

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
PROGRAMA DE PÓS-GRADUAÇÃO EM DEMOGRAFIA

LOURDES MILAGROS MENDOZA VILLAVICENCIO

**Fecundidade do Nordeste Brasileiro: uma abordagem com Modelos Não Linear de  
Efeitos Mistos**

**NATAL/RN**

**2015**

**LOURDES MILAGROS MENDOZA VILLAVICENCIO**

**Fecundidade do Nordeste Brasileiro: uma abordagem com Modelos Não Linear de Efeitos Mistos**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Demografia da Universidade Federal do Rio Grande do Norte, como requisito parcial para a obtenção do título de Mestre em Demografia.

Orientadora: Maria Helena Constantino Spyrides

Co-Orientadora: Lara de Melo Barbosa

Natal  
2015

Catálogo da Publicação na Fonte. UFRN / SISBI / Biblioteca Setorial  
Centro de Ciências Exatas e da Terra – CCET.

Villavicencio, Lourdes Milagros Mendoza.

Fecundidade do nordeste brasileiro: uma abordagem com modelos não linear de efeitos mistos / Lourdes Milagros Mendoza Villavicencio. - Natal, 2015.  
105 f.: il.

Orientadora: Maria Helena Constantino Spyrides.

Coorientadora: Lara de Melo Barbosa.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro de Ciências Exatas e da Terra. Programa de Pós-Graduação em Demografia.

1. Fecundidade – Taxas específicas – Dissertação. 2. Estimativas – Dissertação.  
3. Efeitos mistos – Dissertação. 4. Regressão não linear – Dissertação. I. Spyrides, Maria Helena Constantino. II. Barbosa, Lara de Melo. III. Título.

RN/UF/BSE-CCET

CDU: 618.179

## FOLHA DE APROVAÇÃO

**LOURDES MILAGROS MENDOZA VILLAVICENCIO**

**Fecundidade do Nordeste Brasileiro: uma abordagem com Modelos Não Linear de Efeitos Mistos**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Demografia da Universidade Federal do Rio Grande do Norte, como requisito parcial para a obtenção do título de Mestre em Demografia.

Resultado: \_\_\_\_\_ em: \_\_\_\_/\_\_\_\_/\_\_\_\_

### BANCA EXAMINADORA

---

Profª. Dra. Maria Helena Constantino Spyrides /UFRN

(Orientadora)

---

Profª. Dra. Lára de Melo Barbosa Andrade / UFRN

(Co-orientadora)

---

Prof. Dr. Idemauro Antonio Rodrigues de Lara/USP

(Convidado)

---

Prof. Dr. Moises Calle Aguirre/UFRN

Convidado)

Natal / RN

2015

*Aos meu pais,  
**Cristobal Mendoza e  
Martha Villavicencio**  
pelo amor e confiança em todos os momentos.*

## **AGRADECIMENTOS**

A Jesus meu Senhor, Pai e Salvador, reconheço que se não fosse por ele, nada seria. Por sempre me fazer sentir que eu não estava sozinha, por me dar saúde, perseverança para a concretização de mais um objetivo de vida.

A meus pais, Martha e Cristobal, por estarem sempre perto com aquelas chamadas longas, dando-me ânimos para seguir a caminhada e por serem para mim exemplo de coragem, determinação e esforço. Tudo o que sou é resultado do esforço de vocês. Os erros, claro, ficam por conta das minhas escolhas. Muito Obrigada!

A meu eterno namorado, Felipe Ferreira, por manter-se incessantemente ao meu lado dando-me seus abraços nos momentos mais difíceis. E me proporcionando momentos divertidos que sempre fizeram com que meus dias fossem mais agradáveis.

A meu irmão David, que foi outro pai para mim, e minha motivação a seguir os caminhos acadêmicos.

A meu irmão Rene, que de forma direta ou indireta ensinou-me ser uma pessoa determinada.

A minha irmã Sonia, freira por vocação, sempre me direcionando na vida e orando por minha vida.

À família, Lopes, por que sem me conhecer, quando cheguei neste país, me mostraram sua amabilidade e apoio, obrigada por ser uma família para mim neste país, sinto que jamais terminarei de expressar a minha gratidão.

À família Emmerich, outra família mais que ficam em meu coração para sempre, lembro vocês sendo a testemunha de minha aceitação ao mestrado, pulando todos da felicidade. Pastor George e Jane Obrigada sempre por me cobrir com suas orações.

A minha orientadora e co-orientadora, Maria Helena e Lara, pela amizade, oportunidade, paciência, aprendizado e orientação. Nunca serei capaz de traduzir em palavras a minha gratidão, o meu respeito e a minha admiração por vocês. Ambas muito competentes no saber. Obrigada doutoras por impulsionar o meu crescimento na pesquisa científica.

À Banca, pelas valiosas sugestões que aperfeiçoaram a Dissertação.

À Dra. Julia Toledo, pela enorme amizade, generosidade e humanidade.

Ao Dr. Fidel, Moises, Flavio, Ricardo, pela amizade.

À Dra. Maria Célia Formiga, pela generosidade e amizade.

A meus colegas do CCET: Felipe Azevedo, Pollyanne, Izabelle, Kalline, Isadora, Thiago Nascimento, Eliana Mezquita, Soledad Jaimes, Josivan, Lariça, Priscila, pela troca de experiências e os momentos hilários.

A meu irmão e amigão Guilherme Schott, pela enorme humanidade, pelas palavras de fé, que me ajudaram naqueles momentos difíceis na minha vida, sem dúvida Deus mandou você no momento certo.

A Denise Wingerter, pelas dicas na caminhada antes e durante o Mestrado e pela ajuda na obtenção de dados para a realização deste trabalho. Sua pesquisa foi inspiradora para mim.

A Mario Vinicius por sempre ajudar-me nos pedidos com muita competência, bom humor, carisma e vontade.

A meus irmãos da Igreja Metodista em Natal, em especial, Regina Paiva, obrigada pelas ligações e as conversas longas sem fim, fazendo manter minha fé sempre em pé, Joyce e Ricardo, obrigada pelas orações e cobrir minha vida com elas, Pollyanna Lopes obrigada pelo abrigo quando precisei, Darllan Silveira e pais, obrigada pela ajuda no início de tudo, Andreza Fernanda, obrigada pelo tempo e disposição cada que precisava desabafar, Caio Alcântara e Giana Momo, obrigada pelo carinho, amizade, apoio e as caronas hehe!

A meus *Hermanos da Iglesia Metodista de Pueblo Libre*, em especial Iris, Janeth Cachay, pela amizade e carinho.

Ao Pastor Eduardo e sua esposa Anahí, porque desde quando falei para eles deste objetivo, estiveram orando por mim.

Aos meus amigos distantes e cúmplices: Bill (UNMSM), Yessica (UNMSM), Mirtha (UNMSM), Danilo (UNMSM), Edith (UNMSM), Nivea (UFMG), Daniel de Almeida (UNICAMP), Claudia (UNICAMP), Marco (USP), José Pepe (UFSCAR), que pelas palavras ajudaram a tornar esta dura caminhada mais agradável.

Aos meus amigos distantes da escola de freiras e da escola Pachacutec: Elvia, Erika, Jhon, Sara, Cristhian, Vivi, Luz, Paola, Lau, Anali, Raúl, Pablo, Ada, Gabriela, Fiorella, Rosa, Samantha, etc, pelas minhas noites de solidão fazendo estas engraçadas com as piadas da net, fazendo-me sentir eles perto a mim.

À CAPES pela bolsa concedida durante o curso do Mestrado.

Os agradecimentos são extensivos a todos que de uma forma ou de outra colaboraram para finalização deste trabalho, e a todos que acreditaram em mim e fazem parte da minha vida, o meu

**MUITO OBRIGADA!!!**

*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém  
ainda pensou sobre aquilo que todo mundo vê.”*

*(Arthur Schopenhauer)*

*" Para tudo há um tempo, para cada coisa há um momento debaixo dos céus.  
tempo de nascer, e tempo de morrer;  
tempo de plantar, e tempo de colher;  
tempo de matar, e tempo de curar;  
tempo de derrubar, e tempo de edificar;  
tempo de chorar, e tempo de rir;  
tempo de prantear, e tempo de dançar;  
tempo de espalhar pedras, e tempo de ajuntar pedras;  
tempo de abraçar, e tempo de afastar-se de abraçar;  
tempo de buscar, e tempo de perder;  
tempo de guardar, e tempo de lançar fora;  
tempo de rasgar, e tempo de coser;  
tempo de estar calado, e tempo de falar;  
tempo de amar, e tempo de odiar;  
tempo de guerra, e tempo de paz",  
  
Mas sempre é tempo de amar.*

## **RESUMO**

O objetivo deste trabalho consiste em avaliar a aplicação dos modelos não lineares de efeitos mistos (NLME) para a análise dos níveis e padrões da fecundidade, no âmbito dos estados do Nordeste Brasileiro. Além disso, realiza-se uma análise comparativa do comportamento da fecundidade levando em conta todos os estados para dois momentos no tempo, 2000 e 2010. Para a avaliação da aplicação deste modelo (NLME) foram utilizadas duas fontes de dados, as informações sobre filhos tidos por mulher durante seu período reprodutivo nas faixas etárias de 15 a 45 anos de idade, provenientes do Sistema de Informação de Nascidos Vivos (SINASC) e dos Censos Demográficos realizados em 2000 e 2010. Realizou-se a análise obtendo-se as estimativas de três parâmetros para um modelo assintótico, Taxa de Fecundidade Total (TFT), Taxa Específica no início do período reprodutivo e a idade com que a mulher atinge a metade da sua TFT, cuja interpretação reflete o comportamento da fecundidade das mulheres em idade reprodutiva. Para a estimação dos parâmetros, utilizou-se o método de máxima verossimilhança (MV) já implementado no software R. Os resultados encontrados, indicam que as estimativas obtidas por meio da fonte de dados do SINASC e Censo Demográfico apontaram o estado de Alagoas, em 2000, como aquele com a maior Taxa de Fecundidade Total (TFT) do Nordeste. Os estados do Maranhão e Piauí foram aqueles que atingiram a metade da Taxa de Fecundidade Total (TFT) com a idade de 21 anos. Para o ano 2010, as duas fontes de dados apontaram os estados do Maranhão e de Alagoas como aqueles com a maior TFT, em comparação com outros estados do Nordeste. Constatou-se que em 2010, a população feminina em todos os estados do Nordeste atinge a metade da TFT entre as idades de 20 e 24 anos. Observou-se também que a postergação das mulheres em terem filhos acontece de forma independente do estado de origem. O modelo NLME, traz um novo parâmetro para o comportamento reprodutivo da mulher, pode-se concluir que, os modelos não lineares de efeitos mistos (NLME) tornam-se uma ferramenta promissora dentro do campo da Demografia, permitindo analisar os níveis e padrões da fecundidade da região, levando a um entendimento da realidade e facilitando um diagnóstico para o planejamento de políticas públicas.

**Palavras-chave:** Taxas Específicas de Fecundidade, Estimativas, Efeitos Mistos, Regressão Não Linear.

## **ABSTRACT**

The objective of this study consists in evaluating application of non-linear mixed models (NLME) for the analysis of the levels and patterns of fertility, at the state level of the Brazilian Northeast. Moreover, we make a comparative analysis of fertility behavior taking into account all the states for two points in time, 2000 and 2010. To evaluate the application of this model (NLME) were used two sources of data, information on children born per woman during her reproductive period in the age groups 15-45 years of age, from the Sistema de Informação de Nascidos Vivos (SINASC) and Demographic Census conducted in 2000 and 2010. was performed the analysis obtaining estimates of three parameters for an asymptotic model, Total Fertility Rate (TFR), rate Specifies the beginning of the reproductive period and the age at which women reaches half of its TFT, whose interpretation reflects the behavior of the fertility of women of reproductive age. For the estimation, we used the maximum likelihood (ML) through the R. The results indicate that the estimates obtained by SINASC and Demographic Census showed the state of Alagoas, in 2000, as one with the highest Total Fertility Rate (TFR) of the Northeast. The states of Maranhão and Piauí were those who reached half the Total Fertility Rate (TFR) at the age of 21 years. For the year 2010, the two data sources indicated the states of Maranhão and Alagoas as those with the largest TFT, compared to other states of the Northeast. Found to be in 2010, the female population in all states of the Northeast reaches half the TFT between the ages of 20 and 24, It was also noted that the postponement of the women to have children happens independently of the home state. The NLME model brings a new parameter for the reproductive behavior of women, it can be concluded that non-linear mixed models (NLME) become a promising tool in the field of demography, allowing analyze levels and patterns of fertility in the region, leading to an understanding of reality and easing a diagnosis for planning public policy.

**Keywords:** Specific Fertility Rates, estimates, Mixed Effects, Nonlinear Regression.

## **LISTA DE ABREVIATURAS**

AL: Alagoas

BA: Bahia

CE: Ceará

MA: Maranhão

MS: Ministério da Saúde

NE: Nordeste

NLME: Modelo Não Linear Misto

PB: Paraíba

PE: Pernambuco

PI: Piauí

RN: Rio Grande do Norte

SE: Sergipe

SINASC: Sistema de Informações sobre Nascidos Vivos

TEF: Taxa de Fecundidade Específica

TFT: Taxa de Fecundidade Total

TEFA: Taxa de Fecundidade Específica Acumulada.

## LISTA DE TABELAS

Tabela 1 - Funções de variância contidas na classe varFunc da biblioteca NLME do software R.....	62
Tabela 2 - Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - SINASC, 2000.....	72
Tabela 3 - Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - Censo Demográfico, 2000.....	73
Tabela 4- Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - SINASC, 2010.....	73
Tabela 5- Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - Censo Demográfico, 2010.....	74
Tabela 6- Teste de Normalidade de Shapiro Wilk, segundo fontes de dados, estados da Região Nordeste, 2000 e 2010.....	76
Tabela 7 - Parâmetros estimados e erros-padrão do modelo com efeitos fixos e com efeitos aleatórios, para os Estados da Região Nordeste, SINASC 2000.....	80
Tabela 8- Parâmetros estimados e erros-padrão do modelo com efeitos fixos e com efeitos aleatórios, para os Estados da Região Nordeste, Censo Demográfico 2000.....	80
Tabela 9- Parâmetros estimados e erros-padrão do modelo com efeitos fixos, para os Estados da Região Nordeste, SINASC 2010.....	81
Tabela 10- Parâmetros estimados e erros-padrão do modelo com efeitos fixos, para os Estados da Região Nordeste, Censo Demográfico 2010.....	81
Tabela 11- Parâmetros Não Linear (Asym), Modelo de Regressão Assintótica, por Estados do Nordeste-Brasil, SINASC e CENSO DEMOGRAFICO, 2010.....	82
Tabela 12- Taxa de Fecundidade Total e Variação Porcentual, segundo os Estados da Região Nordeste - SINASC, 2000 e 2010.....	87
Tabela 13- Taxa de Fecundidade Total e Variação Porcentual, segundo os Estados da Região Nordeste - CENSO DEMOGRAFICO, 2000 e 2010.....	88

## LISTA DE FIGURAS

Figura 1 - Taxas de fecundidade total (TFT), Brasil,1960-2010.....	17
Figura 2 - Taxas de fecundidade Total, segundo as Grandes Regiões - 1940-2010.....	18
Figura 3 - Taxas de Fecundidade Total, por estado do Nordeste - 1940-2010.....	19
Figura 4 - Efeito do gompit transformado numa distribuição de Fecundidade.....	26
Figura 5- Representação dos parâmetros da regressão assintótica .....	41
Figura 6 - (a) Modelo de regressão com coeficientes angulares diferentes , (b) modelo de regressão com coeficientes angulares constantes entre os estados do NEB, SINASC 2000.....	66
Figura 7- Avaliação dos resíduos do ajuste da reta de regressão do modelo linear, para os Estados do Nordeste Brasileiro, SINASC 2000.....	67
Figura 8 - Perfis das Taxas Específicas de Fecundidade Acumulada por Estados do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2000. ....	70
Figura 9 - Perfis das Taxas Específicas de Fecundidade Acumulada por Estados do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2010 .....	71
Figura 10 - Intervalos 95% de confiança dos parâmetros por estado do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2000 - 2010.....	75
Figura 11- Teste de normalidade dos resíduos para NLME e análise de resíduos versus valores ajustados - SINASC e Censo Demográfico 2000 .....	77
Figura 12- Teste de normalidade dos resíduos para NLME e análise de resíduos versus valores ajustados - SINASC e Censo Demográfico 2010 .....	78
Figura 13- Padrão das TFT Ajustado com Modelo Não linear de Efeitos Mistos considerando base nas TEFA da SINASC e Censo Demográfico, 2000-2010 .....	79

## SUMÁRIO

1	INTRODUÇÃO .....	16
2	REVISÃO DA LITERATURA.....	22
2.1	Métodos de estimação da fecundidade.....	22
2.1.1	Método P/F de Brass .....	23
2.1.2	Método Relacional de Gompertz (MRG).....	25
2.1.3	Modelo Hadwiger (1940).....	28
2.1.4	Modelo Gamma (1981).....	28
2.1.5	Modelo Beta (1981) .....	28
2.1.6	Modelo de erro ajustado (Adjusted Error Model –AEM) .....	29
3	DADOS E MÉTODOS .....	30
3.1	Dados.....	30
3.2	Métodos.....	32
3.2.1	Modelos de regressão linear .....	32
3.2.2	Modelo Não Linear .....	34
3.2.3	Modelos Lineares de Efeitos Mistos .....	37
3.2.4	Modelo Não Linear de Efeitos Mistos (NLME).....	39
3.2.4.1	Estimação de $\beta$ e b.....	46
3.2.4.2	Estimação de $\theta$ .....	49
3.2.4.2.1	Máxima Verossimilhança.....	49
3.2.4.2.2	Máxima Verossimilhança Restrita (MVR) .....	51
3.2.4.2.3	Aproximações numéricas para o método de estimação.....	52
3.2.4.2.4	Estruturas da matriz de variâncias e covariâncias.....	53
3.2.4.3	Método de Seleção de modelos e qualidade do ajuste .....	55
3.2.4.3.1	Critérios de Informação.....	55
3.2.4.3.2	Teste da Razão de Verossimilhança (TRV) para $\theta$ .....	57
3.2.4.3.3	Intervalos de confiança .....	57
3.2.4.3.4	Estratégia para seleção do modelo .....	58
3.2.4.4	Análise de Resíduos .....	59
3.2.4.4.1	Resíduo Marginal.....	60
3.2.4.4.2	Resíduo condicional.....	60

3.2.4.5	Função da variância.....	61
4	RESULTADOS.....	64
4.1	Análise dos Modelos.....	64
4.2	Análise da Fecundidade no Nordeste a partir do modelo NLME.....	82
5	CONCLUSÕES.....	91
	REFERÊNCIAS.....	94
	APÊNDICE A - Biblioteca NLME.....	102
	APÊNDICE B - Sintaxe do Software R para a ajuste de um NLME.....	104

## 1 INTRODUÇÃO

---

A fecundidade é um componente demográfico importante, pois afeta o crescimento demográfico de uma região e altera a estrutura etária da população. Considera-se que mudanças da estrutura etária numa população vão além dos aspectos demográficos e podem trazer impactos importantes nas políticas públicas relacionadas às áreas de educação, saúde e emprego.

Os níveis de fecundidade têm chamado atenção não somente aos estudiosos da demografia, mas também aos formuladores de políticas públicas, dadas suas implicações sociais como por exemplo, o envelhecimento populacional e a inevitável redução do volume populacional, diante da ausência de medidas que possam reverter este processo (MORGAN, 2003).

A tendência de redução dos níveis de fecundidade e mortalidade no País, chamado de transição demográfica (CARVALHO, 2004; BERQUÓ, CAVENAGUI, 2006) é considerada um fator determinante para a estrutura etária de uma sociedade, sendo de importância sua observação para análise das demandas sociais (SIMÕES, 2006; ARBACHE, 2011).

A partir do final dos anos setenta, o Brasil apresentou um declínio expressivo da fecundidade e esta mudança associada à queda da mortalidade, na qual o Brasil realiza umas das transições demográficas mais rápidas em comparação ao resto de países do mundo. (ARAUJO, 2000; CARVALHO 2004; WONG, 2006; ARABCHE, 2011; WINGERTER, 2013). A queda nos níveis da fecundidade trouxe consigo transformações no crescimento da população e na sua distribuição por idade (MOREIRA, 2001; HORTA, 2003; CARVALHO, 2004).

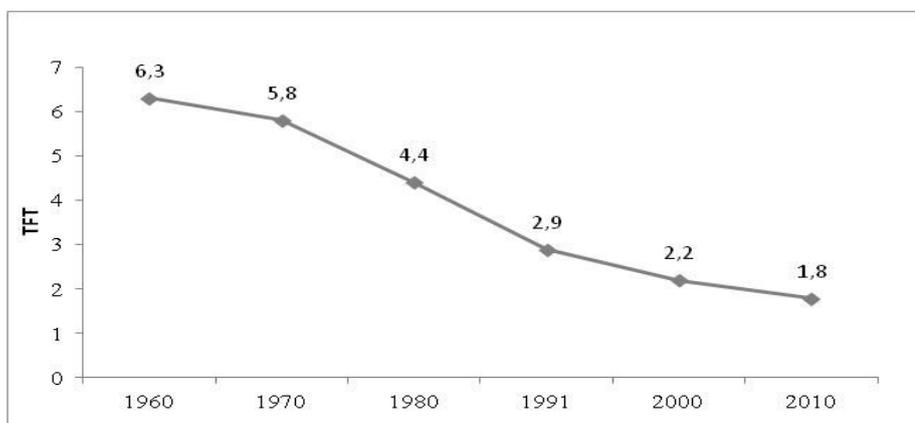
Como bem assinala Berquó (2001, p.28), mudanças demográficas significativas ocorreram no Brasil na segunda metade do século XX: "A mais importante foi a queda

acentuada da fecundidade, que teve impacto direto na desaceleração do crescimento da população, na sua estrutura etária [...] e na redução do tamanho das famílias".

Segundo dados dos Censos Demográficos realizados pelo IBGE, observou-se que o declínio da Taxa de Fecundidade Total (TFT)<sup>1</sup> para o Brasil foi reduzido na década de 1960, acelerou-se nas décadas de 1970 e 1980 e desacelerou ligeiramente nas últimas duas décadas. Entre os anos de 2000 e 2010, a TFT no Brasil apresentou uma queda de 20,1% e este declínio da fecundidade ocorreu em todas as regiões do País, sendo as maiores quedas no Nordeste (23,4%) e no Norte (21,8%) (IBGE,2012a; FRANÇA, 2013).

O País apresentou, em 2010, Taxa de Fecundidade Total de 1,86 filhos por mulher, sendo considerada abaixo do nível de reposição da população (2,1 filhos por mulher), conforme pode ser observado na Figura 1 (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2010).

Figura 1 - Taxas de fecundidade total (TFT), Brasil,1960-2010.



Fonte: Censos demográficos do IBGE

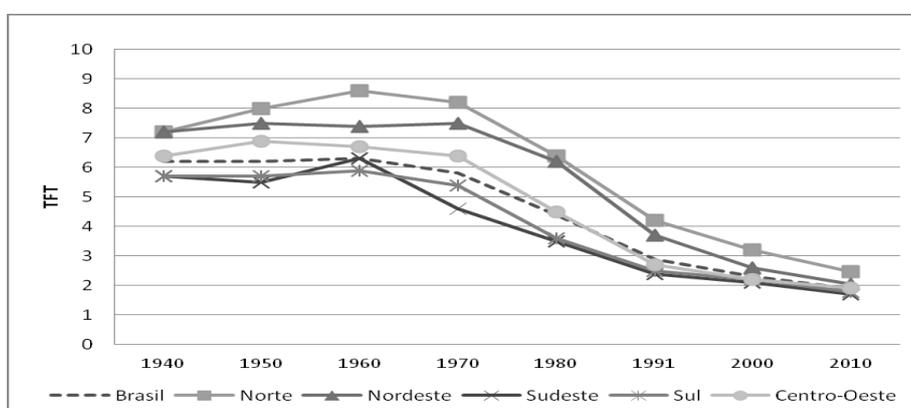
Berquó (2001, p.28) assinala que tais TFT refletem a regulação da fecundidade através de métodos modernos de contracepção, que conferiu maior autonomia ao

<sup>1</sup> A taxa de fecundidade total (TFT) é o principal determinante da dinâmica demográfica e expressa a situação reprodutiva de uma mulher pertencente a uma coorte hipotética, sujeita às taxas específicas de fecundidade (TEF) por idade, observadas na população em estudo, supondo-se a ausência de mortalidade nesta coorte (IBGE, 2009)

exercício da sexualidade e dos direitos reprodutivos, além de novas possibilidades de estilos de vida das diferentes camadas sociais.

O fenômeno da transição da fecundidade ocorreu em todas as regiões do Brasil, no entanto de forma heterogênea entre elas. Existem diferenças regionais quanto ao ritmo do declínio, mais evidentes nos estados das regiões Norte e Nordeste, nos quais as taxas de fecundidade totais são mais altas em comparação com as outras regiões (CARVALHO, 2004; WONG, 2006; GARCIA, 2012). No período 1980-2010 (Figura 2), a região Nordeste teve seu processo de transição iniciado com uma defasagem de aproximadamente 10 anos em relação às regiões Sul, Sudeste e Centro-Oeste, onde apresentaram a maior queda absoluta no nível da fecundidade observada no Brasil, mas, ainda assim, o Nordeste permanece como a segunda região com mais alta Taxa de Fecundidade Total (MOREIRA e FUSCO, 2012).

Figura 2 - Taxas de fecundidade total, segundo as Grandes Regiões - 1940-2010.

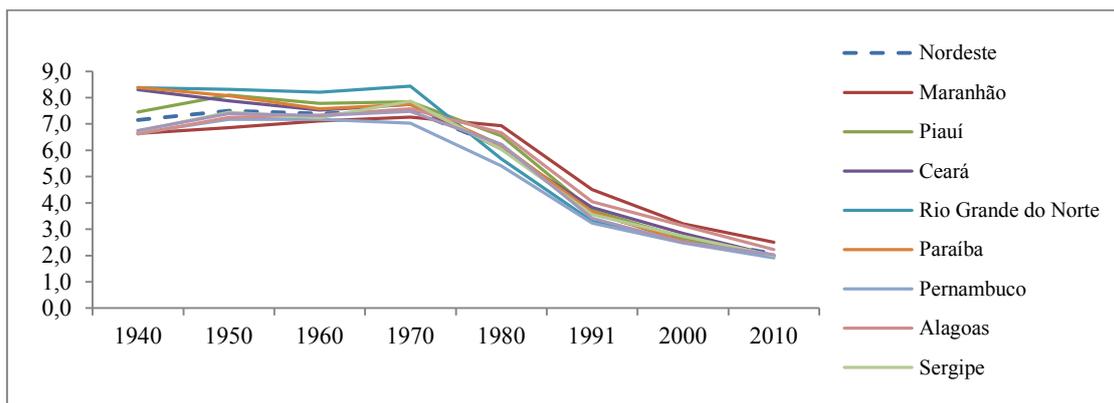


Fonte: IBGE, Censo Demográfico, 1940-2010

Nas três últimas décadas, a região Nordeste caracterizou-se, no que diz respeito aos níveis e tendências da fecundidade nacional, como o conjunto populacional cujas mulheres apresentam os maiores níveis de fecundidade, superados apenas por aqueles encontrados na região Norte. Entretanto, estas regiões apresentaram o declínio mais expressivo no período recente, com o Nordeste passando de 7,2 filhos por mulher em

1940 para 2,01 em 2010 (BARBOSA, FREIRE, 2004; INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2010)

Figura 3 - Taxas de Fecundidade Total, por estado do Nordeste - 1940-2010.



Fonte: IBGE, Censo Demográfico, 1940-2010

Segundo Barbosa (2006) e Wong (2006), existem diferenças no comportamento das taxas de fecundidade para os estados da Região Nordeste (Figura 3). Estas disparidades de valores entre as taxas de fecundidade interestaduais podem ser atribuídas a diferentes fatores, como por exemplo: diferenças entre o comportamento reprodutivo da mulher baseado na faixa etária materna (CARVALHO et al.,1994; BONIFÁCIO, 2011) e fatores sociais, econômicos e demográficos, nos quais observa-se que o segmento da população que detêm menor escolaridade e menores condições econômicas apresentam maiores taxas de fecundidade (JANUZZI, 2001; CARVALHO, BRITO, 2005; RODRIGUES, 2009; MUNIZ,2012), além da possibilidade da subnotificação em algumas localidades, o que impede a real captação das informações (FRIAS et al., 2007; JORGE et al., 2007). Aliás, quanto menores os níveis de desenvolvimento de uma localidade, piores são seus indicadores sociais e maiores são as taxas de fecundidade (CARDOSO et al.2005; WONG, 2006; PAES,SILVA, 2010; ARBACHE, 2011; MUNIZ, 2012; MIRANDA-RIBEIRO,GARCIA, 2012).

Miranda-Ribeiro (2007) comenta que a importância da busca pela medida ideal da fecundidade tem aumentado na medida em que este fenômeno alcança níveis cada

vez mais baixos, sendo importante compreender que a partir de pequenas mudanças com respeito ao declínio da fecundidade, novos desafios se apresentam.

Segundo Horta (2012), torna-se importante a divulgação contínua de levantamentos empíricos de mensuração para diagnóstico do comportamento reprodutivo das mulheres, além da realização de estudos analíticos, com informações válidas, confiáveis e atualizadas sobre o comportamento dos indicadores demográficos.

No Brasil, atualmente o uso da técnica indireta de estimação de fecundidade, denominado Método P/F de Brass, é indispensável para estudos sobre níveis e padrões da fecundidade (MIRANDA-RIBEIRO, 2007). No entanto dentre as diversas metodologias existentes para mensurar a fecundidade, não há aquela que traduza todas as variantes do comportamento reprodutivo da mulher, nesse aspecto a transição da fecundidade gera motivações por metodologias que sejam capazes de interpretar, o nível e padrão, mais próximo da realidade (MIRANDA-RIBEIRO, 2007).

A principal questão que motivou este trabalho, reside no desenvolvimento de um novo modelo paramétrico (Modelos Não Lineares de Efeitos Mistos) de ajuste das curvas de fecundidade e que capte novas estimativas da fecundidade, proporcionando resultados mais plausíveis.

A partir disso neste trabalho foram utilizadas informações oriundas do Censo Demográfico e do SINASC, no qual cada fonte de dados apresenta uma estrutura hierárquica, que refere-se a dois níveis. Para o primeiro nível foram calculadas as Taxas Específicas de Fecundidade Acumulada (TEFA), a partir da Taxa Específica de Fecundidade (TEF) nas diferentes faixas etárias (15-49 anos), pertencendo dessa forma a um mesmo estado, tornando as TEF medidas repetidas, em seguida considerando que estas medidas estejam organizadas de forma aninhada a um segundo nível (Estados).

Reconhecer e levar em conta esta estrutura hierárquica (faixa etária e estados) é importante por várias razões. Em muitos casos, o estudo do "contexto", ou seja, as características das unidades de nível superior (estados) são o principal objetivo da análise (DIEUX-ROU, 1998). Em outros casos, é importante levar em conta a dependência dos dados, pois ao não fazê-la, pode implicar na superestimação dos coeficientes do modelo em estudo (DIEUX-ROU, 1998; SOARES et al 2000). É fato que as TEFA, pertencentes a um mesmo estado, podem estar sujeitas a fatores

semelhantes da fecundidade, isto devido à similaridade das condições socioeconômicas e culturais.

Deve-se, portanto, considerar a dependência entre as observações, para evitar conclusões errôneas e estimativas falsamente significantes, devido à sobrestimação dos erros padrões. Dessa forma, fica reforçada a proposta do uso de modelos multiníveis ou também chamados modelos de efeitos mistos no presente estudo.

Dessa forma, se tem por objetivo avaliar a aplicação do modelo não linear de efeitos mistos para o estudo dos níveis e padrões da fecundidade nos estados do Nordeste do Brasil, apresentando os princípios e supostos básicos, ilustrando sua aplicação para um caso concreto e realizar uma análise comparativa do comportamento da fecundidade entre os estados do Nordeste em dois momentos no tempo, 2000 e 2010.

Assim, este estudo está subdividido em quatro capítulos, além das referências. No primeiro, apresenta-se a introdução da dissertação e os objetivos a serem alcançados. O segundo capítulo refere-se à revisão da literatura de diversos métodos diretos e indiretos da estimação da fecundidade. No terceiro capítulo apresentam-se os dados utilizados e o método, cujo objetivo é usar a metodologia do modelo não linear de efeitos mistos. No quarto capítulo, encontram-se os resultados para o caso dos estados do Nordeste Brasileiro, mostrando a interpretação dos parâmetros de um modelo assintótico no contexto das taxas específicas de fecundidade da população e, por último, o quinto capítulo ressalta as conclusões do estudo.

## **2 REVISÃO DA LITERATURA**

---

### **2.1 Métodos de estimação da fecundidade**

A fecundidade é o resultado concreto da capacidade reprodutiva da mulher, isto refere-se à relação entre nascidos vivos e mulheres em idade reprodutiva (CARVALHO et al., 1998). No Brasil, são disponibilizadas várias fontes de informações que fornecem dados sobre o número de nascimentos, possibilitando o cálculo dos níveis e padrões da fecundidade nacional (BARBOSA, 1999). As principais instituições que disponibilizam as informações relacionadas aos nascimentos vivos são os Cartório de Registro Civil, Censo Demográficos e o Ministério da Saúde (MS). Vale destacar que os Cartórios de Registro Civil e Censo Demográfico são conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2010; BRASIL, 2012; WINGERTER, 2013).

Os padrões de fecundidade podem ser analisados a partir da representação gráfica das taxas específicas de fecundidade observadas para as mulheres em idade fértil num período de tempo concreto (dados do período), ou para um coorte de mulheres concreta (dados de coorte), frente à idade da maternidade.

No campo dos estudos demográficos recorre-se a metodologias para modelar as curvas de fecundidade (CAVENAGHI et al, 1998; MIRANDA-RIBEIRO, 2007; PARRADO, 2014). Trata-se de determinar um certo modelo indireto ou direto que, partindo das taxas específicas de fecundidade, seja capaz de produzir uma curva suave que se ajuste a estas informações da melhor maneira possível de acordo com algum quesito preestabelecido para a qualidade de ajuste.

Na literatura existem diversas técnicas para estimar os níveis e padrões da fecundidade, cada uma destas com especificidades próprias. A seguir, apresentam-se algumas das técnicas utilizadas para a estimação da fecundidade.

### 2.1.1 Método P/F de Brass

O método P/F foi desenvolvido por Brass (1974), trata de uma importante técnica (indireta) demográfica para estimação das taxas de fecundidade. A descrição completa da metodologia original encontra-se no manual X *United Nations* (1983). Quando o interesse do demógrafo é estimar uma estrutura de fecundidade para determinada população com base em fontes de dados de má qualidade, ou seja, quando existe uma deficiência de cobertura nos registros de nascimentos das fontes de estatísticas vitais. Albuquerque et al. (1985) descreveram a técnica de Brass como aquela que fornece a estimativa da fecundidade de um conjunto de mulheres que estão expostas ao risco de ter filhos, aceitando-se o padrão etário das taxas específicas de fecundidade, as quais são derivadas da razão entre o número de filhos tidos nos últimos 12 meses que antecedem a data de referência do censo e o número de mulheres classificadas por grupos quinquenais de idade. De outro lado, o nível de fecundidade é estimado através do número médio de filhos tidos nascidos vivos informado pelo conjunto de mulheres mais jovens.

Este método possui como pressupostos básicos que:

- a) a fecundidade das faixas etárias é considerada constante na série histórica;
- b) a forma (ou padrão) da fecundidade por grupos de idade é corretamente descrita pelas taxas específicas de fecundidade;
- c) o nível da fecundidade é correto para as mulheres mais jovens, uma vez que declararam corretamente o número de filhos nascidos vivos, em comparação às mulheres mais velhas.

Os dados iniciais necessários para os cálculos são: o número total de filhos nascidos vivos por idade da mulher; o número de filhos nascidos vivos no período de

referência de 12 meses, por idade da mulher; e a população feminina por idade, que declarou ter tido filhos nascidos vivos.

Na técnica, a Taxa Específica de Fecundidade de uma determinada faixa etária (nTEFx) não é calculada diretamente dos dados brutos, mas sim baseada nas estatísticas das respostas às perguntas sobre fecundidade. Para cada grupo de idade, define-se:

1) Taxa Específica de Fecundidade por idade indireta ( $f_i$ ): corresponde ao número de filhos nascidos vivos no período de referência por mulher no intervalo  $i$  de acordo com a equação:

$$f_i = \frac{\text{Número de filhos nascidos vivos no período de referência de 12 meses}}{\text{População feminina no grupo, que declarou fecundidade}}$$

Geralmente nas bases de dados o número de filhos nascidos vivos no período de referência é distribuído por idade da mãe na data de referência e não por idade da mãe na ocasião do parto. Essa distorção em  $f_i$  é corrigida pela técnica.

2) Fecundidade acumulada ( $F_i$ ): corresponde ao número médio de filhos que uma mulher teria, ao atingir a idade máxima do grupo  $i$ , de acordo com a equação:

$$F_i = 5 \sum_{j=1}^i f_j$$

3) Parturição média ( $P_i$ ): corresponde ao número médio de filhos para cada mulher no intervalo  $i$ . É uma medida agregada à experiência de fecundidade da mulher até o momento da entrevista e mede a história de fecundidade declarada de uma mulher do grupo. Calcula-se, então, a razão entre a parturição média ( $P_i$ ) e a Fecundidade acumulada ( $F_i$ ).

Como cada uma está baseada em informações vindas de perguntas diferentes, as razões são diferentes em todos os grupos de idade. Outro motivo porque as medidas não são equivalentes é que  $P_i$  é a fecundidade entre mulheres do grupo  $i$  e  $F_i$  é a fecundidade para mulheres que atingem a idade máxima do grupo  $i$ .

A lógica da técnica é estimar  $nN_x$  a partir da nTEFx estimada ( $n\hat{T}EF_x$ ). A taxa específica de fecundidade é estimada por um  $f_i$  ajustado ( $f_i^*$ ). Pelo pressuposto (b),

considera-se que existe um  $i$  que corresponde ao intervalo  $[x, n + x]$ , de acordo com a equação  $nT\hat{E}F_x = k \times fi$ , em que  $k$  é um fator multiplicador de correção, constante para todos os grupos de idade. Este fator de correção é identificado baseando-se na interpolação entre os limites superior e inferior do intervalo da tabela elaborada por Brass para este fim.

Em teoria, os fatores de correção obtidos pela técnica P/F seriam igual a um e estariam representados por uma reta horizontal, caso os pressupostos fossem obedecidos. Porém quando a série P/F é ascendente – há um aumento do valor P/F com o aumento da idade – isto pode estar indicando por um lado, que há erros nos dados e, por outro que está havendo declínio da fecundidade (BRASS e COALE, 1974; WORLD HEALTH ORGANIZATION, 1983).

Albuquerque et al. (1985) destacaram que esta técnica indireta assume imprecisão no período de referência, principal fonte de erro em relação às informações sobre nascimentos ocorridos no ano anterior. As mulheres do grupo de 20 a 24 anos de idade são as que tendem a informar com maior precisão o número de filhos nascidos vivos. Há uma marcante evidência de omissão do total do número de filhos tidos pelas mulheres informantes na medida em que estas vão envelhecendo, devido ao fenômeno de erro de memória. Ocorrendo o mesmo no grupo etário de 15 a 19 anos, mas por omissão da informação.

### 2.1.2 Método Relacional de Gompertz (MRG)

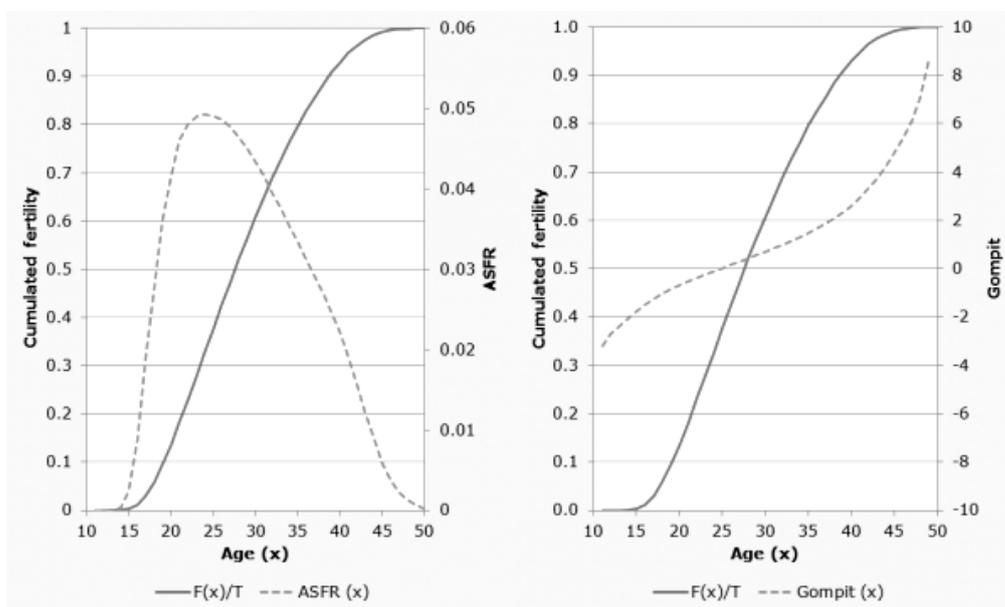
O Modelo Relacional de Gompertz foi proposto por Brass (1978). É uma técnica de estimação indireta na qual calcula as taxas de fecundidade baseando-se na parturição (número total de filhos tidos até à data do censo para cada grupo etário das mulheres). A descrição completa da metodologia original encontra-se no manual X *United Nations* (1983). O método tem como objetivo remediar os erros que comumente encontram-se nos dados de fecundidade associados com poucos nascimentos registrados no período de referência. Este método apresenta semelhanças ao sistema do logito de mortalidade, sendo expresso como uma transformação linear para se obter um padrão estimado, por meio de dois parâmetros (CAMLIN et al., 2004).

Na formulação original do modelo, a distribuição proporcional da parturição ou da fecundidade acumulada é expressa como uma percentagem da taxa da fecundidade total (a distribuição somadas). Por conseguinte, a soma da distribuição é igual a um. As proporções transformadas são conhecidas como *gompits* e são dadas pela seguinte equação:

$$Y(x) = \text{gompit}(F(x)/F) = -\ln(-\ln(F(x)/F))$$

na qual  $F(x)$  é a soma das taxas de fecundidade específicas por idade à idade  $x$  e  $F$  é a Taxa de Fecundidade Total. Exatamente a mesma relação vale para as parturições, substituindo  $F(x)$  com parturição média para os grupos etários e  $F$  com a parturição acumulada à idade de 50 anos. A transformação Gompertz estende o eixo idade original, de modo que os *gompits* traçados contra a idade quase formam uma linha reta. No entanto, a transformação não é perfeita; a linha tende a curvar ligeiramente em ambas as extremidades (Figura 4), que representa graficamente as taxas de fecundidade.

Figura 4 - Efeito do *gompit* transformado numa distribuição de Fecundidade



Fonte: <http://demographicestimation.iussp.org/content/relational-gompertz-model>

Dentre as vantagens oferecidas pelo método têm-se que: o método permite avaliar tendências e qualidades dos dados; está menos sujeito a flutuações de pequenos números; é uma técnica bastante útil para dados truncados ou com diferentes níveis de

confiabilidade (SAWYER, CORREA, 1988). Ao contrario da razão P/F, o modelo MRG não assume que a fecundidade tenha sido constante num passado recente (ZABA, 1981) e usa todas as idades para o ajuste. No entanto, envolve alguma subjetividade na seleção dos pontos que permitem ajustar uma reta para estimar a fecundidade (BRASS, 1996).

À exceção do material de origem que já foi referido, a literatura sobre o modelo de Gompertz relacional é escassa. Enquanto este é, sem dúvida, em parte devido à sua qualificação (BOOTH,1984). Uma exposição coerente de como aplicar o modelo encontra-se no manual *Manual X* ( SLOGGETT, LATÃO, ELDRIDGE et al., 1994 )

Na literatura têm sido propostos diversos modelos paramétricos e não paramétricos com possibilidades de aplicações úteis na pesquisa demográfica. Além de sua utilidade para a criação de taxas hipotéticas para cenários demográficos (previsão e projeção), estes também servem para resumir dados complexos em índices (SCHMERTMANN,2003; PERISTERA e KOSTAKI, 2007). Esses modelos têm sido geralmente criados para os países desenvolvidos do mundo e, geralmente, se encaixam bem à população para a qual o modelo se propõe (HOEM et al., 1981).

Na demografia é muito comum considerar o problema da modelagem das curvas da fecundidade, determinando um certo modelo a partir das taxas específicas da fecundidade, sendo assim capaz de produzir uma nova curva suave que se adapte ou se ajuste da melhor maneira possível às taxas de específicas de fecundidade observadas, tomando em consideração algum critério preestabelecido para o ajuste do modelo.

Hadwiger (1940), Hoem et al., (1981), Chandola et al., (1999) e Peristera e Kostaki (2007), abordaram vários modelos paramétricos para o ajuste de curvas da fecundidade, na qual consideraram uma função  $f$ , que se ajuste de forma adequada às taxas específicas de fecundidade observadas. Alguns exemplos deste tipo de modelos mostraram-se uma continuação, na qual todos dependem da idade à reprodução (representada pela variável  $x$ ).

### 2.1.3 Modelo Hadwiger (1940)

Este modelo foi desenvolvido por Hadwiger (1940), no qual  $a$ ,  $b$  e  $c$  são os parâmetros a serem estimados. De acordo com Chandola et al. (1999), estes parâmetros podem ser interpretados desde um ponto de vista demográfico, no qual  $f(x)$  é a taxa de fecundidade para idade  $x$ . Os parâmetros  $a$ ,  $b$  e  $c$  correspondem, respectivamente, "a" fecundidade total, o parâmetro "b" ajuda a determinar a altura da curva da fórmula  $ab/c$  que está altamente relacionado com a taxa máxima para a idade da fecundidade Específica e o parâmetro "c" está fortemente correlacionada com a idade da maternidade

$$f(x) = \frac{ab}{c} \left(\frac{c}{x}\right)^{\frac{3}{2}} \exp \left\{ -b^2 \left(\frac{c}{x} + \frac{x}{c} - 2\right) \right\}$$

### 2.1.4 Modelo Gamma (1981)

Este modelo foi desenvolvido por Hoem et al. (1981), em que  $d$  representa a idade mais do início da reprodução e  $R$  é o nível da fecundidade. Os parâmetros a serem estimados:  $R$ ,  $b$  e  $c$ , sendo que  $b$  e  $c$ , carentes de interpretação demográfica.

$$f(x) = R \frac{1}{\Gamma(b)c^b} (x-d)^{b-1} \exp \left\{ -\left(\frac{x-d}{c}\right) \right\}$$

### 2.1.5 Modelo Beta (1981)

Hoem et al. (1981) também desenvolveram outro modelo, desta vez usando outros tipos de parâmetros. Estes parâmetros estão relacionados com a media  $\mu$  e a variância  $\sigma^2$  por meio das seguintes expressões:

$$f(x) = R \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} (\beta-\alpha)^{-(A+B-1)} (x-\alpha)^{A-1} (\beta-x)^{B-1}$$

$$B = \left\{ \frac{(\mu - \alpha)(\beta - \mu)}{\sigma^2} - 1 \right\} \frac{\beta - \mu}{\beta - \alpha} \quad A = B \frac{\mu - \alpha}{\beta - \mu}$$

sendo que os parâmetros  $\alpha$  e  $\beta$  são interpretados como o limite inferior e superior, respectivamente, da amplitude plausível de idades da reprodução feminina. O parâmetro  $R$  representa o nível da fecundidade.

Segundo Parrado (2014), estes modelos paramétricos antes mencionados são aplicados quando a curva das taxas específicas de fecundidade em função da idade representa um padrão suave. Sendo que a principal vantagem destes modelos reside em que permitem obter um ajuste razoavelmente bom, fazendo dessa forma o uso de um número reduzido de parâmetros. No entanto sua aplicação é limitada pois estes resultados pioram consideravelmente quando são aplicados sobre novos padrões de fecundidade que foram aparecendo em outras regiões nos últimos anos.

#### 2.1.6 Modelo de erro ajustado (Adjusted Error Model –AEM)

O modelo foi desenvolvido por Gayawan e Cols (2010). Usa uma função de cinco parâmetros, denotados por  $a$ ,  $b$ ,  $\mu$ ,  $s_1$  e  $s_2$  para relacionar as taxas de fecundidade com as idades da mãe. A equação correspondente assume que, para uma mãe com idade  $x$ , a taxa de fecundidade específica a esta idade é dada por:

$$f(x|a, b, \mu, s_1, s_2) = a + b \cdot \begin{cases} Erfc \left[ \left( \frac{x - \mu}{s_1} \right)^2 \right] & \text{se } x \leq \mu \\ Erfc \left[ \left( \frac{x - \mu}{s_2} \right)^2 \right] & \text{se } x > \mu \end{cases}$$

em que,  $Erfc$  representa a função erro complementar explicitada na equação seguinte .

$$Erfc(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-t^2} dt$$

$f(x)$  é o TEF na idade  $x$  da mãe,  $b$  é a taxa de fecundidade modal,  $\mu$  é a média de idade maternidade enquanto  $s_1$  e  $s_2$  são a propagação da distribuição da idade de seus respectivos pontos de referência (GAYAWAN et al, 2010).

### 3 DADOS E MÉTODOS

---

#### 3.1 Dados

As unidades de análises ou níveis de agregação deste trabalho foram os estados do Nordeste Brasileiro. Como fontes de dados foram utilizadas as informações coletadas sobre nascimentos disponíveis no Censo Demográfico dos anos 2000 e 2010 do IBGE e do Sistema de Informações sobre Nascidos Vivos (SINASC) do Ministério da Saúde. No tocante às informações censitárias foram calculadas estimativas de fecundidade utilizando a técnica pelo método indireto P/F de Brass (BRASS, COALE, 1975; CARVALHO, 1982). Os dados do SINASC foram levantados no âmbito do website do DATASUS que disponibiliza as informações de nascidos vivos para todo o território brasileiro. E no que refere aos dados do SINASC foram utilizadas estimativas de fecundidade calculando-se de forma direta as Taxas Específicas de Fecundidade (TEF), conforme a equação:

$$TEF = \frac{{}_n NV_x}{{}_n Popf_x}$$

sendo,  ${}_n TEF_x$  a Taxa Específica de Fecundidade para o grupo etário  $x$ ;  $x+n$ ,  $n$  o valor da amplitude do intervalo de idade), neste caso 5 (grupo etário quinquenal) e  $x$  o valor referente ao limite inferior de cada grupo etário (15, 20, 25, 30, 35, 40, 45).  ${}_n NV_x$ , indica o número de nascidos vivos do ano de referência, provenientes de mulheres entre  $x$  e  $x+n$  anos e  ${}_n Popf_x$  indica a população feminina do grupo etário ( $x$ ;  $x+n$ ).

Barbosa (1999) comenta que o Censo Demográfico é uma importante fonte de informação sobre os níveis da fecundidade dado o seu caráter universal. No entanto autores como Souza e Wong (2002), Souza (2004). Moreira (2012), Wingerter (2013) consideraram que, após duas décadas de existência do SINASC, é de se esperar que este Sistema represente adequadamente o quantitativo de nascimentos existentes no Nordeste.

Frias et al. (2007) e Viacava (2009) citaram que uma das grandes vantagens deste Sistema (SINASC) é que possui cobertura nacional e alta abrangência, uma vez que o preenchimento da Declaração de Nascido Vivo (DNV) é obrigatório para todo qualquer nascimento em território brasileiro, não importando onde ocorra, se em estabelecimento público, particular ou em domicílio.

O período de análise considerado neste trabalho abrange os anos de 2000 e 2010, mas deve-se atentar que o Censo e o SINASC possuem distintos períodos de referência para a coleta das informações. No caso do Censo de 2000, a data de referência (e início da coleta) é a noite de 31 de julho para 1º de agosto de 2000. Para o Censo do ano de 2010, a data de referência é a noite de 31 de julho para 1º de agosto de 2010. Os nascimentos acontecidos nos doze meses anteriores são contabilizados e os nascimentos após a data de referência não são incluídos no recenseamento.

As informações coletadas no SINASC referem-se ao ano em curso (ano civil) e dessa forma, em 2000, as informações coletadas correspondem ao total de nascimentos ocorridos entre 1º de janeiro e 31 de dezembro do ano 2000. Da mesma condição, os dados do SINASC de 2010 dizem respeito aos nascimentos ocorridos ao longo do ano 2010.

Wingerter (2013) comentou que é importante salientar que os dados do SINASC são estáticos, uma vez que esta base não é atualizada com a inclusão de novas informações após sua publicação oficial na página web do DATASUS (Departamento de Informática do SUS- DATASUS, 2013).

## 3.2 Métodos

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística, e envolve a procura de um modelo que seja o mais simples possível, e que descreva bem os dados observados que surgem em diversas áreas do conhecimento como agricultura, demografia, ecologia, economia, medicina, sociologia, entre outras.

A seguir, citam-se algumas técnicas de regressão que serão usadas na análise dos dados com o intuito de encontrar o modelo mais adequado que descreva o comportamento das Taxas Específicas de Fecundidade Acumulada no âmbito do nove estados do Nordeste brasileiro.

### 3.2.1 Modelos de regressão linear

Segundo Turkiman e Silva (2000), o modelo de regressão linear, desenvolvido por Legendre e Gauss no começo do século XIX, foi a principal técnica de modelagem estatística até meados do século XX. Nas diversas áreas do conhecimento como ciências biológicas, física, social e engenharia, os modelos lineares são úteis na fase de análise dos dados resultantes (RENCHER; SCHAALJE, 2008). Estes apresentam grande aplicabilidade por apresentarem maior facilidade de cálculos nos métodos de estimação de parâmetros (SCHABENBERGER; PIERCE, 2001).

O modelo de regressão linear é uma técnica estatística utilizada para estudar a relação entre variáveis. Tanto no caso de duas variáveis (regressão simples) como no caso de mais de duas variáveis (regressão múltipla), a análise de regressão linear pode ser utilizada para explorar e quantificar a relação entre uma variável chamada dependente ou resposta  $Y$  (Taxa Específica de Fecundidade Acumulada) e uma ou mais variáveis chamadas independentes ou preditoras  $X_k$ , ( $K=1$ ) encontra-se subjacente a uma relação do tipo:

$$TEFA = \beta_0 + \beta_1 * Idade \quad (1)$$

sendo que  $a$  e  $b$  são os parâmetros da relação linear procurada. O objetivo pode ser explicativo (demonstrar uma relação matemática que pode indicar, mas não provar, uma relação de causa e efeito) ou preditivo (obter uma relação que nos permita, perante futuras observações da variável  $X=Idade$ , prever o correspondente valor de TEFA).

Supondo que se dispõe de  $n$  conjuntos de medidas com as correspondentes observações, a utilização do modelo incluirá sempre uma parcela de erro. Utilizando o índice  $i$  ( $i=1, \dots, n$ ) para indicar cada conjunto, ter-se-á então:

$$y_i = \beta_0 + \beta_1 * Idade + e_i \quad i=1, \dots, n \quad (2)$$

Independentemente das motivações, a versão da regressão linear que aqui se apresenta consiste em estimar os valores dos parâmetros  $a$  e  $b$ , através da minimização da soma dos quadrados dos desvios.

O modelo tem os seguintes pressupostos:

- i. a relação entre a variável  $Idade$  e a variável  $TEFA$  é linear;
- ii. os valores da variável  $Idade$  são fixos, isto é, a variável  $Idade$  não é uma variável aleatória;
- iii. a média dos erros é nula, isto é:

$$E(e_i) = 0 \quad i=1, 2, \dots, n; \quad (3)$$

- iv. a variância dos erros  $e_i$  é sempre  $\sigma^2$  denotada variância residual, ou seja, erro é homocedástico;
- v. o erro em uma observação é não correlacionado com o erro em qualquer outra observação

$$\text{Var}(e_i) = E(e_i) - [E(e_i)]^2 = E(e_i^2) = \sigma^2; \quad (4)$$

- vi. os erros têm distribuição normal.
- vii. a variação aleatória da variável  $TEFA$  para qualquer valor fixo da  $Idade$  segue uma distribuição Normal, e estes termos de erro são independentes.

### 3.2.2 Modelo Não Linear

A análise de dados através do modelo linear é uma das técnicas mais usadas de estimação. Porém, em diversas situações práticas, algumas de suas suposições, como a normalidade e a linearidade nos parâmetros, não são satisfeitas. Este fato, melhorou o desenvolvimento de novas técnicas estatísticas para os modelos de regressão, surgindo, então, novas classes de modelos, os modelos de regressão não linear e os modelos lineares generalizados (MATTOS, 2013).

Logo, modelos não lineares permitem ajustar relações mais complexas que relações lineares. A principal característica quando utilizam-se estes tipos de modelos é que estes requerem estimativas iniciais dos parâmetros, as quais são melhoradas sucessivamente até que algum critério de convergência seja cumprido e este procedimento pode apresentar alto custo computacional, além de não haver garantia que o algoritmo iterativo convirja para uma solução única (SCHABENBERGER; PIERCE, 2001) ou o mesmo convirja para falsas estimativas dos parâmetros.

Um modelo é classificado como não linear se pelo menos uma das derivadas parciais da função não linear em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo. Uma das principais características desses modelos é que os mesmos em geral são deduzidos a partir de suposições teóricas e os parâmetros resultantes são interpretáveis. Assim, aproximá-los para os modelos lineares, mesmo que sejam alcançados ajustes satisfatórios, prejudicaria bastante a obtenção de estimativas mais realistas dos parâmetros de interesse. Diferentemente dos modelos de regressão lineares, em que a qualidade e principalmente a validade do ajuste são simplesmente avaliadas por meio de diagnósticos usuais, as medidas de não linearidade devem ser calculadas.

Estas medidas, particulares dos modelos de regressão não lineares, são úteis na análise da extensão do comportamento linear, podem ter seus resultados assintóticos invalidados, principalmente em situações em que estão disponíveis apenas pequenas amostras (MATTOS, 2013). Porém, em algumas situações um modelo não linear pode ser mais apropriado, uma vez que muitos fenômenos geralmente não são lineares.

O modelo de regressão não linear é frequentemente usado por pesquisadores em diversas áreas do conhecimento, tais como, agricultura, biologia, econometria, engenharia, química, etc. Na maioria das vezes, as formulações de possíveis modelos são deduzidas a partir de suposições teóricas inerentes ao fenômeno que se tem interesse modelar e os parâmetros resultantes são interpretáveis.

Em matemática os fenômenos observáveis podem ser como uma equação da forma:

$$Y=f(\theta, \alpha) + \varepsilon \quad (5)$$

sendo Y o valor medido de uma resposta (TEFA),  $\varepsilon$  é o erro do experimento associado com esta medida e  $f(\theta, \alpha)$  é uma função que contem o parâmetro  $\theta$  e a variável  $\alpha$  (Idade).

Definição de não linearidade: Descreve-se um procedimento que permite diferenciar um modelo linear de um não linear nos parâmetros. Um modelo linear pode-se escrever da forma:

$$f(\theta, \alpha) = \sum_{i=1}^p \theta_i g_i(\alpha) \quad (6)$$

para alguma função  $g_i$  que depende somente dos valores de  $\alpha$ , mas não dos valores de  $\theta$  (parâmetros). Os modelos que não podem ser escritos desta forma são não lineares nos parâmetros ou simplesmente não lineares. Note-se que a linearidade ou não linearidade do modelo se determina pela forma em que entram os parâmetros ao modelo e não pela forma em que entram as variáveis  $\alpha_1, \dots, \alpha_k$ .

Dessa forma, uma equação quadrática em  $\alpha$  será da forma:

$$f(\theta, \alpha) = \theta_1 + \theta_2 \alpha + \theta_3 \alpha^2 \quad (7)$$

pode-se expressar de forma linear como em (2) fazendo  $g_1(\alpha) = 1$ ,  $g_2(\alpha) = \alpha$  y  $g_3(\alpha) = \alpha^2$

Por outro lado, enquanto o modelo (7):

$$f(\theta, \alpha) = \theta_1 \exp(-\theta_2 \alpha) \quad (8)$$

não pode se fazer esta representação linear. Uma forma simples de ver se um modelo é linear ou não linear é examinando as derivadas de  $f$  com respeito a cada um dos parâmetros  $\theta_i$ . Se  $df/d\theta_i$  não depende de nenhum dos elementos de  $\theta = (\theta_1, \dots, \theta_p)^T$ , o modelo é linear em  $\theta_i$ ; e se é linear em todos os  $p$  parâmetros o modelo se diz linear nos parâmetros ou simplesmente linear. Alguns modelos não lineares, mediante uma apropriada transformação podem ser descritos como uma equação linear. Assim, por exemplo, se o modelo é transformado mediante logaritmo pode ser representado como uma equação linear. As transformações para linearizar, no entanto, têm o efeito de transformar também os erros  $\mathcal{E}$ , alterando assim a relação entre  $f$  e  $\mathcal{E}$ .

Transformações ou reparametrizações podem ser utilizadas para linearizar algumas funções não lineares. Tal situação facilita os cálculos das estimativas dos parâmetros, porém os parâmetros podem perder suas interpretações práticas. Tem-se ainda que as estimativas obtidas na escala original devem sofrer uma transformação inversa, para a escala original e este processo apresenta viés (SCHABENBERGER; PIERCE, 2001).

#### **a) Estimação de Parâmetros**

Segundo Brandão (1996), os métodos mais utilizados na estimação de parâmetros de modelos não lineares são os de Mínimos Quadrados e de Máxima Verossimilhança. Frequentemente o Método de Mínimos Quadrados apresenta algumas variações por conta da suposição sobre a estrutura da matriz de variâncias e covariâncias do problema em estudo (SARTORIO, 2013). Mas detalhes podem ser encontrados em Morettin (2004), Brandão (1996), Riazoshams (2009), Haseman (2005), Streibig (2008).

#### **b) Dados longitudinais**

Uma estrutura diferente de dados surge quando a mesma unidade de observação ou experimentação é medida várias vezes (FREES, 2004). Com dados longitudinais, usa-se observações repetidas do mesmo indivíduo e estas observações tendem a ser correlacionadas. A inclusão do efeito aleatório pertinentes induz uma correlação ou dependência entre as medidas repetidas sobre um mesmo indivíduo. Segundo Fitzmaurice et al. (2008), houve progresso notável em metodologia estatística para a análise de dados longitudinais, nos últimos 25 a 30 anos.

### 3.2.3 Modelos Lineares de Efeitos Mistos

Os modelos lineares mistos, ou não lineares, mistos são também chamados como modelos hierárquicos ou modelos multiníveis, termos comuns usados especialmente nas áreas de educação, ciências sociais e medicina ( GOLDSTEIN et al, 2002)

O modelo linear misto foi proposto por Henderson (1975) e seu uso em estudos longitudinais tornou-se mais frequente com o trabalho de Laird e Ware (1982). Este modelo foi proposto para analisar dados quando os casos estão agrupados em unidades de informação mais amplas e se tomam medidas tanto num nível mais baixo (os casos) como nos níveis mais altos (os grupos).

A estrutura hierárquica neste estudo refere-se em um nível às medidas das taxas específicas acumuladas em cada faixa etária pertencente ao mesmo estado, considerando que haja uma dependência nas medidas entre as idades de um mesmo Estado.

Este fato constitui uma quebra de uma importante pressuposição dos modelos lineares gerais: a independência entre as observações. Os modelos de efeitos mistos permitem abordar este tipo de estrutura hierárquica prestando atenção às covariáveis existentes no ajuste do modelo.

A sua maior flexibilidade, de acordo com Riboldi (2007), é a possibilidade de contemplar a estrutura de interdependência das observações e a estrutura de correlação (intra-indivíduo) presente em medidas repetidas tomadas na mesma unidade experimental. O modelo misto essencialmente considera a existência de efeitos fixos e de efeitos aleatórios, além do erro experimental. De acordo com Hauser et al. (2009) os efeitos são considerados fixos e aleatórios, sendo no caso estudado:

- i. Fixos: refletem a média dos parâmetros da estimativa de fecundidade para cada estado estudado.
- ii. Aleatórios: retratam a variabilidade dos estados com relação aos parâmetros.

**a) Definição do Modelo Linear de Efeitos Mistos**

O modelo linear misto proposto por Laird e Ware (1982) para o  $i$ -ésimo grupo, abrangendo apenas um nível de agrupamento é dado por:

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, i = 1, \dots, M \tag{9}$$

$$\gamma_i \sim N(0, \Psi), \quad \varepsilon_i \sim N(0, \Sigma)$$

em que  $Y_i$  é um vetor  $n_i \times 1$  referente à variável resposta,  $\beta$  é um vetor  $p \times 1$  de parâmetros desconhecidos denominado de efeitos fixos,  $X_i$  é uma matriz  $n_i \times p$  de covariáveis observadas relacionada com os efeitos fixos,  $\gamma_i$  é um vetor  $q \times 1$  de parâmetros denominado de efeitos aleatórios,  $Z_i$  é uma matriz  $n_i \times q$  associada aos efeitos aleatórios e  $\varepsilon_i$  é um vetor  $n_i \times 1$  de erros aleatórios. Assume-se que os valores  $\gamma$  e  $\varepsilon$  são independentes.

Pode-se reescrever o modelo (9) na seguinte forma funcional:

$$Y = X\beta + Z\gamma + \varepsilon \tag{10}$$

em que

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}; Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_q \end{bmatrix};$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}; \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}.$$

De (10) Tem-se que:

$$E(Y) = X\beta$$

$$Var(Y) = V = Z\Psi Z^T + \Sigma$$

sendo que  $\Psi$  é denominada de matriz de variâncias e covariâncias do vetor de efeitos aleatórios sendo uma matriz positiva semi-definida. A matriz  $\Psi$  pode assumir diversas estruturas de covariâncias que estão relacionadas com particularidades do estudo em questão.

## **b) Estimação dos Parâmetros**

Atualmente, pode-se encontrar na literatura diversos métodos para estimação de parâmetros em modelo de regressão, neste caso, dois métodos são os mais utilizados sendo estes: o de Máxima Verossimilhança e o de Máxima Verossimilhança Restrita. Para mais detalhes o leitor pode consultar Laird e Ware (1982), Thompson (1971), entre outros.

### **3.2.4 Modelo Não Linear de Efeitos Mistos (NLME)**

Os primeiros desenvolvimentos associados aos modelos não lineares mistos (NMLE) aparecem com Sheiner e Beal (1980) em estudos de farmacocinética. Têm sido frequentemente utilizados em análises de dados de medidas repetidas de crescimento, (HARTFORD; MEZA; JAFFREZIC; FOULLEY, 2007; WANG, 2007; DAVI-DIAN, 2009). O termo modelo misto surge porque o modelo em análise apresenta uma parte que contém parâmetros de efeito(s) fixo(s) e outra parte que contém parâmetros de efeito(s) aleatório(s). Demidenko (2004) comenta que modelos mistos podem ser utilizados para vários fins como por exemplo: modelar dados longitudinais, dados com múltiplas fontes de variação, para modelar variedade biológica e heterogeneidade, entre outros.

Com o modelo misto, tem-se uma maior versatilidade na aplicação da técnica de análise de dados com medidas repetidas (análise de perfis e de curvas de crescimento), pois:

- a) permitem a modelar estruturas da matriz de covariâncias com um número menor de parâmetros que a dos modelos multivariados gerais;
- b) podem ser usados quando os dados não são balanceados em relação ao tempo;
- c) possibilitam o ajuste de curvas polinomiais de graus diferentes para cada grupo de unidades experimentais.

A sua maior flexibilidade, de acordo com Riboldi (2007), é a possibilidade de contemplar a estrutura de interdependência das observações presente em medidas repetidas tomadas na mesma unidade experimental.

### **Por que o modelo não linear de efeitos mistos?**

Um problema para o uso NLME consiste quando uma população com estrutura hierárquica, as observações individuais não são completamente independentes. Por exemplo, as taxas de fecundidade acumuladas por idades tendem a ter semelhanças entre as mesmas, já que podem proceder da mesma área geográfica (estados).

A utilização de modelos não lineares de efeitos mistos consiste em identificar os efeitos aleatórios, escolher os efeitos fixos e escolher a melhor estrutura para a matriz de covariâncias entre e intra-estados.

Além disso na análise da fecundidade observou-se que as Taxas Específicas de Fecundidade Acumuladas descrevem um comportamento ao longo das faixas etárias (15 até 49 anos) que é bem representada por funções não lineares (Figura5). Assim os modelos estatísticos não lineares são necessários para descrever seu comportamento. Diversas funções não lineares podem ser usadas.

A escolha da função  $f$  não linear adequada dependerá do comportamento dos dados. Spyrides (2004) comenta que o ajuste da função  $f$  pode ser feita utilizando os modelos de Gompertz, Exponencial, Richard, Von Bertalanffy, Logístico Simples, Regressão Assintótico, dentre outros modelos.

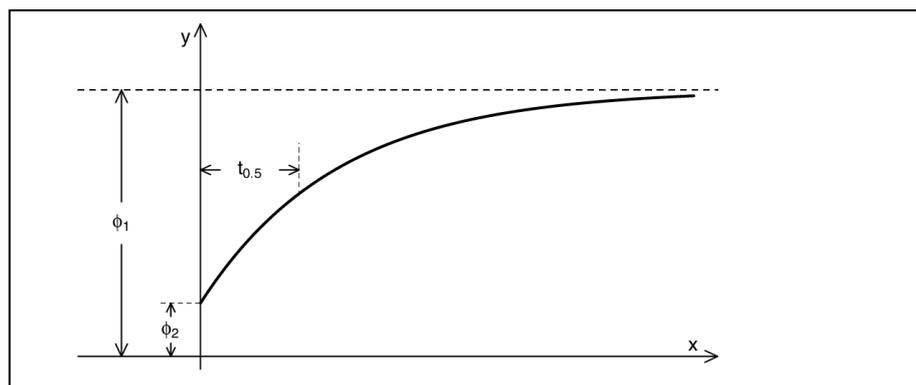
#### **Regressão Assintótica**

Pinheiro e Bates (2000) comentaram que o modelo de regressão assintótica é usado para modelar uma resposta  $y$  que se aproxima de uma assíntota horizontal como  $x \rightarrow \alpha$  Na qual é representada como:

$$y(x) = \phi_1 + (\phi_2 - \phi_1) \exp[-\exp(\phi_3)x]$$

de modo que  $\phi_1$  representa a assíntota  $x \rightarrow \alpha$ ;  $\phi_2$  representa o intercepto  $y(0)$  e o parâmetro  $\phi_3$  representa o logaritmo da taxa constante. Utiliza-se o logaritmo para impor positividade da constante de velocidade assim o modelo aproxima-se a uma assíntota. Estes parâmetros são mostrados na Figura 5.

Figura 5- Representação dos parâmetros da regressão assintótica



Fonte: Pinheiro e Bates (2000)

Quanto à estimação, o procedimento consiste em atribuir valores iniciais mediante a regressão assintótica o qual servirão como semente inicial para o ajuste do modelo não linear misto.

Neste estudo os parâmetros da regressão assintótica podem ser interpretados de acordo com os interesses aplicados à fecundidade. Este modelo permite determinar a idade em que as mulheres dos estados do Nordeste, tendo um comportamento dado pela Figura 5, venham a atingir a metade da taxa específica de fecundidade acumulada (TEFA), ou seja, da assíntota  $\phi_1$ . Portanto,  $t_{0,5}$ , a idade em que as mulheres atingem a metade da TEFA do estado do Nordeste, expressa a velocidade com que aquele estado atinge a metade do número médio de filhos tidos por mulher, é dado por:

$$t_{0,5} = \frac{\ln 2}{\exp(\ln \phi_3)}$$

Com base nos dados observados, o comportamento da taxa Específica de Fecundidade acumulada durante todo seu período reprodutivo da mulher descrevem a forma da função de regressão assintótica, o que significa que a aceleração do crescimento é maior durante os primeiros anos do período reprodutivo da mulher e que este desacelera no fim do período reprodutivo observado. Assim, a função que melhor se ajusta a estes dados prospectivos das taxas de fecundidade específicas acumuladas no período reprodutivo da mulher é o modelo de regressão assintótica, descrito por Pinheiro & Bates (2000).

O uso desta técnica de modelos não lineares de efeitos mistos é particularmente útil quando se lida com medidas repetidas de dados (TEF), sobre a mesma unidade de observação como no caso em estudo os estados do Nordeste Brasileiro (NEB). Tais dados apresentam estrutura hierárquica, uma vez que as medidas repetidas são aninhadas dentro de cada um dos estados.

A estrutura hierárquica deve levar em consideração a suposição de que as observações entre estados sejam independentes e que as aninhadas em cada estado possuam a característica da dependência, isto é, se correlacionam. O modelo não linear de efeitos mistos (ou modelo de efeitos aleatórios) permite analisar dados como o das Taxas Específicas da Fecundidade, incorporando a dependência e a estrutura de correlação dos erros (PINHEIRO & BATES, 2000, TWISK, 2006).

A aplicação mais comum de modelos não lineares mistos (NMLE) é para dados longitudinais. Para o primeiro estágio, no caso mais simples, para um nível, a  $j$ -ésima observação no  $i$ -ésimo indivíduo (grupo) pode ser descrita como:

$$Y_{ij} = f(\phi_i, x_{ij}) + \epsilon_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n_i, \quad (11)$$

em que  $Y_{ij}$  é a variável resposta (Taxa Específica de Fecundidade Acumulada) da  $j$ -ésima observação do  $i$ -ésimo estado do NEB;  $f$  é uma função não linear em pelo menos um parâmetro real e diferenciável de um vetor de parâmetros  $\phi_i$  específico para cada estado e do componente  $x_{ij}$  da matriz de delineamento  $X_i$ , referente ao grupo  $i$ ;  $\epsilon_{ij}$  representa o vetor de erros dentro de estado, sendo normalmente distribuído, ou seja  $\epsilon_{ij} \sim N(0, \sigma^2 R_i)$ ;  $N$  é o número de grupos (estados do nordeste brasileiro) e  $n_i$  é o número de observações dentro do  $i$ -ésimo grupo (estado), ou seja, o número de faixas etárias consideradas ([15-19], [20-24], [25-29], [30-34], [35-39], [40-44] e [45-49]).

No caso em estudo, assumiu-se, para os dados longitudinais das Taxas Específicas de Fecundidade acumulada, o modelo de regressão assintótico para a função  $f$ , dada por:

$$f(\phi_i, x_{ij}) = \text{Asymp} + (R0 - \text{Asymp}) \exp[-\exp(\text{lrc} * \text{idade})] \quad (12)$$

A interpretação dos parâmetros  $\phi_{ij}$  pode ser dada da seguinte forma:

**Asymp**, representa a assíntota correspondendo à Taxa de Fecundidade Total ao final do período reprodutivo da mulher;

**R0**, representa o valor da Taxa Específica de Fecundidade no início do período (15 a 19 anos de idade) reprodutivo da mulher;

**Irc**, cujo logaritmo representa a idade com que a população feminina atinge a metade (50%) da Taxa de Fecundidade Total ou seja, ao final do período reprodutivo (45 a 49 anos de idade).

Este modelo têm um caráter interpretativo muito grande, sendo que os parâmetros do modelo possuem interpretação demográfica e representam coeficientes que explicam o processo de interesse, determinando os níveis e a velocidade com que as mulheres atingem a metade da TFT ao final do período reprodutivo. Dessa forma, o modelo misto não linear torna-se uma alternativa interessante para estudar a fecundidade dos estados.

Neste estudo, utilizou-se o modelo assintótico não linear de efeitos mistos para estimar os níveis da fecundidade dos estados do Nordeste Brasileiro, nos anos de 2000 e 2010. Tanto as inferências sobre os parâmetros de efeito aleatório e de efeito fixo, quanto para as comparações das diferentes estruturas de covariâncias para as matrizes  $D$  e  $R_i$ , serão baseadas nos resultados do Teste da Razão de Verossimilhança e nos critérios de informação de Akaike.

### **Vantagem e desvantagem dos modelos NLME**

A vantagem de utilizar os modelos NLME, em primeiro lugar, é o fato de considerar que uma função não linear possa representar o processo da fecundidade estudado no nordeste brasileiro, oferecendo parâmetros que tenham uma interpretação natural sobre o padrão da fecundidade. Além disso, permite captar mudanças nas respostas individuais do grupo estudado, ou seja, as especificidades de cada estado do Nordeste.

Se corretamente utilizados estes modelos, permitem obter melhores estimações dos coeficientes de regressão e de sua variação que com os modelos tradicionais. Uma grande flexibilidade oferecida por esta abordagem se dá em termos de modelar a estrutura da variância dos dados em função das variáveis explicativas.

A limitação para o uso do método são os aplicativos disponíveis. Alguns tipos de modelos multiníveis passam a ser ajustados com pacotes estatísticos como o SAS, R/SPlus. Os aplicativos especializados, como o MLwiN, são capazes de ajustar toda uma gama de modelos, incluídos os mais complexos. Talvez a mais importante limitante do uso de MLwiN seja a dificuldade de transferência de dados e incapacidade de trabalhar automaticamente com a falta de informação (*missing*).

De acordo com Pinheiro e Bates (2000), os modelos não lineares mistos são uma extensão dos modelos lineares mistos combinada com os modelos não lineares, permitindo que a função de regressão não linear dependa de efeitos fixos e aleatórios.

Para Hartford e Davidian (2000) e Wang (2007), o NLME descreve os dados com medidas repetidas por uma função caracterizando a dependência sistemática das observações sobre as covariáveis e os parâmetros.

Um NLME mais geral possível é apresentado por Pinheiro e Bates (2000) e Davidian (2009), dentre outros, que em sua formulação além de inserirem covariáveis dentro de cada unidade, incluem covariáveis de nível individual.

As aplicações mais comuns para os modelos não lineares de efeitos mistos são em dados com medidas repetidas, particularmente, dados longitudinais. A seguir detalha-se o modelo não linear de efeitos mistos para medidas repetidas proposto por Lindstrom & Bates (1990).

Pode se pensar como um modelo hierárquico. Para o primeiro estágio, no caso mais simples, para um nível da  $j$ -ésima observação no  $i$ -ésimo indivíduo (grupo) tem-se que:

$$y_{ij} = f(\phi_i, \mathbf{x}_{ij}) + \varepsilon_{ij}, \quad i = 1, 2, \dots, N \quad j = 1, 2, \dots, n_i, \quad (13)$$

em que  $f$  é uma função não linear, real e diferenciável de um vetor de parâmetros  $\phi_i$  e do componente  $x_{ij}$  da matriz de delineamento  $X_i$ , referente ao indivíduo  $i$ ;  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathbf{R}_i)$ ;  $N$  é o número de indivíduos e  $n_i$  é o número de observações dentro do  $i$ -ésimo indivíduo.

Pode-se representar matricialmente por

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix} \quad e \quad \boldsymbol{\eta}(\boldsymbol{\phi}_i) = \begin{bmatrix} f(\boldsymbol{\phi}_i, \mathbf{x}_{i1}) \\ f(\boldsymbol{\phi}_i, \mathbf{x}_{i2}) \\ \vdots \\ f(\boldsymbol{\phi}_i, \mathbf{x}_{in_i}) \end{bmatrix},$$

assim, tem-se que

$$\mathbf{y}_i = \boldsymbol{\eta}(\boldsymbol{\phi}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N, \quad (14)$$

sendo  $R_i$  uma matriz de dimensões  $(n_i \times n_i)$  que depende de  $i$  somente por sua dimensão. Por conveniência pode-se suprimir a dependência de  $\eta_i$  sobre o vetor preditor  $x_{ij}$  ( $j=1, \dots, n_i$ ). Em muitas situações  $R_i = I$ , porém esta estrutura pode ser substituída por uma estrutura de covariância marginal não independente, como por exemplo uma estrutura de correlação autorregressiva de primeira ordem AR(1).

No segundo estágio, o vetor de parâmetros específicos do indivíduo ( $\boldsymbol{\phi}_i$ ) é modelado por

$$\boldsymbol{\phi}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad \text{com } \mathbf{b}_i \sim N(0, \sigma^2 \mathbf{D}) \quad (15)$$

em que  $\boldsymbol{\beta}$  um vetor  $p$ -dimensional ( $p \times 1$ ) de efeitos fixos e  $\mathbf{b}_i$  um vetor  $q$ -dimensional ( $q \times 1$ ) de efeitos aleatórios associados ao  $i$ -ésimo estado (que não varia com  $j$ ) com matriz de variâncias-covariâncias  $\sigma^2 \mathbf{D}$ ;  $\mathbf{A}_i$  é a matriz de delineamento associada ao efeito fixo de dimensões  $(n_i \times p)$  e  $\mathbf{B}_i$  é a matriz de delineamento associada ao efeito aleatório com dimensões  $(n_i \times q)$ . Assim, a função  $f$  pode ser descrita como  $f(\mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \mathbf{x}_{ij})$ .

As matrizes  $\mathbf{A}_i$  e  $\mathbf{B}_i$  podem ser utilizadas para simplificar a Especificação do modelo. Por exemplo, considere um estudo em que se tenha dois grupos (ou estados), então a matriz  $\mathbf{A}_i$  poderá ser descrita da forma  $\mathbf{A}_i = [\mathbf{I} \ 0]$  para designar o primeiro grupo como fixo ou  $\mathbf{A}_i = [0 \ \mathbf{I}]$  para designar o segundo grupo como fixo. Se os  $\mathbf{B}_i$  são iguais para todos os indivíduos  $i$ , ambos os grupos ainda teriam efeitos aleatórios a partir da mesma distribuição (LINDSTROM; BATES, 1990). Muitas outras situações podem ser construídas, como o caso de nem todos os parâmetros terem um componente aleatório,

sendo neste caso,  $A_i = I$ , mas  $B_i$  poderá conter apenas algumas de  $A_i$ . Muitas outras combinações são possíveis.

Pode-se agora reescrever o modelo dos  $N$  Estados do Nordeste do Brasil em apenas um. Matricialmente, tem-se que

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} \quad e \quad \boldsymbol{\eta}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{\eta}_1(\phi_1) \\ \boldsymbol{\eta}_2(\phi_2) \\ \vdots \\ \boldsymbol{\eta}_N(\phi_N) \end{bmatrix},$$

$$\tilde{\mathbf{D}} = \text{diag}(\mathbf{D}, \dots, \mathbf{D}) \quad e \quad \mathbf{R} = \text{diag}(\mathbf{R}_1, \dots, \mathbf{R}_N).$$

em que  $\tilde{\mathbf{I}}$  tem dimensões  $(qN \times qN)$  e  $\mathbf{R}$  tem  $(niN \times niN)$ .

Logo, o modelo geral é dado por

$$\mathbf{y}|\mathbf{b} \sim \mathcal{N}(\boldsymbol{\eta}(\boldsymbol{\phi}), \sigma^2 \mathbf{R}), \quad \boldsymbol{\phi} = \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}, \quad (16)$$

$$\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \tilde{\mathbf{D}}), \quad (17)$$

em que

$$\mathbf{B} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_N), \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} \quad e \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_N \end{bmatrix}$$

#### 3.2.4.1 Estimação de $\boldsymbol{\beta}$ e $\mathbf{b}$

Diferentes métodos podem ser utilizados para estimar os parâmetros do NLME como o Método dos Mínimos Quadrados, Máxima Verossimilhança e procedimentos de estimação Bayesiana. Vonesh e Carter (1992) utilizaram no processo de estimação um procedimento de quatro estágios, similar ao procedimento de regressão linear multivariado descrito por Malinvaud (1966) e Gallant (1987), em que os parâmetros da

estrutura de variâncias e covariâncias são estimados pelo método MQO (Mínimos quadrados ordinários) e em seguida, usam-se estas estimativas em um procedimento de estimação de MQG (Mínimos quadrados generalizados).

Contudo, no presente trabalho abordou-se apenas os Métodos de Estimação de Máxima Verossimilhança e Máxima Verossimilhança Restrita. Para maiores detalhes sobre os outros métodos de estimação, Gallant (1987), Davidian e Giltinan (1995) e Vonesh e Chinchilli (1997) são boas referências.

Quando os componentes de variâncias das matrizes R e D são conhecidos e  $\eta$  é uma função linear de b e  $\beta$ :

$$\eta(A_i\beta + B_i b_i) = X_i\beta + Z_i b_i \quad (18)$$

então, define-se

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_n^T \end{bmatrix} \quad \text{e} \quad \mathbf{Z} = \text{diag}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n).$$

O estimador linear de  $\beta$  obtido pelo Método dos Mínimos Quadrados Generalizados é dado por:

$$\hat{\beta}_{lin} = \hat{\beta}_{lin}(\theta) = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{y} \quad (19)$$

e o preditor linear de b pelo Método dos Mínimos Quadrados Generalizados é

$$\hat{\mathbf{b}}_{lin} = \hat{\mathbf{b}}_{lin}(\theta) = \tilde{\mathbf{D}} \mathbf{Z}^T \Sigma^{-1} [\mathbf{y} - \mathbf{X} \hat{\beta}_{lin}(\theta)] \quad (20)$$

em que  $\Sigma = \mathbf{R} + \tilde{\mathbf{L}} \mathbf{Z}^T$  e  $\theta$  contém os elementos das matrizes de covariâncias D e R. O vetor paramétrico  $\theta$  é uma conveniência que permite escrever todos os parâmetros de covariância (exceto  $\sigma^2$ ) em um vetor, sem Especificar a parametrização de R. Note que  $\theta$  é definido em função de D e R. As estimativas de  $\hat{\beta}_{lin}$  e  $\hat{\mathbf{b}}_{lin}$  maximizam conjuntamente a função.

$$g_{lin}(\boldsymbol{\beta}, \mathbf{b}|\mathbf{y}) = -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - \frac{1}{2\sigma^2} \mathbf{b}^T \tilde{\mathbf{D}}^{-1} \mathbf{b} \quad (21)$$

que para  $\boldsymbol{\beta}$  fixo é o logaritmo da densidade *a posteriori* de  $\mathbf{b}$  e para  $\mathbf{b}$  fixo é o logaritmo da verossimilhança para  $\boldsymbol{\beta}$ . Note que (19) é composto por uma soma de quadrados e um termo quadrático em  $\mathbf{b}$ . Transformando o termo quadrático em  $\mathbf{b}$  em um termo equivalente a uma soma de quadrados, pode-se tratar a otimização puramente como um problema de mínimos quadrados. O problema de mínimos quadrados surge com o aumento do vetor de dados com "pseudo-dados", como

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b} + \tilde{\boldsymbol{\varepsilon}} \quad (22)$$

em que

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{X} \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{Z} \\ \tilde{\mathbf{D}}^{-1/2} \end{bmatrix}, \quad \text{e } \tilde{\boldsymbol{\varepsilon}} \sim \mathcal{N}(0, \sigma^2\mathbf{I}) \quad (23)$$

sendo  $\tilde{\mathbf{L}}^{-1/2} = \text{diag}(\mathbf{L}^{-T}, \mathbf{L}^{-T}, \dots, \mathbf{L}^{-T})$  e  $\mathbf{L}$  é o fator de Cholesky de  $\mathbf{D}$  ( $\mathbf{D} = \mathbf{L}^T\mathbf{L}$ ), em que  $\mathbf{L}$  é uma matriz triangular superior e  $\mathbf{L}^{-T}$  é usado para representar a transposta da inversa da matriz  $\mathbf{L}$ . Similarmente,  $\mathbf{R}^{-1/2}$  contém o fator de Cholesky de  $\mathbf{R}_i$ .

Para a situação em que  $\eta$  é uma função não linear, esta idéia é facilmente reproduzida. Em modelo não linear de efeitos mistos o estimador de Máxima Versimilhança  $\hat{\boldsymbol{\beta}}(\theta)$  e a moda *a posteriori*  $\hat{\mathbf{b}}(\theta)$  maximizam a função

$$g(\boldsymbol{\beta}, \mathbf{b}|\mathbf{y}) = -\frac{1}{2\sigma^2}[\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b})]^T \mathbf{R}^{-1}[\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b})] - \frac{1}{2\sigma^2} \mathbf{b}^T \tilde{\mathbf{D}}^{-1} \mathbf{b} \quad (24)$$

para  $\boldsymbol{\beta}$  fixo  $g(\cdot)$  é uma constante mais o logaritmo da função de densidade *a posteriori* de  $\mathbf{b}$ . Assim, é evidente que  $\mathbf{b}$  que maximiza  $g(\cdot)$  para um dado valor de  $\boldsymbol{\beta}$  é a moda *a posteriori*. Logo,  $\hat{\mathbf{b}}$  é um estimador de máxima verossimilhança relativa para uma distribuição marginal aproximada de  $\mathbf{y}$ . Estas estimativas podem ser calculadas como a solução para um problema de mínimos quadrados não linear e é obtido aumentando-se o vetor de dados como "pseudo-dados", sendo

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{\eta}}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) + \tilde{\boldsymbol{\varepsilon}} \quad (25)$$

em que

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{\eta}}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) = \begin{bmatrix} \mathbf{R}^{-1/2}\boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) \\ \tilde{\mathbf{D}}^{-1/2}\mathbf{b} \end{bmatrix}, \quad e \quad \tilde{\boldsymbol{\varepsilon}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) \quad (26)$$

### 3.2.4.2 Estimação de $\theta$

Os métodos de estimação que são frequentemente utilizados para encontrar os valores das estimativas dos parâmetros para minimização da soma de quadrados dos erros de modelos não lineares mistos são os Métodos de Máxima Verossimilhança (MV) e o Método de Máxima Verossimilhança Restrita (MVR).

#### 3.2.4.2.1 Máxima Verossimilhança

Inferências sobre  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$  são baseados no tratamento da densidade condicional de  $y_i|x_i$ . Na prática, assume-se que os efeitos aleatórios  $b_i$  sejam contínuos e usando a independência entre as observações e entre  $\boldsymbol{\varepsilon}_i$  e  $b_i$ , a densidade conjunta de  $\mathbf{y}$  é dada por:

$$p(\mathbf{y}|\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i|x_i; \boldsymbol{\beta}, \boldsymbol{\theta}) \quad (27)$$

em que  $p(y_i|x_i; \boldsymbol{\beta}, \boldsymbol{\theta})$  é a densidade marginal de  $y_i|x_i$ . Então define-se o estimador de máxima verossimilhança para  $\boldsymbol{\theta}$  respeitando a densidade marginal de  $\mathbf{y}$  que é definida por

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{b})p(\mathbf{b})d\mathbf{b} \quad (28)$$

como  $\boldsymbol{\eta}$  não é linear em  $\mathbf{b}$ , não há uma expressão fechada para a densidade e o cálculo das estimativas é muito difícil. Alternativamente, aproxima-se a distribuição condicional de  $\mathbf{y}$  (10) para  $\mathbf{b}$  próximo de  $\hat{\mathbf{b}}(\boldsymbol{\theta})$  por uma Normal multivariada com vetor

de médias que é linear em  $\mathbf{b}$ . Para isso, aproxima-se o erro  $\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b})$  expandido em série de Taylor na vizinhança de  $\hat{\mathbf{b}}$  como

$$\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{b}) \approx \mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Q}}\mathbf{b} - \hat{\mathbf{Q}}\hat{\mathbf{b}} \quad (29)$$

em que

$$\hat{\mathbf{Q}}_i = \hat{\mathbf{Q}}_i(\boldsymbol{\theta}) = \left. \frac{\partial \eta_i}{\partial \mathbf{b}_i^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} = \left( \left. \frac{\partial \eta_i}{\partial \boldsymbol{\phi}^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} \right) \mathbf{B}_i$$

e

$$\hat{\mathbf{Q}} = \hat{\mathbf{Q}}(\boldsymbol{\theta}) = \text{diag}(\hat{\mathbf{Q}}_1, \hat{\mathbf{Q}}_2, \dots, \hat{\mathbf{Q}}_N) = \left. \frac{\partial \boldsymbol{\eta}}{\partial \mathbf{b}^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}}$$

Note que  $\hat{\mathbf{Q}}_i$  é uma função de  $\boldsymbol{\theta}$ . Então

$$\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Q}}\hat{\mathbf{b}} - \hat{\mathbf{Q}}\mathbf{b} | \mathbf{b} \sim N(\mathbf{0}, \sigma^2 \mathbf{R})$$

e a distribuição condicional aproximada de  $\mathbf{y}$  é dada por

$$\mathbf{y} | \mathbf{b} \sim \mathcal{N}(\boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Q}}\hat{\mathbf{b}} - \hat{\mathbf{Q}}\mathbf{b}, \sigma^2 \mathbf{R})$$

Esta expressão, juntamente com a distribuição de  $\mathbf{b}$  permitem aproximar a distribuição marginal de  $\mathbf{y}$  por

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\eta}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\hat{\mathbf{b}}) - \hat{\mathbf{Q}}\hat{\mathbf{b}}, \sigma^2 \hat{\boldsymbol{\Sigma}}) \quad (30)$$

em que

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = \mathbf{R} + \hat{\mathbf{Q}}\tilde{\mathbf{D}}\hat{\mathbf{Q}}^T.$$

é a matriz de variâncias e covariâncias de  $\mathbf{y}$ .

Supondo que a variável resposta tem distribuição normal multivariada, o logaritmo da função de verossimilhança é obtido utilizando a distribuição marginal (30) é dado por:

$$l(\beta, \sigma^2, \theta | \mathbf{y}) = -\frac{1}{2} \log |\sigma^2 \hat{\Sigma}| - \frac{1}{2\sigma^2} [\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Q}}\hat{\mathbf{b}}]^T \hat{\Sigma}^{-1} [\mathbf{y} - \boldsymbol{\eta}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Q}}\hat{\mathbf{b}}] \quad (31)$$

em que  $\log$  é uma função logarítmica; e  $\hat{\mathbf{b}}$  e  $\hat{\mathbf{Q}}$  dependem de  $\theta$ . Logo,  $\hat{\beta}_{EMV}$ ,  $\hat{\sigma}_{EMV}$  e  $\hat{\theta}_{EMV}$  serão os estimadores de máxima verossimilhança de  $\beta$ ,  $\sigma$  e  $\theta$  se maximizarem (31). Segundo Linstrom e Bates (1990), o inverso da matriz de derivadas segundas de (31) fornece uma estimativa para aproximação da matriz de variâncias e covariâncias para  $\hat{\beta}_{EMV}$ ,  $\hat{\sigma}_{EMV}$  e  $\hat{\theta}_{EMV}$ .

#### 3.2.4.2.2 Máxima Verossimilhança Restrita (MVR)

O primeiro método de Máxima Verossimilhança Restrita (MVR) foi proposto por Thompson (1962) e colocado em prática considerando a distribuição Normal em 1971 por Patterson e Thompson (HARVILLE, 1977). Os estimadores dos componentes de variância obtidos pelo método de MVR têm sido utilizados como alternativa ao método de MV para estimação dos componentes da variância em modelos mistos, os quais apresentam a vantagem de considerar a perda dos graus de liberdade envolvidos na estimação dos parâmetros do efeito fixo do modelo. Sendo assim, tais estimativas tendem a ser menos viesadas do que as estimativas fornecidas por MV (VERBEKE; MOLENBERGHS, 2000). Uma discussão sobre algumas das vantagens e desvantagens destes dois métodos de estimação podem ser encontradas em Harville (1977).

Segundo Noh e Lee (2008), o método MRV usado para Modelos Não Lineares de Efeitos Mistos é, muitas vezes, dificultada pela existência de integrais analiticamente intratáveis.

Considerando a distribuição normal, o método para definir os estimadores de MVR é o mesmo utilizado para os estimadores de MV exceto pelo fato de que o logaritmo da função de verossimilhança é dada por:

$$l_R(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\mathbf{y}) = -\frac{1}{2}\log|\sigma^{-2}\hat{\mathbf{P}}^T\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{P}}| + l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\mathbf{y}) \quad (32)$$

em que

$$\hat{\mathbf{P}}_i = \hat{\mathbf{P}}_i(\boldsymbol{\theta}) = \left. \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\beta}^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} = \left( \left. \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\phi}^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}} \right) \mathbf{A}_i$$

e

$$\hat{\mathbf{P}} = \hat{\mathbf{P}}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\mathbf{P}}_1 \\ \hat{\mathbf{P}}_2 \\ \vdots \\ \hat{\mathbf{P}}_N \end{bmatrix} = \left. \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^T} \right|_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}}$$

Isto é, ao invés de otimizar diretamente a verossimilhança das observações, o método da MVR maximiza a parte da função de verossimilhança que é invariante aos efeitos fixos (ALCARDE, 2012). Para maximizar (30) é necessário o uso de algum método iterativo. Os estimadores  $\hat{\boldsymbol{\beta}}_{EMV}$ ,  $\hat{\boldsymbol{\sigma}}_{EMV}$  e  $\hat{\boldsymbol{\theta}}_{EMV}$  são aqueles que maximizam  $l_R$ .

A função de verossimilhança restrita é baseada em n-p contrastes de erros linearmente independentes. No modelo não linear a matriz de derivadas parciais  $\hat{P}$ , que define estes contrastes de erros, depende de  $\hat{\boldsymbol{\beta}}$  e  $\hat{\mathbf{b}}$ . Contudo, desde que o subespaço medido pelas colunas de  $\hat{P}$  depende apenas da não linearidade intrínseca e não da não linearidade de efeitos dos parâmetros (BATES; WATTS, 1980), os resultados serão aproximadamente constantes numa vizinhança das estimativas.

### 3.2.4.2.3 Aproximações numéricas para o método de estimação

Sartorio (2013) comenta que para modelos não lineares de efeitos mistos são necessários métodos iterativos para obter a maximização da função de verossimilhança  $l(\cdot)$ , pois geralmente sua resolução é intratável analiticamente, mesmo quando os efeitos aleatórios seguem uma distribuição normal.

Pinheiro e Bates (2000) citaram a existência de três métodos principais de aproximação empregados para estimação: o primeiro, proposto por Lindstrom e Bates (1990), aproxima a verossimilhança de um modelo não linear misto por um modelo linear misto e é denominado aproximação linear de efeito misto; o segundo método

utiliza aproximação Laplaciana para a função de verossimilhança e é denominado de aproximação Laplaciana modificada. Maiores detalhes deste método podem ser encontrados em Tierney e Kadane (1986). E, por fim, o método de quadratura gaussiana para melhor aproximação de Laplace. Uma descrição completa do método encontra-se disponível em Davidian e Gallant (1992).

#### 3.2.4.2.4 Estruturas da matriz de variâncias e covariâncias.

Na análise de dados longitudinais envolvendo medidas repetidas para um mesmo Estado deve-se levar em conta na modelagem não só a variabilidade entre os Estados, mas também a variação existente entre observações de um mesmo estado. Um passo importante para fazer inferências sobre os parâmetros de um modelo de forma precisa envolve a escolha adequada de uma estrutura de variância.

Spyrides (2004) comenta que as estruturas comumente utilizadas na literatura são: a matriz de covariância uniforme ou simétrica composta (SC), a autorregressiva de primeira ordem e a não estruturada.

##### 1) Matriz de Covariância Uniforme ou Simétrica Composta

Neste modelo, assume-se que existe uma correlação positiva  $\rho$ , entre quaisquer duas medidas de um mesmo indivíduo. Em termos matriciais tem-se:

$$V_0 = \sigma^2 [(1 - \rho)I + \rho J] = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

com  $I$  denotando uma matriz identidade  $n \times n$  e  $J$  uma matriz  $n \times n$  cujos elementos são iguais a 1 e  $\sigma^2$  é a variância em quaisquer instantes e  $\rho$  é a correlação entre duas observações tomadas em quaisquer dois instantes sobre o mesmo Estado. Neste caso, são necessários apenas dois parâmetros independentemente do número de medidas repetidas sobre o mesmo Estado, porém esta estrutura assume não somente que as variâncias sejam constantes como também as covariâncias.

## 2) Matriz de Covariância Não Estruturada

A matriz deste modelo é representada por:

$$V_0 = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n_i} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \cdots & \sigma_{2n_i} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \cdots & \sigma_{3n_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_i 1} & \sigma_{n_i 2} & \sigma_{n_i 3} & \cdots & \sigma_{n_i n_i}^2 \end{bmatrix}$$

na qual  $\sigma^2$  é a variância no instante  $n_i$  e  $\sigma_{uni}$  é a covariância entre as respostas tomadas nos instantes  $n_i$  e  $u$  ( $n_i \neq u$ ). Este modelo não estruturado requer muitos parâmetros à medida que o número de observações por Estado aumenta e, conseqüentemente, requer um tamanho de amostra maior.

## 3) Matriz de Covariância Auto-regressiva de primeira ordem

Como, geralmente, não se dispõe de um tamanho amostral grande e não se espera que uma estrutura simétrica composta seja apropriada para a maioria dos dados longitudinais, é importante considerar uma estrutura parcimoniosa como a auto-regressiva de primeira ordem. Aqui a correlação entre as observações sobre o mesmo indivíduo entre os instantes de tempo é assumido como  $\rho^{ni}$ . Assim:

$$V_0 = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \cdots & \cdots & \cdots & \rho & 1 \end{bmatrix}$$

requerendo  $(n+1)$  parâmetros. Considera-se aqui que cada indivíduo foi medido em intervalos de tempo igualmente espaçados.

O processo de decisão sobre qual a estrutura de covariância mais adequada dependerá de cada estudo e pode ser avaliada usando o Akaike Information Criterion (AIC) ou o Critério Bayesiano de Schwartz (SBC). Estes critérios utilizam os valores do logaritmo da função de verossimilhança penalizados pelo número de parâmetros estimados.

### 3.2.4.3 *Método de Seleção de modelos e qualidade do ajuste*

Sartorio (2013) comentou que a seleção de um modelo deve basear-se em um bom ajuste aos dados amostrais e também na sua coerência com o evento estudado. Dessa forma selecionar o melhor modelo significa selecionar a melhor estrutura para as médias (parte fixa) e a melhor estrutura de covariâncias. A seleção e avaliação de um modelo são tarefas simples, mas de extrema importância, pois determinarão o modelo que melhor prediz a variável dependente de forma parcimoniosa, isto é, com um menor número de parâmetros.

A construção do NLME consta de três etapas: 1) seleção de efeitos fixos; 2) identificação dos efeitos aleatórios e estimação; 3) comparação dos modelos. Os procedimentos descritos por Barbosa (2009) para modelos lineares mistos também servem para modelos não lineares mistos (NLME). Vale salientar que em modelagem, esta é uma das principais etapas na forma de avaliação e verificação da qualidade dos ajustes do modelo. Esta análise oferece evidências sobre possíveis violações nos pressupostos, como por exemplo, homocedasticidade de variância e normalidade dos resíduos, além de indicar a falta de ajuste. Para isso, são utilizadas ferramentas analíticas como critérios de informação, teste de hipóteses e análise de resíduos, descritos a seguir.

#### 3.2.4.3.1 Critérios de Informação

Os critérios de informação são frequentemente usados na literatura para comparação de modelos aninhados ou não aninhados. Como no caso de modelos lineares mistos, o critério de informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC) são utilizados para comparar NLME (PINHEIRO; BATES, 2000).

Estas medidas permitem comparar modelos a fim de selecionar o mais adequado à realidade dos dados e assim quantificar quão mais provável é cada modelo, sendo o cálculo fácil de ser realizado (MOTULSKY; CHRISTOPOULOS, 2003).

Os critérios encontram-se baseados no logaritmo da função de verossimilhança (MV ou MVR), sendo dependentes do número de observações e de parâmetros do modelo em estudo. Não dependem do valor-p ou do conceito de significância estatística, podendo dessa forma ser aplicados a modelos aninhados ou não aninhados.

O AIC (*Akaike Information Criterion*) foi proposto por Akaike (1974), sendo desenvolvido a partir da distância ou informação de Kulback e Leibler (1951). O cálculo do AIC de um determinado modelo é dado por:

$$AIC(\hat{\beta}) = -2 \log L(\hat{\beta}) + 2k. \quad (33)$$

em que  $L(\hat{\beta})$  é o valor da função de verossimilhança do modelo obtido pelas estimativas do processo de maximização e  $k$  é o número total de parâmetros de efeitos fixos e dos efeitos aleatórios. O valor do AIC pode ser positivo ou negativo (SAKAMOTO; ISHIGURO; KITAGAWA, 1986).

Bozdogan (1987) propôs uma correção para o AIC que é uma correção para pequenas amostras, e é dado por:

$$AIC_c = -2 \log L(\hat{\beta}) + 2K + 2 \frac{K(K+1)}{n-K-1}, \quad (34)$$

sendo  $n$  o número de observações utilizadas para estimação do modelo em estudo.

O BIC (*Bayesian Information Criterion*) foi proposto por Schwarz (1978), este critério penaliza os modelos com maior número de parâmetros e o seu uso é recomendado quando se trabalha com pequenas amostras. Pode ser calculado a partir de:

$$BIC = -2 \log L(\hat{\beta}) + k \log(N), \quad (35)$$

Vale salientar que o modelo que apresentar o menor valor de AIC,  $AIC_c$  e /ou de BIC será escolhido e considerado representativo de um melhor ajuste do modelo ao conjunto de dados.

### 3.2.4.3.2 Teste da Razão de Verossimilhança (TRV) para $\theta$

O teste da razão de verossimilhança envolve a comparação dos valores dos logaritmos das funções de verossimilhança maximizada sem restrição e restrita sob  $H_0$ , ou seja, a comparação entre  $L(\hat{\theta})$  e  $L(\hat{\theta}_0)$ .

A comparação de modelos de regressão não linear com qualquer subconjunto de parâmetros pode ser verificada por meio do teste de razão de verossimilhança (REGAZZI, 2003; SOARES, 2007).

Um modelo mais simples é dito aninhado em um modelo mais geral (modelo de referência), quando o espaço paramétrico do primeiro é um espaço do segundo ou quando o espaço paramétrico do primeiro modelo é um subespaço do segundo ou quando os parâmetros do modelo aninhado (modelo com restrição) podem ser obtidos impondo-se certas restrições nos parâmetros do modelo de referência.

As hipóteses testadas pelo TRV são:  $H_0$ : o modelo com restrição é melhor contra  $H_1$ : o modelo de referência é melhor. A estatística do teste é dada por:

$$\text{TRV} = -2 \log \left[ \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \right] = 2[\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] \quad (36)$$

que, sob  $H_0$ :  $\theta = \theta_0$ , segue aproximadamente uma distribuição qui-quadrado com  $g$  graus de liberdade ( $g$  é a diferença dos graus de liberdade dos modelos). Logo, rejeita-se  $H_0$  quando  $\text{TRV} > \chi^2_{p, 1-\alpha}$ , sendo  $\alpha$  o nível de significância considerado.

### 3.2.4.3.3 Intervalos de confiança

Com muita frequência, o propósito do pesquisador consiste em construir intervalos de confiança para os parâmetros. Brandão (1996) citou que intervalos de confiança podem ser usados para indicar a confiabilidade de uma estimativa. Em modelos mistos, auxilia na inserção de um efeito aleatório no parâmetro. Dessa forma, para o vetor de parâmetros  $\beta$ , temos que o intervalo de confiança para um componente  $\beta_k (k=1, \dots, q)$ , é dado por:

$$\hat{\beta}_k - t_{(1-\alpha/2)} [\hat{\Omega}_{ii}]^{1/2} \leq \beta_k \leq \hat{\beta}_k + t_{(1-\alpha/2)} [\hat{\Omega}_{ii}]^{1/2} \quad (37)$$

no qual  $t_{(1-\alpha/2)}$  é o valor crítico da distribuição  $t$ -Student, com  $(nq-r)$  graus de liberdade e um nível de significância  $\alpha$ ;  $r = \text{posto}(\hat{\Omega})$ ,  $\hat{\Omega} = \text{cov}(\hat{\beta})$  é a estimativa da matriz de covariâncias dos estimadores de  $\hat{\beta}$ , dado por:

$$\hat{\Omega} = \left[ \sum_{i=1}^n \mathbf{J}_i^T(\hat{\beta}) \hat{\Sigma}_i^{-1} \mathbf{J}_i(\hat{\beta}) \right]^{-1} \quad (38)$$

em que  $\mathbf{J}(\hat{\beta})$  é o jacobiano da função  $(\hat{\beta})$  e  $\hat{\Sigma}_i = \text{cov}(y_i)$ , e  $\hat{\Omega}_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $\Omega$  calculada em (38).

#### 3.2.4.3.4 Estratégia para seleção do modelo

West; Welch; Galecki (2007) citaram que embora não exista uma estratégia única de se construir modelos mistos, na verdade, determiná-los envolve um equilíbrio entre estatística e o assunto que envolve a modelagem. A seguir são apresentadas algumas estratégias utilizadas.

##### 1) *Top-down* (Estratégia descendente)

Esta estratégia sugere iniciar o procedimento com um modelo que inclui o número máximo de efeitos fixos possíveis, inclusive as interações se for de interesse (modelo maximal), que deverá extrair toda a variação sistemática dos dados. Após considerado o efeito fixo, incluem-se os termos de efeito aleatório, testando-se a importância dos mesmos (VERBEKE; MOLENBERGHS, 2000). A seguir detalha-se a estratégia.

- a) inicia-se com o ajuste do modelo maximal;
- b) seleciona-se uma estrutura de efeitos aleatórios para o modelo. A necessidade de incluir os efeitos aleatórios selecionados pode ser avaliada do TRV e pelos critérios de informação. Vale salientar que nesta etapa, busca-se a melhor estrutura para a matriz de variâncias e covariâncias  $D$ ;
- c) selecione uma estrutura de covariância para explicar a variação intra-indivíduos (Estado do Nordeste Brasileiro). Uma vez que os efeitos fixos e efeitos aleatórios já estão adicionados ao modelo, a variação restante observada é devido ao erro residual.

Assim, uma estrutura de covariâncias adequada para os resíduos deve ser investigada. Nesta etapa, busca-se modelar a matriz de variâncias e covariâncias R.

d) esta etapa envolve o uso de testes estatísticos apropriados para determinar quais parâmetros de efeito fixo são necessários no modelo.

## 2) *Step-up* (Estratégia ascendente)

Esta estratégia inicia no sentido contrário da anterior. Sugere iniciar o processo com o ajuste de um modelo fixo inicial, contendo somente o efeito da média geral. Em seguida são incluídos efeitos fixos de covariáveis e possivelmente efeitos aleatórios associados aos mesmos. A necessidade de inclusão de novos termos no modelo pode ser verificada pelo TRV ou critérios de informação e decidir se existe melhora ou não no ajuste por teste, como em *top-down* (RAUDENBUSH; BRYK, 2000; SNIJDERS, BOSKER, 2012).

### 3.2.4.4 *Análise de Resíduos*

A análise de resíduos é outro passo importante na etapa posterior ao ajuste do modelo. Esta permite verificar pressuposições de qualquer modelo proposto. De maneira geral, a análise de resíduos consiste em verificar a suposição de normalidade, homogeneidade de variâncias. West; Welch; Galecki (2007) relataram que os diagnósticos para modelos mistos são mais difíceis de interpretar, já que o modelo é mais complexo, devido à presença de efeitos aleatórios e de diferentes estruturas de covariâncias. Q-Q plots e envelopes simulados (ATKINSON, 1985) são importantes ferramentas gráficas para avaliar o afastamento da normalidade dos dados, que é uma das suposições feitas sobre a distribuição dos dados.

Pinheiro e Bates (2000), Nobre e Singer (2007) dentre outros, definiram três tipos de resíduos em modelos mistos e cada um deles é útil para avaliar algum tipo de suposição feita no modelo:

#### 3.2.4.4.1 Resíduo Marginal

O resíduo marginal é indicado para avaliar a suposição de linearidade do valor esperado da variável resposta e as covariáveis, além disso permite avaliar o ajuste da estrutura de covariâncias usadas no modelo. É definido da seguinte forma

$$\hat{\xi} = \mathbf{y} - \mathbf{X}\hat{\beta}. \quad (39)$$

em que  $\mathbf{y}$  é o vetor da variável resposta,  $\mathbf{X}$  é a matriz de delineamento e  $\hat{\beta}$  é a estimativa dos parâmetros associados à matriz de delineamento.

#### 3.2.4.4.2 Resíduo condicional

O resíduo condicional é a diferença entre os valores observados e os valores preditos condicionais, sendo utilizados para avaliar a hipótese de normalidade e de homocedasticidade do erro condicional, além da existência de observações discrepantes. Este tipo de resíduo mede o quanto o valor observado desvia da linha de regressão predita associado ao próprio sujeito (amostra), sendo definido como:

$$\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\mathbf{b}}. \quad (40)$$

em que  $\mathbf{Z}$  é a matriz de posto completo e constantes conhecidas que associa  $\mathbf{y}$  a  $\mathbf{b}$  e  $\hat{\mathbf{b}}$  é a estimativa de  $\mathbf{b}$  que contém os coeficientes aleatórios não observados.

Sartorio (2013) comentou que na literatura encontram-se poucos estudos sobre análise de resíduos para modelos não lineares mistos e sugere a possibilidade de estabelecer resíduos mais específicos.

### 3.2.4.5 Função da variância

Além de estruturas matriciais, existe ainda opções como funções de variâncias. Para entender tais funções é necessário, primeiramente, decompor a matriz  $R_i$  em um simples produto de matrizes, como

$$\mathbf{R}_i = \mathbf{V}_i \mathbf{C}_i \mathbf{V}_i \quad (41)$$

sendo  $\mathbf{V}_i$ , é uma matriz diagonal e  $\mathbf{C}_i$  é uma matriz de correlações, definida positiva e com todos os elementos da diagonal iguais a 1. Assim, tem-se que

$$\text{Var}(\epsilon_{ij}) = \sigma^2 [\mathbf{V}_i]_{jj}^2, \quad \text{corr}(\epsilon_{ij}, \epsilon_{jk}) = [\mathbf{C}_i]_{jk} \quad (42)$$

sendo  $\mathbf{V}_i$  a representação das variâncias e  $\mathbf{C}_i$  a representação das correlações do erro dentro de grupo (indivíduo). Pinheiro e Bates (2000) citaram que esta decomposição de  $R_i$  em estruturas de variâncias e correlações é conveniente tanto de forma teórica, como de forma computacional. Dessa forma, a função de variância geral para modelar a heterocedasticidade para o erro dentro de indivíduo pode ser estendido para:

$$\text{Var}(\epsilon_{ij} | \mathbf{b}_i) = \sigma^2 g^2(\mu_{ij}, \mathbf{x}_{ij}, \boldsymbol{\delta}) \quad (43)$$

em que  $\mu_{ij} = E(y_{ij} | \mathbf{b}_i)$ ,  $\mathbf{x}_{ij}$  é um vetor de covariáveis,  $\boldsymbol{\delta}$  é um vetor de parâmetros da variância e  $g(\cdot)$  é a função de variância, contínua em  $\boldsymbol{\delta}$ ,  $i=1, \dots, N$  e  $j=1, \dots, n_i$ . Em geral, os erros intra-indivíduos são assumidos independentes e homocedástico ( $R_i = \sigma^2 \mathbf{I}$ , denominada componente de variância).

A biblioteca NLME, existente no software R, apresenta vários tipos de funções de variância. A tabela 1 apresenta as funções já implementadas na biblioteca NLME.

Tabela 1 - Funções de variância contidas na classe varFunc da biblioteca NLME do software R.

Classe	Descrição
varExp	Exponencial da covariável
varPower	Potência da covariável
varConstPower	Constante somada à potência da covariável
varIdent	Variâncias diferentes por nível de fator
varFixed	Variâncias fixa
varComb	Combinação de funções de variância

Fonte: Pinheiro e Bates (2000)

A classe VarIdent apresenta uma estrutura de variâncias constantes. Se nenhuma forma de agrupamento for inserida, a função será igual a um. Caso incluído um fator,  $h=1, \dots, H$ , de formas de agrupamentos com níveis maiores que um, a função permite diferentes variâncias para cada nível deste fator. Suponha que o fator de agrupamento tenha  $H$  níveis, sendo  $H > 1$ , logo tem-se que

$$\text{Var}(\epsilon_{ij}) = \sigma^2 \delta_{h_{ij}}^2 \quad (44)$$

em que a função de variância corresponde a

$$g(h_{ij}, \boldsymbol{\delta}) = \delta_{h_{ij}}^2$$

Em geral, utiliza-se a restrição  $\delta_1 = 1$ , de modo que  $\delta_s$ ,  $s=2, \dots, H$  representa a razão entre o desvio padrão do  $s^{\text{th}}$  nível com o primeiro nível.

A classe VarPower apresenta uma estrutura de variâncias de potência. O parâmetro  $\delta$  pode modelar aumento ou diminuição da variância de acordo com o fator indicado. Ressalta-se que este tipo de função não pode ser utilizado se a covariável apresentar valores iguais a 0 (PINHEIRO; BATES, 2000)

$$\text{Var}(\epsilon_{ij}) = \sigma^2 |x_{ij}|^{2\delta}$$

em que a função de variância corresponde a

$$g(x_{ij}, \delta) = |x_{ij}|^\delta$$

Outras estruturas podem ser encontradas em Pinheiro e Bates (2000).

## 4 RESULTADOS

---

Esta seção está dividida em duas etapas: a primeira apresenta o desenvolvimento dos ajustes dos modelos descritos na seção de Métodos (3.2) e segunda etapa apresenta em detalhes a análise dos níveis e padrões da fecundidade no âmbito dos estados do Nordeste do Brasil.

### 4.1 Análise dos Modelos

A Taxa de Fecundidade Total (TFT) corresponde ao número médio de filhos que uma mulher teria ao terminar o período reprodutivo (CARVALHO et al, 1998). Dessa forma representa o nível da fecundidade de uma localidade sendo um dos principais determinantes da dinâmica demográfica (MOREIRA, FUSCO, 2012; SIMÕES, 2006; HORTA et al, 2003).

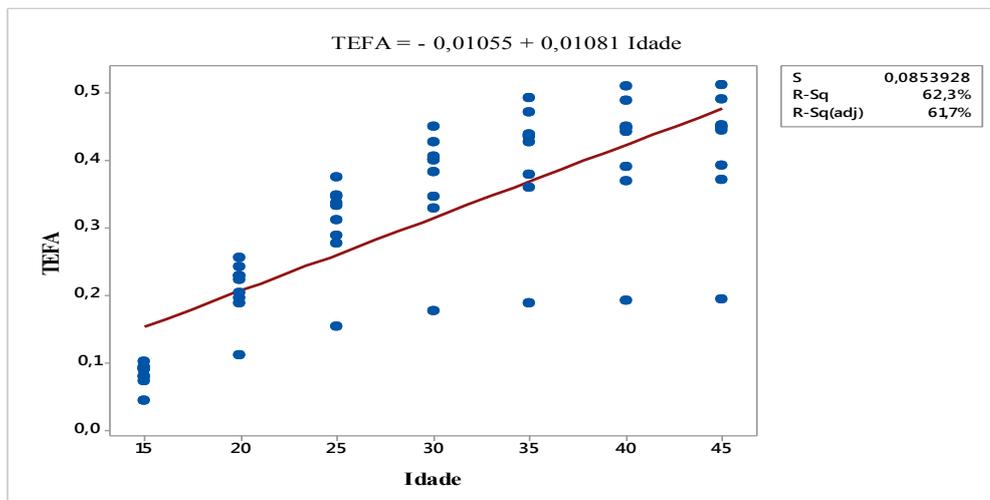
Para ilustração e comparativo dos modelos descritos, em princípio, considerou-se somente as fontes de dados do SINASC 2000. Ajustou-se um modelo sem considerar os efeitos mistos para avaliar se existe alguma relação entre as Taxas Específicas de Fecundidade acumulada (TEFA) e a idade materna. Para isso foi utilizado um modelo de regressão linear assumindo independências das observações de acordo com o seguinte modelo:

$$TEFA = \alpha + \beta \cdot Idade + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2)$$

Observa-se, no Gráfico 1, que os dados apresentam um comportamento quase linear, o modelo aparenta estar bem ajustado sendo que apresenta um  $R^2 = 61,7\%$ , o qual seria aceitável. No entanto, o modelo assume que o coeficiente angular é constante ao longo da vida reprodutiva da mulher, ou seja, que o aumento da TEFA seja constante nas idades.

No entanto, sabe-se que isto não corresponde à realidade, o que seria uma desvantagem deste ajuste para este tipo de estudo. Outra desvantagem de ajustar um modelo de regressão é que, neste caso, os valores estimados (coeficiente linear e angular) seriam os mesmos para toda a região Nordeste sem considerar as nuances de cada Estado.

Gráfico 1- Modelo de Regressão Linear das Taxa Específicas de Fecundidade Acumulada para os Estados do Nordeste Brasileiro, SINASC 2000



Outra possibilidade, então, foi considerar um modelo para cada Estado. Esta forma, implicaria em um número muito grande de parâmetros, ou seja, seriam estimados para cada Estado um coeficiente linear e um angular, resultando dessa forma em 18 parâmetros estimados (dois parâmetros para cada um dos nove Estados). Considerando os estados no modelo de regressão, são obtidas nove retas de regressão, uma para cada estado, todas com o mesmo coeficiente angular, mas diferentes estimativas para o intercepto (Figura 5). Neste caso, o modelo seria dado pela seguinte equação:

$$TEFA_{ij} = \alpha_j + \beta * idade_{ij} + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

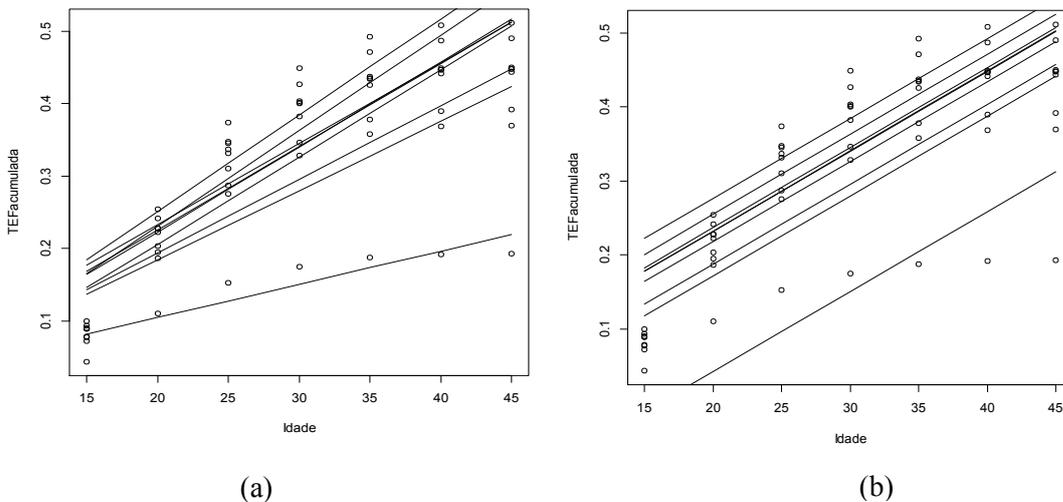
Em outra situação, poder-se-ia assumir diferentes TEF no início do período reprodutivo, ou seja, nas idades e 15 a 20 anos, mas também com ritmos de crescimento das taxas específicas acumuladas para cada Estado. O modelo seria dado pela equação:

$$TEFA_{ij} = \alpha_j + \beta_j * idade_{ij} + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

A Figura 6 ilustra estes dois modelos. Verifica-se que o modelo que contém uma reta de regressão é sem dúvida a mais simples e o modelo com nove retas de regressão com coeficientes específicos para cada um dos nove estados é o mais complexo, pois se teria mais parâmetros envolvidos.

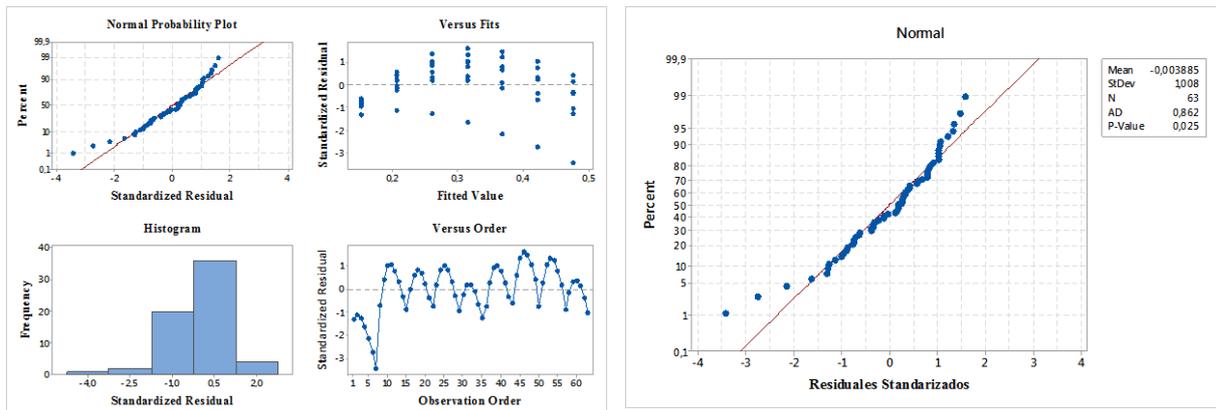
A análise dos resíduos do modelo é feita para verificar a plausibilidade das pressuposições envolvidas, quais sejam: os resíduos sejam normalmente distribuídos, com média zero, independentes (independência) e com variância constante (homogeneidade de variância ou homocedasticidade).

Figura 6 - (a) Modelo de regressão com coeficientes angulares diferentes , (b) modelo de regressão com coeficientes angulares constantes entre os estados do NEB, SINASC 2000.



Avaliando estes modelos via análise de resíduos, na Figura 7, observa-se a não normalidade dos resíduos e a falta da variância constante. Para ter mais certeza aplicou-se um teste de normalidade de Shapiro Wilk nos resíduos considerando as seguintes hipóteses:  $H_0$ : existência de Normalidade contra  $H_1$ : não existência de Normalidade; constatando-se a quebra deste pressuposto (valor- $p=0,025$ ). Portanto, o uso do modelo de regressão clássico mostrou-se não recomendável.

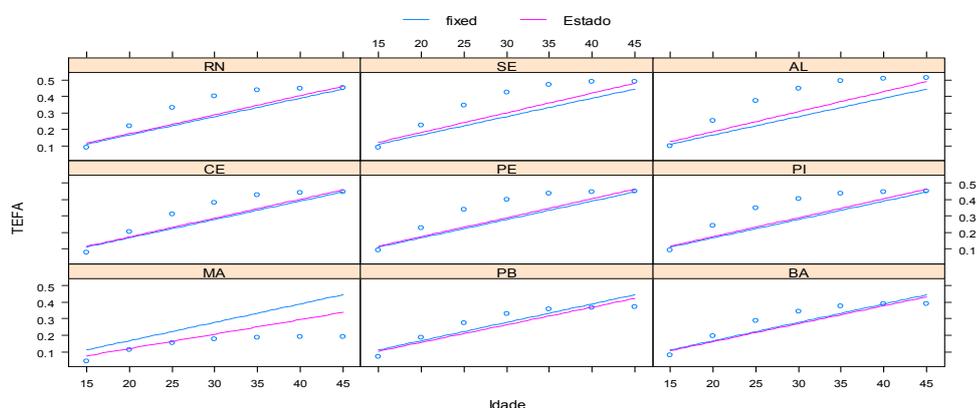
Figura 7- Avaliação dos resíduos do ajuste da reta de regressão do modelo linear, para os Estados do Nordeste Brasileiro, SINASC 2000.



É bastante comum a presença de situações em que os dados apresentam comportamentos dependentes entre medidas repetidas em uma mesma unidade observacional, ou seja, casos em que uma observação depende estocasticamente de outra. De forma geral, são chamados de dados agrupados que compreendem dados longitudinais e medidas repetidas. Neste caso o modelo linear clássico ou não linear acabam por não serem válidos uma vez que a suposição de dados não correlacionados não pode ser verificada.

Com este propósito este estudo tem a finalidade de mostrar a utilização de modelos de efeitos mistos que permitam considerar o intercepto e/ou os coeficientes da regressão como efeitos aleatórios ou fixos. Dessa forma, torna-se possível levar em consideração a presença de dados agrupados e a possível correlação existente entre as observações (no caso as TEFA em cada idade) em uma mesma unidade observacional (os estados do NEB), sendo possível modelar as correlações entre e intra-estados (Gráfico 2).

Gráfico 2- Ajuste do modelo linear de efeitos mistos por Estados do Nordeste Brasileiro, SINASC 2000.



No Gráfico 2, observa-se que o modelo linear misto consegue analisar o comportamento individual e o comportamento médio simultaneamente ao longo das faixas etárias. No entanto, esse modelo assume que o aumento da TEFA seja constante nas idades tanto para o comportamento individual e o comportamento médio. Porém, sabe-se que isto não corresponde à realidade, o que torna uma desvantagem assumir que as TEFA descrevam um comportamento linear ao longo das faixas etárias de uma população feminina no período reprodutivo (Modelo Linear Misto).

Pode-se optar por aplicar alguma transformação aos dados, no entanto, o inconveniente de uma transformação é que, além dos parâmetros perderem suas interpretações, pode alterar a estrutura e distribuição do erro. Por exemplo: se os erros do modelo original satisfizerem as suposições usuais de normalidade, independência e homogeneidade da variância, os erros do novo modelo, em geral, não satisfariam tais suposições (KHURI; CORNELL, 1987).

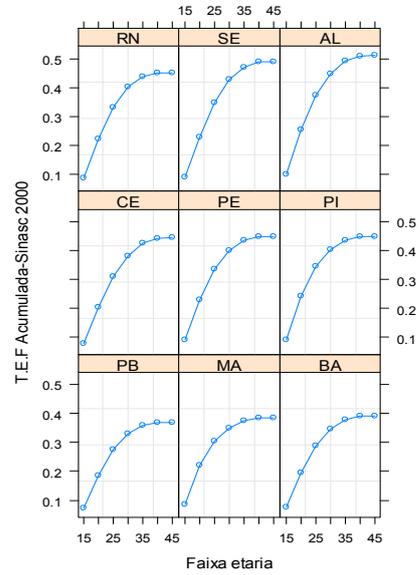
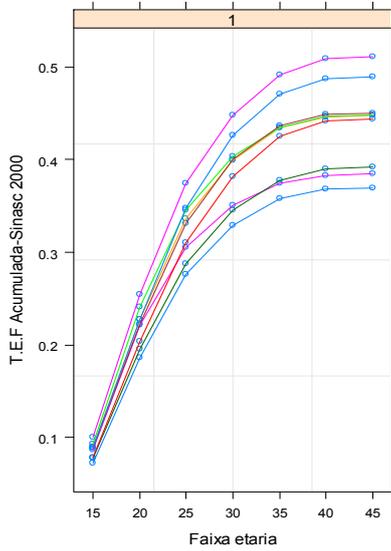
Observando o comportamento dos dados nos gráficos anteriores, percebe-se que os mesmos seguem um padrão não linear. O que remete a busca de padrões e modelos que levem a estimativas de parâmetros mais coerentes com o comportamento das Taxas Específicas de Fecundidade Acumulada para o ano 2000 e 2010.



Figura 8 - Perfis das Taxas Específicas de Fecundidade Acumulada por Estados do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2000.

SINASC 2000

- PB
- BA
- PE
- RN
- AL
- MA
- CE
- PI
- SE



CENSO DEMOGRAFICO 2000

- PE
- SE
- PB
- CE
- MA
- RN
- BA
- PI
- AL

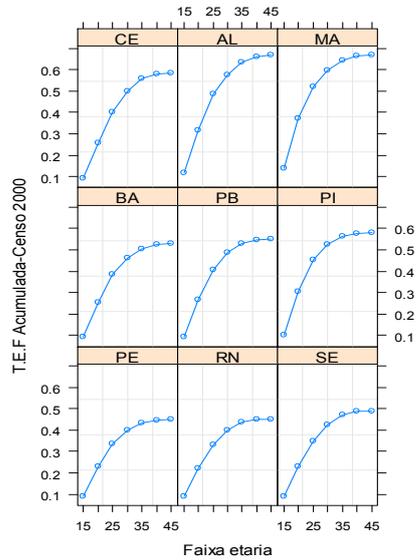
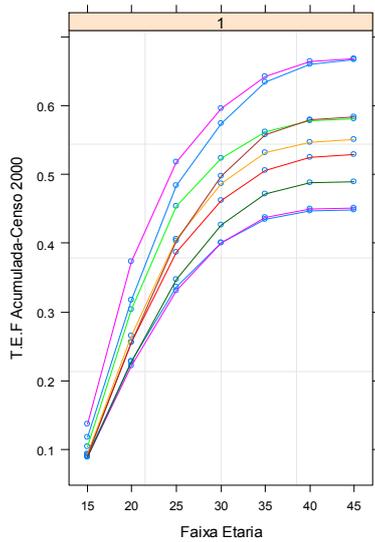
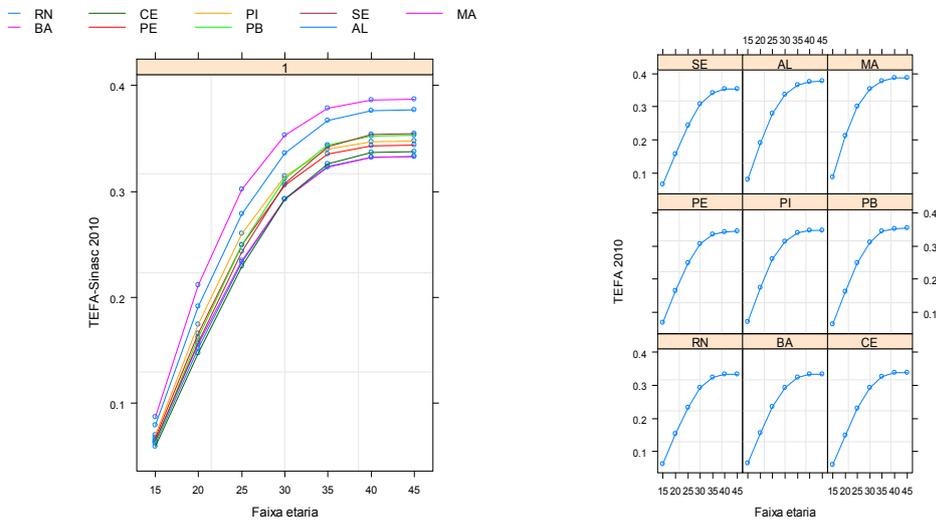
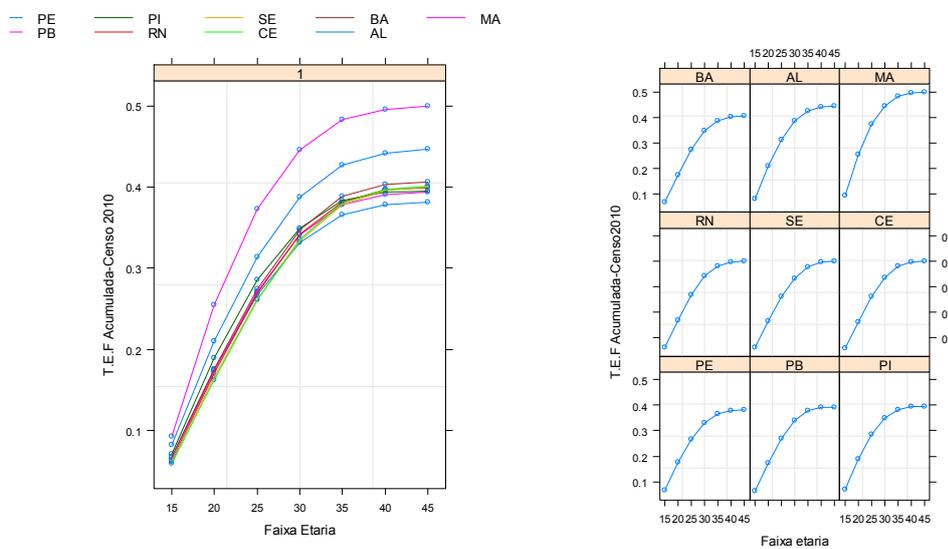


Figura 9 - Perfis das Taxas Específicas de Fecundidade Acumulada por Estados do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2010

SINASC 2010



CENSO DEMOGRAFICO 2010



Nota-se que os perfis possuem comportamentos semelhantes em ambos os anos (2000 e 2010) e ambas fontes de dados, apresentam heterogeneidade de variâncias ao longo das faixas etárias e indicam que um modelo não linear pode representar bem o padrão das TEFA. Dessa forma a escolha do modelo não linear misto usando a regressão assintótica foi motivada pelas Figuras 8 e 9.

Em modelos não lineares mistos, deve-se fornecer valores iniciais para iniciar a estimação dos parâmetros, já que a função de verossimilhança não apresenta solução analítica. Para gerar tais valores iniciais, um modelo não linear só com parâmetros de efeito fixo pode ser utilizado. Assim, pode-se utilizar a função `nls()` da biblioteca `stats` do software R, para ajustar um modelo não linear. Este tipo de função considera todos os parâmetros como de efeito fixo e não permite especificar efeito aleatório.

Dessa forma, para efeito de análise do comportamento dos estados do NEB quanto à fecundidade, aplicou-se o modelo individualmente (modelo não linear de regressão assintótica) para cada estado e a partir daí estimar os parâmetros iniciais (Asym, lrc e R0).

Desde a Tabela 2 até a Tabela 5, são apresentados os parâmetros iniciais com seus respectivos intervalos de confiança para cada um dos parâmetros do modelo (NLME) para ambas fontes de dados em dois momentos no tempo 2000 e 2010, estes intervalos darão uma idéia da variabilidade com 95% de confiança para cada um dos parâmetros.

Tabela 2 - Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - SINASC, 2000

ESTADOS DO NORDESTE	Intervalo de Confiança dos parâmetros, SINASC 2000								
	Asym*5			R0			lrc		
	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior
Maranhão	1,89	1,99	2,09	0,06	0,09	0,11	20	21	22
Piauí	2,24	2,35	2,46	0,07	0,09	0,11	20	21	23
Ceará	2,28	2,44	2,60	0,05	0,07	0,09	22	23	25
Rio Grande do Norte	2,29	2,42	2,56	0,06	0,08	0,10	21	22	24
Paraíba	1,84	1,97	2,10	0,05	0,07	0,09	20	22	24
Pernambuco	2,26	2,38	2,51	0,06	0,08	0,11	21	22	23
Alagoas	2,61	2,74	2,88	0,07	0,09	0,11	21	22	24
Sergipe	2,52	2,67	2,82	0,06	0,08	0,10	21	23	24
Bahia	1,97	2,10	2,23	0,05	0,07	0,09	21	22	24

Asym\*5 = TFT

Tabela 3 - Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - Censo Demográfico, 2000.

ESTADOS DO NORDESTE	Intervalo de Confiança dos parâmetros, Censo Demográfico 2000								
	Asym*5			R0			lrc		
	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior
Maranhão	3,34	3,47	3,60	0,11	0,13	0,16	20	21	22
Piauí	2,90	3,03	3,16	0,07	0,10	0,12	20	21	22
Ceará	3,02	3,22	3,42	0,06	0,08	0,11	22	23	25
Rio Grande do Norte	2,26	2,42	2,59	0,06	0,08	0,11	21	22	24
Paraíba	2,79	2,94	3,10	0,06	0,09	0,11	21	22	23
Pernambuco	2,23	2,38	2,53	0,06	0,08	0,11	20	22	24
Alagoas	3,41	3,58	3,75	0,09	0,11	0,14	21	22	24
Sergipe	2,49	2,67	2,86	0,06	0,08	0,11	21	23	25
Bahia	2,67	2,83	2,99	0,06	0,09	0,11	21	22	24

Asym\*5 = TFT

Tabela 4- Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - SINASC, 2010

ESTADOS DO NORDESTE	Intervalo de Confiança dos parâmetros, SINASC 2010								
	Asym*5			R0			lrc		
	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior
Maranhão	1,91	2,03	2,15	0,06	0,08	0,10	20	21	23
Piauí	1,73	1,86	2,00	0,04	0,06	0,08	20	22	24
Ceará	1,70	1,89	2,07	0,03	0,05	0,07	21	23	26
Rio Grande do Norte	1,67	1,84	2,01	0,03	0,06	0,07	21	23	26
Paraíba	1,78	1,94	2,10	0,03	0,06	0,07	21	23	25
Pernambuco	1,71	1,86	2,02	0,04	0,06	0,08	21	22	25
Alagoas	1,88	2,02	2,16	0,05	0,07	0,09	20	22	24
Sergipe	1,79	1,98	2,16	0,04	0,06	0,08	21	23	26
Bahia	1,66	1,83	2,00	0,04	0,06	0,08	21	23	26

Asym\*5 = TFT

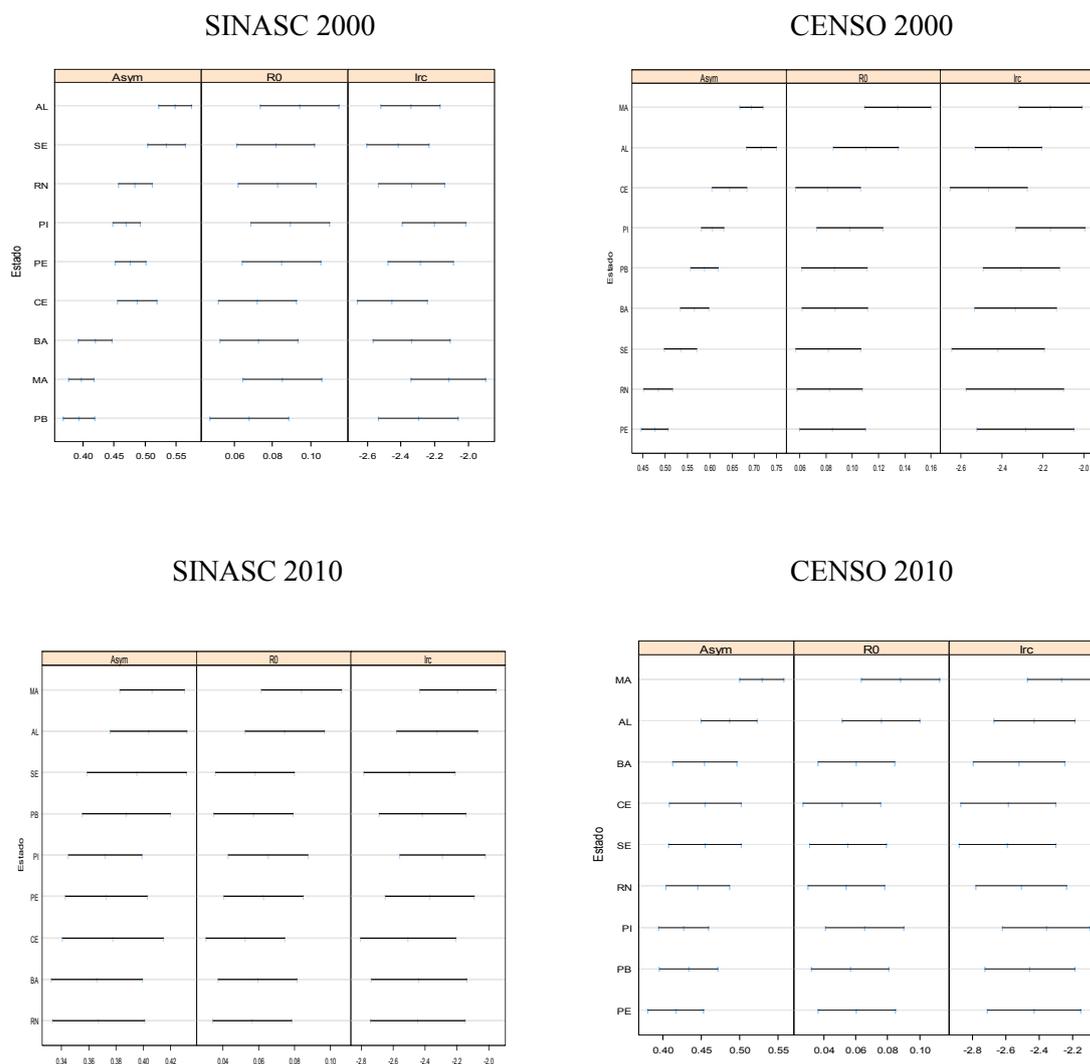
Tabela 5- Intervalos de Confiança dos parâmetros NLME, segundo os Estados da Região Nordeste - Censo Demográfico, 2010

ESTADOS DO NORDESTE	Intervalo de Confiança dos parâmetros, Censo Demográfico 2010								
	Asym*5			R0			Irc		
	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior	Limite inferior	Estimado	Limite superior
Maranhão	2,50	2,65	2,79	0,06	0,09	0,11	20	22	23
Piauí	1,97	2,13	2,30	0,04	0,07	0,09	21	22	25
Ceará	2,05	2,28	2,51	0,03	0,05	0,08	22	24	27
Rio Grande do Norte	2,02	2,23	2,44	0,03	0,05	0,08	21	23	26
Paraíba	1,98	2,17	2,36	0,03	0,06	0,08	21	23	26
Pernambuco	1,90	2,09	2,27	0,04	0,06	0,08	21	23	25
Alagoas	2,25	2,43	2,61	0,05	0,08	0,10	21	23	25
Sergipe	2,04	2,28	2,51	0,03	0,06	0,08	22	24	27
Bahia	2,06	2,27	2,48	0,04	0,06	0,08	22	24	26

Asym\*5 = TFT

A continuação a Figura 10 apresenta os intervalos de 95% de confiança para os três parâmetros (Asym, R0, Irc), para cada um dos nove estados da Região Nordeste e para os anos 2000 e 2010. Neste é possível verificar que, em ambas fontes de dados, nem todos os intervalos se sobrepõem nos três parâmetros. O que remete avaliar a necessidade de incorporar os efeitos aleatórios a cada um dos parâmetros.

Figura 10 - Intervalos 95% de confiança dos parâmetros por estado do Nordeste Brasileiro, SINASC e CENSO DEMOGRAFICO, 2000 - 2010



Utilizando a estratégia *top-down*, inicialmente escolhe-se um modelo com os parâmetros *Asym*, *R0* e *lrc* para todos os estados, considerando a inclusão de efeito fixo e aleatório em ambos. Para os dados oriundos do SINASC 2000 e Censo Demográfico 2000, a aparente variabilidade nestes intervalos sugere a inclusão de efeito aleatório no parâmetro *Asym* do Modelo Não Linear de Efeitos Mistos, algumas estruturas para a matriz de variância e covariância foram testadas e comparando-se o valor do *Akaike Information Criterion* (AIC), constatou-se que a matriz não estruturada para o efeito aleatório foi a que representou o menor valor para AIC sendo de -324,1728 para os dados provenientes do SINASC 2000 e para os dados provenientes do Censo Demográfico 2000 encontrou-se um valor AIC de -272,4124, revelando a importância da inclusão do efeito aleatório para o parâmetro *Asym* no ano 2000 em ambas fontes de

dados. Os testes de comparação entre os modelos fixo e aleatório foram favoráveis à inclusão dos parâmetros de efeitos aleatórios no modelo final para ambas fontes de dados (valor-p < 0,0001).

Da mesma forma no que refere para o ano 2010, par os dados oriundos do SINASC e para os dados do Censo Demográfico, estruturas para a matriz de variância e covariância foram testadas e comparando-se o valor do *Akaike Information Criterion* (AIC), constatou-se que a matriz não estruturada para o efeito aleatório foi a que representou o menor valor para AIC sendo de -321,5236 isto para os dados provenientes do Censo Demográfico 2010, e os dados do SINASC 2010 apresentou um valor AIC de -343,218, dessa maneira revelando a importância da inclusão do efeito aleatório para o parâmetro *Asym* para o ano em referência, em ambas fontes de dados. Os testes de comparação entre os modelos fixo e aleatório foram favoráveis à inclusão dos parâmetros de efeitos aleatórios no modelo final (valor-p < 0,0007).

No que refere à análise de resíduos do modelo não linear de efeitos mistos final, não revelou nenhuma violação dos pressupostos de normalidade em ambas fontes de dados (SINASC e Censo Demográfico) para ambos os anos isto pode ser confirmado com o teste de normalidade de Shapiro-Wilk (Tabela 6). Segundo Sartorio (2013) na literatura encontram-se poucos estudos sobre análise de resíduos para modelos não lineares mistos. Embora, a pressuposição de homocedasticidade dos resíduos foi atendida, pois apresentou uma distribuição em torno de zero, com variância aproximadamente constante (Figura11 e Figura12).

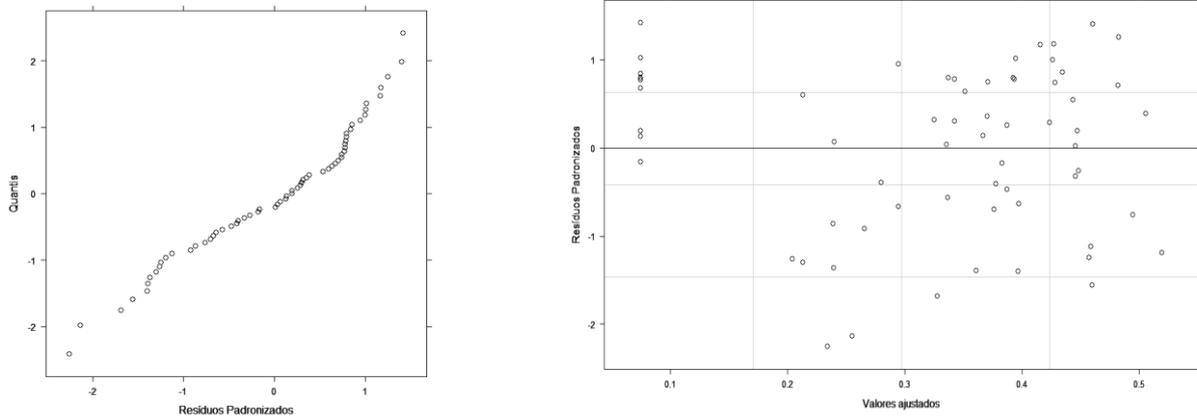
Tabela 6- Teste de Normalidade de Shapiro Wilk, segundo fontes de dados, estados da Região Nordeste, 2000 e 2010

Fontes de dados	Teste Shapiro Wilk	Valor-p
SINASC 2000	0,9708	0,1389
CENSO DEMOGRAFICO 2000	0,9795	0,3759
SINASC 2010	0,9723	0,1666
CENSO DEMOGRAFICO 2010	0,9809	0.4326

A utilização de resíduos para diagnosticar o modelo mais coerente permitiu a visualização de pontos discrepantes, que por opção, não foram retirados. A qualidade do ajuste pode ser visualizada na Figura 13. Observa-se que os valores preditos pelo Modelo Não Linear de Efeitos Mistos ajustaram-se melhor aos valores das TFT observados.

Figura 11- Teste de normalidade dos resíduos para NLME e análise de resíduos versus valores ajustados - SINASC e Censo Demográfico 2000

### SINASC 2000



### CENSO DEMOGRAFICO 2000

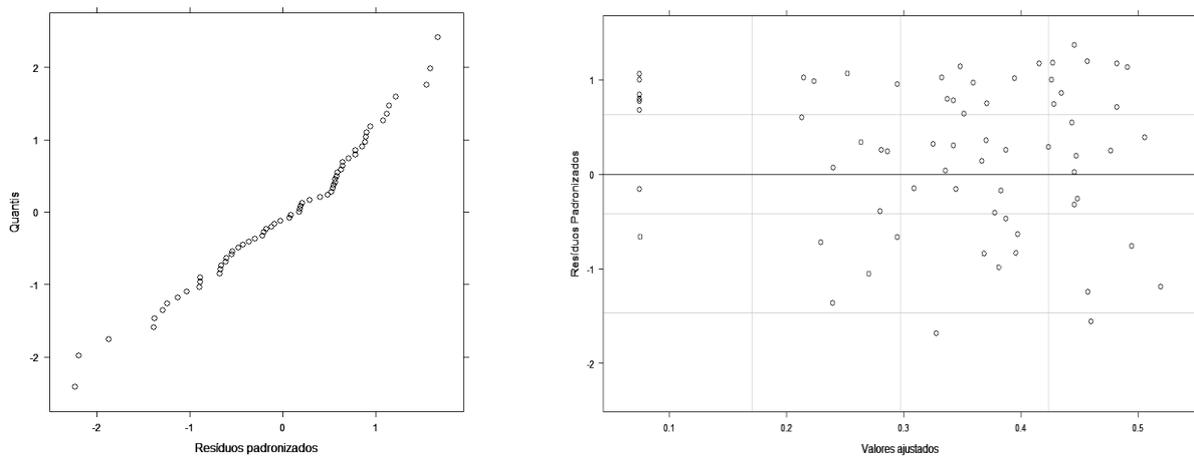
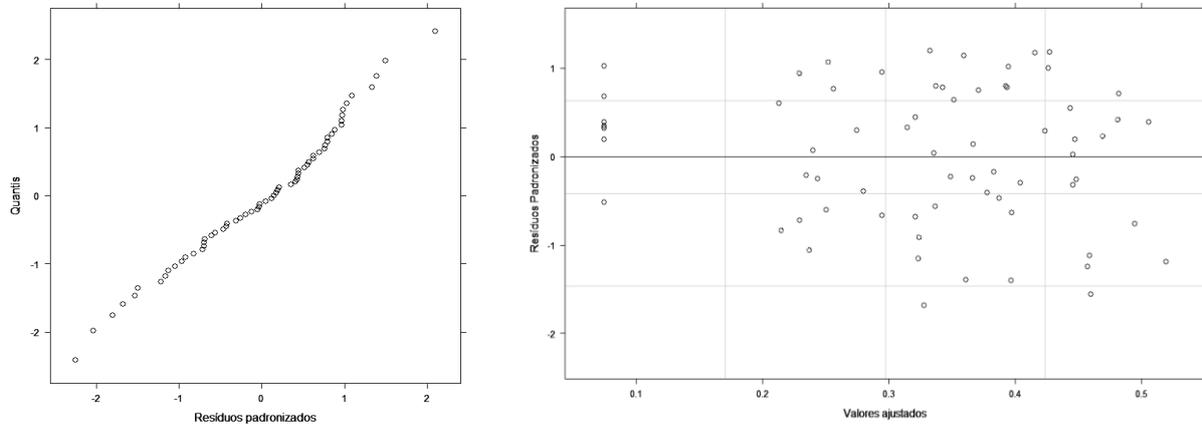
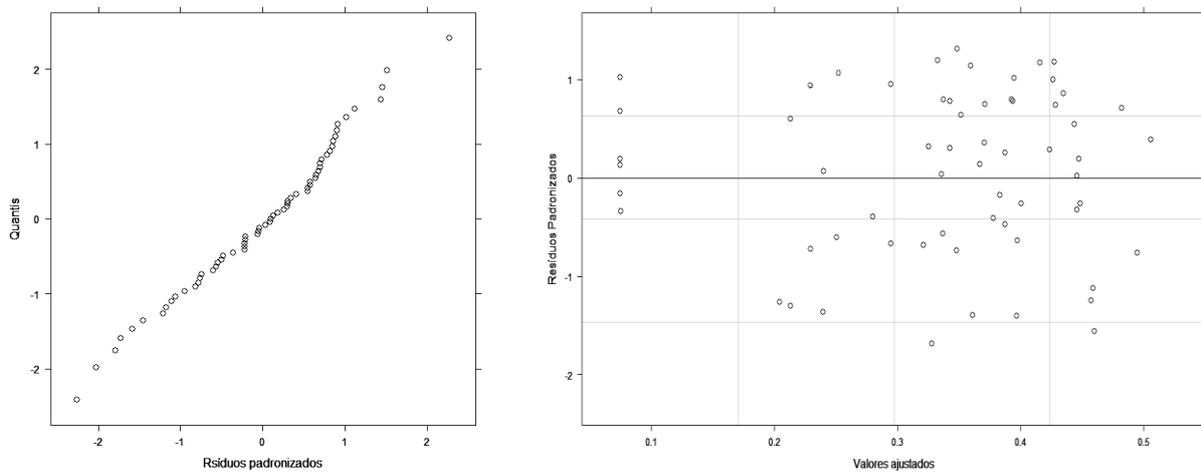


Figura 12- Teste de normalidade dos resíduos para NLME e análise de resíduos versus valores ajustados - SINASC e Censo Demográfico 2010

### SINASC 2010



### CENSO DEMOGRAFICO 2010



Após verificação da qualidade de ajuste, prosseguiu-se a análise das estimativas dos parâmetros do modelo não linear de efeitos mistos apresentados nas Tabelas 7 a 10.

Figura 13- Padrão das TFT Ajustado com Modelo Não linear de Efeitos Mistos considerando base nas TEFA da SINASC e Censo Demográfico, 2000-2010

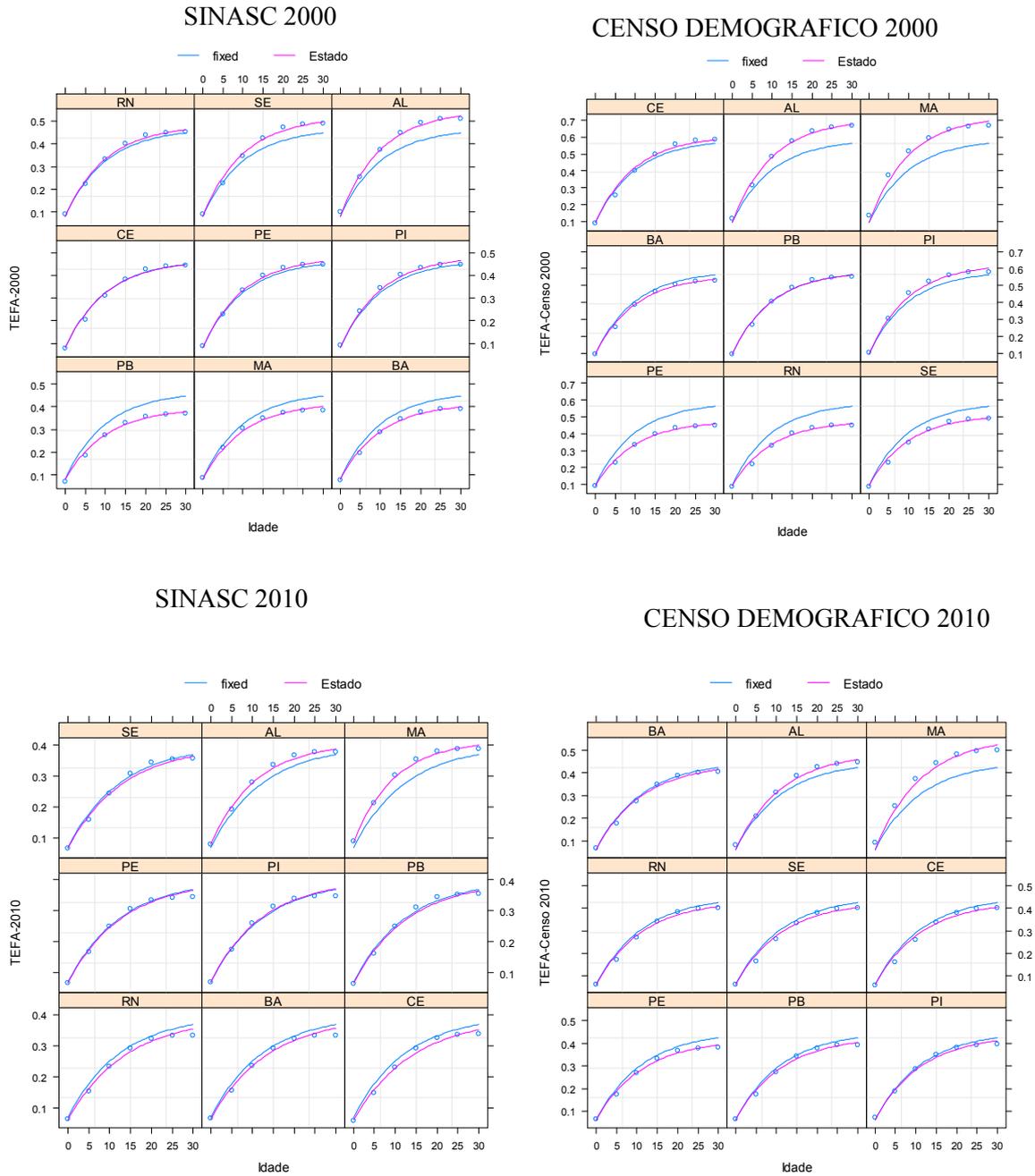


Tabela 7 - Parâmetros estimados e erros-padrão do modelo com efeitos fixos e com efeitos aleatórios, para os Estados da Região Nordeste, SINASC 2000

<b>Efeitos fixos</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>	<b>Erro padrão</b>	<b>Valor-p</b>
Asym	2,15	2,20	2,30	0,017	<.0001
R0	0,06	0,07	0,08	0,004	<.0001
lrc	21,0	21,5	22,0	0,034	<.0001
<b>Intervalos de Confiança 95%</b>					
<b>Efeitos aleatórios</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>		
$\sigma$ (Asym)	0,0296	0,0474	0,0757		
Residual (intra-grupo)	0,0077	0,0218	0,0616		
-2ln verossimilhança	176.6093				
AIC	-343.2186				

Tabela 8- Parâmetros estimados e erros-padrão do modelo com efeitos fixos e com efeitos aleatórios, para os Estados da Região Nordeste, Censo Demográfico 2000

<b>Efeitos fixos</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>	<b>Erro padrão</b>	<b>Valor-p</b>
Asym	2,69	2,75	2,97	0,033	<.0001
R0	0,07	0,08	0,09	0,004	<.0001
lrc	21,1	21,8	22,1	0,037	<.0001
<b>Intervalos de Confiança 95%</b>					
<b>Efeitos aleatorios</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>		
$\sigma$ (Asym)	0,0542	0,0874	0,1411		
Residual (intra-grupo)	0,0134	0,0162	0,0197		
-2ln verossimilhança	141,2062				
AIC	-272,4124				

Tabela 9- Parâmetros estimados e erros-padrão do modelo com efeitos fixos, para os Estados da Região Nordeste, SINASC 2010

<b>Efeitos fixos</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>	<b>Erro padrão</b>	<b>Valor-p</b>
Asym	1,94	1,99	2,06	0,006	<.0001
R0	0,06	0,07	0,09	0,004	<.0001
lrc	22,1	23,1	23,5	0,048	<.0001
<b>Intervalos de Confiança 95%</b>					
<b>Efeitos aleatórios</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>		
$\sigma$ (Asym)	0,0251	0,0411	0,0673		
Residual (intra-grupo)	0,0134	0,0162	0,0197		
-2ln verossimilhança	190,1193				
AIC	-358,2385				

Tabela 10- Parâmetros estimados e erros-padrão do modelo com efeitos fixos, para os Estados da Região Nordeste, Censo Demográfico 2010

<b>Efeitos fixos</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>	<b>Erro padrão</b>	<b>Valor-p</b>
Asym	2,05	2,10	2,18	0,016	<.0001
R0	0,07	0,08	0,09	0,004	<.0001
lrc	21,7	22,5	22,8	0,047	<.0001
<b>Intervalos de Confiança 95%</b>					
<b>Efeitos aleatórios</b>	<b>Limite inferior</b>	<b>Estimativa</b>	<b>Limite Superior</b>		
$\sigma$ (Asym)	0,0270	0,0441	0,0719		
Residual (intra-grupo)	0,0114	0,0139	0,0168		
-2ln verossimilhança	155,6564				
AIC	-301,3129				

## 4.2 Análise da Fecundidade no Nordeste a partir do modelo NLME

A Taxa de Fecundidade Total (TFT) representa o nível da fecundidade de uma localidade ou a condição reprodutiva média das mulheres de um determinado local, sendo considerado um dos principais determinantes da dinâmica demográfica, (CARVALHO et al.,1994; HORTA et al., 2003). Ressalta-se que o parâmetro referente à estimativa da Taxa de Fecundidade Total foi obtido ao multiplicar o parâmetro *Asym* pela amplitude do grupo etário (5 anos).

A Tabela 11 apresenta a Taxa de Fecundidade Total (*Asym*\*5) obtido pelo modelo individual (modelo não linear de regressão assintótica) e as TFT provenientes do SINASC e do Censo Demográfico (Método de estimação da fecundidade de Brass). Vale salientar que as informações do SINASC foram avaliadas pelo cálculo da Taxa de Fecundidade Total (TFT), obtida calculando-se diretamente as Taxas Específicas de Fecundidade – TEFs, como o quociente entre o número de nascimentos vivos de mães em determinado grupo etário e o número de mulheres nesse grupo etário.

Tabela 11- Parâmetros Não Linear (*Asym*), Modelo de Regressão Assintótica, por Estados do Nordeste-Brasil, SINASC e CENSO DEMOGRAFICO, 2010

Estados	TFT 2000				TFT 2010			
	<i>Asym</i> *5 (SINASC)	SINASC	<i>Asym</i> *5 (CENSO )	CENSO (BRASS)	<i>Asym</i> *5 (SINASC)	SINASC	<i>Asym</i> *5 (CENSO )	CENSO (BRASS)
Maranhão	1,99	1,92	3,47	3,21	2,03	1,94	2,65	2,50
Piauí	2,35	2,24	3,03	2,67	1,86	1,74	2,13	1,97
Ceará	2,44	2,22	3,22	2,84	1,89	1,69	2,28	2,00
Rio Grande do Norte	2,42	2,25	2,42	2,54	1,84	1,66	2,23	2,00
Paraíba	1,97	1,85	2,94	2,54	1,94	1,77	2,17	1,97
Pernambuco	2,38	2,24	2,38	2,48	1,86	1,72	2,09	1,91
Alagoas	2,74	2,56	3,58	3,14	2,02	1,88	2,43	2,23
Sergipe	2,67	2,45	2,67	2,74	1,98	1,77	2,28	2,00
Bahia	2,10	1,96	2,83	2,50	1,83	1,67	2,27	2,03

Fonte: SINASC2000/2010, TFT; CENSO2000/2010 ,TFT (BRASS), Wingerter (2013).

Considerando o ano 2000, observa-se que a TFT ( $Asym*5$ ) para os estados de Rio Grande do Norte, Pernambuco e Sergipe mostram semelhanças nas Taxas de Fecundidade Total entre a fonte de dados do SINASC 2000 e a fonte de dados do Censo Demográfico 2000 (Método do Brass). No entanto observa-se que para o estado de Maranhão o parâmetro ( $Asym*5$ ) calculado com ambas fontes de dados (SINASC e CENSO), apresenta grande diferença entre suas TFT. Wingerter (2013) assinalou que a cobertura da TFT calculada pelo SINASC em relação ao Censo para o Nordeste como um todo, no ano 2000 revalaram que os dados do SINASC mostraram-se inadequados para a Região, segundo classificação proposta pela Portaria do Ministério da Saúde N° 201<sup>12</sup>, de 03/11/2009, apresentando uma cobertura de 79,1%.

Os problemas relacionados ao SINASC referem-se ao subregistro das informações, que acontece quando: não há a emissão da Declaração de Nascido Vivo (DNV) ou há o extravio da via branca antes de ser digitada no Sistema, bem como ao período de implantação do SINASC, que para os estados do Nordeste foi posterior a 1990 e que a consolidação do Sistema pode não estar plena para os dados de 2000 (BARBOSA, FREIRE, 2004; SOUZA, 2004; FRIAS et al., 2007; PEDRAZA, 2012; PEREIRA et al., 2013).

Para o ano 2000, em ambas fontes de dados, o estado de Alagoas foi aquele que apresentou a maior TFT do Nordeste, em sentido contrário o estado da Paraíba e Pernambuco foram aqueles que apresentaram as menores TFT a partir dos dados do SINASC e Censo Demográfico respectivamente.

Levando em consideração os dados do ano 2010, observa-se que o parâmetro assintótico ( $Asym*5$ ), calculado a partir de ambas fontes de dados, apresenta em todos os estados do Nordeste uma fecundidade abaixo do chamado "nível de reposição populacional" (2,1 filhos por mulher). Este valor é considerado por garantir a substituição das gerações, isto quer dizer que duas crianças substituem os pais e a fração 0,1 é necessária para compensar os indivíduos que morrem antes de atingir a idade reprodutiva (CARVALHO et al., 1994). Além do envelhecimento futuro da população (CARVALHO, BRITO, 2005; SIMÕES, 2006; MOREIRA, FUSCO, 2012).

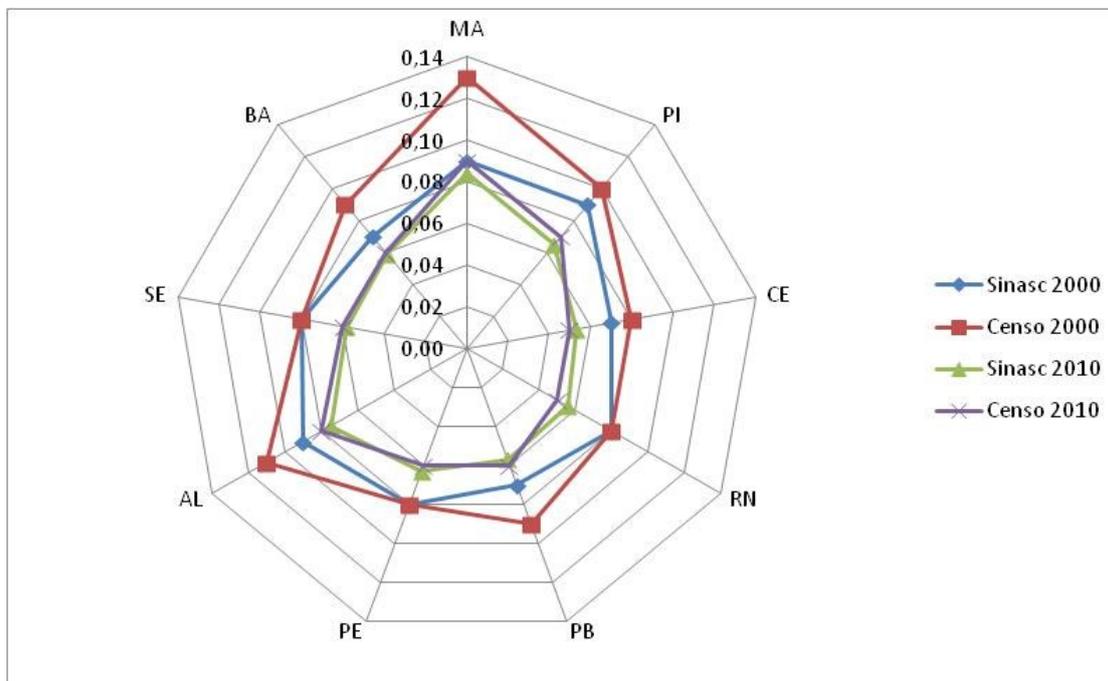
Considerando ambas fontes de dados (Tabela 11), vale salientar que para o ano 2010 o estado de Maranhão e Alagoas foram aqueles que apresentaram a maior TFT, 2,65 e 2,43 respectivamente, em comparação com outros estados do Nordeste. Em sentido contrário os estados que apresentaram as menores TFT foram Bahia e Pernambuco. Segundo dados do Censo Demográfico 2010, Moreira e Fusco (2012) assinalaram que os estados de Maranhão e Alagoas apresentaram Taxas de Fecundidade Total acima do nível de reposição, apenas Pernambuco apresentou estimativa abaixo do nível de reposição considerando os dados do Censo 2010.

Vale salientar que as TFT do modelo NLME, calculadas a partir dos dados do SINASC e Censo Demográfico 2010, ressaltam que a maioria dos estados do Nordeste apresenta estimativas abaixo do nível de reposição. Para Lima (2010) este fenômeno pode ser consequência de diferentes políticas de planejamento familiar que foram implantadas, com a esterilização de muitas mulheres (ligadura de trompas) e uma grande disseminação do uso da pílula como método anticoncepcional que levou a uma acentuação do processo de redução das taxas de fecundidade mais rápido.

A Taxa Específica de Fecundidade -TEF (R0) no início do período reprodutivo das mulheres residentes, corresponde ao número médio de filhos nascidos vivos, tidos por uma mulher, na idade de 15 a 19 anos. Considerando ambas fontes de dados, no ano 2000, observa-se que os estados de Maranhão e Piauí apontaram as maiores  $TEF_{15}$  (R0), há por outro lado evidências de que, para os estados de Rio Grande do Norte, Pernambuco e Sergipe, as  $TEF_{15}$  (R0) revelados pelo SINASC mostram-se similares àquelas obtidas pelos dados censitários (Tabela 2 e 3). No ano 2010, os resultados apontaram o estado de Maranhão como aquele com a maior  $TEF_{15}$  (R0) dentro da região Nordeste, em ambas fontes de dados, no entanto, acontecendo uma diminuição da  $TEF_{15}$  (R0) após uma década. Quando da comparação entre os dados do SINASC de 2010 e o Censo Demográfico de 2010, o SINASC já mostra uma  $TEF_{15}$  (R0) mais próxima daquela revelada pelo Censo Demográfico (Gráfico 4).

Vale destacar o resultado obtido por Moreira e Fusco (2012), que para o ano 2010 o estado de Maranhão foi identificado como o estado de maiores níveis de fecundidade na região, assemelhando-se aos níveis da fecundidade vigentes na região Norte.

Gráfico 4 - Taxa Específica de Fecundidade da mulher de 15-19 anos de idade por Estados do Nordeste Brasileiro, segundo parâmetro R0, SINASC e CENSO 2000 - 2010



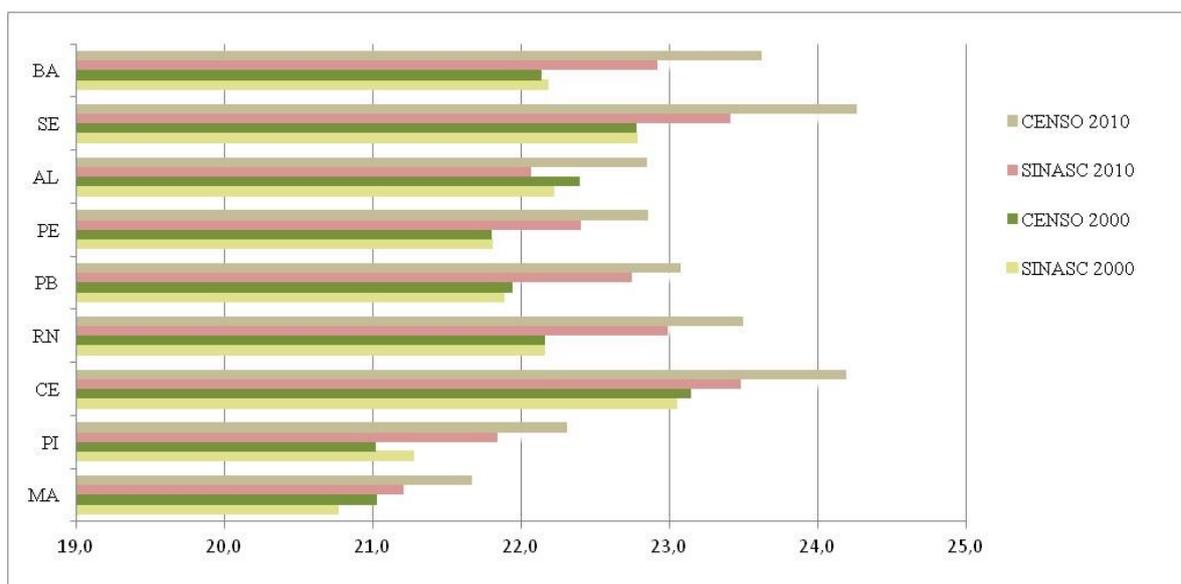
Neste trabalho com uso da metodologia de modelos não lineares mistos (NLME) encontrou-se um parâmetro ( $\text{lrc}$ ), muito importante por refletir a idade com que a mulher atinge a metade da TFT, em todo seu período reprodutivo. Dessa forma, constatou-se que a população feminina em todos os estados do Nordeste atinge a metade da TFT nas idades compreendidas entre os 20 e 24 anos e observou-se também que a postergação das mulheres em ter filhos acontece de forma independente do estado de procedência.

Considerando o ano 2000 e ambas fontes de dados, encontrou-se que os estados de Maranhão e Piauí foram aqueles que atingiram a metade da TFT com a idade mais jovem compreendida entre 21 e 22 anos, por outro lado os estados de Ceará e Sergipe foram aqueles que apresentaram a maior idade (23 anos) para atingir a metade da TFT.

Vale salientar que a década de 1990 foi marcada por uma intensificação do rejuvenescimento da fecundidade no país. Em 1980 cabia às mulheres de 25 a 29 anos o maior número médio de filhos, dentre os grupos etários na faixa reprodutiva. O primeiro deslocamento desta cúspide para o grupo mais jovem, de 20 a 24 anos, ocorre em 1991, mantendo-se em 2000 (BERQUÓ, CAVENAGHI, 2004).

Da mesma forma para o ano 2010 e em ambas fonte de dados observa-se que passado um decênio, novamente o estado de Maranhão foi aquele que atingiu a metade da TFT com a idade mais jovem (21 anos) o qual merece atenção quanto a saúde e desenvolvimento das jovens mães e de seus filhos.

Gráfico 5 - Velocidade (idade) com que a mulher atinge a metade da Taxa de Fecundidade Acumulada por Estados do Nordeste Brasileiro, SINASC e CENSO 2000 - 2010



Para o ano 2010, percebe-se que, além dos estados de Alagoas e Sergipe, o estado do Maranhão também mostrou as maiores TFT da Região Nordeste. O estado do Maranhão apresentou também a maior  $TEF_{15}$ , além de atingir a metade do total de filhos com idades mais jovens aos 21 anos de idade.

Na Tabelas 12 e 13 foram calculadas as variações percentuais (2000/2010) para os parâmetros Asym (TFT) para ambas fontes de dados, segundo os estados da Região Nordeste. Para a fonte de dado do SINASC a variação percentual no ano 2010, em relação ao ano 2000, apresentou os estados de Alagoas e Sergipe como aqueles com a

maior queda das TFT sendo de 26,34% e 26,09% respectivamente. Por outro lado, observa-se que o estado de Maranhão apresentou um incremento de 2,04% em 2010 com relação ao ano 2000, vale destacar o estudo de Wingerter (2013), no qual para a fonte de dados do SINASC o estado de Maranhão apresenta também um incremento da TFT de 1,04% em 2010 com relação ao ano 2000. Apesar de uma melhora constante na qualidade dos dados do SINASC, suspeita-se que a causa disso pode ser devido à falha na captação das informações do SINASC naquele estado, para o ano 2000.

Para a fonte de dado do Censo Demográfico a variação percentual no ano 2010 em relação ao ano 2000 apresentou os estados de Alagoas e Piauí como aqueles com a maior queda das TFT sendo de 32,08% e 29,62% respectivamente, por outro lado o estado do Rio Grande do Norte apresentou a menor queda da TFT sendo de 7,95%

Tabela 12- Taxa de Fecundidade Total e Variação Porcentual, segundo os Estados da Região Nordeste - SINASC, 2000 e 2010

Estados	Taxa de Fecundidade Total		Variação %
	2000	2010	2010
Maranhão	1,99	2,03	2,04
Piauí	2,35	1,86	-20,88
Ceará	2,44	1,89	-22,58
Rio Grande do Norte	2,42	1,84	-24,21
Paraíba	1,97	1,94	-1,67
Pernambuco	2,38	1,86	-21,80
Alagoas	2,74	2,02	-26,34
Sergipe	2,67	1,98	-26,09
Bahia	2,10	1,83	-12,91

Tabela 13- Taxa de Fecundidade Total e Variação Porcentual, segundo os Estados da Região Nordeste - CENSO DEMOGRAFICO, 2000 e 2010

Estados	Taxa de Fecundidade Total		Variação % 2010
	2000	2010	
Maranhão	3,47	2,65	-23,68
Piauí	3,03	2,13	-29,62
Ceará	3,22	2,28	-29,21
Rio Grande do Norte	2,42	2,23	-7,95
Paraíba	2,94	2,17	-26,22
Pernambuco	2,38	2,09	-12,51
Alagoas	3,58	2,43	-32,08
Sergipe	2,67	2,28	-14,90
Bahia	2,83	2,27	-19,66

De acordo com o IBGE(2010) e Moreira e Fusco (2012), o maior decréscimo na TFT entre os anos 2000 e 2010 no Brasil foi observada na Região Nordeste com a maior variação relativa de TFT, apresentando um declínio de 2,69 para 2,07 filhos por mulher, quer dizer, uma queda de 23,4% juntando-se assim, ao final da década, a patamares de fecundidade similares às regiões Sudeste, Sul e Centro-Oeste.

Therborn (2006), propõe que para a compreensão da queda da fecundidade, evidencia-se o senso de domínio pessoal realizado através do controle preventivo da natalidade e o senso dos benefícios de ter ou não filhos, apontando como agentes chaves para que ocorra a mudança da fecundidade a posição social da mulher no sistema familiar; e a mudança cultural do curso da vida pessoal, ou seja, a percepção dos custos e benefícios dos filhos.

Essa condição esta disseminada pelo Brasil e nesse sentido, considerando a fonte de dados do Censo Demográfico, fica reforçada a observação de que, independentemente do estado de procedência das mulheres, houve redução das TFT e TEF, entre 2000 e 2010, confirmando que as mulheres postergaram em média por um ano ter um filho, refletindo no parâmetro Asym (TFT) e lrc em todos os estados do NEB.

Wingerter (2013) apontou que, no ano de 2010, a captação do SINASC melhorou consideravelmente para a Região Nordeste como um todo, o número de nascimentos estimados a partir do Censo foi de aproximadamente um milhão de nascimentos para o Nordeste, e o SINASC captou pouco mais de 841 mil, perfazendo

uma cobertura de 86,09%, porém persistindo o estado do Maranhão com a mais baixa taxa de cobertura (79,87%).

Vale salientar que os resultados permitem concluir que, o Censo Demográfico mostrou TFTs superiores frente àquelas trabalhadas com a fonte de dados do SINASC para a maioria dos estados Nordestinos. Para Moreira e Fusco (2014) a fecundidade nordestina no próximo decênio completaria sua trajetória para níveis abaixo da reposição.

Considerando a fonte de dados do SINASC e ao Nordeste como um todo, observa-se que a TFT ( $Asym*5$ ) no ano 2000 foi de 2,20 e teve uma queda de 10% para o ano 2010, apresentando uma TFT ( $Asym*5$ ) de 1,99, sendo evidente a redução no número médio de filhos tidos por mulher.

Por outro lado, levando em consideração as fontes de dados do Censo Demográfico, observa-se que a TFT em 2000 foi de 2,75 e teve uma queda de 24% no ano 2010, apresentando uma TFT de 2,10, observa-se que os resultados das TFT permitem concluir que, o Censo Demográfico mostrou TFT superiores frente àquelas trabalhadas com a fonte de dados do SINASC para o Nordeste como um todo. Nesse sentido o Nordeste, uma das mais pobres regiões do Brasil, passaria a experimentar decréscimos populacionais adicionais advindos da redução da fecundidade num horizonte relativamente próximo (MOREIRA E FUSCO, 2014)

Com relação à  $TEF_{15}$  (R0), pode-se observar que passada uma década a taxa no início do período reprodutivo das mulheres do Nordeste não apresentaram severas diferenças. No que refere ao parâmetro  $Irc$ , constatou-se que em 2000, a população feminina da Região Nordeste, alcançou a metade da TFT ao redor dos 22 anos de idade e após uma década (2010), subiu para 23 anos de idade, dessa forma observando em média uma leve postergação das mulheres em ter filhos. Leite (1994) comentou que no Nordeste brasileiro as mulheres mais velhas parecem haver aproveitado melhor dos serviços de planejamento familiar do que as mulheres jovens, resultando dessa forma um aumento da contribuição das adolescentes na Taxa de Fecundidade Total.

Entre 2000 e 2010, a Taxa de Fecundidade Total nordestina experimentou uma redução de 23,4%, a maior variação relativa observada no país, declinando de 2,69 para 2,06 filhos por mulher, juntando-se, assim, ao final da década, às regiões Sudeste, Sul e Centro-Oeste, apresentando fecundidade abaixo do nível de reposição (MOREIRA,

2012). Faria (1997) comentou que programas de televisão, especialmente as telenovelas brasileiras, disseminaram imagens que podem ser interpretadas como defensoras da redução da fecundidade.

Considerando a fonte de dados do Censo Demográfico, observou-se que para o ano 2000 nos estados de Ceará, Bahia e Paraíba, as duas curvas ajustadas (populacional e individual) encontram-se sobrepostas, indicando que estes estados do NEB têm um comportamento médio populacional de 2,75 filhos por mulher. Para o ano 2010, os estados de Bahia, Rio Grande do Norte e Piauí apresentaram as curvas ajustadas sobrepostas, dessa forma indicando que estes estados apresentam um número médio de 2,10 filhos por mulher, enquanto que os demais estados não mencionados possuem comportamento médio populacional diferenciado destes e entre estes.

Moreira e Fusco (2014) apontaram que haverá uma maior proporção de mulheres sem filhos no Nordeste, na medida em que a evolução no acesso a educação aumentar continuamente a fração de mulheres com mais elevada escolaridade. Simões (2006) chamou atenção para o fato de que as novas gerações tendem a melhorar seu nível de escolaridade e com tendências a romper com a cultura da esterilização como opção de anticoncepção, ou seja, a tendência esperada é de que no futuro, predomine o padrão etário de fecundidade dos grupos de mais elevada instrução.

Nesse sentido, o IBGE (2003b) advertiu para o fato de que as mulheres de mais elevada instrução estarão em melhores condições de escolher o momento mais apropriado do início de sua vida reprodutiva, contanto que haja mudança no procedimento de controle de reprodução.

Perante os baixos níveis de fecundidade observados na região nordestina na última década, a identificação dos parâmetros neste trabalho torna-se relevante para analisar a baixa fecundidade e reitera como este fenômeno alcançou o Nordeste no decênio passado.

## 5 CONCLUSÕES

---

Este trabalho propôs uma nova metodologia de estimação dos níveis das Taxas Específicas de Fecundidade e das Taxas de Fecundidade Total para os estados do Nordeste em dois momentos no tempo 2000 e 2010.

Foram utilizadas as base de dados do SINASC e Censo Demográfico, permitindo estimar e descrever os diferenciais das taxas de fecundidade do início e final do período reprodutivo da mulher nos estados do Nordeste Brasileiro, como também estimar os diferenciais entre os estados do NEB nas idades em que a população feminina alcança a metade da TFT, ou seja, do número médio de filhos tidos em todo seu período reprodutivo.

A metodologia estatística usando os Modelos Não Lineares de Efeitos Mistos (NLME) por meio do ajuste de uma regressão assintótica, apresentada para a estimação dos níveis e padrões de fecundidade, mostrou-se apropriada para analisar a fecundidade dos estados do Nordeste Brasileiro. As Taxas Específicas de Fecundidade Acumuladas (TEFA) encontradas para os estados do Nordeste Brasileiro descrevem por faixas etárias um comportamento com padrões assintóticos.

A utilização destes modelos em dados longitudinais é bastante útil, pois permite verificar a inserção e modificação de estruturas que melhor representem e expliquem o padrão das TEFA. Esta modelagem não requer a suposição de independência entre as medidas da variável em estudo, pois permitem a inclusão de estruturas de dependência, que proporcionam estimações mais precisas.

Em um contexto de queda da fecundidade, o NLME torna-se uma ferramenta promissória dentro do campo da Demografia, pois permite analisar os níveis e padrões das TFT sem pressupostos de uma população fechada (ausência de migração) e estacionária (onde não há crescimento da população). Acredita-se que os modelos não lineares de efeitos mistos podem levar a um melhor entendimento da realidade facilitando um diagnóstico para a um planejamento de políticas públicas de forma mais eficiente.

Como se enfatizou anteriormente, os modelos não lineares de efeitos mistos oferecem algumas vantagens, como por exemplo: o ajuste de um número reduzido de parâmetros, além de uma interpretação demográfica de seus parâmetros em comparação aos modelos tradicionais, proporcionando uma interpretação real dos parâmetros associada ao comportamento de indicadores demográficos. A utilização de modelos de efeitos mistos em dados longitudinais, como são os utilizados neste trabalho, é bastante útil e promissor, pois permite verificar a inserção e modificação de estruturas que melhor representem e expliquem o padrão das Taxas Específicas de Fecundidade Acumulada a cada faixa etária reprodutiva da população feminina.

A literatura em geral registra que existem diferenças na estrutura etária da fecundidade conforme a condição socioeconômica das mulheres, sendo a fecundidade mais rejuvenescida nos grupos menos instruídos, bem como nos menos favorecidos economicamente (FRANÇA, 2013) de forma que, para o IPEA (2008, p. 10), “dentre todas as variáveis consideradas, a renda é a que parece exercer o maior impacto nos níveis de fecundidade”. Dessa forma outra vantagem muito promissora desta metodologia é que permite ajustar outras variáveis explicativas para o estudo das potenciais causas da queda da fecundidade na região ou no Brasil, como por exemplo: nível de instrução, renda, etc. Como perspectivas futuras de análise de determinantes da fecundidade, esta metodologia (NLME) pode ser aplicada incorporando covariáveis como: raça, grau de urbanização, entre outros.

A desvantagem é a maior complexidade para a análise dos dados, no entanto, vêm crescendo a demanda por estes modelos ao considerar a constituição do grande número de fenômenos de interesse que requerem estruturas hierárquicas nos dados em estudos nas mais diversas áreas do conhecimento como: epidemiológicos, demográficos.

Para Moreira e Fusco (2014) a fecundidade nordestina no próximo decênio completaria sua trajetória para níveis abaixo da reposição, o qual é reforçado neste estudo. O Nordeste já não mais pode ser contemplado, genericamente, como espaço de altos níveis da fecundidade nacional, mas, que apresenta níveis que muito se assemelham às de outras regiões do país nos quais os níveis da fecundidade já são de reposição.

Esta é uma primeira aplicação desta modelagem para estudos do comportamento da fecundidade das populações. Pretende-se futuramente ampliar o potencial desta modelagem para: comparar padrões das TFT do modelo não linear de efeitos mistos versus outros modelos paramétricos existentes na literatura; captar as associações significativas com potenciais determinantes das taxas de fecundidade envolvendo variáveis socioeconômicas e demográficas; realizar um estudo para áreas menores como Microrregiões e Mesorregiões do país.

## REFERÊNCIAS

---

ALBUEQUEQUE. F.R.P.C, OLIVEIRA. J.C, PEREIRA. N.O.M. **Estimativas de fecundidade para o Brasil com base nas técnicas de Brass e de Arriaga**. Rev Bras Estat Pop;46(8):181-2, 1985.

ALLISON PD, **Discrete-time methods for the analysis of event histories**, em: Leinhardt S, ed., Sociological Methodology, San Francisco, CA, EUA: Jossey-Bass, págs. 61–98, , 1982.

ARRIAGA EE. **Estimating fertility from data on children ever born age of mother**. US Bureau of the Census; 1970.

ALVES, J.E.D; CAVENAGUI, S.M. **Transições urbana e da fecundidade e mudanças dos arranjos familiares no Brasil**. Cadernos de Estudos Sociais, v. 27, p.91-114, 2012

BARBOSA, M. **Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos**. 2009. 118 p. Dissertação (Mestrado em Agronomia). Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo, Piracicaba, 2009.

BARBOSA, L. M. **O SINASC como fonte de informações no Nordeste** Mestrado em Demografia, Universidade Federal de Minas Gerais. 1999

BARBOSA. L.M, FREIRE F.H. MA. **O SINASC como fonte de informação sobre fecundidade no Rio Grande do Norte, 2000**. XIV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambú- MG – Brasil, de 20- 24 de Setembro de 2004.

BATES D, PINHEIRO J. **Computacional methods for multinivel modelling bell Labs Technical Memorandum** Memorandum. Disponível em: <<http://stat.bell-labs.com/NLME/CompMulti.pdf>> , 2000.

BERQUÓ E. , Cavenaghi S.M. **Mapeamento sócio-econômico e demográfico dos regimes de fecundidade no Brasil e sua variação entre 1991 e 2000**, Trabalho apresentado no XIV Encontro Nacional de Estudos Populacionais, ABEP, realizado em CaxambuMG – Brasil, de 20- 24 de Setembro de 2004.

BERQUÓ E., Cavenaghi S. **Notas sobre os diferenciais educacionais e econômicos da fecundidade no Brasil**. Revista Brasileira de Estudos de População (impresso), v.31, p. 471-482, 2014.

BERQUÓ, E.; CAVENAGHI, S. M. **A evolução da fecundidade no Brasil**. Ciência Hoje, v. 37, p. 28-33, 2005.

BONATE PL. **Pharmacokinetic-Pharmacodynamic Modeling And Simulation(p. 387)**. New York: Springer.(2005)

BRANDÃO, A.L.S. **Modelos nao-lineares para analise de dados longitudinais**. 1996. 137 p. Dissertação (Mestrado em Estatística). Instituto de matemática, estatística e computação científica, UNICAMP, Campinas, 1996.

BRASIL, Ministério da Saúde. 2012b. **Sistema de Informações sobre Nascidos Vivos - SINASC** Acesso em 20 de Junho, 2014, Disponível em: <<http://portal.saude.gov.br>>. 2012a

BRASS W. **The demography of Tropical Africa**. Princeton University Press; 1968.

BRASS, W. "**The relationalGompertzmodelo/fertilitv by age of woman**" London School of Hygiene and TropicalMedicine, 1978.

CAMLIN CS, GARENNE M, MOULTRIE TA. **Fertility trend and pattern in a rural area of South Africa in the context of HIV/AIDS**. Afr J Reprod Health;8(2):38-54. 2004.

CARVALHO, J. A. M. **Aplicabilidade da técnica de fecundidade de Brass quando a fecundidade está declinando ou quando a população não é fechada**. III Encontro Nacional de Estudos Populacionais, Vitória, 1982

CARVALHO, J. A. M., SAWYER, D. O., RODRIGUES, R. N. "**Introdução a alguns conceitos básicos e medidas em demografia**", ABEP, 2. ed, São Paulo, 1998.

- CAVENAGHI, S. M. . **Acceso a la salud sexual y reproductiva y fecundidad de jóvenes en Brasil: desigualdades territoriales**. Notas de Población (Impresa), v. 96, p. 1-46, 2013.
- CHANDOLA, T., COLEMAN, D., HIORNS, R. **"Recent European fertility patterns: fitting curves to distorted distributions"**, Population Studies 53, 317–329, 1999.
- COALE, A. J. **"Age patterns of marriage"** In: Population Studies 25, p. 193-214, 1971.
- CRUZ, F. **Modelos Multinivel**, Rev. per. epidemiol. Vol 12 N° 3 Diciembre 2008.
- DALVA LÚCIA DE SOUZA FRANÇA Revista GeoAmazônia – ISSN: 2358-1778 (on line) 1980-7759 (impresso), Belém, n. 01, v. 01, p. 01 - 22, jan./jun. 2013.
- DAVIDIAN, M.; GILTINAN, D. M. (n.d.). **Nonlinear mixed effects models** - An overview and update. Retrieved Feb 16, 2011, from [www4.stat.ncsu.edu/~davidian/nlmmreview.pdf](http://www4.stat.ncsu.edu/~davidian/nlmmreview.pdf), 2011.
- DAVIDIAN, M.; GILTINAN, D.M. **Nonlinear Models for Repeated Measurement Data**. London: Chapman & Hall, 1995.
- DIGGLE, P.; HEAGERTY, P.; LIANG, K.; ZEGER, S. **Analysis of Longitudinal Data**. 2nd ed., Oxford University Press, New York, 396 p. 2002.
- DUCHÊNE, J, GUILLET de STEFANO, S. **Ajustement analytique des courbes de fécondité générale**. Population et Famille, 32: 53-93. 1974.
- FARIA VE, **Políticas de governo e regulação da fecundidade: consequências não antecipadas e efeitos perversos**, Revista de Ciência Hoje, págs. 62–103 ,1989
- FITZMAURICE, G.M.; DAVIDIAN, M.; VERBEKE, G.; MOLENBERGHS, G. **Longitudinal data analysis**. New York: Chapman & Hall, 623p. 2008.
- FITZMAURICE, G.M.; LAIRD, N.M.; WARE, J.H. **Applied longitudinal analysis**. New Jersey: John Wiley, 266p., 2004.
- FRANÇA, D. L. S. **Padrões de Fecundidade na região Norte e estado do Pará**. Revista GeoAmazônia – ISSN: 2358-1778 (on line) 1980-7759 (impresso), Belém, n. 01, v. 01, p. 01 - 22, jan./jun. 2013.

FRIAS, Paulo Germano de, PEREIRA, VIDAL, Suely Arruda ; LIRA, Pedro Israel Cabral de. **Avaliação da cobertura do Sistema de Informações Sobre Nascidos Vivos e a contribuição das fontes potenciais de notificação do nascimento em dois municípios de Pernambuco, Brasil.** Epidemiologia e Serviços de Saúde 16(2): 93-101, 2007.

GAYAWAN, E.A; ADEBAYO, S.B; IPINYOMI, R.A; OYEJOLA, B. A. **Modeling fertility curves in Africa.** Demographic Research 22(10): 211-236. 2010.

GALLANT, A.R. **Nonlinear statistical models.** New York: John Wiley,. 610 p, 1987.

GIRALDELLI, B W ; WONG, Laura R **O comportamento do Registro Atrasado de Nascimento (RAN) no estado de São Paulo: um a tentativa de correção do sub-registro.** Informe Demográfico, 13. 1984.

GOLDSTEIN H. BROWNE WJ, RASBASH J. **Multilevel modelling of medical data. Statistics in Medicine.**;21:3291-3315. Disponível em: <http://www.cmm.bristol.ac.uk/team/mmmd.pdf>, 2002.

HADWIGER, H., "Eine analytische reproductions-funktion fur biologische Gesamtheiten", Skandinavisk Aktuarietidskrift 23, 101-113, 1940

HAKKERT, Ralph. **Fontes de dados demográficos.** Belo Horizonte, ABEP. 1996

HOEM, J, MADSEN, D., NIELSEN, J., OHLSEN, E., HANSEN, H., RENNER M-ALM, B., "Experiments in Modelling Recent Danish Fertility Curves", Demography 18, 231-244. 1981.

HORTA, C. J. G.; **Nível e padrão da fecundidade no estado de Minas Gerais e em suas microrregiões de saúde**, XV Seminário Sobre a Economia Mineira, Diamantina, 29 a 31 de agosto de 2012. Anais do XV Seminário Sobre a Economia Mineira, Belo Horizonte, UFMG/CEDEPLAR, 2012.

HOX JJ. **Applied multilevel analysis.** Amsterdam: TT-Publikaties, 1995. Disponível em: [www.geocities.com/joophox/publist/amaboek.pdf](http://www.geocities.com/joophox/publist/amaboek.pdf). 1995

IBGE. 2002. **Divórcios concedidos em 1ª instância a casais sem filhos ou com filhos menores de idade.** Acesso em 23 /06/2014, Disponível em: <http://seriesestatisticas.ibge.gov.br/series.aspx?vcodigo=RC51>. 2002

IBGE. 2009a. **Registro Civil.** Acesso em 18/06/2014, Disponível em: <http://www.registrocivil.ibge.gov.br/>. 2009a

IBGE. 2010. **Censo Demográfico 2010**. Acesso em 13/06/2014, Disponível em: <www.ibge.gov.br> 2010

IBGE. **Indicadores Sociodemográficos e de Saúde no Brasil 2009**. Rio de Janeiro/RJ, IBGE. 2009

IBGE. **Indicadores Sociodemográficos e de Saúde no Brasil 2012**. Rio de Janeiro/RJ, IBGE. 2012.

IBGE 2010: **Censo Demográfico 2010, Resultados Gerais da Amostra**. Rio de Janeiro, RJ: IBGE, 2012a.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Atlas do Censo Demográfico 2000**. Rio de Janeiro: IBGE, 2003a. 127 p.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. PNAD 2007, **Primeiras Análises: Demografia e Gênero**. Brasília: Comunicado da Presidência, 2008.

JORGE, M P M., LAURENTI, Ruy ; GOTLIEB, Sabina. **Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC**. Ciência & Saúde Coletiva. Brasília. 2(3): 643-654, 2007.

LAZO, A C. G., VERDUGO, D F R., DUARTE, Luciano. **Uma Avaliação Crítica das Estatísticas Vitais por Estado Civil: Estado de São Paulo 1986 - 1996**. Encontro. Abep. Caxambu/SP, REBEP. Posters: 2002.

LEITE IC, **The determinants of contraceptive choice in northeast Brazil, 1991**, trabalho apresentado na reunião anual da Population Association of America, Washington, DC, 27–29 de março, 1997

LIMA E.E.C **Voto e fecundidade no Brasil: o efeito do processo eleitoral durante o regime militar sobre a mudança de comportamento reprodutivo brasileiro** Tese apresentada ao curso de Demografia do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, 2010

MAITRE P O., BUHRER M., THOMSON D., STANSKI DR. **A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam**. Journal of pharmacokinetics and biopharmaceutics, 19, 377-384. 1991.

MANDEMA JW., VEROTTA D., SHEINER, L. B. **Building population pharmacokinetic--pharmacodynamic models. I. Models for covariate effects.** Journal of pharmacokinetics and biopharmaceutics, 20(5), 511-28,1992.

MIRANDA-RIBEIRO, **Age sexuality, reproductive behavior, and access to health care in northeast Brazil**, em: Associação Brasileira de Estudos Populacionais, Anais do IX Encontro de Estudos Populacionais, Vol. 3, Caxambu, Brasil, 1994.

MIRANDA-RIBEIRO. A, RIOS-NETO. E. L.G, ORTEGA. J. A. "**Desafíos, oportunidades y prioridades Declínio da Fecundidade no Brasil e México e o nível de reposição: efeito tempo, quantum e parturição** " II Congreso de la Asociación Latinoamericana de Población. ALAP. Guadalajara, México: 2006.

MOREIRA. M. M, FUSCO W. **Mulheres sem filho no Nordeste**, XIX Encontro Nacional de Estudos Populacionais, ABEP, São Pedro/SP – Brasil, de 24 a 28 de novembro de 2014.

MOREIRA. M. M, FUSCO W. **Mapeando a fecundidade Nordestina**. XVIII Encontro Nacional de Estudos Populacionais. Abep. Águas de Lindóia/SP, ABEP: 2012.

NETER, J.; KUTNER, M.H.; NACHTSHEIN, C.J.; WASSERMAN, W. **Applied Linear Statistical Models**. Irwin: McGraw-Hill, 1408 p. 1996.

NEWELL C. **Methods and models in demography**. London: Behaven Press; 1989.

OGLIARI; P.J.; ANDRADE, D.F. **Analysing longitudinal data via nonlinear models in rando-mized block designs**. *Computational Statistics & Data Analysis*, Holanda, v. 36, n. 3, p.319-332, 2001.

PARRADO, B.S. "**Avances metodológicos en Demografía**" Tesis Doctoral. Facultad de Matemáticas, Universidad de Sevilla. 2014

PINHEIRO JC, BATES DM. **Mixed-Effects Models in S and S-PLUS**. Springer-Verllag, New York. 421p. 2000.

POCHMANN, M. **Situação do jovem no mercado de trabalho no Brasil: um balanço dos últimos 10 anos**. Disponível em: Acesso em: 15 jul 2008.

POTTER JE et al., **The spread of television and fertility decline in Brazil: a spatial-temporal analysis, 1970–1991**, trabalho apresentado na 23 a conferência geral da IUSSP, Beijing, 11–17 de outubro, 1997.

Potter, Joseph E. ; Schmertmann, Carl P. ; Assunção, Renato M. ; CAVENAGHI, S. M. **Mapping the Timing, Pace, and Scale of the Fertility Transition in Brazil.** Population and Development Review. v. 36, p. 283-307, 2010.

PREEECE MA, BAINES MJ. **A new family of mathematical models describing the human growth curve.** Ann Hum Biol.: 5: 1-24, 1978.

RIBOLDI J. **Modelos Mistos Aplicados** [52ª RBRAS e 12ª SEAGRO] Santa Maria, Rio Grande do Sul; 2007

RINNE, H. "**The Weibull distribution**". Boca Raton: CRC Press. 2009.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike Information Criterion Statis-tics**, Dordrecht: Reidel,. 290 p. 1986

SÁNCHEZ-CANTALEJO E, OCAÑA-RIOLA R. **Los modelos multinivel o la importancia de la jerarquía.** Gac Sanit.;13(5):391-8. Disponível em [http://www.elsevier.es/revistas/ctl\\_servlet? f=7064 &articuloid=13008388](http://www.elsevier.es/revistas/ctl_servlet? f=7064 &articuloid=13008388), 1999.

SARTORIO, S.D. **Modelos não lineares mistos em estudos de degradabilidade ruminal in situ** . 2013. 198p. Tese (Doutorado em Ciências) - Escola Superior de Agricultura Luiz de Queiroz", Universidade de Sao Paulo, Piracicaba, 2013.

SILVA, Ricarly S, BONFIM, Cristine, OLIVEIRA, Conceição ; FERREIRA, Daniela K S. **Avaliação da qualidade do Sistema de Informação sobre Nascidos Vivos, nos estados da Região Nordeste do Brasil, 2000-2009.** XVIII Encontro Nacional de Estudos Populacionais. Abep. Águas de Lindóia/SP, ABEP: 2012.

SOARES T. M., MENDONÇA M.M. **Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000** Pesquisa Operacional, v.23, n.3, p.421-441, Setembro a Dezembro de 2003.

SOARES, A.P.M. **Ajuste do modelo de Orskov & McDonald (1979) a dados de degradação ruminal in situ utilizando mínimos quadrados ponderados.** 2007. 65 p. Dissertação (Mestre em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

SOUZA, L. **Avaliação do Sistema de Informação sobre Nascidos Vivos – SINASC, Minas Gerais e Mesoregiões**, Dissertação (Mestrado em Demografia), Centro de Desenvolvimento e Planejamento Regional/ Universidade Federal de Minas Gerais, Belo Horizonte, 2000.

SCHMERTMANN, Carl P. ; CAVENAGHI, S. M. ; ASSUNÇÃO, Renato Martins ; Potter, Joseph E. . **Bayes plus Brass: Estimating total fertility for many small areas from sparse census data**. Population Studies. v. 67, p. 1-19, 2013.

SPYRIDES, M.H. C. **Padrão de Crescimento Infantil e Amamentação, Rio de Janeiro, 1999 – 2001**. 2004. Tese (Doutorado em Saúde Pública). Escola Nacional de Saúde Pública Fundação Oswaldo Cruz. Rio de Janeiro. 2004.

THERBORN, G. **Sexo e poder: a família no mundo, 1900-2000**. São Paulo, S. P.: Contexto, 2006.

UN Population Division. 1983. **Manual X: Indirect Techniques for Demographic Estimation**. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. Acesso em 20/06/2014 Disponível em: <<http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html>>

WAHLBY U., JONSSON E N, KARLSSON M O. **Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis**. AAPS pharmSci, 4(4), E27. 2002

WEST, B.T.; WELCH, K.B., GALECKI, A.T. **Linear mixed models: a practical guide using statistical software**. New York: Chapman & Hall, 376p. 2007.

WINGERTER, D. G. **Registro da Informação Sobre Nascidos Vivos no Nordeste: Uma Avaliação da Evolução do Sinasc entre 2000 e 2010**. 2013. Dissertação (Mestrado em Demografia). Universidade Federal do Rio Grande do Norte.UFRN.Natal.2014.

Wong LR, **A queda da fecundidade no Nordeste: uma aproximação aos determinantes, em: BEMFAM, Fecundidade, Anticoncepção e Mortalidade Infantil,Rio de Janeiro, Brasil: BEMFAM, págs. 9–36, 1994.**

WOOD C, Carvalho JAM, **The Demography of In-equality in Brazil**,Cambridge, Reino Unido: Cambridge University Press, 1988.

ZEVIANI, W.M. **Avaliação de modelos de regressão não linear na cinética de liberação de potássio de resíduos orgânicos**. 2009. 87 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2009.

## APÊNDICE A - BIBLIOTECA NLME

---

A biblioteca nlme foi desenvolvida por Pinheiro e Bates (2000) primeiramente no software S-PLUS e depois adaptada para o software R. O software R possui uma linguagem de programação similar ao S-PLUS e entre as vantagens é um software livre e de código aberto, que além de poder ser explorado, permite que o mesmo possa ser modificado de acordo com a proposta do pesquisador.

A desvantagem é que por ser um software livre, não se tem o suporte técnico que softwares privados oferecem aos seus usuários. No entanto, é uma excelente opção para uso acadêmico. O software pode ser obtido no site do CRAN (<http://cran.r-project.org/>), no qual é possível baixar também os pacotes ou bibliotecas opcionais.

Nesta biblioteca, várias funções são disponibilizadas, dentre as quais a função nlme que ajusta modelos não lineares mistos usando a formulação apresentada por Lindstrom e Bates (1990), de forma prática e fácil. Vários argumentos podem ser utilizados com esta função, sendo o mais típico:

```
> nlme(model, data, fixed, random, start, correlation, weights, method)
```

em que o argumento

- **model:** Especifica o modelo não linear;
- **data:** indica o nome do arquivo que contém as variáveis;
- **fixed:** descreve a parte do modelo referente ao efeito fixo;
- **random:** descreve a parte do modelo referente ao efeito aleatório;
- **start:** define valores iniciais para os efeitos fixos e aleatórios, que serão utilizados no método numérico para maximização da verossimilhança;
- **correlation:** argumento opcional utilizado para descrever a estrutura de correlação intra-indivíduos. Uma lista de opções está disponível na classe corStruct. Por *default* admite-se correlações nulas intra-indivíduos;
- **weights:** argumento opcional utilizado para descrever a estrutura heterocedástica intra-indivíduos. Uma lista de opções está disponível na classe varFunct. Por *default* admite-se a homocedasticidade dos erros intra-indivíduos;
- **method:** Especifica o método a ser utilizado na estimação do modelo não linear misto. Por *default*, method = ML, o modelo será ajustado pelo método da

Máxima Verossimilhança e se `method=REML`, o modelo será ajustado pelo Método da Máxima Verossimilhança Restrita.

## APÊNDICE B - SINTAXE DO SOFTWARE R PARA A AJUSTE DE UM NLME

---

```
BancoDados= read.table(file= "dados.csv", ";", header=T)
library(nlme)
attach(dados)
names(dados)
require(geepack)
require(nlme)
require(MASS)
require(car)
library(lmtest)
library(lattice)
idade= Fetaria-15
CENSO2000 = data.frame(dados,idade)
#####criando un groupedata#####
tefestado = groupedData(TEFacumulada~idade|Estado, data = CENSO2000,
labels=list(x="Faixa etaria",y= "TEF acumulada",order.groups=T))
plot(tefestado, ylab= "T.E.F Acumulada-Censo2000", xlab="Idade",layout = c(3,3))
plot(tefestado,layout = c(3,3),ylab= "T.E.F Acumulada-Censo 2000", xlab="Faixa
etaria")
plot(tefestado,outer= ~1,ylab= "T.E.F Acumulada-Censo 2000", xlab="Faixa Etaria")
tefestado.nls <- nls(TEFacumulada~ SSasymp(idade,Asym, R0, lrc), data = tefestado)
summary(tefestado.nls)
plot( tefestado.nls, Estado~resid(.), abline = 0 )
dados.lis <- nlsList(TEFacumulada~SSasymp(idade,Asym, R0, lrc),data=tefestado)
plot( intervals(dados.lis),layout = c(3,1) )
summary(dados.lis)
fixed(dados.lis)
intervals(dados.lis)
summary(dados)
anova(dados)
intervals(dados)
qqnorm(dados, xlab="Resíduos padronizados", ylab="Quantis" )
residuales = residuals(dados)
shapiro.test(residuales)
```

```
dado<- nlme(TEFacumulada~SSasyp(idade,Asym, R0, lrc),
data= tefestado,
fixed= Asym+R0+lrc~1,
random= Asym~1,
method= 'REML', start= c(0.58, 0.09,-2.31),weights=varExp())
plot(fitted(dados4),
residuals(dados, type = "n"),
ylab= "Resíduos Condicionais Studentizados",
xlab = "Valores Preditos",pch= 19, cex=0.5)
abline(a=-1, b=0, lty=1, col=1, lwd=1)
abline(a=1, b=0, lty=1, col=1, lwd=1)
```