

CEA – USP – RAE 18P03

**RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO “AGLOMERAÇÃO E
DISPERSÃO DAS REDES DE *FAST FOOD* NO BRASIL”**

VIVIANA GIAMPAOLI

PIERO CONTI KAUFFMANN

WILSON DOS ANJOS

São Paulo, junho de 2018

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP**RELATÓRIO DE ANÁLISE ESTATÍSTICA**

TÍTULO: Relatório de análise estatística do projeto: “Aglomeração e dispersão de redes de *fast-food* no Brasil”

PESQUISADORA: Adriana Beatriz Madeira

INSTITUIÇÃO: Faculdade de Economia - Universidade Presbiteriana Mackenzie

FINALIDADE: Artigo Científico

RESPONSÁVEIS PELA ANÁLISE: Viviana Giampaoli

Piero Conti Kauffmann

Wilson dos Anjos

REFERÊNCIA DESTE TRABALHO:

GIAMPAOLI, V.; KAUFFMANN, P. e ANJOS, W., **Relatório de análise estatística sobre o do projeto: “Aglomeração e dispersão de redes de *fast-food* no Brasil”**. São Paulo, IME-USP, 2018. (RAE-CEA-18P03).

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS

1. EAE BUSINESS SCHOOL (2014). Fast food consumption in Spain will rise by 50% over the next five years (2014). Disponível em: <https://en.eae.es/news/fast-food-consumption-in-spain-will-rise-by-50-over-the-next-five-years>. Acesso em: 21.05.2018. Uniformizar!
2. MADEIRA, A. B.; GIAMPAOLI, V., (2017), Agglomeration of fast food companies in Brazil. **Management Research: Journal of the Iberoamerican Academy of Management**.
3. MARSHAL A. (1920), **Principles of Economy**, McMillan, London.
4. ALVES, L. G.; RIBEIRO, H. V.; RODRIGUES, F. A., (2017), The role of city size and urban metrics on crime modeling
5. KAUFFMAN, D.; KRAAY, A.; ZOIDO-LOBATÓN, P. (1999), Governance Matters, Policy Research Working Paper 2196, **The World Bank Development Research Group Macroeconomics and Growth and World Bank Institute Governance, Regulation, and Finance**.
6. KAUFFMAN, D.; KRAAY, A.; ZOIDO-LOBATÓN, P. (2002), Governance Matters II: updated indicators for 2000-01, Policy Research Working Paper 2772, **The World Bank Development Research Group and World Bank Institute Governance, Regulation, and Finance**.
7. GLAESER, E. L.; LA PORTA, R.; LOPEZ-DE-SILANES, F.; SHLEIFER, A. (2004), Do institutions cause growth?, **Journal of Economic Growth**, Vol. 9, No. 3, pp. 271-303.
8. ROBLES, F. (2015), **Agglomeration and institutional determinants of franchised fast food networks in Latin America**, 9th Iberoamerican Academy of Management Conference, From Developing to Developed Economies: the future of entrepreneurship and innovation in Iberoamerica, Universidad del Desarrollo, Santiago, Chile, December 3-5.
9. DAS, K.; QUIRK, T. (2016), Which institutions promote growth? Revisiting evidence, **Economic Papers: A journal of Applied Economics and Policy**, Vol. 35 No. 1, pp. 37-58.

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Microsoft Excel for Windows (versão 2007)

Microsoft Word for Windows (versão 2007)

R 4.3

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Outros (07:990)

ÁREA DE APLICAÇÃO

Econometria (14:070)

SUMÁRIO

Resumo	6
1.Introdução	7
1.1 Objetivo	9
2.Descrição do estudo	9
3.Descrição das variáveis de interesse	9
4.Análise descritiva	14
4.1.Análise em nível de cidade	14
4.1.1.Análise de valores faltantes	14
4.1.2.Análise univariada	16
4.1.3.Análise bivariada	17
4.2.Análise em nível de loja	20
4.2.1.Análise univariada	21
4.2.2.Análise bivariada	21
5.Análise inferencial	22
5.1 Modelo para o Índice de Herfindahl de uma cidade	23
5.1.1 Ajuste do modelo para o Índice de Herfindahl Nacional	24
5.1.2 Ajuste do modelo para o Índice de Herfindahl Estrangeiro	25
5.2 Modelo para o total de lojas de uma empresa em uma cidade	26
5.2.1 Ajuste do modelo para empresas nacionais	27
5.2.2 Ajuste do modelo para empresas estrangeiras	28
6.Conclusões	30
Apêndice A: Análise descritiva unidimensional em nível de cidade	31
Apêndice B: Análise descritiva multidimensional em nível de cidade	37
Apêndice C: Análise descritiva em nível de loja	43
Apêndice D: Gráficos de resíduos	49
Apêndice E: Distribuições de probabilidade	52
Apêndice F: Modelos	53

Resumo

Redes de comida rápida (*fastfood*) são principalmente caracterizadas pela rápida distribuição de seus produtos e alto nível de sistematização do processo de produção. Essa padronização é devida ao chamado “sistema de franquias” de uma rede de *fastfood*, na qual determina-se um acordo entre o produtor e um investidor independente. Neste, o produtor cede direitos da marca ao empresário para conduzir o negócio de uma maneira específica. Por diversos fatores logísticos, essas redes tendem a se aglomerar geograficamente, e esse fenômeno de aglomeração, no caso das indústrias, tem sido amplamente estudado por economistas ao redor do mundo. No Brasil, o quarto país que mais consome *fastfood* no mundo, observa-se uma aglomeração bem concentrada nas capitais do país. Neste estudo, avaliamos a dispersão geográfica de 7653 lojas, de 103 marcas, presentes em 542 cidades brasileiras em 2015. O objetivo do estudo é avaliar quais fatores referentes aos municípios presentes no estudo, favoreceram no processo de aglomeração no país. Os resultados principais desta análise apontaram que a quantidade de homicídios e a população do município tem impacto significativo na aglomeração de lojas estrangeiras e nacionais nos municípios.

1. Introdução

O Brasil é o quarto país que mais consome comidas rápidas (*fastfood*) no mundo, diz o estudo da EAE Business Scholl (2016) [1]. O mesmo estudo indica que o consumo no país, em 2014, foi de R\$ 265,00 por habitante, e que este deve crescer 30,88% até 2019. No entanto, conforme visto neste trabalho, essas franquias não se distribuem uniformemente pelo território nacional. De fato, através de 7653 unidades de redes de *fastfood* aqui analisadas, observa-se o fenômeno conhecido por aglomeração geográfica. Isto motiva um estudo a cerca dos fatores referentes a municípios, que favorecem a instalação dessas unidades de franquias de *fastfood* em seus territórios.

As redes de *fastfoods* são principalmente caracterizadas pela rápida preparação e entrega de seus produtos. Isto é devido ao alto nível de padronização e mecanização do processo de produção. Em grande maioria, essas redes formam parte de um sistema denominado franquia (*franchises*), em que a franquia é um acordo entre o produtor e um investidor independente, no qual o produtor cede os direitos da marca ao empresário para conduzir o negócio de uma maneira específica com um prazo de retorno de investimentos. Atualmente esses contratos incluem todo o plano de negócio como estratégias de marketing, sistema de controle de qualidade, entre outros. Segundo o relatório de Desempenho da Associação Brasileira de Franchising (ABF), o setor alimentício foi responsável por 20% do faturamento total neste tipo de empresa em 2015. O mesmo relatório indica a geração de 90.000 empregos neste período, o que representa um crescimento de 8,9% no setor.

Este projeto visa dar continuidade a um estudo anterior de Madeira e Giampaoli (2017) no qual modelos Beta inflacionados mistos foram ajustados para se encontrar uma relação entre as variáveis analisadas e o índice de Herfindahl Normalizado, um indicador de aglomeração bastante consolidado na literatura de medidas econômicas. Vide Marshal (1920) e estudos de aglomeração industrial como Kaufmann et al. (1999, 2002), Glaeser et al. (2004), Robles (2015) e Das e Quirk (2016).

Visto o caráter retrospectivo do estudo, visa-se a entender quais características referentes às cidades contempladas explicam o processo de aglomeração e dispersão

geográficas das redes de *fastfood* no Brasil. As economias de aglomeração podem ser definidas como os ganhos econômicos advindos da concentração geográfica das atividades produtivas. Tais ganhos podem se manifestar de diferentes formas: por meio da difusão local do conhecimento, da redução dos custos logísticos, do surgimento de atividades complementares, do adensamento do mercado de trabalho, entre outros.

Uma forma de se avaliar o grau de aglomeração de uma cidade é calculando o índice de Herfindahl como em Madeira e Giampaoli (2017), a partir da fórmula:

$$H_r = \sum_{i=1}^{N_r} (S_{ri})^2.$$

em que:

$$S_{ri} = \frac{E_{ri}}{\sum_{k=1}^N E_{rk}}$$

$$E_{ri} = \text{Total de lojas da marca } i \text{ na cidade } r,$$

$$TRE_r = \sum_{i=1}^{N_r} E_{ri} = \text{Número total de lojas na cidade } r,$$

$$N_r = \text{Número de marcas na cidade } r.$$

$$N = \text{Número total de lojas}$$

H_r é um valor entre $1/N_r$ e 1, em que, um valor próximo de um representa uma cidade na qual uma única marca domina o mercado. Valores próximos de zero representam cidades onde as marcas estão igualmente distribuídas em termos de quantidade de lojas.

Este projeto visa avaliar a capacidade de explicação do Índice de Herfindahl como variável resposta, assim como outros métodos que serão mencionados posteriormente.

1.1 Objetivo

O objetivo deste trabalho é avaliar o efeito de características institucionais e populacionais das cidades no processo de aglomeração de redes de *fastfood* ao redor do país. O estudo contribui com o conhecimento do mercado de *fastfood* brasileiro, assim como revela as condições socioeconômicas e de infraestrutura que favorecem o crescimento do setor no país.

2. Descrição do estudo

Esse estudo analisou a dispersão geográfica de 103 redes de *fastfood*, 94 nacionais e 9 multinacionais, distribuídas em 542 cidades brasileiras. Os dados das marcas foram fornecidos pela Associação Brasileira de *Franchising* (ABF), em que foram disponibilizados o total de lojas, tanto franquizadas quanto próprias, em todas as cidades nas quais as marcas estavam presentes no ano de 2015. Desta maneira os pesquisadores buscaram de diversas fontes, variáveis referentes às cidades que já eram habitualmente utilizadas na literatura de aglomeração industrial e econômica. Dentre elas, variáveis relacionadas a índices de segurança, infraestrutura, educação, renda, condições ambientais e governança das cidades.

3. Descrição das variáveis de interesse

Tomaremos as seguintes variáveis como resposta no ajuste dos modelos:

Variáveis resposta

- ÍNDICE DE HERFINDAHL:
 - ÍNDICE DE HERFINDAHL GERAL: valor do índice para cada cidade;
 - ÍNDICE DE HERFINDAHL NACIONAL: valor do índice utilizando-se apenas empresas nacionais para cada cidade;

- ÍNDICE DE HERFINDAHL ESTRANGEIRO: valor do índice utilizando-se apenas empresas estrangeiras para cada cidade;
- TOTAL DE LOJAS: total de lojas de cada uma das marcas em cada cidade;
- TOTAL DE LOJAS NA CIDADE: total de lojas de todas as marcas na cidade;

Variáveis explicativas quantitativas

- IDHM – RENDA: indicador da renda municipal per capita, isto é, a soma da renda de todos os residentes dividida pelo número de pessoas que moram no município – inclusive crianças e pessoas sem registro de renda;
- IDHM – LONGEVIDADE: medida pela expectativa de vida ao nascer, calculado por método indireto, a partir dos dados dos Censos Demográficos do IBGE. Ele indica o número médio de anos que uma pessoa de determinado município vive;
- IDHM – EDUCAÇÃO: medido por meio de dois indicadores. A escolaridade da população adulta é medida pelo percentual de pessoas de 18 anos ou mais de idade com ensino fundamental completo – peso 1. É medido também o fluxo escolar da população jovem que equivale a média aritmética do percentual de crianças de 5 a 6 anos frequentando a escola, do percentual de jovens de 11 a 13 anos frequentando os anos finais do ensino fundamental, do percentual de jovens de 15 a 17 anos com ensino fundamental completo e do percentual de jovens de 18 a 20 anos com ensino médio completo – peso 2 (elevado ao quadrado). Os dois índices estão entre 0 e 1, dividindo-se a porcentagem encontrada por 100.
- IDHM – Geral: Média geométrica dos índices IDHM – RENDA, IDHM – LONGEVIDADE e IDHM – EDUCAÇÃO

- MATRICULAS PROFISSIONAIS TÉCNICAS: quantidade de matrículas profissionais técnicas em 2010, de cada cidade.

Fonte: O Atlas do Desenvolvimento Humano no Brasil,
<http://www.atlasbrasil.org.br/>

- TAXA DE HOMICÍDIOS: número de homicídios dividido pela população, multiplicado por 100.000 (taxa por cem mil habitantes), considerando o ano de 2011.
- NUM DE HOMICÍDIOS: número absoluto de homicídios, considerando o ano de 2011.

Fonte: Segurança Pública SUMÁRIO – 9º ANUÁRIO BRASILEIRO DE SEGURANÇA PÚBLICA,
<http://www.forumseguranca.org.br/produtos/anuario-brasileiro-de-seguranca-publica/9o-anuario-brasileiro-de-seguranca-publica>

- TAXA BENEFICIADOS PAT: porcentagem de trabalhadores beneficiados pelo Programa de Alimentação do Trabalhador (PAT) – serviços próprios de alimentação, distribuição de alimentos, inclusive não preparados (cesta básica), cartões eletrônicos que permitem a aquisição de refeições ou produtos de gênero alimentício em estabelecimentos comerciais – em relação à estimativa populacional de 2015;
- NUM. DE BENEFICIADOS PAT: número absoluto de trabalhadores beneficiados pelo Programa de Alimentação do Trabalhador (PAT) – no ano de 2015;

Fonte: Ministério do Trabalho e Emprego,
<http://consulta.mte.gov.br/atlas/pat.asp>

- TOTAL DE HOTÉIS: quantidade de hotéis (inclusive hotéis históricos, resorts e hotéis-fazenda) por município no ano de 2011;

Fonte: Pesquisa de Serviços e Hospedagem – IBGE,
<http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=p&o=15&i=P&c=3440;>

- PIB MUNICIPAL: valor do Produto Interno Bruto (PIB) de cada município em unidades de R\$1.000,00, no ano de 2010.

Fonte: Produto Interno Bruto dos Municípios – IBGE, <http://www.ibge.gov.br/home/estatistica/economia/pibmunicipios/2010>;

- POPULAÇÃO 2010: População municipal no ano de 2010;
- POPULAÇÃO 2015: Estimativa populacional municipal no ano de 2015;

Fonte: Estimativa da População 2010 e 2015 – IBGE, <http://www.ibge.gov.br/home/estatistica/populacao/>

Variáveis explicativas qualitativas

- CIDADE: Cidade contemplada no estudo;
- ESTADO: Estado da federação de cada cidade;
- MARCA: Redes de *fastfood* contempladas no estudo;

Fonte: Associação Brasileira de *Franchising* (ABF);

- CIDADE DE ORIGEM: Variável binária assumindo valor 1 para marcas de origem estrangeira e 0 para marcas de origem nacional;
- INDICADOR DE ORIGEM DA MARCA: Variável binária assumindo valor 1 para marcas de origem estrangeira e 0 para marcas de origem nacional;

Fonte: Ambas as informações referentes às marcas foram obtidas de seus respectivos websites.

Variáveis explicativas adicionadas posteriormente

Todas as variáveis descritas até aqui foram obtidas previamente a esse estudo. Houve-se, no entanto a possibilidade de incluir algumas variáveis que pudessem ajudar na análise estatística e no entendimento do problema, uma vez observada a alta correlação entre as cidades de origem das marcas e suas expansões geográficas. São estas:

- LATITUDE DO MUNICÍPIO: Latitude do centroide do município

- LONGITUDE DO MUNICÍPIO: Longitude do centroide do município

Fonte: API do Google Maps de Georeferenciamento – Versão 3.33, <https://developers.google.com/maps/?hl=pt-br>

Definindo-se o centroide do município como a média aritmética das coordenadas do polígono que delimita a região administrativa do município. Este representa geograficamente o local central do município.

4. Análisedescritiva

A análise descritiva foi dividida em duas partes principais; a primeira delas se refere à análise descritiva das variáveis medidas em nível de cidade, que tem como objetivo compreender melhor a estrutura de inter-relações entre as variáveis explicativas para uma melhor caracterização das cidades. A segunda parte da análise, que avalia os dados em nível de loja/marca, busca descrever o comportamento comum das empresas, sob aspectos geográficos, e as suas diferenças e particularidades.

4.1. Análise em nível de cidade

4.1.1. Análise de valores faltantes

Como análise preliminar realizou-se a contagem do número de cidades em que não se observou alguma das variáveis de interesse, os resultados estão apresentados na Tabela 1. Variáveis sem valores faltantes – *missingvalues* – não constam na tabela.

Tabela 1: Contagem de valores não disponíveis em cada variável.

Variável	Número de observações disponíveis	Número de observações faltantes	Proporção de valores faltantes (%)
ÍNDICE DE HERFINDAHL GERAL	534	8	1,48
ÍNDICE DE HERFINDAHL ESTRANGEIRO	503	39	7,20
ÍNDICE DE HERFINDAHL NACIONAL	319	223	41,14
NUM. DE BENEFICIADOS PAT	482	60	11,07
TOTAL DE HOTÉIS	139	403	74,35

Da Tabela 1, verifica-se que apenas cinco variáveis possuem algum valor faltante na base de dados. As variáveis referentes ao Índice de Herfindahl apresentam valores faltantes devido à ocorrência de cidades em que não existe nenhuma loja instalada por qualquer uma das redes de *fastfood* em algumas das categorias de interesse (geral, estrangeiro ou nacional).

No que diz respeito aos valores ausentes do Índice de Herfindahl Geral, a presença de cidades sem nenhuma loja ocorre devido ao fato de serem apenas cidades de origem de fundação de algumas das redes de *fastfood* estudadas.

A variável “Total de Hotéis” é a variável com menos valores disponíveis, presente em apenas 139 das 542 cidades. A Figura 1 apresenta os *boxplots* da Estimativa da População (2015, em escala logarítmica base 10) segundo a presença ou não das variáveis “Total de Hotéis” e “Número beneficiados PAT”. Nesta figura pode se verificar que as cidades com valores indisponíveis em alguma destas duas variáveis possuem estimativa populacional (para o ano 2015) mediana inferior às cidades com valores disponíveis, evidenciando que a indisponibilidade destas informações está relacionada ao tamanho da cidade.

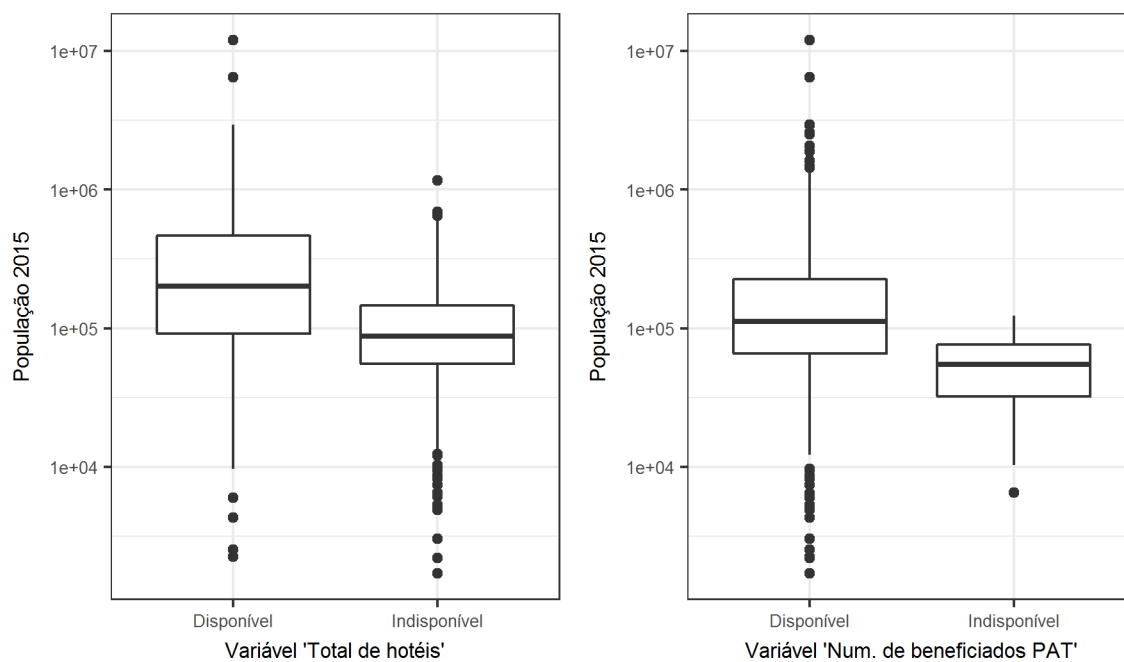


Figura 1: Box plot da Estimativa da População (2015, em escala logarítmica base 10) segundo a presença ou não das variáveis “Total de Hotéis” e “Num. Beneficiados PAT”.

4.1.2. Análise univariada

A análise univariada estudou inicialmente as variáveis resposta: “Total de lojas na cidade” e “Índice de Herfindahl Geral”. Na Figura A.1 e Tabela A.1 nota-se que a distribuição da variável “Total de lojas na cidade” possui comportamento altamente assimétrico à direita, com um valor máximo de 1.218 lojas e uma mediana de apenas 2 lojas. Isto é, há uma alta frequência de cidades com poucas lojas e baixa frequência de cidades com muitas lojas. A transformação logarítmica $\log_{10}(\text{Total de lojas na cidade} + 1)$ facilitou a visualização da distribuição dos dados (Figura A.1, segundo painel), sendo necessário o acréscimo de uma unidade na variável para evitar indefinição matemática da função logaritmo no ponto zero.

A Figura A.2 apresenta o histograma do Índice de Herfindahl (Geral, Nacional e Estrangeiro). Nota-se que para os três índices a faixa de valores $[0,9; 1,0]$ possui a maior frequência. Da Tabela A.2 verifica-se que há cidades em que o Índice de Herfindahl assume o máximo de 1, o que indica a presença de apenas uma marca ou rede de *fastfood* na cidade. Os maiores valores de média e mediana foram observados no caso do Índice de Herfindahl estrangeiro, o qual também assumiu um valor menor de desvio padrão em relação ao Índice de Herfindahl nacional.

Em seguida, são estudadas as distribuições univariadas das variáveis socioeconômicas IDHM (Figura A.3 e Tabela A.3), População (2010 e 2015) (Figura A.4 e Tabela A.4) e PIB Municipal (Figura A.5).

Os Índices IDHM, e suas dimensões Educação, Renda e Longevidade variam numa escala de 0 a 1, sendo que quanto maior mais desenvolvido é o município. Verifica-se que a distribuição dos Índices IDHM é bem concentrada, visto que as observações centrais – entre o primeiro e terceiro quartil – estão contidas em uma faixa de valores cuja diferença não excede 0,1 (Figura A.3) sendo que os valores mais altos correspondem ao IDHM-Longevidade seguido de IDHM-Renda. Porém, os desvios padrões das três dimensões são similares (Tabela A.4). A presença de observações extremas ou *outliers* comuns para um conjunto de dois ou mais índices também é observada, por exemplo, a cidade Tobias Barreto (SE) é simultaneamente um

outlier para os Índices IDHM-Geral, IDHM-Educação e IDHM-Longevidade; observa-se que nesta cidade foram observadas os valores mínimos destes índices.

A análise univariada das variáveis “População 2010” e “População 2015” mostrou que a distribuição de ambas é muito parecida, como esperado.

Além disso, notou-se o caráter simétrico do logaritmo das variáveis População e PIB Municipal, indicando a possibilidade destas variáveis serem distribuídas segundo uma distribuição log-normal.

Da Tabela A.5, verifica-se que a distribuição das variáveis “Taxa de homicídios”, “Matrícula profissional técnica” e “Taxa beneficiados pelo Programa de Auxílio ao Trabalhador (PAT)” são altamente assimétricas. Porém, a Figura A.6 indica que a transformação logarítmica destas variáveis produz uma distribuição aproximadamente simétrica e relativamente similar a uma distribuição normal. Verifica-se também que existe uma porcentagem considerável de cidades (14,39%) com nenhuma matrícula profissional técnica, que indica as cidades que podem não oferecer mão de obra técnica.

4.1.3. Análise bivariada

A análise bivariada em nível de cidade teve como um dos objetivos principais avaliara relação entre as variáveis de interesse e o tamanho populacional. Em particular, um tópico relevante foi obter uma forma de padronizar as variáveis altamente correlacionadas ao tamanho da população (como Número de homicídios ou PIB) para permitir uma comparação entre cidades de tamanhos populacionais muito distintos, sem o viés da variável populacional.

Como já discutido em Alves et al. (2017), o comportamento entre variáveis econômicas, criminais e indicadores de educação frequentemente segue uma relação de proporcionalidade descrita por:

$$y \propto C \cdot \text{População}^{\beta} \quad (1)$$

Para verificar isso, foram feitos os gráficos de dispersão (com escala logarítmica nos dois eixos) presentes nas Figuras B.1 a B.7, apresentados no Apêndice B.

Considerando o logaritmo na relação (1) fica estabelecida a relação linear $\log(y) \propto \log(C) + \beta \log(\text{População})$, a qual esclarece as relações que serão posteriormente formuladas.

Na Figura B.1, observa-se um comportamento linear entre o logaritmo da variável População 2010 e o logaritmo do PIB Municipal 2010. Um comportamento parecido é observado nas Figuras B.2 e B.3, que mostram a relação entre o tamanho populacional e as variáveis “Número de homicídios” e “Matriculas profissional técnica”. A seguinte padronização para estas variáveis foi adotada com o objetivo de se retirar o possível efeito de multicolinearidade no ajuste dos modelos:

$$y' = \frac{\log(y)}{\log(\text{População})}, \text{ ou } y'' = \frac{\log(y+1)}{\log(\text{População})} \quad (2)$$

A transformação y' ou y'' (para o caso em que y assume valores nulos) permite o cálculo de correlações entre variáveis, sem que esta seja excessivamente afetada pelo efeito de confundimento do tamanho populacional de cada cidade.

Vale notar que algumas das variáveis socioeconômicas foram medidas em anos cuja população não está disponível no conjunto de dados, para essas variáveis adotou-se a padronização usando a estimativa populacional mais próxima do período. Por exemplo, para a variável “Número de Homicídios” (referente ao ano de 2011), o denominador da equação (2) utilizado foi a variável “População 2010”.

As Figuras B.6 e B.7 mostram que parece existir uma relação não linear entre os logaritmos das variáveis e a População. Verificou-se alta correlação entre as variáveis resposta medidas de aglomeração: Índice de Herfindahl e Total de lojas na cidade. Diferentemente das variáveis medidas anteriormente, a relação aparenta ser constante para cidades com menos de 100 mil habitantes, exibindo um padrão de crescimento linear apenas para cidades com mais de 100 mil habitantes ou mais (Figuras B.6 e B.7).

Nesta análise, o mesmo padrão de não linearidade foi observado na variável Total de Hotéis e Taxa de beneficiados pelo Programa de Auxílio ao Trabalhador (PAT), presentes no Apêndice B (Figuras B.4 e B.5).

No final do Apêndice B as correlações das variáveis explicativas e respostas estão apresentadas na matriz de correlações de *Pearson* (Figura B.8). Também foram avaliadas as correlações de *Pearson* após a transformação de variáveis proposta (Figura B.9 e Equação (2) de 4.1.3). Uma análise comparativa entre as Figuras B.8 e B.9 sugere que parte da correlação observada entre as variáveis resposta e as variáveis explicativas era devido ao efeito de confundimento com o tamanho populacional.

4.2. Análise em nível de loja

Para compreender melhor os comportamentos tanto individuais para cada empresa como comuns a todas as redes, uma análise foi feita sob nível de loja, com foco nas variáveis geográficas e referentes à origem, disponíveis para cada empresa. Para isso, tomou-se o subgrupo de 94 redes de *fastfood* nacionais, visto que apenas estas contam com informações referentes ao estado de origem de fundação da rede, que permitem as análises espaciais seguintes. Os nomes das empresas foram codificados com um identificador próprio e as empresas foram divididas entre estrangeiras e nacionais, como consta na Tabela 2.

Tabela 2: Listagem das redes por categoria (Nacional ou Estrangeira) com contagem de redes em cada categoria.

Estrangeiras		Nacionais
Identificador de Rede	A26, A27, A44, A52, A58, A65, A67, A84, A92	A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, A12, A13, A14, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25, A28, A29, A30, A31, A32, A33, A34, A35, A36, A37, A38, A39, A40, A41, A42, A43, A45, A46, A47, A48, A49, A50, A51, A53, A54, A55, A56, A57, A59, A60, A61, A62, A63, A64, A66, A68, A69, A70, A71, A72, A73, A74, A75, A76, A77, A78, A79, A80, A81, A82, A83, A85, A86, A87, A88, A89, A90, A91, A93, A94, A95, A96, A97, A98, A99, A100, A101, A102, A103
Número de redes	9	94

4.2.1. Análise univariada

A Figura C.1, apresentada no Apêndice C, representa a proporção de cidades contempladas por uma rede que pertencem ao estado de origem de fundação da rede; na figura nota-se que grande parte (23 de 94) das redes de *fastfood* analisadas não deixam o estado de origem. Isso naturalmente sugere um questionamento se este fenômeno está diretamente relacionado às redes com poucas lojas. Porém, a Figura C.3 indica que não parece haver relação direta entre esse questionamento.

Na Figura C.2 é visível que a distribuição do número total de lojas por empresa possui comportamento altamente assimétrico, em que cerca de 16% das empresas possui no total 5 lojas ou menos. Conclusão similar pode ser obtida da observação da Figura C.3 em que se representa o Número total de lojas instaladas por cada empresa nacional versus a proporção de lojas no estado de origem da rede.

A distribuição da distância calculada entre os centroides das cidades contempladas por uma empresa e a cidade de origem da mesma empresa, está representada na Figura C.4, indicando que o fator proximidade à cidade de origem parece ser importante para uma empresa se instalar em uma cidade.

4.2.2. Análise bivariada

Na análise bivariada, foi possível observar o padrão particular de proximidades para cada empresa, avaliando-se a distribuição da variável “Distância da cidade de origem” por empresa (Figura C.5). A análise deste gráfico revela empresas que não possuem lojas nas suas cidades de origem respectivas, como por exemplo a empresa A49. O gráfico reforça a noção de que redes diferentes possuem comportamentos muito diferentes entre si em relação à distribuição de lojas, causados provavelmente por uma série de fatores econômicos (como recursos, capital disponível e parcerias econômicas), fatores administrativos (como postura da empresa diante ao mercado, apetite ao risco e capacidade de atração de parceiros comerciais), e fatores geográficos (como proximidade a metrópoles e fatores ambientais).

Com o intuito de estudar melhor a dispersão das lojas de cada rede, o gráfico presente na Figura C.6 foi construído, permitindo a visualização, no mapa do Brasil, da aglomeração e dispersão geográfica de cada uma das marcas. Vale destacar a presença de algumas marcas em regiões específicas do país, como Norte e Nordeste, sem diversificação. Outro fato que pode ser visualizado nestes mapas é que na distribuição de lojas as marcas parecem ter tido preferências pelas cidades do litoral do país.

Uma outra linha de análise compara a relação entre o total de lojas de cada empresa com a quantidade de anos de atividade da empresa no mercado (Figura C.7). É possível verificar a existência de uma correlação positiva entre estas duas variáveis, ainda que sutil.

A Figura C.8 indica correlação fraca ou inexistente entre o alcance, definido como a distância máxima presente entre uma cidade instalada e a cidade de origem da empresa. É possível sugerir então que o processo de expansão de redes de *fastfood* não seja relacionado diretamente com a quantidade de anos de atividade. Porém, não é possível testar a validade desta conjectura, pois não se tem informação do ano de instalação de cada loja e nem dos fatores macroeconômicos e sociais no período, que permitiriam esta análise.

5. Análise Inferencial

A etapa de análise inferencial foi dividida em dois itens principais que correspondem a duas abordagens diferentes de modelagem estatística para estudar o problema de aglomeração de redes de *fastfood*.

A primeira abordagem tem como objetivo quantificar a relação da aglomeração, medida por meio do Índice de Herfindahl, com as variáveis explicativas de cada uma das cidades.

O segundo item de análise estuda a relação entre o número total de lojas de cada uma das empresas em cada uma das cidades com as variáveis explicativas disponíveis tanto das empresas como das cidades.

5.1 Modelo para o Índice de Herfindahl

Para esta finalidade, foram ajustados modelos lineares aos dados com o objetivo de explicar a variável referente ao Índice de Herfindahl de uma cidade com as variáveis socioeconômicas relativas a essa cidade. Levando em conta as diferenças observadas entre as distribuições dos Índices de Herfindahl Estrangeiro e Nacional (Figura A.2), adotou-se a utilização de dois modelos de regressão para analisar estas duas variáveis respostas separadamente.

Foi considerado um modelo aditivo generalizado de escala e forma (*gamlssGeneralizedAdditiveModels for LocationScaleandShape*) considerando que a resposta tem distribuição Beta Inflacionada em 1 (descrita no Apêndice E.1) para os Índices de Herfindahl, devido ao fato do índice de Herfindahl poder assumir valores no intervalo (0,1].

Na Tabela 8, são listadas as variáveis explicativas utilizadas em cada um dos modelos.

Além das variáveis presentes na base de dados, a variável “Distância mediana à origem” foi calculada para cada cidade pelo seguinte procedimento: I. São identificadas todas as empresas presentes na cidade II. São calculadas as distâncias entre a cidade e todas as cidades de origem das empresas presentes na cidade. III. Calcula-se o valor da mediana desse conjunto de distâncias.

Tabela 3: Variáveis explicativas dos dois modelos considerados. As variáveis indicadas com (*) foram padronizadas segundo o procedimento descrito na seção 4.1.3 equação (2). A terceira coluna indica a abreviação utilizada nas tabelas de ajuste dos modelos.

	Modelo para o Índice de Herfindahl Nacional	Modelo para o Índice de Herfindahl Estrangeiro	Abreviação
Variáveis explicativas	log10(População 2015)	log10(População 2015)	Lpop
	PIB Municipal(*)	PIB Municipal(*)	PIBp
	IDHM Geral	IDHM Geral	IDHM
	Número de homicídios (*)	Número de homicídios (*)	HOMp
	Número de matrículas profissionais técnicas (*)	Número de matrículas profissionais técnicas (*)	MATRp
	Número de trabalhadores beneficiados PAT (*)	Número de trabalhadores beneficiados PAT (*)	TRABp
	Distância Mediana à origem	-	distmd

A variável “Distância Mediana à origem” foi calculada apenas para as empresas nacionais, pelo fato de apenas estas possuírem informações referentes à cidade de origem.

5.1.1 Ajuste do modelo para o Índice de Herfindahl Nacional

O modelo definido para o Índice de Herfindahl Nacional segue descrito no Modelo F.1, apresentado no Apêndice F. O ajuste deste modelo de regressão com todas as variáveis explicativas, presentes na Tabela 3, foi realizado com auxílio da biblioteca *gamlss* do pacote estatístico R. Após a aplicação do Algoritmo F.1 de seleção de variáveis, foi obtido um modelo reduzido, cujo ajuste segue apresentado em sequência na Tabela 4.

Os testes de Wald de inexistência de efeito, efetuados pelo Algoritmo F.1, indicaram que apenas os coeficientes das variáveis Lpop, IDHM, HOMp, TRABp são estatisticamente significantes adotando-se um nível de significância de 5%.

Tabela 4: Ajuste do modelo reduzido para o Índice de Herfindahl Nacional. São representadas as estimativas dos coeficientes, assim como seus erros padrões e os respectivos Valores P do teste de Wald de inexistência dos efeitos. As estimativas dos parâmetros adicionais da distribuição Beta Inflacionada em 1 para o modelo são $\hat{\sigma} = 0,25$ e $\hat{p}_1 = 0,29$.

Variável abreviada	Estimativa do Coeficiente	Erro padrão	Valor P
(Intercepto)	5,40	0,86	< 0,01
Lpop	-0,71	0,10	< 0,01
IDHM	-3,11	1,09	<0,01
HOMp	0,40	0,16	0,01
TRABp	-1,58	0,32	< 0,01

Os coeficientes obtidos indicam que acréscimos sofridos nos termos Lpop, IDHM e TRABp causam diminuição no Índice de Herfindahl Nacional, isto é, um índice de aglomeração maior, como era esperado. Outro resultado que atendeu às expectativas foi o coeficiente da variável HOMp, que obteve valor positivo e estatisticamente significativo, indicando que um aumento do número de homicídios padronizado implica num aumento da média do índice de Herfindahl Nacional.

A análise de diagnóstico dos resíduos do modelo, conforme pode ser vista na Figura D.1 do Apêndice D, indicaram que a distribuição proposta se adequa aos dados. Porém, nota-se um padrão desigual de dispersão dos resíduos na mesma. Tentou-se modelar o parâmetro de dispersão σ em função das variáveis explicativas com a finalidade de atenuar este efeito indesejado, mas não obtivemos um resultado efetivo.

5.1.2 Ajuste do modelo para o Índice de Herfindahl Estrangeiro

Para o Índice de Herfindahl Estrangeiro adotou-se o modelo de regressão que está descrito no Apêndice F.2. O ajuste do modelo após a aplicação do procedimento de seleção de variáveis (algoritmo F.1) segue apresentado na Tabela 5 a seguir.

Tabela 5: Ajuste do modelo para o Índice de Herfindahl Estrangeiro. São representadas as estimativas dos coeficientes, assim como seus erros padrões e os respectivos Valores P do teste de Wald de inexistência dos efeitos. As estimativas dos parâmetros adicionais da distribuição Beta Inflacionada em 1 para o modelo são $\hat{\sigma} = 0,17$ e $\hat{p}_1 = 0,56$.

Variável	Estimativa do coeficiente	Erro padrão	Valor P
Intercepto	0,50	0,25	0,05
Lpop	-0,12	0,05	0,03
HOMp	0,23	0,09	0,01

Pode-se observar que o procedimento de seleção de variáveis selecionou apenas as variáveis Lpop e HOMp. Observou-se novamente para as empresas estrangeiras que a variável Lpop influencia negativamente no índice de Herfindahl, enquanto a variável HOMp influencia positivamente. Uma comparação entre os modelos para o Índice de Herfindahl Nacional e Estrangeiro mostra que para as empresas nacionais, as variáveis IDHM e TRABp são relevantes para o processo de aglomeração.

A análise de diagnóstico dos resíduos do modelo, conforme pode ser vista na Figura D.2 do Apêndice D, corroborou com as premissas do modelo (conforme definido no Apêndice F.2).

5.2 Modelo para o total de lojas de uma empresa em uma cidade

Novamente, para esta etapa da análise foram considerados modelos distintos para as empresas nacionais e estrangeiras. Diferentemente do modelo apresentado anteriormente, iremos agora adotar a distribuição Binomial Negativa (descrita no Apêndice E.2) para descrever o número total de lojas instaladas por cada uma das empresas em cada uma das cidades. Vale ressaltar que neste tópico as observações possuem dependência, isso decorre do fato de que em todas as cidades são medidos o número total de lojas referentes a cada uma das empresas.

Para incluir essa dependência foi adicionado um efeito aleatório referente a cada uma das cidades no modelo. Esperava-se que esse efeito representasse o fenômeno

de aglomeração, visto que este atuaria no sentido de aumentar ou diminuir o número médio de lojas em uma determinada cidade igualmente entre todas as empresas.

Devido ao fato dos dados estarem representados ao nível de loja, é possível incluir as variáveis: distância da cidade de origem (representada pela abreviação *dist*) e tempo de atividade da empresa (representada pela abreviação *AT*).

5.2.1 Ajuste do modelo para empresas nacionais

Para as empresas nacionais foi feito o ajuste do modelo, descrito no Apêndice F.3, e logo em seguida foi aplicado o procedimento de seleção de variáveis. Os resultados do ajuste do modelo final seguem apresentados na Tabela 6. Com exceção da variável IDHM, todas as variáveis restantes foram consideradas significantes segundo um nível de 5% de significância.

A interpretação dos coeficientes pode ser feita em função da terceira coluna da Tabela 6, em que cada valor tabulado corresponde ao fator multiplicativo aplicado à média do número total de lojas quando esta variável é acrescida em uma unidade, para uma mesma cidade. Por exemplo, se uma cidade com 10 mil habitantes passasse a ter 100 mil habitantes (sem que nenhuma das outras variáveis se altere) espera-se que a média do número total de lojas da cidade após a alteração seja multiplicado pelo fator 19,10 (linha 3, coluna 3 da Tabela 6). Vale comentar que este cálculo é condicional àquela cidade específica, visto que se assume que o efeito aleatório de cidade não se altera.

Tabela 6: Ajuste do modelo para o total de lojas de empresas nacionais. São representadas as estimativas dos coeficientes, assim como seus erros padrões e os respectivos Valores P do teste de Wald de inexistência dos efeitos.

Variável	Estimativa do coeficiente	exp(Estimativa do coeficiente)	Erro padrão	Valor P
Intercepto	-14,207	-	0,997	< 0,001
Lpop	2,950	19,105	0,071	< 0,001
PIBp	-3,963	0,019	0,697	< 0,001
HOMp	-1,263	0,283	0,106	< 0,001
MATRp	-0,714	0,489	0,273	0,010
TRABp	4,044	57,054	0,287	< 0,001
Distância da origem	-0,001	0,999	4e-5	< 0,001
Tempo de atividade	0,055	1,057	0,287	< 0,001

A Figura D.3 (Apêndice D) indica concordância entre as suposições do modelo e o ajuste. Porém a interpretação do efeito aleatório incluído no modelo parece não corresponder à expectativa de quantificação do fenômeno de aglomeração, como visto na Figura D.4, que indica não existir correlação entre o Índice de Herfindahl Nacional e o valor predito dos efeitos aleatórios.

5.2.2 Ajuste do modelo para empresas estrangeiras

A partir do modelo para as empresas estrangeiras descrito no Apêndice F.4, cujo ajuste é apresentado na Tabela 7, nota-se que todas as variáveis explicativas foram estatisticamente significantes segundo um nível de 5% significância. A interpretação dos parâmetros deste modelo é equivalente à apresentada na seção anterior.

Tabela 7: Ajuste do modelo para o total de lojas de empresas nacionais. São representadas as estimativas dos coeficientes, assim como seus erros padrões e os respectivos Valores P do teste de Wald de inexistência dos efeitos.

Variável	Estimativa do coeficiente	exp(Estimativa do coeficiente)	Erro padrão	Valor P
Intercepto	-16,207		0,997	< 0,001
IDHM	8,55	5174,17	0,929	< 0,001
Lpop	1,72	5,60	0,071	< 0,001
PIBp	-2,61	0,07	0,697	0,01
HOMp	-0,51	0,60	0,106	< 0,001
MATRp	-1,20	0,30	0,273	< 0,001
TRABp	2,83	17,06	0,287	< 0,001
Tempo de atividade	0,06	1,07	0,287	< 0,001

O resultado da análise de resíduos (Figura D.5) se mostrou adequado a todas as suposições do modelo. Novamente, a interpretação do efeito aleatório não seguiu o idealizado (Figura D.6), assim como no modelo da Seção 5.2.1.

Verificou-se, porém, que para as empresas estrangeiras, a variável IDHM apresentou efeito significativo no total de lojas abertas por uma empresa em cada uma das cidades. Esse resultado é diferenciado obtido para o modelo do Índice de HerfindahlEstrangeiro, em que a variável IDHM não teve efeito estatisticamente significativo de aumentar a aglomeração em cada cidade. Isso indica que a variável IDHM parece influenciar positivamente na média do número total de lojas abertas por cada empresa, porém, apesar do número bruto de lojas aumentar, a aglomeração parece se manter constante.

6. Conclusões

Da análise inferencial, observa-se que a padronização das variáveis explicativas reduziu o efeito de colineariedade induzida pela variável População. Verificou-se que os sinais dos coeficientes dos modelos ajustados para os Índices de Herfindahl Nacional e Estrangeiro são condizentes com o que era esperado.

Nos modelos para os totais de lojas (Seção 5.2), as variáveis referentes às empresas (distância à origem e tempo de atividade) se mostraram significantes, concordando assim com o resultado obtido da análise descritiva. Também pode-se concluir que os efeitos aleatórios de cidade não foram capazes de absorver por completo o efeito do fenômeno de aglomeração.

Em ambas as abordagens verificou-se que a variável Homicídios tem impactonegativo significativo na aglomeração, ou seja, espera-se que cidades com maiores taxas de homicídios tenham menores concentrações de redes de *fastfood*. A variável População também se mostrou importante em todos os modelos propostos, atuando de maneira a aumentar a aglomeração de maneira significativa.

Apêndice A: Análise descritiva unidimensional em nível de cidade

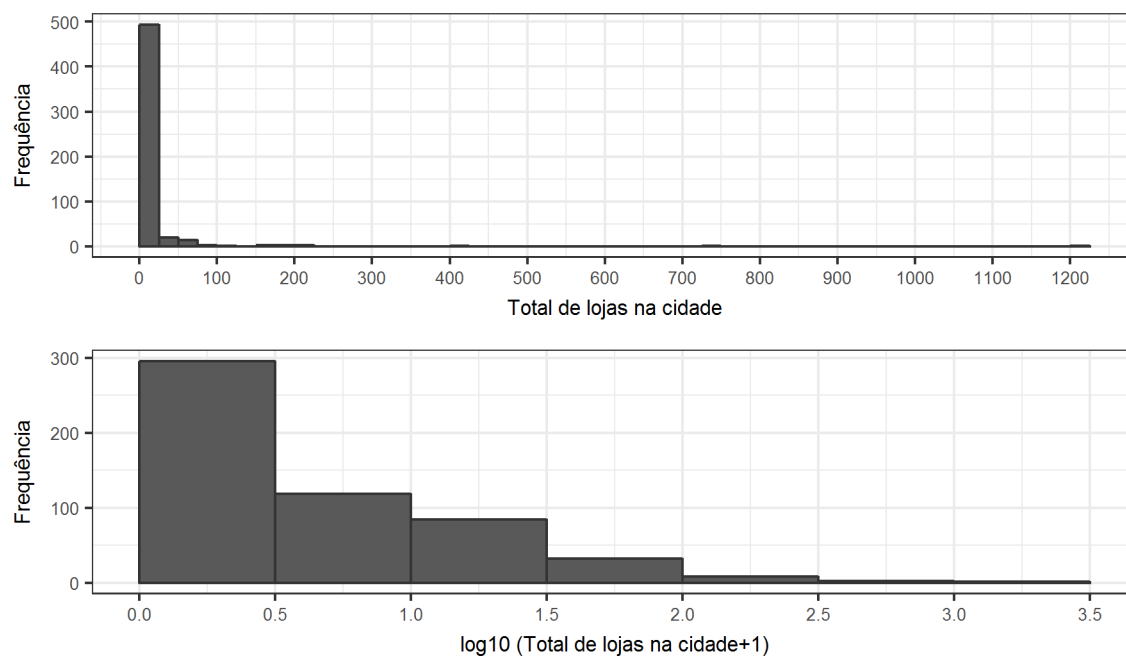


Figura A.1: Distribuição da variável “Total de lojas na cidade” sem nenhuma transformação (topo) e na transformação logarítmica base 10 (baixo).

Tabela A.1: Medidas descritivas da variável “Total de lojas na cidade”.

	Total de lojas na cidade	$\log_{10}(\text{Total de lojas na cidade} + 1)$
Mínimo	0,00	0,00
Máximo	1218,00	3,09
Mediana	2,00	0,48
Média	14,12	0,68
Variância	4535,75	0,24
Desvio padrão	67,35	0,49

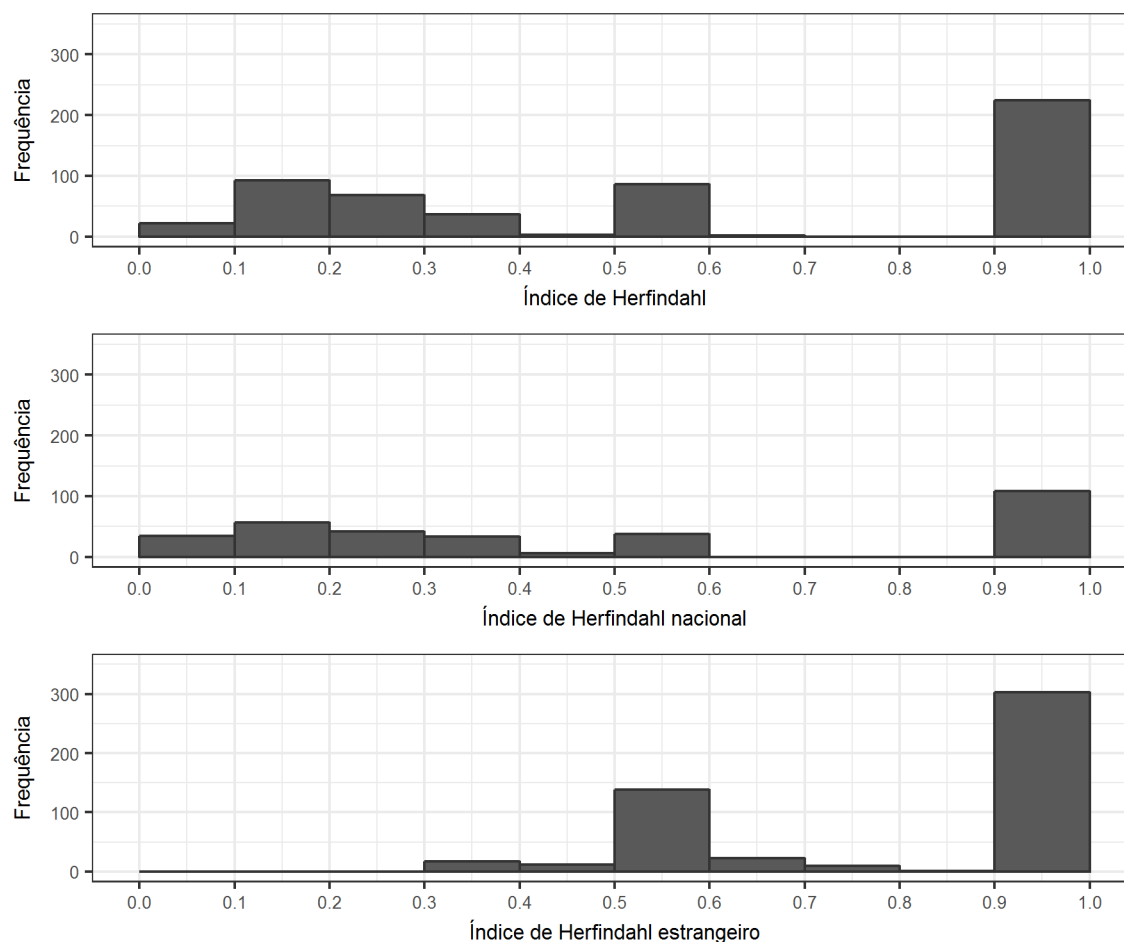


Figura A.2: Histograma do índice de Herfindahl geral, nacional e estrangeiro.

Tabela A.2 Medidas descritivas do índice de Herfindahl geral, nacional e estrangeiro.

	Índice de Herfindahl geral	Índice de Herfindahl nacional	Índice de Herfindahl estrangeiro
Número de observações não disponíveis	8	223	39
Número de observações com o valor máximo	224	108	303
Mínimo	0,06	0,05	0,30
Máximo	1,00	1,00	1,00
Mediana	0,5	0,33	1,00
Média	0,59	0,51	0,81
Variância	0,14	0,14	0,05
Desvio padrão	0,37	0,37	0,24

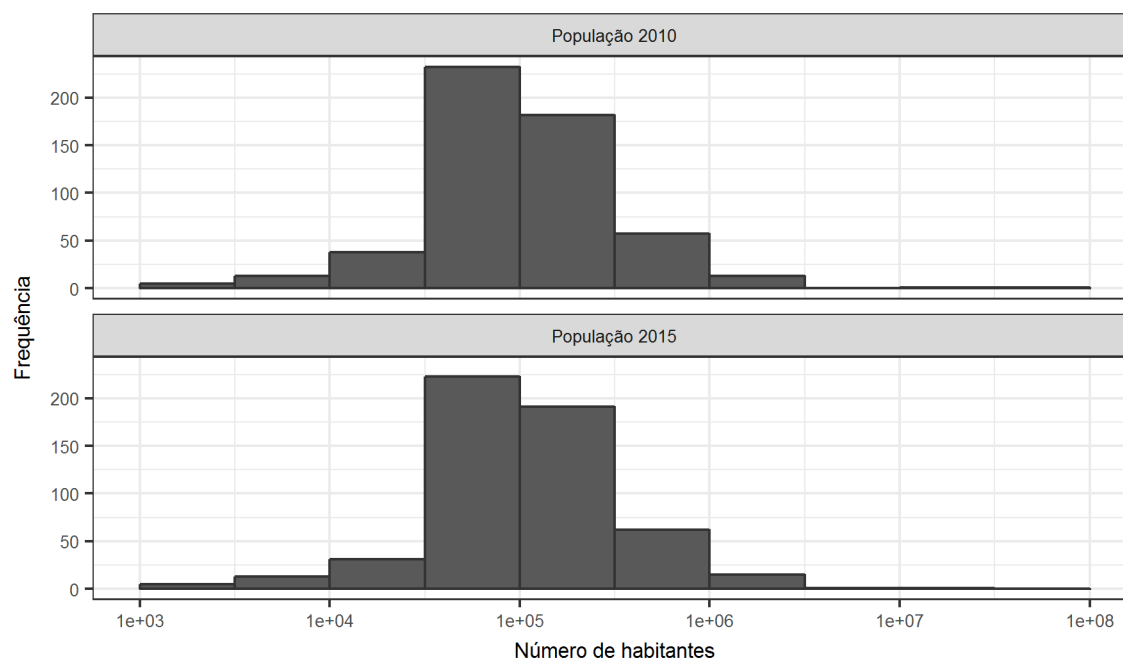


Figura A.4: Histograma da População (2015) e População (2010), em escala logarítmica base 10.

Tabela A.4: Medidas descritivas para População 2010 e População 2015

	População 2010	População 2015
Mínimo	1,61e+03	1,71e+03
Máximo	1,35e+07	1,19e+07
Mediana	9,25e+04	9,92e+04
Média	2,26e+05	2,28e+05
Desvio-padrão	8,02e+05	6,57e+05

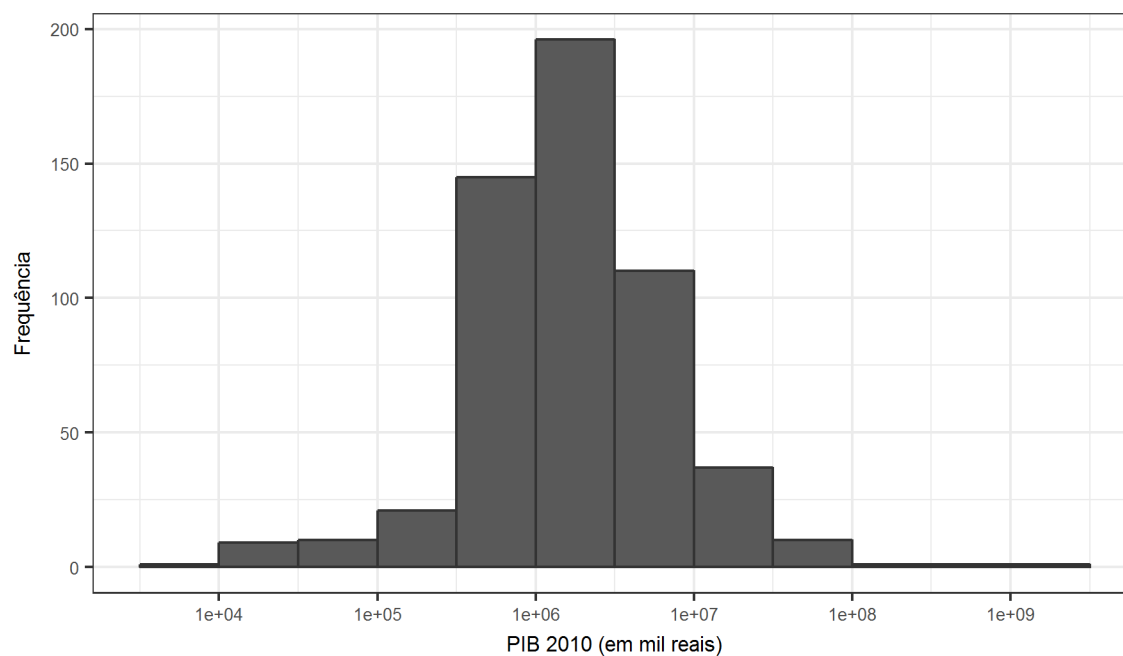


Figura A.5: Histograma do PIB Municipal (em mil reais) sob escala logarítmica de base 10.

Tabela A.5: Medidas descritivas para as variáveis “Taxa de homicídios 2011”, “Matrícula profissional técnica 2010” e “Número de beneficiados pelo Programa de Auxílio ao Trabalhador(PAT) 2015”.

	Taxa de homicídios 2011	Matrícula profissional técnica 2010	Número de beneficiados pelo Programa de Auxílio ao Trabalhador (PAT) 2015
Número de observações não disponíveis	0,00	0,00	60,00
Mínimo	0,00	0,00	1,00
Máximo	1.812,00	352.053,00	71.623
Mediana	21,00	656,50	223,50
Média	73,50	2.260,35	946,08
Variância	34.462,16	2,51e+08	1,44e+07
Desvio padrão	185,64	15.853,54	3.799,69

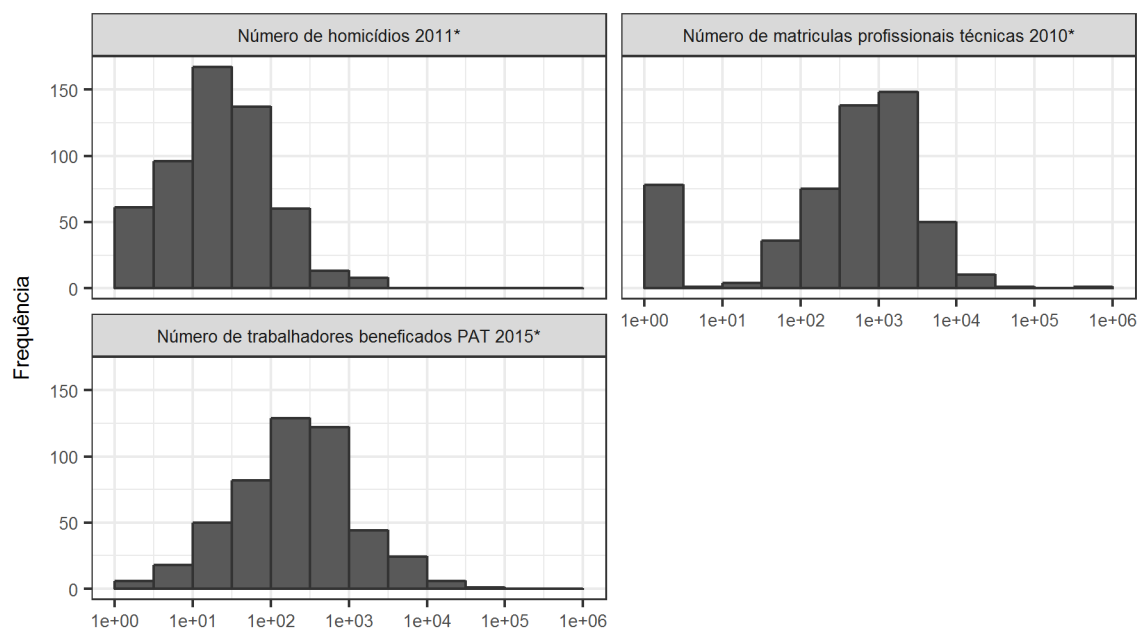


Figura A.6: Distribuição das variáveis “Taxa de homicídios 2011”, “Matrícula profissional técnica 2010” e “Número de beneficiados pelo Programa de Auxílio ao Trabalhador(PAT)” em escala logarítmica base 10.

* As variáveis foram acrescidas em uma unidade para permitir a transformação logarítmica.

Apêndice B: Análise descritiva multidimensional em nível de cidade

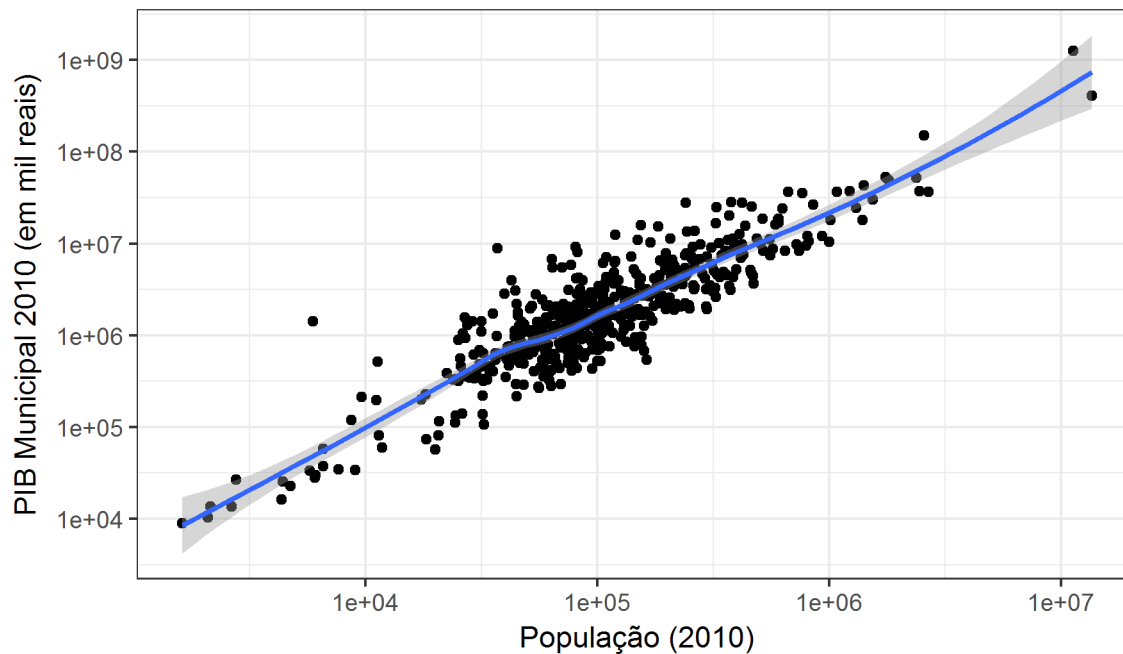


Figura B. 1: População (2010) vs. PIB municipal 2010 (em mil reais), em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *span*¹ de 0,5. A área rachurada representa um intervalo de 95% de confiança para o suavizador.

¹ O parâmetro *span* é definido como a fração de pontos usada para ajustar cada regressão local do suavizador.

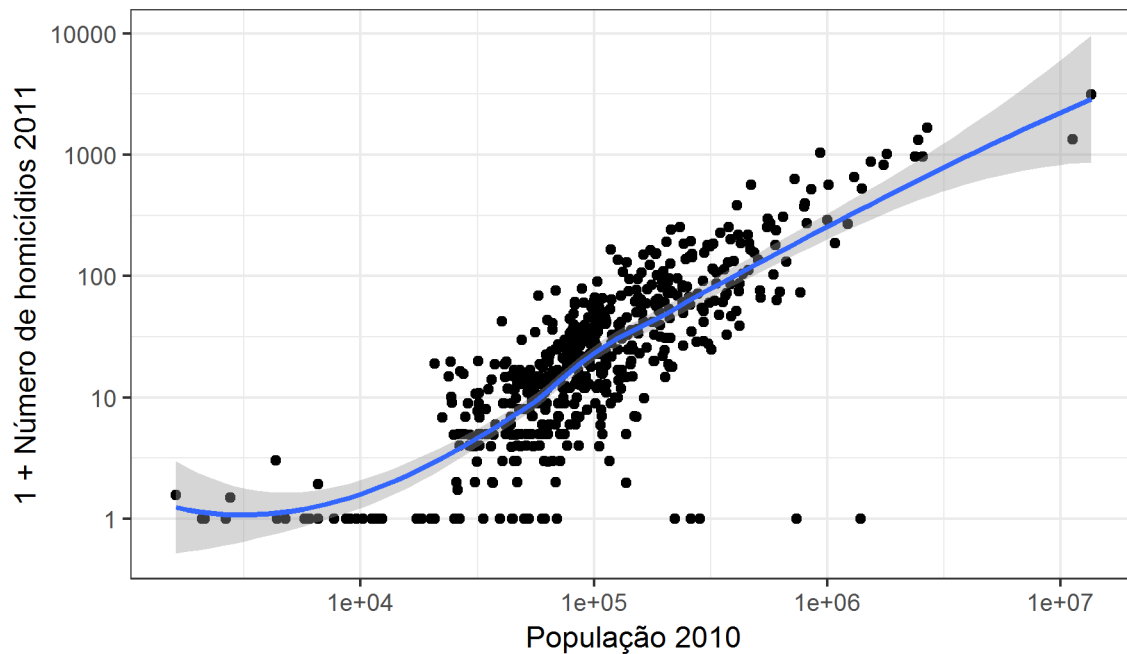


Figura B. 2: População (2010) vs. Número de homicídios (2011) + 1, em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *spande* 0,5. A área rachurada representa um intervalo de 95% de confiança para o suavizador.

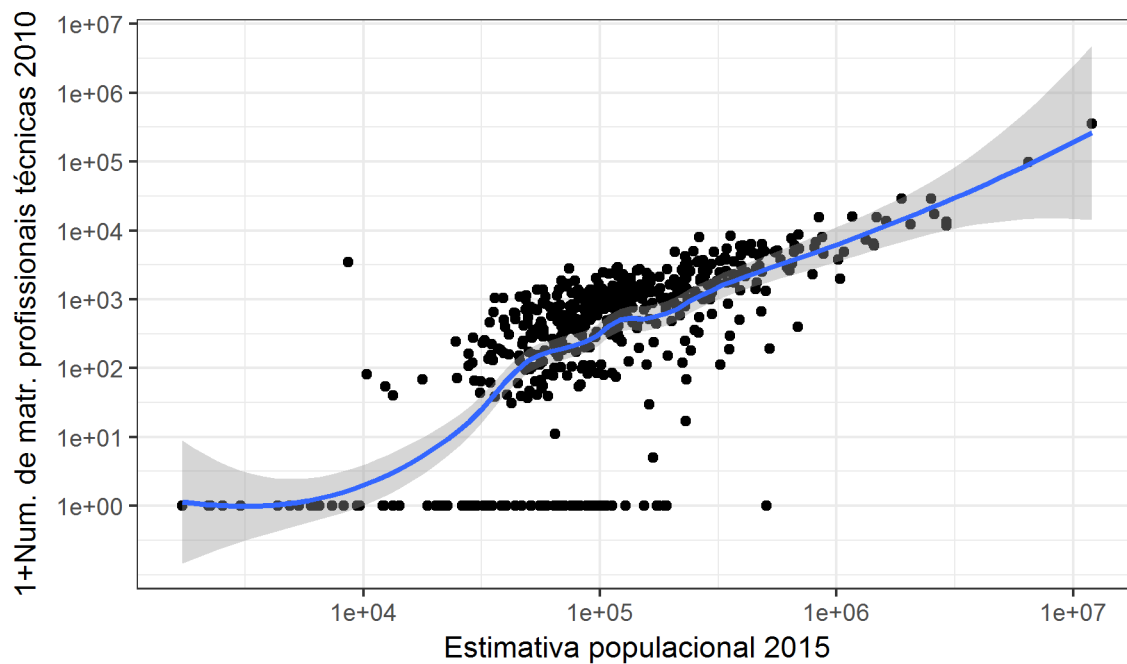


Figura B. 3: Estimativa populacional 2015 vs. (1 + Número de matrículas profissionais técnicas 2011), em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *spande* 0,5. A área rachurada representa um intervalo de 95% confiança para o suavizador.

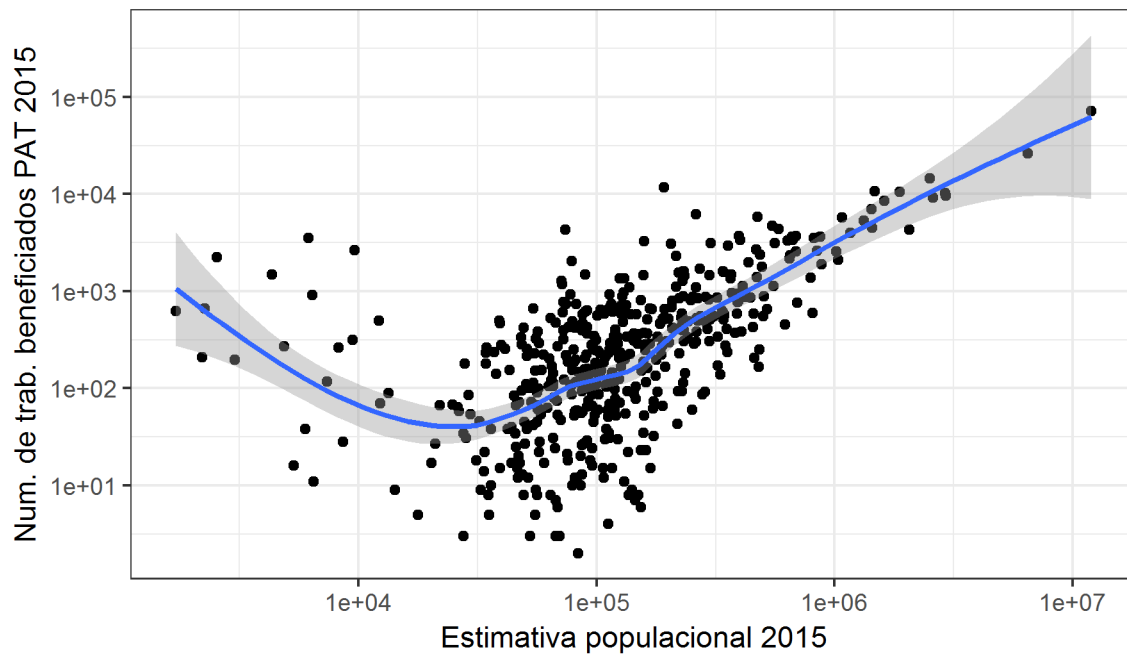


Figura B. 4: Estimativa populacional 2015 vs. (1 + Número de trabalhadores beneficiados PAT 2015), em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *spande* 0,5. A área rachurada representa um intervalo de 95% confiança para o suavizador.

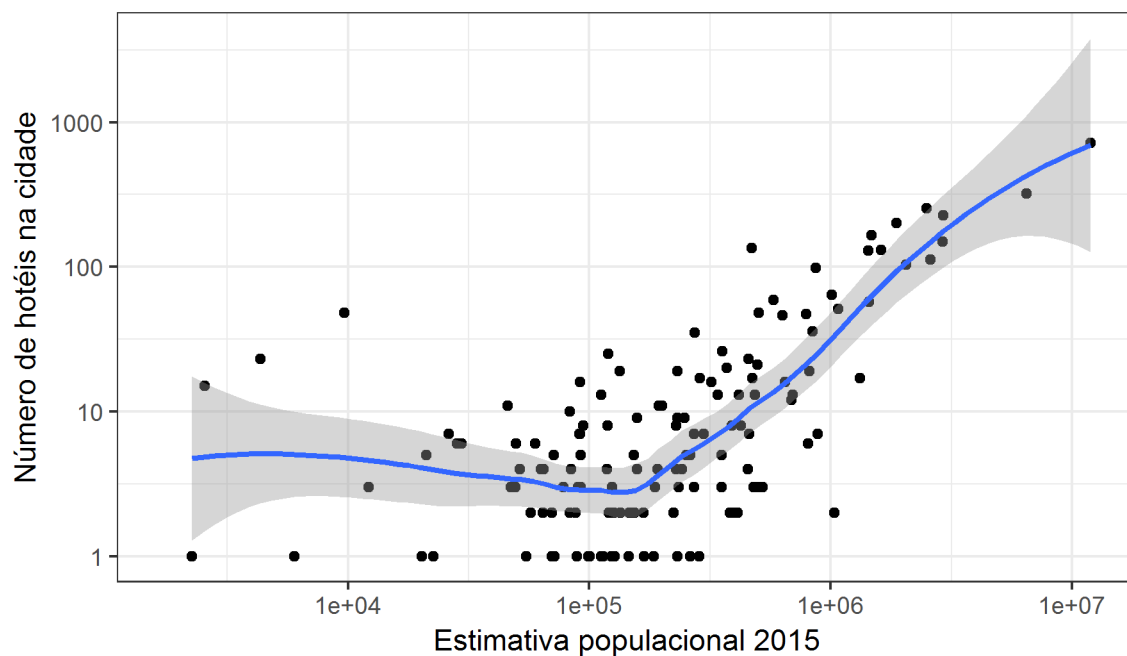


Figura B. 5: Estimativa populacional 2015 vs. Número de hotéis na cidade, em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *spande* 0,5. A área rachurada representa um intervalo de 95% confiança para o suavizador.

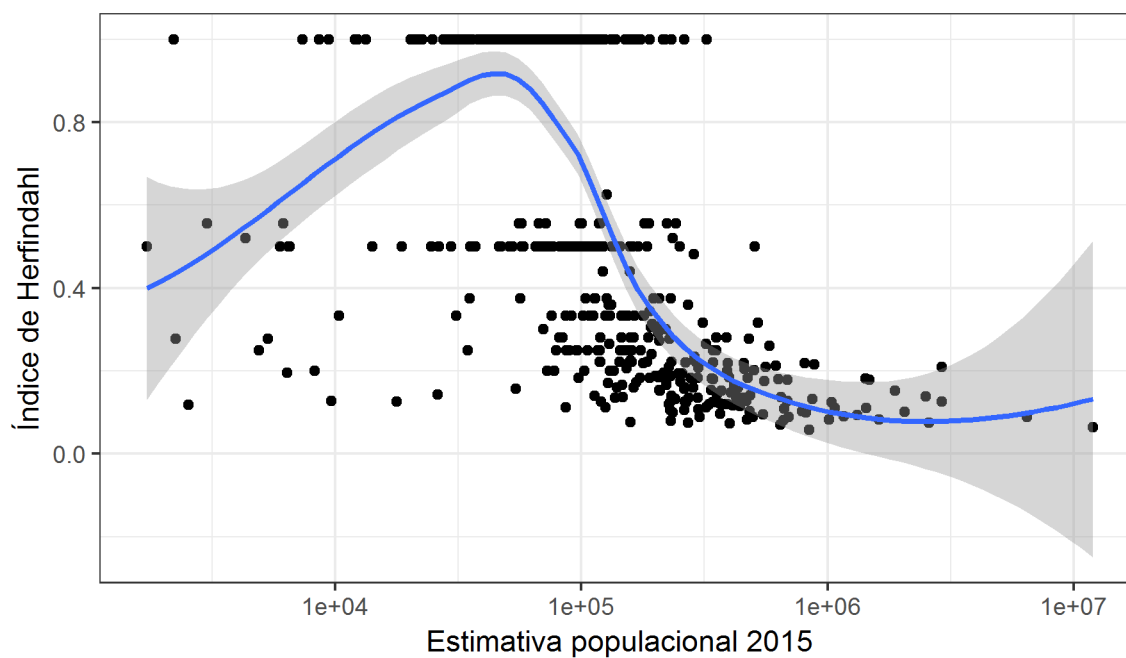


Figura B. 6: Estimativa populacional 2015 vs. Índice de Herfindahl, em escala logarítmica no eixo vertical e com suavizador LOESS, com parâmetro *span* de 0,5. A área rachurada representa um intervalo de 95% confiança para o suavizador.

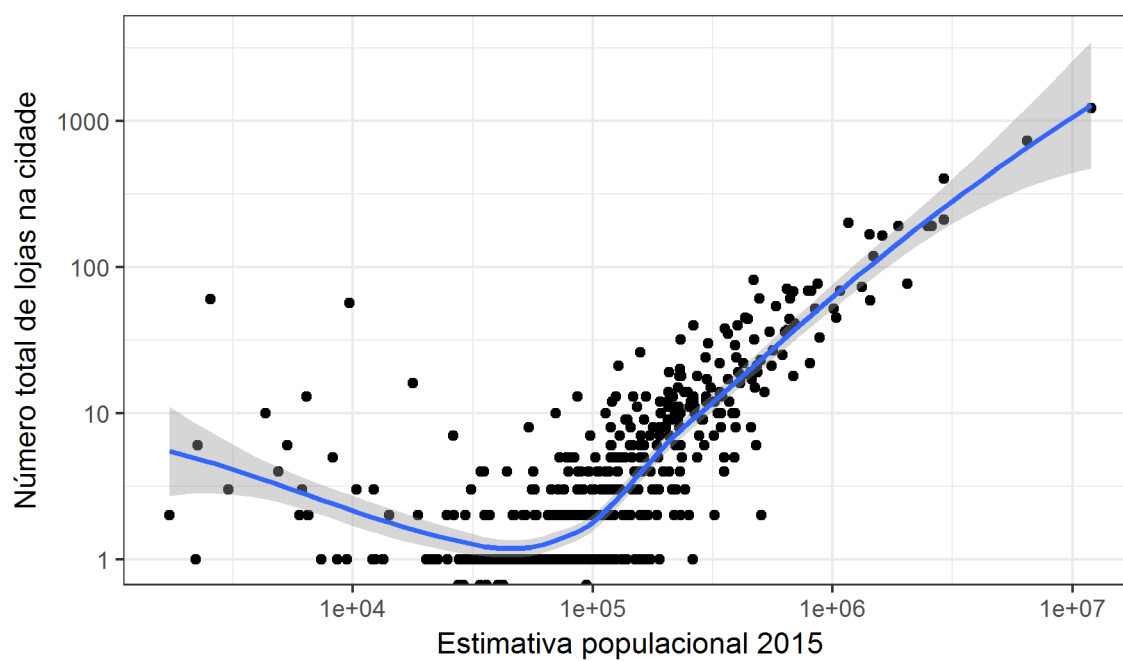


Figura B. 7: População 2015 vs. Total de lojas, em escala logarítmica em ambos os eixos e com suavizador LOESS, com parâmetro *span* de 0,5. A área rachurada representa um intervalo de 95% confiança para o suavizador.

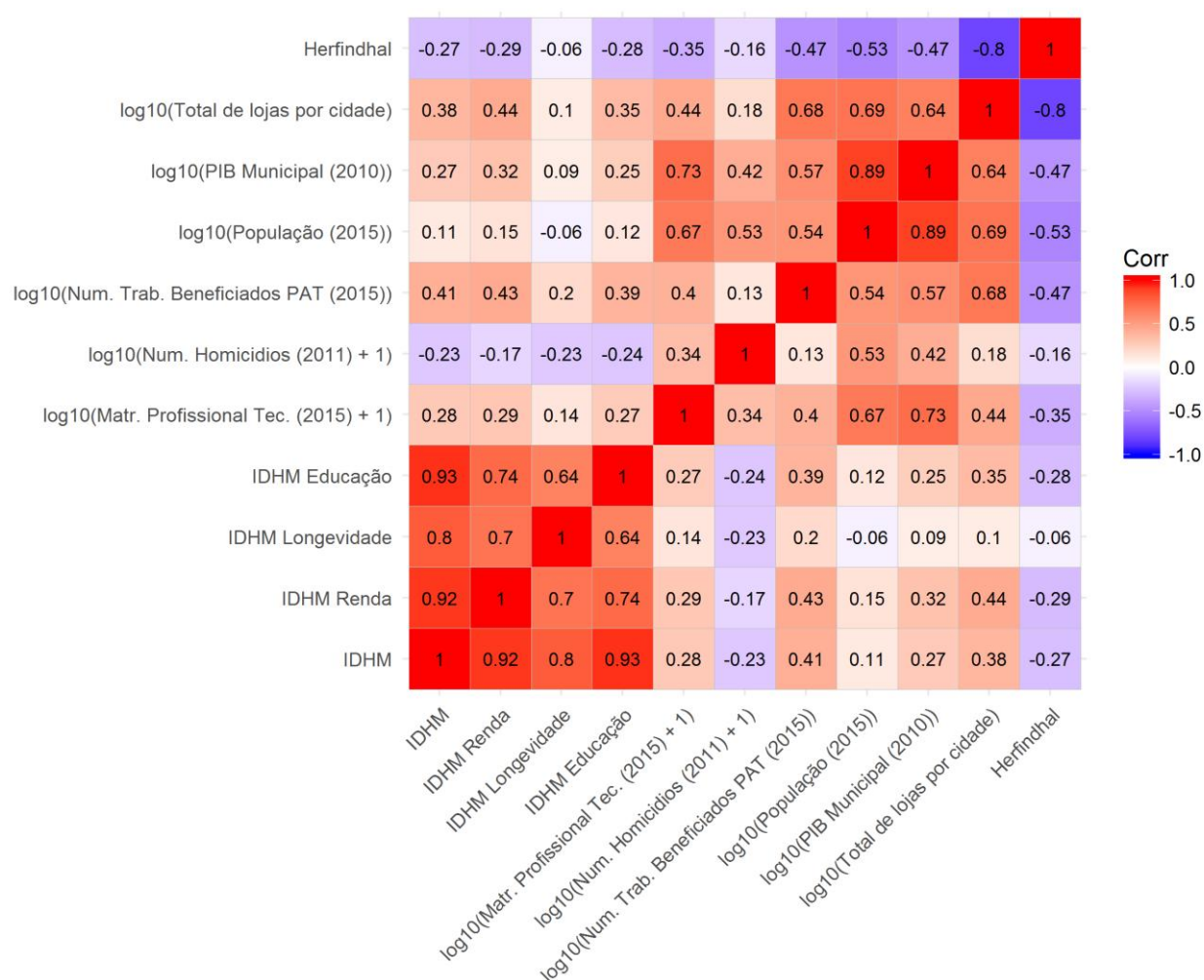


Figura B. 8: Matriz de correlações de *Pearson*, para as variáveis explicativas e resposta. A intensidade da cor vermelha representa uma correlação positiva, a intensidade da cor azul representa uma correlação negativa.

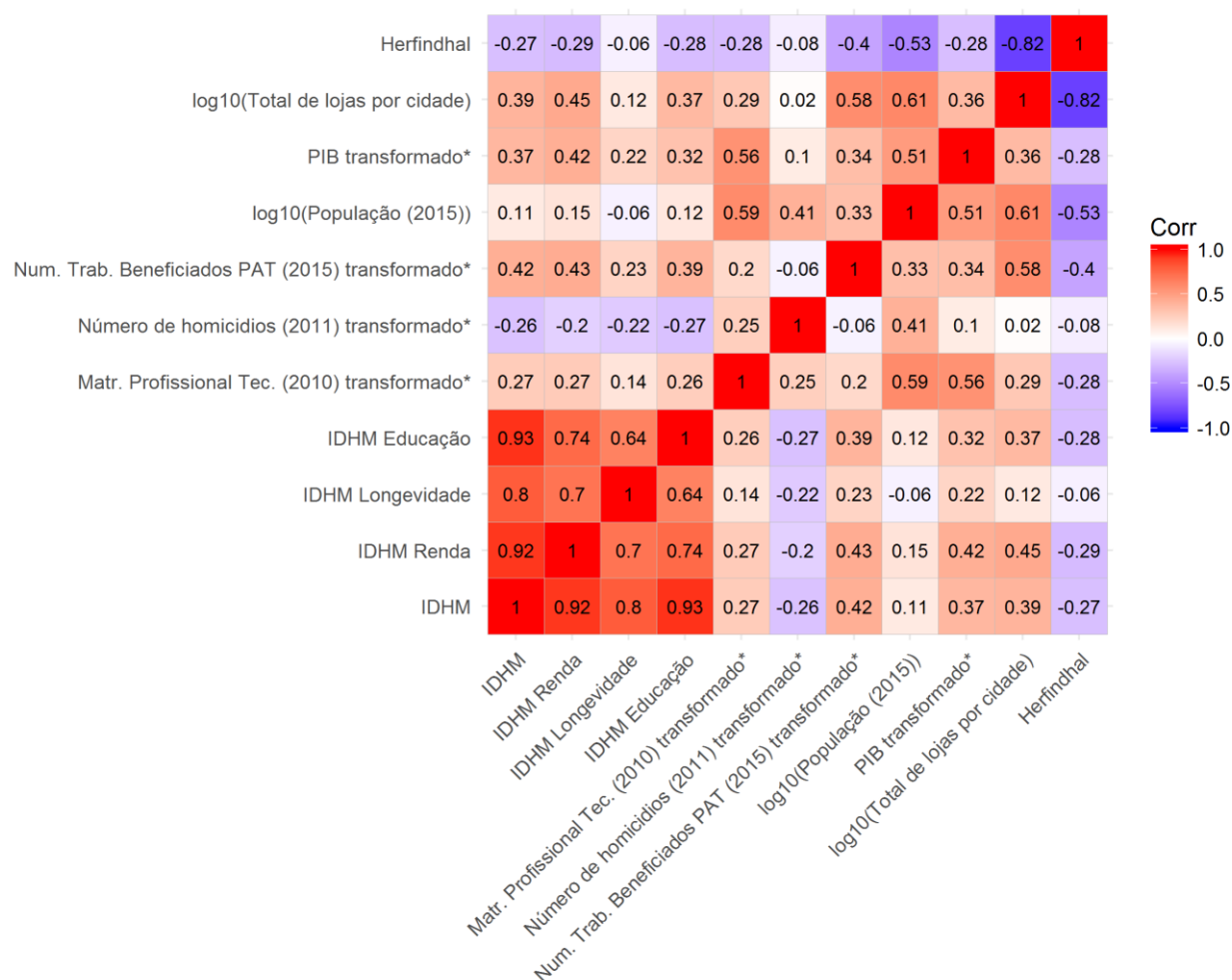


Figura B. 9: Matriz de correlações de *Pearson*, para as variáveis explicativas e resposta. Algumas variáveis foram transformadas (como indicado no gráfico) segundo a transformação discutida na seção 4.1.3 deste trabalho. A intensidade da cor vermelha representa uma correlação positiva, a intensidade da cor azul representa uma correlação negativa.

Apêndice C: Análise descritiva em nível de loja

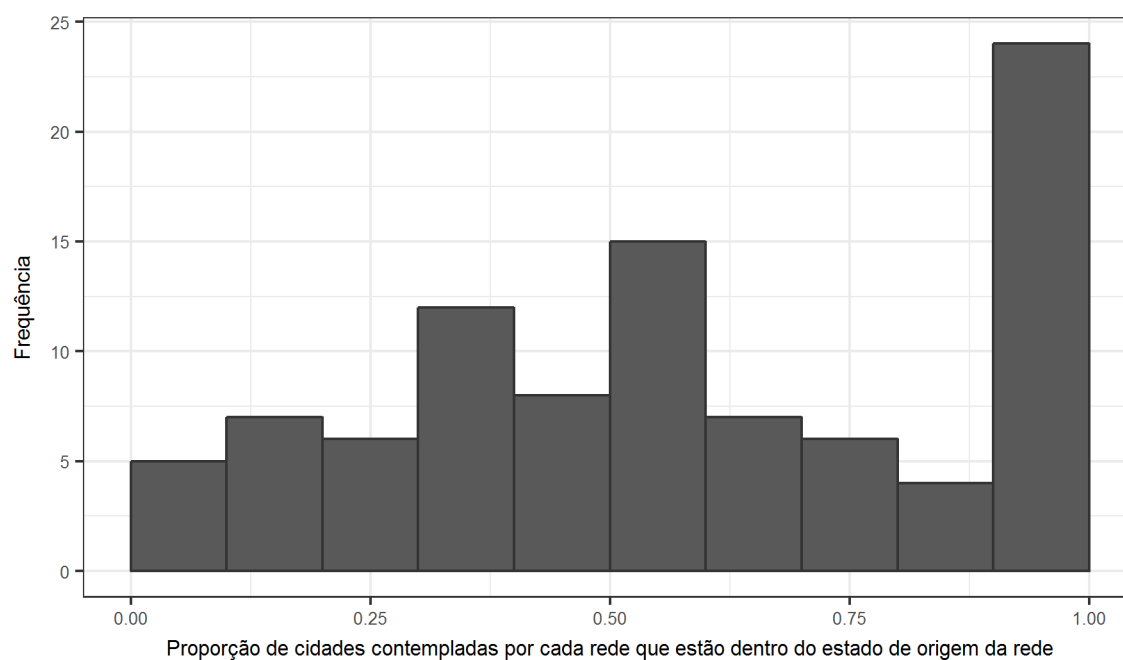


Figura C. 1: Histograma do calculo feito para cada uma das 94 redes nacionais, em que se verificou a proporção de cidades contempladas por esta rede que pertencem ao estado de origem de fundação da mesma.

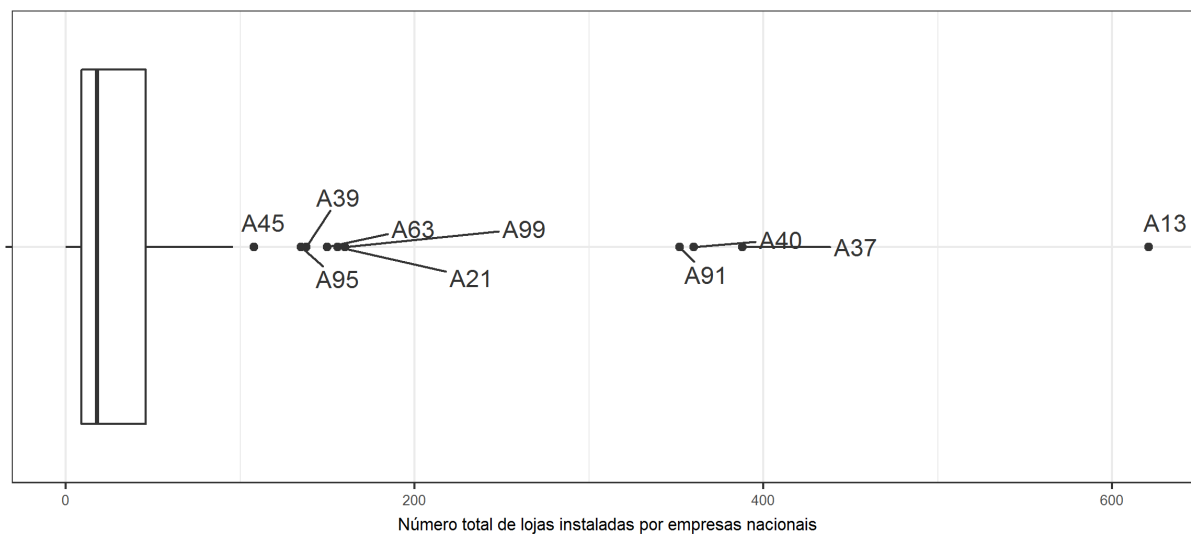


Figura C. 2: Histograma do número de Total por empresa, com *outliers* indicados (empresas).

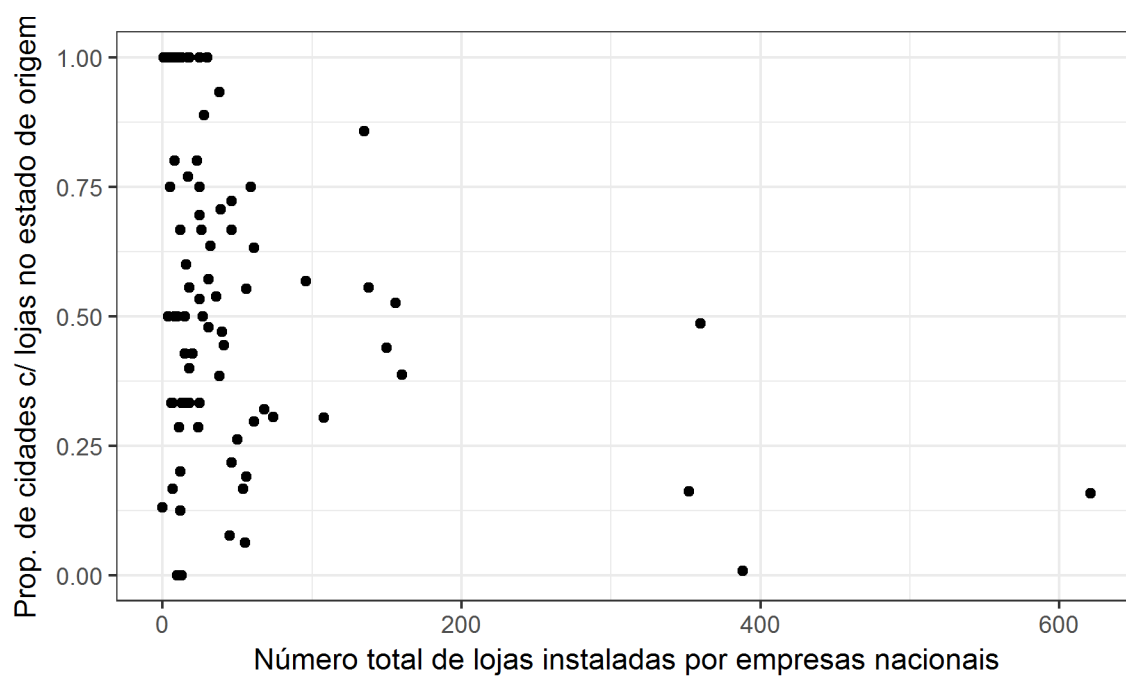


Figura C. 3: Total de lojas instaladas por cada empresa nacional vs. proporção de lojas no estado de origem da rede.

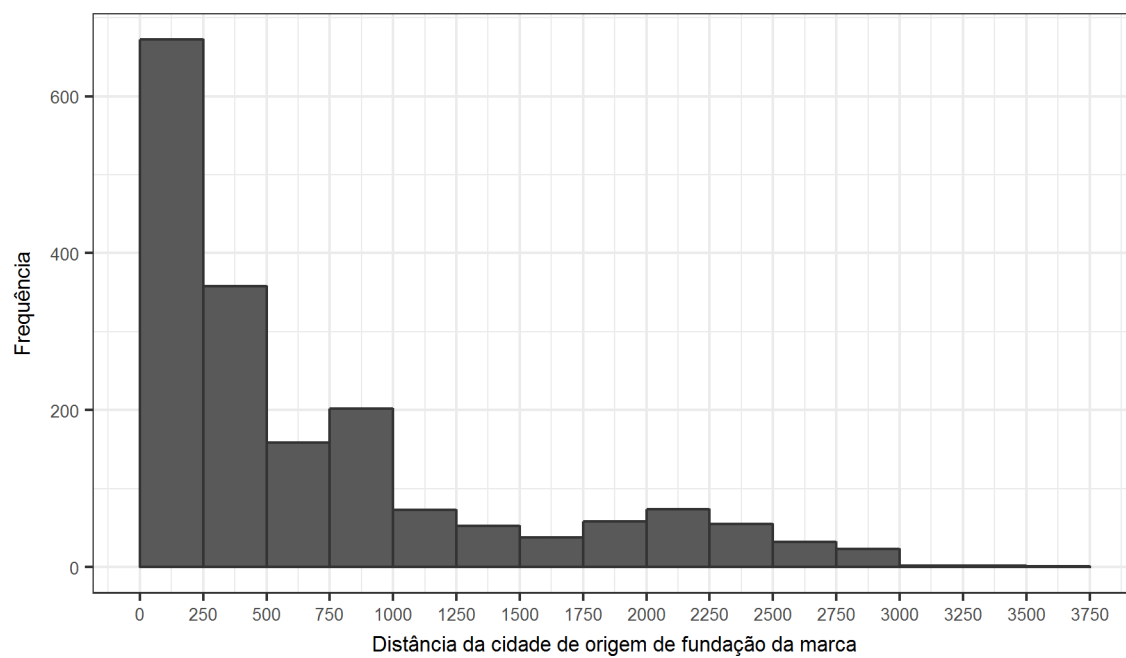


Figura C. 4: Histograma da distância calculada entre todas as cidades contempladas por uma rede e sua cidade de origem de fundação, excluindo-se do calculo as cidades de origem (distância = 0)

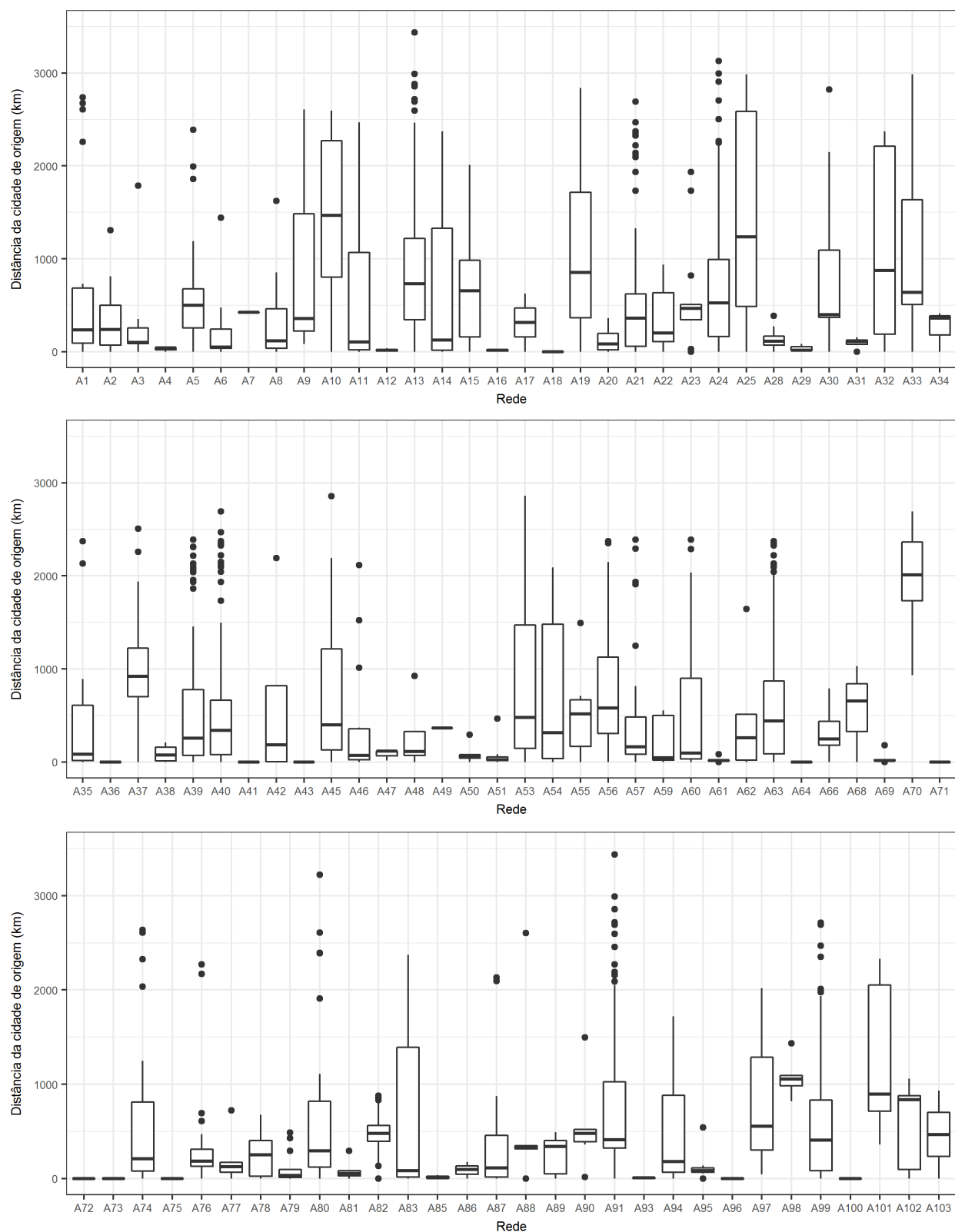


Figura C. 5: Distribuição das distâncias calculadas entre as cidades contempladas e a cidade de origem de cada rede.

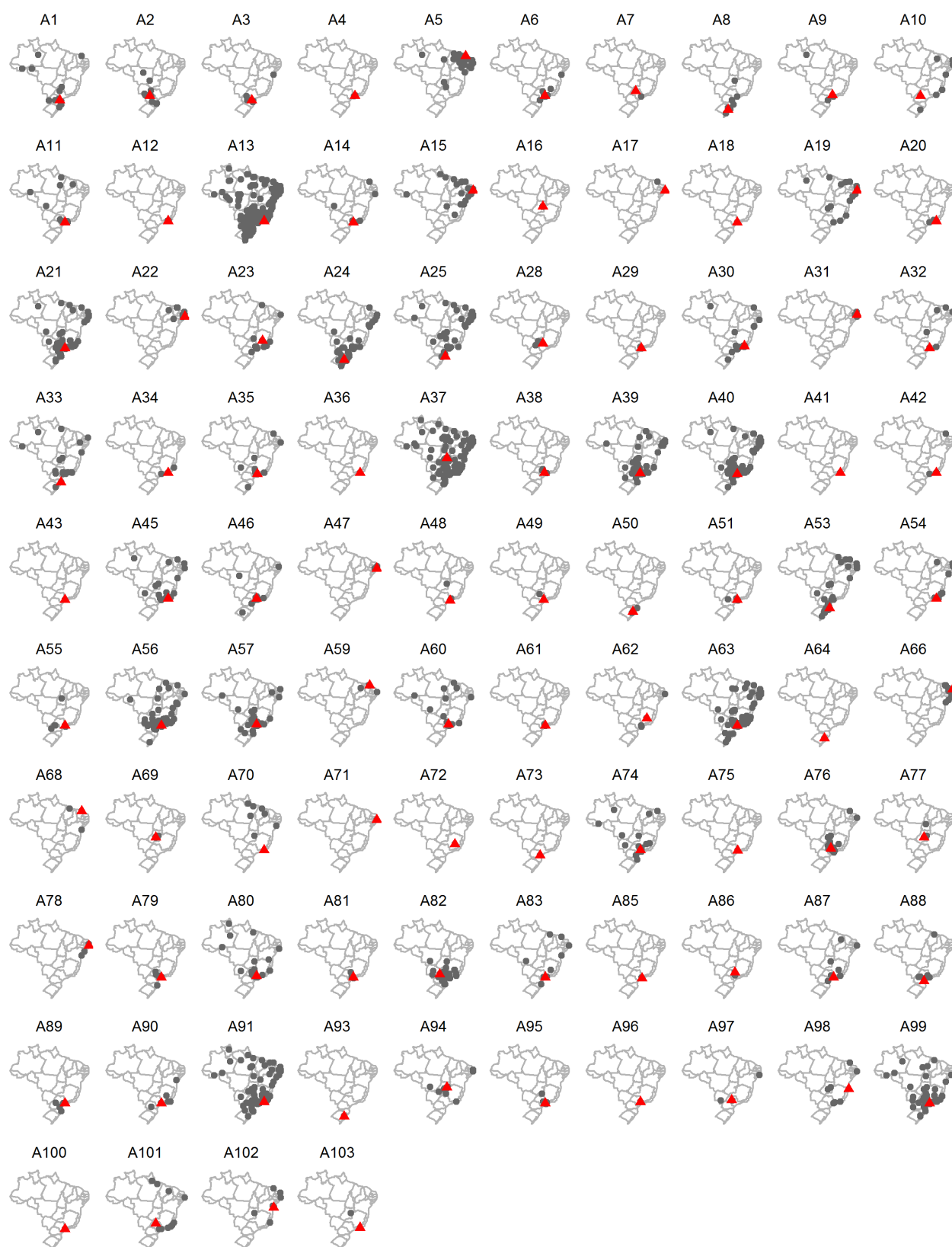


Figura C. 6: Distribuição geográfica das lojas de cada uma das redes nacionais. Cada ponto corresponde à uma cidade com lojas e as cidades de origem estão representadas por um triângulo vermelho.

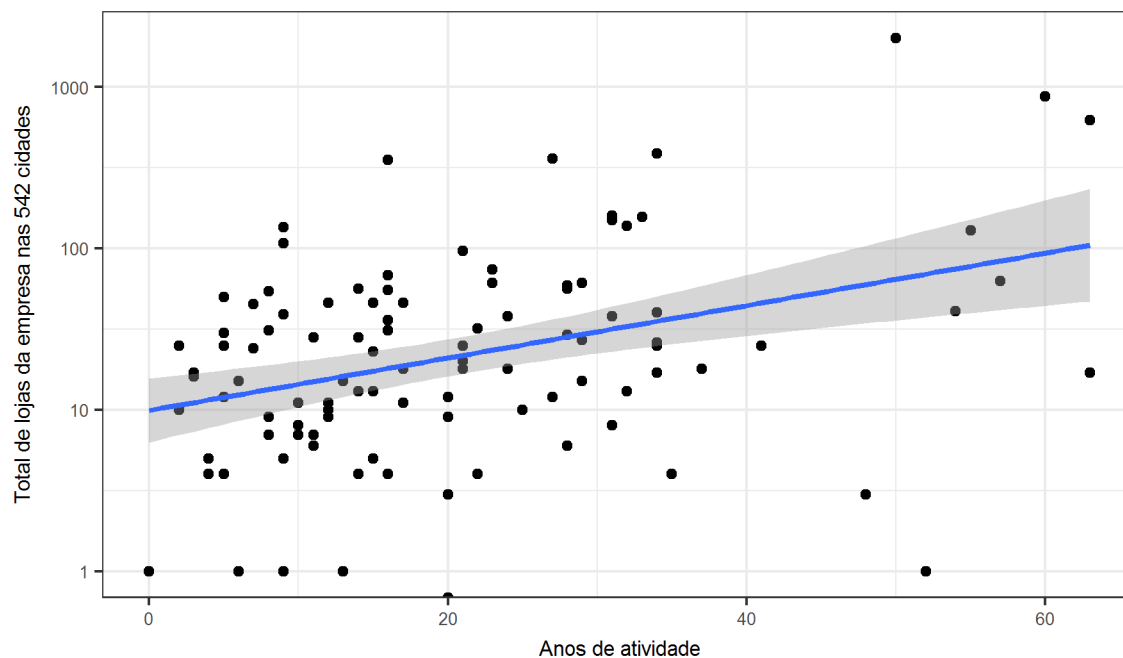


Figura C. 7: $\log_{10}(\text{Total agregado de lojas da rede}^*)$ vs. anos de atividade. O coeficiente de correlação linear de Pearson para este par de variáveis vale 0,39.

* A rede “A33” não consta no gráfico, pois não possui nenhuma loja, apenas quiosques.

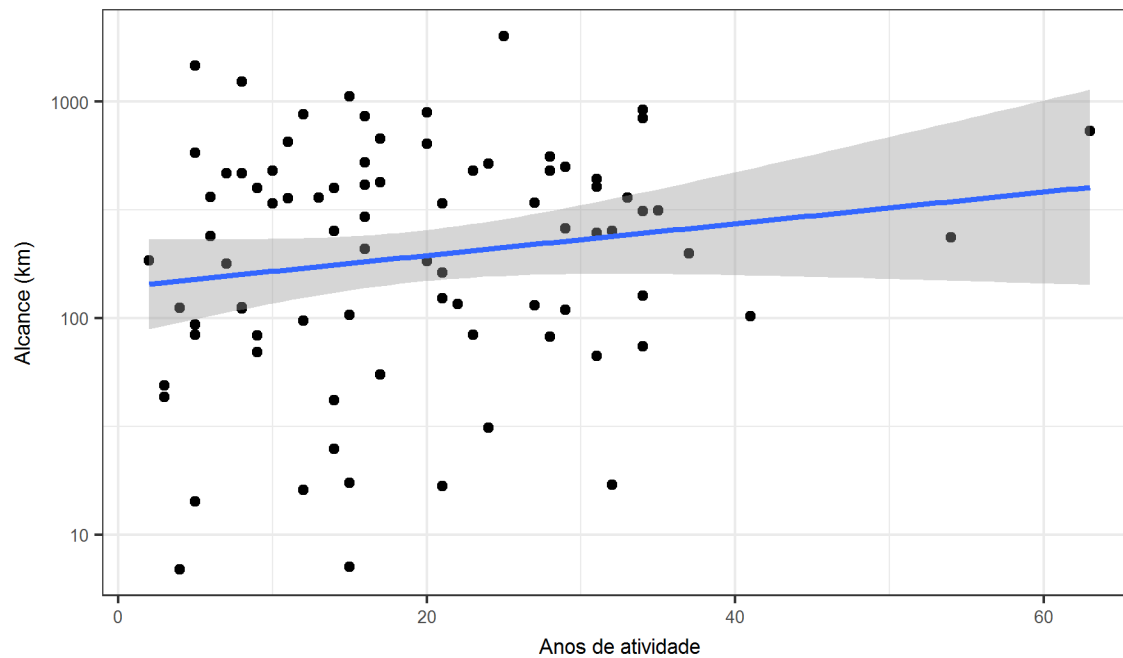


Figura C. 8: $\log_{10}(\text{distância máxima de uma cidade com lojas/quiosques da origem, por rede}^*)$ vs. anos de atividade. O coeficiente de correlação linear de Pearson para este par de variáveis vale 0,16.

Apêndice D: Gráficos de resíduos

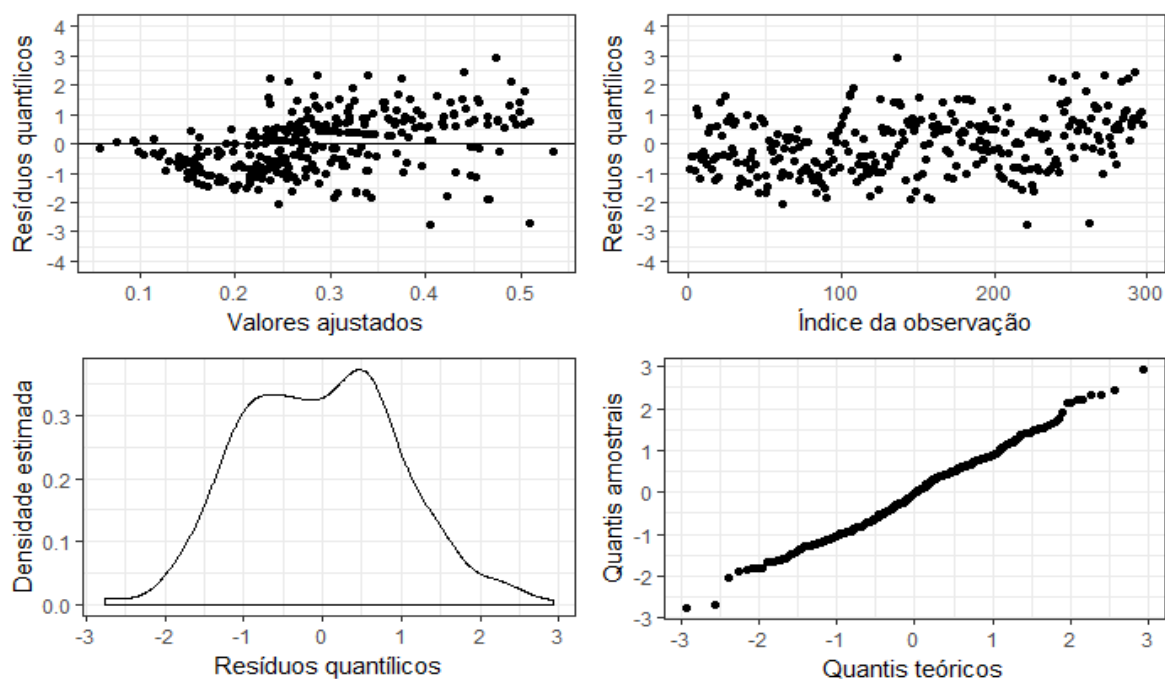


Figura D. 1: Gráficos de diagnóstico para os resíduos quantílicos do modelo de 5.1.1: Valores ajustados vs. Resíduos (canto superior esquerdo), Índice da observação vs Resíduos (canto superior direito), densidade estimada dos resíduos (canto inferior esquerdo) e gráfico Q-Q (referente à uma distribuição normal) dos resíduos.

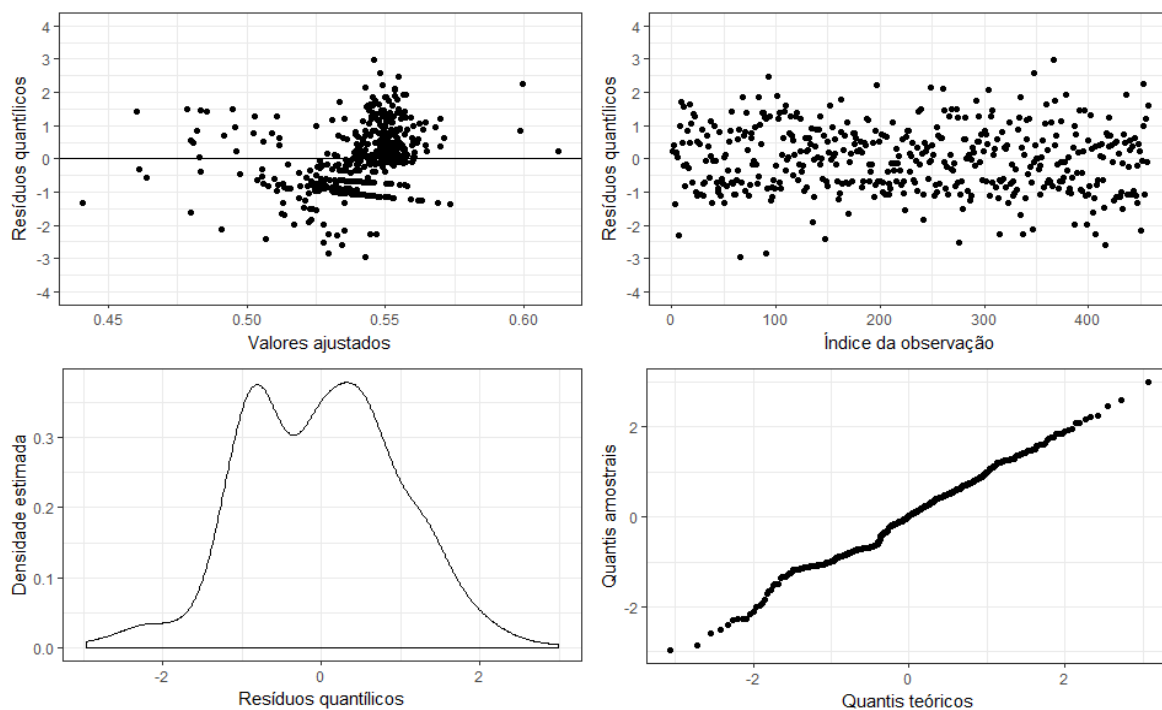


Figura D.2: Gráficos de diagnóstico para os resíduos quantílicos do modelo de 5.1.1: Valores ajustados vs. Resíduos (canto superior esquerdo), Índice da observação vs Resíduos (canto superior direito), densidade estimada dos resíduos (canto inferior esquerdo) e gráfico Q-Q (referente à uma distribuição normal) dos resíduos.

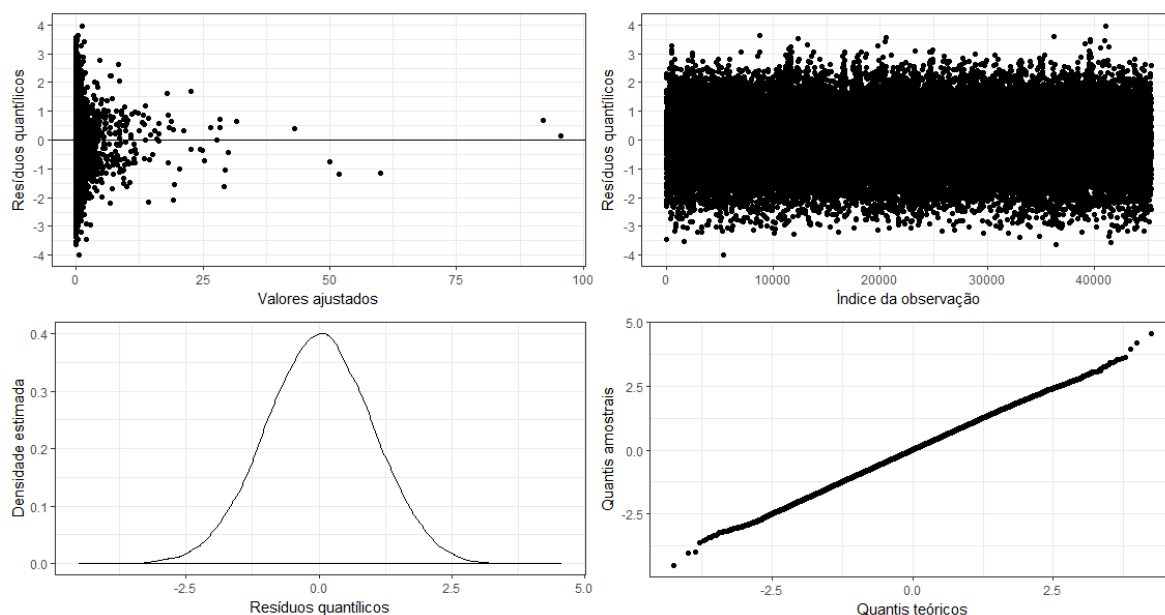


Figura D. 3: Gráficos de diagnóstico para os resíduos quantílicos do modelo de 5.2.1: Valores ajustados vs. Resíduos (canto superior esquerdo), Índice da observação vs Resíduos (canto superior direito), densidade estimada dos resíduos (canto inferior esquerdo) e gráfico Q-Q (referente à uma distribuição normal) dos resíduos.

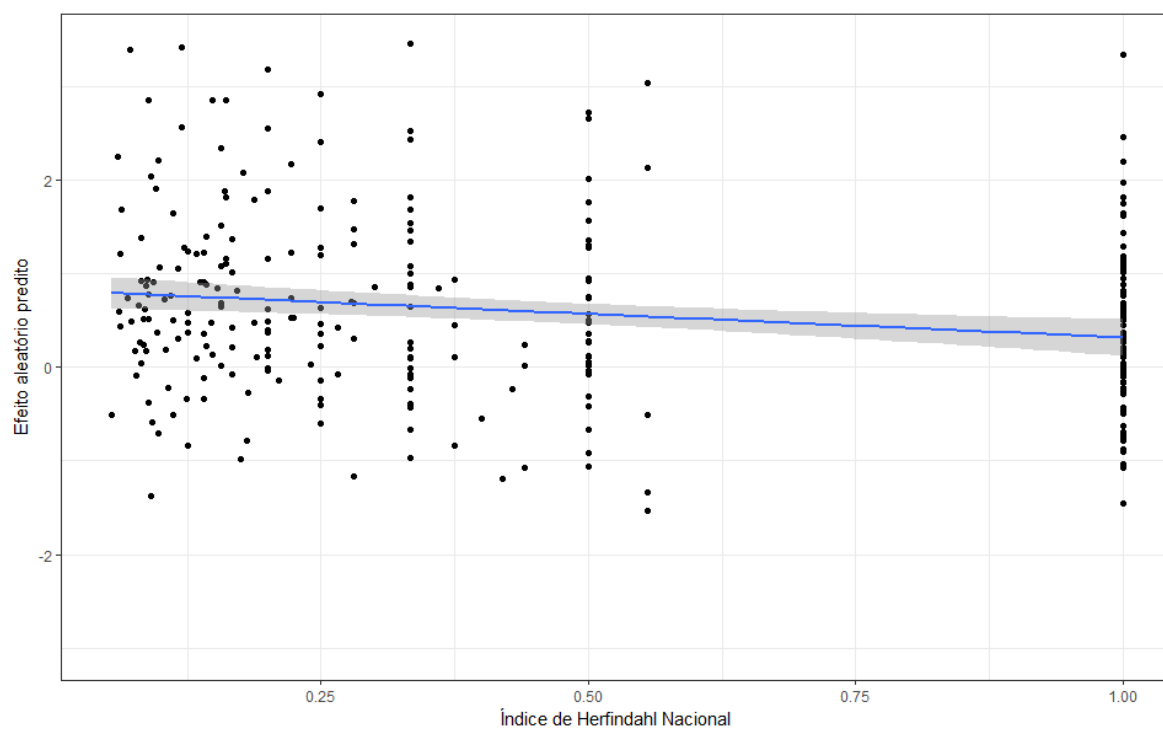


Figura D. 4: Índice de Herfindahl Nacional vs. Efeito aleatório predito do modelo descrito na seção 5.2.1 e reta de regressão ajustada.

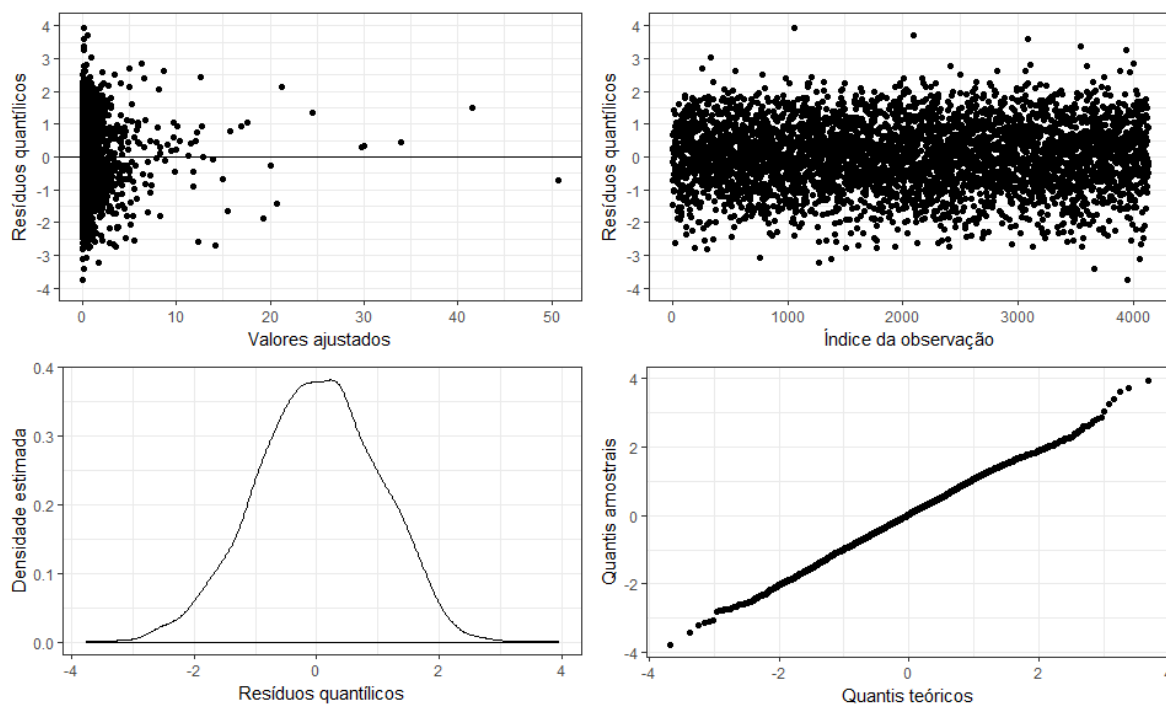


Figura D. 5: Gráficos de diagnóstico para os resíduos quantílicos do modelo de 5.2.2: Valores ajustados vs. Resíduos (canto superior esquerdo), Índice da observação vs Resíduos (canto superior direito), densidade estimada dos resíduos (canto inferior esquerdo) e gráfico Q-Q (referente à uma distribuição normal) dos resíduos.

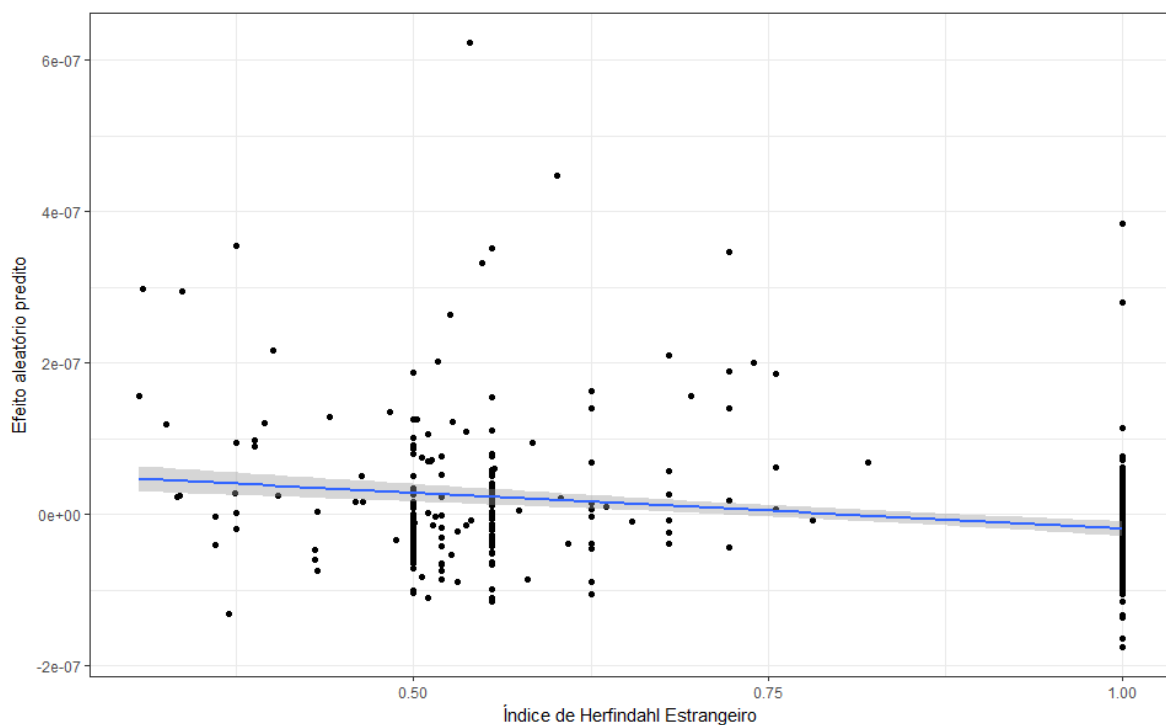


Figura D. 6: Índice de HerfindahlEstrangeiro vs. Efeito aleatório predito do modelo descrito na seção 5.2.2 e reta de regressão ajustada.

Apêndice E – Distribuições de probabilidade

Distribuição E.1 – Distribuição Beta Inflacionada em 1

Se Y é uma variável aleatória com distribuição Beta Inflacionada em 1 de parâmetros μ, σ, p_1 , então a sua função de densidade de probabilidade é dada por

$$f_Y(y|\mu, \sigma, p_1) = \begin{cases} (1-p_1) [B(\alpha, \beta)]^{-1} y^{\alpha-1} (1-y)^{\beta-1} & \text{se } 0 < y < 1 \\ p_1 & \text{se } y = 1 \end{cases}$$

Onde $\alpha = \mu(1 - \sigma^2)/\sigma^2$, $\beta = (1 - \mu)(1 - \sigma^2)/\sigma^2$ e $B(\alpha, \beta)$ é a função Beta

Com $0 < \mu < 1$, $0 < \sigma < 1$.

O parâmetro μ (que pode ser denominado “parâmetro de escala”) pode ser interpretado como o valor médio da distribuição Y condicional à informação de que $Y < 1$. Similarmente, é possível demonstrar que a variância da distribuição Y condicional à informação de que $Y < 1$ é diretamente proporcional à σ^2 (que pode ser denominado “parâmetro de dispersão”).

Distribuição E.2 – Distribuição Binomial Negativa

Se Y é uma variável aleatória com distribuição Binomial Negativa de parâmetros μ, σ , então a sua função de densidade de probabilidade é dada por

$$f_Y(y|\mu, \sigma) = \frac{\Gamma\left(y + \frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \left(\frac{\sigma\mu}{1 + \sigma\mu}\right)^y \left(\frac{1}{1 + \sigma\mu}\right)^{1/\sigma}$$

Onde $\Gamma(\alpha)$ é a função Gama

Com $\mu > 0$, $\sigma > 0$ e $y = 1, 2, \dots$

O parâmetro μ (que pode ser denominado “parâmetro de escala”) pode ser interpretado como o valor médio da distribuição Y . É possível demonstrar também que a variância da distribuição Y é dada pela expressão $\mu + \sigma\mu^2$. O parâmetro σ pode ser denominado “parâmetro de dispersão”.

Apêndice F – Modelos

Algoritmo F.1 Procedimento de seleção de variáveis

Passo A_0 : Ajusta-se um modelo inicial M_0 com todas as variáveis disponíveis

Repetir até $|S_i| = 0$

- Passo R. Remove-se a variável com maior nível descritivo para o teste de Wald de inexistência de efeito do modelo M_{i-1}
- Passo A: Repete-se o ajuste para o modelo M_i sem a variável

Onde o conjunto de variáveis com nível descritivo superior a 5% na i -ésima iteração do algoritmo é representado por S_i .

ModeloF.1 – Modelo para o Índice de Herfindahl Nacional (HNA)

Assume-se que o HNA_r da r -ésima cidade tem distribuição $BINF1(\mu_r, \sigma, \tau)$, onde

$$\log \frac{\mu_r}{1 - \mu_r} = \beta_0 + \beta_1 Lpop_r + \beta_2 PIBp_r + \beta_3 IDHM_r + \beta_4 HOMp_r + \beta_5 MATRp_r + \beta_6 TRABp_r + \beta_7 distmd_r$$

$$(Cidade)r = 1, 2, \dots, 319$$

Supõe-se adicionalmente que HNA_{r_i} e HNA_{r_j} são independentes para quaisquer $r_i \neq r_j$.

ModeloF.2 – Modelo para o Índice de Herfindahl Estrangeiro (HES)

Assume-se que o HES_r da r -ésima cidade tem distribuição $BINF1(\mu_r, \sigma, \tau)$, onde

$$\log \frac{\mu_r}{1 - \mu_r} = \beta_0 + \beta_1 Lpop_r + \beta_2 PIBp_r + \beta_3 IDHM_r + \beta_4 HOMp_r + \beta_5 MATRp_r + \beta_6 TRABp_r + \beta_7 distmd_r$$

$$(Cidade)r = 1, 2, \dots, 503$$

Supõe-se adicionalmente que HNA_{r_i} e HNA_{r_j} são independentes para quaisquer $r_i \neq r_j$

Modelo F.3 – Modelo para o total de lojas de empresas nacionais

Assume-se que o $Total\ de\ lojas_{ir}$ da i -ésima empresa na r -ésima cidade tem distribuição $BN(\mu_{ir}, \sigma)$, onde

$$\log \mu_{ir} = a_r + \beta_0 + \beta_1 Lpop_r + \beta_2 PIBp_r + \beta_3 IDHM_r + \beta_4 HOMp_r + \beta_5 MATRp_r + \\ \beta_6 TRABp_r + \beta_7 dist_{ir} + \beta_8 AT_{ir}$$

$$a_r \sim N(0, \sigma_r^2),$$

(Empresa) $i = 1, 2, \dots, 94$

(Cidade) $r = 1, 2, \dots, 542$

Supõe-se adicionalmente que a_{ri} e a_{rj} são independentes para quaisquer $r_i \neq r_j$

Modelo F.4 – Modelo para o total de lojas de empresas estrangeiras

Assume-se que o *Total de lojas*_{ir} da i-ésima empresa na r-ésima cidade tem distribuição $BN(\mu_{ir}, \sigma)$, onde

$$\log \mu_{ir} = a_r + \beta_0 + \beta_1 Lpop_r + \beta_2 PIBp_r + \beta_3 IDHM_r + \beta_4 HOMp_r + \beta_5 MATRp_r + \\ \beta_6 TRABp_r + \beta_7 dist_{ir} + \beta_8 AT_{ir}$$

$$a_r \sim N(0, \sigma_r^2),$$

(Empresa) $i = 1, 2, \dots, 9$

(Cidade) $r = 1, 2, \dots, 542$

Supõe-se adicionalmente que a_{r1} e a_{r2} são independentes para quaisquer $r_1 \neq r_2$