

**Como a composição de gênero da coorte de
ensino médio influencia a escolha de
graduação dos estudantes?**

Uma abordagem de ciência de dados

Dayanne Cristina Pereira Gomes

RELATÓRIO APRESENTADO AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA UNIVERSIDADE DE SÃO PAULO
PARA EXAME DE QUALIFICAÇÃO DE
MESTRADO EM CIÊNCIAS

Programa: Ciência da Computação
Orientador: Prof. Dr. Fabio Kon

São Paulo
Dezembro de 2023

**Como a composição de gênero da coorte de
ensino médio influencia a escolha de
graduação dos estudantes?**

Uma abordagem de ciência de dados

Dayanne Cristina Pereira Gomes

Esta é a versão original do texto de
qualificação elaborado pela candidata
Dayanne Cristina Pereira Gomes, tal
como submetido à Comissão Julgadora.

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Do. Or do not. There is no try.

— Mestre Yoda

[illegible]

Resumo

Dayanne Cristina Pereira Gomes. **Como a composição de gênero da coorte de ensino médio influencia a escolha de graduação dos estudantes?: Uma abordagem de ciência de dados.** Exame de Qualificação (Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo, 2023.

[illegible]

Palavras-chave: Palavra-chave1. Palavra-chave2. Palavra-chave3.

Abstract

Dayanne Cristina Pereira Gomes. **How does the gender composition of a high school cohort influences students' major choice?: A data science approach.** Qualifying Exam (Master's). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

[illegible]

Keywords: Keyword1. Keyword2. Keyword3.

Lista de Abreviaturas

SISU	Sistema de Seleção Unificada (<i>Continuous Fourier Transform</i>)
IBGE	Instituto Brasileiro de Geografia e Estatística (<i>Discrete Fourier Transform</i>)
UNICAMP	Universidade Estadual de Campinas (<i>Electron-Ion Interaction Potentials</i>)
STFT	Transformada de Fourier de tempo reduzido (<i>Short-Time Fourier Transform</i>)
ABNT	Associação Brasileira de Normas Técnicas
URL	Localizador Uniforme de Recursos (<i>Uniform Resource Locator</i>)
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo

Lista de Símbolos

ω	Frequência angular
ψ	Função de análise <i>wavelet</i>
Ψ	Transformada de Fourier de ψ

Lista de Figuras

2.1	Tipos de estudos observacionais	3
-----	---	---

Lista de Tabelas

3.1	Estatísticas descritivas das variáveis dependentes por gênero. A coluna "Diferença" reporta o coeficiente do teste-t. P-valor: $p < 0,01$. Desvio padrão em parênteses. Adaptado de SILVA BORGES (2021)	11
-----	--	----

Lista de Programas

Sumário

1	Introdução	1
1.0.1	Motivação e Objetivos	2
2	Fundamental teórico	3
2.1	Estudo de coorte	3
2.2	Estratégias empíricas: Variações idiossincráticas na composição de gênero	5
2.2.1	Efeitos de gênero em escolhas de graduação	6
3	Trabalhos relacionados	9
4	Metodologia	13
4.1	Bases de dados	13
4.1.1	ENEM	13
4.1.2	SISU	14
4.1.3	Censo Escolar	15
4.2	Pré-processamento	15
4.2.1	Redução de dados	16
4.2.2	16
4.3	Análise exploratória de dados	16

Apêndices

Anexos

A	Perguntas Frequentes sobre o Modelo	17
----------	--	-----------

	Referências	19
--	--------------------	-----------

Capítulo 1

Introdução

O momento da escolha de carreira profissional é um acontecimento de grande importância na vida de um indivíduo. Esse processo é determinante na definição de características individuais e coletivas de pessoas que se veem na responsabilidade de tomar uma decisão impactante (AKOSAH-TWUMASI *et al.*, 2018) cujas consequências podem influenciar uma vida inteira. As implicações de optar por um determinado caminho, tanto no âmbito acadêmico como profissional, vão além da preferência por uma área do conhecimento. Elas podem influenciar de forma significativa na qualidade de vida futura, já que sua ocupação tem poder de influenciar personalidade, nível de renda, status social e grupos sociais nos quais os sujeitos se caracterizarão (SHAHID KAZI e AKHLAQ, 2017).

As mudanças ocasionadas pelas transformações sociais fizeram com que se percebesse a necessidade de entender como a juventude é afetada no mundo contemporâneo (UNESCO, 2006). A responsabilidade de tomar decisões significativas não é alheia a esses jovens. Segundo GATI e SAKA (2001), adolescentes estão envolvidos diretamente no processo de escolhas complexas. Preocupações com questões educacionais, como estudos e carreira, mostram um nível de entendimento dos riscos e implicações envolvidos, além das dificuldades que podem surgir.

Esses temas podem ser observados de forma muito expressiva em estudantes do ensino médio. Para aqueles que estão vivendo o término da adolescência e início da vida adulta, a transição envolve não só o amadurecimento etário, mas também a passagem da educação secundarista para a superior. Para uma grande maioria desses estudantes, há a aspiração de adentrar em algum tipo de curso de graduação (VENEZIA e JAEGER, 2013). A escolha de frequentar a universidade pode começar cedo durante a jornada educacional, e nesse caminho até adentrar o curso desejado existem diversas etapas (CABRERA e NASA, 2000). Em todo o mundo, exames educacionais competitivos são aplicados para medir o desempenho de alunos secundaristas. Os resultados desses exames podem ser utilizados para ranquear os melhores candidatos e convocá-los para admissão em uma instituição. Alguns exemplos são ACT e SAT, nos Estados Unidos, Baccalauréat, na França e GCSE, no Reino Unido.

No Brasil, tais exames são conhecidos como vestibular. Os participantes costumam realizá-lo no último ano do ensino médio, quando estão encerrando seus estudos, ou até que consigam a classificação para a vaga pretendida. As instituições têm autonomia para

adotar uma prova específica, como o vestibular tradicional, cuja aplicação é realizada pela própria instituição. Outra opção é a adoção do **ENEM** como forma de entrada. O Exame Nacional do Ensino Médio é uma prova anual com abrangência em todo território brasileiro. A nota desse exame, além de avaliar o desempenho de estudantes secundaristas, pode ser utilizada para adentrar algum curso superior através de programas governamentais, como SISU, ProUni e FIES.

Dadas diversas formas de ingresso, ainda é parte da escolha do estudante qual curso de graduação ele deseja seguir. Múltiplos fatores devem ser considerados nesse processo. Para **BORCHERT (2001)**, existem três grandes áreas que afetam a escolha de carreira: oportunidade, personalidade e ambiente. Segundo **ABBAGNANO (2012)**, ambiente pode ser definido como um complexo conjunto de relações entre mundo natural e ser vivo, que influem na vida e no comportamento do mesmo. O gênero tem um grande papel nas questões relacionadas ao ambiente no qual os jovens se encontram. No cenário brasileiro, as questões de gênero são relevantes e consideradas no processo de levantamento de indicadores sociais. Um exemplo é o estudo **Estatísticas de gênero: indicadores sociais das mulheres no Brasil**, do Instituto Brasileiro de Geografia e Estatística (IBGE). Nele, são levantados diferentes aspectos da vida da população, incluindo desigualdades no mercado de trabalho e na educação.

1.0.1 Motivação e Objetivos

Neste contexto, queremos descrever o perfil dos estudantes, em especial de mulheres, e entender como um fator específico, o gênero, impacta na decisão de carreira. Para o processo de análise das escolhas de um curso de graduação, em particular cursos com histórica baixa representatividade feminina, bem como a composição de turmas do ensino médio, levantamos as seguintes questões de pesquisa:

- **Q1** Há diferença no desempenho de homens e mulheres nas disciplinas do ENEM?
- **Q2** Homens e mulheres que realizam o processo seletivo do SISU fazem escolhas distintas? Em quais cursos estão mais concentrados? Há uma diferença na proporção daqueles que selecionam cursos STEM?
- **Q3** Quais são as características das escolas de ensino médio nas quais os estudantes se formaram?
- **Q4** A composição de gênero da coorte do ensino médio influencia a escolha de graduação das estudantes? Mulheres em escolas com mais colegas do sexo feminino são mais ou menos propensas a selecionarem cursos STEM? Há um efeito sobre a probabilidade das mulheres escolherem o curso de ciência da computação?

Para alcançar os objetivos dessa pesquisa, realizaremos um estudo de coortes levando em consideração os efeitos da composição de gênero. Serão aplicadas técnicas de Ciência de Dados para alcançar os objetivos deste trabalho. A Ciência de Dados é uma área multidisciplinar que envolve técnicas estatísticas, matemáticas e computacionais para resolver problemas de um determinado domínio. Os dados utilizados nesse processo passam uma série de etapas, como tratamento, análise e visualização, a fim de dar suporte à tomada de decisões estratégicas. Trabalharemos com dados educacionais a nível nacional de múltiplas fontes, incluindo ENEM e SISU.

Capítulo 2

Fundamental teórico

2.1 Estudo de coorte

A palavra coorte tem origem no latim *cohors*. Esse termo era utilizado para nomear uma unidade militar do Império Romano, que compunha uma legião. Pode ser considerada equivalente ao conceito moderno de batalhão. No contexto científico, uma coorte pode ser definida como um grupo de pessoas que possuem uma característica ou experiência em comum. Coortes podem ser estabelecidas com vários propósitos. Um deles é a análise de grupos dentro de um determinado domínio, como em estudos econométricos, epidemiológicos e demográficos. Dentro das metodologias utilizadas para explorar uma questão específica estão os estudos observacionais. O que diferencia os estudos de caráter observacional de outros estudos é a realização de intervenções. Nesses estudos, os pesquisadores não interferem nos fenômenos estudados, apenas os observam, fazendo com que a variável considerada não esteja sob controle (SONG e CHUNG, 2010).

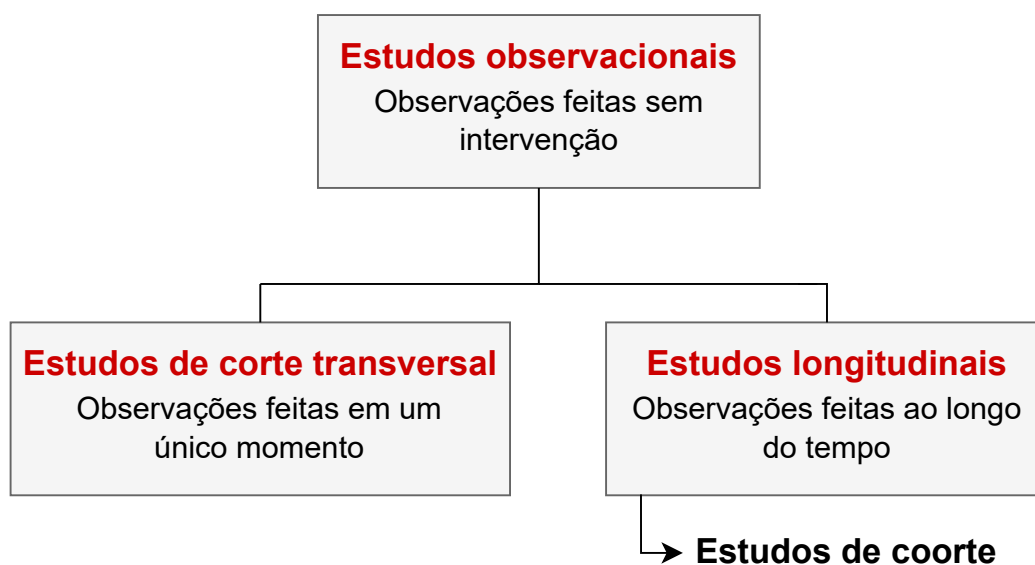


Figura 2.1: Tipos de estudos observacionais

Estudos observacionais podem ser divididos de acordo com o período de coleta de dados. Quando as observações são feitas em um momento específico, coletando dados em um curto intervalo de tempo, chamamos de estudo de corte transversal. Eles são úteis para observar o estado e as condições atuais dos participantes, sem que haja um acompanhamento dos mesmos (ZANGIROLAMI-RAIMUNDO *et al.*, 2018). Alguns exemplos de estudos que seguem essa estrutura no Brasil são o **Censo Demográfico**, realizado a cada 10 anos pelo IBGE, e o **Censo Escolar**, realizado anualmente pelo INEP.

Já quando os indivíduos são acompanhados por longos períodos de tempo, os estudos são classificados como longitudinais. São realizadas coletas de dados contínuas ou repetidas em intervalos regulares, como dias, meses e anos. Como os dados levam em consideração um grupo pré-definido, é possível ajustar os métodos estatísticos para mudanças observadas ao longo do tempo para o grupo como um todo ou para sujeitos específicos (CARUANA *et al.*, 2015). Temos como exemplos o **índice Bovespa**, disponibilizado pela B3 e **taxa Selic**, disponibilizada pelo Banco Central.

Os estudos de coorte são um tipo de estudo longitudinal que leva em consideração uma segmentação específica da população, que é a coorte. Um exemplo de estudo de coorte é a observação do desenvolvimento de doenças em uma população. A amostra populacional pode ser dividida em duas coortes: a coorte 1, de expostos à doença e a coorte 2, de não expostos. Assim, é possível comparar a ocorrência da doença entre os grupos. Estudos de coorte podem ser conduzidos de duas maneiras principais:

- Nos estudos prospectivos, os indivíduos são acompanhados do início do estudo, no presente, durante um período de tempo, para analisar os dados recolhidos no futuro;
- Nos estudos retrospectivos, os dados são coletados e mensurados no passado, acompanhados por um determinado período de tempo, para a realização de uma análise no presente.

Na área da educação, estudos de coorte podem ser utilizados para analisar como questões específicas impactam sujeitos inseridos no sistema educacional, como professores, estudantes e outros envolvidos. BJÖRKENSTAM *et al.* (2011) investigou a associação da performance escolar com taxas de suicídio em estágios posteriores da vida, utilizando uma coorte de nascidos entre 1972 e 1981 na Suécia. Outro exemplo foi o trabalho de ENSMINGER e SLUSARCICK (1992), que examinou os caminhos de desenvolvimento de uma coorte de estudantes negros de uma escola de Chicago entre 1966 e 1977. No sentido dessa pesquisa, pode-se delimitar uma coorte como um grupo de estudantes que estão inseridos no mesmo ambiente escolar. Eles interagem entre si, podem desenvolver relacionamentos e compartilhar experiências. A realização de estudos de coorte educacionais, em particular os retrospectivos, como é o atual, é facilitada pela disponibilização de dados públicos de diferentes abrangências. No Brasil, há a iniciativa do **Portal Brasileiro de Dados Abertos**, com dados a nível federal e local de múltiplas áreas.

2.2 Estratégias empíricas: Variações idiossincráticas na composição de gênero

Variações idiossincráticas podem ser definidas como particularidades e diferenças individuais únicas que podem influenciar resultados ou comportamentos. Esses fatores comuns ou gerais que afetam um grupo de indivíduos também influenciam os sistemas nos quais estes estão inseridos (MEISTER, 1991). Essas variações podem ser diversas, abrangendo uma ampla gama de características, experiências e situações pessoais. Alguns exemplos de variações são:

- **Contexto pessoal:** Dinâmica familiar, status socioeconômico, crenças, *background* cultural, valores familiares;
- **Carreira:** Aspirações e objetivos profissionais, experiência prévia de trabalho, habilidades técnicas;
- **Educação:** Qualidade escolar, atividades extracurriculares, exposição prematura a tópicos avançados.

Por serem altamente individualizadas e frequentemente envolverem múltiplos fatores, é um desafio conduzir uma pesquisa que isole e entenda o impacto de fatores específicos, como o efeito de pares. O efeito de pares (do inglês *peer effect*) se refere à influência e impactos que indivíduos dentro do círculo próximo (familiar, social ou educacional) têm nas suas decisões e comportamentos. Um trabalho muito relevante no entendimento dos efeitos sociais do grupo que uma pessoa se insere é o de MANSKI (1993). Ele aponta a existência de efeitos de pares correlatos, que são comportamentos similares em pessoas do mesmo grupo por conta da semelhança de características individuais ou ambientes institucionais.

Isso é especialmente importante no contexto educacional, já que estudantes do ensino médio compartilham o ambiente escolar durante sua formação. Inseridos em classes diversificadas, eles devem conviver diariamente com outros adolescentes por uma parte relevante de suas vidas. Os colegas de escola podem ser uma importante força social não só na performance acadêmica, mas também nas aspirações profissionais e decisões de seguir uma área específica (TANG *et al.*, 2008).

A literatura observa que os pares têm grande relevância em comportamentos, escolhas e resultados educacionais (SACERDOTE, 2014; ZIMMERMAN, 2003). Para entender melhor esse efeito, é importante considerar como estão estruturados os grupos de referências de estudantes, já que eles são altamente vulneráveis à influência uns dos outros pela exposição contínua e proximidade. Os efeitos dos pares podem se apresentar tanto de maneira positiva, quanto negativa, que se refletem em pontuações de prova, motivação e hábitos de estudo, por exemplo. Uma maneira de fazer isso é através da observação da composição das turmas, isto é, como diferentes distribuições dos pares nas classes de aula afetam os alunos.

Um importante precursor dessas observações é o estudo de HOXBY (2000). Ela identificou e mensurou a existência de efeitos dos pares em coortes escolares que diferem na composição de fatores específicos, como o gênero. Seus resultados sugerem que um grupo

de pares com maioria feminina eleva as pontuações de meninos e meninas em matemática e leitura.

Nesse sentido, múltiplos trabalhos exploram o problema com uma metodologia similar, denominada de variações idiossincráticas na composição de gênero (SCHØNE *et al.*, 2019; LAVY e SCHLOSSER, 2011; SCHNEEWEIS e ZWEIMÜLLER, 2012; BRENØE e ZÖLITZ, 2020; SILVA BORGES, 2021). Eles utilizam estratégias empíricas para levantar hipóteses que ajudem a estabelecer um modelo econométrico para investigar os efeitos de gênero entre os pares.

Os modelos econométricos, apesar de semelhantes, diferem-se por conta de serem estimados utilizando conjuntos de dados educacionais de diferentes países (Noruega, Israel, Áustria, Dinamarca e Brasil), que se configuram em sistemas de ensino distintos, bem como se apoiam em outros recursos, como registros demográficos. Além disso, são consideradas diferentes etapas da educação básica, como anos iniciais e finais do fundamental e médio. Uma lista não exaustiva de variáveis analisadas inclui parcela de estudantes do gênero feminino observadas na coorte escolar, características individuais e escolares e efeitos fixos de escola e coorte.

Como parte da nossa metodologia se propõe a analisar o efeito da composição de gênero nas escolhas de graduação, utilizaremos como referência o trabalho de SILVA BORGES, 2021, que se volta para coortes de escolas de ensino médio brasileiras, um contexto similar ao nosso. Para esse propósito, ela emprega dados do Censo Escolar e do vestibular da Universidade Estadual de Campinas (UNICAMP). As definições teóricas apresentadas, assim como a equação definida, foram extraídas do capítulo *Gender peer effects on major choice* (SILVA BORGES, 2021).

2.2.1 Efeitos de gênero em escolhas de graduação

A hipótese da estratégia empírica é de que o efeito de pares tem um papel significativo na influência das escolhas do curso de graduação. Pais e alunos podem escolher suas turmas potenciais, que pode ser um problema por conta do viés da auto-seleção (*self-selection bias*). Apesar das escolhas pessoais das famílias, elas não conseguem prever corretamente as variações coorte a coorte na composição de gênero dos estudantes, que seguem processos aleatórios. Assim, explorando a variação idiossincrática de coortes na proporção de alunas, temos a estimação do seguinte modelo econométrico, onde i indexa estudante, e escolas, c coorte e t ano do vestibular:

$$y_{iect} = \gamma_0 + \beta_1 \text{Feminino}_i + \beta_2 \text{Prop_FemEM}_{iec} + \beta_3 \text{Feminino}_i \times \text{Prop_FemEM}_{iec} + \alpha_e + \alpha_t + X_i \omega + Z_{ec} \delta + \gamma_{et} + \epsilon_{iect} \quad (2.1)$$

A equação representa um modelo de regressão para estimar o impacto de alguns fatores na variável dependente y_{iect} . Ela é composta por:

- Feminino_i , indicador de gênero;
- Prop_FemEM_{iec} , proporção de colegas do gênero feminino na escola e no ano de

conclusão do ensino médio c ;

- α_e , efeitos fixos da escola;
- α_t , efeitos fixos do ano do vestibular;
- X_i , vetor de características individuais do aluno (notas do ENEM, idade, indicador de trabalho durante o ensino médio);
- Z_{ec} , variáveis da coorte escolar: tamanho da turma, proporção de alunos que frequentam aulas diurnas, idade média dos colegas;
- γ_{et} , tendência linear de tempo específica da escola;
- β_1 , coeficiente de diferenças de gênero nas variáveis de resultados;
- β_2 , coeficiente de efeitos de pares de gênero aplicáveis tanto a homens quanto a mulheres;
- β_3 , coeficiente do impacto diferencial de mulheres terem uma proporção maior de colegas do sexo feminino.

As variáveis de resultado (y_{iect}) podem ser agrupadas de acordo com o objetivo principal. Para analisar se maiores proporções de colegas do sexo feminino durante o ensino médio influenciam a escolha de graduação das alunas, são consideradas duas variáveis: um indicador de área intensiva em matemática e física, outro de escolha de curso STEM. Já para analisar o efeito das colegas do sexo feminino na composição de gênero de um curso, são mensurados se o curso tem maioria de mulheres (*female-dominated*), homens (*male-dominated*) ou é balanceado entre os gêneros (*gender-balanced*) e a média de aplicantes do gênero feminino. Por fim, para analisar o impactos das colegas na probabilidade de escolher cursos competitivos, foram consideradas as médias de candidatos por vaga e de notas de corte.

De forma similar, utilizaremos a metodologia de [SILVA BORGES \(2021\)](#) para analisar os efeitos da composição de gênero das coortes de ensino médio nas opções de carreira dos estudantes, dessa vez numa perspectiva mais abrangente, que é proporcionada pelo conjunto de dados do SISU. Abordaremos com mais detalhes no [Capítulo 4](#) como isso será feito.

Capítulo 3

Trabalhos relacionados

Um estudo de grande relevância para o entendimento de como os pares afetam os colegas na sala de aula é o de [HOXBY \(2000\)](#). Nessa pesquisa, a economista Caroline Hoxby explora como a composição de uma turma escolar pode influenciar a experiência de aprendizado dos estudantes e suas conquistas acadêmicas. São utilizados dados de alunos do 4º ao 7º ano do ensino fundamental de escolas públicas do Texas, Estados Unidos na década de 1990. A autora aborda duas fontes de variação idiossincrática, sendo elas as mudanças na composição de gênero e raça de uma turma escolar em anos adjacentes. Duas estratégias empíricas discutem essas variações: na primeira, são avaliados os efeitos de ter um grupo com maioria feminina e diferente composição racial; na segunda, são avaliados os efeitos das conquistas dos pares em grupos masculinos e femininos.

De forma similar, outros trabalhos abordam os efeitos da composição de gênero através de variações idiossincráticas. [SCHØNE *et al.* \(2019\)](#) visaram entender como a composição de gênero afeta escolhas educacionais. Para isso, foram utilizados dados de alunos dos anos finais do ensino fundamental da Noruega, entre os anos de 2003 e 2008. Nesse cenário, são considerados os efeitos da parcela feminina de pares nas escolhas de áreas e disciplinas do ensino médio, mais especificamente naqueles orientados a STEM. Algumas das hipóteses levantadas investigam se há efeitos na performance escolar, na perpetuação de estereótipos de gênero e na competitividade.

[LAVY e SCHLOSSER \(2011\)](#) apresentam a extensão dos efeitos da composição de gênero na função de produção educacional. Nesse trabalho, são investigados as conquistas educacionais de meninos e meninas em diferentes estágios do sistema educacional. Para isso, são utilizados dados de estudantes de Israel do ensino fundamental, entre os anos de 2002 a 2005, e médio, de 1993 a 2000. São utilizados como resultados as notas de disciplinas e performance em exames de entrada. Alguns dos mecanismos destacados são a disrupção e violência na sala de aula, interação entre estudantes, relacionamento aluno-professor e senso de fadiga de professores na sua profissão.

Outro aspecto considerado pela literatura é a diferença no impacto de escolas segregadas por sexo e coeducacionais, onde meninos e meninas são ensinados juntos, como no trabalho de [SCHNEEWEIS e ZWEIMÜLLER \(2012\)](#). Schneeweis e Zweimüller identificam o impacto causal da composição de gênero na escolha de um campo acadêmico. Os dados cobrem

os anos de 1988 e 2006 da cidade de Linz, Áustria, do ensino fundamental, com foco no último ano. No contexto austríaco, os estudantes podem escolher uma área dentro de uma escola vocacional, que prepara para uma vaga de trabalho, ou seguir estudos superiores na universidade. Além de levar em consideração o interesse vocacional, também é estimado o impacto de estudar em turmas com mais colegas do gênero feminino, que pode levar a escolhas de áreas mais técnicas, a depender da composição.

No trabalho de [BRENØE e ZÖLITZ \(2020\)](#) é observado o efeito dos pares a longo prazo. A utilização de dados de registros de estudantes que ingressam na linha de matemática no ensino médio entre os anos de 1980 e 1994 possibilitou o acompanhamento de toda essa população pelo período de 20 anos. Assim, a investigação dos efeitos da composição de gênero na participação na área STEM na Dinamarca pode acompanhar não só as escolhas educacionais, mas as consequências diretas e retardadas ao longo do tempo em turmas de ensino médio. Leva-se em consideração se a exposição a mais pares femininos está correlacionada com a disparidade de gênero. São observadas probabilidade da entrada e finalização de um curso STEM, bem como os ganhos salariais e a fertilidade de homens e mulheres em diferentes etapas da vida.

Ainda sobre estudos que observam variações idiossincráticas na composição de gênero, temos um recente trabalho que considera o contexto nacional. O trabalho de [SILVA BORGES \(2021\)](#) utiliza essa metodologia para avaliar se a composição de gênero de coortes do ensino médio influencia a escolha de curso de graduação de estudantes, em especial as mulheres. Borges levantou algumas questões específicas, como meninas estudantes expostas a maiores proporções de colegas de seu gênero são mais prováveis de escolher áreas de estudo focadas em matemática. Também são consideradas a seleção em cursos balanceados ou com maioria em relação ao gênero e a competitividade, ilustrada através das taxas de admissão dos cursos. Essa análise foi realizada com dados do vestibular entre os anos 2000 a 2008 de uma universidade pública, a UNICAMP, que foram relacionados aos dados do Censo Escolar. Sendo assim, foi possível traçar as escolhas e os perfis dos estudantes inseridos em turmas de ensino médio. O modelo econométrico de Borges identificou que mulheres são menos prováveis a se aplicarem a áreas focadas em matemática e cursos STEM. Outra observação foi que mulheres estão mais concentradas em cursos de maioria feminina, além de escolherem cursos com alto número de concorrentes por vaga, mas com nota de corte inferior à dos homens. Esses resultados podem ser vistos na ??.

Outros estudos também utilizaram dados educacionais brasileiros para analisar diferentes fatores. [MACHADO e SZERMAN \(2021\)](#) utilizou dados entre os anos de 2010 e 2017 do SISU e do Censo Escolar para investigar os impactos de sistemas de admissão centralizados na composição de estudantes. Machado observou características dos estudantes como gênero, idade, etnia e migração, além de características das escolas para mensurar os efeitos do SISU na atração de candidatos de diversos perfis. [MELLO \(2022\)](#) analisa como reformas educacionais que expandiram o acesso à educação superior impactaram na admissão de estudantes de baixa renda. Essas políticas incluem a expansão da centralização de aplicações com o SISU e mais oferta de cotas de ações afirmativas. São utilizados dados do Censo da Educação Superior dos anos de 2010 a 2015 e do ENEM dos anos de 2009 a 2014.

Ainda sobre estudos que observam variações idiossincráticas na composição de gênero,

Variáveis dependentes	Todos	Mulheres	Homens	Diferença
Matemática ou física são disciplinas prioritárias	0,42 (0,49)	0,26 (0,44)	0,60 (0,49)	-0,33
Curso STEM	0,47 (0,50)	0,34 (0,47)	0,60 (0,49)	-0,27
Área com maioria masculina	0,26 (0,44)	0,08 (0,26)	0,45 (0,50)	-0,38
Área com maioria feminina	0,16 (0,37)	0,26 (0,44)	0,06 (0,23)	0,21
Área balanceada entre os gêneros	0,58 (0,49)	0,66 (0,47)	0,49 (0,50)	0,17
Média de participação de candidatas na carreira	0,50 (0,24)	0,62 (0,18)	0,38 (0,24)	0,24
Média de candidatos por vaga	33,45 (25,94)	36,49 (27,70)	30,25 (23,53)	6,24
Média de nota de corte	525,66 (72,80)	522,14 (78,60)	529,38 (65,93)	-7,24
Observações	139896	71742	68154	139896

Tabela 3.1: Estatísticas descritivas das variáveis dependentes por gênero. A coluna "Diferença" reporta o coeficiente do teste-t. P-valor: $p < 0,01$. Desvio padrão em parênteses. Adaptado de [SILVA BORGES \(2021\)](#)

temos um recente trabalho que considera o contexto nacional. O trabalho de [SILVA BORGES \(2021\)](#) utiliza essa metodologia para avaliar se a composição de gênero de coortes do ensino médio influencia a escolha de curso de graduação de estudantes, em especial as mulheres. Borges levantou algumas questões específicas, como meninas estudantes expostas a maiores proporções de colegas de seu gênero são mais prováveis de escolher áreas de estudo focadas em matemática. Também são consideradas a seleção em cursos balanceados ou com maioria em relação ao gênero e a competitividade, ilustrada através das taxas de admissão dos cursos. Essa análise foi realizada com dados do vestibular entre os anos 2000 a 2008 de uma universidade pública, a UNICAMP, que foram relacionados aos dados do Censo Escolar. Sendo assim, foi possível traçar as escolhas e os perfis dos estudantes inseridos em turmas de ensino médio. O modelo econométrico de Borges identificou que mulheres são menos prováveis a se aplicarem a áreas focadas em matemática e cursos STEM. Outra observação foi que mulheres estão mais concentradas em cursos de maioria feminina, além de escolherem cursos com alto número de concorrentes por vaga, mas com nota de corte inferior à dos homens. Esses resultados podem ser vistos na [Tabela 3.1](#).

Capítulo 4

Metodologia

Neste capítulo, . As bases de dados utilizadas estão descritas na Seção 4.1.

4.1 Bases de dados

Nesta seção, descrevemos as bases de dados utilizadas nesta etapa da pesquisa. Para realizar a análise das escolhas de graduação, utilizamos dados educacionais dos estudantes. Uma das fontes é o Exame Nacional do Ensino Médio (ENEM). Através dele, conseguimos obter informações sobre alunos concludentes e que já concluíram o ensino médio, bem como de suas escolas. A outra fonte utilizada é o Sistema de Seleção Unificada (SISU). Com ela, obtemos informações relativas à inscrição dos alunos em cursos de nível superior, além de detalhes das instituições e cursos ofertados. Uma versão resumida dos dicionários de dados, que explicam as variáveis das bases, está disponível no Apêndice A.

4.1.1 ENEM

O Exame Nacional do Ensino Médio (ENEM) é um exame realizado pelo Ministério da Educação cujo objetivo é avaliar o desempenho escolar no final da educação básica (INEP, 2023b). Desde 2009, ele passou a ser utilizado como mecanismo de ingresso à educação superior, cujas notas podem ser aproveitadas no Sistema de Seleção Unificada (SISU) e Programa Universidade para Todos (ProUni). O exame também possibilita o pleito de certificação do ensino médio. Os participantes realizam provas em quatro áreas de conhecimento: linguagens, ciências humanas, ciências da natureza e matemática. Além disso, eles devem desenvolver um texto dissertativo-argumentativo dada uma situação-problema, conhecido como redação.

Os dados do Exame Nacional do Ensino Médio foram obtidos através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Os dados abertos do Inep incluem os microdados do ENEM, que reúnem um conjunto de informações relativas ao exame. Os microdados são definidos como o menor nível de desagregação de dados recolhidos (INEP, 2023a), sendo disponibilizados dados dos anos de 1998 a 2022. Nesta etapa preliminar da pesquisa, utilizamos os dados do ano de 2016. Além dos microdados relativos às edições anuais, também são disponibilizados outros arquivos relevantes, como

dicionário de dados, documentos técnicos, provas, gabaritos e programa para leitura da base.

Com o passar dos anos, os microdados foram se diferenciando à medida que eram incluídas ou retiradas determinadas variáveis, mas pode-se observar uma estrutura comum entre as edições. Em 2016, os dados estão divididos nas categorias:

- Dados do participante;
- Dados da escola;
- Dados dos pedidos de atendimento especializado;
- Dados dos pedidos de atendimento específico;
- Dados dos pedidos de recursos especializados e específicos para realização das provas;
- Dados dos pedidos de certificação do ensino médio;
- Dados do local de aplicação da prova;
- Dados da prova objetiva;
- Dados da redação;
- Dados do questionário socioeconômico.

Como parte da política adotada pela Lei Geral de Proteção de Dados Pessoais (LGPD), os dados passam por um tratamento antes de serem publicados (INEP, 2023c). Isso significa que dados cadastrais e sensíveis, como nome, endereço, RG, etc, não são disponibilizados ou passam por uma máscara para anonimizá-lo, como é o caso do número de inscrição. Os arquivos de microdados são disponibilizados no formato .csv (valores separados por vírgulas). Os dados de 2016 constituem-se por apenas um com uma tabela. Cada linha da tabela representa a inscrição de um candidato de forma individual, bem como as colunas são as variáveis definidas anteriormente, que caracterizam o participante.

4.1.2 SISU

O Sistema de Seleção Unificada (SISU) é um sistema eletrônico do Ministério da Educação, no qual instituições públicas de ensino superior de todo o Brasil oferecem vagas para estudantes participantes do Exame Nacional do Ensino Médio (MEC, 2023). Durante o período da oferta de vagas, os alunos são ranqueados de acordo com as notas no exame e, aqueles com melhor classificação, são selecionados. Em cada processo seletivo do SISU, que tem duas aberturas anuais, o candidato pode escolher até duas opções de curso. É possível verificar informações sobre as vagas oferecidas, como cursos, instituições e localizações, turnos e modalidade de concorrência.

No ato da inscrição, o sistema recupera as notas da edição mais recente anterior do ENEM. Por exemplo, o SISU 2023 leva em consideração a edição do ENEM 2022. Apenas aqueles que obtiveram nota superior a zero na redação e não têm o status de treineiro no ENEM podem se inscrever. O processo é totalmente digital e gratuito, sendo o estudante o responsável por acompanhar o status da sua inscrição durante o mesmo. Quando não há a

aprovação em uma das duas opções selecionadas, conhecido por chamada regular, ainda é possível a disputa por vaga através da lista de espera.

Os dados do Sistema de Seleção Unificada foram obtidos através do **Portal de Dados Abertos do MEC**. O portal é uma plataforma que disponibiliza dados e informações públicas do Ministério da Educação, que podem ser usadas no desenvolvimento de aplicativos e ações. Além do SISU, é possível observar conjuntos de dados de outros programas como FIES, ProUni e PRONATEC. São disponibilizados dados relacionados às inscrições realizadas nos processos seletivos dos anos de 2017 a 2022. Nesta etapa preliminar da pesquisa, utilizamos os dados do ano de 2017.

São fornecidas informações detalhadas sobre o participante, como dados pessoais e desempenho nas provas do ENEM, a vaga para qual ele se inscreve, além da classificação e aprovação. Diferente do ENEM, não há especificação de categoria dos dados no dicionário fornecido. Também há a aplicação de máscara para anonimizar dados sensíveis, como CPF e número de inscrição. Os arquivos também são disponibilizados no formato .csv. Desta vez, são constituídos por múltiplas tabelas. Cada tabela representa uma etapa do processo de convocação dos candidatos, sendo divididas entre chamada regular e lista de espera. Ocorre duas chamadas regulares ao longo do ano, uma em cada semestre. Já a quantidade de listas de espera varia, conforme o preenchimento de vagas nas etapas anteriores. Nesta etapa, utilizamos apenas os dados das chamadas regulares. Cada linha da tabela representa uma inscrição de um candidato, sendo possível que um candidato tenha múltiplas inscrições por conta das duas aberturas do processo ao longo do ano e por poder se inscrever em mais de um curso.

4.1.3 Censo Escolar

O Censo Escolar é um levantamento de informações da educação básica brasileira em escolas e instituições de ensino por todo o país. Essa ferramenta demográfica realiza coletas anuais em colaboração entre o Inep e as secretarias estaduais e municipais de educação, contando com a participação de todas as escolas públicas (federais, estaduais e municipais) e privadas da rede de ensino. O Censo abrange diferentes etapas e modalidades de ensino da educação básica e profissional. Ele permite a obtenção de dados individualizados, em diversos aspectos, de estudantes, professores, turmas e escolas. A pesquisa é realizada em duas etapas: a primeira coleta informações sobre os estabelecimentos de ensino, gestores, turmas, alunos e profissionais escolares em sala de aula; a segunda, informações sobre o movimento e o rendimento escolar dos alunos.

Ele é um importante recurso no planejamento e

4.2 Pré-processamento

Na etapa da metodologia de pré-processamento, os dados serão tratados a fim de serem melhor processados em termos de poder computacional (tempo de execução, utilização de memória). Com isso, otimizamos as etapas posteriores através de obtenção de um conjunto de dados que seja mais relevante para a pesquisa, facilitando o processo de análise. As técnicas aplicadas estão agrupadas em grupos de acordo com o objetivo em comum entre

elas. São elas redução de dados, integração de dados, limpeza de dados e transformação de dados. Nas subseções subsequentes, apresentaremos como essas técnicas foram utilizadas, relacionando seu propósito com as características das fontes de dados. Utilizamos como referência os trabalhos de

É importante frisar que o processo de ciência de dados não é rígido e estático, mas sim um processo que se flexibiliza à medida que novas necessidades vão surgindo durante o projeto. Assim, as técnicas aplicadas no pré-processamento serão utilizadas além dessa etapa inicial. Para isso, utilizamos a linguagem Python, com as bibliotecas pandas e NumPy para análise e manipulação de dados.

4.2.1 Redução de dados

Os datasets originais possuem um grande volume de dados, tanto pela quantidade de participantes inscritos, quanto pela quantidade de informações armazenadas sobre eles, expressas pelas colunas (ou features). Uma consequência disso é uma maior utilização de recursos computacionais para processá-los, seja nos processos de leitura e escrita de arquivo quanto no armazenamento. Isso será particularmente importante nas etapas futuras por conta da inclusão de novas fontes de dados, como será abordado na seção Plano de trabalho. Assim, as técnicas de redução de dados visam diminuir a quantidade de dados pouco relevantes para a pesquisa, gerando dados menos volumosos e mais representativos.

Todos os datasets possuem o mesmo tipo de arquivo, o formato .csv. O csv é muito popular para uso geral por sua compatibilidade com softwares populares de análise e visualização de dados, como Microsoft Excel. Apesar disso, ele nem sempre é o mais adequado para projetos de ciência de dados. Por isso, optamos por adotar o formato Parquet¹. Comparado ao csv, ele oferece vantagens para grandes datasets. A sua abordagem orientada a colunas permite que sejam lidas apenas as features que serão utilizadas na análise, o que melhora questões de performance, além de oferecer uma eficiente compressão e codificação de dados, gerando arquivos menores e mais fáceis de serem armazenados.

Uma particularidade da biblioteca pandas é a utilização de grandes tipos de dados. Por exemplo, uma variável numérica do tipo inteiro pode ser representada pelos tipos

4.2.2

4.3 Análise exploratória de dados

¹ <https://parquet.apache.org/>

Anexo A

Perguntas Frequentes sobre o Modelo²

- **Não consigo decorar tantos comandos!**

Use a colinha que é distribuída juntamente com este modelo (gitlab.com/ccsl-usp/modelo-latex/raw/master/pre-compilados/colinha.pdf?inline=false).

- **Por que tantos arquivos?**

O preâmbulo \LaTeX deste modelo é muito longo; as partes que normalmente não precisam ser modificadas foram colocadas no diretório `extras`, juntamente com alguns arquivos acessórios. Já os arquivos de conteúdo (capítulos, anexos etc.) foram divididos de maneira que seja fácil para você atualizar o modelo (copiando os novos arquivos ou com um sistema de controle de versões) sem que alterações no conteúdo de exemplo (este texto que você está lendo) causem conflitos com o seu próprio texto.

- **As figuras e tabelas são colocadas em lugares ruins.**

Veja a discussão a respeito na Seção ??.

- **Estou tendo problemas com caracteres acentuados.**

Versões modernas de \LaTeX usam UTF-8, mas arquivos antigos podem usar outras codificações (como ISO-8859-1, também conhecido como `latin1` ou `Windows-1252`). Nesses casos, use `\usepackage[latin1]{inputenc}` no preâmbulo do documento. Você também pode representar os caracteres acentuados usando comandos \LaTeX : `\'a` para á, `\c{c}` para cedilha etc., independentemente da codificação usada no texto³.

- **Existe algo específico para citações de páginas web?**

Biblatex define o tipo “online”, que deve ser usado para materiais com título, autor etc., como uma postagem ou comentário em um blog, um gráfico ou mesmo uma mensagem de email para uma lista de discussão. Bibtex, por padrão, não tem um tipo específico para isso; com ele, normalmente usa-se o campo “howpublished” para especificar

² Esta seção não é de fato um anexo, mas sim um apêndice; ela foi definida desta forma apenas para servir como exemplo de anexo.

³ Você pode consultar os comandos desse tipo mais comuns em en.wikibooks.org/wiki/LaTeX/Special_Chapters. Observe que a dica sobre o pingo do `i` não é mais válida atualmente; basta usar `\i`.

que se trata de um recurso *online*. Se o que você está citando não é algo determinado com título, autor etc. mas sim um sítio (como uma empresa ou um produto), pode ser mais adequado colocar a referência apenas como nota de rodapé e não na lista de referências; nesses casos, algumas pessoas acrescentam uma segunda lista de referências especificamente para recursos *online* (bibtex permite criar múltiplas bibliografias). Já artigos disponíveis *online* mas que fazem parte de uma publicação de formato tradicional (mesmo que apenas *online*), como os anais de um congresso, devem ser citados por seu tipo verdadeiro e apenas incluir o campo “url” (não é nem necessário usar o comando `\url{}`), aceito por todos os tipos de documento do bibtex/bibtex.

- **Aparece uma folha em branco entre os capítulos.**

Essa característica foi colocada propositalmente, dado que todo capítulo deve (ou deveria) começar em uma página de numeração ímpar (lado direito do documento). Se quiser mudar esse comportamento, acrescente “openany” como opção da classe, i.e., `\documentclass[openany,...]{book}`.

- **É possível resumir o nome das seções/capítulos que aparece no topo das páginas e no sumário?**

Sim, usando a sintaxe `\section[mini-titulo]{titulo enorme}`. Isso é especialmente útil nas legendas (*captions*) das figuras e tabelas, que muitas vezes são demasiadamente longas para a lista de figuras/tabelas.

- **Existe algum programa para gerenciar referências em formato bibtex?**

Sim, há vários. Uma opção bem comum é o JabRef; outra é usar Zotero ou Mendeley e exportar os dados deles no formato .bib.

- **Posso usar pacotes L^AT_EX adicionais aos sugeridos?**

Com certeza! Você pode modificar os arquivos o quanto desejar, o modelo serve só como uma ajuda inicial para o seu trabalho.

- **Como faço para usar o Makefile (comando make) no Windows?**

Lembre-se que a ferramenta recomendada para compilação do documento é o latexmk, então você não precisa do make. Mas, se quiser usá-lo, você pode instalar o MSYS2 (www.msys2.org) ou o Windows Subsystem for Linux (procure as versões de Linux disponíveis na Microsoft Store). Se você pretende usar algum dos editores sugeridos, é possível deixar a compilação a cargo deles, também dispensando o make.

- **Como eu faço para...**

Leia os comentários dos arquivos “tese.tex” e outros que compõem este modelo, além do tutorial (Capítulo ??) e dos exemplos do Capítulo ??; é provável que haja uma dica neles ou, pelo menos, a indicação da *package* relacionada ao que você precisa.

Referências

- [ABBAGNANO 2012] Nicola ABBAGNANO. *Dicionário de filosofia*. Mar. de 2012 (citado na pg. 2).
- [MEC 2023] Portal Único de ACESSO AO ENSINO SUPERIOR. SISU. 2023. URL: <https://accessunico.mec.gov.br/sisu> (acesso em 13/10/2023) (citado na pg. 14).
- [AKOSAH-TWUMASI *et al.* 2018] Peter AKOSAH-TWUMASI, Theophilus I. EMETO, Daniel LINDSAY, Komla TSEY e Bunmi S. MALAU-ADULI. “A systematic review of factors that influence youths career choices—the role of culture”. Em: *Frontiers in Education* 3 (jul. de 2018). DOI: [10.3389/feduc.2018.00058](https://doi.org/10.3389/feduc.2018.00058). URL: <https://doi.org/10.3389/feduc.2018.00058> (citado na pg. 1).
- [BJÖRKENSTAM *et al.* 2011] Charlotte BJÖRKENSTAM *et al.* “School grades, parental education and suicide—a national register-based cohort study”. Em: *Journal of Epidemiology & Community Health* 65.11 (2011), pgs. 993–998. ISSN: 0143-005X. DOI: [10.1136/jech.2010.117226](https://doi.org/10.1136/jech.2010.117226). eprint: <https://jech.bmj.com/content/65/11/993.full.pdf>. URL: <https://jech.bmj.com/content/65/11/993> (citado na pg. 4).
- [BORCHERT 2001] Michael BORCHERT. “Career choice factors of high school students”. Em: *Career choice factors* (nov. de 2001). URL: <https://minds.wisconsin.edu/bitstream/handle/1793/40311/2002borchertm.pdf?sequence=1> (citado na pg. 2).
- [BRENØE e ZÖLITZ 2020] Anne Ardila BRENØE e Ulf ZÖLITZ. “Exposure to more female peers widens the gender gap in STEM participation”. Em: *Journal of Labor Economics* 38.4 (out. de 2020), pgs. 1009–1054. DOI: [10.1086/706646](https://doi.org/10.1086/706646). URL: <https://doi.org/10.1086/706646> (citado nas pgs. 6, 10).
- [CABRERA e NASA 2000] Alberto F. CABRERA e Steven M. La NASA. “Understanding the college-choice process”. Em: *New Directions for Institutional Research* 2000.107 (2000), pgs. 5–22. DOI: [10.1002/ir.10701](https://doi.org/10.1002/ir.10701). URL: <https://doi.org/10.1002/ir.10701> (citado na pg. 1).
- [CARUANA *et al.* 2015] Edward Joseph CARUANA, Marius ROMAN, Jules HERNÁNDEZ-SÁNCHEZ e Piergiorgio SOLLI. “Longitudinal studies”. Em: *Journal of Thoracic Disease* 7.11 (2015). ISSN: 2077-6624. URL: <https://jtd.amegroups.org/article/view/5822> (citado na pg. 4).

- [ENSMINGER e SLUSARCICK 1992] Margaret E. ENSMINGER e Anita L. SLUSARCICK. “Paths to high school graduation or dropout: a longitudinal study of a first-grade cohort”. Em: *Sociology of Education* 65.2 (abr. de 1992), pg. 95. DOI: [10.2307/2112677](https://doi.org/10.2307/2112677). URL: <https://doi.org/10.2307/2112677> (citado na pg. 4).
- [INEP 2023a] Instituto Nacional de ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. *Enem*. 2023. URL: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem> (acesso em 13/10/2023) (citado na pg. 13).
- [INEP 2023b] Instituto Nacional de ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. *Exame Nacional do Ensino Médio (Enem)*. 2023. URL: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem> (acesso em 13/10/2023) (citado na pg. 13).
- [INEP 2023c] Instituto Nacional de ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. *Tratamento de Dados Pessoais*. 2023. URL: <https://www.gov.br/inep/pt-br/aceso-a-informacao/tratamento-de-dados-pessoais> (acesso em 13/10/2023) (citado na pg. 14).
- [GATI e SAKA 2001] Itamar GATI e Noa SAKA. “High school students’ career-related decision-making difficulties”. Em: *Journal of Counseling & Development* 79.3 (jul. de 2001), pgs. 331–340. DOI: [10.1002/j.1556-6676.2001.tb01978.x](https://doi.org/10.1002/j.1556-6676.2001.tb01978.x). URL: <https://doi.org/10.1002/j.1556-6676.2001.tb01978.x> (citado na pg. 1).
- [HOXBY 2000] Caroline HOXBY. *Peer Effects in the Classroom: Learning from Gender and Race Variation*. Rel. técn. Ago. de 2000. DOI: [10.3386/w7867](https://doi.org/10.3386/w7867). URL: <https://doi.org/10.3386/w7867> (citado nas pgs. 5, 9).
- [LAVY e SCHLOSSER 2011] Victor LAVY e Analía SCHLOSSER. “Mechanisms and impacts of gender peer effects at school”. Em: *American Economic Journal: Applied Economics* 3.2 (abr. de 2011), pgs. 1–33. DOI: [10.1257/app.3.2.1](https://doi.org/10.1257/app.3.2.1). URL: <https://doi.org/10.1257/app.3.2.1> (citado nas pgs. 6, 9).
- [MACHADO e SZERMAN 2021] Cecilia MACHADO e Christiane SZERMAN. “Centralized college admissions and student composition”. Em: *Economics of Education Review* 85 (2021), pg. 102184. ISSN: 0272-7757. DOI: <https://doi.org/10.1016/j.econedurev.2021.102184>. URL: <https://www.sciencedirect.com/science/article/pii/S027277572100100X> (citado na pg. 10).
- [MANSKI 1993] Charles F. MANSKI. “Identification of endogenous social effects: the reflection problem”. Em: *The Review of Economic Studies* 60.3 (1993), pgs. 531–542. ISSN: 00346527, 1467937X. URL: <http://www.jstor.org/stable/2298123> (acesso em 24/10/2023) (citado na pg. 5).

- [MEISTER 1991] “Chapter 6 - idiosyncratic variables”. Em: *Psychology of System Design*. Ed. por David MEISTER. Vol. 17. Advances in Human Factors/Ergonomics. Elsevier, 1991, pgs. 245–265. DOI: <https://doi.org/10.1016/B978-0-444-88378-0.50011-4>. URL: <https://www.sciencedirect.com/science/article/pii/B9780444883780500114> (citado na pg. 5).
- [MELLO 2022] Ursula MELLO. “Centralized admissions, affirmative action, and access of low-income students to higher education”. Em: *American Economic Journal: Economic Policy* 14.3 (ago. de 2022), pgs. 166–197. DOI: [10.1257/pol.20190639](https://doi.org/10.1257/pol.20190639). URL: <https://doi.org/10.1257/pol.20190639> (citado na pg. 10).
- [SACERDOTE 2014] Bruce SACERDOTE. “Experimental and quasi-experimental analysis of peer effects: two steps forward?” Em: *Annual Review of Economics* 6.1 (ago. de 2014), pgs. 253–272. DOI: [10.1146/annurev-economics-071813-104217](https://doi.org/10.1146/annurev-economics-071813-104217). URL: <https://doi.org/10.1146/annurev-economics-071813-104217> (citado na pg. 5).
- [SCHNEEWEIS e ZWEIMÜLLER 2012] Nicole SCHNEEWEIS e Martina ZWEIMÜLLER. “Girls, girls, girls: gender composition and female school choice”. Em: *Economics of Education Review* 31.4 (ago. de 2012), pgs. 482–500. DOI: [10.1016/j.econedurev.2011.11.002](https://doi.org/10.1016/j.econedurev.2011.11.002). URL: <https://doi.org/10.1016/j.econedurev.2011.11.002> (citado nas pgs. 6, 9).
- [SCHØNE *et al.* 2019] Pål SCHØNE, Kristine von SIMSON e Marte STRØM. “Peer gender and educational choices”. Em: *Empirical Economics* 59.4 (abr. de 2019), pgs. 1763–1797. DOI: [10.1007/s00181-019-01697-2](https://doi.org/10.1007/s00181-019-01697-2). URL: <https://doi.org/10.1007/s00181-019-01697-2> (citado nas pgs. 6, 9).
- [SHAHID KAZI e AKHLAQ 2017] Asma SHAHID KAZI e Abeeda AKHLAQ. “Factors affecting students’ career choice”. Em: *Journal of Research and Reflections in Education* 11 (dez. de 2017), pgs. 187–196 (citado na pg. 1).
- [SILVA BORGES 2021] Bruna Pugiali da SILVA BORGES. “Gender in higher education”. Tese de dout. 2021. DOI: [10.11606/t.12.2021.tde-27052021-215611](https://doi.org/10.11606/t.12.2021.tde-27052021-215611). URL: <https://doi.org/10.11606/t.12.2021.tde-27052021-215611> (citado nas pgs. viii, 6, 7, 10, 11).
- [SONG e CHUNG 2010] Jae W. SONG e Kevin C. CHUNG. “Observational studies: cohort and case-control studies”. Em: *Plastic and Reconstructive Surgery* 126.6 (dez. de 2010), pgs. 2234–2242. DOI: [10.1097/prs.0b013e3181f44abc](https://doi.org/10.1097/prs.0b013e3181f44abc). URL: <https://doi.org/10.1097%2Fprs.0b013e3181f44abc> (citado na pg. 3).
- [TANG *et al.* 2008] Mei TANG, Wei PAN e Mark D. NEWMAYER. “Factors influencing high school students’ career aspirations”. Em: *Professional School Counseling* 11.5 (jun. de 2008), pg. 2156759X0801100. DOI: [10.1177/2156759x0801100502](https://doi.org/10.1177/2156759x0801100502). URL: <https://doi.org/10.1177/2156759x0801100502> (citado na pg. 5).
- [UNESCO 2006] UNESCO. *Juventude e contemporaneidade: possibilidades e limites*. pt. 2006. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000154569> (citado na pg. 1).

- [VENEZIA e JAEGER 2013] Andrea VENEZIA e Laura JAEGER. “Transitions from high school to college”. Em: *The Future of Children* 23.1 (2013), pgs. 117–136. ISSN: 10548289, 15501558. URL: <http://www.jstor.org/stable/23409491> (acesso em 20/10/2023) (citado na pg. 1).
- [ZANGIROLAMI-RAIMUNDO *et al.* 2018] Juliana ZANGIROLAMI-RAIMUNDO, Jorge De Oliveira ECHEIMBERG e Claudio LEONE. “Research methodology topics: cross-sectional studies”. Em: *Journal of Human Growth and Development* 28.3 (nov. de 2018), pgs. 356–360. DOI: [10.7322/jhgd.152198](https://doi.org/10.7322/jhgd.152198). URL: <https://doi.org/10.7322%2Fjhgd.152198> (citado na pg. 4).
- [ZIMMERMAN 2003] David J. ZIMMERMAN. “Peer effects in academic outcomes: evidence from a natural experiment”. Em: *Review of Economics and Statistics* 85.1 (fev. de 2003), pgs. 9–23. DOI: [10.1162/003465303762687677](https://doi.org/10.1162/003465303762687677). URL: <https://doi.org/10.1162/003465303762687677> (citado na pg. 5).

Índice Remissivo

B

biblatex, [18](#)

bibtex, [17](#)

C

Captions, *veja* Legendas

Código-fonte, *veja* Floats

E

Equações, *veja* Modo Matemático

F

Figuras, *veja* Floats

Floats

Algoritmo, *veja* Floats, Ordem

Fórmulas, *veja* Modo Matemático

I

Inglês, *veja* Língua estrangeira

L

Legendas, [18](#)

M

Mendeley, [18](#)

P

Palavras estrangeiras, *veja* Língua estrangeira

R

Rodapé, notas, *veja* Notas de rodapé

S

Subcaptions, *veja* Subfiguras

Sublegendas, *veja* Subfiguras

T

Tabelas, *veja* Floats

V

Versão corrigida, *veja* Tese/Dissertação, versões

Versão original, *veja* Tese/Dissertação, versões

Z

Zotero, [18](#)