

# Crítica aos modelos de inferência estatística na análise dos determinantes das desigualdades na acessibilidade

## *Critique of statistical inference models to analyze the determinants of accessibility inequalities*

Maria Cristina Cavalcante Belo<sup>1</sup>, Davi Garcia Lopes Pinto<sup>1</sup>, Carlos Felipe Grangeiro Loureiro<sup>1</sup>

<sup>1</sup>Universidade Federal do Ceará, Fortaleza, Ceará, Brasil

Contato: criscb.maria@gmail.com,  (MCCB); davi.garcia@det.ufc.br,  (DGLP); felipe@det.ufc.br,  (CFGL)

### Recebido:

16 de novembro de 2023

### Revisado:

21 de maio de 2024

### Aceito para publicação:

29 de julho de 2024

### Publicado:

15 de outubro de 2024

### Editor de Área:

Cira Souza Pitombo, Universidade de São Paulo, Brasil

### Palavras-chave:

Desigualdades na acessibilidade.  
Modelos de inferência estatística.  
Análise de causalidade.

### Keywords:

Accessibility inequalities.  
Statistical inference models.  
Causal analysis.

DOI: 10.58922/transportes.v32i3.2972



### 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 vies 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), ao colocar a acessibilidade - que está ligada à ideia das atividades que podem ser acessadas e de como se dão as condições desse acesso - como objeto central de análise, reconhecendo a necessidade de integração entre os subsistemas de transportes e de uso do solo (Bertolini et al., 2005). Nesse novo paradigma de planejamento da

acessibilidade sustentável é possível analisar como as características desses subsistemas podem restringir o acesso às oportunidades de atividades de trabalho, educação, saúde, cultura, compras e lazer, possibilitando o planejamento integrado do sistema urbano que incentive padrões de mobilidade mais equitativos e 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 da problemática urbana deve ser considerada na primeira fase do nível estratégico do planejamento (Ministério das Cidades, 2006), também chamada de diagnóstico ou avaliação estratégica dos problemas, com foco nas desigualdades socioespaciais na acessibilidade, e guiada por princípios de equidade e sustentabilidade (Garcia, 2016). Uma das etapas da fase de diagnóstico estratégico consiste na análise agregada das relações de causa e efeito hipotetizadas acerca da problemática da acessibilidade e mobilidade urbana. 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, as limitações teóricas e práticas de se extrair interpretações causais a partir da modelagem estatística em transportes (Brathwaite e Walker, 2018), dadas as dificuldades de se incorporar em suas formulações toda a complexidade das relações envolvidas no fenômeno, como a representação de caminhos de *backdoors* que podem ocasionar endogeneidade (Pearl, 2000), além da dificuldade em se estimar efeitos indiretos das variáveis explicativas (Shiple, 2016). Considerando, portanto, que os objetivos estabelecidos no planejamento urbano integrado do uso do solo e dos transportes buscam resolver problemas de desigualdades socioespaciais na acessibilidade, a análise agregada no nível estratégico das diversas causas dessa problemática, antes de se propor soluções no nível tático, pode evitar o esforço de se implementar modificações no sistema de transportes que não sejam eficazes (Soares, 2022). Assim, o diagnóstico estratégico deve ser bem sistematizado para permitir uma adequada compreensão da problemática urbana.

De acordo com Bollen (1989), a formulação teórica de uma relação de causa-efeito possui três componentes: associação, isolamento e direção de influência. A associação entre duas variáveis não é suficiente para definir uma relação de causa-efeito, já que correlação não implica necessariamente em causalidade – embora o inverso seja sempre verdadeiro (Pearl et al., 2016). 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: i) existe correlação estatística entre as duas variáveis; ii) não existe influência de nenhuma outra variável; iii) quando  $x$  varia,  $y$  também varia, mas o contrário não é válido. Consequentemente, 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 (Louvriere et al., 2005). O controle de variáveis endógenas não precisa ser físico necessariamente, 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 (Brathwaite e Walker, 2018). Westreich e Greenland (2013) ainda alertam que, mesmo quando os modelos estatísticos são adequadamente 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 em análises de causalidade, o objetivo deste trabalho é, portanto, discutir criticamente a aplicação desse tipo de modelagem na etapa de diagnóstico estratégico dos efeitos das inadequações no uso do solo e das restrições da oferta de transportes sobre a problemática das desigualdades socioespaciais na acessibilidade, utilizando a contextualização urbana e os dados observacionais

de Fortaleza para ilustrar situações de possíveis fontes de endogeneidade. Para tanto, 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 encontrados na literatura científica, apontando-se como os respectivos efeitos das restrições sobre a acessibilidade foram hipotetizados. A Seção 4 apresenta uma proposta metodológica de análise crítica dos modelos estatísticos de diagnóstico estratégico selecionados, com base em uma aplicação de simulação com dados agregados em zonas no município de Fortaleza para investigar 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 de uso do solo e transportes, construído 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.

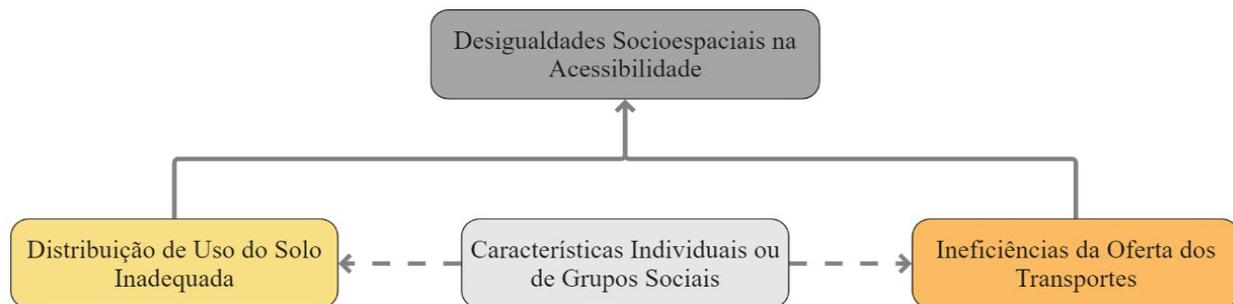
## **2. CAUSALIDADE NO DIAGNÓSTICO DOS PROBLEMAS DE DESIGUALDADES SOCIOESPACIAIS NA ACESSIBILIDADE**

Segundo a teoria de Inferência Causal de Pearl (2000), a construção de uma 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, propondo-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 para se deslocar. Por isso, facilitar os deslocamentos, especialmente pelos modos mais sustentáveis, deve ser o foco do planejamento urbano integrado, fazendo da acessibilidade o principal bem a ser distribuído pelas políticas e projetos de transportes (Martens et al., 2012). Reconhecendo a acessibilidade como o resultado da interação entre os subsistemas urbanos (Soares, 2022), esses devem se constituir como o objeto central de análise, ou seja, devem ser a fonte de investigação e coleta de dados. Assim, de acordo com Garcia et al. (2018), os problemas de desigualdades socioespaciais na acessibilidade são causados devido às distribuições inadequadas do uso do solo e às ineficiências no desempenho das redes de transportes, além das restrições impostas por características dos indivíduos ou grupos sociais, como ilustrado na Figura 1. Compreender como as desigualdades na acessibilidade são formadas requer, portanto, entender quais são os fatores (restrições) dos subsistemas urbanos que possuem influência nos níveis de acessibilidade de distintas regiões do território, e quais são as formas (ou estados) que esses fatores podem assumir que geram tais desigualdades. Por exemplo, um problema comum em grandes cidades brasileiras é o das elevadas distâncias e tempos de deslocamentos 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

(Pinto et al., 2023). 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 urbano.



**Figura 1.** Causas dos problemas de desigualdades socioespaciais na acessibilidade [adaptado de: Garcia et al., 2018].

## 2.2. Evolução da problemática das desigualdades na acessibilidade em Fortaleza

O caso de Fortaleza ilustra bem essa problemática urbana. Como é comum ocorrer em metrópoles brasileiras, as oportunidades de empregos se concentram em sua região Central (Figura 2), observando-se paralelamente um fenômeno de espraiamento urbano, com tendência de adensamento residencial nas periferias. Esse processo é acompanhado pela segregação socioespacial, uma vez que os centros são ocupados predominantemente pela população de alta renda, enquanto os domicílios de baixa renda são empurrados involuntariamente para áreas distantes das oportunidades, intensificando o problema de *spatial mismatch* domicílios-empregos (Andrade et al., 2020). Nas últimas décadas, a população mais vulnerável tem ocupado com mais intensidade a região Oeste da cidade, enquanto os grupos de maior renda têm migrado para o Leste, onde houve investimentos em serviços e infraestrutura, além do aumento na densidade de atividades e no número de empregos. Assim, mesmo se afastando dos centros, processo denominado autosegregação, segmentos de alta renda ainda permanecem mais próximos 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 (Lopes et al., 2020). Em Fortaleza, conforme representado no mapa da Figura 2, 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 sua maior centralidade (Pinto et al., 2023). Já o transporte público foi expandido no lado Oeste, tanto através de corredores de ônibus quanto de linhas metroferroviárias, ambos reforçando a ligação centro-periferia (Braga et al., 2023). Porém, a rede de ônibus possui uma configuração tronco-alimentadora, aumentando o tempo de viagem de quem vem das regiões periféricas, pois precisam passar por terminais físicos de integração. Dessa forma, enquanto as regiões Central, Nordeste e Sudeste possuem melhores condições de malha viária, possibilitando um tempo médio de acesso aos empregos de 20 minutos por automóvel, no lado Oeste esse acesso é em média 75 minutos por ônibus (Lima et al., 2021). Assim, a melhor infraestrutura viária do lado Leste de Fortaleza beneficia ainda mais novos investimentos, aumentando a geração de empregos nessas regiões (Pinto et al., 2023). 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 alta renda reside nos locais com melhores condições de acesso e que estimula a criação de mais empregos.

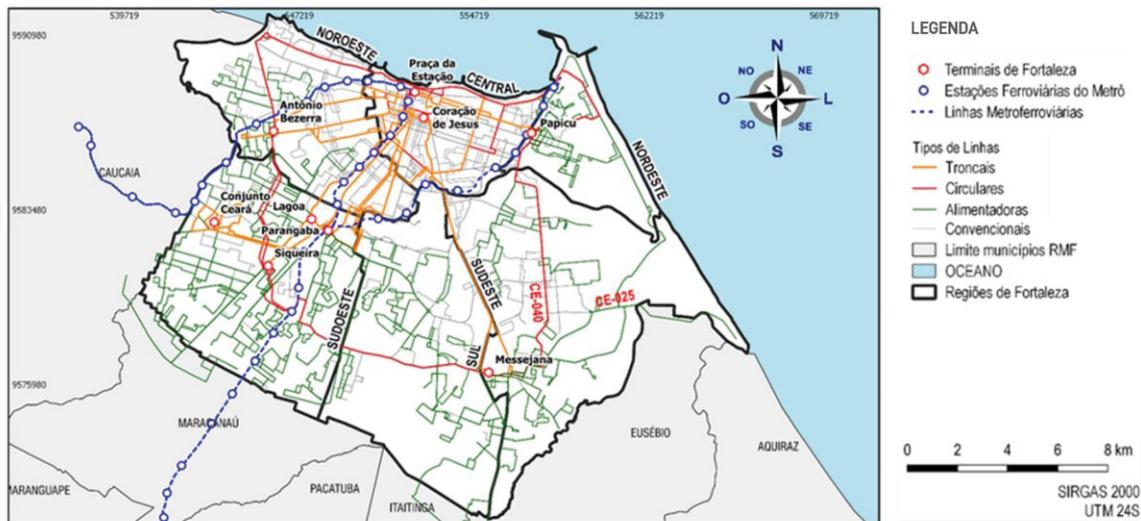


Figura 2. Divisão espacial do território de Fortaleza [adaptado de: Lima et al., 2021].

Com base no contexto resumido acima, Siqueira et al. (2022) realizaram uma análise de causalidade, agregada no nível de zonas, do impacto das restrições dos subsistemas de uso do solo e transportes, assim como das características socioeconômicas dos grupos populacionais, sobre as desigualdades socioespaciais na acessibilidade ao trabalho em Fortaleza. Para isso, os autores construíram um diagrama causal, justificado a partir de mapas mentais de hipóteses de causalidade, que ilustra a complexa estrutura das relações de causa-efeito, explicitando caminhos diretos e indiretos, bem como as possíveis fontes de endogeneidade. Partindo então da estrutura de causalidade proposta por Siqueira et al. (2022), foi elaborado o diagrama causal da Figura 3, que sistematiza as hipóteses das relações de causalidade entre acessibilidade e seus determinantes no contexto da cidade de Fortaleza, de modo a embasar o método de análise crítica da modelagem de inferência estatística proposto na Seção 4. O diagrama causal apresentado na Figura 3, com indicadores formulados nos trabalhos de Sousa (2019) e Siqueira et al. (2022), explicita as hipóteses de relações diretas e indiretas entre os componentes de uso e do solo e da oferta de transportes sobre a acessibilidade aos empregos, bem como caminhos de *backdoors*, que possam vir a ser fontes de endogeneidade nas estimativas dos efeitos causais.

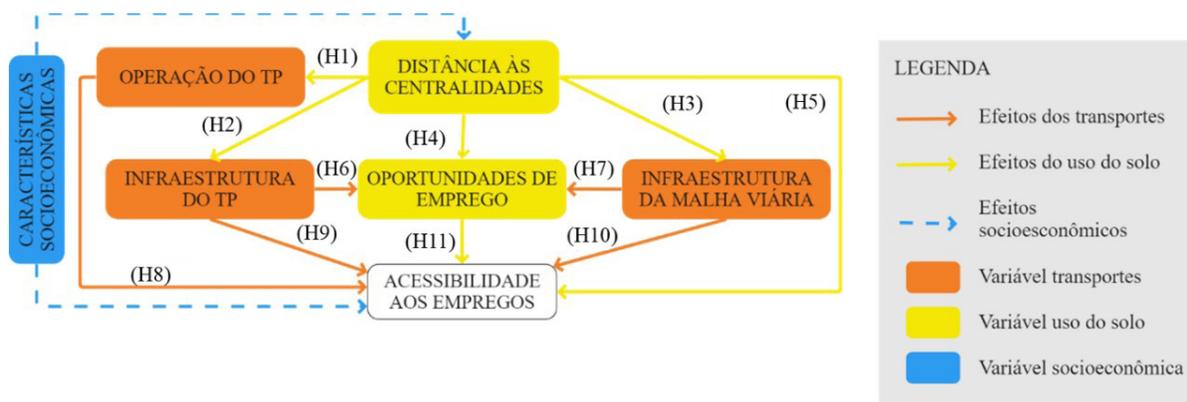


Figura 3. Diagrama causal dos efeitos de uso do solo e transportes na acessibilidade [adaptado de: Siqueira et al., 2022].

Inicialmente, com base nas evidências empíricas de que houve um viés espacial no provimento de infraestrutura de transportes em Fortaleza nas últimas décadas, com maiores investimentos em transporte público no lado Oeste e aumento de capacidade viária nas regiões Central e Leste, reconhece-se a Distância às Centralidades, indicador utilizado para representar o efeito do *spatial mismatch* sem considerar as redes de transportes, medido por meio de distâncias euclidianas (Sousa, 2019), como causa direta dos indicadores: i) Operação do TP (H1); ii) Infraestrutura do TP (H2); e iii) Infraestrutura da Malha Viária (H3). Similarmente, a partir das evidências de que há uma forte concentração de oportunidades de empregos na região Central, foi estabelecida a relação direta entre Distância às Centralidades e Oportunidades de Emprego (H4), já que quanto mais distante do Centro menor é a densidade de empregos.

Por sua vez, reconhecendo que melhores condições de infraestrutura de transportes acarretam numa valorização das áreas atendidas, atraindo os setores de comércio e serviços e conseqüentemente gerando novas oportunidades de empregos, foi estabelecida uma relação direta entre Infraestrutura do TP e da Malha Viária com Oportunidades de Emprego (H6 e H7). Como causas diretas da Acessibilidade aos Empregos, mensurada por meio de indicador baseado em logsum proposto por Sousa (2019), foram estabelecidas as hipóteses de que esta é resultante dos atributos das redes de transportes (H8, H9 e H10) e das Oportunidades de Emprego (H11), que assim como Distância às Centralidades causa impacto direto na Acessibilidade aos Empregos (H5).

Por fim, no contexto de metrópoles como Fortaleza, há uma clara distinção na distribuição espacial de domicílios de diferentes grupos socioeconômicos, de modo que grupos de baixa renda residem mais distantes das centralidades do que grupos de alta renda. Além disso, a variável renda domiciliar condiciona o acesso aos modos de transportes e a própria percepção dos níveis de acessibilidade. Dessa forma, ao impactar simultaneamente Distâncias às Centralidades e Acessibilidade aos Empregos (seta azul tracejada na Figura 3), a variável de Condição Socioeconômica representa uma fonte de viés nas estimativas, de modo que se propõe um controle amostral ao se estimar modelos distintos para cada grupo socioeconômico.

### 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 modelados (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, 2000). 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 (Pearl et al., 2016).

Para exemplificar essas associações, considere o exemplo fictício da Figura 4. O diagrama causal (a), onde tempo de espera nas paradas (*tem*) e número de empregos (*emp*) em uma área possuem efeito sobre acessibilidade (*aces*), estabelece que a variável *aces* sofre efeito direto das variáveis *emp* e *tem*. Além disso, há um efeito indireto de *emp* sobre *aces*, que é mediado por *tem*, através do caminho  $emp \rightarrow tem \rightarrow aces$ . No diagrama (a), o efeito total de *tem* sobre *aces* é equivalente ao efeito direto  $tem \rightarrow aces$ . Já para se contabilizar o efeito total de *emp* sobre *aces* deve-se considerar o efeito direto  $emp \rightarrow aces$  somado ao efeito indireto no caminho  $emp \rightarrow tem \rightarrow aces$ .

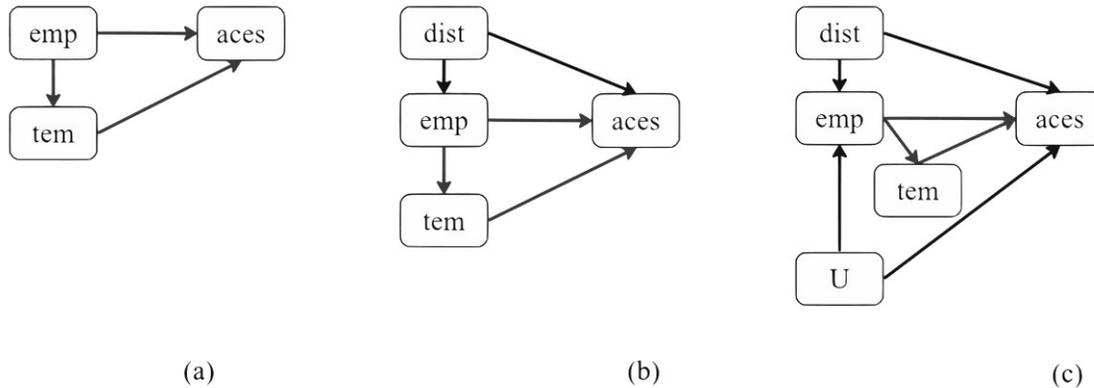


Figura 4. Exemplos de possíveis fontes de endogeneidade entre variáveis.

No exemplo acima, ao se estimar um modelo de regressão com a seguinte formulação:

$$aces = \beta_0 + \beta_1 tem + \beta_2 emp \quad (1)$$

interpretam-se os coeficientes  $\beta_1$  e  $\beta_2$  como os efeitos totais de *tem* e *emp*, respectivamente, sobre *aces*. Essa interpretação, apesar de correta para a variável *tem*, é equivocada para a variável *emp*, pois nessa formulação o coeficiente  $\beta_2$  está incorporando apenas o efeito de direto de *emp* sobre *aces*, ignorando a parcela do efeito indireto pelo caminho  $emp \rightarrow tem \rightarrow aces$ .

Por sua vez, ao se considerar o cenário do diagrama causal (b) da Figura 4, em que foi incluída uma nova variável *dist* como causa simultânea de *emp* e *aces*, 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 *backdoor*. Quando não controladas, variáveis endógenas causam viés nas estimativas, resultando em conclusões não confiáveis. Nesse novo cenário, ao se regressar novamente a Equação 1, sem controlar a variável *dist*, o coeficiente  $\beta_2$  estaria incorporando viés do caminho de *backdoor*  $emp \rightarrow dist \rightarrow aces$ . Já o efeito direto de *tem* sobre *aces*, representado por  $\beta_1$ , não seria afetado, pois o efeito de *dist* sobre *tem* seria bloqueado ao se incluir na equação (controlar) a variável *emp*. Dessa forma, efeito direto de *emp* sobre *aces* só seria corretamente estimado ao se incorporar a variável *dist* na equação, uma vez que o caminho de *backdoor* estaria sendo “fechado”.

Pearl (2000) define ainda um *collider* como uma outra possível forma de confundimento da relação causal. 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 4, 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 *backdoor* 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. Modelos estatísticos de diagnóstico da acessibilidade selecionados para crítica

Com o propósito de se utilizar o diagrama causal da Figura 3 como ferramenta de análise crítica das hipóteses de causalidade, buscou-se na literatura acadêmica por trabalhos recentes com formulação de modelos de inferência estatística para diagnóstico, no nível estratégico, das determinantes das desigualdades socioespaciais na acessibilidade, calibrados a partir de dados agregados sobre o território urbano. Foram encontrados, além da modelagem baseada em inferência causal proposta por Siqueira et al. (2022), os trabalhos desenvolvidos por Garcia et al. (2018) e Sousa (2019), tendo como estudo de caso, respectivamente, as cidades de Lisboa e Fortaleza. Nesses dois esforços de modelagem, foram formuladas hipóteses de relações de causa-efeito entre a acessibilidade e suas restrições determinantes, sendo essas restrições classificadas como derivadas do subsistema de uso do solo, do subsistema de transportes ou das características socioeconômicas grupos populacionais. Vale destacar que a busca bibliográfica realizada não tinha como objetivo fazer uma revisão sistemática da literatura sobre modelos de inferência causal e estatística das determinantes da acessibilidade; seu propósito era unicamente selecionar alguns esforços de modelagem estatística, baseados em estruturas de causalidade semelhantes à da Figura 3, para ilustrar possíveis fontes de endogeneidade em análises de diagnóstico estratégico das desigualdades socioespaciais na acessibilidade urbana.

Garcia et al. (2018) utilizaram modelos clássicos e espaciais globais de regressão linear múltipla, 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 os dados por modo de transporte - transporte público (TP) e veículo privado (VP). A hipótese de causalidade era que as características físicas das vias impactavam diretamente a acessibilidade na malha viária; enquanto a acessibilidade na rede de TP era impactada por suas características operacionais. Já as características do uso do solo foram consideradas como restrição para os dois modos. A Figura 5 ilustra a estrutura causal adotada.

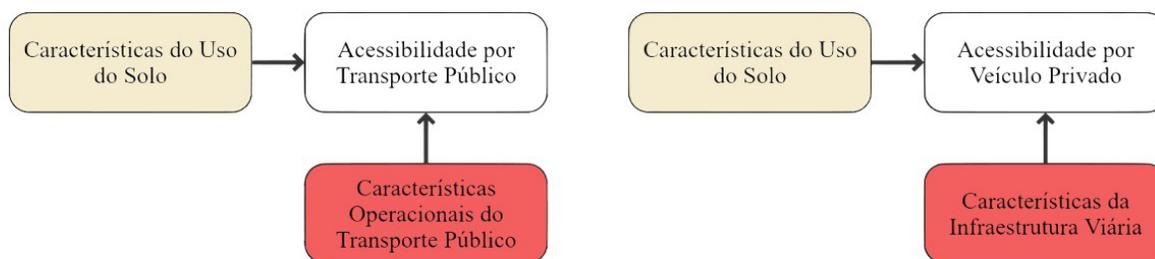
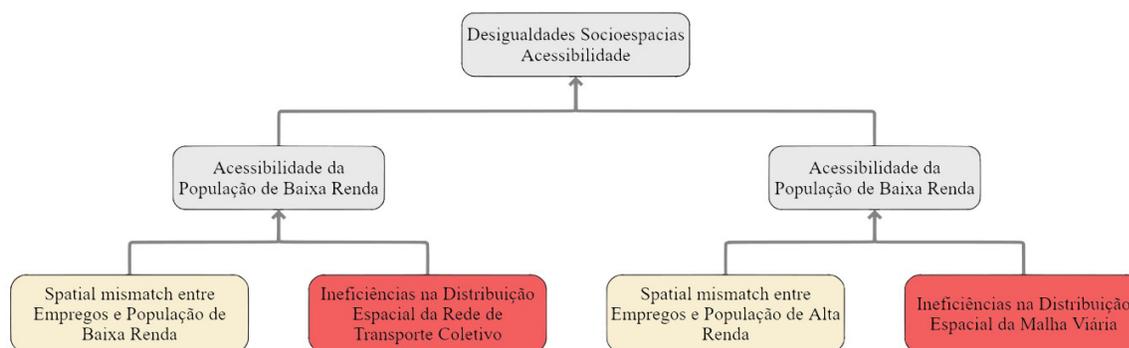


Figura 5. Representação das hipóteses formuladas por Garcia et al. (2018).

Uma das limitações da modelagem realizada por Garcia et al. (2018) 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, impossibilitando 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.

Por sua vez, Sousa (2019), ao diagnosticar no nível estratégico as desigualdades socioespaciais na acessibilidade ao trabalho 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 foram atribuídas ao *spatial mismatch* domicílios-empregos. Já as ineficiências do subsistema de transportes foram 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; assumindo-se que a população de baixa renda é majoritariamente cativa do transporte público. As hipóteses de causalidade levantadas foram 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 6. Assim como na modelagem de Garcia et al. (2018), o autor não considerou possíveis interrelações entre as restrições, não permitindo a identificação de relações indiretas ou endógenas.



**Figura 6.** Representação da problemática das desigualdades socioespaciais na acessibilidade ao trabalho em Fortaleza [adaptado de: 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 exclusivamente o ferramental de inferência estatística, foram utilizados dados de Fortaleza para estimativas das relações especificadas no diagrama causal da Figura 3 através de modelos de regressão linear. Não se pretende aqui 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 3, 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 3, a *infraestrutura de transporte público* e a *malha viária* são endógenas a 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 *distância às centralidades*, regrediu-se uma equação apenas com essa variável, pois não existe nenhuma outra que seja endógena a essa relação, fora a variável moderadora *características socioeconômicas*, 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.

Por fim, na Situação 4 buscou-se estimar o viés gerado por controle indevido de *collider*. De acordo com o diagrama causal da Figura 3, 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 3 foram espacializadas e agregadas nas 241 zonas de análise de tráfego de Fortaleza (Lima et al., 2021). 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 da manhã, 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 km<sup>2</sup>. 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, dividida 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 bases do IPEA (Pereira et al., 2022), divididos 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 malha 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 identificar quais as restrições mais relevantes para cada grupo, 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 *distância às centralidades* como variável de controle. Isso ocorre devido ao bloqueio dos efeitos indiretos que essa variável causa na *densidade de empregos*. 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 <sup>2</sup>
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.

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 às centralidades* 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* aumenta 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.

## 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 analisada, que também existe na modelagem de Sousa (2019). De acordo com o diagrama causal proposto, o indicador de *densidade de empregos* é um *collider* entre a *infraestrutura de transporte público* e *da 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 do transporte público* e *da 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. Portanto, 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 proposto 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 socioespaciais 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 mensurar efeitos 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 no nível de zonas, deixando-se como recomendação que futuros trabalhos possam investigar tais efeitos.

## AGRADECIMENTOS

Os autores agradecem à FUNCAP e CAPES pela concessão de bolsas de mestrado e doutorado, assim como ao CNPq pela concessão de bolsa de produtividade em pesquisa, fundamentais ao desenvolvimento deste trabalho.

## REFERÊNCIAS

- Andrade, B.R.; C.F.G. Loureiro; F.F.L.M. Sousa et al. (2020) Efeitos da periferização nos níveis de acessibilidade aos empregos da população de baixa renda em Fortaleza. *Transportes*, v. 28, n. 3, p. 46-60. DOI: 10.14295/transportes.v28i3.1810.
- Banister, D. (2008) The sustainable mobility paradigm. *Transport Policy*, v. 15, n. 2, p. 73-80. DOI: 10.1016/j.tranpol.2007.10.005.
- Bertolini, L.; F. le Clercq and 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, n. 3, p. 207-220. DOI: 10.1016/j.tranpol.2005.01.006.
- Bollen, K.A. (1989) *Structural Equations with Latent Variables* (2nd ed.). Chichester: John Wiley & Sons. DOI: 10.1002/9781118619179.
- Braga, C.K.V.; C.F.G. Loureiro and R.H.M. Pereira (2023) Evaluating the impact of public transport travel time inaccuracy and variability on socio-spatial inequalities in accessibility. *Journal of Transport Geography*, v. 109, p. 103590. DOI: 10.1016/j.jtrangeo.2023.103590.
- Brathwaite, T. and J.L. Walker (2018) Causal inference in travel demand modeling (and the lack thereof). *Journal of Choice Modelling*, v. 26, p. 1-18. DOI: 10.1016/j.jocm.2017.12.001.
- Garcia, C.S.H.F. (2016) *Strategic Assessment of Accessibility on Urban Mobility Networks*. Tese (doutorado). Instituto Superior Técnico, Lisboa. Disponível em: <<https://scholar.tecnico.ulisboa.pt/records/55Obv83c9xtc1BGqqVoDYtxLUuWhSJo2WMKz>> (acesso em 12/10/2023).
- Garcia, C.S.H.F.; R. Macário; E.D.A.G. Menezes et al. (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. DOI: 10.1007/s11067-018-9391-4.
- Lima, L.S.; C.F.G. Loureiro; F.F.L.M. Sousa et al. (2021) Espriamento urbano e seus impactos nas desigualdades socioespaciais da acessibilidade ao trabalho em Fortaleza. *Transportes*, v. 29, n. 1, p. 229-246. DOI: 10.14295/transportes.v29i1.2348.
- Lopes, A.S.; C.B. Cavalcante; D.S. Vale et al. (2020) Convergence of planning practices towards LUT integration: seeking evidences in a developing country. *Land Use Policy*, v. 99, p. 104842. DOI: 10.1016/j.landusepol.2020.104842.
- Louviere, J.; K. Train; M. Ben-Akiva et al. (2005) Recent progress on endogeneity in choice modeling. *Marketing Letters*, v. 16, n. 3, p. 255-265. DOI: 10.1007/s11002-005-5890-4.
- Martens, K.; A. Golub and 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. DOI: 10.1016/j.tra.2012.01.004.
- Meyer, M.D. (2016) *Transportation Planning Handbook* (4th ed.). New Jersey: John Wiley & Sons. DOI: 10.1002/9781119174660.
- Ministério das Cidades (2006) *Gestão Integrada da Mobilidade Urbana*. Brasília: Secretaria de Transporte e da Mobilidade Urbana.
- Pearl, J. (2000) *Causality: Models, Reasoning, and Inference* (1st ed.). New York: Cambridge University Press.
- Pearl, J.; M. Glymour and N.P. Jewell (2016) *Causal Inference in Statistics: a Primer*. Chichester: John Wiley & Sons.
- Pereira, R.H.M.; C.K.V. Braga; D. Herszenhut 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). Brasília: Ipea.
- Pinto, D.G.; C.F.G. Loureiro; F.F.L.M. Sousa et al. (2023) The effects of informality on socio-spatial inequalities in accessibility to job opportunities: evidence from Fortaleza, Brazil. *Journal of Transport Geography*, v. 108, p. 103577. DOI: 10.1016/j.jtrangeo.2023.103577.
- Shiple, B. (2016) *Cause and Correlation in Biology* (2nd ed.). New York: Cambridge University Press. DOI: 10.1017/CB09781139979573.
- Siqueira, M.F.; D.G.L. Pinto; M.C.C. Belo et al. (2022) Análise dos efeitos causais das restrições de uso do solo e transportes sobre as desigualdades socioespaciais na acessibilidade em Fortaleza. In *Anais do XXXVI Congresso de Pesquisa e Ensino em Transportes*. Fortaleza: ANPET.

- Soares, F.D.P. (2022) *A Problem Oriented Approach to Urban Transportation Planning*. Toronto: Department of Civil and Mineral Engineering, University of Toronto. Disponível em: <<https://tspace.library.utoronto.ca/handle/1807/123590>> (acesso em 22/09/2023).
- Sousa, F.F.L.M. (2019) *Diagnóstico Estratégico das Desigualdades Socioespaciais ao Trabalho em Fortaleza*. Tese (doutorado). Universidade Federal do Ceará, Fortaleza. Disponível em: <<https://repositorio.ufc.br/handle/riufc/53751>> (acesso em 12/07/2023).
- Westreich, D. and 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. DOI: 10.1093/aje/kws412. PMID:23371353.