



**Publicação
Preliminar**

INADEQUAÇÕES HABITACIONAIS E INFORMALIDADE NO MERCADO DE HABITAÇÃO BRASILEIRO. EVIDÊNCIAS EMPÍRICAS A PARTIR DO CENSO 2010.

Autora: Vanessa Gapriotti Nadalin
Produto editorial: Texto para Discussão (TD)
Cidade: Brasília
Editora: Instituto de Pesquisa Econômica Aplicada (Ipea)
Ano: 2022
Edição 1ª

O Ipea informa que este texto não foi objeto de padronização, revisão textual ou diagramação pelo Editorial e será substituído pela sua versão final uma vez que o processo de editoração seja concluído.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

1. INTRODUÇÃO

Como todos os mercados de bens, o mercado de habitação apresenta segmentações. Quais são os segmentos do mercado de habitação no Brasil? Tradicionalmente, os segmentos do mercado de uma cidade estão baseados em preços semelhantes em um mesmo “bairro”, ou porção territorial da cidade (GOODMAN; THIBODEAU, 1998). Ao mesmo tempo, muito se fala sobre a cidade dual, a cidade formal e informal. O Fórum Urbano Mundial organizado pela ONU no Rio de Janeiro em 2010, por exemplo, tinha no título: “Bridging the Urban Divide”, propondo uma ponte para essa divisão urbana. Este trabalho busca caracterizar o mercado formal e informal brasileiro e possíveis segmentações dentro dessas duas primeiras divisões.

O principal objetivo é explicitar as semelhanças e diferenças, e as heterogeneidades dentro de cada grupo. Levando em consideração a inadequação habitacional, há uma parte do mercado formal que é semelhante ao mercado informal. Assim como há uma parte do mercado informal que é semelhante ao mercado formal, e, portanto, livre de inadequações habitacionais mais graves. Portanto, o trabalho auxilia a responder esta pergunta essencial: o quanto a informalidade influencia a disponibilização de moradia digna?

Os preços dos imóveis são resultado do equilíbrio do mercado para este bem heterogêneo: da interação entre oferta e demanda. Refletem a valorização dos seus principais componentes: da estrutura física (número de dormitórios, se é apartamento, por exemplo); das amenidades urbanas (áreas verdes e paisagismo do bairro, proximidade a equipamentos de lazer) e da inserção urbana (acessibilidade a oportunidades de emprego, equipamentos de saúde, educacionais). (ROSEN, 1974) Em termos das diferenças dentre mercados, entre diferentes cidades, outras características agregadas explicam a variação dos preços, como os salários dos trabalhadores e amenidades ambientais das cidades (estar na beira do mar, por exemplo). (ROBACK, 1982).

A princípio o mercado habitacional informal apresentaria componentes que se valorizariam de forma semelhante. O surgimento, crescimento e principais características das favelas no Brasil estão em sintonia com a ocupação de áreas não disponíveis para o mercado formal, ou de menor preço, mas ainda próximas aos pólos econômicos das cidades. Ao mesmo tempo as favelas também se identificam simplesmente com ocupações irregulares, que por serem irregulares são opções de moradia mais baratas, na periferia. (BRUECKNER; MATION; NADALIN, 2019; CAVALCANTI; DA MATA; SANTOS, 2019; SMOLKA; BIDERMAN, 2011) Ou seja, em alguma medida as favelas brasileiras são submercados de baixa renda que competem com o mercado formal de baixa renda. Há evidências de que o processo de valorização de localidades mais centrais criam condições para que bairros informais bem localizados sejam formalizados, fazendo com que a boa inserção urbana cada vez mais seja característica mais presente no mercado formal. (HENDERSON; REGAN; VENABLES, 2020) verificam empiricamente e modelam teoricamente esse processo para Nairobi, no Kenia. Portanto, bairros informais mais centrais tenderiam a ser mais diferentes de seus vizinhos formais que os bairros informais periféricos. A partir do momento que os bairros informais centrais se formalizam, e se gentrificam, essas diferenças desapareceriam.

Foi utilizado como definição de bairro informal o conceito de aglomerado subnormal (AGSN) do IBGE para o Censo 2010. (IBGE, 2010a) A definição de AGSN do IBGE é adequada, porém inacurada para muitos aspectos da informalidade. Assim, tomamos esta classificação como uma medida proxy da informalidade.

¹ Este trabalho inova ao caracterizar os dois mercados a partir de informações do questionário da amostra do Censo 2010, cuja identificação de pertencimento a um aglomerado subnormal foi feita dentro da sala do sigilo no IBGE. Duas informações importantes disponibilizadas são as referentes ao aluguel e tempo de

¹ A pesquisa “Núcleos Urbanos Informais”, capitaneada pelo IPEA, em parceria com diversas instituições e encomentada pelo Ministério do Desenvolvimento Regional investiga os limites da definição de AGSN do IBGE para captar o fenômeno da informalidade habitacional urbana.

deslocamento. Além disso, também é inovadora a análise conjunta da proximidade a empregos formais geolocalizados a partir das informações dos empregos formais da RAIS, e declividade média do setor censitário e distância dos setores censitários ao centro de negócios de cada cidade.

Utilizamos como amostra do estudo 510 municípios que ou apresentavam AGSN ou pertenciam a um arranjo populacional (IBGE, 2015) em que algum dos municípios possuía AGSN. O total de população segundo o Censo 2010 destes 510 municípios correspondia a 56,2% da população do Brasil em 2010, sendo a população em AGSN 11,4% desta população (IBGE, 2010b) ². São feitas análises descritivas utilizando-se, dentre outros, os indicadores de autocorrelação espacial local (LISA), gráficos de dispersão e regressão (*binscatter regressions*) e uma análise de clusters para setores censitários nas metrópoles do Rio de Janeiro e São Paulo. Também analisamos os agrupamentos urbanos que congregam estes 510 municípios em 70 aglomerados e 106 municípios isolados. Estes agrupamentos são chamados de arranjos populacionais. Os arranjos populacionais são “agrupamentos de dois ou mais municípios com forte integração populacional, devido aos movimentos pendulares para trabalho ou estudo, ou à contiguidade entre manchas urbanas.” (IBGE, 2015).

Exploramos as diferenciações dos mercados em cinco níveis geográficos: 1. o conjunto das cidades com AGSN (média dos 510 municípios); 2. os mercados habitacionais das maiores cidades (os 13 maiores arranjos populacionais); 3. os 106 municípios isolados e 70 arranjos populacionais (representados em mapas do Brasil) ; 4. As diferenças entre os 510 municípios (visualizando a diferença entre municípios sede, do entorno e isolados) e por fim, 5. os setores censitários dentro de São Paulo e Rio de Janeiro (na análise de clusters de setores censitários). Assim, a estrutura deste trabalho segue os níveis geográficos, do mais abrangente para o menos abrangente.

2. CONJUNTO DAS CIDADES COM AGSN

Esta seção traz as estatísticas descritivas provenientes das informações de todos domicílios da amostra dentro dos 510 municípios da amostra: 2.089.661 amostrados, desses, 196.150 em AGSN. Bom “resumo” das diferenças nas características dessa divisão.

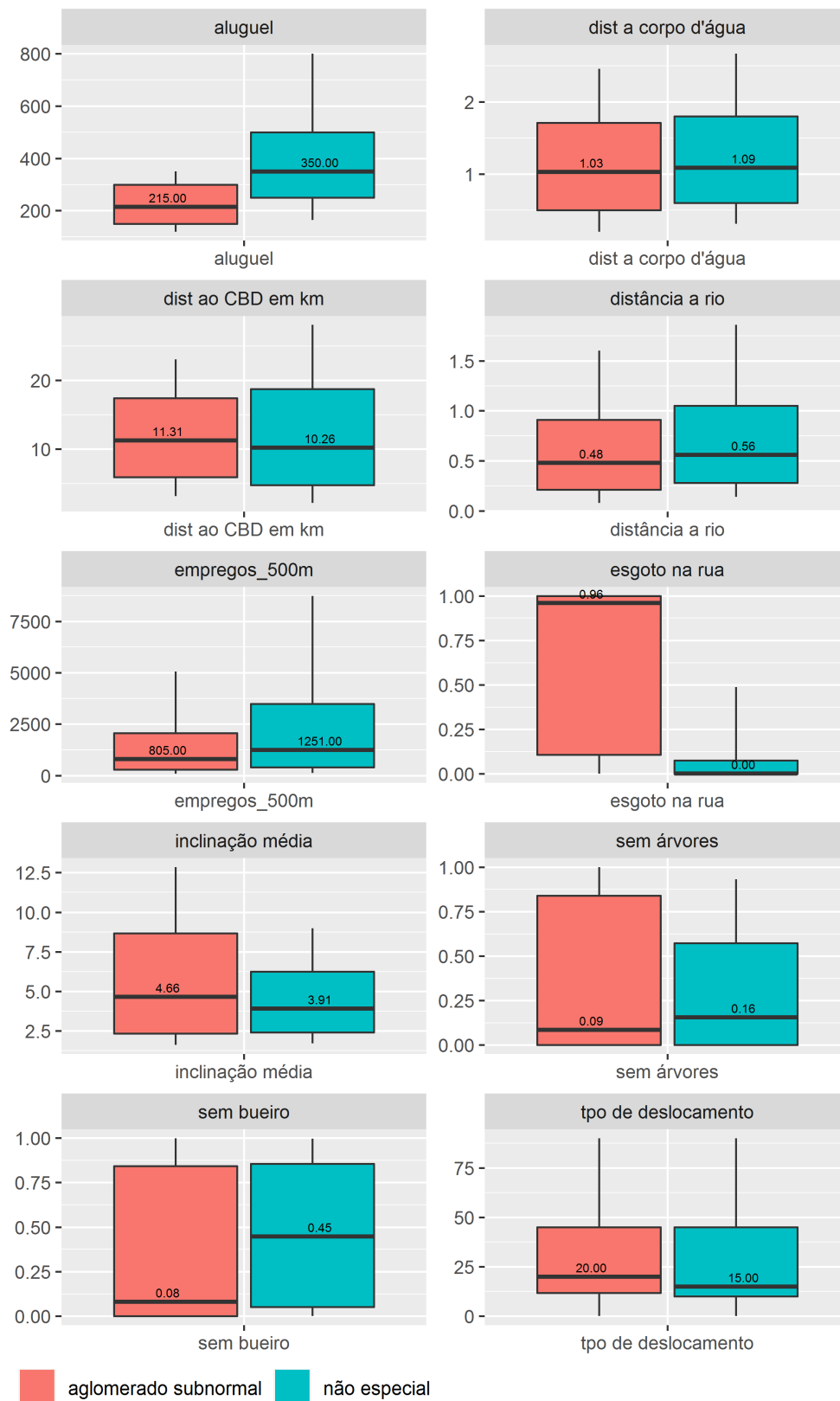
A figura 1 mostra a inserção urbana dos AGSN, condizente com descrições anteriores. São regiões da cidade que ocupam mais os terrenos de alta declividade, mas também os de menor declividade. Com mais frequência, os AGSN se localizam mais periféricamente, estão próximos de menos empregos formais e levam mais tempo no deslocamento de casa para o trabalho. A mediana da distância ao CBD é um pouco maior para os AGSN, mas esta distância varia menos que os não especiais.

A declividade média do terreno é maior, com intervalo interquartil (IQR)³ maior. Já o tempo de deslocamento é maior, com IQR menor e a proximidade a empregos formais menor, com IQR menor. A mediana do aluguel é bem menor, IQR com pouca sobreposição ao formal. (150 a 300 vs. 250 a 500), sendo a primeira evidência de que o mercado de habitação informal seria um mercado separado do mercado formal. Das variáveis do entorno, a que mais chama a atenção é a presença de esgoto na rua: a mediana é 96% contra 0,4%, sendo que o percentil 25 dos agsn é maior (10,7%) que o percentil 75 dos domicílios em áreas não especiais (7,5%). Evidencia uma condição urbana muito pior para os AGSN.

² População do Brasil 178.360.003, população dos 510 municípios: 100.229.069, população em AGSN: 11.413.993.

³ O intervalo interquartil mede a diferença entre o terceiro quartil e o primeiro quartil; que também são os limites da “caixa” do gráfico boxplot. Os quartis são os valores dos limites de intervalos ao ordenar a variável em ordem crescente e dividir a amostra em quatro partes iguais.

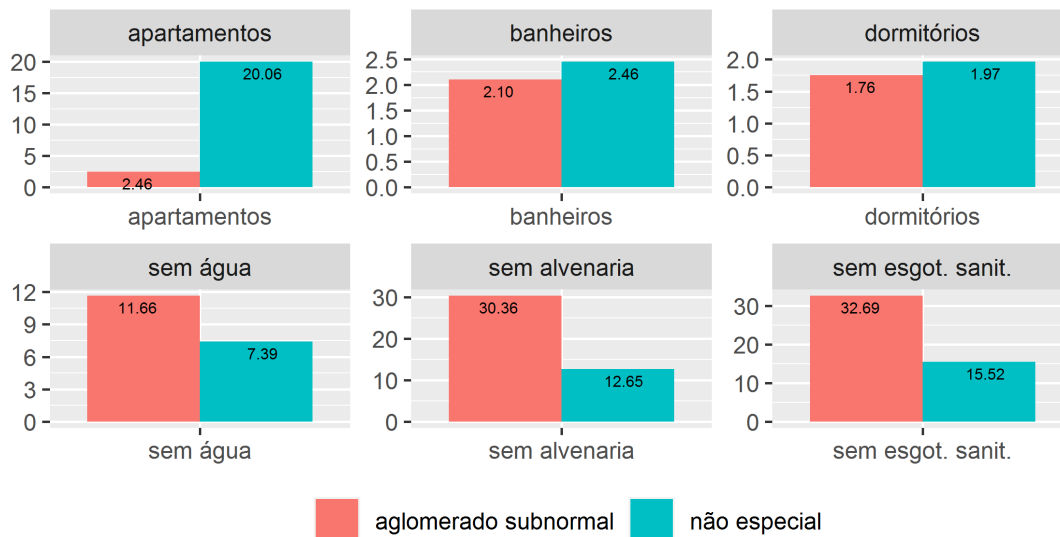
Figura 1: Boxplot da Inserção urbana, aluguel e características do entorno. Total da amostra



Fonte: Censo 2010. Empregos: RAIS 2017. Informações geográficas: projeto OpenStreetMap. Topografia do terreno sensor de satélite STRM.

Nota: Aluguéis em R\$ de 2010; distâncias em km; esgoto na rua, sem árvores e sem bueiro: proporção de domicílios com as características; inclinação média: em porcentagem; tempo de deslocamento em minutos. Corpos d'água são acumulações de água como mares, lagos e reservatórios.

Figura 2: Características dos domicílios: médias



Fonte: Censo 2010.

Nota: Ver definições de água e esgotamento inadequados em www.odsbrasil.com.br, unidades: proporção de domicílios com as características à excessão do número médio de dormitórios e banheiros.

Já a figura 2 mostra as médias das características das estruturas físicas dos domicílios. Domicílios em AGSN na grande maioria não são apartamentos, têm menos banheiros, menos dormitórios, menos paredes de alvenaria, menos abastecimento de água e esgotamento adequados (definições segundo o ODS 11⁴).

3. OS MERCADOS HABITACIONAIS DAS MAIORES CIDADES

Este trabalho explora a ideia de que o mercado de habitação é segmentado, no caso do Brasil entre mercado formal e informal. As segmentações mais comuns são em bairros, regiões da cidade que compartilham de preços semelhantes. Mesmo assim, os diferentes segmentos de mercado ou submercados de uma mesma metrópole também podem ser visualizados na sua totalidade, e considerados como um mercado habitacional único, específico a cada cidade. Isto acontece porque as diferentes localidades dentro de uma mesma aglomeração urbana pertencem a um mesmo mercado de trabalho, onde potencialmente é possível trabalhar e morar em qualquer par de localizações dentro dessa região.

Portanto, vale a pena visualizar as principais características dos mercados habitacionais das principais cidades do país. Dadas as diferentes trajetórias de desenvolvimento institucional e econômico, os diferentes suportes geográficos e as diferentes localizações relativas há grandes desigualdades entre estes mercados habitacionais.

Mostraremos essas características para os 13 maiores arranjos populacionais do IBGE. Nos parece que estes agrupamentos de municípios propostos pelo IBGE são boas delimitações das maiores regiões metropolitanas

⁴ Ver www.odsbrasil.com.br

brasileiras, e boas aproximações do que seriam os grandes mercados de trabalho e de habitação correspondentes a essas grandes “cidades”, ou regiões metropolitanas. Nesta seção chamaremos de “cidades” esses arranjos populacionais. As regiões metropolitanas brasileiras oficiais são delimitadas por lei estadual e não seguem critérios homogêneos para a inclusão de municípios. Para as grandes metrópoles não há diferenças muito significativas, pois, as diferenças entre os conjuntos de municípios considerados acabam correspondendo a municípios muito pequenos.

a. Inserção urbana

Começamos com as diferenças na estrutura urbana, visíveis a partir da distância dos domicílios amostrados ao centro de negócios (CDB, Central Business District) e da quantidade de empregos formais existentes em um *buffer* de 500 metros a partir dos limites do setor censitário no qual o domicílio está inserido. As diferenças entre os domicílios em setores não especiais e domicílios em AGSN exemplificam as desigualdades na inserção urbana destes domicílios. As figuras mostram em formato de boxplot os valores dos percentis 10, 25, 50, 75 e 90.⁵ Em primeiro lugar, sobressaem-se as cidades com maiores valores: Brasília e Rio de Janeiro, cidades que ocupam uma extensão muito ampla de território. Brasília muito espalhada por conformação do tecido urbano fragmentado e descontínuo, o RJ muito espalhado devido à geografia, barreiras físicas à ocupação urbana.

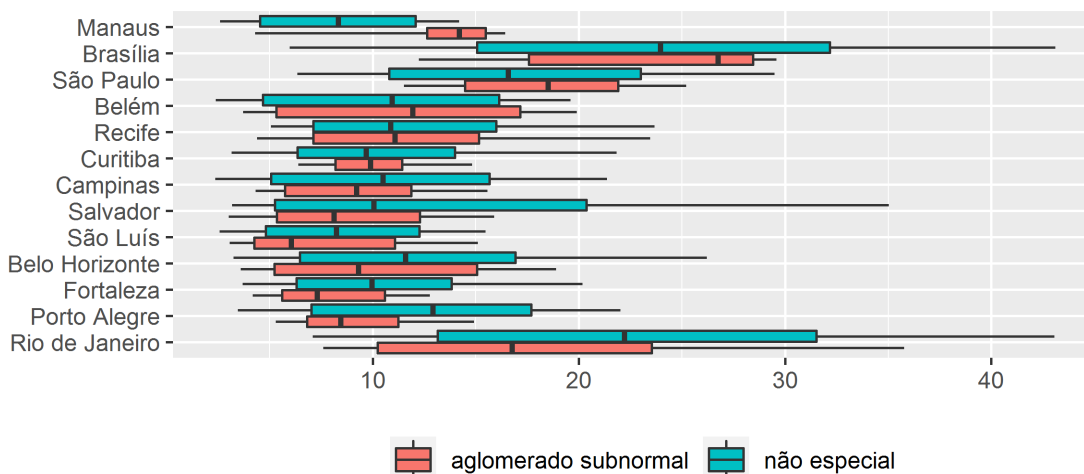
Evidentemente que quanto maior a população da cidade, maior a distância ao CDB, pois há a necessidade de usar mais terreno para construir mais casas para abrigar mais população. Mas mesmo cidades de porte semelhante, como Brasília e Fortaleza, São Paulo e Rio de Janeiro ou Curitiba e Salvador apresentam diferenças amplas nas medianas e nos demais percentis da distribuição das distâncias ao centro dos domicílios amostrados pelo Censo, indicando diferenças na estrutura interna urbana como diferentes valores de densidade interna e configurações intraurbanas de pólos de concentração de população.

Outra característica muito diferenciada entre as cidades é a inserção urbana dos AGSN. Em geral, os AGSN se distribuem em distâncias ao centro menores que os domicílios em setores não especiais. A exceção claríssima é Manaus, em que uma minoria dos domicílios em AGSN fica próximo ao CBD; Belém também apresenta AGSN mais longínquos. Por outro lado, no Rio de Janeiro, Porto Alegre e Fortaleza há uma boa porção de domicílios em AGSN localizados mais centralmente que os domicílios não especiais. Já em SP e Brasília, os AGSN estão mais distantes do centro. Note que na seção anterior, para o conjunto dos domicílios amostrados pelo Censo, a mediana da distância dos AGSN era superior a dos domicílios não especiais. Aqui a informação segregada por cidade traz um retrato mais detalhado, ilustrando a grande heterogeneidade que o dado agregado estava ocultando.

Este padrão da proximidade ao CBD também se repete na distribuição dos totais de empregos formais próximos aos domicílios. No entanto, este indicador também é sensível a o quanto os empregos estão concentrados, e quantos subcentros de empregos existem na metrópole, trazendo Belém como uma cidade com boa proximidade de empregos e domicílios, por exemplo. Em Manaus, a maior parte dos domicílios em AGSN fica próximo a menos empregos que as áreas não especiais, assim como em Belém. Já Brasília, que não apresentava AGSN discrepantes no quesito distância ao centro aqui apresenta domicílios em AGSN com nítida desvantagem aos não especiais no quesito proximidade a empregos. São Paulo, que na Figura 3, de distância ao centro, mostrou distâncias bem menores que o Rio de Janeiro, por exemplo, aqui apresenta AGSN com bem menos empregos formais próximos.

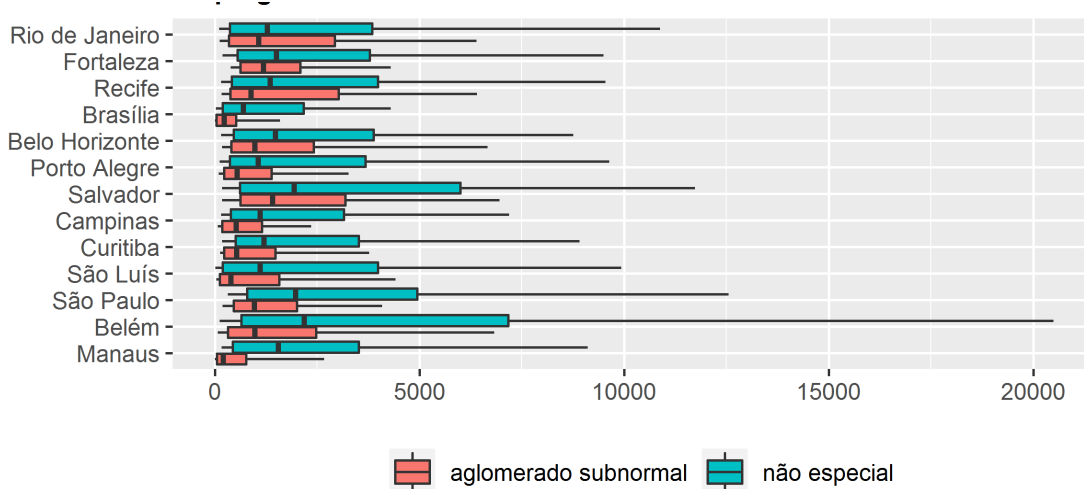
⁵ Os percentis se referem aos valores que uma variável toma quando ordenamos em uma fila os indivíduos de uma amostra do menor valor ao maior valor da esquerda (menor) para a direita (maior), por exemplo. O décimo percentil, por exemplo, corresponde ao valor do indivíduo que está na posição da fila em que 10% dos indivíduos estão à sua esquerda, apresentando valores menores que o dele e 90% dos indivíduos estão à direita, estes apresentando valores maiores que o dele. O percentil 50 é a mediana da distribuição da variável.

Figura 3: Distância ao centro de negócios em Km. AGSN e setores não especiais.



Fonte: Censo 2010.

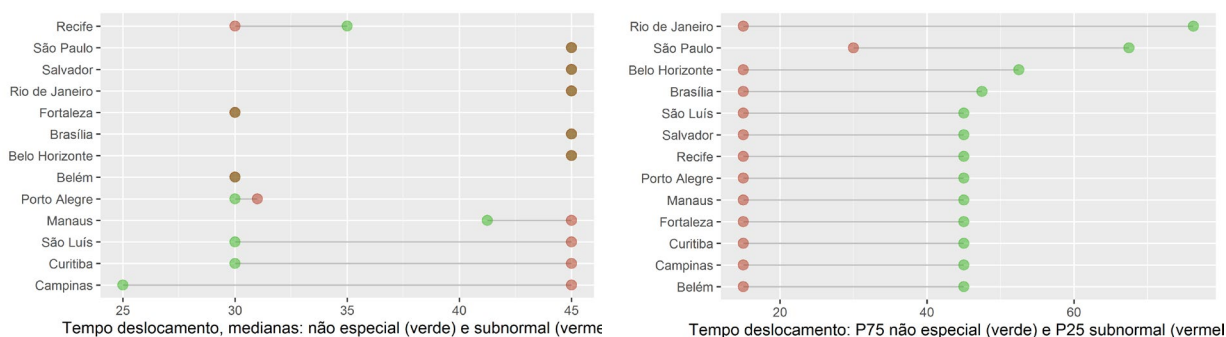
Figura 4: Proximidade e empregos dentro de buffer de 500m. AGSN e setores não especiais.



Fonte: Censo 2010.

Este quadro tanto da estrutura intraurbana diferenciada entre as cidades quanto da inserção urbana desigual entre domicílios não especiais e subnormais também se reflete no tempo de deslocamento casa-trabalho, uma das variáveis medidas no questionário da amostra do Censo. A mediana do tempo de deslocamento é mostrada na Figura 5, à esquerda, deixando claro como as desvantagens dos AGSN em Manaus, descritas acima, se traduzem em maior tempo de deslocamento, por exemplo. As demais cidades com medianas subnormais superiores: Porto Alegre, São Luís, Campinas e Curitiba também apresentam desvantagens em inserção urbana coerentes com esse tempo de deslocamento. Já a Figura 5, painel da direita, apresenta a comparação da “melhor” porção dos AGSN (percentil 25, menor tempo de deslocamento) com a “pior” porção dos não especiais (percentil 75, maior tempo de deslocamento). Os melhores subnormais sempre gastam menos tempo se deslocando, (no mínimo 25 minutos a menos) sendo a diferença maior para o Rio de Janeiro. Isto nos mostra o quanto existe de heterogeneidade interna aos AGSN. Mesmo em Manaus há AGSN muito bem localizados, muito melhores que os não especiais, apesar de serem poucos como vimos acima. Em Recife a vantagem de inserção urbana dos AGSN é mais homogênea, pois a própria mediana do tempo de deslocamento dos AGSN é inferior à dos domicílios não especiais.

Figura 5: Medianas do tempo de deslocamento de AGSN e não especiais (esquerda). Tempo de deslocamento: Percentil 75 não especial e Percentil 25 AGSN (direita)



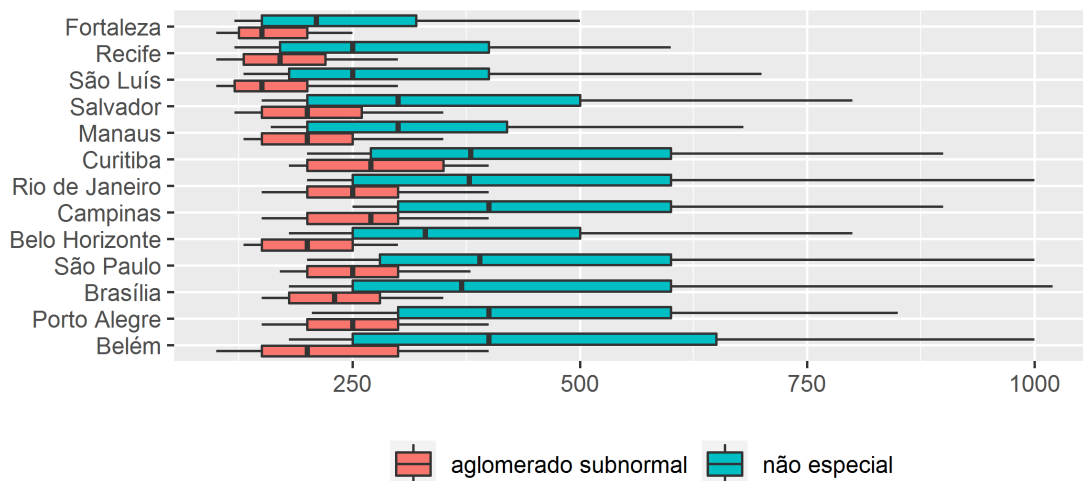
Fonte: Censo 2010.

Nota: Quando há sobreposição dos pontos verdes (não especial) com os pontos vermelhos (AGSN), visualiza-se o ponto na cor marrom.

b. Aluguel e amenidades urbanas

O questionário da amostra do Censo 2010 perguntou o valor do aluguel para imóveis alugados. A Figura 6 traz essa preciosa informação para os mercados formal e informal nas 13 cidades aqui analisadas. O mercado não especial apresenta uma amplitude maior de valores e valores de percentil 90 muito mais amplos que os valores de percentil 75. Indicando uma desigual distribuição de renda e de riquezas já muito conhecida e retratada por diversas fontes de dados. Os valores do mercado informal são menores, assim como sua amplitude (o intervalo interquartilico) é menor. Os valores mais baixos do mercado informal não são alcançados pelo mercado formal.

Figura 6: Distribuição de valores de aluguéis. AGSN e setores não especiais.



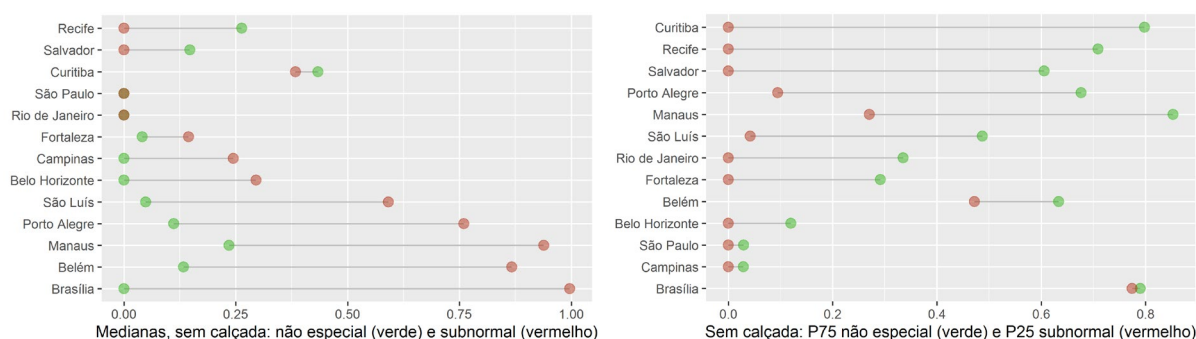
Fonte: Censo 2010.

Isto indica que há uma faixa de preços para os quais há domicílios alugados tanto no mercado formal, quanto no informal. Em São Paulo parece haver uma menor proporção de domicílios nessa faixa que em Curitiba, por exemplo. Em Belo Horizonte e Porto Alegre menos ainda. Para essas faixas de preços, ambos mercados competem pelo mesmo consumidor, sendo interessante investigar as demais características desses mercados para compreender o que há de vantajoso na opção pelo mercado informal. A análise da inserção urbana, feita acima, indica que há uma porção dos melhores subnormais que possui melhor inserção urbana que os piores domicílios em setores não especiais, nestas cidades, por exemplo. O caso do Rio de Janeiro indicou

boas inserções urbanas dos AGSN, e de fato, da mediana para cima do mercado informal carioca parece estar sobreposto ao mercado formal. Para ilustrar o quanto as amenidades de vizinhanças acompanham essa variação do aluguel, a Figura 7 traz informações relativas ao entorno dos domicílios: a proporção de domicílios em ruas com calçadas no setor censitário do domicílio, coletadas pelos recenseadores do Censo 2010.

A diferença entre as medianas dos AGSN e setores não especiais nos mostra como em Recife, Salvador e Curitiba, há maior proporção de AGSN com calçadas, que proporção de domicílios com calçadas em setores não especiais. Já o que era de se esperar, nas demais cidades o entorno dos subnormais é pior que os setores não especiais. A comparação da “melhor” porção dos AGSN (percentil 25) com a “pior” porção dos não especiais (percentil 75) mostra que em todas as cidades os melhores subnormais estão em setores censitários com menos domicílios sem calçadas que os piores não especiais. No Rio de Janeiro e em Curitiba, locais onde vimos uma proporção grande de domicílios em faixas comuns de preços ao mercado formal e informal a vantagem dos melhores subnormais é maior, algo que pode explicar a escolha por morar em AGSN em detrimento de um setor não especial de mesmo preço.

Figura 7: Domicílios em ruas sem calçadas: Medianas (esquerda). P75 não especial e P25 AGSN (direita)

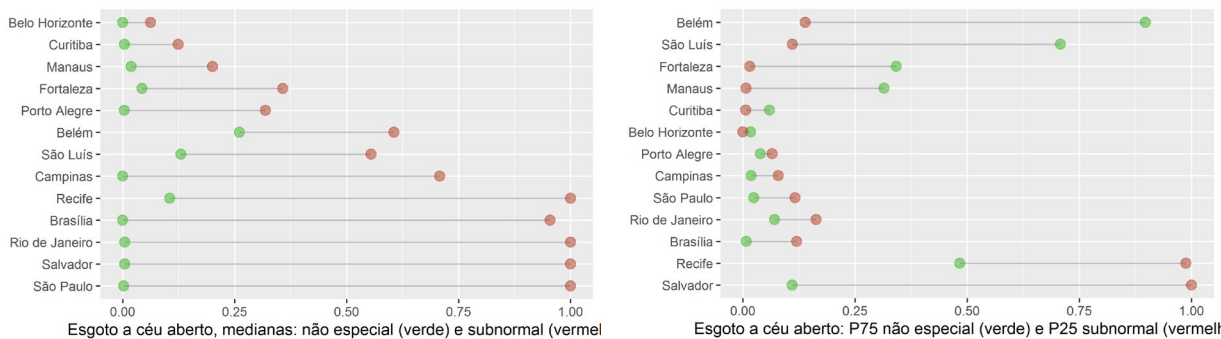


Fonte: Censo 2010

Já a figura 8 nos mostra outra característica do entorno dos domicílios: a presença de esgoto a céu aberto. Esta característica, à diferença da existência de calçada é mais generalizadamente presente nos subnormais. Mesmo assim para Belém, Fortaleza, São Luís, Manaus, Curitiba e Belo Horizonte os “melhores” AGSN são melhores que os “piores” não especiais.

Essas figuras sustentam um retrato da desigualdade tanto entre o mercado formal e informal quanto dentro de cada um desses mercados. O mercado informal possui características piores e aluguel inferior. A desigualdade de inserção urbana, de aluguel e de qualidade da vizinhança dentro do mercado informal chama a atenção uma vez que comprova a heterogeneidade interna aos AGSN e justifica que haja alguns bairros classificados como AGSN que são “concorrentes” do mercado formal, inclusive com amenidades urbanas melhores que estes.

Figura 8: Esgoto a céu aberto: Medianas (esquerda). P75 não especial e P25 AGSN (direita)



Fonte: Censo 2010.

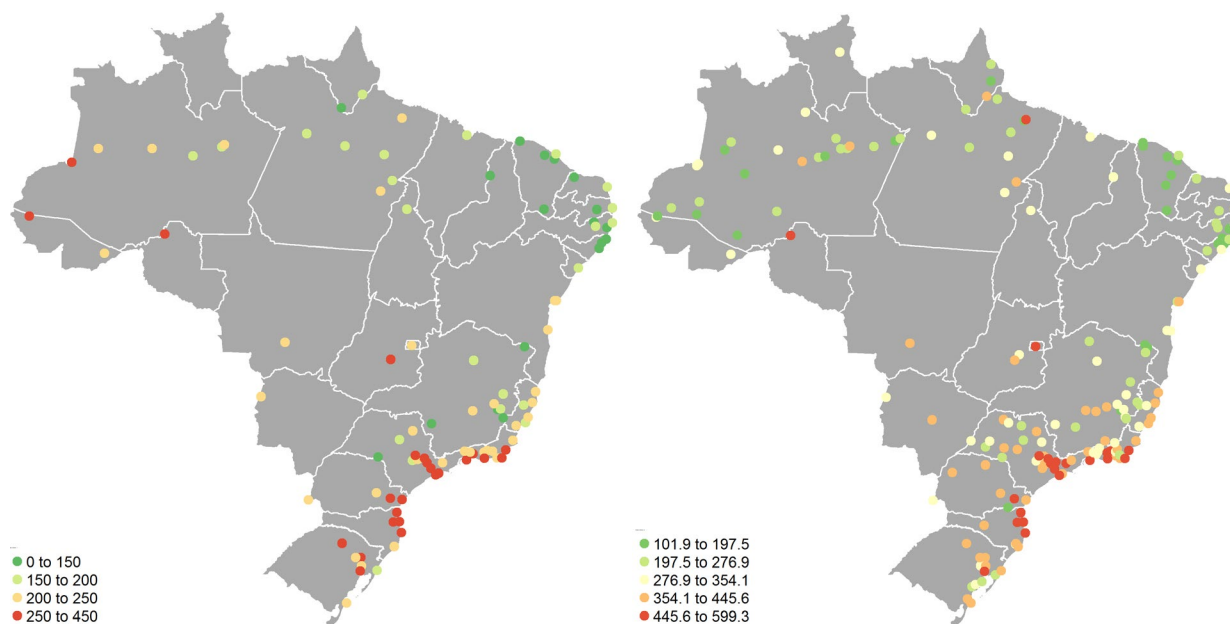
Fica registrado também a heterogeneidade dentre as 13 “cidades, ou arranjos populacionais, analisadas. Há combinações diferentes de desigualdades de inserção urbana/ valor do aluguel e amenidades urbanas. Por exemplo, em Manaus há clara desvantagem na inserção urbana dos AGSN e das amenidades urbanas, mas não tão grande diferença no valor do aluguel. Já em Belo Horizonte há grande diferença no aluguel e pouca diferença na inserção urbana e amenidades urbanas. O caso do Rio de Janeiro ilustra bem a presença de muitos domicílios em AGSN bem localizados, enquanto o de São Paulo o de AGSN não bem localizados e com amenidades urbanas piores que os setores não especiais.

4. MAPAS DOS 70 ARRANJOS POPULACIONAIS E 106 MUNICÍPIOS ISOLADOS

Aproveitando as informações de todos os 176 núcleos urbanos, sendo 70 arranjos populacionais e 106 municípios isolados com presença de aglomerados subnormais, visualizaremos as principais características no mapa do Brasil, buscando por padrões regionais de variação destas características. Começaremos pelo mapa de aluguel e de renda do chefe do domicílio, na Figura 9. Os mapas da distribuição dos valores dos alugueis formais e informais sugerem aluguéis mais altos no Sul e Sudeste e menores no Norte Nordeste. Note como as faixas de cores dos valores médios dos alugueis informais correspondem a valores menores que as faixas de valores das mesmas cores do mapa de aluguéis médios formais, estes intervalos são calculados utilizando o método de jenks. O mapa formal, inclusive, possui uma faixa de cores de 445 a 599 reais de aluguel médio que nem existe para o mercado informal.

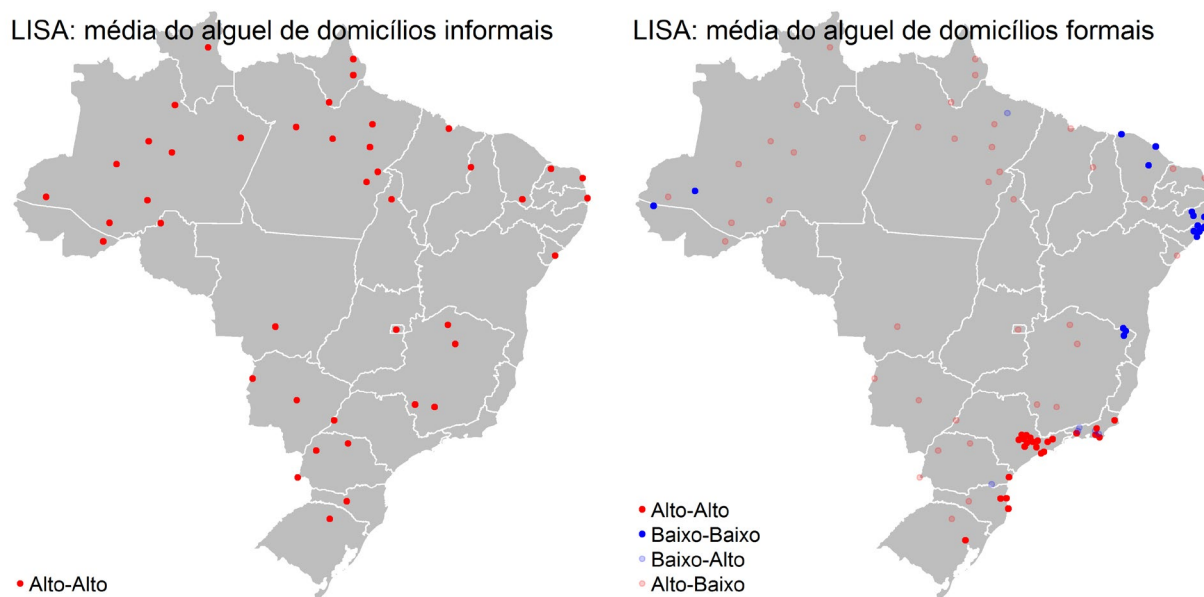
No entanto, buscamos corroborar este padrão espacial estatisticamente, para além da inspeção visual, calculando o indicador de correlação espacial local (LISA) (ANSELIN, 1995). Este indicador, além de quantificar o quanto observações próximas possuem valores próximos indica se esta associação espacial é estatisticamente significativa, pois compara essa distribuição dos valores no espaço com possíveis distribuições “aleatórias” dos mesmos valores no espaço.

Figura 9: Mapa coroplético da distribuição dos valores dos Aluguéis AGSN (esquerda) e não especiais (direita). 70 arranjos populacionais e 106 municípios isolados.



Fonte: Censo 2010.

Figura 10: Mapa dos “clusters” LISA dos valores dos aluguéis médios em AGSN (esquerda) e setores não especiais (direita). 70 arranjos populacionais e 106 municípios isolados.



Fonte: Censo 2010.

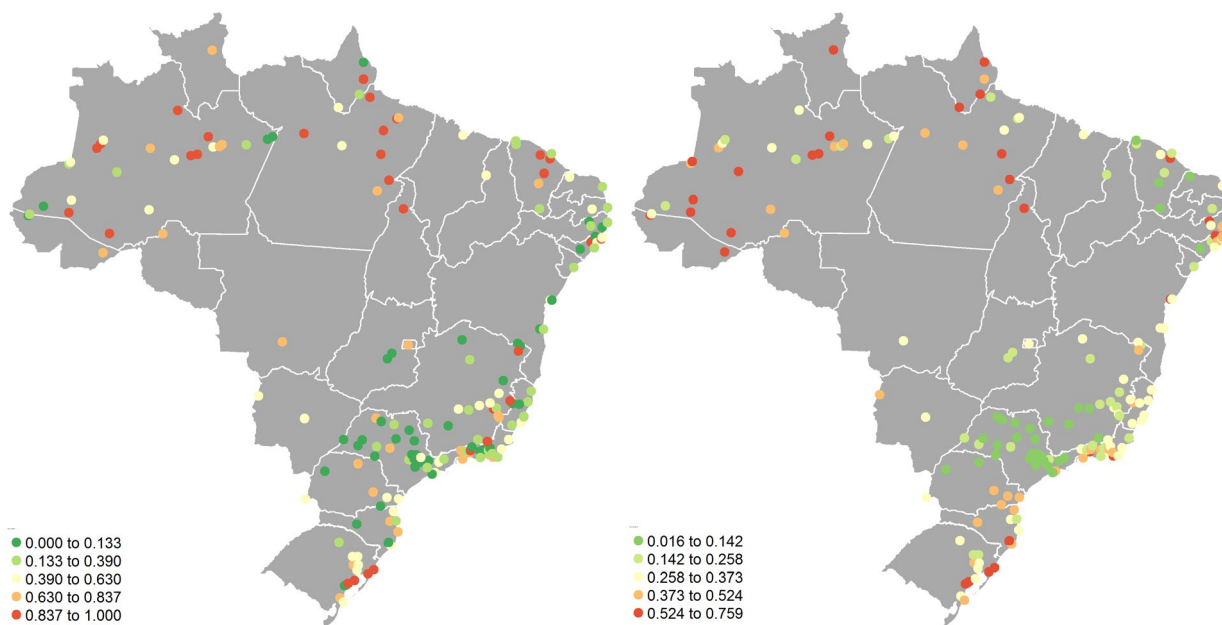
A análise do indicador LISA indicou a existência de regiões que concentram valores altos de aluguéis formais no Sul e Sudeste e valores baixos de aluguéis no Nordeste, mas não encontramos padrão espacial no aluguel informal. Seguem esses resultados na figura 10. Talvez isso aconteça porque não há muita variância nos valores dos aluguéis informais, uma vez que a porção mais humilde da população não é tão diferenciada entre as regiões brasileiras quanto a porção mais abastada. Há ricos muito mais ricos nas regiões mais ricas do país, enquanto os pobres não são tão diferenciados assim. A relação empírica entre os rendimentos das famílias, as amenidades urbanas e os valores dos aluguéis está documentada na literatura. (CRAVO; RESENDE; CRUZ, 2019) encontram o padrão regional compatível ao da figura 10 ao investigar esta relação.

Já para a distribuição das diferenças nas médias das amenidades urbanas dentre os 176 núcleos urbanos analisados, o padrão espacial Sul/Sudeste com boas amenidades e Norte/Nordeste com piores amenidades é encontrado tanto para o mercado formal como para o mercado informal e confirmado para ambos os mercados pela análise de LISA. A Figura 11 mostra os dois mapas da distribuição dos valores das médias de domicílios sem calçadas, e indica exatamente esse padrão de valores piorando com menores latitudes.

A Figura 12 mostra esses clusters para domicílios com esgoto a céu aberto, mostrando clusters de valores baixos (com melhor amenidade urbana) nos estados de São Paulo, Rio de Janeiro e Minas Gerais, e concentração de valores altos no Norte e Nordeste.

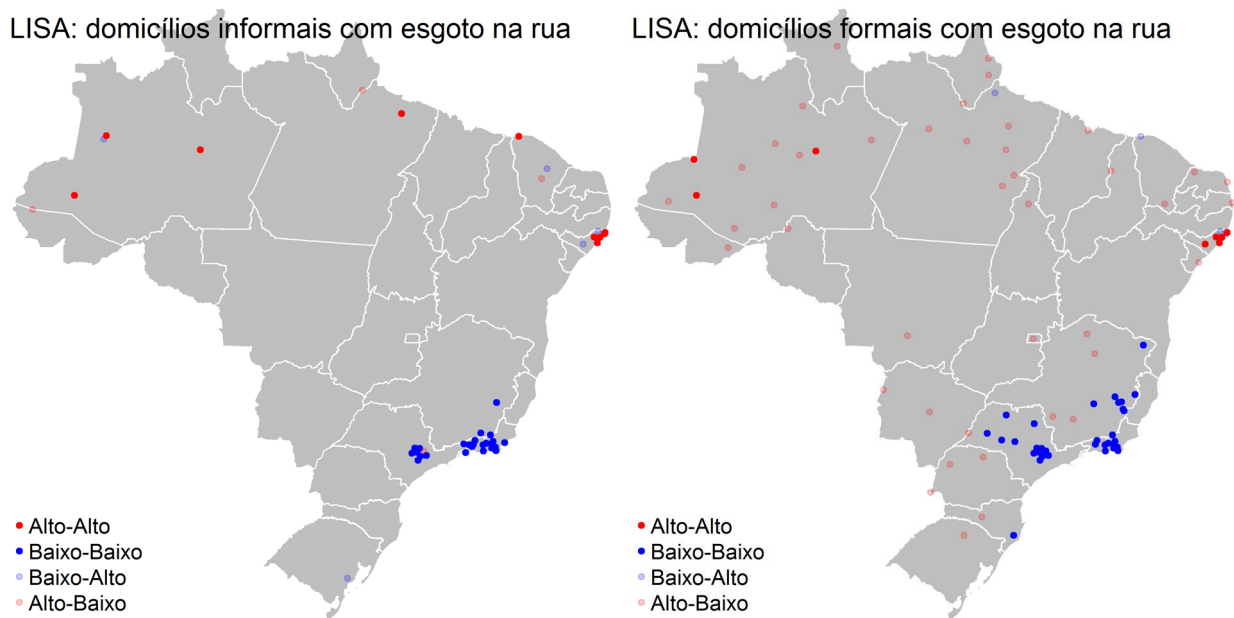
Além disso, a Figura 13 mostra um padrão espacial semelhante para clusters da variável de disponibilidade de esgotamento sanitário adequado, com a diferença que para os domicílios formais há clusters de valores baixos (melhor amenidade) no estado do Rio Grande do Sul. Note-se a ausência de associação espacial significativa para os domicílios em AGSN em toda a região Sul.

Figura 11: Mapa coroplético da distribuição das proporções de domicílios em ruas sem calçada em AGSN (esquerda) e não especiais (direita). 70 arranjos populacionais e 106 municípios isolados.



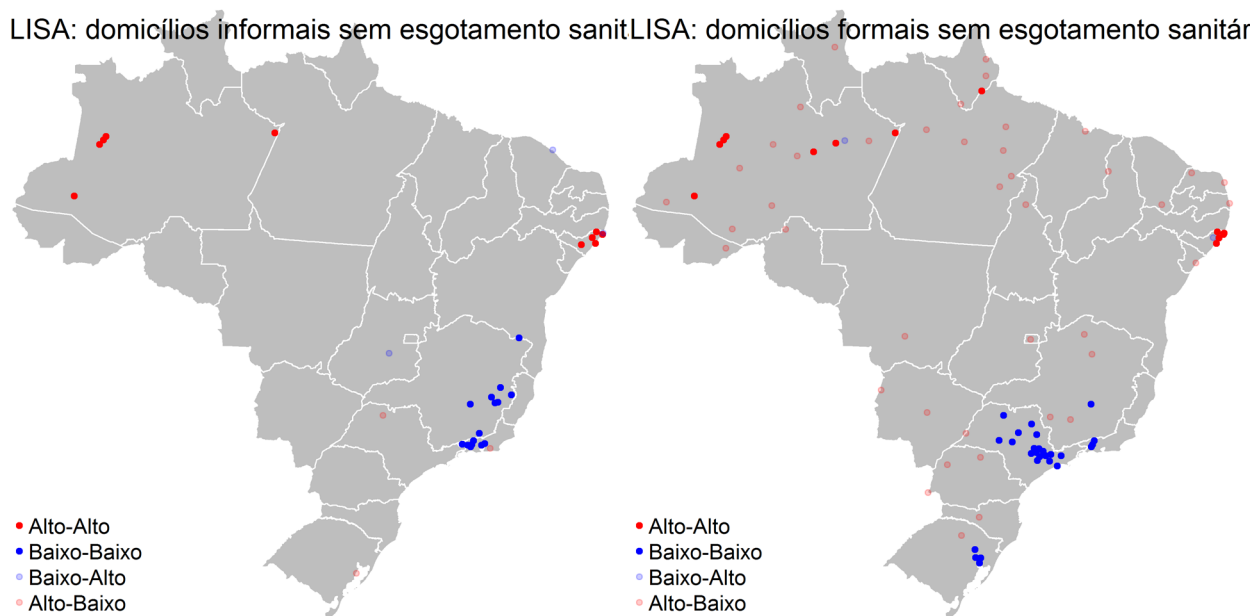
Fonte: Censo 2010.

Figura 12: Mapa dos “clusters” LISA da proporções médias de domicílios em ruas com esgoto a céu aberto em AGSN (esquerda) e setores não especiais (direita). 70 arranjos populacionais e 106 municípios isolados.



Fonte: Censo 2010.

Figura 13: Mapa dos “clusters” LISA das médias de domicílios sem esgotamento sanitário em AGSN (esquerda) e setores não especiais (direita). 70 arranjos populacionais e 106 municípios isolados.



Fonte: Censo 2010.

5. OS MERCADOS HABITACIONAIS DOS MUNICÍPIOS

Esta seção explora as informações para cada um dos 510 municípios da nossa amostra: os 106 municípios isolados, os 70 municípios sede e os 334 municípios no “entorno” dos municípios sede dos arranjos populacionais, conforme definição feita pelo IBGE.(IBGE, 2010b). Analisaremos as diferenças na variação das características dos mercados habitacionais dentre estes municípios. Apresentaremos os resultados da análise de regressão em gráficos de dispersão por grupos de intervalos (*binscatter regressions*).

A Tabela 1 ilustra a distribuição dos domicílios particulares permanentes nas categorias de municípios da nossa amostra. Apesar de apenas 70 municípios serem sede, estes concentram 47% dos domicílios de todos os 510 municípios. Por outro lado, notamos como os municípios isolados são pequenos, pois os 106 municípios concentram somente 11% dos domicílios. A proporção de domicílios em AGSN varia de 7% a 11%. Os municípios sede apresentam a maior proporção de AGSN, como era de se esperar, pois nesses municípios no mercado formal os preços são mais elevados devido à concentração de famílias de maior renda, assim como a concentração de melhores amenidades urbanas e oportunidades de emprego.

Assim como notado na diferença da média do conjunto dos municípios e dos 13 maiores arranjos populacionais (seções 2 e 3 deste trabalho), as características das estruturas físicas dos domicílios variam bastante entre AGSN e domicílios em setores não especiais; e aqui notamos como variam também bastante dentre os mercados dos municípios analisados. A Figura 14 nos mostra essas distribuições a partir de estimativas de curvas de densidade de kernel, que nada mais são que uma suavização do tradicional histograma.

Nota-se que, dentre os municípios, os mercados formais apresentam maior qualidade e menor variabilidade na ausência de alvenaria e ausência de esgotamento sanitário. Essas características, que indicam um nível

de qualidade bastante baixo, são bem pouco frequentes nos mercados formais, enquanto há mercados informais melhores e piores nesses quesitos.

Tabela 1: Distribuição dos domicílios particulares permanentes nas categorias de municípios da amostra.

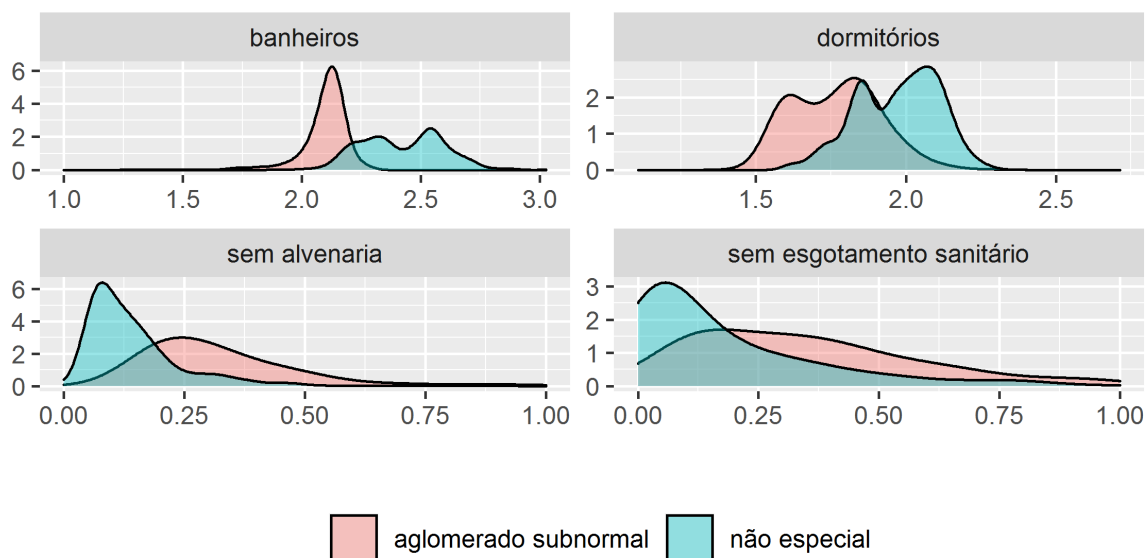
	não especial		aglomerado subnormal		Total	
	(A)	(A/C)	(B)	(B/C)	(C)	% do total
entorno	21.645.540	93%	1.654.420	7%	23.299.960	42%
isolado	5.601.519	92%	509.782	8%	6.111.301	11%
sede	2.863.495	89%	2.825.481	11%	25.688.976	47%
510 municípios	50.110.554	91%	4.989.683	9%	55.100.237	100%

Fonte: Censo 2010

Já a distribuição do número médio de dormitórios parece ter variabilidade semelhante, sendo em geral 0,25 dormitórios por domicílio maior nos domicílios não especiais. A distribuição de número médio de banheiros dos AGSN fica bem concentrada ao redor de 2,1 banheiros, enquanto as médias dos domicílios não especiais se concentram entre 2,2 e 2,7 banheiros por domicílio.

A qualidade dos domicílios em termos de número de dormitórios e banheiros varia mais dentre os domicílios em setores não especiais, o que é condizente com a ideia de que o mercado formal abrange tanto os segmentos de menor renda, quanto de maior renda da população.

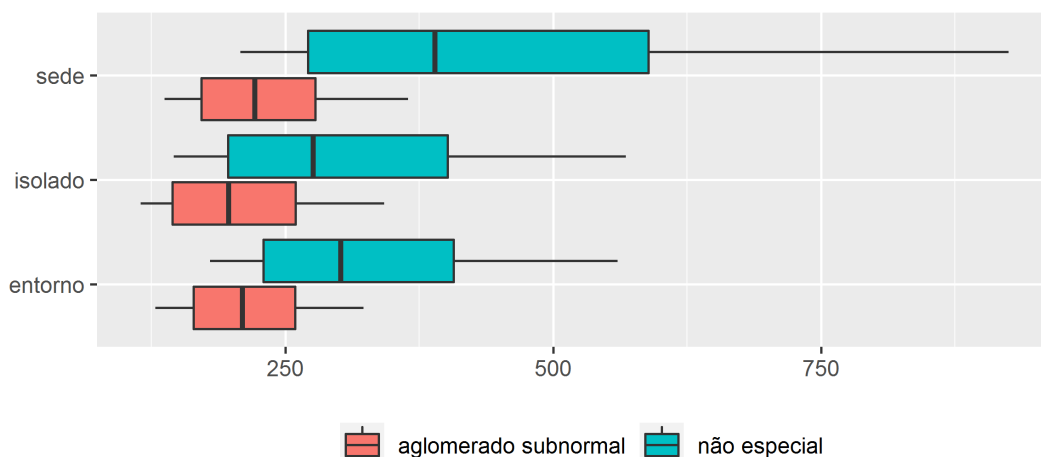
Figura 14: Densidade de kernel da distribuição de médias dos municípios de número de banheiros, dormitórios, domicílios sem alvenaria e sem esgotamento sanitário adequado. Municípios da amostra.



Fonte: Censo 2010.

Já a figura 15 introduz uma diferenciação importante entre os mercados dos municípios sede, do entorno e isolados, a média dos aluguéis. Nota-se como os aluguéis dos municípios sede são em geral bem mais frequentes que os dos demais tipos de municípios, e concentram os valores mais altos. É de se esperar, pois esses municípios são os que apresentam melhores amenidades urbanas, melhor qualidade urbanística, mais oportunidades de emprego, educação e cultura. Nota-se, também, como os valores dos aluguéis de municípios do entorno e dos municípios isolados não são muito pronunciadamente diferentes.

Figura 15: Boxplot da distribuição dos valores de aluguéis nos municípios da amostra: sede, entorno e municípios isolados



Fonte: Censo 2010.

Com relação aos aluguéis dos AGSN, notamos como a diferença de valores entre não-especial e AGSN é muito mais ampla para os municípios sede que para os demais municípios. E como a distribuição dos valores desses aluguéis em AGSN não parece se diferenciar muito dentre municípios sede, isolados e do entorno. (Figura 15)

Tabela 2: Regressão de comparação de aluguéis dos municípios isolados

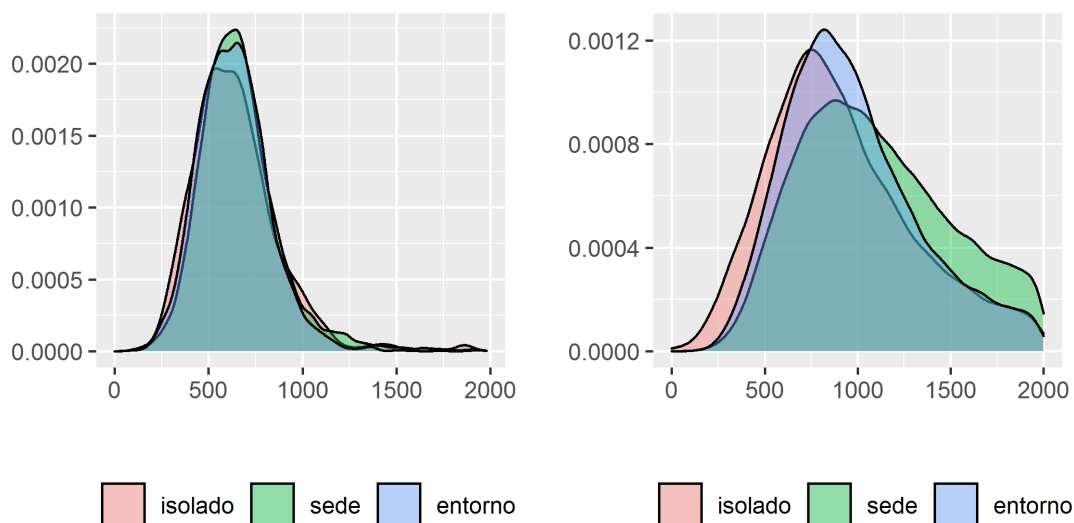
<i>Variável Dependente:</i>				
	Mediana não especial	Mediana AGSN	Média não especial	Média AGSN
Municípios sede	105,307*** (74,616, 135,998)	38,980*** (17,325, 60,636)	142,770*** (103,797, 181,743)	45,355*** (23,953, 66,757)
Municípios do entorno	23,313 (-3,889, 50,515)	24,877** (5,716, 44,038)	25,835 (-8,708, 60,377)	23,019** (4,082, 41,956)
Constante	251,889*** (229,519, 274,259)	175,333*** (159,549, 191,117)	292,438*** (264,032, 320,844)	190,137*** (174,538, 205,736)
Observações	190	191	190	191
R ²	0,169	0,046	0,195	0,061
Estatística F	18,965*** (df = 2; 187)	4,483** (df = 2; 188)	22,580*** (df = 2; 187)	6,082*** (df = 2; 188)

Nota: *p<0,1; **p<0,05; ***p<0,01

Fonte: Censo 2010.

A tabela 2 mostra o resultado de regressões lineares simples de comparação das medianas e médias dos aluguéis formais e informais. Esses valores são regredidos contra as variáveis binárias *dummies* indicando a condição de município sede ou do entorno, tomando como base os municípios isolados. De fato, as diferenças entre municípios isolados e do entorno no mercado formal não são estatisticamente significantes, mas são significantes entre municípios isolados e municípios sede, e num montante grande, de R\$ 105 e R\$142. Já no mercado informal, todas diferenças são estatisticamente diferentes, porém a significância da diferença é muito fraca, e os montantes das diferenças muito pequenas, entre R\$23 e R\$45.

Figura 16: Densidade de kernel da renda média dos chefes dos domicílios dos setores censitários dos municípios da amostra. Por categoria de município e setores em AGSN (esquerda) e setores não especiais (direita).



Fonte: Censo 2010.

Uma das variáveis que está mais envolvida na formação dos aluguéis é a renda das famílias ((ROBACK, 1982). Seriam estes rendimentos também mais homogêneos nos AGSN? Dispomos de informação dos rendimentos médios dos chefes dos domicílios por setor censitário. A figura 16 mostra a distribuição dessa renda nos 510 municípios de análise, densidade de kernel, ponderada por número de domicílios em cada setor censitário e separada por setores que pertencem aos municípios sede, do entorno e isolados. A figura é bem clara em ilustrar como também essa variável é mais homogênea nos setores subnormais, e como nos setores não especiais os chefes dos domicílios dos municípios sede apresentam em geral renda superior à dos municípios do entorno, e isolados.

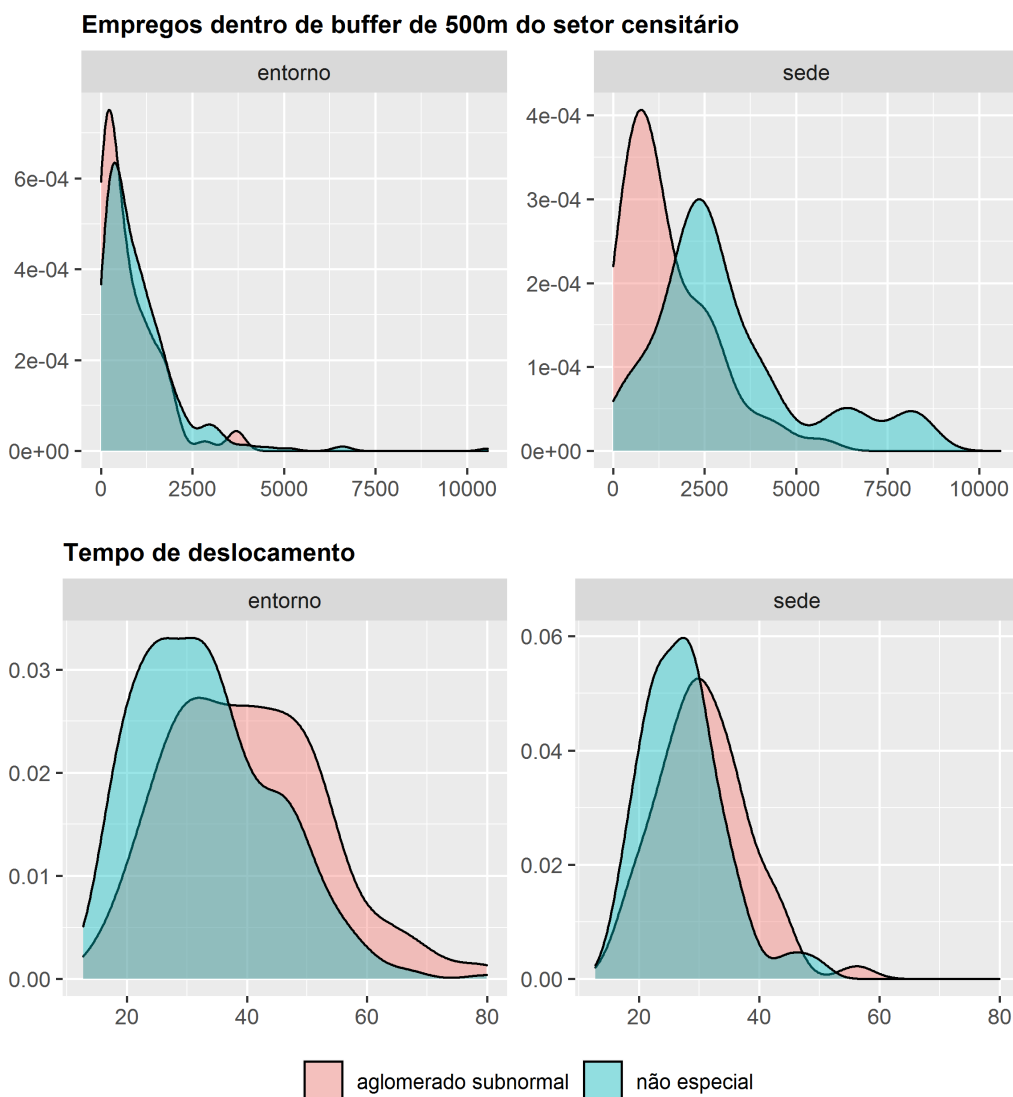
Esse resultado já havia sido encontrado na análise dos padrões regionais dos aluguéis dos 70 arranjos populacionais e dos 106 municípios isolados. Ou seja, há uma homogeneidade regional no valor do aluguel em AGSN, assim como há uma homogeneidade em termos de categoria do município, se sede, isolado ou entorno.

Em seguida, passamos a comparar a distribuição somente dos municípios sede e do entorno com relação às variáveis sensíveis às diferenças na estrutura urbana das cidades. Nem é necessário falar da distância ao centro das cidades, pois por definição esta é menor nos municípios sede. Mas, com relação à posição relativa dos domicílios e dos empregos formais presentes no núcleo urbano, como era de se esperar, os domicílios do município sede estão próximos a mais empregos que os do entorno.

Outra característica importante, que nos mostra a Figura 17: Empregos dentro de um buffer de 500 m, é que nos municípios do entorno praticamente não há diferenças entre AGSN e setores não especiais na quantidade de empregos próximos aos domicílios. Já nos municípios sede, há menos variabilidade da quantidade de empregos próximos e menor quantidade para os domicílios em AGSN. Nas sedes, os domicílios em AGSN em geral estão próximos a 1000 empregos menos que nos domicílios em setores não especiais.

Como a inserção urbana dos domicílios e sua posição relativa aos empregos impacta diretamente o tempo de deslocamento casa-trabalho, a Figura 17: Tempo de deslocamento também mostra uma situação semelhante à da figura anterior: o tempo de deslocamento é maior, em geral, no entorno, e é pior para AGSN.

Figura 17: (a) Proximidade a empregos e (b) tempo de deslocamento. Municípios sede e do entorno da amostra.



Fonte: Censo 2010.

No entanto esta informação mostra uma diferença entre AGSN e não especiais nos municípios do entorno que na proximidade dos empregos não havia. Parece que há poucos empregos disponíveis no entorno, então para o buffer utilizado de 500 metros, há poucos empregos próximos tanto de AGSN como de não especiais. A diferença do tempo de deslocamento, no entanto, reforça que os AGSN ocupam bairros mais distantes, ou com menos acesso a infraestrutura de mobilidade urbana.

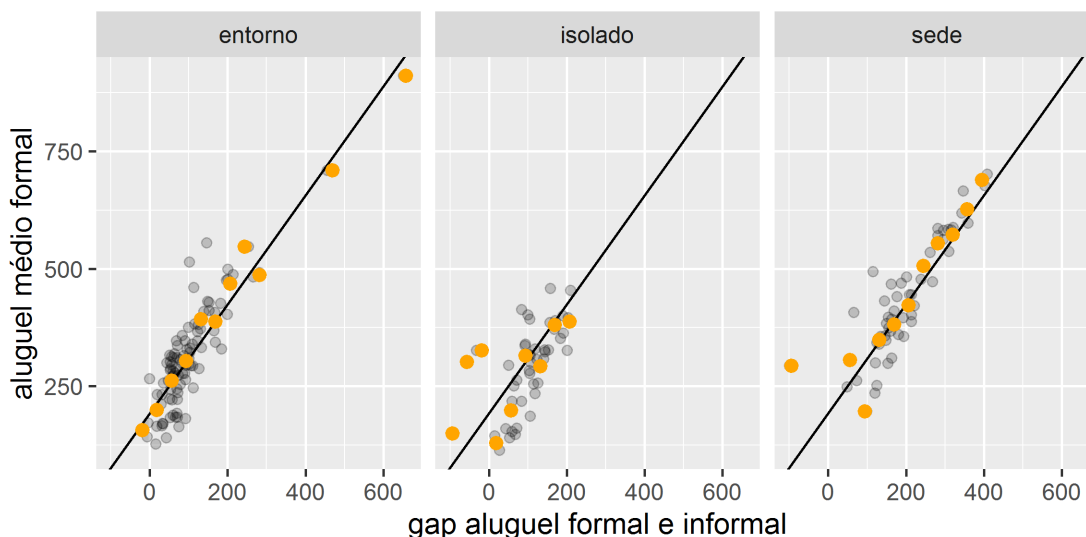
c. *Binned Scatterplot*

Nesta subseção apresentaremos os resultados da análise quantitativa que indica uma correlação positiva entre o valor médio do aluguel formal e a diferença de médias do aluguel formal e informal em 190 municípios da nossa análise, para os quais há AGSN e há informação suficiente de aluguel informal. A figura 18 traz um gráfico de dispersão simples ilustrando a correlação para os municípios sede, do entorno e isolados.

Este achado é relevante, pois indica algum mecanismo de ligação entre ambos mercados que deve ser melhor investigado. A figura apresenta em amarelo as médias ressaltadas para grupos de intervalos e uma reta que corresponde a uma regressão linear simples da diferença de médias dos aluguéis formal e informal na média do aluguel formal. Esta regressão simples pode ser entendida como uma regressão de preços hedônicos, não tanto como em (ROSEN, 1974), mas sim como em (ROBACK, 1982), pois cada unidade da regressão é um município, um mercado (ou submercado) distinto e não estão presentes na análise as características da moradia que compõem os modelos hedônicos. No entanto, essa correlação simples não teria interpretação causal devido à explicação já conhecida do viés de variável omitida, como lembrado por (GREENSTONE, 2017). No caso, uma correlação entre características não mensuradas dos municípios e a diferença dos aluguéis, por exemplo, poderia estar “mascarando” o efeito das características do mercado informal no formal, por exemplo.

Os principais determinantes das médias dos aluguéis formais que estão de fora da regressão simples da Figura 18, e que poderiam estar correlacionados com o “gap” do mercado formal e informal são três: as características das estruturas físicas formais; as amenidades ambientais dos bairros formais e as diferenças dentre os municípios de renda e atividade econômica (os salários do modelo de (ROBACK, 1982)). Mas, necessário seria aprofundar o conhecimento das possíveis relações entre os mercados, para determinar como seriam as correlações entre esses determinantes do mercado formal e o mercado informal.

Figura 18: Gráfico de dispersão do aluguel médio formal e diferença de médias do mercado formal e informal



Fonte: Censo 2010.

Os mercados formal e informal são mercados substitutos em alguma medida, mas de que maneira? Há modelos econômicos teóricos que nos indicam que ambos mercados disputam os mesmos tipos de terra e que o mercado informal diminui a oferta de terra a ser ocupada pelo mercado formal, num contexto de oferta fixa de terra (BRUECKNER; SELOD, 2009), ou oferta flexível de terra (SELOD; TOBIN, 2018). (SMOLKA; BIDERMAN, 2011) descrevem para a América Latina que este seria o caso, pois, seria mais rentável para o loteador não seguir os padrões urbanísticos formais. Mesmo assim há ocupação de áreas que não seriam ocupáveis pelo mercado formal, dando alguma “barreira de entrada” ao mercado formal, como por exemplo, as áreas com riscos de desastres ocupadas por favelas. Diferentes custos de formalização (CAVALCANTI; DA MATA; SANTOS, 2019), diferentes riscos de remoção, diferentes possibilidades de acesso a serviços urbanos de saneamento básico e segurança pública e diferentes possibilidades de inserção urbana influenciam a relação entre os mercados formal e informal, certamente.

A técnica utilizada para investigar a robustez da correlação da figura 19 é o gráfico de dispersão em grupos de intervalos (*binned scatterplots*); uma técnica de visualização da correlação entre duas variáveis que aprimora o gráfico já mostrado. Os grupos de intervalos (*bins*) não são tão essenciais no nosso caso, uma vez que não dispomos de uma quantidade de observações tão grande que faça a nuvem de pontos se tornar ilegível. O que é mais essencial, a técnica permite a inserção de variáveis de controle.

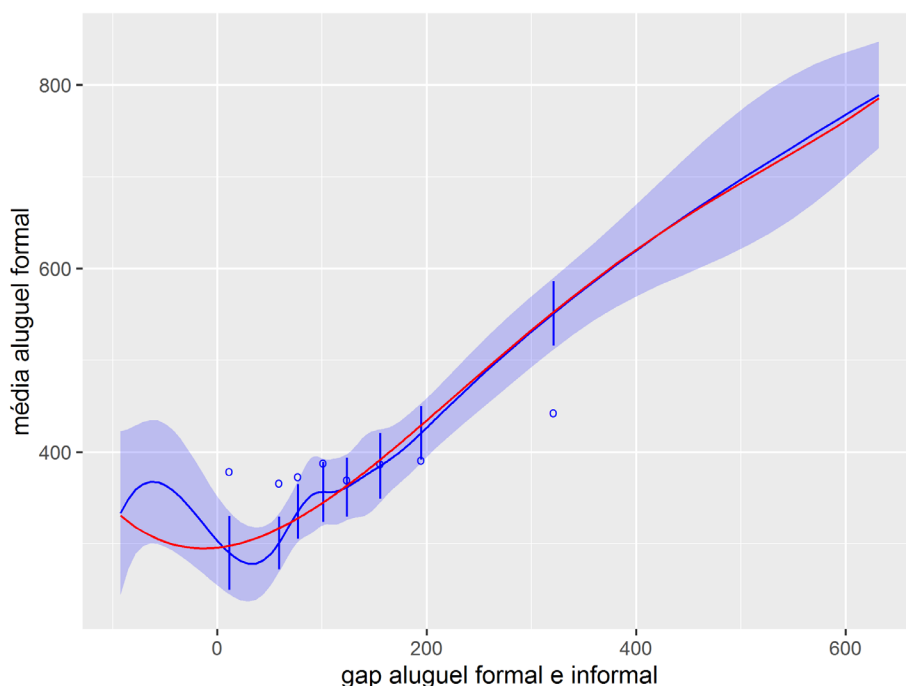
Além disso, a técnica permite avaliar a função de esperança condicional de uma variável com relação à outra, avaliando se a relação é de fato linear, como pressuposto em uma regressão linear. (CATTANEO; CRUMP; FARRELL; FENG, 2021) Além de ser uma técnica de suavização (não paramétrica), permite o cálculo de bandas de confiança, intervalos de confiança para uma função. Trata-se de uma função, pois, a suavização permite flexibilidade para que a relação entre as duas variáveis não seja linear, e sim seja uma “função”. O pacote utilizado é o *binsreg* no software R.

Infelizmente, não há estudos ou modelos que tratem as relações de causalidades acima descritas, envolvendo as duas variáveis de interesse, de maneira a construir um grafo acíclico direcionado de causalidade (PEARL, 1995), que nos desse embasamento teórico para contornar assertivamente o problema de viés de variável omitida, indicando quais variáveis de controle são “bons controles”, ou seja, amenizam o viés de variáveis omitidas mas não trazem viés de variável confundidora, por exemplo (PEARL, 1995).

No entanto, visto a discussão acima e a literatura citada, sabemos que as diferenças na qualidade das unidades habitacionais e as diferenças de renda explicam os preços/aluguéis formais. Podemos argumentar que a qualidade das unidades formais médias parece estar descolada do funcionamento do mercado informal, visto que somente uma parte do mercado formal é substituída do mercado informal, dada a grande desigualdade de renda Brasileira. Já o padrão regional de desigualdade de renda não parece explicar as características do mercado informal, um resultado da nossa análise de LISA da seção 4 deste trabalho. Isto nos encoraja a incluir variáveis dicotômicas indicando pertencimento às grandes regiões como controle no *binscatter*, para levar também em consideração o papel do desenvolvimento econômico, ou diferenças em salários como no modelo de (ROBACK, 1982). O controle pela condição de municípios sede, ou do entorno também capta a diferenciação de renda intrametropolitana.

A Figura 19 traz o resultado. No nosso caso, apesar da flexibilidade em assumir qualquer forma funcional, a relação entre as duas variáveis: a média do aluguel formal e a diferença das médias do aluguel formal e informal, parece de fato linear e aparece mesmo após a inserção das variáveis de controle. Esta “função” de correlação foi calculada levando em consideração como controles a média do número de banheiros dos domicílios formais, a qual grande região pertencia o município e se se encontrava na sede, no entorno ou se era município isolado. A regressão linear simples que corresponde a essa correlação linear é mostrada na Tabela 3, indicando que uma diferença de R\$ 100 entre o aluguel formal e informal está associada a um aluguel médio formal R\$ 79 maior.

Figura 19: Gráfico de dispersão em grupos de intervalos (*binned scatterplots*) da média do aluguel formal e da diferença das médias do aluguel formal e informal. Municípios da amostra.



Nota: Variáveis de controle: média do número de banheiros dos domicílios formais, a qual grande região pertencia e se o município se encontrava na sede, no entorno ou se era município isolado. A linha vermelha corresponde a um modelo global de regressão polinomial cúbica de grau 3. A linha azul corresponde à regressão polinomial nos grupos de intervalos (*bins*). Já os pontos correspondem ao gráfico de dispersão em *bins* tradicional. A mancha azul é a banda de confiança. Fonte: Censo 2010.

Tabela 3: Resultados da Regressão linear da média do aluguel formal, com covariáveis

	<i>Variável Dependente:</i>
	Média do aluguel formal
Diferença das médias do aluguel formal e informal	0.79*** (0.68, 0.89)
Número médio de banheiros dos domicílios formais	162.18*** (101.46, 222.91)
Nordeste	-103.01*** (-125.88, -80.15)
Sudeste	-21.40 (-43.91, 1.11)
Centro-Oeste	5.07 (-21.18, 31.33)
Sul	-20.73 (-62.63, 21.17)
Municípios sede*	31.66*** (12.71, 50.61)
Municípios do entorno*	29.95*** (14.88, 45.01)
Constante	-123.43 (-246.80, -0.06)
Observações	190
R ²	0.87
R ² Ajustado	0.86
Estatística F	148.99*** (df = 8; 181)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. * Em comparação aos municípios isolados.

Fonte: Censo 2010.

6. CLUSTERS DE SETORES CENSITÁRIOS NO RIO DE JANEIRO E SÃO PAULO. OS AGSN SÃO UM CLUSTER SEPARADO?

Esta seção tem como objetivo investigar quais seriam os segmentos do mercado habitacional nas duas maiores metrópoles brasileiras, Rio de Janeiro e São Paulo agrupando os setores censitários de acordo com suas características. Seriam segmentos de “alta renda”, de “média renda” e de “baixa renda”? Tomando os AGSN como uma mensuração da cidade “informal”, seriam os AGSN um segmento de mercado distinto do segmento de “baixa renda”, ou “baixíssima renda” formal?

O mercado de habitação pode ser segmentado com base em preços semelhantes em um mesmo “bairro”, ou porção da cidade (GOODMAN; THIBODEAU, 1998). Vamos partir do princípio oposto: checar como pequenas porções do território (os setores censitários) compartilham características semelhantes e observar se estes grupos de setores censitários semelhantes estão aglomerados em territórios suficientemente contíguos. O princípio é semelhante, porém o ponto de vista o inverso, pois, os preços das habitações “resumem” as características tanto dos imóveis quanto da vizinhança.

Uma técnica de análise exploratória de dados, os algoritmos de clusterização de aprendizado de máquina, é ferramenta que propicia fazer uma partição de um conjunto de dados de maneira a obter casos mais semelhantes entre si dentro de um mesmo grupo e grupos suficientemente distintos entre si. À diferença dos demais algoritmos de aprendizado de máquina, na clusterização não se sabe de antemão a quais grupos cada observação pertence, pois o importante é “descobrir” quais são os grupos.(IRIZARRY, 2019)

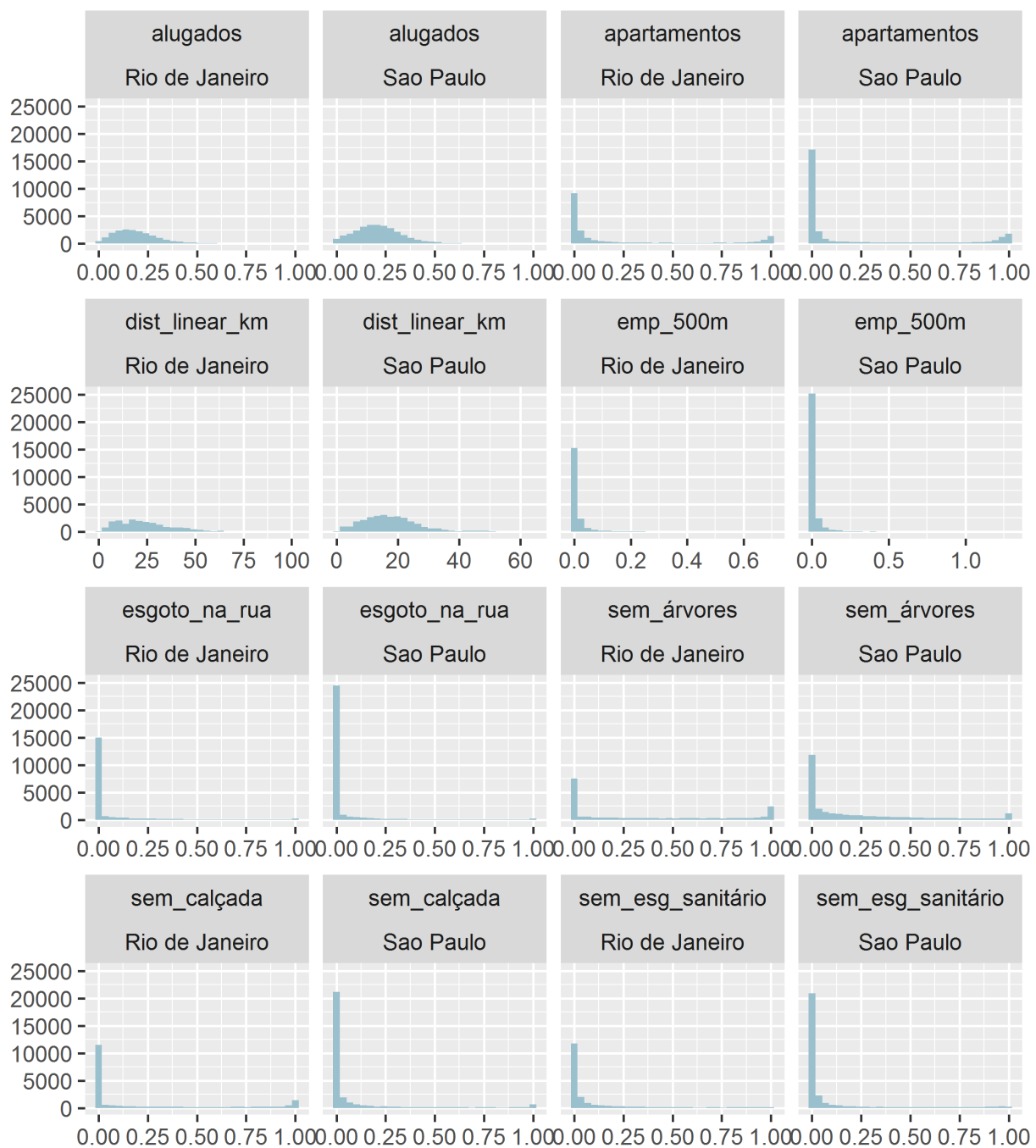
No nosso caso, as observações serão os setores censitários do Censo 2010, e os “previsores”, variáveis explicativas, ou características de cada grupo, serão as variáveis provenientes do questionário do universo do Censo 2010, enriquecidas com as variáveis que descrevem o entorno dos domicílios (IBGE, 2010b) e três variáveis externas ao Censo 2010, a distância do setor censitário ao ponto central da cidade: Sé em São Paulo e Praça Mauá no Rio de Janeiro; a declividade do setor censitário e a quantidade de empregos formais dentro de um *buffer* de 500 metros de cada setor censitário.⁶

Os valores observados passam por uma normalização para que variáveis de diferentes escalas possam ser comparadas. Esses valores são utilizados para calcular a distância euclidiana entre as observações, nossa medida de semelhança. Faremos uma breve análise de componentes principais para utilizar as características que “mais variam” entre os setores censitários. Com relação aos diferentes métodos, que utilizam diferentes passos para aglutinar as observações em grupos, testaremos duas técnicas de clusterização: kmeans e cluster hierárquico.(IRIZARRY, 2019).

Por fim, vale a pena visualizar o formato das distribuições dos previsores utilizados na separação dos clusters. Quase todas as variáveis são proporções de domicílios com determinada característica com relação ao total de domicílios de cada setor censitário. Na Figura 20, somente a distância ao centro não é uma porcentagem. Nota-se muita polarização em muitas das variáveis, ou pouca variação. A polarização parece ser maior em São Paulo que no Rio de Janeiro.

⁶ As informações de empregos formais geocodificados estão descritas em (PEREIRA; BRAGA; SERRA; NADALIN, 2020), a distância ao centro foi calculada pela autora utilizando os centroides dos setores censitários como referência, calculando a distância euclidiana.

Figura 20: Histograma das variáveis descritoras selecionadas



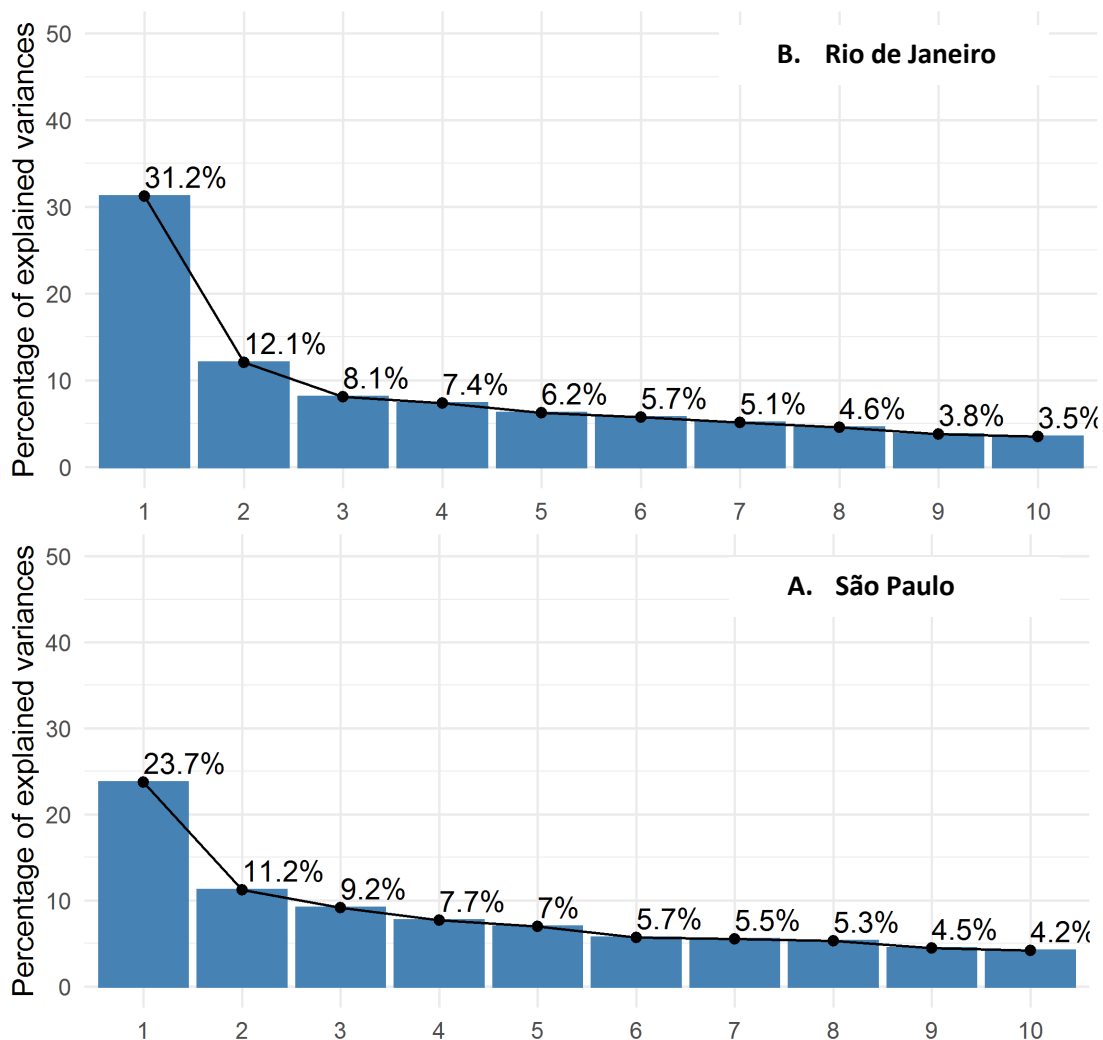
Fonte: Censo 2010.

d. Análise de componentes principais

Partimos de uma seleção de 16 variáveis que descrevem os setores censitários: 7 do entorno dos domicílios, 6 do questionário do universo e 3 externas (distância ao centro, declividade e proximidade a empregos formais). Algumas destas variáveis contêm informações redundantes entre si, ou seja, acrescentá-las na análise não confere mais poder explicativo, somente adiciona mais ruído. Portanto, vale a pena investigar essas redundâncias entre os “previsores”, para conseguir melhor performance na análise de cluster. Quanto mais ruído dentre os previsores, maior a tendência de existir um grupo que concentra quase todas as observações.

Dentro da linguagem do software R, utilizamos os pacotes factoextra e FactoMineR para fazer as análises aqui descritas. Os dois painéis da Figura 21 a seguir mostram a porcentagem da variância explicada de cada componente principal encontrado. Os componentes principais são combinações das 16 variáveis. Em SP os 5 primeiros componentes explicam 58,7% da variância, enquanto no Rio de Janeiro os 5 primeiros componentes principais explicam 65% da variância.

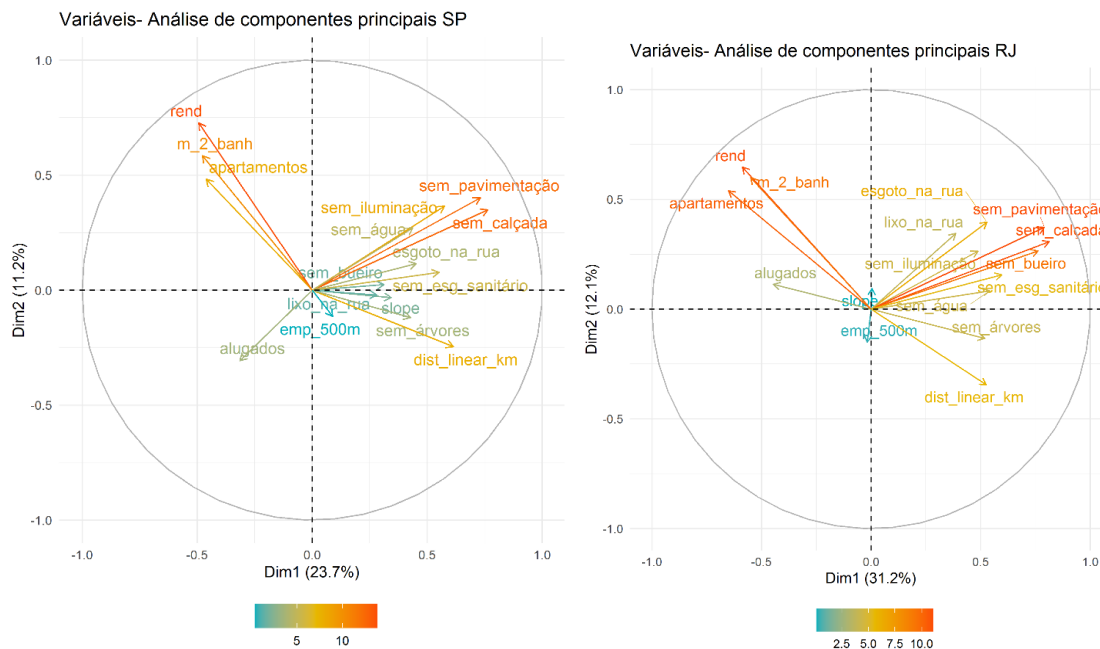
Figura 21: Scree plot: Gráfico da variância explicada a cada componente principal adicionado.



Fonte: Censo 2010.

Ao analisar a contribuição de cada uma das 16 variáveis para os 2 primeiros componentes principais, é possível visualizar quais variáveis dão a mesma contribuição para estes componentes (flechas muito próximas entre si) sendo, portanto redundantes, e quais contribuem de forma diferenciada. A Figura 22 traz estas visualizações para SP e RJ.

Figura 22: Variáveis que contribuem para os dois primeiros componentes principais



Fonte: Censo 2010.

Em congruência com estas análises descritivas, e com as principais variáveis que contribuem para os primeiros componentes principais, escolhemos como “previsores” para serem utilizados na análise de cluster os seguintes conjuntos de variáveis:

SP: "emp_500m", "dist_linear_km", "sem árvores", "lixo na rua", "sem calçada", "mais de 2 banheiros" e "alugados"

RJ: "apartamentos", "emp_500m", "dist_linear_km", "sem árvores", "sem calçada" e "esgoto_na_rua".

e. Resultado da análise de clusters e descrição dos clusters encontrados

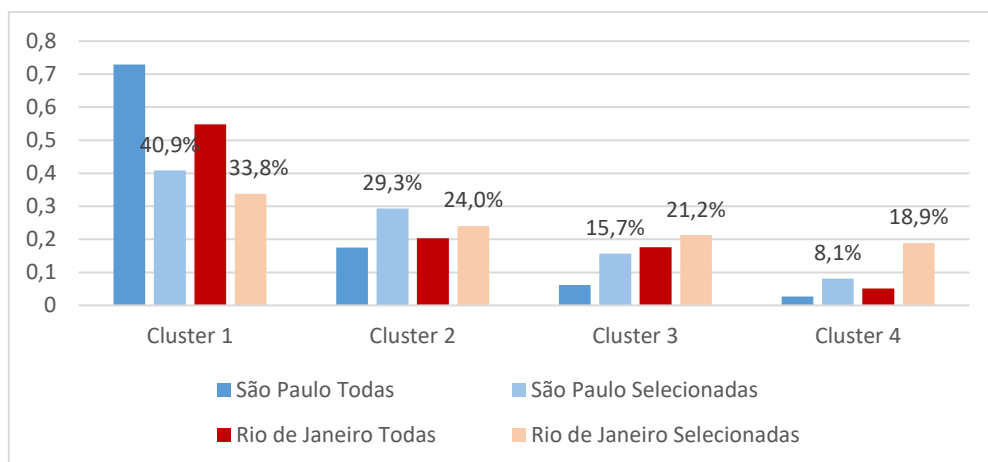
Se as informações dos setores censitários fossem totalmente aleatórias não seria possível identificar clusters, ou subgrupos, com características distintas. Então para investigar se de fato há uma estrutura que diferencia os setores censitários foram feitos testes que tentam descobrir qual seria o número “ótimo” de clusters presente nos dados, o teste do “cotovelo” (“Elbow”) e da silhueta (“Silhouette”). Eles se diferenciam entre si na maneira de medir o quanto os clusters são homogêneos internamente e heterogêneos entre os grupos⁷. Estas análises foram feitas utilizando as variáveis escolhidas na seção acima. Para São Paulo e Rio de Janeiro

⁷ O teste de cotovelo investiga a porcentagem da variância explicada como uma função do número de clusters utilizados. A partir de certo ponto, adicionar mais um cluster vai adicionar bem menos porcentagem de variância explicada. Este é o ponto do “cotovelo” do gráfico. Já o teste da silhueta procura o número de clusters que maximiza o valor médio da silhueta. Este valor corresponde a o quanto cada observação é similar ao seu cluster, e distante dos demais clusters.

valida-se a existência de 7 clusters, os gráficos são mostrados no anexo. Mesmo assim, escolhemos separar SP e RJ em 6 clusters, já que não há perda significativa de poder explicativo.

A redução da dimensionalidade de fato serviu para tirar ruído das variáveis “previsoras”. Utilizando o método kmeans para formar os clusters com todas as 16 variáveis, a Figura 23 mostra a proporção de moradores nos 4 maiores clusters formados. Nota-se como os moradores se concentram demasiadamente em um cluster só no caso de SP (73%). Aplicando o algoritmo com as variáveis escolhidas há mais diferenciação das observações, com 41% dos moradores no cluster maior. No RJ o problema não seria tão grave (de 55% para 34%).

Figura 23: Proporção de moradores em cada cluster utilizando todas as 16 variáveis disponíveis ou as variáveis selecionadas via análise de componentes principais



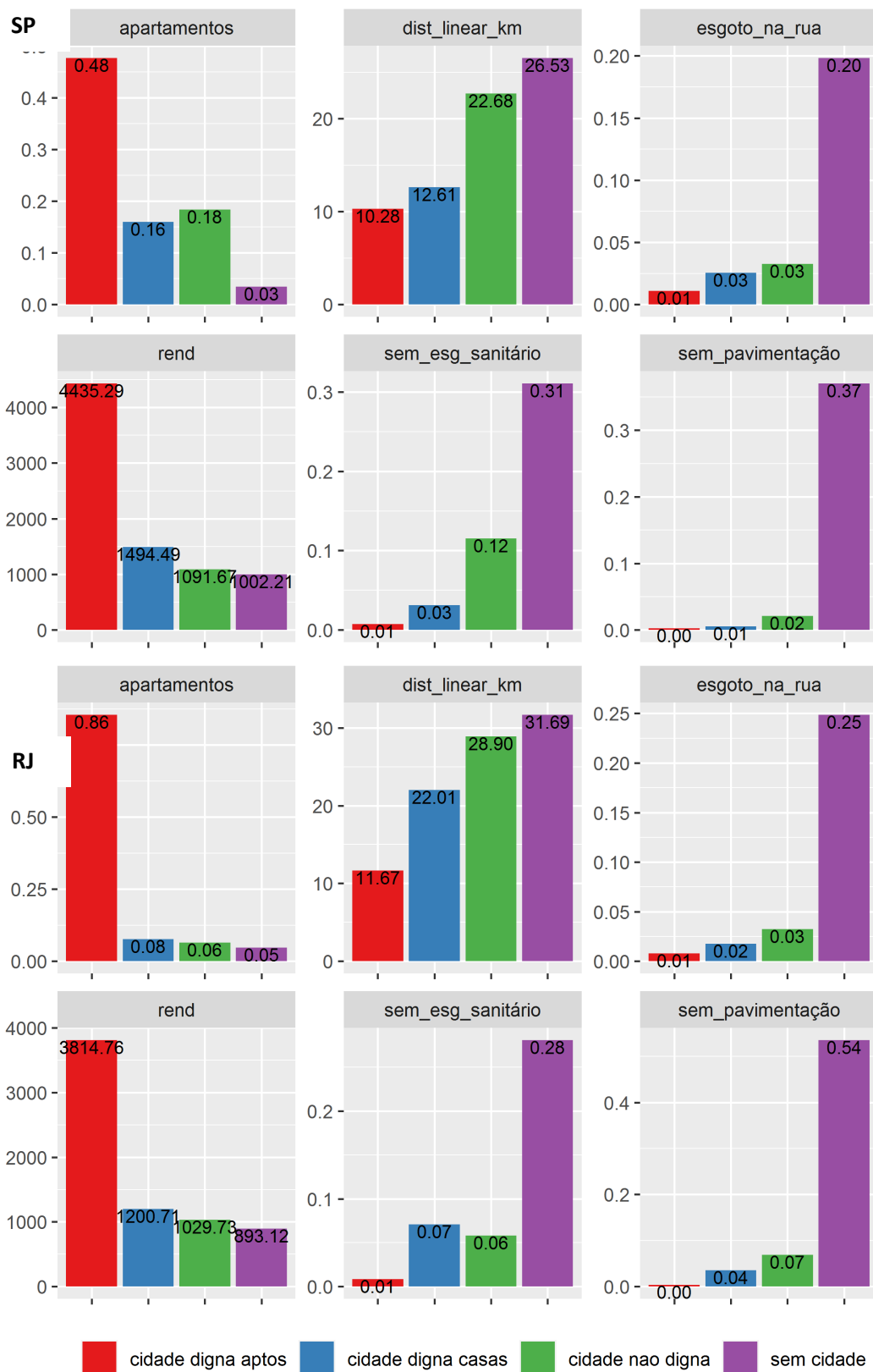
Fonte: Censo 2010.

A figura 23 também nos indica que as proporções finais dos clusters são semelhantes entre as 2 cidades. Em seguida mostramos as diferenças entre as médias ponderadas pelo número de domicílios das variáveis “previsoras” para os diferentes clusters. As combinações de características nos levaram a escolher os nomes para os clusters. No caso de SP os 2 menores clusters formados foram desconsiderados. No RJ 2 clusters de características muito próximas foram aglutinados e um desconsiderado, o de menor tamanho. Escolhemos os seguintes nomes: “cidade digna casas”; “cidade não digna”, “cidade digna apartamentos” e “sem cidade”.

As médias ponderadas por total de domicílios dos previsores varia bastante entre esses clusters. De fato, estes nomes foram escolhidos por serem condizentes com as principais características de cada cluster. A figura 24 mostra essas diferenças para 6 das variáveis consideradas, as demais são apresentadas no anexo 2.

Por exemplo, o cluster “cidade digna apartamentos” é composto em sua maioria por apartamentos (48% dos domicílios em SP e 86% no RJ), apresenta renda muito maior que os demais clusters, e todos indicadores de infraestrutura urbana e serviços urbanos adequados. Em SP ele corresponde a 15,7% do total de moradores, e no RJ, 19%. Este cluster seria uma delimitação apropriada da cidade dos “ricos”, ou formal. Já o cluster “cidade digna casas”, não é composto por apartamentos (16% em SP e 8% no RJ). É o maior cluster no RJ: 33% dos moradores, mas o segundo maior em SP: 29% dos moradores. Apresenta a segunda maior renda média do chefe do domicílio e indicadores de infraestrutura (pavimentação, calçada e iluminação) muito próximos ao cluster da cidade mais rica. Já os indicadores de serviços como esgotamento, abastecimento de água, lixo na rua e esgoto na rua, podem ser parecidos com a cidade rica ou piores.

Figura 24: Variáveis selecionadas: médias ponderadas por número de domicílios por clusters.



Fonte: Censo 2010.

O nome “cidade não digna” implica um limite arbitrário de inadequação que este cluster teria ultrapassado. Lembrando que os valores limites das características de cada cluster são resultado do algoritmo de agrupamento. A “cidade digna de casas” apresenta indicadores de inadequação às vezes próximos à “cidade não digna”, mas este cluster é sempre pior que aquele. Características como a falta de calçadas, pavimentação e iluminação públicas são consideravelmente piores no RJ. A falta de esgotamento sanitário e abastecimento de água adequados são consideravelmente piores em SP.

Por outro lado, o cluster “sem cidade” apresenta algumas características muito destoantes dos demais, claramente concentrando inadequações. A falta de esgotamento sanitário e a falta de abastecimento de água em SP são notórias, maiores que no RJ. A magnitude da inadequação do entorno do domicílio é maior no RJ: 78% de domicílios sem calçadas, 25% com esgoto na rua, 60% sem árvores (73%, 20% e 55% respectivamente, em SP). Em São Paulo 8,8% da população está neste cluster, enquanto no RJ 24%.

f. Localização intraurbana dos clusters

Nesta seção apresentamos a localização dos clusters. Os mapas, nas Figura 25 e 26, mostram os centroides dos setores censitários diferenciados por cores que distinguem a que cluster pertencem. Como dito anteriormente, uma definição de submercado de habitação fala explicitamente em uma subdivisão do território de uma cidade, em que os preços sejam suficientemente semelhantes. No nosso caso desconsideramos a estrutura espacial da cidade, e as relações de vizinhança física entre os setores censitários. As variáveis de distância ao centro e declividade média do terreno são as únicas que descrevem de alguma maneira as propriedades geográficas dos setores.

Mesmo assim, os clusters formados apresentam uma estrutura espacial visualmente identificável, conforme mostram as figuras em seguida, assim como a localização separada de cada cluster no anexo. Como era de se esperar, os setores censitários do cluster “cidade digna apartamentos” se concentram nas zonas centrais do RJ e SP, correspondendo aos bairros mais valorizados e de melhor infraestrutura e serviços urbanos. A “cidade digna casas” ocupa os bairros centrais próximos aos da “cidade digna aptos”, mas de bairros mais identificados com a classe média, como a Zona Leste em SP e Tijuca, Vila Isabel e Meier no RJ. A “cidade não digna” e “sem cidade” estão claramente localizados periféricamente, sendo que esta última tende a ser mais periférica que a anterior.

Os limites não são perfeitamente definidos, e as regiões de maior concentração de um cluster muitas vezes se sobrepõem às regiões de concentração de outro cluster.

Assim, a partir desta classificação seria impossível criar uma partição do território, gerando uma divisão da cidade em submercados, como na definição mais tradicional de submercados de habitação. Inclusive, no caso do RJ, o cluster “sem cidade” está muito mais entremeado nas regiões centrais em que os demais clusters se concentram que em SP. Em SP “sem cidade” está bem claramente localizado nas zonas mais periféricas possíveis. A distância ao centro média é, de fato, maior no cluster “sem cidade” em SP. No RJ essa média é maior, porém a variância muito ampla⁸, havendo muitos setores do tipo “sem cidade” muito próximos aos do cluster “cidade digna apartamentos”.

⁸ O desvio padrão da distância para os “sem cidade” no RJ é de 13,6 km enquanto em SP 9,1 km.

Figura 25: Localização Intraurbana dos clusters. São Paulo

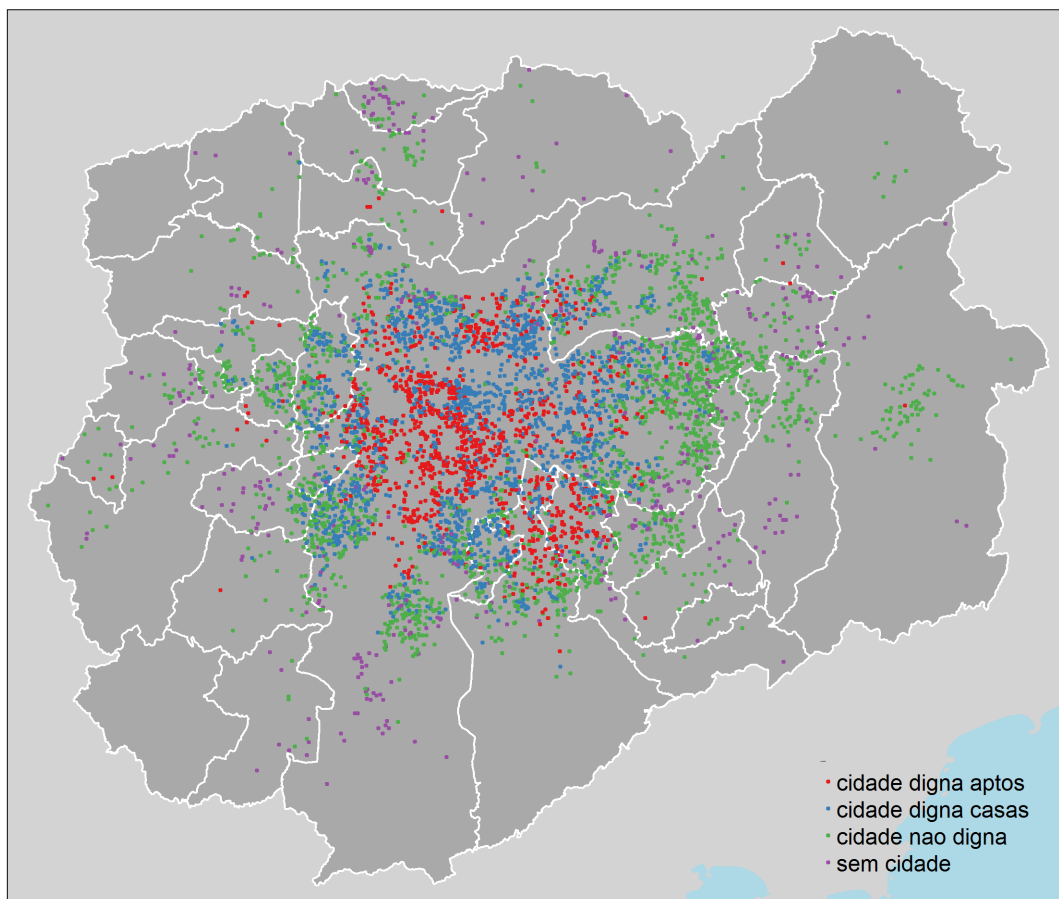
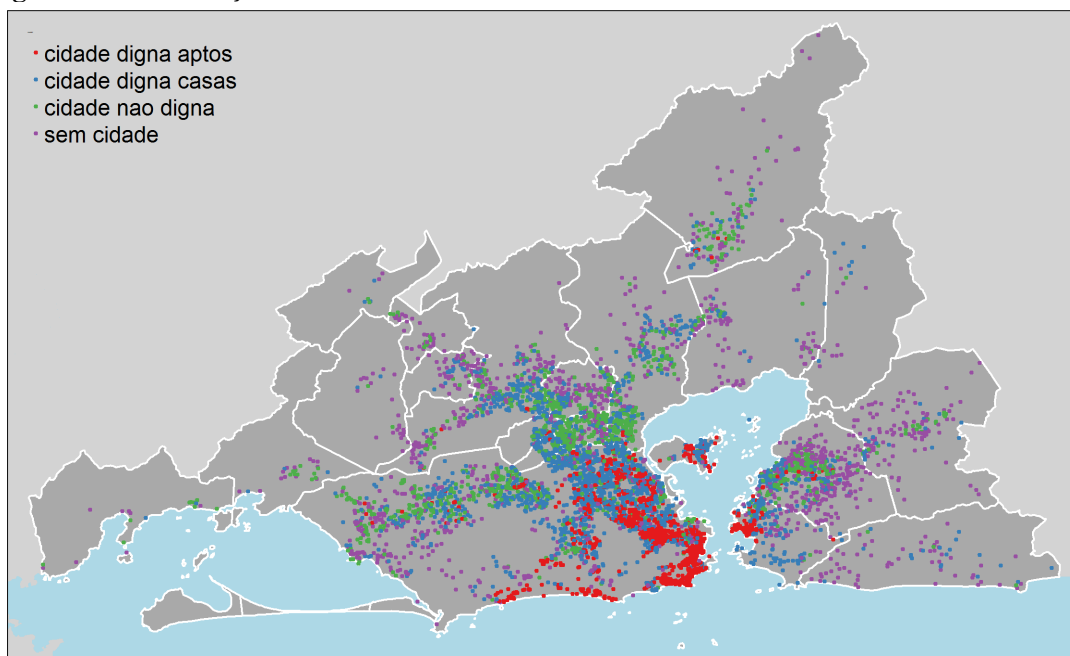


Figura 26: Localização Intraurbana dos clusters. Rio de Janeiro



Fonte: Censo 2010.

g. Correspondência com os setores subnormais.

Já que os setores do cluster “sem cidade” são muito piores em termos de características urbanas e de renda, seria esse cluster o correspondente à cidade “informal”, ou às favelas? Na nossa classificação disponível, aos setores subnormais? No RJ os “sem cidade” estão até localizados mais centralmente, visivelmente nos morros dos bairros nobres da cidade, onde ficam as favelas.

A tabela 4 mostra o quanto os moradores dos setores censitários classificados como AGSN se distribuíram nos clusters criados. No RJ os moradores em AGSN se concentraram mais no cluster “cidade digna casas”, enquanto em SP se concentraram mais em “cidade não digna”. Pouquíssimos moradores de AGSN estão em setores do cluster “cidade digna apartamentos”. Mas, apesar da concentração naqueles clusters, os moradores de AGSN não correspondem à maioria dos totais de moradores em nenhum cluster.

Tabela 4: Distribuição de moradores e moradores em AGSN nos clusters encontrados

		A	B	A/B	$A/\sum A$	$B/\sum B$
		Total de moradores em AGSN	Total de moradores	% de moradores em AGSN	% do total de moradores em AGSN	% do total de moradores
Rio de Janeiro	cidade digna aptos	46.283	2.272.588	2%	3%	19%
	cidade digna casas	1.117.262	4.060.509	28%	64%	34%
	cidade nao digna	196.658	2.555.299	8%	11%	21%
	sem cidade	365.619	2.889.327	13%	21%	24%
São Paulo	cidade digna aptos	11.535	3.059.159	0%	1%	16%
	cidade digna casas	202.563	5.709.026	4%	9%	29%
	cidade nao digna	1.476.500	7.972.023	19%	68%	41%
	sem cidade	355.058	1.585.994	22%	16%	8%

Fonte: Censo 2010

No caso do cluster “sem cidade”, tanto em SP como no RJ a grande maioria dos moradores de “sem cidade” não estão em setores classificados como AGSN. As grandes inadequações urbanas identificadas nesse cluster demonstram como é ele que concentra as necessidades habitacionais das duas metrópoles estudadas. Devido à definição de AGSN incluir algumas inadequações habitacionais: iluminação elétrica, coleta de lixo, água e esgotamento sanitário, esperaríamos que ambas classificações fossem mais próximas. Cabe notar que na definição de AGSN do IBGE há dois aspectos que não são mensurados pelos questionários do Censo: a urbanização fora do padrão e a ocupação de áreas restritas, como faixas de domínio de rodovias e áreas contaminadas, por exemplo. Essas informações são coletadas antes da aplicação dos questionários, quando da classificação dos setores censitários em subnormais ou não.

Este resultado ressalta mais uma vez a heterogeneidade dentre os AGSN. Alguns estudos empíricos já evidenciaram essa heterogeneidade (BRUECKNER; MATION; NADALIN, 2019; MATION; NADALIN; KRAUSE, 2014). Há AGSN mais bem servidos, e menos bem servidos. Porém essa classificação não corresponde à concentração de inadequações urbanas. Talvez a questão chave para interpretar este resultado seja que, nestas metrópoles analisadas, o fator que de fato distingue os AGSN seja a inadequação fundiária, a urbanização fora do padrão e a ocupação de áreas restritas, informações não disponíveis, e não utilizadas na análise aqui apresentada.

Encontramos evidências de que tanto o mercado formal, quanto o mercado informal (aqui visto como os AGSN) são segmentados, e que há regiões das metrópoles analisadas que se diferenciam bastante em termos de inadequações habitacionais e urbanas. Essas diferenciações podem não ser captadas pelos AGSN.

7. TABELA COM PRINCIPAIS RESULTADOS

Variável	Fatos estilizados encontrados	Literatura com achado semelhante
Distância ao centro	heterogeneidade entre cidades: poucas em que a maioria dos AGSN são mais próximos	(CATALÁ; CARMO, 2021; NADALIN; MATION, 2018)
	para o conjunto do Brasil, a mediana geral dos AGSN é maior.	inédito
	SP e Brasília AGSN mais distantes, RJ e BH AGSN mais próximos do centro	(BRUECKNER; MATION; NADALIN, 2019; CATALÁ; CARMO, 2021; NADALIN; MATION, 2018)
Total de empregos em <i>buffer</i> de 500m do setor censitário	AGSN estão próximos a menos empregos formais	(LINKE; SERRA; GARREFA; ARAÚJO <i>et al.</i> , 2016; NADALIN, 2020; NAKAMURA; AVNER, 2018)
	Em RJ e Recife os AGSN são próximos a mais empregos que SP e Manaus	inédito
	Diferença é maior para municípios sede de arranjos populacionais que para municípios do entorno.	inédito
Tempo de deslocamento	Medianas de AGSN e não especiais coincidem para muitas cidades. Em geral, AGSN apresentam tempos maiores.	(CELHAY; GIL, 2020)
	Os tempos dos "melhores" AGSN, percentil 25, são sempre bem menores que os dos "melhores" não especiais, percentil 75.	inédito
	Há diferença nos tempos entre AGSN e não especiais, tanto nos municípios sede dos arranjos populacionais quanto nos municípios do entorno.	inédito
Valores de Aluguel	aluguéis de setores não especiais são maiores no Sul e Sudeste	(CRAVO; RESENDE; CRUZ, 2019)
	não há padrão regional nos aluguéis em AGSN.	inédito
	Há associação linear entre aluguel médio formal e diferença de aluguéis médios de setores não especiais e de AGSN.	inédito
	A diferença entre aluguéis médios de setores não especiais e de AGSN é maior para municípios sede de arranjos populacionais.	inédito
Análise de cluster	O cluster que concentra as inadequações urbanas não é aquele em que se concentram os moradores de AGSN.	(CATALÁ; CARMO, 2021)

8. COMENTÁRIOS FINAIS

Este texto buscou caracterizar dois segmentos do mercado habitacional brasileiro, o formal e informal, conforme a definição dos aglomerados subnormais do IBGE. As principais semelhanças e diferenças foram explicitadas através da análise de cinco níveis geográficos. Encontramos desigualdades tanto regionais quanto intraurbanas, assim como dentro da hierarquia das cidades. O que ficou mais claro é que os AGSN são muito heterogêneos, e que não são, necessariamente, sinônimo de inadequações habitacionais. Passamos a resumir os principais resultados.

Com relação à inserção urbana, distância ao centro, proximidade a empregos e tempo de deslocamento verificamos que as cidades se diferenciam bastante de acordo com diferenças na estrutura intraurbana. Mesmo assim, em geral os AGSN possuem distâncias ao centro menores, menos proximidade a empregos e mais tempo de deslocamento. No entanto, os melhores subnormais estão sempre melhor inseridos na cidade que os piores domicílios em setores não especiais.

Já com relação aos aluguéis, a distribuição dos valores em AGSN são sempre em valores menores que do mercado formal, mas há superposições de valores de ambos mercados. De acordo com a cidade há maiores porções de ambos mercados superpostas. Em geral, as características da vizinhança e de estrutura física dos AGSN são piores, justificando os alugueis menores, sendo que a ausência de revestimento em alvenaria e a presença de esgoto a céu aberto são as características que mais os distinguem. Nas cidades em que há mais domicílios em ambos mercados com valores superpostos, ou semelhantes, os domicílios em AGSN apresentam melhores características, relativamente às características dos domicílios em setores não especiais.

Há um padrão regional nos aluguéis e características da estrutura física e de vizinhança em que em geral Norte e Nordeste são piores que as demais regiões. Mesmo assim, encontramos que esse padrão espacial não é estatisticamente significativo para os aluguéis informais, somente para os formais. Isto nos indicaria que as condições de pobreza, e pouca capacidade de pagamento, não se diferenciam tanto entre as regiões do país quanto as condições de riqueza. Os maiores aluguéis de cada região apresentam variabilidade muito maior que os aluguéis informais, em geral. Este padrão de desigualdade maior para o mercado formal que para o mercado informal também é encontrado para as diferenças intraurbanas, ao considerar a diferenciação dos municípios sede, do entorno e municípios isolados. A desigualdade intraurbana de inserção urbana também é mais pronunciada nos municípios sede, onde os AGSN são menos parecidos aos setores não especiais.

A primeira análise quantitativa apresentada no estudo encontrou uma correlação entre a média de aluguéis formais e na diferença das médias dos aluguéis formais e informais para 190 municípios para os quais haviam dados disponíveis, levando em consideração como variáveis de controle a região, se era município sede e a média de banheiros dos domicílios formais. Esta correlação robusta é linear e documenta o quanto ambos mercados são substitutos.

Já a segunda análise quantitativa, utilizando algoritmos de clusterização de aprendizado de máquina, agrupou os setores censitários das metrópoles do Rio de Janeiro e de São Paulo de acordo com as suas características. Os grupos resultantes estão localizados em porções do território bastante definidas, o grupo de melhor qualidade correspondendo aos bairros sabidamente mais valorizados, por exemplo. Isto corroborou com a definição de segmentos de mercados que são porções do território com preços homogêneos.

No entanto, os AGSN não têm a maioria dos moradores em nenhum dos clusters, nem sequer no cluster que claramente concentra mais inadequações habitacionais. Os AGSN não são um cluster separado. Mesmo assim, encontram-se com maior frequência nos clusters de menor qualidade e quase não estão presentes no cluster de melhor qualidade. Este resultado corrobora a visão de que os AGSN são muito heterogêneos, e de que para diminuir as inadequações habitacionais as políticas públicas não devem restringir o foco, tratando apenas da questão da informalidade, ou dos AGSN.

Este trabalho teve cunho principalmente descritivo, mas contribui por trazer informações muito ricas e inéditas que contribuem para as reflexões necessárias à definição das políticas habitacionais e de desenvolvimento urbano no nosso país. Como futuros desenvolvimentos, ressaltamos a necessidade de resumir os achados teóricos das modelagens já disponíveis na literatura sobre o fenômeno em um grafo causal acíclico direcionado, de maneira a esclarecer os canais causais envolvidos. Assim, políticas mais assertivas e sustentáveis podem ser desenhadas.

9. REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L. Local indicators of spatial association—LISA. **Geographical analysis**, 27, n. 2, p. 93-115, 1995.

BRUECKNER, J. K.; MATION, L.; NADALIN, V. G. Slums in Brazil: Where are they located, who lives in them, and do they ‘squeeze’ the formal housing market? **Journal of Housing Economics**, 44, p. 48-60, 2019/06/01/ 2019.

BRUECKNER, J. K.; SELOD, H. A Theory of Urban Squatting and Land-Tenure Formalization in Developing Countries. **American Economic Journal: Economic Policy**, 1, n. 1, p. 28-51, 2009.

CATALÁ, L. S.; CARMO, R. L. D. O conceito de aglomerado subnormal do IBGE e a precariedade dos serviços básicos de infraestrutura urbana. **Revista Brasileira de Estudos de População**, 38, 2021.

CATTANEO, M.; CRUMP, R.; FARRELL, M.; FENG, Y. **On Binscatter**. 2021.

CAVALCANTI, T.; DA MATA, D.; SANTOS, M. On the determinants of slum formation. **The Economic Journal**, 129, n. 621, p. 1971-1991, 2019.

CAVALCANTI, T.; DA MATA, D.; SANTOS, M. On the Determinants of Slum Formation. **The Economic Journal**, 129, n. 621, p. 1971-1991, 2019.

CELHAY, P. A.; GIL, D. The function and credibility of urban slums: Evidence on informal settlements and affordable housing in Chile. **Cities**, 99, p. 102605, 2020.

CRAVO, T. A.; RESENDE, G. M.; CRUZ, B. D. O. Mensurando as disparidades regionais no Brasil: Salários, custo de vida e amenidades locais. **Revista Brasileira de Economia**, 73, p. 1-28, 2019.

GOODMAN, A. C.; THIBODEAU, T. G. Housing Market Segmentation. **Journal of Housing Economics**, 7, n. 2, p. 121-143, 1998/06/01/ 1998.

GREENSTONE, M. The continuing impact of Sherwin Rosen’s “Hedonic prices and implicit markets: product differentiation in pure competition”. **Journal of Political Economy**, 125, n. 6, p. 1891-1902, 2017.

HENDERSON, J. V.; REGAN, T.; VENABLES, A. J. Building the City: From Slums to a Modern Metropolis. **The Review of Economic Studies**, 88, n. 3, p. 1157-1192, 2020.

IBGE, C. D. D. E. D. D. I. **Censo Demográfico. Aglomerados subnormais. Primeiros resultados**. Rio de Janeiro: Ministério do Planejamento, Orçamento e Gestão 2010a.

IBGE, C. D. D. E. D. D. I. **Censo Demográfico, Resultados do universo**. Rio de Janeiro: Ministério de Planejamento, Orçamento e Gestão, 2010b.

IBGE, C. D. G. **Arranjos populacionais e concentrações urbanas do Brasil**. Rio de Janeiro: IBGE, 2015. 9788524044069.

IRIZARRY, R. A. **Introduction to data science: Data analysis and prediction algorithms with R**. CRC Press, 2019. 1000707733.

LINKE, C. C.; SERRA, B.; GARREFA, F.; ARAÚJO, D. C. *et al.* Inserção Urbana de Habitação de Interesse Social: um olhar sobre mobilidade cotidiana e uso do solo. 2016.

MATION, L.; NADALIN, V.; KRAUSE, C. **Favelização no Brasil Entre 2000 e 2010: Resultados de Uma Classificação Comparável**. Instituto de Pesquisa Econômica Aplicada-IPEA. 2014.

NADALIN, V. Slums, accessibility to opportunities and economic outcomes. An empirical analysis for São Paulo, Brazil. *In: ANPEC-48º Encontro Nacional de Economia*, 2020.

NADALIN, V. G.; MATION, L. F. **Localização intraurbana das favelas brasileiras: O papel dos fatores geográficos**. Texto para Discussão. 2018.

NAKAMURA, S.; AVNER, P. Spatial Distributions of Job Accessibility, Housing Rents, and Poverty in Nairobi, Kenya. The World Bank 2018.

PEARL, J. Causal diagrams for empirical research. *Biometrika*, 82, n. 4, p. 669-688, 1995.

PEREIRA, R. H.; BRAGA, C. K. V.; SERRA, B.; NADALIN, V. G. **Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras—2019**. Brasília. 2020.

ROBACK, J. Wages, rents, and the quality of life. *Journal of political Economy*, 90, n. 6, p. 1257-1278, 1982.

ROSEN, S. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82, n. 1, p. 34-55, 1974.

SELOD, H.; TOBIN, L. The informal city. **World Bank Policy Research Working Paper**, n. 8482, 2018.

SMOLKA, M. O.; BIDERMAN, C. Housing informality: An economist's perspective on urban planning. *In: BROOKS, N.; DONAGHY, K., et al (Ed.). The Oxford handbook of urban economics and planning: Oxford University Press., 2011.*

