

CRÍTICA AOS MODELOS DE INFERÊNCIA ESTATÍSTICA NA ANÁLISE DOS DETERMINANTES DAS DESIGUALDADES NA ACESSIBILIDADE

Maria Cristina Cavalcante Belo

Davi Garcia Lopes Pinto

Carlos Felipe Grangeiro Loureiro

Programa de Pós-Graduação em Engenharia de Transportes
Universidade Federal do Ceará

RESUMO

O diagnóstico estratégico das desigualdades socioespaciais na acessibilidade utiliza comumente ferramental de modelagem estatística para analisar causalidade. Porém, os modelos de regressão linear formulados podem não ser adequados, gerando viés nas estimativas e erros na interpretação dos efeitos causais. Portanto, o objetivo deste trabalho é criticar os modelos estatísticos que analisam as relações de causa-efeito entre as restrições dos subsistemas de uso do solo e transportes com os níveis de acessibilidade no planejamento urbano estratégico. Para isso, foram utilizados dados de Fortaleza para se exemplificar os problemas que podem ocorrer ao se realizar um diagnóstico sem estabelecer os possíveis caminhos indiretos entre a acessibilidade e suas restrições, não considerando as fontes de endogeneidade. Foi possível constatar que a análise de fenômenos complexos através de regressão linear pode se beneficiar da utilização de diagramas causais, possibilitando um melhor entendimento dos caminhos causais entre as variáveis, com o controle adequado da endogeneidade.

ABSTRACT

The strategic diagnosis of socio-spatial inequalities in accessibility commonly uses statistical modeling to analyze causality. However, the formulated linear regression models may not be adequate, generating bias in the estimates and errors in the interpretation of the causal effects. Therefore, the objective of this work is to criticize the statistical models that analyze the cause-effect relationships between land-use and transport subsystems' restrictions with accessibility levels in the strategic urban planning. For this, data from Fortaleza were used to exemplify the problems that can occur when carrying out a diagnosis without establishing the possible indirect paths between accessibility and its restrictions, in addition to not considering the sources of endogeneity. It was possible to verify that the analysis of complex phenomena through linear regression can benefit from the use of causal diagrams, allowing a better understanding of the causal paths between the variables, with the adequate control of endogeneity.

1. INTRODUÇÃO

Na busca por soluções para as externalidades negativas decorrentes de padrões de mobilidade centrados no uso de modos individuais motorizados, o planejamento urbano evoluiu com foco no conceito de mobilidade sustentável (Banister, 2008), colocando a acessibilidade, que está ligada à ideia do que pode ser acessado e de como se dão as condições do acesso, como objeto central de planejamento, reconhecendo a necessidade de integração entre os subsistemas de transportes e de uso do solo (Bertolini *et al.*, 2005). Nesse paradigma, é possível analisar como as características de uso do solo e dos sistemas de transportes podem restringir o acesso às oportunidades, possibilitando o planejamento de um sistema urbano que incentive padrões de mobilidade mais equitativos e o uso de modos menos poluentes.

Baseado nesse novo conceito de planejar, a compreensão das causas dos problemas deve vir antes de se propor sua solução (Meyer, 2016). Essa análise deve ser considerada como uma das primeiras fases do nível estratégico do planejamento (Ministério das Cidades, 2006), também chamada de diagnóstico ou avaliação dos problemas, com foco na problemática das desigualdades socioespaciais na acessibilidade, e guiada por princípios de equidade e sustentabilidade (Garcia, 2016). Uma das etapas da fase de diagnóstico consiste na análise das relações de causa e efeito hipotetizadas acerca da problemática de interesse. As ferramentas de

inferência estatística são as mais comumente empregadas para análises exploratórias e/ou confirmatórias dessas relações, como modelos de regressão linear e espacial, por exemplo (Garcia *et al.*, 2018; Sousa, 2019). Reconhece-se, entretanto, a limitação de se extrair interpretações causais a partir desses modelos, dadas as dificuldades de se incorporar em suas formulações toda a complexidade das relações envolvidas no fenômeno, como a representação de *backdoors*, que podem ocasionar endogeneidade (Shiple, 2016), além da dificuldade em se estimar efeitos indiretos das variáveis explicativas. Considerando que os objetivos estabelecidos no planejamento buscam resolver os problemas de desigualdades na acessibilidade, a análise das causas desses problemas, antes de se propor sua solução, pode evitar o esforço de se implementar modificações no sistema que não serão efetivas (Garcia, 2016). Assim, o diagnóstico deve ser bem elaborado e compreendido.

De acordo com Bollen (1989), a definição de causa-efeito possui três componentes: associação, isolamento e direção de influência. A associação entre duas variáveis não é suficiente para definir causalidade, já que correlação não implica necessariamente em causalidade (Bollen, 1989). Além de associadas, as variáveis devem estar isoladas de quaisquer outras influências e o fluxo de ação deve estar bem estabelecido. Assim, considerando duas variáveis x e y , diz-se que x causa y quando: existe correlação estatística entre as duas variáveis; não existe influência de nenhuma outra variável; quando x varia, y também varia, mas o contrário não é válido. Quando mal especificado, um modelo pode apresentar endogeneidade, situação que pode ser definida como a existência de correlação entre uma variável explicativa e o erro (Louviere *et al.*, 2005). O controle de variáveis endógenas não precisa ser necessariamente físico, podendo ser estatístico (Shiple, 2016). Na área de transportes, o controle de variáveis geralmente depende do controle estatístico, já que muitas vezes o controle físico ou não é possível ou não é ético. Westreich e Greenland (2013) ainda alertam que, mesmo quando os modelos são corretamente especificados, ainda podem ocorrer problemas quanto à interpretação dos coeficientes estimados, levando a conclusões incorretas sobre o fenômeno.

Visando à compreensão das principais limitações dos modelos baseados em inferência estatística, o objetivo deste trabalho é fazer uma análise crítica desse tipo de modelagem na etapa de diagnóstico dos efeitos de restrições no uso do solo e na oferta de transportes sobre a problemática das desigualdades socioespaciais na acessibilidade, utilizando dados de Fortaleza para ilustrar esses casos. Na Seção 2 é realizada uma revisão conceitual acerca da estrutura causal da problemática analisada, incluindo uma análise do fenômeno na cidade de Fortaleza, trazendo contexto à aplicação. A Seção 3 discute na teoria como os resultados das estimativas de inferência estatística podem ser mal interpretados quando não há uma base para se apoiar as relações de causa e efeito. Também é feita uma análise sobre a formulação dos modelos de diagnóstico das desigualdades na acessibilidade, apontando-se como os efeitos das restrições sobre a acessibilidade foram hipotetizados. A Seção 4 apresenta o método de análise crítica dos modelos estatísticos, que faz uma aplicação de dados coletados na cidade de Fortaleza para identificar como a modelagem apresentada nos trabalhos analisados pode gerar estimativas enviesadas ou interpretadas de forma incorreta. Para isso, é utilizado um diagrama causal que estabelece as associações entre a acessibilidade e suas restrições, apresentado na Seção 2. Os resultados são apresentados e discutidos na Seção 5. Na Seção 6 são apresentadas as contribuições do trabalho, bem como suas limitações e sugestões para trabalhos futuros.

2. CAUSALIDADE NO DIAGNÓSTICO DOS PROBLEMAS DE DESIGUALDADES SOCIOESPACIAIS NA ACESSIBILIDADE

A construção da estrutura causal que explicita as relações de causa-efeito de um determinado fenômeno requer sua compreensão conceitual e empírica *a priori*, considerando o contexto local no qual está inserido. Portanto, nesta seção será realizada uma revisão conceitual dos determinantes da problemática das desigualdades socioespaciais na acessibilidade e de como ela evoluiu no contexto da cidade de Fortaleza, obtendo-se ao final um diagrama causal que explicita a estrutura dessas relações.

2.1. Determinantes das desigualdades socioespaciais na acessibilidade

O engajamento de pessoas em atividades é o principal motivo de se deslocar. Por isso, facilitar os deslocamentos, especialmente pelos modos mais sustentáveis, deve ser o foco do planejamento urbano, fazendo da acessibilidade o principal bem a ser distribuído pelos projetos de transportes (Martens *et al.*, 2012). Reconhecendo a acessibilidade como o resultado da interação entre os subsistemas urbanos (Soares, 2022), esses devem formar o objeto de análise, ou seja, devem ser a fonte de investigação e coleta de dados. Assim, de acordo com Garcia *et al.* (2018), os problemas na acessibilidade são causados devido às distribuições inadequadas do uso do solo e às ineficiências nos transportes, além das características dos indivíduos ou grupos sociais, como ilustrado na Figura 1. Compreender como as desigualdades na acessibilidade são formadas requer então entender quais são os fatores (restrições) dos subsistemas que possuem influência na acessibilidade e quais são as formas (ou problemas) que esses fatores podem assumir que geram as desigualdades. Por exemplo, um problema comum em grandes cidades brasileiras é o das distâncias entre residências de baixa renda aos centros de emprego, condição chamada de *spatial mismatch*, fator que agrava as condições de acessibilidade da população mais vulnerável. Já as ineficiências no subsistema de transportes podem ser atribuídas a má distribuição de infraestrutura do transporte público e da malha viária sobre o território.



Figura 1: Causas dos problemas de acessibilidade. Fonte: Adaptado de Garcia *et al.* (2018)

2.2. Evolução da problemática em Fortaleza

O caso de Fortaleza ilustra bem essa problemática (Sousa, 2019). Como é comum ocorrer em grandes cidades latino-americanas, as oportunidades de empregos se concentram em sua região central. Observa-se um fenômeno de espraiamento urbano, com predominância de residências nas periferias. Esse processo é acompanhado pela segregação socioespacial, uma vez que os centros são ocupados pela população de alta renda, enquanto os domicílios de baixa renda são empurrados para áreas distantes das oportunidades, caracterizando o problema de *spatial mismatch* (Pinto *et al.*, 2023). A população mais vulnerável passou a ocupar com mais

intensidade a região Oeste, enquanto os grupos de maior renda migraram para o Leste, devido aos investimentos em serviços e infraestrutura nessa região, que atraiu grande número de empregos. Assim, mesmo se afastando dos centros, processo denominado autosegregação, a classe de alta renda ainda permaneceu mais próxima das oportunidades e infraestruturas (Lima *et al.*, 2021).

É reconhecido que o subsistema de transportes, através da rede de transporte público e da malha viária, é um indutor de transformações no subsistema de uso do solo (Lima *et al.*, 2021). Em Fortaleza, o investimento no sistema de transportes das últimas décadas levou à construção de malha viária de alta capacidade no lado Leste, ampliando a conexão dessas regiões com o centro. Já o transporte público foi expandido na região Oeste, tanto através de linhas rodoviárias quanto ferroviárias, também realizando uma ligação centro-periferia (Siqueira *et al.*, 2022). Porém, a rede de ônibus possui uma configuração tronco-alimentadora, aumentando o tempo de viagem de quem vem das periferias, pois precisam passar pelos terminais físicos de integração. Assim, enquanto o setor Leste possui melhores condições de malha viária, possibilitando um tempo médio de acesso aos empregos de 20 minutos, no lado Oeste o mesmo tipo deslocamento dura média de 75 minutos, por ônibus. Essa melhor infraestrutura do setor Leste beneficia ainda mais novos investimentos, aumentando a geração de empregos nessas regiões (Lima *et al.*, 2021). Dessa forma, a desigualdade tende a aumentar, já que a população de baixa renda se vê segregada em locais com poucos empregos e baixa acessibilidade, enquanto a população de baixa renda reside nos locais com melhores condições de acesso e que estimula a criação de mais empregos.

De acordo com o contexto descrito, Siqueira *et al.* (2022) realizaram um diagnóstico do impacto das restrições nos subsistemas urbanos sobre as desigualdades na acessibilidade para a cidade de Fortaleza. Os autores elaboraram um diagrama causal que ilustra a estrutura complexa dessas relações, explicitando caminhos diretos e indiretos, bem como as possíveis fontes de endogeneidade. Um indicador de distâncias às centralidades foi utilizado para representar o efeito de *spatial mismatch* e periferação das zonas residenciais, sendo causa direta e indireta da acessibilidade. Reconheceu-se que a condição socioeconômica, ao impactar as oportunidades de empregos e a acessibilidade, seria uma possível fonte de viés no contexto de Fortaleza, de modo que se realizou um controle amostral ao se estimar modelos distintos para cada grupo socioeconômico. Considerando a revisão conceitual e empírica desse fenômeno para Fortaleza, a Figura 2 apresenta uma adaptação do diagrama causal elaborado por Siqueira *et al.* (2022). A adequação foi realizada ao se notar que os investimentos em infraestrutura de transportes geram atração na quantidade de oportunidades disponíveis em uma região, o que não foi levando em conta no trabalho dos autores.

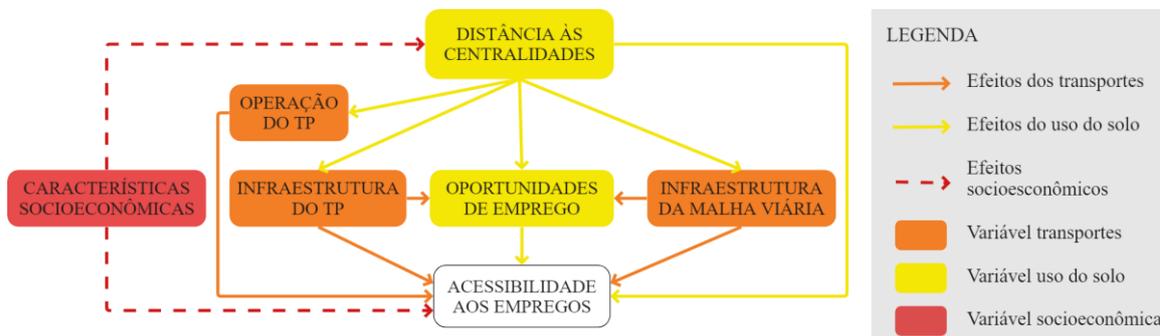


Figura 2: Diagrama causal dos efeitos de uso do solo e transportes na acessibilidade. Fonte: adaptado de Siqueira *et al.*, 2022

3. CRÍTICA TEÓRICA AOS MODELOS DE INFERÊNCIA ESTATÍSTICA

3.1. Interpretação causal dos efeitos estimados

Utilizar a inferência estatística para se obter interpretações causais dos coeficientes estimados, sem levar em consideração mecanismos para avaliar adequadamente o isolamento de variáveis e direção das relações analisadas, pode levar a uma má interpretação dos efeitos obtidos (Westreich e Greenland, 2013). O caminho que se forma entre duas variáveis de interesse é chamado de *front-door*, podendo ser direto e/ou indireto, passando por uma variável mediadora (Pearl *et al.*, 2016). Pode-se chamar de efeito total o resultado da medida do conjunto com todos os caminhos causais entre duas variáveis de interesse, e de efeito direto àquele obtido após se controlar todos os caminhos indiretos (Westreich e Greenland, 2013).

Para exemplificar essas associações, pode-se supor como verdadeiro o diagrama (a) da Figura 3, onde tempo de espera nas paradas (*tem*) e número de empregos (*emp*) em uma área possuem efeito sobre acessibilidade (*aces*). A variável empregos possui um efeito sobre tempo ao se levantar a hipótese de que locais com mais empregos possuem uma maior frequência de linhas de ônibus, diminuindo o tempo de espera. O caminho $emp \rightarrow tem \rightarrow aces$ é um efeito indireto da variável empregos sobre acessibilidade, mediado por tempo. Enquanto o caminho $tem \rightarrow aces$ gera o efeito direto e o total de tempo sobre acessibilidade, o caminho $emp \rightarrow aces$ gera apenas o efeito direto de empregos sobre acessibilidade.

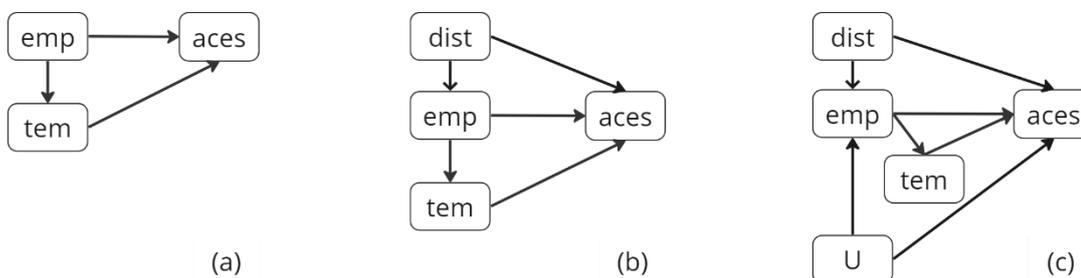


Figura 3: Exemplos de possíveis fontes de endogeneidade em um diagrama causal

Ou seja, ao se estimar uma regressão do tipo:

$$aces = \beta_0 + \beta_1 tem + \beta_2 emp \tag{1}$$

os coeficientes não podem ser entendidos da mesma forma, pois enquanto β_1 expressa o efeito total de *tem*, β_2 apresenta apenas um efeito direto, possuindo ainda um efeito indireto que não

está sendo considerado na equação. Para estimar esse efeito, uma equação onde conste apenas a variável *emp* deve ser estimada, podendo então esse coeficiente ser comparado com o β_1 . Considerando que o modelo (a) da Figura 3 está corretamente especificado, os efeitos diretos das variáveis de interesse podem ser obtidos sem nenhum viés. Porém, ao considerar um cenário em que a distância (*dist*) até os centros comerciais de uma cidade possui influência sobre a acessibilidade e a quantidade de empregos, como foi ilustrado pelo diagrama (b) na Figura 3, cria-se uma situação de endogeneidade. Conforme proposto originalmente por Pearl (2000), o caminho entre a variável endógena e as variáveis de interesse se chama de *back-door*. Quando não controladas, variáveis endógenas causam viés nas estimativas, resultando em conclusões não confiáveis. Nesse novo cenário, ao se regressir a Equação 1, o efeito direto de *tem* sobre *aces* não seria afetado, pois o efeito de *dist* sobre *tem* seria bloqueado ao se incluir na equação (controlar) a variável *emp*. Já o valor do efeito direto de *emp* sobre *aces* sofreria um viés, pois existe um caminho entre essas duas variáveis que não foi controlado. O efeito de *emp* sobre *aces* só seria corretamente estimado ao se incorporar a variável *dist* na equação.

Pearl (2000) define ainda um *collider* como uma outra possível forma de confundimento. Um *collider* ocorre quando uma variável se encontra em um caminho não direcionado (não segue a direção apontada pelas setas) entre duas outras variáveis. Por exemplo, no diagrama (c) da Figura 3, a variável *U*, além de causar endogeneidade nas estimativas de *emp*, também faz dessa variável um *collider* no caminho não direcionado $dist \rightarrow emp \leftarrow U$. Portanto, quando indevidamente controlado, o *collider* se torna uma variável de confundimento. Considerando agora que o diagrama (c) foi corretamente modelado, mas apenas as variáveis *tem*, *dist* e *emp* foram consideradas como explicativas, ocorre o seguinte: a estimativa de *tem* não é afetada, pois a influência que poderia sofrer por *U* está bloqueada ao se controlar *emp*; *U* causa viés nas estimativas de *emp*; a estimativa de *dist* sofre viés, pois o *collider* *emp* foi controlado, criando uma correlação com *U*. Para evitar o viés em *emp* e *dist*, é necessário que *U* seja também incluído como um determinante. Uma terceira fonte de viés pode ocorrer devido a uma variável moderadora, que é aquela que está em um caminho de *back-door* e modifica o efeito entre a variável dependente e as variáveis independentes. Em outras palavras, existem subgrupos na população de interesse que, apesar de apresentarem a mesma estrutura causal, possuem estimativas diferentes a depender do valor da variável moderadora (Pearl *et al.*, 2016).

3.2. Relações causais no diagnóstico das desigualdades socioespaciais na acessibilidade

Na literatura, além da modelagem baseada em inferência causal proposta por Siqueira *et al.* (2022), foram encontrados esforços de modelagem baseada em inferência estatística, propostos por Garcia *et al.* (2018) e Sousa (2019) para avaliar hipóteses de relações causais no diagnóstico das desigualdades socioespaciais na acessibilidade. Todos os trabalhos formularam hipóteses de relação de causa e efeito entre a acessibilidade e suas restrições, sendo as restrições classificadas como derivadas do subsistema de uso do solo, do subsistema de transportes ou das características socioeconômicas dos indivíduos ou grupos populacionais.

Garcia *et al.* (2018) utilizaram modelos de regressão para estimar o impacto de características de uso do solo e de transportes sobre as desigualdades na acessibilidade ao trabalho na cidade de Lisboa. Foi realizado um controle amostral, ao se segmentar a amostra por modo de transporte - transporte público (TP) e veículo privado (VP). A hipótese de causalidade era que as características das vias impactavam diretamente a acessibilidade por veículo privado,

enquanto a acessibilidade por transporte público é impactada por suas características operacionais. As características do uso do solo foram consideradas como restrição para os dois modos. A Figura 4 ilustra a estrutura causal adotada.



Figura 4: Representação das hipóteses formuladas por Garcia *et al.* (2018)

Uma das limitações da modelagem realizada foi a adoção da premissa de que os indicadores de transporte público e malha viária seriam independentes entre si e de qualquer outra variável que pudesse afetar a acessibilidade que estaria contida no erro. Dessa forma, os efeitos diretos seriam também efeitos totais, possibilitando sua comparação. Ou seja, não foi considerada nenhum tipo de endogeneidade nas relações de interesse, nem caminhos indiretos entre as variáveis. Por se tratar de um fenômeno complexo, onde há associação entre os elementos, essa é uma premissa muito forte. Outra limitação é que as características dos indivíduos não foram controladas, o que pode ter ocasionado vieses nas estimativas obtidas.

Sousa (2019), ao diagnosticar as desigualdades socioespaciais na acessibilidade para a cidade de Fortaleza, superou essa limitação, ao segmentar a amostra utilizada para a modelagem estatística em grupos de alta e baixa renda. As inadequações do uso do solo são atribuídas ao *spatial mismatch* domicílios-empregos. Já as ineficiências do subsistema de transportes são creditadas à má distribuição da rede de transporte público e da infraestrutura viária sobre o espaço, afetando a população de baixa e alta renda, respectivamente, pois se assume que a população de baixa renda é majoritariamente cativa do transporte público. As hipóteses de causalidade são de que os indicadores dessas restrições são causas diretas da acessibilidade, sendo o subsistema de transportes separado em indicadores da infraestrutura viária para a alta renda e da rede de transporte público para baixa renda, como apresentado na Figura 5. Assim como na modelagem de Garcia *et al.* (2018), o autor não considerou as possíveis relações entre as restrições, não permitindo a identificação de relações indiretas ou endôgenas.

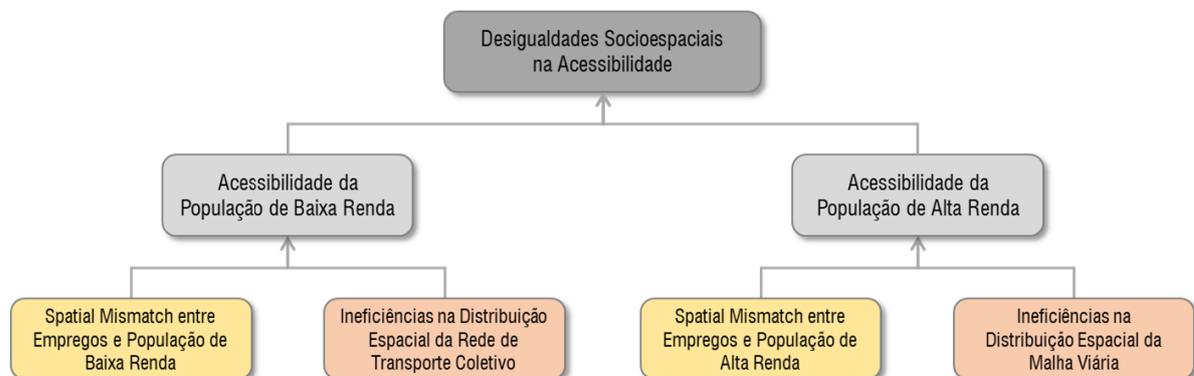


Figura 5: Representação da problemática das desigualdades socioespaciais na acessibilidade ao trabalho em Fortaleza. Fonte: Sousa, 2019

4. MÉTODO DE ANÁLISE CRÍTICA DOS MODELOS ESTATÍSTICOS

Para ilustrar como os modelos formulados por Garcia *et al.* (2018) e Sousa (2019) podem gerar vieses e má interpretação nos resultados ao se utilizar a inferência estatística, foram utilizados dados de Fortaleza para estimativas das relações especificadas na Figura 2 através de modelos de regressão linear. Não se pretende fazer uma análise fenomenológica sobre a cidade, mas apenas exemplificar com dados reais qual seria a forma de modelar as equações para impedir vieses e interpretar as estimativas de maneira adequada. Foram analisadas quatro situações, descritas a seguir.

4.1. Modelagem das equações

A Situação 1 buscou mostrar qual o viés causado por variáveis moderadoras. Através do diagrama causal da Figura 2, pode-se notar que as características socioeconômicas se constituem como variável moderadora, pois como justificado na Seção 2, as oportunidades que estão disponíveis para um indivíduo dependem de sua classe socioeconômica, levando heterogeneidade para os dados da população no geral. Foram então comparadas as estimativas de uma regressão sem segmentação da amostra por classe socioeconômica com as estimativas nas quais se segmentou a amostra em grupos de baixa, média e alta renda.

A Situação 2 buscou evidenciar a ocorrência de vieses gerados por variáveis endógenas. Para isso, utilizou-se como relação de interesse o efeito da densidade de empregos sobre a acessibilidade aos empregos da população de baixa renda. De acordo com o diagrama causal da Figura 2, a infraestrutura de transporte público e da malha viária são endógenas à essa relação, assim como a distância às centralidades, já que são causas tanto da acessibilidade quanto da densidade de empregos. Foram então estimados coeficientes de regressão para o indicador densidade de empregos para situações com diferentes variáveis de controle.

A Situação 3 representa como pode ocorrer má interpretação das relações causais de interesse. Na Situação 1, foi estimada uma equação para cada grupo socioeconômico, onde *acessibilidade aos empregos* foi a variável explicada e foram incluídos todos os indicadores de restrições como variáveis explicativas. Os efeitos estimados são totais para as variáveis *oportunidades de empregos* e *infraestrutura do TP*, mas apenas direto para o restante dos indicadores. Para se encontrar o efeito total do indicador *distâncias às centralidades*, regrediu-se uma equação apenas com essa variável, pois não existe nenhuma outra que seja endógena a essa relação, fora as características dos indivíduos, que já está sendo controlada. Para calcular os efeitos totais de *infraestrutura do TP* e *infraestrutura da malha viária*, foi controlada a variável *distância às centralidades*, que é endógena a essas relações.

A Situação 4 buscou estimar o viés gerado por controle indevido de *collider*. De acordo com o diagrama causal, a variável *oportunidades de emprego* é um *collider* em três caminhos não direcionais: entre *distância às centralidades* e *infraestrutura do TP*, entre *distância às centralidades* e *infraestrutura da malha viária* e entre *infraestrutura do TP* e *infraestrutura da malha viária*. Foi utilizado como exemplo um caso em que se quer saber apenas o efeito de *oportunidades de empregos*, *distância às centralidades* e *infraestrutura do TP* sobre *acessibilidade aos empregos*. Dessa forma, uma regressão onde consta apenas essas restrições como variáveis explicativas foi estimada para a classe de baixa renda.

4.2. Formulação dos indicadores

As variáveis utilizadas para a representação de cada indicador da Figura 2 foram espacializadas e agregadas em 241 zonas de tráfego. Os indicadores foram mensurados a partir de dados observacionais e modelados, descritos na Tabela 1. Os indicadores de acessibilidade e do subsistema de uso do solo escolhidos permitem a divisão em grupos socioeconômicos. O indicador de acessibilidade foi especificado como a quantidade de empregos que podem ser acessados por transporte público, a partir de um certo ponto e no horário de pico, em um tempo de viagem de até 30 minutos, disponibilizado no site do projeto Acesso a Oportunidades, do IPEA (Pereira *et al.*, 2022). Os empregos foram subdivididos entre baixa, média e alta escolaridade, sendo utilizados como aproximação para representar a população de baixa, média e alta renda, respectivamente. A acessibilidade foi calculada a partir de hexágonos de 0,11 km². Assim, para agregar os dados nas zonas de análise, utilizou-se como referência a quantidade de empregos correspondente ao terceiro quartil de empregos, na pretensão de considerar a maior quantidade de empregos que poderia ser acessada a partir das zonas, mas descartando-se possíveis outliers.

Tabela 1: Indicadores utilizados na modelagem

Variável	Indicador	Fonte
Acessibilidade	Quantidade de empregos que podem ser acessados dentro de 30 min.	IPEA
Densidade de empregos	Quantidade de empregos na zona dividido pela área da zona	
Distância às centralidades	Distância euclidiana média ponderada pelas oportunidades de empregos das centralidades	Sousa (2019)
Infraestrutura da rede viária	Densidade de vias arteriais e expressas (vias de alta capacidade)	Sousa (2019) SEUMA-For
Infraestrutura do TP	Soma da frequência das linhas de alta capacidade na hora-pico	ETUFOR
Operação do TP	Tempo médio de espera nas paradas	CBTU

Para representar os problemas do subsistema de uso do solo, foram mensurados os indicadores de densidade de empregos e distância às centralidades. As densidades foram obtidas através dos dados de empregos em cada zona, a partir de dados do IPEA (Pereira *et al.*, 2022), dividido pela área da zona. Para as distâncias, 49 zonas foram consideradas centrais (Sousa, 2019). Foi então calculado o somatório das distâncias entre cada zona e os centros, ponderada pelos empregos nas zonas centrais. Para a representação do subsistema de transportes, foram utilizados indicadores da infraestrutura da rede viária e medidas da infraestrutura e desempenho da rede de transporte público. Para a infraestrutura da malha viária foi considerada a densidade de vias de grande capacidade, como as arteriais e expressas. Já para o transporte público, foram utilizados indicadores do tempo de espera como medida da operação e a soma de frequência das linhas de alta capacidade, na hora-pico, para a infraestrutura.

5. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta estimativas dos efeitos das restrições sobre a acessibilidade para dados sem segmentação da população (Geral) e para os grupos de baixa (BR), média (MR) e alta (AR) renda, resultado da Situação 1. Pode-se notar que existem diferenças nas estimativas entre grupos que são ignoradas ao se analisar os dados sem considerar a heterogeneidade na população, especialmente para os indicadores de uso do solo. Ao se segmentar a população, é

possível entender quais as restrições mais relevantes para cada grupos, permitindo um melhor entendimento sobre a problemática.

Tabela 2: Efeitos diretos das restrições com e sem divisão por grupo socioeconômico

Indicador	Geral	BR	MR	AR
Distância aos centros	-0,68*	-0,12*	-0,45*	-0,12*
Densidade de empregos	0,25*	0,19*	0,25*	0,29*
Operação do TP	-0,21*	-0,05*	-0,14*	-0,03**
Infraestrutura do TP	0,04	0,01*	0,03	0,01
Infraestrutura da MV	0,33*	0,07*	0,24*	0,04

* $\alpha = 5\%$ de significância; ** $\alpha = 10\%$ de significância

A Tabela 3 apresenta os resultados para a Situação 2, que aponta como variáveis endógenas podem afetar as estimativas em uma relação de interesse. Ao se regredir uma equação com a *densidade de empregos* como a única variável explicativa, a estimativa é enviesada, pois existem três fontes de endogeneidade agindo sobre a relação. Pode-se notar que seu valor cai 42% ao se incluir a *distância às centralidades* como variável de controle. Isso ocorre devido ao bloqueio dos efeitos indiretos que essa variável causa na *densidade*. O valor continua caindo ao se incluir a *infraestrutura de transporte público e da malha viária*, mas não muda ao se incluir a variável de *operação de transporte público*, pois essa última não possui influência sobre a relação de interesse.

Tabela 3: Estimativas da densidade de empregos ao se incluir as variáveis de controle

Variáveis de controle	Dens_emp	R ²
Nenhuma	0,45*	0,52
Distância	0,26*	0,74
Distância, infraestrutura de TP	0,23*	0,75
Distância, infraestrutura de TP e MV	0,19*	0,76
Todas as variáveis	0,19*	0,77

* $\alpha = 5\%$ de significância; ** $\alpha = 10\%$ de significância

Os efeitos das variáveis de *distância às centralidades* e *infraestrutura do transporte público e da malha viária* apresentados na Tabela 2 são apenas diretos, e por isso não podem ser comparados com os demais, que são efeitos totais. Os resultados da Situação 3 estão na Tabela 4, que apresenta os efeitos totais para todas as variáveis. Ao contar com os efeitos indiretos, a intensidade dos indicadores sobre a variável explicada aumenta. O efeito total da *distância aos centros* é 73% maior do que seu efeito direto para média renda e 75% maior para a baixa e alta renda. O efeito total para *infraestrutura do transporte público* sobe quase duas vezes em relação ao seu efeito direto para a baixa renda, e passa a ser significativo para as classes de média e alta renda. A estimativa do efeito total do indicador da *infraestrutura da malha viária* para a baixa renda chega próximo ao efeito da *densidade de empregos*, e ultrapassa essa variável para a média renda. Assim, pode-se notar que apresentar apenas o efeito direto subestima o real efeito das variáveis explicativas sobre a explicada.

Tabela 4: Estimativa dos efeitos totais das restrições sobre a acessibilidade

Indicador	Baixa renda	Média renda	Alta renda
Distância aos centros	-0,21*	-0,78*	-0,21*
Densidade de empregos	0,19*	0,25*	0,29*
Operação do TP	-0,05*	-0,14*	-0,03**
Infraestrutura do TP	0,03*	0,11*	0,03*
Infraestrutura da MV	0,18*	0,65*	0,17*

* $\alpha = 5\%$ de significância; ** $\alpha = 10\%$ de significância

Os resultados referentes à Situação 4, em que ocorre o controle indevido de um *collider*, estão na Tabela 5. Ao se controlar a variável *densidade de empregos*, mas não a *infraestrutura da malha viária*, se “abre” um caminho entre essa última variável e *infraestrutura de transporte público*, criando um viés. Os indicadores do exemplo possuem baixo valor, e por isso apresentaram também apenas uma pequena divergência.

Tabela 5: Estimativas de infraestrutura de transporte público com e sem viés devido ao *collider*

Variáveis de controle	Infraestrutura de transporte público
Distância, densidades	0,02*
Todas as variáveis	0,01*

* $\alpha = 5\%$ de significância; ** $\alpha = 10\%$ de significância

6. CONCLUSÕES E RECOMENDAÇÕES

O ferramental de regressão linear clássica, quando utilizado para investigar relações de causalidade, pode chegar a estimativas enviesadas caso não seja apoiado por um diagrama causal, justificado por uma base conceitual bem definida (Pearl, 2000). Este trabalho buscou apresentar uma crítica metodológica aos modelos que utilizam inferência estatística para estabelecer as causas das desigualdades na acessibilidade, ilustrando essa problemática através de dados para Fortaleza. A análise foi apoiada por um diagrama causal proposto, que mostra as relações diretas e indiretas entre os indicadores.

A condição socioeconômica foi reconhecida como uma variável moderadora, pois os empregos disponíveis para um indivíduo variam de acordo com sua condição socioeconômica, gerando um caráter heterogêneo aos indicadores do uso do solo e acessibilidade. Não considerar essa variável no modelo, como foi o caso em Garcia *et al.* (2018), gera estimativas enviesadas, já que os efeitos das restrições sobre a acessibilidade são diferentes de acordo com o grupo socioeconômico. O controle indevido de *colliders* foi outra fonte de viés, que também existe na modelagem de Sousa (2019). De acordo com o diagrama, o indicador de *densidade de empregos*, é um *collider* entre a *infraestrutura de transporte público* e a *malha viária*. Assim, ao não incluir o indicador de infraestrutura da malha viária no modelo para diagnosticar a acessibilidade por transporte público, existe um viés devido ao controle do *collider*, que no caso ilustrado não se mostrou expressivo.

No caso exemplificado neste trabalho, o problema do *spatial mismatch* gera influência sobre os indicadores de *infraestrutura de transporte público* e *malha viária* e das *densidades de empregos*. Ao não representar esse elemento na modelagem, como ocorre em Garcia *et al.*

(2018), surge um viés. No exemplo apresentado, a estimativa para a *densidade de empregos* sobe 42% ao não incluir na modelagem o indicador de *distância às centralidades*, representante do *spacial mismatch*. Mesmo após modelar corretamente os caminhos de causalidade, ainda existe o desafio em se compreender as estimativas. Os efeitos obtidos ao se considerar todas as variáveis explicativas na regressão não são equivalentes. Existem indicadores que possuem caminhos diretos e indiretos até a variável explicada, como é o caso das *distâncias aos centros e infraestrutura de transporte público e da malha viária* da aplicação. O efeito total dessas variáveis sobre a *acessibilidade* deve ser obtido ao se estimar uma equação onde não constem as variáveis que estão nos caminhos indiretos, mas incluindo as devidas variáveis endógenas.

O diagrama causal se mostrou essencial para se compreender os caminhos diretos e indiretos entre as restrições e a acessibilidade, evidenciando as relações de endogeneidade. Analisar relações de causalidade em fenômenos complexos como o de desigualdades na acessibilidade sem esse tipo de recurso, através unicamente do ferramental de inferência estatística, leva ao risco de se obter resultados com diversas fontes de viés. Deve-se admitir que dificilmente seria possível elaborar um diagrama sem erros, havendo ainda alguns efeitos endógenos sobre as relações de interesse, mesmo que menos relevantes. Contudo, o objetivo do diagnóstico não é de prever resultados, mas sim explicar as relações causais. Na teoria da inferência causal, são geralmente utilizados modelos de equações estruturais (SEM) para se encontrar as relações de causalidade. Apesar de suas vantagens em relação à regressão linear, como a estimação de equações diretas e indiretas e a incorporação de variáveis latentes, o SEM ainda apresenta risco de endogeneidade se o modelo for mal especificado. Por fim, cabe destacar que este trabalho possui limitações relativas à investigação da endogeneidade causada devido a autocorrelação espacial que existe em fenômenos com estrutura espacial, como é o caso da acessibilidade e suas restrições quando agregadas ao nível de zonas, deixando-se como recomendação que futuros trabalhos possam investigar seus efeitos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Banister, D. (2008) The Sustainable Mobility Paradigm. *Transport Policy*, v. 15, n. 2, p. 73-80.
- Bertolini L.; F. le Clercq e L. Kapoen (2005) Sustainable accessibility: a conceptual framework to integrate transport and land use plan-making. Two Test-Applications in The Netherlands and a Reflection on the Way Forward. *Transport Policy*, v. 12, p. 207-220.
- Bollen, K. A. (1989) *Structural Equations with Latent Variables*. Ed. John Wiley & Sons (2° ed.).
- Garcia, C. S. H. F. (2016) *Strategic Assessment of Accessibility on Urban Mobility Networks*. Tese (Doutorado em Engenharia de Transportes) – Instituto Superior Técnico, Lisboa.
- Garcia, C. S. H. F.; M. R. M. R. Macário; E. D. A. G. Menezes e C. F. G. Loureiro (2018). Strategic Assessment of Lisbon's Accessibility and Mobility Problems from an Equity Perspective. *Networks and Spatial Economics*, v. 18, n. 2, p. 415-439.
- Lima, L. S.; C. F. G. Loureiro; F. F. L. M. Sousa e A. S. Lopes (2021) Espreadimento Urbano e seus Impactos nas Desigualdades Socioespaciais da Acessibilidade ao Trabalho em Fortaleza. *Transportes*, v. 29, n. 1, p. 229-246.
- Louviere, J.; K. Train; M. Ben-Akiva; C. Bhat; D. Brownstone; T. A. Cameron; R. T. Carson; J. R. Deshazo; D. Fiebig; W. Greene; D. Hensher e D. Waldman (2005) Recent Progress on Endogeneity in Choice Modeling. *Marketing Letters*, v. 16, n. 3, p. 255-265.
- Martens, K.; A. Golub; G. Robinson (2012) A justice-theoretic approach to the distribution of transportation benefits: Implications for transportation planning practice in the United States. *Transportation Research Part A: Policy and Practice*, v. 46, n. 4, p. 684-695.
- Meyer, M. D. (2016) *Transportation Planning Handbook*. Institute of Transportation Engineers. (4° ed). John Wiley & Sons, New Jersey.

- Ministério das Cidades (2006) *Gestão Integrada da Mobilidade Urbana*. Secretaria de Transporte e da Mobilidade Urbana. – Brasília, DF.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference* (1° ed.). Cambridge University Press, New York.
- Pearl, J.; Glymour, M.; Jewell, N. P. (2016) *Causal inference in statistics: a primer*. Ed. Wiley.
- Pereira, R. H. M. et al. (2022) Estimativas de acessibilidade a empregos e serviços públicos via transporte ativo, público e privado nas 20 maiores cidades do Brasil em 2017, 2018, 2019. Texto para Discussão. Ipea - Instituto de Pesquisa Econômica Aplicada.
- Pinto, D. G. L.; C. F. G. Loureiro; F. F. L. M. Sousa e B. Motte-Baumvol (2023) The effects of informality on socio-spatial inequalities in accessibility to job opportunities: Evidence from Fortaleza, Brazil. *Journal of Transport Geography*, v. 108.
- Shipley, B. (2016) *Cause and Correlation in Biology* (2° ed.). Cambridge University Press, New York.
- Siqueira, M. F.; D. G. L. Pinto; M. C. C. Belo; C. F. G. Loureiro e F. F. L. M. Sousa (2022) Análise dos Efeitos Causais das Restrições de Uso do Solo e Transportes sobre as Desigualdades Socioespaciais na Acessibilidade em Fortaleza. *Anais do XXXVI Congresso de Pesquisa e Ensino em Transportes, ANPET*, Fortaleza, v. 1.
- Soares, F. D. P. (2022) A Problem Oriented Approach to Urban Transportation Planning. University of Toronto. Department of Civil and Mineral Engineering.
- Sousa, F. F. L. M. (2019) *Diagnóstico Estratégico das Desigualdades Socioespaciais ao Trabalho em Fortaleza*. Tese (Doutorado em Engenharia de Transportes). Universidade Federal do Ceará, Ceará.
- Westreich, D. e S. Greenland (2013) The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology*. v. 177, n. 4. p. 292–298.

Maria Cristina Cavalcante Belo (mcristina.belo@det.ufc.br)
Davi Garcia Lopes Pinto (davi.garcia@det.ufc.br)
Carlos Felipe Grangeiro Loureiro (felipe@det.ufc.br)
Departamento de Engenharia de Transportes, Universidade Federal do Ceará
Campus do Pici, Bloco 703, CEP 60455-760 – Fortaleza, CE, Brasil