



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE - MESTRADO

**APLICAÇÃO DE DADOS EM PAINEL PARA TRATAR
INFORMAÇÕES DE MORTALIDADE INFANTIL**

Maizza Micaelle Carlos Euclides Fernandes

**João Pessoa
2019**

MAIZZA MICAELLE CARLOS EUCLIDES FERNANDES

**APLICAÇÃO DE DADOS EM PAINEL PARA TRATAR
INFORMAÇÕES DE MORTALIDADE INFANTIL**

Dissertação apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde nível Mestrado do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como requisito regulamentar para obtenção do título de mestra.

Linha de Pesquisa: Modelos de Decisão

Orientadores:

Prof. Dr. João Agnaldo do Nascimento

Prof. Dr. Rodrigo Pinheiro de Toledo Vianna

João Pessoa
2019

Catálogo na publicação
Seção de Catalogação e Classificação

F363a Fernandes, Maizza Micaelle Carlos Euclides.
APLICAÇÃO DE DADOS EM PAINEL PARA TRATAR INFORMAÇÕES DE
MORTALIDADE INFANTIL / Maizza Micaelle Carlos Euclides
Fernandes. - João Pessoa, 2019.
63f. : il.

Orientação: JOÃO AGNALDO DO NASCIMENTO, RODRIGO
PINHEIRO DE TOLEDO VIANNA.
Dissertação (Mestrado) - UFPB/CCEN.

1. Taxa de Mortalidade Infantil. 2. Dados em Painel. 3.
Modelos Pooled. 4. Efeito Fixo e Efeito Aleatório. I.
NASCIMENTO, JOÃO AGNALDO DO. II. VIANNA, RODRIGO
PINHEIRO DE TOLEDO. III. Título.

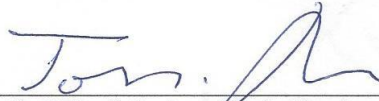
UFPB/

MAIZZA MICAELLE CARLOS EUCLIDES FERNANDES

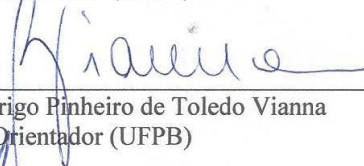
**APLICAÇÃO DE DADOS EM PAINEL PARA TRATAR
INFORMAÇÕES DE MORTALIDADE INFANTIL**

João Pessoa, 28 de fevereiro de 2019

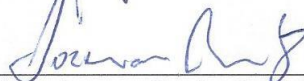
BANCA EXAMINADORA



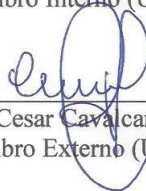
Prof. Dr. João Agnaldo do Nascimento
Orientador (UFPB)



Prof. Dr. Rodrigo Pinheiro de Toledo Vianna
Orientador (UFPB)



Prof. Dr. Jozemar Pereira dos Santos
Membro Interno (UFPB)



Prof. Dr. Cesar Cavalcanti da Silva
Membro Externo (UFPB)



Profa. Dra. Aurilene Josefa Cartaxo Gomes de Arruda
Membro Externo (UFPB)

Dedico este trabalho a Deus, por ser essencial em minha vida, é Ele o autor de meu destino, o meu guia. Aos meus eternos orientadores João Agnaldo, Cesar Cavalcanti e Rodrigo Pinheiro; aos meus pais: Demuntier e Iêda; a minha irmã Marília; a minha prima irmã Clara Costa; ao secretário e segundo pai Chico, e aos meus amigos irmãos: Alisson Santos e Juan Demétrios pelos seus ensinamentos, paciência e confiança ao longo desses anos de convivência diária.

Gratidão, Senhor!

“ Aprendi a agradecer a Deus por todas as coisas.
As lutas me ensinaram a ser forte.
As dificuldades me ensinaram a ser grande.
E em todos os momentos, Deus me ensinou a
viver.
Seja sorrindo, seja chorando, é Deus nos fazendo
crescer.”

[Yla Fernandes]

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me dado forças e coragem para não desistir dos meus sonhos. À minha família que é a base de tudo.

Aos meus pais, Demuntier e Iêda pelo amor e carinho.

À minha irmã Marília, pela sua paciência, dedicação, conselhos, palavras de carinho e forças em momentos que pensei que não fosse capaz de conseguir meus objetivos.

À minha mãe de criação Elizângela, que se dedicou boa parte da sua vida pra cuidar de mim, concedendo amor e força passada mesmo por telefone e por torcer e estar sempre do meu lado em todos os momentos da minha vida desde o meu nascimento.

À minha vizinha Valdetrudez (in memoriam), que sempre foi o meu apoio em tudo, deixando muitas saudades e encorajamento para lutar pelos meus objetivos.

À minha prima irmã Clara Costa, seu esposo Hermano e meu sobrinho Pedro Campos por me apoiarem em cada decisão tomada, pela paciência nos momentos que mais precisei da compreensão, do amor, atenção e por fazerem meus dias mais felizes ao lado de vocês.

Ao meu amigo irmão de coração: Alisson por todos os conhecimentos trocados, pela paciência, alegrias, dificuldades, desesperos, choros, tristezas, companheirismo, forças para prosseguir nesta caminhada árdua ao longo desse mestrado e por ser meu verdadeiro irmão que eu sempre pude contar de verdade e que a estatística me deu esse maior presente que foi conhecê-lo.

Ao meu querido amigo Juan que sempre esteja tão presente na minha vida, pela sua amizade, apoio, palavras de incentivo e perseverança.

Ao meu pai de coração Chico pelo seu apoio dado, pelos conselhos, pela suas palavras de carinho e por acreditar no meu esforço e que eu era capaz de conseguir tudo que almejava. O meu muito obrigada!

Aos meus orientadores: Cesar Cavalcanti, João Agnaldo e Rodrigo Pinheiro; que devo a minha gratidão, pelas experiências passadas, pelos choros trocados nos dias de desespero e por me darem coragem para não desistir dos meus sonhos; pela força que me impulsionou a não desistir do mestrado, por mais que os obstáculos fossem grandes; pela ajuda no desenvolvimento deste trabalho, pela paciência, amor, carinho, dedicação, pelos conselhos, palavras de carinho nos

momentos mais difíceis dessa jornada, enfim, a minha gratidão a vocês são eternas. Não existem palavras concretas para expressar o quanto são importantes na minha vida.

À todos os professores do programa PPGMDS pelo valioso conhecimento transmitido, pelas disponibilidades em ajudar e por contribuírem para o meu crescimento profissional.

Aos colegas de mestrado e doutorado da turma 2017, em especial à Bárbara, Geane, Karlene, Kersinha, Mirian e Sauana, pela amizade, cumplicidade, momentos compartilhados, alegrias vivenciadas, angústias, tristezas, desesperos e encorajamentos.

À Capes pelo apoio financeiro necessário ao longo da pesquisa.

A todos a minha eterna gratidão, pois vocês foram essências para a concretização desse trabalho.

RESUMO

Este trabalho objetiva determinar uma relação linear ou não linear entre a Taxa de Mortalidade Infantil (TMI) e um conjunto de variáveis socioeconômicas e de cobertura observadas por unidade federativas do Brasil no período de 2005 a 2010 utilizando o modelo de dados em painel (*Panel Data*) linear e não linear considerando os efeitos fixo e aleatório. Os resultados mostram que o modelo mais adequado é o de efeito fixo com transformação logarítmica nas variáveis dependente e independentes. A variável dependente foi a TMI e as independentes: 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. Os modelos não lineares Binomial Negativa não apresentaram convergência satisfatória na estimação dos parâmetros para nenhum modelo. O modelo não linear de Poisson apresentou convergência apenas nos modelos *Pooled OLS* (Mínimos Quadrados Ordinários) e *Randon* (*Modelo de Efeito Aleatório*). A conclusão obtida é que as variáveis independentes que causam impacto significativo na TMI são taxa de analfabetismo, PIB per capita e proporção de pessoas com baixa renda. Utilizou-se o software R para realizar esta análise de dados com os pacotes *plm* e *pglm*.

Palavras-chave: Taxa de Mortalidade Infantil; Dados em Painel; Modelos Pooled, Efeito Fixo e Efeito Aleatório.

ABSTRACT

This work aims to determine a linear or non-linear relationship between the Infant Mortality Rate (IMR) and a set of socioeconomic and coverage variables observed by the federative unit of Brazil from 2005 to 2010 using the panel data model (Panel Data) linear and non-linear considering the fixed and random effects. The results show that the most suitable model is the fixed effect with logarithmic transformation in the dependent and independent variables. The dependent variable was the IMR and the independent ones: 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. The nonlinear Binomial Negative models did not present satisfactory convergence in the estimation of the parameters for any model. The Poisson nonlinear model showed convergence only in the Pooled OLS and Random (Random Effect Model) models. The conclusion obtained is that the independent variables that have a significant impact on IMR are illiteracy rate, per capita GDP and proportion of people with low income. R software was used to perform this data analysis with the plm and pglm packages.

Key words: Child mortality rate; Panel Data; Pooled Models, Fixed Effect and Random Effect.

LISTA DE FIGURAS

Figura 4.1 - Conexão entre as hipóteses nulas e alternativas dos testes aplicados ao Modelo Dados em Painel.....	36
Figura 5.1 - Gráfico dos Resíduos para o Modelo OLS Pooled.....	46
Figura 5.2 - Gráfico de Normalidade do Efeito Fixo.....	47
Figura 5.3 - Gráfico de Normalidade do Efeito Aleatório.....	49

LISTA DE TABELAS

Tabela 5.1 - Medidas descritivas das variáveis pertencentes aos indicadores: demográfico, socioeconômicos e de coberturas entre os anos de 2005 a 2010	44
Tabela 5.2 - Resultado da Regressão OLS Pooled (Teste de Chow).....	45
Tabela 5.3 - Resultado da Regressão Linear Múltipla do Modelo de Efeito Fixo.....	47
Tabela 5.4 - Resultado da Regressão Linear Múltipla do Modelo de Efeito Aleatório.....	48
Tabela 5.5 - Resultado dos Testes de Breusch-Pagan e Hausman.....	50
Tabela 5.6 - Resultado do Teste de Efeito Individual dos Estados.....	50
Tabela 5.7 - Resultado da Regressão Binomial Negativo do Modelo de Efeito Fixo.....	52
Tabela 5.8 - Resultado da Regressão Poisson do Modelo Pooled.....	53
Tabela 5.9 - Resultado da Regressão Poisson do Modelo de Efeito Fixo.....	53

LISTA DE ABREVIATURA E SIGLAS

CCEN	Centro de Ciências da Exatas e da Natureza
CMI	Coeficiente de Mortalidade Infantil
DNV	Declaração de Nascimento Vivo
DO	Declaração de Óbito
DPA	Dados em Painéis
FUNASA	Fundação Nacional de Saúde
GLS	Mínimos Quadrados Generalizados
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
INPC	Índice Nacional de Preços ao Consumidor
MI	Mortalidade Infantil
MRBN	Modelo de Regressão Binomial Negativa
MRLS	Modelo de Regressão Linear Simples
MRP	Modelo de Regressão de Poisson
MS	Ministério da Saúde
OLS	Mínimos Quadrados Ordiniais
PIB	Produto Interno Bruto
PNAD	Pesquisa Nacional por Amostra de Domicílios
RC	Registro Civil
SIM	Sistema de Informações sobre Mortalidade
SINASC	Sistema de Informação sobre Nascidos Vivos
SIS	Sistema de Informações
SUS	Sistema Único de Saúde
TCC	Trabalho de Conclusão de Curso
TMI	Taxa de Mortalidade Infantil
UF	Unidades de Federação
UFPB	Universidade Federal da Paraíba

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	18
2.1	GERAL.....	18
2.2	ESPECÍFICOS.....	18
3	REFERENCIAL TEÓRICO	19
3.1	MODELO PARA DADOS EM PAINEL.....	19
3.1.1	Modelo de Efeitos Fixos	20
3.1.2	Modelo de Efeitos Aleatórios	21
3.2	MODELO DE REGRESSÃO.....	21
3.2.1	Modelo de Regressão de Poisson	21
3.2.2	Modelo de Regressão Binomial Negativa	22
3.3	MORTALIDADE INFANTIL.....	27
4	REFERENCIAL METODOLÓGICO	27
4.1	TIPOLOGIA DA PESQUISA.....	27
4.2	FONTE DE DADOS.....	27
4.3	CONSTRUÇÃO E ORGANIZAÇÃO DAS BASES DE DADOS.....	28
4.3.1	Variáveis de estudo	28
4.4	MODELO PARA DADOS EM PAINÉIS.....	30
4.4.1	Modelo de efeitos fixos	31
4.4.2	Modelo de efeitos aleatórios	32
4.5	MODELO POOLED.....	35
4.6	MODELO DE REGRESSÃO LINEAR SIMPLES.....	36
4.6.1	Modelo de regressão múltipla	37
4.6.2	Modelo de Regressão de Poisson (MRP)	39
4.6.3	Modelo de Regressão Binomial Negativa (MRBN)	40
5	RESULTADOS E DISCUSSÃO	42
5.1	ANÁLISE DESCRITIVA DAS VARIÁVEIS.....	42
5.2	MODELOS: OLS POOLED, EFEITO FIXO E EFEITO ALEATÓRIO.....	44
5.3	MODELOS DE REGRESSÃO BINOMIAL NEGATIVA E REGRESSÃO DE POISSON.....	51
6	CONSIDERAÇÕES FINAIS	54
	REFERÊNCIAS	55
	Apêndice A – Script do Programa implementado no software R	59

1 INTRODUÇÃO

A *Teoria dos Dados em Painel* baseia-se na análise de fatores ou objetos para dois ou mais intervalos de tempo. Permitem o uso de mais observações, aumentando o número de graus de liberdade e diminuindo a colinearidade entre as variáveis explicativas. De acordo com Hsião (1986), os modelos para dados em painel oferecem uma série de vantagens em relação aos modelos de corte transversal ou aos de séries temporais sendo que, se referem ao fato de que esses modelos controlam a heterogeneidade presente nos indivíduos.

Para tanto, os dados em painel verificam os efeitos fixos ou aleatórios que não se percebem com clareza no modelo tradicional e caso não seja utilizado tem-se os fatores, que proporcionam estimativas mais explícitas dos vieses nas estimativas que podem sofrer influências diversas da multicolinearidade.

Percebe-se, portanto, uma ampliação e difusão do Método de Dados em Painel (DPA) os quais ganham contornos de significativa importância para a Estatística devido a necessidade de constantes mudanças verificadas nas diversas taxas de controle e nos processos de trabalho por elas gerado. Por falar da dinamicidade no uso de DPA, no campo da saúde esse método pode gerar significativa eficácia, principalmente no que diz respeito ao uso dele no manejo dos dados inerentes ao sistema de informação em saúde. Inclusive, atualmente no Brasil vários estudos apontam a necessidade de ampliar a utilização de métodos estatísticos para trabalhar dados referentes ao Sistema de Informação em Saúde (SIS), tais como o Sistema de Informação de Mortalidade (SIM).

O SIM é uma fonte alternativa de dados de óbitos no país, criado na década de 1970, a partir da implantação da declaração de óbito (DO) em todo o território nacional, tem como principal objetivo fornecer subsídios para traçar um perfil de mortalidade no país. (SCHRAMM. J.M.A; SZWARCWALD.CL, 2000). Assim, essas informações geradas poderão possibilitar o planejamento de ações de saúde que combatam ou diminuam determinadas causas de óbitos, como por exemplo, os precoces ou evitáveis, na população brasileira, por exemplo.

A partir do SIM, é possível calcular, a taxa de mortalidade infantil, que representa o número de óbitos de menores de um ano de idade, dividido por mil nascidos vivos, na população residente em determinado espaço geográfico, verificado durante o período de um ano. Ele é avaliado como um bom indicador de qualidade de vida e das condições de saúde da população,

pois, é possível estimar o número de crianças sobreviventes ao primeiro ano de vida, além de identificar também, a eficácia das políticas públicas nos setores da saúde, educação, saneamento, bem como geração e distribuição de renda no país. (SOUSA T.R.V; LEITE FILHO P.A.M, 2008).

Embora o Brasil tenha apresentado nas últimas décadas um avanço significativo no que se refere a mudanças sócio econômicas, inclusive melhorando de ranking em 2005 no Índice de Desenvolvimento Humano (IDH), o que significa uma redução da taxa de mortalidade infantil nosso país é rico em diversidades regionais que precisam ser consideradas, tanto no que se refere a sua extensão territorial com graves desigualdades sociais, raciais, entre outras. (SOUSA T.R.V; LEITE FILHO P.A.M, 2008)

Nesse sentido, o estudo em questão tem como **finalidade**, desvendar novas possibilidades de uso para o Método dos Dados em Painel (DPA) com vistas a ampliação de suas possibilidades para aplicação em estudos estatísticos e epidemiológicos, nas áreas das ciências exatas, da saúde e da educação. Parte-se do **Pressuposto** que, as ferramentas estatísticas em uso, acompanharam, sem a devida atualização, as necessidades verificadas no dia-a-dia dos processos de trabalho, trazendo para a tomada de decisão, em particular, uma importante lacuna a ser preenchida com a ampliação das possibilidades de uso destas ferramentas.

O **Problema** que se deseja superar com base nos resultados desta investigação é a subutilização deste método nas mais diversas áreas, em que sua aplicação significaria um diferencial no âmbito da coleta e do uso das informações, como nos processos de tomada de decisões para a redução da mortalidade infantil, por exemplo.

A tomada de decisão envolve raciocínio lógico, sendo, portanto, eficaz quando baseada em processos cientificamente adequados para esse fim. Para Chiavenato (2011), no âmbito do processo de trabalho, a tomada de decisão não é um atributo privativo dos administradores, porquanto pertença a todas as pessoas, em todos os lugares, nas diversas situações previsíveis e imprevisíveis.

O estudo se **justifica** pela necessidade de ampliar as possibilidades de uso do Modelo de Dados em Painel. Durante meu trabalho de conclusão de curso foi possível determinar uma relação linear entre a Taxa de Mortalidade Infantil (TMI) e um conjunto de variáveis socioeconômicas observadas em unidades federativas no período de 2005 a 2010, utilizando o modelo de dados em painel de efeitos fixos e aleatório. Identificou-se a possibilidade de

ampliação de coleta, exposição e difusão dos dados, sendo este o interesse pelo objeto do presente estudo.

O projeto integra a linha de pesquisa “Modelos de Decisão” do Programa de Pós-graduação em Modelos de Decisão e Saúde do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba (PPGMDS/CCEN/UFPB). Esta linha desenvolve pesquisa sobre o processamento de informações em várias áreas da Saúde, abrangendo o uso de modelos matemáticos, estatísticos e computacionalmente inteligentes sob uma ótica de suporte a tomada de decisão, estimulando a investigação interdisciplinar.

A **Importância do Estudo** se dá pela possibilidade de ampliar o escopo de utilização do DPA com o adição de outros modelos, tais como a Regressão Binomial Negativa e a Regressão de Poisson na utilização para avaliar as taxas de mortalidade infantil nos Estados Brasileiros.

O **Objeto de Estudo** desta investigação são Dados em Painel, aplicado em toda sua potencialidade para dados relativos a Taxa de Mortalidade Infantil em um determinado período. Na busca por mudanças positivas e potentes nos processos de trabalho, tomada de decisões e ferramentas estatísticas para sua verificação, tomar-se-ão como **Questões Norteadoras do Estudo** as assertivas a seguir:

- a) Há na literatura especializada a descrição de utilização do Método dos Dados em Painel em toda sua potencialidade?
- b) O Método dos Dados em Painel Linear ou com o uso de Regressão Binomial Negativa ou com uso de Regressão Poisson atendem a necessidade de uma modelagem mais precisa para explicar a ligação entre a variável dependente e as variáveis independentes?

Para responder a estas questões formulam-se os seguintes **objetivos**:

2 OBJETIVOS

2.1 GERAL

Utilizar os modelos de Dados em Painel Linear, Poisson e Binomial Negativa, que melhor expliquem a influência na Taxa de Mortalidade Infantil (Variável dependente) para o conjunto de variáveis independentes: taxa de analfabetismo, PIB per capita, baixa renda, abastecimento de água e coleta de lixo.

2.2 ESPECÍFICOS

1. Descrever os modelos de dados em painel Linear, Poisson e Binomial Negativa.
2. Apresentar as estratégias utilizadas para decisão da escolha do modelo de dados em painel.
3. Fazer uma aplicação do modelo de dados em painel Linear, Poisson e Binomial Negativa para descrever a relação entre a mortalidade infantil e o conjunto de variáveis independentes.

3 REFERENCIAL TEÓRICO

Nesta seção iremos apresentar teoricamente alguns métodos importantes para construção dessa dissertação, como definições de modelos de dados em painéis para dados fixos e aleatórios e os modelos de regressão linear, binomial negativa e Poisson.

3.1 MODELOS PARA DADOS EM PAINEL

Dados que se caracterizam por medidas repetidas observadas em um período de tempo não longo, é o que se chama de dados em painel. Apresenta uma componente tempo (longitudinal) e uma relação a ser descrita por uma variável dependente e um conjunto de variáveis independentes. Conhecido na literatura por outros nomes Crosssection, Panel data, Estudo de grupos em dados longitudinais, regressão segmentas longitudinal e outros.

Neste modelo, ao se comparar grupos avaliados ao longo do tempo de forma desagregada se produziria inúmeras equações de regressão e acarretaria maior complexidade em sua interpretação. No entanto, a simplificação de estruturar a informação em um único modelo Dados em Painel para obter informações sobre os grupos e a influência do tempo, leva a perda da suposição básica de independência das observações nos distintos grupos ao longo do tempo, ou seja: a suposição de independência do modelo clássico não se verifica e por esta razão desenvolveu-se uma teoria específica para modelar a relação entre a variável dependente e as independentes nas situações onde as observações variam no grupo de forma temporal e esta teoria elaborou dois tipos de dados em painel: modelo de efeito aleatório e modelo de efeito fixo (HSIÃO, 2003).

O autor prossegue afirmando que, um corte longitudinal, ou painel, é um conjunto de dados, que segue uma dada amostra de indivíduos ao longo do tempo, e, assim, fornece várias observações em cada indivíduo na amostra. Dados em painel tornaram-se amplamente disponíveis em países desenvolvidos e em desenvolvimento.

Conforme Hsião (1986), o método de dados em painel (DPA) permite o uso de mais observações, aumentando o número de graus de liberdade e diminuindo a colinearidade entre as variáveis explicativas. Sabe-se que, quando existe multicolinearidade torna-se difícil estabelecer se um regressor individual influencia uma variável resposta. Uma vez eliminado esse problema,

pode-se obter uma melhora na qualidade da estimação dos parâmetros. Esse tipo de método relaciona-se com uma secção transversal de observações, sejam, indivíduos, grupos, países ou regiões repetida ao longo de vários períodos de tempo.

No modelo com dados em painel, a abordagem no que diz respeito ao tipo de experimento, considera-se modelo de efeito fixo e modelo de efeito aleatório. Esta escolha, como dito antes, é decidida pelo teste de Housman e depende do conjunto de dados que fornecerão evidência estatística para que se decida entre os dois modelos.

Um importante benefício da aplicação dos modelos de dados em painel relaciona-se com a inclusão da heterogeneidade individual, isto é, pode-se mensurar o efeito individual que ocorre nas diferenças existentes com as observações ao longo do tempo. (MARQUES, 2000).

Para Marques (2000), os dados em painéis contribuem para uma menor colinearidade nas variáveis independentes, maior número de grau de liberdade nos resíduos e melhor eficácia na estimação. Portanto a utilização do modelo de dados em painéis no lugar do modelo de regressão linear múltipla representa melhor a realidade dos dados sendo portanto uma modelagem estatística mais comprometida com a realidade.

3.1.1 Modelos de efeitos fixos

O modelo de efeito fixo, também denominado variável *dummy* de mínimos quadrados ou LSDV (*Least Square Dummy Variable*) é um modelo constante no intercepto e inclui uma variável *dummy* para os efeitos das variáveis omissas, que são constante no tempo. Os efeitos individuais podem ser espontaneamente correlacionados com os outros regressores. O objetivo do efeito fixo é verificar os impactos das variáveis independentes variando entre indivíduos e também ao longo do tempo. Para este fim, se supõe que o intercepto varia de um indivíduo para outro, no entanto, é constante ao longo tempo.

A expressão “efeitos fixos” é uma expressão inadequada, uma vez que, tanto no modelo de efeito fixo quanto no modelo de efeito aleatório, tem-se que os efeitos nos indivíduos são variáveis (não são fixos). De acordo com Cameron; Trivedi (2009), os modelos de efeitos fixos mostram quais regressores estão correlacionados com os efeitos devido ao fator (grupos). Para que, uma estimativa consistente dos parâmetros do modelo seja possível é necessário uma

eliminação ou controle dos efeitos fixos. O termo erro da regressão depende dos fatores e do tempo e devem ter variância constante.

3.1.2 Modelos de efeitos aleatórios

Este modelo estima as componentes de variância para os grupos (ou tempos) e o erro, com a suposição de que, os interceptos e inclinações são iguais. O termo erro tem duas componentes: uma sistemática e outra não sistemática, sendo ambas não correlacionadas com os regressores. As suposições para aplicação dos Mínimos Quadrados Ordinários (OLS) são violadas (homocedasticidade). A diferença entre os grupos ou períodos está representada na variância do termo erro e não nos interceptos como no caso do modelo de efeito fixo.

O intercepto não muda de um fator para outro, nem ao longo do tempo. Os parâmetros de inclinação deste modelo são constantes para todos os fatores e todos os períodos de tempo.

3.2 MODELOS DE REGRESSÃO

Os modelos de regressão são aplicados em diversas áreas do conhecimento com o intuito de descobrir as características que define o objeto em estudo, por exemplo, na saúde utilizam-se para construção de indicadores relacionando a desigualdade social com os tipos de violências; um perfil de pacientes submetidos a determinado ensaio clínico para se verificar a evolução ou retardamento de uma patologia. O uso de regressão para dados em contagem é indicado quando a variável dependente representa a contagem de um evento, como o número de carros que passam em determinada avenida em certo horário; o número de horas em que um usuário fica online, assumindo assim uma distribuição de probabilidade de Poisson. A Regressão de Poisson e a Regressão Binomial Negativa possui uma qualidade de ajuste melhor com os dados em contagem do que a Regressão Linear que precisa assumir o pressuposto de normalidade dos dados.

3.2.1 Modelo de Regressão de Poisson (MRP)

O modelo de Regressão de Poisson pertence à classe especial de modelos lineares generalizados, que é uma extensão do modelo de regressão linear. Ela é utilizada nas situações

em que se tem uma variável dependente com dados de contagem e se deseja estudar a relação com variáveis independentes. Utilizada para criar modelos para dados longitudinais em que suas observações podem ser ponderadas de acordo com o tempo de acompanhamento ou o tamanho das unidades observacionais, como por exemplo, número de indivíduos numa coorte de estudo longitudinal (GODOY, 2008).

Um dos pressupostos para o ajuste do MRP se deve a ocorrência da independência entre os eventos ao longo do tempo, visto que pode haver correlação entre os eventos, isto é, o fato de um evento ocorrer pode aumentar a probabilidade de eventos subsequentes, ou seja, pode existir uma forma de dependência dinâmica entre a ocorrência de eventos sucessivos, como por exemplo, uma paciente com doença crônica tende a visitar o médico com certa frequência. (GODOY, 2008).

3.2.2 Modelo de Regressão Binomial Negativa (MRBN)

A distribuição binomial negativa é indicada para a análise de dados de contagem quando ocorre o caso da super-dispersão, ou seja, quando a variância for maior que a média. Nesses casos, o modelo de Poisson é inapropriado mostrando ajustes com resíduos maiores do que os limites permitidos. (SCHMIDT, 2003)

O MRBN analisa a debilidade do modelo de Poisson, ao adicionar um parâmetro que representa a heterogeneidade não-observada entre as observações. Segundo Frees (2004), o MRBN podem ser aplicados para dados de painel ou longitudinais, no qual o elemento de estudo tem observações repetidas em um período de tempo.

3.3 MORTALIDADE INFANTIL

De acordo com Silva et al. (2008), ressalta-se que a mortalidade é motivada por fatores incompatíveis tais quais: as parcialidades sociais e diversidades nas condições de vida; o desenvolvimento da urbanização; transformações na estrutura de idade da população, na formação de trabalhadores e no mercado de trabalho, no nível de educação da população e modificações na ordenação dos serviços de saúde.

De acordo com Paes (2007), as estatísticas de mortalidade são componentes do sistema de estatísticas vitais, subutilizadas em vários países da América Latina em virtude de seus limites por serem incompletos, desatualizados e dispersos. A subutilização se deve da mesma forma à crença injustificada de que esses limites comprometeriam permanentemente qualquer resultado que deles poderiam ser derivados.

O crescimento da mortalidade estabelece um considerável componente para verificação e percepção do crescimento populacional, fruto também no caso do Brasil, pela violência crescente que assola a realidade de nosso país. No entanto, como é foco do nosso estudo a mortalidade na infância iremos destacar aqui os estudos feitos especialmente por Giorgio Mortara que explora por meio dos censos demográficos a mortalidade global e infantil para o Brasil.

Segundo o autor, para analisar o crescimento da mortalidade global e infantil no Brasil, é necessário conhecer a população em estudo, o número total de óbitos que ocorrem, englobando os menores de um ano de idade e o número de nascidos vivos. Ainda que se conheça normalmente a estimativa da população, não se possui outros elementos, em decorrência das imprecisões de registros de óbitos e especialmente de nascimentos.

A mortalidade infantil (MI) elevada é conhecida desde os anos 1970, por infraestrutura de saúde ineficiente. Porém, ela pode ser evitável desde que seja estabelecido uma melhor condição na qualidade da atenção a saúde (RUTSTEIN, 1976). Nos países onde há um número alto em relação ao risco de morrer, com menos de um ano de idade, tem-se mostrado mais difícil de conseguir estimadores com capacidade de mostrar com clareza os fatores que influenciam a mortalidade infantil. É como se o atraso cultural e da qualidade de vida impedissem de observar detalhadamente este grande problema que fica escondido por trás de um ineficiente sistema de informações de saúde. Diante do exposto, discutir o tema da mortalidade infantil é essencial não só como reflexo positivo ou negativo da situação socioeconômica e médico-sanitária de uma sociedade, mas também, esse indicador possibilita planejar e construir novas metas e ações em saúde com intuito de melhorar a qualidade de vida de uma determinada população.

Segundo Araújo (1973), em Salvador, capital do estado Bahia, por exemplo, ocorreu melhorias nos últimos anos nos dados do registro de óbitos, porém, ainda existe um sério problema quanto à evasão do registro civil de nascimento. Diante disso, o Serviço de Bioestatística da Secretaria da Saúde Pública adotou o padrão de calcular a mortalidade infantil considerando como denominador, o número estimado de nascidos vivos em cada ano. Esta

estimativa é feita a partir da população total e do índice de natalidade, calculados com base nos dados dos últimos censos. A estimativa de mortalidade infantil tem o proveito de assegurar uma maior comparação de ano a ano, sem correr os riscos de erros resultantes das mudanças no registro civil.

A origem das estatísticas de mortalidade é a Declaração de Óbito (DO) que, além de dar origem à certidão (documento de caráter jurídico), determina diversas utilidades, tais como: benefício para programas de saúde pública; ponto de casos para investigação clínico-epidemiológica; documento inicial para estudos retrospectivos; desfecho para estudos prospectivos; uso em seguros de vida e estudo demográficos de tendências da mortalidade (CASTRO; CERGOLE; CARNEIRO; MUCINHATO; SERVO, 2002)

O tema mortalidade é fundamental para identificar o perfil epidemiológico de uma população que é disponibilizado por meio da Declaração de Óbito (DO), dessa forma o estudo de seus registros auxilia aos investigadores na preparação do quadro epidemiológico. A morte mesmo sendo um fato marcante no ponto jurídico e médico, passou por modificações na sua descrição no transcorrer do tempo.

A morte pode ser juridicamente entendida como o ‘desaparecimento permanente de todo sinal de vida, em um momento qualquer depois do nascimento’, ou, conforme definição da Organização Mundial de Saúde, ‘o desaparecimento de todos os sinais de vida ou a cessação das funções vitais, sem a possibilidade de ressuscitar’ (LAURENTI; MELLO JORGE, 1992).

Em 1975 o Ministério da Saúde aprovou a DO uniformizada para todo o país, visto que em comparação às causas de morte conserva-se o padrão que foi proposto em 1948 pela Organização Mundial de Saúde (MELLO JORGE; LAURENTI; GOTLIEB, 2002). A DO é o documento encarregado pelo registro da morte e “o documento oficial hábil para que a pessoa seja considerada morta, permitindo, portanto, que seja iniciada toda uma sequência de atos jurídicos, principalmente dentro do Direito de Família e do Direito das Sucessões” (MELLO JORGE, 2006).

No decorrer dos anos, o modelo da padronização da DO sofreu modificações e integração de variáveis. Nos dias de hoje esse documento é expedido em três vias numeradas que são impressas pelo Ministério da Saúde e são distribuídas para as Secretarias Estaduais de Saúde que ficam responsáveis para repassar as Secretarias Municipais de Saúde, que entregam cópias as

associações de saúde, Instituto de Medicina Legal, Serviços de Verificação de Óbito e Cartórios de Registro Civil (MELLO JORGE; LAURENTI; GOTLIEB, 2002). Contudo, em conformidade com a Fundação Nacional de Saúde (FUNASA) a sucessão varia conforme o tipo de óbito e onde ocorreu.

O Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde (MS), consiste em uma fonte opcional de dados de óbitos no país. Em 1976 foi criada a partir da fundação do modelo padronizado da DO em todo o território nacional, o SIM tem a finalidade indispensável de oferecer auxílio para traçar o perfil de mortalidade no país (BRASIL, 1982). Além da disponibilidade de dois sistemas de informações sobre óbitos em nível nacional, o do Registro Civil (RC) e o SIM, uma das grandes complicações que ainda restringem as análises de mortalidade no Brasil é o sub-registro de óbitos, que tem relevância significativa, principalmente, nas regiões Norte e Nordeste, com predomínio entre as crianças com menos de um ano de idade. O sub-registro de óbitos diz respeito, sobretudo, à ocorrência de funeral sem a determinação da certidão, nos designados cemitérios clandestinos, sendo relacionado à pobreza e predominante na área rural (MELLO-JORGE, 1983, 1988).

Os dados sem retificação representam principalmente a ocorrência de mortalidade das regiões mais desenvolvidas. Do mesmo modo é benéfico para os dados do litoral do Nordeste em comparação ao sertão nordestino, das capitais com relação ao interior dos estados, dos centros urbanos em comparação às áreas rurais. Os dados de registro vão sempre refletir às áreas que melhor informam a mortalidade. Mesmo que retifiquem as informações de registro por razões diferenciadas por região, está corrigindo apenas aquelas relacionadas às áreas que, de uma certa forma, notificam a mortalidade e que, supostamente, possuem as melhores condições de saúde.

Em 1990, no que está relacionado aos dados referentes aos nascimentos, o MS estabeleceu o Sistema de Informação sobre Nascidos Vivos (SINASC). Esse sistema tem como alicerce o uso da Declaração de Nascimento Vivo (DNV), documento cuja emissão é obrigatória no serviço de saúde onde ocorreu o parto. Diferente do sistema de nascimentos do Registro Civil, dos quais o propósito primordial é a contagem do número de registros de nascimentos, o SINASC tem como finalidade evidenciar as condições de nascimento (MELLO-JORGE et al., 1996).

Apesar da cobertura do SINASC está progredindo, principalmente no que se refere a qualidade da informação desde a sua implantação, percebe-se que a cobertura do sistema ainda não é total em virtude das diferenças regionais e peculiaridades inerentes aos estados brasileiros.

O confronto entre o número de nascimentos ocorridos nos hospitais do Sistema Único de Saúde (SUS) e os declarados no SINASC, demonstra ainda insuficiências na cobertura do SINASC em alguns estados do país, ainda que de menor relevância que as mostradas pelo SIM (SCHRAMM; SZWARCOWALD, 2000).

Na presença de deficiências das fontes de informações, o IBGE, que é o instituto responsável por propiciar a estimativa da mortalidade infantil no Brasil, tem utilizado a técnicas demográficas, ditas de classificação indireta, para representar as estimativas clássicas (IBGE, 2001). Assim, as técnicas são fundamentadas em dados de entrevistas domiciliares que são executadas por viabilidade dos Censos Demográficos ou das Pesquisas Nacionais de Amostra por Domicílio (PNAD).

O coeficiente de mortalidade infantil (CMI) é um dos indicadores mais utilizado para avaliar níveis de saúde e de crescimento social de um lugar. A mortalidade infantil verifica o risco de um nascido vivo morrer no seu primeiro ano de vida. Nesse sentido, o coeficiente abaixo de 20 por 1000 é considerado baixo; 50 ou mais por 1000 nascidos vivos é considerado alto.

A concepção de que o CMI é um indicador excessivamente sensível, precisa ser atualizada. As vertentes fortemente procedentes da mortalidade infantil, que vem sendo percebidas no país como um todo, estariam, segundo concepções usuais, indicando uma melhoria significativa das condições de saúde e, portanto, da qualidade de vida das populações. Ainda assim, quando se procede a análise de outros indicadores socioeconômicos, percebe-se que estes não evidenciam uma evolução coerente com a queda da mortalidade infantil e muitas vezes chegam a comprovar a degradação das condições e da qualidade de vida da população.

A taxa de mortalidade infantil (TMI) é identificado não só como medida das situações de saúde da população inferior a 1 ano, elas servem para descrever as circunstâncias total da vida, sendo conhecida como, do bem-estar, qualidade de vida e do grau de crescimento da população. Dessa maneira, é provável determinar ligações entre a magnitude da TMI e as características do modelo de desenvolvimento de uma população.

A TMI é definida por:

$$TMI = \frac{\text{Número de óbitos de residentes com menos de um ano de idade}}{\text{Número de nascidos vivos de mães residentes}} \times 1000$$

4 REFERENCIAL METODOLÓGICO

4.1 TIPOLOGIA DA PESQUISA

Trata-se de um processo investigativo exploratório, descritivo e inferencial mediante uma abordagem quantitativa. Segundo Gil (1991) as pesquisas exploratórias são aquelas que têm como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias, com vistas à formulação de problemas mais precisos ou questionamentos/hipóteses pesquisáveis para estudos posteriores. As pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar uma visão geral sobre determinados fatos. Este tipo de pesquisa é realizado especialmente quando o tema escolhido é pouco explorado e torna-se difícil formular questionamentos/hipóteses precisas e operacionalizáveis sobre ele.

Para Martins (1994) as pesquisas descritivas são aquelas em que o pesquisador observa, registra, analisa e correlaciona fatos ou fenômenos. Neste tipo de pesquisa busca-se: frequências; natureza dessas frequências; características do fenômeno a ser pesquisado; causas, relações e associações entre estes fenômenos.

A abordagem quantitativa envolve a obtenção de dados obtidos mediante a aplicação de testes estatísticos sobre a quantidade de eventos ou fenômenos estudados. Nesta abordagem a aparência do fenômeno é mais importante que sua essência. Trabalha com dados primários (informações brutas colhidas diretamente no cenário da pesquisa) e dados secundários (informações já trabalhadas através de testes específicos).

4.2 FONTES DE DADOS

As fontes oficiais de referência com relação a 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 estados e municípios brasileiros.

Os dados relacionados a mortalidade infantil e os indicadores socioeconômico e de cobertura, usados para este trabalho foram extraído 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.

4.3 CONSTRUÇÃO E ORGANIZAÇÃO DAS BASES DE DADOS

Os dados obtidos dos sistemas de informações foram por Unidade de Federação, sendo organizados em planilhas eletrônicas do Microsoft Office Excel 2007. Foram extraídos a taxa de mortalidade infantil; Taxa de analfabetismo: percentual da população de 15 ou mais anos não alfabetizada; PIB per capita; Proporção de pessoas com baixa renda; Abastecimento de água; Coleta de lixo das UFs que é composto por 27 estados brasileiros.

Os dados para os anos 2005 à 2010 foram retirados do Sistema de Informação sobre Mortalidade (SIM) para as UFs. As variáveis do estudo com seus referentes valores foram transferidos para um banco de dados nos software estatísticos R, sendo executadas as análises dos dados por meio de aplicações de dados em painel.

4.3.1 Variáveis de estudo

Dos indicadores e dados básicos para avaliar a mortalidade infantil nas UFs, selecionou-se 5 (cinco) variáveis pertencentes aos indicadores socioeconômicos e indicadores de coberturas, sendo 3 (três) pertencentes a Indicadores Socioeconômico: Taxa de analfabetismo; produto interno bruto (PIB) per capita; proporção de pessoas com baixa renda e 2 (dois) pertencentes a 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 definição é feita da seguinte forma:

1) Taxa de analfabetismo

Analfabetismo são aquelas pessoas que não sabem ler nem escrever. A taxa de analfabetismo é a porcentagem da população de 15 ou mais anos não alfabetizada das Unidades de Federação.

2) Produto interno bruto (PIB) per capita

O PIB é a soma em valores, de todo o capital e serviços finais produzidos numa cidade, estado, país. Os valores do PIB per capita em reais correntes e valores do PIB em milhões de reais correntes.

3) Proporção de pessoas com baixa renda

São pessoas com renda domiciliar mensal per capita de até um salário mínimo. O salário mínimo do último ano para o qual a série está sendo calculada torna-se a referência para toda a série. Esse valor é corrigido para todos os anos anteriores com base no INPC (Índice Nacional de Preços ao Consumidor) de setembro do último ano, alterando o valor da linha de pobreza e consequentemente a proporção de pobres. Nesta tabela, o valor de referência, salário mínimo de 2012, é de R\$ 622,00. Por se tratar de uma pesquisa amostral, o valor do indicador pode não ter significância estatística quando desagregado para segmentos populacionais específicos, tais como indígenas, amarelos e pretos, pois estes grupos são muito pequenos em alguns estados e regiões. Os valores das PNAD 2001 a 2012 estão ponderados considerando os pesos amostrais disponibilizados após a publicação do Censo 2010.

4) Proporção da população servida por rede de abastecimento de água

Mensura o revestimento de serviços de abastecimentos satisfatório de água para população por meio de um conjunto de obras, equipamentos e serviços designado ao abastecimento de água potável para o consumo da população. Esta proporção da população servida por rede de abastecimento de água tem os valores das PNAD 2001 a 2012 estão ponderados considerando os pesos amostrais disponibilizados após a publicação do Censo 2010.

5) Proporção da população servida por coleta de lixo

Mensura o revestimento de serviços de coleta domiciliar de lixo de uma população. Esta proporção da população servida por coleta de lixo tem os valores das PNAD 2001 a 2012 estão ponderados considerando os pesos amostrais disponibilizados após a publicação do Censo 2010.

4.4 MODELO DADOS EM PAINEL

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

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

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.

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} x_{1i1} & x_{2i1} & \dots & x_{ki1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \dots & x_{kiT} \end{bmatrix}, \boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i1} & \beta_{1i1} & \beta_{2i1} & \dots & \beta_{ki1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{0iT} & \beta_{1iT} & \beta_{2iT} & \dots & \beta_{kiT} \end{bmatrix}, \mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}$$

Em que \mathbf{y}_i e \mathbf{e}_i são vetores de dimensão $(T \times 1)$ e contém respectivamente, as T variáveis dependentes e os T erros. \mathbf{X}_i é uma matriz de dimensão $(k \times T)$ 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, $\boldsymbol{\beta}_i$ é a matriz dos parâmetros a serem estimados.

Neste modelo geral, o intercepto e os parâmetros resposta são distintos para cada indivíduo e para cada período de tempo. Encontrando assim, mais parâmetros desconhecido do que observações, não sendo possível, neste caso, estimar os seus parâmetros. Assim, é necessário identificar suposições em relação ao modelo geral, afim de transforma-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.

4.4.1 Modelos de Efeito Fixo

Em conformidade com Hill, Griffiths e Judge (1999), as suposições do modelo são:

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

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), será:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \cdot \\ \cdot \\ \cdot \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{kiT} \end{bmatrix} \times \begin{bmatrix} \beta_{0i1} & \beta_{1i1} & \cdots & \beta_{ki1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0iT} & \beta_{1iT} & \cdots & \beta_{kiT} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \cdot \\ \cdot \\ \cdot \\ e_{iT} \end{bmatrix}$$

Assim, neste modelo o α_i refere-se aos interceptos a serem estimados, um para cada indivíduo. Assim como os parâmetros respostas não muda entre os indivíduos e nem ao longo do tempo, todas as diferenças de procedimentos entre os indivíduos deve 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. 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:

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

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 $\beta_0 = \gamma_1$. Para $i \geq 2$, o intercepto é dado por $\beta_0 + \gamma_i$. Através do OLS é estimado o modelo de efeito fixos. As estimativas alcançadas são não tendenciosas e consistentes, por sua vez o modelo supõe que os erros tem distribuição normal, variância constante e não são correlacionados.

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 nula e alternativa são as seguintes:

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

$$H_1 : \text{os interceptos } \beta_{0i} \text{ não são todos iguais}$$

As hipóteses 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.

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. 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.

4.4.2 Modelos de Efeito Aleatório

O modelo de efeito aleatório, as variáveis tem 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-se que os indivíduos possui dados que são amostras aleatórias de indivíduos com uma população maior. Assim como proposto por Hill, Willian, Griffiths e Judge (1993), serão modelados os n interceptos da seguinte forma:

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

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, $\bar{\beta}_0$, refere-se ao intercepto populacional. O modelo geral de efeitos aleatórios é dado da seguinte forma:

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

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

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

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

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 do 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 o 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 (LM), 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. Quando não, é necessário atribuir que o modelo de efeitos aleatórios é mais adequado para os dados que estão sendo abordados.

Segundo Wooldridge (2002), argumenta que a principal razão para escolher entre o modelo de efeito fixos e o modelo de efeitos aleatórios é o efeito não observado α_i . O modelo de efeitos aleatórios é o mais apropriado em circunstâncias que α_i não é correlacionado com todas as variáveis explicativas. Do contrário, se α_i for correlacionado com algumas variáveis explicativas, logo deve ser aplicado o modelo de efeitos fixos. Portanto, nesse sentido, o modelo de efeitos aleatórios gera estimadores inconsistentes.

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:

$$H_0: \alpha_i \text{ não é correlacionado com as variáveis explicativas}$$

$$H_1: \alpha_i \text{ é correlacionado com as variáveis explicativas}$$

Em Greene (1997) 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 utiliza o modelo de efeitos fixos.

4.5 MODELO POOLED

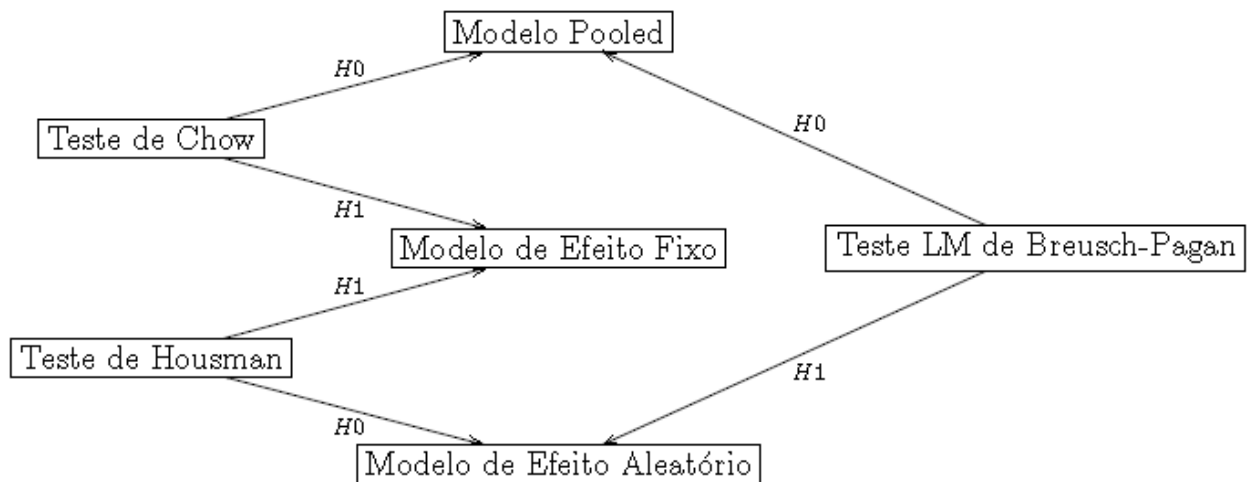
A especificação do modelo *pooled cross-section* básica é:

$$Y_{it} = a + b_1X_{it} + e_{it}$$

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. 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. Pelo teste rejeita-se a hipótese nula de que a variância do coeficiente individual estimado é igual a 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. A figura abaixo mostra a conexão entre as hipóteses nulas e alternativas dos testes aplicados na análise do modelo de Dados em Painel, onde são essenciais o teste de Chow, Bruesch-Pagan e Housman.

Figura 4.1: Conexão entre as hipóteses nulas e alternativas dos testes aplicados ao Modelo de Dados em Painel



4.6 MODELO DE REGRESSÃO LINEAR SIMPLES

Um modelo de regressão linear simples (MRLS) descreve uma relação entre duas variáveis X e Y se admitirmos que Y é função linear de X , podemos estabelecer uma regressão linear simples, cujo modelo estatístico é:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Onde X é uma variável independente (explicativa), Y é uma variável dependente (resposta), β_0 é o coeficiente linear da reta, conhecido como termo constante da equação de regressão, β_1 é o coeficiente angular da reta, denominado coeficiente de regressão e ε é o erro aleatório.

Ao estabelecer o modelo de regressão linear simples, se pressupõe que:

- 1 - A relação entre X e Y é linear.
- 2 - Os valores de X são fixos, isto é, X não é uma variável aleatória.
- 3 - A média do erro é nula, isto é, $E(\varepsilon) = 0$.
- 4 - Para um dado valor de X , a variância do erro ε é σ^2 para todos os valores. Esta variância é

denominada de variância residual, dada por:

$$E(\varepsilon_i^2) = \sigma^2$$

ou

$$\sigma^2 = E [Y_i - E (Y_i | X_i)]^2.$$

Tendo em vista que a variância do erro não varia para cada valor de X, diz-se que a variável aleatória erro é homocedástica ou que possui homocedasticia (na variável erro ou na variável dependente).

5- O erro de uma observação é não-correlacionado com o erro em outra observação, isto é: $E(\varepsilon_i \varepsilon_j) = 0$ para $i \neq j$.

6- Os erros têm distribuição normal.

4.6.1 Modelo de Regressão Linear Múltipla

O modelo de regressão linear múltipla ocorre quando se admite que o valor da variável dependente é função linear de duas ou mais variáveis independentes. Este modelo para o caso de k variáveis independentes é dado por:

$$Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \varepsilon_j, \quad j = 1, \dots, n.$$

ou,

$$Y_j = \alpha + \sum_{i=1}^k \beta_i X_{ij} + \varepsilon_j.$$

Utilizando notação matricial o modelo pode ser escrito na forma

$$Y = X\beta + \varepsilon$$

onde,

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad \text{e} \quad \varepsilon_i = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}.$$

As pressuposições agora são as seguintes:

- 1) a variável dependente (Y_i) é função linear das variáveis explanatórias (X_{ij} , $i = 1, \dots, k$);
- 2) os valores das variáveis explanatórias são fixos;
- 3) $E(\varepsilon) = \mathbf{0}$, onde $\mathbf{0}$ representa um vetor de zeros;
- 4) os erros são homocedásticos, isto é, $E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$;
- 5) os erros são não-correlacionados entre si, isto é, $E(\varepsilon_j\varepsilon_h) = 0$ para $j \neq h$;
- 6) os erros têm distribuição normal multivariada ou seja: $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

4.6.2 Modelo de Regressão de Poisson (MRP)

O modelo de regressão de Poisson é utilizado quando os dados a ser analisados são em forma de contagem, sendo y_i uma variável aleatória dependente que indica o número de vezes que um determinado evento ocorre, por exemplos: número de mortalidade por ano, número de acidentes de moto por ano para o indivíduo i . Logo, ela assume números inteiros que variam de 0 a $+\infty$, $y = \{0, 1, 2, \dots\}$. Portanto, se y tem uma distribuição de probabilidade de Poisson:

$$P(y|\mu) = \frac{e^{-\mu} \mu^y}{y!}, \text{ para } y = 0, 1, 2, \dots$$

Uma relevante propriedade da variável aleatória (y) é que a média e a variância são iguais, e indica se a variação da distribuição de Poisson está definida pela distribuição (JONES, 2007).

$$E(y) = \mu \quad \text{e} \quad \sigma^2(y) = \mu$$

Além do mais, é importante observar que μ deve atribuir valores positivos, pois retrata a média ou o número esperado das contagens, $\mu = \exp(x'\beta)$. Uma vez que μ seja muito pequeno, nessa situação espera-se que existam muitos casos de zero, e a distribuição tem o seu viés positivo. Logo, com o aumento médio dos eventos a distribuição aproxima-se de uma distribuição normal. Por fim, sempre que o número médio de eventos aumenta, a variância também aumenta por meio da população.

O MRP pode ser representado como (CAMERON E TRIVEDI, 1998):

$$Y_i = E(y_i) + \mu_i = \mu_i + \mu_i$$

Onde Y é uma variável aleatória independentemente distribuída com média μ_i para cada indivíduo, e :

$$\mu_i = \exp(x'_i \beta), \text{ onde } i = 1, \dots, n$$

Onde, x representa as variáveis dependentes.

O MRP estende a distribuição de Poisson, admitindo que cada observação tenha um diferente valor de μ . Assim, no modelo de regressão de Poisson atribui-se que a contagem observada para o indivíduo i é derivada de uma distribuição de Poisson com média μ_i , onde μ_i é estimado pelas características observáveis. Dessa forma, tem a seguinte equação estrutural:

$$\mu_i = E(y_i|x_i) = \exp(x_i' \beta)$$

Considerando a exponencial de $x\beta$, μ será e positivo, significando que a contagem só pode assumir zero ou um número positivo.

4.6.3 Modelo de Regressão Binominal Negativa (MRBN)

A Distribuição Binomial negativa é indicada para a análise de dados de contagem quando ocorre o caso da super-dispersão, ou seja, quando a variância for maior que a média. Nesses casos, o MRP é inapropriado mostrando ajustes com resíduos maiores do que os limites permitidos. A distribuição é obtida supondo que o parâmetro de intensidade do MRP tem um componente estocástico v_i , onde v_i assume uma distribuição Gamma, possibilitando que a variância seja diferente da média amostral. Sendo assim, o MRBN pode ser especificado da seguinte forma:

$$h(y|\mu, \alpha) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left[\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right]^{\alpha^{-1}} \left[\frac{\mu}{\mu + \alpha^{-1}} \right]^y, y = 0, 1, 2, 3, \dots$$

Com média e variância iguais a:

$$E(Y|\mu, \alpha) = \mu \quad \text{e} \quad \sigma^2(y|\mu, \alpha) = \mu(1 + \alpha \mu)$$

Onde, α é o parâmetro de dispersão, medida da extensão da super-dispersão, $\mu_i = \exp(x_i' \beta)$, $\Gamma(\cdot)$ é uma função gamma.

Se $\alpha \rightarrow 0$, $\text{Var}(Y) \rightarrow \mu$ a distribuição binomial negativa converge para a distribuição de Poisson (média = variância), isto é, o MRP é um caso especial do MRBN. Se $\alpha > 0$, esse parâmetro é utilizado para corrigir o desvio-padrão dos coeficientes de regressão. É importante notar que u_i representa um erro aleatório. Ou seja, introduz-se na equação, que define a média amostral, um termo de erro. Assim, indivíduos com o mesmo valor de x_i observado podem não ter a mesma média, devido à heterogeneidade não observável, a média continua igual a Poisson. Porém, a variância muda com a adição α , admitindo desta maneira que a variância ultrapasse a média. Esta mudança faz com que o MRBN seja mais flexível do que o MRP.

5 RESULTADOS E DISCUSSÃO

5.1 ANÁLISE DESCRITIVA DAS VARIÁVEIS

Nessa seção apresenta-se a aplicação dos dados em painel utilizando modelos de regressão linear múltipla, binomial negativa e Poisson sobre um conjunto de variáveis, que são: Taxa de analfabetismo, Produto interno bruto (PIB) per capita, Proporção de pessoas com baixa renda, Proporção da população servida por rede de abastecimento de água e Proporção da população servida por coleta de lixo. Os dados são formados por uma série temporal de 2005 a 2010 das 27 UFs sendo equilibrado no sentido de todas as unidades estarem presentes em todos os anos. Inicialmente apresenta-se uma análise descritiva (média, valor mínimo e valor máximo) das variáveis em estudo por anos, em que podemos observar na (Tabela 5.1).

A tabela 5.1 mostra os resultados das estatísticas descritivas das variáveis pertencentes aos indicadores: demográfico, socioeconômicos e de coberturas para todos os anos em estudo. O indicador demográfico informa que a Taxa de mortalidade infantil em 2005 alcançou a maior média, com 22,27 número de óbitos infantis (menores de um ano) por 1.000 nascidos vivos, com o valor mínimo de 12,90 e o máximo de 28,60 óbitos infantis por 1.000 nascidos vivos. Já no ano de 2010 obteve a menor média com 17,27 óbitos infantis, valor mínimo de 11,20 e o máximo de 25,40. Conforme observado (Tabela 5.1), o nível da Taxa de Mortalidade Infantil (TMI) baixou ao longo do período (2005-2010).

A Taxa de Analfabetismo obteve uma menor média de 13,91% em 2005, com o valor mínimo de 4,66% e o máximo de 29,27%. O ano de 2010 obteve a maior média da Taxa de Analfabetismo com 11,83% de pessoa que não sabem ler nem escrever, valor mínimo de 3,59% e o máximo de 23,64%. Conforme observado (Tabela 5.1), o nível da Taxa de Analfabetismo diminuiu ao longo do período (2005-2010).

O PIB per capita mostra uma menor média de R\$ 9.836,00 reais no ano de 2005, com o valor mínimo de R\$ 3.701,00 reais e o máximo de R\$ 34.515,00 reais. O ano de 2010 mostra um maior crescimento no PIB per capita com uma média de R\$ 16.810,00 reais, valor mínimo R\$6.883,00 reais e o máximo R\$ 58.326,00 reais. De acordo com a Tabela 5.1, o PIB per capita aumentou ao longo do período dos anos.

A Variável de Proporção de Pessoas com baixa renda mostra que no ano de 2005 obteve-se a sua maior média (54,32%) entre as populações das UF's com renda domiciliar mensal per capita de até um salário mínimo, com o valor mínimo de 23,27% e o máximo de 77,56%. O ano de 2010 obteve um declínio em relação ao ano de 2005, com média de 41,10% da população. Diante disso, observa-se que houve uma diminuição de pessoas com baixa renda ao longo dos anos (2005-2010).

A Variável de Proporção da População servida por rede de abastecimento de água indica que no ano de 2005 obteve-se a menor média (73,79%) entre o período de anos, o valor mínimo de 36,24% e o máximo de 95,87%. O ano de 2009 obteve um maior percentual em relação aos anos analisados, com média de 76,70% da população, valor mínimo 39,05% e máximo 96,95%.

A Variável de Proporção da População servida por coleta de lixo obteve-se a menor média (78,13%) no ano de 2005, com o valor mínimo de 48,98% e o máximo de 98,38%. O ano de 2009 mostra um maior percentual em relação aos anos analisados, com média de 82,83% da população, valor mínimo 54,95% e máximo 98,96%.

As mudanças significativas podem ser observadas para as variáveis ao longo dos anos. Acompanhando a tendência observada para as variáveis nas UF's, a mortalidade infantil apresentou um declínio importante do ano de 2005 com relação ao ano de 2010. Em 2005 era de 22,27 óbitos infantis (menores de um ano) por 1.000 nascidos vivos, obtendo uma diminuição para 11,27 óbitos infantis no ano de 2010, conforme mostra a tabela 5.1.

Tabela 5.1: Medidas descritivas das variáveis pertencentes aos indicadores: demográfico, socioeconômicos e de coberturas entre os anos de 2005 a 2010

Variáveis	Medidas descritivas	Anos					
		2005	2006	2007	2008	2009	2010
Taxa de Mortalidade Infantil	<i>Média</i>	22,27	21,60	20,50	19,30	18,30	17,27
	<i>Mínimo</i>	12,90	12,60	12,00	11,70	11,20	11,20
	<i>Máximo</i>	28,60	27,80	30,20	29,00	28,80	25,40
Taxa de Analfabetismo	<i>Média</i>	13,91	12,97	12,51	12,30	11,85	11,83
	<i>Mínimo</i>	4,66	3,77	3,73	4,02	3,26	3,59
	<i>Máximo</i>	29,27	26,44	25,14	25,74	24,57	23,64
PIB per capita	<i>Média</i>	9.836,0	10.680,0	11.863,0	13.718,0	14.600,0	16.810,0
	<i>Mínimo</i>	3.701,0	4.212,0	4.611,0	5.372,0	6.051,0	6.883,0
	<i>Máximo</i>	34.515,0	37.599,0	41.062,0	45.978,0	50.438,0	58.326,0
Proporção de Pessoa com baixa renda	<i>Média</i>	54,32	49,64	47,53	44,15	42,18	41,10
	<i>Mínimo</i>	23,27	18,62	16,88	17,77	13,81	13,86
	<i>Máximo</i>	77,56	72,59	67,44	66,49	65,93	64,70
Proporção da População servida por rede de abastecimento de água	<i>Média</i>	73,79	74,71	75,18	76,63	76,70	74,65
	<i>Mínimo</i>	36,24	39,65	40,30	41,64	39,05	37,64
	<i>Máximo</i>	95,87	96,21	96,56	96,33	96,95	94,87
Proporção da População servida por coleta de lixo	<i>Média</i>	78,13	79,24	80,76	81,91	82,83	81,07
	<i>Mínimo</i>	48,98	49,64	53,17	54,46	54,95	53,55
	<i>Máximo</i>	98,38	98,37	98,64	98,46	98,96	98,18

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

5.2 MODELOS: OLS POOLED, EFEITO FIXO E EFEITO ALEATÓRIO

Apresentaremos aqui uma análise completa da aplicação dos dados em painel nos modelos de Regressão Linear Múltipla: OLS Pooled, Efeito Fixo e Efeito Aleatório. A tabela (5.2) abaixo apresenta os resultados das variáveis explicativas do modelo estimado pelo método Pooled, que não leva em conta a especificação do efeito individual omitido, tendo um total de 162 observações.

Tabela 5.2: Resultado da Regressão OLS Pooled (Teste de Chow)

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Intercepto	0,1532	3,0391	5,0419	$0,1265 \times 10^{-5}***$
Taxa de Analfabetismo	-0,3126	0,0589	-5,2997	$0,3898 \times 10^{-6}***$
PIB per capita	$8,1289 \times 10^{-5}$	$3,3951 \times 10^{-5}$	-2,3943	0.0178*
Proporção de Pessoa com baixa renda	0,3132	0,0264	11,8650	$0,0022 \times 10^{-13}**$
Proporção da População servida por rede de abastecimento de água	-0,0651	0,1899	-3.4275	$0,0779 \times 10^{-2}***$
Proporção da População servida por coleta de lixo	-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: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

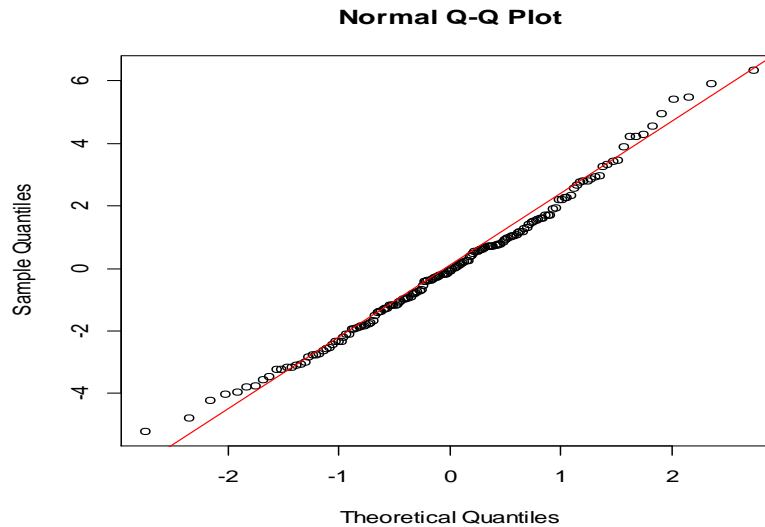
Os resultados estimados com o modelo OLS Pooled apresentados na tabela 5.2 mostra que as variáveis Taxa de analfabetismo, PIB per-capita, Percentual de pessoas com baixa renda e X_4 = Percentual da população servida por rede de Abastecimento de água foram significativas para este modelo ao nível de 5%. Somente a variável Proporção da população servida por coleta de lixo não apresentou contribuição significativa na TMI.

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 0,7998 e o R^2 ajustado foi de 0,7698; mostrando bom ajuste.

A figura 5.1, mostra que há desvios de normalidade nas extremidades (caudas pesadas). Para confirmar se há normalidade dos resíduos aplicou-se o teste Kolmogorov-Smirnov cujo p-

valor = 0,5797. Logo, não temos evidências estatísticas para rejeitar a hipótese de normalidade dos dados, portanto admite-se distribuição normal para os resíduos do modelo OLS Pooled.

Figura 5.1: Gráfico dos Resíduos para o Modelo OLS pooled



A tabela (5.3) 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 per capita e percentual servida por coleta de lixo, não apresentaram evidência estatística significativa. Porém, 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 25,55%, a proporção de pessoas com baixa renda se reduz em 34,24% e o abastecimento de água em 12,64%. Neste modelo, R^2 foi igual a 0,7380.

Tabela 5.3: Resultado da Regressão Linear Múltipla do Modelo de Efeito Fixo

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Taxa de Analfabetismo	0,2555	$1,2000 \times 10^3$	2.1306×10^{-1}	0.0350 *
PIB per capita	$4,7765 \times 10^{-5}$	$5,3000 \times 10^{-1}$	0.9014	0.3691
Proporção de Pessoa com baixa renda	0,3424	$4,2500 \times 10^2$	80.521	4.5130×10^{-13} ***
Proporção da População servida por rede de abastecimento de água	0,1264	$4,8500 \times 10^2$	26.057	1.0240×10^{-2} *
Proporção da População servida por coleta de lixo	- 0,0364	$6,0900 \times 10^2$	-0.5978	0.5511
R^2	0,7380		R^2 Ajustado	0,59219
Estatística F	73,2209		p-valor	$0,0222 \times 10^{-14}$

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

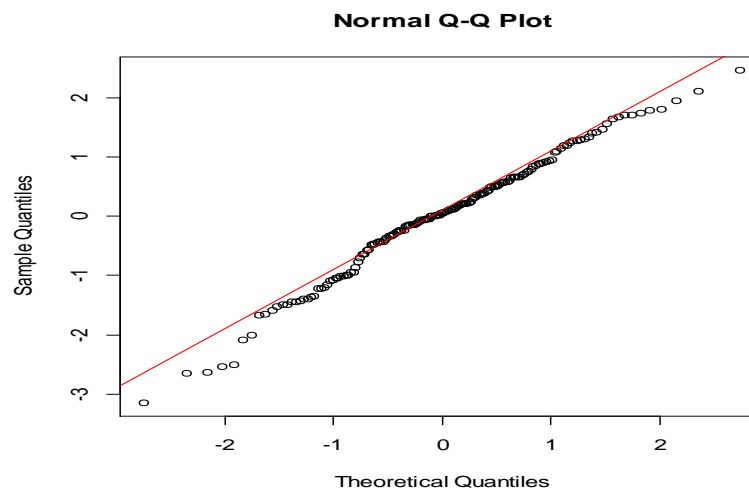
Figura 5.2: Gráfico de Normalidade do Efeito Fixo

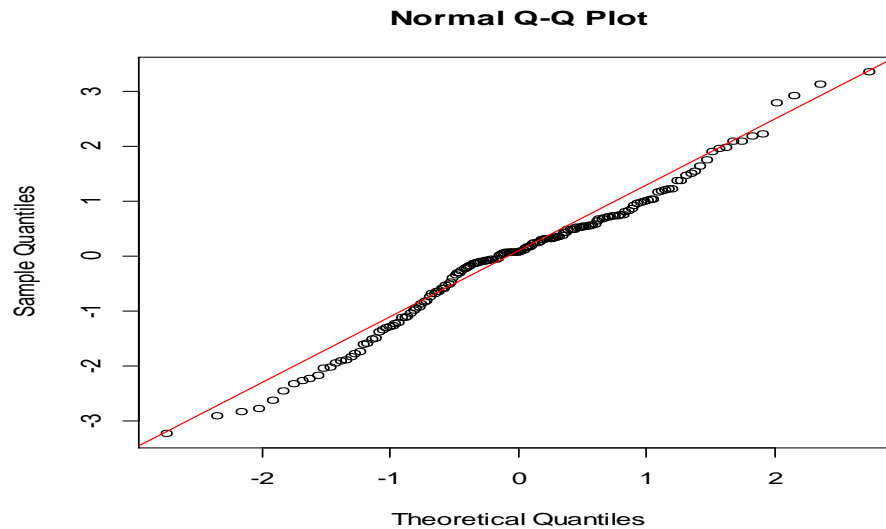
Tabela 5.4: Resultado da Regressão Linear Múltipla do modelo de Efeito Aleatório

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Intercepto	5,9149	4,9104	1,2046	0,2302
Taxa de Analfabetismo	$-1,6521 \times 10^{-1}$	$8,2347 \times 10^{-2}$	- 2,0063	0,0466 *
PIB per capita	$2,2580 \times 10^{-5}$	$4,6133 \times 10^{-5}$	0,4895	0,6252
Proporção de Pessoa com baixa renda	$3,6210 \times 10^{-1}$	$3,5006 \times 10^{-2}$	1.03437×10^{-1}	$2,0000 \times 10^{-16}$ ***
Proporção da População servida por rede de abastecimento de água	$-2,2103 \times 10^{-2}$	$3,2027 \times 10^{-2}$	-0,6901	0,4911
Proporção da População servida por coleta de lixo	$7,1262 \times 10^{-3}$	$5,0725 \times 10^{-2}$	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: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

A figura 5.2, mostra que tem distorções nos resíduos. Logo, para confirmar se há normalidade, aplicou-se o teste Kolmogorov-Smirnov, com p-valor = 0.0408, logo, temos evidências estatísticas para rejeitar a hipótese de normalidade dos resíduos deste modelo, portanto não há normalidade dos resíduos no modelo de efeito fixo. A tabela (5.4) apresenta a estimação para o modelo de efeito aleatório.

Para o modelo com efeito aleatório, as variáveis PIB, abastecimento de água, coleta de lixo não foram significativos para o modelo. Logo, a variável de percentual 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. Este modelo apresenta um $R^2 = 0,7076$.

Figura 5.3: Gráfico de Normalidade do Efeito Aleatório

A figura (5.3), mostra que há distorções nos resíduos. Logo, para 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. Em todos os modelos estimados as variáveis taxa de analfabetismo e renda apresentaram-se significância ao nível de 5%.

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 (5.5) 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 de acordo com a figura (4.1) decide-se pelo modelo de efeito fixo. As soluções dos testes estão na tabela (5.5) seguinte:

Tabela 5.5: Resultados dos testes de Breusch-Pagan e Hausman

Teste de Breusch-Pagan	Teste de Hausman
124,323	35,609
$2,200 \times 10^{-16}$	$1,137 \times 10^{-6}$

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

Com base nos resultados obtidos dos modelos estimados e dos teste executados o modelo de dados em painel mais adequado para explicar a redução da TMI das UFs foi o modelo de efeitos fixos. Assim sendo, o percentual 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 proporciona uma educação de boa qualidade.

A tabela (5.6) apresenta o resultado para o teste do modelo de efeitos individuais dos estados. Considerando o modelo com as variáveis originais não apresentaram 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 (5.6), onde se apresenta os efeitos individuais dos estados.

Tabela 5.6: Resultado do Teste de Efeito Individual dos Estados

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Z_1	-0,1127	0,0487	-2,3161	0,0221 *
Z_2	- 0,2024	0,0518	-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	0,64812		R^2 Ajustado	0,5201
Estatística F	47,889		p-valor	$0,0222 \times 10^{-14}$

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

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: X1 (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 chamada 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 = ao aumentar uma unidade no PIB per capita esta variável provoca uma diminuição 0,20% na TMI.

5.3 MODELOS DE REGRESSÃO BINOMIAL NEGATIVA E REGRESSÃO DE POISSON

Nessa seção apresenta-se a aplicação dos dados em painel utilizando modelos de regressão binomial negativa e Poisson sobre um conjunto de variáveis demográfica, sócio econômica e de cobertura.

A tabela (5.7) apresenta os resultados a Regressão Binomial Negativa do modelo de efeito fixo, Taxa de analfabetismo, PIB per capita, Percentual da população servida por rede de Abastecimento de água Proporção da população servida por coleta de lixo, não apresentaram evidência estatística significativa. Porém, o intercepto e Percentual de pessoas com baixa renda, indicam evidências estatísticas para a redução da TMI com 5% de significância. Dessa forma, quando o percentual de pessoas com baixa renda diminui com 5%, então: a TMI se reduz com 2,05%.

Tabela 5.7: Resultado da Regressão Binomial Negativo do Modelo de Efeito Fixo

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Intercepto	33,66	1,287	26,165	$0,0002 \times 10^{-12}$ ***
Taxa de Analfabetismo	$0,1428 \times 10^{-1}$	0,0251	0,5690	0.5694
PIB per capita	$0,1163 \times 10^{-5}$	$0,1042 \times 10^{-4}$	0,1120	0.9111
Proporção de Pessoa com baixa renda	$0,1928 \times 10^{-1}$	$0,8323 \times 10^{-2}$	2,3160	0.0205 *
Proporção da População servida por rede de abastecimento de água	$0,8028 \times 10^{-2}$	$0,1114 \times 10^{-1}$	0,7210	0.4710
Taxa de Analfabetismo	$0,6342 \times 10^{-2}$	$0,1126 \times 10^{-1}$	0,5630	0.5732

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

A tabela (5.8) apresenta os resultados a Para a Regressão Poisson do modelo Pooled, PIB per capita, Percentual da população servida por rede de Abastecimento de água e Proporção da população servida por coleta de lixo, não apresentaram evidência estatística significativa. Porém, o intercepto, Taxa de analfabetismo e Percentual de pessoas com baixa renda, 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 0,32%. A Tabela (5.9) apresenta os resultados a Para a Regressão Poisson do modelo de Efeito Fixo, logo, nesse modelo nenhuma das variáveis apresentaram evidência estatística significativa. Portanto, esse modelo não conseguiu convergir para as variáveis em estudo.

Tabela 5.8: Resultado da Regressão Poisson do Modelo Pooled

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
Intercepto	2.749	0,2979	9,2260	$0,0002 \times 10^{-12}***$
Taxa de Analfabetismo	$-0,1603 \times 10^{-1}$	$0,5444 \times 10^{-2}$	-2,9440	0,0032 **
PIB per capita	$-0,7623 \times 10^{-5}$	$0,4156 \times 10^{-5}$	-1,8340	0,0666
Proporção de Pessoa com baixa renda	$0,1562 \times 10^{-1}$	$0,2672 \times 10^{-2}$	5,8460	$0,0503 \times 10^{-7}***$
Proporção da População servida por rede de abastecimento de água	$-0,2987 \times 10^{-2}$	$0,1760 \times 10^{-2}$	-1,6970	0,0897
Proporção da População servida por coleta de lixo	$0,1456 \times 10^{-3}$	$0,3202 \times 10^{-2}$	0,0450	0,9637

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE)

A tabela (5.9) apresenta os resultados a Para a Regressão Poisson do modelo de Efeito Fixo, logo, nesse modelo nenhuma das variáveis apresentaram evidência estatística significativa. Portanto, esse modelo não conseguiu convergir para as variáveis em estudo.

Tabela 5.9: Resultado da Regressão Poisson do Modelo de Efeito Fixo

Variáveis	Estimativa	Erro Padrão	t-Student	P-valor
X_1	$0,8579 \times 10^{-2}$	$0,2185 \times 10^{-1}$	0,3930	0.6946
X_2	$-0,3239 \times 10^{-5}$	$0,1192 \times 10^{-4}$	-0,2720	0.7859
X_3	$0,1538 \times 10^{-1}$	$0,8320 \times 10^{-2}$	1,8490	0.0645
X_4	$0,5067 \times 10^{-2}$	$0,8871 \times 10^{-2}$	0,5710	0.5679
X_5	$0,5865 \times 10^{-4}$	$0,1120 \times 10^{-1}$	0,0050	0.9958

Fonte: Elaboração própria, com base nos dados do Ministério da Saúde (MS), Sistema de Informação sobre Mortalidade (SIM), Sistema de Informações sobre Nascidos Vivos (SINASC) e o Instituto Brasileiro de Geografia e Estatística (IBGE).

6 CONSIDERAÇÕES FINAIS

A utilização dos modelos de Dados em Painel que mais adéqua a investigação do conjunto de variáveis sócio econômica e de cobertura que influenciam na TMI a tomada de decisão no contexto dessa problemática. Ao comparar os modelos de Dados em Painel com Regressão Linear Múltipla de Efeito Fixo e Regressão de Poisson com modelo Pooled, observou-se o modelo de Regressão Linear com Efeito Fixo obteveram-se um melhor ajuste nas variáveis em estudo.

Dentro os modelos analisados são: Regressão Linear Múltipla (OLS Pooled, efeito fixo e aleatório), Regressão Binomial Negativa (OLS Pooled, efeito fixo e aleatório) e Regressão Poisson (OLS Pooled, efeito fixo e aleatório) decidiu-se pelo modelo Regressão Linear Múltipla 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.

O modelo Regressão Linear Múltipla de efeito fixo não possui resíduos normais e tornou-se necessário uma transformação logarítma no modelo nas variáveis dependente e independentes. Este modelo transformado apresentou um melhor ajuste e resíduos com distribuição normal.

A contribuição deste trabalho foi mostrar os passos necessários para a análise de DP com Regressão Linear e Não Linear 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 numa maior distribuição de renda.

REFERÊNCIAS

- AIDIL Jesus da Silveira Barros, NEIDE Aparecida de Souza Lehfeld. Fundamentos de metodologia científica { 3a Ed. { São Paulo: Pearson Prentice Hall, 2007.
- 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.
- ANTONOVSKI, A. & BERNSTEIN, J. Social class and infant mortality. Soc. Sci. Med., 11:453-70, 1977.
- ARAUJO, José Duarte de. A mortalidade infantil no estado da Bahia, Brasil. Rev. Saúde Pública [online]. 1973, vol.7, n.1, pp.29-36. ISSN 0034-8910. <http://dx.doi.org/10.1590/S0034-89101973000100003>.
- BAROSSO Filho, M. e Braga, M. Bobik, Metodologia da Econometria IN: Vasconcellos et al, Manual de Econometria, Editora Atlas, São Paulo, 2000.
- BRASIL, 1982. O Subsistema de Informações sobre Mortalidade. In: Estatísticas de Mortalidade: Brasil, 1979 (Ministério da Saúde, org.), pp. 2-56, Brasília: Centro de Documentação, MS.
- BRASS W. Uses of census or survey data for the estimation of vital rates. 1964. Paper prepared for the African Seminar on Vital Statistics, 14{19 December, 1964. Addis Ababa.
- BRASS W. Methods for estimating fertility and mortality from limited and defective data. Chapel Hill, North Carolina: Carolina Population Center, Laboratories for Population Statistics; 1975.
- BREUSCH, T. & Pagan, A. (1980), 'The lagrange multiplier test and its applications to model specification in econometrics', *Review of Economic Studies* **47**(1), 239-253.
- CASTRO, P.M.G.; CERGOLE, M.C.; CARNEIRO, M.H.; MUCINHATO, C.M.D.; SERVO; G.J.M. 2002 Crescimento, mortalidade e taxa de exploração do goete, *Cynoscion jamaicensis* (PERCIFORMES: SCIAENIDAE), na região Sudeste/Sul do Brasil. Boletim do Instituto de Pesca, 28(2): 141-153.
- CAMERON, Adrian Colin, TRIVEDI, Pravin K. Regression analysis of count data. Cambridge, UK; New York, NY, USA: Cambridge University Press. 1998.
- CAMERON, A. C.; TRIVEDI, P. K. Microeconometrics using Stata. College Station: Stata CAMPOS, Tatiana P.; SA CARVALHO, M.; BARCELLOS, Christovam C. Mortalidade infantil no Rio de Janeiro, Brasil: _areas de risco e trajet_oria dos pacientes at_e os servi_cos de sa_ude. PAN AMERICAN JOURNAL OF PUBLIC HEALTH, v. 8, p. 164-171, 2000.
- CAMERON, A.C; TRIVEDI,P.K. Microeconometrics Using Stata. Stata Press, 2009.

CAMPOS, TATIANA P.; SA CARVALHO, M.; BARCELLOS, Christovam C. Mortalidade infantil no Rio de Janeiro, Brasil: áreas de risco e trajetória dos pacientes até os serviços de saúde. PAN AMERICAN JOURNAL OF PUBLIC HEALTH, v. 8, p. 164-171, 2000.

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.

CHIAVENATO, Idalberto. Introdução à teoria geral da administração. 8. ed. Rio de Janeiro: Elsevier, 2011.

COSTA, Maria da Conceição Nascimento et al. Mortalidade infantil no Brasil em períodos recentes de crise econômica. Revista de Saúde Pública, v. 37, n. 6, p. 699-706, 2003.

DATASUS - Departamento de Informática do Sistema Único de Saúde: Disponível em: <<http://datasus.saude.gov.br/>>.

DE ARAUJO, Jose Duarte. A mortalidade infantil no Estado da Bahia, Brasil. Revista de Saúde Pública, v. 7, n. 1, p. 29-36, 1973.

DEATON, Angus. Panel data from time series of cross-sections. Journal of econometrics, v. 30, n. 1, p. 109-126, 1985.

DUARTE, P.C.; LAMOUNIER, W.M.; TAKAMATSU, R.T. Modelos Econométricos para Dados em Painel: Aspectos Teóricos e Exemplos de Aplicação - a Pesquisa em Contabilidade e Finanças.

FREES, E. W. Longitudinal and panel data: analysis and applications in the social sciences. Cambridge: Cambridge University Press, 2004.

GIL, Antonio Carlos. Como elaborar projetos de pesquisa. São Paulo: Atlas, 1991.

GODOY, Marcia Regina Regulamentação dos planos de saúde e risco moral: aplicação da regressão quantílica para dados de contagem. Tese (Doutorado em Economia) - Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de PósGraduação em Economia, Porto Alegre, 2008.

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

GRIFFITHS, William E.; R. Carter HILL; e George G. JUDGE, *Learning and Practicing Econometrics*, John Wiley and Sons, New York, 1993.

HILL Carter; Willian Griffiths; George Judge, Livro Econometria, 1999.

HSIAO, C. Analysis of panel data. Cambridge: UP, 1999. (Econometric Society HSIAO, Cheng. Analysis of panel data. Cambridge: Cambridge University Press, 1986.

HSIAO, Cheng. Analysis of panel data. Cambridge: Cambridge University Press, 1986

HSIAO, C. **Analysis of panel data**. 2. ed. Cambridge: Cambridge University Press, 2003.

IBGE { Instituto Brasileiro de Geografia e Estatística: Disponível em:
<<http://www.ibge.gov.br/home/>>.

IBGE (Fundação Instituto Brasileiro de Geografia e Estatística), 1998. Estatísticas do Registro Civil, 1996. v. 23. Rio de Janeiro: IBGE.

IBGE (Fundação Instituto Brasileiro de Geografia e Estatística), 2001. População. Indicadores Sociais. <<http://www.ibge.gov.br>>.

JONES, Andrew. Applied Econometrics for Health Economists: a practical guide. Second Edition. OHE, 2007.

LAURENTI R. Análise da informação em saúde: 1893-1993, cem anos da Classificação Internacional de Doenças. Rev Saúde Pública 1991; 25:407-17.

MARTINS, Gilberto de Andrade. Manual para elaboração de monografias e dissertações. 2. ed. São Paulo: Atlas, 1994

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. {6a Ed. {São Paulo: Atlas, 2001. Metodologia do Trabalho Científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos.

MARQUES, L. D. Modelos dinâmicos com dados em painel: revisão da literatura. Série Working Papers do Centro de Estudos Macroeconômicos e Previsão (CEMPRE) da Faculdade de Economia do Porto, Portugal, n. 100, 2000.

MELLO JORGE, M. H. P. et al. O Sistema de Informações sobre Mortalidade: passado, presente e futuro. São Paulo: CBCD; 2006.

MELLO, Jorge Maria Helena Prado de Mello; GOTLIEB, Sabina Lea Davidson and LAURENTI, Ruy. O sistema de informações sobre mortalidade: problemas e propostas para o seu enfrentamento I - Mortes por causas naturais. Rev. bras. epidemiol. [online]. 2002, vol.5, n.2, pp. 197-211.

MELLO-JORGE, M. H. P., 1983. Sub-registro dos eventos vitais. Revista de Saúde Pública, 17:148-151.

MELLO JORGE, M. H. P. Investigação sobre a mortalidade por acidentes e violências na infância. São Paulo, 1988. [Tese de Livre-Docência Faculdade de Saúde Pública da USP].

MELLO-JORGE, M. H. P.; GOTLIEB, S. L. D. & OLIVEIRA, H., 1996. O Sistema de Informação sobre Nascidos Vivos: Primeira avaliação dos dados brasileiros. Informe Epidemiológico do SUS, 2:15-48. Monographs, n. 11).

MORTARA, Giorgio | Estimativa do número de óbitos e da taxa de mortalidade geral para o Brasil (1950). In: INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa sobre a mortalidade no Brasil: 2.a série. Rio de Janeiro, 1956. p. 56-7. (Estudos de Estatística Teórica e Aplicada. Estatística Demográfica, 20).

PAES, Neir Antunes. Qualidade das estatísticas de óbitos por causas desconhecidas dos Estados brasileiros. Rev. Saúde Pública [online]. 2007, vol.41, n.3, pp. 437.

PNUD/FAO/IBAMA/UFPB/GOV. PARA_IBA. Diagnóstico do setor florestal do Estado da Paraíba. João Pessoa: PNUD / FAO / BRA. 87- 007, 1994. 84 p. Press, 2009.

RAJARATNAM JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. PLoS Med. 2010;7(4):e1000253. doi: 10.1371/journal.pmed.1000253. [PMC free article] [PubMed] 32.

RUTSTEIN, D. D.; BERENBERG, W.; CHALMERS, T. C.; CHILD, C. G.; FISHMAN, A. P.; PERRIN, E. B. Measuring the quality of medical care: a clinical method. N. Engl. J. Med., 294: 582-8, 1976.

SANTOS SILVA, João M. C.; WINDMEIJER, Frank, 2001. Two-part multiple spell models for health care demand. Journal of Econometrics, Elsevier, vol. 104(1), pages 67-89, August, 2001.

SCHMIDT, C. M. C. Modelo de regressão de Poisson aplicado à área da saúde. 2003. 98f. Dissertação (Mestrado em Modelagem Matemática) – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí.

SCHRAMM, J. M. A. & SZWARCOWALD, C. L., 2000. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e a natimortalidade. Revista de Saúde Pública, 34:272-279. Spatial Panel data models in R.

SOUSA, T. R. V.; LEITE FILHO, P. A. M. Análise por dados em painel do status de saúde no Nordeste Brasileiro. Revista de Saúde Pública, v. 42, n. 5, p. 796-804, 2008.

STOCKWELL, E.G. Infant mortality and socioeconomic status: a changing relationship. Milbank mem. Fd Quart., 11:101-11, 1962.

Vallin J, Pollard JH, Heligman L, editors. (1984) Methodologies for the collection and analysis of mortality data. Ordina, Liege: International Union for the Scientific Study of Population.

VALLIN. J. & BEHM, H. Mortality differentials among human groups. [Paper presented to the Seminar on Biological and Social Aspects of Mortality and the Length of Life. International Union for the Scientific Study of Population. Fiuggi Terme, Italy, 1980].

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

Apêndice A - Script do Programa implementado no software R

```

#=====
#-----# Dados #-----#
#=====#

library(plm) ## pacote plm##
Maizza<-
read.csv("C:/Users/MAIZZAMICAELLE/Documents/Dissertação(orientadorJoãoAgnaldo)/Maiz
za3.txt",sep=";",header = T)
Maizza      ##arquivo do Banco de dados##
attach(Maizza)
View(Maizza)
names(Maizza)
length(names(Maizza)) # Tamanho
dim(Maizza)      # Linhas e Colunas
summary(Maizza)

#=====#

                        Aplicação dos Modelos

#=====#

####Teste de Chow##

gpooling<-plm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), model =
"pooling")
summary(gpooling)

qqnorm(gpooling$residuals)
abline(0.1,2.3,col="red")

### Teste Lilliefors###
library(nortest)
lillie.test(gpooling$residuals)

####Modelo de Efeito Fixo###

gfixed<-plm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), model =
"within")
summary(gfixed)

qqnorm(gfixed$residuals)

```

```
abline(0.1,1.0,col="red")
```

```
### Teste Lilliefors###
```

```
library(nortest)
```

```
lillie.test(gfixed$residuals)
```

```
### Modelo de Efeito Aleatório ###
```

```
grandom<-plm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), model = "random")
```

```
summary(grandom)
```

```
qqnorm(grandom$residuals)
```

```
abline(0.1,1.2,col="red")
```

```
## Teste Lilliefors ##
```

```
library(nortest)
```

```
lillie.test(gradon$residuals)
```

```
## Teste de Housman##: Ho: O efeito de modelo aleatório é apropriado x #H1:Modelo de efeito fixo é apropriado
```

```
phtest(gfixed, grandom)
```

```
### Teste Breusch Pagan ###
```

```
##Random effect and Pooling ##: Ho: modelo Pooled OLS model is apropriado versus H1: # Modelo aleatório é apropriado
```

```
plmtest(gpooling, type=c("bp"))
```

```
##Teste para saber se existe o efeito individual dos estados###
```

```
## Ho= O efeito individual não existe x H1 = Existe o efeito individual
```

```
grandom<-plm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), model = "random", effect=c("individual"))
```

```
summary(grandom)
```

```
#=====
```

Transformação Logarítma dos Resíduos do Modelo de Efeito Fixo

```
#=====
```

```
Maizza<-
```

```
read.csv("C:/Users/MAIZZAMICAELLE/Documents/Dissertação(orientadorJoãoAgnaldo)/Maizza3.txt", sep=";", header = T)
```

```
Maizza ##arquivo do Banco de dados##
```

```
attach(Maizza)
```

```

names(Maizza)
length(names(Maizza)) # Tamanho
dim(Maizza)          # Linhas e Colunas
Estado= Maizza$Estado
Ano = Maizza$Ano
log(X1)
Z1=log(X1)
Z1
Z2=log(X2)
Z2
Z3=log(X3)
Z3
Z4=log(X4)
Z4
Z5=log(X5)
Z5
Z=log(Y)
Z

```

```

#=====#
# Teste para avaliar o efeito das variáveis com relação ao fator tempo (ano) #
#=====#

```

```

library(plm)
library(foreign)

Maizza<-read.csv2("D:/Maizza1.csv", header=T)
Ano<-Maizza$Ano
Ano<-as.factor(Ano)
Estado<-Maizza$Estado
Estado<-as.factor(Estado)
y<-Maizza$Y
x1<-Maizza$X1
x2<-Maizza$X2
x3<-Maizza$X3
x4<-Maizza$X4
x5<-Maizza$X5
fixed.dum<-lm(y~x1+x2+x3+x4+x5+factor(Ano)-1, data=Maizza)
fixef(fixed)
pFtest(fixed,ols)
plmtest(fixed,c("Ano"),type=("bp"))

```

```
#=====#
#Teste para avaliar o efeito das variáveis com relação ao fator (Estados)#
#=====#
```

```
library(plm)
library(foreign)

Maizza<-read.csv2("D:/Maizza1.csv", header=T)
Ano<-Maizza$Ano
Ano<-as.factor(Ano)
Estado<-Maizza$Estado
Estado<-as.factor(Estado)
y<-Maizza$Y
x1<-Maizza$X1
x2<-Maizza$X2
x3<-Maizza$X3
x4<-Maizza$X4
x5<-Maizza$X5
fixed.dum<-lm(y~x1+x2+x3+x4+x5+factor(Estado)-1, data=Maizza)
fixef(fixed)
pFtest(fixed,ols)
plmtest(fixed,c("Ano"),type=("bp"))
```

```
#=====#
```

Aplicação dos Modelo Binomial Negativa

```
#=====#
```

```
# binomial negativa
# modelo within (dentro)

bin1<-pglm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), family =
negbin, model = "within", print.level=3, method="nr")
summary(bin1)
```

```
#=====
```

Aplicação do Modelo Poisson

```
#=====
```

```
# poisson
```

```
# modelo pooling
```

```
poi1<-pglm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), family =  
poisson, model = "pooling", print.level=3, method="nr")
```

```
summary(poi1)
```

```
#=====
```

Aplicação do Modelo Aleatório (*Randon*)

```
#=====
```

```
#modelo random (aleatório)
```

```
poi2<-pglm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), family =  
poisson, model = "random", print.level=3, method="nr")
```

```
summary(poi2)
```

```
#=====
```

Aplicação do Modelo Within (Efeito Fixo)

```
#=====
```

```
#modelo within (dentro)
```

```
poi4<-pglm(Y~X1 + X2 + X3 + X4 + X5, data = Maizza, index= c("Estado","Ano"), family =  
poisson, model = "within", print.level=3, method="nr")
```

```
summary(poi4)
```