

ESTUDO DA MORTALIDADE INFANTIL NOS ESTADOS BRASILEIROS UTILIZANDO APLICAÇÃO DE DADOS EM PAINEL

Recebido em: 25/09/2023

Aceito em: 25/10/2023

DOI: 10.25110/arqsaude.v27i10.2023-036

Maizza Micaelle Carlos Euclides Fernandes ¹

João Agnaldo do Nascimento ²

Rodrigo Pinheiro de Toledo Vianna ³

RESUMO: Este trabalho tem como objetivo determinar uma relação linear entre a Taxa de Mortalidade Infantil (TMI) e um conjunto de variáveis socioeconômicas observadas por unidades federativas no período de 2005 à 2010 utilizando o modelo de dados em painel de efeitos fixo e aleatório. Metodologia: trata-se de um estudo descritivo com abordagem quantitativa, com utilização dos Sistema de Informação sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (SINASC) e em seguida utilizou-se o software R para realizar esta análise de dados com a função plm. Resultados: os estudos mostram que o modelo mais adequado é o de efeito fixo com transformação logarítmica nas variáveis independentes e na variável dependente que foram as seguintes: TMI, taxa de analfabetismo, PIB per capita, proporção pessoas com baixa renda, percentual da população servida por rede de abastecimento de água e a proporção da população servida por coleta de lixo. Conclusão: As variáveis independentes que causam impacto significativo na TMI são taxa de analfabetismo, PIB per capita e proporção de pessoas com baixa renda.

PALAVRAS-CHAVE: Taxa de Mortalidade Infantil; Dados em Painel; Modelo de Efeito Fixo e Aleatório.

STUDY OF CHILD MORTALITY IN BRAZILIAN STATES USING PANEL DATA APPLICATION

ABSTRACT: This work aims to determine a linear relationship between the Infant Mortality Rate (IMR) and a set of socioeconomic variables observed by federative units in the period from 2005 to 2010 using the fixed and random effects panel data model. Methodology: this is a descriptive study with a quantitative approach, using the Mortality Information System (SIM) and the Live Birth Information System (SINASC) and then using the R software to perform this data analysis with the plm function. Results: studies show that the most appropriate model is the fixed effect model with logarithmic transformation in the independent variables and the dependent variable, which were as follows: IMR, illiteracy rate, GDP per capita, proportion of people with low income, percentage of the population served by water supply network and the proportion of the population served by garbage collection. Conclusion: The independent variables that have

¹ Mestre em Modelos de Decisão e Saúde. Universidade Federal da Paraíba.

E-mail: maizzamicaelle@hotmail.com ORCID: <https://orcid.org/0000-0003-2076-1256>

² Doutor em Estatística. Universidade Federal da Paraíba.

E-mail: joao.agnaldo@academico.ufpb.br ORCID: <https://orcid.org/0000-0002-7314-4844>

³ Doutor em Saúde Coletiva. Universidade Federal da Paraíba.

E-mail: vianna@ccs.ufpb.br ORCID: <https://orcid.org/0000-0002-5358-1967>

a significant impact on IMR are the illiteracy rate, GDP per capita and the proportion of people with low income.

KEYWORDS: Infant Mortality Rate; Panel Data; Fixed and Random Effect Model.

ESTUDIO DE MORTALIDAD INFANTIL EN ESTADOS BRASILEÑOS MEDIANTE APLICACIÓN DE DATOS DE PANEL

RESUMEN: Este trabajo tiene como objetivo determinar una relación lineal entre la Tasa de Mortalidad Infantil (TMI) y un conjunto de variables socioeconómicas observadas por las unidades federativas en el período 2005 a 2010 utilizando el modelo de datos de panel de efectos fijos y aleatorios. Metodología: se trata de un estudio descriptivo con enfoque cuantitativo, utilizando el Sistema de Información de Mortalidad (SIM) y el Sistema de Información de Nacidos Vivos (SINASC) y luego utilizando el software R para realizar este análisis de datos con la función plm. Resultados: los estudios muestran que el modelo más adecuado es el modelo de efectos fijos con transformación logarítmica en las variables independientes y la variable dependiente, las cuales fueron las siguientes: TMI, tasa de analfabetismo, PIB per cápita, proporción de personas con bajos ingresos, porcentaje de la población atendida por red de suministro de agua y la proporción de la población atendida por recolección de basura. Conclusión: Las variables independientes que tienen un impacto significativo en la TMI son la tasa de analfabetismo, el PIB per cápita y la proporción de personas con bajos ingresos.

PALABRAS CLAVE: Tasa de Mortalidad Infantil; Datos de Panel; Modelo de Efectos Fijos y Aleatorios.

1. INTRODUÇÃO

A Mortalidade Infantil é um tema complexo de ser avaliado em sua completude, podendo ser qualificado através de indicadores em determinadas regiões conhecidas de acordo com elementos primordiais para o perfil do cenário de satisfação social de uma determinada população. Dessa forma, dentre os indicadores o mais clássico é a taxa de mortalidade infantil (TMI), que por sua vez, caracteriza não apenas as condições de saúde de uma sociedade, como também o nível socioeconômico da população que tem sido empregado como padrão sintetizador de desenvolvimento constante para caracterizar as condições de vida de uma população (BRASIL, 2021).

No entanto, a inconsistência das estatísticas de Registro Civil (RC) para os nascimentos quanto para os óbitos, tem colocado grandes restrições para o cálculo da TMI no Brasil (HILL; JUDGE; GRIFFITHS, 1999). Assim sendo, fazendo uso das fontes de dados: os censos demográficos, portanto, ao se fazer uso das fontes de dados seja estas dos censos demográficos ou da Pesquisa Nacional por Amostra de Domicílios (PNAD), torna-se necessário métodos de estimação da TMI, por esta razão há fragilidades na divulgação

destas informações que podem se distanciar da nossa realidade. Neste trabalho, não houve correções para a TMI e utilizou-se a base de dados do Datasus.

A Taxa de Mortalidade Infantil (TMI) é um indicador essencial para verificar a qualidade dos serviços de saúde e de saneamento básico oferecidos a uma população. Sua importância pode ser expressa através de seu uso constante em estudos das mais diversas áreas, na avaliação da cobertura e da qualidade dos serviços de saúde (COSTA; BORGES, 2022). Ao decorrer dos anos, a diminuição dos óbitos inferiores a um ano de idade, caracteriza uma das importantes metas na área de saúde em inúmeros países, que tem como consequência a redução da mortalidade infantil (DUARTE, 2007). As variações na TMI serão avaliadas pelo Estado e ao longo do tempo pelo modelo Dados em Painel (DP). É importante utilizar modelos econométricos que mostrem a influência de variações que representem sua contribuição e influência na TMI, uma vez que a análise destas influências poderão decidir estratégias de gestão de saúde para melhorar as condições de vida de uma população.

Modelos econométricos podem descrever uma variável dependente como função de variáveis independentes representadas por informações associadas às condições socioeconômicas e de saúde pública. Podem medir de forma indireta o avanço obtido com investimentos públicos realizados ou oferecer um direcionamento para estabelecer quais metas estariam mais ligadas ao plano de diminuir o TMI (BRASIL, 2021).

Os países mais desenvolvidos diminuíram a TMI a um ponto onde o mesmo permanece estável em um valor pequeno quando comparado com os valores da TMI em países do terceiro mundo. O limite da redução esbarra na importância do conhecimento de como tratar a mortalidade infantil decorrente de doenças (ALVES; COELHO, 2021). Então por muitos anos a TMI refletirá o desenvolvimento de um povo, o produto de sua luta e o combate e a manutenção das condições de vida.

Sabe-se que utilizam números para avaliar a qualidade de vida de determinada população, a começar por um dado bastante sensível; da mortalidade infantil na região em que estar inserido. Sendo assim, a questão que norteia o trabalho é: Quão relevantes e assertivas estão as variáveis para a Taxa de Mortalidade Infantil no Brasil?

O objetivo deste artigo é descrever uma relação linear entre a variável dependente (TMI) e as variáveis independentes por grupo, cujas observações ocorreram ao longo de alguns anos, caracterizando um estudo longitudinal. Além de descrever o modelo de

Dados em Painel com efeito fixo e aleatório através de testes estatísticos aplicados no modelo Pooled.

2. MÉTODOS

As fontes oficiais de referência com relação à mortalidade infantil são dadas pelo Ministério da Saúde (MS), por intermédio do Sistema de Informação sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (SINASC). A Fundação Instituto Brasileiro de Geografia e Estatística (Fundação IBGE) fornece dados Estatísticas do Registro Civil, com disponibilidade anual para todos os estados e municípios brasileiros (IBGE, 2022).

Os dados relacionados à mortalidade infantil e os indicadores socioeconômico e de cobertura usado para este trabalho foram extraídas do SIM e do IBGE/ Pesquisa Nacional por Amostra de Domicílios (PNAD) para os anos de 2005 e 2010, por Unidade de Federação (UF), acessados através do site do Departamento de Informática do SUS (DATASUS). O SIM foi criado pelo DATASUS para captação dos dados de mortalidade no país e é a única fonte que divulga os resultados por causa básica de morte no Brasil, sendo este um dos principais motivos que justifica seu uso (BRASIL, 2021).

Dos indicadores e dados básicos para avaliar a mortalidade infantil nas UFs, selecionou-se 6 (seis) variáveis pertencentes aos indicadores: mortalidade, socioeconômicos e de coberturas, sendo um (1) pertencente ao indicador de mortalidade, sendo ele a Taxa de mortalidade infantil (TMI), 3 (três) pertencentes aos Indicadores Socioeconômico, sendo eles: Taxa de analfabetismo; produto interno bruto (PIB) per capita; proporção de pessoas com baixa renda e 2 (dois) pertencentes aos Indicadores de Cobertura: proporção da população servida por rede de abastecimento de água e a proporção da população servida por coleta de lixo, para os anos 2005 à 2010, cuja descrição está no tabela (1) abaixo, expressa as variáveis que serão utilizadas no modelo, definido a seguir:

Tabela 1: Descrição das Bases de Dados do DataSUS.

Variáveis	Descrição	Fonte de Dados
TMI	Taxa de Mortalidade Infantil até um ano de vida	Datasus
Analfabetismo	Taxa de analfabetismo para pessoas com 15 anos ou mais de idade	Datasus
PIB per capita	Produto Interno Bruto	Datasus
Baixa renda	Proporção de pessoas com baixa renda	Datasus
Abastecimento de água	Proporção da população servida por rede de abastecimento de água	Datasus
Coleta de lixo	Proporção da população servida por coleta de lixo	Datasus

Fonte: Elaborado pelos autores (2018)

Dados em painel (DP) são medidas repetidas em diferentes períodos de tempo. Nesses dados deve-se considerar o efeito de grupo e do tempo que são analisados como efeito fixo ou aleatório. Para isto, é necessário que se tenha dois tipos de variáveis: uma referente ao objeto, por exemplo, instituição, estado ou grupo, e outra referente ao tempo, podendo ser mês, trimestre, semestre ou ano. Os modelos de DP são classificados como efeito fixo em que as variáveis *dummy* são consideradas parte do intercepto ou aleatório em que as variáveis *dummy* são consideradas como erros aleatórios (FERNANDES, 2016). Então para se decidir qual tipo de modelo deve-se utilizar, devemos aplicar o teste de Housman (1978), que decide sobre o efeito a ser aceito. A hipótese nula desse teste é que os efeitos individuais não estão correlacionados com os demais regressores do modelo, ou seja, o modelo de efeito aleatório deve ser acatado, caso contrário ao rejeitamos a hipótese nula utilizamos o modelo, com efeito, fixo (CHASE, 1969). Portanto, deve-se atingir uma conclusão que consiga representar uma relação linear múltipla entre uma variável dependente e um conjunto de variáveis independentes, permitindo avaliar essa situação utilizando o método de dados em painel.

2.1 Modelos Dados em Painel

O modelo geral para dados em painel é descrito da forma seguinte:

$$y_{it} = \beta_{0it} + \beta_{1it}x_{1it} + \dots + \beta_{kit}x_{kit} + e_{it}$$

Neste caso, o subscrito *i* expressa os diferentes indivíduos e o subscrito *t* indica o período de tempo que está sendo estudado. β_0 relaciona-se ao parâmetro de intercepto e β_k ao coeficiente angular equivalente à *k*-ésima variável explicativa do modelo (CARVALHO; GÓES, 2018).

$$y_i = [y_{i1} \ y_{i2} \ \dots \ y_{iT}] \quad X_i = [x_{1i1} \ x_{2i1} \ \dots \ x_{ki1} \ \vdots \ \vdots \ x_{1iT} \ x_{2iT} \ \dots \ x_{kiT}] \quad \beta_i = [\beta_{0i1} \ \beta_{1i1} \ \beta_{2i1} \ \dots \ \beta_{ki1} \ \vdots \ \vdots \ \beta_{0iT} \ \beta_{1iT} \ \beta_{2iT} \ \dots \ \beta_{kiT}] \quad e_i = [e_{i1} \ e_{i2} \ \dots \ e_{iT}]$$

Em que,

y_i e e_i são vetores de dimensão (T x 1) e contém respectivamente, as T variáveis dependentes e os T erros. X_i é uma matriz de dimensão (K x T) com as variáveis explicativas do modelo. Desta forma, o elemento x_{kit} refere-se à *k*-ésima variável explicativa para o indivíduo *i* no instante de tempo *t*. Por fim, β_i é a matriz dos parâmetros a serem estimados (CARVALHO; GÓES, 2018).

Neste modelo geral, o intercepto e os parâmetros de resposta são distintos para cada indivíduo e para cada período de tempo. Encontrando assim, mais parâmetros desconhecidos do que observações, não sendo possível, neste caso, estimar os seus parâmetros. Assim, é necessário identificar suposições em relação ao modelo geral, a fim de transformá-lo em operacional. Dentre os modelos que condiz com dados de séries temporais e dados em corte transversal, dois deles serão utilizados, Modelo de Efeitos Fixos e Modelos de Efeitos Aleatórios.

2.2 Modelos de Efeito Fixo

Em conformidade com Hill, Griffiths e Judge (1999), as suposições do modelo são (GRIFFITHS; HILL; JUDGE, 1993).

$$\beta_{0it} = \beta_{0i} \quad \text{e} \quad \beta_{1it} = \beta_1 \dots \beta_{kit} = \beta_k$$

O modelo de efeitos fixos, no entanto, é dado da seguinte forma:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + e_{it}$$

A forma matricial para o i -ésimo indivíduo, como sugerido por Griffiths, Hill e Judge (1993), é dado por (GRIFFITHS, HILL; JUDGE, 1993):

$$\begin{bmatrix} y_{i1} & y_{i2} & \dots & y_{iT} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \alpha_i + \begin{bmatrix} x_{1i1} & x_{2i1} & \dots & x_{ki1} & \vdots & \vdots & x_{1iT} & x_{2iT} & \dots & x_{kiT} \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_k \end{bmatrix} + \begin{bmatrix} e_{i1} & e_{i2} & \dots & e_{iT} \end{bmatrix}$$

Assim, neste modelo o α_i refere-se aos interceptos a serem estimados, um para cada indivíduo. Assim como os parâmetros de respostas não mudam entre os indivíduos e nem ao longo do tempo, todas as diferenças de procedimentos entre os indivíduos devem ser percebido pelo intercepto. Dessa maneira, α_i pode ser explicado como o efeito das variáveis omissas no modelo.

Outro pressuposto relevante do modelo efeito fixo é que o parâmetro é fixo e desconhecido que compreende as particularidades dos indivíduos da amostra. Logo, as inferências feitas com relação ao modelo são apenas a respeito dos indivíduos dos quais se utiliza de dados (CARVALHO, GÓES, 2018).

Pode-se realizar uma classificação do modelo de efeitos fixos empregando variáveis binárias para mostrar os interceptos próprios para cada indivíduo. Assim sendo, a equação geral será expressa por:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \dots + \gamma_n D_{ni} + e_{it}$$

Segundo Carvalho e Góes (2018) o dado que D_{ni} consiste em uma variável binária dummy para cada indivíduo e é um quando $i = n$ e quando não, será zero. Contudo, a equação expõe uma variável binária para cada indivíduo (n variáveis binárias), tornando-se no problema econométrico de multicolinearidade excelente. Com o intuito de retirar a multicolinearidade é necessário omitir uma variável binária. Dessa forma, o modelo apresentado por Stock e Watson (2004) que será determinada como:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \dots + \gamma_n D_{ni} + e_{it}$$

Neste sentido,

$\beta_0, \beta_1, \dots, \beta_k, \gamma_2, \dots, \gamma_n$ são os parâmetros que serão estimados. Quando $i=1$, o intercepto é dado por $\beta_0 = \alpha_i$. Para $i \geq 2$, o intercepto é dado por $\beta_0 + \gamma_i$. Através do OLS é estimado o modelo de efeitos fixos. As estimativas alcançadas são não tendenciosas e consistentes, por sua vez que o modelo supõe que os erros têm distribuição normal, variância constante e não são correlacionados.

A estimativa do modelo com variáveis binárias produz os mesmos resultados da estimação de efeitos fixos (WOOLDRIDGE, 2002). No entanto, o emprego de variáveis dummy não é prático. Por isso, mesmo assim com total baixo de indivíduos no modelo, cresce muito o número de parâmetros a serem estimados. Em alguns pacotes econométricos quando se tem uma quantidade alta de indivíduos, essa regressão pode ser improvável de ser executada (WOOLDRIDGE, 2002).

Para verificar se as suposições sobre o modelo estão adequadas é satisfatório realizar um teste para investigar se os interceptos são diferentes entre os indivíduos. De acordo com Griffiths, Hill e Judge (1993), as hipóteses nulas e alternativa são as seguintes:

$$H_0: \beta_{01} = \beta_{02} = \dots = \beta_{0k}$$

H_1 : os interceptos β_{0i} não são todos iguais

As hipóteses são testadas através da estatística F. Quando rejeita-se H_0 , conclui-se que os interceptos não são todos iguais, atendendo a suposição do modelo de n interceptos diferentes.

O estimador do intercepto do modelo, $\hat{\alpha}_i$, além de não viesado, é inconsistente quando a quantidade de indivíduos é muito grande ($N \rightarrow \infty$). Isso ocorre porque cada observação em corte seccional adicionada aumenta um novo coeficiente α_i . Dessa forma, os estimadores de α_i serão melhores quando o número de períodos de tempo observado for grande.

Quando o intercepto α_i é correlacionado com as variáveis explicativas em qualquer período de tempo, o modelo de efeitos fixos é mais adequado para modelar dados em painel (WOOLDRIDGE, 2002). Além disso, como o intercepto do modelo é abordado como um parâmetro fixo, além disso, utiliza efeitos fixos quando as observações são adquiridas de toda população e o que se pretende realizar são inferências para os indivíduos dos quais propõe os dados.

2.3 Modelos de Efeito Aleatório

O modelo de efeito aleatório, as variáveis têm as mesmas suposições do modelo de efeito fixo, ou seja, o intercepto muda de um indivíduo para outro, porém não ao decorrer do tempo, e os parâmetros resposta são constantes para todos os indivíduos e perante todas as etapas de tempo. O que muda entre o modelo de efeito fixo e o modelo de efeito aleatório é o tratamento do intercepto. O modelo de efeitos aleatórios analisa os interceptos como variáveis aleatórias, assim, o modelo entende que os indivíduos possuem dados que são amostras aleatórias de indivíduos com uma população maior (CARVALHO; GÓES, 2018). Assim como proposto por Hill, Griffiths e Judge (1993), serão modelados os n interceptos da seguinte forma:

$$\beta_{0i} = \underline{\beta}_0 + \alpha_i \quad i = 1, \dots, n$$

Consegue-se observar que este intercepto é formado pelo intercepto do modelo de efeitos fixos α_i que verifica as diferenças do comportamento dos indivíduos e por um outro componente, $\underline{\beta}_0$, refere-se ao intercepto populacional. O modelo geral de efeitos aleatórios é dado da seguinte forma:

$$y_{it} = \underline{\beta}_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + v_{it}$$

Logo,

$v_{it} = e_{it} + \alpha_i$ caracteriza o erro.

A forma matricial desse modelo, para i -ésimo indivíduo, será dado por:

$$\begin{bmatrix} y_{i1} & y_{i2} & \dots & y_{iT} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \underline{\beta}_0 + \begin{bmatrix} x_{1i1} & x_{2i1} & \dots & x_{ki1} & \dots & x_{1iT} & x_{2iT} & \dots & x_{kiT} \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_k \end{bmatrix} + \begin{bmatrix} e_{i1} & e_{i2} & \dots & e_{iT} \end{bmatrix}$$

Hil, Griffiths, Judge (1999) demonstra as quatro propriedades do termo estocástico

v_{it} :

- 1) $E(v_{it}) = 0$
- 2) $\text{var}(v_{it}) = \sigma_e^2 + \sigma_\alpha^2$
- 3) $\text{cov}(v_{it}, v_{is}) = \sigma_\alpha^2 \quad \forall t \neq s$
- 4) $\text{cov}(v_{it}, v_{jt}) = 0 \quad \forall i \neq j$

Desse modo, as duas primeiras propriedades significam que v_{it} tem média zero e variância constante, ou seja, o erro é homocedástico. Na terceira propriedade, tem-se que os erros do mesmo indivíduo em diferentes períodos de tempo são correlacionados, determinando uma autocorrelação. Por fim, a quarta propriedade expressa que os erros de diferentes indivíduos no mesmo instante de tempo não são correlacionados, assim, não ocorre correlação contemporânea.

Assim como ocorre correlação entre os indivíduos dos mesmos indivíduos em períodos de tempo diferentes, o método OLS não é o mais adequado para estimar os coeficientes do modelo de efeitos aleatórios. Assim sendo, o método que fornece os melhores estimadores é os mínimos quadrados generalizados (GLS).

Para testar se o modelo de efeitos aleatórios é apropriado, Breusch e Pagan (1980) apud Greene (1997) desenvolveram um teste baseado no multiplicador de Lagrange, determinando-se as seguintes hipóteses:

$$H_0 : \sigma_\alpha^2 = 0$$

$$H_1 : \sigma_\alpha^2 \neq 0$$

A estatística do teste é dada por:

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n [\sum_{t=1}^T e_{it}]^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2$$

Conforme a hipótese nula (H_0), LM tem distribuição Qui-Quadrado com um grau de liberdade. Caso a hipótese H_0 não rejeite, o modelo de efeitos fixos é preferível ao modelo de efeitos aleatórios (DUARTE; LAMOUNIER; TAKAMATSU, 2007). Quando não, é necessário atribuir que o modelo de efeitos aleatórios é mais adequado para os dados que estão sendo abordados.

Greene (1997) propõe o teste de Hausman e formula as hipóteses nula e alternativa para verificar se existe correlação entre α_i e as variáveis explicativas:

H_0 : α_i não é correlacionado com as variáveis explicativas

H_1 : α_i é correlacionado com as variáveis explicativas

Em Greene (1993) acha-se a estatística do teste. Quando a hipótese nula não rejeita, não há evidências que α_i seja correlacionado com as variáveis explicativas. Dessa forma, deve-se aplicar o modelo de efeitos aleatórios. No entanto, se rejeita a hipótese nula, logo se utiliza o modelo de efeitos fixos.

2.4 Modelo Pooled

A especificação do modelo *pooled cross-section* básica é:

$$Y_{it} = a + b_1 X_{it} + e_{it}$$

Confirme Y (taxa de mortalidade infantil) a variável dependente, X é o vetor de variáveis que estabelece o Y, o subscrito t indica o tempo (2005-2010) e o i representa as UFs. Se e_{it} for um ruído branco, pode-se estimar por POLS, ou seja, o estimador OLS ampliado para o caso de painel ((GRIFFITHS, HILL; JUDGE, 1993). Logo, neste modelo a estimação é feita atribuindo que os parâmetros a e b são comuns para todas as UFs.

Para decidir qual o modelo é o mais apropriado para indicar a redução da taxa de mortalidade infantil neste trabalho, serão aplicados dois testes: o teste de Breusch-Pagan, que é utilizado para determinar qual dos modelos é o mais adequado: *Pooled* ou o de Painel (GRIFFITHS, HILL; JUDGE, 1993). Pelo teste rejeita-se a hipótese nula de que a variância do coeficiente individual estimado é igual à zero; e o teste de Hausman com a finalidade de decidir entre os modelos de efeito aleatório e efeito fixo, que tem como hipótese nula a não existência de correlação com as variáveis explicativas.

3. RESULTADOS

O modelo de dados em painel é formado por uma série temporal 2005 a 2010 e pelas 27 UFs, sendo equilibrado no sentido de todas as unidades estão presentes em todos os anos. A tabela (2) apresenta os resultados das variáveis explicativas do modelo estimado por POLS, que não leva em conta a especificação do efeito individual omitido, tendo um total de 162 observações.

Tabela 2 Resultado da Regressão OLS Pooled (Teste de Chow)

Variáveis	Estimativa	Erro Padrão	t valor	P-valor
Intercepto	0,1532	3.0391	5.0419	$0,1265 \times 10^{-5}$ ***
X_1	-0,3126	0,0589	-5,2997	$0,3898 \times 10^{-6}$ ***
X_2	$-8,1289 \times 10^{-5}$	$3,3951 \times 10^{-5}$	-2,3943	0,0178 *
X_3	0,3132	0,0264	11,8650	$0,0022 \times 10^{-13}$ ***
X_4	-0,0651	0,1899	-3.4275	$0,0779 \times 10^{-2}$ ***
X_5	0,0016	0,0344	-0.0470	0,9625
R^2	0.7994		R^2 Ajustado	0.7698
Estatística F:	124,323		p-valor	$0,0222 \times 10^{-14}$

Fonte: Elaborado pelos autores (2018)

A tabela (3) apresenta os resultados da estimação do modelo de Efeito Fixo, tendo um total de 162 observações. Para o modelo de efeito fixo, as variáveis PIB e percentual da população servida por coleta de lixo não apresentaram evidências estatísticas significativas. Já a taxa de analfabetismo, percentual de pessoas com baixa renda e percentual da população servida por rede de abastecimento de água indicam evidências estatísticas para a redução da TMI com 5% de significância. Dessa forma, quando a taxa de analfabetismo diminui com 5%, então: a TMI se reduz com 25,552%, a proporção de pessoas com baixa renda se reduz em 34,242% e o abastecimento de água em 12,635%. Neste modelo, R^2 foi igual a 0,73796 (DUARTE; LAMOUNIER; TAKAMATSU, 2007)

Tabela 3: Resultado da Regressão do modelo de Efeito Fixo

Variáveis	Estimativa	Erro Padrão	t valor	P-valor
X_1	0,2555	$1,2000 \times 10^3$	$2,1306 \times 10^1$	0,0350 *
X_2	$4,7765 \times 10^{-5}$	$5,3000 \times 10^1$	0,9014	0.3691
X_3	0,3424	$4,2500 \times 10^2$	80,521	$4,5130 \times 10^{-13}$ ***
X_4	0,1264	$4,8500 \times 10^2$	26,057	$1,0240 \times 10^{-2}$
X_5	-0,0364	$6,0900 \times 10^2$	-0,5978	0,5511
R^2	0,73796		R^2 Ajustado	0,5922
Estatística F:	73.2209		p-valor	$0,0222 \times 10^4$

Fonte: Elaborado pelos autores (2018).

Para o modelo de efeito fixo, as variáveis PIB e % da população servida por coleta de lixo não apresentaram evidências estatísticas significativas. Já a taxa de analfabetismo,

% de pessoas com baixa renda e % da população servida por rede de abastecimento de água indicam evidências estatísticas para a redução da TMI com 5% de significância. Dessa maneira, quando a taxa de analfabetismo diminui com 5%, então: a TMI se reduz com 25,552%, a proporção de pessoas com baixa renda se reduz em 34,242% e o abastecimento de água em 12,635%. Neste modelo, R^2 foi igual a 0,73796. A tabela (4) apresenta a estimação para o modelo de efeitos aleatórios.

Tabela 4: Resultado da Regressão do modelo de Efeito Aleatório

Variáveis	Estimativa	Erro Padrão	t valor	P-valor
Intercepto	5,9149	4,9104	1,2046	0,2302
X_1	$-1,6521 \times 10^{-1}$	$8,2347 \times 10^{-2}$	-2,0063	0,0466 *
X_2	$2,2580 \times 10^{-5}$	$4,6133 \times 10^{-5}$	0,4895	0,6252
X_3	$3,6210 \times 10^{-1}$	$3,5006 \times 10^{-2}$	$1,0344 \times 10^{-1}$	2×10^{-16} ***
X_4	$-2,2103 \times 10^{-2}$	$3,2027 \times 10^{-2}$	-0,6901	0,4911
X_5	$7,1262 \times 10^{-3}$	$5,0725 \times 10^{-2}$	0,1405	0,8885
R^2	0,7076		R^2 Ajustado	0,6814
Estatística F:	75,518		p-valor	$0,0222 \times 10^{-14}$

Fonte: Elaborado pelos autores (2018).

Para o modelo com efeito aleatório, as variáveis PIB, abastecimento de água, coleta de lixo não foram significativas para o modelo. Logo, a variável de % de pessoas com baixa renda teve evidência estatística para redução da mortalidade infantil, com maior contribuição, em seguida a taxa de analfabetismo também apresenta contribuição importante para diminuir a mortalidade infantil (FERNANDES; et al, 2022). Este modelo apresenta um $R^2 = 0,7076$; tendo um modelo razoável para explicar os valores observados. Em todos os modelos estimados a variável taxa de analfabetismo e renda apresentou-se significância ao nível de 5%.

4. DISCUSSÃO

Os resultados estimados com o modelo OLS Pooled apresentados na tabela (5), mostra que as variáveis X_1 = taxa de analfabetismo, X_2 = PIB per capita, X_3 = % de pessoas com baixa renda e X_4 = % da população servida por rede de Abastecimento de água foram significativas para este modelo ao nível de 5%. Somente a variável X_5 = % proporção da população servida por coleta de lixo é que não foi significativa. A taxa de analfabetismo e proporção de pessoas de baixa renda são as que mais contribuíram para diminuir a taxa de mortalidade infantil nas UFs, pelos OLS pooled; uma vez que essas variáveis apresentaram coeficientes mais elevados. Pelo modelo, ao ser excluído o analfabetismo (e aumentando os serviços de educação básica) e baixa renda (tendo salários melhores), a

esperança de diminuir a taxa de mortalidade será mais satisfatória. Neste modelo o R^2 foi de 0.7994 e o R^2 ajustado foi de 0.7698, mostrando bom ajuste.

Salienta-se a existência de desvios de normalidade nas extremidades (caudas pesadas) de regressão OLS Pooled (Teste de Chow). Assim, para confirmar se há normalidade dos resíduos aplicou-se o teste Kolmogorov-Smirnov, com p-valor = 0.5797, logo, não temos evidências estatísticas para rejeitar a hipótese de normalidade dos dados, considera-se, portanto, distribuição normal para os resíduos do modelo OLS Pooled.

Ressalta-se ainda que, a partir das distorções nos resíduos de regressão do modelo de Efeito Fixo, para confirmar se há normalidade, aplicou-se o teste Kolmogorov-Smirnov, com p-valor = 0,04082, evidenciando-se estatisticamente formas para rejeitar a hipótese de normalidade dos resíduos deste modelo, portanto não há normalidade dos resíduos no modelo de efeito fixo.

Por fim, através das distorções nos resíduos de normalidade do efeito aleatório, com o fim de confirmar se há normalidade entre os dados aplicou-se o teste Kolmogorov-Smirnov, com p-valor igual a 0,0004, logo, temos evidências estatísticas para rejeitar a hipótese de normalidade dos dados, portanto não há normalidade dos erros no modelo aleatório.

Ainda, em conformidade com os testes executados, o teste de Breusch-Pagan, que serve para escolher entre o modelo de OLS pooled e modelo de dados em painel de efeito aleatório, teve a hipótese nula rejeitada, com 5% de significância, por consequência, o modelo de Dados em Painel de efeito aleatório é o mais apropriado. Em seguida, realizou-se o teste de Hausman que decide sobre o efeito fixo ou aleatório. A sua hipótese nula deste teste é que o efeito é aleatório, ou seja, não há correlação entre as variáveis explicativas e a componente erro do modelo de efeitos aleatório. Após a aplicação deste teste, e o teste de Breusch-Pagan cuja hipótese nula é que o modelo é o OLS pooled versus modelo de efeito aleatório. A tabela (5) mostra que o teste de Housman decide para o modelo de efeito fixo e o de Breusch-Pagan para o modelo de efeito aleatório. Portanto, decide-se pelo modelo de efeito fixo.

Assim sendo, a % de pessoas com baixa renda tem sido uma variável decisiva para diminuição da TMI, visto que uma proporção pequena de pessoas com baixa renda faz com que beneficie melhor a qualidade e expectativa de vida com mais assistência médica proporcionando uma educação de boa qualidade. Considerando que o modelo com as variáveis originais não apresentou resíduos com provável distribuição normal, admitiu-

se realizar uma transformação logarítmica ($Z_i = \ln x$ e $Z = \ln y$) nas variáveis dependentes e independentes do modelo, obtendo-se os resultados da tabela 5 onde se apresenta os efeitos individuais dos estados.

Tabela 5: Resultado do Teste de Efeito Individual dos Estados

	Estimativa	Erro Padrão	t valor	P-valor
Z_1	- 0,1127	0,0487	- 2,3161	0,0221 *
Z_2	-0,2024	0,051824	-3,9052	0,0002 ***
Z_3	0,5331	0,0641	8,3216	$1,0310 \times 10^{-13}$ ***
Z_4	-0,1259	0,0641	-1,9648	0,0516
Z_5	0,1844	0,1244	1,4823	0,1407
R^2	0,6481		R^2 Ajustado	0,5201
Estatística F	47,889		p-valor	$0,0222 \times 10^{-14}$

Fonte: Elaborado pelos autores (2018).

O resultado do teste F geral permite decidir que é provável a existência do efeito individual e as variáveis que mais contribuem para isso são: X_1 (taxa de analfabetismo) e % de pessoas com baixa renda são responsáveis. Portanto, fica comprovado a variação significativa da TMI por estados.

Com base na transformação feita para os resíduos do modelo de efeito fixo com logaritmo, sendo as variáveis chamadas de Z_1 , Z_2 , Z_3 , Z_4 e Z_5 . A variável Z_1 = aumentando uma unidade na taxa de analfabetismo diminui 0.11% na TMI; Z_2 = aumentando uma unidade no PIB per capita diminui 0.20% na TMI. Isto pode ser decorrente da elevada concentração que não permite o benefício ser compartilhado por toda população baixa. Este dado mostra de forma irracional o nosso crescimento econômico onde a justiça social não recebe o benefício do enriquecimento do Estado.

5. CONCLUSÃO

Dentro os modelos analisados são: OLS pooled, efeito fixo e aleatório decidiu-se pelo modelo de efeito fixo, onde as variáveis que causam impacto significativo na TMI foram taxa de analfabetismo, % de pessoas com baixa renda e o PIB per capita, sendo que esta última variável influencia a TMI de forma inversa, ou seja; quanto maior o PIB per capita menor a TMI parecendo absurdo, porém, isto reflete a previsão do capitalismo brasileiro que produz a doença social da má distribuição da riqueza de cada estado. Isto é confirmado com os imutáveis índices de distribuição de renda do Brasil ao longo dos anos. Sendo assim, as variáveis evidenciam a necessidade de políticas públicas mais assertivas e enfáticas para as populações mais pobres e vulneráveis do país.

A contribuição deste trabalho foi mostrar os passos necessários para a análise de DP de forma simples em um programa em linguagem R, tornando acessível realizar esta aplicação, uma vez que, os textos básicos de Análise de Regressão não abordam este tema.

Deixa-se a contribuição de como efetuar uma análise de dados em DP com a utilização do software R e a convicção da sensibilidade de poder diminuir a TMI com investimentos no combate ao analfabetismo e com ações na gestão administrativa que permite uma maior distribuição de renda.

Este trabalho trouxe como base de estudos a pesquisa estatística para fundamentação teórica e prática diante da temática abordada, as limitações estão a partir do dados do DataSUS para conseguinte facilitar a tabulação os dados no programa R. Com base nisso, não se esgota a necessidade de novos estudos que venham a enfatizar as variáveis com impacto significativo na TMI no Brasil, fornecendo subsídios para o cuidado com a saúde da população.

REFERÊNCIAS

DATASUS. Departamento de Informática do Sistema Único de Saúde [internet]. Brasil; 2021. [Acesso em: 08 fev. 2021]. Disponível em: <http://datasus.saude.gov.br/>

HILL C.; GRIFFITHS, W.; JUDGE, G. *Econometria*. Rio de Janeiro: Saraiva; 1999.

COSTA, L. D.; BORGES, L. de M. Características epidemiológicas da mortalidade neonatal e infantil em uma regional de saúde. **Arquivos de Ciências da Saúde da UNIPAR**, Umuarama, v. 26, n. 1, p. 57-64, jan./abr. 2022. Disponível em: <https://ojs.revistasunipar.com.br/index.php/saude/article/view/8250/4198>. Acesso em: 30 abr. 2023.

DUARTE, C. M. R. Reflexos das políticas de saúde sobre as tendências da mortalidade infantil no Brasil: revisão da literatura sobre a última década. **Cad. Saúde Pública**, Rio de Janeiro, v. 23, n. 7, p. 1511-1528, July 2007. Disponível em: <https://www.scielo.br/j/csp/a/RJQFMNBMr7ThwvQnmj84N4p/abstract/?lang=pt>. Acesso em: 10 mar. 2023.

ALVES, T. F.; COELHO, A. B. Mortalidade infantil e gênero no Brasil: uma investigação usando dados em painel. **Ciência & Saúde Coletiva**, [S.L.], v. 26, n. 4, p. 1259-1264, abr. 2021. <http://dx.doi.org/10.1590/1413-81232021264.04022019>. Disponível em: <https://www.scielo.br/j/csc/a/nMq54VMxLCKDSMhsPhK6JYG/>. Acesso em: 15 mai. 2023.

ALMEIDA, A. S. - Estimativa da mortalidade infantil em 50 cidades do Estado da Bahia. *Rev. Serv. Saúde pública*, Rio de Janeiro, 8:153-68, 1955.

IBGE. Instituto Brasileiro de Geografia e Estatística [internet]: [Acesso em: 08 fev. 2021]. Disponível em: <http://www.ibge.gov.br/home/>.

FERNANDES, M. M. C. E. Estudo Da Mortalidade Infantil Nos Estados Brasileiros Utilizando Aplicação De Dados Em Painel. 2016. [Acesso em: 08 fev. 2021]. Disponível em: <http://www.de.ufpb.br/graduacao/tcc/TCC2015p2Maizza.pdf>

CHASE, H.C. Ranking countries by infant mortality rates. *Publ. Hlth Rep.*, 84:19-27, 1969. CMCCD (Committee on Medical Certification of Causes of Death), 1958. Problems in the medical certification of causes of death. *American Journal of Public Health*, 48:71-80.

CARVALHO, A. X. Y.; GÓES, G. S. Introdução ao Software R e à Análise Econométrica. 2018. [Acesso em: 08 fev. 2021]. Disponível em: https://repositorio.enap.gov.br/bitstream/1/3452/5/Aula%203b%20%20Geraldo%20Goes%20e%20Alexandre%20Ywata%20-%20Dados_Painel_v1.pdf.

FERNANDES, M. M. C. E.; *et al.* Fatores que influenciam a mortalidade infantil. **Arquivos de Ciências da Saúde da Unipar**, [S.L.], v. 27, n. 6, p. 2353-2364, 5 jun. 2023. <https://doi.org/10.25110/arqsaude.v27i6.2023-015>. Disponível em: <https://ojs.revistasunipar.com.br/index.php/saude/article/view/10101/4807>. Acesso em: 10 jun. 2023.

GRIFFITHS, W. E.; HILL, C.; JUDGE, G. *Learning and Practicing Econometrics*, John Wiley and Sons, New York, 1993.

WOOLDRIDGE, J. (2002), *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, Massachusetts.

BAROSSO, F. M. e BRAGA, M. B. *Metodologia da Econometria IN: Vasconcellos et al, Manual de Econometria*, Editora Atlas, Sao Paulo, 2000.

GREENE, William H. *Econometric Analysis*. 2nd ed New York: MacMillan, 1993.

DUARTE, P.C.; LAMOUNIER, W.M.; TAKAMATSU, R.T. *Modelos Econométricos para Dados em Paineis: Aspectos Teóricos e Exemplos de Aplicação _a Pesquisa em Contabilidade e Finanças*, v. 4, 2007.