

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PRISCILLA MARIA DE LUCA ZAUPA COSTA

EMPREENDEDORISMO ESTADO-INDEPENDENTE

PONTA GROSSA

2023

PRISCILLA MARIA DE LUCA ZAUPA COSTA

EMPREENDEDORISMO ESTADO-INDEPENDENTE

State-Independent Entrepreneurship

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia de Produção da Universidade Tecnológica Federal do Paraná.

Orientador: Prof^a. Dr^a Giane Gonçalves Lenzi

PONTA GROSSA

2023



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

21/02/2023 15:28



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Ponta Grossa



PRISCILLA MARIA DE LUCA ZAUPA COSTA

EMPREENDEDORISMO ESTADO-INDEPENDENTE

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestra Em Engenharia De Produção da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Gestão Industrial.

Data de aprovação: 14 de fevereiro de 2023

Dra. Giane Goncalves Lenzi, Doutorado Universidade Tecnológica Federal do Paraná

Dr. Angelo Marcelo Tusset, Doutorado Universidade Tecnológica Federal do Paraná

Dra. Augusta Pelinski Raiher, Doutorado Universidade Estadual de Ponta Grossa

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 14/02/2023.

https://sistemas2.utfpr.edu.br/dpls/sistema/aluno06/mpCADEDocsAssinar.pcRecuperarDoc?p_hash_documento=081AF85B1C61E9605CF0B34EB57169133756C68F

Esta dissertação é dedicada ao meu marido,
Celso, e à minha filha, Manuela, pelo apoio
fundamental para a realização deste trabalho.

AGRADECIMENTOS

À minha orientadora Professora Dra. Giane Gonçalves Lenzi, pela amizade, oportunidade e confiança depositada em mim, por sua dedicação e ensinamentos e pelo exemplo de pesquisadora e pessoa.

Ao coordenador do mestrado e doutorado em engenharia de produção Professor Dr. Angelo Marcelo Tuset, pela amizade e suporte durante esta caminhada.

À Professora Dra. Augusta Pelinski Raiher, pelos comentários e sugestões valiosas sobre esta dissertação.

Aos professores, que com muita paciência e dedicação, ensinaram-me não somente o conteúdo programado, mas também o sentido da amizade e do respeito.

À minha família, pelo apoio incondicional, não apenas neste momento, mas em toda a vida.

À minha amiga Franciele, pelas longas conversas e trocas de informações sobre as disciplinas, as normas etc.

À Universidade Tecnológica Federal do Paraná (UTFPR), pela estrutura e qualidade de ensino e pesquisa.

Aos funcionários da UTFPR por sempre estarem à disposição para resolver questões burocráticas.

Aos amigos, que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período em que me dediquei a este trabalho.

A todos que participaram, direta ou indiretamente do desenvolvimento deste trabalho de pesquisa, enriquecendo o meu processo de aprendizado.

Por fim, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio de bolsa DS, código de financiamento 001.

RESUMO

No ano de 2020, em meio à pandemia da Covid-19 foi registrado o maior número de empreendedores da história do Brasil, sendo que a vocação empreendedora pode não ter sido a principal motivação, mas a perda de renda. Ou seja, esta crise teria impulsionado indivíduos não-empreendedores para a atividade empresarial em busca de renda. Desta forma, o objetivo deste trabalho é analisar se o estado (situação econômica) levaria indivíduos a abrirem empresas, mesmo não tendo um perfil empreendedor. Isto é, seria o empreendedorismo estado-independente? Para tanto, usou-se a técnica econométrica *Propensity Score Matching* em empresas localizadas no município de Ponta Grossa – PR, abertas entre 2019 e 2020. Os resultados apontam que o balanceamento, além de resolver o problema do viés de seleção, ajuda na confirmação da hipótese assumida neste trabalho, o empreendedorismo é estado-independente.

Palavras-chave: empreendedorismo; propensity score matching.

ABSTRACT

In 2020, in the midst of the Covid-19 pandemic, the largest number of entrepreneurs in the history of Brazil was recorded, and the entrepreneurial vocation may not have been the main motivation, but the loss of income. In other words, this crisis would have driven non-entrepreneurs to enter business in search of income. In this way, the objective of this work is to analyze whether the state (economic situation) would lead individuals to open companies, even if they do not have an entrepreneurial profile. That is, is entrepreneurship state-independent? For this, the econometric technique Propensity Score Matching was used in companies located in the municipality of Ponta Grossa - PR, opened between 2019 and 2020. The results indicate that balancing, in addition to solving the problem of selection bias, helps to confirm the hypothesis assumed in this work, entrepreneurship is state-independent.

Keywords: entrepreneurship; propensity score matching.

LISTA DE FIGURAS

Figura 1 – Organograma da inserção deste trabalho na Engenharia de Produção.	11
Figura 2 – Pessoa tendo dois caminhos a seguir.	26
Figura 3 – Pessoa escolhendo um dos dois caminhos.	27
Figura 4 – Construção do contrafactual clonado da pessoa.	27
Figura 5 – Contrafactual clonado da pessoa e as covariadas.	28
Figura 6 – Exemplo de amostra sem balanceamento.	28
Figura 7 – Cálculo do ATT e do ATE na amostra sem balanceamento.	29
Figura 8 – Balanceamento da amostra.	30
Figura 9 – Cálculo do ATT e do ATE na amostra com balanceamento.	30
Figura 10 – A distribuição normal dos resíduos $E(\varepsilon_i) = 0$ indica que o modelo linear é apropriado.	38
Figura 11 – A forma parabólica dos resíduos indicam que o modelo linear não é apropriado.	39
Figura 12 – Gráficos das funções de densidade cumulativa empírica para os dados antes e depois do balanceamento – linha cinza para unidades de controle e a preta para unidades tratadas.	51
Figura 13 – Distribuições dos grupos de tratamento e controle antes e depois do balanceamento – linha cinza para unidades de controle e a preta para unidades tratadas.	52
Figura 14 – Diferença das médias padronizada.	53

LISTA DE TABELAS

Tabela 1 – Modelo Logit – variável dependente: probabilidade da firma ter sido aberta em 2020 (grupo tratado).	50
Tabela 2 – Resumo estatístico para os dados antes do balanceamento.	50
Tabela 3 – Resumo estatístico para os dados balanceados.	51
Tabela 4 – Melhora percentual pelo balanceamento.	52
Tabela 5 – MQO sem o balanceamento. Variável dependente: Ind_Empreendedorismo.	53
Tabela 6 – MQO com o balanceamento. Variável dependente: Ind_Empreendedorismo.	54
Tabela 7 – Base de dados	65

LISTA DE QUADROS

Quadro 1 – Variáveis usadas no modelo econométrico.	25
--	-----------

SUMÁRIO

1	INTRODUÇÃO	10
2	ECONOMIA PONTA-GROSSENSE E O EMPREENDEDORISMO	13
2.1	Sala do Empreendedor de Ponta Grossa	14
2.2	Ponta Grossa no ranking sobre ambiente de negócios	17
3	REFERENCIAL TEÓRICO	19
3.1	Empreendedorismo	19
3.2	Propensity Score Matching	21
4	METODOLOGIA	23
4.1	Estratégia empírica	23
4.2	Método de análise de dados	25
4.2.1	Um exemplo "ingênuo" do problema do viés de solução e como resolvê-lo	25
4.2.2	Propensity Score Matching	30
4.2.3	Modelo Logit	35
4.2.4	Mínimos Quadrados Ordinários	37
5	ANÁLISE DOS RESULTADOS	49
6	CONCLUSÕES	55
	REFERÊNCIAS	56
	APÊNDICE A - SCRIPT DO R	61
	APÊNDICE B - PACOTE MATCHIT	63
	ANEXO A - BASE DE DADOS	65
	ANEXO B - PUBLICAÇÃO NOS ANAIS DO XII CONBREPO	69
	ANEXO C - SÚMULA CURRICULAR	71
	ÍNDICE REMISSIVO	72

1 INTRODUÇÃO

A pandemia do COVID-19 declarada pela Organização Mundial de Saúde (OMS) no dia 11 de março de 2020 tem se apresentado como uma das maiores crises que o mundo já enfrentou, pois não se trata apenas de uma crise sanitária, mas também econômica, pois o tratamento não-farmacológico teve forte impacto no aumento do desemprego e na diminuição da renda. Para se ter uma ideia do quão importante é este tema, nos nove primeiros meses de 2020, o número de microempreendedores individuais (MEIs), no país, cresceu 14,8%, na comparação com o mesmo período de 2019, chegando a 10,9 milhões de registros, segundo dados do Portal do Empreendedor do Governo Federal. Somado às mais de 7,5 milhões de micro e pequenas empresas, esse setor representa 99% dos negócios privados e 30% do Produto Interno Bruto (NITAHARA, 2020).

O governo brasileiro até tentou arrefecer os efeitos do desemprego. Um exemplo foi o Programa Emergencial de Manutenção do Emprego e da Renda (Medida Provisória nº 936 de 2020) cujo objetivo era reduzir a jornada de trabalho e, proporcionalmente, o salário, mediante acordo individual ou negociação coletiva (DIEESE, 2020). Mesmo assim, a taxa de desemprego chegou a 14,9% no primeiro trimestre de 2021 (VEJA, 2021). A solução encontrada por indivíduos que perderam renda, nesta pandemia, foi tornar-se microempresário. Contudo, a atividade empresarial não é algo simples, principalmente, para não-empresários, uma vez que, segundo a pesquisa Sobrevivência de Empresas realizada pelo SEBRAE¹, três em cada 10 MEIs fecham as portas antes de cinco anos (BEDÊ, 2016).

Portanto, reconhecendo a importância da sobrevivência de microempresas para a atividade econômica, o objetivo deste trabalho é analisar se o empreendedorismo é estado-independente, ou seja, se as condições econômicas resultantes de crises interferem ou não na decisão de abrir uma firma. Pois, espera-se que um empreendedor não dependa do estado da economia. Em outras palavras, os não-empresários não entrariam na atividade empresarial em períodos de não- crise. Para atingir o objetivo proposto, usou-se a ferramenta econométrica *Propensity Score Matching* – técnica que busca estimar o efeito de um tratamento contabilizando as covariadas que preveem o recebimento do tratamento. Assim, tenta reduzir o viés de uma estimativa do efeito do tratamento obtida pela simples comparação de resultados entre unidades que receberam o tratamento versus aquelas que não receberam.

Em termos práticos, o procedimento de estudo foi usar uma base de dados de empresas localizadas no município de Ponta Grossa², entre os anos de 2019 e 2020, diferenciando

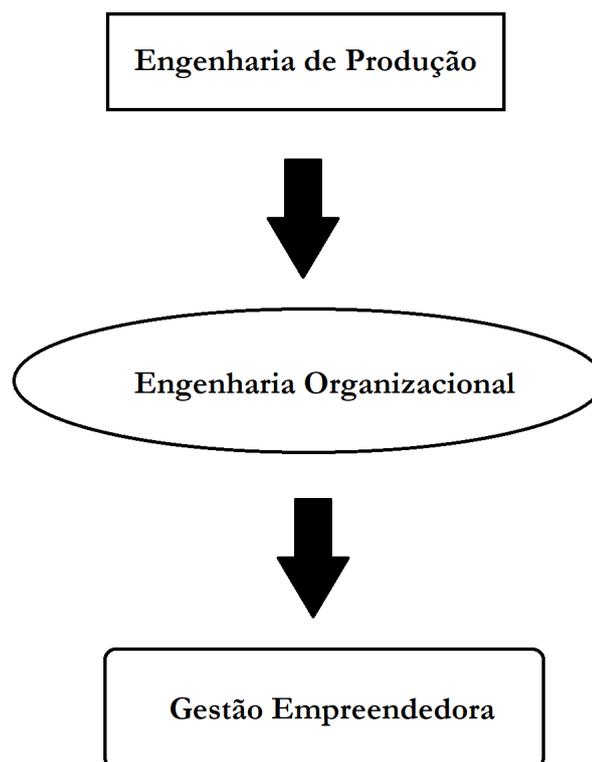
¹ Serviço Brasileiro de Apoio às Micro e Pequenas Empresas.

² Segundo o Índice de Concorrência dos Municípios, publicado em dezembro de 2022 pela Secretaria de Acompanhamento Econômico do Ministério da Economia, pelo Governo Federal, coloca Ponta Grossa na terceira colocação nacional na pontuação geral do índice com 603,13, atrás apenas de duas capitais: Porto Alegre (RS) com 654,21; e Belo Horizonte (MG) com 618,64. Além disso, teve o maior score em relação ao empreendedorismo, 92,39, a segunda colocada foi Brasília com 79,4 (GOVERNO FEDERAL, 2022).

em grupo de tratamento (25 unidades) – empresas abertas em momentos de crises (durante a atual pandemia) – e em grupo de controle (94 unidades) – empresas abertas em momentos de não-cries. Em que, a intuição é que o grupo de controle represente os "verdadeiros" empreendedores. A justificativa para este estudo é que tendo conhecimento que "verdadeiros" empreendedores possuem uma maior probabilidade de permanecerem na atividade empresarial, políticas públicas poderiam ser desenvolvidas para evitar que não-empreendedores se aventurem em algo fora da sua vocação, e isso, diminuiria as perdas econômicas dadas por falências precoces de microempresas.

É importante mencionar que esta dissertação contribui com o conjunto de conhecimentos relacionados à gestão das organizações, englobando em seus tópicos o planejamento estratégico e operacional, as estratégias de produção, a gestão empreendedora, a propriedade intelectual, a avaliação de desempenho organizacional, os sistemas de informação e sua gestão e os arranjos produtivos. A figura 1 apresenta o organograma da inserção deste trabalho na engenharia de produção.

Figura 1 – Organograma da inserção deste trabalho na Engenharia de Produção.



Fonte: Autoria própria (2022).

Nos primeiros quatro meses de 2020, o saldo líquido de empresas abertas foi o maior da última década. Neste aspecto, é possível perceber que, no Brasil, há uma grande necessidade de se empreender para superar as dificuldades apresentadas no período da pandemia, visto que muitas pessoas perderam seus empregos e necessitam de uma fonte de renda para sobreviver (GOVERNO FEDERAL, 2020).

Como uma alternativa em meio à crise econômica e ao desemprego, surge o conceito de “empreendedorismo por necessidade”, que, segundo o SEBRAE, acontece quando não há oferta de trabalho para o indivíduo e este, para garantir sua própria sobrevivência e, às vezes, de seus familiares, encontra saída abrindo uma empresa (SEBRAE, 2020). Portanto, o empreendedorismo por necessidade resulta num atraso para o desenvolvimento econômico social, visto que não há qualquer estudo ou preparo desses indivíduos para se instalarem no mercado (NASSIF; GHOBIL; AMARAL, 2009).

Vale destacar que Schumpeter nega a importância do desenvolvimento econômico nas preferências dos consumidores, pois acredita que os produtores que exercem mudanças relevantes. Em outras palavras, o empreendedorismo é o catalisador do desenvolvimento econômico por meio da organização da produção, das novas formas de produzir e ao combinar diferentes insumos e habilidades, gerando novos produtos, ou os mesmos produtos com melhor qualidade e menor custo (SCHUMPETER, 1982).

Neste trabalho, os resultados obtidos apresentam que o pareamento usando o *propensity score* melhora, consideravelmente, o balanceamento da amostra. Isso, além de ser importante para eliminar o viés de seleção, também ajuda a confirmar a hipótese deste estudo, o empreendedorismo é estado-independente. Pois, a regressão usando a amostra não balanceada não apresentou significância estatística para o coeficiente relacionado à variável de tratamento. Contudo, isto não ocorreu com a amostra balanceada.

Além desta introdução, esta dissertação está estruturada tendo o capítulo dois como uma descrição do município de Ponta Grossa, destacando a sua importância regional, o que justificaria a sua escolha nesta análise. O capítulo três traz o referencial teórico sobre empreendedorismo e sobre o *Propensity Score Matching* – estratégia empírica usada para eliminar o viés de seleção usada nesta dissertação. A metodologia – como a base de dados foi trabalhada e uma discussão mais formal sobre o *Propensity Score Matching* – é apresentada no capítulo quatro. E o capítulo cinco apresenta a análise dos resultados.

2 ECONOMIA PONTA-GROSSENSE E O EMPREENDEDORISMO

Ponta Grossa é um município brasileiro localizado no Estado do Paraná, sendo o quarto mais populoso do Estado, com estimativa de 360 mil habitantes, em 2021. Conta com a nona maior população do Sul do Brasil e a septuagésima sexta do país. Com uma área é 2.054,732 km^2 , Ponta Grossa é conhecida como Princesa dos Campos. A distância até Curitiba, capital administrativa estadual, é de, aproximadamente, 100 quilômetros, e de Brasília, capital federal, é de 1.320 quilômetros (IBGE, 2023).

A economia de Ponta Grossa é uma das mais importantes do Estado do Paraná e se destaca na região Sul por seu posicionamento estratégico na rota do agronegócio. Desde a sua origem, a cidade sempre foi conhecida por estar bem localizada entre grandes centros econômicos e ter forte vocação industrial. Como principal polo financeiro da região dos Campos Gerais, Ponta Grossa não para de subir no ranking das principais economias do Sul do país, registrando forte expansão em seu PIB e índices de desenvolvimento humano.

Este município possui grandes potencialidades e tem diversos investimentos em infraestrutura e transportes. Impulsionada pelo crescimento da indústria e do comércio, a economia de Ponta Grossa já é a quinta maior do Estado do Paraná e uma das mais importantes da região Sul. Um grande diferencial é sua proximidade com grandes centros econômicos, como São Paulo e Curitiba, além de estar localizado em um dos mais importantes corredores de escoamento do agronegócio nacional.

Dessa forma, pelo seu complexo rododiferroviário são transportados produtos essenciais, tais como: grãos (soja, milho, trigo, café); carnes; fertilizantes; matérias-primas para a indústria e para a construção civil; e veículos e peças automotivas. As principais rodovias que passam pelo Município são a PR-151, a BR-376 e a BR-277, que cruzam o Estado do Paraná de norte a sul e atendem parte dos estados do Mato Grosso do Sul, de São Paulo e de Santa Catarina, além do Paraguai. Essas estradas também ligam os polos agrícolas do Centro-Oeste do país e do Norte do Paraná ao Porto de Paranaguá, que escoam a produção dessas regiões e faz o desembarque do maior volume de fertilizantes usados nas lavouras brasileiras. Além do mais, passam por Ponta Grossa duas ferrovias de grande relevância, que dão fluidez ao transporte de cargas do norte do Estado ao Porto de Paranaguá e a Curitiba (PREFEITURA DE PONTA GROSSA, 2023).

A economia de Ponta Grossa também se beneficia de um dos mais importantes entroncamentos ferroviários do Sul do Brasil, passando pelo município toda a produção agrícola destinada ao Porto de Paranaguá, para exportação. Esse sistema ferroviário é constituído pela Rede Ferroviária Federal S/A, com suas linhas em direção ao Porto de Paranaguá, à Curitiba e ao sul do país. E pela Estrada de Ferro Central do Paraná, que faz ligação com o Norte do Estado, através de Apucarana, além da ligação ao Norte Velho, via Wenceslau Braz, com ramificação até Ourinhos (SP). Além disso, Ponta Grossa conta com o aeroporto Sant'Ana, localizado a 10 km do centro da cidade. Essa é uma estrutura homologada pelo Departamento

de Aeronáutica Civil e possui pista pavimentada em condições de receber aviões executivos de pequeno e médio porte, usados principalmente por empresários da região.

A indústria e o comércio, de Ponta Grossa, são importantes para toda a economia dos Campos Gerais. Sendo que a maior parte dessa indústria está concentrada no beneficiamento da soja e de cereais, bem como no processamento de alimentos, de bebidas, de fertilizantes, de insumos florestais (madeira, papel e celulose) e de metalurgia. Com destaque para as seguintes indústrias: Tetra Pak; Ambev; Heineken; Cargill Agrícola; Crown; Continental; Frísia; Arauco; LP do Brasil; e Yara Fertilizantes. Além disso, devido Ponta Grossa estar na região de maior densidade populacional do Estado do Paraná, com mais de 360 mil habitantes em 2021, por ter o maior PIB per capita da sua microrregião, acima de R\$ 48 mil (IBGE, 2023), é natural que o município registre uma atividade comercial relevante, com grandes centros comerciais e redes varejistas e atacadistas. Entre os comércios, pode-se destacar: Supermercados Tazetto; Supermercados Condor; Liquigás Distribuidora; Havan; AGL Ceregato Automóveis; Ultragás; Coca-Cola; e Grupo Pitangueiras.

No setor de serviços e turismo, o Município apresenta uma rica e diversificada rede de serviços, com destaque para sua ampla infraestrutura de transportes, grande rede hospitalar e importantes instituições de ensino técnico e superior, como a Universidade Estadual de Ponta Grossa e a Universidade Tecnológica Federal do Paraná. Além disso, pode-se destacar: Costa Teixeira Transportes; Vivo; Expresso Princesa dos Campos; Rumo; Viação Iapó; e Novamel Transporte.

O turismo também tem posição de destaque na economia ponta-grossense, com um diversificado conjunto de atrações naturais, históricas e culturais e espaços para a prática de esportes radicais e contato com o meio ambiente, além de realizar inúmeros eventos comerciais e de entretenimento ao longo do ano que atraem um público expressivo de toda a região. Entre as principais atrações turísticas de Ponta Grossa estão: Parque Estadual de Vila Velha; Parque Nacional dos Campos Gerais; Cachoeira da Mariquinha; Casa da Memória Paraná; Estação Arte; Estação Saudade; Museu Campos Gerais, Museu Época; Cine-Teatro Ópera; Teatro Marista; Mosteiro da Ressurreição; Igreja do Rosário; Capela de Santa Bárbara do Pitangui; e o Cemitério São José (PREFEITURA DE PONTA GROSSA, 2023).

2.1 Sala do Empreendedor de Ponta Grossa

A Sala do Empreendedor de Ponta Grossa começou em um espaço pequeno e sem grandes pretensões, mas, hoje, é referência no Paraná e no Brasil pelos projetos inovadores desenvolvidos no apoio aos novos negócios. Com mais de 250 mil atendimentos e 28 mil capacidades para microempresários da cidade, a Sala do Empreendedor de Ponta Grossa funciona desde 2011 e, a partir de 2015, o apoio ao empreendedorismo passou a se organizar como uma política pública municipal, com espaço próprio para atendimento ao empreendedor local.

Nos dias atuais, integra a Coordenadoria de Fomento ao Empreendedorismo e Inovação, da Secretaria da Fazenda, e também o ecossistema Vale dos Trilhos.

Nas palavras da prefeita de Ponta Grossa Elizabeth Schmidt:

É com muito orgulho que celebramos esta data, porque a Sala do Empreendedor já se destacou como referência no Paraná e no Brasil, com reconhecimentos e premiações pelo desempenho e projetos inovadores que desenvolveram neste período. São 10 anos de história de sucesso, com projetos que realmente fizeram a diferença na vida de muitas empresas, que encontraram na Sala o caminho e a orientação para formalização e capacitação. E para os próximos anos, vamos fortalecer e ampliar este importante trabalho (PREFEITURA DE PONTA GROSSA, 2021, n.p).

Para entender o quão grande é o trabalho desenvolvido pela Sala do Empreendedor de Ponta Grossa, é importante contar um pouco da sua história. No início, em 2011, a sua estrutura contava com um pequeno espaço na estrutura da Secretaria de Indústria, Comércio e Qualificação Profissional. Através da lei 12.182/2015, a Prefeitura estabeleceu as disposições sobre microempreendedor individual, instituindo uma política pública com objetivo de formalizar pequenos empreendimentos e a inclusão social e previdenciária, sendo vedado impor restrições ao MEI relativamente ao exercício da profissão ou participação em licitações, em função da sua respectiva natureza jurídica.

Dado o seu sucesso inicial, a demanda por serviços voltados ao microempreendedor individual cresceu de forma exponencial. Por isso, em 2017, a Sala do Empreendedor de Ponta Grossa passou a contar com um novo espaço, maior e exclusivo, na parte térrea do Paço Municipal David Federmann. Neste mesmo ano, foi instituído "Minha Empresa Legal", programa com o objetivo de incentivar a legalização e formalização de negócios informais e de novos empreendimentos por meio da prestação de serviços públicos gratuitos ou por valores subsidiados, destinados a constituição da pessoa jurídica e acompanhamento técnico-contábil, planejamento e assessoramento empresarial.

Segundo a presidente da Agência de Inovação de Desenvolvimento, Tonia Mansani:

Este é o foco da Sala do Empreendedor: incentivar a formalização dos negócios e dar aos pequenos empresários o suporte necessário para que ele fortaleça seu negócio e cresça. E foi com este propósito que ampliamos os serviços e projetos realizados pela nossa equipe, ouvindo quais eram as necessidades que os empresários estavam trazendo até nós. Por isso, além da orientação e formalização, também buscamos parcerias para cursos, oficinas e mentorias que possam agregar aos pequenos negócios, especialmente os MEIs, em que uma única pessoa cuida das vendas, do administrativo, do financeiro e do marketing. Queremos que estes negócios prosperem (PREFEITURA DE PONTA GROSSA, 2021, n.p).

Em 2019, o projeto foi ampliado, e a Sala do Empreendedor se tornou um dos braços da Coordenadoria de Fomento ao Empreendedorismo e Inovação, vinculada à Secretaria Municipal da Fazenda. Segundo Cláudio Grokoviski – Secretário da Fazenda :

Ao longo destes 10 anos de história, a Sala desenvolveu projetos através de parcerias com empresas e instituições locais, e é uma das iniciativas que fazem de Ponta Grossa um dos celeiros de inovação no Paraná e no Brasil. Além de todo o suporte prestado pela Sala do Empreendedor para a abertura de negócios e capacitação, as mentorias, acelerações e demais orientações vêm desempenhando um importante papel para que os pequenos negócios se desenvolvam, com a possibilidade de crescer. Estes negócios têm um importante impacto em nossa economia e a Sala do Empreendedor certamente é fundamental neste processo (PREFEITURA DE PONTA GROSSA, 2021, n.p).

Olhando para o futuro, a Sala do Empreendedor de Ponta Grossa apostou nos serviços digitais, criando a Sala Digital: uma plataforma com proposta de agregar serviços e acessos que facilitam o dia a dia do empresário, diminuindo a distância entre empreendedores e soluções como emissão de guias, tutoriais sobre serviços, informações sobre formalização, entre outros. Além deste serviço, a Coordenadoria de Fomento ao Empreendedorismo e Inovação também ampliou os pontos de atendimento através dos Centros de Referência de Assistência Social (CRAS)³ de Ponta Grossa, estando habilitados para realizar o atendimento integrado de serviços básicos da Sala do Empreendedor. Isto é relevante, pois descentraliza os atendimentos da Sala do Empreendedor. O resultado deste trabalho pode ser observado pelo número de formalizações, que passou de 3.715, em 2013, para cerca de 20 mil, em 2020 (PREFEITURA DE PONTA GROSSA, 2021).

Um programa da Sala do Empreendedor com destaque é o "Voe Aceleradora de Negócios", criado em 2017, este foi uma iniciativa inédita no Paraná, com orientações e foco no autoconhecimento, capacitação e inovação, auxiliando no desenvolvimento dos negócios de forma consistente e integrada, capacitando os participantes como gestores de forma gratuita. Também, em 2020, teve início o programa "Decola: A arte do começo", voltado para estruturar ideias de negócios para os potenciais novos empreendedores, isto é, para pessoas que querem empreender e não sabem por onde começar. Em outras palavras, este programa tem como propósito auxiliar os futuros empreendedores a partir da utilização de metodologias inovadoras.

É importante destacar que o programa "Voe Aceleradora de Negócios" foi reconhecido como o melhor programa na categoria econômica de grandes empresas, em 2019, na 12^a edição do prêmio Ozires Silva de Empreendedorismo Sustentável, premiação a nível nacional pro-

³ Além dos serviços básicos no atendimento diário, como orientações e encaminhamentos sobre abertura de empresas e formalização, os Cras também recebem serviços avançados, como formalização de novos empreendedores, com cronograma de atendimento itinerante.

movida pelo ISAE Escola de Negócios. A "Voe Aceleradora de Negócios" também garantiu para Ponta Grossa o Prêmio Gestor Público - PR 2020, na categoria Empreendedorismo.

Já a ferramenta "Sala Digital" conquistou o reconhecimento do Sebrae no X Prêmio Prefeito Empreendedor, na categoria "Políticas Públicas para o Desenvolvimento dos Pequenos Negócios", regional Sul. Esta premiação reconhece iniciativas de gestões públicas que tenham o estímulo ao desenvolvimento dos pequenos negócios como uma das prioridades, impactando no empreendedorismo local. Além disso, antes dessa premiação, Ponta Grossa já havia sido contemplada com o prêmio em nível estadual.

2.2 Ponta Grossa no ranking sobre ambiente de negócios

Em 2019, a Prefeitura de Ponta Grossa já havia recebido dois reconhecimentos estaduais colocando o Município em posição de destaque no Estado pelas políticas de incentivo ao empreendedorismo e ao desenvolvimento de micro e pequenas empresas. Durante o Encontro Estadual de Agentes de Crédito e de Desenvolvimento, promovido pelo Sebrae e pela Fomento Paraná, Ponta Grossa foi contemplada com o 1º lugar no ranking das cidades acima de 200 mil habitantes no Prêmio Fomento Paraná e Sebrae/PR de Microcrédito e também com o reconhecimento do Sebrae, na categoria Ouro por cinco anos consecutivos, de Agentes de Desenvolvimento do Programa Cidade Empreendedora (AREDE, 2023).

Mas o reconhecimento nacional veio em 2022. Ponta Grossa ganhou destaque nacional no Índice de Concorrência dos Municípios, ranking desenvolvido pelo Ministério da Economia, do Governo Federal, por meio da Secretaria de Acompanhamento Econômico. Entre as 119 cidades pesquisadas, o Município ficou na terceira pontuação geral, atrás somente de Porto Alegre (RS) e Belo Horizonte (MG). Além disso, Ponta Grossa ficou com o 1º lugar geral, com 92,39 pontos, no tema "Empreendendo no Município", que engloba abertura de empresas e questões econômicas. E em relação ao tema "Concorrência em Serviços Públicos", a cidade também ficou em 1º lugar na Região Sul.

Segundo a Prefeita Elizabeth Schmidt,

Esse relatório é uma importante ferramenta de gestão, demonstrando o ambiente empreendedor que Ponta Grossa tem. Isso mostra que estamos no caminho certo, pois a cidade se destacou a nível nacional em diversos âmbitos, como liberdade econômica, poder público e infraestrutura (AREDE, 2023, n.p).

Segundo a presidente da Agência de Inovação de Desenvolvimento, Tonia Mansani:

Ficar em primeiro lugar em âmbito nacional vem corroborar que o Vale dos Trilhos é um ambiente propício e fértil para negócios inovadores. A criação da agência de inovação trouxe essa importante agenda para o governo (AREDE, 2023, n.p).

Ponta Grossa também teve destaque na avaliação da infraestrutura, aparecendo na 7^a colocação nacional, com nota de 67,93, sendo a 2^a melhor do Paraná, atrás apenas de São José dos Pinhais. Sendo que os dois primeiros colocados foram Porto Alegre (76,46) e Florianópolis (74,04), e a média nacional foi 51,87. Além disso na categoria Construção, Ponta Grossa aparece entre as 40 melhores, com uma nota 40, também acima da média nacional (33,9). Em Qualidade da Regulação Urbanística, Ponta Grossa ficou em 3^o lugar geral, com nota 72,75, próximo de Piracicaba e Curitiba – 1^o e 2^o lugares. Com relação à Infraestrutura, o município alcançou o 7^o lugar geral. Já na categoria “Liberdade Econômica”, ficou em 4^o lugar na Região Sul, e, em 8^o lugar geral. Em “Contratando com o Poder Público” ao Município ficou com a 5^a colocação e a 7^a em “Tributação”.

3 REFERENCIAL TEÓRICO

Este capítulo desenvolve o referencial teórico relevante para esta dissertação. Primeiro, traz o desenvolvimento histórico do que é empreendedorismo, e na sequência, fala sobre a estratégia empírica usada, *Propensity score matching*.

3.1 Empreendedorismo

A palavra "empreendedor" é derivada do francês "*entreprendre*", que significa "empreender" ou "alcançar". Portanto, empreendedor é aquele que se compromete a organizar, gerenciar e assumir os riscos de um negócio. Uma vez que o mundo moderno está em constante transformação, a definição de empreendedor tem se tornado mais ampla. Nos dias atuais, um empreendedor é um inovador ou desenvolvedor que reconhece e aproveita oportunidades; transforma essas oportunidades em ideias viáveis ou comercializáveis; agrega valor, esforço, dinheiro ou habilidades ao longo do tempo; assume os riscos de mercado para implementar essas ideias e colher os frutos desses esforços (RONSTADT, 1984).

De uma forma geral, não existe uma definição formal de empreendedor, entretanto, pesquisas tem buscado cada vez mais identificar as características intrínsecas do empreendedorismo. Para se ter uma ideia, as forças sociais e econômicas da atividade empreendedora existem há muito tempo. Uma vez que, o reconhecimento dos empreendedores remonta à França no século 18, quando o economista Richard Cantillon associou as atividades de "risco" a algo empreendedor. Naquela época, ocorria a Revolução Industrial, e o empreendedor desempenhou um papel relevante na assunção de riscos e transformação de recursos. Desde então, o empreendedorismo tem atraído o interesse da pesquisa econômica (KIRZNER, 1983).

Durante muito tempo, prevaleceu a percepção de que as características de um empreendedor não poderiam ser ensinadas ou aprendidas. Contudo, atualmente, o empreendedorismo é visto como uma disciplina com modelos, processos e estudos de casos. Portanto, o desenvolvimento econômico tem o empreendedorismo como catalisador mais relevante. Pois, empreendedores são considerados "agentes da mudança", por fornecerem ideias criativas e inovadoras (KENT; SEXTON; VESPER, 1982).

Já para Schumpeter, empreendedorismo consiste em fazer coisas que não costumam ser feitas no curso normal na rotina dos negócios. Pois, trata-se, essencialmente, de um fenômeno que vem sob o aspecto mais amplo da liderança (SCHUMPETER, 2017). E de acordo com Cole (1959), empreendedorismo constitui uma ponte entre a sociedade como um todo. Depois de analisar a evolução do empreendedorismo e examinar suas diferentes definições, pode-se dizer que empreendedorismo é o processo dinâmico de criação de riqueza adicional criada por indivíduos que assumem os maiores riscos em termos de capital, tempo ou comprometimento com o desafio de criar valor com algum produto ou serviço (RONSTADT, 1984).

Normalmente, empreendedores são indivíduos que reconhecem oportunidades onde outros somente percebem caos. A literatura procura sintetizar as principais características de um empreendedor em: iniciativa pessoal; capacidade de mobilização de recursos; competências de gestão; desejo de autonomia; tomada de risco; ousadia; competitividade; comportamento orientado a um objetivo; confiança; senso de oportunidade; intuição; ações baseadas na realidade e na capacidade de aprender com os erros; e de empregar habilidades de relações humanas (KENT; SEXTON; VESPER, 1982).

Neste sentido, um amplo estudo utilizando dados de mais de 1400 proprietários de negócios em 11 países identificou características relacionadas aos atributos pessoais do empreendedor: aprovação; independência; desenvolvimento pessoal; segurança; e auto-realização (CARTER *et al.*, 2003; LEZANA; TONELLI, 2004; BIRLEY; WESTHEAD, 1992; WESTHEAD; WRIGHT, 2006). Posteriormente Shane, Kolvereid e Westhead (1991) reapplicaram esta pesquisa, apenas para Grã-Bretanha, Nova Zelândia e Noruega. Como resultado, obtiveram quatro características que os motivaram empreendedores a abrirem novos negócios: reconhecimento; independência; aprendizado; e reproduzir padrões. Já para Birley e Westhead (1992), trabalhando com uma amostra de 408 proprietários de negócios sediados na Grã-Bretanha, encontraram as seguintes características relacionadas com a motivação em abrir um negócio: aprovação pessoal; independência; desenvolvimento pessoal; autorealização; e segurança financeira.

Para ter sucesso com um empreendimento, o empreendedor reconhece que capital humano é extremamente importante, pois quanto maior o conhecimento maior a percepção de como funciona a atividade empresarial (UCBASARAN; WRIGHT; WESTHEAD, 2003). Segundo Lezana e Tonelli (2004), aspectos técnicos relacionados com o negócio, a experiência na área comercial, a escolaridade, a experiência em empresas, a formação complementar e a vivência em situações novas são essenciais para um empreendedor.

Muitas vezes, o aprendizado tem origem na experiência da atividade empresarial, pois verdadeiros empreendedores aprendem mesmo quando fracassam (MINNITI; BYGRAVE, 2001). Foster e Lin (2003) estudaram as diferenças individuais na aprendizagem empreendedora, e concluíram que conhecimentos obtidos anteriormente e não ligados a estudos empresariais, fazem diferença no sucesso do negócio. Assim, Kolb (1984) classifica quatro formas da experiência apoiar o sucesso de um empreendedor: experimentando; observação reflexiva; conceituação e abstração; e através das suas ações.

A "sobrevivência" não é algo simples para um empreendimento, pois são inúmeras as habilidades necessárias para o bom funcionamento de uma empresa (Lezana; Tonelli, 2004). Assim, há um papel preponderante na formação das competências de qualquer indivíduo, sintetizados como a capacidade maximizar os resultados a partir de recursos escassos (GUISSO; SCHIVARDI, 2005). Tonelli (2003) destaca quais seriam essas habilidades: autodisciplina; objetividade; flexibilidade; criatividade; e comunicação.

Outra característica importante de um empreendedor é a identificação de oportunidades (ALSOS; KAIKKONEN, 2004; ARDICHVILI; CARDOZO; RAY, 2003; ECKHARDT; SHANE,

2003; GAGLIO; KATZ, 2001; PARK, 2005; SAKS; GAGLIO, 2002; SIPILÄ, 2006; STUART; SORENSON, 2005). Segundo Sarasvathy *et al.* (2002), pode-se definir oportunidade empreendedora como a existência de uma circunstância favorável e adequada para se poder agir criando algo que, de forma positiva, tenha valor econômico. Ou seja, novas ideias ou invenções que podem levar ao alcance de um ou mais objetivos econômicos, possíveis através dessas ideias ou invenções. Além disso, crenças acerca de coisas favoráveis ao alcance desses objetivos e ações que implementem esses objetivos também caracterizam a identificação de uma oportunidade.

3.2 Propensity Score Matching

O método *Propensity Score Matching* foi desenvolvido por Rosenbaum e Rubin (1985) para medir a probabilidade condicional de empregar um tratamento. Assim, este é um método estatístico que busca estimar o efeito de um tratamento usando a covariadas que predizem quem pode receber o tratamento.

Dessa forma, é definido um vetor de variáveis observáveis para identificar os indivíduos que apresentam características que devem ser tratadas. A teoria do *propensity score* é suficiente para remover um possível viés de seleção. Rosenbaum e Rubin (1985) definiram o método de amostragem por meio da combinação *propensity score*. O *Propensity Score Matching* pode ser entendido como uma generalização da correspondência discriminante, sendo ajustado pela subclassificação no *propensity score*, assim subclasses iguais são usadas para estimar os efeitos do benefício para todas as variáveis das subpopulações.

Conforme Khandker *et al* (2010), o *Propensity Score Matching* desenvolve um grupo de controle para confrontar com o grupo tratado. O primeiro grupo representa aqueles indivíduos que poderiam participar do tratamento, pois apresentam características observáveis semelhantes as do grupo tratado, porém não receberam o tratamento. Os indivíduos que não apresentarem base de comparação, por não terem nenhuma correspondência são eliminados da amostra.

Apesar de eliminar indivíduos sem pares, reduzindo a quantidade de elementos da amostra, o *Propensity Score Matching* não deixa de ter validade, pois é fundamentado em dois conceitos: independência condicional; e suporte comum. Para Ribeiro e Cacciamali (2012), a hipótese de independência condicional garante a independência das variáveis de pré-tratamento, isto é, as variáveis observáveis não dependem do tratamento. Assim, o grupo de tratamento e os impactos potenciais do tratamento são independentes e os fatores observáveis não comprometem a participação.

A hipótese de suporte comum tem importância no sentido de definir que os indivíduos do grupo que receberam o tratamento possam ter correspondentes no grupo de controle, desta forma para cada probabilidade estimada para os integrantes do grupo que recebeu o tratamento haverá uma probabilidade semelhante no grupo de controle, isto representa um suporte comum entre beneficiários e não beneficiários do tratamento (RIBEIRO; CACCIAMALI, 2012).

Segundo Khandker *et al* (2010), o *Propensity Score Matching* é semelhante aos métodos de correspondência, pois busca desenvolver um grupo contrafactual o mais parecido possível com o do grupo de tratamento em termos de características observáveis. Assim, cada participante do grupo de tratamento é combinado com um participante do grupo de controle – dadas as características semelhantes –, em seguida, é tomada a diferença média dos resultados dos dois grupos para obter o efeito do tratamento.

Neste sentido, o efeito médio do tratamento é então calculado como a média da diferença dos resultados destes dois grupos. Contudo, durante a composição do grupo de controle, o viés de seleção pode ser identificado, visto que características observáveis e não-observáveis no grupo de tratamento podem ser diferentes. Uma forma de eliminar esse problema de viés consiste na correta identificação de grupos semelhantes, porém encontrar indivíduos exatamente semelhantes em dois grupos considerando características parecidas requer certo rigor (KHANDKER *et al.*, 2010).

4 METODOLOGIA

A metodologia de pesquisa desta dissertação é usar uma estratégia empírica para eliminar o viés de seleção da amostra, e assim, avaliar se o empreendedorismo é estado-independente. Para tanto, este capítulo apresenta como a base de dados foi construída e método de análise utilizado. Inicialmente, será apresentado em exemplo simples do problema do viés de seleção e como resolvê-lo. Então, apresenta-se o *propensity score matching* – ferramenta para resolver o viés de seleção –, o modelo logit – usado para construir o *propensity score* – e método de mínimos quadrados ordinários – usado para estimar a regressão linear para o cálculo do ATT.

4.1 Estratégia empírica

As evidências empíricas deste trabalho serão geradas a partir dos dados sobre 119 microempresas⁴ abertas em 2019 (94 microempresas) e em 2020 (25 microempresas) da base de dados da Sala do Empreendedor de Ponta Grossa⁵. Além das variáveis desta base, foi construído um "Índice de Empreendedorismo", por meio de um questionário, com objetivo de capturar a vocação empreendedora do microempresário.

Desde as contribuições de McClelland (1961), sobre o papel da necessidade de realização como motivação no comportamento dos empreendedores, e de Miner (1997), sobre as características de sucesso de um empreendedor, a literatura tem buscado reconhecer ações que exercem determinados traços pessoais sobre o desejo de empreender. Neste sentido, Hornaday (1982) encontra quarenta e duas características que com frequência são listadas como próprias de indivíduos empreendedores⁶. Deste total, dezenove características são listadas como as mais frequentes: auto-confiança; perseverança, determinação; energia, diligência; de-

⁴ A base de dados inicialmente fornecida pela Sala do Empreendedor de Ponta Grossa apresentava 2397 cadastros (572 abertas, em 2020, e 1825 abertas, em 2019), deste total, apenas 119 microempresas responderam o questionário ou estavam com o cadastro completo e correto.

⁵ As Salas do Empreendedor são locais de atendimento das prefeituras municipais que facilitam os processos de abertura de empresas, regularização e baixa; bem como serviços exclusivos aos microempreendedores individuais (MEI).

⁶ A saber: confiança; perseverança, determinação; energia, diligência; desembaraço; habilidade para assumir riscos calculados; dinamismo, liderança; otimismo; necessidade de realização; versatilidade, conhecimento do produto, mercado, maquinaria e tecnologia; criatividade; habilidade para influenciar outros; habilidade para conviver com outras pessoas; iniciativa; flexibilidade; inteligência; orientado para aclarar metas; hábil gestão do tempo, eficiência; habilidade para tomar decisões rapidamente; capacidade de responder positivamente a desafios; independência; honestidade, integridade; maturidade, equilíbrio; sensibilidade a sugestões e críticas; responsabilidade; previsão; exatidão, precisão; cooperatividade; orientação para o lucro; habilidade de aprender com seus erros; senso de poder; personalidade agradável; egotismo; coragem; imaginação, perceptividade; tolerância com a ambiguidade; agressividade; capacidade para divertir-se; eficácia; confiança; habilidade para demonstrar confiança nos subordinados; e sensível aos outros.

sembaraço; habilidade para assumir riscos calculados; necessidade de realização; criatividade; iniciativa; flexibilidade; capacidade de responder positivamente a desafios; independência; previsão; dinamismo, liderança; versatilidade, conhecimento do produto, mercado, maquinaria e tecnologia; habilidade para conviver com outras pessoas; sensibilidade a sugestões e críticas; orientação para o lucro; perceptividade; e otimismo. Outra característica envolve a idade dos empreendedores, por exemplo, nos Estados Unidos, o grupo de empreendedores com idade entre 55 e 64 anos é predominante, pois são mais confiantes em suas habilidades para iniciar negócios do que o grupo com idade entre 18 e 44 anos (KELLEY; SINGER; HERRINGTON, 2016).

Dadas as características de um empreendedor apresentadas no referencial teórico e nos comentários anteriores, a base de dados da Sala do Empreendedor de Ponta Grossa contribui com algumas variáveis, mas também é necessário ter uma variável de "Índice de Empreendedorismo". Para tanto, foi aplicado um questionário simples, visando a facilidade em responder as questões, mas suficiente para obter as principais características de um empreendedor propostas pela literatura. Assim, o questionário aplicado foi:

1. Você sempre sonhou em ter um negócio próprio?
2. Algum dos seus pais possui negócio próprio?
3. Você estava empregado quando abriu o seu negócio próprio?
4. Se o seu negócio não desse certo, você tentaria um outro negócio próprio?

Essas perguntas buscam identificar algumas características básicas de um empreendedor. A primeira procura reconhecer se o indivíduo pensava em empreender, independentemente, o momento econômico do país. Além disso, esta pergunta está relacionada com a necessidade de realização pessoal do empreendedor (McCLELLAND, 1961; CARTER *et al.*, 2003; LEZANA; TONELLI, 2004; BIRLEY; WESTHEAD, 1992; WESTHEAD; WRIGHT, 2006). Na segunda, o objetivo é detectar se abrir um negócio próprio teria origem nos pais. A terceira pergunta seria a mais importante para identificar a necessidade de abrir um negócio, pois a intuição é que um "verdadeiro" empreendedor não espera perder o emprego para abrir o próprio negócio (RONSTADT, 1984). Por fim, na quarta pergunta, o objetivo é identificar a persistência e a consciência de risco do indivíduo, uma vez que, verdadeiros empreendedores aprendem mesmo quando fracassam (MINNITI; BYGRAVE, 2001).

Resumindo a ideia por trás do valor do "Índice de Empreendedorismo", cada resposta do questionário contribui com valor 1 (um), se a resposta é "SIM", e com 0 (zero), se a resposta é "NÃO", ou seja,

$$0 \leq \text{Índice de Empreendedorismo} \leq 4.$$

Em palavras, se o microempresário responde "SIM" nas quatro perguntas do questionário, o seu índice recebe o valor de 4, se responde "SIM" em três perguntas, o valor do índice é 3, e

assim por diante. Isto é, este índice apresenta valor máximo 4 (quatro) e um valor mínimo 0 (zero).

E as variáveis usadas no modelo econométrico estão apresentadas no quadro 1.

Quadro 1 – Variáveis usadas no modelo econométrico.

Variável	Descrição
Variável de tratamento	Variável binária que assume valor igual a 1 se a MEI foi aberta em 2020 e 0 se foi aberta em 2019
Idade	Idade do microempresário em 2022
Escolaridade	Escolaridade do microempresário
Índice de Empreendedorismo	Construído por meio do questionário aplicado aos microempresários (min=0 e max = 4, ou seja, cada resposta tem peso 1)

Fonte: Autoria própria (2022).

4.2 Método de análise de dados

Nesta seção, apresenta-se o método de análise de dados desta dissertação, ou seja, qual foi a estratégia empírica para verificar se o empreendedorismo é estado-independente.

4.2.1 Um exemplo "ingênuo" do problema do viés de solução e como resolvê-lo

Quando a distribuição aleatória não é possível, pode-se buscar um quase-experimento. Esta estratégia empírica representa contornos de pesquisas que não têm distribuição aleatória nos grupos de tratamento e de controle. Como a comparação entre as condições de tratamento e não-tratamento deve sempre ser feita com grupos não equivalentes, se não é possível distribuir aleatoriamente pessoas (ou grupos) pelas condições experimentais, perde-se a capacidade de controlar o que ocorre com esses elementos. Entretanto, ainda é possível observar o que ocorre, quando ocorre e a quem ocorre; e, ao decidir o que e quando medir, poderá planejar um quase-experimento.

Se por um lado, perde-se o controle por não usar uma distribuição aleatória, por outro lado, poderá realizar uma pesquisa e analisar relações de causa-efeito sem um experimento. Ao coletar dados criteriosamente em situações, pode-se criar um quase-experimento. Em outras palavras, em um experimento, é necessário analisar o efeito de um indivíduo receber o tratamento e não receber o tratamento. Mas, em Ciências Sociais Aplicadas, isto é praticamente impossível, pois se um indivíduo recebe um tratamento, por exclusão ele não pode estar no

grupo que não recebeu o tratamento (grupo de controle). Alguém poderia dizer que é possível medir o efeito em um indivíduo antes do tratamento, aplicar o tratamento, e medir novamente o mesmo indivíduo. Contudo, isso não pode ser feito ao mesmo tempo, e condições externas podem afetar o indivíduo antes e depois do tratamento. Assim, se o momento temporal não é o mesmo, um indivíduo não é o mesmo nos dois momentos.

Para usar um exemplo simples (ingênuo) e explicar a estratégia empírica deste trabalho, imagine uma pessoa que possui dois caminhos a seguir: abrir a firma em 2019; ou abrir a firma em 2020⁷. Esta situação pode ser vista na figura 2, em que, uma pessoa tem que decidir qual caminho seguir, ou seja, se abre a firma em 2019, ou em 2020.

Figura 2 – Pessoa tendo dois caminhos a seguir.



Fonte: Autoria própria (2022) baseado em Angrist [s.d.].

Dessa forma, ao decidir abrir a firma em 2020, evidentemente, a firma não poderá ser aberta em 2019 (figura 3). Alguém poderia dizer, mas a firma poderia ser aberta em 2019, fechar ainda no mesmo ano, e ser aberta novamente em 2020. Contudo, mesmo que isto seja possível, juridicamente, as condições de abertura, no tempo, serão diferentes, e assim, as firmas serão consideradas diferentes.

Nesta situação, não é possível realizar um experimento, como é feito em laboratórios de física, química etc. É necessário usar um quasi-experimento, ou seja, usar estratégias empíricas para deixar os grupos balanceados. Para tanto, por meio de covariadas – variáveis que moderam o impacto das variáveis independentes na variável dependente – é possível "construir" um contrafactual clonado do indivíduo (figura 4).

A figura 5 apresenta duas pessoas diferentes: Renata; e Cristina. Evidentemente, são pessoas diferentes. Além disso, a Renata pertence ao grupo de tratamento, pois abriu a sua firma em 2020, e a Cristina pertence ao grupo de controle, visto que abriu a sua firma em 2019. Contudo, por meio das covariadas – idade e escolaridade – pode-se assumir que a Cristina

⁷ Neste trabalho, as firmas abertas em 2020 representam o grupo de tratamento, enquanto, as firmas abertas em 2019 representam o grupo de controle.

Figura 3 – Pessoa escolhendo um dos dois caminhos.



Fonte: Autoria própria (2022) baseado em Angrist [s.d.].

Figura 4 – Construção do contrafactual clonado da pessoa.

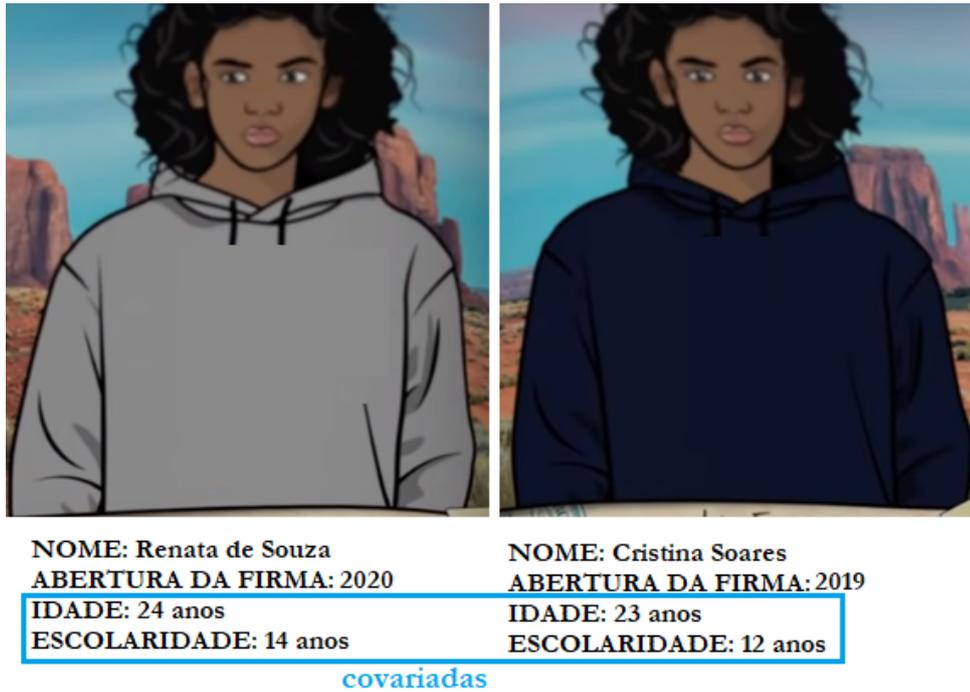


Fonte: Autoria própria (2022) baseado em Angrist [s.d.].

é o clone contrafactual da Renata, dado que, ambas possuem a mesma faixa de idade e o mesmo nível de escolaridade. Resumindo, apesar da impossibilidade de usar um experimento sobre o nível de empreendedorismo entre os grupos de tratamento e controle, é possível usar um quasi-experimento, isto é, comparar o clone contrafactual (Cristina) com o indivíduo base (Renata).

Uma vez entendido como usar um quasi-experimento para contornar a impossibilidade do uso de um experimento em Ciências Sociais Aplicadas. O próximo passo é entender a diferença entre amostras balanceadas e não-balanceadas, e entre o efeito médio do tratamento (ATE) e o efeito médio do tratamento no grupo de tratamento (ATT). Em poucas palavras, o objetivo de um trabalho empírico com grupos de controle e tratamento é calcular o efeito médio do tratamento (ATE), mas isso só é possível em experimentos. Então, em trabalhos em Ciên-

Figura 5 – Contrafactual clonado da pessoa e as covariadas.

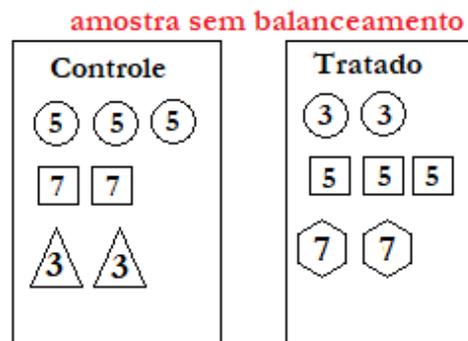


Fonte: Autoria própria (2022) baseado em Angrist [s.d.].

cias Sociais Aplicadas, sempre calcula-se o efeito médio do tratamento no grupo de tratamento (ATT). Mas, como será visto, usando covariadas, é possível ter os grupos de tratamento e controle balanceados, e assim, assume-se que o ATT possui o mesmo valor do ATE⁸.

A figura 6 apresenta um exemplo de amostra não balanceada, ou seja, os grupos de controle e tratado não são equivalentes. Nota-se que o grupo de controle é formado por três círculos, dois quadrados e dois triângulos, com um total de sete elementos. Já o grupo tratado, apesar de ter o mesmo número de elementos (sete), possui dois círculos, três quadrados e dois hexágonos. Ou seja, este último grupo, em relação ao primeiro, possui um círculo a menos, um quadrado a mais, dois hexágonos a mais e dois triângulos a menos. Em outras palavras, dadas as formas geométricas dos dois grupos, não é possível dizer que os grupos estão balanceados.

Figura 6 – Exemplo de amostra sem balanceamento.

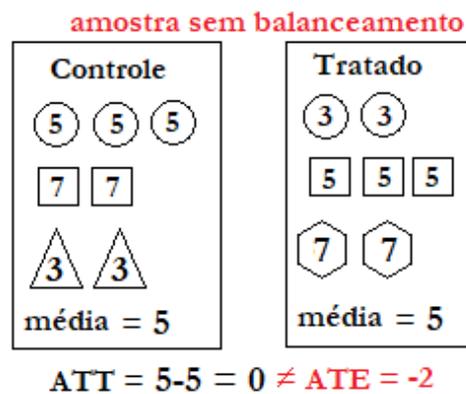


Fonte: Autoria própria (2022).

⁸ Na próxima seção, será apresentada uma explicação formal para o $ATT \approx ATE$.

Mas qual é o problema da amostra não ser balanceada? Suponha que o valor dentro de cada figura geométrica seja o valor de interesse de estudo do efeito do tratamento. Por exemplo, se pegar apenas um círculo de cada grupo, e compará-los, percebe-se que o efeito do tratamento é $3-5=-2$. Da mesma forma, se pegar apenas um quadrado de cada grupo, e compará-los, percebe-se que o efeito do tratamento é $3-5=-2$. Em outras palavras, o efeito médio do tratamento (ATE) é -2 . Contudo, usando todos os elementos dos grupos para este cálculo, chega-se em: $\frac{2 * 3 + 3 * 5 + 2 * 7}{7} - \frac{3 * 5 + 2 * 7 + 2 * 3}{7} = 5 - 5 = 0$. Em outras palavras, dada a amostra sem balanceamento, o efeito médio do tratamento no grupo tratado (ATT) é zero, mas, como foi visto, o efeito médio do tratamento é -2 , ou seja $ATT \neq ATE$ (figura 6). A conclusão é que se a amostra não for balanceada, o resultado do quasi-experimento não será válido.

Figura 7 – Cálculo do ATT e do ATE na amostra sem balanceamento.

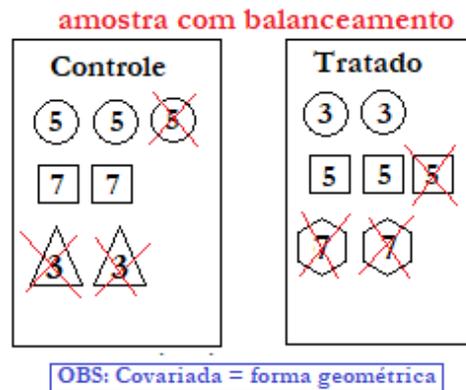


Fonte: Autoria própria (2022).

Para deixar os grupos balanceados é preciso eliminar os elementos em excesso em cada grupo. Ou seja, os elementos somente ficam na amostra se encontrar um par no outro grupo. Por exemplo, se os dois círculos do grupo tratado procurar um par no grupo de controle, encontrarão, contudo, o círculo residual no grupo de controle deve ser retirado da amostra. Já dos três quadrados do grupo tratado, apenas dois encontram pares no grupo de controle, assim, o quadrado residual do grupo tratado deve ser retirado da amostra. Como triângulos aparecem apenas no grupo de controle e hexágonos aparecem apenas no grupo tratado, esses dois tipos de elementos devem ser retirados da amostra. Assim, restaram apenas quatro elementos em cada grupo (dois círculos e dois quadrados) (figura 8). Resumindo, a covariada neste exemplo é a forma geométrica – usada para deixar a amostra balanceada.

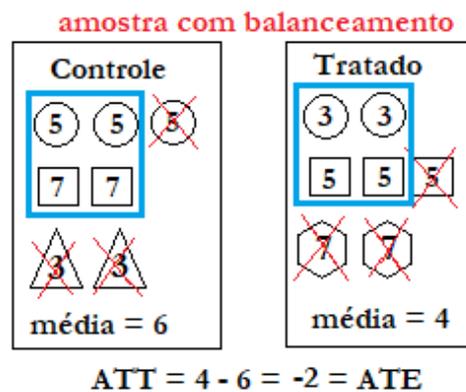
De posse da amostra balanceada, é possível realizar o quasi-experimento. Assim, o cálculo do ATT será: $\frac{2 * 3 + 2 * 5}{4} - \frac{2 * 5 + 2 * 7}{4} = 4 - 6 = -2$. Portanto, o valor do ATT é o mesmo do ATE. Resumindo, apesar de não ser possível realizar experimentos em Ciências Sociais Aplicadas, balanceando a amostra, é possível realizar um quasi-experimento e assim calcular o ATT e assumir que $ATT \approx ATE$.

Figura 8 – Balanceamento da amostra.



Fonte: Autoria própria (2022).

Figura 9 – Cálculo do ATT e do ATE na amostra com balanceamento.



Fonte: Autoria própria (2022).

4.2.2 Propensity Score Matching

Sem dúvida alguma o método dos mínimos quadrados ordinários (MQO) é uma das ferramentas mais clássicas da econometria. Contudo, ferramentas de pareamento estatístico tal como o *Propensity Score Matching* têm sido amplamente utilizadas em pesquisas na área de avaliação de políticas públicas ou na mensuração do impacto de algum tratamento. O *Propensity Score Matching* é utilizado quando a hipótese de distribuição aleatória do tratamento não é verificada. Neste caso, ao usar o MQO, o coeficiente estimado não será o verdadeiro valor esperado, pois carregará no valor calculado a correlação de não ser uma distribuição aleatória. Assim, no MQO,

$$Y = \beta X + \delta \text{trat} + \mu$$

O valor esperado para δ será:

$$\mathbb{E}(\delta) = \frac{\sum_{i=0}^n (\text{trat}_i - \bar{\text{trat}}_i) y_i}{\sum_{i=0}^n (\text{trat}_i - \bar{\text{trat}}_i)^2}$$

onde $trat$ é a variável de tratamento, X é uma matriz com n variáveis e i indivíduos e β é um vetor com as inclinações dos parâmetros estimados. Neste caso, quando o objetivo da análise é apenas calcular o resultado de um tratamento, a condição de primeira ordem traz δ como o parâmetro de interesse:

$$\frac{\partial Y}{\partial trat} = \delta$$

Em outras palavras,

$$\text{Efeito Médio} = \mathbb{E}(Y|trat = 0) - \mathbb{E}(Y|trat = 1) = \delta$$

Assim, para obter esse resultado, o tratamento não pode ser correlacionado com a ação de receber ou não o tratamento. A suposição de distribuição aleatória do tratamento implica que o valor esperado de ambos os grupos seria o mesmo se recebesse ou não o tratamento. Ou seja,

$$\mathbb{E}(Y|trat = 0) - \mathbb{E}(Y|trat = 1) = 0$$

Ao usar o *Propensity Score Matching* é importante refletir se seria razoável supor que os grupos de tratamento e de controle possuem a mesma qualidade esperada. Assim, ao comparar os dois grupos implica que, sem o tratamento, o resultado esperado de ambos deve ser o mesmo. Em outras palavras, firmas que abriram em períodos de não-criSES e firmas que abriram em crises devem ser semelhantes no desenvolvimento da pesquisa. Por exemplo, se na base de dados há apenas duas informações: momento de abertura da firma; e se ao abrir a firma, o indivíduo estaria desempregado. O *Propensity Score Matching* separa em grupos as firmas em que o indivíduo estaria desempregado na sua abertura de um lado, e do outro, as firmas em que o indivíduo não estaria desempregado na sua abertura.

Angrist and Pischke (2009) apontam que a estimativa pelo *Propensity Score Matching* ocorre em duas etapas: primeiro, estima-se com algum modelo paramétrico, como Probit ou Logit. Então, a estimativa do efeito do tratamento passa a ser calculada combinando o valor encontrado na primeira etapa ou usando algum esquema de peso, compara os efeitos nos grupos pareados, diferindo apenas pela variável de controle definida. Assim, o pareamento por *Propensity Score Matching* é uma ferramenta de balanceamento estatístico que tenta estimar o efeito médio do tratamento (Average Treatment Effect, ATE), assumindo um provável viés de seleção. Sendo que a probabilidade condicional de receber tratamento é calculada para tentar "imitar" uma distribuição aleatória do tratamento, ou seja, produzir um experimento análogo à randomização, pareando indivíduos semelhantes com base nas características observadas, diferenciando-se apenas em relação ao tratamento.

Enquanto a regressão por mínimos quadrados ordinários requer que o tratamento seja distribuído de forma randomizada, o *Propensity Score Matching* supõe apenas uma hipótese de independência condicional. Esta suposição permite contornar o viés de seleção. Portanto, para observações com os mesmos *propensity score*, a distribuição de covariadas deve ser a mesma

nos dois grupos. Dependendo do *propensity score*, cada indivíduo tem a mesma probabilidade de participar do tratamento, simulando assim a distribuição aleatória do tratamento. Ou seja,

$$\text{tratamento} = E \left[\frac{(t - Pr(x))y}{\delta(t - Pr(x))} \right]$$

$$\text{tratamento} \tilde{=} = \frac{\sum_{i=0}^n (t_i - \tilde{p}(x))y_i}{\sum_{i=0}^n [\tilde{\delta}(t_i - \tilde{p}(x))]}$$

onde $\delta = Pr(t = 1)$ e:

$$\tilde{\delta} = \frac{1}{n} \sum_{i=0}^n t_i$$

Dessa forma, qualquer diferença no resultado entre os indivíduos tratados e não tratados não é atribuível ao viés de seleção. Em seguida, calcula-se a diferença média dos resultados nos dois grupos para obter o verdadeiro efeito esperado do tratamento do programa.

Quando o ganho potencial da técnica do *Propensity Score Matching* está relacionado a uma característica intrínseca e não observada, a técnica de pareamento não resolve o problema do viés de seleção e não torna a modelagem mais próxima do verdadeiro valor esperado. Assim, o método de pareamento pelo *Propensity Score Matching* é limitado às variáveis presentes no banco de dados.

Em outras palavras, a ideia é comparar observações tratadas e não tratadas que possuem as mesmas características (x), ou seja, que tenham as características balanceadas. No entanto, há um problema de dimensionalidade em fazer essa comparação. Por exemplo, suponha que o objetivo seja comparar indivíduos tendo 20 variáveis binárias que os caracterizam. Portanto, há $2^{20} = 1.048.576$ possibilidade de grupos possíveis com o mesmo X . Para eliminar o problema de dimensionalidade das variáveis explicativas ao compará-las, usa-se o *Propensity Score Matching*. O *propensity score* (PS) é estimador da probabilidade de seleção ao tratamento condicional a um vetor de covariadas, $P(w = 1|x)$, estimado usando modelos Probit ou Logit. Note que o $p(x)$ resume todas as informações do vetor x de cada observação (vetor de covariadas pertencente ao \mathbb{R}^K e o PS, ou $p(X)$, pertence a \mathbb{R}).

A hipótese de identificação deste método apoia-se no fato de o mecanismo de seleção para o tratamento ser determinado por variáveis observáveis. Em outras palavras, não há nada de sistemático que faça com que um indivíduo seja tratado ou não, após condicionar a x , sobre apenas a aleatoriedade. Rosenbaum (2002) mostrou que indivíduos com o mesmo $p(x)$ possuem a mesma distribuição em x (ou seja, são pares homogêneos). Assim, se a seleção para o tratamento e as covariadas são condicionalmente independentes da do $p(x)$, pode-se escrever a diferença esperada entre resultados observados do tratamento e do controle da seguinte forma

$$ATE[p(x)] = E[E(Y(1)|p(x), w = 1)] - E[E(Y(0)|p(x), w = 0)]$$

$$\begin{aligned}
 &= E[Y(1) - Y(0)|p(x)] \\
 &= ATT[p(x)]
 \end{aligned}$$

Resumidamente, a análise do *Propensity Score Matching* pode ser resumida em quatro etapas:

Etapa I: Testes de balanceamento

Antes do tratamento, deve-se fazer o teste bivariado das covariadas usando os grupos de tratamento e de controle. Se o teste rejeita H_0 , i.e., se há diferença entre as covariadas dos grupos é preciso incluí-las no PS. Após o pareamento (*matching*), também realiza-se o teste bivariado entre os pares. Se não rejeita-se H_0 , se não há diferenças entre os grupos, o *propensity score* (PS) explica as diferenças observadas. Caso contrário, deve-se reestruturar o PS.

Para testar as diferenças entre os grupos, pode-se usar uma variedade de testes de duas populações (bivariadas). Os mais utilizados são os testes de diferença de médias, ou variâncias, o teste Wilcoxon-Mann-Whitney Rank-Sum (que compara a soma do rank das variáveis de grupo de controle e de tratamento em uma sequência única) e o teste de Kolmogorov-Smirnov (este teste compara as distribuições de amostras aleatórias independentes).

Etapa II: Estimando o *Propensity Score*

Para estimar o *Propensity Score*, escreve-se a probabilidade condicional de receber tratamento:

$$E(w_i|x_i) = P[w_i = 1|x_i] = f(x_i\beta)$$

Como conhece-se a natureza da distribuição da variável dependente (tem uma distribuição de Bernoulli), pode-se estimar os β 's por Máxima Verossimilhança. É importante sempre fazer o diagnóstico usual (verificar multicolinearidade, ajuste do modelo, análise de sensibilidade), mas isso não é suficiente para responder se $\hat{p}(x)$ representa o verdadeiro *propensity score*. Um bom modelo é condição necessária, mas insuficiente, para um bom PS. O melhor PS é aquele cujas variáveis incluídas tem respaldo teórico e que propõe um bom balanceamento entre os dois grupos nas covariadas⁹

Etapa III: Pareamento

Uma vez estimado o $p(X)$, pareiam-se os grupos de tratamento e controle (*matching* das uni-

⁹ A especificação correta é crucial para estimar o PS e para identificar ATE e ATT. A escolha das variáveis é fundamental porque pode mudar a análise. Parte da literatura foca na relevância teórica das variáveis. Assim, é muito importante especificar o modelo com justificativas para as covariadas, pois os resultados são sensíveis à especificação (SMITH; TODD, 2005).

dades para obtenção do contrafactual¹⁰). Pode-se usar vários critérios para parear as observações:

- **Menor distância:** Ordena-se aleatoriamente os participantes e calcula-se a distância¹¹ entre o primeiro tratado e todos os controles. A unidade j com a menor distância de i é escolhida como par de i . Este procedimento é repetido até que pares para todos os i sejam encontrados.
- **Vizinho mais próximo:** Sejam p_i e p_j os PS de I_1 e I_0 os conjuntos das unidades tratadas e não tratadas, respectivamente. Uma vizinhança contém um participante não tratado $j \in I_0$ como par de $i \in I_1$ se a diferença absoluta dos PS é a menor entre todos os possíveis pareamentos de i com unidades não tratadas. Uma vez encontrado o par de i , este é retirado de I_0 (*matching* sem reposição). Para cada i há um j (1-1 *matching*).
- **Matching com threshold:** O método de vizinho mais próximo pode gerar vizinhos distantes, pois não há limite para a distância entre os elementos a serem pareados. Assim, pode-se definir uma tolerância (*threshold*) $\epsilon > 0$ para calibrar o modelo tal que j é para de i somente se $|P_i - P_j| < \epsilon$ ¹².
- **Vizinho mais próximo com threshold e distância Mahalanobis:** Considere a combinação das técnicas anteriores:
 1. Ordena-se aleatoriamente os tratados e seleciona-se um primeiro tratado i .
 2. Verifica-se os indivíduos não tratados dentro do *threshold*.
 3. Calcula-se as distâncias Mahalanobis entre esses participantes e o tratado com base em um número (pequeno) de covariadas.
 4. Escolhe-se j de modo a minimizar d_{ij}
 5. As unidades pareadas são removidas da base.
 6. Retoma-se a análise até que todas as unidades tratadas tenham encontrado seus pares.

Etapa IV: Análise pós-pareamento

Logo após o pareamento é importante checar se há balanceamento das variáveis x na amostra pareada. Para tal, aplica-se os testes de comparação de populações sobre a amostra pareada, sendo uma amostra a dos indivíduos tratados e a outra dos indivíduos não tratados que foram pareados (teste de média, teste Wilcoxon-Mann-Whitney, ou teste Kolmogorov-Smirnov).

¹⁰ Casos que possuem probabilidade similar de serem selecionadas para o tratamento.

¹¹ Distância euclidiana, ou a métrica de Mahalanobis, por exemplo.

¹² Rosenbaum e Rubin (1985) sugerem $\epsilon \leq 0,25\hat{\sigma}_p$, sendo que $\hat{\sigma}_p$ denota o erro-padrão de \hat{p} .

Verificado o balanceamento do pareamento, pode-se estimar o ATE e ATT. Foi visto que o ATE será a média ponderada da diferença das médias entre tratados e controle apenas sobre a amostra de tratados, ou seja, usa-se apenas a amostra pareada para fazer esse cálculo. Especificamente, pode-se usar a seguinte regressão:

$$y_t = \alpha + \beta x_t + \varepsilon_t \quad (1)$$

onde α é o intercepto, β é a inclinação e ε é o termo de erro. Ou seja, o valor obtido por β será o ATT estimado. Em outras palavras, o valor do β representa o efeito médio do tratamento sobre o grupo tratado.

Mesmo após o pareamento, alguns valores de covariadas podem ser observados apenas em um dos grupos. Isso é comum quando há poucos tratados e muitos controles ou vice-versa. Nestes casos, pode-se selecionar uma subamostra de controles potenciais (ou tratamento potenciais) com base na informação pré-tratamento, isto é, busca-se uma subamostra que tenha maior sobreposição entre os grupos (excluindo unidades por *trimming*)¹³.

4.2.3 Modelo Logit

Em geral, os modelos econométricos utilizam variáveis dependentes contínuas. No entanto, muitas vezes, as escolhas que indivíduos fazem não podem ser medidas por uma variável resultado contínua. Para tanto, é preciso ter ferramentas que possam ter a variável dependente limitada, ou seja, seu intervalo de valores é restrito.

O modelo logit tem o objetivo de produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável binária, a partir de uma série de variáveis explicativas contínuas ou binárias. Em outras palavras, em comparação com as técnicas comuns de regressão, em especial a regressão linear, o modelo logit distingue-se essencialmente pelo fato da variável resposta ser binária.

O modelo logit analisa dados distribuídos binomialmente da forma

$$Y_i \sim B(p_i, n_i), \text{ for } i = 1, \dots, m,$$

¹³ Quando se calcula o efeito para uma subamostra, perde-se representatividade da análise (diminui o que é chamado de validade externa), mas ganha-se robustez/precisão nas estimativas (aumento o que é chamado de validade interna). Nesse sentido, há um *trade-off* entre uma resposta crível para uma subpopulação e uma resposta controversa para toda a amostra. A primeira opção costuma ser a mais atrativa. Neste sentido, a ideia de procedimento de *trimming* é excluir unidades para as quais $p(X_i)$ é próximo de 0 ou 1. Considera-se, portanto, as observações dentro do intervalo: $\alpha < p(X_i) < 1 - \alpha$, o que melhora a habilidade em estimar precisamente o efeito. Crump *et al.* (2009) estima que $\hat{\alpha} \approx 0,1$.

onde os números de ensaios de Bernoulli n_i são conhecidos e as probabilidades de êxito p_i são desconhecidas.

O modelo é obtido com base de que cada ensaio (valor de i) e o conjunto de variáveis explicativas/independentes possa informar acerca da probabilidade final. Estas variáveis independentes podem ser apresentadas como um vector X_i k -dimensional:

$$p_i = E \left(\frac{Y_i}{n_i} \middle| X_i \right).$$

As probabilidades binomiais desconhecidas são modeladas como uma função linear dos X_i :

$$L(p_i) = \ln \left(\frac{p_i}{1 - p_i} \right) = Y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}.$$

onde $L(p_i)$ é logaritmo da razão de chances para uma unidade de mudança que não é somente linear em X_i , mas também é linear nos parâmetros. $L(p_i)$ é denominado o logit e, consequentemente, o modelo anterior é denominado modelo logit (GUJARATI, 2011).

Note-se que um elemento particular de X_i pode ser ajustado a 1 para todo o i obtendo o intercepto no modelo. Os parâmetros desconhecidos β_j são habitualmente estimados através de máxima verossimilhança.

A interpretação dos valores estimados do parâmetro β_j é similar aos efeitos aditivos em logaritmo da razão de chances para uma unidade de mudança na j -ésima variável explicativa. O modelo tem uma formulação equivalente dada por:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}.$$

Segundo Gujarati (2011), as principais características do modelo logit são:

1. Quando p passa de 0 a 1, o logit varia de $-\infty$ para $+\infty$. Ou seja, embora as probabilidades se situem entre 0 e 1, os logits não estão submetidos a essa restrição.
2. Embora L seja linear em X , as próprias probabilidades não o são.
3. Se L é positivo, isto significa que, quando o valor de um ou mais regressores aumenta, as chances de que o regressando seja igual 1 aumenta. Se L for negativo, as chances de que o regressando seja igual 1 diminui com X aumentando. Em outras palavras, o logit se torna negativo e cada vez maior à medida que a razão de chances diminui de 1 para 0 e se torna cada vez maior e positivo à medida que a razão de chances aumenta de 1 até infinito.

4.2.4 Mínimos Quadrados Ordinários

O método de mínimos quadrados ordinários¹⁴ (MQO) ou OLS (do inglês *Ordinary Least Squares*) é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (resíduos). Esta técnica é a mais comum entre os métodos de estimação econométrica. Em que, consiste em um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma a maximizar o grau de ajuste do modelo aos dados observados (GUJARATI, 2011).

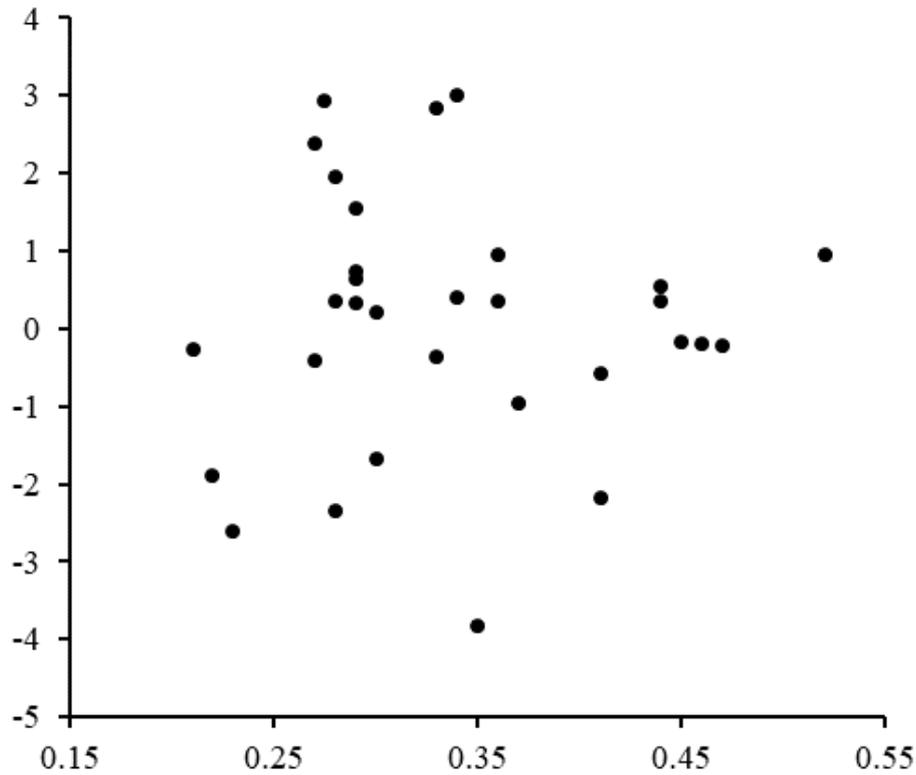
O primeiro requisito para usar o método dos mínimos quadrados ordinários é que o modelo seja linear nos parâmetros, isto é, as variáveis apresentam uma relação linear entre si. Outro requisito é que o termo de erro (resíduo) seja normalmente distribuído. A figura 10 apresenta um exemplo de distribuição normal dos resíduos. Já a figura 11 apresenta um exemplo em que o modelo linear não é o mais apropriado. O requisito anterior é uma premissa razoável, dado que o termo de erro seguirá uma distribuição normal, pelo teorema do limite central. Contudo, se isso não for satisfatório, sempre é possível testar a hipótese de que os resíduos sejam normais, o teste de Jarque-Bera é muito utilizado para este objetivo. Este utiliza os resultados para os momentos da distribuição normal, em particular os coeficientes de assimetria e de curtose (SARTORIS NETO, 2017). Algo importante é que o Teorema Gauss-Markov garante que o estimador de mínimos quadrados seja o estimador não-enviesado de mínima variância linear na variável resposta. Para tanto, um modelo deve seguir as seguintes hipóteses básicas:

1. $E(\varepsilon_i) = 0$ (os resíduos têm média zero).
2. Os resíduos são normalmente distribuídos.
3. Os x_i são fixos (não estocásticos).
4. $var(\varepsilon_i) = \sigma^2$ (a variância dos resíduos é constante).
5. $E(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$ (erros não são autocorrelacionados).
6. Cada variável independente x_i não pode ser combinação linear das demais.

A primeira exposição clara e concisa do método dos mínimos quadrados foi publicada por Legendre, em 1805. Mas foi em 1822, Gauss foi capaz de afirmar que a abordagem de mínimos quadrados para análise de regressão é ótima no sentido de que em um modelo linear onde os erros têm média zero, não são correlacionados e têm variâncias constante, assim, o melhor estimador linear é o estimador de mínimos quadrados. Este resultado é conhecido como teorema de Gauss-Markov (MEMÓRIA, 2004).

¹⁴ Atribui-se a Carl Friedrich Gauss as bases fundamentais desse método.

Figura 10 – A distribuição normal dos resíduos $E(\varepsilon_i) = 0$ indica que o modelo linear é apropriado.



Fonte: Autoria própria (2022).

O objetivo do método de mínimos quadrados ordinários consiste em ajustar os parâmetros de uma função para que seja o melhor ajuste a um conjunto de dados. Um conjunto de dados simples consiste em n pontos (pares ordenados)

$$(x_i, y_i), i = 1, \dots, n,$$

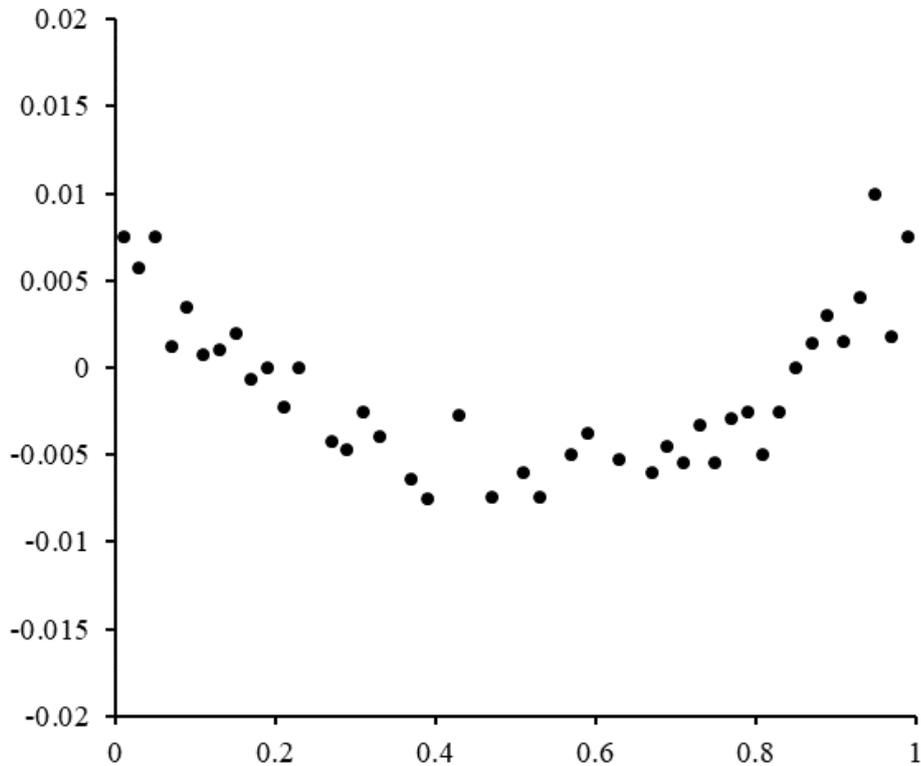
onde x_i é a variável independente e y_i é a variável dependente. Assim, o objetivo é encontrar os valores dos parâmetros para o modelo que "melhor" se ajusta aos dados. O ajuste do modelo é feito por seu resíduo, definido como a diferença entre o valor real da variável dependente e o valor previsto pelo modelo:

$$r_i = y_i - f(x_i, \beta)$$

O método de mínimos quadrados, então, encontra os valores dos parâmetros ótimos, minimizando a soma dos quadrados dos resíduos, S :

$$S = \sum_{i=1}^n r_i^2$$

Figura 11 – A forma parabólica dos resíduos indicam que o modelo linear não é apropriado.



Fonte: Autoria própria (2022).

Um exemplo de modelo em duas dimensões é uma linha reta. Denotando o intercepto em y , α , e a inclinação β , a função do modelo é dada por:

$$y = \alpha + \beta x$$

Dessa forma, o objetivo é estimar valores de uma determinada variável y . Para isso, consideram-se os valores de outra variável x que acredita-se ter poder de explicação sobre y , dado por:

$$y = \alpha + \beta x + \varepsilon$$

onde α é a constante (porque não depende de x), β é o coeficiente de variação de x e ε é o termo de erro (resíduo) – representa a variação de y que não é explicada pelo modelo.

Para uma base de dados com n valores observados de y e de x . Perceba que y e x são vetores, ou seja, representam uma lista de valores. O método de mínimos quadrados ordinários ajuda a encontrar as estimativas de α e β . Como o nome diz, serão somente estimativas desses parâmetros, porque o valor real dos parâmetros são desconhecidos. Portanto, ao fazer

a estimativa, muda-se algumas notações:

$$\alpha \rightarrow \hat{\alpha}$$

$$\beta \rightarrow \hat{\beta}$$

Deste modo, ao estimar o modelo usando a base de dados:

$$y_i = \hat{\alpha} + \hat{\beta}x_i$$

onde i indica cada uma das n observações da base de dados e $E(\varepsilon_i) = 0$.

Como mencionado, o método de mínimos quadrados minimiza a soma dos quadrado dos resíduos, ou seja, minimiza $\sum_{i=1}^n e_i^2$. A ideia por trás dessa técnica é que, minimizando a soma do quadrado dos resíduos, encontra-se $\hat{\alpha}$ e $\hat{\beta}$ fornecendo a menor diferença entre a previsão de y e o y realmente observado. Substituindo e_i por $y_i - \hat{\alpha} - \hat{\beta}x_i$, chega-se:

$$S = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

A minimização se dá ao derivar S em relação a $\hat{\alpha}$ e $\hat{\beta}$:

$$\frac{\partial S}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

$$\frac{\partial S}{\partial \hat{\beta}} = -2 \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

Distribuindo e dividindo a primeira expressão por $2n$ tem-se:

$$\frac{-2 \sum_{i=1}^n y_i}{2n} + \frac{2 \sum_{i=1}^n \hat{\alpha}}{2n} + \frac{2 \sum_{i=1}^n \hat{\beta}x_i}{2n} = \frac{0}{2n}$$

$$\frac{-\sum_{i=1}^n y_i}{n} + \frac{\sum_{i=1}^n \hat{\alpha}}{n} + \frac{\hat{\beta} \sum_{i=1}^n x_i}{n} = 0$$

$$-\bar{y} + \hat{\alpha} + \hat{\beta}\bar{x} = 0$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

onde \bar{y} é a média amostral de y e \bar{x} é a média amostral de x .

Substituindo esse resultado na segunda expressão¹⁵:

$$\begin{aligned}
 -2 \sum_{i=1}^n x_i (y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i) &= 0 \\
 \sum_{i=1}^n [x_i (y_i - \bar{y}) + x_i \hat{\beta} (\bar{x} - x_i)] &= 0 \\
 \sum_{i=1}^n x_i (y_i - \bar{y}) + \hat{\beta} \sum_{i=1}^n x_i (\bar{x} - x_i) &= 0 \\
 \hat{\beta} &= \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})}
 \end{aligned}$$

A regressão múltipla apresenta um funcionamento semelhante ao da regressão simples, porém, leva em consideração diversas variáveis independentes x influenciando y ao mesmo tempo:

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \dots + x_k\beta_k + \varepsilon$$

Ao usar a base de dados com k variáveis explicativas e n observações, o modelo pode ser escrito na forma matricial:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ 1 & x_{14} & x_{24} & \dots & x_{k4} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \times \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ \dots \\ e_n \end{pmatrix}$$

onde x_{ji} representa o valor da j -ésima variável da i -ésima observação. A fórmula também pode ser escrita na forma resumida:

$$y = Xb + e$$

A solução de mínimos quadrados continua sendo alcançada através da minimização da soma do quadrado dos resíduos $\sum_{i=1}^n e_i^2$, que pode ser reescrita como $e'e$.

¹⁵ Alguns autores também usam uma equação diferente, mas com o mesmo resultado:

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Substituindo e por $y - Xb$, chega-se:

$$\begin{aligned} S(b) &= (y - Xb)^T (y - Xb) \\ &= y^T y - y^T Xb - b^T X^T y + b^T X^T Xb \end{aligned}$$

A minimização pode ser obtida ao derivar $S(b)$ em relação a b e igualar a zero. O primeiro termo não depende de b , os segundo e terceiro termos são iguais e o terceiro termo é uma forma quadrática dos elementos de b .

$$\begin{aligned} \frac{\partial S}{\partial b} &= -2X^T y + 2X^T Xb = 0 \\ X^T Xb &= X^T y \\ b &= (X^T X)^{-1} X^T y \end{aligned}$$

Acima, foram apresentadas hipóteses básicas que o modelo de regressão deve seguir. Entretanto, em muitas situações essas hipóteses não são verificadas, principalmente, quando o estudo pertence à Ciências Sociais Aplicadas, pois os dados não resultam de um experimento controlado. Sendo mais específico, as hipóteses com maior chance de não se verificar são: que as variáveis independentes sejam fixas; que a variância dos resíduos seja constante; que os erros não sejam autocorrelacionados; e que cada variável independente não seja uma combinação linear das demais.

Multicolinearidade é a situação em que, ao menos, uma variável independente é combinação linear de, ao menos, uma outra variável independente. Quando isso ocorre, a variação na variável é decorrente da variação na outra variável. Por exemplo:

$$x_2 = 3x_1$$

$$x_3 = x_1 + x_2$$

$$x_4 = 3x_2 - 5$$

Quando há multicolinearidade da forma $x_2 = 3x_1$, x_2 é o triplo de x_1 , assim, qualquer variação em x_1 terá uma variação proporcionalmente idêntica em x_2 . Portanto, é impossível diferenciar a influência dessas variáveis independentes na variável dependente. Em Ciências Sociais Aplicadas, esse tipo de interação entre as variáveis independentes é muito comum (SARTORIS, 2017).

As consequências da multicolinearidade é que o teste t de cada parâmetro pode resultar em não-significante, ainda que as variáveis independentes expliquem o comportamento da variável dependente. Isto ocorre, pois o desvio-padrão de cada $\hat{\beta}$ aumenta na presença de multicolinearidade. Especificamente, as variâncias dos coeficientes na regressão múltipla são

dadas por: $S_{\hat{\beta}}^2 = S^2 (\mathbf{X}'\mathbf{X})^{-1}$. Se o coeficiente de correlação for próximo de 1 ou -1, o valor do determinante da matriz \mathbf{X} , conseqüentemente, da matriz $\mathbf{X}'\mathbf{X}$, será muito pequeno e, assim, $S_{\hat{\beta}}^2$ será muito grande. Fato importante é que mesmo na presença de multicolinearidade, as propriedades usuais do estimador de mínimos quadrados são mantidas. Ou seja, o estimador continua sendo não-viesado, eficiente e consistente (SARTORIS, 2017).

Uma maneira usual de identificar multicolinearidade é quando o teste t de cada parâmetro é pouco significativo, mas o teste F é bastante significativo ou o R^2 é alto. Sinais dos parâmetros invertidos ao esperado, também, é outro indicativo de multicolinearidade. Além disso, também é possível analisar, previamente, se há correlações entre as variáveis independentes. Contudo, isto não deve funcionar se tiver mais de duas variáveis independentes, pois a correlação é calculada usando duas a duas variáveis, mas a multicolinearidade pode ser o resultado da combinação de três ou mais variáveis independentes. Por isso, modelos com muitas variáveis não é aconselhável, pois é muito difícil não ter uma relação entre as variáveis quando o estudo é das Ciências Sociais Aplicadas.

Quando for detectado o problema da multicolinearidade, a providência mais simples é retirar as variáveis correlacionadas do modelo. O motivo de retirar variáveis correlacionadas, é que a variável dependente já estaria captando o efeito desejado, tornando a outra variável independente desnecessária. Mais especificamente, suponha o seguinte modelo: $Y = \beta_1 X_1 + \beta_2 X_2$, sendo $X_2 = 2X_1$, assim, $Y = \beta_1 X_1 + \beta_2 2X_1 = (\beta_1 + 2\beta_2) X_1$, definindo: $\beta = \beta_1 + 2\beta_2$, o modelo resultaria em: $Y = \beta X_1$. Em palavras, dado que X_1 e X_2 são correlacionadas, deve-se usar apenas uma das duas variáveis para captar o efeito em Y .

Contudo, o critério de retirar as variáveis independentes correlacionadas pode não ser interessante para a proposta da pesquisa. Pois, talvez o objetivo seja de verificar o efeito individual de cada variável independente na variável dependente. Outra solução é aumentar a amostra, pois alguns efeitos podem ser intrínsecos da amostra e não pertencer à população. Também poderia reespecificar o modelo. Por fim, ainda há a alternativa de não se fazer nada, pois o estimador de mínimos quadrados mantém as propriedades de não viés, eficiência e consistência, mesmo na presença de multicolinearidade. Assim, se o objetivo for fazer previsões a respeito da variável dependente, retirar variáveis independentes correlacionadas irá reduzir a eficácia da previsão (SARTORIS, 2017). Resumindo, em Ciências Sociais Aplicadas, é muito difícil evitar o problema da multicolinearidade, mais isso é mais uma questão numérica que pode ser deixada de lado nas situações justificáveis.

Outra hipótese que deve ser preservada é a da não autocorrelação dos erros. Em geral, isto significa a correlação de uma variável com os valores defasados dela mesma, ou seja, a variável independente x_t tem correlação com x_{t-1} , por exemplo. É importante destacar que esta situação é algo associado a séries temporais. Outra advertência é que a hipótese diz "não autocorrelação dos erros", o termo de erro (ou resíduo) não é uma variável específica do modelo, mas um conjunto de influências que fogem do controle do modelador. Assim, no caso de uma variável independente importante ser omitida do modelo, a omissão desta variável

jogará a sua influência para o termo de erro. Portanto, a omissão de uma variável importante pode fazer com que ocorra autocorrelação dos erros.

Outro motivo para a existência da autocorrelação dos erros é a má especificação funcional do modelo. Por exemplo, se a suposição é um modelo linear, mas a verdadeira relação entre as variáveis dependente e independente é exponencial, o termo de erro apresentaria um comportamento sistemático relacionado com a verdadeira relação entre as variáveis. Em muitos casos, a autocorrelação dos erros tem origem no próprio processo da variável, pois o resultado provocado nesta variável pode demorar algum tempo para ser absorvido, assim, os efeitos seriam distribuídos em mais de um período.

Na presença de autocorrelação dos erros, a demonstração do Teorema Gauss-Markov não é válida, assim, o estimador de mínimos quadrados ordinários não será o melhor estimador linear não-viesado, assim, este estimador não será o de menor variância entre todos os estimadores possíveis. Contudo, a hipótese para que o estimador seja não-viesado e consistente não é violada. Portanto, na maioria dos casos¹⁶, ainda que não tenha a menor variância, o estimador continua não-viesado e consistente, mesmo na presença de autocorrelação dos erros.

A forma mais comum de identificar autocorrelação dos erros é por meio do teste Durbin-Watson. Contudo, este teste não pode ser usado quando o modelo não incluir o intercepto, ou quando incluir defasagens da variável dependente entre as variáveis independentes. Para resolver este problema, o primeiro passo é identificar a causa da autocorrelação do erro. Se o problema for de especificação, pode-se resolver isso com a inclusão de mais variáveis, ou com a alteração da forma funcional do modelo. Contudo, se a autocorrelação dos erros é parte integrante do modelo, corrigir o problema passa pelo conhecimento de como é a estrutura dessa autocorrelação. Por exemplo, assumo o modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

sendo que existe autocorrelação dos erros na forma de um AR(1) $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$. E suponha que ρ seja conhecido, multiplicando a defasagem de um período do modelo anterior fica:

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho\varepsilon_{t-1},$$

agora, subtraindo o segundo modelo do primeiro,

$$Y_t - \rho Y_{t-1} = \beta_1 - \rho\beta_1 + \beta_2 X_t - \rho\beta_2 X_{t-1} + \varepsilon_t - \rho\varepsilon_{t-1}$$

¹⁶ Quando a defasagem da variável dependente é variável independente não há garantia da não existência de viés.

ou,

$$\tilde{Y}_t = \tilde{\beta}_1 + \beta_2 \tilde{X}_t + u_t$$

onde $\tilde{Y}_t = Y_t - \rho Y_{t-1}$, $\tilde{X}_t = X_t - \rho X_{t-1}$ e $\tilde{\beta}_1 = \beta_1(1 - \rho)$.

Portanto, o modelo anterior não tem autocorrelação dos erros. E o parâmetro ρ pode ser conhecido usando o termo de erro, resultante da estimação do modelo com autocorrelação, para estimar o modelo AR(1) $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, ou pode-se usar a estatística do teste Durbin-Watson (DW), ou seja, $\rho \approx 1 - \frac{DW}{2}$ (SARTORIS, 2017).

Homocedasticidade é o nome que se dá quando a variância dos erros é constante. Assim, quando isto não ocorre, há um problema de heterocedasticidade. Este problema ocorre quando para incrementos pequenos na variável independente a variância dos erros é pequena, mas para incrementos grandes na variável independente a variância dos erros é grande. Em outras palavras, na medida que há aumentos na variável independente, a variável dependente aumenta, mas com taxas de crescimento crescentes, isto é, com aumento na variância dos erros.

O problema da heterocedasticidade é semelhante ao apresentado sobre a autocorrelação dos erros, ou seja, é uma hipótese necessária para a demonstração do Teorema Gauss-Markov. Dessa forma, o estimador de mínimos quadrados ordinários continuam não-viesados, mas já não é o de menor variância. Para identificar a heterocedasticidade é comum usar o teste de Goldfeld e Quandt, este teste utiliza um valor de inflexão da variável independente, \bar{X} , de tal forma, que seja possível separar a regressão em duas partes (como se fossem modelos separados), a primeira com valores da variável independente inferior a \bar{X} , e a segunda com valores da variável independente superior a \bar{X} . Então, é feito um teste F para comparar as variâncias das partes separadas. Havendo diferença nas variâncias das duas regressões, a hipótese nula de homocedasticidade é rejeitada.

Havendo heterocedasticidade, o procedimento mais simples é se o padrão da causa do problema é conhecido. Por exemplo, suponha o seguinte modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

seja a variância dos erros dada por $var(\varepsilon_t) = \sigma_t^2 = z_t \sigma^2$. Ou seja, a variância dos erros não é constante, mas é uma variável de z_t . Então, para ter homocedasticidade o objetivo é retirar o z_t da variância. Se o modelo anterior for dividido por $\sqrt{z_t}$, chega-se em:

$$\frac{Y_t}{\sqrt{z_t}} = \beta_1 \frac{1}{\sqrt{z_t}} + \beta_2 \frac{X_t}{\sqrt{z_t}} + \frac{\varepsilon_t}{\sqrt{z_t}}$$

ou,

$$\hat{Y}_t = \hat{\beta}_1 + \beta_2 \hat{X}_t + \hat{\varepsilon}_t$$

onde $\widehat{Y}_t = \frac{Y_t}{\sqrt{z_t}}$, $\widehat{\beta}_1 = \beta_1 \frac{1}{\sqrt{z_t}}$, $\widehat{X}_t = \frac{X_t}{\sqrt{z_t}}$ e $\widehat{\varepsilon}_t = \frac{\varepsilon_t}{\sqrt{z_t}}$.

Assim, a variância dos erros do novo modelo será dado por: $var(\widehat{\varepsilon}_t) = var\left(\frac{\varepsilon_t}{\sqrt{z_t}}\right) = \frac{1}{z_t} var(\varepsilon_t) = \frac{1}{z_t} z_t \sigma^2 = \sigma^2$. Portanto, o novo modelo é homocedástico.

Outra hipótese necessária do Teorema Gauss-Markov é a de que os as variáveis independentes sejam fixos, isto é, não-estocásticos. Quando isto não ocorre, o problema é chamado de simultaneidade. Em Ciências Sociais Aplicadas, é comum ter duas equações simultâneas. Um caso típico é o modelo de determinação do preço de um produto qualquer. Em particular, pelo lado da oferta, o produtor produzirá mais quanto maior for o preço do bem, pelo lado da demanda, o consumidor comprará mais quanto menor for o preço do bem.

Suponha que a quantidade produzida (Q^o) seja função apenas do preço (P), então, a equação da oferta será dada por:

$$Q_t^o = \alpha_1 + \alpha_2 P_t + \mu_t, \text{ com } \alpha_1 > 0.$$

Suponha ainda, que a quantidade de produto demandada pelos consumidores (Q^d) (equação da demanda) seja função dos preços (P) e da renda (R):

$$Q_t^d = \beta_1 + \beta_2 P_t + \beta_3 R_t + \nu_t, \text{ com } \beta_2 < 0.$$

Em equilíbrio, $Q^o = Q^d = Q$, assim, as duas equações anteriores ficam:

$$Q_t = \alpha_1 + \alpha_2 P_t + \mu_t, \text{ com } \alpha_1 > 0.$$

$$Q_t = \beta_1 + \beta_2 P_t + \beta_3 R_t + \nu_t, \text{ com } \beta_2 < 0.$$

Neste modelo, as variáveis Q e P se determinam mutuamente, isto é, são chamadas de variáveis endógenas. E R é chamada de variável exógena. A regressão por mínimos quadrados ordinários apresentará um estimador viesado e inconsistente, já que uma variável independente é uma variável endógena, e assim, estará correlacionada com o termo de erro.

Igualando as duas equações anteriores,

$$\alpha_1 + \alpha_2 P_t + \mu_t = \beta_1 + \beta_2 P_t + \beta_3 R_t + \nu_t$$

chega-se em,

$$P = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3}{\alpha_2 - \beta_2} R_t + \frac{\nu_t - \mu_t}{\alpha_2 - \beta_2}$$

A equação anterior apresenta o preço em função apenas de variáveis exógenas. Nota-se, olhando esta equação, que o preço é correlacionado com os dois termos de erro. Substi-

tuindo a equação anterior na equação da oferta, chega-se:

$$Q_t = \alpha_1 + \alpha_2 \left[\frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3}{\alpha_2 - \beta_2} R_t + \frac{\nu_t - \mu_t}{\alpha_2 - \beta_2} \right] + \mu_t$$

reagrupando-a,

$$Q_t = \left[\alpha_1 + \frac{\alpha_2 \beta_1 - \alpha_2 \alpha_1}{\alpha_2 - \beta_2} \right] + \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} R_t + \left[\mu_t + \frac{\alpha_2 \nu_t - \alpha_2 \mu_t}{\alpha_2 - \beta_2} \right]$$

ou,

$$Q_t = \frac{\alpha_2 \beta_1 - \alpha_1 \beta_2}{\alpha_2 - \beta_2} + \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} R_t + \frac{\alpha_2 \nu_t - \mu_t \beta_2}{\alpha_2 - \beta_2}$$

Essa equação também apresenta Q apenas em função de variáveis exógenas. Há um novo sistema de equações, que isola as variáveis endógenas de cada equação, essas equação são chamadas de equações na forma reduzida. Ou seja,

$$P = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3}{\alpha_2 - \beta_2} R_t + \frac{\nu_t - \mu_t}{\alpha_2 - \beta_2}$$

$$Q_t = \frac{\alpha_2 \beta_1 - \alpha_1 \beta_2}{\alpha_2 - \beta_2} + \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} R_t + \frac{\alpha_2 \nu_t - \mu_t \beta_2}{\alpha_2 - \beta_2}$$

ou,

$$P = \gamma_1 + \gamma_2 R_t + \frac{\nu_t - \mu_t}{\alpha_2 - \beta_2} + \xi_t$$

$$Q_t = \gamma_3 + \gamma_4 + \zeta_t$$

onde $\gamma_1 = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2}$, $\gamma_2 = \frac{\beta_3}{\alpha_2 - \beta_2}$, $\gamma_3 = \frac{\alpha_2 \beta_1 - \alpha_1 \beta_2}{\alpha_2 - \beta_2}$, $\gamma_4 = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2}$, $\xi_t = \frac{\nu_t - \mu_t}{\alpha_2 - \beta_2}$ e $\zeta_t = \frac{\alpha_2 \nu_t - \mu_t \beta_2}{\alpha_2 - \beta_2}$.

Nota-se que as equações na forma reduzida não têm mais o problema de que um ou mais variáveis independentes são correlacionados com o termo de erro, então, podem ser perfeitamente estimados por mínimos quadrados ordinários. Apenas, há a questão de que os valores estimados na forma reduzida são os parâmetros γ e não α e β . Esta forma de estimação é conhecida como mínimos quadrados indiretos. A ressalva que ele só pode ser usado nos casos do modelo ser perfeitamente identificável, que é o caso anterior – duas variáveis endógenas para duas equações.

Se o modelo é subidentificado, não dá para estimá-lo. Se for superidentificado, que não é um problema, não dá para ter uma relação um para um entre os parâmetros da forma estrutural e reduzida que dê apenas uma única solução. Mas é possível usar o chamado método dos mínimos quadrados de dois estágios, que consiste em estimar as equações na forma reduzida. Então, encontrar os valores estimados para as variáveis endógenas. Como são valores estima-

dos, não incluem os resíduos, portanto, não tem correlação com o termo de erro (SARTORIS, 2017).

5 ANÁLISE DOS RESULTADOS

Neste capítulo, será apresentado os resultados deste trabalho. Com objetivo de evitar o viés de seleção, foi realizado o pareamento, ou seja, a seleção de subconjuntos e a sub-classificação para criar grupos de tratamento e controle balanceados nas covariadas incluídas. Assim, os procedimentos para melhorar os modelos estatísticos paramétricos por meio do pré-processamento de dados com métodos de pareamento não paramétricos seguiram Ho, Imai, King e Stuart (2007) que implementaram uma ampla variedade de métodos de pareamento, tornando possível reduzir a dependência de inferências causais em suposições de modelagem estatística difíceis de justificar.

Diversos métodos de pareamento envolvem a distância entre as unidades tratadas e controle. Entre as opções incluem a distância de Mahalanobis, distância com o *propensity score*, entre outras. Assim, neste trabalho, o *propensity score* foi escolhido como medida de "distância". Para tanto, usou-se um modelo de Logit.

De posse do *propensity score*, o passo seguinte foi balancear a amostra. Para esta finalidade, usou-se o método ótimo de pareamento das unidades, ou seja, a soma das distâncias absolutas entre pares na amostra pareada é a menor possível, com uma taxa de pareamento 1 (para cada unidade tratada foi escolhida uma unidade de controle). A vantagem do pareamento ótimo é que a ordem de pareamento não precisa ser especificada, assim, é menos provável que as distâncias extremas entre pares sejam grandes, ao contrário do pareamento do vizinho mais próximo. Contudo, como é um método de seleção de subconjunto, o pareamento ótimo tende a funcionar de maneira semelhante ao pareamento do vizinho mais próximo, pois subconjuntos semelhantes de unidades serão selecionados para serem combinados.

Como mencionado, o *propensity score* foi construído usando um modelo logit binomial, a tabela 1 apresenta os resultados desta estimação, sendo a variável dependente a variável tratamento (se abriu firma em 2020) e escolaridade, idade e os quadrados dessas duas variáveis como variáveis independentes. Nota-se que os resultados da estatística z indicaram um baixo nível de significância das variáveis independentes. Contudo, o mais relevante é o PS realizar um bom balanceamento da amostra (esses resultados serão apresentados a seguir), além disso, as relações entre cada variável e o seu valor ao quadrado estiveram conforme o esperado, pois os sinais estão invertidos, indicando que o crescimento da variável tem um sentido, mas com incremento decrescente.

Portanto, mesmo com o problema do valor z dos coeficientes, é possível analisar os sinais desses coeficientes. Segundo o modelo, indivíduos com escolaridade menor, e idade maior, teriam uma probabilidade maior de fazerem parte do grupo de tratamento, mas dado o valor dos coeficientes das variáveis ao quadrado, esses efeitos seriam decrescentes. Em outras palavras, dada a suposição de que o grupo de tratamento seria formado por menos empreendedores, os resultados indicam que os empreendedores teriam uma escolaridade maior e seriam mais jovens.

Tabela 1 – Modelo Logit – variável dependente: probabilidade da firma ter sido aberta em 2020 (grupo tratado).

	Coeficiente	Erro Padrão	valor z	$Pr(> z)$
Constante	-1,841774	3,784997	-0,487	0,627
Escolaridade	-0,554807	0,486109	-1,141	0,254
Idade	0,120840	0,174244	0,694	0,488
<i>Escolaridade</i> ²	0,032952	0,030819	1,069	0,285
<i>Idade</i> ²	-0,001326	0,002119	-0,626	0,532
Média var. depend.		0,21	D. P. var. depend.	0,41
<i>R</i> ² de McFadden		0,02	<i>R</i> ² ajustado	-0,06
Qui-quadrado(4)		2,47		

Fonte: Autoria própria (2022).

A tabela 2 apresenta uma descrição estatística das covariadas e do *propensity score* calculado diferenciando entre os grupos de tratamento e de controle. No grupo de tratamento, foram obtidos 0,23, 7,6 e 40,8 para PS, escolaridade e idade, respectivamente. Já para o grupo de controle os resultados para essas variáveis foram 0,2, 7,98 e 38,7, respectivamente. Com diferenças das médias padronizadas de 0,36, -0,12 e 0,2, respectivamente, para as variáveis acima. Portanto, nota-se que existe diferença nas variáveis entre os dois grupos. As duas últimas colunas da tabela 2 representam o teste de Kolmogorov-Smirnov (KS)¹⁷, segundo esses resultados, para um $\alpha = 0,05$, não é possível rejeitar a hipótese nula de que as amostras possuem a mesma distribuição, para escolaridade e para idade, pois todos os valores são inferiores ao valor crítico de 0,238, mas deve-se rejeitar a hipótese nula para PS.

Tabela 2 – Resumo estatístico para os dados antes do balanceamento.

	Média tratados	Média controle	Dif média padr	Taxa de var	KS médio	KS Max
Propensity score (PS)	0,2255	0,206	0,3557	0,8622	0,1016	0,2468
Escolaridade	7,6	7,9787	-0,1218	1,3144	0,0391	0,1153
Idade	40,8	38,734	0,2	0,9376	0,0642	0,2149
<i>Escolaridade</i> ²	67,04	70,9362	-0,0804	1,2740	0,0391	0,1153
<i>Idade</i> ²	1767,12	1612,9681	0,1801	0,9892	0,0642	0,2149

Fonte: Autoria própria (2022).

Analisando os resultados depois do balanceamento (tabela 3), percebe-se que as diferenças das médias padronizadas diminuíram significativamente (figura 14), isto é, o balanceamento deixou a distribuição dos dois grupos mais semelhantes, isto pode ser visto nas estatísticas KS, que diminuíram consideravelmente. Na figura 12, são apresentados os gráficos das funções de densidade cumulativa empírica. Esses gráficos são criados para cada covariada antes e depois do balanceamento. Cada ponto nas linhas corresponde à proporção de unidades menor ou igual ao valor da covariada correspondente (no eixo x). Desvios entre as linhas no mesmo gráfico indicam desequilíbrio da distribuição entre os grupos de tratamento para a co-

¹⁷ A estatística de Kolmogorov-Smirnov quantifica a distância entre a função distribuição empírica da amostra e a função distribuição acumulada da distribuição de referência. A hipótese nula diz que as amostras são retiradas da mesma distribuição.

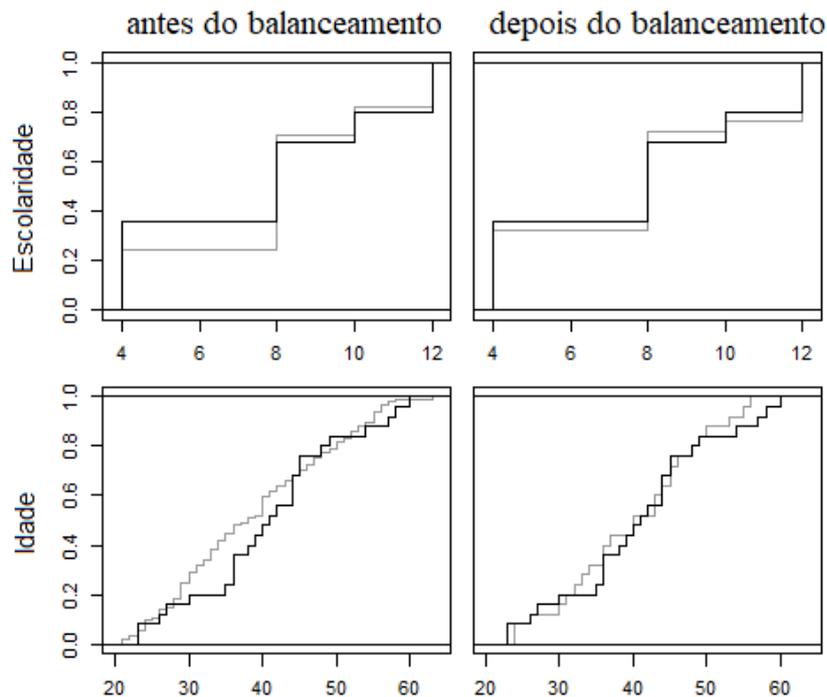
variada. Assim, falando de forma simples, o teste KS é calculado medindo a distância entre as funções dos dois grupos. Nota-se que, depois do balanceamento, as duas curvas ficaram mais próximas, e por isso, percebe-se que houve uma melhora substancial na estatística KS, isto fica evidente analisando a figura 13, em que, depois do balanceamento, as distribuições ficaram mais próximas de uma distribuição normal.

Tabela 3 – Resumo estatístico para os dados balanceados.

	Média tratados	Média controle	Dif média padr	Taxa de var	KS médio	KS Max
Propensity score	0,2255	0,2254	0,0025	0,9917	0,0105	0,04
Escolaridade	7,6	7,76	-0,0515	1,0240	0,03	0,04
Idade	40,8	40	0,0774	1,2235	0,037	0,12
<i>Escolaridade</i> ²	67,04	69,28	-0,0462	0,9819	0,03	0,04
<i>Idade</i> ²	1767,12	1683,76	0,0974	1,3154	0,037	0,12

Fonte: Autoria própria (2022).

Figura 12 – Gráficos das funções de densidade cumulativa empírica para os dados antes e depois do balanceamento – linha cinza para unidades de controle e a preta para unidades tratadas.

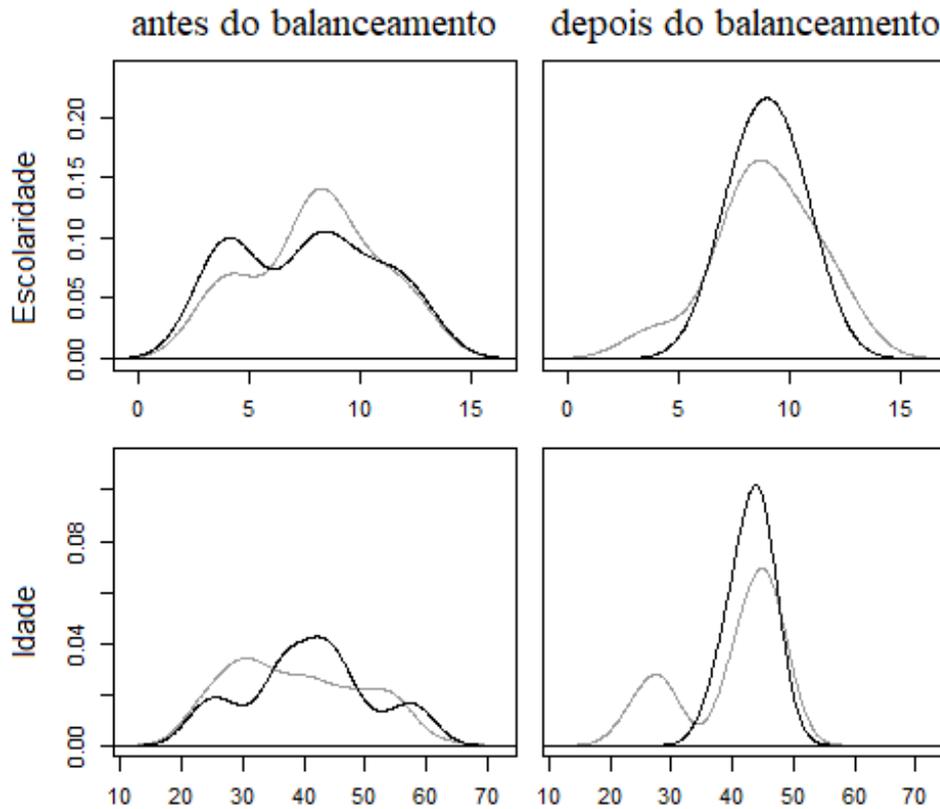


Fonte: Autoria própria (2022).

Já a porcentagem de melhora do balanceamento está apresentada na tabela 4, o maior e o menor refinamentos são vistos no PS e na escolaridade, 99,3% e 57,8%, respectivamente. Este resultado fica mais evidente olhando a figura 14.

Pelos resultados apresentados, percebe-se que o balanceamento foi eficiente em evitar o viés de seleção. Então, o próximo passo do trabalho foi verificar a questão principal deste estudo, ou seja, se o empreendedorismo é estado-independente. Para tanto, foram calculadas

Figura 13 – Distribuições dos grupos de tratamento e controle antes e depois do balanceamento – linha cinza para unidades de controle e a preta para unidades tratadas.



Fonte: Autoria própria (2022).

Tabela 4 – Melhora percentual pelo balanceamento.

	Dif. média pad.	taxa Var.	KS médio	KS Max
Propensity score	99,3	94,4	89,7	83,8
Escolaridade	57,8	91,3	23,4	65,3
Idade	61,3	-213,1	42,4	44,2
$Escolaridade^2$	42,5	92,5	23,4	65,3
$Idade^2$	45,9	-2432,7	42,4	44,2

Fonte: Autoria própria (2022).

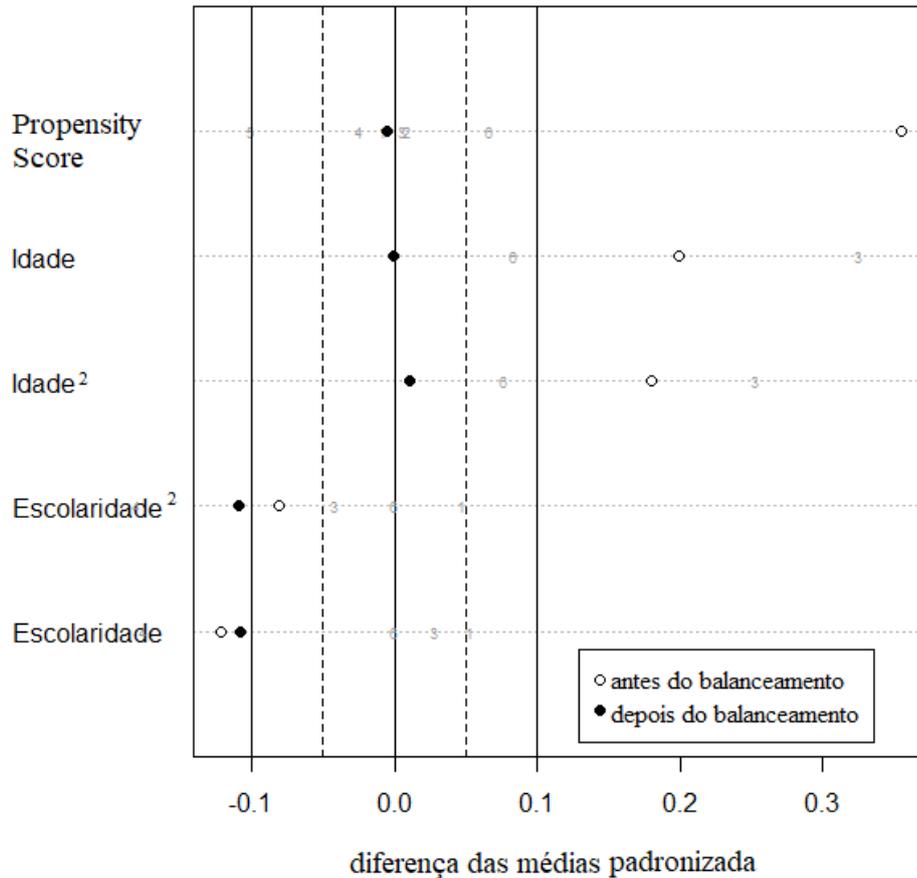
duas regressões usando o método dos mínimos quadrados ordinários para as amostras antes e depois do balanceamento. O modelo estimado foi:

$$\text{índice de empreendedorismo}_t = \alpha + \beta \text{ tratamento}_t + \varepsilon_t \quad (2)$$

onde α é o intercepto, β é a inclinação e ε é o termo de erro.

O resultado da regressão da amostra sem balanceamento está apresentado na tabela 5. Apesar da constante ser estatisticamente significativa, o coeficiente relacionado ao tratamento apresentou um p-valor de 0,131. Em outras palavras, usando a amostra sem o balanceamento, não é possível dizer que o indivíduo que abriu firma em 2020 possui um índice de empreen-

Figura 14 – Diferença das médias padronizada.



Fonte: Autoria própria (2022).

dedorismo inferior aos indivíduos que abriram firmas em 2019 em mais de 13% dos casos, ou seja, o empreendedorismo não seria estado-independente.

Tabela 5 – MQO sem o balanceamento. Variável dependente: Ind_Empreendedorismo.

	Coefficiente	Erro Padrão	razão-t	p-valor
constante	3,3	0,13	26,2	0,0
tratamento	-0,42	0,27	-1,52	0,131
Soma resid. quadrados		174,3	E. P. da regressão	1,22
R^2		0,0194	R^2 ajustado	0,0112
F(1,117)		2,32	P-valor (F)	0,131

Fonte: Autoria própria (2022).

Já a tabela 6 apresenta os resultados da regressão usando a amostra depois do balanceamento. Diferentemente, do resultado anterior, nesta estimação, o coeficiente relacionado com a variável de tratamento (β) apresentou um p-valor de 0,012, ou seja, este resultado possui uma significância estatística de 98,8%. Em outras palavras, o empreendedorismo é estado-independente. Isto significa que o indivíduo com maior índice de empreendedorismo não dependeria do estado da economia para seguir a sua vocação. Além disso, dado que o coeficiente β teve resultado -0,8 e o coeficiente α 3,68, o índice de empreendedorismo médio para quem

abriu firma em 2019 é 3,68 e para quem abriu firma em 2020 é 2,88 (3,68-0,8), e o ATT (ou ATE) seria -0,8.

Tabela 6 – MQO com o balanceamento. Variável dependente: Ind_Empreendedorismo.

	Coeficiente	Erro Padrão	razão-t	p-valor
constante	3,68	0,22	17,02	0,0
tratamento	-0,8	0,31	-2,62	0,012
Soma resid. quadrados		56,08	E. P. da regressão	1,081
R^2		0,125	R^2 ajustado	0,11
F(1,48)		6,85	P-valor (F)	0,012

Fonte: Autoria própria (2022).

6 CONCLUSÕES

O objetivo deste trabalho era estudar se o empreendedorismo é estado-independente, ou seja, se as condições econômicas resultantes de crises interferem ou não na decisão de abrir uma firma. Pois, espera-se que um empreendedor não dependa do estado da economia. Em outras palavras, os não-empresários não entrariam na atividade empresarial em períodos de não-criese. Para tanto, foi usado o *propensity score* como ferramenta de pareamento entre unidades de tratamento e de controle.

Os resultados obtidos indicam que o pareamento melhorou o balanceamento da amostra. E isso tornou a regressão, que buscava analisar se o tratamento explicava o índice de empreendedorismo, estatisticamente significativa, ou seja, segundo a hipótese deste trabalho, o empreendedorismo é estado-independente. Portanto, o balanceamento dos dados, além de eliminar o viés de seleção, também foi importante em validar a proposta deste trabalho.

Para Schumpeter uma economia sem inovação é descrita como um sistema circular de equilíbrio. Nesta situação, o empresário não seria um "verdadeiro" empreendedor, uma vez que, não exerce qualquer papel relevante. Por outro lado, o agente empreendedor impulsiona uma mudança espontânea e descontínua no sistema anterior. Ou seja, o "verdadeiro" empreendedor é o responsável pelo desenvolvimento econômico. Assim, uma economia, capaz de evitar que não-empresários entrem na atividade empresarial, terá um desenvolvimento econômico maior.

REFERÊNCIAS

- ALSOS, G. A.; KAIKKONEN, V. Opportunities and prior knowledge: a study of experienced entrepreneurs. **Frontiers of Entrepreneurship Research**. Wellesley: Babson College, 2004. Disponível em: < http://www.babson.edu/entrep/fer/FER_2004/web-content/Section%20XIII/P1/XIII-P1.html>. Acesso em: 2 dez. 2021.
- ANGRIST, J. **Course mastering econometrics with Joshua Angrist**, [s.d.]. Disponível em: <<https://mru.org/mastering-econometrics-joshua-angrist>>. Acesso em: 20 de jan. de 2023.
- ANGRIST, J. D.; PISCHKE, J. **Mostly harmless econometrics: an empiricist's companion**. New Jersey: Princeton university press, 2009.
- ARDICHVILI, A.; CARDOZO, R.; RAY, S. A theory of entrepreneurial opportunity identification and development. **Journal of Business Venturing**, v. 18, p. 105-123, 2003.
- AREDE. ARede, 2023. Disponível em: <<https://arede.info/ponta-grossa/454425/pg-e-a-3-cidade-mais-competitiva-do-pais-em-ambiente-de-negocios?d=1>> Acesso em: 25 de fevereiro de 2023.
- BEDÊ, M. A. **Sobrevivência das empresas no Brasil**. Brasília : Sebrae, 2016. Disponível em: <<https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/sobrevivencia-das-empresas-no-brasil-102016.pdf>> Acesso em: 25 de fevereiro de 2022.
- BIRLEY, S; WESTHEAD, P. A comparison of new firms in "assisted"and "non"assisted areas in Great Britain. **Entrepreneurship and Regional Development**, v. 4, n. 4, p. 299-338, 1992.
- CARTER, N. M.; GARTNER, W. B.; SHAVER, K. G.; GATEWOOD, E. J. The career reasons of nascent entrepreneurs. **Journal of Business Venturing**, v. 18, n. 1, p. 13-39, 2003.
- COLE A. **Business enterprise in its social setting**. Cambridge, MA: Harvard University Press, p.27-28, 1959.
- CRUMP, R. K.; HOTZ, V. J.; IMBENS, G. W.; MITNIK, O. A. Dealing with limited overlap in estimation of average treatment effects. **Biometrika**, 96(1), 187-199, 2009.
- DIEESE. DIEESE, 2020. Nota técnica 232. Disponível em: <<https://www.dieese.org.br/notatecnica/2020/notaTec232ProgramaEmergencialGoverno.pdf>>. Acesso em: 20 de fev. de 2022.
- ECKHARDT, J. T.; SHANE, S. A. Opportunities and entrepreneurship. **Journal of Management**, v. 19, n. 3, p. 333-349, 2003.
- FOSTER, J.; LIN, A. Individual differences in learning entrepreneurship and their implications for web-based instruction in e-business and e-commerce. **British Journal of Educational Technology**, v. 34, n. 4, 2003.

GAGLIO, C. M.; KATZ, J. The psychological basis of opportunity identification: entrepreneurial alertness. **Small Business Economics**, v. 16. p. 95-111, 2001.

GOVERNO FEDERAL. Governo Federal, 2020. Disponível em: <<https://www.gov.br/ptbr/noticias/financas-impostos-e-gestao-publica/2020/06/brasil-registra-saldo-de-quase-700-milempresas-abertas-nos-primeiros-quatro-meses-doano>> Acesso em: 25 de fevereiro de 2022.

GOVERNO FEDERAL. Governo Federal, 2022. Disponível em: <<https://www.gov.br/economia/pt-br/aceso-a-informacao/reg/indice-de-concorrencia-dos-municipios-brasileiros-icm>> Acesso em: 15 de fevereiro de 2023.

GUJARATI, D. N. **Econometria básica**. 5. ed. Rio de Janeiro: Mcgraw Hill. 2011.

HO, D. E.; IMAI, K.; KING, G.; STUART, E. A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. **Political analysis**, 15(3), 199-236, 2007.

HORNADAY, J. A. Research about living entrepreneurs. In: KENT, C. A.; SEXTON, D. L.; VESPER, K. H. (Ed.). **Encyclopedia of entrepreneurship**. Englewood Cliffs: Prentice-Hall, 1982. p. 20-38.

IBGE. IBGE, 2023. Disponível em:<<https://cidades.ibge.gov.br/brasil/pr/ponta-grossa/panorama>> Acesso em: 15 de fevereiro de 2023.

KELLEY, Donna J.; SINGER, Slavica; HERRINGTON, Mike. The global entrepreneurship monitor. **2011 Global Report, GEM 2011**, v. 7, p. 2-38, 2012.

KENT, Calvin A.; SEXTON, Donald L.; VESPER, Karl H. Encyclopedia of entrepreneurship. **University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship**, 1982.

KHANDKER, Shahidur R.; KOOLWAL, Gayatri B.; SAMAD, Hussain A. **Handbook on impact evaluation: quantitative methods and practices**. World Bank Publications, 2009.

KIRZNER, I. M. **Perception, opportunity, and profit**. Chicago University Press, 1983.

KOLB, D. A. **Experiential learning: experience as the source of learning and development**. Englewood Cliffs: Prentice Hall, 1984.

LEZANA, Á. G. R.; TONELLI, A. O comportamento do empreendedor. In: DE MORI, F. (Org.). **Empreender: identificando, avaliando e planejando um novo negócio**. Florianópolis: ENE, 2004.

MCCLELLAND, David C.; MAC CLELLAND, David Clarence. **Achieving society**. Simon and Schuster, 1961.

MEMÓRIA, J. M. P. **Breve história da estatística**. Brasília: Embrapa. Texto para discussão (21). 2004.

MINER, J. B. **A psychological typology of successful entrepreneurs**. Westport: Quorum Books, 1997.

MINNITI, M.; BYGRAVE, W. A dynamic model of entrepreneurial learning. **Entrepreneurship Theory and Practice**, v. 25, n. 3, p. 5-15, 2001.

NITAHARA, A. **Desemprego na pandemia continua subindo e chega a 13,7%**. AGENCIA BRASIL, 2020. Disponível em: <<https://agenciabrasil.ebc.com.br/economia/noticia/2020-08/desemprego-na-pandemia-continua-subindo-e-chega-137>>. Acesso em: 20 de jun. de 2022.

PARK, J. S. Opportunity recognition and product innovation in entrepreneurial hitech start-ups: a new perspective and supporting case study. **Technovation**, v. 25, p. 739-752, 2005.

PREFEITURA DE PONTA GROSSA. Prefeitura de Ponta Grossa, 2021. Disponível em:<<https://www.pontagrossa.pr.gov.br/node/47898>> Acesso em: 15 de fevereiro de 2023.

PREFEITURA DE PONTA GROSSA. Prefeitura de Ponta Grossa, 2023. Disponível em:<<https://turismo.pontagrossa.pr.gov.br/>> Acesso em: 15 de fevereiro de 2023.

RIBEIRO, Rosana; CACCIAMALI, Maria Cristina. Impactos do programa bolsa-família sobre os indicadores educacionais. **Revista Economia**, v.13, n.2, maio./ago. 2012.

RONSTADT, R. **Entrepreneurship**: text, cases and notes. Dover, MA: Lord Publishing, 1984.

ROSENBAUM, P. R. **Observational studies**. New York: Springer, 2002.

ROSENBAUM, P. R.; RUBIN, D. B. The bias due to incomplete matching. **Biometrics**, 103-116, 1985.

SAKS, Na. T.; GAGLIO, C. M. Can opportunity identification be taught? **Journal of Enterprising Culture**, v. 10, n. 4, p. 313-358, 2002.

SARAVASTHY, S. D.; DEW, N.; VELAMURI, S. R.. **A testable typology of entrepreneurial opportunity**: extensions of Shane and Venkataraman. College Park: University of Maryland, 2002 (Working Paper).

SARTORIS NETO, A. **Estatística e introdução à econometria**. 2. ed. São Paulo: Editora Saraiva. 2017.

SCHUMPETER, J. A. **Teoria do desenvolvimento econômico**: uma investigação sobre lucros, capital, crédito, juro e ciclo econômico. São Paulo: Abril Cultural, 1982.

SCHUMPETER, J. A. **Essays: on entrepreneurs, innovations, business cycles and the evolution of capitalism**. Routledge, 2017.

SHANE, S.; KOLVEREID, L.; WESTHEAD, P. An exploratory examination of the reasons leading to new firm formation across country and gender. **Journal of Business Venturing**, v. 6, p. 431–446, 1991.

SIPILÄ, Miika. Opportunity recognition in entrepreneurial studies. In: **Seminar in Business Strategy and International Business**. 2006.

SMITH, J. A.; TODD, P. E. Does matching overcome LaLonde's critique of nonexperimental estimators? **Journal of econometrics**, 125(1-2), 305-353, 2005.

STUART, T. E.; SORENSON, O. Social networks and entrepreneurship. In: ALVAREZ, S. A.; AGARWAL, R.; SORENSON, O. (Ed.). **Handbook of entrepreneurship research: disciplinary perspectives**. New York: Springer, 2005. p. 233-252.

TONELLI, A. **Avaliação de competências em curso de especialização a distância via WEB**. Florianópolis, 2003. 211p. Tese (Doutorado em Engenharia de Produção) - Programa de Pós Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2003.

UCBASARAN, D.; WESTHEAD, P. Does entrepreneurial experience influence opportunity identification? **Frontiers of Entrepreneurship Research**. Wellesley: Babson College, 2002.

VEJA. Veja, 2021. Disponível em: <<https://veja.abril.com.br/economia/ibge-desemprego-durante-a-pandemia-foi-maior-que-o-estimado/>> Acesso em: 25 de fevereiro de 2022.

WESTHEAD, P.; WRIGHT, M. Novice, portfolio and serial founders: are they different? **Frontiers of Entrepreneurship Research**. Wellesley: Babson College, 1997. Disponível em: <<http://www.babson.edu/entrep/fer/papers97/westhead/es.htm>>. Acesso em: 2 dec. 2006.

APÊNDICE A - SCRIPT DO R

```

install.packages("MatchIt") library(MatchIt)
Data=read.csv("BaseFinal.csv",sep=";",dec=',')
Data$Escolaridade2<- (Data$Escolaridade)2
Data$Idade2 < -(Data$Idade)2
summary(lm( AbriuFirma2020  Masculino + Escolaridade + Idade + Escolaridade2 + Idade2,
data = Data))
#fazendo o matching modmatch = matchit(AbriuFirma2020  Escolaridade + Idade + Escolari-
dade2 + Idade2, data = Data, method = "optimal", ratio = 1, estimand = "ATT")
#"exact", "cem", "nearest", "optimal", "full", "genetic", "subclass", or "cardinality", or "mahalano-
bis"# ratio é a razão entre o número de observações do (grupo de controle / grupo tratado), 1
significa que para cada elemento do grupo dos tratados, há 1 elemento no grupo de controle.
#vamos verificar quão bom foi o processo de matching, distance é o propensity score sum-
mary(modmatch) plot(modmatch, type = "jitter")
# A primeira parte fornece informação para a base inteira e para a base com matching
#os valores eQQ se referem as diferenças de valores em funções quantílicas empíricas, para
cada uma das covariadas. Basicamente, quão mais próximo de zero, melhor.
# a parte "percent balance improvement"mensura a melhora percentual ao se utilizar a base
com matching em relação a base total
#agora vamos pegar a base que tem apenas as observações com matching baseprop=
match.data(modmatch)
#agora, podemos rodar a regressão com a base com o matching
reg1= lm(IndEmpreendedorismo1  AbriuFirma2020, data = baseprop)
# e com base original reg2= lm(IndEmpreendedorismo1  AbriuFirma2020, data = Data)
summary(reg1)
summary(reg2)

```

APÊNDICE B - PACOTE MATCHIT

O pacote do R usado neste trabalho é o MatchIt desenvolvido por Daniel Ho (<<https://orcid.org/0000-0002-2195-5469>>) e Kosuke Imai (<<https://orcid.org/0000-0002-2748-1022>>), cujo título é Pré-processamento não paramétrico para inferência causal paramétrica (*Nonparametric Preprocessing for Parametric Causal Inference*). Este pacote seleciona amostras combinadas dos grupos tratado e de controle originais com distribuições de covariáveis semelhantes – pode ser usado para corresponder exatamente em covariáveis, em pontuações de propensão ou executar uma variedade de outros procedimentos de correspondência. O pacote também implementa uma série de recomendações oferecidas em Ho, Imai, King e Stuart (2007).

As funções, do pacote MatchIt, usadas neste trabalho foram: `matchit`; e `match.data`. O `matchit` (Correspondência para Inferência Causal) é a principal função do MatchIt e realiza pareamento, seleção de subconjunto e subclassificação com o objetivo de criar grupos de tratamento e controle balanceados nas covariáveis incluídas. MatchIt implementa as sugestões de Ho, Imai, King e Stuart (2007) para melhorar modelos estatísticos paramétricos por meio do pré-processamento de dados com métodos de correspondência não paramétricos. MatchIt implementa uma ampla gama de métodos de correspondência sofisticados, tornando possível reduzir bastante a dependência de inferências causais em suposições de modelagem estatística difíceis de justificar, mas comumente feitas.

O software também se encaixa facilmente nas práticas de pesquisa existentes, pois, após o pré-processamento com o MatchIt, os pesquisadores podem usar qualquer modelo paramétrico que usariam sem o MatchIt, mas produzir inferências com substancialmente mais robustez e menos sensibilidade às suposições de modelagem. Esta página documenta o uso geral de `matchit()`, mas para detalhes de como `matchit()` funciona com métodos de correspondência individuais, consulte as páginas individuais vinculadas na seção Detalhes abaixo.

Já o `match.data` constrói um conjunto de dados correspondente a partir de um objeto `matchit`, ou seja, cria um conjunto de dados com variáveis adicionais para a medida de distância, pesos correspondentes e subclasses após a correspondência. Este conjunto de dados pode ser usado para estimar os efeitos do tratamento após a correspondência ou subclassificação.

O conjunto criado será idêntico ao conjunto de dados fornecido, exceto que várias novas colunas serão adicionadas contendo informações relacionadas à correspondência. Quando `drop.unmatched = TRUE`, o padrão, as unidades com peso zero, que são aquelas unidades que foram descartadas pelo suporte comum ou pelo caliper ou simplesmente não correspondidas, serão descartadas do conjunto de dados, deixando apenas o subconjunto de unidades correspondentes. A ideia é que a saída seja usada como entrada do conjunto de dados para estimar os efeitos do tratamento na amostra correspondente. Assim, é importante incluir os pesos na estimativa do efeito e seu erro padrão.

ANEXO A - BASE DE DADOS

Tabela 7 – Base de dados

Indice	IndEmpreendedorismo	AbriuFirma2020	Idade	Escolaridade
1	5	0	63	4
2	3	1	60	4
3	3	1	58	4
4	5	0	58	8
5	4	1	57	4
6	3	0	57	4
7	1	0	56	10
8	5	0	56	8
9	4	0	56	12
10	5	0	55	8
11	3	0	55	4
12	3	0	55	12
13	3	0	55	8
14	4	1	54	12
15	4	0	54	12
16	4	0	53	12
17	4	0	53	4
18	5	0	52	4
19	2	0	52	4
20	1	0	52	8
21	2	0	51	12
22	3	0	50	4
23	1	0	50	12
24	3	0	50	8
25	3	1	49	4
26	4	0	49	8
27	4	0	48	10
28	3	0	48	8
29	3	1	48	8
30	4	0	47	4
31	3	0	47	4
32	5	0	47	8
33	3	0	46	4
34	3	0	46	8
35	5	1	45	10
36	3	1	45	4
37	3	0	45	8
38	4	0	45	4
39	4	0	44	8
40	3	0	44	8

continuação ...

Indice	IndEmpreendedorismo	AbriuFirma2020	Idade	Escolaridade
41	3	1	44	8
42	0	1	44	8
43	4	0	43	4
44	5	0	43	8
45	2	0	42	12
46	3	1	42	8
47	3	0	42	8
48	3	0	41	10
49	0	1	41	4
50	3	0	41	12
51	1	0	40	8
52	3	0	40	8
53	4	0	40	8
54	4	0	40	8
55	3	1	40	8
56	4	0	40	8
57	4	0	40	10
58	4	0	40	8
59	2	1	39	10
60	4	0	39	4
61	3	0	38	4
62	3	0	38	8
63	2	1	38	12
64	1	0	37	12
65	4	1	36	8
66	4	0	36	12
67	3	1	36	12
68	3	0	36	8
69	3	0	36	4
70	2	1	36	12
71	0	0	35	8
72	4	0	35	8
73	5	0	35	4
74	3	0	34	4
75	3	0	34	4
76	4	0	34	12
77	3	0	33	4
78	0	0	33	8
79	2	0	33	8
80	3	0	33	8

continuação ...

Indice	IndEmpreendedorismo	AbriuFirma2020	Idade	Escolaridade
81	2	0	32	12
82	4	0	32	12
83	1	0	31	10
84	4	0	31	4
85	4	0	31	10
86	1	0	30	12
87	1	1	30	12
88	4	0	30	12
89	5	0	30	10
90	2	0	30	10
91	5	0	29	8
92	5	0	29	8
93	4	0	29	4
94	4	0	29	8
95	4	0	29	8
96	4	0	29	12
97	4	0	28	8
98	3	0	28	8
99	2	0	28	12
100	3	1	27	4
101	3	0	27	8
102	2	0	26	8
103	4	0	26	8
104	3	0	26	8
105	2	0	25	10
106	5	0	24	4
107	4	0	24	8
108	5	0	24	8
109	5	0	24	10
110	3	1	23	10
111	3	0	23	8
112	3	0	23	8
113	1	0	22	10
114	3	0	21	4
115	4	0	21	8
116	5	1	35	8
117	3	1	26	8
118	4	1	23	4
119	3	1	44	4

Fonte: Autoria própria (2022)

ANEXO B - PUBLICAÇÃO NOS ANAIS DO XII CONBREPO

O trabalho

Empreendedorismo Estado-Independente

sob autoria de

Priscilla Costa, Giane Gonçalves Lenzi, Angelo Marcelo Tusset

foi publicado nos anais do XII CONBREPRO.

Realizado de 30 de novembro a 02 de dezembro de 2022.



e1b90346c92331860b1391257a106bb1

Os dados desta certificação são fiéis e autênticos, conforme nossos registros e podem ser verificados no endereço <https://inscricao.aprepro.org.br/index.php/check/index>.
Certificado registrado online no dia 07/12/2022. Para maiores informações entre em contato com o suporte.


Prof. Dr. João Luiz Kovaleski
Presidente da APREPRO


Prof. Dr. Gilberto Zammar
Presidente da Comissão Científica

Realização

Apoio



ANEXO C - SÚMULA CURRICULAR

Súmula Curricular

Nome: Priscilla Maria De Luca Zaupa Costa

1) Formação

Ano	Título ou atividade	Instituição
2001	Graduação	Faculdade de Administração de Empresas de Catanduva
2020	Especialização	Uninter
2023	Pós-Doutoramento	Universidade Tecnológica Federal do Paraná

2) Histórico profissional.

- Lívia Veículos e Peças Ltda – representante comercial, de 1998 a 2004.
- Posto Flamingo – gerente geral, de 2005 a 2011.
- Grupo SHC – analista de crédito, de 2011-2012.

3) Produção

3.1) Artigos completos publicados em periódicos

TUSSET, ANGELO M.; INACIO, DANILO; FUZIKI, MARIA E. K.; COSTA, PRISCILLA M. L. Z.; LENZI, GIANE G. Dynamic Analysis and Control for a Bioreactor in Fractional Order. Symmetry-Basel. , v.14, p.1609 - 1620, 2022.

3.2) Trabalhos publicados em anais de eventos (completo)

COSTA, P. M. L. Z.; LENZI, G. G. Empreendedorismo Estado-Independente In: XII Congresso Nacional de Engenharia de Produção, 2022, online. XII Congresso Nacional de Engenharia de Produção. , 2022.

4) Link Currículo Lattes

<https://lattes.cnpq.br/9727932628247570>

ÍNDICE REMISSIVO

ATE, 27–29, 31, 33, 35, 54

ATT, 23, 27–29, 33, 35, 54

balanceamento do pareamento

 pareamento, 35

covariadas, 10, 21, 26, 28, 31–35, 49, 50, 61

COVID-19, 10

empreendedor, 10, 19, 20, 23, 24

empreendedores, 11, 16, 19, 20

empreendedorismo, 10, 12, 14, 19, 27

 empreendedorismo por necessidade,
 12

 sala do empreendedor, 14

 índice de empreendedorismo, 23, 53, 55

grupo de controle, 11, 21, 22, 26, 28, 29, 50

grupo de tratamento, 11, 21, 22, 26, 49, 50

grupo tratado, 21, 28, 29

pareamento, 12, 30–33, 35, 49

propensity score, 12, 21, 23, 31–33, 49, 50,
55

Propensity Score Matching, 10, 12, 19, 21,
22, 30–33

Schumpeter, 12, 19, 55

viés de seleção, 12, 21–23, 31, 32, 49, 51,
55