

Rapunzel

**Uma aplicação do pacote *Blinder-Oaxaca*
Decomposition em R para estimar o prêmio
salarial STEM**

RESUMO

O papel desempenhado pela força de trabalho, que produz e difunde o avanço tecnológico, tem ganhado destaque desde que os modelos de crescimento econômico introduzem a noção de que é a associação entre o capital humano e o capital físico que explica o crescimento sustentado das nações modernas. Neste sentido, o presente trabalho se propõe a realizar uma aplicação do pacote *Blinder-Oaxaca Decomposition* em linguagem R, para quantificar e analisar as principais características dos trabalhadores formais que atuam em ocupações de aplicação direta de ciência e tecnologia - STEM (Science Technology Engineering and Math) no estado de Santa Catarina. Investiga-se ainda o comportamento dos diferenciais de remuneração desses trabalhadores em comparação com os demais – Não-STEM. Empregando a metodologia da decomposição de Oaxaca-Blinder, são avaliados os componentes internos e externos dos determinantes salariais, destacando características de idade e gênero. Foi verificada a existência do prêmio salarial para os trabalhadores do grupo STEM para homens e mulheres, sendo maior a diferença na remuneração para o gênero masculino. Em contrapartida, a característica de discriminação salarial ocorre em maior grau nas ocupações STEM e Não-STEM das mulheres. Verificou-se ainda que o componente de idade apresenta significativa distinção na determinação dos salários para os trabalhadores de aplicação de tecnologia em relação aos demais. A análise estatística por grupos de idade indicou que há crescimento do prêmio salarial STEM na medida em que se elevam as faixas de idade dos trabalhadores.

Palavras-chaves: Diferenciais salariais. Oaxaca-Blinder. Ciência e Tecnologia. STEM.

ABSTRACT

The role played by the labor force, that produces and disseminates technological advancement, has gained prominence since the models of economic growth introduced the notion that it is the association between human capital and physical capital that explains the sustained growth of modern nations. In this sense, the present work proposes to carry out an application of the Blinder-Oaxaca Decomposition package in R language, to quantify and analyze the main characteristics of formal workers who work in occupations with direct application of science and technology - STEM (Science Technology Engineering and Math) in Santa Catarina - Brazil. It also investigates the behavior of these workers' pay differentials compared to the others - Non-STEM. Using the Blinder-Oaxaca decomposition methodology, the internal and external components of wage determinants are evaluated, highlighting age and gender characteristics. The existence of the wage premium for workers in the STEM group for men and women was verified, with a greater difference in pay for men. In contrast, the wage discrimination feature occurs to a greater extent in women's STEM and non-STEM occupations. It was also found that the age component presents a significant distinction in determining wages for workers applying technology in relation to the others. The statistical analysis by age groups indicated that there is an increase in the STEM wage premium as workers' age ranges increase.

Keywords: Wage differentials. Blinder-Oaxaca. Science and technology. STEM.

LISTA DE ILUSTRAÇÕES

Figura 1 – Histograma do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018	27
Figura 2 – Função densidade do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018	28
Figura 3 – qqplot do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018	28
Figura 4 – Função densidade do logaritmo do salário-hora dos trabalhadores STEM e Não-STEM de Santa Catarina para o ano de 2018	29
Figura 5 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores de Santa Catarina para o ano de 2018	30
Figura 6 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores Não-STEM de Santa Catarina para o ano de 2018	31
Figura 7 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores STEM de Santa Catarina para o ano de 2018	31
Figura 8 – Relação de idade e salário-hora por gênero dos trabalhadores de Santa Catarina para o ano de 2018	32
Figura 9 – Relação de idade e salário-hora por gênero dos trabalhadores Não-STEM de Santa Catarina para o ano de 2018	33
Figura 10 – Relação de idade e salário-hora por gênero dos trabalhadores STEM de Santa Catarina para o ano de 2018	34
Figura 11 – Função densidade por idade e salário-hora dos trabalhadores de Santa Catarina para o ano de 2018	35
Figura 12 – Função densidade do salário-hora por categorias de idade dos trabalhadores de Santa Catarina para o ano de 2018	36
Figura 13 – Relação de idade e salário-hora dos trabalhadores STEM e Não-STEM de Santa Catarina para o ano de 2018	37
Figura 14 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM	38
Figura 15 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores homens e mulheres Não-STEM	40
Figura 16 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores homens e mulheres STEM	41
Figura 17 – Componentes de idade e ensino superior da decomposição da diferença salarial entre trabalhadoras STEM e Não-STEM	43

Figura 18 – Componentes de idade e ensino superior da decomposição da diferença salarial entre trabalhadores homens STEM e Não-STEM	44
Figura 19 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM	45
Figura 20 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM até 25 anos	47
Figura 21 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 26 e 35 anos	48
Figura 22 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 36 e 45 anos	49
Figura 23 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 46 e 55 anos	50
Figura 24 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM acima de 56 anos	51

LISTA DE TABELAS

Tabela 1 – Variáveis explicativas do modelo	21
Tabela 2 – Variáveis dummies de categoria de idade	22
Tabela 3 – Estatísticas descritivas dos trabalhadores formais em Santa Catarina - 2018	26
Tabela 4 – Estatísticas descritivas dos trabalhadores formais em Santa Catarina por faixa etária - 2018	26
Tabela 5 – Estatísticas descritivas do salário-hora em <i>log</i>	27
Tabela 6 – Estimativa da vantagem salarial do grupo STEM pela decomposição de Oaxaca	38
Tabela 7 – Desvantagem salarial feminina: trabalhadores Não-STEM	40
Tabela 8 – Desvantagem salarial feminina no grupo de trabalhadores STEM	41
Tabela 9 – Decomposição do salário das trabalhadoras STEM em relação às Não- STEM	42
Tabela 10 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM	43
Tabela 11 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando os intervalos de idade	45
Tabela 12 – Decomposição do salário dos trabalhadores STEM em relação aos Não- STEM considerando o intervalo até 25 anos de idade	46
Tabela 13 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 26 a 35 anos de idade	47
Tabela 14 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 36 a 45 anos de idade	48
Tabela 15 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 46 a 55 anos de idade	50
Tabela 16 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo acima de 55 anos de idade	51
Tabela 17 – Output Modelo Geral	59
Tabela 18 – Output Modelo Não-STEM	59
Tabela 19 – Output Modelo STEM	59
Tabela 20 – Output Modelo Feminino	60
Tabela 21 – Output Modelo Masculino	60
Tabela 22 – Output Modelo Geral agecat	60
Tabela 23 – Output Modelo agecat1	61
Tabela 24 – Output Modelo agecat2	61

Tabela 25 – Output Modelo agecat3	61
Tabela 26 – Output Modelo agecat4	62
Tabela 27 – Output Modelo agecat5	62
Tabela 28 – Teste do Fator de Inflação da Variância (FIV) para o Modelo Geral . .	63
Tabela 29 – Teste do Fator de Inflação da Variância (FIV) para modelo por categoria de idades	63
Tabela 30 – Teste do Fator de Inflação da Variância (FIV) para modelo com inserção de categorias de idade	64

SUMÁRIO

1	INTRODUÇÃO	1
2	REVISÃO DE LITERATURA	4
2.1	A moderna teoria do crescimento econômico	5
2.2	A teoria econômica dos diferenciais de salários	8
2.3	Hiato salarial entre gênero	9
2.4	Força de trabalho STEM	11
3	METODOLOGIA	14
3.1	Correspondência ocupacional	14
3.2	Decomposição do diferencial de rendimentos: Método de Oaxaca-Blinder	15
3.2.1	Abordagem Tríplice ou <i>Threefold</i>	16
3.2.2	Evitando o problema de multicolinearidade	17
3.2.3	Estimativa de decomposição: o comando <i>oaxaca()</i>	18
3.3	Tratamento dos dados	19
3.4	Modelos estimados	22
4	ANÁLISE DOS RESULTADOS	25
4.1	Descrição dos dados	25
4.2	Análise estatística do prêmio STEM em Santa Catarina	37
4.3	Inserção das categorias de idade para STEM e Não-STEM	45
4.4	Prêmio STEM para as categorias de idade	46
5	CONCLUSÃO	53
	REFERÊNCIAS	55
	APÊNDICE A – OUTPUTS DA DECOMPOSIÇÃO DE OAXACA-BLINDER	59
	APÊNDICE B – TESTES DE FATOR DE INFLAÇÃO DA VARIÂNCIA (FIV)	63
	APÊNDICE C – PROGRAMAÇÃO EM LINGUAGEM R	65
	ANEXO A – OCUPAÇÕES STEM	80

1 INTRODUÇÃO

O crescimento de longo prazo das economias é um dos problemas centrais da teoria econômica. No pioneiro modelo de crescimento de [Solow \(1956\)](#), a dinâmica de acumulação de capital respondia aos fatores de capital e trabalho, sendo o progresso tecnológico uma variável exógena ao modelo. Nesta configuração, o avanço tecnológico tem o papel de aumentar a efetividade do trabalho, elevando a produtividade e o nível de renda das economias. Além disso, o fator de progresso técnico seria o principal responsável pelo crescimento da renda per capita no longo prazo (crescimento persistente).

Desde então, o papel da tecnologia e do progresso técnico nos modelos de crescimento vêm ganhando cada vez mais destaque. Por estar alinhada à produtividade do trabalho, a qualidade da mão de obra das atividades ganharam ênfase. Em a Teoria do Capital Humano, [Schultz \(1967b\)](#) aponta o conhecimento como capital e a escolha de investimento de qualificação do indivíduo como uma resolução para o desenvolvimento e a produtividade.

Na mesma linha, [Becker \(1992\)](#) define o capital humano como um conjunto das habilidades incorporadas no conhecimento e na qualificação dos indivíduos. Assim, variáveis como a escolaridade, idade e tempo no emprego são determinantes de produtividade, de forma que o investimento em educação favorece o crescimento econômico e tem efeito sobre os rendimentos individuais.

Ao incrementar o modelo de crescimento neoclássico, a qualidade e a quantidade da mão de obra das atividades de aplicação direta de ciência e tecnologia passam a ganhar relevância para a difusão do progresso técnico e sucessivo desenvolvimento econômico dos países. Segundo classificação da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), estas atividades de aplicação direta da ciência e tecnologia nas atividades produtivas podem ser agrupadas pelo acrônimo STEM (Science, Technology, Engineering and Mathematics) – definição dada, por exemplo, por [Langdon et al. \(2011\)](#) e por [Marginson et al. \(2013\)](#) para uma comparação internacional.

Concebida por [Schultz \(1967b\)](#), o investimento em educação e na formação de capital humano passou a ser peça fundamental para a produtividade e crescimento das economias. Neste sentido, para alcançar maior progresso na corrida tecnológica, tanto os países da OCDE quanto os países emergentes têm buscado incentivar sua força produtiva STEM. Na China, por exemplo, o governo estabeleceu o objetivo de se equiparar aos principais centros inovadores até 2025 e, para tanto, tem expandido sua capacidade produtiva com formação nas carreiras STEM. Em 2016, o número de pessoas graduadas nas áreas STEM foi oito vezes maior do que nos Estados Unidos ([NOONAN, 2017](#)).

Além desta importância para o crescimento de longo prazo, a literatura tem reportado algumas características relevantes da força de trabalho STEM. Uma primeira é

o crescimento persistente da demanda por esses trabalhadores, inclusive em ritmo mais intenso do que a demanda por trabalhadores nas demais áreas. Esse crescimento forte tem garantido aos profissionais com qualificação nestas áreas maior média salarial do que o restante do mercado de trabalho, vantagem esta que tem sido chamada de prêmio salarial STEM.

Uma segunda característica observada, é a reduzida participação feminina entre os trabalhadores. Essa suposta perda do talento feminino para inovação tem levado os formuladores de políticas educacionais a prover incentivos para atrair o sexo feminino para as carreiras STEM.

Uma terceira característica do trabalho nas áreas de ciência e tecnologia é que o prêmio STEM das mulheres é maior do que o dos homens, o que implica que o *gap* salarial entre homens e mulheres seja menor do que na média do mercado de trabalho (BEEDE *et al.*, 2011). Em algumas ocupações STEM nos Estados Unidos, a desvantagem feminina é, de fato, revertida (DEY; HILL, 2007).

Do ponto de vista da oferta de trabalho, expoentes autores como Mincer (1958), Becker (1962) e Schultz (1961) argumentam que há relação positiva entre educação e renda pessoal, em que a qualificação amplia a capacidade produtiva e a remuneração dos trabalhadores. Por outro lado, o papel da discriminação no mercado de trabalho é reconhecido por Becker (1957), que discute as preferências discriminatórias por parte de consumidores, trabalhadores e empregadores. Desta maneira, é reconhecido na teoria econômica que a remuneração do fator trabalho pode ser tratada de maneira distinta, com base em critérios que não interferem na produtividade.

Desta forma, o problema de pesquisa norteador deste trabalho trata de avaliar os determinantes da remuneração do trabalho nas ocupações de ciência e tecnologia, destacando as diferenças de gênero e os papéis da qualificação e da experiência do trabalhador na definição de salário. Justifica-se o estudo pela relevância das ocupações STEM para a difusão da inovação e do desenvolvimento econômico e a problematização da remuneração destes trabalhadores, em que é reconhecida a crescente demanda por estes profissionais, a existência de um prêmio STEM, a baixa participação feminina nesta categoria e possível discriminação salarial.

Segundo o Tech Report 2020, o estado catarinense possui destaque na área de tecnologia do Brasil, sendo a quarta Unidade da Federação com maior faturamento e a terceira com maior número de trabalhadores (ACATE, 2020). Os números do ensino superior, do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), também sustentam o destaque catarinense, sendo o estado com maior participação de alunos em cursos STEM¹ do Brasil, com 31% (INEP, 2018).

¹ Houve mudança de metodologia do INEP na classificação dos cursos para 2018. A ACATE considerou os cursos de Agricultura, Engenharia, Ciências, Computação e Matemáticas, de acordo com os códigos CINE 51 a 83.

Desta forma, o objetivo principal desse trabalho é quantificar e analisar os determinantes da remuneração dos trabalhadores STEM em Santa Catarina no ano de 2018, informação mais recente dos dados do Ministério da Economia. Como objetivos secundários, coloca-se a descrição do perfil de gênero e a estimativa do prêmio salarial dos trabalhadores STEM para homens e mulheres. Além disso, objetiva-se analisar o prêmio STEM em diferentes faixas etárias, de modo a determinar a relevância do fator experiência na determinação das remunerações no estado de Santa Catarina.

A análise é realizada a partir do método Oaxaca-Blinder de [Oaxaca \(1973\)](#) e [Blinder \(1973\)](#). Esta metodologia permite a decomposição da diferença salarial em coeficientes relacionados às variáveis explicativas da remuneração e relacionados a possível discriminação entre grupos.

O trabalho está dividido em cinco capítulos, sendo a introdução o primeiro destes. No capítulo seguinte é apresentada a literatura que fundamenta a teoria do crescimento econômico e que discute o papel da força de trabalho STEM na economia. No terceiro capítulo é explicada a metodologia utilizada e posteriormente a análise dos resultados encontrados. Adiante, no quinto capítulo, serão apresentadas as considerações finais, seguidamente pelas referências utilizadas no presente trabalho. Adicionalmente, fazem parte do trabalho três apêndices. O primeiro conta com resultados da decomposição dos modelos abordados no capítulo 4, outro integra os testes para identificação de multicolinearidade e o último descreve os códigos em linguagem R utilizados na execução do trabalho. Por fim, um anexo traz os códigos que identificam as ocupações STEM.

2 REVISÃO DE LITERATURA

As principais abordagens teóricas que fundamentam esta proposta de trabalho têm origem na teoria do crescimento econômico e na teoria do capital humano. Em primeiro lugar, a teoria do crescimento desenvolvida a partir do modelo neoclássico é apresentada resumidamente como a fundamentação teórica acerca do desempenho econômico de longo prazo das nações. Em segundo lugar, coloca-se em discussão a baixa capacidade do modelo neoclássico de sustentar motivos que fazem com que nações já desenvolvidas sigam crescendo em ritmo maior que países em desenvolvimento. Cabe lembrar que o modelo neoclássico sugere a existência de um estado estacionário para cada economia, a qual irá crescer de acordo com a capacidade de rendimentos do capital. Assim, países em desenvolvimento teriam menores taxas de acúmulo de capital e, portanto, rendimentos marginais significativamente maiores que as nações desenvolvidas.

Um avanço na teoria do crescimento é a formulação dos modelos de crescimento endógeno, cujo ponto de partida é a assimilação da teoria do capital humano. Neste contexto, o entendimento deste fenômeno de crescimento sustentado dos países desenvolvidos é possibilitado pela introdução da teoria do capital humano no modelo de crescimento neoclássico. Ao incrementar o modelo de crescimento neoclássico, nos modelos de crescimento endógeno, a qualidade e a quantidade da mão de obra das atividades de aplicação direta de ciência e tecnologia, passam a determinar a capacidade das economias modernas de gerar e difundir inovações. Por sua vez, essa capacidade determina a posição de uma nação em termos da corrida tecnológica para se manter na liderança.

Para a economia brasileira, não há ainda uma literatura que investiga as áreas STEM enquanto um segmento específico do mercado de trabalho, ou do ponto de vista do processo de formação e qualificação profissional. A pesquisa sobre formação e trabalho nestas áreas ainda é incipiente, mas uma primeira aproximação aos dados sobre esse mercado indica que a economia brasileira apresenta características que replicam o perfil do trabalho STEM em outros países.

Justificam-se, então, duas fortes tendências observadas mundialmente, sobre as quais se fundamenta este trabalho. A primeira é o contínuo crescimento da demanda por educação em nível superior e a segunda é o incentivo à educação e ao trabalho nas atividades de aplicação direta de ciência e tecnologia, uma vez que estas são atividades que renderão os melhores frutos em termos de crescimento econômico. Além de sua importância na geração e difusão da inovação através do sistema produtivo e, portanto, na manutenção de um país no ritmo de crescimento econômico, outros aspectos desta força de trabalho têm sido reportados na literatura, que agrupam estas ocupações no acrônimo STEM.

Nas seções a seguir, estas teorias são brevemente apresentadas e contextualizadas

na discussão do papel do capital humano como determinante do crescimento econômico de longo prazo das economias modernas. Na sequência, a força de trabalho STEM é abordada e integrada ao contexto da teoria do capital humano.

2.1 A moderna teoria do crescimento econômico

O modelo de crescimento de Solow (1956) é uma das primeiras construções teóricas acerca do crescimento econômico de longo prazo das economias. O ponto de partida é uma função de produção que especifica o produto da economia sendo produzido a partir de bens de capital físico (tais como edifícios, máquinas, computadores, entre outros) e trabalho. Ao descrever a dinâmica da acumulação de capital físico e sua interação com a força de trabalho, o modelo propõe demonstrar que o produto per capita é uma função crescente da razão entre capital e trabalho.

Romer (1994) expõe a formulação de Solow a partir de quatro variáveis: produto (Y), capital (K), trabalho (L) e conhecimento (A), que resultam em um nível de produção.

$$Y(t) = F(K(t), A(t)L(t)) \quad (2.1)$$

Nota-se que o nível de produto só será alterado em um determinado período quando forem alterados os volumes de insumo da produção. Outro destaque é termo multiplicativo sobre o fator trabalho, que resulta que esse insumo passe a impactar o produto produzido já em termos de trabalho efetivo (AL). Seguindo o modelo de crescimento exógeno de Solow (1956), a função de produção apresenta retornos constantes de escala, de modo que os produtos marginais do capital e do trabalho são decrescentes. Uma função de produção que atende a estes requisitos específicos é a função de produção Cobb-Douglas a seguir:

$$F(K, AL) = K^\alpha (AL)^{1-\alpha}, 0 < \alpha < 1. \quad (2.2)$$

Ainda que considere as alterações tecnológicas (A) como fator explicativo do crescimento econômico, o modelo de Solow compreende este fator como exógeno ao desenvolvimento das economias. Oreiro (1999) ressalta que a eficiência do fator trabalho não é constante, ela pode aumentar com o passar do tempo, contudo, essa medida é determinada exogenamente. Além disso, outros pressupostos do modelo são que o progresso tecnológico aumenta a produtividade de todo o estoque de capital existente e, por condições de retornos constantes de escala, a renda per capita irá crescer à mesma taxa que o estoque de capital per capita. Na ausência de progresso tecnológico, a renda per capita seria constante no longo prazo. Mas na presença desse fator, ele se torna o principal elemento de crescimento da renda per capita no longo prazo (crescimento persistente).

Na função de produção com rendimentos constantes de escala (1), os fatores de produção possuem retornos marginais decrescentes. Assim, o crescimento de longo prazo

das economias deveria ser mais intenso em países com menor acúmulo de capital, pois estes ainda obteriam taxas de rendimento mais altas. Esta conclusão é resultado das condições básicas do modelo neoclássico acima - ver por exemplo [Lucas Jr. \(1990\)](#) para maiores esclarecimentos.

Ao contrário do resultado gerado pelo modelo de [Solow \(1956\)](#), [Romer \(1994\)](#) demonstra que nas últimas décadas os países menos desenvolvidos apresentaram crescimento menor, em média, que os países mais desenvolvidos. Além disso, verifica-se que este fluxo de capitais em direção a países de menor renda não tem sido a regra das transações econômicas, o que se situa em desacordo com o modelo de Solow. [Romer \(1994\)](#) sugere que a diferença de capital humano e as imperfeições de mercado são elementos que ajudam a explicar este fenômeno.

A solução para explicar o crescimento persistente das nações desenvolvidas começa a ser vislumbrada a partir das formulações de [Lucas Jr. \(1990\)](#) e [Romer \(1994\)](#), em que sinaliza para três possibilidades: i) admitir a existência de externalidades positivas de um incremento de capital sobre a tecnologia; ii) sugerir que o nível de tecnologia difere entre países por questões históricas e que a difusão desse conhecimento é lento e; iii) alocar um novo fator, capital humano (H), ao lado de capital e trabalho.

Para [Oreiro \(1999\)](#), a dificuldade de aplicação de Solow para explicar as diferenças entre as taxas de renda per capita pode ser explicada pela exogeneidade da tecnologia. De acordo com este autor, em modelos de crescimento endógeno, a taxa de renda per capita se torna uma variável que pode ser determinada dentro do modelo, ao invés de ser um parâmetro. [Romer \(1994\)](#) afirma que a distinção destes modelos está na ênfase de que o crescimento da renda é um efeito endógeno do sistema econômico e não no resultado de forças externas, como afirma Solow. Por esta razão, mudanças externas e não explicadas da tecnologia deixam de ser elementos que explicam as diferenças entre as rendas dos países.

Por conta dessas limitações, [Oreiro \(1999\)](#) afirma que o modelo de Solow foi expandido ao ser acrescentado um novo insumo à função de produção, o capital humano. Neste contexto, a moderna teoria do crescimento econômico pressupõe que a capacidade dos países em gerar prosperidade de forma contínua — ou seja, o crescimento sustentado — é resultado da capacidade da força de trabalho em compreender e estabelecer as novas tecnologias na atividade econômica. Em outras palavras, o modelo de crescimento econômico deve incorporar as noções de capital humano.

Assim, entende-se como a Moderna Teoria do Crescimento Econômico a área que passou a compreender a abordagem da Teoria do Capital Humano — estabelecida por [Becker \(1957\)](#) e [Schultz \(1967b\)](#) — integrada ao modelo básico neoclássico de crescimento econômico inicialmente descrita por [Solow \(1956\)](#).

Esta Teoria do Capital Humano foi concebida por Theodore Schultz a partir dos anos 50. Nela, o autor aponta o conhecimento como capital e assim, a escolha de investir na qualificação do indivíduo passa a ser uma resolução própria ou das partes interessadas em

desenvolver a produtividade. Schultz (1967b) atribui ao fator trabalho um peso maior, uma vez que o trabalhador que investir em conhecimento passará a ter maiores rendimentos, portanto o indivíduo investirá em educação na expectativa de retornos maiores. Para que a educação resulte em crescimento econômico, Schultz evidencia que os investimentos devem ser de boa qualidade e corretos (SCHULTZ, 1967b).

Schultz (1967b) ainda caracteriza a educação como um investimento, que tanto pode afetar as futuras despesas como as futuras rendas. Ao analisar os custos deste investimento, o autor descreve que há outros elementos envolvidos, uma vez que pessoas que antes apenas trabalhariam passam a optar pelo estudo - diminuindo a oferta de trabalho no curto prazo e deixando de obter rendimentos de salário neste período. Assim, Schultz concluiu que há custos de oportunidades para as pessoas e eventualmente para o governo, ao pagar a educação das pessoas sem que elas produzam:

“[...] o componente de produção, decorrente da instrução, é um investimento em habilidades e conhecimento que aumenta futuras rendas e, desse modo, assemelha-se a um investimento em (outros) bens de produção.” (SCHULTZ, 1967a, p.23)

Com o conceito de capital humano consolidado por Schultz, a educação se torna fundamental no contexto de expansão econômica. Como resultado, diversos pesquisadores passam a mostrar a importância do capital humano para o crescimento econômico, e qual sua relação com a educação e renda.

Com uma abordagem mais estruturada, Becker (1992) observou a relação custo-benefício das atividades que exploram a produtividade. O autor define capital humano como o conjunto das habilidades do indivíduo ligadas à capacidade produtiva, que são incorporadas no conhecimento e na qualificação para determinadas atividades. Para o autor, a escolaridade, idade e tempo no emprego são determinantes de produtividade, de forma que o investimento em educação favorece o crescimento econômico e tem efeito sobre os rendimentos individuais.

Conseqüentemente, diferentes níveis de capital humano são refletidos nas diferenças das remunerações, visto que no modelo neoclássico o salário deve ser análogo à produtividade, em virtude disto há uma identidade entre as habilidades e o salário.

Assim, a teoria do capital humano demonstra que investimentos em educação formal contribuem para o desenvolvimento das aptidões e habilidades cognitivas do indivíduo, de modo a torná-lo mais produtivo. Em um contexto de desenvolvimento, isto permite que exista uma relação positiva entre as habilidades e a produtividade do trabalhador, ligando-se à teoria do crescimento econômico por influenciar diretamente nas taxas de crescimento econômico das nações.

2.2 A teoria econômica dos diferenciais de salários

A teoria econômica apresenta algumas abordagens para explicar as desigualdades salariais que podem ser encontradas na teoria do capital humano através de [Schultz \(1961\)](#), [Becker e Chiswick \(1966\)](#) e [Arrow et al. \(1973\)](#). Uma delas é a teoria da segmentação ou mercado dual por [Doeringer e Piore \(1985\)](#) e [Vietorisz e Harrison \(1973\)](#). Outra abordagem é a teoria da discriminação estatística com trabalhos de [Phelps \(1972\)](#) e [Oaxaca e Dickinson \(2006\)](#). Por fim, na teoria da discriminação com [Oaxaca \(1973\)](#) e [Blinder \(1973\)](#), por exemplo. Dessa forma, estudos sobre a determinação e as diferenças de salários dos trabalhadores podem ser guiados tendo como ponto de partida argumentos teóricos firmados pelas diferenças de atributos produtivos, através da segmentação dos postos de trabalho e por meio da discriminação econômica e estatística.

A teoria de capital humano, do ponto de vista da oferta de trabalho, que tem como expoentes [Mincer \(1958\)](#), [Becker \(1962\)](#) e [Schultz \(1961\)](#), afirma a existência de uma relação positiva entre educação e renda pessoal, ou seja, os investimentos realizados pelos indivíduos em educação e qualificação ampliam a capacidade produtiva destes, fazendo com que haja no mercado diferentes dotações de salários. Neste caso, as diferenças salariais são resultantes das decisões individuais de acumulação de capital humano, de forma que estes investimentos em capital humano são fundamentais para a compreensão de desigualdades de rendas entre indivíduos, uma vez que a produtividade e os rendimentos de um trabalhador crescem em função do aumento da escolaridade e de suas habilidades adquiridas com a experiência ([BECKER, 1962](#)).

Em relação à discriminação no mercado de trabalho, [Becker \(1957\)](#) introduziu a concepção de preconceito dentro da noção de preferências e procurou mensurar os reflexos disso para a dinâmica do mercado de trabalho. Destacando a relevância da discriminação em relação a distribuição de salários entre gêneros, o autor pressupõe que a segregação ocupacional pode ser determinada pelo grau de discriminação por parte do empregador, evidenciando forte correlação entre segregação e discriminação¹.

Esta teoria econômica da discriminação argumenta que os trabalhadores são tratados de maneiras distintas com base em algum critério sem importância por causa dos grupos a que pertencem ([BECKER, 1962](#); [BECKER; CHISWICK, 1966](#)). Para [Loureiro \(2003\)](#), a discriminação ocorre quando um grupo de indivíduos possuem habilidades, treinamento adequado, educação, produtividade e experiência iguais, mas recebem salários e tratamentos diferentes devido à cor da pele, sexo, condições econômicas ou sociais.

Na teoria da segmentação do mercado de trabalho, também chamado de mercado duplo – ver por exemplo definição empregada por [Lima \(1980\)](#) – a diferença salarial ocorre quando os trabalhadores são tratados de maneira diferente devido às características

¹ O autor ainda discute as preferências discriminatórias em três diferentes situações: (i) discriminação por parte dos consumidores, (ii) discriminação por parte dos trabalhadores e (iii) discriminação por parte dos empregadores.

do trabalho. Um aspecto da explicação é fornecido por [Doeringer e Piore \(1985\)](#), que destacam as características pessoais e o comportamento dos indivíduos como determinantes do tipo de mercado em que eles serão inseridos e sua remuneração. Ou seja, a explicação das disparidades salariais ocorre pelo ajuste alocativo, no qual os atributos pessoais dos indivíduos determinam em qual mercado o indivíduo irá trabalhar, através do dualismo tecnológico ([VIETORISZ; HARRISON, 1973](#)), ou até mesmo por diferenças de classes sociais, que intensificam o comportamento e diferenças pessoais nas características da força de trabalho ([REICH; GORDON; EDWARDS, 1973](#)). Estes três aspectos da teoria da segmentação se complementam e mantêm a perspectiva de demanda e trabalho, fazendo com que existam inúmeras evidências empíricas das diferenças salariais no mercado de trabalho e com diferentes contribuições e fatores para explicar essas desigualdades, relacionadas ao capital humano, questões de gênero e raça, setores econômicos ou entre regiões geográficas.

Então, a desigualdade de salários pode resultar de condições como a heterogeneidade dos trabalhadores no que tange as suas características produtivas, podendo também estar relacionada com o fato de trabalhadores com igual qualificação obterem remunerações diferentes, seja por discriminação - termo empregado para tratamento diferenciado respaldado em atributos não produtivos - seja porque o mercado de trabalho é segmentado - situação em que os postos de trabalho prestigiam os atributos dos trabalhadores de modo distinto.

Assumidas então as características de produtividade dos trabalhadores e mediante a hipótese de ausência de salários compensatórios, a permanência de lacunas salariais é atribuída pela discriminação. Segundo [Arrow et al. \(1973\)](#), há discriminação econômica para indivíduos de um grupo sempre que as remunerações sejam inferiores, já havendo o desconto das diferenças pelas habilidades individuais provocadas pelo empregador ou pelo funcionário. Estudos como o de [Phelps \(1972\)](#) ainda apresentam a discriminação estatística derivada da disfunção de informações imperfeitas do mercado sobre as características produtivas e o potencial do empregado onde, nesse caso, é atribuído valor ao trabalhador tendo como base a média do grupo a que este pertence.

2.3 Hiato salarial entre gênero

Nos últimos anos, conforme a mulher passou a conquistar maiores oportunidades na esfera do mercado de trabalho, constatou-se a concepção de outras pesquisas que trazem uma reflexão sobre as discrepâncias nos rendimentos segundo gênero. A literatura que aborda este tratamento de diferencial de salários, estabelece que a existência do hiato de rendimentos entre homem e mulher pode ocorrer devido à segmentação ocupacional no mercado de trabalho ou às evidências de discriminação presentes.

Na literatura empírica que discute a discriminação de gênero, [Blau e Kahn \(2000\)](#) descrevem a evolução do hiato salarial e da discriminação nos Estados Unidos no período compreendido entre 1978 e 1998. Eles mostram que houve uma atenuação nas diferenças

salariais entre os gêneros, embora houvesse persistência da discriminação feminina. Tal estudo, que também inclui variáveis que captam as condições de mercado de trabalho, revela uma mudança nos retornos à qualificação favorecendo os trabalhadores mais qualificados. Como as mulheres ocupavam trabalhos menos qualificados, os autores sustentam que essa mudança desfavorecia as mulheres no período estudado.

Um estudo de [Cobb-Clark e Tan \(2011\)](#) do início da década de 2000 obteve um resultado significativo oriundo dos efeitos das habilidades não cognitivas² sobre a probabilidade de emprego no mercado de trabalho australiano, muito embora as mulheres obtenham salários mais baixos do que os homens empregados na mesma ocupação, caracterizando a segregação no mercado de trabalho.

Do mesmo modo ocorre em outros países – ver por exemplo [Biltagy \(2014\)](#) para estimações de disparidade salarial no Egito por meio do método de decomposição de Oaxaca-Blinder. Já [Fonseca \(2017\)](#) aborda uma diferença salarial que favorece os homens no período entre 2007 e 2012 em Portugal, onde parte é explicada principalmente por práticas discriminatórias, apesar da redução na discriminação salarial, devido às melhores dotações produtivas das trabalhadoras. Ainda mediante às perspectivas das diferenças salariais entre os sexos, alguns estudos empíricos nacionais como [Souza \(1978\)](#), [Cugini et al. \(2014\)](#), [Salvato et al. \(2008\)](#) comprovam a diferenciação e a discriminação de salários no mercado de trabalho entre homens e mulheres. Os estudos também reforçam a hipótese de que as desigualdades salariais e a discriminação de gênero são diferentes conforme o grau de heterogeneidade do país, fazendo com que estas discriminações sejam menores em regiões menos desenvolvidas economicamente.

A revisão de literatura apresentada até aqui buscou esclarecer as explicações para a disparidade salarial entre os gêneros e fatores tradicionalmente enfatizados, assim como as explicações oferecidas por trabalhos de autores expoentes. Discussões sobre a notável reversão da diferença de gênero presente hoje nas universidades, bem como o crescente compromisso das mulheres com a força de trabalho remunerada também são temas abordados por estes autores. Outro enfoque relevante que agrega maior valor aos resultados está relacionado ao conceito de segregação ocupacional, no qual homens e mulheres estão inseridos em distintas ocupações de trabalho. Esta segregação ocupacional favorece a diferença salarial total entre gênero, mas não necessariamente se classifica como discriminação. Uma vez que não haja controle desse fator, então, pode ocorrer superestimação do efeito da discriminação salarial. Diversos estudos abordam discussões de diferenças salariais de gênero residuais, que não podem ser explicadas facilmente. Possivelmente essas diferenças inexplicáveis indicam onde está a discriminação salarial.

O recente estudo de [Stein, Sulzbach e Bartels \(2015\)](#) analisou 100 mil salários de

² De acordo com [Lipnevich e Roberts \(2012\)](#) as habilidades não cognitivas são habilidades que compreendem construtos de diferentes categorias, como atitudes, crenças, qualidades emocionais e sociais e traços de personalidade.

trabalhadores no Rio Grande do Sul e concluiu que as mulheres brasileiras ganham 20% menos que os homens — mas apenas 7% não podem ser explicados pela diferença de produtividade. Segundo os autores, o diferencial das previsões dos logaritmos dos salários é de 0,202, o que pode ser interpretado como um diferencial salarial de 20,2% entre homens e mulheres. Destes 20,2%, 13,5 pontos percentuais (p.p.) são explicados pelas diferenças nas características individuais, ou seja, aproximadamente dois terços (66,8%) do diferencial salarial se devem às covariadas utilizadas no exercício empírico, de modo que apenas 6,7 p.p dizem respeito a fatores não observados. A conclusão deste estudo converge com os dados encontrados em [Goldin \(2006\)](#) - para os Estados Unidos, encontrou-se uma porcentagem um pouco menor (5%) que não é explicada pela produtividade.

Em vista de uma persistente diferença salarial que não pode ser explicada, deu-se início a uma investigação buscando observar se essas diferenças também podem ser encontradas de forma homogênea entre diferentes ocupações. Assim, alguns trabalhos sobre análise ocupacional da estrutura salarial buscaram observar o comportamento dos salários para carreiras de aplicação direta à tecnologia. Como apontado na seção anterior, esse segmento do mercado de trabalho está diretamente ligado à difusão da ciência e tecnologia e de sua aplicação na atividade produtiva. Assim, o grupo de ocupações STEM tem papel relevante no desenvolvimento econômico. Por conta disso, a próxima seção busca conceituar o grupo de trabalho STEM e elencar as principais características associadas a ele.

2.4 Força de trabalho STEM

Na força laboral que atua nas áreas de aplicação direta da ciência e tecnologia estão os profissionais responsáveis pela produção e difusão de novas tecnologias ([GOLDIN et al., 1992](#)). A importância desse segmento da força produtiva de uma nação se associa ao fato de que, como apontado na seção anterior, os modelos de crescimento endógeno — [Romer \(1990\)](#) e [Lucas Jr. \(1988\)](#) — sugerem que a geração de progresso tecnológico e a transmissão das mudanças tecnológicas determinam formalmente a persistência do crescimento dos países. Desse modo, a inovação, através da invenção, desenvolvimento e profusão de novas tecnologias é a fonte fundamental do progresso econômico, o que se intensifica na atuação das ocupações que requerem o conhecimento da área STEM.

Assim, as economias fortemente baseadas nas atividades STEM apresentam um melhor desempenho em indicadores econômicos e em termos de inovação e geração de novas vagas de empregos. As taxas de emprego, de patentes registradas, de salários e até mesmo de exportações são mais elevadas nessas economias ([ROTHWELL, 2013](#)).

Do ponto de vista do mercado de trabalho, o crescimento persistente da demanda por trabalhadores STEM em ritmo mais intenso do que os demais — enquanto que a oferta de trabalhadores com qualificação nestas áreas cresce no mesmo ritmo das demais áreas

— tem garantido a estes profissionais maior média salarial do que o restante da força de trabalho. [Beede et al. \(2011\)](#) têm abordado esta vantagem salarial como prêmio STEM.

A força de trabalho STEM também se caracteriza pela maior estabilidade empregatícia, uma vez que trabalhadores com formação nas áreas STEM apresentam menores taxas de desemprego. Além disso, durante as recessões, o aumento da taxa de desemprego nas atividades STEM é menor do que no conjunto das demais ocupações ([BONINI; CUSTODIO; SILVA, 2019](#)).

Outro fato observado na força de trabalho STEM é a baixa representatividade das mulheres nessas áreas ([BEEDE et al., 2011](#)). Essa suposta perda do talento feminino para inovação tem levado os formuladores de políticas educacionais a prover incentivos para atrair o gênero feminino para as carreiras STEM. A convergência de capital humano entre gêneros é parte da mudança na estrutura socioeconômica causada pelo exercício de atividade econômica por parte das mulheres. [Goldin \(2006\)](#) descreve as etapas desse processo, que se iniciou ainda no final do século XIX e se tornou uma “revolução silenciosa” ainda em curso atualmente. Cabe destacar que, inicialmente, a entrada das mulheres na força de trabalho teve caráter hegemonicamente econômico e prevaleceu em atividades de baixa qualificação e remuneração.

Enquanto a população feminina também buscava melhor capacitação, passou a ampliar o contingente de matrículas nas universidades americanas e, subsequentemente, europeias. Tal fenômeno demarcou que a entrada das mulheres na força economicamente produtiva das nações seria uma mudança que viria a ultrapassar a esfera econômica. Assim, as mulheres puderam expandir seus horizontes de escolha, reforçar seu papel nos domínios público e privado e se habilitaram a tomar decisões, tanto na esfera pessoal quanto no ambiente de trabalho.

A presença das mulheres nas universidades é uma das principais conquistas femininas do último século, sendo maioria entre os estudantes matriculados em cursos de graduação. De acordo com o relatório *Education at a Glance 2019* da OCDE, enquanto 18% dos homens brasileiros de 25 a 34 anos têm ensino superior, essa porcentagem sobe para 25% entre as mulheres da mesma faixa etária. Apesar disso, o Brasil ainda se encontra abaixo das médias dos países da OCDE, de 38% para homens e 51% para mulheres, segundo dados de 2018. Tal disparidade se observa em outros países e tem aumentado nas gerações mais novas, constatando a existência de uma tendência global de mulheres constituírem maioria no ensino superior. Ressalta-se que a diferença pode ser observada tanto no acesso ao curso, onde mais mulheres entram, quanto na conclusão. ([OECD, 2019](#)).

Ainda de acordo com o relatório, para o mercado de trabalho brasileiro, a empregabilidade de mulheres brasileiras de 25 a 34 anos com ensino superior é de 82% e cai para 63% entre mulheres com ensino técnico e para 45% entre mulheres sem essa capacitação. Já entre homens brasileiros, esses índices são mais altos, tornando a taxa de empregabilidade dos que têm ensino superior de 89%; de 76% dos que têm ensino técnico e 76% dos que não

tem nenhuma formação superior. O que explicaria, então, o fato de que mulheres tenham mais dificuldades em se inserir no mercado de trabalho é que elas estão significativamente mais representadas nos campos de educação e ciências sociais, enquanto que homens estão mais alocados em campos como tecnologias da informação e engenharias.

No Brasil especificamente, observa-se que a taxa de emprego dos homens aumenta pouco com maiores níveis de escolaridade. Já para mulheres, a taxa de emprego aumenta consideravelmente - podendo ser esse um incentivo maior para elas cursarem o ensino superior.

Ainda que não haja uma maneira aparente de incentivo ou coerção, as escolhas das mulheres resultam numa especialização por gênero na universidade e no mercado de trabalho, fazendo com que as mulheres se afastem de algumas áreas técnicas, tais como engenharias e computação. Como consequência das escolhas de cursos superiores, a ocupação dos postos de trabalho por parte das mulheres também ocorre de maneira heterogênea fazendo com que as mulheres apresentem participação no grupo de ocupações STEM bastante abaixo da média do mercado como um todo (GOLDIN, 2006). Este fenômeno tem motivado órgãos governamentais e empresas a buscarem mecanismos para incentivar a população feminina para essas áreas.

Nos Estados Unidos, a representação feminina em ocupações STEM aumentou desde a década de 1970, mas ainda permanece significativamente sub-representada em ocupações de engenharia e informática - ocupações responsáveis por mais de 80% de todos empregos STEM. As décadas mais recentes mostram menor crescimento nos empregos STEM entre as mulheres mais jovens. O máximo de crescimento da participação das mulheres abaixo de 40 anos nestas áreas ocorreu entre 1970 e 1990. (LANDIVAR, 2013). O mesmo relatório conclui que entre os graduados em ciências e engenharias, os homens que possuem empregos em uma ocupação STEM são dobro da taxa de mulheres: 31% contra 15% - cerca de 1 em cada 5 mulheres graduadas em ciências e engenharias estão fora da força de trabalho, em comparação com menos de 1 em cada 10 homens de mesma formação.

3 METODOLOGIA

Este capítulo tem como objetivo apresentar os procedimentos metodológicos que serão adotados neste trabalho e descrever o banco de dados analisado. Na seção 3.1 se apresenta a correspondência ocupacional utilizada pela Relação Anual de Informações Sociais (RAIS), do Ministério da Economia, por meio da Classificação Brasileira de Ocupações (CBO), onde foram identificadas as profissões de acordo com seus campos de estudo. A seção 3.2 traz o desenvolvimento e aplicação do método de decomposição de diferencial de rendimentos Oaxaca-Blinder. É fornecido ainda o tratamento das informações e suas adequações aos modelos propostos. A seção 3.3 apresenta a descrição dos dados e, por fim, a seção 3.4 aborda como serão estimados os modelos.

Considerando o objetivo proposto, a metodologia adotada nesta pesquisa é de natureza teórica-empírica e do tipo exploratória-descritiva, com abordagem transversal. Um estudo transversal é um método de pesquisa que analisa as variações nas características de uma população com dados coletados em um único período. (CRESWELL, 2010; SAMPIERI et al., 2010)

Esta definição tem como fundamento as definições de Gil (2002, p. 41), que afirma que “pesquisas exploratórias têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses”, enquanto pesquisas descritivas têm por objetivo descrever as características de determinada população ou de algum fenômeno (GIL, 2002).

As pesquisas exploratórias objetivam a familiarização com um assunto que é pouco explorado. Em pesquisas deste tipo, geralmente há pouco conhecimento do tema a ser abordado. As pesquisas descritivas, por sua vez, têm por objetivo descrever as características de uma população de um fenômeno ou de uma experiência (GIL, 2002).

3.1 Correspondência ocupacional

Os dados utilizados neste trabalho têm como origem a Relação Anual de Informações Sociais (RAIS), do Ministério da Economia, para o ano de 2018. Nas informações sobre a ocupação do trabalhador, disponibilizadas pela RAIS, considera-se a Classificação Brasileira de Ocupações (CBO), estabelecida em 2002 pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A CBO 2002 é um documento que normaliza, nomeia e codifica os títulos e os conteúdos das ocupações do mercado de trabalho brasileiro, sendo ao mesmo tempo uma classificação enumerativa e descritiva.

A estrutura que compõe a CBO é hierárquica e se divide em 10 grandes grupos, além de outras divisões em subgrupos e famílias ocupacionais, que compreendem o total de 2.422

ocupações e de 7.258 títulos sinônimos. Dentre elas, identificam-se 164 códigos de ocupações que correspondem ao grupo STEM. São consideradas as ocupações de suporte profissional e técnico nos campos de ciências da computação, matemática, engenharias, ciências físicas e da terra (BEEDE et al., 2011), enquanto áreas como programação de computadores, engenharia mecânica e ciências ambientais são consensualmente pertencentes ao conjunto STEM¹.

A definição do grupo de ocupações que compõem o acrônimo STEM para o Brasil segue Seemann e Bonini (2017), também empregada por Custodio e Bonini (2019), que são os primeiros estudos com objetivo de sistematizar um grupo de ocupações STEM dentro da classificação adotada no Brasil. Seemann e Bonini (2017) seguem os critérios do Departamento de Estatística do governo americano - Economics and Statistics Administration (ESA), mas como sugere Davies (2014), não há definição padrão para o conjunto de áreas de formação e atuação profissional STEM.

3.2 Decomposição do diferencial de rendimentos: Método de Oaxaca-Blinder

Para estimar o prêmio salarial das ocupações STEM, a metodologia a ser utilizada é a decomposição Oaxaca-Blinder, introduzida pela primeira vez pelos economistas Ronald Oaxaca e Alan Blinder em Oaxaca (1973) e Blinder (1973) e adaptada para comparar os grupos STEM e Não-STEM. Este modelo se originou e tem sido amplamente utilizado no estudo da discriminação no mercado de trabalho. Embora as decomposições de Oaxaca-Blinder tenham sido um dos pilares da pesquisa empírica sobre discriminação, elas podem ser, em princípio, aplicadas para explicar as diferenças em qualquer resultado contínuo entre dois grupos.

A decomposição de Oaxaca-Blinder é um método econométrico que decompõe diferenças nos resultados médios entre dois grupos, sendo que esta primeira parte é devida à diferenças de grupo nos níveis de variáveis explicativas e a outra parte se deve às magnitudes diferenciais dos coeficientes de regressão - ver por exemplo Jann (2008) e Hlavac (2014).

O pacote empregado no tratamento de dados nesta pesquisa trabalha com distinções entre grupos. Para apresentação da metodologia, por escolha da autora, os grupos serão denominados como Grupo A e Grupo B, somente para fins de diferenciação. Nos modelos gerais, o grupo A será identificado como Não-STEM e o grupo B como STEM, enquanto que para os modelos com distinção de gênero, o grupo A será composto por homens e o grupo B por mulheres.

¹ De acordo com Langdon et al. (2011), há quem considere ainda as ocupações relacionadas à medicina, arquitetura e ciências sociais, mas estas aplicações não são consenso entre os pesquisadores e não são consideradas nesta pesquisa.

Desta forma, a diferença média de resultados a ser explicada é a diferença dos resultados médios das observações no Grupo A e no Grupo B:

$$\Delta\bar{Y} = \bar{Y}_A - \bar{Y}_B \quad (3.1)$$

3.2.1 Abordagem Tríplice ou *Threefold*

No contexto de uma regressão linear, o resultado médio para o Grupo $G \in \{A, B\}$ pode ser expresso como $\bar{Y}_G = \bar{X}'_G \hat{\beta}_G$, onde \bar{X}_G contém os valores médios das variáveis explicativas e $\hat{\beta}_G$ são os coeficientes de regressão estimados. Assim, a expressão acima (3.1) pode ser reescrita como:

$$\Delta\bar{Y} = \bar{X}'_A \hat{\beta}_A - \bar{X}'_B \hat{\beta}_B \quad (3.2)$$

Essa expressão pode, por sua vez, ser escrita como a soma dos três termos a seguir:

$$\Delta\bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' \hat{\beta}_B}_{\text{endowments}} + \underbrace{\bar{X}'_B (\hat{\beta}_A - \hat{\beta}_B)}_{\text{coefficients}} + \underbrace{(\bar{X}_A - \bar{X}_B)' (\hat{\beta}_A - \hat{\beta}_B)}_{\text{interaction}} \quad (3.3)$$

Esta é a decomposição tríplice ou *threefold* de Oaxaca-Blinder da diferença média de resultados. O termo *endowments* representa a contribuição das diferenças nas variáveis explicativas de cada grupo, e o termo dos *coefficients* (coeficientes) é a parte que se deve às diferenças de grupo nos coeficientes. Por fim, o termo de interação (*interaction*) explica o fato de que diferenças entre grupos em variáveis e coeficientes explicativos podem ocorrer ao mesmo tempo.

A decomposição tríplice também pode ser estimada separadamente para cada variável explicativa:

$$\underbrace{(\bar{X}_A - \bar{X}_B)' \hat{\beta}_B}_{\text{endowments}} = \underbrace{(\bar{X}_{1A} - \bar{X}_{1B}) \hat{\beta}_{1B}}_{\text{variable 1}} + \underbrace{(\bar{X}_{2A} - \bar{X}_{2B}) \hat{\beta}_{2B}}_{\text{variable 2}} + \dots \quad (3.4)$$

$$\underbrace{\bar{X}'_B (\hat{\beta}_A - \hat{\beta}_B)}_{\text{coefficients}} = \underbrace{\bar{X}_{1B} (\hat{\beta}_{1A} - \hat{\beta}_{1B})}_{\text{variable 1}} + \underbrace{\bar{X}_{2B} (\hat{\beta}_{2A} - \hat{\beta}_{2B})}_{\text{variable 2}} + \dots \quad (3.5)$$

$$\underbrace{(\bar{X}_A - \bar{X}_B)' (\hat{\beta}_A - \hat{\beta}_B)}_{\text{interaction}} = \underbrace{(\bar{X}_{1A} - \bar{X}_{1B}) (\hat{\beta}_{1A} - \hat{\beta}_{1B})}_{\text{variable 1}} + \underbrace{(\bar{X}_{2A} - \bar{X}_{2B}) (\hat{\beta}_{2A} - \hat{\beta}_{2B})}_{\text{variable 2}} + \dots \quad (3.6)$$

A equação (3.6) resultará em formulações gerais do modelo, apresentando os coeficientes e os desvios-padrão. Ou seja, os resultados da abordagem *threefold* são apresentados no agregado das variáveis. Estes dois valores são apresentados para os tipos *endowments*, *coefficients* e *interaction*. Nisto consiste a decomposição de Oaxaca-Blinder, que separa os valores dos coeficientes estimados (betas) a partir da contribuição dos diferentes fatores.

Os resultados intitulados "*endowments*" são, portanto, as diferenças de salário explicadas pela produtividade do trabalhador ou variáveis internas do modelo. Adotando as formulações da teoria do capital humano, nesse trabalho, essas dotações de capital humano de cada grupo corresponderão a características de idade, o grau de escolaridade e o tempo no emprego. Já os resultados intitulados "*coefficients*" podem ser entendidos como o componente do salário não explicado pelas características de produtividade, uma vez que corresponde à parte não explicada pelas variáveis internas do modelo. No caso desse trabalho, esse componente dos resultados deverá estar associado ao prêmio STEM e à desvantagem salarial de gênero. Por fim, o grupo "*interaction*" é a interação destes dois, que trará o quanto cada variável (contemplando tanto as diferenças das variáveis internas como os coeficientes) contribui para a formação salarial de cada grupo.

3.2.2 Evitando o problema de multicolinearidade

Para os resultados das decomposições foram consideradas variáveis *dummies* de identificação de trabalhadores com ensino superior incompleto, completo, pós-graduação, gênero e grupo de ocupação STEM e Não-STEM. De acordo com Gujarati (2000), o uso destas variáveis categóricas pode acarretar em multicolinearidade perfeita. Este problema não gera estimativas viesadas ou altera a forma da matriz de variância-covariância, mas superestima os desvios padrões de cada coeficiente. Para evitar a colinearidade perfeita, recomenda-se a construção de uma linha de base, de modo que a omissão de todas as variáveis *dummies* representem esta categoria base. Em função disso, os modelos de decomposição de salários deste trabalho utilizam como base os trabalhadores que tenham até a escolaridade básica completa. Para os modelos de grupos STEM e Não-STEM, o modelo base se refere a trabalhadores Não-STEM, enquanto que os modelos de gênero, a base indica trabalhadores do gênero masculino.

Um modelo de regressão linear que contém uma variável explicativa categórica pode, portanto, ter a seguinte forma geral:

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \dots + \beta_{k-1} D_{k-1} + X' \gamma + \varepsilon \quad (3.7)$$

onde D_i tal que $i = 1, \dots, k - 1$ são variáveis indicadoras que representam níveis individuais de uma variável categórica. A categoria k é a linha de base omitida.

Para garantir que os resultados da decomposição de Oaxaca-Blinder sejam invariáveis à escolha da categoria de linha de base omitida, ainda é implementado um procedimento

que transforma o modelo de regressão acima em:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 D_1 + \tilde{\beta}_2 D_2 + \tilde{\beta}_3 D_3 + \dots + \tilde{\beta}_{k-1} D_{k-1} + \tilde{\beta}_k D_k + X' \gamma + \varepsilon \quad (3.8)$$

De forma que:

$$\alpha = \sum_{j=1}^{k-1} \beta_j / k \quad (3.9)$$

Assim, o valor do ajuste é adicionado ao intercepto original:

$$\tilde{\beta}_0 = \beta_0 + \alpha \quad (3.10)$$

E subtraído de cada um dos outros coeficientes da regressão:

$$\tilde{\beta}_i = \beta_i - \alpha \quad (3.11)$$

para $i = 1, \dots, k$. Os coeficientes ajustados, bem como os resultados das decomposições detalhadas de variável por variável, permanecerão os mesmos, independentemente da escolha da categoria omitida² k .

3.2.3 Estimativa de decomposição: o comando *oaxaca()*

Nesta subseção será apresentada a metodologia para obtenção dos resultados utilizando o pacote *Blinder-Oaxaca Decomposition* de [Hlavac \(2014\)](#) para linguagem R. Todos os códigos desenvolvidos para esta monografia estão no Apêndice C.

A função principal *oaxaca()* executa as variantes tríplices e duplas da decomposição de Oaxaca-Blinder usando observações da base de dados fornecidas no argumento *data*. O modelo de regressão linear para esta decomposição é especificado por meio do argumento *formula*. Definem-se então a variável dependente (*y*), as variáveis explicativas (*x1*, *x2*, *x3*, etc.), bem como um indicador variável (*z*) que indica se uma observação pertence ao Grupo A (quando *z* é FALSO ou 0) ou Grupo B (quando é igual a VERDADEIRO ou 1). Essas variáveis, junto com a forma funcional do modelo, ocorrem no argumento *formula* no objeto da classe *Formula* do pacote de [Zeileis e Croissant \(2010\)](#).

² Este caso de categoria omitida é comum quando se trata de variáveis explicativas categóricas. Sugestões para correção do problema são propostas por [Oaxaca e Ransom \(1999\)](#) e [Yun \(2005\)](#), [Yun \(2008\)](#).

Tipicamente, o modelo assume a seguinte forma funcional:

$$y \sim x_1 + x_2 + x_3 + \dots | z \quad (3.12)$$

Se o modelo de regressão contém *dummies* que representam uma variável categórica (d1, d2, d3, etc.), elas podem ser especificadas adicionando outra parte à fórmula:

$$y \sim x_1 + x_2 + x_3 + \dots | z | + d_1 + d_2 + d_3 + \dots \quad (3.13)$$

Quando variáveis categóricas são especificadas, a função *oaxaca ()* irá ajustar automaticamente as estimativas para serem invariáveis em relação à escolha da categoria de base omitida. Caso não haja nenhum outro argumento, o comando *oaxaca ()* irá estimar a decomposição com base em regressões de Mínimos Quadrados Ordinários (MQO), estimado por meio da função *lm³ ()* padrão, e calculará os erros padrão com base em 100 replicações de *bootstrap*⁴.

3.3 Tratamento dos dados

As variáveis explicativas na equação de determinação do salário são: idade do trabalhador, tempo de emprego na empresa, escolaridade, gênero e ocupação. Inclui-se ainda no modelo a idade ao quadrado para capturar o fato de que o retorno associado à idade apresenta retornos marginais decrescentes.

O modelo econométrico de regressão típico decorrente da equação minceriana, desenvolvido por [Mincer \(1974\)](#) segue a forma funcional:

$$\ln w = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot exp + \beta_3 \cdot exp^2 + \gamma' \cdot x + \varepsilon \quad (3.14)$$

onde:

- $\ln w$ é o salário recebido pelo indivíduo;
- β_0 é o coeficiente de intercepto, que representa a heterogeneidade atribuída a variáveis não observadas no modelo e comuns a todos indivíduos;
- $\beta_1 educ$ é a escolaridade, geralmente medida por anos de estudo;
- $\beta_2 exp$ é experiência, geralmente aproximada pelo idade do indivíduo;
- $\gamma' x$ é um vetor de características observáveis do indivíduo;

³ *lm*: Linear Model

⁴ *Bootstrap* consiste em um método de reamostragem proposto por Bradley Efron em 1979. Utiliza-se para aproximar distribuição na amostra de um levantamento estatístico. Esta ferramenta é utilizada frequentemente para aproximar o viés ou a variância de um conjunto de dados estatísticos, assim como para construir intervalos de confiança ou realizar contrastes de hipóteses sobre parâmetros de interesse.

- ε é um erro estocástico.

Este é um modelo de regressão no formato log-linear: a variável dependente (salário), está em formato logarítmico e a variável independente mais relevante, a escolaridade, está em nível. Portanto, o coeficiente β_1 mede quanto um ano a mais de escolaridade causa de variação proporcional no salário no indivíduo. Em termos gerais, por exemplo, se β_1 é estimado em 0,12, isto quer dizer que cada ano a mais de estudo está relacionado, em média, com um aumento de salário de 12%.

Então, derivando a equação (3.14), encontra-se que:

$$\frac{\partial \ln w}{\partial educ} = \beta_1 \quad (3.15)$$

Enquanto que ao aplicar a regra da cadeia é obtido:

$$\frac{\partial \ln w}{\partial educ} = \frac{\partial w}{\partial educ} \cdot \frac{1}{w} = \frac{\frac{\partial w}{\partial educ}}{w} \quad (3.16)$$

Logo:

$$\beta_1 = \frac{\frac{\partial w}{\partial educ}}{w} \quad (3.17)$$

corresponde, portanto, a variação percentual do salário decorrente de cada acréscimo unitário de ano de estudo.

O coeficiente da regressão minceriana, quando somente com a constante e uma variável específica (tal como a variável *educação*) fornecerá o prêmio bruto relativo - não controlado - em termos de variação de renda. Já o coeficiente de uma variável de uma regressão minceriana multivariada (log-linear com constante e uma série de variáveis adicionais) produzirá o prêmio marginal bruto relativo controlado em termos de variação de renda. Então, procura-se separar o efeito desta variável das possíveis correlações com as demais variáveis consideradas.

Há duas razões para que transformações logarítmicas sejam aplicadas aos dados: uma estatística e uma proporcional (COX, 2007). Para o motivo estatístico, assumindo que a distribuição da variável em questão possui viés, ou seja, uma das extremidades elevadas e uma cauda longa, podem ser muito influenciadas pelo pico da distribuição ou pela presença de *outliers*. A aplicação da transformação pode reduzir o efeito do viés. Já no que tange à justificativa proporcional acontece porquê alguns conceitos são melhor

compreendidos quando tratados sobre a proporção dos objetos do que sobre a diferença entre eles. Neste caso a transformação logarítmica, por exemplo, pode ser utilizada para comparar resultados de uma mesma grandeza, de forma mais justa - podendo ajudar a deixar a relação entre os dados mais prática.

Para o trabalho proposto, a variável dependente é o *log* do salário-hora das pessoas empregadas nas atividades selecionadas. A escolaridade será medida em intervalos. O tempo de emprego na empresa será medido em meses. A idade dos trabalhadores será medida em anos. Os modelos incluirão as variáveis *dummies* de ocupação (STEM ou Não-STEM) e de sexo (homem ou mulher). Seus coeficientes estimados indicarão as estimativas do prêmio salarial ocupacional e de desvantagem salarial de gênero, respectivamente.

Nos modelos semilogarítmicos com variáveis *dummies*, utiliza-se a técnica sugerida por Halvorsen, Palmquist et al. (1980), assumindo o antilogaritmo (de base *e*) do coeficiente binário estimado. Logo, subtrai-se 1 e se multiplica o resultado por 100, analisando então o coeficiente também em termos percentuais. Se por exemplo:

$$[e(\text{coef}) - 1] = 0,31 * 100. \quad (3.18)$$

assume-se que a mediana⁵ do salário (variável dependente) dos indivíduos cuja característica associa o valor 1 da variável *dummy* (por exemplo, STEM), seja 31% maior do que a mediana dos indivíduos cuja *dummy* seja igual a 0 (Não-STEM). Para coeficientes negativos a interpretação é inversa.

O grau de escolaridade será dividido em quatro intervalos: i) ensino médio completo (categoria de base omitida); ii) superior incompleto; iii) superior completo e; iv) pós-graduação, conforme a tabela 1.

Tabela 1 – Variáveis explicativas do modelo

Nome	Descrição	Categoria
<i>tempo</i>	Tempo de emprego em meses	Contínua
<i>idade</i>	Idade	Contínua
<i>idade2</i>	Idade ao quadrado	Contínua
<i>supinc</i>	Escolaridade superior incompleta	<i>Dummy</i>
<i>sup</i>	Escolaridade superior completa	<i>Dummy</i>
<i>pos</i>	Pós-graduação completa	<i>Dummy</i>
<i>stem</i>	Ocupação direta em ciência ou tecnologia	<i>Dummy</i> , STEM = 1
<i>genero</i>	Gênero	<i>Dummy</i> , Mulher = 1

Fonte - Elaboração própria.

A variável *idade* exerce dois papéis nos modelos. Em um primeiro momento, utiliza-se das observações de anos de vida dos trabalhadores como uma variável contínua, o que permitirá avaliar o quanto um ano a mais de vida exercerá de influência sobre os salários.

⁵ Por se tratar de interpretação e modelagem com variável *dummy* (0,1).

Em um segundo momento, a variável também é utilizada para separar os trabalhadores em 5 intervalos de idade, que constituem os anos onde há maior nível de ocupação profissional. As faixas utilizadas para avaliação das categorias de idade são apresentadas na tabela 2.

Tabela 2 – Variáveis dummies de categoria de idade

Categoria de intervalo de idade	Nome da variável dummy
Até 25 anos	agecat1
Entre 26 e 35 anos	agecat2
Entre 36 e 45 anos	agecat3
Entre 46 e 55 anos	agecat4
Acima de 56 anos	agecat5

Fonte - Elaboração própria.

Testes adicionais de normalidade dos resíduos são feitos para garantir que os valores dos testes de significância individual e coletiva dos modelos possam ser interpretados de maneira consistente e, adicionalmente, testam-se as hipóteses aqui assumidas para garantir estimativas eficientes e consistentes dos parâmetros dos modelos.

Para testar o grau de multicolinearidade das variáveis presentes nos modelos, aplica-se o teste de *Variance Inflation Factor* (VIF), já na estimação robusta. A regra prática para o teste é considerar que o modelo possui níveis aceitáveis de multicolinearidade quando o resultado para cada variável for menor que 10 ou quando a média geral for menor que 6. Os testes apontaram níveis de multicolinearidade aceitáveis nas variáveis de idade quando utilizadas todas as categorias de idade. O nível de aceitabilidade da normalidade dos dados, contudo, foi extrapolado nos modelos que utilizam recortes de faixas de idade, conforme a tabela 29 no Apêndice B. Isto ocorre pois os valores ao quadrado das idades se tornam drasticamente mais semelhantes ao seu formato em nível quando os intervalos de idade são menores.

3.4 Modelos estimados

Para a primeira etapa deste trabalho foram desenvolvidos dois modelos. Este primeiro modelo é utilizado de duas maneiras. O banco de dados é dividido, separando os grupos de trabalhadores STEM e não-STEM e será avaliado o grau de discriminação salarial por gênero em cada um dos dois grupos de ocupação. O objetivo dessa análise é avaliar se as mulheres que atuam nas carreiras de ciência e tecnologia desfrutam de menor desvantagem salarial do que as que atuam fora desse conjunto de ocupações. O modelo pode ser descrito assumindo a seguinte forma funcional:

$$\begin{aligned} \ln w = & \beta_0 + \beta_1 \cdot idade + \beta_2 \cdot idade^2 + \beta_3 \cdot tempo + \beta_4 \cdot supinc \\ & + \beta_5 \cdot sup + \beta_6 \cdot pos + \varepsilon \mid gênero \end{aligned} \quad (3.19)$$

onde:

- gênero é uma variável binária, sendo 0 para homens e 1 para mulheres. Neste modelo, atua como componente de decomposição dos salários.

O segundo modelo é ajustado ao banco de dados inteiro para estimar o prêmio salarial do grupo de trabalhadores que atuam em ocupações STEM. Este modelo busca avaliar o prêmio salarial para as ocupações relacionadas à tecnologia. O modelo possui a seguinte forma funcional:

$$\begin{aligned} \ln w = & \beta_0 + \beta_1 \cdot idade + \beta_2 \cdot idade^2 + \beta_3 \cdot tempo + \beta_4 \cdot supinc \\ & + \beta_5 \cdot sup + \beta_6 \cdot pos + \varepsilon \mid stem \end{aligned} \quad (3.20)$$

onde:

- $\ln w$ onde a variável dependente sempre será vetor com os logaritmos naturais (\log) do salário-hora, em reais, dos trabalhadores, representado por w ;
- β_0 é o coeficiente de intercepto, que representa a heterogeneidade atribuída a variáveis não observadas no modelo e comuns a todos indivíduos;
- $idade$ representa os vetores com as idades dos trabalhadores, medida em anos;
- $idade^2$ representa os vetores com as idades ao quadrado. Utiliza-se o quadrado da idade por assumir que os retornos marginais no salário são decrescentes para anos de vida;
- $tempo$ é uma variável contínua, expressa como vetor do tempo de emprego atual dos indivíduos, mensurado em meses;
- $supinc$ é uma variável *dummy* de escolaridade que assume o valor 1 quando o indivíduo possui ensino superior incompleto;
- sup é uma variável *dummy* de escolaridade que assume o valor 1 quando o indivíduo possui ensino superior completo;
- pos indica todos os indivíduos que possuem ao menos algum curso de pós-graduação;
- $stem$ é uma variável dicotômica que é 0 quando o indivíduo atua numa ocupação que não pertence ao STEM e assume o valor 1 quando a ocupação é do grupo STEM. Neste modelo, este é o componente de decomposição dos salários.

- Assume-se que os termos do vetor ε são independentes, identicamente distribuídos e não correlacionados com as variáveis explicativas. Esta variável procura incluir todas as influências no comportamento da variável de rendimento que não podem ser explicadas pelos comportamentos das variáveis de controle apresentadas.

Na terceira parte dessa primeira etapa, o modelo da equação (3.20) é ajustado para estimar o prêmio STEM dos homens e das mulheres, separadamente. Essas estimativas do prêmio STEM feminina e masculina terão como variáveis de controle a idade, o grau de escolaridade e o tempo no emprego de cada um dos grupos e fornecerão na próxima seção o *insight* de prêmio salarial feminino para as carreiras de aplicação direta de tecnologia - como apresentado por [Beede et al. \(2011\)](#).

Já a segunda etapa deste estudo consiste no ajuste dos dados a seis modelos, agora com distinções de intervalos de idade. Logo, no presente trabalho foi escolhido para categoria de base o primeiro intervalo de idade, composto por trabalhadores com até 25 anos - categoria *agecat1*. Como o foco dessa análise é o prêmio salarial das ocupações STEM, não serão abordadas as distinções de gênero. O objetivo de ajustar os dados a estes modelos é investigar em que medida o prêmio salarial STEM varia ao longo da estrutura etária da população estudada. A formulação geral desta primeira equação, pode ser representada por:

$$\begin{aligned} \ln w = & \beta_0 + \beta_1 \cdot \text{tempo} + \beta_2 \cdot \text{supinc} + \beta_3 \cdot \text{sup} + \beta_4 \cdot \text{pos} + \beta_5 \cdot \text{agecat2} \\ & + \beta_6 \cdot \text{agecat3} + \beta_7 \cdot \text{agecat4} + \beta_8 \cdot \text{agecat5} \mid \text{stem} \end{aligned} \quad (3.21)$$

onde foram acrescentadas ao modelo variáveis *dummies* de idade, representadas pelas categorias apresentadas na tabela 2.

Este modelo considera todos os intervalos de idade entre os grupos STEM e Não-STEM. Dadas estas categorias, além da análise mediante regressão seguindo o formato geral indicado em (3.21), foram estabelecidos outros cinco modelos, agora diferenciando cada um deles pelo corte amostral por categoria de idade dos trabalhadores e considerando a variável binária *stem* (0 para o grupo Não-STEM e 1 para STEM), de modo que apresentem a mesma forma assumida pela equação (3.20), mas ajustado para coortes de idade.

No modelo geral a idade é tratada como variável contínua. Esta forma funcional é repetida para as categorias: *agecat1*, *agecat2*, *agecat3*, *agecat4* e *agecat5*, onde cada modelo assumirá o nome do grupo correspondente.

Os resultados dos modelos são apresentados no capítulo seguinte, que se inicia com a apresentação descritiva dos dados e, na sequência, são apresentados os resultados dos modelos de decomposição salarial, conforme a metodologia supracitada.

4 ANÁLISE DOS RESULTADOS

A abordagem principal desta seção será a análise dos diferenciais de rendimento no mercado de trabalho formal catarinense de acordo com suas características e ocupações. Para tanto, primeiro são apresentadas as características descritivas das variáveis em questão, destacando as remunerações dos grupos STEM e Não-STEM e dentre as categorias de idade que serão avaliadas. Entende-se que tal avaliação é essencial para a compreensão dos diferenciais de rendimento das áreas de tecnologia, considerando as diferenças entre gêneros e categorias de idade.

Após a apresentação descritiva, a análise dos resultados será abordada a partir de duas etapas. A primeira etapa consiste em analisar os resultados das estimativas obtidas a partir das equações (3.20) e (3.19), avaliando cinco aspectos dos diferenciais salariais em Santa Catarina, sendo eles: a diferença de salários entre os trabalhadores STEM e Não-STEM, separando os determinantes internos da remuneração dos coeficientes ligados ao prêmio STEM e à possível discriminação. São analisadas também as diferenças na remuneração de homens e mulheres para o grupo STEM e para o grupo Não-STEM, verificando os prêmios salariais das mulheres e dos homens.

A segunda etapa avalia possíveis diferenciações nas remunerações de acordo com as faixas etárias dos trabalhadores. Cita-se que entre os modelos, avaliam-se os grupos de até 25 anos (Agecat1), entre 26 e 35 anos (Agecat2), de 36 a 45 anos (Agecat3), dos 46 aos 55 anos (Agecat4) e após os 56 anos (Agecat5). Sendo que todas as variáveis apresentaram significância estatística nos modelos avaliados, a nível de 95% de confiança.

4.1 Descrição dos dados

Com base nos microdados da Relação Anual de Informações Sociais (RAIS) para Santa Catarina e ano de 2018, o total de trabalhadores formalmente registrados no estado são 2.103.738 empregados. O grupo que identifica os trabalhadores que não são classificados como STEM é formado por 2.076.031 profissionais (98,2%) e o grupo dos trabalhadores pertencentes ao grupo STEM somam 27.707 (1,8%) observações.

A tabela 3 apresenta a frequência de trabalhadores, média de idade e o salário-hora médio entre os grupos STEM e Não-STEM, destacando a diferença entre os gêneros. O número de trabalhadores homens são superiores aos das mulheres em ambas as categorias, sendo 20% superior entre as ocupações não relacionadas à ciência e tecnologia e tendo mais que o triplo de profissionais entre as ocupações STEM.

Além do maior volume, os trabalhadores de gênero masculino também apresentam, em média, mais anos de idade e maior salário-hora. Ambas as variáveis apresentam o

mesmo comportamento de ampliação da diferença no grupo STEM.

Tabela 3 – Estatísticas descritivas dos trabalhadores formais em Santa Catarina - 2018

Grupo	Gênero	Trabalhadores	Média de idade	Salário-hora médio
STEM	Homens	21.909	35,89	R\$ 47,08
	Mulheres	5.798	33,81	R\$ 38,26
	Total	27.707	35,45	R\$ 45,23
Não-STEM	Homens	1.134.069	35,96	R\$ 18,60
	Mulheres	941.962	35,43	R\$ 16,77
	Total	2.076.031	35,72	R\$ 17,77
Total		2.103.738	35,72	R\$ 18,13

Fonte - Elaboração própria com base na RAIS 2018.

Separando os grupos STEM e Não-STEM por categorias de idade, conforme a tabela 4, percebe-se que o maior volume de trabalhadores se encontra na faixa de e 26 a 35 anos de idade, para ambos os grupos. Destaca-se o fato de que cerca de 75% dos trabalhadores STEM se concentram nas faixas dos 26 a 45 anos, restando 12% na faixa mais jovem e 14% na categoria acima de 46 anos. Estes valores são sensivelmente maiores que a composição dos demais trabalhadores, onde o percentual de jovens e de pessoas com mais de 46 anos corresponde a 22%, deixando a faixa de 26 a 45 anos com 56% do total.

Os valores das faixas de idade também mostram que o valor médio de salário-hora cresce na medida em que aumentam as faixas de idade, alcançando valores próximos a R\$ 100,00 por hora trabalhada no conjunto de trabalhadores STEM com mais de 56 anos de idade, enquanto a mesma categoria de ocupações não relacionadas à tecnologia são remunerados em uma quantia 73% inferior.

Tabela 4 – Estatísticas descritivas dos trabalhadores formais em Santa Catarina por faixa etária - 2018

Grupo	Faixas de idade	Trabalhadores	Média de idade	Salário-hora médio
STEM	Até 25 anos	3.272	22,94	R\$ 18,67
	26 a 35 anos	12.380	30,55	R\$ 37,04
	36 a 45 anos	8.181	39,53	R\$ 53,83
	46 a 55 anos	2.677	49,77	R\$ 67,26
	Acima de 56 anos	1.197	60,51	R\$ 94,57
	Total	27.707	35,45	R\$ 45,23
Não-STEM	Até 25 anos	462.207	21,62	R\$ 10,94
	26 a 35 anos	647.297	30,47	R\$ 16,36
	36 a 45 anos	517.169	40,06	R\$ 20,89
	46 a 55 anos	325.825	50,03	R\$ 23,00
	Acima de 56 anos	123.533	60,10	R\$ 25,47
	Total	2.076.031	35,72	R\$ 17,77
Total		2.103.738	35,72	R\$ 18,13

Fonte - Elaboração própria com base na RAIS 2018.

Conforme mencionado na seção 3.3, os valores de salário-hora foram convertidos para sua forma logarítmica, seguindo as referências bibliográficas de decomposição salarial. Assim, apresentam-se as medidas estatísticas da variável de interesse deste trabalho na tabela 5, se destaca a superioridade das remunerações STEM em relação aos demais trabalhadores.

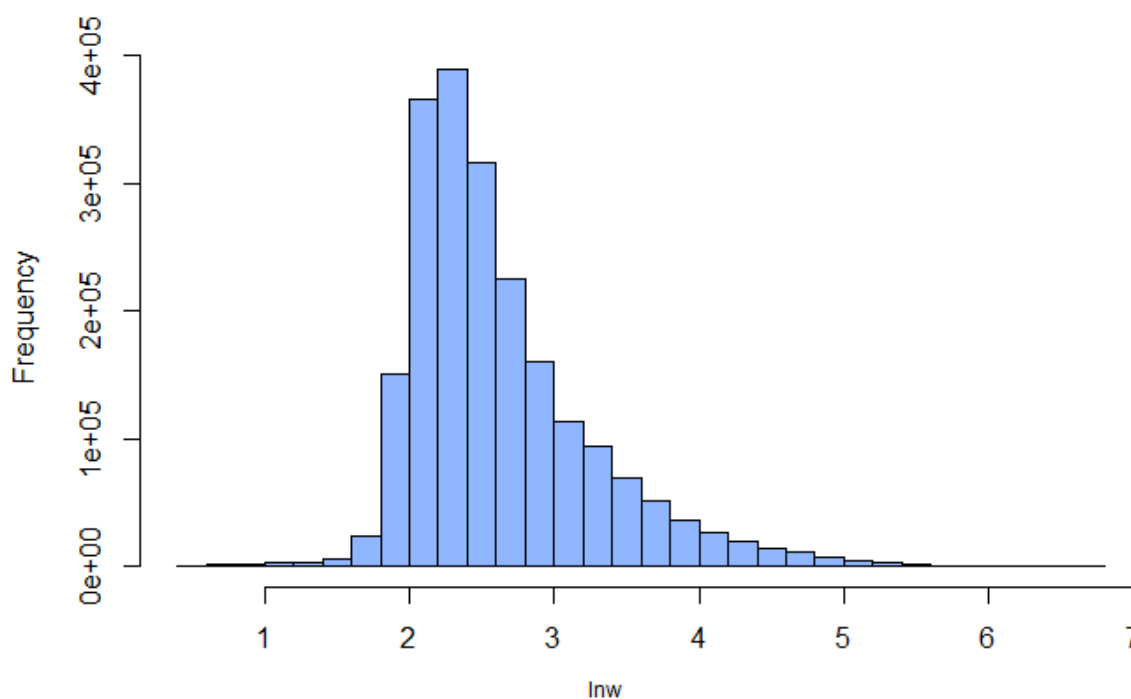
Tabela 5 – Estatísticas descritivas do salário-hora em *log*

Grupo	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo
STEM	0,6127	3,0876	3,6308	3,5766	4,0509	6,6130
Não-STEM	0,4862	2,1805	2,4520	2,6108	2,8825	6,6045
Total	0,4862	2,1839	2,4590	2,6235	2,9028	6,6130

Fonte - Elaboração própria com base na RAIS 2018.

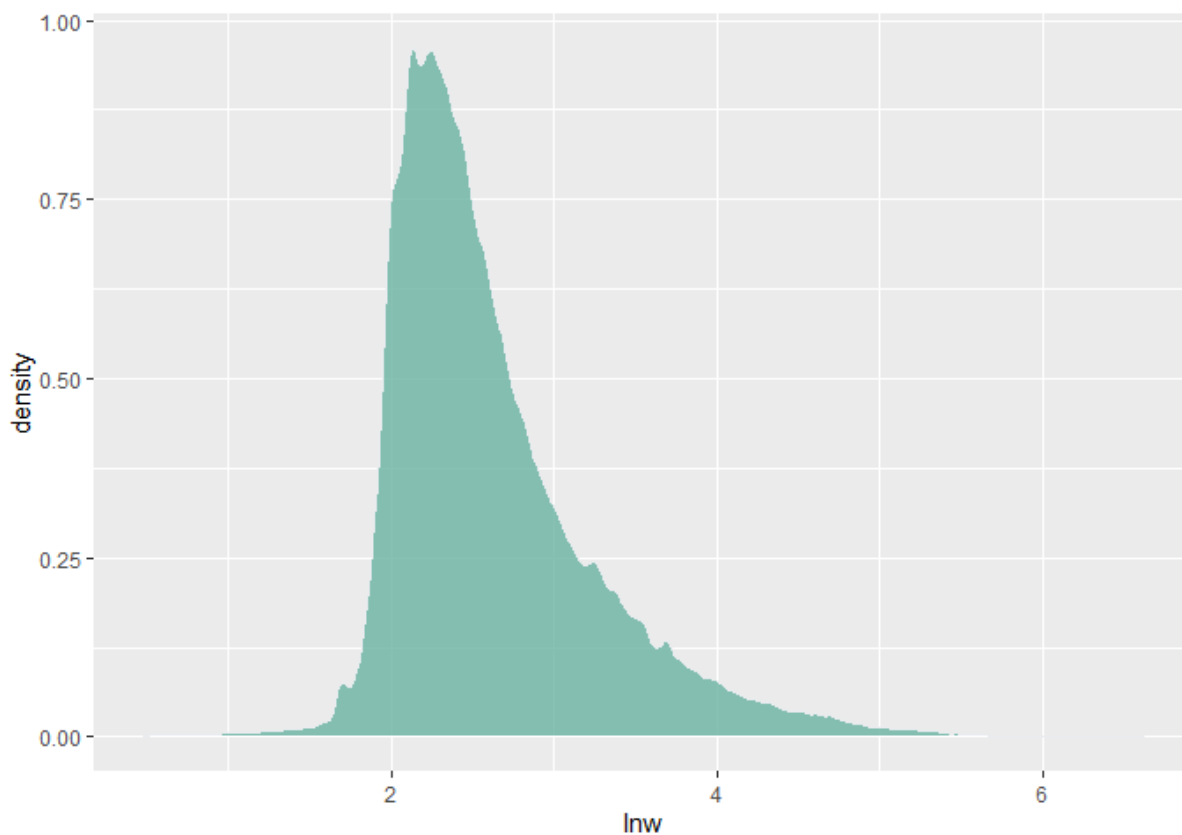
A distribuição dos valores de salário-hora em *log* também podem ser observados a partir do histograma e da função densidade da variável, conforme as figuras 1 e 2. As figuras mostram elevação da frequência de remuneração em faixas de salário superiores ao salário mínimo, que em *log* equivale a 1,8 reais a hora. A existência de observações com valores inferiores ao mínimo se deve principalmente aos registros formais de estágios e jovens-aprendizes.

Figura 1 – Histograma do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

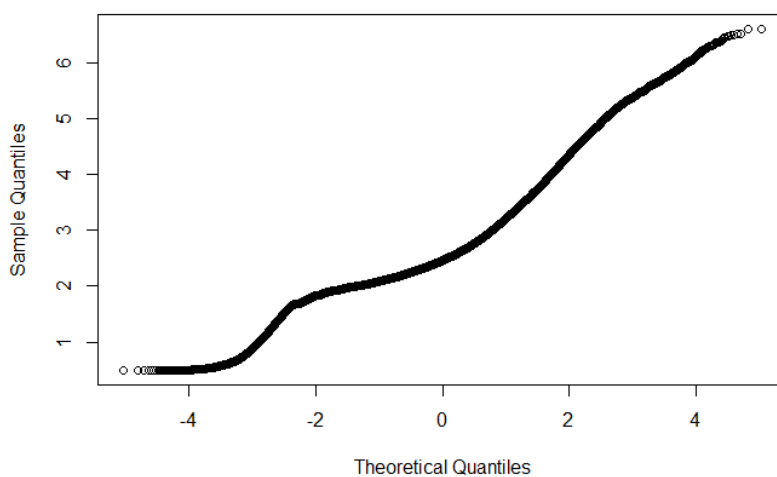
Figura 2 – Função densidade do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Esta elevação da frequência de trabalhadores a partir do salário mínimo é ainda mais visível na figura 3, que representa graficamente os quantis do salário-hora em *log*.

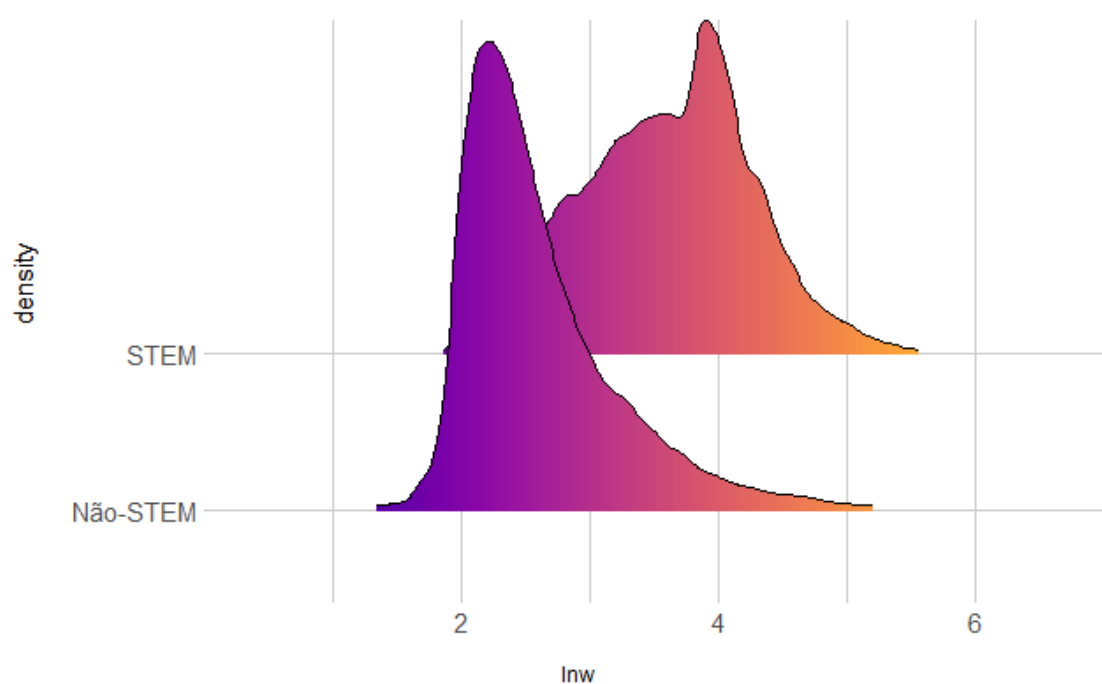
Figura 3 – qqplot do logaritmo do salário-hora dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Apesar de ser um fator relevante para a remuneração dos trabalhadores catarinenses, de modo geral, a elevação da frequência em função do salário mínimo não apresenta efeito prático para a função densidade das ocupações ligadas à ciência e tecnologia, como pode ser observado na figura 4. Neste grupo, remunerações mais elevadas se tornam mais comuns, existindo, inclusive, um aumento a partir da remuneração de 4 reais a hora, em *log*.

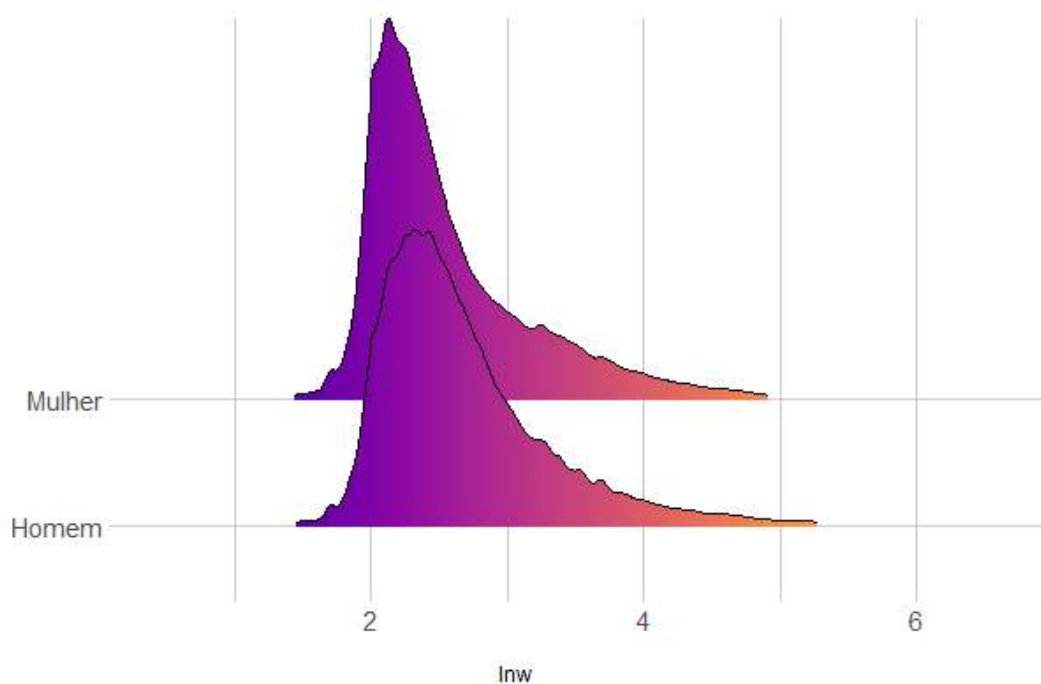
Figura 4 – Função densidade do logaritmo do salário-hora dos trabalhadores STEM e Não-STEM de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Para explicar as diferenças nas remunerações STEM e Não-STEM, este trabalho avalia os determinantes dos componentes de gênero, idade, experiência e escolaridade. Além das justificativas teóricas, a avaliação descritiva destas variáveis tornam ainda mais relevante a sua inclusão no modelo. No quesito de gênero, as funções densidade da remuneração de cada um dos grupos apresenta uma concentração elevada em faixas menores de salário para o gênero feminino, conforme a figura 5. Além disso, a densidade do salário para os homens se estende mais em faixas mais elevadas de remuneração.

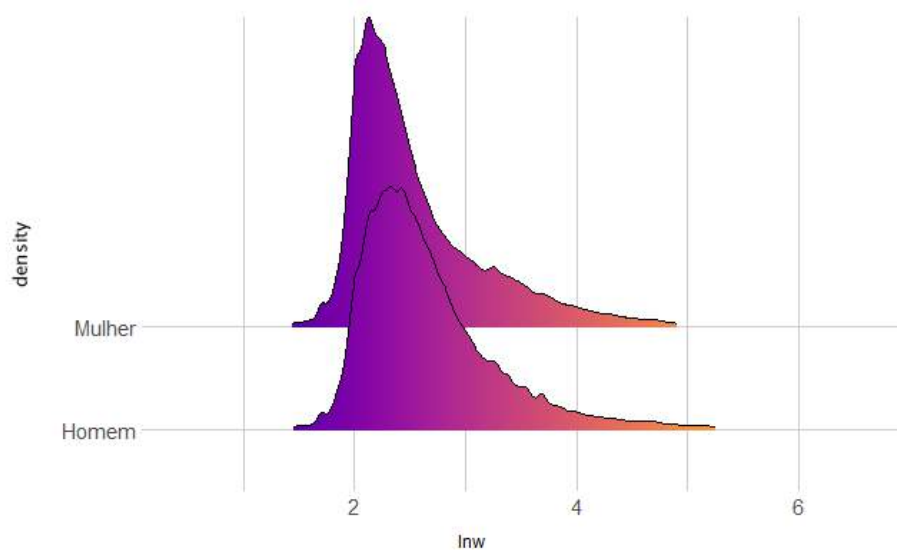
Figura 5 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

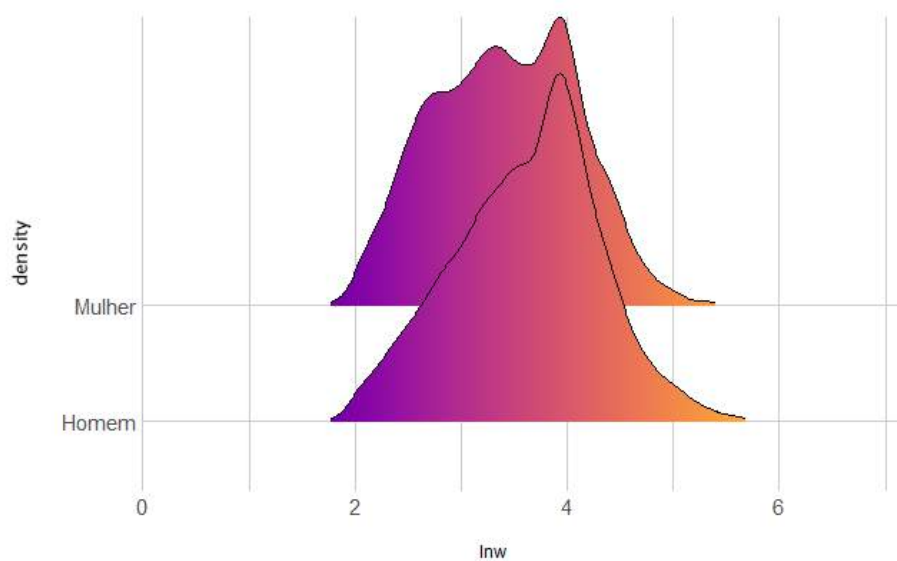
Ainda pela ótica de gênero, ao observar a figura 6 é notável que para o grupo Não-STEM a distribuição dos salários ao longo da função densidade é similar à figura 5, uma vez que este grupo é predominante entre os trabalhadores, correspondendo cerca de 98,2% da força laboral do estado. Entretanto, para os profissionais da tecnologia, o grupo feminino possui densidade mais elevada em faixas salariais menores, o que demonstra a maior frequência de trabalhadoras com salário inferior ao dos homens, que possuem densidade mais extensa à direita, refletindo sua predominância entre os mais bem remunerados. Além disso, verifica-se o mesmo pico salarial em 4 reais a hora (*log*) para ambos os sexos, encontrado na figura 4.

Figura 6 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores Não-STEM de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Figura 7 – Função densidade do logaritmo do salário-hora por gênero dos trabalhadores STEM de Santa Catarina para o ano de 2018

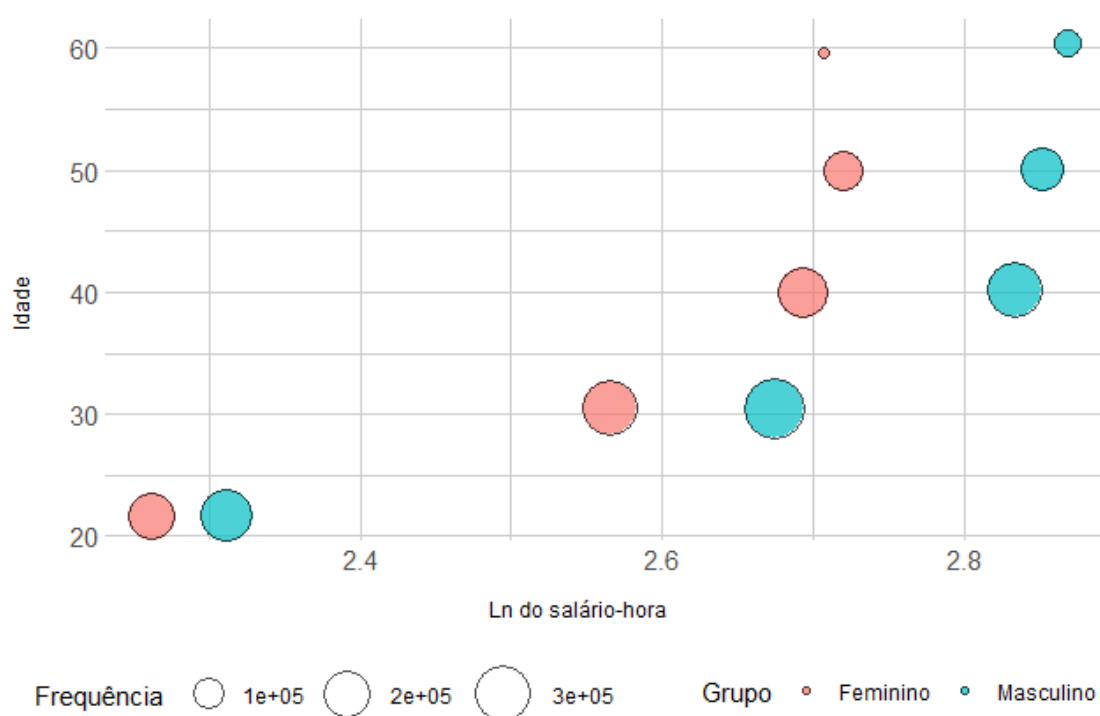


Fonte - Elaboração própria com base na RAIS 2018.

Outra observação relevante na avaliação da diferença salarial por gênero reside da comparação dos valores médios de remuneração para as cinco categorias de idade avaliadas

neste trabalho. A figura 8 apresenta que há diferenças salariais entre gênero para todas as categorias de idade e que estes valores se tornam cada vez mais discrepantes a medida em que são elevadas as categorias de idade. Entre os mais jovens, a remuneração média dos trabalhadores homens é 5% superior ao das mulheres, já para os profissionais com mais de 56 anos, esta diferença supera a faixa de 17%. Resultado similar é apresentado por [Blau e Kahn \(2017\)](#), indicando que nos Estados Unidos, evidências empíricas consideráveis mostram impacto dos filhos sobre os salários das mulheres, uma vez que foram obtidos resultados de um efeito negativo dos filhos sobre os salários, mesmo em análises que controlam a experiência no mercado de trabalho. Esta relação negativa entre a presença de filhos e salários das mulheres (*gap*) é conhecido como *motherhood wage penalty*, ou pena salarial por maternidade. O resultado é plausível no sentido de que na maioria das famílias americanas, as mulheres ainda são as principais responsáveis pelo trabalho doméstico e pelo cuidado dos filhos, refletindo empregos de meia jornada e menores remunerações ([BLAU; KAHN, 2007](#)).

Figura 8 – Relação de idade e salário-hora por gênero dos trabalhadores de Santa Catarina para o ano de 2018

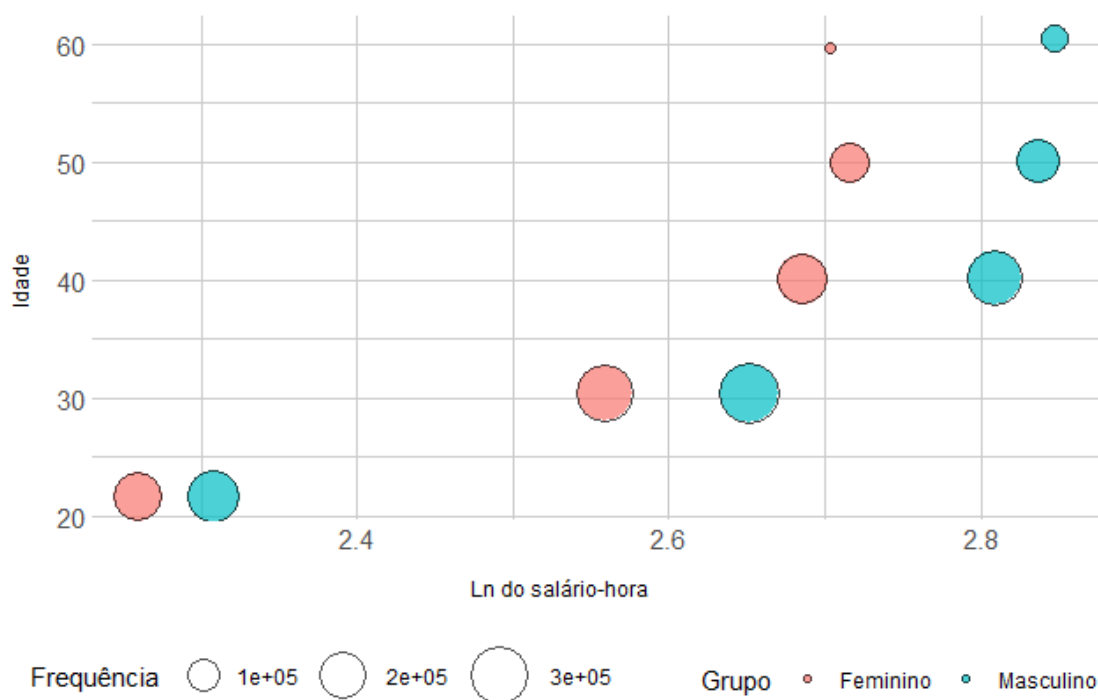


Fonte - Elaboração própria com base na RAIS 2018.

Ainda para a visualização das categorias de idade em função do salário, é possível observar esta relação para o grupo STEM e Não-STEM. Este último (figura 9) é ainda

muito similar à figura acima, por conta destes trabalhadores serem maioria na força laboral, mantendo o *gap* crescente entre os gêneros conforme a idade dos profissionais aumenta.

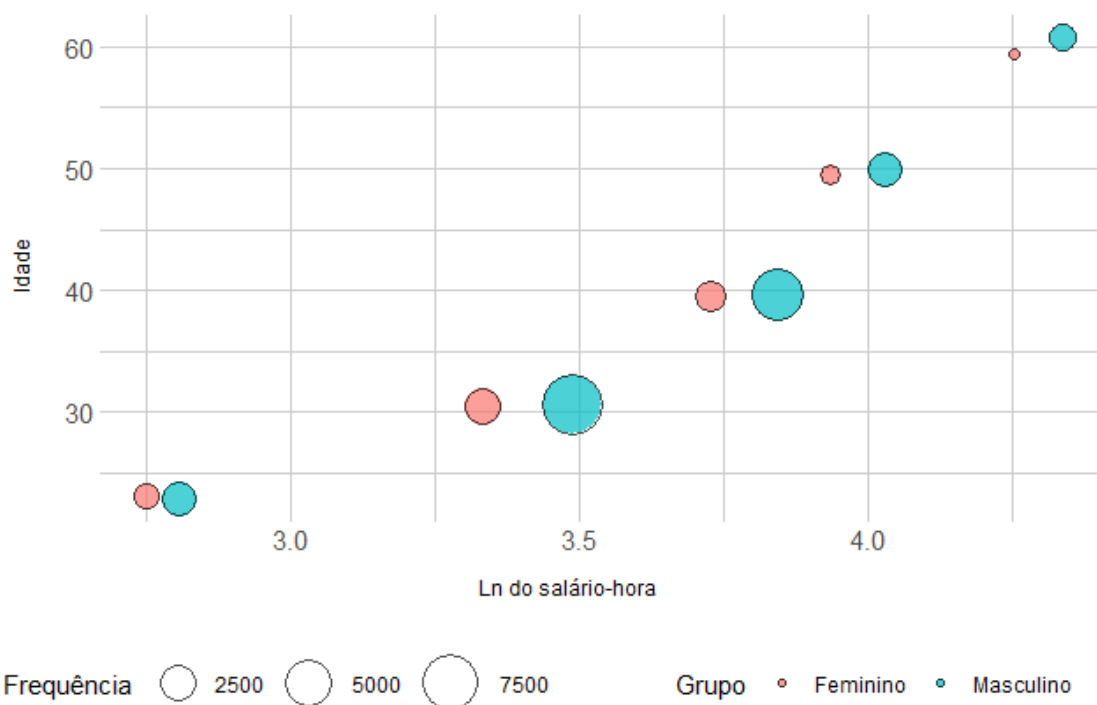
Figura 9 – Relação de idade e salário-hora por gênero dos trabalhadores Não-STEM de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Já para o grupo de ocupações de ciência e tecnologia, a figura 10 mostra que, diferentemente das demais ocupações, os profissionais STEM são capazes de reduzir esta diferença salarial entre gênero, convergindo para o fato de que a disparidade salarial de gênero entre homens e mulheres é menor nestas profissões do que em outras carreiras.

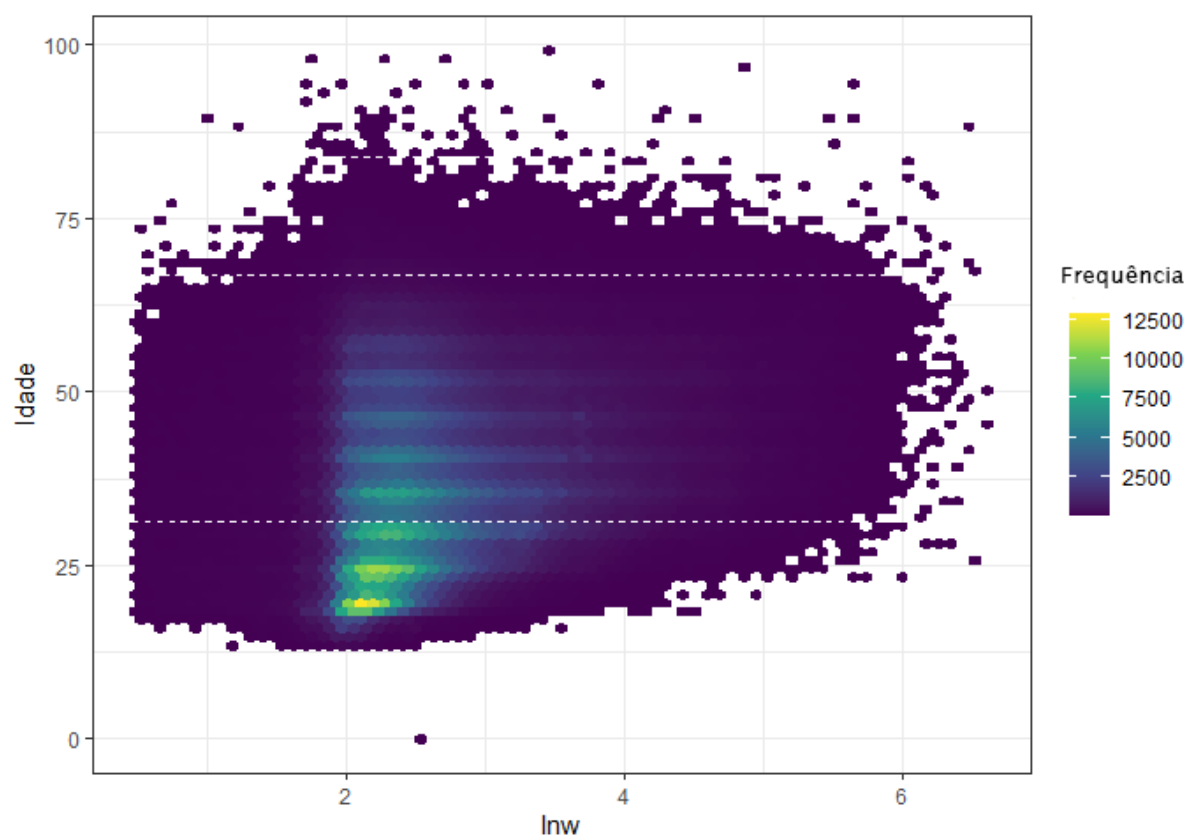
Figura 10 – Relação de idade e salário-hora por gênero dos trabalhadores STEM de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

A relação do fator de idade com a remuneração também pode ser observada a partir da figura 11, que apresenta a densidade da remuneração (em *log*) e da idade dos trabalhadores de Santa Catarina. Nesta figura, além da concentração dos trabalhadores nas remunerações a partir do salário mínimo, conforme apresentado nas funções densidade e histograma, percebe-se uma inclinação positiva na dispersão das observações a medida que aumentam as idades. Dessa forma, a concentração dos trabalhadores mais jovens em remunerações menores é notavelmente maior que a concentração em faixas etárias mais elevadas.

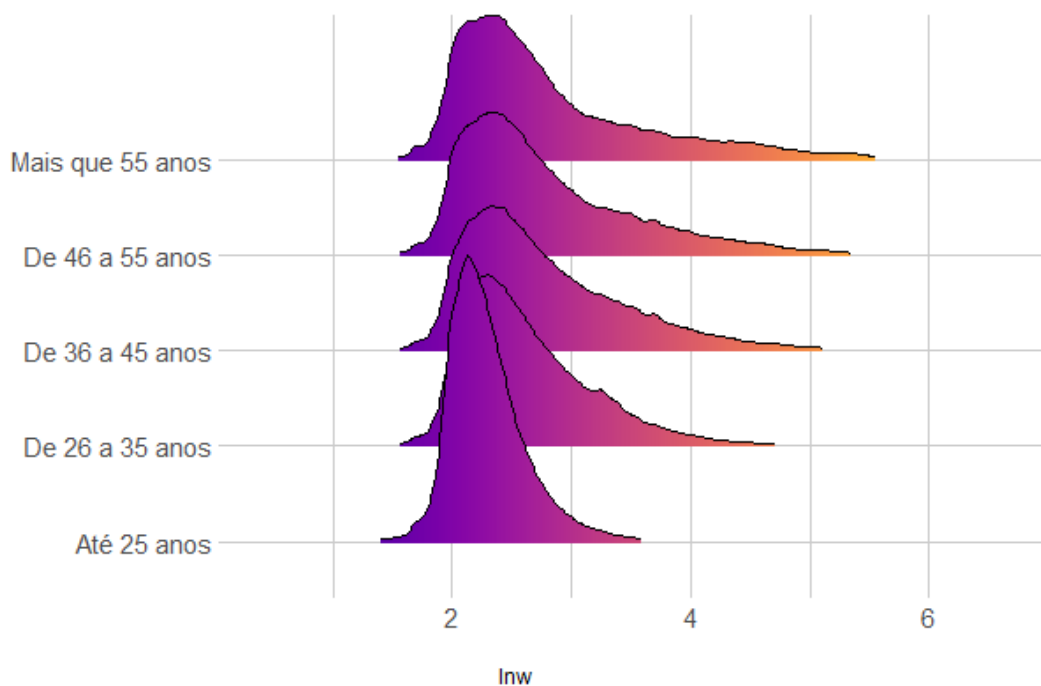
Figura 11 – Função densidade por idade e salário-hora dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Esta relação das faixas de idade e concentração das remunerações são mais visíveis na figura 12, em que as funções densidade do salário-hora estão separadas por cada uma das categorias etárias. Nota-se, portanto, que as faixas mais jovens estão mais concentradas em valores próximos ao salário mínimo, apresentando cauda mais curta para salários maiores. A figura 12 mostra ainda que esta cauda ganha mais tamanho à medida que são considerados os trabalhadores de mais idade.

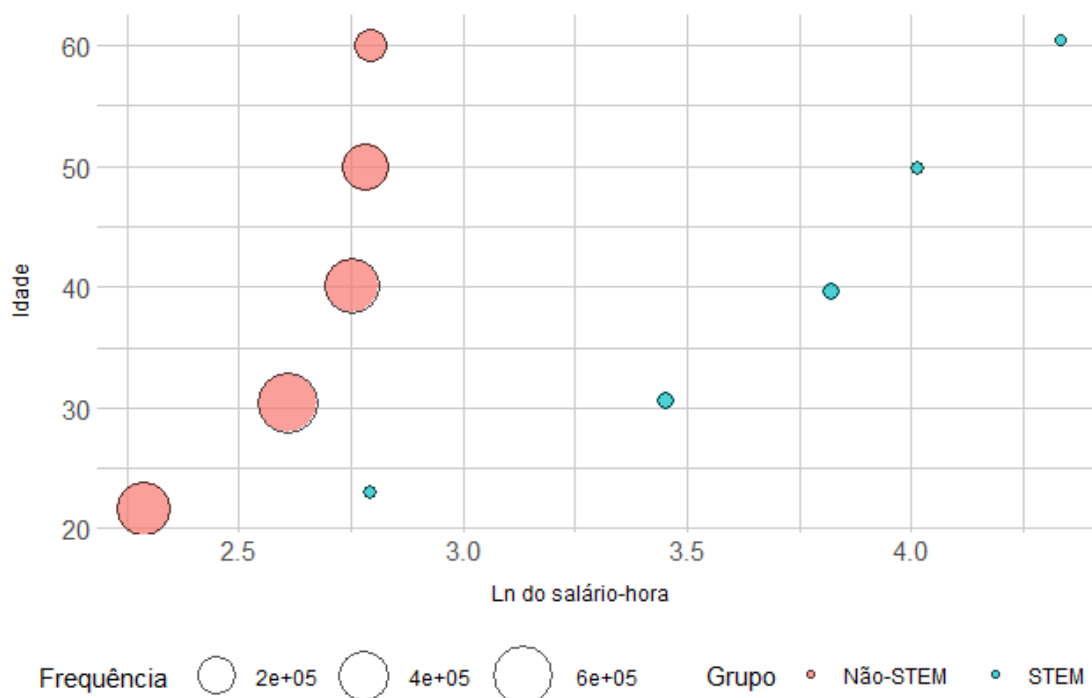
Figura 12 – Função densidade do salário-hora por categorias de idade dos trabalhadores de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

As categorias de idade ainda demonstram uma outra característica relevante quando relacionadas aos grupos STEM e Não-STEM, que pode ser observada na figura 13. Além de confirmar a maior remuneração, em média, para as faixas de maior idade, a figura demonstra comportamentos diferentes entre os grupos. De um lado, as ocupações não ligadas à ciência e tecnologia apresentam ganhos marginais no salário para faixas etárias maiores, enquanto as diferenças de remuneração entre as categorias de idade nos trabalhadores STEM se mantêm significativas mesmo entre trabalhadores de mais idade. Com isso, é possível que o prêmio STEM também siga esta tendência de elevação entre profissionais de mais idade.

Figura 13 – Relação de idade e salário-hora dos trabalhadores STEM e Não-STEM de Santa Catarina para o ano de 2018



Fonte - Elaboração própria com base na RAIS 2018.

Para avaliar os determinantes dos salários e a existência de prêmio STEM entre gêneros e faixas de idade, as próximas seções apresentam os resultados das decomposições de Oaxaca-Blinder, de acordo com a metodologia proposta.

4.2 Análise estatística do prêmio STEM em Santa Catarina

Nesta seção serão apresentados os resultados do ajuste dos dados ao modelo das equações (3.19) e (3.20). Sendo escolaridade, idade e tempo no emprego determinantes de produtividade de acordo com a teoria do capital humano - estabelecida por [Becker \(1957\)](#), estas serão, portanto, variáveis de controle para estimativa do prêmio STEM.

A estimativa do modelo de avaliação entre ocupações de ciência e tecnologia gera uma diferença negativa de 61,9% entre o (ln) salário do grupo Não STEM e grupo STEM. Ou seja, os trabalhadores que não estão ligados à ciência ou tecnologia recebem cerca de 62% a menos que os STEM. O inverso disso apresenta que os STEM recebem, em média, cerca de 162% do valor das demais ocupações.

Desta diferença, a decomposição de Oaxaca-Blinder apresenta que o fator de dotações internas é responsável por uma redução salarial de 39,9% do grupo (ou 63,8%

em favor dos STEM), enquanto o fator não explicado, que neste caso é definido como prêmio STEM, é de 45,3% - Tabela 6. Portanto, as dotações internas dos grupos explicam cerca de metade da diferença salarial dos trabalhadores ligados à ciência e tecnologia com os demais, enquanto outros 40% não são explicados (e o restante se deve à interação de fatores internos e externos).

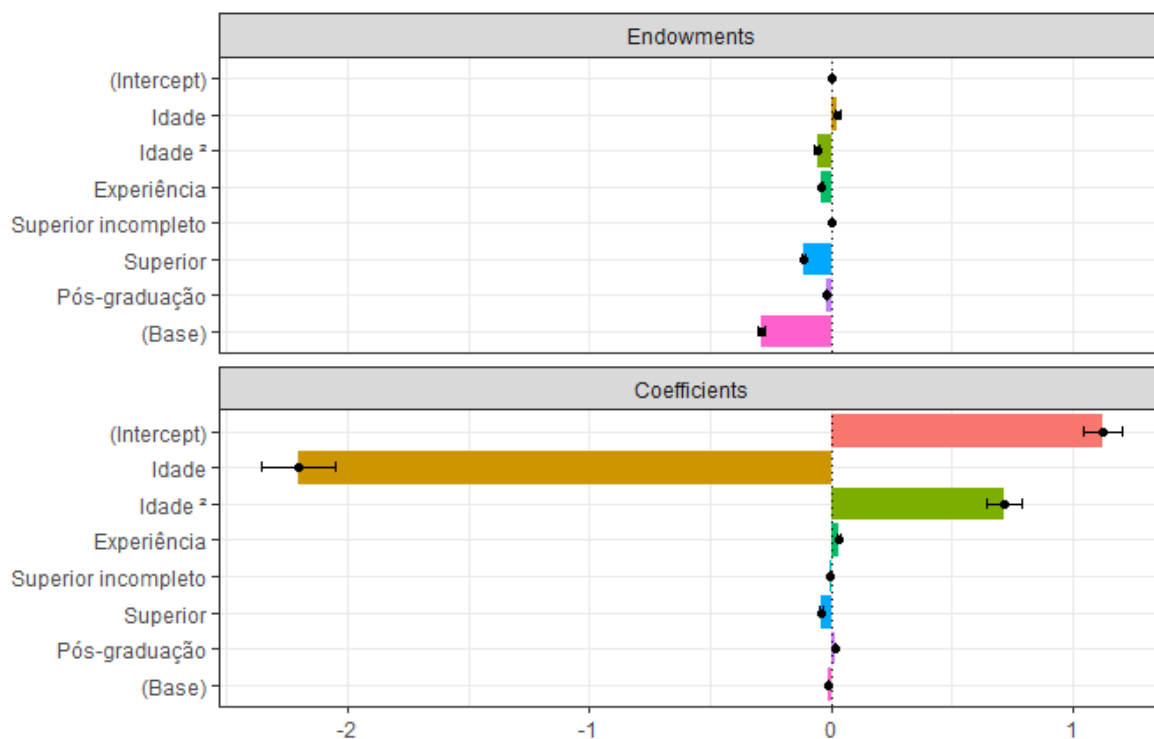
Tabela 6 – Estimativa da vantagem salarial do grupo STEM pela decomposição de Oaxaca

	Diferença Salarial	Explicado	Não explicado
Vantagem STEM	162,7%	63,8%	45,3%
Participação	-	51,1%	38,7%

Fonte - Elaboração própria com base na RAIS 2018.

Entre os principais determinantes para este resultado proveniente de características internas do modelo (*endowments*), a variável de maior relevância para aumento de salário no grupo STEM é a identificação de trabalhadores com ensino superior completo. Em seguida, a segunda variável com maior relevância é a de idade ao quadrado (*idade2*), que capta o retorno associado à idade e apresenta retornos marginais decrescentes, conforme a figura 14.

Figura 14 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM



Fonte - Elaboração própria com base na RAIS 2018.

Assim, significativa parte do aumento de salário em favor das ocupações STEM advém da maior escolaridade destes profissionais. De modo geral, a presença de escolaridade superior aumenta o salário, em média, em 106,4% para ambos os grupos. No caso analisado, o percentual de trabalhadores Não-STEM que possuem ensino superior completo é de 18,55%, enquanto o percentual para trabalhadores do grupo STEM é de 82,62%. Portanto, a maior presença de trabalhadores com ensino superior completo entre as ocupações STEM explica, em parte, o porquê da maior remuneração deste grupo em relação aos demais.

Além da escolaridade, outra informação relevante é o fator negativo da idade na remuneração das ocupações STEM. Esta condição está ligada ao fato de que a experiência é um elemento de ganho salarial, mas, em média, os trabalhadores STEM são mais jovens e, por isso, teriam um peso negativo em relação à idade na sua remuneração.

Neste caso, o impacto de cada ano a mais na idade do trabalhador corresponde a um aumento de 3,54% no salário. Sendo que a média de idade do grupo Não-STEM é de 36 anos, enquanto que a média de idade do grupo STEM é de 35 anos. Assim, esta variável trabalha em movimento contrário, induzindo a uma maior remuneração ao grupo de trabalhadores mais experientes.

Na parte não explicada pelas variáveis internas, o fator de maior diferença nos coeficientes é a idade, isto significa que nas ocupações STEM, as pessoas de mais idade são mais bem remuneradas, em um nível superior à valorização do grupo Não-STEM. Cabe mencionar que isto não ocorre na variável que identifica o tempo de experiência na empresa.

Para o grupo Não-STEM, o modelo sugere que para cada ano a mais na idade do empregado, o salário aumenta em média 3,50%, enquanto que para o grupo STEM este aumento é de 9,72%, em média.

Por outro lado, as variáveis de escolaridade, como a identificação de ensino superior completo e incompleto não são tão críticos para o salário do grupo STEM em comparação ao grupo Não-STEM, o que pode estar relacionado ao fato de que o ensino superior é uma condição mais comum neste primeiro grupo, fazendo que com este diferencial tenha menos impacto na remuneração do trabalhador.

Este comportamento pode ser verificado nas regressões isoladas, sendo que possuir o ensino superior completo no grupo de maior amplitude faz com que o salário aumente em média 106,3% e no grupo STEM este aumento é de 80,8%.

Para avaliar a discrepância do salário em função do gênero, foram aplicadas decomposições de Oaxaca com os grupos masculino e feminino para as ocupações STEM e Não-STEM, comparando descritivamente os resultados obtidos.

Diferentemente do modelo STEM x Não-STEM, a diferença de salário no caso do gênero para trabalhadores que não estão ligados à ciência e tecnologia tem maior parcela sendo explicada por coeficientes externos, e não pelas variáveis internas. A decomposição de Oaxaca-Blinder mostra que o homens recebem, em média, 10,5% mais que as mulheres.

Desta diferença, a parcela correspondente às diferenças nas variáveis internas é negativo, ou seja, se os valores de salário dependessem apenas dos fatores de idade, escolaridade e experiência, as mulheres deveriam receber valores superiores aos homens.

Tabela 7 – Desvantagem salarial feminina: trabalhadores Não-STEM

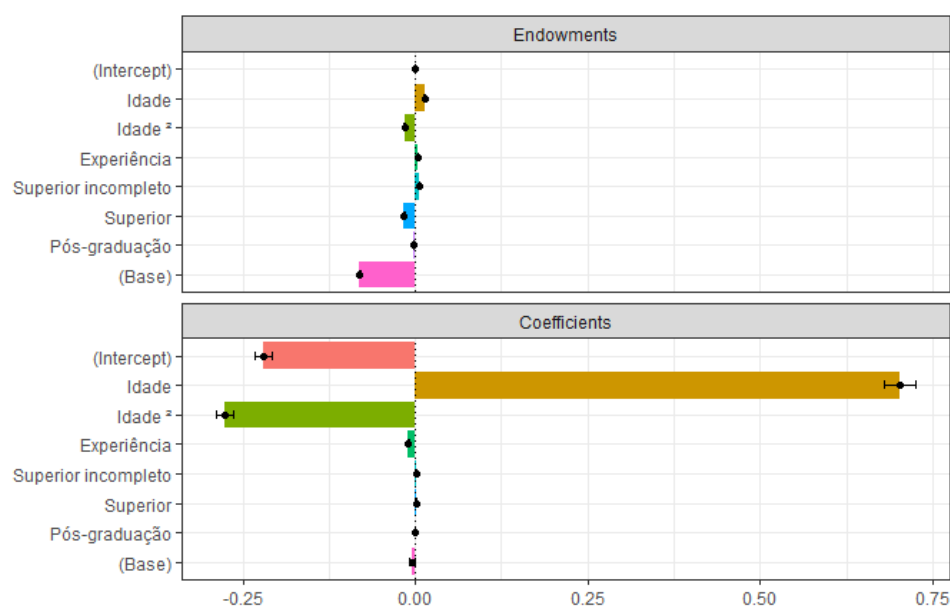
	Diferença Salarial	Explicado	Não Explicado
Variação no salário	10,5%	-8,7%	21,2%
Participação	-	-91,4%	193,4%

Fonte - Elaboração própria com base na RAIS 2018.

Esta primeira conclusão é resultado principalmente da condicionante de ensino superior completo, que de modo geral aumenta a remuneração em 114,4% e é uma condição mais presente no gênero feminino. Entre os trabalhadores do gênero masculino, 13,36% possuem ensino superior completo, já para as mulheres o valor é de 24,69%.

Assim, a diferença em prol dos homens é carregada pela diferença nos condicionantes não explicados. Neste caso, a variável de maior relevância é a idade, o que pode ser observado na figura 15. Foi identificado que cada ano a mais na idade equivale a um aumento salarial de 3,57%, em média. Esta condição, contudo, é significativamente diferente quando verificada as condições de gênero, pois, para o grupo de homens, o aumento percentual de cada ano de idade na remuneração é de 4,46% e o das mulheres é de 2,48%. Este fator, aliado a uma média de idade ligeiramente maior para os homens (35,96 no gênero masculino e 35,43 para o feminino) é o principal componente de discriminação na remuneração feminina.

Figura 15 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores homens e mulheres Não-STEM



Fonte - Elaboração própria com base na RAIS 2018.

No caso dos trabalhadores STEM, continua existindo diferença na remuneração média entre os gêneros. As mulheres recebem, em média, valores inferiores aos homens de cerca de 22,5%.

Tabela 8 – Desvantagem salarial feminina no grupo de trabalhadores STEM

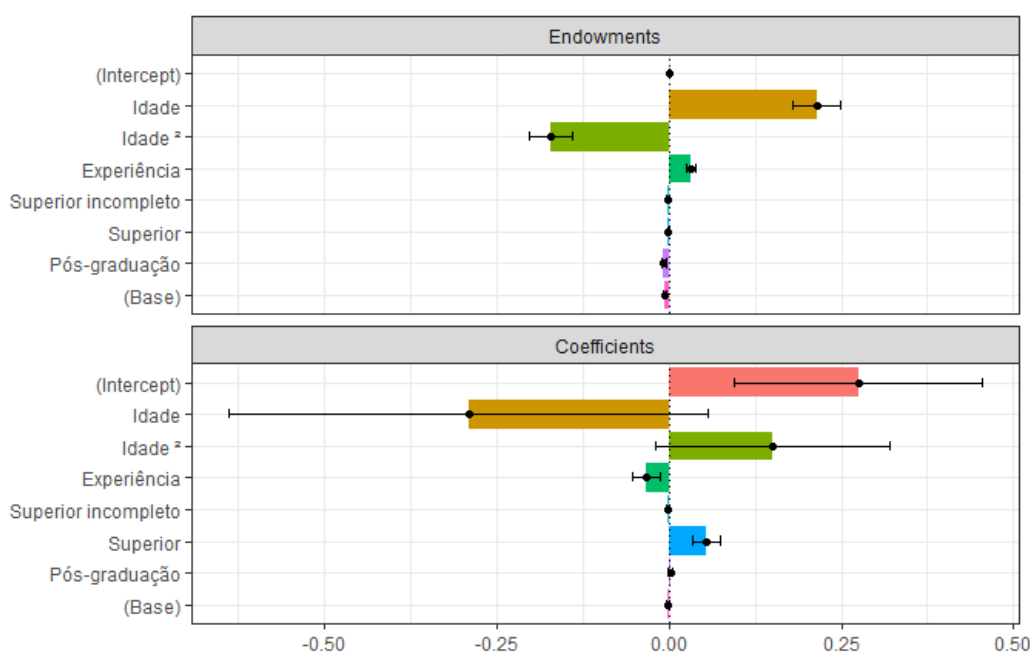
	Diferença Salarial	Explicado	Não explicado
Variação no salário	22,5%	5,8%	16,4%
Participação	-	27,9%	74,8%

Fonte - Elaboração própria com base na RAIS 2018.

Neste caso, a decomposição do modelo mostra que tanto as condições internas como os coeficientes são responsáveis pela menor remuneração da mulher. Do lado das condições internas, foi identificado que as mulheres STEM são notavelmente mais jovens que os homens, onde a diferença de idade neste caso chega a 2,07 anos de idade. Este fator faz com que a remuneração dos homens naturalmente seja maior.

Por outro lado, a diferença correspondente à discriminação continua sendo a principal causa da discrepância de salário, explicando aproximadamente três quartos da diferença salarial dos gêneros. A figura 16 mostra que entre as variáveis mais representativas deste fenômeno está a maior valorização salarial para homens formados com ensino superior completo, que tem um impacto de 62,15% nas suas remunerações, enquanto as para mulheres formadas com ensino superior se observa um aumento salarial de 51,57%.

Figura 16 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores homens e mulheres STEM



Fonte - Elaboração própria com base na RAIS 2018.

Cabe notar que o caso da diferença de gênero nas ocupações STEM não apresentam discriminação de salário em favor de homens mais experientes. Aqui, os valores mostraram sinal contrário, favorecendo, portanto, as mulheres com mais experiência. De qualquer modo, esta valorização superior das mulheres em função da idade não foi suficiente para que os coeficientes continuem sendo favoráveis ao gênero masculino.

Outro fator relevante nas avaliações de gênero entre os grupos STEM e Não-STEM está no fato de que o grau de impacto da discriminação em favor dos homens ser maior no caso do grupo que não está ligado à ciência e tecnologia. Este fator se destaca ainda mais quando considerado que a diferença salarial de gênero é inferior para os trabalhadores em geral, o que acaba invertendo quando se trata do fator discriminação. Assim, o coeficiente externo apresenta um estímulo de 21,2% para os homens Não-STEM, enquanto o grau de discriminação para os trabalhadores STEM é de 16,4%.

Para avaliar o prêmio STEM entre os trabalhadores homens e mulheres, também foi realizada a decomposição dos salários nos grupos ligados ou não à ciência e tecnologia conforme o gênero dos trabalhadores.

Iniciando pelo conjunto de observações que compõe todas as mulheres registradas formalmente no mercado de trabalho, a diferença salarial entre as mulheres que não ocupam cargos STEM e as que ocupam é de -57,7%, valor inferior ao encontrado no modelo contendo todos os trabalhadores.

A diferença na remuneração das mulheres entre os dois grupos é explicada, em sua maior parte, pelas diferenças nas variáveis internas, (44%), mas a parcela referente à discriminação no grupo (onde aqui se trata do prêmio STEM) também é um fator relevante e influencia para baixo a remunerações das trabalhadoras não ligadas à ciência ou tecnologia em -25,9% (ou 35% em favor aos STEM).

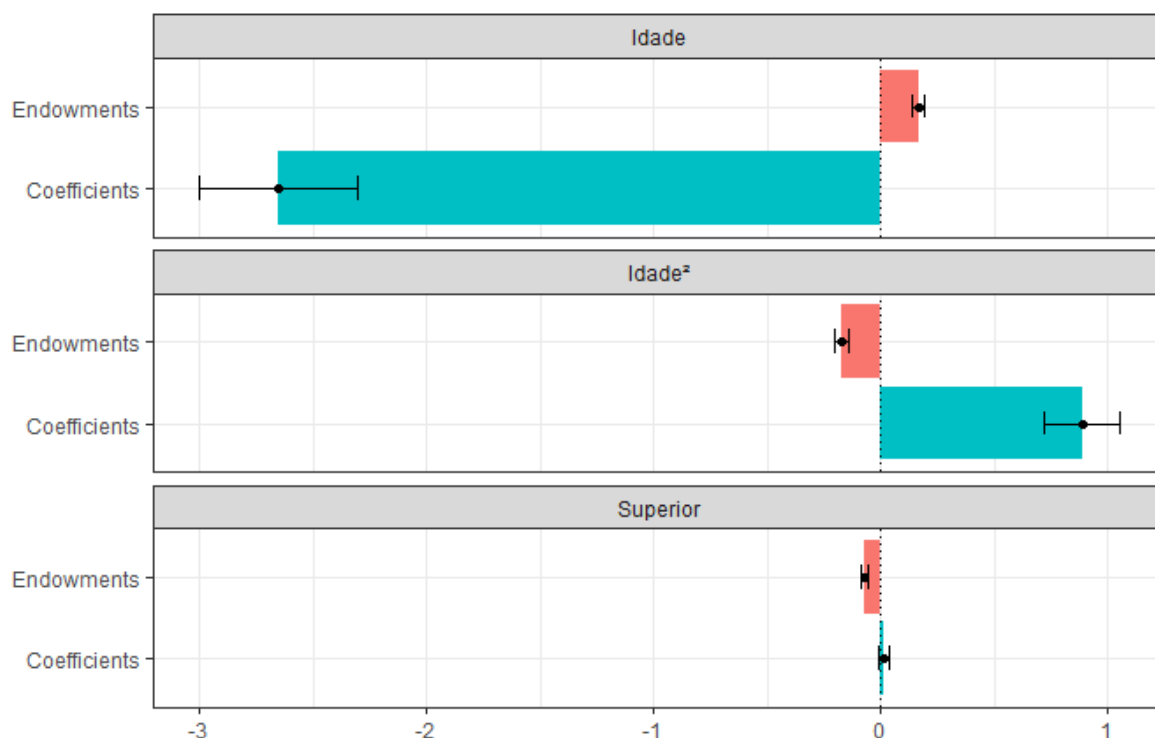
Tabela 9 – Decomposição do salário das trabalhadoras STEM em relação às Não-STEM

	Diferença Salarial	Explicado	Não explicado
Variação no salário	136,3%	46%	35%
Participação	-	44%	34,9%

Fonte - Elaboração própria com base na RAIS 2018.

Em ambos os fatores (*endowments* e *coefficients*), a variável de maior relevância e discrepância das remunerações é a idade, o que pode ser verificado na figura 17. Já foi mencionado neste trabalho que a idade é um fator positivo à remuneração dos trabalhadores, o que se repete neste modelo. Assim, a variável de idade deveria favorecer a remuneração do grupo com maior média de anos de idade, que se trata do grupo Não-STEM, que possui 1,6 ano a mais em relação às trabalhadoras STEM. Resulta desta condição, contudo, o movimento de maior discriminação, sendo que trabalhadoras mais jovens com ocupação STEM são mais bem remuneradas que às de maior experiência no grupo não relacionado à ciência ou tecnologia.

Figura 17 – Componentes de idade e ensino superior da decomposição da diferença salarial entre trabalhadoras STEM e Não-STEM



Fonte - Elaboração própria com base na RAIS 2018.

Já a diferença salarial para os homens que não ocupam cargos STEM e os do grupo STEM é de -61,8%, sendo que os fatores internos são os maiores responsáveis pela diferença. Assim como no caso geral, os trabalhadores STEM que possuem ensino superior completo tem uma representação muito mais significativa que os demais. Neste caso, 82% dos trabalhadores ligados à ciência ou tecnologia possuem formação superior completa, enquanto que dos demais trabalhadores, apenas 13% alcançaram este patamar.

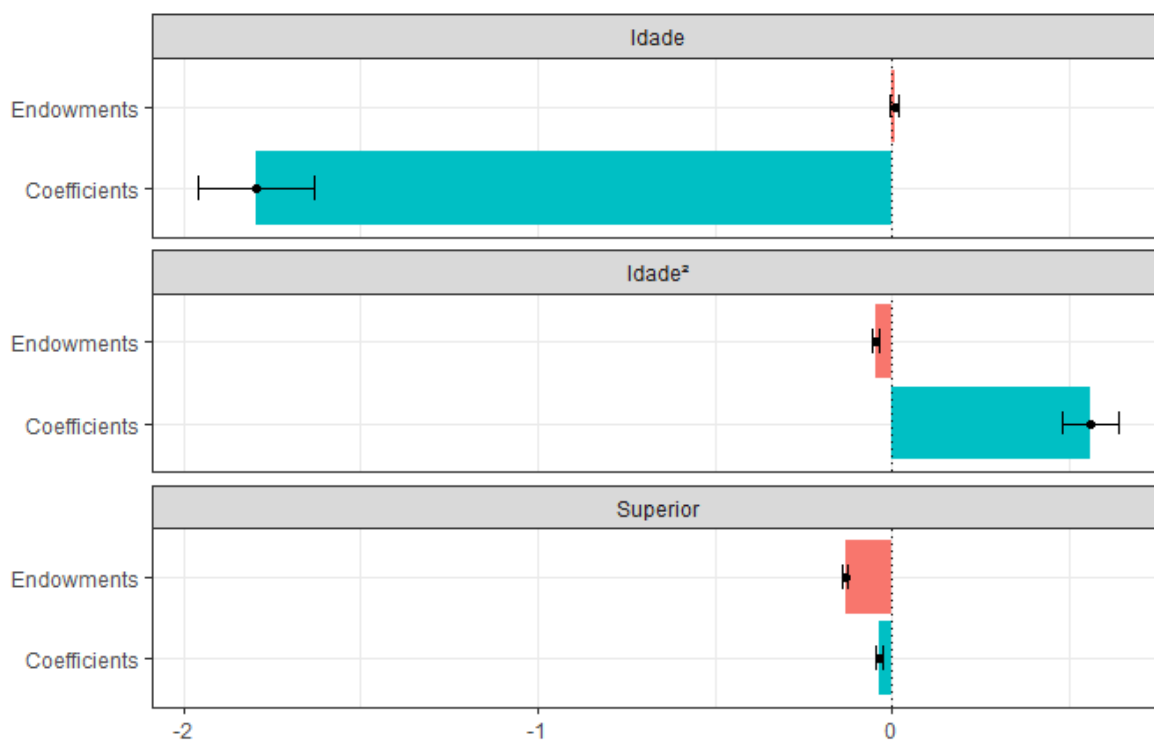
Tabela 10 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM

	Diferença Salarial	Explicado	Não explicado
Variação no salário	162%	65,9%	34,2%
Participação	-	57,4%	30,5%

Fonte - Elaboração própria com base na RAIS 2018.

Por outro lado, a idade novamente exerce o maior fator de discriminação entre os grupos. Em média, os trabalhadores Não-STEM têm mais idade que os STEM, o que os faria ter uma remuneração maior, contudo, o fato de os trabalhadores receberem maiores valores faz com que o fator de discriminação apareça nesta variável, conforme a figura 18.

Figura 18 – Componentes de idade e ensino superior da decomposição da diferença salarial entre trabalhadores homens STEM e Não-STEM



Fonte - Elaboração própria com base na RAIS 2018.

Contando com essas influências, apesar da diferença no salário dos homens nos grupos STEM e Não-STEM ser maior que o das mulheres, o fator relacionado aos fatores externos, ou seja, o prêmio STEM, é menor no caso dos trabalhadores de gênero masculino. Este fator reduz a remuneração do grupo Não-STEM em 25,5%, enquanto que para as mulheres ele é de 25,9%.

Em função da alta relevância do fator idade nos modelos apresentados acima, foram realizadas novas abordagens de diferenciação salarial incluindo variáveis *dummies* para intervalo de idades, onde cada uma destas categorias entram na equação de salário (omitindo-se a *dummy* referente ao intervalo que será a base de comparação, neste caso a variável descrita por *agecat1*).

Assim como o modelo anterior, a inserção das categorias de idade para STEM e Não-STEM abrange todos os trabalhadores formalmente registrados em Santa Catarina no ano de 2018, totalizando 2.103.738 empregados. Os resultados deste modelo são apresentados na seção seguinte.

4.3 Inserção das categorias de idade para STEM e Não-STEM

Por se tratar do mesmo conjunto de trabalhadores divididos nos mesmos grupos (STEM e Não-STEM), a diferença na remuneração média permanece em 61,9%. Por outro lado, a inserção das *dummies* que identificam as faixas etárias aumentam o poder explicativo do modelo, fazendo com que a participação das características internas aumente para 52%. Como resultante disso, o prêmio STEM passa de 45,3%, do modelo geral apresentado acima, para 44,9%.

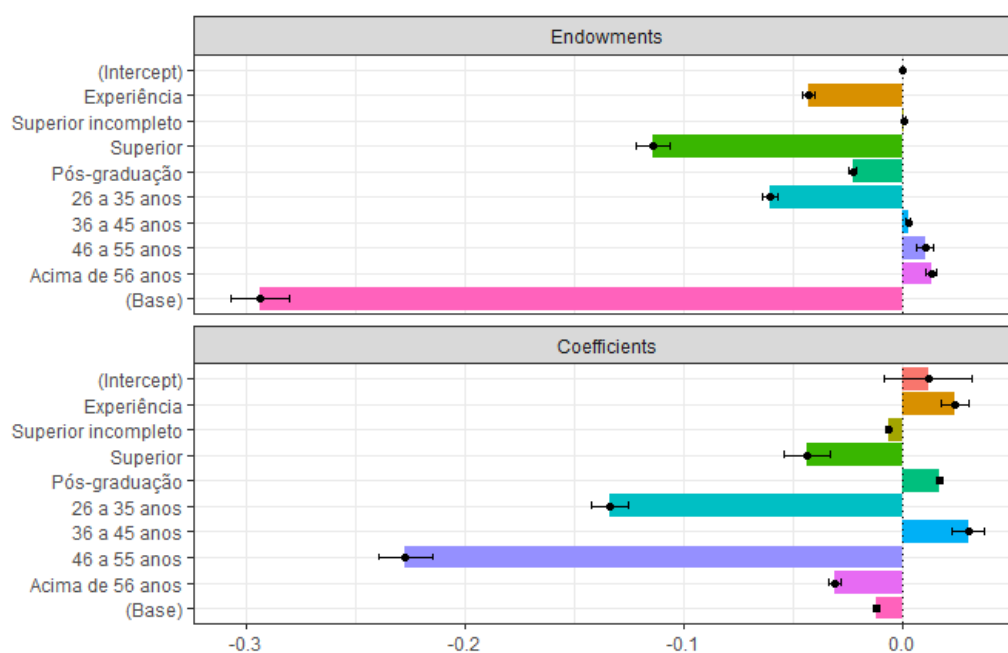
Tabela 11 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando os intervalos de idade

	Diferença Salarial	Explicado	Não explicado
Variação no salário	162,7%	73,9%	44,9%
Participação	-	52,5%	30,5%

Fonte - Elaboração própria com base na RAIS 2018.

Entre as variáveis avaliadas, a figura 19 mostra que o componente de maior relevância dentre os fatores internos é a identificação dos trabalhadores que possuem ensino superior completo. A segunda variável com maior relevância é a *dummy* que sinaliza o efeito do aumento de salário para trabalhadores STEM com idade entre 25 e 35 anos (*agecat2*).

Figura 19 – Componentes explicados e não explicados da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM



Fonte - Elaboração própria com base na RAIS 2018.

Esta superioridade das *dummies* que identificam pessoas com ensino superior e com idade entre 25 a 35 anos revelam características importantes do grupo de trabalhadores STEM que afetam na determinação do salário. Assim, parte do aumento de salário em favor das ocupações STEM advém da maior escolaridade destes profissionais e da maior produtividade atrelada aos trabalhadores de faixa de idade intermediária. De modo geral, pertencer a este grupo de trabalhadores aumenta o salário, em média, em 16% frente aos trabalhadores da categoria omitida (*agecat1*) – com menos de 25 anos.

Por outro lado, ainda que as categorias de idade superiores a esta também tenham influência positiva na determinação dos salários, as faixas etárias são menos presentes entre as ocupações STEM, o que favorece a remuneração dos demais trabalhadores. Outra variável que demonstra uma característica das ocupações ligadas à ciência e tecnologia é o tempo de trabalho na empresa, que é significativamente mais influente para os profissionais STEM, o que revela um *turn over* maior nas demais ocupações de trabalho.

No que concerne aos coeficientes não explicados, a variável de experiência em termos de tempo de trabalho apresenta uma discriminação negativa para as ocupações de ciência e tecnologia, enquanto que o aumento da faixa etária de trabalho e a *dummy* de ensino superior completo têm efeito contrário. A sequência da análise consta com modelos que individualizam as categorias de idade para avaliação do prêmio STEM para cada uma delas.

4.4 Prêmio STEM para as categorias de idade

Com o recorte da primeira faixa etária, o modelo *Agecat1* apresenta diferença salarial de 39,6%. Neste caso, as diferenças de remuneração explicadas pelos componentes internos *endowments* demonstram ter uma maior influência na determinação do salário. Já em relação aos componentes externos, a *dummy* com maior influência positiva é a de idade, em favor aos profissionais não ligados à ciência ou à tecnologia.

Tabela 12 – Decomposição do salário dos trabalhadores STEM em relação aos Não-STEM considerando o intervalo até 25 anos de idade

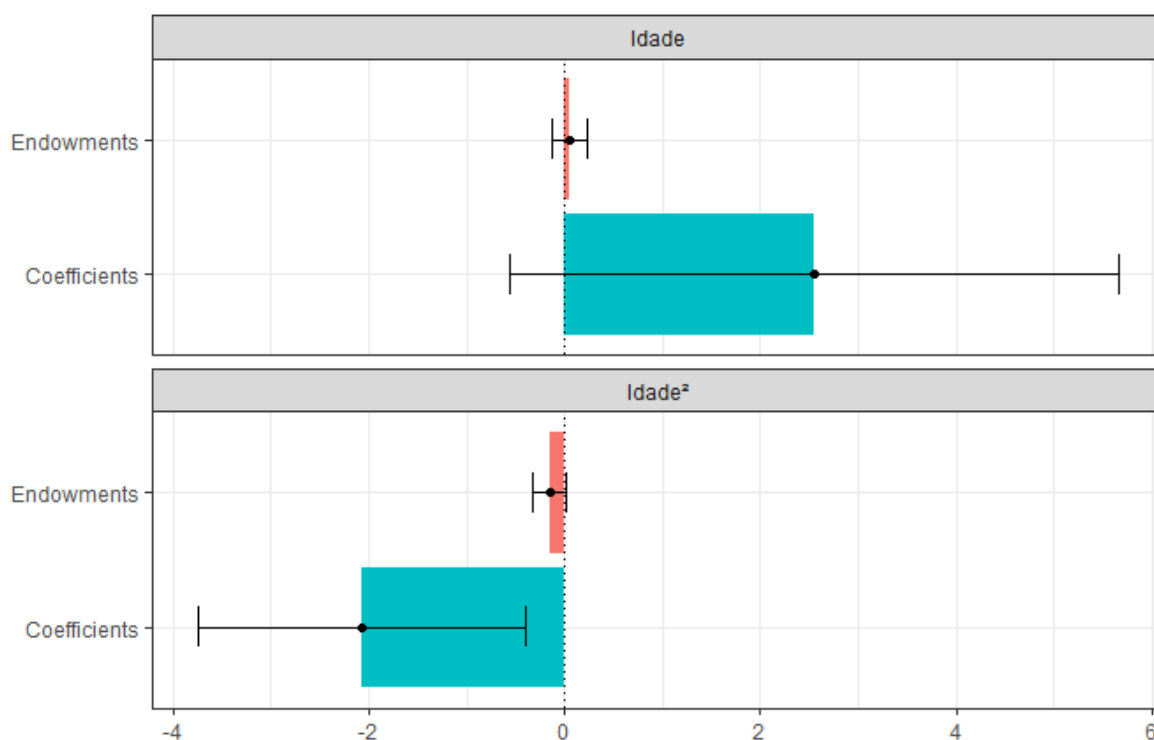
	Diferença Salarial	Explicado	Não explicado
Variação no salário	65,7%	40,5%	30,4%
Participação	-	67,3%	52,5%

Fonte - Elaboração própria com base na RAIS 2018.

Esta situação é contrária ao do modelo geral de idades e decorre de duas condições, a primeira é que a média de idade de trabalhadores STEM (dentro da faixa até 25 anos de idade) é maior que a de trabalhadores Não-STEM, que possuem cerca de 1,3 anos a menos. Além disso, a idade se apresenta no grupo STEM como um fator negativo na composição do salário deste grupo, enquanto que para os demais ele é positivo.

A figura 20 mostra os resultados da estimativa para cada variável, juntamente com barras de erro (linhas pretas) que indicam intervalos de confiança de 95%.

Figura 20 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM até 25 anos



Fonte - Elaboração própria com base na RAIS 2018.

Cabe notar que a idade do trabalhador está inclusa neste modelo apenas na faixa etária de análise, que neste caso compõe o grupo com até 25 anos de idade. Em condições lineares, o aumento de idade dos trabalhadores é melhor remunerado (discriminação positiva) para o grupo STEM, o que explica maior parcela da diferença salarial entre as ocupações.

Em relação ao prêmio STEM, o grupo de trabalhadores não ligados à ciência e de até 25 anos apresentam uma remuneração -23,3% àqueles que são considerados STEM, valor inferior à média encontrada no modelo geral da segunda seção deste capítulo.

Tabela 13 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 26 a 35 anos de idade

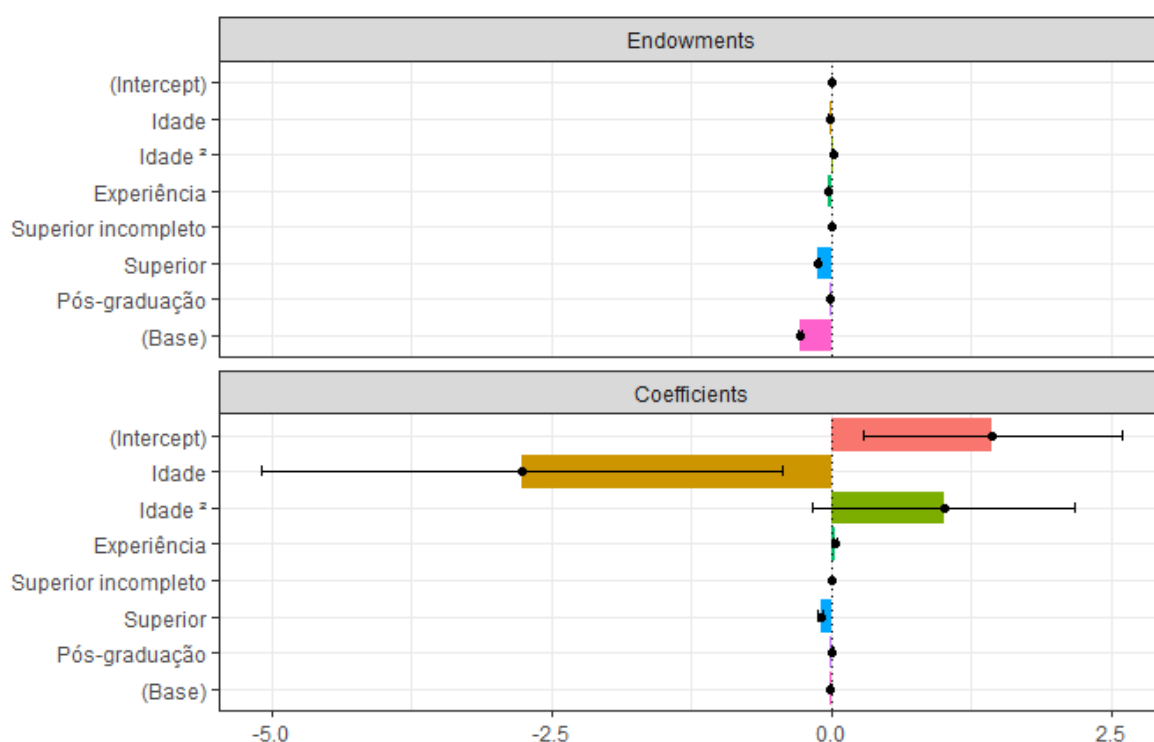
	Diferença Salarial	Explicado	Não explicado
Variação no salário	133%	56,7%	47,5%
Participação	-	53,1%	45,9%

Fonte - Elaboração própria com base na RAIS 2018.

No segundo caso, considerando trabalhadores de 26 e 35 anos, a diferença salarial é de -57,1%, resultante de uma influência de -36,2% dos coeficientes internos e de -32,2% do prêmio STEM.

A figura 21 mostra que as variáveis pertencentes às diferenças explicadas têm pouco destaque, porém as variáveis dos *coefficients* se mostram significativas a favor do grupo STEM quanto à idade e existência de curso superior, enquanto que a variável de idade ao quadrado desfavorece, mais uma vez, os trabalhadores Não-STEM.

Figura 21 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 26 e 35 anos



Fonte - Elaboração própria com base na RAIS 2018.

Para o terceiro grupo (*Agecat3*) com idades entre 36 e 45 anos, a diferença média de salário entre STEM e Não-STEM nesta categoria é de cerca de -65,7%, que reflete em um valor atrelado às características internas de -40,1% e a um prêmio STEM de ordem -34,3%.

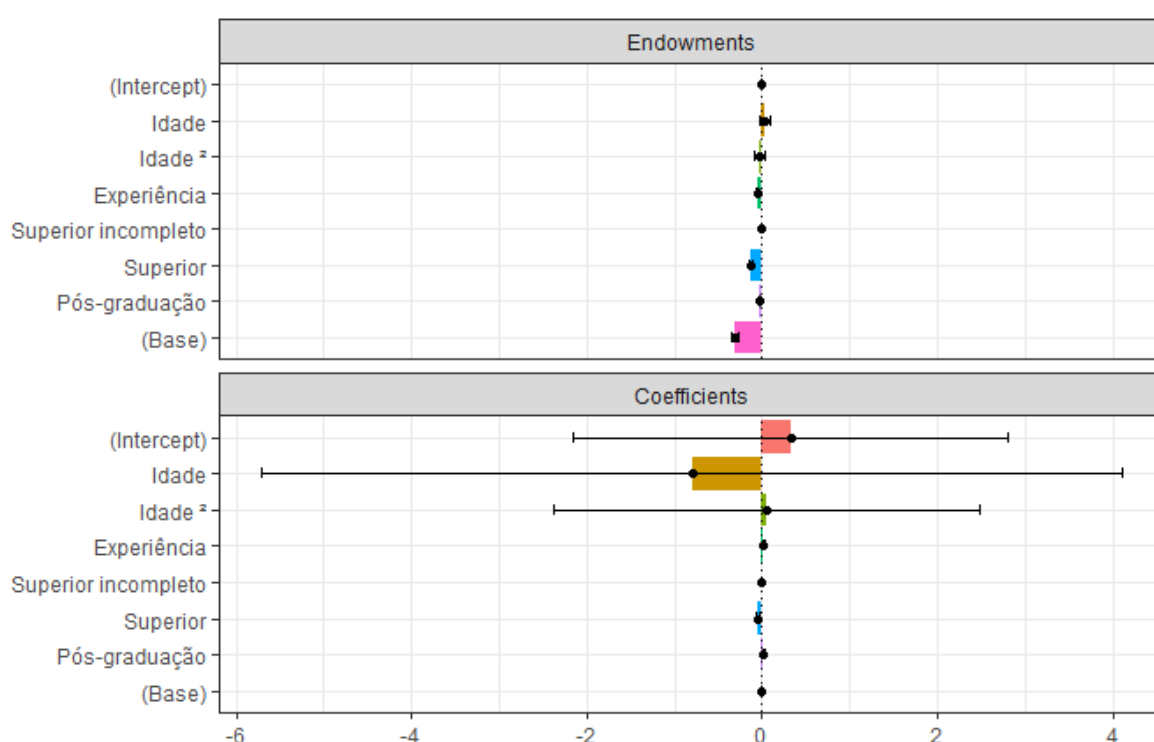
Tabela 14 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 36 a 45 anos de idade

	Diferença Salarial	Explicado	Não explicado
Variação no salário	191,6%	66,9%	52,1%
Participação	-	43,9%	34,7%

Fonte - Elaboração própria com base na RAIS 2018.

De acordo com a figura 22, a variável de ensino superior é a mais relevante a favor do grupo STEM dentre as diferenças explicadas, mesmo que estas se apresentem pouco significativas. Já para a parte não explicada (*coefficients*) a variável de idade continua sendo a com maior capacidade de discriminação que desfavorece as ocupações Não-STEM, seguida pela variável de ensino superior, com menor relevância. Para este caso, a variável de idade ao quadrado tem baixa influência positiva para os profissionais Não-STEM, indicando moderada vantagem ligada à experiência.

Figura 22 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 36 e 45 anos



Fonte - Elaboração própria com base na RAIS 2018.

O quarto grupo categorizado por faixa etária (*Agecat4*), constituído por profissionais com idades entre 46 e 55 anos, apresenta a segunda maior diferença de remuneração mensal dentre os demais, com percentual de -70,8% aos trabalhadores Não-STEM. Assim como os demais, segue-se a tendência de aumento do prêmio STEM, alcançando -34,8%. Os valores atrelados às características influenciam os resultados em -41,8%.

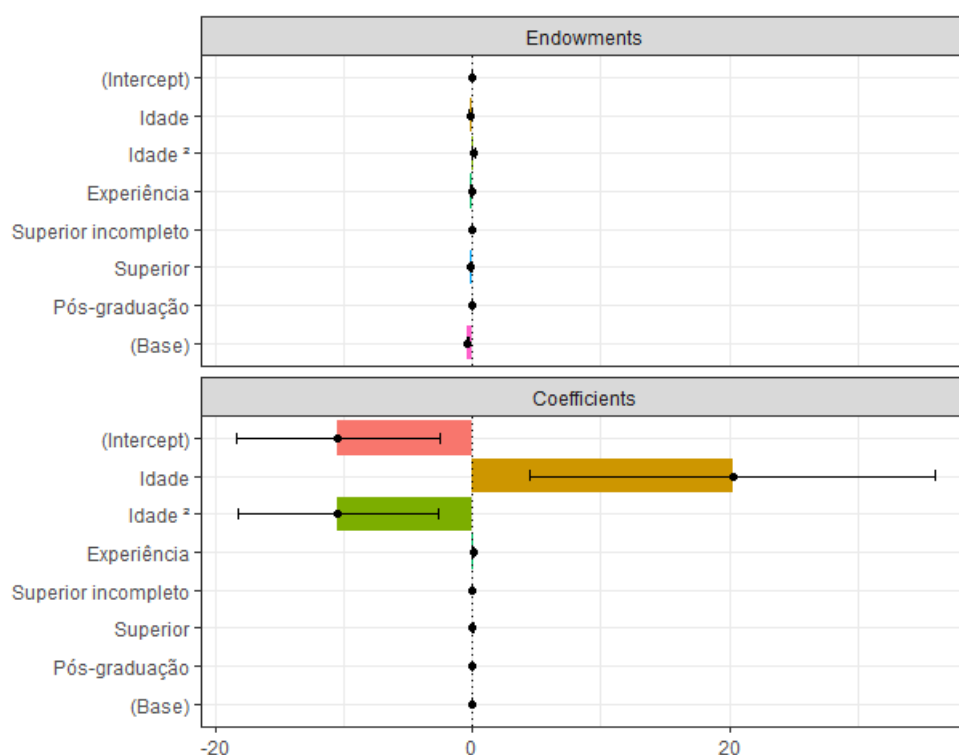
Tabela 15 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo de 46 a 55 anos de idade

	Diferença Salarial	Explicado	Não explicado
Variação no salário	242,6%	71,7%	53,4%
Participação	-	43,9%	34,7%

Fonte - Elaboração própria com base na RAIS 2018.

A variável de idade é que mais se destaca nos coeficientes tanto de *endowments*, como um fator positivo aos trabalhadores Não-STEM, pois este grupo possui cerca de 0,3 anos a mais de idade, em média, que o grupo ligado às ocupações de ciência e tecnologia. Já no quesito não explicado (*coefficients*), a idade se apresenta como um fator desfavorável aos Não-STEM, o que é resultado de um aumento mais significativo de salário dos trabalhadores STEM em decorrência de mais anos de idade.

Figura 23 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM entre 46 e 55 anos



Fonte - Elaboração própria com base na RAIS 2018.

Por fim, o quinto grupo (*Agecat5*) tem a maior diferenciação de salário, com -78,5% e o maior prêmio STEM, equivalente a -36,8%. Este grupo é constituído somente por indivíduos com mais de 56 anos, e tem as variáveis internas como principais determinantes da diferenciação salarial, assim como todos os demais grupos acima. Neste caso, os *endowments* influenciam o salário dos Não-STEM em -57,7%.

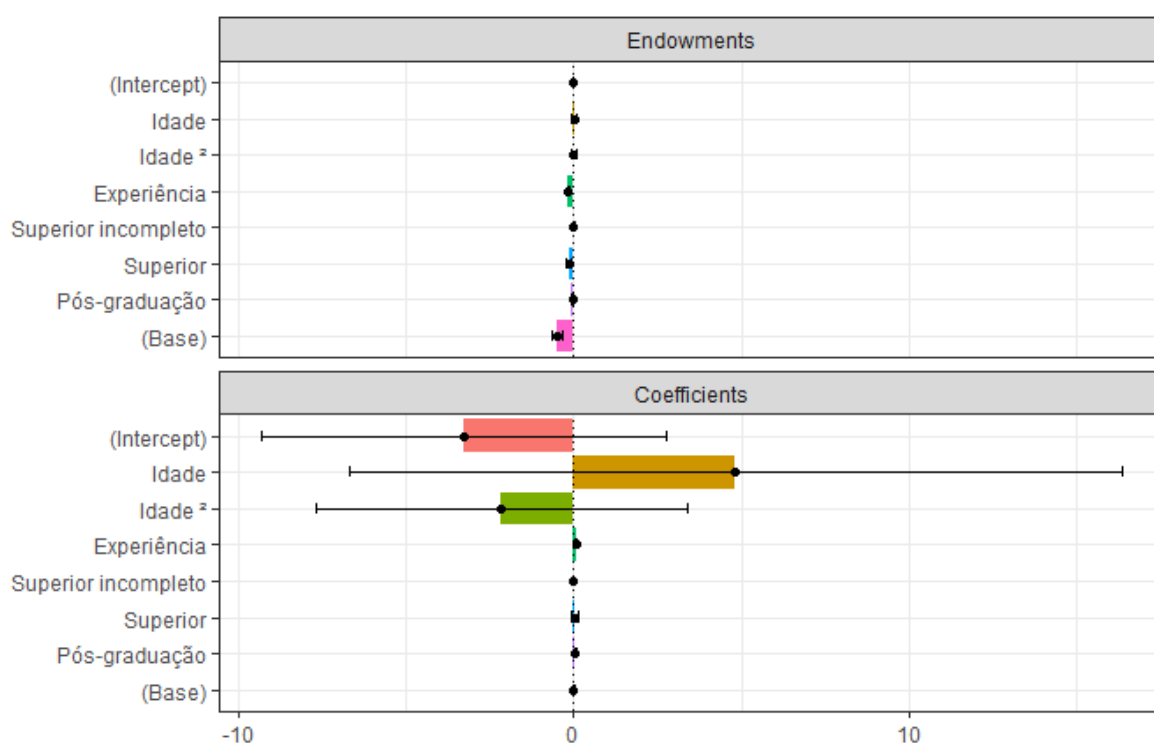
Tabela 16 – Decomposição do salário dos trabalhadores Não-STEM em relação aos STEM considerando o intervalo acima de 55 anos de idade

	Diferença Salarial	Explicado	Não explicado
Variação no salário	365,3%	136,5%	58,2%
Participação	-	56%	29,8%

Fonte - Elaboração própria com base na RAIS 2018.

Este caso se assemelha ao grupo de menor faixa de idade (até 25 anos de idade), que também apresenta discriminação positiva de idade em favor dos Não-STEM. Tal fato também decorre de uma relação negativa entre o salário e a idade para o grupo STEM, ou seja, a cada ano a mais de vida entre estes trabalhadores, haveria uma diminuição do salário. Esta relação é contrária ao modelo geral de idades e nas demais categorias de idade, existindo apenas nas faixas mais jovens e mais maduras.

Figura 24 – Componentes de idade da decomposição da diferença salarial entre trabalhadores STEM e Não-STEM acima de 56 anos



Fonte - Elaboração própria com base na RAIS 2018.

De qualquer modo, a decomposição de Oaxaca-Blinder seguiu os resultados encontrados na avaliação descritiva deste trabalho, em que o aumento das faixas de idade eram acompanhados por maior diferenciação salarial. Este aumento se refletiu em elevações em ambos os componentes da decomposição, os *endowments* e os *coefficients*. Resultou,

portanto, em um aumento do prêmio STEM a medida que se é contemplado maior nível de idade dos trabalhadores.

5 CONCLUSÃO

As ocupações de ciência e tecnologia têm ganhado relevância no contexto do desenvolvimento econômico, uma vez que a produtividade e a inovação são fatores fundamentais para elevar o crescimento de longo prazo das nações. Neste sentido, este trabalho visa avaliar os determinantes do componente salarial das ocupações STEM em Santa Catarina, decompondo as suas remunerações a partir de componentes como a experiência, idade e escolaridade, além da agregação em grupos de gênero e de faixas de idade.

Através da revisão de literatura é reconhecida a relevância da teoria de crescimento econômico como sustentação para os princípios de diferenciais de salários, bem como para a compreensão dos hiatos salariais entre gênero. Juntas, essas teorias corroboram para justificar o estudo da força de trabalho STEM e seu comportamento frente às demais ocupações.

Para investigar a vantagem salarial dos trabalhadores das áreas de ciência e tecnologia - reunidos no acrônimo STEM - foi aplicada a metodologia de decomposição de Oaxaca-Blinder. Partiu-se da hipótese de que em Santa Catarina ocorre um prêmio salarial associado às ocupações STEM, bem como desvantagem salarial de gênero, e que esse prêmio varia também de acordo com a idade. A aplicação desta técnica discorreu sobre a sua construção e formulação por meio do pacote *Blinder-Oaxaca Decomposition* em linguagem R, utilizado também para o tratamento dos dados e estimação dos modelos.

Após o tratamento dos dados, a análise descritiva resultou que o número de trabalhadores homens é superior ao de mulheres, tanto nas ocupações relacionadas à ciência e tecnologia quanto nas demais. Além da maior quantidade, o gênero masculino também apresenta, em média, mais anos de idade e maior salário-hora.

A densidade dos salários-hora, em *log*, confirmaram que há uma concentração mais elevada de ocupações Não-STEM em faixas salariais menores. Isto também ocorreu para as mulheres e para as pessoas mais jovens, com especial concentração dos jovens em profissões de baixo nível salarial. Para todas as faixas de idade, as mulheres recebem, em média, salários inferiores aos dos homens, o que se intensifica na medida em que se aumentam os anos de vida. Além disso, para ambos os gêneros foram encontrados picos na densidade salarial das ocupações STEM. Esta intensificação de trabalhadores em um nível salarial mais elevado chama atenção e pode ser objeto de estudo para pesquisadores interessados no comportamento salarial das ocupações de ciência e tecnologia.

Quanto às análises estatísticas do pacote *Blinder-Oaxaca*, foram encontrados resultados onde a diferença salarial de profissionais STEM com os demais tem um componente não explicado pelas variáveis da pesquisa em um nível de 63,8%. Em outras palavras, apenas cerca de 45% da vantagem salarial dos trabalhadores de ciência e tecnologia é

explicada por diferenças de escolaridade, experiência e idade, enquanto que o restante está atrelado a componentes não explicados, denominado prêmio STEM.

Alinhado ao que vem sendo apontado pela literatura internacional, os resultados apontam que os homens recebem em média mais que as mulheres em ambos os grupos (STEM e Não-STEM), mas a discriminação é maior entre os trabalhadores Não-STEM. Diferente do caso de ocupações de ciência e tecnologia. Neste último, foi encontrado o resultado de que as características de escolaridade, experiência e idade deveriam conduzir a remuneração mais elevada para as mulheres, o que não ocorre devido ao fator não explicado pelo modelo.

Para as coortes de idade, a vantagem salarial do grupo STEM aumenta conforme as categorias de idade se elevam, entretanto a parte não explicada vai diminuindo ao longo dessas categorias. Porém este *gap* se reduz para os trabalhadores ligados à tecnologia.

Além dos resultados acima citados, não foi explorada ainda a questão da análise ao longo de toda a distribuição quantílica, como técnica de decomposição do diferencial de rendimentos. O método *Recentered Influence Function (RIF)* por exemplo, permite a realização da análise dos rendimentos além da média, podendo acontecer em qualquer estatística de distribuição, como por exemplo quantis. Além disso, esse método pode ter seu procedimento facilmente interpretado economicamente, uma vez que, ao contrário de outros tipos de abordagens distributivas, não exige intensivo processamento computacional ou re-amostragem múltipla. Essas especificações, aplicadas ao contexto STEM entre gênero e idades podem ser exploradas em trabalhos futuros.

REFERÊNCIAS

- ACATE. TECH REPORT 2020: Panorama do Setor de Tecnologia Catarinense. Associação Catarinense de Empresas de Tecnologia, Florianópolis, 2020.
- ARROW, K. et al. The theory of discrimination. *Discrimination in labor markets*, Princeton, v. 3, n. 10, p. 3–33, 1973.
- BECKER, G. S. *The Economics of Discrimination*. [S.l.]: Chicago: University of Chicago Press, 1957.
- BECKER, G. S. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, The University of Chicago Press, v. 70, n. 5, Part 2, p. 9–49, 1962.
- BECKER, G. S. *The Economic Way of Looking at Life*. Citeseer, 1992.
- BECKER, G. S.; CHISWICK, B. R. Education and the Distribution of Earnings. *The American Economic Review*, JSTOR, v. 56, n. 1/2, p. 358–369, 1966.
- BEEDE, D. N. et al. Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, n. 04-11, 2011.
- BILTAGY, M. Estimation of gender wage differentials in Egypt using Oaxaca decomposition technique. *Topics in Middle Eastern and African Economies*, v. 16, n. 1, p. 17–42, 2014.
- BLAU, F. D.; KAHN, L. M. Gender differences in pay. *Journal of Economic Perspectives*, v. 14, n. 4, p. 75–99, 2000.
- BLAU, F. D.; KAHN, L. M. The gender pay gap: Have women gone as far as they can? *Academy of Management Perspectives*, Academy of Management Briarcliff Manor, NY 10510, v. 21, n. 1, p. 7–23, 2007.
- BLAU, F. D.; KAHN, L. M. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, v. 55, n. 3, p. 789–865, 2017.
- BLINDER, A. S. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, JSTOR, p. 436–455, 1973.
- BONINI, P.; CUSTODIO, C.; SILVA, F. Quanto temos de trabalho STEM em Santa Catarina? In: *XIII Encontro de Economia Catarinense*. Criciúma, Brasil: [s.n.], 2019. p. 340–365.
- COBB-CLARK, D. A.; TAN, M. Noncognitive skills, occupational attainment, and relative wages. *Labour Economics*, Elsevier, v. 18, n. 1, p. 1–13, 2011.
- COX, N. J. *Transformations: an introduction*. Durham University, 2007.
- CRESWELL, J. W. Projeto de pesquisa: Métodos Qualitativo, Quantitativo e Misto. 3 ed. In: . [S.l.]: Artmed Editora, Porto Alegre, 2010.

- CUGINI, S. C. B. et al. A força de trabalho feminina no mercado de trabalho brasileiro: Discriminação salarial por gênero em 2002 e 2011. *Encontro De Economia Paranaense*, v. 11, 2014.
- CUSTODIO, C.; BONINI, P. Educação superior e trabalho em Santa Catarina: um enfoque nas carreiras de aplicação direta de ciência e tecnologia. *Textos de Economia*, v. 22, n. 1, p. 82–112, 2019.
- DAVIES, P. Getting to the Root of STEM. In: *Regional Business Economics Newspaper*. Minneapolis, Estados Unidos: Federal Reserve Bank of Minneapolis, 2014.
- DEY, J. G.; HILL, C. *Behind the pay gap*. [S.l.]: Education Resources Information Center, 2007.
- DOERINGER, P. B.; PIORE, M. J. *Internal labor markets and manpower analysis*. [S.l.]: ME Sharpe, 1985.
- FONSECA, L. F. Discriminação salarial por gênero no mercado de trabalho português. 2017.
- GIL, A. C. Como elaborar projetos de pesquisa. *São Paulo*, v. 5, n. 61, p. 16–17, 2002.
- GOLDIN, C. The quiet revolution that transformed women’s employment, education, and family. *American Economic Review*, v. 96, n. 2, p. 1–21, 2006.
- GOLDIN, C. et al. Understanding the gender gap: An economic history of American women. *OUP Catalogue*, Oxford University Press, 1992.
- GUJARATI, D. N. Econometria Básica. 3 Edição. *Editora: Makron Books*, p. 530–531, 2000.
- HALVORSEN, R.; PALMQUIST, R. et al. The interpretation of dummy variables in semilogarithmic equations. *American economic review*, American Economic Association, v. 70, n. 3, p. 474–475, 1980.
- HLAVAC, M. Oaxaca: Blinder-Oaxaca decomposition in R. *Disponível em: SSRN 2528391*, 2014.
- INEP. *Censo da Educação Superior*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 2018. Disponível em: <<http://portal.inep.gov.br/educacao-superior>>.
- JANN, B. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, SAGE Publications Sage CA: Los Angeles, CA, v. 8, n. 4, p. 453–479, 2008.
- LANDIVAR, L. C. Disparities in STEM employment by sex, race, and Hispanic origin. *Education Review*, v. 29, n. 6, p. 911–922, 2013.
- LANGDON, D. et al. STEM: Good Jobs Now and for the Future. ESA Issue Brief# 03-11. *US Department of Commerce*, ERIC, 2011.
- LIMA, R. Mercado de trabalho: o capital humano e a teoria da segmentação. Instituto de Pesquisa Econômica Aplicada (Ipea), 1980.

- LIPNEVICH, A. A.; ROBERTS, R. D. Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences*, Elsevier Science, v. 22, n. 2, p. 173–177, 2012.
- LOUREIRO, P. R. Uma resenha teórica e empírica sobre economia da discriminação. *Revista Brasileira de Economia*, SciELO Brasil, v. 57, n. 1, p. 125–157, 2003.
- LUCAS Jr., R. E. On the mechanics of economic development. *Journal of Monetary Economics*, Elsevier, v. 22, n. 1, p. 3–42, 1988.
- LUCAS Jr., R. E. Why doesn't capital flow from rich to poor countries? *American Economic Review*, v. 80, n. 2, p. 92–96, 1990.
- MARGINSON, S. et al. STEM: Country comparisons: international comparisons of Science, Technology, Engineering and Mathematics (STEM) education. Final Report. Australian Council of Learned Academies, 2013.
- MINCER, J. Investment in human capital and personal income distribution. *Journal of Political Economy*, The University of Chicago Press, v. 66, n. 4, p. 281–302, 1958.
- MINCER, J. Schooling, experience, and earnings. human behavior & social institutions no. 2. ERIC, 1974.
- NOONAN, R. STEM Jobs: 2017 Update. ESA Issue Brief 02-17. *US Department of Commerce*, Education Resources Information Center, 2017.
- OAXACA, R. Male-female wage differentials in urban labor markets. *International Economic Review*, JSTOR, p. 693–709, 1973.
- OAXACA, R. L.; DICKINSON, D. L. Statistical Discrimination in Labor Markets: An Experimental Analysis. IZA Discussion Paper, 2006.
- OAXACA, R. L.; RANSOM, M. R. Identification in detailed wage decompositions. *Review of Economics and Statistics*, MIT Press, v. 81, n. 1, p. 154–157, 1999.
- OECD. *Education at a Glance 2019*. [S.l.]: OECD Indicators Paris, Organisation for Economic Co-operation and Development, 2019.
- OREIRO, J. L. Progresso tecnológico, crescimento econômico e as diferenças internacionais nas taxas de crescimento da renda per capita. Uma crítica aos modelos neoclássicos de crescimento. *Economia e Sociedade*, v. 8, n. 1, p. 41–67, 1999.
- PHELPS, E. S. The statistical theory of racism and sexism. *The American Economic Review*, JSTOR, v. 62, n. 4, p. 659–661, 1972.
- REICH, M.; GORDON, D. M.; EDWARDS, R. C. A theory of labor market segmentation. *The American Economic Review*, JSTOR, v. 63, n. 2, p. 359–365, 1973.
- ROMER, P. M. Endogenous technological change. *Journal of political Economy*, The University of Chicago Press, v. 98, n. 5, Part 2, p. S71–S102, 1990.
- ROMER, P. M. The Origins of Endogenous Growth. *Journal of Economic perspectives*, v. 8, n. 1, p. 3–22, 1994.

- ROTHWELL, J. *The hidden STEM economy*. [S.l.]: Metropolitan Policy Program at Brookings, 2013.
- SALVATO, M. A. et al. Mercado de trabalho em Minas Gerais e Bahia: Considerações sobre uma análise da discriminação de raça e gênero. *Anais do XIII Seminário sobre a Economia Mineira*. Cedeplar, Universidade Federal de Minas Gerais, 2008.
- SAMPIERI, R. H. et al. *Metodologia de Pesquisa*. [S.l.]: México: McGraw-Hill, 2010.
- SCHULTZ, T. W. Investment in human capital. *The American Economic Review*, JSTOR, v. 51, n. 1, p. 1–17, 1961.
- SCHULTZ, T. W. *O Valor Econômico da Educação*. [S.l.]: Columbia University Press, 1967.
- SCHULTZ, T. W. The rate of return in allocating investment resources to education. *The Journal of Human Resources*, JSTOR, v. 2, n. 3, p. 293–309, 1967.
- SEEMANN, M. W.; BONINI, P. Trabalho STEM no Brasil de acordo com a Classificação Brasileira de Ocupações (CBO). In: *Seminário de Iniciação Científica*. Florianópolis, Brasil: [s.n.], 2017.
- SOLOW, R. M. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, MIT Press, v. 70, n. 1, p. 65–94, 1956.
- SOUZA, M. C. C. d. Mercado de trabalho: abordagens duais. *Revista de Administração de Empresas*, SciELO Brasil, v. 18, n. 1, p. 59–69, 1978.
- STEIN, G.; SULZBACH, V.; BARTELS, M. Relatório sobre o mercado de trabalho do Rio Grande do Sul - 2001-13. *Porto Alegre: FEE*, 2015.
- VIETORISZ, T.; HARRISON, B. Labor market segmentation: positive feedback and divergent development. *The American Economic Review*, JSTOR, v. 63, n. 2, p. 366–376, 1973.
- YUN, M.-S. A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, Wiley Online Library, v. 43, n. 4, p. 766–772, 2005.
- YUN, M.-S. Identification problem and detailed Oaxaca decomposition: a general solution and inference. *Journal of Economic and Social Measurement*, IOS Press, v. 33, n. 1, p. 27–38, 2008.
- ZEILEIS, A.; CROISSANT, Y. Extended model formulas in r: Multiple parts and multiple responses. *Journal of Statistical Software*, v. 34, n. 1, 2010.

APÊNDICE A – OUTPUTS DA DECOMPOSIÇÃO DE OAXACA-BLINDER

Modelos Etapa 1

Tabela 17 – Output Modelo Geral

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	3.541765e-02	1.648751e-04	2.148151e+02	0.000000e+00
<i>idade2</i>	-4.034280e-04	2.107052e-06	-1.914656e+02	0.000000e+00
<i>tempo</i>	2.837134e-03	4.974677e-06	5.703153e+02	0.000000e+00
<i>supinc</i>	2.053874e-01	1.436347e-03	1.429928e+02	0.000000e+00
<i>sup</i>	7.248750e-01	8.514797e-04	8.513121e+02	0.000000e+00
<i>pos</i>	1.437550e+00	3.763141e-03	3.820079e+02	0.000000e+00
<i>Endowments</i>	-0.493401240	0.009600391		
<i>Coefficients</i>	-0.373843679	0.003613962		
<i>Interaction</i>	-0.098591840	0.009216102		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 18 – Output Modelo Não-STEM

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	3.578562e-02	1.616076e-04	2.214352e+02	0.000000e+00
<i>idade2</i>	-4.122682e-04	2.065886e-06	-1.995600e+02	0.000000e+00
<i>tempo</i>	2.793929e-03	4.893346e-06	5.709650e+02	0.000000e+00
<i>supinc</i>	0.22704831	0.00141249	160.74335018	0.00000000
<i>sup</i>	7.628874e-01	8.450383e-04	9.027843e+02	0.000000e+00
<i>pos</i>	1.491764e+00	3.839123e-03	3.885689e+02	0.000000e+00
<i>Endowments</i>	-0.0908961712	0.0006178208		
<i>Coefficients</i>	0.1924220332	0.0007035109		
<i>Interaction</i>	-0.0020507433	0.0003082063		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 19 – Output Modelo STEM

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	0.096320847	0.002058619	46.789062594	0.000000000
<i>idade2</i>	-9.301504e-04	2.551319e-05	-3.645763e+01	2.740842e-284
<i>tempo</i>	2.356058e-03	4.748680e-05	4.961501e+01	0.000000e+00
<i>supinc</i>	1.506170e-01	1.761174e-02	8.552081e+00	1.270247e-17
<i>sup</i>	0.60306021	0.01257363	47.96229574	0.00000000
<i>pos</i>	0.94840452	0.01871659	50.67185448	0.00000000
<i>Endowments</i>	0.056495473	0.006962778		
<i>Coefficients</i>	0.151570856	0.006566305		
<i>Interaction</i>	-0.005305496	0.002769501		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 20 – Output Modelo Feminino

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	2.506035e-02	2.446844e-04	1.024191e+02	0.000000e+00
<i>idade2</i>	-3.045023e-04	3.176972e-06	-9.584669e+01	0.000000e+00
<i>tempo</i>	2.911270e-03	7.381615e-06	3.943948e+02	0.000000e+00
<i>supinc</i>	0.20717931	0.00188733	109.77375856	0.00000000
<i>sup</i>	7.590162e-01	1.090784e-03	6.958447e+02	0.000000e+00
<i>pos</i>	1.487884e+00	4.985295e-03	2.984546e+02	0.000000e+00
<i>Endowments</i>	-0.378477820	0.019514164		
<i>Coefficients</i>	-0.300189737	0.006479279		
<i>Interaction</i>	-0.181178403	0.019161563		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 21 – Output Modelo Masculino

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	4.495450e-02	2.154848e-04	2.086203e+02	0.000000e+00
<i>idade2</i>	-5.038383e-04	2.724012e-06	-1.849618e+02	0.000000e+00
<i>tempo</i>	2.716496e-03	6.479158e-06	4.192668e+02	0.000000e+00
<i>supinc</i>	2.392452e-01	2.077708e-03	1.151486e+02	0.000000e+00
<i>sup</i>	7.718796e-01	1.302641e-03	5.925495e+02	0.000000e+00
<i>pos</i>	1.445394e+00	5.380919e-03	2.686146e+02	0.000000e+00
<i>Endowments</i>	-0.553070451	0.010842121		
<i>Coefficients</i>	-0.293883595	0.004153005		
<i>Interaction</i>	-0.116177628	0.010364601		

Fonte - Elaboração própria com base na RAIS 2018.

Modelos Etapa 2

Tabela 22 – Output Modelo Geral agecat

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>tempo</i>	2.875751e-03	4.956202e-06	5.802327e+02	0.000000e+00
<i>supinc</i>	0.20602566	0.00143889	143.18377492	0.00000000
<i>sup</i>	7.253355e-01	8.542080e-04	8.491322e+02	0.000000e+00
<i>pos</i>	1.439049e+00	3.768632e-03	3.818493e+02	0.000000e+00
<i>agecat2</i>	1.483098e-01	8.963469e-04	1.654603e+02	0.000000e+00
<i>agecat3</i>	0.04530981	0.00104040	43.55035018	0.00000000
<i>agecat4</i>	1.634489e-01	1.126107e-03	1.451452e+02	0.000000e+00
<i>agecat5</i>	0.115560912	0.001560793	74.039858739	0.000000000
<i>Endowments</i>	-0.506064204	0.009249981		
<i>Coefficients</i>	-0.371175882	0.003318162		
<i>Interaction</i>	-0.088596673	0.009155392		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 23 – Output Modelo agecat1

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	7.284564e-02	2.883617e-03	2.526190e+01	1.043672e-140
<i>idade2</i>	-1.084241e-03	6.816386e-05	-1.590639e+01	5.923978e-57
<i>tempo</i>	3.504028e-03	2.496117e-05	1.403792e+02	0.000000e+00
<i>supinc</i>	0.096575701	0.001526233	63.277182734	0.000000000
<i>sup</i>	3.303473e-01	1.933324e-03	1.708701e+02	0.000000e+00
<i>pos</i>	5.612105e-01	3.263724e-02	1.719540e+01	3.013525e-66
<i>Endowments</i>	-0.339830185	0.012185336		
<i>Coefficients</i>	-0.265108573	0.007390071		
<i>Interaction</i>	0.100117832	0.011725576		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 24 – Output Modelo agecat2

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	7.947323e-02	4.656614e-03	1.706674e+01	2.710737e-65
<i>idade2</i>	-1.082376e-03	7.626545e-05	-1.419221e+01	1.039754e-45
<i>tempo</i>	3.576463e-03	1.418017e-05	2.522157e+02	0.000000e+00
<i>supinc</i>	0.205756277	0.002328511	88.363867894	0.000000000
<i>sup</i>	6.088905e-01	1.356709e-03	4.487996e+02	0.000000e+00
<i>pos</i>	1.284833e+00	6.972699e-03	1.842662e+02	0.000000e+00
<i>Endowments</i>	-0.449302761	0.011826691		
<i>Coefficients</i>	-0.388733618	0.005117444		
<i>Interaction</i>	-0.007994271	0.011336001		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 25 – Output Modelo agecat3

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	4.556810e-02	7.755225e-03	5.875793e+00	4.210754e-09
<i>idade2</i>	-5.808662e-04	9.610181e-05	-6.044280e+00	1.501796e-09
<i>tempo</i>	2.857142e-03	1.061575e-05	2.691417e+02	0.000000e+00
<i>supinc</i>	0.291045390	0.003917287	74.297687379	0.000000000
<i>sup</i>	8.031009e-01	1.693387e-03	4.742572e+02	0.000000e+00
<i>pos</i>	1.495777e+00	6.595603e-03	2.267840e+02	0.000000e+00
<i>Endowments</i>	-0.512298302	0.021250305		
<i>Coefficients</i>	-0.419632099	0.005968103		
<i>Interaction</i>	-0.138217514	0.021030524		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 26 – Output Modelo agecat4

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	-0.03073184	0.01253739	-2.45121597	0.01423797
<i>idade2</i>	0.0002803490	0.0001245808	2.2503391033	0.0244280862
<i>tempo</i>	2.404472e-03	9.242433e-06	2.601557e+02	0.000000e+00
<i>supinc</i>	0.365545154	0.006767259	54.016726436	0.000000000
<i>sup</i>	0.92339560	0.00233034	396.24922733	0.00000000
<i>pos</i>	1.567367e+00	8.687448e-03	1.804174e+02	0.000000e+00
<i>Endowments</i>	-0.1152434709	0.0009508567		
<i>Coefficients</i>	0.2531143094	0.0013594965		
<i>Interaction</i>	-0.0002926128	0.0004400081		

Fonte - Elaboração própria com base na RAIS 2018.

Tabela 27 – Output Modelo agecat5

Variável	Coefficiente	Desvio-padrão	t	P > t
<i>idade</i>	2.897127e-02	6.363388e-03	4.552806e+00	5.298510e-06
<i>idade2</i>	-2.418015e-04	5.008104e-05	-4.828204e+00	1.379329e-06
<i>tempo</i>	2.260521e-03	1.309244e-05	1.726584e+02	0.000000e+00
<i>supinc</i>	4.184413e-01	1.303703e-02	3.209636e+01	4.127537e-225
<i>sup</i>	1.059143e+00	4.171647e-03	2.538909e+02	0.000000e+00
<i>pos</i>	1.61400028	0.01455914	110.85824164	0.00000000
<i>Endowments</i>	-0,86076066	0.10607674		
<i>Coefficients</i>	-0.45886802	0.01673746		
<i>Interaction</i>	-0.21793643	0.10516098		

Fonte - Elaboração própria com base na RAIS 2018.

APÊNDICE B – TESTES DE FATOR DE INFLAÇÃO DA VARIÂNCIA (FIV)

Tabela 28 – Teste do Fator de Inflação da Variância (FIV) para o Modelo Geral

Variável	Modelo Geral					
	STEM x Não-STEM	Não-STEM	STEM	Prêmio STEM feminino	Prêmio STEM masculino	
<i>idade</i>	35,8911	35,8838	39,4008	37,8620		34,8754
<i>idade</i> ²	35,7556	35,7473	39,1724	37,7807		34,7192
<i>tempo</i>	1,3021	1,3013	1,4688	1,3434		1,2775
<i>supinc</i>	1,0289	1,0300	1,7355	1,0427		1,0200
<i>sup</i>	1,1280	1,1125	2,3818	1,1317		1,1540
<i>pos</i>	1,0140	1,0089	1,7288	1,0150		1,0146
<i>stem ou gênero</i> *	1,0413	1,0268	1,0126	1,0151		1,0854

Fonte - Elaboração própria. *As variáveis são, respectivamente: stem - genero - genero - stem - stem

Tabela 29 – Teste do Fator de Inflação da Variância (FIV) para modelo por categoria de idades

	Modelo por categoria de idades				
	Até 25 anos	Entre 26 e 35 anos	Entre 36 e 45 anos	Entre 46 e 55 anos	Acima de 56 anos
<i>idade</i>	249,0741	579,2466	1020,7252	1582,7995	282,4628
<i>idade</i> ²	249,3357	579,1943	1020,7585	1582,6912	282,4199
<i>tempo</i>	1,1112	1,0417	1,0718	1,1338	1,1439
<i>supinc</i>	1,0191	1,0249	1,0144	1,0063	1,0044
<i>sup</i>	1,1008	1,0796	1,1070	1,1510	1,1689
<i>pos</i>	1,0005	1,0089	1,0141	1,0175	1,0255
<i>stem</i>	1,0405	1,0468	1,0407	1,0270	1,0390
Média FIV	71,9545	166,2347	292,3902	452,97	81,4663

Fonte - Elaboração própria.

Tabela 30 – Teste do Fator de Inflação da Variância (FIV) para modelo com inserção de categorias de idade

Modelo com inserção de categorias de idade	
Variável	FIV
<i>tempo</i>	1,2887
<i>supinc</i>	1,0296
<i>sup</i>	1,1320
<i>pos</i>	1,0140
<i>agecat2</i>	1,7159
<i>agecat3</i>	2,0122
<i>agecat4</i>	3,0341
<i>agecat5</i>	1,3481
<i>stem</i>	1,0416
Média FIV	1,5129

Fonte - Elaboração própria.

APÊNDICE C – PROGRAMAÇÃO EM LINGUAGEM R

Neste apêndice estão todos os códigos desenvolvidos em linguagem R para a realização desta monografia. Alguns códigos foram adaptados com quebras de linhas para melhor adequação ao arquivo final, portanto para reprodução do *script* é sugerido que as quebras sejam retiradas do código.

```
#####
# Instalando e carregando pacotes #
#####

#pacote para leitura de dados
library('data.table')

#pacote para rodar modelo oaxaca
library('oaxaca')

#pacote para testar multicolinearidade
library('car')

#pacotes para graficos
library('ggribes')
library('ggplot2')
library('viridis')
library('hrbrthemes')
library('hexbin')

#####
# Lendo e Organizando os dados #
#####

#definicao do local
setwd('C:\\Dados\\e\\Scripts\\Rais\\Bases\\Baixadas')

#leitura dos dados
rais <- fread("Sul.txt", sep = ';')
#visualizar o nome das colunas
```

```
names(rais)

#selecionando as colunas
rais <- rais[,c(7,8,22,24,26,32,43,44,47,48)]

#visualizar as primeiras observacoes
head(rais)

#fazer um filtro de quem tem vinculo ativo
rais <- subset(rais, `Vinculo Ativo 31/12` == 1)

#para trocar a virgula por ponto na nova coluna tempo
rais$tempo <- gsub(",", ".", rais$`Tempo Emprego`)

#para trocar de caracter para numerico
rais$tempo <- as.numeric(rais$tempo)

#para trocar a virgula por ponto na nova coluna salario
rais$salario <- gsub("\\.", "", rais$`Vl Remun Dezembro Nom`)
rais$salario <- gsub(",", ".", rais$salario)

#para trocar de caracter para numerico
rais$salario <- as.numeric(rais$salario)

#mantendo somente salarios e horas trabalhadas significantes
rais <- subset(rais, salario > 0)

rais <- subset(rais, `Qtd Hora Contr` > 19)

rais$salhora <- (rais$salario/(rais$`Qtd Hora Contr`*4))

#criando o ln do salario por hora trabalhada
rais$lnw <- log(rais$salario/(rais$`Qtd Hora Contr`)*4)

#verificar os valores da coluna escolaridade com o dicionario
table(rais$`Escolaridade apos 2005`)

#criarvariavel correspondente a Ensino Superior Incompleto, denominada de supinc
rais$supinc <- ifelse(rais$`Escolaridade apos 2005`==8,1,0)

#verificar os valores da coluna nova supinc
```

```

table(rais$supinc)

#criar variavel correspondente a Ensino Superior Completo, denominada de sup
rais$sup <- ifelse(rais$`Escolaridade apos 2005`==9,1,0)

#criar variavel correspondente a Pos-graduacao Completa, denominada de pos
rais$pos <- ifelse(rais$`Escolaridade apos 2005`>9,1,0)

#criar variavel de idade ao quadrado, denominada de idade2
rais$idade2 <- (rais$Idade^2)

#dummy pra genero, sendo o grupo feminino identificado como 1
rais$genero <- (rais$`Sexo Trabalhador`-1)

#gerando as variaveis para os intervalos de idade
#ate 25 anos de idade
rais$agecat1 <- ifelse(rais$`Idade`<=25,1,0)
#de 26 a 35 anos
rais$agecat2 <- ifelse(rais$`Idade`>25 & rais$`Idade`<= 35,1,0)
#de 36 a 45 anos
rais$agecat3 <- ifelse(rais$`Idade`>35 & rais$`Idade`<= 45,1,0)
#de 46 a 55 anos
rais$agecat4 <- ifelse(rais$`Idade`>45 & rais$`Idade`<= 55,1,0)
#acima de 56 anos
rais$agecat5 <- ifelse(rais$`Idade`>55,1,0)

#definir as cbos que sao stem
cbo <- fread('cbo_stem.csv')

#para trocar de int para caracter
cbo$`CBO 2002` <- as.character(cbo$`CBO 2002`)
rais <- left_join(rais, cbo)
table(rais$stem)

#chamar o nulo (nao stem) de zero
rais$stem[is.na(rais$stem)] <- 0

#removendo a base cbo
rm(cbo)

```

```
#####
# Plotando graficos #
#####

#Avaliar o comportamento da variavel salario
summary(rais$lnw)
summary(subset(rais , stem == 0)$lnw)
summary(subset(rais , stem ==1)$lnw)

hist(rais$lnw)
qqnorm(rais$lnw)

#Funcao densidade do lnw
rais %>%
  ggplot(aes(x=lnw)) +
  geom_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)

#Funcao densidade do lnw por categorias de idade

rais$id <- ifelse(rais$`Idade` <=25,"Ate_25_anos",
  ifelse(rais$`Idade` >25 & rais$`Idade` <= 35, "De_26_a_35_anos",
  ifelse(rais$`Idade` >35 & rais$`Idade` <= 45, "De_36_a_45_anos",
  ifelse(rais$`Idade` >45 & rais$`Idade` <= 55, "De_46_a_55_anos",
  ifelse(rais$`Idade` >55, "Mais_que_55_anos", "NA"))))

ggplot(rais , aes(x = lnw, y = id, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "Temp.[F]", option = "C") +
  labs(title = 'Densidade_do_salario_hora_em_log_por_categoria_de_idade') +
  theme_ipsum() +
  theme(
    legend.position="none",
    panel.spacing = unit(0.1, "lines"),
    strip.text.x = element_text(size = 8)
  )

#Funcao densidade do lnw por grupo STEM e nao-STEM

rais$tecnologia <- ifelse(rais$`stem`==1,"STEM", "Nao-STEM")
```

```
ggplot(rais , aes(x = lnw , y = tecnologia , fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "Temp. [F]", option = "C") +
  labs(title = 'Densidade do salario hora em log por grupo STEM e Nao-STEM') +
  theme_ipsum() +
  theme(
    legend.position="none" ,
    panel.spacing = unit(0.1, "lines"),
    strip.text.x = element_text(size = 8)
  )
```

#Funcao densidade do lnw por genero

```
rais$sexo <- ifelse(rais$`genero`==1,"Mulher","Homem")
```

```
ggplot(rais , aes(x = lnw , y = sexo , fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01, alpha=0.7) +
  scale_fill_viridis(name = "Temp. [F]", option = "C") +
  labs(title = 'Densidade do salario hora em log por genero') +
  theme_ipsum() +
  theme(
    legend.position="none" ,
    panel.spacing = unit(0.1, "lines"),
    strip.text.x = element_text(size = 8)
  )
```

#Grafico de escolaridade e salario por grupo Stem e nao stem

```
ggplot(raisagecat5 , aes(x=lnw, y=`Escolaridade apos 2005`, color=tecnologia)) +
  geom_point(size=1) +
  theme_ipsum()
```

#Grafico de densidade 2d do lnw com idade

```
ggplot(rais , aes(x=lnw , y=Idade) ) +
  geom_hex(bins = 70) +
  scale_fill_continuous(type = "viridis") +
  theme_bw()
```

#Grafico de densidade 2d do lnw com tempo

```
ggplot(rais , aes(x=lnw , y=tempo) ) +
  geom_hex(bins = 70) +
  scale_fill_continuous(type = "viridis") +
```

```

theme_bw()

#criando sintese para grafico de bolha STEM e nao-STEM

tabela <- data.frame()

sfreq <- table(subset(raais , stem == 1)$id)
nsfreq <- table(subset(raais , stem == 0)$id)
slnw <- aggregate(x = subset(raais , stem == 1)$lnw ,
by = list(subset(raais , stem == 1)$id), FUN = mean)
nslnw <- aggregate(x = subset(raais , stem == 0)$lnw ,
by = list(subset(raais , stem == 0)$id), FUN = mean)
sid <- aggregate(x = subset(raais , stem == 1)$Idade ,
by = list(subset(raais , stem == 1)$id), FUN = mean)
nsid <- aggregate(x = subset(raais , stem == 0)$Idade ,
by = list(subset(raais , stem == 0)$id), FUN = mean)

stabela <- cbind(slnw , sid , sfreq)
stabela <- stabela[,-c(3,5)]
names(stabela) <- c('categoria ', 'lnw ', 'idade ', 'freq ')
stabela$Grupo <- 'STEM'

nstabela <- cbind(nslnw , nsid , nsfreq)
nstabela <- nstabela[,-c(3,5)]
names(nstabela) <- c('categoria ', 'lnw ', 'idade ', 'freq ')
nstabela$Grupo <- 'Nao-STEM'

tabela <- rbind(stabela , nstabela)

rm(sfreq , nsfreq , slnw , nslnw , sid , nsid , stabela , nstabela)

ggplot(tabela , aes(x=lnw , y=idade , size = freq , fill=Grupo)) +
  geom_point(alpha=0.7 , shape = 21 , color="black") +
  scale_size(range = c(2,12) , name= "Frequencia") +
  theme_ipsum() +
  theme(legend.position="bottom") +
  ylab("Idade") +
  xlab("Ln do salario-hora")

#criando sintese para grafico de bolha para generos

```

```

tabela <- data.frame()

hfreq <- table(subset(rais , genero == 0)$id)
mfreq <- table(subset(rais , genero == 1)$id)
hlnw <- aggregate(x = subset(rais , genero == 0)$lnw ,
by = list(subset(rais , genero == 0)$id), FUN = mean)
mlnw <- aggregate(x = subset(rais , genero == 1)$lnw ,
by = list(subset(rais , genero == 1)$id), FUN = mean)
hid <- aggregate(x = subset(rais , genero == 0)$Idade ,
by = list(subset(rais , genero == 0)$id), FUN = mean)
mid <- aggregate(x = subset(rais , genero == 1)$Idade ,
by = list(subset(rais , genero == 1)$id), FUN = mean)

htabela <- cbind(hlnw , hid , hfreq)
htabela <- htabela[ , -c(3 , 5)]
names(htabela) <- c('categoria ' , 'lnw ' , 'idade ' , 'freq ')
htabela$Grupo <- 'Masculino '

mtabela <- cbind(mlnw , mid , mfreq)
mtabela <- mtabela[ , -c(3 , 5)]
names(mtabela) <- c('categoria ' , 'lnw ' , 'idade ' , 'freq ')
mtabela$Grupo <- 'Feminino '

tabela <- rbind(mtabela , htabela)

rm(hfreq , mfreq , hlnw , mlnw , hid , mid , htabela , mtabela)

ggplot(tabela , aes(x=lnw , y=idade , size = freq , fill=Grupo)) +
  geom_point(alpha=0.7 , shape = 21 , color="black") +
  scale_size(range = c(2 , 12) , name= "Frequencia") +
  theme_ipsum() +
  theme(legend.position="bottom") +
  ylab("Idade") +
  xlab("Ln do salario - hora")

#salvar os results do oaxaca por categoria de idade (omitindo cat de base)
regage <- oaxaca(lnw ~ tempo + supinc + sup + pos + agecat2 + agecat3 +
agecat4 + agecat5 | stem | supinc + sup + pos , data = rais)

summary(regage$reg$reg.pooled.2)$coefficients["tempo" , ]

```



```

summary(regage$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["pos" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["agecat2" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["agecat3" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["agecat4" ,]
summary(regage$reg$reg.pooled.2)$coefficients ["agecat5" ,]

#salvar os results do oaxaca por categoria de idade
raisagecat1 <- subset(rais , agecat1 == 1)
raisagecat1$salhora <- (raisagecat1$salario/(raisagecat1$`Qtd Hora Contr`*4))
regagecat1 <- oaxaca(salhora ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisagecat1)

#Avaliando a relevancia estatistica dos modelos
summary(regagecat1$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regagecat1$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regagecat1$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regagecat1$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regagecat1$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regagecat1$reg$reg.pooled.2)$coefficients ["pos" ,]

raisagecat2 <- subset(rais , agecat2 == 1)
raisagecat2$salhora <- (raisagecat2$salario/(raisagecat2$`Qtd Hora Contr`*4))
regagecat2 <- oaxaca(salhora ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisagecat2)

#Avaliando a relevancia estatistica dos modelos
summary(regagecat2$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regagecat2$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regagecat2$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regagecat2$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regagecat2$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regagecat2$reg$reg.pooled.2)$coefficients ["pos" ,]

raisagecat3 <- subset(rais , agecat3 == 1)
raisagecat3$salhora <- (raisagecat3$salario/(raisagecat3$`Qtd Hora Contr`*4))
regagecat3 <- oaxaca(salhora ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisagecat3)

#Avaliando a relevancia estatistica dos modelos

```

```

summary(regagecat3$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regagecat3$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regagecat3$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regagecat3$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regagecat3$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regagecat3$reg$reg.pooled.2)$coefficients ["pos" ,]

raisagecat4 <- subset(rais , agecat4 == 1)
raisagecat4$salhora <- (raisagecat4$salario/(raisagecat4$`Qtd Hora Contr`*4))
regagecat4 <- oaxaca(salhora ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisagecat4)

#Avaliando a relevancia estatistica dos modelos
summary(regagecat4$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regagecat4$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regagecat4$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regagecat4$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regagecat4$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regagecat4$reg$reg.pooled.2)$coefficients ["pos" ,]

raisagecat5 <- subset(rais , agecat5 == 1)
raisagecat5$salhora <- (raisagecat5$salario/(raisagecat5$`Qtd Hora Contr`*4))
regagecat5 <- oaxaca(lnw ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisagecat5)

reg#Avaliando a relevancia estatistica dos modelos
summary(regagecat5$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regagecat5$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regagecat5$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regagecat5$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regagecat5$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regagecat5$reg$reg.pooled.2)$coefficients ["pos" ,]

#oaxaca
#install.packages('oaxaca')

#carregar o pacote
library(oaxaca)

rais$amostra <- sample(c(0,1), size=nrow(rais), replace=T, prob=c(.9, .1))
#table(rais$amostra)

```

```
backup ← rais
#rais ← backup

#rais ← subset(rais, amostra == 1)

#salvar os results do oaxaca
reg ← oaxaca(lnw ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = rais)

#Avaliando a relevancia estatistica dos modelos
summary(reg$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(reg$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(reg$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(reg$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(reg$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(reg$reg$reg.pooled.2)$coefficients ["pos" ,]

raisf ← subset(rais, genero == 1)

#salvar os results do oaxaca para premio stem mulher
regf ← oaxaca(lnw ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = raisf)

#Avaliando a relevancia estatistica dos modelos
summary(regf$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(regf$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regf$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regf$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regf$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regf$reg$reg.pooled.2)$coefficients ["pos" ,]

rais0 ← subset(rais, genero == 0)

#salvar os results do oaxaca para premio stem homem
reg0 ← oaxaca(lnw ~ Idade + idade2 + tempo + supinc + sup +
pos | stem | supinc + sup + pos, data = rais0)

#Avaliando a relevancia estatistica dos modelos
summary(reg0$reg$reg.pooled.2)$coefficients ["Idade" ,]
```

```
summary(regm$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(regm$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(regm$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(regm$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(regm$reg$reg.pooled.2)$coefficients ["pos" ,]
```

```
plot(reg)
```

```
library("foreign")
write.dta(rais, "rais.dta")
```

```
summary(rais$lnw)
```

```
reg
```

```
#Plot padrao
```

```
plot(reg)
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
```

```
plot(reg, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo" =
"Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))
```

```
summary(reg$reg$reg.pooled.2)$coefficients ["Idade" ,]
```

```
plot(reg, decomposition = "threefold", group.weight = -1,
      unexplained.split = TRUE, components = c("endowments", "coefficients"),
      component.labels = c("endowments" = "Explicado", "coefficients" =
"Discriminacao"), component.left = TRUE,
      variables = c("Idade", "tempo", "sup"), variable.labels =
c("Idade" = "Years_of_Age", "tempo" = "Female",
"sup" = "College_Education"))
```

```
reg$x$x.mean.diff['sup']
```

```
reg_genns
```

```
plot(reg_genns)
```

```
summary(reg_genns$reg$reg.pooled.2)$coefficients ["Idade" ,]
```

```
summary(reg_genns$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(reg_genns$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(reg_genns$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(reg_genns$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(reg_genns$reg$reg.pooled.2)$coefficients ["pos" ,]
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(reg_genns, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo" =
"Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior", "pos"
= "Pos-graduacao"))
```

```
reg_gens
plot(reg_gens)
```

```
summary(reg_gens$reg$reg.pooled.2)$coefficients ["Idade" ,]
summary(reg_gens$reg$reg.pooled.2)$coefficients ["idade2" ,]
summary(reg_gens$reg$reg.pooled.2)$coefficients ["tempo" ,]
summary(reg_gens$reg$reg.pooled.2)$coefficients ["supinc" ,]
summary(reg_gens$reg$reg.pooled.2)$coefficients ["sup" ,]
summary(reg_gens$reg$reg.pooled.2)$coefficients ["pos" ,]
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(reg_gens, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo" =
"Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(regf, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo" =
"Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))
```

```
plot(regf, decomposition = "threefold", group.weight = -1,
      unexplained.split = TRUE, components = c("endowments", "coefficients"),
      component.left = TRUE, variables = c("Idade", "idade2", "sup"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade2",
"sup" = "Superior"))
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(regm, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo" =
        "Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
        "pos" = "Pos-graduacao"))
```

```
plot(regm, decomposition = "threefold", group.weight = -1,
      unexplained.split = TRUE, components = c("endowments", "coefficients"),
      component.left = TRUE, variables = c("Idade", "idade2", "sup"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade2",
        "sup" = "Superior"))
```

```
regf
```

```
plot(regage)
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(regage, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_ ", "tempo"
        = "Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
        "pos" = "Pos-graduacao",
        "agecat2" = "26_a_35_anos", "agecat3" = "36_a_45_anos", "agecat4" =
        "46_a_55_anos", "agecat5" = "Acima_de_56_anos"))
```

```
plot (regagecat1)
```

```
#Plot apenas dos endowments e coeficientes, corrigindo os nomes das variaveis
plot(regagecat1, components = c("endowments", "coefficients"),
      variable.labels = c("Idade" = "Idade", "idade2" = "Idade_2", "tempo"
        = "Experiencia", "supinc" = "Superior_incompleto", "sup" = "Superior",
        "pos" = "Pos-graduacao", "agecat2" = "26_a_35_anos", "agecat3" =
        "36_a_45_anos", "agecat4" = "46_a_55_anos",
        "agecat5" = "Acima_de_56_anos"))
```

```
plot(regagecat1, decomposition = "threefold", group.weight = -1,
      unexplained.split = TRUE, components = c("endowments", "coefficients"),
      component.left = TRUE, variables = c("Idade", "idade2"),
```

```

variable.labels = c("Idade" = "Idade", "idade2" = "Idade2"))

plot (regagecat2)

plot(regagecat2, components = c("endowments", "coefficients"),
variable.labels = c("Idade" = "Idade", "idade2" = "Idade□2", "tempo" =
"Experiencia", "supinc" = "Superior□incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))

plot (regagecat3)
plot(regagecat3, components = c("endowments", "coefficients"),
variable.labels = c("Idade" = "Idade", "idade2" = "Idade□2", "tempo" =
"Experiencia", "supinc" = "Superior□incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))

plot (regagecat4)
plot(regagecat4, components = c("endowments", "coefficients"),
variable.labels = c("Idade" = "Idade", "idade2" = "Idade□2", "tempo" =
"Experiencia", "supinc" = "Superior□incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))

plot (regagecat5)
plot(regagecat5, components = c("endowments", "coefficients"),
variable.labels = c("Idade" = "Idade", "idade2" = "Idade□2", "tempo" =
"Experiencia", "supinc" = "Superior□incompleto", "sup" = "Superior",
"pos" = "Pos-graduacao"))

#####
#Modelos Etapa 1#
#####

#Teste VIF para Modelo Geral
> vif(reg$reg$reg.pooled.2)

#Teste VIF para Modelo Nao-STEM
> vif(reg_genns$reg$reg.pooled.2)

```

```
#Teste VIF para Modelo STEM
> vif(reg_gens$reg$reg.pooled.2)

#Teste VIF para Modelo Feminino
> vif(regf$reg$reg.pooled.2)

#Teste VIF para Modelo Masculino
> vif(regm$reg$reg.pooled.2)

#####
#Modelos Etapa 2#
#####

#Teste VIF para Modelo Geral agecat
> vif(regage$reg$reg.pooled.2)

#Teste VIF para Modelo agecat1
> vif(regagecat1$reg$reg.pooled.2)

#Teste VIF para Modelo agecat 2
> vif(regagecat2$reg$reg.pooled.2)

#Teste VIF para Modelo agecat 3
> vif(regagecat3$reg$reg.pooled.2)

#Teste VIF para Modelo agecat 4
> vif(regagecat4$reg$reg.pooled.2)

#Teste VIF para Modelo agecat 5
> vif(regagecat5$reg$reg.pooled.2)
```


ANEXO A – OCUPAÇÕES STEM

Esta relação se encontra no trabalho de [Custodio e Bonini \(2019\)](#) e pode ser acessada neste [endereço](#).

A identificação dos profissionais ocorreu de forma que, dentre as 2.511 ocupações da Classificação Brasileira de Ocupações (CBO), foram identificados 164 códigos que correspondem ao grupo STEM. São códigos de apenas dois grandes grupos, *1 e o 2*. Do *Grande Grupo 1*, apenas os gerentes de apoio (142), cujas famílias de ocupação contêm os gerentes de tecnologia da informação, correspondendo às ocupações de gerente de rede; gerente de desenvolvimento de sistemas, gerente de produção de Tecnologia da Informação; gerente de projetos de TI, Gerente de Segurança da TI, Tecnólogo de Gestão de TI. Do *Grande Grupo 2*, que designa Profissionais de Ciências e das Artes, foram considerados 3 subgrupos principais: todas as subdivisões do subgrupo 20 (pesquisadores e profissionais policientíficos, todas as divisões do subgrupo principal 21 (Profissionais das Ciências Exatas, Física e da Engenharia) e ocupações do subgrupo principal 22 (agrônomos e afins).

Códigos CBO identificados como STEM

142005	203025	212215	213205	214245	214415	214750
142510	203105	212305	213210	214250	214420	214805
142515	203110	212310	213215	214255	214425	214810
142520	203115	212315	213305	214260	214430	214905
142525	203120	212320	213310	214265	214435	214910
142530	203125	212405	213315	214270	214505	214915
142535	203205	212410	213405	214275	214510	214920
142605	203210	212415	213410	214280	214515	214925
142610	203215	212420	213415	214305	214520	214930
201105	203220	213105	213420	214310	214525	214935
201110	203225	213110	213425	214315	214530	214940
201115	203230	213115	213430	214320	214535	221105
201205	203405	213120	213435	214325	214605	221205
201210	203410	213125	213440	214330	214610	222105
201215	203415	213130	214005	214335	214615	222110
201220	203420	213135	214010	214340	214705	222115
201225	211110	213140	214205	214345	214710	222120
202105	211115	213145	214210	214350	214715	222205
202110	211120	213150	214215	214355	214720	222215
202115	211205	213155	214220	214360	214725	203015
202120	211210	213160	214225	214365	214730	203010
203005	211215	213165	214230	214370	214735	212205
213170	214235	214405	214740			