	0,6300
Sul	0,2007	0,2027	-0,0051	.	0,0021	0,0021	0,5600
Centro-Oeste	0,1094	0,1108	-0,0044	.	0,0014	0,0014	0,4800
Ano 2016	0,1487	0,1503	-0,0044	.	0,0016	0,0016	0,5300
Ano 2017	0,1456	0,1483	-0,0077	.	0,0027	0,0027	0,5400
Ano 2018	0,1499	0,1505	-0,0017	.	0,0006	0,0006	0,5600
Ano 2019	0,1565	0,1605	-0,0110	.	0,0040	0,0040	0,5700
Ano 2020	0,1245	0,1183	0,0186	.	0,0062	0,0062	0,5100
Ano 2021	0,1289	0,1263	0,0078	.	0,0026	0,0026	0,5300
Ano 2022	0,1459	0,1458	0,0002	.	0,0001	0,0001	0,5400
<b>Observações</b>	<b>58.497</b>	<b>333.715</b>					

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela 8 – Resultados após o pareamento – Homens

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima	Dist. Padrão dos Pares
Distância	0,3673	0,3673	0,0000	1,0000	0,0000	0,0003	0,0000
Idade	42,9709	42,9910	-0,0016	1,0451	0,0037	0,0126	0,7420
Idade ao quadrado	2010,8593	2005,4752	0,0045	1,0516	0,0037	0,0126	0,7200
Habitantes no domicílio	3,1066	3,0769	0,0231	1,0738	0,0017	0,0153	0,7860
Filhos(as)	1,0632	1,0569	0,0064	1,0721	0,0021	0,0107	0,7700
Anos de estudo	15,3245	15,3267	-0,0012	1,0408	0,0011	0,0078	0,0400
Raça Branca/Amarela	0,6261	0,6425	-0,0338	.	0,0163	0,0163	0,6556
Chefe de família	0,6017	0,6149	-0,0270	.	0,0132	0,0132	0,6920
Região metropolitana	0,4236	0,4232	0,0008	.	0,0004	0,0004	0,6440
Setor privado	0,6913	0,7031	-0,0257	.	0,0119	0,0119	0,6540
Cônjuge	0,6417	0,6530	-0,0236	.	0,0113	0,0113	0,7200
Período integral	0,6736	0,6801	-0,0140	.	0,0065	0,0065	0,2600
Norte	0,1179	0,1086	0,0290	.	0,0094	0,0094	0,5310
Nordeste	0,1975	0,1916	0,0148	.	0,0059	0,0059	0,6130
Sudeste	0,3609	0,3707	-0,0206	.	0,0099	0,0099	0,6510
Sul	0,2093	0,2158	-0,0159	.	0,0065	0,0065	0,5460
Centro-Oeste	0,1145	0,1134	0,0035	.	0,0011	0,0011	0,5010
Ano 2016	0,1490	0,1509	-0,0052	.	0,0019	0,0019	0,4970
Ano 2017	0,1458	0,1494	-0,0102	.	0,0036	0,0036	0,5130
Ano 2018	0,1488	0,1514	-0,0072	.	0,0025	0,0025	0,5130
Ano 2019	0,1582	0,1524	0,0160	.	0,0058	0,0058	0,5180
Ano 2020	0,1274	0,1260	0,0043	.	0,0014	0,0014	0,4850
Ano 2021	0,1242	0,1224	0,0056	.	0,0019	0,0019	0,4800
Ano 2022	0,1464	0,1475	-0,0031	.	0,0011	0,0011	0,5230
<b>Observações</b>	<b>37,280</b>	<b>546,336</b>					

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela 9 – Resultados após o pareamento – Todos

Covariáveis	Média do Tratado	Média do Controle	Diferença das médias padronizadas	Variância	Densidade Acumulada Média	Densidade Acumulada Máxima	Dist. Padrão dos Pares
Distância	0,4590	0,4590	0,0000	1,0000	0,0000	0,0002	0,0000
Idade	42,0869	42,0860	0,0001	1,0170	0,0015	0,0056	0,8582
Idade ao quadrado	1909,0787	1906,6931	0,0022	1,0292	0,0015	0,0056	0,8358
Habitantes no domicílio	3,1514	3,1071	0,0345	1,0764	0,0021	0,0171	0,8447
Filhos(as)	1,1060	1,0855	0,0213	1,0641	0,0021	0,0116	0,8320
Anos de estudo	15,5215	15,5266	-0,0035	1,0257	0,0005	0,0015	0,0617
Gênero	0,6108	0,5902	0,0422	.	0,0206	0,0206	0,6190
Raça Branca/Amarela	0,5990	0,6174	-0,0376	.	0,0184	0,0184	0,7200
Chefe de família	0,4690	0,4757	-0,0136	.	0,0068	0,0068	0,7652
Região metropolitana	0,4692	0,4560	0,0263	.	0,0131	0,0131	0,7469
Setor privado	0,5537	0,5618	-0,0162	.	0,0080	0,0080	0,4492
Cônjuge	0,6041	0,6151	-0,0227	.	0,0111	0,0111	0,7628
Período integral	0,5986	0,6013	-0,0054	.	0,0027	0,0027	0,4788
Norte	0,1182	0,1112	0,0217	.	0,0070	0,0070	0,5425
Nordeste	0,2264	0,2199	0,0155	.	0,0065	0,0065	0,6698
Sudeste	0,3399	0,3514	-0,0243	.	0,0115	0,0115	0,6927
Sul	0,2040	0,2092	-0,0128	.	0,0052	0,0052	0,5732
Centro-Oeste	0,1114	0,1082	0,0101	.	0,0032	0,0032	0,5019
Ano 2016	0,1488	0,1516	-0,0079	.	0,0028	0,0028	0,5434
Ano 2017	0,1457	0,1494	-0,0106	.	0,0037	0,0037	0,5530
Ano 2018	0,1495	0,1503	-0,0023	.	0,0008	0,0008	0,5561
Ano 2019	0,1571	0,1575	-0,0009	.	0,0003	0,0003	0,5625
Ano 2020	0,1256	0,1209	0,0142	.	0,0047	0,0047	0,5143
Ano 2021	0,1271	0,1232	0,0118	.	0,0039	0,0039	0,5196
Ano 2022	0,1461	0,1470	-0,0027	.	0,0010	0,0010	0,5530
<b>Observações</b>	<b>95.777</b>	<b>880.051</b>					

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Os elementos-chave obtidos com o pareamento são: os pesos computados pelo *matching*, o par de observações pareadas (subclasse)<sup>17</sup>, o *propensity score* (distância padrão dos pares) e a matriz do *matching*, que mostra com qual observação do grupo de controle cada observação do grupo de tratamento foi pareada. Além disso, quando o método permite a possibilidade de reposição, o tamanho efetivo da amostra é calculado, ou seja, uma estimativa do número de observações do grupo de controle que pode ser comparável com o grupo de tratamento (MCCAFFREY; RIDGEWAY; MORRAL, 2004)<sup>18</sup>.

Após a estimação dos valores de *propensity scores*, a qualidade do *matching* precisa ser investigada. Em primeiro lugar, as covariáveis dos grupos de controle e de tratamento precisam estar balanceadas (GREIFER, 2022). Thoemmes e Kim (2011) apontam para a importância de reportar o balanceamento com o intuito de demonstrar a adequação das análises e a validade dos efeitos causais.

Geralmente, o balanceamento das covariáveis é avaliado e reportado a partir da análise de medidas como a diferença das médias padronizadas, índices de variância, teste t ou teste de Kolmogorov-Smirnov<sup>19</sup>. A partir dos dados reportados nas Tabelas 10, 11 e 12, podemos observar valores das diferenças das médias padronizadas e das densidades acumuladas médias e máximas mais próximos a zero, indicando uma melhora no balanceamento das covariáveis. Stuart, Lee e Leacy (2013) apontam que valores das diferenças das médias padronizadas menores do que 0,1 sugerem um balanceamento satisfatório. Para a variância, o *threshold* adequado é de 2 (RUBIN, 2001). As Tabelas 10, 11 e 12 trazem essa informação. Pode-se observar que todas as covariáveis estão balanceadas. A melhoria após o pareamento também pode ser visualizada graficamente a partir dos Gráficos 4, 5 e 6.

---

<sup>17</sup> Para mais detalhes sobre quais profissões foram pareadas com o grupo de tratamento, ver Tabela A1 nos Apêndices.

<sup>18</sup> Considerando a amostra total, o grupo de controle é composto por 880.051 observações e o grupo de tratamento por 95.777. Foram pareadas 52.004 observações e o tamanho efetivo da amostra foi de 32,694,87. As observações não pareadas foram 828.047. Quando consideramos só a parcela da amostra contendo mulheres, 333.715 observações formam o grupo de controle e 58.497 o grupo de tratamento. Foram pareadas 28.208 observações e o tamanho efetivo da amostra foi de 16.835,5. As observações não pareadas foram 305.507. A base contendo apenas homens, por sua vez, possui um grupo de controle e um grupo de tratamento formados por 546.336 e 37.280 observações respectivamente. Foram pareadas 23.357 observações e o tamanho efetivo da amostra foi de 16.261,04. As observações não pareadas foram 522.979.

<sup>19</sup> Austin (2011), entretanto, aponta dois motivos pelos quais o teste t para analisar o balanceamento das covariáveis após o tratamento não é adequado. O primeiro motivo diz respeito ao tamanho da amostra, uma vez que após o pareamento o tamanho da amostra normalmente diminui e isso pode afetar a validade do teste. Já o segundo fator se relaciona com o fato de que o balanceamento é propriedade de uma amostra em particular e que a generalização para a população pode ser inadequada.

Tabela 10 – Balanceamento das covariáveis após o pareamento – Mulheres

Covariáveis	Tipo	Dif. Médias Pad. Antes*	Variância**	Dif. Médias Pad. Depois
Distância	Distância	2,2092	1,0000	0,0000
Idade	Contínua	0,0697	1,0045	0,0066
Idade ao quadrado	Contínua	0,0332	1,0080	0,0068
Habitantes no domicílio	Contínua	-0,1597	1,0643	0,0369
Filhos(as)	Contínua	-0,1443	1,0417	0,0250
Anos de estudo	Contínua	4,4122	1,0409	-0,0117
Raça Branca/Amarela	Binária	0,1573		-0,0186
Chefe de família	Binária	-0,0306		0,0087
Região metropolitana	Binária	-0,0800		0,0143
Setor privado	Binária	-0,4091		-0,0127
Cônjuge	Binária	0,0167		-0,0027
Período integral	Binária	-0,0758		-0,0084
Norte	Binária	0,0030		0,0024
Nordeste	Binária	-0,0243		0,0042
Sudeste	Binária	0,0236		-0,0032
Sul	Binária	0,0025		-0,0021
Centro-Oeste	Binária	-0,0048		-0,0014
Ano 2016	Binária	-0,0113		-0,0016
Ano 2017	Binária	-0,0148		-0,0027
Ano 2018	Binária	-0,0085		-0,0006
Ano 2019	Binária	-0,0018		-0,0040
Ano 2020	Binária	0,0131		0,0062
Ano 2021	Binária	0,0161		0,0026
Ano 2022	Binária	0,0071		0,0001

Nota: \* *Threshold* da Média, Balanceado, <0,1 para todas as variáveis. \*\* *Threshold* da Variância, Balanceado, <2 para todas as variáveis.

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela 11 – Balanceamento das covariáveis após o pareamento – Homens

Covariáveis	Tipo	Dif. Médias Pad. Antes*	Variância**	Dif. Médias Pad. Depois
Distância	Distância	1,9570	1,0000	0,0000
Idade	Contínua	0,1444	1,0451	-0,0016
Idade ao quadrado	Contínua	0,1202	1,0516	0,0045
Habitantes no domicílio	Contínua	-0,2685	1,0738	0,0231
Filhos(as)	Contínua	-0,2302	1,0721	0,0064
Anos de estudo	Contínua	3,2780	1,0408	-0,0012
Raça Branca/Amarela	Binária	0,2262		-0,0163
Chefe de família	Binária	0,0316		-0,0132
Região metropolitana	Binária	-0,2266		0,0004
Setor privado	Binária	-0,2281		-0,0119
Cônjuge	Binária	-0,0200		-0,0113
Período integral	Binária	-0,0978		-0,0065
Norte	Binária	-0,0166		0,0094
Nordeste	Binária	-0,0839		0,0059

(continua)

(continuação)

Covariáveis	Tipo	Dif. Médias Pad. Antes*	Variância**	Dif. Médias Pad. Depois
Sudeste	Binária	0,0788		-0,0099
Sul	Binária	0,0188		-0,0065
Centro-Oeste	Binária	0,0029		0,0011
Ano 2016	Binária	-0,0136		-0,0019
Ano 2017	Binária	-0,0134		-0,0036
Ano 2018	Binária	-0,0075		-0,0025
Ano 2019	Binária	0,0032		0,0058
Ano 2020	Binária	0,0118		0,0014
Ano 2021	Binária	0,0088		0,0019
Ano 2022	Binária	0,0107		-0,0011

Nota: \* *Threshold* da Média, Balanceado, <0,1 para todas as variáveis. \*\* *Threshold* da Variância, Balanceado, <2 para todas as variáveis.

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

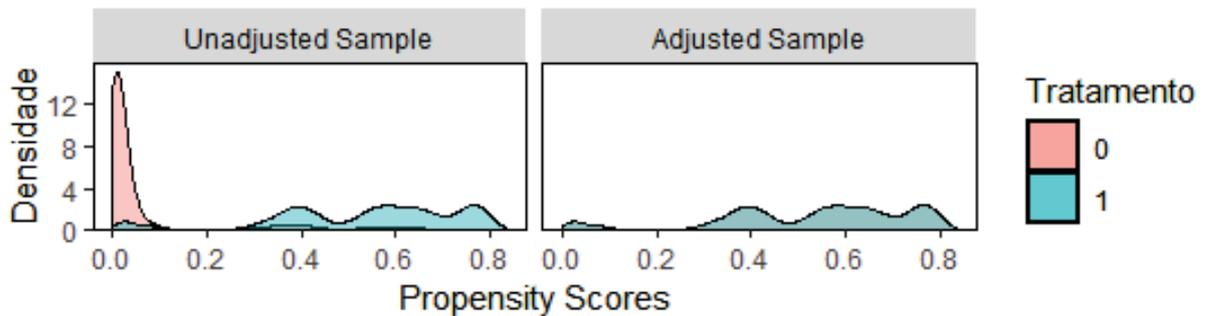
Tabela 12 – Balanceamento das covariáveis após o pareamento – Todos

Covariáveis	Tipo	Dif. Médias Pad. Antes*	Variância**	Dif. Médias Pad. Depois
Distância	Distância	2,0917	1,0000	0,0000
Idade	Contínua	0,0940	1,0170	0,0001
Idade ao quadrado	Contínua	0,0591	1,0238	0,0022
Habitantes no domicílio	Contínua	-0,2144	1,0764	0,0345
Filhos(as)	Contínua	-0,1830	1,0641	0,0213
Anos de estudo	Contínua	3,9171	1,0257	-0,0035
Gênero	Binária	0,2316		0,0206
Raça Branca/Amarela	Binária	0,1898		-0,0184
Chefe de família	Binária	-0,0423		-0,0068
Região metropolitana	Binária	-0,1537		0,0131
Setor privado	Binária	-0,3489		-0,0080
Cônjuge	Binária	-0,0203		-0,0111
Período integral	Binária	-0,1178		-0,0027
Norte	Binária	-0,0090		0,0070
Nordeste	Binária	-0,0503		0,0065
Sudeste	Binária	0,0499		-0,0115
Sul	Binária	0,0106		-0,0052
Centro-Oeste	Binária	-0,0012		0,0032
Ano 2016	Binária	-0,0128		-0,0028
Ano 2017	Binária	-0,0140		-0,0037
Ano 2018	Binária	-0,0076		-0,0008
Ano 2019	Binária	0,0009		-0,0003
Ano 2020	Binária	0,0116		0,0047
Ano 2021	Binária	0,0127		0,0039
Ano 2022	Binária	0,0092		-0,0010

Nota: \* *Threshold* da Média, Balanceado, <0,1 para todas as variáveis. \*\* *Threshold* da Variância, Balanceado, <2 para todas as variáveis.

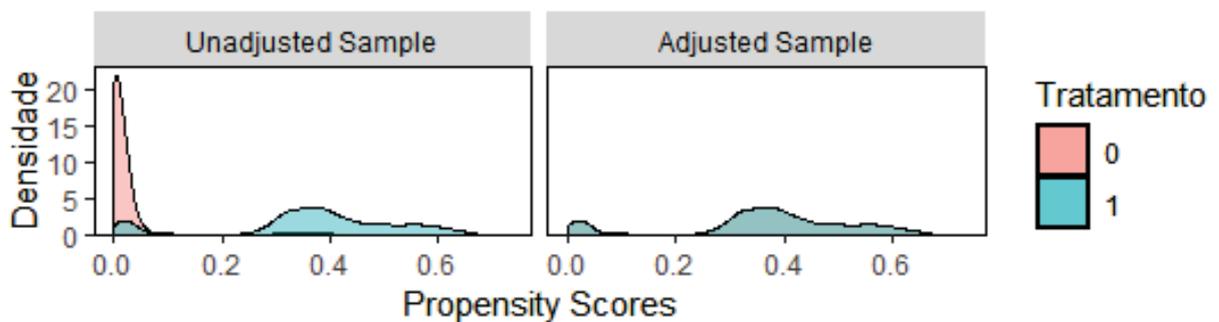
Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Gráfico 4 – Densidade da distribuição dos valores dos *propensity scores* para cada grupo antes e depois do pareamento – Mulheres



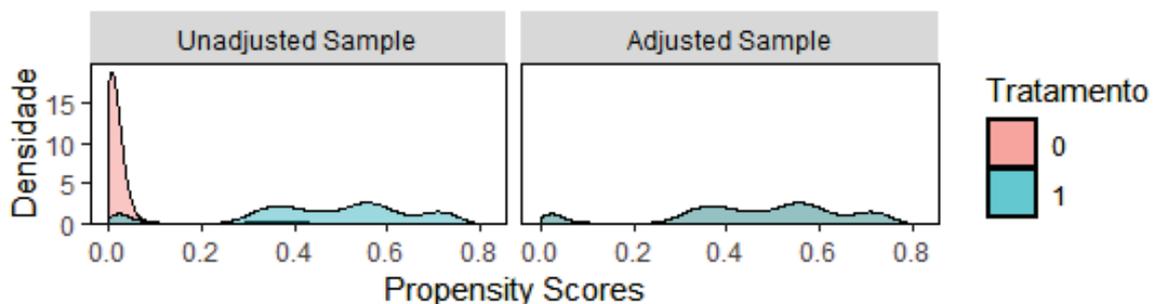
Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Gráfico 5 – Densidade da distribuição dos valores dos *propensity scores* para cada grupo antes e depois do pareamento – Homens



Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Gráfico 6 – Densidade da distribuição dos valores dos *propensity scores* para cada grupo antes e depois do pareamento - Todos



Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Após a avaliação e validação da qualidade do *matching*, o próximo e último passo consiste em estimar o efeito do tratamento e seus erros-padrão. Para isso, foi ajustado um modelo de regressão linear com o salário/hora sendo a variável dependente e as covariáveis e

os valores do *propensity score* estimados como as variáveis independentes<sup>20</sup>. A Tabela 13 a seguir reporta os valores encontrados.

Tabela 13 – Efeito médio do tratamento sobre os tratados (ATT)

	ATT	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	5,954***	0,493	$< 2,22e-16$	4,987	6,920
Homens	10,750***	0,778	$< 2,22e-16$	9,223	12,270
Todos	7,614***	0,429	$< 2,22e-16$	6,772	8,456

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

O valor encontrado para o efeito médio do tratamento sobre os tratados (ATT) foi de 7,614 (SE = 0,4429;  $p < 0,001$ ), ou seja, ser intelectual atua de forma a aumentar o valor do salário/hora em aproximadamente R\$ 7,61 de todos os indivíduos da base de dados.

Se não houvesse diferenças de gênero no prêmio salarial por fazer parte do grupo de intelectuais, então o resultado da amostra completa deveria ser estatisticamente idêntico ao resultado para mulheres e homens avaliados separadamente. Visando testar essa hipótese, quando analisamos separadamente para cada gênero, encontramos os valores de 10,750 (SE = 0,778;  $p < 0,001$ ) para os homens e 5,954 (SE = 0,463;  $p < 0,001$ ) para as mulheres. Ou seja, enquanto ser intelectual atua de forma a aumentar o salário/hora recebido pelos homens em cerca de R\$ 10,75, o valor estimado para as mulheres é de apenas R\$ 5,95, aproximadamente 55% do valor dos homens. A significância dessa diferença fica ainda evidente quando analisamos o intervalo de confiança dos resultados. Enquanto o efeito para os homens pode variar de um limite inferior de R\$ 9,22 e superior de R\$ 12,27 o das mulheres varia entre R\$ 4,98 e R\$ 6,92. Ou seja, os valores mais altos que as mulheres tendem a ganhar por serem intelectuais ficam abaixo dos valores mais baixos que os homens tendem ganhar por estarem na mesma categoria, indicando que as duas distribuições não se sobrepõem.

### 2.4.3 Análise de robustez

Algumas hipóteses podem explicar a diferença no prêmio salarial entre mulheres e homens por atuarem em ocupações do grande grupo de “Profissionais das Ciências e Intelectuais”. Uma delas está relacionada ao viés de variáveis omitidas. A base de dados

<sup>20</sup> Os coeficientes estimados das regressões estão reportados nas Tabelas A2, A3 e A4 do Apêndice.

utilizada não permite acompanhar os indivíduos ao longo do tempo, e, por consequência, não é possível eliminar, por exemplo, o efeito de suas habilidades não observadas. Não obstante, assumindo que as habilidades são igualmente distribuídas entre mulheres e homens, o mesmo viés estaria presente em ambas as estimações e pouco afetaria a diferença observada no prêmio salarial entre os gêneros. Ainda assim, podem existir outras variáveis cuja distribuição é distinta entre mulheres e homens e que afetam os salários de ambos. Nesse sentido, as limitações da base de dados não permitem qualquer tipo de avanço.

Por outro lado, algumas hipóteses podem ser testadas a partir dos dados em mãos. A principal delas diz respeito à heterogeneidade dos indivíduos que compõem a categoria dos “Profissionais das Ciências e Intelectuais”. Nesse sentido, foram realizados sete testes para investigar essa questão e seus resultados estão detalhados nas próximas seções.

O primeiro teste analisou se o prêmio salarial por pertencer ao grupo de tratamento ainda se manteria caso apenas os indivíduos que auferissem maiores remunerações (10%, 5% e 1%) fossem considerados. O segundo teste estimou os resultados separadamente para os indivíduos que trabalham meio período e para aqueles que trabalham período integral enquanto o terceiro teste analisou os resultados separadamente para indivíduos que trabalham no setor público e para aqueles que atuam no setor privado. O quarto teste considerou apenas os indivíduos que possuem ensino superior. O quinto teste, por sua vez, transferiu os indivíduos da categoria “Ensino” para grupo de controle (com exceção dos profissionais do “Ensino Superior”) por serem aqueles que recebem menores salários. O mesmo teste foi realizado para o subgrupo principal da “Saúde” no qual os profissionais da saúde, com exceção dos “Médicos(as)” e “Dentistas”, foram transferidos para o grupo de controle. O sexto teste analisou o efeito do tratamento sobre os seis subgrupos da categoria de intelectuais separadamente, na tentativa de investigar se existem evidências apontando para a existência de segregação ocupacional entre os gêneros em cada subgrupo. E, por fim, o sétimo e último teste analisou apenas os anos de 2016 a 2019 para contornar possíveis efeitos da pandemia da COVID-19.

#### *2.4.3.1 Maiores remunerações*

O primeiro teste considerou apenas os indivíduos que estão entre os 10% com maiores salários. Os resultados da estimação estão reportados na Tabela 14 a seguir.

Tabela 14 – Efeito médio do tratamento sobre os tratados para o décimo decil de renda

	Efeito (ATT)	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	-2,219	1,586	0,1619	-5,328	0,890
Homens	3,952**	1,664	0,0175	0,690	7,214
Todos	1,050	1,161	0,3657	-1,225	3,326

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Pode-se observar a partir dos resultados da Tabela 14 que o diferencial por gênero no prêmio salarial por pertencer ao grupo de “Profissionais das Ciências e Intelectuais” ainda se mantém, com diferenças significativas apenas para os homens. Além disso, tais valores apontam para a existência do efeito teto de vidro<sup>21</sup> nessa categoria. Isso porque os homens que estão entre os 10% mais bem remunerados recebem um prêmio salarial de aproximadamente R\$ 4 por hora enquanto o valor estatisticamente não significativo para as mulheres sugere que não há prêmio salarial para aquelas que estão nessa categoria de remuneração.

Também foram realizados testes com os indivíduos que estão entre os 5% e 1% com maiores salários. No primeiro caso apenas o resultado para os homens foi estatisticamente significativo, com um valor de -0,036 (SE = 0,0149,  $p < 0,05$ ). No segundo caso nenhum resultado apresentou significância estatística.

#### 2.4.3.2 Formação superior

Uma das características produtivas mais importante na determinação dos salários é o nível de educação dos indivíduos. Essa observação conduz a segunda hipótese da análise de robustez: os diferenciais podem ser explicados por diferentes requerimentos no nível de qualificação dos indivíduos. Sendo assim, na tentativa de investigar se o nível de escolaridade está afetando os resultados, o ATT foi estimado considerando apenas indivíduos que declararam ter formação superior (graduação, especialização, mestrado ou doutorado). A Tabela 15 a seguir contém os valores encontrados.

Pode-se observar que os resultados encontrados são similares aos resultados do modelo geral, indicando que o nível de escolaridade não está afetando os valores, tal qual sugere Bütikofer, Jensen e Salvanes (2018). Além disso, os resultados sugerem que o subgrupo de

<sup>21</sup> O teto de vidro ocorre quando o diferencial salarial entre mulheres e homens é maior na parcela superior da curva de distribuição dos salários, isto é, uma barreira artificial que dificulta a ascensão profissional de indivíduos que já ocupam bons cargos (CHRISTOFIDES; POLYCARPOU; VRACHIMIS, 2013).

ensino possui características específicas, corroborando a ideia de que outros fatores, expliquem as diferenças entre os subgrupos dentro de “Profissionais das Ciências e Intelectuais”.

Tabela 15 – Efeito médio do tratamento sobre os tratados considerando a base de dados apenas com indivíduos que declararam ter formação superior

	Efeito (ATT)	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	5,487***	0,548	$< 2,22e-16$	4,412	6,562
Homens	11,200***	0,912	$< 2,22e-16$	9,409	12,980
Todos	7,820***	0,478	$< 2,22e-16$	6,881	8,758

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

#### 2.4.3.3 *Profissionais de ensino e saúde*

A terceira hipótese da análise de robustez está relacionada à diversidade de ocupações dentro do grupo de intelectuais. Ou seja, uma parte da diferença nos salários entre mulheres e homens pode decorrer da concentração em ocupações específicas que foram agrupadas em “Profissionais das Ciências e Intelectuais”, a saber: “Profissionais das Ciências e das Engenharias” (Exatas), “Profissionais da Saúde” (Saúde), “Profissionais do Ensino” (Ensino), “Especialistas em Organização da Administração Pública e de Empresas” (Administração), “Profissionais de Tecnologias da Informação e Comunicações” (TI) e “Profissionais em Direito, em Ciências Sociais e Culturais” (Sociais).

A partir dos valores reportados pela Tabela 16 a seguir, é possível observar que cerca de 52% das mulheres que declararam ocupar uma função do grande grupo de intelectuais faz parte do subgrupo principal “Profissionais do Ensino”. Os dados mostram também que o salário/hora das mulheres desse subgrupo principal é o mais baixo entre todos os subgrupos e, aproximadamente, 72% do salário dos homens da mesma categoria.

O segundo subgrupo principal que mais concentra mulheres é o “Profissionais da Saúde” com 18%, sendo que nesse caso elas ganham cerca de 59% dos salários dos homens da mesma categoria.

Tabela 16 – Estatísticas descritivas para o grande grupo de “Profissionais das Ciências e Intelectuais”

Subgrupos principais	Média do salário/hora		Desvio-Padrão		Observações		% entre os subgrupos principais		% dentro dos subgrupos principais		Dif. Sal
	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	
Exatas	40,00	54,70	0,82	1,10	2.496	5.477	4%	15%	31%	69%	73%
Saúde	46,00	78,10	0,64	1,60	10.316	4.902	18%	13%	68%	32%	59%
Ensino	30,70	42,70	0,28	0,87	30.692	8.971	52%	24%	77%	23%	72%
Administração	39,70	57,30	1,80	1,99	5.783	6.442	10%	17%	47%	53%	69%
TI	44,40	48,20	1,33	0,90	647	2.656	1%	7%	20%	80%	92%
Sociais	46,20	52,20	0,81	1,13	8.563	8.832	15%	24%	49%	51%	89%
Total					58.497	37.280	100%	100%	61%	39%	

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

De acordo com a Classificação de Ocupação para Pesquisas Domiciliares da PNAD Contínua, o subgrupo principal “Profissionais do Ensino” pode ser dividido em cinco subgrupos: “Professores de Universidades e do Ensino Superior”; “Professores de Formação Profissional”; “Professores do Ensino Médio”; “Professores do Ensino Fundamental e Pré-Escolar” e; “Outros Profissionais do Ensino”<sup>22</sup>.

A Tabela 17 a seguir mostra que, com exceção dos profissionais que atuam no ensino superior, o restante dos subgrupos apresenta médias baixas de salário/hora quando comparados aos valores das outras categorias do grande grupo de intelectuais. Nota-se também uma representatividade feminina acentuada nos subgrupos de “Ensino Médio”, “Outros” e, principalmente, no “Ensino Fundamental e Pré-Escolar” no qual cerca de 87% dos indivíduos que declararam ocupar uma função desse subgrupo são mulheres. Além disso, os dados mostram também que apesar da representatividade significativa das mulheres nesses subgrupos, seus salários médios ainda são inferiores aos dos homens.

---

<sup>22</sup> O subgrupo “Outros profissionais de ensino” é composto por “Especialistas em métodos pedagógicos”; “Educadores para necessidades especiais”; “Outros professores de idiomas”; “Outros professores de música”; “Outros professores de artes”; “Instrutores em tecnologia da informação” e “Profissionais do ensino não classificados anteriormente.

Tabela 17 – Estatísticas descritivas para o subgrupo principal “Profissionais do Ensino”

Subgrupos	Média do salário/hora		Desvio-Padrão		Observações		% entre os subgrupos		% dentro dos subgrupos		Dif. Sal
	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	
Superior	72,90	89,60	1,98	3,95	1.332	1.442	4%	16%	48%	52%	81%
Profissionalizante	25,10	30,60	1,35	3,27	300	380	1%	4%	44%	56%	82%
Médio	35,90	37,80	1,22	0,77	4.450	2.658	14%	30%	63%	37%	95%
Fundamental e Pré-Escolar	27,50	30,80	0,26	0,95	19.886	2.935	65%	33%	87%	13%	89%
Outros	27,40	32,80	0,45	2,18	4.724	1.556	15%	17%	75%	25%	84%
Total					30.692	8.971	100%	100%	77%	23%	

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

O subgrupo principal de “Profissionais da Saúde” é subdividido em outros seis subgrupos de acordo com a Classificação de Ocupações da PNAD Contínua: “Médicos”; “Profissionais de enfermagem e partos”; “Profissionais da medicina tradicional e alternativa”; “Paramédicos”; “Veterinários” e “Outros profissionais da saúde”<sup>23</sup>. Devido ao fato de alguns subgrupos terem poucas observações, as mesmas foram reclassificadas em sete subgrupos: “Médicos(as)”; “Dentistas”; “Profissionais da enfermagem e partos”; “Veterinários(as)”; “Fisioterapeutas”; “Farmacêuticos(as)”; e “Outros profissionais da saúde”<sup>24</sup>.

A Tabela 18 a seguir mostra que a média do salário/hora do subgrupo da saúde é de aproximadamente R\$ 33,76 para as mulheres e 35,74 para os homens quando são desconsiderados do cálculo os subgrupos de “Médicos(as)” e “Dentistas” por serem aqueles que apresentam maiores médias salariais. Além disso, entre os subgrupos da saúde, “Profissionais da enfermagem e partos” é o que mais concentra mulheres em relação aos homens, com 30% e 11%, respectivamente. Nessa categoria, a parcela feminina ganha cerca de 113% do salário da parcela masculina.

Os subgrupos de “Fisioterapeutas” e “Farmacêuticos(as)” apresentam valores mais igualitários entre os gêneros no que diz respeito ao salário recebido por hora, porém o número de indivíduos atuantes é bastante distinto, com as mulheres sendo a maioria. O subgrupo de “Veterinários(as)”, por sua vez, apresenta quantidade similar de indivíduos entre os gêneros, mas, em termos salariais, as mulheres ganham apenas 82% do salário dos homens, aproximadamente.

Quando os subgrupos que auferem maiores níveis salariais são analisados, os dados mostram que as mulheres médicas e dentistas ganham, respectivamente, cerca de 80% e 82% dos salários de seus colegas homens.

---

<sup>23</sup> De acordo com a PNAD Contínua, o subgrupo principal “Outros profissionais da saúde” é formado por “Dentistas”; “Farmacêuticos”; “Profissionais da saúde e da higiene laboral e ambiental”; “Fisioterapeutas”; “Dietistas e nutricionistas”; “Fonoaudiólogos”; “Optometristas” e “Profissionais da saúde não classificados anteriormente”.

<sup>24</sup> O presente estudo reclassificou o subgrupo principal “Outros profissionais da saúde” como sendo formado pelos seguintes subgrupos: “Profissionais da medicina tradicional e alternativa”; “Paramédicos”; “Profissionais da saúde e da higiene laboral e ambiental”; “Dietistas e nutricionistas”; “Fonoaudiólogos”; “Optometristas” e “Profissionais da saúde não classificados anteriormente”.

Tabela 18 – Estatísticas descritivas para o subgrupo principal “Profissionais da Saúde”

Subgrupos	Média do salário/hora		Desvio-Padrão		Observações		% entre os subgrupos principais		% dentro dos subgrupos principais		Dif. Sal
	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	
Médicos(as)	101,00	126,00	3,31	2,76	1611	2006	16%	41%	45%	55%	80%
Dentistas	52,20	63,30	1,65	2,56	1455	972	14%	20%	60%	40%	82%
Enfermeiros(as)	34,90	30,80	2,02	0,48	3151	541	30%	11%	85%	15%	113%
Veterinários(as)	32,60	40,00	1,67	1,62	420	363	4%	7%	54%	46%	82%
Fisioterapeutas	35,90	36,00	1,44	2,57	1351	375	13%	8%	78%	22%	100%
Farmacêuticos(as)	31,80	31,30	1,58	1,15	1011	435	10%	9%	70%	30%	102%
Outros	33,60	40,60	1,08	4,10	1374	153	13%	3%	90%	10%	83%
Total					10373	4845	100%	100%	68%	32%	

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nesse sentido, pode-se perceber que tanto os subgrupos principais “Ensino” quanto “Saúde” são bastante heterogêneos no que diz respeito aos subgrupos que os compõem. Os “Professores(as) do ensino superior”, “Médicos(as)” e “Dentistas” apresentam médias salariais que destoam das demais categorias.

Diante dessas informações, foram realizados dois testes: o primeiro analisou o que aconteceria com o diferencial de salário recebido por hora caso fosse retirado do grupo de tratamento e transferido para o grupo de controle o subgrupo principal “Profissionais do Ensino” (com exceção do subgrupo “Profissionais do Ensino Superior”), por se tratar dos subgrupos que auferem menores níveis salariais. O segundo teste, por sua vez, seguiu o mesmo princípio e transferiu para o grupo de controle os “Profissionais da Saúde” (com exceção dos “Médicos(as)” e “Dentistas”). A Tabela 19 a seguir apresenta os valores encontrados no primeiro caso e a Tabela 20 no segundo.

Tabela 19 – Efeito médio do tratamento sobre os tratados considerando a base de dados sem os subgrupos do Ensino que auferem menores níveis salariais

	ATT	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	16,320***	0,7668	$< 2,22e-16$	14,810	17,820
Homens	15,770***	0,9496	$< 2,22e-16$	13,910	17,630
Todos	16,680***	0,5735	$< 2,22e-16$	15,550	17,800

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Quando a parcela dos profissionais do ensino que auferem menores salários é transferida para o grupo de controle, os resultados reportados na Tabela 19 mostram que as disparidades entre os prêmios salariais não se mantêm. Isso indica que o diferencial salarial estimado é parcialmente explicado pela concentração de mulheres em ocupações que pagam salários médios menores. Excluindo esse subgrupo, o resultado indica que os salários entre mulheres e homens são estatisticamente semelhantes para os demais “Profissionais das Ciências e Intelectuais”.

Sendo assim, há indícios de que a existência de diferenciais de salário entre mulheres e homens no grupo de “Profissionais das Ciências e Intelectuais” é decorrência da segregação ocupacional de mulheres em ocupações com menores salários, seja por características produtivas (não observadas) diferentes, seja porque são ocupações menos valorizadas no mercado de trabalho.

De acordo com Lassibile e Gomez (2020), em todo o mundo, as mulheres ocupam atualmente cerca de 66% e 57% dos postos de professores no ensino primário e no ensino médio

secundário, respectivamente. As autoras apontam também que a transformação do ensino em uma profissão feminina em muitos países remete ao início da expansão do sistema escolar. Fatores como baixos salários, a associação do ensino com o cuidado e com as habilidades que a sociedade normalmente atribui às mulheres e o pouco prestígio da profissão são exemplos de fatores que podem estar contribuindo para essa realidade.

Além do Ensino, a área da Saúde também está relacionada ao cuidado. Por isso, foi testado o que aconteceria caso os “Profissionais da Saúde” (com exceção dos(as) “Médicos(as)” e “Dentistas”) fossem transferidos para o grupo de controle. A Tabela 20 a seguir traz os resultados.

Tabela 20 – Efeito médio do tratamento sobre os tratados considerando a base de dados sem os subgrupos da Saúde que auferem menores níveis salariais

	ATT	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	5,882***	0,5114	$< 2,22e-16$	4,819	6,824
Homens	11,150***	0,8283	$< 2,22e-16$	9,522	12,770
Todos	7,757***	0,4794	$< 2,22e-16$	6,817	8,697

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Os resultados encontrados seguem à tendência dos valores do modelo geral, indicando que o subgrupo principal da Saúde não apresenta tanta heterogeneidade quanto o do Ensino no que diz respeito às categorias que os compõem.

Entretanto, vale ressaltar segregação ocupacional ainda existe, mesmo que não seja tão acentuada. Os dados da Tabela 21 mostram que enquanto as mulheres da área da saúde recebem em média um incremento de R\$ 10,48 no salário/hora, os homens recebem aproximadamente R\$ 16,28 por desempenhar uma função dessa categoria. Quando se investiga mais detalhadamente quais profissões constituem esse subgrupo, os resultados apontam mais uma vez para a segregação ocupacional entre os gêneros com as mulheres se concentrando em trabalhos que remuneram menos.

O subgrupo de “Enfermeiros(as) e profissionais dos partos” é o que apresenta o maior diferencial na quantidade de mulheres e homens atuantes. Além de ser uma área socialmente vista com menor prestígio e que auferem menores salários quando comparadas à medicina, por exemplo, é também potencialmente perigosa para as mulheres. Conforme Kingma (2001), as enfermeiras estão mais sujeitas a sofrerem algum tipo de violência do que guardas de penitenciárias. Brophy, Keith e Hurley (2018), por sua vez, afirmam que o problema da violência na área da saúde é recorrente, subnotificado e frequentemente ignorado. Dessa forma,

esse resultado aponta para a segregação ocupacional existente dentro dessa categoria, novamente com as mulheres ocupando atividades menos valorizadas, que pegam menores salários, e cujas características estão associadas aos estereótipos de gênero. Silveira e Siqueira (2021) apontam que os estereótipos de gênero promovem a ideia de que certas ocupações como enfermeiras, cuidadoras, trabalhadoras de serviços administrativos e professoras de educação básica são mais bem executadas por mulheres do que por homens, embora não exista nenhum tipo de justificativa que corrobore esse pensamento.

#### 2.4.3.4 Diferenças entre grupos

Para analisar a questão acerca da segregação ocupacional nas outras áreas, foram testados pareamentos nos quais o grupo de controle passou a ser composto por todos os indivíduos que declararam ser intelectuais e o grupo de tratamento passou a ser aquele no qual os indivíduos declararam ser intelectuais de determinado subgrupo. Ou seja, foram investigados os diferenciais dentro do grupo de “Profissionais da Ciência e Intelectuais”. Para tanto, foi estimado o efeito médio do tratamento sobre os tratados em separado para cada subgrupo (Exatas, Saúde, Ensino, Administração, TI e Sociais). A Tabela 21 a seguir apresenta os valores encontrados.

Tabela 21 – Efeito médio do tratamento sobre os tratados para os subgrupos individualmente

Área	Caso	ATT	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Exatas	Mulheres	5,465***	1,159	$< 2,22e-16$	3,194	7,737
	Homens	3,308**	1,472	0,024572	0,424	6,193
	Todos	3,856***	1,092	0,000415	1,715	5,996
Saúde	Mulheres	10,48***	0,806	$< 2,22e-16$	8,896	12,060
	Homens	16,28***	3,083	1,29E-07	10,240	22,320
	Todos	13,45***	0,931	$< 2,22e-16$	11,620	15,270
Ensino	Mulheres	-22,35***	1,339	$< 2,22e-16$	-24,970	-19,720
	Homens	-24,17***	1,766	$< 2,22e-16$	-27,630	-20,710
	Todos	-31,95***	1,759	$< 2,22e-16$	-35,400	-28,510
Administração	Mulheres	6,933***	1,998	0,000465	3,077	10,910
	Homens	4,161*	2,428	0,086553	-0,598	8,920
	Todos	4,984**	1,588	0,001699	1,871	8,097
TI	Mulheres	12,45***	1,502	$< 2,22e-16$	9,501	15,390
	Homens	-1,491	1,611	0,354680	-4,647	1,666
	Todos	3,538***	0,990	0,000350	1,598	5,477
Sociais	Mulheres	7,755***	0,994	6,21E-15	5,806	9,704
	Homens	-3,368	2,112	0,110760	-7,507	0,771
	Todos	3,459***	1,129	0,002191	1,246	5,672

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022. Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%.

A partir dos resultados reportados na Tabela 21, é possível observar evidências estatísticas que apontam para a presença de heterogeneidade entre todos os subgrupos e isso, por sua vez, corrobora a hipótese de existência de segregação ocupacional entre os gêneros.

Na área de Exatas, enquanto as mulheres recebem um prêmio salarial de cerca de R\$ 5,46 por hora por desempenhar uma função desse grupo, o valor para os homens é de aproximadamente R\$ 3,31. O intervalo de confiança muito mais amplo para os homens do que para as mulheres, com o limite inferior próximo a zero, sugere que os salários dos homens nas exatas podem não ser diferentes de outras ocupações dentro de “Profissionais das Ciências e Intelectuais”.

O subgrupo principal de TI apresentou resultados significativos apenas para as mulheres e foi também a área exibiu o maior prêmio salarial para elas. Ou seja, entre todas as áreas que compõem o grande grupo de intelectuais, fazer parte do subgrupo de TI é o que garante o maior prêmio salarial para as mulheres. Contudo, é preciso ponderar que apenas 1% das mulheres estão concentradas na área de TI e que, dentro da área, elas são apenas 20%.

A área de Saúde apresentou significância estatística e sinal positivo em todos os casos. A magnitude dos valores aponta que, entre todas as áreas, a saúde é a que garante maiores prêmios salariais para os homens. De acordo com a Tabela 20, 40% dos homens da saúde são médicos.

O grupo de Ensino, por sua vez, apresentou retornos negativos, como esperado, de acordo com as análises realizadas no teste de robustez anterior.

As áreas de Administração e Sociais apresentaram resultados similares. Nos dois casos, considerando a separação por gênero, os resultados foram significativos apenas para as mulheres na área de Sociais e mais significativo na área de Administração, em relação aos homens, o que também aponta para a existência da segregação ocupacional.

Nesse sentido, os resultados encontrados indicam que para as mulheres apenas fazer parte do grupo de “Profissionais das Ciências e Intelectuais” não garante um diferencial positivo de salários, além disso, a escolha ocupacional importa. Grande parte das mulheres, ainda que sejam “Profissionais das Ciências e Intelectuais”, possuem o nível de qualificação necessária para tal e tenham tido a oportunidade de escolher a carreira científica ou intelectual, ainda permanecem concentradas em ocupações com menores salários e menos valorizadas pelo mercado de trabalho.

#### 2.4.3.5 Jornada de trabalho: período integral e meio período

O quinto teste estimou os resultados separadamente para os indivíduos que trabalham meio período e para aqueles que trabalham período integral. A Tabela 22 a seguir traz os resultados estimados.

Tabela 22 – Efeito médio do tratamento sobre os tratados separadamente por jornada de trabalho

	Efeito (ATT)	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
<b>Período integral</b>					
Mulheres	6,195***	0,241	< 2,22e-16	5,721	6,668
Homens	9,075***	0,423	< 2,22e-16	8,245	9,905
Todos	7,305***	0,234	< 2,22e-16	6,846	7,764
<b>Meio período</b>					
Mulheres	4,805***	1,181	4,73E-05	2,491	7,120
Homens	13,850***	2,388	6,65E-09	9,168	18,530
Todos	7,895***	1,114	1,34E-12	5,713	10,080

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Conforme os valores reportados na Tabela 22, pode-se notar que os resultados para aqueles que trabalham em período integral são semelhantes aos do modelo original. Quando se analisa a jornada de meio período, entretanto, os valores apontam para uma diferença bastante acentuada entre os gêneros, com as mulheres recebendo aproximadamente 35% do prêmio salarial dos homens. Isso é um problema na medida em que muitas mulheres acabam tendo que recorrer a empregos com jornadas flexíveis de modo a conciliar suas vidas profissionais com seus trabalhos não remunerados de cuidados com o lar e a família (GOLDIN, KATZ, 2011; GOLDIN, 2014).

#### 2.4.3.6 Esfera de atuação: setor público e setor privado

O sexto teste estimou os resultados separadamente para os indivíduos que trabalham no setor público e para aqueles que atuam na esfera privada. Isso porque o Brasil é caracterizado por apresentar um sistema particular de empregos públicos no que diz respeito ao ingresso, progressão de carreira e estrutura salarial. A Tabela 23 a seguir apresenta os resultados.

Tabela 23 – Efeito médio do tratamento sobre os tratados separadamente por esfera de trabalho

	Efeito (ATT)	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
<b>Setor público</b>					
Mulheres	23,52***	1,516	< 2,22e-16	20,55	26,500
Homens	36,78***	2,512	< 2,22e-16	31,850	41,700
Todos	29,23***	1,330	< 2,22e-16	26,62	31,830
<b>Setor privado</b>					
Mulheres	13,77***	0,629	< 2,22e-16	12,53	15,000
Homens	10,71***	0,932	< 2,22e-16	8,881	12,540
Todos	12,6***	0,575	< 2,22e-16	11,47	13,720

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Conforme os resultados reportados na Tabela 23, pode-se observar um prêmio salarial significativamente maior para os indivíduos atuantes em carreiras científicas na esfera pública. Isso pode acontecer devido ao fato de o setor público apresentar, normalmente, carreiras mais atrativas, tanto em termos de remuneração (principalmente iniciais) quanto em termos de estabilidade. Contudo, era esperado que o diferencial entre homens e mulheres atuantes no âmbito público fosse menor, na medida em que não se espera distinção de remuneração entre eles. Dessa forma, os resultados encontrados evidenciam novamente a questão da segregação ocupacional, com as mulheres concentradas em funções que remuneram menos, ainda que atuantes na esfera pública, muito provavelmente concentradas em ocupações relacionadas ao ensino, em especial ao ensino fundamental e médio.

#### 2.4.3.7 Pandemia COVID-19

O sétimo e último teste analisou apenas os anos de 2016 a 2019 para contornar possíveis efeitos da pandemia do Covid-19. A Tabela 24 a seguir apresenta os resultados.

Tabela 24 – Efeito médio do tratamento sobre os tratados considerando a base de dados entre os anos de 2016 e 2019

	ATT	Erro Padrão	Pr ( $> z $ )	2.5%	97.5%
Mulheres	5,116***	0,503	< 2,22e-16	4,130	6,101
Homens	9,828***	1,085	< 2,22e-16	7,701	11,950
Todos	6,375***	0,550	< 2,22e-16	5,297	7,452

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%

Os resultados reportados na Tabela 24 mostram que, comparativamente com o resultado do modelo geral, os anos da pandemia não afetaram negativamente o prêmio salarial dos

indivíduos por fazerem parte do grande grupo de intelectuais. Os homens foram levemente beneficiados na medida em que o intervalo de confiança dos seus resultados aumentou de 7,70 a 11,95 entre 2016 e 2019 para 9,22 a 12,27 com a adição dos anos de 2021 e 2022.

## 2.5 CONSIDERAÇÕES FINAIS

Apesar de estar sendo observada nos últimos anos a elevação da participação feminina, tanto na força de trabalho no geral quanto em ocupações dentro de carreiras científicas, e também a diminuição no diferencial salarial entre os gêneros, a literatura ainda aponta que fatores como, por exemplo, a segregação ocupacional, persiste e favorece a permanência das mesmas em situações desvantajosas no que diz respeito às suas escolhas – ou imposições – profissionais.

Diante desse cenário, o presente ensaio teve como objetivo analisar a diferença por gênero no prêmio salarial por esta empregado em uma função do grande grupo de “Profissionais das Ciências e Intelectuais” da Pesquisa Nacional por Amostra de Domicílios Contínua.

Os resultados encontrados indicam a existência de um prêmio salarial para os indivíduos que declararam ocupar uma função do grupo de intelectuais e que esse valor é diferente entre os gêneros, uma vez que as mulheres recebem um prêmio salarial aproximadamente 55% inferior quando comparada aos homens. Além disso, os valores mais altos que as mulheres tendem a ganhar por serem intelectuais ficam abaixo dos valores mais baixos que os homens tendem a ganhar por estarem na mesma categoria, indicando que as duas distribuições não se sobrepõem. Na tentativa de analisar a robustez desse resultado, e compreendê-lo com mais detalhes, outras sete possibilidades de agrupamento foram investigadas.

Quando apenas os indivíduos com melhores níveis de remuneração são analisados, verifica-se que essa tendência se mantém, porém os diferenciais para as mulheres perdem significância, indicando evidências estatísticas para a existência do efeito teto de vidro.

Em um cenário no qual apenas os indivíduos com ensino superior são considerados, nota-se que os resultados encontrados são similares aos resultados do modelo geral, indicando, dessa forma, que o nível de escolaridade não está afetando o diferencial de salários entre os gêneros.

Quando se retira do grupo de tratamento o subgrupo que pertence à categoria de Ensino (com exceção do Ensino Superior) por auferirem os menores níveis salariais do grupo, os

resultados indicam que o diferencial salarial entre mulheres e homens no grupo de intelectuais é decorrente da segregação ocupacional de mulheres em ocupações com menores salários, seja por características produtivas diferentes, seja porque são ocupações que possuem menor prestígio, como as profissões relacionadas ao ensino. Quando o mesmo raciocínio é aplicado à categoria da Saúde, por também estar relacionada ao cuidado, os resultados encontrados semelhantes aos do modelo geral sugerem que a categoria não é tão heterogênea quanto a do Ensino.

Quando os subgrupos são analisados individualmente, também se encontram resultados que apontam para a presença de heterogeneidade entre todos os subgrupos, corroborando, dessa forma, com a hipótese de existência de segregação ocupacional entre mulheres e homens, com estas se concentrando em ocupações menos valorizadas no mercado de trabalho.

Além disso, quando a jornada de trabalho é analisada, os resultados sugerem que as mulheres são significativamente penalizadas por atuarem em trabalhos mais flexíveis. A mesma penalização aparece para as mulheres atuantes no setor público, em comparação aos homens do mesmo setor, muito provavelmente impulsionada pelas professoras do ensino primário. Por fim, os resultados que desconsideram os anos da pandemia do Covid-19 se mostraram semelhantes aos do modelo geral.

Uma das implicações dos resultados obtidos é que, apesar de importante, o incentivo para as mulheres e meninas a entrarem e perseguirem carreiras científicas não é suficiente para reduzir a desigualdade salarial das mesmas vis-à-vis seus colegas homens. Mesmo que o diferencial salarial em algumas áreas seja pequeno e até mesmo insignificante, existem áreas nas quais as mulheres continuam sendo mais penalizadas. As duas que se mais se destacam no presente estudo são as áreas da Saúde e Educação, com ênfase na educação infantil. Não por coincidência, são esferas muito relacionadas ao cuidado que, por sua vez, está extremamente relacionado aos estereótipos de gênero como exemplos de tarefas “femininas”.

Nesse sentido, vale ressaltar, como apontado por Vaz (2013), que os resultados encontrados não são apenas oriundos da discriminação contra as mulheres. As normas sociais e os estereótipos de gênero têm um papel fundamental na explicação desses fatores, posto que fomentam uma divisão do trabalho que ainda atribui a maior parte das responsabilidades pelo cuidado da família e do lar às mulheres, gerando um conflito cotidiano para que elas consigam conciliar suas vidas pessoais e profissionais.

### 3 ENSAIO 2 – A DEDICAÇÃO VALEU A PENA? UMA ANÁLISE DA RELAÇÃO ENTRE GÊNERO, DESEMPENHO ACADÊMICO E MERCADO DE TRABALHO FORMAL

#### 3.1 INTRODUÇÃO

A atuação feminina no mercado de trabalho apresentou mudanças significativas nos últimos anos e alcançou, recentemente, uma posição curiosa: as mulheres, apesar de passarem mais tempo fora do mercado de trabalho e auferirem menores salários, estão investindo mais em capital humano do que os homens. Em 2010, mais mulheres completavam o ensino superior na maior parte dos países desenvolvidos e em muitos países em desenvolvimento quando comparadas aos homens (BECKER; HUBBARD; MURPHY, 2010). Em 2011, o público feminino ultrapassou o masculino na obtenção de ensino superior completo em 29 dos 34 países da OCDE (AUTOR *et al.*, 2016). De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (2023), os dados do último Censo da Educação Superior evidenciaram que as mulheres eram maioria no ensino superior no Brasil em 2020.

Os efeitos da educação nos resultados no mercado de trabalho podem ser analisados a partir de duas teorias principais: a teoria do capital humano (BECKER, 1964) e a teoria da sinalização (SPENCE, 1973). De acordo com a teoria do capital humano, a educação influencia positivamente os salários uma vez que aumenta a produtividade dos indivíduos. Na medida em que se espera que a universidade ensine aos estudantes habilidades relevantes para desempenhar funções no mercado de trabalho, e que as notas são um reflexo de quão bem os estudantes absorveram e desenvolveram essas habilidades, o desempenho acadêmico pode ser entendido como um reflexo da preparação dos estudantes para o mercado de trabalho<sup>25</sup> (QUADLIN, 2018).

A teoria da sinalização, por sua vez, sugere que os trabalhadores, em especial aqueles pertencentes a grupos discriminados no mercado de trabalho, como as mulheres, encontram na educação uma forma de sinalização, que vem como resposta à expectativa de que suas características produtivas não sejam corretamente avaliadas. A teoria da sinalização pode ser entendida como uma extensão da teoria do capital humano. Contudo, enquanto essa foca no papel da aprendizagem como determinante dos retornos da educação, aquela se concentra nas

---

<sup>25</sup> Se as notas refletem ou não o quanto os estudantes absorvem do conhecimento e das habilidades é tema de discussão. Porém, foge do escopo do presente ensaio.

maneiras pelas quais a educação atua como sinal ou filtro para as diferenças de produtividade que os empregadores não conseguem avaliar diretamente (WEISS, 1995).

Não obstante a existência de estudos que investigam os retornos da educação aos salários, poucos são aqueles que levam em consideração não apenas a educação em si, mas também o desempenho acadêmico dos indivíduos. Nesse ensejo, o presente estudo busca auxiliar no preenchimento dessa lacuna e vai um pouco além, na medida em que considera, também, as disparidades entre os gêneros.

A tendência de as mulheres apresentarem melhores desempenhos acadêmicos em relação aos homens já foi previamente encontrado na literatura, desde níveis iniciais de educação, como jardim de infância (DIPRETE; JENNINGS, 2012; ENTWISLE; ALEXANDER; OLSON, 2007; PERKINS *et al.*, 2004), passando para níveis intermediários do ensino médio e, também, níveis de formação superior acadêmica (DIPRETE; BUCHMANN, 2013).

Não existe, contudo, consenso acerca do papel da performance acadêmica. Alguns estudos sugerem que o melhor desempenho apresentado pelas mulheres em níveis iniciais e intermediários de educação contribuem, entre outras coisas, para que elas sejam mais prováveis de ingressar e concluir um curso de nível superior quando comparadas aos homens (BUCHMANN; DIPRETE, 2006; CARBONARO; ELLISON; COVAY, 2011; RIEGLE-CRUMB, 2010). Em relação ao mercado de trabalho, na tentativa de contornar as percepções e os julgamentos discriminatórios de que as mulheres são menos capazes, menos preparadas e menos comprometidas em seus empregos (RIDGEWAY, 2011; KELLY *et al.*, 2010), muitas acabam tentando neutralizar esses fatores evidenciando o bom desempenho acadêmico em seus currículos (QUADLIN, 2018). Quando esse é o caso, a literatura sugere dois possíveis resultados. Por um lado, as mulheres com altas performances são altamente demandadas por empregos, em especial naqueles comprometidos com a diversidade de gênero no ambiente de trabalho (CORRELL *et al.*, 2020). Por outro lado, o alto nível de desempenho das mulheres também pode ser interpretado de forma negativa como ganância e frieza (CUDDY; FISKE; GLICK, 2008; EAGLY; CARLI, 2007).

Dessa forma, estudos que investiguem a relação entre desempenho acadêmico e resultados no mercado de trabalho são importantes, especialmente para as mulheres, tendo em vista todas as desvantagens que enfrentam nesse aspecto: dupla jornada de trabalho, segregação ocupacional, estereótipos de gênero, assédio, dentre tantos outros. De acordo com dados do IBGE (2023), em 2022 as mulheres ganhavam apenas 78% do salário dos homens no Brasil, mesmo sendo maioria com ensino superior completo.

Nesse contexto, o presente ensaio propõe avaliar se, durante o período de formação universitária, mulheres e homens possuem comportamentos distintos no que diz respeito ao desempenho acadêmico, e se esse comportamento se reflete ou não nos resultados no mercado de trabalho.

Para tanto, foi construída uma base de dados que integra informações sobre os salários, as ocupações, os setores de atuação e o tempo de experiência da base de dados da Relação Anual de Informações Sociais Identificada (RAIS Identificada), com informações sobre alunos egressos da Universidade Federal do Paraná (UFPR) como gênero, etnia, nacionalidade, estado civil, curso, turno, data de início e conclusão do curso, forma de evasão, média dos créditos cursados, quantidade de faltas e situação (aprovação, reprovação por nota, reprovação por frequência). O período da análise compreendeu os anos entre 2003 e 2021 e a abordagem empírica utilizou o método de Mínimos Quadrados Ordinários (MQO).

O presente estudo avança em relação à literatura existente em duas frentes principais. A primeira delas diz respeito à utilização de uma base de dados única, com acompanhamento dos mesmos indivíduos em duas fases distintas, formação acadêmica e mercado de trabalho. A segunda, por sua vez, remete à investigação da relação entre o desempenho acadêmico, os resultados no mercado de trabalho e as disparidades de gênero em um contexto de um país em desenvolvimento, com elevado *gap* salarial e crescente aumento da qualificação das mulheres.

O ensaio está organizado em outras quatro seções, além dessa introdução. A seção dois apresenta a revisão de literatura que norteia as análises. A seção três traz a metodologia e o tratamento de dados. A seção quatro, por sua vez, contém os resultados e a análise de duas robustezes. Por fim, a seção cinco apresenta as considerações finais do ensaio.

## **3.2 REVISÃO DE LITERATURA**

A revisão de literatura a seguir foca em duas frentes principais. A primeira diz respeito às teorias do capital humano e da sinalização, na medida em que relacionam a educação à produtividade e aos resultados no mercado de trabalho. A segunda, por sua vez, traz informações acerca de estudos que analisaram não somente a influência da educação, mas também os efeitos do desempenho acadêmico dos indivíduos no mercado de trabalho.

### 3.2.1 A teoria do capital humano e da sinalização e o desempenho acadêmico

De acordo com a teoria do capital humano, que tem como um de seus maiores expoentes Gary Becker (1964), o capital humano pode ser entendido não apenas como uma medida da educação dos indivíduos, mas também como uma medida das habilidades desenvolvidas ao longo do tempo. Quando um indivíduo investe em educação ou acumula experiência desempenhando determinada função, isso se traduz em aumento de capital humano e, assim como o capital, contribui para a produção na medida em que promove o aumento da produtividade, e, dessa forma, do retorno sobre o investimento (HADLEY, 2019). Em suma, “[w]hat the theory of human capital says is that people... can develop themselves in various ways” (BECKER; EWALD; HARCOURT 2012, p. 18).

O trabalho de Spence (1973), por sua vez, foi pioneiro ao trazer à tona a ideia de que o investimento em educação pode ser entendido como um sinal de produtividade para os empregadores. Apesar de alguns investimentos em capital humano não serem observáveis, o investimento na educação, por outro lado, é observável e pode sinalizar produtividade. Segundo Spence (1973), ainda que a educação não aumente a produtividade dos indivíduos, se apenas indivíduos mais produtivos conseguem adquirir esse sinal, então ela passa a distingui-los dos demais, e pode ser utilizada como um *sinal* de produtividade. A magnitude da sinalização e sua importância para que o indivíduo se diferencie dos demais é tão maior quanto menos confiável for a observação direta da produtividade por parte dos empregadores (LANG; MANOVE, 2011). Nesse sentido, a investigação acerca da sinalização se torna ainda mais relevante quando aplicada a grupos minoritários e/ou que sofrem algum tipo de discriminação.

O arcabouço teórico da economia sugere dois modelos básicos para analisar a discriminação no mercado de trabalho: o modelo de preferências por discriminação (BECKER, 1971) e o modelo de discriminação estatística desenvolvido inicialmente por Edmund Phelps (1972) e Kenneth Arrow (1973). O primeiro inclui na função de utilidade dos agentes econômicos (i.e., trabalhadores, empregadores, consumidores) um desejo de evitar membros de certos grupos sociais. No segundo modelo, por sua vez, os empregadores analisam as características médias de um grupo para prever os atributos individuais dos trabalhadores do grupo como forma de reduzir a assimetria de informação existente no mercado de trabalho. Assim, mesmo que os empregadores não tenham a priori nenhum tipo de preconceito, na falta de melhores informações, avaliariam negativamente os indivíduos que pertencem a grupos com características produtivas médias piores (SCHWAB, 1986).

Estudos recentes, como Arcidiacono, Bayer e Hizmo (2010), Lang e Manove (2011) e Nielsson e Steingrimsdottir (2015) Lang e Manove (2011), sugerem que os indivíduos que sofrem algum tipo de discriminação, como negros e mulheres, por exemplo, investem mais em capital humano de forma a compensar o fato de pertencerem à grupos discriminados e sinalizarem sua real produtividade. De acordo com Lang e Manove (2011), existem evidências de que os empregadores consideram mais difícil realizar a avaliação de candidatos negros vis-à-vis candidatos brancos. Dessa forma, quando os investimentos observáveis em capital humano estão disponíveis, é provável que a discriminação estatística irá fazer com que pessoas negras invistam mais em si mesmas do que pessoas brancas<sup>26</sup>.

A questão sobre raça e sinalização também é analisada por Arcidiacono, Bayer e Hizmo (2010). De acordo com os autores, existem evidências acerca da existência de discriminação estatística no mercado de trabalho contra negros recém-formados no ensino médio, ou seja, pessoas negras têm incentivos maiores a adentrar no ensino superior, e concluí-lo, de modo a sinalizar suas habilidades. Além disso, segundo o estudo, o papel desempenhado pela educação na determinação de salários vai além da sinalização de produtividade. Os autores sugerem que o retorno da educação é maior para pessoas recém-graduadas no ensino superior imediatamente após a entrada no mercado de trabalho e não se altera à medida que ganham mais experiência no emprego. O inverso ocorre para as pessoas recém-formadas no ensino médio: o retorno da educação é baixo no início e aumenta com o tempo e a experiência.

Por sua vez, segundo Blair e Chung (2020), em empregos que requerem licenciamento profissional (médicos, engenheiros, arquitetos, dentistas, por exemplo), o peso dado por parte dos empregadores às características dos candidatos a determinado posto de emprego, como raça e gênero, é menor. Como resultado, negros e mulheres, por exemplo, que possuem licença ocupacional enfrentam menores níveis de *gap* salarial vis-à-vis seus pares não licenciados.

Enquanto Arcidiacono, Bayer e Hizmo (2010) e Lang e Manove (2011) focam na questão racial, Nielsson e Steingrimsdottir (2015) investigam se a educação é mais importante para sinalizar as habilidades entre as mulheres do que entre os homens. A partir de um modelo de discriminação racial baseado em Lang e Manove (2011), os autores encontram que as mulheres, em média, obtêm 0,5 a 0,7 anos extras de educação quando comparadas aos homens com o mesmo nível de habilidade. Além disso, os autores sugerem que, dado o mesmo nível de

---

<sup>26</sup> Vale ressaltar, que essa constatação não está em consonância com as proposições iniciais dos modelos de discriminação estatística (i.e. ARROW, 1973), segundo os quais as pessoas discriminadas investiriam menos em si mesmas do que indivíduos não discriminados, uma vez que recebem menos retorno desse investimento (LANG; MANOVE, 2011).

habilidade entre mulheres e homens, estas ganham salários mais baixos e que esse *gap* é maximizado para níveis intermediários de educação.

Não obstante a existência de estudos que investigam a relação entre a educação e os resultados no mercado de trabalho, poucos ainda são aqueles que buscam entender os efeitos não apenas da educação, mas também do desempenho acadêmico dos indivíduos, com foco nas diferenças entre os gêneros.

O estudo de Buchmann e DiPrete (2006), por exemplo, sugere que as diferenças de gênero em relação ao *background* familiar e ao desempenho acadêmico influenciaram a crescente vantagem observada para as mulheres no que diz respeito à conclusão de cursos superiores nos Estados Unidos. Um dos motivos que explicam essa tendência é a desvantagem observada para filhos homens de famílias cujo pai possui baixo nível de escolaridade ou é ausente. Em outro estudo, utilizando dados dos Estados Unidos entre 1964 e 2002, DiPrete e Buchmann (2006) sugerem que os retornos da educação superior para as mulheres, no que diz respeito ao aprimoramento do padrão de vida e à ampliação da segurança contra pobreza, aumentou mais rapidamente quando comparada aos homens, indicando que a vantagem feminina na conclusão de um curso universitário pode estar relacionada, pelo menos em parte, com as mudanças nos retornos da educação entre os gêneros.

A partir de um estudo experimental baseado no envio de currículos profissionais para a candidatura a vagas de empregos e as respectivas respostas a eles, Quadlin (2018) analisa o desempenho acadêmico de mulheres e homens no ensino superior e sua influência, em termos de remuneração, posteriormente no mercado de trabalho. A autora encontra que quando as mulheres apresentam uma performance acadêmica moderada são beneficiadas no mercado de trabalho. Entretanto, se a performance feminina foi alta, as mulheres são prejudicadas e esse resultado é potencializado quando a área de formação é em matemática. Além disso, a autora verifica que os empregadores geralmente buscam qualidades como comprometimento e competência entre os candidatos homens, mas privilegiam fatores como a simpatia nas candidatas mulheres. Esse padrão acaba por beneficiar as mulheres com desempenho acadêmico moderado, geralmente descritas como sociáveis e extrovertidas, mas prejudicam as que apresentam desempenho elevado, cuja personalidade é vista com mais ceticismo. Os resultados para os homens são aproximadamente iguais, independentemente do nível dos seus desempenhos. O estudo sugere que os resultados encontrados evidenciam a existência dos estereótipos de gênero, que penalizam as mulheres com melhores performances no ensino superior, criando, dessa forma, retornos desiguais no mercado de trabalho.

Por sua vez, o trabalho de Castagnetti e Rosti (2009) analisa as influências do desempenho acadêmico no mercado de trabalho para italianos no ano de 2004. A partir de um modelo OLS, os resultados encontrados corroboram o fato de as mulheres apresentarem melhores desempenhos acadêmicos e piores resultados no mercado de trabalho em comparação aos homens. Além disso, ao decompor as diferenças entre mulheres e homens no que diz respeito ao desempenho acadêmico em fatores observáveis e não observáveis, as autoras encontram que esse explica aproximadamente 56% da diferença. Partindo do princípio de que as habilidades são igualmente distribuídas entre os gêneros, as autoras sugerem, então, que a melhor performance acadêmica das mulheres se deve ao maior esforço individual exercido por elas.

Já o estudo de Dekhtyar *et al.* (2018) acompanha indivíduos da Suécia do ensino superior ao mercado de trabalho com o intuito de investigar se as diferenças nas habilidades durante o curso superior influenciam as escolhas profissionais futuras de forma distinta entre mulheres e homens. Os autores encontram, entre outras coisas, que as mulheres apresentam mais habilidades em áreas verbais e de linguagem enquanto os homens são mais habilidosos em áreas técnicas e numéricas. Além disso, quando os indivíduos saíram do ensino superior e foram para o mercado de trabalho, tanto as mulheres quanto os homens buscaram carreiras que se relacionavam com suas habilidades mais proeminentes durante o curso superior. Entretanto, grande parte das mulheres que apresentava habilidades em áreas técnicas e numéricas evitou empregos desse ramo no mercado de trabalho. Diante disso, os autores concluem que, muito embora a segregação de gênero tenha sido associada às escolhas individuais a partir de pontos fortes acadêmicos, as discrepâncias nos resultados acerca das carreiras profissionais entre mulheres e homens podem apenas em parte ser atribuídas às diferenças no desempenho acadêmico.

Em suma, os trabalhos indicam que a qualificação e o desempenho acadêmico de mulheres e homens possuem efeitos distintos no mercado de trabalho, que tendem a se refletir em diferenciais de salários persistentes. A interligação entre esses fatores será o foco de análise do presente ensaio, ou seja, buscar-se-á verificar se existe uma diferença de desempenho acadêmico por gênero entre os estudantes egressos da UFPR e se essa eventual diferença tem impactos posteriores no *gap* salarial entre mulheres e homens no mercado de trabalho.

### 3.3 METODOLOGIA E TRATAMENTO DE DADOS

Apesar de existirem estudos que exploram os impactos da educação nos salários, são poucos aqueles que consideram não apenas a educação em si, mas também o desempenho acadêmico dos indivíduos, especialmente ao analisar as disparidades de gênero.

Diante disso, o presente estudo busca avaliar se a performance de estudantes durante seus cursos de graduação e como essa performance se traduz em resultados no mercado de trabalho para mulheres e homens. Na medida em que se espera que a universidade ensine aos estudantes habilidades relevantes para desempenhar funções no mercado de trabalho, e que as notas são um reflexo de quão bem os estudantes absorveram e desenvolveram essas habilidades, o desempenho acadêmico pode ser entendido como um reflexo da preparação dos estudantes para o mercado de trabalho (QUADLIN, 2018).

As melhores notas durante a formação acadêmica podem refletir dois aspectos: habilidades ou esforço. Indivíduos mais habilidosos terão maiores notas independente de se esforçarem ou não. Ainda assim, indivíduos menos habilidosos, mas com maior nível esforço poderão obter boas notas. Em ambos os casos, seja por maiores habilidade ou por maior esforço, ao acompanhar os mesmos indivíduos espera-se que essa característica permaneça no mercado de trabalho e se reflita em melhores salários. Não obstante, espera-se que não exista um viés de gênero do indivíduo nesse efeito. Dessa forma, tal qual Castagnetti e Rosti (2009), o presente ensaio considera o desempenho acadêmico como *proxy* para habilidade dos indivíduos.

#### 3.3.1 Bases de dados e recorte amostral

A condução das análises parte da integração entre os dados da Relação Anual de Informações Sociais (RAIS) identificada, disponibilizados pelo Ministério da Fazenda e de dados administrativos da Universidade Federal do Paraná, concedidos Agência de Tecnologia da Informação e Comunicação (AGTIC). Nos dados da RAIS foi possível identificar variáveis como o salário, a ocupação, o setor de atuação, e o tempo de experiência no mesmo trabalho para cada egresso da universidade. Quanto aos dados da UFPR, foram utilizadas as seguintes informações: gênero, etnia, nacionalidade, estado civil<sup>27</sup>, curso, turno<sup>28</sup>, data de início e

---

<sup>27</sup> A informação sobre o estado civil é única e se refere ao momento de ingresso no curso.

<sup>28</sup> A informação sobre o turno foi disponibilizada apenas por curso no momento do ingresso e não por matéria efetivamente cursada.

conclusão do curso, forma de evasão, média dos créditos cursados, quantidade de faltas e situação (aprovação, reprovação por nota, reprovação por frequência). A integração entre as bases de dados foi possível por meio do CPF de cada discente egresso<sup>29</sup>.

### 3.3.1.1 O ensino superior: dados administrativos da UFPR

O processo de organização da base de dados da UFPR contou com as seguintes etapas. Inicialmente, dentre o total de informações disponibilizadas, foram considerados apenas os dados de alunos do nível de graduação da UFPR (cursos de 15 semanas, cursos de 18 semanas, cursos de 20 semanas, cursos anuais e cursos da graduação do campus do litoral) que cursaram disciplinas entre 1986 e 2020. Para contabilizar as notas de cada discente, foram consideradas apenas disciplinas para as quais estavam disponíveis a situação de aprovação, reprovação por nota ou reprovação por frequência. Foram desconsideradas todas as matérias nas quais os alunos foram aprovados e a média constava como zero (a maior parte desses casos se concentra nas matérias de Condicionamento Físico, Iniciação Esportiva, Desporto de Livre Escolha, Prática Desportiva e Educação Física). Ademais, nos casos em que a informação sobre data de fim ou a data de conclusão do curso estava faltando, foi considerada a data da última matéria cursada pelo discente. Após essas considerações, a base de dados resultante contou com 95.255 observações. A Tabela 20 a seguir traz as estatísticas descritivas por gênero e o Gráfico 7 apresenta informações sobre as médias das notas de cada gênero.

Tabela 25 – Estatísticas descritivas e Teste t – base de dados UFPR

Variáveis	Mulheres		Homens		Teste t
	Média	Desvio-padrão	Média	Desvio-padrão	
<b>Raça</b>					
Brancos(as) e amarelos(as)	0,2541	0,4353	0,2624	0,4400	2,9476***
Pretos(as), pardos(as) e indígenas	0,0161	0,1258	0,0207	0,1425	5,3506***
Não declarada	0,7298	0,4441	0,7168	0,4506	-4,4887***
<b>Nacionalidade</b>					
Brasileiros(as)	0,9923	0,0875	0,9895	0,1021	-4,5823***
<b>Estado civil</b>					
Solteiros(as)	0,6339	0,4818	0,6495	0,4771	5,0194***
Casados(as)	0,0851	0,2790	0,0679	0,2516	-9,9318***

(continua)

<sup>29</sup> Os dados identificados só foram disponibilizados após o projeto dessa pesquisa ter sido submetido a um comitê de ética, a fim de garantir a segurança e a proteção das informações particulares dos indivíduos. Número do parecer: 5.867.174.

(continuação)

<u>Variáveis</u>	<b>Mulheres</b>		<b>Homens</b>		<b>Teste t</b>
	<b>Média</b>	<b>Desvio-padrão</b>	<b>Média</b>	<b>Desvio-padrão</b>	
Outros*	0,0457	0,2088	0,0490	0,2159	2,4221**
Não declarada	0,2354	0,4242	0,2336	0,4231	-0,6537
<b>Idade</b>					
Nascimento 1930-1959	0,0323	0,1768	0,0277	0,1642	-4,1519***
Nascimento 1960-1969	0,1635	0,3698	0,1781	0,3826	5,9868***
Nascimento 1970-1979	0,4019	0,4903	0,3997	0,4898	-0,6671
Nascimento 1980-1989	0,3978	0,4895	0,3900	0,4877	-2,4855**
Nascimento 1990-1999	0,0044	0,0661	0,0042	0,0649	-0,3651
<b>Turno</b>					
Integral	0,4709	0,4992	0,5183	0,4997	14,6580***
Matutino	0,2492	0,4326	0,1789	0,3833	-26,4540***
Noturno	0,2074	0,4055	0,2537	0,4352	16,9980***
Vespertino/Diurno	0,0725	0,2593	0,0490	0,2159	-15,1150***
<b>Áreas**</b>					
Agrárias	0,0509	0,2198	0,0691	0,2536	11,8580***
Biológicas	0,0431	0,2030	0,0219	0,1465	-18,2960***
Engenharias	0,0663	0,2487	0,1950	0,3962	60,5500***
Exatas e Terra	0,1294	0,3356	0,2387	0,4263	44,1410***
Humanas	0,1961	0,3971	0,0746	0,2627	-55,2050***
Linguística, Letras e Artes	0,0682	0,2521	0,0340	0,1811	-23,8950***
Saúde	0,2138	0,4100	0,1202	0,3251	-38,8260***
Sociais Aplicadas	0,2323	0,4223	0,2466	0,4311	5,1937***
<b>Situação acadêmica</b>					
Média	63,5937	25,1890	52,4131	27,0040	-66,1130***
Faltas	15,4379	22,1737	22,6250	27,4383	44,6200***
Aprovação	0,7828	0,3202	0,6550	0,3611	-57,9090***
Reprovação por nota	0,1358	0,2449	0,2104	0,2819	43,7110***
Reprovação por frequência	0,0814	0,2106	0,1346	0,2598	34,8380***
Forma de evasão formatura	0,6723	0,4694	0,5547	0,4970	-0,1237***
Tempo graduação	4,4769	2,0140	4,6682	2,3547	13,5070***
<b>Total de observações</b>	<b>45.609</b>		<b>49.646</b>		

Fonte: Elaboração própria a partir de dados da UFPR.

\* Outros: separado(a), divorciado(a), desquitado(a), viúvo(a), outros.

\*\* Os cursos foram categorizados em Áreas conforme a Tabela de Áreas do Conhecimento da CAPES. Para mais informações, ver Tabela A6 no Apêndice.

A partir dos dados reportados na Tabela 25, pode-se observar uma tendência semelhante para mulheres e homens no que diz respeito à raça, nacionalidade, estado civil e idade: são indivíduos majoritariamente brasileiros, solteiros, que não declararam a respectiva raça ou etnia e que nasceram nas décadas de 1970 e 1980.

Em relação aos turnos, 47% das mulheres e 52% dos homens estão em cursos de turno integral. Cursos de turno noturno concentram 25% dos homens e 21% das mulheres enquanto cursos de turno matutino concentram 25% das mulheres e 18% dos homens. Os cursos de turno

vespertino/diurno, por sua vez, são os que reúnem menos indivíduos: 7% das mulheres e 5% dos homens. Uma vez que a informação acerca do turno foi disponibilizada apenas por curso e não por matéria efetivamente cursada, não foi possível analisar possíveis efeitos de mudança de turno e também não é possível garantir que os indivíduos de fato realizaram todo ou a maior parte do curso no turno informado.

No que diz respeito às áreas, Engenharias, Exatas e Terra, Humanas e Saúde são aquelas que apresentam os maiores diferenciais entre a participação de mulheres e homens. Enquanto as Engenharias (6,6% mulheres e 19,5% homens) e Exatas e Terra (13% mulheres e 24% homens) são predominantemente masculinas, as áreas de Humanas (19,6% mulheres e 7,4% homens) e Saúde (21,3% mulheres e 12,2% homens) são caracterizadas por maior presença feminina. Esses padrões obedecem a tendência de sub-representação de mulheres em áreas *STEM* (ciência, tecnologia, engenharia e matemática) e sua predominância em campos relacionados às humanidades, realidade amplamente influenciada pelos estereótipos de gênero (BARONE; ASSIRELLI, 2018; SAHOO; KLASSEN, 2018; GONZÁLEZ-PÉREZ; CABO; SÁINZ, 2020; BUCKLEY; FARRELL; TYNDALL, 2021).

Em relação ao desempenho acadêmico, notam-se resultados melhores para as mulheres vis-à-vis seus colegas homens. Enquanto elas apresentam 63,6 pontos em média para as notas das matérias cursadas, o valor dos homens é de 52,4. Os dados também mostram que os homens faltam, em média, mais do que as mulheres: 22,6 faltas para eles em comparação com 15,4 faltas para elas. Além disso, a parcela feminina apresenta melhores resultados no que diz respeito aos níveis de aprovação, reprovação por nota ou reprovação por frequência – 78% das mulheres são aprovadas nas matérias e 65% dos homens, 13,6% das mulheres são reprovadas por notas e 21% dos homens e 8% das mulheres reprovam por frequência contra 13% dos homens.

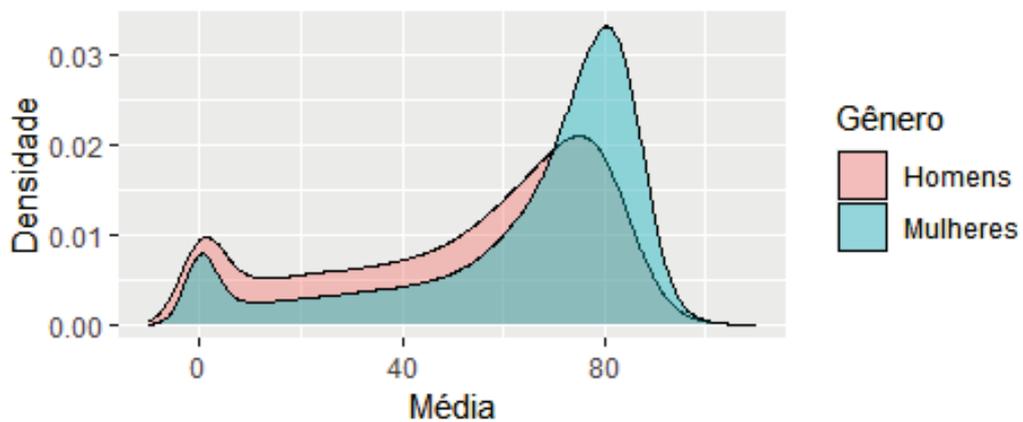
A tendência de as mulheres apresentarem melhores desempenhos acadêmicos em relação aos homens já foi previamente encontrado na literatura, desde níveis iniciais de educação, como jardim de infância (DIPRET; JENNINGS, 2012; ENTWISLE; ALEXANDER; OLSON, 2007; PERKINS *et al.*, 2004), passando para níveis intermediários do ensino médio e, também, níveis de formação superior acadêmica (DIPRETE; BUCHMANN, 2013; DOWNEY; VOGT YUAN, 2005; DUMAIS, 2002).

Nota-se também que o tempo de duração dos cursos (calculada pela diferença entre o ano mínimo e o ano máximo das matérias cursadas) entre mulheres e homens é semelhante. Contudo, quando consideramos apenas os indivíduos que apresentam “Formatura” como forma de evasão da UFPR, os resultados são mais distintos: 67% das mulheres se formaram em

comparação a 55% dos homens. É importante ressaltar, contudo, que parte da amostra não tem informações acerca da forma de evasão.

A partir das informações apresentadas no Gráfico 7, pode-se observar que as mulheres estão mais concentradas do que os homens na parcela da distribuição que apresenta as maiores notas.

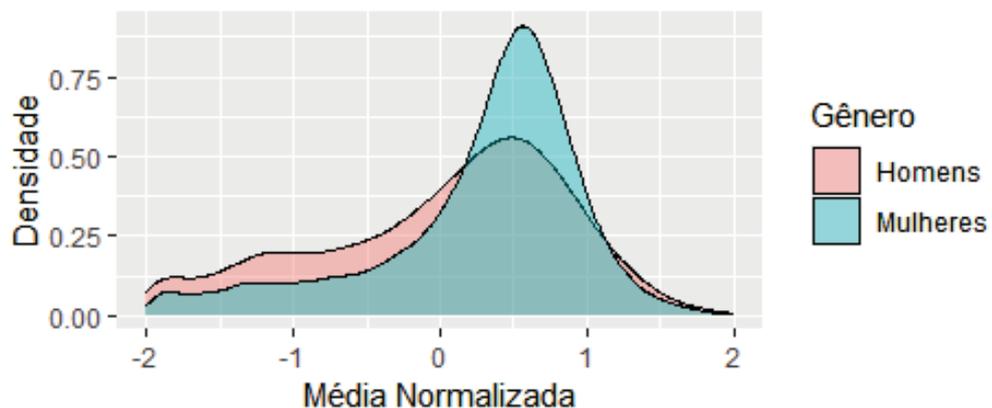
Gráfico 7 – Densidade da média das notas dos alunos por gênero



Fonte: Elaboração própria a partir de dados da UFPR.

Entretanto, como cada curso possuem especificidades em relação à atribuição das notas e à avaliação do desempenho acadêmico, buscou-se trazer informações adicionais a respeito das tendências dos valores das notas a partir de seus valores normalizados por curso, conforme o Gráfico 8.

Gráfico 8 – Densidade da média das notas normalizadas por curso para mulheres e homens

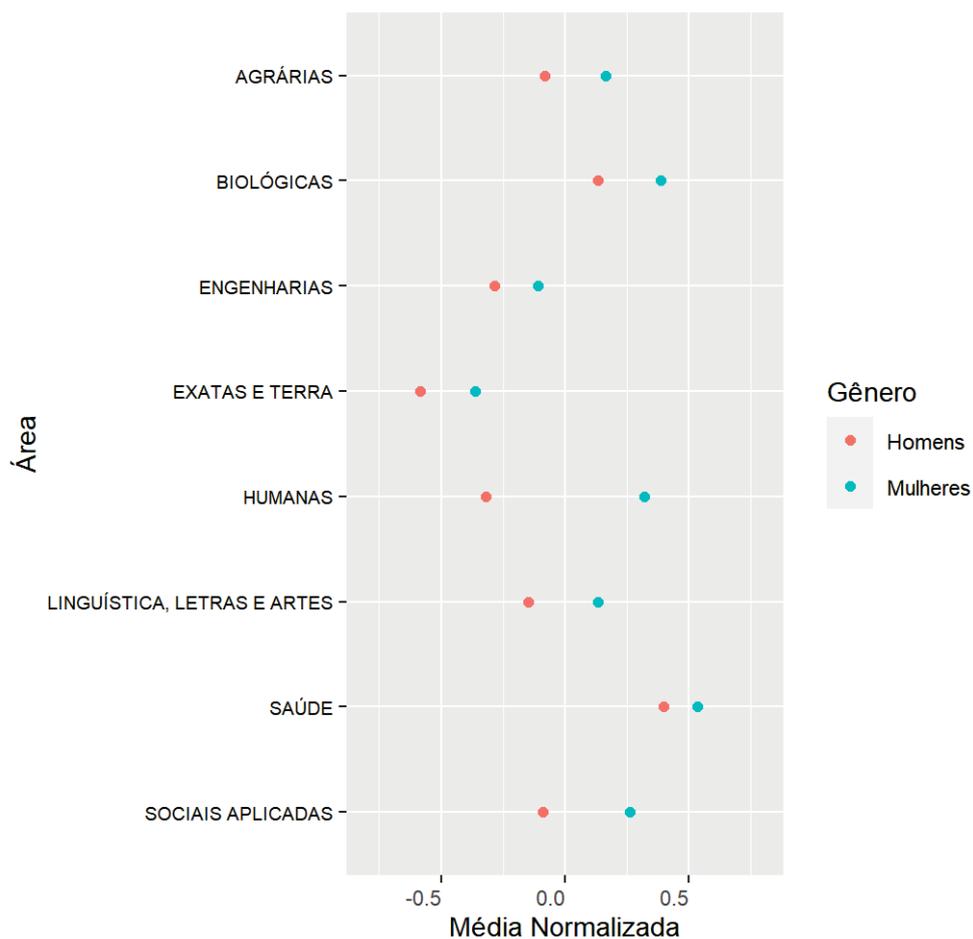


Fonte: Elaboração própria a partir de dados da UFPR

O Gráfico 8 apresenta a densidade das médias das notas de mulheres e homens normalizadas de acordo com cada curso. Pode-se observar que existe uma maior concentração de notas acima da média para as mulheres em comparação aos homens.

O Gráfico 9, por sua vez, analisa separadamente para cada uma das oito áreas mais abrangentes (Agrárias; Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde; Sociais Aplicadas).

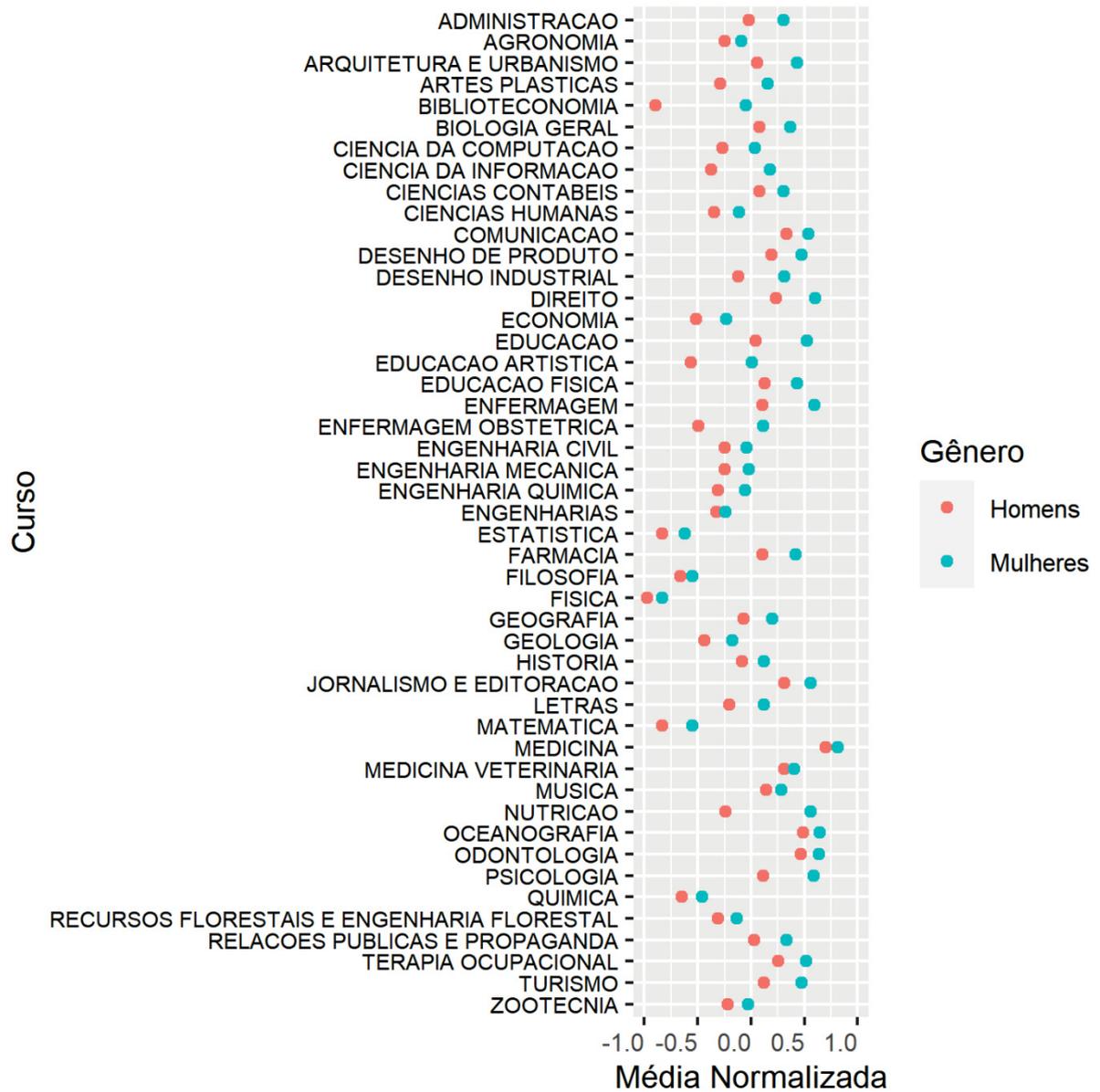
Gráfico 9 – Posição da média normalizada de cada gênero por área mais abrangente



Fonte: Elaboração própria a partir de dados da UFPR.

Primeiramente, pode-se observar que em nenhuma área os homens possuem médias maiores do que as mulheres e que a diferença entre os gêneros é maximizada na área de Humanas seguida pela área das Sociais Aplicadas. A área da Saúde, por sua vez, é a que apresenta o menor diferencial seguida pelas Engenharias. O Gráfico 10 a seguir traz essas informações para áreas menos abrangentes (ou cursos).

Gráfico 10 – Posição da média normalizada de cada gênero por curso



Fonte: Elaboração própria a partir de dados da UFPR.

A partir das informações apresentadas pelo Gráfico 10, pode-se notar, novamente, que mesmo considerando áreas menos abrangentes, as mulheres continuam apresentando notas mais elevadas do que os homens em todos os casos. Os cursos de Biblioteconomia e Nutrição apresentam os maiores diferenciais de notas entre os gêneros enquanto os cursos de Engenharias, Medicina, Medicina Veterinária e Filosofia exibem os menores diferenciais.

### 3.3.1.2 O mercado de trabalho formal: dados da UFPR e da RAIS

Posteriormente à análise do perfil dos alunos egressos e de seus desempenhos acadêmicos, foi realizada a busca desses mesmos indivíduos na base de dados da RAIS identificada com o objetivo de investigar suas informações salariais. Mais especificamente, buscou-se entender se o melhor desempenho acadêmico apresentado pelas mulheres está ou não se traduzindo em melhores remunerações no mercado de trabalho posteriormente às suas conclusões do ensino superior. Das 95.255 observações da base de dados da UFPR, foi possível encontrar 31.553 indivíduos na RAIS entre os anos de 2003 e 2021. A Tabela 21 a seguir contém as estatísticas descritivas da nova amostra.

Tabela 26 – Estatísticas descritivas e Teste t - base de dados UFPR e RAIS

<u>Variáveis</u>	<b>Mulheres</b>		<b>Homens</b>		<b>Teste t</b>
	<b>Média</b>	<b>Desvio-padrão</b>	<b>Média</b>	<b>Desvio-padrão</b>	
<b>Ensino superior - UFPR</b>					
<b>Raça</b>					
Branco(as) e amarelo(as)	0,3784	0,4850	0,3901	0,4878	2,1371**
Pretos(as), pardos(as) e indígenas	0,0450	0,2072	0,0593	0,2362	5,7465***
Não declarada	0,5766	0,4941	0,5505	0,4975	-4,6691***
<b>Nacionalidade</b>					
Brasileiros(as)	0,9946	0,0735	0,9881	0,1086	-6,2422***
<b>Estado civil</b>					
Solteiros(as)	0,6948	0,4605	0,6910	0,4621	-0,7339
Casados(as)	0,0814	0,2734	0,0779	0,2680	-1,1443
Outros	0,0435	0,2041	0,0570	0,2319	5,4894***
Não declarada	0,1803	0,3844	0,1741	0,3792	-1,4380
<b>Idade</b>					
Nascimento 1930-1959	0,0230	0,1500	0,0219	0,1463	-0,6985
Nascimento 1960-1969	0,1339	0,3405	0,1340	0,3407	0,0355
Nascimento 1970-1979	0,2734	0,4457	0,2907	0,4541	3,4188***
Nascimento 1980-1989	0,5582	0,4966	0,5436	0,4981	-2,6157***
Nascimento 1990-1999	0,0115	0,1067	0,0099	0,0989	-1,4136
<b>Turno</b>					
Integral	0,4782	0,4995	0,4723	0,4992	-1,0552
Matutino	0,2392	0,4266	0,1940	0,3954	-9,7469***
Noturno	0,2223	0,4158	0,2735	0,4458	10,5460***
Vespertino/Diurno	0,0603	0,2380	0,0603	0,2380	-0,0151
<b>Áreas</b>					
Agrárias	0,0564	0,2307	0,0564	0,2307	3,3350***

(continua)

(continuação)

<b>Variáveis</b>	<b>Mulheres</b>		<b>Homens</b>		<b>Teste t</b>
	<b>Média</b>	<b>Desvio-padrão</b>	<b>Média</b>	<b>Desvio-padrão</b>	
Biológicas	0,0536	0,2253	0,0342	0,1818	-10,7910***
Engenharias	0,0812	0,2731	0,1613	0,3678	3,0886***
Exatas e Terra	0,1635	0,3698	0,2508	0,4335	25,8900***
Humanas	0,1497	0,3568	0,0991	0,2989	-20,873***
Linguística, Letras e Artes	0,0759	0,2648	0,0478	0,2133	-13,967***
Saúde	0,2070	0,4052	0,1162	0,3204	-26,813***
Sociais Aplicadas	0,2127	0,4092	0,2342	0,4235	39,595***
<b>Situação acadêmica</b>					
Média das notas	62,6251	25,5803	54,5770	26,4850	-27,458***
Faltas	14,0585	20,5039	17,2302	22,2657	13,17***
Aprovação	0,7637	0,3253	0,6710	0,3497	-24,383***
Reprovação por nota	0,1307	0,2279	0,1731	0,2426	16,006***
Reprovação por frequência	0,1057	0,2265	0,1559	0,2649	18,138***
Forma de evasão formatura	0,6561	0,4750	0,5630	0,4960	-17,029****
Tempo graduação	4,6365	2,1176	4,8234	2,4052	73,324***
<b>Mercado de trabalho - RAIS</b>					
<b>Raça</b>					
Branco(as) e amarelo(as)	0,7250	0,4465	0,7031	0,4569	-4,3206***
Pretos(as), pardos(as) e indígenas	0,0725	0,2593	0,0802	0,2716	2,5890***
Não declarada/Indefinida	0,2025	0,4019	0,2167	0,4120	31,072***
<b>Região de trabalho</b>					
Sul	0,8862	0,3176	0,8579	0,3491	-7,5288***
Sudeste	0,0813	0,2732	0,0843	0,2778	0,9793
Centro-Oeste	0,0235	0,1514	0,0433	0,2035	9,8364***
Nordeste	0,0051	0,0713	0,0075	0,0864	2,7022***
Norte	0,0039	0,0627	0,0070	0,0831	3,6418***
<b>Salário e jornada</b>					
Salário	6.557,77	8.776,79	8.737,49	12.481,20	18,002***
Meses empregado(a)	40,7337	50,9431	47,0096	58,2718	10,196***
Média de jornada de horas de trabalho	35,3565	8,5049	36,7246	7,9343	14,7600***
Jornada até 20 horas	0,2486	0,4322	0,2033	0,4024	-9,6277***
Jornada de 21 a 30 horas	0,1955	0,3966	0,1610	0,3675	-8,0061***
Jornada de 31 a 44 horas	0,5560	0,4969	0,6358	0,4812	14,484***
Trabalhava antes de entrar na graduação	0,0623	0,2417	0,0810	0,2729	6,4511***
Trabalhava antes de cursar a última matéria	0,1887	0,3913	0,2366	0,4250	10,422***
<b>Total de observações</b>	<b>15.612</b>		<b>16.097</b>		

Fonte: Elaboração própria a partir de dados da UFPR e RAIS.

A partir dos dados reportados na Tabela 21, pode-se observar que as estatísticas descritivas da nova amostra seguem as tendências da base original. Entretanto, as variáveis sobre a raça sofreram alterações, uma vez que no caso da RAIS é o empregador quem declara a raça dos empregados. Isso é um grande problema na medida em que não há garantias de que a raça declarada de uma pessoa é a mesma que ela considera ter. Além disso, a mesma pessoa pode ter sua raça declarada de formas diferentes por diferentes empregadores. Na tentativa de

contornar essa questão, foi utilizada a raça de maior ocorrência para cada indivíduo. Por exemplo, uma pessoa foi considerada branca se, dentre todos os empregos que ela teve no setor formal, mais empregadores a declararam como branca. Sendo assim, levando em consideração esses fatores, tem-se que 72% das mulheres e 70% dos homens são brancos ou amarelos. As mulheres negras, pardas e indígenas somam pouco mais de 7%, e os homens pretos, pardos e indígenas pouco mais de 8%. Para 20% das mulheres e 22% dos homens não constam informações sobre a declaração da raça ou ela é indeterminada (quando, por exemplo, uma pessoa teve dois empregos no qual um deles a declarou como branca e o outro como negra, parda ou indígena). Considerando os problemas relacionados a essa variável, optou-se por limitar a sua análise na presente seção de estatísticas descritivas, não as utilizando posteriormente nas regressões econométricas.

As variáveis de estado civil e idade também apresentam tendências parecidas com a base de dados anterior, com a maior parte dos indivíduos declarando ser solteiros e nascidos nas décadas de 1970 e 1980. Contudo, como a informação acerca do estado civil possui elevado nível de informações faltantes (“não declarado”), é única e determinada no ato de ingresso no ensino superior, não é possível garantir que o estado civil dos indivíduos não tenha se alterado com o passar do tempo. Por conta disso, também se optou por não utilizar as variáveis sobre estado civil posteriormente no exercício empírico.

Em relação às variáveis sobre o mercado de trabalho, nota-se que a maior parte dos indivíduos trabalha na região Sul: 89% das mulheres e 86% dos homens, seguida pelas regiões Sudeste, Centro-Oeste, Nordeste e Norte. Para chegar a tais resultados foi utilizada a mesma estratégia da *dummy* de raça, ou seja, um indivíduo foi considerado na região Sul se, dentre todos os empregos que ele teve no setor formal, a maior ocorrência foi de um emprego da região Sul.

No que diz respeito à remuneração, a média salarial das mulheres é de aproximadamente R\$ 6.500 reais enquanto a dos homens é de aproximadamente R\$ 8.700 reais.

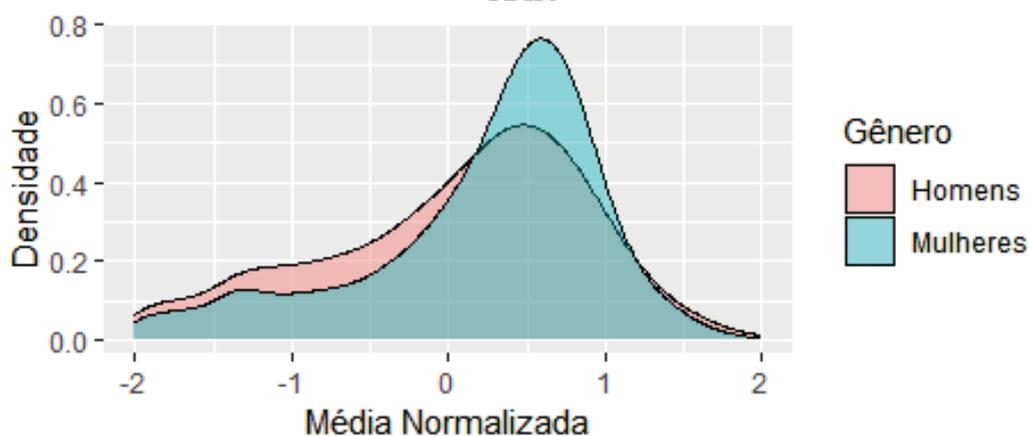
A média de jornada dos homens é ligeiramente maior do que a das mulheres, 36,7 e 35,3 horas respectivamente. Devido ao fato de a base de dados conter muitas observações com zero nas horas trabalhadas declaradas, optou-se por utilizar o salário mensal e adicionar controles de jornada de trabalho por horas. Nesse sentido, os dados expõem que indivíduos estão concentrados em jornadas de trabalho entre 31 e 44 horas semanais. Contudo, nota-se que as mulheres estão mais concentradas em jornadas até 20 horas e de 21 a 30 horas comparativamente aos homens. Novamente, para estipular em qual intervalo o indivíduo se encontra, foi considerado aquele com a maior ocorrência, assim como para a raça e para a

região. Conforme indica a literatura (GOLDIN, KATZ, 2011; GOLDIN, 2014), essa diferença pode ser explicada devido ao fato de que a parcela feminina ainda precisa dividir seu tempo entre tarefas remuneradas no mercado de trabalho e não remuneradas em casa e muitas vezes acabam tendo que optar por trabalhos que oferecem uma jornada mais flexível. Nesse sentido, as mulheres são penalizadas, também, por demandarem empregos que apresentam estrutura de horários mais flexível, de forma que consigam conciliar o trabalho remunerado com o trabalho não remunerado do cuidado.

Ademais, os dados também mostram que 6% das mulheres e 8% dos homens já estavam no mercado de trabalho formal quando iniciaram o curso superior na UFPR e que 19% das mulheres e 24% dos homens começaram a trabalhar antes de cursar a última matéria.

Os Gráficos 12 e 13 a seguir trazem as informações acerca da média normalizada das notas por curso e da média normalizada dos salários.

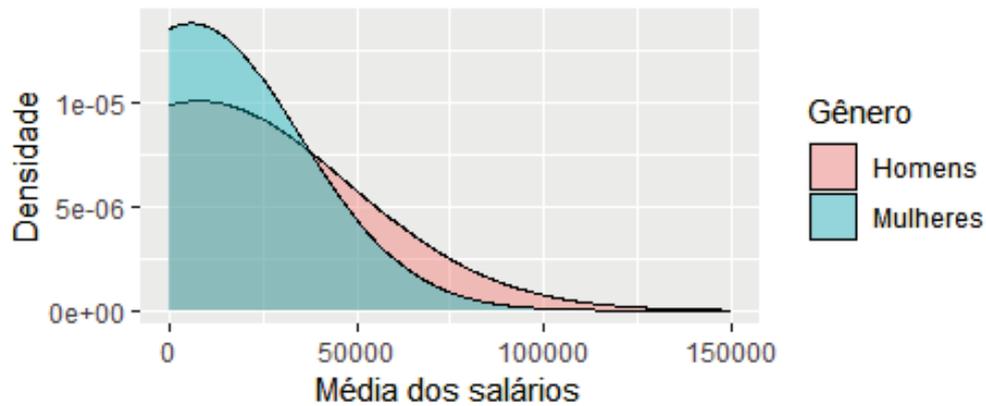
Gráfico 12 – Densidade da média normalizada das notas por curso - base de dados UFPR e RAIS



Fonte: Elaboração própria a partir de dados da UFPR e RAIS.

Pode-se notar a partir do Gráfico 12 que as mulheres permanecem mais concentradas do que os homens na parcela da distribuição com maiores notas, como na base anterior. Contudo, quando os salários são analisados, a partir do Gráfico 13 a seguir, essa tendência se inverte.

Gráfico 13 – Densidade da média normalizada dos salários em reais - base de dados UFPR e RAIS



Fonte: Elaboração própria a partir de dados da UFPR e RAIS.

Pode-se observar que mulheres estão relativamente mais concentradas do que os homens na parte da distribuição com menores salários. Esse resultado joga luz na seguinte hipótese, também apontada por Quadlin (2018), que será investigada na próxima seção: mesmo com maiores habilidades ou maior nível de esforço, as mulheres enfrentam resultados relativamente desfavoráveis no mercado de trabalho apesar de apresentarem melhores desempenhos acadêmicos?

### 3.3.2 Descrição das variáveis e das hipóteses

A principal questão a ser investigada pelo presente ensaio é se as mulheres, apesar de apresentarem, em média, melhores desempenhos em comparação aos homens, estão enfrentando os piores resultados em termos de remuneração no mercado de trabalho.

O Quadro 2 a seguir traz as descrições das variáveis consideradas.

Quadro 2 – Descrição das variáveis

Variável	Descrição
Gênero	<i>Dummy</i> igual a um se o indivíduo é mulher e igual a zero se é homem.
Média das notas	Média normalizada das notas por área das matérias cursadas na UFPR.
Faltas	Média das faltas agregadas na UFPR.
Aprovação	Média das aprovações na UFPR.

(continua)

(continuação)

Variável	Descrição
Reprovação por nota	Média das reprovações por nota.
Reprovação por frequência	Média das reprovações por frequência.
Turno integral; matutino*; noturno e diurno/vespertino	<i>Dummies</i> de turnos.
Forma de evasão formatura	<i>Dummy</i> igual a um se o indivíduo tem "Formatura" como forma de evasão e igual a zero caso contrário.
Tempo graduação	Tempo calculado como a diferença entre o primeiro e o último ano das matérias cursadas.
Agrárias; Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde; Sociais Aplicadas*	<i>Dummies</i> de áreas.
Média das notas x Média das notas	Variável de interação entre a média normalizada por área das notas das matérias cursadas.
Gênero x Nota	Variável de interação entre o gênero e a média normalizada por área das notas.
Salário	Logaritmo da média do salário em reais recebido pelos indivíduos aos preços de dezembro de 2021.
Meses empregado(a)	Média dos meses empregados de cada indivíduo.
Jornada de trabalho até 20 horas*	<i>Dummy</i> igual a um se a jornada de trabalho do indivíduo é de até 20 horas semanais e igual a zero caso contrário.
Jornada de trabalho de 21 a 30 horas	<i>Dummy</i> igual a um se a jornada de trabalho do indivíduo é de 21 a 30 horas semanais e igual a zero caso contrário.
Jornada de trabalho de 31 a 44 horas	<i>Dummy</i> igual a um se a jornada de trabalho do indivíduo é de 31 a 44 horas semanais e igual a zero caso contrário.
Nascidos entre 1930 - 1959; 1960-1969; 1970-1979; 1980-1989; 1990-1999*	<i>Dummies</i> de décadas de nascimento.
Regiões Sul*; Sudeste; Centro-Oeste; Nordeste; Norte	<i>Dummies</i> de regiões.
Estava no mercado de trabalho formal antes de iniciar o curso	<i>Dummy</i> igual a um se o indivíduo já estava no mercado de trabalho formal antes de entrar na graduação e igual a zero caso contrário.
Estava no mercado de trabalho antes de terminar o curso	<i>Dummy</i> igual a um se o indivíduo entrou no mercado de trabalho formal antes de cursar a última matéria na UFPR e igual a zero caso contrário.
Inversa de Mills	Razão Inversa de Mills.

Fonte: Elaboração própria.

\* Categorias de base.

### 3.3.3 Correção de Heckman

Um dos problemas mais comuns de estudos que utilizam dados amostrais não aleatórios diz respeito ao viés de seleção. Conforme Heckman (1979), o viés de seleção pode surgir tanto a partir da auto seleção por parte dos indivíduos analisados quanto a partir dos pesquisadores e

do desenho amostral das pesquisas. Dessa forma, o presente ensaio utiliza a correção de Heckman (1979) uma vez que não é possível auferir os salários e as horas trabalhadas dos alunos egressos que não estão presentes na base de dados da RAIS. Em outras palavras, a RAIS só capta trabalhadores no mercado formal e este grupo é positivamente selecionado.

A equação de seleção de Heckman estima a probabilidade de os indivíduos estarem no mercado de trabalho formal a partir de algumas variáveis explicativas, com o resíduo corrigido para o viés de seleção amostral.

Conforme Kassouf (1994), considerando uma variável  $L$  que representa a participação no mercado de trabalho formal, tem-se que:

$$L^*_i = \gamma'Z_i + u_i \quad (5)$$

Em que  $Z_i$  é um vetor de variáveis que explicam a participação no mercado de trabalho.  $L^*$  não é observada na prática. O que se observa é  $L$ , tal que:

$$L_i = 1 \text{ se } L^*_i > 0 \quad (6)$$

$$L_i = 0 \text{ se } L^*_i \leq 0 \quad (7)$$

Considerando  $W$  como os salários, tem-se que:

$$W_i = \beta'R_i + v_i \quad (8)$$

Em que  $R_i$  é um vetor de variáveis explicativas que determinam os salários e  $v_i$  é o vetor de erros. A variável  $W$  será observada somente quando  $L > 0$ .

Assumindo que:

$$u_i \sim N(0, \sigma_u),$$

$$v_i \sim N(0, \sigma_v),$$

$$\text{corr}(u_i, v_i) = \rho$$

Então:

$$\begin{aligned} E(W_i | W_i \text{ observado}) &= E(W_i | L^*_i > 0) = E(W_i | u_i > -\gamma'Z_i) = \\ &\beta'R_i + E(v_i | u_i > -\gamma'Z_i) = \beta'R_i + \rho\sigma_v\lambda_i(\alpha_u) \end{aligned} \quad (9)$$

Em que:

$$\lambda_i(\alpha_u) = \frac{\phi\left(\frac{\gamma'Z_i}{\sigma_u}\right)}{\Phi\left(\frac{\gamma'Z_i}{\sigma_u}\right)} \quad (10)$$

E  $\phi$  é a função densidade de probabilidade e  $\Phi$  é a função de distribuição. A função  $\lambda_i(\alpha_u)$ , por sua vez, é denominada inverso da razão de Mills, que incluída na equação de salários resulta em:

$$W|L > 0 = \beta'R_i + \beta_\lambda\lambda_i(\alpha_u) + \varepsilon_i \quad (11)$$

A partir da Equação 11, pode-se notar que a regressão dos salários em função das variáveis explicativas a partir de mínimos quadrados ordinários somente para indivíduos que estão no mercado de trabalho formal resultaria em um estimador de  $\beta$  inconsistente na medida em que estaria omitindo a variável  $\lambda$ .

Para contornar esse problema, pode-se utilizar o procedimento de correção de Heckman (1979), cuja estimação consiste basicamente em dois passos. Primeiro, um modelo probit é estimado por máxima verossimilhança, no qual a variável dependente assume o valor igual a um se o indivíduo participar do mercado de trabalho formal e igual a zero caso contrário. Posteriormente, a partir dos resultados do primeiro cálculo, é possível obter as estimativas para  $\lambda$ . A partir disso, pode-se, então, estimar a regressão de  $W$  em função de  $R$  e  $\lambda$  por meio do MQO e obter estimativas consistentes de  $\beta$ .

A equação de seleção utilizada segue a Equação 12:

$$\begin{aligned} \text{Mercado de trabalho} = & \beta_0 + \beta_1 \text{Gênero}_i + \beta_2 \text{Idade de formação}_i + \\ & \beta_3 \text{Idade de formação}_i^2 + \beta_4 \text{Tempo de graduação}_i + \beta_5 \text{Dummies de turno}_i + \\ & \beta_6 \text{Média das notas} + \varepsilon_i \end{aligned} \quad (12)$$

Em que: *Mercado de trabalho*: probabilidade do egresso estar no mercado de trabalho formal; *Gênero*: *dummy* igual a um se o indivíduo é mulher e igual a zero caso contrário; *Idade de formação*: idade que o indivíduo concluiu o curso ou cursou a última matéria; *Idade de formação*<sup>2</sup>: idade que o indivíduo concluiu o curso ou cursou a última matéria elevada ao quadrado; *Tempo de graduação*: tempo entre o início e término do curso – ou entre a primeira e a última matéria cursada; *Dummies de turno*: *dummy* igual a um se o indivíduo ingressou em um curso com turno integral/matutino/noturno/diurno ou vespertino; *Média das notas*: média das notas dos indivíduos.

As equações foram estimadas para a amostra completa considerando todos os indivíduos e também apenas para os formados, juntos e separadamente para mulheres e homens. Os resultados das estimações estão reportados na Tabela A6 nos Apêndices.

### 3.4 RESULTADOS E DISCUSSÕES

Os modelos consideraram como variável independente o logaritmo do salário em reais deflacionado de acordo com o IPCA tendo como período base dezembro de 2021. As estimações foram realizadas a partir do método econométrico de Mínimos Quadrados Ordinários (MQO).

Foram estimadas sete especificações, com a adição gradual das seguintes variáveis explicativas: gênero, média normalizada das notas por curso, média normalizada ao quadrado das notas por curso, variável de interação entre o gênero e a média normalizada das notas por curso e a razão da inversa de Mills para mitigar o problema do viés de seleção. Além disso, também foram considerados controles de características pessoais (décadas de nascimento), controles de características da formação (turnos dos cursos: matutino, integral, noturno e diurno/vespertino; *dummy* para formados, tempo de graduação, áreas dos cursos), controles de características do trabalho (região do trabalho, jornadas de trabalho em horas semanais, média de meses empregado(a)) e outros controles (*dummy* para quem já estava no mercado de trabalho formal antes de ingressar no curso e *dummy* para quem entrou no mercado de trabalho formal antes de terminar o curso).

Primeiramente, os valores foram estimados considerando a base de dados completa (31.553 observações) com os indivíduos juntos e também individualmente para mulheres e homens. Posteriormente, apenas aqueles cuja forma de evasão constava como “Formatura” (19.202 observações) foram considerados.

A última especificação, mais completa, segue a Equação 13. Os resultados para todos os indivíduos estão reportados nas Tabelas 27. A Tabela 28, por sua vez, traz os valores estimados para as amostras de mulheres e a Tabela 29 para a amostra de homens. A Tabela 27 a seguir enfatiza os principais resultados. Para as estimações completas, ver Tabelas A7 a A12 nos Apêndices.

$$\ln \text{salários}_i = \beta_0 + \beta_1 \text{Gênero}_i + \beta_2 \text{Notas}_i + \beta_3 \text{Notas}_i^2 + \beta_4 \text{Gênero}_i \times \text{Nota}_i + \beta_5 \text{IMR}_i + \beta_6 \Gamma + \beta_7 \Pi + \beta_8 \Delta + \beta_9 \text{H} + \varepsilon_i \quad (13)$$

Em que:  $\ln \text{salários}_i$ : logaritmo natural dos salários recebidos;  $Gênero_i$  : variável *dummy* igual a 1 para as mulheres e igual a 0 para os homens;  $Notas_i$ : média normalizada por curso das notas dos indivíduos;  $Notas_i^2$ : média normalizada por curso das notas ao quadrado dos indivíduos;  $Gênero_i \times Nota_i$ : variável de interação entre Gênero e Nota;  $\Gamma$ : vetor de controles de características pessoais;  $\Pi$ : vetor de controles de características da formação;  $\Delta$ : vetor de controles de características do trabalho;  $H$ : vetor de outros controles;  $\varepsilon_i$ : termo de erro.

Tabela 27 – Resultados das estimações considerando todos os indivíduos

Variáveis	1	2	3	4	5	6	7
Intercepto	8.630*** (0.007)	8.640*** (0.007)	8.612*** (0.008)	8.609*** (0.008)	8.660*** (0.050)	8.601*** (0.053)	9.420*** (0.104)
Gênero	-0.259*** (0.010)	-0.280*** (0.010)	-0.284*** (0.010)	-0.284*** (0.010)	-0.286*** (0.010)	-0.270*** (0.010)	-0.197*** (0.009)
Média das notas normalizada por área		0.120*** (0.005)	0.157*** (0.006)	0.135*** (0.008)	0.115*** (0.008)	0.106*** (0.009)	0.061*** (0.008)
(Média das notas normalizada por área) <sup>2</sup>			0.030*** (0.003)	0.031*** (0.003)	0.029*** (0.003)	0.029*** (0.003)	0.020*** (0.003)
Gênero x Média das notas normalizadas por área				0.048*** (0.010)	0.044*** (0.010)	0.041*** (0.010)	0.021*** (0.009)
IMR							-0.858*** (0.073)
Controles de características pessoais <sup>30</sup>					✓	✓	✓
Controles de características da formação <sup>31</sup>						✓	✓
Controles de características do trabalho <sup>32</sup>							✓

(continua)

<sup>30</sup> Características pessoais: décadas de nascimento (1930-1959; 1960-1969; 1970-1979; 1980-1989; 1990-1999\*).

<sup>31</sup> Características da formação: turnos do curso (noturno, integral, vespertino/diurno, matutino\*); *dummy* para formados; tempo de graduação; área do curso (Agrárias, Biológicas, Engenharias, Exatas e Terra, Humanas, Linguística, Letras e Artes, Saúde, Sociais Aplicadas\*).

<sup>32</sup> Características do trabalho formal: região (Norte, Nordeste, Sudeste, Centro-Oeste, Sul\*); *dummies* de jornadas de trabalho semanais em horas (até 20 horas\*; 21-30; 31-44); média de meses empregado(a).

Variáveis	(continuação)						
	1	2	3	4	5	6	7
Outros controles <sup>33</sup>							✓
Observações	31,553	31,553	31,553	31,553	31,553	31,553	31,553
Log Likelihood	-42,524.390	-42,261.850	-42,215.000	-42,204.560	-41,577.230	-41,325.490	-37,476.700
Akaike Inf. Crit.	85,052.780	84,529.710	84,437.990	84,419.130	83,172.460	82,692.99	75,015.400

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

<sup>33</sup> Outros controles: *dummy* para quem estava no mercado de trabalho formal antes de terminar o curso; *dummy* para quem estava no trabalho formal antes de iniciar o curso.

Como esperado, o sinal da variável do *Gênero* é negativo em todos os modelos testados, indicando que ser mulher impacta negativamente os salários recebidos comparativamente aos homens.

O sinal da variável *Média das notas normalizadas por área* também se portou como o esperado, na medida em que sugere evidências estatísticas de que o salário dos indivíduos é positivamente influenciado por seus desempenhos acadêmicos em termos de notas. Ou seja, indivíduos cujo desempenho acadêmico é melhor no ensino superior têm retornos maiores em termos salariais no mercado de trabalho. O sinal positivo e estatisticamente significativo da variável das *Média das notas normalizadas por área ao quadrado* indica que sua taxa de mudança está acelerando à medida que os salários aumentam.

O sinal positivo e estatisticamente significativo da variável de interação entre o gênero e a nota está sugerindo que existem evidências estatísticas de que o melhor desempenho acadêmico apresentado pelas mulheres no ensino superior está se traduzindo em melhores salários posteriormente no mercado de trabalho formal, ainda que a sua magnitude indique um efeito pequeno. Ou seja, as mulheres que possuem notas mais elevadas durante a graduação tendem a conseguir reduzir parcialmente o efeito negativo de gênero sobre seus salários.

A variável que representa a inversa da razão de Mills apresentou significância estatística e sinal positivo, indicando que o problema de viés de seleção da base foi mitigado.

As Tabelas 28 e 29 a seguir trazem os resultados das estimações feitas separadamente para as amostras de mulheres e homens, respectivamente.

Tabela 28 – Resultados das estimações para as mulheres

Variáveis	1	2	3	4	5
Intercepto	8.358*** (0.007)	8.325*** (0.009)	8.439*** (0.066)	8.394*** (0.070)	7.239*** (0.156)
Média normalizada por área	0.142*** (0.007)	0.183*** (0.009)	0.155*** (0.009)	0.145*** (0.011)	0.099*** (0.010)
(Média normalizada por área) <sup>2</sup>		0.031*** (0.004)	0.026*** (0.004)	0.026*** (0.004)	0.018*** (0.004)
IMR					0.922*** (0.115)
Controles de características pessoais			✓	✓	✓
Controles de características da formação				✓	✓

(continua)

(continuação)					
Variáveis	1	2	3	4	5
Controles de características do trabalho					✓
Outros controles					✓
Observações	15,456	15,456	15,456	15,456	15,456
Log Likelihood	-20,189.53	-20,163.440	-19,823.400	-19,699.850	-17,405.720
Akaike Inf. Crit.	40,383.050	40,332.870	39,660.790	39,437.700	34,871.450

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela 29 – Resultados das estimações para os homens

Variáveis	1	2	3	4	5
Intercepto	8.638*** (0.008)	8.609*** (0.009)	8.590*** (0.074)	8.544*** (0.078)	7.464*** (0.159)
Média normalizada por área	0.101*** (0.007)	0.136*** (0.009)	0.118*** (0.009)	0.107*** (0.011)	0.071*** (0.009)
(Média normalizada por área) <sup>2</sup>		0.031*** (0.005)	0.031*** (0.004)	0.032*** (0.005)	0.022*** (0.004)
IMR					0.896*** (0.117)
Controles de características pessoais			✓	✓	✓
Controles de características da formação				✓	✓
Controles de características do trabalho					✓
Outros controles					✓
Observações	16,097	16,097	16,097	16,097	16,097
Log Likelihood	-22,034.55	-22,011.040	-21,718.240	-21,514.610	-19,103.420
Akaike Inf. Crit.	44,073.110	44,028.080	43,450.480	43,067.220	38,266.830

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Quando as estimações são feitas separadamente para as mulheres e para os homens, nota-se que tanto os sinais quanto as magnitudes dos valores permanecem semelhantes aos das estimações considerando todos os indivíduos juntos. Nesses casos, o efeito da média normalizada das notas por curso é ligeiramente mais acentuado para as mulheres (0,099) do que para os homens (0,071).

### *3.4.1 Análise de robustez e heterogeneidade*

Na tentativa de explorar a robustez dos resultados reportados acima, algumas estimações adicionais foram realizadas. Primeiramente, os valores foram estimados considerando apenas os indivíduos que apresentaram “Formatura” como forma de evasão da UFPR, uma vez que pode existir o efeito diploma nos resultados anteriores. Em seguida, foram investigadas outras variáveis de desempenho, como as faltas, as reprovações por nota e por frequência e as aprovações. Uma vez que as áreas são heterogêneas entre si, o terceiro e último teste analisou separadamente os efeitos para cada uma das oito áreas (Agrárias; Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde e Sociais Aplicadas).

#### *3.4.1.1 Formados*

Posteriormente às estimações com todos os indivíduos, foram estimados os resultados considerando apenas aqueles que apresentaram “Formatura” como forma de evasão. Isso porque pode existir um efeito diploma nos resultados anteriores.

As Tabelas 30, 31 e 32 trazem os resultados para as novas amostras.

Tabela 30 – Resultados das estimações considerando todos os indivíduos formados

Variáveis	1	2	3	4	5	6	7
Intercepto	8.704*** (0.010)	8.719*** (0.010)	8.697*** (0.011)	8.696*** (0.011)	8.793*** (0.062)	8.533*** (0.070)	8.782*** (0.125)
Gênero	-0.250*** (0.014)	-0.280*** (0.014)	-0.280*** (0.014)	-0.280*** (0.014)	-0.281*** (0.013)	-0.244*** (0.013)	-0.205*** (0.012)
Média das notas normalizada por área		0.127*** (0.007)	0.143*** (0.007)	0.124*** (0.010)	0.114*** (0.010)	0.144*** (0.010)	0.033*** (0.015)
(Média das notas normalizada por área) <sup>2</sup>			0.022*** (0.004)	0.021*** (0.004)	0.021*** (0.004)	0.020*** (0.004)	0.017*** (0.004)
Gênero x Média das notas normalizadas por área				0.037*** (0.014)	0.014 (0.013)	0.008 (0.013)	0.025*** (0.012)
Inversa de Mills							-0.285***
Controles de características pessoais					✓	✓	✓
Controles de características da formação						✓	✓
Controles de características do trabalho							✓
Outros controles							✓
Observações	19,202	19,202	19,202	19,202	19,202	19,202	19,202
Log Likelihood	-26,168.910	-25,996.720	-25,983.220	-25,979.630	-25,447.700	-25,250.270	-22,912.130
Akaike Inf. Crit.	52,341.810	51,999.450	51,974.450	51,969.260	50,913.390	50,540.540	45,884.250

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela 31 – Resultados das estimações individuais para as mulheres formadas

Variáveis	1	2	3	4	5
Intercepto	8.456*** (0.009)	8.438*** (0.010)	8.611*** (0.080)	8.359*** (0.090)	9.827*** (0.261)
Média normalizada por área	0.118*** (0.009)	0.134*** (0.010)	0.109*** (0.010)	0.133*** (0.010)	0.049*** (0.012)
(Média normalizada por área) <sup>2</sup>		0.018*** (0.005)	0.017*** (0.005)	0.017*** (0.005)	0.013*** (0.004)
Inversa de Mills					-1.261*** (0.188)
Controles de características pessoais			✓	✓	✓
Controles de características da formação				✓	✓
Controles de características do trabalho					✓
Outros controles					✓
Observações	10,006	10,006	10,006	10,006	10,006
Log Likelihood	-13,089.43	-13,083.080	-12,835.750	-12,707.200	-11,433.740
Akaike Inf. Crit.	26,182.870	26,172.160	25,685.500	25,450.400	22,923.490

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela 32 – Resultados das estimações individuais para os homens formados

Variáveis	1	2	3	4	5
Intercepto	8.697*** (0.010)	8.673*** (0.013)	8.677*** (0.096)	8.467*** (0.107)	9.456*** (0.228)
Média normalizada por área	0.126*** (0.010)	0.139*** (0.011)	0.123*** (0.011)	0.156*** (0.012)	0.074*** (0.011)
(Média normalizada por área) <sup>2</sup>		0.025*** (0.007)	0.026*** (0.007)	0.024*** (0.007)	0.018*** (0.006)
Inversa de Mills					-0.975***
Controles de características pessoais			✓	✓	✓
Controles de características da formação				✓	✓
Controles de características do trabalho					✓
Outros controles					✓
Observações	9,196	9,196	9,196	9,196	9,196
Log Likelihood	-12,889.39	-12,883.500	-12,594.930	-12,516.220	-11,400.530
Akaike Inf. Crit.	25,782.770	25,773.010	25,203.860	25,068.450	22,857.070

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Os valores estimados considerando apenas os formados mantêm a tendência dos resultados para a base completa, tanto em termos de sinais quanto em termos de magnitudes, nos três casos indicando robustez nos resultados do primeiro modelo, ainda que o diferencial entre homens e mulheres não seja significativo em especificações intermediárias, passa a ser significativo e de magnitude semelhante na especificação mais completa. Ou seja, melhores desempenhos no ensino superior, em termos de maiores notas, estão apresentando melhores retornos no mercado de trabalho para ambos os gêneros.

### 3.4.1.2 Outras variáveis de desempenho

Além das notas, as variáveis que representam as faltas, as reprovações por nota e as reprovações e por falta e as aprovações também podem ser consideradas como determinantes do desempenho acadêmico dos indivíduos. Nesse sentido, a Tabela 33 a seguir apresenta os resultados das estimações considerando essas outras variáveis. A primeira coluna analisa a variável que representa as faltas, a segunda coluna a variável de reprovação por nota, a terceira coluna a variável de reprovação por frequência e a quarta coluna as aprovações. Todas as variáveis são também analisadas interagindo com o gênero dos indivíduos.

Tabela 33 – Resultados das estimações para todos os indivíduos com outras variáveis de desempenho

Variáveis	1	2	3	4
Intercepto	7.581*** (0.111)	7.578*** (0.111)	7.584*** (0.110)	7.317*** (0.113)
Gênero	-0.186*** (0.011)	-0.188*** (0.011)	-0.192*** (0.010)	-0.231*** (0.021)
Faltas	-0.001** (0.0003)			
Gênero x Faltas	-0.001 (0.0004)			
Reprovações por nota		-0.090*** (0.028)		
Gênero x Reprovações por nota		-0.052 (0.037)		

(continua)

(continuação)				
Variáveis	1	2	3	4
Reprovações por frequência			-0.171*** (0.026)	
Gênero x Reprovações por frequência			-0.051 (0.036)	
Aprovações				0.216*** (0.022)
Gênero x Aprovações				0.040 (0.026)
Inversa de Mills	0.832*** (0.082)	0.833*** (0.082)	0.857*** (0.082)	0.940*** (0.082)
Controles de características pessoais	✓	✓	✓	✓
Controles de características da formação	✓	✓	✓	✓
Controles de características do trabalho	✓	✓	✓	✓
Outros controles	✓	✓	✓	✓
Observações	31,553	31,553	31,553	31,553
Log Likelihood	-36,663.120	-36,657.480	36,630.340	-36,594.440
Akaike Inf. Crit.	73,388.240	73,376.950	73,322.680	73,250.880

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

A partir dos resultados da Tabela 33, pode-se observar que as variáveis de faltas, reprovações por nota e por frequência são estatisticamente significativas e com sinal negativo, indicando que o salário dos indivíduos é negativamente afetado por elas. A variável que representa as aprovações foi estatisticamente significativa e com sinal positivo, conforme esperado. Os resultados não significativos das variáveis de interação com o gênero sugerem que não existem evidências estatísticas de diferença nesse efeito entre mulheres e homens.

### 3.4.1.3 Análises por áreas

Uma vez que as áreas do conhecimento são heterogêneas e possuem especificidades, o terceiro teste estima os valores separadamente para cada uma das oito áreas consideradas (Agrárias; Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde e Sociais Aplicadas). A Tabela 31 a seguir traz os resultados encontrados

Tabela 34 – Resultados das estimações para todos os indivíduos para cada área individualmente

Variáveis	Linguística, Letras e Artes						
	Agrárias	Biológicas	Engenharias	Exatas e Terra	Humanas	Saúde	Sociais Aplicadas
Intercepto	8.570*** (0.566)	6.366*** (0.573)	10.135*** (0.516)	8.944*** (0.211)	8.267*** (0.261)	8.562*** (0.408)	9.550*** (0.211)
Gênero	-0.279*** (0.040)	-0.202*** (0.044)	-0.277*** (0.029)	-0.268*** (0.020)	-0.082*** (0.025)	-0.055 (0.034)	-0.213*** (0.019)
Média das notas normalizada por área	0.157*** (0.042)	0.179*** (0.049)	0.031 (0.024)	0.084*** (0.014)	0.151*** (0.024)	0.123*** (0.034)	0.075*** (0.018)
(Média das notas normalizada por área) <sup>2</sup>	0.027* (0.015)	0.026 (0.018)	0.032** (0.014)	0.052*** (0.010)	0.062*** (0.012)	0.064*** (0.018)	0.052*** (0.008)
Gênero x Média das notas normalizadas por área	-0.024 (0.039)	-0.079* (0.042)	0.049* (0.029)	0.007 (0.019)	0.017 (0.024)	0.049 (0.033)	0.038** (0.019)
Inversa de Mills	-0.338 (0.374)	1.724*** (0.406)	-1.124*** (0.271)	-0.733*** (0.145)	-0.213 (0.176)	-0.695*** (0.253)	-1.020*** (0.152)
Controles de características pessoais	✓	✓	✓	✓	✓	✓	✓
Controles de características da formação	✓	✓	✓	✓	✓	✓	✓
Controles de características do trabalho	✓	✓	✓	✓	✓	✓	✓
Outros controles	✓	✓	✓	✓	✓	✓	✓
Observações	1,780	1,380	3,851	6,564	3,910	1,942	7,057
Log Likelihood	-2,158.26	-1,551.17	-4,706.72	-7,513.48	-4,362.200	-2,094.82	-8,211.431
Akaike Inf. Crit.	4,358.538	3,146.353	9,459.449	15,074.98	8,772.400	4,235.646	16,470.860

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Conforme os resultados reportados na Tabela 31, quando a variável do gênero é analisada separadamente para cada área, nota-se que os salários das mulheres são mais penalizados nas áreas de Exatas e Terra seguida por Agrárias, Engenharias, Sociais Aplicadas, Biológicas, Saúde e Humanas. A área de Linguística, Letras e Artes não apresentou resultado estatisticamente significativo.

Em relação ao desempenho, o efeito positivo das melhores notas sobre os salários é mais acentuado na área de Biológicas seguida por Agrárias, Humanas, Linguística, Letras e Artes, Exatas e Terra e Sociais Aplicadas. A área da Saúde apresentou um efeito negativo das notas sobre os salários enquanto o resultado para as Engenharias não foi estatisticamente significativo. No caso específico da Saúde, vale ressaltar que existem muitos profissionais que atuam por conta própria e são pessoas jurídicas. Dessa forma, seus rendimentos não estão contabilizados na RAIS, o que pode explicar o resultado contrário dos demais.

Quando o retorno das notas é analisado por gênero, nota-se que o melhor desempenho no ensino superior das mulheres é recompensado posteriormente no mercado de trabalho formal nas áreas da Saúde, Engenharias e Sociais Aplicadas. A área de Biológicas foi a única que apresentou sinal negativo, indicando que as mulheres com melhor desempenho dessa área são prejudicadas em termos salariais posteriormente no mercado de trabalho. Os resultados para as demais áreas não apresentaram significância estatística.

### *3.4.2 Limitações metodológicas*

A primeira limitação metodológica do presente ensaio diz respeito ao fato de que investiga apenas o mercado de trabalho formal e pessoas físicas. Não se pode, portanto, obter informações acerca das pessoas jurídicas e, principalmente, daquelas que estão no mercado de trabalho informal. Isso é um problema na medida em que, de acordo com o IBGE (2023), o Brasil apresentou uma taxa de informalidade de aproximadamente 22,8% entre os indivíduos com ensino superior completo no ano de 2022.

A segunda limitação metodológica se refere aos problemas das informações sobre a raça e o sexo dos indivíduos. Em muitos casos, pôde-se notar que as variáveis sobre a raça e sexo sofreram alterações, uma vez que no caso da RAIS é o empregador quem as declara. Isso é um grande problema na medida em que não há garantias de que a raça ou o sexo declarado de uma pessoa são os mesmos que ela considera ter. Além disso, a mesma pessoa pode ter sua raça ou sexo declarados de formas diferentes por diferentes empregadores. Dessa forma, optou-se por não incluir nas análises essas informações advindas da base de dados da RAIS. A informação

do sexo dos indivíduos foi utilizada da base de dados da UFPR enquanto a raça não foi considerada devido ao fato de muitas observações constarem como “não definida” na base da UFPR. Contudo, vale ressaltar a importância da busca por alternativas para contornar esse problema de natureza estrutural, com vistas a garantir que os indivíduos não tenham suas raças e/ou sexos, aspectos tão sensíveis à autoidentificação em sociedade, declarados de forma equivocada.

A terceira limitação, por sua vez, remete aos limites espaciais impostos pela própria base de dados. O Brasil é um país conhecido por seu tamanho e suas diferenças regionais. Dessa forma, os resultados encontrados utilizando apenas egressos da Universidade Federal do Paraná, localizada na capital de um Estado da Região Sul, devem ser interpretados com cautela, na medida em que não necessariamente irão refletir a realidade dos demais indivíduos do país.

### **3.5 CONSIDERAÇÕES FINAIS**

Os últimos anos vêm sendo marcados pela presença relativamente mais acentuada das mulheres nas universidades quando comparadas aos homens – mais mulheres entram e mais mulheres concluem um curso de ensino superior. Além disso, a literatura aponta também para o fato de que a parcela feminina apresenta, em média, desempenhos melhores do que a masculina, tanto em níveis iniciais e intermediários de educação quanto em níveis superiores. Mas, até que ponto essa dedicação vale a pena? Para além da formação superior, as mulheres enfrentam resultados relativamente desfavoráveis no mercado de trabalho apesar de apresentarem melhores desempenhos acadêmicos?

Na tentativa de investigar acerca dessa questão, o presente ensaio analisou se durante o período de formação universitária, mulheres e homens possuem comportamentos distintos no que diz respeito ao desempenho acadêmico, e se esse comportamento se reflete ou não nos resultados no mercado de trabalho.

A partir de uma base de dados que integrou informações acadêmicas de egressos da Universidade Federal do Paraná com informações acerca de suas remunerações no mercado de trabalho formal da base de dados da RAIS identificada, os resultados encontrados sugerem que os salários dos indivíduos no mercado de trabalho são influenciados positivamente por seus desempenhos no ensino superior. Quando os gêneros são analisados separadamente, encontra-se, também, que as mulheres com melhores performances ganham maiores retornos salariais, ainda que a magnitude do efeito seja pequena, e insuficiente para compensar a desigualdade

entre os gêneros. Essa tendência se mantém quando apenas aqueles que concluíram o ensino superior são considerados.

Quando outras variáveis que refletem o desempenho são consideradas, como as faltas, e as reprovações por notas e por frequência, os resultados obtidos sugerem que todas elas impactam negativamente os salários dos indivíduos, não apresentando diferenças entre mulheres e homens.

Por fim, quando os resultados são estimados separadamente para cada uma das áreas de formação (Agrárias; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes, Saúde e Sociais Aplicadas), encontrou-se que o melhor desempenho no ensino superior das mulheres é recompensado posteriormente no mercado de trabalho formal nas áreas da Saúde, Engenharias e Sociais Aplicadas.

Vale ressaltar, contudo, que esses resultados devem ser interpretados com cautela, na medida em que utilizam dados de egressos de apenas uma universidade, de modo que, considerando o tamanho e as diferenças regionais presentes no Brasil, não necessariamente irão refletir a realidade dos demais indivíduos egressos de outras universidades e/ou regiões do país. De todo modo, os resultados sugerem que o esforço vale a pena, pois reduz parcialmente o *gap* salarial entre gêneros, mas não é suficiente para que as mulheres se posicionem tão bem quanto os homens no mercado de trabalho. Além disso, indicam que as escolhas acerca da área de formação superior e da atuação no mercado de trabalho importam, corroborando os resultados do primeiro ensaio. Nesse sentido, fica novamente evidente a importância de políticas públicas e ações afirmativas que neutralizem a desigualdade de gênero, em especial nas áreas historicamente dominadas pelos homens, uma vez que, conforme apontou Levanon, England e Allison (2009), não são as mulheres que procuram piores empregos, e sim que as áreas perdem valor quando concentram um grande número da parcela feminina, indicando que é o trabalho mal pago que procura as mulheres e não o contrário.

## 4 ENSAIO 3: AS IMPLICAÇÕES DO MISMATCH EDUCACIONAL ENTRE MULHERES E HOMENS NO MERCADO DE TRABALHO FORMAL BRASILEIRO

### 4.1 INTRODUÇÃO

A média de anos de estudo da força de trabalho no Brasil experimentou um aumento significativo nas últimas décadas. Esse padrão se reflete, também, no aumento constante da participação de trabalhadores com educação superior. Conforme dados do Censo Demográfico, na década de 1980, apenas 5% dos indivíduos ocupados no Brasil possuíam ensino superior. Esse percentual aumentou para 8,5% na década de 1990, passando para 9,5% no início do novo milênio e chegando em 15% em 2010 (REIS, 2021). O crescimento observado a partir dos anos 2000 foi promovido, entre outras coisas, a partir de políticas sociais voltadas para atender a parcela marginalizada e de baixa renda da população que não tinha condições de frequentar níveis mais elevados de ensino. Por exemplo, o Programa Universidade Para Todos (PROUNI) concedeu aos alunos de baixa renda financiamento total ou parcial para o ingresso em universidades particulares; já o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), por sua vez, fomentou o investimento público no ensino superior e objetivou torná-lo mais inclusivo ao expandir o número de vagas e reduzir as desigualdades regionais. Além disso, contribuíram, também, as políticas afirmativas de cotas para alunos das escolas públicas e para a população negra. Conforme o INEP, essas políticas foram responsáveis por aumentar em aproximadamente 45% o número de matrículas no ensino superior entre 2003 e 2013 (MARIONI, 2020).

Contudo, quando a demanda por trabalho qualificado não acompanha o aumento da oferta de mão de obra educada, surge o problema do *mismatch* educacional. Este pode ser tanto de natureza vertical, quando um indivíduo está trabalhando em uma ocupação que requer mais (*undereducation* – subeducação) ou menos (*overeducation* – sobre-educação) anos de estudo do que efetivamente possui, quanto de natureza horizontal, que ocorre quando o indivíduo está empregado em uma ocupação não relacionada com a sua área de formação acadêmica. O presente ensaio focará nos sobre-educados (*mismatch* vertical) e no *mismatch* horizontal.

De acordo com Carvalho e Reis (2023), cerca de 25% dos trabalhadores ocupados no Brasil em 2012 eram sobre-educados e essa porcentagem aumentou em 11 pontos percentuais

em 2022. Os autores apontam também que, considerando o nível educacional dos indivíduos, os sobre-educados com ensino médio completo eram aproximadamente 42% em 2012, aumentando em 9 pontos percentuais em 2022 e os sobre-educados com ensino superior completo eram cerca de 28%, aumentando em 7 pontos percentuais dez anos depois.

A sobre-educação é um problema na medida em que indica que parte do investimento em educação não está sendo aproveitado pelo mercado de trabalho. A literatura existente aponta que os indivíduos sobre-educados não são remunerados da mesma forma que aqueles que possuem a escolaridade adequada à função que desempenham (BAUER, 2002; KORPI, TÄHLIN, 2009; TSAI, 2010; REIS, 2017; MARIONI, 2020). Além disso, eles são também mais propensos a ser insatisfeitos em relação aos seus trabalhos (VERHOFSTADT; DE WITTE; OMEY, 2007; ALLEN; VAN DER VELDEN, 2001) e a apresentar maior rotatividade de empregos (SICHERMAN, 1991).

De acordo com Reis (2021), a participação das mulheres no total de trabalhadores com nível superior completo aumentou de 41% em 1980 para 57% em 2010. Além disso, o autor destaca que, entre 1980 e 2000, a taxa de sobre-educação era maior para as mulheres. No entanto, essa dinâmica se inverteu em 2010, quando os homens se tornaram relativamente mais sobre-educados.

Os resultados encontrados pela literatura existente que combina a análise do *mismatch* educacional com questões gênero parecem depender bastante das técnicas utilizadas. Por isso, muitos estudos apontam para a importância do controle da heterogeneidade não observada dos indivíduos nas análises. Existem pesquisas que apontam evidências de que o *mismatch* impacta tanto homens quanto mulheres, mas que estas são penalizadas de maneira mais significativa (SALINAS-JIMÉNEZ, RAHONA-LÓPEZ; MURILLO-HUERTAS, 2013). Outras pesquisas indicam que o *mismatch* educacional horizontal afeta exclusivamente as mulheres (MAHUTEAU *et al.*, 2015), e também existem resultados que mostram que o *mismatch* não afeta nem mulheres nem homens (BOTO-GARCÍA; ESCALONILLA, 2022).

Diante desse breve contexto, apesar do *mismatch* educacional ser um problema atual no Brasil, caracterizado por ser um país em desenvolvimento e apresentar uma tendência crescente do número de indivíduos, especialmente as mulheres, com ensino superior completo, poucos ainda são os estudos que objetivam analisá-lo com foco em questões de gênero. Nesse sentido, o presente ensaio busca auxiliar no preenchimento dessa lacuna e se debruça na investigação da hipótese que as mulheres são mais penalizadas em termos salariais no mercado de trabalho formal por estarem em situação de *mismatch* educacional quando comparadas aos homens.

Para tal, os objetivos dividem-se em duas partes. Primeiramente, pretende-se investigar a proporção de indivíduos egressos da Universidade Federal do Paraná (UFPR) que se encontram em situações de *mismatch* educacional vertical e horizontal. Em um segundo momento, a partir de um painel de dados que permite controlar as características individuais não observáveis e invariantes no tempo, busca-se analisar se os efeitos decorrentes do *mismatch* educacional vertical e horizontal exercem impactos diferentes sobre os salários de mulheres e homens no mercado de trabalho formal.

As investigações do presente ensaio se diferenciam das demais existentes na literatura uma vez que realiza uma análise dinâmica a partir de uma base de dados que integra informações da Relação Anual de Informações Sociais (RAIS) identificada e da Classificação Brasileira de Ocupações (CBO) com informações de egressos da UFPR, possibilitando a identificação das áreas de formação desses indivíduos e suas posteriores ocupações e remunerações no mercado de trabalho formal ao longo do tempo. O período de análise vai do ano de 2003 ao ano de 2021.

O ensaio está estruturado em outras quatro seções, além desta introdução. A seção dois traz a revisão de literatura. A seção três apresenta a base de dados e a metodologia utilizada na condução das análises. Os resultados e discussões estão expostos na seção quatro. Por fim, a seção cinco apresenta as considerações finais do ensaio.

## 4.2 REVISÃO DE LITERATURA

A revisão de literatura a seguir sintetiza os principais trabalhos encontrados sobre o tema. A partir dela, pode-se ressaltar dois aspectos principais. Em primeiro lugar, os resultados sugerem, na grande maioria, efeitos negativos do *mismatch* educacional no salário dos indivíduos. Em segundo lugar, contudo, os estudos apontam para a importância do controle da heterogeneidade não observada dos indivíduos e para a necessidade de cautela na interpretação dos resultados de trabalhos que não tiveram a possibilidade de fazê-lo.

### 4.2.1 *Literatura internacional*

O trabalho de Bauer (2002) aplica o modelo OLS *pooled* a um painel de dados de indivíduos da Alemanha entre os anos de 1984 e 1998. Os resultados encontrados apontam que os trabalhadores sobre-educados ganham menos, e trabalhadores subeducados ganham mais do

que trabalhadores com o mesmo nível de formação educacional, mas que trabalham em ocupações que requerem plenamente seus níveis de educação. Contudo, quando a heterogeneidade não observada dos indivíduos é controlada, os resultados se alteram e, na maioria dos casos, os efeitos do *mismatch* educacional desaparecem completamente. Dessa forma, o autor acende alerta para a importância do controle da heterogeneidade não observada nas análises.

Um outro exemplo de estudo que atenda para essa questão é Tsai (2010). A partir de um painel de dados de 1979 a 2005 para indivíduos dos Estados Unidos, o autor não encontra resultados significativos que apontem para efeitos negativos do *mismatch* educacional sobre os salários quando a heterogeneidade não observada dos indivíduos é levada em consideração.

Por sua vez, Korpi e Tåhlin (2009), analisam o *mismatch* educacional a partir de modelos estáticos e dinâmicos em painel com controle de efeitos fixos aplicados a dados de indivíduos da Suécia entre os anos de 1974 e 2000. Os resultados encontrados pelos autores sugerem que os trabalhadores sobre-educados são penalizados desde cedo por uma taxa inferior de retorno ao ensino do qual não se recuperam.

Já os trabalhos de Allen e van der Velden (2001) e Verhofstadt, De Witte e Omey (2007) analisam a questão do *mismatch* educacional com a satisfação profissional, verificada a partir de um questionário respondido pelos indivíduos analisados. Allen e van der Velden (2001) exploram a relação entre *mismatch* educacional e o *mismatch* de habilidades, caracterizado pela incompatibilidade entre as habilidades adquiridas na formação superior e as requeridas no mercado de trabalho. A partir de um modelo *crosssection* aplicado a uma base de dados de onze países da Europa e do Japão para o ano de 1998, os autores encontram que o *mismatch* educacional afeta significativamente o salário dos indivíduos e apontam para a importância da distinção entre escolaridade e habilidade, uma vez que essa se relaciona fortemente com a satisfação e a procura por trabalhos.

Por sua vez, Verhofstadt, De Witte e Omey (2007), a partir de um modelo Logit aplicado a dados de indivíduos da Bélgica em seus primeiros empregos nos anos de 1999 e 2000, encontra que os trabalhadores mais educados estão mais satisfeitos do que seus colegas menos educados, porque possuem empregos de melhor qualidade. Quando se controla por todas as características do trabalho, o estudo encontra uma relação negativa, com os trabalhadores mais educados relatando menor satisfação no trabalho e sugere que tal resultado pode ser devido ao fato de que trabalhadores com mais educação têm expectativas mais altas em relação aos empregos, particularmente quando se trata do primeiro.

Os estudos de Salinas-Jiménez, Rahona-López e Murillo-Huertas (2013), Boto-García e Escalonilla (2022) e Mahuteau *et al.* (2015) são exemplos que combinam a análise do *mismatch* educacional com questões de gênero.

Salinas-Jiménez, Rahona-López e Murillo-Huertas (2013) analisam os efeitos do *mismatch* educacional sobre os salários de mulheres e homens na Espanha e se o *gap* salarial entre os gêneros é explicado por diferenças nas características produtivas ou por retornos diferentes associados a tais características. A partir da abordagem proposta por Mincer (1974), do modelo ORU (*Over, Required and Under Education*) de Duncan e Hoffman (1981) e da decomposição de Blinder-Oaxaca (Blinder, 1973; Oaxaca 1973), as autoras concluem que o *mismatch* educacional contribui para a diminuição dos salários de ambos os gêneros, mas que as mulheres sofrem penalidades mais significativas. Além disso, ao considerar o nível educacional, o estudo revela que a diferença salarial entre os gêneros é menor para indivíduos com baixos níveis de educação, e a porção desse *gap* relacionada às diferenças nos retornos é mais proeminente nesse grupo. Em contrapartida, a disparidade salarial de gênero é mais acentuada entre trabalhadores altamente educados; contudo, nesse cenário, a maior parte do *gap* salarial é atribuído às diferenças nas características produtivas.

Na mesma linha, Boto-García e Escalonilla (2022) investigam a questão do *mismatch* separadamente para mulheres e homens. A partir de um modelo Probit aplicado a uma base de dados de graduados na Espanha nos anos de 2009 e 2010, os autores não encontramos evidências de que homens e mulheres apresentam riscos diferentes estarem em *mismatch* educacional condicional às suas características observáveis.

Por sua vez, Mahuteau *et al.* (2015), examinam os impactos dos *mismatches* vertical e horizontal nos salários na Austrália, com atenção especial às disparidades de gênero. Utilizando um painel de dados com efeitos fixos, os autores concluem que o *mismatch* vertical resulta em redução salarial tanto para homens quanto para mulheres. Enquanto isso, o *mismatch* horizontal não afeta os salários por hora dos homens, mas causa uma redução significativa para as mulheres. No entanto, são as pessoas que experimentam *mismatches* vertical e horizontal simultaneamente que enfrentam as penalidades mais severas, especialmente as mulheres.

#### 4.2.2 *Literatura nacional*

Em relação ao Brasil, o trabalho de Marioni (2020) analisa o *mismatch* educacional para o mercado de trabalho brasileiro a partir de uma variação do modelo proposto por Verdugo e Verdugo (1989) aplicado a um painel de dados com efeitos fixos e informações da RAIS. A

autora encontra proporções semelhantes em relação ao número de indivíduos em situação de *mismatch* educacional vertical – subqualificados e superqualificados. Além disso, o estudo aponta que os trabalhadores superqualificados (subqualificados) ganham significativamente menos (mais) em comparação àqueles que se não estão em *mismatch* e a que a penalidade para a superqualificação é equivalente ao prêmio pela subqualificação.

A partir da utilização dos dados do Censo e da CBO, Reis (2018) fornece uma medida contínua para representar a distância entre as habilidades adquiridas no ensino superior e aquelas exigidas na ocupação de um indivíduo. Essa medida é baseada na semelhança entre as atividades geralmente realizadas na ocupação de cada indivíduo e aquelas necessárias na ocupação considerada a mais intimamente relacionada ao seu campo de estudo. A partir da sua utilização, o autor encontra que indivíduos que têm ocupações com habilidades semelhantes às adquiridas em suas áreas de estudo ganham mais do que aqueles cujas ocupações e área de formação não se relacionam. Por fim, o autor acende alerta para a importância de considerar descrições mais precisas de trabalhadores e ocupações para caracterizar melhor o *mismatch* entre a ocupação e a área de formação e seus impactos nos salários.

O estudo de Reis e Fortes (2019) analisa os efeitos do *mismatch* educacional vertical especificamente para duas modalidades de educação profissional no Brasil: qualificação profissional e curso técnico de nível médio. Os autores encontram que, entre os indivíduos que estão em ocupações que não demandam a conclusão de um curso de educação profissional, aqueles que a concluíram ganham salários mais altos do que aqueles que não a concluíram. Porém, entre os indivíduos que completaram um curso de educação profissional, aqueles que estão em ocupações que exigem essa qualificação ganham mais do que aqueles que estão em ocupações que não exigem. Por fim, entre os indivíduos que estão em ocupações que exigem a conclusão de um curso de educação profissional, aqueles que de fato a possuem ganham mais do que aqueles que não a possuem. Os resultados são observados tanto para cursos de qualificação profissional quanto para cursos técnicos de nível médio.

Por sua vez, Reis (2020) investiga acerca da probabilidade de estar sobre-educado no Brasil. De acordo com os resultados, entre os indivíduos com o ensino médio completo, aqueles egressos de escolas da rede pública têm maior probabilidade de sobre-educação em comparação com os de escolas privadas. Para os que completaram um curso superior, os resultados mostram o contrário, ou seja, os egressos da rede pública possuem uma probabilidade menor de serem sobre-educados. O autor sugere que instituições que apresentam a qualidade média pior na educação estão associadas a uma probabilidade relativamente maior de formarem indivíduos sobre-educados.

A partir de informações de quatro edições do Censo Demográfico, Reis (2021) analisa o comportamento do *mismatch* educacional vertical entre trabalhadores com ensino superior no Brasil entre 1980 e 2010. Conforme os resultados encontrados pelo autor, a proporção de indivíduos com educação superior classificados como sobre-educados apresentou um aumento significativo entre 1980 e 2000, embora já fosse elevada no período inicial da análise. Em 1980, aproximadamente 25% dos trabalhadores estavam em situação de *mismatch* vertical, aumentando em cerca de 10 pontos percentuais vinte anos depois.

Carvalho e Reis (2023) procuram investigar acerca da sobre-educação no mercado de trabalho brasileiro em períodos mais recentes. A partir de dados da PNAD Contínua dos anos entre 2012 e 2022, os autores encontram como resultados iniciais o crescimento da proporção de trabalhadores sobre-educados, sendo este mais intenso entre os trabalhadores com ensino médio completo. Além disso, os autores apontam que o requerimento de escolaridade das ocupações pouco aumentou ao longo dos últimos dez anos, expondo, dessa forma, que o crescimento da demanda por trabalho qualificado cresceu em ritmo bem inferior à oferta.

Em suma, os estudos apontam para os efeitos negativos do *mismatch* educacional nos resultados do mercado de trabalho. Nesse sentido, o presente ensaio busca contribuir para a literatura nacional a partir da utilização de uma base de dados única com ênfase nas diferenças entre os gêneros.

### 4.3 METODOLOGIA E TRATAMENTO DE DADOS

#### 4.3.1 Base de dados e recorte amostral

A condução das análises do presente ensaio parte da integração de dados da Relação Anual de Informações Sociais Identificada (RAIS Identificada) disponibilizados pelo Ministério da Fazenda, da Classificação Brasileira de Ocupações (CBO) e de dados administrativos da Universidade Federal do Paraná (UFPR) concedidos em parceria com a Agência de Tecnologia da Informação e Comunicação (AGTIC). A integração entre as bases de dados foi possível por meio do CPF de cada discente egresso<sup>34</sup>

---

<sup>34</sup> Os dados identificados só foram disponibilizados após o projeto dessa pesquisa ter sido submetido a um comitê de ética, a fim de garantir a segurança e a proteção das informações particulares dos indivíduos. Número do parecer: 5.867.174.

A partir da combinação entre essas bases foi possível criar uma amostra que permite identificar a área de formação na UFPR de um indivíduo, suas posteriores ocupações e remunerações no mercado de trabalho formal e a escolaridade máxima requerida para desempenhar as funções nessas ocupações. A amostra conta com 14.853 observações (7.879 mulheres e 6.974 homens) que iniciaram e concluíram um curso de ensino superior na UFPR entre os anos de 1986 e 2020.

Devido à falta de algumas informações da base de dados da UFPR, em alguns casos os anos mínimos e máximos foram calculados a partir da data da primeira e da última matéria cursada, respectivamente, apenas para os indivíduos que apresentaram “Formatura” como forma de evasão. Ainda, foram considerados apenas níveis de graduação (15 semanas, 18 semanas, 20 semanais e cursos anuais). Para as pessoas que se formaram em mais de um curso nesse período foi considerado apenas o primeiro. Além disso, foram considerados apenas os indivíduos que possuem ensino superior completo como escolaridade máxima, excluindo aqueles com mestrado e doutorado. Por fim, assim como no segundo ensaio, os cursos foram classificados em áreas do conhecimento conforme a CAPES.

Os indivíduos presentes na base de dados da UFPR foram posteriormente encontrados na base de dados da RAIS identificada, de modo que fosse possível obter informações acerca de suas remunerações e suas ocupações no mercado de trabalho formal. Para fins da construção do painel de dados, foram considerados apenas o primeiro emprego daqueles indivíduos que trocaram de emprego no mesmo ano ou que estavam empregados em mais de um emprego no mesmo ano. Além disso, foram considerados apenas empregos imediatamente posteriores ao ano de formação na UFPR. O período considerado vai do ano de 2003 ao ano de 2021. Por fim, os dados da RAIS foram combinados com as informações sobre a escolaridade requerida de cada ocupação, conforme a Classificação Brasileira de Ocupações (CBO)<sup>35</sup>. O presente ensaio considerou como constante a necessidade educacional de cada ocupação ao longo do período analisado, tal qual Carvalho e Reis (2023).

Uma vez que a análise prévia dos dados apontou para a presença de *outliers* no que diz respeito às remunerações, foi utilizada a técnica de Z-scores para identificação e exclusão dos indivíduos que recebiam salários muito distantes da média geral. Para estimar os Z-scores, foi subtraído da média o valor de cada observação e o resultado foi dividido pelo desvio-padrão. A distribuição resultante apresentou média zero e a distância entre as observações passou a ser em unidades de desvio-padrão. Quanto maior for o valor do Z-score, em valores absolutos,

---

<sup>35</sup> A escolaridade requerida segue o trabalho de Reis (2020), cujos dados detalhados foram gentilmente fornecidos pelos autores.

maior é a distância entre uma determinada observação e a média geral (FIGUEIREDO FILHO *et al.*, 2023). Em conformidade com a literatura existente, o presente ensaio definiu com *outliers* os casos que apresentaram valores superiores a 3 e inferiores a -3 unidades de desvio-padrão (ATKINSON; MULIRA, 1993; LEWIS; BARNETT, 1978; WALFISH, 2006).

#### 4.3.2 Descrição das variáveis e das hipóteses

O Quadro 3 a seguir traz as informações das variáveis que foram utilizadas para investigar a hipótese de que as mulheres em situação de *mismatch* educacional são mais penalizadas em termos de remuneração no mercado de trabalho formal do que os homens na mesma situação.

Quadro 3 – Descrição das variáveis

Variável	Descrição
Gênero	Variável <i>dummy</i> igual a 1 se o indivíduo for mulher e igual a 0 se for homem.
<i>Mismatch</i> vertical	Variável <i>dummy</i> igual a 1 se o indivíduo está em <i>mismatch</i> vertical e igual a zero caso contrário.
<i>Mismatch</i> horizontal	Variável <i>dummy</i> igual a 1 se o indivíduo está em <i>mismatch</i> horizontal e igual a zero caso contrário.
<i>Mismatch</i> Vertical x <i>Mismatch</i> Horizontal	Variável <i>dummy</i> igual a 1 se o indivíduo está em <i>mismatch</i> vertical e horizontal ao mesmo tempo e igual a zero caso contrário.
Áreas	<i>Dummies</i> de áreas de formação (Agrárias; Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde; Sociais Aplicadas*).
Regiões	<i>Dummies</i> de regiões do Brasil (Sul*, Sudeste, Centro-Oeste, Nordeste e Norte).
Horas contratadas	Variável contínua de horas contratadas para cada emprego.
Tempo empregado	Variável contínua para o tempo empregado em cada emprego.

Fonte: Elaboração própria.

\* Variáveis de referência.

#### 4.3.3 *Mismatches* vertical e horizontal

Os *mismatches* educacionais podem ser classificados em vertical e horizontal. O *mismatch* vertical acontece quando um indivíduo está trabalhando em uma ocupação que requer

mais (*undereducation* – subeducação) ou menos (*overeducation* – sobre-educação) anos de estudo do que efetivamente possui. O *mismatch* horizontal, por sua vez, ocorre quando o indivíduo está empregado em uma ocupação não relacionada com a sua respectiva área de formação acadêmica (PARK, 2021).

Para a classificação do *mismatch* vertical foram utilizadas as informações referentes ao nível de escolaridade considerado necessário para desempenhar as atividades de cada ocupação, conforme a Classificação Brasileira de Ocupações (CBO), ao nível de quatro dígitos. Uma vez que todos os indivíduos da base de dados utilizada nesse ensaio possuem nível superior completo e o máximo de escolaridade que a CBO utiliza nas classificações também se refere ao nível superior completo (15 anos de educação formal), apenas os sobre-educados foram analisados, ou seja, todos aqueles que possuem um número de anos de estudo maior do que o máximo considerado necessário para a sua ocupação, conforme definido em Carvalho e Reis (2023).

A estratégia de classificação do *mismatch* horizontal, por sua vez, foi baseada no cálculo da moda das ocupações em cada área conforme Mahuteau *et al.* (2015). Em outras palavras, foi identificada para cada área a ocupação que concentra a maior parte dos indivíduos. Assim, foram considerados como horizontalmente em *mismatch* todos aqueles empregados em uma ocupação diferente daquela que concentra a maior parte dos indivíduos de sua respectiva área (Agrárias, Biológicas; Engenharias; Exatas e Terra; Humanas; Linguística, Letras e Artes; Saúde; Sociais Aplicadas).

#### 4.3.4 Estratégia de identificação e modelo empírico

A especificação utilizada no presente ensaio é uma variação do modelo proposto por Verdugo e Verdugo (1989), com a diferença que, ao invés de definir os trabalhadores como sobre-educados a partir do valor do desvio-padrão acima da média do nível de escolaridade dos indivíduos de cada ocupação, utiliza a comparação entre o nível de escolaridade dos indivíduos e aquele requerido como o máximo para trabalhar em determinada ocupação, conforme a CBO. Dessa forma, para estimar o efeito dos *mismatches* vertical e horizontal sobre os salários, são adicionadas *dummies* no modelo *pooled* OLS conforme a Equação 14 a seguir:

$$\ln W_{it} = \sum_{k=1}^n \beta_k MV_{it} + \beta_2 MH_{it} + \beta_3 MVH_{it} + \beta_4 \Gamma_{it} + \varepsilon_{it} \quad (14)$$

Em que:  $\ln W_{it}$ : logaritmo natural dos salários em reais do indivíduo  $i$  no tempo  $t$ ;  $MV_{it}$ : variável *dummy* igual a 1 se o indivíduo  $i$  está em *mismatch* vertical no tempo  $t$ ;  $MH_{it}$ : variável *dummy* igual a 1 se o indivíduo  $i$  está em *mismatch* horizontal no tempo  $t$ ;  $MVH_{it}$ : variável *dummy* igual a 1 se o indivíduo está em *mismatch* vertical e horizontal no tempo  $t$ ;  $\Gamma_{it}$ : vetor de variáveis observáveis;  $\varepsilon_{it}$ : termo de erro.

Conforme Ashenfelter, Harmon e Oosterbeek (1999), a escolaridade dos indivíduos pode estar correlacionada com o termo de erro causando o problema da endogeneidade e, conseqüentemente, estimações inconsistentes. Dessa forma, para mitigar o problema da heterogeneidade não observada entre os indivíduos, o presente ensaio utiliza o painel de dados com efeitos fixos conforme a Equação 15 a seguir.

$$\ln W_{it} = \sum_{k=1}^n \beta_k MV_{it} + \beta_2 MH_{it} + \beta_3 MVH_{it} + \beta_4 \Gamma_{it} + c_i + \varepsilon_{it} \quad (15)$$

Em que  $c_i$ : termo de características não observáveis.

A principal vantagem das estimações utilizando efeitos fixos é a possibilidade de eliminação do efeito das características individuais que não variam com o tempo. As características não observáveis do indivíduo podem afetar tanto os salários quanto o *mismatch*. Entre essas características estão as habilidades não observáveis do indivíduo que, se consideradas fixas ao longo do tempo, não influenciam o resultado obtido nos salários e/ou no *mismatch*. Após as estimações dos modelos *pooled*, de efeitos aleatórios e de efeitos fixos, foi realizado o Teste de Hausman que indicou que a heterogeneidade não observada possui um comportamento fixo em todos os casos, fazendo com que os resultados obtidos a partir do modelo com efeitos fixos sejam preferíveis aos obtidos a partir do modelo com efeitos aleatórios.

#### 4.4 RESULTADOS E DISCUSSÃO

A seção de resultados e discussão a seguir traz as estatísticas descritivas, os resultados das estimações do painel com efeitos fixos e a análise de heterogeneidade.

## 4.4.1 Estatísticas descritivas

A Tabela 35 a seguir contém as estatísticas descritivas e os testes t das amostras separadas entre mulheres e homens.

Tabela 35 - Estatísticas descritivas e teste t

Variáveis	Mulheres		Homens		Teste t
	Média	Desvio-Padrão	Média	Desvio-Padrão	
Salário	6.522	7.151	8.168	5.945	41,202***
<i>Mismatch</i> vertical	0,3455	0,4795	0,3582	0,4755	4,414***
<i>Mismatch</i> horizontal	0,8193	0,3795	0,8256	0,3848	2,721***
<i>Mismatch</i> vertical e horizontal	0,2754	0,4467	0,2919	0,4546	6,0826***
Idade	35,4785	8,4871	35,8237	8,6602	6,672***
Sul	0,9180	0,2880	0,9087	0,2744	-5,455***
Sudeste	0,0596	0,2364	0,0594	0,2367	-0.114
Centro-Oeste	0,0160	0,1499	0,0230	0,1256	8,282***
Nordeste	0,0039	0,0660	0,0044	0,0627	1,0980
Norte	0,0024	0,0670	0,0045	0,0494	5,740***
Horas contratadas	34,6896	10,3302	35,9537	10,6977	19,933***
Tempo empregado(a)	58,1967	67,7217	59,3674	66,6510	2,884***
Agrárias	0,0485	0,2565	0,0708	0,2147	15,562***
Biológicas	0,0529	0,1732	0,0309	0,2238	-18,326***
Engenharias	0,0634	0,3956	0,1943	0,2436	65,0340***
Exatas e Terra	0,1013	0,3895	0,1865	0,3017	40,1830***
Humanas	0,1821	0,2888	0,0918	0,3859	-44,277***
Linguística, Letras e Artes	0,0700	0,1922	0,0384	0,2552	-23,378***
Saúde	0,2576	0,3406	0,1339	0,4373	-52,700***
Sociais Aplicadas	0,2243	0,4349	0,2533	0,4171	11,247***
<b>Total de observações</b>	<b>7879</b>		<b>6974</b>		

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Conforme os valores reportados na Tabela 35, pode-se observar que as mulheres ganham aproximadamente 80% do salário dos homens. Além disso, a maior parte dos indivíduos estão concentrados em ocupações na Região Sul e apresentam valores semelhantes em relação à média de idade (34,5 anos para as mulheres e 35,8 anos para os homens); à média de horas contratadas no mercado de trabalho (34,7 horas para as mulheres e 36 horas para os homens) e à média de tempo empregado(a) (58 meses para as mulheres e 59 meses para os homens).

A área da Saúde é a que concentra a maior parte das mulheres (26%), seguida pelas áreas de Sociais Aplicadas (22%), Humanas (18%), Exatas e Terra (10%), Linguística, Letras e Artes (7%), Engenharias (6%), Biológicas (5%), e Agrárias (5%). Os homens, por sua vez,

estão mais concentrados na área de Sociais Aplicadas (25%), seguida pelas áreas de Engenharias (19%), Exatas e Terra (19%), Saúde (13%), Humanas (9%), Agrárias (7%), Linguística, Letras e Artes (4%) e Biológicas (3%). Dessa forma, pode-se perceber que as áreas com as maiores diferenças entre a concentração de mulheres e homens são a Saúde e as Humanas em favor das mulheres e as áreas das Engenharias e Exatas e Terra em favor dos homens. Vale ressaltar que esses valores apontam para a sub-representação de mulheres nas áreas STEM (ciência, tecnologia, engenharia e matemática), em conformidade com a literatura existente (BARONE E ASSIRELLI, 2018; SAHOO E KLASSEN, 2018; GONZÁLEZ-PÉREZ, CABO, SÁINZ, 2020; BUCKLEY, FARRELL, TYNDALL, 2021).

Em relação aos *mismatches*, pode-se notar que a proporção de indivíduos com *mismatch* vertical e horizontal é semelhante entre mulheres e homens: aproximadamente 35% das mulheres e 36% dos homens estão em *mismatch* vertical e 82% das mulheres e 83% dos homens em *mismatch* horizontal. O mesmo acontece para os indivíduos que estão tanto em *mismatch* vertical quanto em horizontal: 28% das mulheres e 29% dos homens. É possível notar que a tendência de os homens se tornarem mais sobre-educados do que as mulheres a partir de 2010, conforme exposto em Reis (2021), também se reflete nos dados analisados neste ensaio.

#### 4.4.2 Resultados das estimações do painel de dados

A Tabela 36 a seguir apresenta os resultados das estimações do painel de dados. Primeiramente, foram estimados os coeficientes considerando todos os indivíduos. Depois, as estimações foram feitas separadamente para mulheres e homens. As colunas 1 consideram apenas a variável de *mismatch* vertical, as colunas 2 consideram apenas o horizontal, e as colunas 3 consideram as variáveis dos dois tipos de *mismatch* concomitantemente com uma variável de interação entre os *mismatches*. Todas as especificações comparam os indivíduos em *mismatch* vertical (sobre-educados) e/ou *mismatch* horizontal àqueles que possuem a mesma quantidade de anos de estudo formal, mas que estão alocados em ocupações mais adequadas aos seus níveis de escolaridade (vertical) e às suas áreas de formação (horizontal). O período das análises vai do ano 2009 ao ano de 2021.

Uma vez que o Teste de Hausman apontou para os modelos de efeitos fixos como os mais adequados em todos os casos, apenas estes estão reportados na Tabela 36 a seguir. Os resultados das estimações com os modelos *OLS pooled* e de efeitos variáveis podem ser encontrados nas Tabelas A13 a A21 nos Apêndices.

Tabela 36 - Resultados das estimações do painel de dados com efeitos fixos

Variáveis	Todos			Mulheres			Homens		
	1	2	3	1	2	3	1	2	3
<i>Mismatch</i> vertical	-0.142*** (0.006)		-0.263*** (0.013)	-0.155*** (0.008)		-0.277*** (0.018)	-0.127*** (0.009)		-0.250*** (0.020)
<i>Mismatch</i> horizontal		-0.029*** (0.007)	-0.105*** (0.009)		-0.029*** (0.009)	-0.110*** (0.012)		-0.030*** (0.010)	-0.101*** (0.013)
<i>Mismatch</i> vertical x <i>Mismatch</i> horizontal			0.134*** (0.014)			0.135*** (0.019)			0.136*** (0.021)
Controles	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observações	108.799	108.799	108.799	57.980	57.980	57.980	50.819	50.819	50.819
R <sup>2</sup>	0.101	0.096	0.102	0.110	0.104	0.112	0.091	0.087	0.093
R <sup>2</sup> Ajustado	-0.041	-0.047	-0.039	-0.029	-0.037	-0.028	-0.053	-0.058	-0.051

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Notas: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão. Estatística F significativa a 1% em todos os casos.

A principal vantagem das estimações utilizando efeitos fixos é a possibilidade de eliminação do efeito das características individuais que não variam com o tempo. Dessa forma, os coeficientes podem ser interpretados como o efeito médio nos salários devido a mudanças no status de *mismatch* vertical e/ou horizontal dos indivíduos ao longo do tempo (MARIONI, 2020).

A partir dos resultados reportados na Tabela 36, pode-se notar que estar em *mismatch* vertical e horizontal afeta negativamente os salários em todos os casos, em consonância com os resultados encontrados por Bauer (2002), Korpi e Tåhlin (2009), Marioni (2020), Mahuteau *et al.* (2015). Considerando todos os indivíduos, as colunas 1 e 2 mostram que o *mismatch* vertical diminui o salário em aproximadamente 14% enquanto o efeito do horizontal é uma diminuição de cerca de 3%. Já os indivíduos que estão tanto em *mismatch* vertical quando em horizontal têm seus salários afetados negativamente em cerca de 13%. Vale ressaltar que, conforme Bauer (2002), os resultados estimados por efeitos fixos são significativamente menores daqueles estimados pelo modelo *pooled* (Tabelas A13 a A21 nos Apêndices), sustentando a importância do controle da heterogeneidade não observada dos indivíduos e a cautela ao se interpretar os resultados de estimações que não tiveram a possibilidade de fazê-lo.

Analisando separadamente para mulheres e homens, é possível observar que os resultados são semelhantes, tanto na direção do sinal quanto em termos de magnitude: enquanto o *mismatch* vertical afeta negativamente o salário dos homens em cerca de 13% e o horizontal 3%, o efeito é ligeiramente pior para as mulheres: aproximadamente 15% e 3%, respectivamente, em conformidade com os resultados encontrados por Mahuteau *et al.* (2015).

#### 4.4.3 Análise de heterogeneidade

De acordo com Mahuteau *et al.* (2015), é possível que o *mismatch* educacional horizontal seja mais provável entre os indivíduos que concluíram uma formação superior em uma área que fornece habilidades mais gerais e menos provável entre aqueles de áreas que oferecem habilidades mais específicas para uma ocupação. Isso porque, diferentemente do *mismatch* vertical, o *mismatch* horizontal é mais difícil preservar parte do capital humano específico que os indivíduos obtêm de determinadas áreas. No entanto, os efeitos salariais podem ser maiores no último caso para aqueles que não conseguem encontrar uma correspondência adequada. Sendo assim, na tentativa de investigar acerca das heterogeneidades, os coeficientes foram estimados separadamente para cada uma das oito áreas, conforme Tabela 37 a seguir.

Tabela 37 - Resultados das estimações por gênero e área do conhecimento

Áreas	Casos	Mismatch vertical	Mismatch horizontal	Mismatch vertical e horizontal
Agrárias	Todos	-0.233*** (0.027)	-0.008 (0.025)	-0.080 (0.057)
	Mulheres	-0.247*** (0.041)	0.007 (0.035)	-0.172*** (0.036)
	Homens	-0.221*** (0.036)	-0.022 (0.035)	-0.186*** (0.033)
Biológicas	Todos	-0.019 (0.031)	0.024 (0.030)	0.025 (0.029)
	Mulheres	-0.025 (0.037)	0.007 (0.039)	0.008 (0.036)
	Homens	0.039 (0.059)	0.045 (0.050)	0.083* (0.049)
Engenharias	Todos	-0.148*** (0.014)	-0.171*** (0.019)	-0.131*** (0.014)
	Mulheres	-0.184*** (0.029)	-0.196*** (0.037)	-0.175*** (0.028)
	Homens	-0.134*** (0.016)	-0.157*** (0.022)	-0.113*** (0.016)
Exatas e Terra	Todos	-0.124*** (0.015)	0.027 (0.018)	-0.124*** (0.015)
	Mulheres	-0.114*** (0.025)	0.011 (0.028)	-0.114*** (0.025)
	Homens	-0.134*** (0.019)	0.041* (0.023)	-0.134*** (0.019)
Humanas	Todos	-0.161*** (0.017)	-0.090*** (0.018)	-0.144*** (0.017)
	Mulheres	-0.215*** (0.020)	-0.121*** (0.020)	-0.184*** (0.020)
	Homens	-0.036 (0.032)	0.031 (0.041)	-0.052* (0.032)
Linguística, Letras e Artes	Todos	-0.126*** (0.029)	-0.079*** (0.023)	-0.082*** (0.025)
	Mulheres	-0.161*** (0.034)	-0.037 (0.028)	-0.104*** (0.031)
	Homens	-0.036 (0.052)	-0.156*** (0.040)	-0.037 (0.042)
Saúde	Todos	-0.166*** (0.016)	-0.353*** (0.022)	-0.166*** (0.016)
	Mulheres	-0.209*** (0.018)	-0.325*** (0.032)	-0.209*** (0.018)
	Homens	-0.054 (0.033)	-0.390*** (0.032)	-0.054 (0.033)
Sociais Aplicadas	Todos	-0.112*** (0.011)	0.121*** (0.012)	-0.024** (0.011)
	Mulheres	-0.086*** (0.014)	0.106*** (0.016)	-0.003 (0.015)
	Homens	-0.143*** (0.016)	0.140*** (0.019)	-0.047*** (0.017)

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Notas: \*, \*\*, \*\*\* Significante a 10%, 5% e 1%, respectivamente. Em entre parênteses estão os erros-padrão.

Em relação aos efeitos nos salários, pode-se observar que, para todos os indivíduos, as áreas caracterizadas por habilidades mais específicas, como a Saúde e Engenharias, apresentam efeitos negativos significativamente mais acentuados no que diz respeito ao *mismatch* horizontal em comparação às demais áreas de conhecimentos menos específicos, como Humanas, Linguística; Letras e Artes e Sociais Aplicadas, por exemplo, em conformidade com Mahuteau *et al.* (2015). O mesmo padrão, contudo, não aparece quando se analisam separadamente cada gênero. Enquanto as mulheres são mais penalizadas do que os homens por estarem em *mismatch* horizontal nas áreas de Engenharias e Humanas, os homens são mais penalizados nas áreas de Linguística Letras e Artes, Saúde e Sociais Aplicadas.

Em relação àqueles que estão em situação de *mismatch* vertical e *mismatch* horizontal, as mulheres são mais penalizadas na maior parte dos casos: Agrárias, Humanas, Engenharias, Linguística, Letras e Artes e Saúde. Os homens, por sua vez, são mais penalizados nas áreas de Exatas e Terra e Sociais Aplicadas.

#### 4.5 CONSIDERAÇÕES FINAIS

Os últimos anos foram marcados por um aumento da média de anos de estudo e da participação de indivíduos com ensino superior na força de trabalho no Brasil. Entretanto, quando a demanda por mão de obra qualificada não acompanha o aumento da sua oferta, surge o problema do *mismatch* educacional, ou seja, os indivíduos estão desempenhando funções em ocupações que não requerem o nível de escolaridade que possuem (*mismatch* vertical) ou, então, estão em ocupações não relacionadas com as suas formações acadêmicas (*mismatch* horizontal).

Apesar do *mismatch* educacional ser um problema atual no Brasil, caracterizado por ser um país em desenvolvimento e apresentar uma tendência crescente do número de indivíduos, especialmente as mulheres, com ensino superior completo, poucos ainda são os estudos que objetivam analisá-lo com foco em questões de gênero. Nesse sentido, o presente ensaio buscou auxiliar no preenchimento dessa lacuna e se debruçou na investigação da hipótese que as mulheres são mais penalizadas em termos salariais no mercado de trabalho formal por estarem em situação de *mismatch* educacional quando comparadas aos homens.

A partir de um painel de dados que integrou informações de egressos da Universidade Federal do Paraná, como área de formação, com suas informações sobre ocupação e

remuneração no mercado de trabalho da base de dados da RAIS identificada, os resultados encontrados apontam que os *mismatches* vertical e horizontal afetam negativamente o salário dos indivíduos.

Quando os gêneros são analisados separadamente, observa-se que os resultados são semelhantes, tanto na direção do sinal quanto em termos de magnitude, com as mulheres sendo ligeiramente mais penalizadas em comparação aos homens, tanto no *mismatch* vertical quanto no horizontal, corroborando os resultados encontrados por Mahuteau *et al.* (2015) e Salinas-Jiménez, Rahona-López, Murillo-Huertas, (2013).

Por fim, quando as análises são feitas separadamente para cada área, os resultados indicam que áreas caracterizadas por habilidades mais específicas, como a Saúde e Engenharias, apresentam efeitos negativos significativamente mais acentuados no que diz respeito ao *mismatch* horizontal em comparação às demais áreas de habilidades menos específicos, como Humanas, Linguística; Letras e Artes e Sociais Aplicadas. Porém, esse padrão não se repete quando as análises são feitas separadamente entre mulheres e homens. Essa é uma das contribuições centrais do presente ensaio, posto que os resultados separados por áreas do conhecimento de indivíduos que foram acompanhados ao longo do tempo não foram previamente encontrados na literatura. Ademais, vale ressaltar que outras alternativas de análises, principalmente em relação ao cálculo *mismatch* horizontal, poderiam tornar os resultados mais robustos. Uma delas diz respeito à autoavaliação, na qual os próprios indivíduos determinam, a partir de uma análise subjetiva, se exercem funções no mercado de trabalho relacionadas as suas formações acadêmicas.

## 5 CONCLUSÃO GERAL

Como o papel social destinado para cada sexo interfere nas escolhas – ou imposições – educacionais e ocupacionais que, por sua vez, impactam na situação socioeconômica das mulheres?

Há 75 anos, Simone de Beauvoir (2019 [1949]) descreveu a mulher como “o segundo sexo”. Segundo ela, a mulher é tudo o que o homem não é, mas é também tudo aquilo que o homem depende para poder ser quem é. Marçal (2017) foi um pouco além e apontou que, assim como existe um “segundo sexo” aos moldes de Beauvoir, existe uma “segunda economia”. O trabalho tradicionalmente executado pelos homens é o que conta e é o que define a visão do mundo econômico. O trabalho executado pela mulher é o “outro”, é tudo o que o homem não faz, mas de que depende para poder fazer o que faz.

Em 1949, Beauvoir afirmou que as mulheres eram, em conjunto, inferiores aos homens, pois as circunstâncias lhes ofereciam menos possibilidades (BEAUVOIR, 2019 [1949]) e já naquela época a autora acendia um alerta sobre as implicações da perpetuação dessa realidade. Infelizmente, mais de setenta anos depois e de diversos avanços, mulheres e homens ainda não gozam de igualdade de condições, sejam elas econômicas, sociais ou políticas.

Nesse sentido, apesar de estar sendo observada nos últimos anos uma elevação da participação feminina na força de trabalho e uma diminuição nas diferenças salariais entre os gêneros, ainda existem diversos fatores, intimamente relacionados às normas sociais e aos estereótipos de gênero, que continuam contribuindo para que as mulheres permaneçam em situação de desvantagem no que diz respeito às suas escolhas – ou imposições – profissionais.

Essas normas sociais atribuem o trabalho invisível do cuidado às mulheres e muitas vezes também agem para que no mercado de trabalho elas se concentrem em profissões que são entendidas como extensão da casa. Nesse sentido, os resultados encontrados no primeiro ensaio apontam que, dentre as ocupações caracterizadas como carreiras intelectuais e científicas pela PNAD Contínua, aquelas que mais concentram mulheres e que pior remuneram estão relacionadas ao ensino, principalmente nos níveis básico e fundamental. É importante ressaltar, como aponta Levanon, England e Allison (2009), que não são as mulheres que procuram piores empregos, e sim que as áreas perdem valor quando concentram um grande número da parcela feminina, indicando que é o trabalho mal pago que procura as mulheres e não o contrário. E isso se agrava ainda mais quando a raça entra na análise e o teto de vidro se torna teto de

concreto, uma vez que as mulheres pretas e pardas são duplamente penalizadas: por seu gênero e por sua cor.

Os problemas da segregação entre os gêneros podem ser vistos também no aspecto da formação educacional, na medida em que ainda existem mais mulheres do que homens em cursos relacionados às humanidades enquanto as áreas das exatas são mais dominadas pelo público masculino. Entretanto, independentemente de áreas, estudos apontam que as mulheres apresentam melhores desempenhos acadêmicos em relação aos homens desde níveis iniciais e intermediários de educação (DIPRETE E JENNINGS, 2012; ENTWISLE, ALEXANDER E OLSON, 2007; PERKINS et al., 2004) até níveis superiores de formação acadêmica (DIPRETE E BUCHMANN 2013). Nesse ensejo, o segundo ensaio buscou investigar a seguinte questão: as mulheres enfrentam resultados relativamente desfavoráveis no mercado de trabalho apesar de apresentarem melhores desempenhos acadêmicos? Os resultados encontrados sugerem evidências estatísticas de que as mulheres com melhores performances de fato ganham maiores retornos salariais no mercado de trabalho. Porém, é importante ressaltar que a magnitude dos valores encontrados é pequena e pode variar bastante a depender das áreas de formação em questão. Em qualquer um dos casos, o efeito positivo do desempenho é insuficiente para eliminar o diferencial de gêneros nos salários.

Além disso, apesar da importância da conclusão do ensino superior para um indivíduo, o terceiro ensaio analisou problemas que podem surgir depois da formação superior, na medida em que quando a demanda por mão de obra qualificada não acompanha o aumento da sua oferta, surge o problema do *mismatch* educacional. Os resultados sugerem que esse também é um problema que penaliza as mulheres de forma mais acentuada do que os homens corroborando os resultados encontrados por Mahuteau *et al.* (2015) e Salinas-Jiménez, Rahona-López, Murillo-Huertas, (2013).

Nesse sentido, os resultados encontrados nos três ensaios jogam luz sobre o fato que, apesar dos avanços recentes, muito ainda precisa ser feito para alcançar a igualdade entre mulheres e homens no mercado de trabalho. De acordo com a Organização Internacional do Trabalho (2019), a desigualdade entre os gêneros no mercado de trabalho se deve, principalmente, à dupla jornada das mulheres, na medida em que são responsabilizadas pelos cuidados domésticos e da família. Calcula-se que sejam necessários ainda 209 anos para reverter essa situação.

Dessa forma, políticas públicas e ações afirmativas que visem neutralizar as desigualdades entre mulheres e homens são de fundamental importância: concursos públicos com cotas para as mulheres, incentivos à ocupação de cargos em comissão, cargos de

liderança<sup>36</sup> e cargos políticos<sup>37</sup>, flexibilização da jornada de trabalho, incentivos para empresas investirem em diversidade de gênero e raça, políticas que combatam o assédio moral e sexual, melhor divisão das licenças maternidade e paternidade<sup>38</sup>, oferta suficiente de creches, e igualdade salarial para mulheres e homens que ocupem os mesmos cargos<sup>39</sup>.

Contudo, é importante ressaltar que as políticas públicas e as ações afirmativas, apesar de extremamente importantes e necessárias, estão trabalhando em um problema cujas raízes se encontram nas estruturas da sociedade. De acordo com Claudia Goldin (2021, p. 5) “*The gender earnings gap is a result of the career gap; the career gap is at the root of couple inequity*”. A desigualdade entre casais, por sua vez, se origina no fato de que desde crianças, meninas e meninos são tratados de formas diferentes e discriminatórias por seus pais ou responsáveis, que, por sua vez, muito provavelmente também foram tratados de forma discriminatória por seus pais ou responsáveis.

É dessa forma, portanto, que o papel social destinado para cada sexo interfere nas escolhas – ou imposições – educacionais e ocupacionais que, por sua vez, impactam na situação socioeconômica das mulheres. Cabe a nós continuarmos a desafiar e a quebrar esses padrões para que, um dia, conforme idealizou Beauvoir (2022 [1960]), nada nos defina, nada nos sujeite, que a liberdade possa ser a nossa própria substância.

---

<sup>36</sup> De acordo com dados do IBGE (2021), em 2019 apenas 37,4% dos cargos de liderança no Brasil eram ocupados por mulheres.

<sup>37</sup> De acordo com a Emenda Constitucional nº 117 de 2022: “Altera o art. 17 da Constituição Federal para impor aos partidos políticos a aplicação de recursos do fundo partidário na promoção e difusão da participação política das mulheres, bem como a aplicação de recursos desse fundo e do Fundo Especial de Financiamento de Campanha e a divisão do tempo de propaganda gratuita no rádio e na televisão no percentual mínimo de 30% (trinta por cento) para candidaturas femininas” (CÂMARA DOS DEPUTADOS, 2022).

<sup>38</sup> No Brasil, a Constituição Federal de 1988 garante às mães trabalhadoras urbanas e rurais licença maternidade de 120 dias, sem prejuízo do emprego e do salário, para filhos biológicos ou adotados. Para os pais, a licença paternidade, que ainda não é regulamentada por lei, é de apenas 5 dias (BRASIL, 1988 [2016]).

<sup>39</sup> A Lei 14.611/2023 garantiu a igualdade salarial entre mulheres e homens que exercem trabalho de igual valor ou atuam na mesma função no Brasil.

## REFERÊNCIAS

- AGRAWAL, T. Gender segregation and wage differentials in India: the role of educational attainment and occupational choices. **International Journal of Manpower**, 42(1): 1–20, 2021.
- ALLEN, J.; VAN DER WELDEN, R. Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. **Oxford Economic Papers**, 53(3): 434–452, 2001.
- ARAÚJO, V. F.; RIBEIRO, E. P. Diferenciais de salários por gênero no Brasil: uma análise regional. Porto Alegre: PPGE/UFRGS, **Texto para Discussão n° 11**, 2001.
- ARCIDIACONO, P.; BAYER, P.; HIZMO, A. Beyond signaling and human capital: education and the revelation of ability. **American Economic Journal: Applied Economics**, 2(4): 76–107, 2010.
- ARORA, D.; BRAUNSTEIN, E.; SEGUINO, S. A macro analysis of gender segregation and job quality in Latin America. **World Development**, 164, 106153, 2023.
- ARROW, K. The theory of discrimination. IN: ASHENFELTER, O.; REES, A. **Discrimination in labor markets**. Princeton: Princeton University Press, 1973.
- ASHENFELTER, O.; HARMON, C.; OOSTERBEEK, H. A review of estimates of the schooling/earnings relationship, with tests for publication bias. **Labour Economics**, 6(4): 453–470, 1999.
- ATKINSON, A.; MULIRA, H.-M. The stalactite plot for the detection of multivariate outliers. **Statistics and Computing**, 3(1): 27–35, 1993.
- AUSTIN, P. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. **Multivariate Behavioral Research**, 46(3): 399–424, 2011.
- AUSTIN, P.; STUART, E. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. **Statistics in Medicine**, 34(28): 3661–3679, 2015
- AUTOR, D.; FIGLIO, D.; KARBOWNIK, K.; ROTH, J.; WASSERMAN, M. School quality and the gender gap in educational achievement. **The American Economic Review**, 106(5): 289–295, 2016.
- BARONE, C.; ASSIRELLI, G. Gender segregation in higher education: an empirical test of seven explanations. **Higher Education**, 79: 55–78, 2020.
- BAUER, T. Educational mismatch and wages: a panel analysis. **Economics of Education Review**, 21: 221–229, 2002.
- BEAUVOIR, S. De. **O segundo sexo: fatos e mitos**. 5ª Edição. Rio de Janeiro: Nova Fronteira, 2019 [1949].

BEAUVOIR, S. De. **A força da idade**. 8ª Edição. Rio de Janeiro: Nova Fronteira, 2022 [1960].

BECKER, G. **Human capital**. New York: National Bureau of Economic Research, 1964.

BECKER, G. **The economics of discrimination**. Chicago: The University of Chicago Press, 1971.

BECKER, G.; EWALD, F.; HARCOURT, B. “Becker on Ewald on Foucault on Becker”: American neoliberalism and Michel Foucault’s 1979 *Birth of Biopolitics* lectures. **Coase-Sandor Working Paper Series in Law and Economics**, University of Chicago, 2012.

BECKER, G. S.; HUBBARD, W. H.; MURPHY, K. M. Explaining the worldwide boom in higher education of women. **Journal of Human Capital**, 4(3): 203–241, 2010.

BLAIR, P.; CHUNG, B. Job market signaling through occupational licensing. **NBER Working Paper Series** 24791, 2020. Disponível em: <<http://www.nber.org/papers/w24791>>.

BLAU, F. D.; BRUMMUND, P.; LIU, A. Y. Trends in Occupational Segregation by Gender 1970–2009: Adjusting for the Impact of Changes in the Occupational Coding System. **Demography**, 50(1): 471–492, 2013.

BLINDER, A. Wage discrimination: reduced form and structural estimates. **Journal of Human Resources**, 8(4): 436–455, 1973.

BORROWMAN, M.; KLASSEN, S. Drivers of Gendered Sectoral and Occupational Segregation in Developing Countries. **Feminist Economics**, 26(3): 1–33, 2019.

BOTASSIO, D. C.; VAZ, D. V. Segregação ocupacional por sexo no mercado de trabalho brasileiro: uma análise de decomposição para o período 2004-2015. **Revista Brasileira de Estudos Populacionais**, 37(1): 1–30, 2020.

BOTO-GARCÍA, D.; ESCALONILLA, M. University education, mismatched jobs: are there gender differences in the drivers of overeducation? **Economia Política**, 39: 861–902, 2022.

BRASIL. [Constituição (1988)]. Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal, 2016.

BRASIL. **Lei Nº 11.770, de 09 de setembro de 2008**. Cria o Programa Empresa Cidadã, destinado à prorrogação da licença-maternidade mediante concessão de incentivo fiscal, e altera a Lei nº 8.212, de 24 de julho de 1991. Brasília, DF: Diário Oficial da União, 2008.

BRASIL. **Lei Nº 14.611, de 3 de julho de 2023**. Dispõe sobre a igualdade salarial e de critérios remuneratórios entre mulheres e homens; e altera a Consolidação das Leis do Trabalho, aprovada pelo Decreto-Lei nº 5.452, de 1º de maio de 1943. Brasília, DF: Diário Oficial da União, 2023.

BRASIL. Senado Federal. **Proposta de Emenda à Constituição nº 18, de 2021**. Brasília, DF: Senado Federal, 2021.

BRASIL. Câmara dos Deputados. **Proposta de Emenda à Constituição nº 117, de 2022**. Brasília, DF: Câmara dos Deputados, 2022.

BROPHY, J. T.; KEITH, M. M.; HURLEY, M. Assaulted and unheard: violence against healthcare staff. **New Solutions**, 27(4): 581–606, 2018.

BUCHMANN, C.; DIPRETE, T. The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement. **American Sociological Review**, 71(4): 515–541, 2006.

BUCKLEY, C.; FARRELL, L.; TYNDALL, I. Brief stories of successful female role models in science help counter gender stereotypes regarding intellectual ability among young girls: A pilot study. **Early Education and Development**, 33(4): 1–12, 2021.

BÜTIKOFER, A.; JENSEN, S.; SALVANES, K. The role of parenthood on the gender gap among top earners. **European Economic Review**, 109(C):103–123, 2018.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. **Journal of Economic Surveys**, 22(1): 31–72, 2008.

CARBONARO, W.; ELLISON, B.; COVAY, E. Gender inequalities in the college pipeline. **Social Science Research**, 40(1): 120–135, 2011.

CARVALHO, S.; REIS, M. Evolução da sobre-educação no mercado de trabalho no Brasil entre 2012 e 2022: primeiros resultados. **IPEA – Nota Técnica, Mercado de Trabalho 75**, 2023.

CASTAGNETTI, C.; ROSTI, L. Effort allocation in tournaments: the effect of gender on academic performance in Italian universities. **Economic of Education Review**, 28(1): 357-369, 2009.

CHRISTOFIDES, L. N.; POLYCARPOU, A.; VRACHIMIS, K. Gender wage gaps, ‘sticky floors’ and ‘glass ceilings’ in Europe. **Labour Economics**, 21(1): 86–102, 2013.

CORRELL, S.; WEISSHAAR, K.; WYNN, A.; WEHNER, J. Inside the black box of organizational life: the gendered language of performance assessment. **American Sociological Review**, 85(6): 1022–1050, 2020.

CUDDY, A.; FISKE, S.; GLICK, P. Warmth and competence as universal dimensions of social perception: the stereotype content model and the BIAS map. **Advances in Experimental Social Psychology**, 40: 61–149, 2008.

DEKHTYAR, S.; WEBER, D.; HELGERTZ, J.; HERLITZ, A. Sex differences in academic strengths contribute to gender segregation in education and occupation: A longitudinal examination of 167,776 individuals. **Intelligence**, 67: 84–92, 2018.

DIPRETE, T.; BUCHMANN, C. Gender-specific trends in the value of education and the emerging gender gap in college completion. **Demography**, 43(1): 1–24, 2006.

DIPRETE, T.; BUCHMANN, C. The rise of women: the growing gender gap in education and what it means for American schools. **Russell Sage Foundation Executive Summary**, 2013.

DIPRETE, T.; JENNINGS, J. Social and behavioral skills and the gender gap in early educational achievement. **Social Science Research**, 41: 1–15, 2012.

DOLADO, J. J.; FELGUEROSO, F.; JIMENO, J. F. Where do women work?: Analyzing patterns in occupational segregation by gender. **Annales D'Économie et Statistique**, 71/72: 293–315, 2003.

DOWNEY, D.; VOGT YUAN, A. Sex differences in school performance during high school: puzzling patterns and possible explanations. **The Sociological Quarterly**, 46: 299–321, 2005.

DUMAIS, S. Cultural capital, gender, and school success: the role of habitus. **Sociology of Education**, 75(1): 44–68.

DUNCAN, G.; HOFFMAN, S. The incidence and wage effects of overeducation. **Economics of Education Review**, 1(1): 75–86, 1981.

EAGLY, A.; CARLI, L. **Through the labyrinth: the truth about how women become leaders**. Boston: Harvard Business School Press, 2007.

ELSEVIER GENDER REPORT. **Gender in the global research landscape**. Elsevier, 2017.

ENTWISLE, D.; ALEXANDER, K.; OLSON, L. Early schooling: the handicap of being poor and male. **Sociology of Education**, 80(2): 114–138

FEDERICI, SILVIA. ENTREVISTA DA 2ª [Entrevista concedida à Folha de São Paulo]. Úrsula Passos. Folha de São Paulo, São Paulo. 14. out. 2019. Disponível em: <<https://www1.folha.uol.com.br/ilustrada/2019/10/o-que-eles-chamam-de-amor-nos-chamamos-de-trabalho-nao-pago-diz-silvia-federici.shtml>>. Acesso em: 20/01/2024.

FERNANDEZ-MATEO, I.; FERNANDEZ, R. M. Bending the pipeline? Executive search and gender inequality in hiring for top management jobs. **Management Science**, 62(12): 1–20, 2016.

FIGUEIREDO FILHO, D.; SILVA, L.; PIRES, A.; MALAQUIAS, C. Living with outliers: how to detect extreme observations in data analysis. **Revista Brasileira de Informação Bibliográfica em Ciências Sociais**, 99: 1–24, 2023.

FRANCESCONI, M.; PAREY, M. Early gender gaps among university graduates. **IZA – Institute of Labor Economics**. Discussion paper series n° 113161, 2018.

GOLDIN, C. A grand gender convergence: its last chapter. **American Economic Review**, 104(4): 1091–1119, 2014.

GOLDIN, C. **Career and family: women's century-long journey toward equity**. Princeton, New Jersey, Princeton University Press, 2021.

GOLDIN, C.; KATZ, L. The cost of workplace flexibility for high-powered professionals. **The Annals of the American Academy of Political and Social Science**, 638: 45–67, 2011.

GOLDIN, C.; ROUSE, C. Orchestrating impartiality: the impact of “blind” auditions on female musicians. **The American Economic Review**, 90(4): 715-741, 2000.

GONZÁLEZ-PÉREZ, S.; CABO, R.; SÁINZ, M. Girls in STEM: is it a female role-model thing? **Frontiers in Psychology**, 11: 1–21, 2020.

GREIFER, N. MatchIt: Getting Started. 2022. Disponível em: < <https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html> >. Acesso em: 27/01/2023.

GRIFFITH, A. L.; MAIN, J. B. The role of the teaching assistant: female role models in the classroom. **Economics of Education Review**, 85: 102179.

GUO, S.; FRASER, M. W. **Propensity Score Analysis. Statistical Methods and Applications**. 2nd Edition. New York: SAGE Publications, 2015.

HADLEY, E. Human capital. IN: HADLEY, E.; JAFFE, A.; WINTER, S. **From political economy to economics through nineteenth-century literature**. Cham: Palgrave Macmillan, 2019.

HECKMAN, J. Sample selection bias as a specification error. **Econometrica**, 47(1): 153–161.

HECKMAN, J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. **The Journal of Economic Resources**, 32(3): 441–462, 1997.

HECKMAN, J. J.; LOCHNER, L. J.; TODD, P. E. Earnings functions, rates of return and treatment effects: the Mincer equation and beyond. In: *Handbook of the Economics of Education*. (Eds) Eric A. Hanushek and Finis Welch, Elsevier, 2006.

HO, D. E.; IMAI, K.; KING, G.; STUART, E. A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. **Political Analysis**, 15(3): 199–236, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Censo Demográfico. 2022. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html>>. Acesso em: 20/01/2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Estatísticas de gênero – indicadores sociais das mulheres no Brasil. 2021. Disponível em: <<https://www.ibge.gov.br/estatisticas/multidominio/genero/20163-estatisticas-de-genero-indicadores-sociais-das-mulheres-no-brasil.html>>. Acesso em: 13/11/2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Síntese de Indicadores. 2023. Disponível em: < <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/38543-em-2022-rendimento-hora-dos-trabalhadores-brancos-r-20-0-era-61-4-maior-que-o-dos-pretos-ou-pardos-r-12-> >



MARÇAL, K. **O lado invisível da economia: uma visão feminista**. 1ª Edição. Alaúde Editorial Ltda. São Paulo, 2017.

MARIONI, L. Overeducation in the labour market: evidence from Brazil. **Education Economics**, 29(1): 53–72, 2020

MCCAFFREY, D.; RIDGEWAY, G.; MORRAL, A. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. **Psychological Methods**, 9(4): 403–425, 2004.

MINCER, J. **Schooling, experience, and earnings**. New York, 1974.

NIELSSON, U; STEINGRIMSDOTTIR, H. The signaling value of education across genders. **Empirical Economics**, 54(4): 1827–1854, 2018.

OAXACA, R. Male-female wage differentials in urban labor markets. **International Economic Review**, 14(3): 693–709.

PARK, K. Education-job mismatch and gender wage gap: evidence from recent college graduates in Korea. **Asian Women**, 37(1): 1–24, 2021.

PEREZ, C. C. **Mulheres invisíveis: o viés dos dados em um mundo projetado para homens**. 1ª edição. Rio de Janeiro: Intrínseca, 2022.

PERKINS, R.; KLEINER, B.; ROEY, S.; BROWN, J. **The high school transcript study**. Washington, DC: National Center for Education Statistics.

PHELTS, E. The statistical theory of racism and sexism. **The American Economic Review**, 62(4): 659–661, 1972.

PORTER, C.; SERRA, D. Gender differences in the choice of major: the importance of female role models. **American Economic Journal: Applied Economics**, 12(3): 226–254, 2020.

QUADLIN, N. The mark of woman's record: gender and academic performance in hiring. **American Sociological Review**, 00(0): 1-30, 2018.

REIS, M. Education mismatch and labor earnings in Brazil. **International Journal of Manpower**, 38(2): 180–197, 2017.

REIS, M. Measuring the mismatch between field of study and occupation using a task-based approach. **Journal for Labour Market Research**, 52(9): 1–15.

REIS, M. Os ensinos público e privado no Brasil e a incidência de sobre-educação no mercado de trabalho. **Economia Aplicada**, 24(3): 367–392, 2020.

REIS, M. Educação superior e sobre-educação no Brasil entre 1980 e 2010. **IPEA – Texto para Discussão 2655**, 2021.

REIS, M.; FORTES, M. Educação profissional, exigências da ocupação e rendimentos do trabalho no Brasil. **IPEA – Texto para Discussão 2446**, 2019.

REUBEN, E.; SAPIENZA, P.; ZINGALES, L. How stereotypes impair woman's careers in Science. **PNAS Early Edition**, 111(12): 4403–4408, 2014.

RIDGEWAY, C. **Framed by gender: how gender inequality persists in the modern world**. New York: Oxford University Press, 2011.

RIEGLE-CRUMB, C. More girls go to college: exploring the social and academic factor behind the female postsecondary advantage among Hispanic and White students. **Research in Higher Education**, 51(6): 573–593, 2010.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, 70(1): 41–55, 1983.

RUBIN, D. B. Using propensity scores to help design observational studies: application to the tobacco litigation. **Health Services & Outcomes Research Methodology**, 2(1): 169–188, 2001.

SAHOO, S.; KLASSEN, S. Gender segregation in education and its implications for labour market outcomes: evidence from India. **IZA – Institute of Labor Economics**. Discussion paper n° 11660, 2018.

SALINAS-JIMÉNEZ, M.; RAHONA-LÓPEZ, M. MURILLO-HUERTAS, I. Gender wage differentials and educational mismatch: an application to the Spanish case. **Applied Economics**, 45(30): 4226–4235, 2013.

SANTOS, J. A. F. Classe social e desigualdade de gênero no Brasil. **Revista de Ciências Sociais**, 51(2): 353–402, 2008.

SCHIEBINGER, L.; GILMARTIN, K. Household is an academic issue. **Academe**, 96: 39–44, 2010.

SCHNEEWEIS, N.; ZWEIMÜLLER, M. Girls, girls, girls: gender composition and female school choice. **Economics of Education Review**, 31(4): 482–500, 2012.

SCHWAB, S. Is statistical discrimination efficient? **The American Economic Review**, 76(1): 228–234, 1986.

SICHERMAN, N. Overeducation in the labor market. **Journal of Labor Economics**, 9(2): 101–122, 1991.

SILVA, Y. G.; VAZ, D. V. Por que as ocupações femininas pagam menos? Um estudo longitudinal. **Revista Brasileira de Estudos de População**, v. 39, pp. 1–28, 2022.

SILVEIRA, L. S.; SIQUEIRA, N. L. Segregação ocupacional e diferenciais de renda por gênero e raça no Brasil: uma análise para grupos etários. **Revista Brasileira de Estudos Populacionais**, 38: 1–22, 2021.

SOARES, S. S. D. O Perfil da discriminação no mercado de trabalho: homens negros, mulheres brancas e mulheres negras. Texto para discussão n° 789, **Instituto de Pesquisa Econômica e Aplicada**, 2000.

SOOKRAM, S.; STROBL, E. The role of educational choice in occupational gender segregation: Evidence from Trinidad and Tobago. **Economics of Education Review**, 28(1): 1–10, 2009.

SPENCE, M. Job Market signaling. **Quarterly Journal of Economics**, 87(3): 355–374, 1973.

STUART, E. A.; LEE, B. K.; LEACY, F. P. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. **Journal of Clinical Epidemiology**, 66(1): S84-S90, 2013.

THOEMMES, F. J.; KIM, E. S. A systematic review of propensity score methods in the social sciences. **Multivariate Behavioral Research**, 46(1): 90–118, 2011.

TSAI, Y. Returns to overeducation: a longitudinal analysis of the US labor market. **Economics of Education Review**, 29(4): 606–617, 2010.

VALENTOVA, J. V.; OTTA, E.; SILVA, M. L.; McELLOGOTT, A. G. Underrepresentation of women in the senior levels of Brazilian science. **PeerJ** 5: e4000, 2017.

VAZ, D. V. O teto de vidro nas organizações públicas: evidências para o Brasil. **Economia e Sociedade**, 22(3): 765–790, 2013.

VAZ, D. V.; BOTASIO, D. C. Occupational feminization and pay: the case of Brazil. **Brazilian Review of Econometrics**, 42(1): 1–20, 2022.

VAZ, D. V.; HOFFMAN, R. Occupational gender segregation in the Brazilian public sector between 19995 and 2008. **Revista ABET**, v. X(1): 120–141, 2011.

VERDUGO, R.; VERDUGO, N. The impact of surplus schooling on earnings: some additional findings. **Journal of Human Resources**, 24(4): 629–643.

VERDUGO-CASTRO, S.; GARCÍA-HOLGADO, A.; SÁNCHEZ-GÓMEZ, M. C.; GARCÍA-PEÑALVO, F. J. Multimedia analysis of Spanish female role models in Science, Technology, Engineering and Mathematics. **Sustainability**, 13: 12612, 2021.

VERHOFSTADT, E.; DE WITTE, H.; OMEY, E. Higher educated workers: better jobs but less satisfied? **International Journal of Manpower**, 28: 135–151, 2007.

WALFISH, S. A review of statistical outlier methods. **Pharmaceutical Technology**, 30(11): 82–88.

WEISS, A. Human capital vs. signaling explanations of wages. **Journal of Economic Perspectives**, 9(4): 133–154.

## APÊNDICES

Quadro A1 – Categorias de “Profissionais das Ciências e Intelectuais” da Classificação de Ocupação para Pesquisas Domiciliares da PNAD Contínua

<b>PROFISSIONAIS DAS CIÊNCIAS E DA ENGENHARIA</b>
Físicos, químicos e afins
Físicos e astrônomos
Meteorologistas
Químicos
Geólogos e geofísicos
Matemáticos, atuários e estatísticos
Matemáticos, atuários e estatísticos
Profissionais em ciências biológicas
Biólogos, botânicos, zoólogos e afins
Agrônomos e afins
Profissionais da proteção do meio ambiente
Engenheiros (exclusive eletrotécnicos)
Engenheiros industriais e de produção
Engenheiros civis
Engenheiros de meio ambiente
Engenheiros mecânicos
Engenheiros químicos
Engenheiros de minas, metalúrgicos e afins
Engenheiros não classificados anteriormente
Engenheiros eletrotécnicos
Engenheiros eletricitas
Engenheiros eletrônicos
Engenheiros em telecomunicações
Arquitetos, urbanistas, agrimensores e desenhistas
Arquitetos de edificações
Arquitetos paisagistas
Desenhistas de produtos e vestuário
Urbanistas e engenheiros de trânsito
Cartógrafos e agrimensores
Desenhistas gráficos e de multimídia
<b>PROFISSIONAIS DA SAÚDE</b>
Médicos
Médicos gerais
Médicos especialistas
Profissionais de enfermagem e partos
Profissionais de enfermagem
Profissionais de partos

(continua)

(continuação)

**PROFISSIONAIS DA SAÚDE**

Profissionais da medicina tradicional e alternativa  
 Profissionais da medicina tradicional e alternativa  
 Paramédicos  
 Paramédicos  
 Veterinários  
 Veterinários  
 Outros profissionais da saúde  
 Dentistas  
 Farmacêuticos  
 Profissionais da saúde e da higiene laboral e ambiental  
 Fisioterapeutas  
 Dietistas e nutricionistas  
 Fonoaudiólogos e logopedistas  
 Optometristas  
 Profissionais da saúde não classificados anteriormente

**PROFISSIONAIS DO ENSINO**

Professores de universidades e do ensino superior  
 Professores de universidades e do ensino superior  
 Professores de formação profissional  
 Professores de formação profissional  
 Professores do ensino médio  
 Professores do ensino médio  
 Professores do ensino fundamental e pré-escolar  
 Professores do ensino fundamental  
 Professores do ensino pré-escolar  
 Outros profissionais do ensino  
 Especialistas em métodos pedagógicos  
 Educadores para necessidades especiais  
 Outros professores de idiomas  
 Outros professores de música  
 Outros professores de artes  
 Instrutores em tecnologias da informação  
 Profissionais de ensino não classificados anteriormente

**ESPECIALISTAS EM ORGANIZAÇÃO DA ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS**

Especialistas em finanças  
 Contadores  
 Assessores financeiros e em investimentos  
 Analistas financeiros  
 Especialistas em organização de administração  
 Analistas de gestão e administração  
 Especialistas em políticas de administração  
 Especialistas em políticas e serviços de pessoal e afins  
 Especialistas em formação de pessoal

(continua)

(continuação)

**ESPECIALISTAS EM ORGANIZAÇÃO DA ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS**

Profissionais de vendas, comercialização e relações públicas

Profissionais da publicidade e da comercialização

Profissionais de relações públicas

Profissionais de vendas técnicas e médicas (exclusive tic)

Profissionais de vendas de tecnologia da informação e comunicações

**PROFISSIONAIS DE TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÕES**

Desenvolvedores e analistas de programas e aplicativos (software) e multimídia

Analistas de sistemas

Desenvolvedores de programas e aplicativos (software)

Desenvolvedores de páginas de internet (web) e multimídia

Programadores de aplicações

Desenvolvedores e analistas de programas e aplicativos (software) e multimídia não classificados anteriormente

Especialistas em base de dados e em redes de computadores

Desenhistas e administradores de bases de dados

Administradores de sistemas

Profissionais em rede de computadores

Especialistas em base de dados e em redes de computadores não classificados anteriormente

**PROFISSIONAIS EM DIREITO, EM CIÊNCIAS SOCIAIS E CULTURAIS**

Profissionais em direito

Advogados e juristas

Juízes

Profissionais em direito não classificados anteriormente

Arquivologistas, curadores de museus, bibliotecários e afins

Arquivologistas e curadores de museus

Bibliotecários, documentaristas e afins

Especialistas em ciências sociais e teologia

Economistas

Sociólogos, antropólogos e afins

Filósofos, historiadores e especialistas em ciência política

Psicólogos

Assistentes sociais

Ministros de cultos religiosos, missionários e afins

Escritores, jornalistas e linguistas

Escritores

Jornalistas

Tradutores, intérpretes e linguistas

Artistas criativos e interpretativos

Artistas plásticos

Músicos, cantores e compositores

Bailarinos e coreógrafos

Diretores de cinema, de teatro e afins

(continua)

(continuação)

<b>PROFISSIONAIS EM DIREITO, EM CIÊNCIAS SOCIAIS E CULTURAIS</b>
Atores
Locutores de rádio, televisão e outros meios de comunicação
Artistas criativos e interpretativos não classificados anteriormente

Fonte: Elaboração própria a partir de informações da PNAD Contínua.

Tabela A1 – Relação das ocupações do grupo de controle pareadas com o grupo de tratamento

Número de observações pareadas	Denominação
13.678	Escriturários gerais
3.779	Comerciantes de lojas
3.136	Profissionais de nível médio do direito e serviços legais e afins
2.739	Dirigentes de administração e de serviços não classificados anteriormente
2.518	Balconistas e vendedores de lojas
2.514	Profissionais de nível médio de enfermagem
2.269	Secretários (geral)
1.807	Dirigentes de serviços de educação
1.763	Secretários executivos e administrativos
1.658	Secretários(geral)
1.656	Vendedores a domicilio
1.560	Dirigentes de vendas e comercialização
1.499	Trabalhadores comunitários da saúde
1.470	Dirigentes financeiros
1.397	Gerentes de sucursais de bancos, de serviços financeiros e de seguros
1.387	Gerentes de comércios atacadistas e varejistas
1.302	Trabalhadores de limpeza de interior de edifícios, escritórios, hotéis e outros estabelecimentos
1.258	Recepcionistas em geral
1.174	Instrutores de educação física e atividades recreativas
1.173	Condutores de automóveis, taxis e caminhonetes
1.143	Cozinheiros
1.127	Supervisores de secretaria
999	Cuidadores de crianças
944	Dirigentes superiores da administração pública
908	Agentes imobiliários
876	Policiais
865	Especialistas em tratamento de beleza e afins
860	Ajudantes de professores
830	Agricultores e trabalhadores qualificados em atividades da agricultura (exclusive hortas, viveiros e jardins)
809	Guardas de segurança
768	Trabalhadores dos serviços domésticos em geral
753	Criadores de gado e trabalhadores qualificados da criação de gado
740	Oficiais das forças armadas
662	Cabeleireiros
614	Agentes e corretores de bolsa, câmbio e outros serviços financeiros

(continua)

(continuação)

Número de observações pareadas	Denominação
608	Padeiros, confeitadores e afins
588	Diretores gerais e gerentes gerais
572	Contabilistas e guarda livros
565	Caixas de banco e afins
559	Inspetores de polícia e detetives

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Tabela A2 – Resultados estimados das regressões – Mulheres

Variáveis	Estimações	Erro padrão	Teste t	Pr(> t )
Intercepto	-17,2655	8,0940	-2,1330	0,032917 *
Intelectual	-13,2577	9,7658	-1,3580	0,174606
Idade	0,2565	0,2640	0,9720	0,331255
Idade ao quadrado	0,0032	0,0030	1,0640	0,287382
Habitantes	-1,3531	0,5462	-2,4770	0,013240 *
Número de filhos	0,7496	0,7068	1,0610	0,288838
Anos de estudo	3,2433	0,3667	8,8460	< 2e-16 ***
Raça Branca/Amarela	9,5126	0,9241	10,2940	< 2e-16 ***
Chefe de família	3,5186	0,9270	3,7960	0,000147 ***
Região metropolitana	-13,4291	0,8455	-15,8830	< 2e-16 ***
Setor privado	-8,4798	0,8693	-9,7550	< 2e-16 ***
Cônjuge	4,5439	0,8689	5,2290	1,70e-07 ***
Tempo integral	-14,8461	0,8260	-17,9740	< 2e-16 ***
Região Norte	-13,5984	1,7158	-7,9250	2,30e-15 ***
Região Nordeste	-14,9554	1,4860	-10,0640	< 2e-16 ***
Região Sudeste	-6,6106	1,4238	-4,6430	3,44e-06 ***
Região Sul	-8,6839	1,5780	-5,5030	3,74e-08 ***
Região Centro-Oeste	NA	NA	NA	NA
Ano 2016	4,3608	1,4927	2,9210	0,003484 **
Ano 2017	3,5475	1,4959	2,3710	0,017721 *
Ano 2018	4,4628	1,4891	2,9970	0,002728 **
Ano 2019	1,5406	1,4646	1,0520	0,292851
Ano 2020	8,0502	1,5858	5,0760	3,85e-07 ***
Ano 2021	-0,2423	1,5545	-0,1560	0,876154
Ano 2022	NA	NA	NA	NA
Erro padrão residual	67,8900			
R <sup>2</sup>	0,0523			
Teste F	111,200***			

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%.

Tabela A3 – Resultados estimados das regressões – Homens

Variáveis	Estimações	Erro padrão	Teste t	Pr(> t )
Intercepto	-22,4213	9,4162	-2,3810	0,017262 *
Intelectual	1,9936	12,0073	0,1660	0,868133
Idade	0,3220	0,3216	1,0010	0,316683
Idade ao quadrado	0,0050	0,0034	1,4800	0,138879
Habitantes	-1,2338	0,9078	-1,3590	0,174154
Número de filhos	2,4232	1,1603	2,0880	0,036765 *
Anos de estudo	3,7251	0,3516	10,5940	< 2e-16 ***
Raça Branca/Amarela	11,7718	1,4228	8,2740	< 2e-16 ***
Chefe de família	5,4289	1,4469	3,7520	0,000175 ***
Região metropolitana	-14,2043	1,2783	-11,1120	< 2e-16 ***
Setor privado	-11,2245	14,1351	-7,9410	2,04e-15 ***
Cônjuge	9,4193	1,5046	6,2600	3,87e-10 ***
Tempo integral	-22,5930	1,3384	-16,8810	< 2e-16 ***
Região Norte	-11,2588	2,6151	-4,3050	1,67e-05 ***
Região Nordeste	-15,0094	2,3014	-6,5220	6,99e-11 ***
Região Sudeste	-1,2532	2,0988	-0,5970	0,550461
Região Sul	-3,6970	2,3319	-1,5850	0,112883
Região Centro-Oeste	NA	NA	NA	NA
Ano 2016	3,3919	2,2398	1,5140	0,129938
Ano 2017	6,0229	2,2420	2,6860	0,007225 **
Ano 2018	3,7438	2,2335	1,6760	0,093694
Ano 2019	7,8319	2,2285	3,5140	0,000441 ***
Ano 2020	7,2314	2,3420	3,0880	0,002018 **
Ano 2021	-1,1394	2,3568	-0,4830	0,628769
Ano 2022	NA	NA	NA	NA
Erro padrão residual	93,0800			
R <sup>2</sup>	0,0792			
Teste F	121,200***			

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%.

Tabela A4 – Resultados estimados das regressões – Todos

Variáveis	Estimações	Erro padrão	Teste t	Pr(> t )
Intercepto	-12,4474	6,0088	-2,0720	0,038313 *
Intelectual	-2,7866	7,4173	-0,3760	0,707152
Idade	0,0087	0,2034	0,0430	0,966013
Gênero	-13,9115	0,7407	-18,7820	< 2e-16 ***
Idade ao quadrado	0,0070	0,0022	3,1130	0,001855 **
Habitantes	-1,1595	0,4789	-2,4210	0,015468 *
Número de filhos	2,1022	0,6196	3,3930	0,000692 ***
Anos de estudo	3,6335	0,2459	14,7750	< 2e-16 ***
Raça Branca/Amarela	11,1672	0,7937	14,0700	< 2e-16 ***
Chefe de família	5,6846	0,7691	7,3910	1,46e-13 ***
Região metropolitana	-13,4714	0,7170	-18,7880	< 2e-16 ***
Setor privado	-9,0639	0,7421	-12,2140	< 2e-16 ***
Cônjuge	7,1777	0,7546	9,5120	< 2e-16 ***
Tempo integral	-17,4934	0,7240	-24,1620	< 2e-16 ***
Região Norte	-11,1385	1,4877	-7,4870	7,10e-14 ***

(continua)

(continuação)				
Variáveis	Estimações	Erro padrão	Teste t	Pr(> t )
Região Nordeste	-14,2656	1,2940	-11,0240	< 2e-16 ***
Região Sudeste	-5,0952	1,2143	-4,1960	2,72e-05 ***
Região Sul	-6,3323	1,3454	-4,7070	2,52e-06 ***
Região Centro-Oeste	NA	NA	NA	NA
Ano 2016	3,8988	1,2663	3,0790	0,002079 **
Ano 2017	6,8168	1,2696	5,3690	7,92e-08 ***
Ano 2018	4,5275	1,2668	3,5740	0,000352 ***
Ano 2019	4,8361	1,2521	3,8630	0,000112 ***
Ano 2020	8,3016	1,3415	6,1880	6,11e-10 ***
Ano 2021	0,0517	1,3333	0,0390	0,969085
Ano 2022	NA	NA	NA	NA
Erro padrão residual	78,6900			
R <sup>2</sup>	0,0725			
Teste F	256,700***			

Fonte: Elaboração própria a partir de microdados da PNADC de 2016 a 2022.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%.

Tabela A5 – Cursos e áreas da UFPR

Curso	Área Capes
Administração	Sociais Aplicadas
Agronomia	Agrárias
Arquitetura e Urbanismo	Sociais Aplicadas
Artes plásticas	Linguística, Letras e Artes
Biblioteconomia	Sociais Aplicadas
Biologia geral	Biológicas
Ciências da computação	Exatas e Terra
Ciências da informação	Sociais Aplicadas
Ciências contábeis	Sociais Aplicadas
Ciências humanas	Humanas
Comunicação	Sociais Aplicadas
Desenho de produto	Sociais Aplicadas
Desenho industrial	Sociais Aplicadas
Direito	Sociais Aplicadas
Economia	Sociais Aplicadas
Educação	Humanas
Educação artística	Humanas
Educação física	Saúde
Enfermagem	Saúde
Enfermagem obstétrica	Saúde
Engenharia civil	Engenharias
Engenharia mecânica	Engenharias
Engenharia química	Engenharias

(continua)

(continuação)

<b>Curso</b>	<b>Área Capes</b>
Engenharias	Engenharias
Estatística	Exatas e Terra
Farmácia	Saúde
Filosofia	Humanas
Física	Exatas e Terra
Geografia	Exatas e Terra
Geologia	Exatas e Terra
História	Humanas
Jornalismo de editoração	Sociais Aplicadas
Letras	Linguística, Letras e Artes
Matemática	Exatas e Terra
Medicina	Saúde
Medicina veterinária	Agrárias
Música	Linguística, Letras e Artes
Nutrição	Saúde
Oceanografia	Biológicas
Odontologia	Saúde
Psicologia	Humanas
Química	Exatas e Terra
Recursos florestais e engenharia florestal	Exatas e Terra
Relações públicas e propaganda	Sociais Aplicadas
Terapia ocupacional	Saúde
Turismo	Sociais Aplicadas
Zootecnia	Agrárias

Fonte: Elaboração própria a partir de dados da UFPR e CAPES.

Tabela A6 – Resultados das estimações das equações de seleção

<b>Variáveis</b>	<b>Todos</b>	<b>Formados</b>	<b>Mulheres formadas</b>	<b>Homens formados</b>
Intercepto	-1.760*** (0.067)	-4.599*** (0.103)	-2.335*** (0.171)	-1.704*** (0.186)
Gênero	-0.013 (0.009)			
Idade de formação	0.068*** (0.004)	0.074*** (0.006)	0.056*** (0.009)	0.055*** (0.010)
(Idade de formação) <sup>2</sup>	-0.001*** (0.0001)	-0.001*** (0.0001)	-0.001*** (0.0001)	-0.001*** (0.0002)

(continua)

(continuação)

Variáveis	Todos	Formados	Mulheres formadas	Homens formados
Tempo de graduação	0.010*** (0.002)	0.127*** (0.003)	0.046*** (0.006)	0.030*** (0.006)
Curso integral	-0.011 (0.019)	0.174*** (0.024)	0.255*** (0.035)	-0.245*** (0.043)
Curso matutino	0.037* (0.020)	0.164*** (0.025)	0.276*** (0.036)	-0.110** (0.046)
Curso noturno	0.031 (0.019)	0.103*** (0.025)	0.185*** (0.036)	-0.229*** (0.045)
Média agregada	0.001*** (0.0002)	0.027*** (0.103)	0.006*** (0.001)	0.004*** (0.001)
Observações	96,18	95,769	30,91	27,807
Log Likelihood	-60,346.970	-41,260.860	-19,350.230	-17,512.380
Akaike Inf. Crit.	120,711.900	82,537.720	38,716.450	35,040.750

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A7 – Resultados das estimações considerando todos os indivíduos

Variáveis	1	2	3
<b><u>Características pessoais</u></b>			
Nascimento 1930-1959	-0.225*** (0.060)	-0.143** (0.060)	-0.872*** (0.055)
Nascimento 1960-1969	-0.328*** (0.051)	-0.285*** (0.051)	-0.759*** (0.046)
Nascimento 1970-1979	-0.217*** (0.050)	-0.182*** (0.050)	-0.501*** (0.045)
Nascimento 1980-1989	0.113** (0.050)	0.139*** (0.049)	-0.042 (0.044)
<b><u>Características da formação</u></b>			
Turno noturno		0.094*** (0.016)	0.037** (0.015)
Turno integral		0.121*** (0.019)	0.098*** (0.017)
Turno vespertino/diurno		-0.038 (0.024)	-0.010 (0.022)

(continua)

	(continuação)		
Variáveis	1	2	3
Formados		0.039*** (0.015)	0.069*** (0.013)
Tempo de graduação		0.003 (0.003)	-0.013*** (0.003)
Agrárias		-0.245*** (0.027)	-0.178*** (0.024)
Biológicas		-0.090*** (0.028)	-0.127*** (0.025)
Engenharias		-0.001 (0.022)	0.075*** (0.019)
Exatas e Terra		-0.145*** (0.016)	-0.102*** (0.015)
Humanas		-0.238*** (0.019)	-0.221*** (0.017)
Linguística, Letras e Artes		-0.112*** (0.024)	-0.098*** (0.021)
Saúde		-0.096*** (0.021)	-0.041** (0.019)
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.631*** (0.061)
Região Nordeste			0.497*** (0.056)
Região Sudeste			0.411*** (0.016)
Região Centro-Oeste			0.437*** (0.025)
Jornada 31 a 44 horas semanais			0.141*** (0.011)
Jornada 21 a 30 horas semanais			0.058*** (0.014)
Média de meses empregado(a)			0.007*** (0.0001)

(continua)

Variáveis	(continuação)		
	1	2	3
<b>Outros controles</b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.141***
Estava no mercado de trabalho formal antes de iniciar o curso			-0.230*** (0.022)
Observações	31,553	31,553	31,553
Log Likelihood	-41,577.230	-41,325.490	-37,476.700
Akaike Inf. Crit.	83,172.460	82,692.990	75,015.400

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A8 – Resultados das estimações considerando apenas as mulheres

Variáveis	1	2	3
<b>Características pessoais</b>			
Nascimento 1930-1959	-0.306*** (0.080)	-0.246*** (0.080)	0.006 (0.079)
Nascimento 1960-1969	-0.401*** (0.068)	-0.362*** (0.068)	0.028 (0.067)
Nascimento 1970-1979	-0.277*** (0.067)	-0.247*** (0.066)	-0.046 (0.061)
Nascimento 1980-1989	0.053 (0.066)	0.075 (0.066)	0.004 (0.057)
<b>Características da formação</b>			
Turno noturno		0.080*** (0.022)	0.056*** (0.019)
Turno integral		0.155*** (0.026)	0.073*** (0.022)
Turno vespertino/diurno		0.056* (0.034)	-0.025 (0.029)
Formados		0.032 (0.021)	0.048*** (0.018)
Tempo de graduação		0.002 (0.004)	0.004 (0.004)
Agrárias		-0.335*** (0.037)	-0.200*** (0.032)

(continua)

(continuação)			
Variáveis	1	2	3
Biológicas		-0.058 (0.036)	-0.057* (0.031)
Engenharias		-0.129*** (0.033)	-0.002 (0.029)
Exatas e Terra		-0.213*** (0.024)	-0.153*** (0.021)
Humanas		-0.170*** (0.024)	-0.196*** (0.021)
Linguística, Letras e Artes		-0.052* (0.030)	-0.052** (0.026)
Saúde		-0.054** (0.027)	-0.014 (0.024)
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.614*** (0.096)
Região Nordeste			0.413*** (0.084)
Região Sudeste			0.392*** (0.022)
Região Centro-Oeste			0.565*** (0.040)
Jornada 31 a 44 horas semanais			0.158*** (0.015)
Jornada 21 a 30 horas semanais			0.058*** (0.018)
Média de meses empregado(a)			0.006*** (0.0001)
<b><u>Outros controles</u></b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.326*** (0.021)
Estava no mercado de trabalho formal antes de iniciar o curso			-0.287*** (0.032)
Observações	15,456	15,456	15,456
Log Likelihood	-19,823.400	-19,699.850	-17,405.720
Akaike Inf. Crit.	39,660.790	39,437.700	34,871.450

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A9 – Resultados das estimações considerando apenas os homens

<b>Variáveis</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b><u>Características pessoais</u></b>			
Nascimento 1930-1959	-0.137 (0.089)	-0.043 (0.088)	0.053 (0.084)
Nascimento 1960-1969	-0.249*** (0.077)	-0.205*** (0.076)	0.122* (0.072)
Nascimento 1970-1979	-0.150** (0.075)	-0.110 (0.074)	0.039 (0.067)
Nascimento 1980-1989	0.180** (0.075)	0.193*** (0.074)	0.105* (0.064)
<b><u>Características da formação</u></b>			
Turno noturno		0.098*** (0.023)	0.024 (0.021)
Turno integral		0.089*** (0.027)	0.021 (0.024)
Turno vespertino/diurno		-0.102*** (0.036)	-0.141*** (0.031)
Formados		0.042** (0.021)	0.122*** (0.018)
Tempo de graduação		0.005 (0.003)	-0.005 (0.004)
Agrárias		-0.160*** (0.038)	-0.124*** (0.033)
Biológicas		-0.129*** (0.044)	-0.137*** (0.038)
Engenharias		0.069** (0.029)	0.099*** (0.025)
Exatas e Terra		-0.106*** (0.022)	-0.105*** (0.019)
Humanas		-0.338*** (0.029)	-0.271*** (0.025)
Linguística, Letras e Artes		-0.191*** (0.037)	-0.174*** (0.032)
Saúde		-0.183*** (0.032)	-0.051* (0.028)

(continua)

(continuação)			
Variáveis	1	2	3
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.632*** (0.075)
Região Nordeste			0.524*** (0.073)
Região Sudeste			0.390*** (0.023)
Região Centro-Oeste			0.364*** (0.031)
Jornada 31 a 44 horas semanais			0.168*** (0.016)
Jornada 21 a 30 horas semanais			0.080*** (0.021)
Média de meses empregado(a)			0.006*** (0.0001)
<b><u>Outros controles</u></b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.308*** (0.021)
Estava no mercado de trabalho formal antes de iniciar o curso			-0.366*** (0.030)
Observações	16,097	16,097	16,097
Log Likelihood	-21,718.240	-21,514.610	-19,103.420
Akaike Inf. Crit.	43,450.480	43,067.220	38,266.830

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A10 – Resultados das estimações considerando todos os indivíduos formados

Variáveis	1	2	3
<b><u>Características pessoais</u></b>			
Nascimento 1930-1959	-0.248*** (0.080)	-0.105 (0.080)	-0.762*** (0.072)
Nascimento 1960-1969	-0.425*** (0.065)	-0.332*** (0.064)	-0.719*** (0.058)
Nascimento 1970-1979	-0.322*** (0.063)	-0.254*** (0.063)	-0.500*** (0.056)
Nascimento 1980-1989	0.090 (0.062)	0.140** (0.062)	-0.0004 (0.055)

(continua)

	(continuação)		
Variáveis	1	2	3
<b><u>Características da formação</u></b>			
Turno noturno		0.073*** (0.022)	0.052*** (0.020)
Turno integral		0.084*** (0.025)	0.049** (0.022)
Turno vespertino/diurno		0.006 (0.036)	0.016 (0.032)
Tempo de graduação		0.036*** (0.005)	0.008 (0.007)
Agrárias		-0.187*** (0.033)	-0.130*** (0.030)
Biológicas		-0.069** (0.034)	-0.127*** (0.030)
Engenharias		0.063** (0.028)	0.138*** (0.028)
Exatas e Terra		-0.086*** (0.023)	-0.042* (0.022)
Humanas		-0.269*** (0.025)	-0.257*** (0.023)
Linguística, Letras e Artes		-0.159*** (0.032)	-0.127*** (0.029)
Saúde		-0.061** (0.026)	-0.051** (0.024)
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.692*** (0.071)
Região Nordeste			0.557*** (0.069)
Região Sudeste			0.425*** (0.021)
Região Centro-Oeste			0.487*** (0.032)
Jornada 31 a 44 horas semanais			0.137*** (0.014)

(continua)

	(continuação)		
Variáveis	1	2	3
Jornada 21 a 30 horas semanais			0.025 (0.018)
Média de meses empregado(a)			0.008*** (0.0001)
<b><u>Outros controles</u></b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.123*** (0.019)
Estava no mercado de trabalho formal antes de iniciar o curso			-0.351*** (0.036)
Observações	19,202	19,202	19,202
Log Likelihood	-25,447.700	-25,250.270	-22,912.130
Akaike Inf. Crit.	50,913.390	50,540.540	45,884.250

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A11 – Resultados das estimações considerando apenas as mulheres formadas

Variáveis	1	2	3
<b><u>Características pessoais</u></b>			
Nascimento 1930-1959	-0.341*** (0.101)	-0.159 (0.101)	-0.863*** (0.091)
Nascimento 1960-1969	-0.475*** (0.084)	-0.339*** (0.084)	-0.762*** (0.075)
Nascimento 1970-1979	-0.376*** (0.082)	-0.276*** (0.081)	-0.556*** (0.072)
Nascimento 1980-1989	-0.006 (0.081)	0.074 (0.080)	-0.077 (0.071)
<b><u>Características da formação</u></b>			
Turno noturno		0.095*** (0.029)	0.136*** (0.028)
Turno integral		0.102*** (0.032)	0.085*** (0.028)
Turno vespertino/diurno		0.093** (0.045)	0.313*** (0.055)
Tempo de graduação		0.034*** (0.006)	-0.015* (0.009)

(continua)

	(continuação)		
<b>Variáveis</b>	<b>1</b>	<b>2</b>	<b>3</b>
Agrárias		-0.248*** (0.046)	-0.187*** (0.041)
Biológicas		-0.074* (0.042)	-0.125*** (0.037)
Engenharias		0.127*** (0.044)	0.166*** (0.039)
Exatas e Terra		-0.168*** (0.035)	-0.135*** (0.031)
Humanas		-0.264*** (0.030)	-0.288*** (0.027)
Linguística, Letras e Artes		-0.162*** (0.039)	-0.119*** (0.035)
Saúde		-0.044 (0.033)	-0.065** (0.029)
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.814*** (0.118)
Região Nordeste			0.429*** (0.099)
Região Sudeste			0.411*** (0.028)
Região Centro-Oeste			0.558*** (0.045)
Jornada 31 a 44 horas semanais			0.123*** (0.019)
Jornada 21 a 30 horas semanais			0.035 (0.023)
Média de meses empregado(a)			0.008*** (0.0002)
<b><u>Outros controles</u></b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.201*** (0.026)

(continua)

	(continuação)		
Variáveis	1	2	3
Estava no mercado de trabalho formal antes de iniciar o curso			-0.384*** (0.050)
Observações	10,006	10,006	10,006
Log Likelihood	-12,835.750	-12,707.200	-11,433.740
Akaike Inf. Crit.	25,685.500	25,450.400	22,923.490

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A12 – Resultados das estimações considerando apenas os homens formados

Variáveis	1	2	3
<b><u>Características pessoais</u></b>			
Nascimento 1930-1959	-0.088 (0.130)	-0.012 (0.130)	-0.997*** (0.121)
Nascimento 1960-1969	-0.368*** (0.100)	-0.317*** (0.099)	-0.821*** (0.091)
Nascimento 1970-1979	-0.255*** (0.098)	-0.216** (0.097)	-0.527*** (0.087)
Nascimento 1980-1989	0.203** (0.097)	0.226** (0.096)	0.039 (0.086)
<b><u>Características da formação</u></b>			
Turno noturno		0.031 (0.035)	0.074** (0.033)
Turno integral		0.049 (0.039)	0.131*** (0.038)
Turno vespertino/diurno		-0.093 (0.057)	-0.167*** (0.052)
Tempo de graduação		0.037*** (0.007)	0.004 (0.008)
Agrárias		-0.153*** (0.048)	-0.124*** (0.043)
Biológicas		-0.069 (0.057)	-0.139*** (0.051)
Engenharias		0.033 (0.040)	0.084** (0.036)

(continua)

	(continuação)		
Variáveis	1	2	3
Exatas e Terra		-0.040 (0.033)	-0.026 (0.029)
Humanas		-0.293*** (0.045)	-0.264*** (0.041)
Linguística, Letras e Artes		-0.157*** (0.057)	-0.213*** (0.052)
Saúde		-0.099** (0.042)	-0.028 (0.038)
<b><u>Características do trabalho formal</u></b>			
Região Norte			0.609*** (0.091)
Região Nordeste			0.679*** (0.098)
Região Sudeste			0.434*** (0.031)
Região Centro-Oeste			0.437*** (0.046)
Jornada 31 a 44 horas semanais			0.164*** (0.022)
Jornada 21 a 30 horas semanais			0.008 (0.029)
Média de meses empregado(a)			0.007*** (0.0002)
<b><u>Outros controles</u></b>			
Estava no mercado de trabalho formal antes de terminar o curso			-0.120*** (0.028)
Estava no mercado de trabalho formal antes de iniciar o curso			-0.420*** (0.054)
Observações	9,196	9,196	9,196
Log Likelihood	-12,594.930	-12,516.220	-11,400.530
Akaike Inf. Crit.	25,203.860	25,068.450	22,857.070

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A13 – Resultados das estimações com efeitos fixos para toda a base

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.142*** (0.006)			-0.263*** (0.013)
<i>Mismatch</i> horizontal		-0.029*** (0.007)		-0.105*** (0.009)
<i>Mismatch</i> vertical x horizontal			-0.106*** (0.006)	0.134*** (0.014)
Norte	0.169*** (0.058)	0.175*** (0.058)	0.176*** (0.058)	0.167*** (0.058)
Nordeste	0.243*** (0.051)	0.246*** (0.051)	0.241*** (0.051)	0.249*** (0.051)
Centro-Oeste	0.321*** (0.027)	0.310*** (0.028)	0.314*** (0.028)	0.318*** (0.027)
Sudeste	0.213*** (0.016)	0.212*** (0.016)	0.215*** (0.016)	0.214*** (0.016)
Horas contratadas	0.020*** (0.0002)	0.020*** (0.0002)	0.020*** (0.0002)	0.020*** (0.0002)
Tempo empregado(a)	0.002*** (0.0001)	0.002*** (0.0001)	0.002*** (0.0001)	0.002*** (0.0001)
Observações	108,799	108,799	108,799	108,799
R <sup>2</sup>	0.101	0.096	0.099	0.102
R <sup>2</sup> Ajustado	-0.041	-0.047	-0.043	-0.039
Estatística F	1,508.243***	1,419.930***	1,469.389***	1,190.431***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A14 – Resultados das estimações com efeitos aleatórios para toda a base

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.307*** (0.011)			-0.483*** (0.026)
<i>Mismatch</i> horizontal		-0.075*** (0.013)		-0.196*** (0.018)
<i>Mismatch</i> vertical x horizontal			-0.253*** (0.011)	0.192*** (0.027)
Norte	0.353*** (0.102)	0.372*** (0.105)	0.370*** (0.103)	0.357*** (0.102)

(continua)

(continuação)				
Variáveis	1	2	3	4
Nordeste	0.313*** (0.091)	0.328*** (0.093)	0.313*** (0.091)	0.323*** (0.090)
Centro-Oeste	0.517*** (0.048)	0.495*** (0.050)	0.507*** (0.048)	0.509*** (0.048)
Sudeste	0.295*** (0.027)	0.294*** (0.028)	0.300*** (0.027)	0.298*** (0.027)
Horas contratadas	0.021*** (0.0004)	0.019*** (0.0004)	0.020*** (0.0004)	0.021*** (0.0004)
Tempo empregado(a)	0.004*** (0.0001)	0.003*** (0.0001)	0.003*** (0.0001)	0.004*** (0.0001)
Constante	7.561*** (0.027)	7.556*** (0.042)	7.539*** (0.041)	7.718*** (0.031)
Observações	108,799	108,799	108,799	108,799
R <sup>2</sup>	0.311	0.225	0.287	0.324
R <sup>2</sup> Ajustado	0.311	0.224	0.287	0.324
Estatística F	4,330.469***	3,310.935***	4,007.058***	4,523.015***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A15 – Resultados das estimações OLS *pooled* para toda a base

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.578*** (0.005)			-0.759*** (0.011)
<i>Mismatch</i> horizontal		-0.169*** (0.006)		-0.298*** (0.007)
<i>Mismatch</i> vertical x horizontal			-0.541*** (0.005)	0.203*** (0.012)
Norte	0.453*** (0.038)	0.608*** (0.040)	0.495*** (0.038)	0.467*** (0.037)
Nordeste	0.364*** (0.035)	0.468*** (0.037)	0.386*** (0.035)	0.378*** (0.034)
Centro-Oeste	0.633*** (0.016)	0.682*** (0.017)	0.633*** (0.017)	0.620*** (0.016)

(continua)

(continuação)				
Variáveis	1	2	3	4
Sudeste	0.387*** (0.009)	0.448*** (0.010)	0.414*** (0.009)	0.398*** (0.009)
Horas contratadas	0.026*** (0.0002)	0.020*** (0.0002)	0.024*** (0.0002)	0.027*** (0.0002)
Tempo empregado(a)	0.005*** (0.00003)	0.005*** (0.00003)	0.005*** (0.00003)	0.005*** (0.00003)
Constante	7.515*** (0.008)	7.637*** (0.009)	7.527*** (0.008)	7.704*** (0.009)
Observações	108,799	108,799	108,799	108,799
R <sup>2</sup>	0.321	0.234	0.302	0.332
R <sup>2</sup> Ajustado	0.321	0.234	0.302	0.331
Estatística F	7,345.949***	4,750.393***	6,721.881***	5,994.333***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A16 – Resultados das estimações com efeitos fixos para as mulheres

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.155*** (0.008)			-0.277*** (0.018)
<i>Mismatch</i> horizontal		-0.029*** (0.009)		-0.110*** (0.012)
<i>Mismatch</i> vertical x horizontal			-0.116*** (0.008)	0.135*** (0.019)
Norte	-0.083 (0.097)	-0.067 (0.098)	-0.066 (0.097)	-0.081 (0.097)
Nordeste	0.204*** (0.067)	0.214*** (0.068)	0.208*** (0.067)	0.207*** (0.067)
Centro-Oeste	0.338*** (0.040)	0.328*** (0.040)	0.333*** (0.040)	0.335*** (0.040)
Sudeste	0.202*** (0.021)	0.199*** (0.021)	0.204*** (0.021)	0.201*** (0.021)
Horas contratadas	0.021*** (0.0003)	0.020*** (0.0003)	0.021*** (0.0003)	0.021*** (0.0003)

(continua)

(continuação)				
Variáveis	1	2	3	4
Tempo empregado(a)	0.002*** (0.0001)	0.002*** (0.0001)	0.002*** (0.0001)	0.002*** (0.0001)
Observações	57,98	57,98	57,98	57,98
R <sup>2</sup>	0.110	0.104	0.108	0.112
R <sup>2</sup> Ajustado	-0.029	-0.037	-0.033	-0.028
Estatística F	887.501***	830.574***	862.589***	700.521***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A17 – Resultados das estimações com efeitos aleatórios para as mulheres

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.318*** (0.015)			-0.492*** (0.035)
<i>Mismatch</i> horizontal		-0.080*** (0.018)		-0.207*** (0.024)
<i>Mismatch</i> vertical x horizontal			-0.267*** (0.016)	0.187*** (0.037)
Norte	0.160 (0.169)	0.189 (0.174)	0.189 (0.169)	0.174 (0.167)
Nordeste	0.223* (0.121)	0.255** (0.124)	0.235* (0.121)	0.233* (0.120)
Centro-Oeste	0.566*** (0.069)	0.550*** (0.072)	0.562*** (0.070)	0.553*** (0.069)
Sudeste	0.277*** (0.036)	0.275*** (0.037)	0.284*** (0.036)	0.280*** (0.036)
Horas contratadas	0.021*** (0.001)	0.019*** (0.001)	0.020*** (0.001)	0.021*** (0.001)
Tempo empregado(a)	0.004*** (0.0001)	0.003*** (0.0001)	0.004*** (0.0001)	0.004*** (0.0001)
Constante	7.486*** (0.036)	7.486*** (0.045)	7.465*** (0.044)	7.649*** (0.040)
Observações	57,98	57,98	57,98	57,98
R <sup>2</sup>	0.306	0.220	0.284	0.321
R <sup>2</sup> Ajustado	0.306	0.220	0.284	0.321
Estatística F	2,415.570***	1,859.155***	2,252.581***	2,535.709***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A18 – Resultados das estimações OLS *pooled* para as mulheres

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.563*** (0.006)			-0.730*** (0.014)
<i>Mismatch</i> horizontal		-0.169*** (0.008)		-0.306*** (0.010)
<i>Mismatch</i> vertical x horizontal			-0.534*** (0.007)	0.180*** (0.016)
Norte	0.335*** (0.059)	0.470*** (0.062)	0.372*** (0.059)	0.343*** (0.058)
Nordeste	0.284*** (0.047)	0.403*** (0.050)	0.326*** (0.048)	0.311*** (0.047)
Centro-Oeste	0.690*** (0.024)	0.779*** (0.026)	0.702*** (0.024)	0.662*** (0.024)
Sudeste	0.351*** (0.012)	0.398*** (0.013)	0.372*** (0.013)	0.361*** (0.012)
Horas contratadas	0.024*** (0.0003)	0.018*** (0.0003)	0.022*** (0.0003)	0.026*** (0.0003)
Tempo empregado(a)	0.005*** (0.00004)	0.005*** (0.00005)	0.005*** (0.00004)	0.005*** (0.00004)
Constante	7.487*** (0.010)	7.629*** (0.012)	7.507*** (0.011)	7.673*** (0.012)
Observações	57,98	57,98	57,98	57,98
R <sup>2</sup>	0.317	0.231	0.300	0.329
R <sup>2</sup> Ajustado	0.317	0.231	0.300	0.329
Estatística F	3,837.712***	2,488.700***	3,542.966***	3,158.394***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A19 – Resultados das estimações com efeitos fixos para os homens

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.127*** (0.009)			-0.250*** (0.020)
<i>Mismatch</i> horizontal		-0.030*** (0.010)		-0.101*** (0.013)
<i>Mismatch</i> vertical x horizontal			-0.094*** (0.008)	0.136*** (0.021)

(continua)

(continuação)				
Variáveis	1	2	3	4
Norte	0.293*** (0.072)	0.293*** (0.072)	0.295*** (0.072)	0.289*** (0.072)
Nordeste	0.293*** (0.077)	0.288*** (0.077)	0.284*** (0.077)	0.303*** (0.077)
Centro-Oeste	0.315*** (0.038)	0.303*** (0.038)	0.307*** (0.038)	0.313*** (0.038)
Sudeste	0.227*** (0.023)	0.227*** (0.023)	0.227*** (0.023)	0.230*** (0.023)
Horas contratadas	0.019*** (0.0003)	0.019*** (0.0003)	0.019*** (0.0003)	0.020*** (0.0003)
Tempo empregado(a)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
Observações	50,819	50,819	50,819	50,819
R <sup>2</sup>	0.091	0.087	0.090	0.093
R <sup>2</sup> Ajustado	-0.053	-0.058	-0.055	-0.051
Estatística F	630.629***	598.089***	615.993***	498.074***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A20 – Resultados das estimações com efeitos aleatórios para os homens

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.300*** (0.016)			-0.485*** (0.039)
<i>Mismatch</i> horizontal		-0.070*** (0.020)		-0.188*** (0.026)
<i>Mismatch</i> vertical x horizontal			-0.242*** (0.017)	0.205*** (0.040)
Norte	0.445*** (0.128)	0.458*** (0.132)	0.454*** (0.129)	0.444*** (0.128)
Nordeste	0.428*** (0.137)	0.424*** (0.141)	0.413*** (0.138)	0.439*** (0.136)
Centro-Oeste	0.479*** (0.067)	0.454*** (0.069)	0.465*** (0.067)	0.474*** (0.066)
Sudeste	0.319*** (0.040)	0.319*** (0.041)	0.322*** (0.040)	0.324*** (0.040)

(continua)

(continuação)				
Variáveis	1	2	3	4
Horas contratadas	0.020*** (0.001)	0.019*** (0.001)	0.020*** (0.001)	0.021*** (0.001)
Tempo empregado(a)	0.003*** (0.0001)	0.003*** (0.0001)	0.003*** (0.0001)	0.003*** (0.0001)
Constante	7.655*** (0.034)	7.643*** (0.044)	7.630*** (0.044)	7.807*** (0.041)
Observações	50,819	50,819	50,819	50,819
R <sup>2</sup>	0.315	0.227	0.291	0.326
R <sup>2</sup> Ajustado	0.315	0.227	0.291	0.326
Estatística F	1,974.316***	1,482.074***	1,796.333***	2,044.370***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.

Tabela A21 – Resultados das estimações OLS *pooled* para os homens

Variáveis	1	2	3	4
<i>Mismatch</i> vertical	-0.592*** (0.007)			-0.773*** (0.016)
<i>Mismatch</i> horizontal		-0.166*** (0.009)		-0.278*** (0.011)
<i>Mismatch</i> vertical x horizontal			-0.549*** (0.007)	0.209*** (0.018)
Norte	0.487*** (0.049)	0.654*** (0.052)	0.530*** (0.050)	0.503*** (0.049)
Nordeste	0.442*** (0.051)	0.531*** (0.054)	0.442*** (0.051)	0.444*** (0.050)
Centro-Oeste	0.561*** (0.022)	0.578*** (0.024)	0.550*** (0.023)	0.560*** (0.022)
Sudeste	0.427*** (0.014)	0.502*** (0.015)	0.458*** (0.014)	0.437*** (0.014)
Horas contratadas	0.026*** (0.0003)	0.021*** (0.0003)	0.025*** (0.0003)	0.027*** (0.0003)
Tempo empregado(a)	0.005*** (0.00005)	0.005*** (0.0001)	0.005*** (0.00005)	0.005*** (0.00005)

(continua)

(continuação)				
Variáveis	1	2	3	4
Constante	7.584*** (0.012)	7.677*** (0.015)	7.585*** (0.013)	7.770*** (0.014)
Observações	50,819	50,819	50,819	50,819
R <sup>2</sup>	0.326	0.236	0.306	0.335
R <sup>2</sup> Ajustado	0.326	0.236	0.305	0.335
Estadística F	3,511.961***	2,247.878***	3,193.693***	2,839.907***

Fonte: Elaboração própria a partir de dados da UFPR e da RAIS.

Nota: \* Significante a 10%, \*\* Significante a 5%, \*\*\* Significante a 1%, em entre parênteses estão os erros-padrão.