

Trabalho_Final_Mardegan

Gabriel_Mardegan

2023-12-09

Trabalho final da disciplina “Métodos Quantitativos de Pesquisa II”

Gabriel Rodrigues Mardegan (no USP: 8939269)

Professor Manoel Galdino

Replicação do estudo “City size and public service access: evidence from Brazil and Indonesia” (POST; KUIPERS, 2023) link

INTRODUÇÃO

O artigo selecionado como referência para o trabalho final da disciplina “Métodos Quantitativos de Pesquisa II”, intitulado “City size and public service access: evidence from Brazil and Indonesia” (POST; KUIPERS, 2023), tem como questão central avaliar como o tamanho populacional dos municípios está relacionado à oferta de serviços públicos (em termos de provisão e não de qualidade). Os autores questionam certa literatura de federalismo fiscal que aponta que as cidades maiores teriam melhor desempenho na provisão, por conta da economia de escala. Para isso, apresentam dados do Brasil e da Indonésia que indicam que cidades menores tendem a possuir, proporcionalmente, mais equipamentos públicos básicos como UBSs e escolas. São esses dados, referentes à população das cidades e a densidade de serviços públicos, que serão analisados no presente artigo.

As autoras enfatizam que o tamanho da população importa na análise da provisão de serviços públicos locais. Alegam que essa variável geralmente é deixada de lado em pesquisas de política econômica e políticas públicas, sendo apenas utilizada como variável controle. Quando ela é levada em conta, no caso da literatura de federalismo fiscal ou nos estudos sobre processos de descentralização, afirma-se que as cidades maiores tendem a ofertar mais serviços públicos, proporcionalmente.

Diante desse cenário, os dados analisados no artigo são o ponto de partida para questionar essa leitura predominante, e fornecem relevantes subsídios e evidências de que a população municipal é sim uma variável preditora importante para se avaliar provisão de equipamentos. Apresentados e discutidos esses dados, os autores analisam achados de pesquisas de campo realizadas em duas cidades do interior paulista. A partir de todo esse material, buscam teorizar os motivos para cidades menores terem mais prédios/equipamentos públicos de serviços básicos disponíveis para a população, em comparação com cidades mais populosas.

A regressão é usada enquanto modelo preditivo e, portanto, não há a pretensão de afirmar que o tamanho da cidade é a causa da densidade de serviços públicos, mas sim de apontar uma correlação entre essas duas variáveis. As análises realizadas pelas autoras, e que serão replicadas a seguir, indicam que quanto menor a cidade, maior a oferta proporcional de determinados bens públicos.

METODOLOGIA

Dados utilizados

Os autores do artigo disponibilizaram os códigos e os dados utilizados no desenvolvimento do trabalho no repositório “Harvard Dataverse”. Esses arquivos serviram de base para esta replicação.

O trabalho original envolve dados de dois países de renda baixa/média (LMICs na sigla em inglês), foco da pesquisa: Brasil e Indonésia. Devido a multiplicidade de variáveis envolvidas, tanto de variáveis dependentes analisadas separadamente (os serviços públicos), quanto de variáveis controle, optou-se limitar a replicação aos dados do Brasil, para melhor enfoque das análises dos dados e dos modelos.

Desenho de pesquisa

Deste modo, o trabalho será dividido em quatro análises principais: a. replicação da análise do quantitativo proporcional de UBSs (unidades básicas de saúde) como variável resposta; b. replicação da análise do quantitativo proporcional de escolas como variável dependente (sendo esses dois primeiros serviços o foco dos autores); c. replicação da análise da proporção da oferta de energia elétrica nos municípios como variável predita; d. utilização do quantitativo de trabalhadores da categoria de serviço social na área de assistência social do município como nova variável dependente, sendo essa última uma extensão do trabalho original. A taxa/percentual de energia elétrica foi escolhida, pois pode ser analisada pela perspectiva de um teste placebo, a ser explicado adiante.

Serão criados os bancos de dados referentes às variáveis elencadas acima, e serão realizadas análises críticas acerca dos pressupostos do modelo ajustado.

Criando o banco de dados

Primeiramente, foi criado um nome para o caminho do arquivo de dados referentes ao Brasil utilizados no estudo (em formato .csv). A partir desse arquivo foi criado o banco de dados “br_dados”, contendo 43 variáveis para as 5566 observações (municípios). Grande parte desse conjunto de variáveis foi utilizada em análises complementares do artigo que não serão abordadas nessa replicação, cujo foco são as regressões lineares desenvolvidas.

Em seguida, foram filtrados os municípios considerados “urbanizados” de acordo com uma categorização binária do IBGE. Realizada a seleção, o banco de dados ficou com 1588 municípios. Como o dataframe original do repositório não apresenta a população dos municípios em escala logarítmica, foi criada uma nova coluna “pop_log”.

Com o dataframe “br_dados” pronto para ser utilizado, foram criadas novas bases específicas para cada regressão a ser rodada.

No caso da UBS como variável resposta principal, foram selecionadas 9 variáveis: população em escala logarítmica (variável preditora), quantidade de UBSs por 1000 habitantes (principal variável resposta), e sete variáveis controles. Essas variáveis controle foram incluídas no estudo com a finalidade de descartar explicações teóricas alternativas, e foram categorizadas pelos autores em quatro grupos: processos de urbanização, diversidade, desenvolvimento econômico e partidarismo.

Deste novo dataframe, nomeado “dados_ubs”, foram omitidos os valores não observados (NAs), resultando em 1510 observações para as 9 variáveis. O mesmo foi feito para o banco de dados das escolas per capita e também para a taxa de fornecimento de eletricidade, criando, respectivamente, as bases “dados_esc” e “dados_elec”, ambas também com 1510 observações para 9 variáveis. Na tabela com os resultados das regressões, o paper aponta que o número de observações utilizado foi de 1509, portanto uma a menos que a utilizada nesta replicação. Não foi encontrado o motivo para essa diferença.

Para complementar os achados do artigo, esse trabalho optou por avaliar também a correlação entre serviços de assistência social e o tamanho populacional do município, abordando, portanto, um novo setor da esfera pública. O foco do paper é nas instalações físicas de serviços públicos (UBSs, escolas) devido à teorização de que a inauguração desses equipamentos estaria ligada a reivindicação de crédito político por vereadores, prefeitos e outros agentes políticos. Entretanto, os autores também avaliam a relação entre dados de recursos

humanos (professores e médicos per capita) com a população. O SUAS, Sistema Único de Assistência Social, prevê a instalação de centros de referências, os CRASs, cujos dados poderiam ser utilizados neste estudo.

Entretanto, como não é um serviço tão capilarizado quanto escolas e postos de saúde, optou-se por utilizar dados de recursos humanos do SUAS, considerando que seria uma variável que pode refletir melhor a provisão de serviços nessa área. Primeiramente, foi cogitado a utilização do quantitativo de profissionais de ensino superior trabalhando na assistência social, uma vez que é um trabalho fortemente multiprofissional, mas como esse dado é coletado como intervalo de valores, utilizou-se o quantitativo de assistentes sociais por município.

Para a criação do dataframe “dados_suas” foram utilizadas informações disponibilizadas no Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome, mais especificamente do censo SUAS 2017. Foi utilizado esse censo específico, pois além de ser o mais recente disponível no site do ministério, é um ano próximo aos dos outros dados utilizados pelos autores do artigo.

A planilha do ministério foi unificada com o dataframe “dados_br” através do código identificador do IBGE. Após a seleção das variáveis de interesse, a criação de nova variável de assistentes sociais/1000hab, e exclusão dos NAs, obteve-se um banco de dados com 1497 observações, próximo ao utilizado com as outras variáveis.

RESULTADOS

a. UBS como variável resposta

Table 1: Relação entre população(log) e UBS per capita

	prop_comm_health
pop_log	−0.124*** (0.008)
pop_growth	−0.082*** (0.022)
density	−0.004 (0.004)
div_index	−0.008 (0.037)
avg_salary_mo_2010	−0.00004 (0.00003)
ill_rate_2010	0.415*** (0.081)
elec_index	0.038 (0.034)
align	−0.011 (0.009)
Constant	0.804*** (0.046)
N	1,510
R ²	0.320
Adjusted R ²	0.316
Residual Std. Error	0.139 (df = 1501)

*p < .1; **p < .05; ***p < .01

A tabela 1, acima, aponta que os coeficientes obtidos na regressão linear são os mesmos apresentados pelos autores na tabela ao final do artigo. Primeiramente, isso confirma que os dados utilizados nesta replicação foram os mesmos, e os códigos similares. O coeficiente referente ao preditor população (em log) indica que há, de fato, uma correlação negativa entre população do município e oferta de UBSs (em termos proporcionais). Tudo o mais constante, ou seja, com todas as variáveis controle fixadas, o aumento de uma unidade na população (em log) representa uma diminuição em 0.124 na proporção de UBSs no município. Esse resultado se alinha às interpretações do artigo de que os municípios menores possuem, proporcionalmente, mais serviços públicos à disposição de seus cidadãos.

Para aprofundar e contextualizar a análise do coeficiente, calculamos a média nacional de UBSs por habitante: 0.27/mil habitantes (aproximadamente 1 UBS para cada quatro mil habitantes). Caso o efeito da população sobre a oferta de serviços fosse zero, uma cidade de 10000 habitantes (log 10k = 4) teria aproximadamente 3 UBSs. Já uma cidade de 100000 habitantes (log 100k = 5) teria 27 UBSs, em média. No

entanto, o efeito encontrado na regressão é que a proporção diminui quanto maior a população. Portanto, comparar uma cidade de 1000 habitantes ($\log 1k = 3$) com uma cidade de 100000 habitantes ($\log 100k = 5$), uma diferença de 2 log na população, implicaria, em média, a uma diminuição de 0,24 na proporção de UBSs. Ao invés de 27 UBSs da média calculada, a oferta esperada seria reduzida para aproximadamente 20 UBSs.

Foi também gerado um gráfico de dispersão simples com as duas variáveis de interesse para melhor observação da distribuição das variáveis (apêndice 1). Nota-se a tendência de municípios maiores terem menor proporção de UBSs, sendo que as observações vão se concentrando mais em torno do eixo x na parte direita do gráfico. Poucos municípios maiores de 100mil habitantes tem mais de 0,25 UBSs para 1000 habitantes, ou 2,5 UBSs para 10000 habitantes. Uma questão que poderia ser levantada é que cidades menores tendem a ter mais dispersão dos habitantes no território, e por isso constroem mais UBSs, o que não seria necessário em cidades grandes mais adensadas. Essa hipótese foi considerada pelos autores, que levantaram dados de georreferenciamento para testá-la. Entretanto, esse aspecto não será abordado neste trabalho.

Ainda sobre o gráfico de dispersão, percebe-se que na parte esquerda do gráfico, referente aos municípios menos populosos, existem “ranhuras” decrescentes. A base de dados foi analisada com maior atenção para tentar entender a origem desses padrões peculiares, porém não foi levantada nenhuma hipótese consistente. Algo a ser avaliado futuramente é se esses traços referem-se a certas determinações do SUS para implementação de UBSs, o que poderia levar a tais padrões. A partir do segundo terço do gráfico, esses padrões desaparecem, e as observações parecem distribuídas de maneira mais coesa.

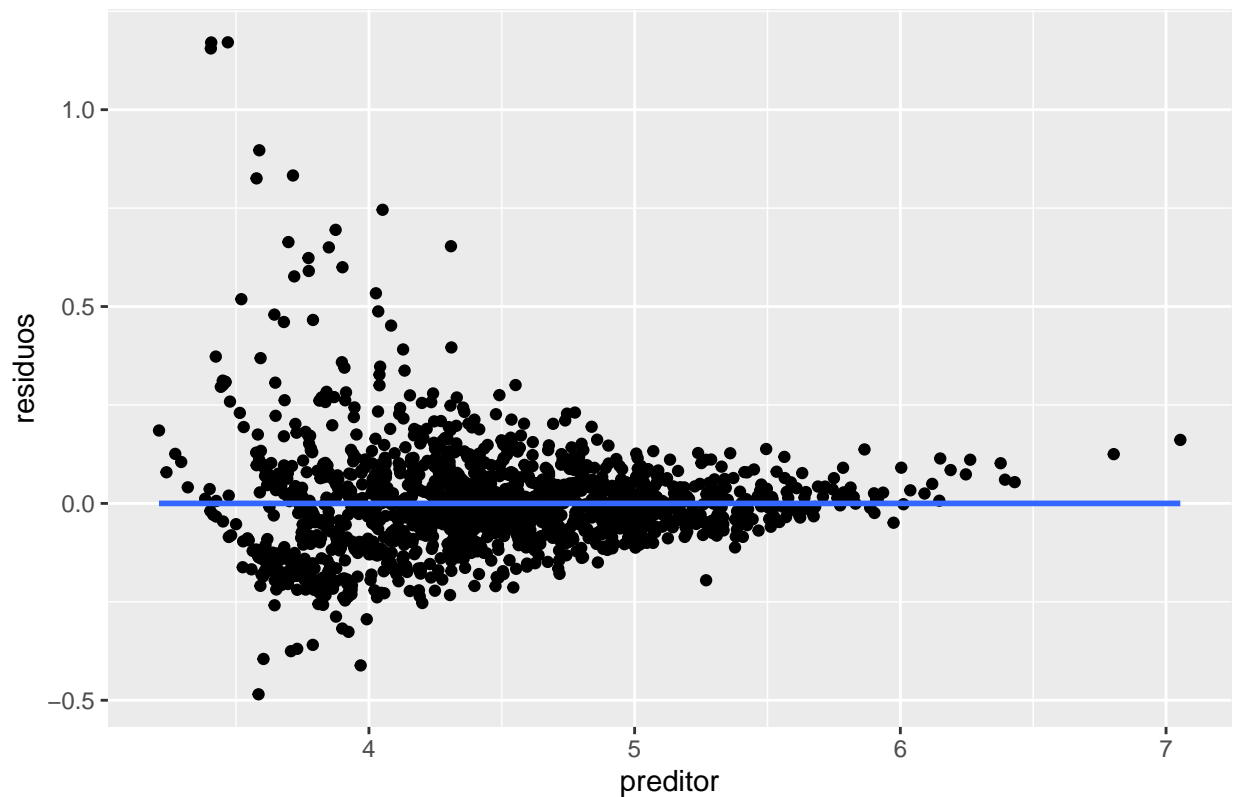
checagem dos pressupostos do modelo linear

Apresentados os coeficientes obtidos na regressão, é preciso analisar os dados em relação aos pressupostos de um modelo linear. Essa análise ocorre fundamentalmente a partir da observação dos resíduos das observações em relação ao modelo.

Uma das propriedades dos resíduos é que eles devem ter esperança zero condicional aos preditores. No gráfico abaixo, a reta horizontal azul em cima do eixo x indica que a média dos resíduos dessa primeira regressão é zero, contemplando, portanto, esse pressuposto. Entretanto, percebe-se que não há uma homogeneidade na dispersão dos resíduos, havendo um padrão evidente na parte inicial do gráfico, referente aos municípios de menor população. Muitos desses municípios apresentaram resíduos positivos em maior quantidade do que resíduos negativos, que estão mais concentrados em torno da reta azul. Ou seja, há muito mais UBSs nessas cidades (em termos proporcionais) do que prevê o modelo obtido. Trata-se de um indício de que o modelo linear não se ajusta tão bem para essa parte das observações.

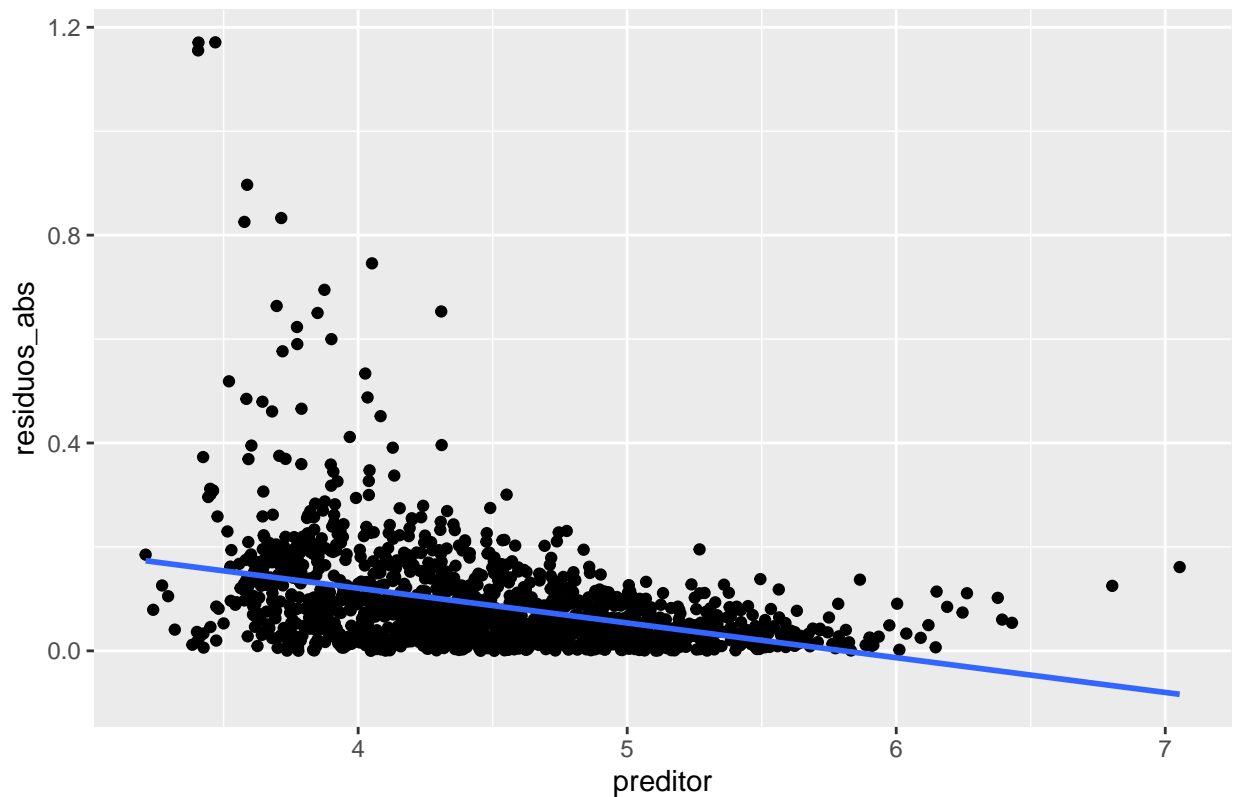
No meio do gráfico, por sua vez, é possível dizer que não há um padrão sistemático nos resíduos, uma vez que eles estão bem distribuídos em torno da reta. Nesse trecho é possível afirmar que o modelo linear funciona bem, ou seja, consegue fazer uma previsão com menos erros dos dados observados. Na parte final do gráfico, há poucas observações (provavelmente São Paulo e Rio de Janeiro, seguidas de outras capitais), também com maior densidade de UBSs do que prevê o modelo, mas não com uma discrepância como a que se percebe no início do gráfico.

Gráfico 1: resíduos contra o preditor



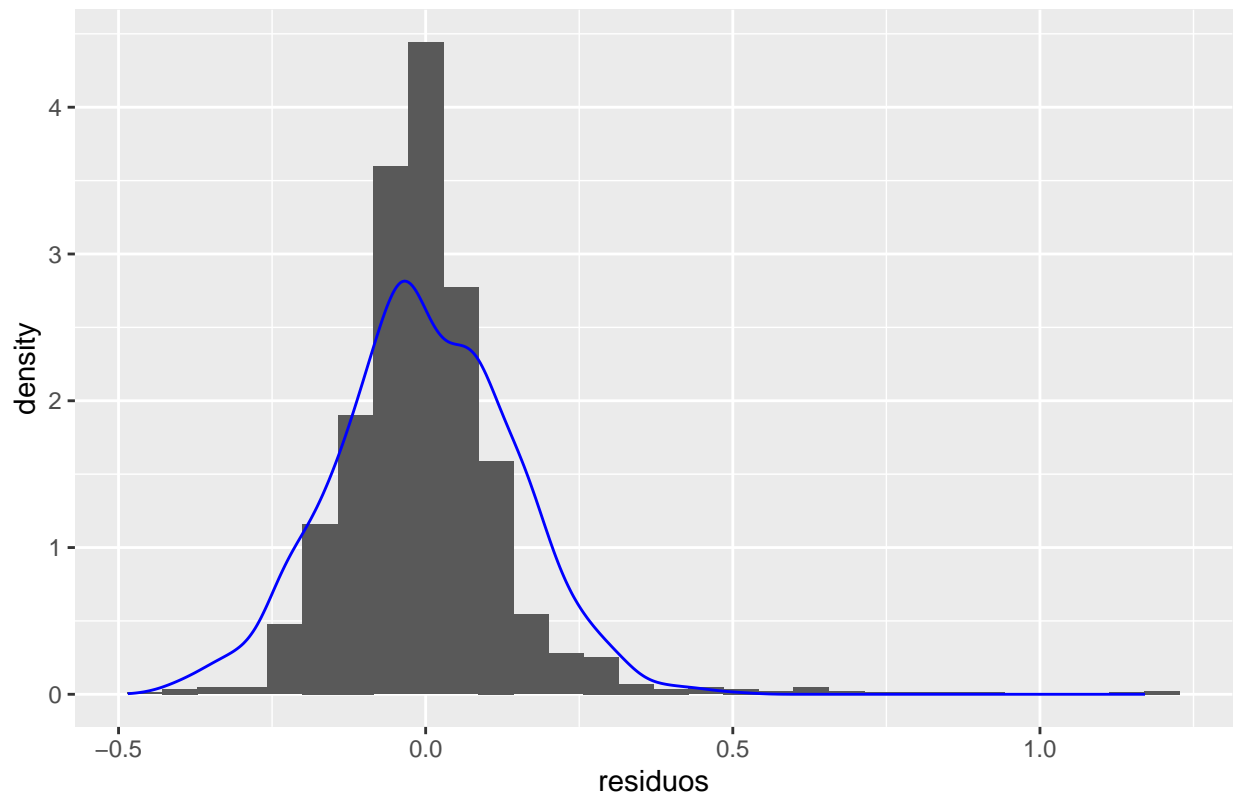
Outra análise acerca do modelo de regressão linear refere-se à suposição de homocedasticidade. A homocedasticidade prevê que os resíduos devem ter variância constante em relação ao preditor. No gráfico 2 abaixo, plotado com o valor absoluto dos resíduos (em vez do quadrado) o primeiro aspecto a ser realçado é que a reta azul não é horizontal, o que já acusa heterocedasticidade. Além disso, a distribuição dos resíduos também deixa patente que a variação é muito maior na parte inicial do gráfico. O meio do gráfico tem um padrão mais coeso de variâncias. A violação desse pressuposto prejudica a inferência estatística relativas aos dados da regressão.

Gráfico 2: homocedasticidade x heterocedasticidade

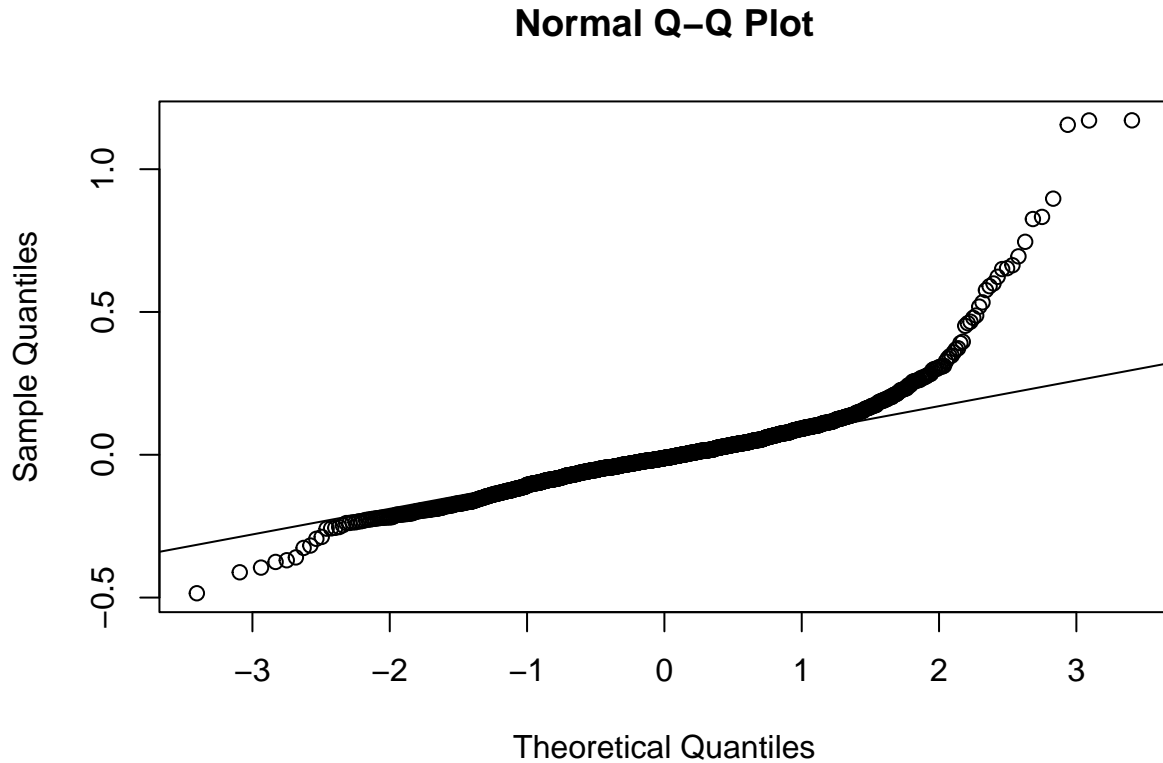


O gráfico 3 abaixo refere-se à verificação do pressuposto da normalidade dos resíduos. Observado o histograma dos resíduos, é possível perceber uma distribuição que se assemelha à normal. No entanto, nota-se um alongamento assimétrico na cauda direita, evidenciando comentário sobre o primeiro pressuposto de que há muitos municípios com mais UBSs do que prevê o modelo. Por conta dessa observação, não é possível dizer que o pressuposto foi atendido.

Gráfico 3: normalidade dos resíduos



O gráfico QQplot (quantile-quantile plot), abaixo, também é usado para avaliar se os resíduos do modelo seguem uma distribuição normal. A linha reta no centro do gráfico representa a distribuição teórica normal esperada. Se os pontos estiverem próximos ou alinhados com essa linha, os resíduos podem ser considerados aproximadamente normais. Os pontos representam os resíduos do seu modelo, comparados com os quantis esperados da distribuição normal.



Na região central (próxima de 0 no eixo dos quantis teóricos), os pontos estão relativamente alinhados com a linha reta, indicando que os resíduos nesta faixa central seguem aproximadamente a distribuição normal. Porém, nas extremidades (caudas), especialmente à direita, há um desvio enorme em relação à linha de referência. Isso sugere a presença de resíduos extremos (outliers) ou uma cauda longa à direita, ou seja, uma distribuição assimétrica dos resíduos (viés positivo).

b. escolas como variável resposta

Table 2: Relação entre população(log) e escolas per capita

	<u>schools_prop</u>
pop_log	-0.353*** (0.033)
pop_growth	-0.319*** (0.095)
density	0.001 (0.017)
div_index	0.732*** (0.162)
avg_salary_mo_2010	-0.0004*** (0.0001)
ill_rate_2010	3.673*** (0.358)
elec_index	0.089 (0.150)
align	-0.033 (0.038)
Constant	2.113*** (0.201)
N	1,510
R ²	0.343
Adjusted R ²	0.339
Residual Std. Error	0.613 (df = 1501)

*p < .1; **p < .05; ***p < .01

Table 3: Coeficientes padronizados

var_resposta	coeficiente_pad
UBSs_por_1000	-0.42
escolas_por_1000	-0.27

A tabela 2 confirma os coeficientes encontrados pelos autores do artigo quando a variável dependente são as escolas per capita. Assim como nas UBSs, há uma correlação negativa entre densidade de escolas e população, o que sustenta o argumento central do paper. Todos os outros fatores constantes, a cada uma unidade de aumento na população municipal (em escala logarítmica), espera-se um decréscimo de 0.35 na proporção de escolas. Os coeficientes padronizados das UBSs e escolas serão discutidos e comparados a frente.

Também foi criado um gráfico de dispersão para a variável escola (apêndice 2). Percebe-se padrão similar ao das UBSs, com uma dispersão maior na parte esquerda. Os padrões de “ranhuras” também aparecem para as escolas, mas de maneira menos significativa em comparação com os postos de saúde.

avaliação dos pressupostos

A checagem dos pressupostos para essa variável será realizada a partir do conjunto de gráficos gerado pela biblioteca “easystats” (LÜDECKE et al., 2022), apresentado no apêndice 3.

O gráfico da linearidade tem uma linha de referência reta e horizontal, o que indica que o pressuposto foi contemplado. Entretanto, na parte final do gráfico, à direita, há alguns pontos que se afastam da reta, indicando que o modelo não funciona bem para certos municípios com número elevado de escolas per capita.

O gráfico “homogeneidade da variância” aponta que o pressuposto da homocedasticidade foi violado, uma vez que a linha de referência é curva, e há um comportamento destoante das variâncias na parte final do gráfico, à direita, com resíduos positivos.

O gráfico de normalidade dos resíduos sugere que esse pressuposto foi atendido, pois os pontos acompanham a linha de referência, com exceção de alguns valores extremos à direita.

comparando os coeficientes padronizados da “reg_ubs” e “reg_esc”

Os coeficientes apresentados na tabela 3, acima, são resultado da padronização dos dados utilizados nas regressões. Foram encontrados os mesmos coeficientes apresentados na figura 3 do artigo: -0.42 para as UBSs; -0.27 para as escolas. Com a padronização, é possível comparar esses dois resultados. Em ambos os setores há uma correlação negativa em relação à população, em consonância com o debate realizado pelos autores. Os dados indicam, ainda, que a densidade dos serviços básicos de saúde tende a diminuir com mais intensidade conforme aumenta a população dos municípios, em comparação com as escolas.

Há, portanto, uma diferença no padrão de provimento desses dois setores da esfera pública, o que poderia suscitar uma linha específica de pesquisa: por qual razão as escolas tendem a ser mais bem distribuídas (em termos populacionais, e não territoriais) do que as UBSs? Na realidade, talvez seja um reflexo do propósito e da utilização desses serviços. Uma UBS recebe muito mais usuários do que uma única escola, o que demandaria maior oferecimento de escolas, fato que responderia a questão acima.

c. percentual de provisão de energia elétrica como variável resposta

Os coeficientes da regressão linear, apresentados acima, são os mesmos daqueles obtidos no artigo original. O coeficiente indica que o tamanho populacional afeta de maneira quase irrisória a provisão de energia elétrica nos municípios. Esse achado está em consonância com a hipótese dos autores de que, por se tratar de bem público fornecido pelo governo estadual, os fenômenos políticos relacionados ao tamanho da população não atuam nessa provisão em particular. Pontua-se que os autores investigaram os efeitos das dinâmicas políticas locais sobre a provisão de serviços, e a oferta de serviços estaduais estaria fora dessa lógica.

O uso dessa variável no estudo pode ser considerado um teste placebo. Testes placebo são ferramentas úteis para avaliar a plausibilidade das suposições subjacentes a um desenho de pesquisa. Servem para

Table 4: Relação entre população(log) e provisão de energia elétrica

	electr_share_2010
pop_log	−0.002*** (0.001)
pop_growth	−0.002 (0.002)
density	0.0003 (0.0003)
div_index	−0.010*** (0.003)
avg_salary_mo_2010	0.00001*** (0.00000)
ill_rate_2010	−0.062*** (0.006)
elec_index	0.004 (0.003)
align	−0.0001 (0.001)
Constant	1.002*** (0.003)
N	1,510
R ²	0.244
Adjusted R ²	0.240
Residual Std. Error	0.011 (df = 1501)

*p < .1; **p < .05; ***p < .01

checar um possível desvio dessas suposições (EGGERS; TUÑÓN; DAFOE, 2023). No caso em tela, como os pesquisadores levantam a hipótese de que a provisão de serviços públicos está conectada com dinâmicas políticas locais (e não são apenas reflexo de uma economia de escala), o tamanho populacional dos municípios não deveria ter efeito sobre esse serviço, uma vez que é de responsabilidade de outro ente federativo.

Foi gerado um gráfico de dispersão para essa variável (apêndice 4). O gráfico evidencia o que os autores mencionam no artigo: em muitos municípios o fornecimento de energia elétrica já é universalizado, sobretudo em cidades urbanizadas, que é a amostra do estudo. Há dois padrões diferentes quando se compara as duas metades do gráfico. Na primeira metade, há maior dispersão das observações, indicando que no grupo dos municípios menores há maior número de cidades que não alcançaram a universalização desse serviço. Já na metade direita do gráfico, a grande maioria dos pontos se concentra próximo ao eixo do 100%.

Nota-se no gráfico que a reta ajustada é levemente positiva. Entretanto, o coeficiente, mesmo que muito pequeno, é negativo, o que indica a influência das variáveis controle.

avaliação dos pressupostos

A avaliação dos pressupostos da regressão da variável resposta eletricidade será realizada através dos gráficos gerados pelo pacote “easystats” (apêndice 5).

O gráfico atinente à normalidade dos resíduos indica que esse pressuposto foi atendido, apesar de dados destoantes na parte esquerda do gráfico, que fogem à reta de referência. Esse aspecto está em consonância com o que foi discutido acerca do gráfico de dispersão, uma vez que há uma quantidade razoável de municípios que escapam ao modelo por não terem alcançado a universalização da eletrificação, aumentando os resíduos negativos.

O gráfico “homogeneidade da variância” aponta que o pressuposto da homocedasticidade foi violado. A linha de referência é curva, e as variâncias apresentam um padrão destoante no meio do gráfico.

Em relação à linearidade, a linha de referência é reta e horizontal. Esse pressuposto foi atendido, mesmo havendo um padrão destoante das observações no meio do gráfico, o que indica que o modelo não se ajusta bem a certos municípios, aqueles ainda distantes de alcançar valores próximos a 100% da eletrificação.

d. assistentes sociais como variável resposta

Essa terceira regressão utilizou uma nova variável para avaliar as suposições do artigo. Assim como a regressão referente à provisão de UBSs, observa-se uma correlação negativa entre o tamanho populacional dos municípios e o número de assistentes sociais per capita. Tudo o mais constante, a cada aumento de

Table 5: Relação entre população(log) e profissionais SUAS per capita

	assist_prop
pop_log	-0.193*** (0.009)
pop_growth	0.044 (0.027)
density	0.013*** (0.005)
div_index	-0.293*** (0.046)
avg_salary_mo_2010	0.0002*** (0.00004)
ill_rate_2010	0.310*** (0.102)
elec_index	0.101** (0.043)
align	0.007 (0.011)
Constant	1.103*** (0.057)
N	1,497
R ²	0.265
Adjusted R ²	0.261
Residual Std. Error	0.174 (df = 1488)

*p < .1; **p < .05; ***p < .01

Table 6: Coeficiente padronizado assist. social/1000hab

var_resposta	Coeficiente_pad
assist_social_por_1000	-0.55

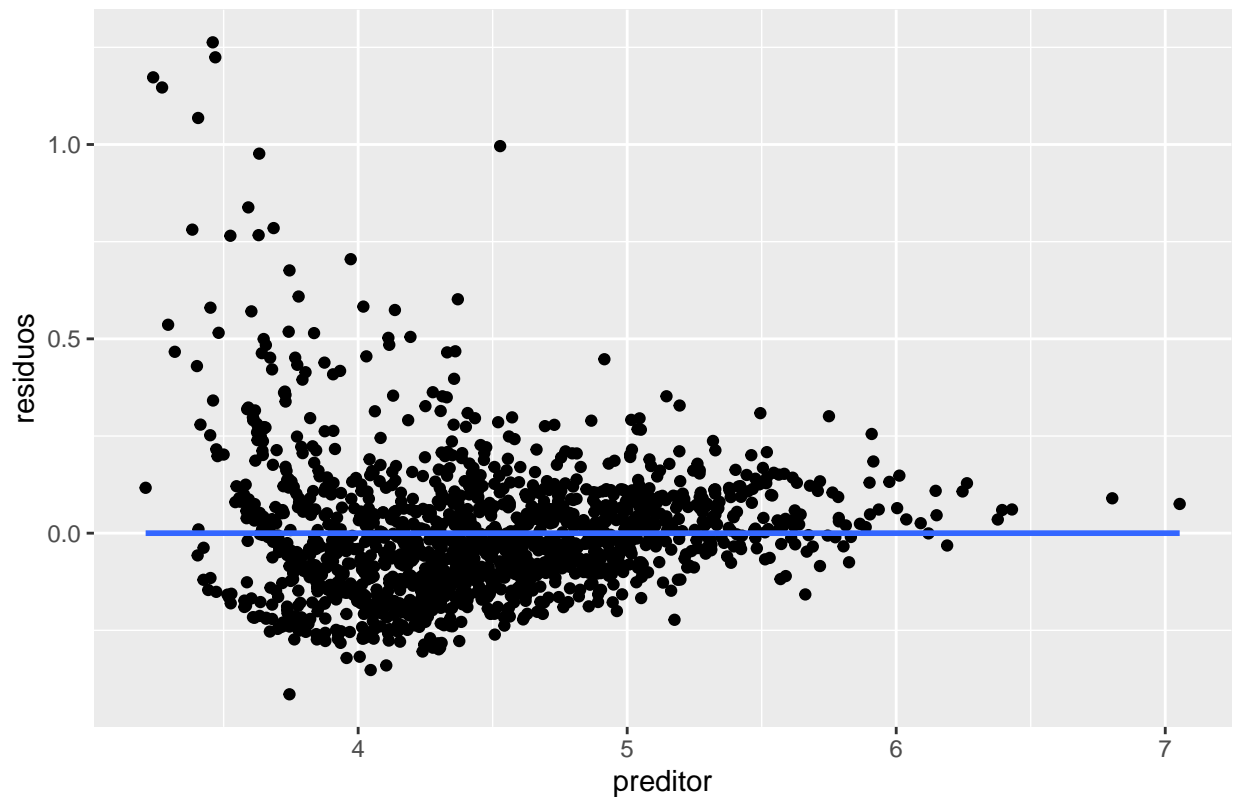
uma unidade da população em escala logarítmica, há uma diminuição de 0.193 no quantitativo de assistentes sociais na equipe municipal para 1000 habitantes.

A tabela 6, acima, apresenta o coeficiente padronizado do preditor população(log): -0.55. Em termos absolutos, esse coeficiente é maior do aqueles encontrados pelos autores para as variáveis médico per capita e professores per capita, ambas próximas de 0.3. Essa diferença indica que o tamanho dos municípios impacta mais a disponibilidade de recursos humanos no setor da assistência social, do que nos setores da educação e da saúde. Uma hipótese para tal achado, é de que se trata de um sistema novo, o que dificultaria sua expansão e capilaridade em grandes cidades, enquanto que cidades menores teriam fundos de outras esferas que permitiram uma estruturação mais rápida do SUAS.

testagem dos pressupostos

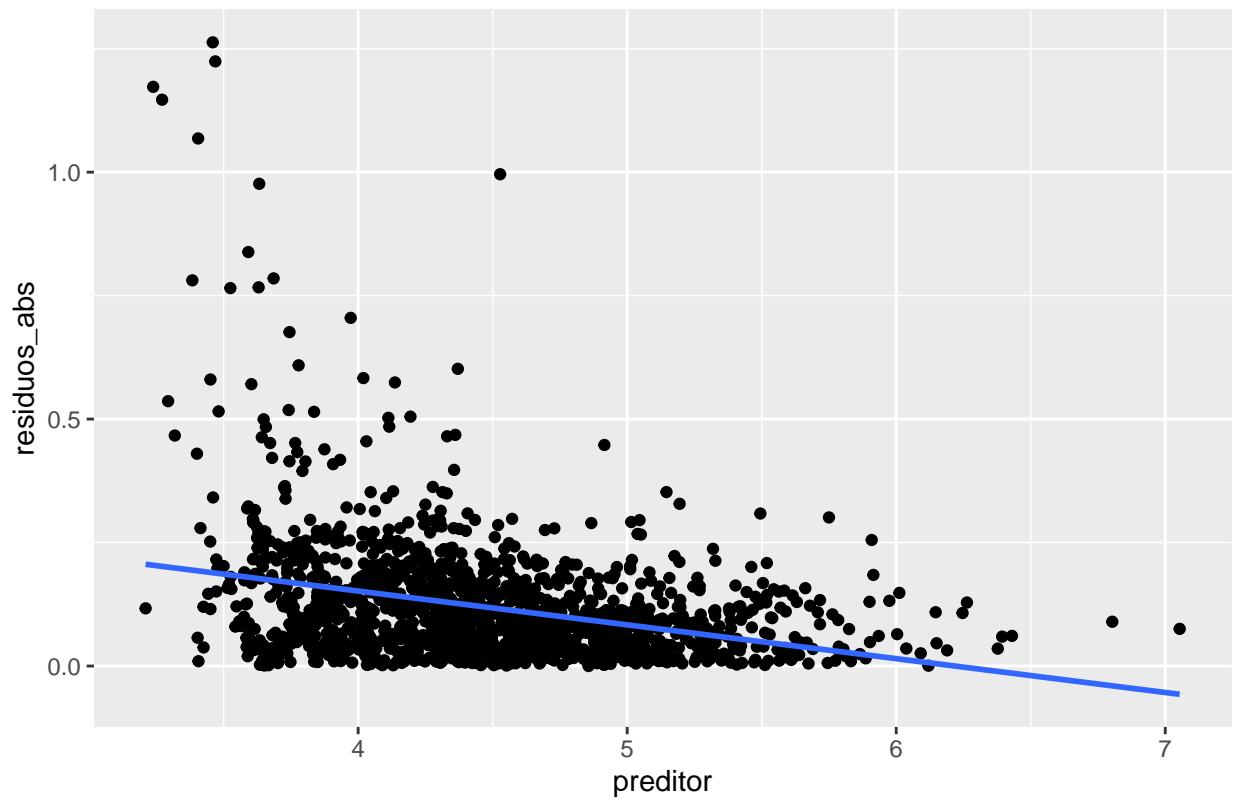
A checagem do modelo linear para a regressão com a variável assistentes sociais será realizada com o mesmo passo a passo da checagem com a variável UBS.

Gráfico 4: resíduos contra o preditor



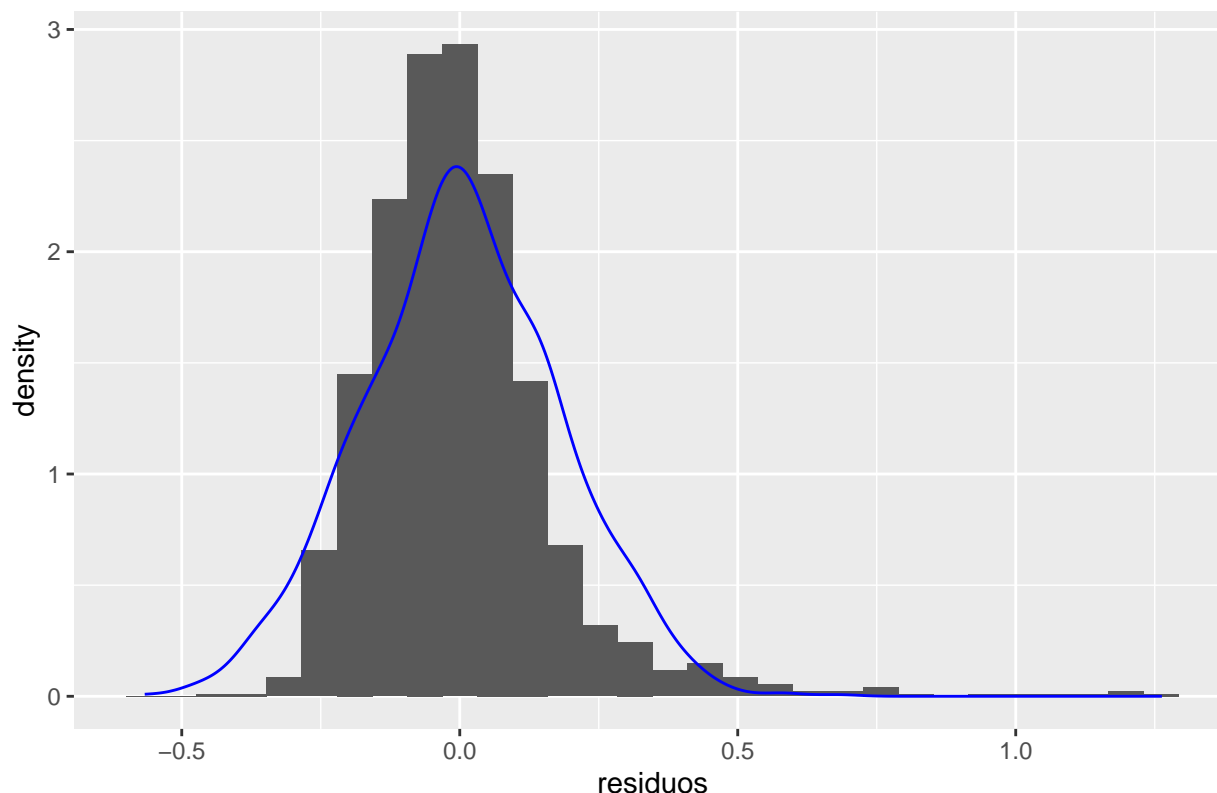
O gráfico 4 evidencia que o pressuposto da linearidade foi contemplado: a linha de referência é horizontal e segue o eixo x. Entretanto, a dispersão dos resíduos mostra diferentes padrões de distribuição de acordo com o valor do preditor. A primeira parte do gráfico, dos municípios menores, tem muitos valores distantes da reta de referência, com resíduos positivos. É o mesmo padrão encontrado nos outros serviços públicos. Indica que a variância desse subgrupo de cidades não se encaixa tão bem a esse modelo linear, pois muitas observações se distanciam dos valores esperados. Na parte central do gráfico, os valores tendem a se concentrar em torno da linha de referência, o que indica que para esse grupo de cidades o modelo linear funciona bem.

Gráfico 5: homocedasticidade x heterocedasticidade



O gráfico 5 indica violação do pressuposto de homocedasticidade das variâncias. A reta de referência não é horizontal, e não há uma coesão das variâncias em torno da reta, em quase nenhum trecho do eixo dos preditores. Novamente, salta à vista o padrão destoante no início do gráfico, com vários municípios menos populosos cuja variância dos resíduos é muito alta.

Gráfico 6: normalidade dos resíduos



Percebe-se uma normalidade na distribuição dos resíduos, mas também se nota uma cauda para a direita, dos resíduos positivos, inclusive com a presença de alguns outliers. Apesar de poucos casos, os extremos nesse lado da curva indicam a presença de municípios com muito mais assistenciais sociais per capita do que prevê o modelo linear ajustado.

CONSIDERAÇÕES FINAIS

Para além das questões didáticas próprias a este trabalho e à disciplina de metodologia quantitativa da qual ele é avaliação final, a replicação apresentada acima foi exitosa em alcançar os mesmos parâmetros calculados no artigo original, tanto para as duas variáveis centrais (UBSs e escolas) quanto para uma variável que pode ser analisada enquanto teste placebo (eletricidade). Foi possível também ampliar o trabalho original a partir da análise de uma nova variável que, mesmo não sendo uma estrutura física, reflete oferta de serviços públicos: quantitativo per capita de assistentes sociais trabalhando no Sistema Único de Assistência Social. Também foi exitosa na realização da verificação dos pressupostos do modelo ajustado, pois permitiu ampliar a discussão sobre os dados trabalhados, ajudando a complexificar a análise original.

Em todas as regressões, o coeficiente calculado para a população como variável preditora foi negativo, confirmando a tese principal dos autores: o tamanho das cidades importa para avaliar a oferta de serviços públicos; cidades menores tendem a oferecer, proporcionalmente, mais serviços públicos para seus municípios. A comparação dos coeficientes padronizados das variáveis UBS e escola indica, porém, que esses setores se comportam de maneira diferente, sendo que no caso da saúde, o tamanho da cidade tende a impactar mais a proporção de postos de saúde, o que pode ser um reflexo de questões inerentes à própria política pública.

A análise do fornecimento de energia elétrica foi importante para verificar o teste placebo em face da teoria dos autores. O acréscimo de uma nova variável (profissionais do SUAS) pode suscitar novas pesquisas setoriais e de campo para compreender se ocorrem dinâmicas próprias para esse setor do serviço público.

A análise do modelo linear nas quatro variáveis abordadas leva a algumas reflexões pertinentes sobre os dados trabalhados no estudo. O ajuste do modelo linear para tratar tais dados mostrou-se útil para evidenciar o cerne da pesquisa: há uma correlação negativa entre população e oferta de serviços públicos em termos proporcionais. A partir desse achado, foi possível aprofundar o tema da pesquisa, e também seguir com os estudos de campo que geraram o paper. Entretanto, a violação de alguns pressupostos, discutido acima, aponta que para certos agrupamentos de dados o modelo linear talvez não seja o mais adequado. Em todas as variáveis, foi apontado que em municípios menores a distribuição das observações é mais dispersa, e ficam distantes aos valores esperados na reta de regressão. Portanto, para esses municípios, muita informação se perde em relação ao modelo linear.

Nos gráficos para análise do pressuposto de linearidade, foi notado certo padrão peculiar também no agrupamento de municípios menores. As observações se configuram em traços curvos e descendentes, nomeados aqui de “ranhuras”. Não foi possível teorizar sobre tal padrão apenas com a observação dos dados, mas podem ser reflexo de algum aspecto das políticas públicas analisadas, e/ou da maneira como os dados são coletados/configurados. Trata-se, portanto, de fenômeno a ser considerado em pesquisas futuras.

REFERÊNCIAS

EGGERS, A. C.; TUÑÓN, G.; DAFOE, A. Placebo Tests for Causal Inference. *American Journal of Political Science*, 2023.

HLAVAC; MAREK. stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. 2022 <https://CRAN.R-project.org/package=stargazer>

LÜDECKE, D.; PATIL, I.; BEN-SACHAR, M. S.; WIERNIK, B. M.; BACHER, E.; THÉRIAULT, R.; MAKOWSKI, D. Easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting. CRAN, 2022. Available from <https://easystats.github.io/easystats/>

POST, A. E.; KUIPERS, N. City Size and Public Service Access: Evidence from Brazil and Indonesia. *Perspectives on Politics*, vol. 21, n. 3, p. 811-830, 2023.

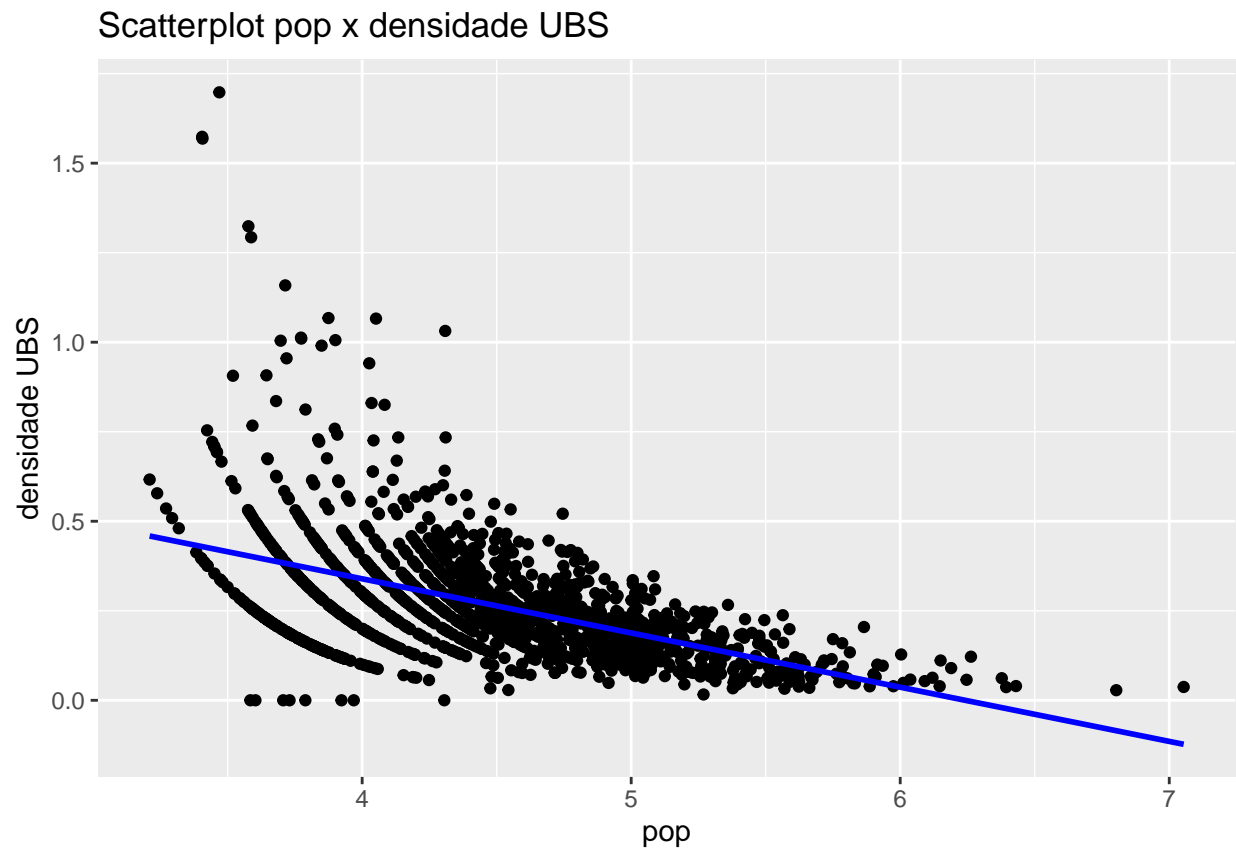
FONTE DOS DADOS

MINISTÉRIO DO DESENVOLVIMENTO E ASSISTÊNCIA SOCIAL, FAMÍLIA E COMBATE À FOME. Censo SUAS 2017, microdados/assistência social, 2017. Disponível em: <https://www.gov.br/mds/pt-br/noticias-e-conteudos/dados-e-ferramentas-informacionais/microdados> (acesso em 07 de janeiro de 2024)

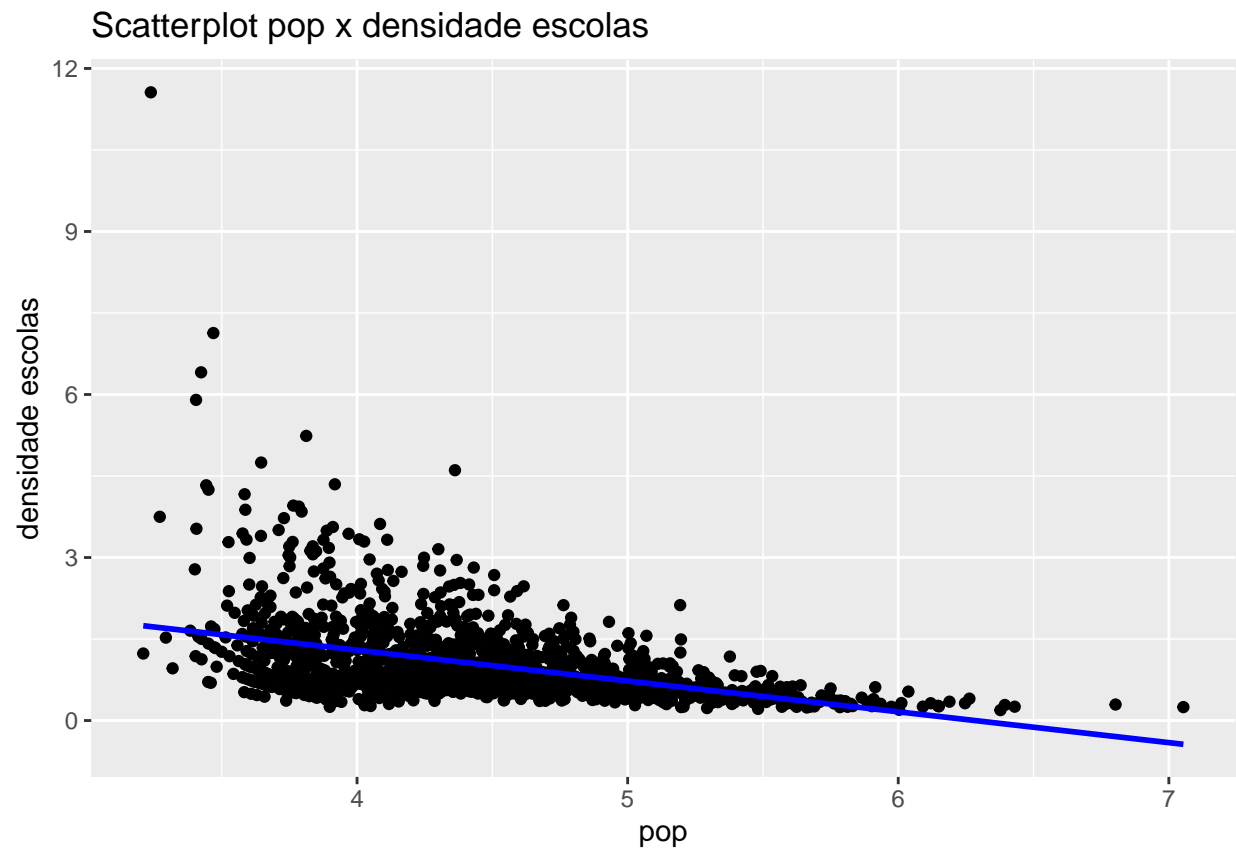
POST, A.; KUIPERS, P. Replication Files for “City Size and Public Service Access: Evidence from Brazil and Indonesia.”. Harvard Dataverse, 2023. Disponível em: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KGTRVF> (acesso em 07 de janeiro de 2024)

APÊNDICES

Apêndice 1: gráfico de dispersão população(log) x UBS per capita



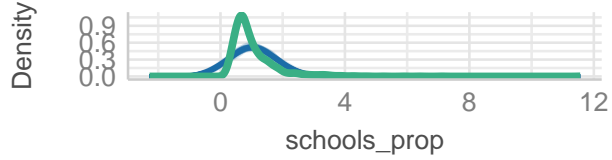
Apêndice 2: gráfico de dispersão população(log) x UBS per capita



Apêndice 3: gráficos para checagem do modelo linear - escolas como v. resposta

Posterior Predictive Check

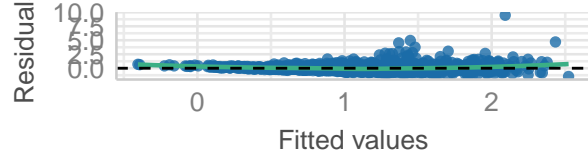
Model-predicted lines should resemble observed data



— Observed data — Model-predicted data

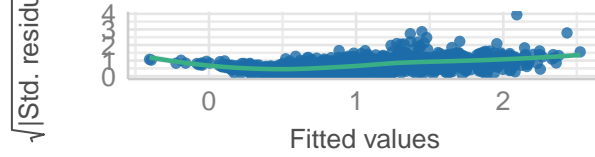
Linearity

Reference line should be flat and horizontal



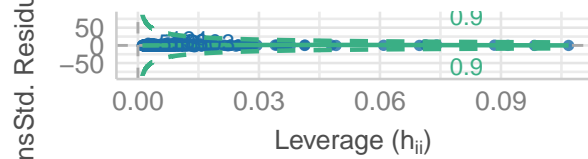
Homogeneity of Variance

Reference line should be flat and horizontal



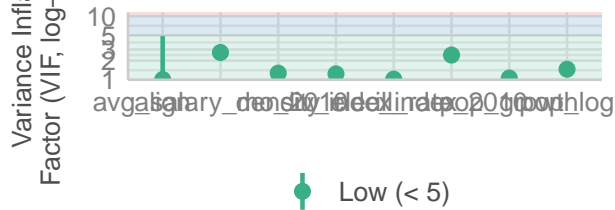
Influential Observations

Points should be inside the contour lines



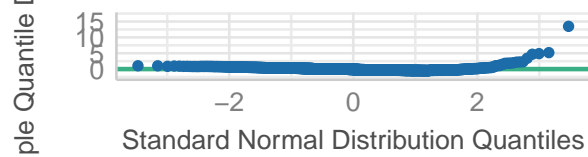
Collinearity

High collinearity (VIF) may inflate parameter uncertainty

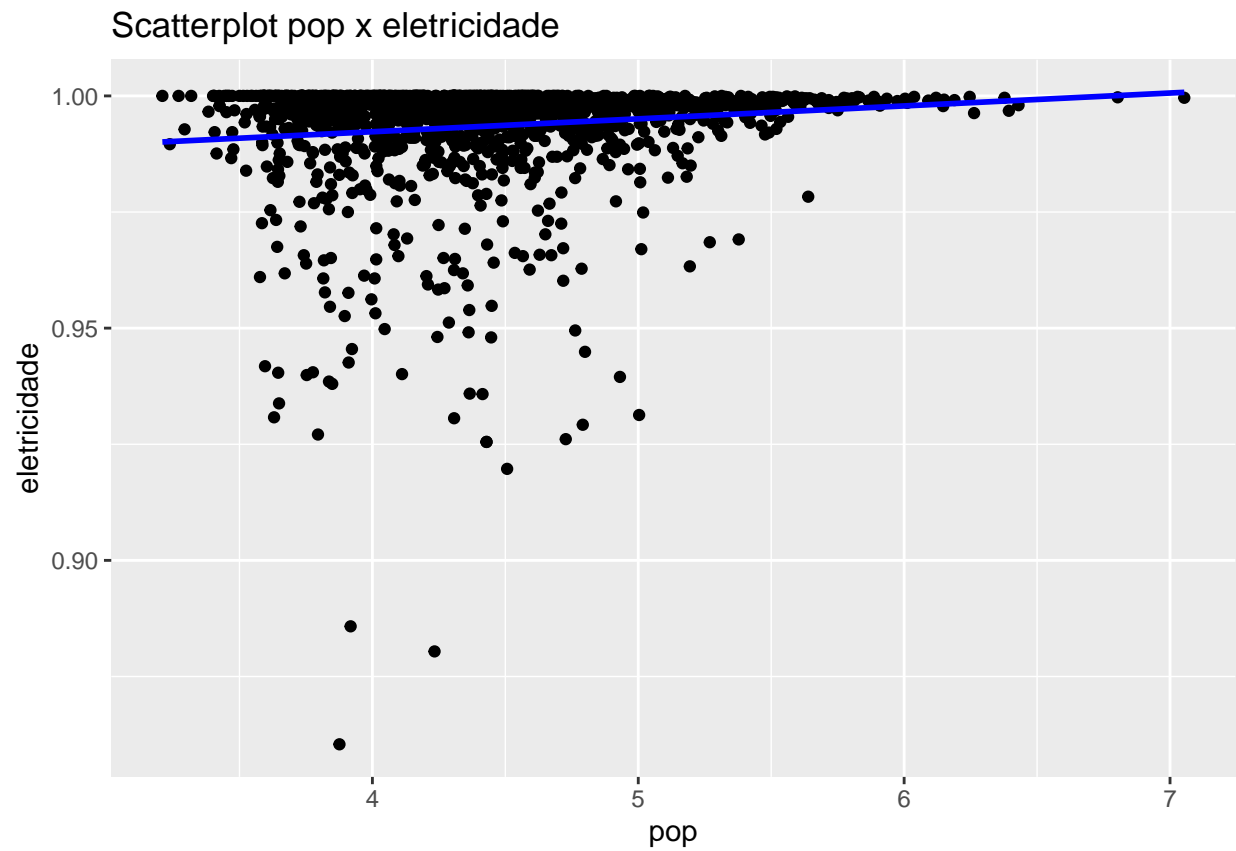


Normality of Residuals

Points should fall along the line



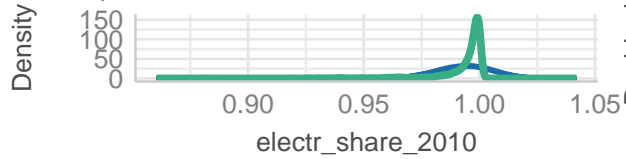
Apêndice 4: gráfico de dispersão população(log) x provisão eletricidade



Apêndice 5: gráficos para checagem do modelo linear - eletricidade como v. resposta

Posterior Predictive Check

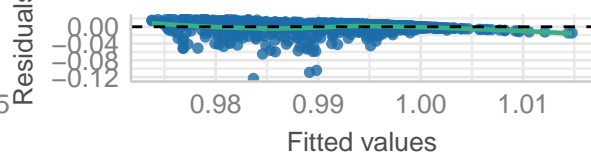
Model-predicted lines should resemble observed data



— Observed data — Model-predicted data

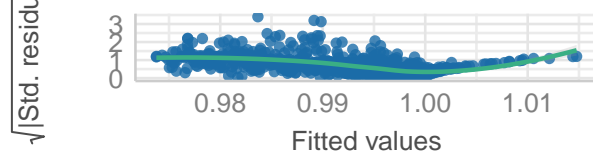
Linearity

Reference line should be flat and horizontal



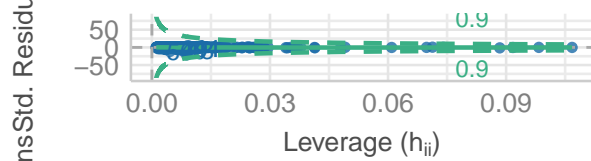
Homogeneity of Variance

Reference line should be flat and horizontal



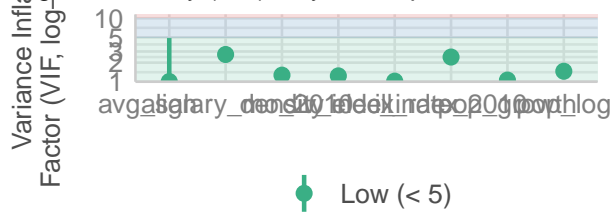
Influential Observations

Points should be inside the contour lines



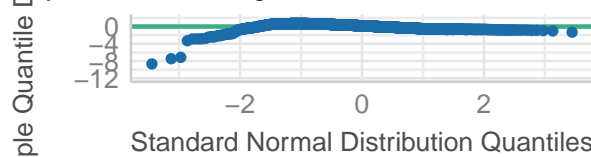
Collinearity

High collinearity (VIF) may inflate parameter uncertainty



Normality of Residuals

Points should fall along the line



Apêndice 6: gráfico de dispersão população(log) x assistente social per capita

Scatterplot pop x assist soc per capita

