

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

Alisson dos Santos Silva

CENÁRIOS PROSPECTIVOS DA
COBERTURA DOS ÓBITOS NOS
ESTADOS DO SEMIÁRIDO BRASILEIRO

João Pessoa, 10 Dezembro de 2015

Alisson dos Santos Silva

Cenários Prospectivos da Cobertura dos Óbitos nos Estados do Semiárido Brasileiro

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Demografia.

Orientador: Prof^o. Dr. NEIR ANTUNES PAES.

João Pessoa, 10 Dezembro de 2015

Alisson dos Santos Silva

Cenários Prospectivos da Cobertura dos Óbitos nos Estados do Semiárido Brasileiro

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Demografia.

João Pessoa, 10 de Dezembro de 2015

BANCA EXAMINADORA

Prof^o. Dr. NEIR ANTUNES PAES
Orientador (UFPB)

Prof^o. Dr. Jozemar Pereira dos Santos
(UFPB)

Prof^a. Dr^a Maria Lídia Coco Terra
(UFPB)

*A DEUS que iluminou o meu caminho
durante esta longa jornada.*

*Aos meus pais, Antônio e Maria
pelo incentivo e força ao longo dessa graduação.*

Agradecimentos

Agradeço primeiramente à DEUS, por sempre ter guiado todos os meus passos, me dando força, coragem e discernimento ao longo dessa jornada.

Agradeço a minha família, meus pais Antônio da Silva Santos e Maria de Fátima dos Santos Silva, meu irmão Alan dos Santos Silva e minha tia Maria Alice Guedes de Arruda por incentivar e apoiar em todos os momentos ao longo dessa formação.

Agradeço a minha amiga Maizza Micaelle que considero como irmã, pela amizade, carinho e dedicação desde quando entrei na UFPB, me dando apoio em todas as etapas da minha graduação.

Agradeço aos meus amigos e colegas da graduação, entre eles, Andreza, Aldine, Camila Ribeiro, Camila Ravena, Elaine, Geisislane, Henrique, Jodavid, Marina, Marília, Michelle, Ramon, Saul e tantos outros, obrigado por estarem próximos nos momentos de estudo e de descontração.

Aos professores do DE-UFPB que foram tão importantes na minha formação acadêmica, por engrandecer meus conhecimentos ao longo desses anos.

Aos amigos e colegas do LED todos aqueles que estão atualmente ou passaram no laboratório pela troca de experiência e aprendizado, entre eles, Clarissa, Diego, Everlane, Iane, Jéssica, Lígia, Luísa, Telmo.

Ao Professor Neir Antunes Paes, por ter me orientado por mais de dois anos dentro do LED. Em que gostaria de agradecer imensamente pela paciência e dedicação, sendo ele, um dos principais responsáveis pela minha formação acadêmica e profissional e pela realização desse trabalho de conclusão de curso.

Ao CNPq, pelo apoio financeiro.

*Nada te perturbe,
Nada te espante,
Tudo passa,
Deus não muda,
A paciência tudo alcança;
Quem a Deus tem,
Nada lhe falta,
Só Deus basta.
Teresa D'Ávila*

Resumo

A elaboração de projeções de mortalidade é de fundamental importância para estimar demandas de serviços públicos e de públicos-alvo nas políticas públicas, particularmente em áreas como na saúde e na previdência social. No entanto, antes de se projetar a mortalidade faz-se necessário conhecer o grau de qualidade dos registros de óbitos. Ao serem abordadas estas questões para regiões como o Semiárido brasileiro esbarra-se com um recorte geográfico espacial que convive com um distanciamento acentuado dos padrões de desenvolvimento do país e é marcado por grandes desigualdades sociais. Diante de uma situação de incompletude dos óbitos associados a uma região com níveis de desenvolvimento baixo indaga-se sobre quando a cobertura dos óbitos atingirá a universalização dos seus registros de óbitos. O presente trabalho faz uso de métodos matemáticos de projeções (Logístico, Gompertz e Holt) para estimar em que ano a cobertura atingirá 100% para os Estados do Semiárido brasileiro. Entre as modelagens, Holt apresentou melhor adaptação ao padrão dos dados das séries descrevendo as previsões em menores tempos e erros de estimação para o estudo.

Palavras - Chaves: Cobertura dos Óbito, Semiárido Brasileiro, Projeções da Mortalidade.

Lista de Figuras

3.1	Curva de Ajuste do Modelo Logístico	11
3.2	Curva de Ajuste do Modelo de Gompertz	12
3.3	Cobertura dos Óbitos do Nordeste Brasileiro	13
4.1	Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Nordeste	28
4.2	Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Maranhão	28
4.3	Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Piauí .	29
4.4	Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Ceará	29
4.5	Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Rio Grande do Norte	30
4.6	Análise gráfica dos resíduos do modelo Logístico e Gompertz para a Paraíba	30
4.7	Análise gráfica dos resíduos do modelo Logístico e Gompertz para Pernambuco	31
4.8	Análise gráfica dos resíduos do modelo Logístico e Gompertz para Alagoas	31
4.9	Análise gráfica dos resíduos do modelo Logístico e Gompertz para Sergipe .	32
4.10	Análise gráfica dos resíduos do modelo Logístico e Gompertz para a Bahia	32
4.11	Projeção das coberturas de óbitos estimadas para o Nordeste Brasileiro. . .	34
4.12	Projeção das coberturas de óbitos estimadas para os Estados do Maranhão, Piauí e Ceará.	35
4.13	Projeção das coberturas de óbitos estimadas para os Estados do Rio Grande do Norte, Paraíba e Pernambuco.	36
4.14	Projeção das coberturas de óbitos estimadas para os Estados do Alagoas, Sergipe e Bahia.	37

Lista de Tabelas

4.1	Cobertura dos Óbitos para o Brasil, Nordeste e Estados do Nordeste, de 1991 a 2011.	23
4.2	Medidas de Precisão para estimativas das coberturas completas de óbitos para o Nordeste e os Estados do Nordeste.	25
4.3	Estimativas do valor-p residual dos testes Shapiro-Wilk, Breusch-Pagan e Durbin-Watson para o Nordeste e Estados do Nordeste	27
4.4	Estimativas do valor-p dos testes Dickey-Fuller e Wilcoxon para o Nordeste e Estados do Nordeste	33
4.5	Estimativas das Coberturas Completas de óbitos para o Nordeste e Estados do Nordeste.	38
4.6	Estimativas e Modelos Adotados para o Nordeste e Estados do Nordeste.	39
4.7	Cobertura de óbitos do espaço geográfico dos Estados do Semiárido Brasileiro entre 1991, 2000 e 2010.	40
4.8	Cobertura dos Óbitos para o espaço geográfico dos Estados do Semiárido brasileiro, de 1991 a 2010.	42
4.9	Previsão para as estimativas das coberturas completas de óbitos para os Estados do Semiárido brasileiro.	43
4.10	Medidas de precisão para as estimativas das coberturas completas de óbitos para os Estados do Semiárido brasileiro.	44
4.11	Estimativas e Modelos Adotados para os espaços geográficos dos Estados do Semiárido brasileiro.	44

Lista de Siglas

AL - Estado do Alagoas

AEH - Alisamento Exponencial de Holt

BA - Estado da Bahia

BR - Brasil

CE - Estado do Ceará

DATASUS - Departamento de Informática do SUS

EAM - Erro Absoluto Médio

EPM - Erro Percentual Médio

EQM - Erro Quadrático Médio

IBGE - Instituto Brasileiro de Geografia e Estatística

IDH - Índice de Desenvolvimento Humano

MA - Estado do Maranhão

MS - Ministério da Saúde

NE - Região Nordeste

PB - Estado da Paraíba

PE - Estado de Pernambuco

PEA - População Economicamente Ativa

PI - Estado do Piauí

RIPSA - Rede Interagencial de Informação para a Saúde

RN - Estado do Rio Grande do Norte

SE - Estado de Sergipe

SIM - Sistema de Informação da Mortalidade

Sumário

1	Introdução	1
2	Referencial teórico	3
2.1	Qualidade dos registros de óbitos	3
2.2	Métodos de Projeção da Cobertura de Óbitos	5
3	Metodologia	7
3.1	Fonte de Dados	7
3.2	Qualidade dos Óbitos dos Estados do Nordeste	8
3.3	Métodos de Projeção da Cobertura de Óbitos	10
3.3.1	Modelo Logístico	10
3.3.2	Modelo de Gompertz	11
3.3.3	Modelo de Suavização Exponencial de Holt	14
3.4	Medidas de Diagnósticos	15
3.5	Análise Residual	16
3.5.1	Erro Absoluto Médio	17
3.5.2	Erro Percentual Médio	17
3.5.3	Erro Quadrático Médio	18
3.6	Cobertura dos Óbitos para o Semiárido	18
3.7	Critérios de Seleção dos Modelos	20
4	Resultados e Discussões	22
4.1	Cobertura dos Óbitos para os Estados do Nordeste brasileiro	22
4.2	Análise Residual	25
4.3	Medidas de Diagnósticos	27
4.4	Projeções das Coberturas dos Óbitos do Nordeste Brasileiro	34

4.5 Cobertura e projeções dos Óbitos do espaço geográfico dos Estados do Semiárido brasileiro	40
5 Conclusão	45
Referências Bibliográficas	47
A Rotina no R - Modelo Logístico	51
B Rotina no R - Modelo de Gompertz	54
C Rotina no R - Algoritmo de Holt	57
D Rotina no R - Cobertura dos Estados do Semiárido	60

Capítulo 1

Introdução

A elaboração de projeções demográficas é de fundamental importância para estimar demandas de serviços públicos e de públicos-alvo nas políticas públicas, em áreas como na educação, saúde, trabalho e planejamento urbano, onde faz-se necessário o uso de técnicas matemáticas. O Ministério da Educação, por exemplo, necessita saber o tamanho, a evolução e a localização da população em idade escolar. O planejamento urbano precisa de informações sobre a população total de uma localidade para poder ampliar a rede de postos de saúde e de serviços médicos na extensão da cobertura da rede de abastecimento de água e saneamento básico e na implantação de projetos sociais entre outros. O Ministério do Trabalho precisa acompanhar a evolução atual e futura dos variados subgrupos de populações-meta, como a PEA e a Previdência Social verificando seu nível para um período específico (MARTINE & LIMA, 1997).

Naturalmente, como todo processo de formulação de políticas sociais envolve a avaliação quantitativa da população-alvo a que ela se destina, no horizonte de tempo em que se planeja sua atuação, a produção de projeções populacionais constitui parte imprescindível das estatísticas disponíveis para a sociedade (SEADE, 1995). No entanto, antes de se projetar a população faz-se necessário conhecer o comportamento dos componentes demográficos (mortalidade, fecundidade e migração) e suas influências nos padrões da população. Particularmente, projetar a mortalidade é essencial no planejamento de políticas e ações voltadas ao atendimento de demandas sociais e de saúde.

Uma questão que merece destaque quanto ao uso dos componentes demográficos diz respeito a qualidade dos dados que servirão de base para as projeções. Especificamente, interessa saber a qualidade dos registros de óbitos quando se pretende fazer projeções de

mortalidade. Ao serem abordadas estas questões para regiões como o Semiárido brasileiro esbarra-se com um recorte geográfico espacial que convive com um distanciamento acentuado dos padrões de desenvolvimento do país e é marcado por grandes desigualdades sociais (PAES & REGIS, 2010). O Índice de Desenvolvimento Humano (IDH) do Semiárido é considerado baixo para aproximadamente 82% dos municípios, que possuem IDH até 0,65. Isto significa um déficit em relação aos indicadores de renda, educação e longevidade para 62% da população do Semiárido (PORTAL ASA BRASIL, 2010).

Diante de uma situação de incompletude dos óbitos associado a uma região com níveis de desenvolvimento precário há de indagar-se quando a cobertura de óbitos do semiárido atingirá o patamar de 100%, ou seja, a universalização de suas estatísticas de óbitos. Por outro lado, uma outra questão surge ao se questionar sobre os rumos da mortalidade no Semiárido brasileiro, já que o subregistro implica em correção dos níveis da mortalidade, elevando sua magnitude. Na busca de resposta para esses questionamentos, formulam-se os objetivos a seguir:

Objetivo Geral:

Traçar cenários prospectivos sobre a cobertura dos registros dos óbitos para os espaços geográficos dos Estados do Semiárido brasileiro.

Objetivos Específicos:

- Projetar a cobertura dos óbitos para o Nordeste e seus Estados até atingir a universalização;
- Elaborar a Série da Cobertura dos Óbitos para os Estados do Semiárido.
- Projetar a cobertura dos óbitos para os Estados do Semiárido brasileiro até atingir a universalização;

Capítulo 2

Referencial teórico

2.1 Qualidade dos registros de óbitos

As estatísticas de mortalidade refletem a situação da saúde de uma população, as quais são utilizadas na formulação de indicadores demográficos, como Taxa Bruta de Mortalidade (TBM), Razão de Mortalidade Materna (RMM), Proporção de Óbitos por Causas Mal Definidas, entre outros. No entanto, a evolução dos registros vitais no Brasil é marcada por fatores políticos e instabilidades administrativas que comprometeram sua qualidade e portanto sua utilidade (PAES, 2005).

Devido a este comprometimento, o Ministério da Saúde (MS) tem optado por uma metodologia que utiliza estimativas indiretas desses indicadores, criando possíveis situações que retratem o desenvolvimento na saúde brasileira, dispensando o uso direto das estimativas vitais (JORGE; GOTLIEB, 2001). No Brasil existem vários órgãos e instituições que produzem e/ou utilizam estimativas indiretas, entre eles:

- A Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) que produz e faz estimativas. Em 1974 iniciou suas publicações anuais das estatísticas vitais para todas as unidades de federação constituindo-se como principal provedor de dados e informações do País, que atendem às necessidades dos mais diversos segmentos da sociedade civil, bem como dos órgãos das esferas governamentais federal, estadual e municipal;
- O Sistema de Informação de Mortalidade (SIM) apenas produz dados. Foi desenvolvido em 1975 pelo Ministério da Saúde (MS) para obter dados regulares sobre

mortalidade no País. O SIM foi concebido para suprir as falhas do Sistema de Registro Civil e permitir conhecer o perfil epidemiológico da mortalidade em todo o país;

- A Rede Interagencial de Informação para a Saúde (RIPSA) não produz dados, mas calcula indicadores de mortalidade. Foi criada em 1996, cuja finalidade foi tornar disponíveis todas as bases de dados importantes para os setores a níveis federal, estadual e municipal.

As informações apresentadas tanto pelo SIM como pelo IBGE em sua maioria são deficientes, as quais impedem uma cobertura completa dos dados de óbitos em pequenas áreas e que somado ao mal preenchimento da Declaração de Óbito dificultam o processo de notificação e contagem. No Brasil está ocorrendo uma melhora bastante lenta e gradativa nos registros de óbitos, no entanto algumas regiões ainda estão distantes de uma cobertura completa.

Registros de óbitos confiáveis classificados por idade ou quaisquer outras características populacionais ainda representam um entrave na construção de indicadores de mortalidade. No entanto, dimensionar esses óbitos significa medir a sua cobertura, ou seja, medir o seu subregistro. Sabe-se que a cobertura é diferenciada por sexo e idade, cujo grau de cobertura pode ser estimado por vários métodos, além de que os óbitos nas primeiras idades são inferiores aos adultos, limitando assim o uso desses dados (PAES, 2007). Contudo suas aplicações sugerem que mesmo dados deficientes podem ser úteis para estimar os níveis e tendências de mortalidade, cuja diferença entre suas aplicações se remetem aos seus supostos, exigências e grau de precisão.

Diversas técnicas foram propostas para dimensionar a cobertura dos óbitos. Entre elas destaca-se o Método da Equação de Balanço do Crescimento que foi desenvolvido por Brass (1975). Ele mostrou que em uma população estável, a equação do balanço do crescimento pode ser expressa como uma relação linear entre a taxa de mortalidade e a taxa de natalidade para os diferentes grupos etários, cujo coeficiente angular especifica o fator de correção, ou seja, o subregistro de óbitos e o coeficiente de interseção determina a taxa de crescimento da população de estudo (BRASS, 1975).

Existem muitos obstáculos e dificuldades para atingir uma universalização pelos sistemas atuais como vencer desafios em controlar o fluxo de informações pelos cartórios e unidade de saúde, em regiões de extrema pobreza no interior do Nordeste. No entanto,

a universalização dos registros de óbitos é uma meta alcançável e, nesta perspectiva sabe-se que a medida que os níveis de cobertura aumentam, os registros tanto para homens como mulheres devem aumentar em conjunto, rumo a meta de 100% de completude (PAES, 2005). Abordagens nessa linha são raras, apenas um trabalho foi encontrado na literatura, realizado por Paes e Regis (2010), ao investigar quando os Estados do Nordeste iriam atingir a universalização dos registros de óbitos.

2.2 Métodos de Projeção da Cobertura de Óbitos

Não existem métodos específicos para se projetar a cobertura dos óbitos. No entanto, a literatura disponibiliza modelos matemáticos que podem ser adaptados para situações de projeções da cobertura de óbitos, desde que não violem a natureza de seus critérios. Entre eles, faz-se referências aos seguintes modelos: Modelo Logístico, Modelo de Gompertz e Modelo de Suavização Exponencial de Holt.

Em 1825, o matemático Benjamin Gompertz com o intuito de descrever a taxa de mortalidade de seres humanos, desenvolveu uma equação diferencial ordinária não linear denominada de Curva de Gompertz, tendo como princípio uma relação que mostra a taxa de crescimento de um fenômeno biológico com o passar do tempo, em que após certo estágio de maturidade sexual, o risco de morte passa a se comportar de forma exponencial (GOMPERTZ, 1825).

Nesse estudo, utilizou-se o modelo de Gompertz para medir a relação entre o nível das coberturas dos óbitos e os anos subsequentes de uma série temporal estimada pela RIPSAs, em que espera-se que a cobertura aumente exponencialmente com o passar dos anos, mediante o melhoramento na captação dos dados de óbitos. Atualmente, curvas derivadas desta função não linear são intensamente utilizadas para a análise de dados longitudinais na área da saúde, onde medidas de natureza contínua são estudadas ao longo do tempo. Por exemplo, Lo (2007) e Albano e Giorno (2006) utilizaram para estudar prospectivamente a progressão do tamanho de tumores sólidos, permitindo uma melhor compreensão da evolução da doença. Na oftalmologia, Augusteyn (2007) utilizou curvas de Gompertz para analisar o crescimento do cristalino humano a partir do acúmulo de peso úmido, em função da idade (SOUZA, 2010).

Em 1838, o matemático Verhulst criou um modelo para expressar a lei do crescimento de populações humanas denominado de Curva Logística. Ela descreve o crescimento populacional durante um determinado período de tempo. O matemático belga observou que, no início, uma população cresceria exponencialmente até atingir o limite de recursos críticos. Ele supôs que o crescimento populacional seria freado por alguma função proporcional à superabundância populacional. Essa diferença no crescimento geraria uma curva de crescimento no formato de S.

A curva logística aponta que o crescimento populacional é o produto da articulação de três fatores: crescimento constante relativo, chamado de parâmetro malthusiano; o tamanho real da população no início do período; e a diferença entre a capacidade de suporte e as necessidades da população. A curva logística tem sido largamente empregada para a representação de dados empíricos de crescimento de animais e vegetais, de crescimento de populações humanas e de adoção de novos bens econômicos. A curva logística também supõe que as taxas de natalidade e de mortalidade são contínuas, ou seja, não sofrem alterações aleatórias ao longo da sua trajetória. No entanto, a população representada pela curva logística apresenta uma distribuição etária estável, ou seja, não há diferença entre os indivíduos daquela população, e não há predominância de faixa etária (BEZERRA, 2008).

Por sua vez, o Alisamento Exponencial de Holt (AEH) é um modelo aplicado em uma série temporal que apresente tendência linear positiva ou negativa. O AEH utiliza duas constantes de alisamento diferentes que alisam diretamente os valores do nível e da tendência da série. Segundo Morettin e Tolo (2006), se existem indicações de que uma série temporal exibe uma tendência linear (e nenhuma sazonalidade), o método de Holt geralmente oferece bons resultados. O AEH é aplicável em diversas áreas, como por exemplo, nas ciências sociais. Oliveira (2012) utilizou este método para prever uma série temporal de homicídios dolosos no Rio de Janeiro; na saúde Silva (2012) utilizou para fazer uma previsão de demanda para fios cirúrgicos, utilizados em um hospital escola. Marinho e Paes (2010) em trabalho pioneiro fizeram uso desse método para fazer previsões de quando a cobertura de óbitos dos Estados do Nordeste iriam atingir 100% de cobertura.

Capítulo 3

Metodologia

3.1 Fonte de Dados

No Brasil, a produção e utilização de informações sobre as estatísticas vitais (óbitos e nascimentos) são processadas em um contexto muito complexo de relações institucionais, compreendendo variados mecanismos de gestão e financiamento. Existem duas fontes oficiais de informações que são responsáveis pela produção das estatísticas vitais de óbitos: O Ministério da Saúde, por meio do Sistema de Informação sobre Mortalidade e a Fundação Instituto Brasileiro de Geografia e Estatística, através das Estatísticas do Registro Civil, com disponibilidade anual para todos os municípios brasileiros. Ambos são sistemas de vigilância epidemiológica nacional, cujo objetivo consiste em captar dados sobre os óbitos do País para fornecer informações sobre mortalidade para todas as instâncias do sistema de saúde.

Na projeção da cobertura dos óbitos, utilizou-se a série temporal anual de 1991 a 2011 (a única disponível) da Razão entre Óbitos informados e estimados pela Rede Interagencial de Informações para a Saúde (RIPSA, 2012) para o Nordeste Brasileiro, acessados através do site do Departamento de Informática do SUS (DATASUS). A RIPSA tem como propósito promover a disponibilidade adequada e oportuna de dados básicos, indicadores e análises sobre as condições de saúde e suas tendências, visando aperfeiçoar a capacidade de formulação, gestão e avaliação de políticas e ações públicas pertinentes. É importante destacar que não há disponibilidade de estimativas da cobertura de óbitos especificamente para os espaços geográficos do Semiárido brasileiro.

Para construção e análises das estimativas apresentadas nesse trabalho utilizou-se o Ambiente R. O R é uma linguagem de programação orientada a objetos onde o usuário pode criar suas próprias funções, e sua própria rotina na análise de dados. Ele é uma importante ferramenta na manipulação e análise de dados, com testes paramétricos e não paramétricos, modelagem linear e não linear, análise de séries temporais, entre outros. O R é um ambiente de software livre para computação estatística e gráficos, onde é apresentado em versões de acordo com o sistema operacional UNIX, Windows ou Macintosh, apresenta código fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos por qualquer usuário a qualquer momento.

3.2 Qualidade dos Óbitos dos Estados do Nordeste

O indicador de qualidade dos dados de mortalidade (cobertura dos óbitos) possibilita analisar variações geográficas e temporais dos dados coletados pelo SIM, com o objetivo de avaliar a sua consistência. Indica áreas geográficas com subregistro expressivo de dados do SIM e contribui para o desenvolvimento operacional do sistema. É útil para orientar a escolha da melhor fonte de dados de óbitos no cálculo de indicadores. Além disso, subsidia o aperfeiçoamento de estimativas obtidas por métodos demográficos indiretos e contribui para a melhoria da qualidade da informação em saúde. Segundo Jorge (2007) existem diversos métodos usados para estimar a cobertura:

- Número de municípios com informação regular de mortalidade na população coberta pelo sistema e número de óbitos informados pelo total de municípios, bem como por aqueles com informações regulares. Esta metodologia foi usada no início da implantação do SIM e abandonado a partir dos dados de 1995;
- Comparação dos seus dados com aqueles dos Cartórios (via IBGE), embora se saiba que estes também têm falhas nas áreas menos desenvolvidas do país, onde enterramentos são feitos sem o competente registro;
- Comparação entre os dados do SIM e projeções demográficas feitas pelo IBGE. Desde 1996, com a criação da RIPSAs, essas estimativas têm sido usadas como referência para avaliação do SIM;

- Análise do coeficiente de mortalidade qualquer que seja o nível de saúde de uma população. O pressuposto é de que esse indicador deve flutuar entre 6 e 12 óbitos por mil habitantes.

Nesse trabalho tratar-se-á do uso do terceiro método. Para o cálculo da cobertura realizado pela RIPSAs (2012) foi utilizado o número de óbitos notificados ao Sistema de Informações de Mortalidade do Ministério da Saúde, em relação ao número estimado pelo IBGE, na população residente dos Estados do Nordeste brasileiro, para os anos de 1991 a 2011. Esse indicador mede a relação quantitativa entre óbitos informados no SIM e os estimados por projeções demográficas, refletindo a cobertura do SIM. Valores de cobertura próximos a 100 indicam coincidência entre a frequência dos óbitos apurados pelo SIM e as estimativas demográficas; valores acima de 100 sugerem que a estimativa demográfica está subestimada e valores abaixo de 100 que há subregistro do SIM. A cobertura de óbitos é dada pela seguinte equação:

$$C = \frac{OI}{OE} \cdot 100$$

Em que:

C: Cobertura dos Óbitos;

OI: Número informado de óbitos de residentes;

OE: Número estimado de óbitos de residentes.

Como todo método derivado de estimação está sujeito a erros, a RIPSAs aponta alguns erros metodológicos:

- A estimativa do número de óbitos para anos intercensitários pode, em alguns casos, não refletir o padrão demográfico atual, por estar baseada em tendências passadas;
- No caso das estimativas, envolve as dificuldades metodológicas e imprecisões inerentes às técnicas utilizadas, cujos pressupostos podem não se cumprir por mudanças da dinâmica demográfica. A imprecisão é maior no caso de pequenas populações;
- Há possibilidade de variação no valor da razão devido a imprecisões no registro do local de residência, na Declaração de Óbito.

Apesar destas restrições, Paes (1999) aponta que as estimativas da cobertura de óbitos são razoavelmente bem representadas e chama à atenção para os erros, ao lembrar que eles são inerentes a quaisquer estimativas.

3.3 Métodos de Projeção da Cobertura de Óbitos

A regressão não linear é uma forma de análise em que os dados são modelados por uma função que é uma combinação não linear dos parâmetros do modelo. Entre esse tipo de modelagem de regressão, a função Logística e a função de Gompertz são bastante usuais na estimação do crescimento de populações humanas (HOFFMAN & VIEIRA, 1998). Em qualquer tipo de modelagem, além do ajuste realizado, é necessário fazer inferências sobre os parâmetros em estudo.

Uma das melhores alternativas que se pode fazer para assegurar uma análise não linear bem sucedida é obter bons valores iniciais para os parâmetros, a partir dos quais a convergência poderá ser obtida rapidamente. Um número crescente de pesquisadores compartilham o sentimento de que as relações entre variáveis biológicas são melhores descritas por funções não lineares. Processos como crescimento, decaimento, nascimento, mortalidade, abundância, competição e produção raramente são relacionadas linearmente às variáveis explicativas. Nesse trabalho usaremos esses dois tipos de modelagem não linear para estimar o ano de completude das coberturas dos óbitos.

3.3.1 Modelo Logístico

A curva logística faz parte das chamadas curvas de saturação. Ela é aplicada a fenômenos em que a taxa de crescimento de suas observações acumulada cresce até um determinado valor e, a partir desse ponto, começa a cair com a mesma intensidade do crescimento, tendendo para um valor estacionário no longo prazo. Assim, no ajuste logístico, está implícito que o ritmo de aumento das taxas de crescimento é igual ao do decréscimo, o que, no longo prazo, dificilmente acontece com a tendência de projeções humanas. Fazendo as devidas adaptações nas variáveis do modelo, a equação é dada por:

$$Y = \frac{\alpha}{1 + e^{-\gamma(x-\beta)}}$$

Em que:

Y : coberturas dos óbitos;

x : tempo em anos das coberturas dos óbitos;

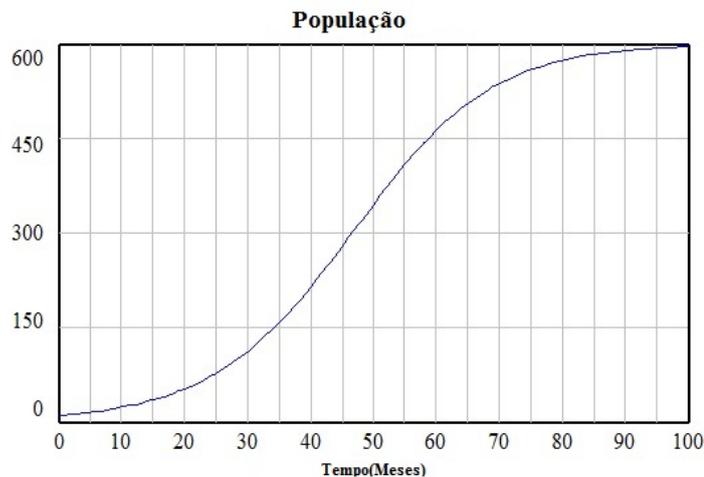
α : corresponde ao valor assintótico, indicando o valor de estabilização da variável dependente em relação ao tempo;

β : corresponde ao parâmetro de localização;

γ : medida relativa a taxa de crescimento da curva.

Em que α , β e γ são parâmetros, $\alpha > 0$ e $\gamma > 0$. A função que expressa este modelo foi indicada para o estudo descritivo do crescimento de populações humanas por Verhulst, que a denominou de "curva logística". Verifica-se que a função é monotonicamente crescente e fica entre duas assíntotas horizontais que são o eixo das abcissas e a reta ordenada constante e igual a α , cujo ponto de inflexão para a abcissa $X = \beta/\gamma$, quando $Y = \alpha/2$. A Figura 3.1 ilustra a curva do modelo Logístico.

Figura 3.1: Curva de Ajuste do Modelo Logístico



3.3.2 Modelo de Gompertz

A função de Gompertz foi inicialmente utilizada para descrever a taxa de mortalidade de uma população por Gompertz em 1825, cuja aplicabilidade é sugerida também em modelos sigmodais de crescimento. Baseia-se em que a taxa de crescimento relativo é decrescente de forma exponencial. A curva é assimétrica em relação a seu ponto de inflexão, sendo o ritmo de desenvolvimento mais lento depois de alcançado esse ponto, ou seja, a função de Gompertz foi desenvolvida sob a suposição de que a taxa de crescimento específico cai exponencialmente com o tempo. A equação do modelo é dada por:

$$Y = \alpha e^{-e^{\gamma(x-\beta)}}$$

Em que, feitas as devidas adaptações, tem-se que:

Y : coberturas dos órbitos;

x : tempo em anos das coberturas dos órbitos;

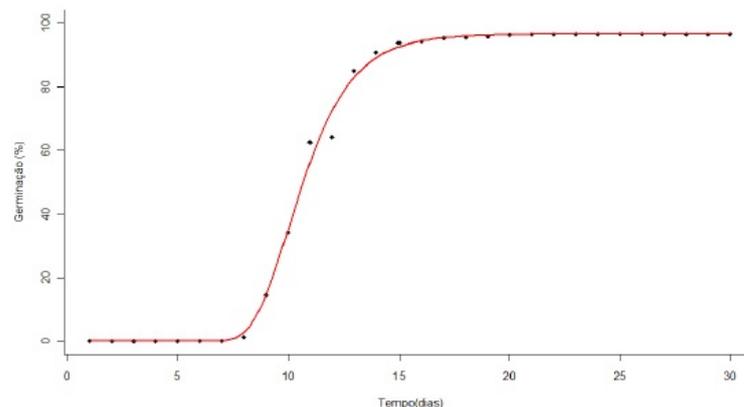
α : corresponde ao valor assintótico, indicando o valor de estabilização da variável dependente em relação ao tempo;

β : corresponde ao parâmetro de localização;

γ : medida relativa a taxa de crescimento da curva.

A curva de Gompertz é frequentemente usada em situações em que o crescimento não é simétrico em relação ao ponto de inflexão que ocorre em $x = \beta$ quando $Y = \frac{\alpha}{e}$, sendo $W_{mx} = \gamma \frac{\alpha}{e}$. A função de Gompertz é monotonicamente crescente e fica entre duas assíntotas horizontais: o eixo das abcissas e a reta de ordenada. Pode-se verificar que a função de Gompertz tem ponto de inflexão quando $X = -\ln(-\beta)/\ln\gamma$ e $Y = e^{\alpha-1}$. A Figura 3.2 ilustra a curva do modelo de Gompertz.

Figura 3.2: Curva de Ajuste do Modelo de Gompertz



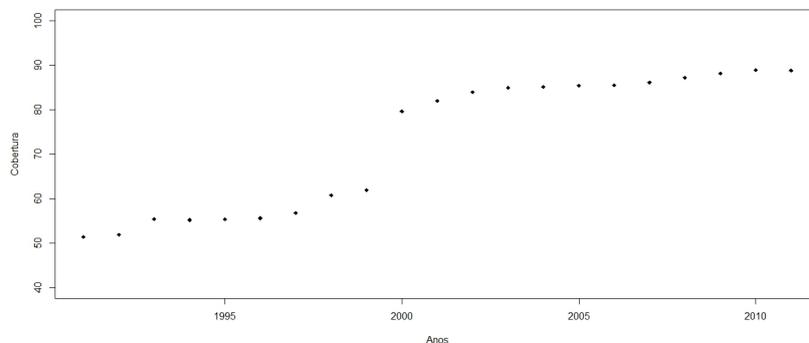
Além dos métodos de regressão não linear, foi abordado o Alisamento Exponencial de Holt, que é bastante utilizado para projetar estimativas em uma coorte através de uma série temporal, mas que foi usado para projetar a cobertura de órbitos. Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes ao

estar interessados em analisar e modelar esta dependência. Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável (MORETTIN&TOLOI, 2006).

Dize-se que uma série é sazonal quando os fenômenos que ocorrem durante o tempo se repetem a cada período idêntico de tempo. Por exemplo, é natural esperar que as vendas mensais de brinquedos terão um pico no mês de dezembro e talvez um pico secundário em outubro. Tendência é o comportamento de longo prazo da série, que pode ser causada pelo crescimento (ou decrescimento) de vários possíveis padrões (EHLERS, 2009).

Segundo Morettin&Toloi (2006), um primeiro passo na análise de uma série temporal é a construção de seu gráfico, que revelará características importantes, como tendência, sazonalidade, observações atípicas (outliers) e etc. Como por exemplo, na Figura 3.3, mostra-se uma tendência crescente no seu comportamento com o passar dos anos.

Figura 3.3: Cobertura dos Óbitos do Nordeste Brasileiro



Além dessa inspeção gráfica, é possível utilizar testes de hipóteses estatísticos para verificar a presença de tendência na série. Isto pode ser feito de duas maneiras: Antes ou depois da estimação de T_t . Todavia, é aconselhável estabelecer se existe tendência na série antes de aplicar qualquer procedimento para sua estimação. Existem alguns testes não-paramétricos que são úteis para testar se há tendência em um conjunto de observações. Antes da análise gráfica foram realizados testes com embasamentos não paramétricos: os testes de Dickey-Fuller e Wilcoxon. O teste de Dickey-Fuller possui as seguintes hipóteses:

H_0 : A série não é estacionária ($\gamma = 0$) ou possui raiz unitária ($\rho = 1$)

H_1 : A série é estacionária ($\gamma < 0$) ou não possui raiz unitária ($\rho < 1$)

As hipóteses para o teste de Wilcoxon são dadas por:

H_0 : A diferença entre as médias é zero ($\theta = 0$)

H_1 : A diferença entre as médias não é zero ($\theta \neq 0$)

O primeiro testa se as observações temporais de uma série são estacionárias e o outro detecta a presença de tendência através do sinal de pares ordenados de uma série temporal, cujas hipóteses nulas são rejeitadas quando $p - valor \leq 0,05$.

3.3.3 Modelo de Suavização Exponencial de Holt

Os modelos de Suavização Exponencial de Holt (SEH) são utilizados satisfatoriamente em séries temporais que apresentam tendência de crescimento linear. Considere o caso de uma série temporal que é composta localmente de soma de nível, tendência e resíduo com média zero e variância constante σ_a^2 , isto é

$$Z_t = \mu_t + T_t + a_t, t = 1, \dots, N,$$

Em que:

μ_t é o nível do modelo;

T_t é a tendência modelo;

a_t é o erro aleatório ocorrido no tempo t .

Os valores do nível e da tendência da série serão estimados por

$$\bar{Z}_t = AZ_t + (1 - A)(\hat{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1, t = 2, \dots, N,$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1, t = 2, \dots, N,$$

A, C são as constantes de suavização. As previsões dos valores futuros da série para esse procedimento é dado por:

$$\bar{Z}_t(h) = \bar{Z}_t + h\hat{T}_t, \forall h > 0$$

ou seja, a previsão é feita adicionando-se ao valor básico (\bar{Z}_t) a tendência multiplicativa pelo número de passos à frente que se deseja prever (h).

3.4 Medidas de Diagnósticos

Em regressão não linear, a análise dos resíduos de um modelo é feita para verificar a plausibilidade das pressuposições envolvidas. Para a verificação dos pressupostos, pode-se utilizar análise gráfica dos resíduos, sendo este um método informal de análise que envolvem os gráficos de resíduos em relação às variáveis independentes e aos valores preditos, ou por meio de testes estatísticos, que é uma maneira mais objetiva de se analisar os resíduos por fornecer uma medida numérica para algumas das discrepâncias previamente descritas.

Para a verificação do pressuposto de normalidade, várias formas gráficas têm sido propostas como histogramas dos resíduos, em que se podem visualizar desvios grosseiros da normalidade ou "Q-Q plot" que é um gráfico que testa a conformidade entre a distribuição empírica e uma dada distribuição teórica, apresentando como regra de decisão se a observação gráfica entre os quantis da variável aparece alinhados com os quantis da distribuição. O teste estatístico de Shapiro-Wilk é um dos mais utilizados para a verificação do pressuposto em que rejeita a hipótese nula quando o $p - valor \leq 0,05$, suas hipóteses são:

$$H_0 : \text{Possui normalidade } (F(z) = \phi(z), \forall z)$$

$$H_1 : \text{Não possui normalidade } (F(z) \neq \phi(z), \text{ para algum } z)$$

Um gráfico dos resíduos contra os valores estimados pode ser utilizado para examinar se as variâncias dos erros são constantes, observando se a dispersão dos resíduos ocorre aleatoriamente em torno da média zero. O teste estatístico frequentemente utilizado para heterocedasticidade é chamado de Breusch-Pagan, em que rejeita -se a hipótese nula quando $p - valor \leq 0,05$, suas hipóteses são:

H_0 : Possui homocedasticidade ($\alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_n$)

H_1 : Não possui homocedasticidade (*algum* $\alpha_i \neq 0$)

Resíduos relacionados com o tempo são chamados auto correlacionados e para verificar a existência de correlação entre os resíduos, pode-se plotá-los contra o tempo ou qualquer variável de interesse. Quando os erros são independentes, espera-se que os mesmos flutuem aleatoriamente em torno da média zero. O teste de Durbin-Watson será utilizado para verificar a existência de autocorrelação de primeira ordem, suas hipóteses são:

H_0 : Não há correlação entre os resíduos ($\rho = 0$)

H_1 : Há correlação entre os resíduos ($\rho \neq 0$)

Tem-se que $p - valor \leq 0,05$, a hipótese nula é rejeitada.

3.5 Análise Residual

O Viés estatístico (V) é a medida mais simples e menos representativa. É dado pela média da diferença entre a série estimada e a observada, e indica o erro sistemático (GOMES, 2003).

$$V = \frac{\sum_{i=1}^N (V_{o_i} - V_{s_i})}{N}$$

Em que:

V_o : é o valor da variável observada;

V_s : o valor da variável estimada;

N : o período de tempo.

O Viés pode ter qualquer valor real, onde os valores mais próximos de zero, por definição, representam menor erro. É medido com a mesma unidade usada nas séries.

3.5.1 Erro Absoluto Médio

O Erro Absoluto Médio (EAM) é considerado uma medida de erro mais justa. Na mesma os desvios negativos não se cancelam com os positivos, o que o torna mais representativo que o Viés.

$$EAM = \frac{\sum_{i=1}^N |Vo_i - Vs_i|}{N}$$

Em que:

Vo : é o valor da variável observada;

Vs : o valor da variável estimada/prevista;

N : o período de tempo.

Pela definição, só pode assumir valores não negativos, sendo que quanto menor o valor maior semelhança entre as séries. É medido com a mesma unidade usada nas séries.

3.5.2 Erro Percentual Médio

O Erro Percentual Médio (EPM) é calculado dividindo o Erro Percentual (EP) pelo número de períodos N . O EP mede a porcentagem do erro em relação ao valor real.

$$EPM = \frac{\sum_{i=1}^N \left(\frac{Vo_i - Vs_i}{Vo_i} \right)}{N} \times 100$$

Em que:

Vo : é o valor da variável observada;

Vs : o valor da variável estimada/prevista;

N : o período de tempo.

Se os erros positivos forem compensados pelos erros negativos, o resultado deve ser aproximadamente nulo.

3.5.3 Erro Quadrático Médio

O Erro Quadrático Médio (EQM) é definido pela soma dos quadrados das diferenças entre resultados estimados/previstos e as observações.

$$EQM = \frac{\sum_{i=1}^N (V_{o_i} - V_{s_i})^2}{N}$$

Em que:

V_o : é o valor da variável observada;

V_s : o valor da variável estimada/prevista;

N : o período de tempo.

É a medida de erro mais comumente usada. É muito sensível a grandes desvios entre os valores das séries comparadas, o que se torna mais relevante quando se trata de avaliação de grandes erros. Pode assumir qualquer valor não negativo e tem a mesma unidade de medida das séries. É maior a semelhança entre as séries estimada/prevista e a observada quanto mais próximo de zero for o seu valor.

3.6 Cobertura dos Óbitos para o Semiárido

Ao longo do tempo, vários Métodos foram propostos para calcular a cobertura de óbitos entre eles Brass (1975), Courbaje e Fargues (1979), Preston et al., (1980). O método proposto por Brass se baseia na teoria das populações estáveis e configura-se como robusto, mesmo em situação de descumprimento de certos supostos. Para a aplicação do método é necessário o conhecimento da distribuição da população por faixa etária além da distribuição dos óbitos. Estimar a cobertura de óbitos de uma determinada região equivale ao ajuste do seguinte modelo de regressão:

$$\frac{N(a)}{N(a+)} = r + K \cdot \frac{D'(a+)}{N(a+)}$$

Onde,

$N(a)$ =população na idade exata a ;

$N(a+)$ =somatório de pessoas que estão na idade exata até um limite de idade “ w ”

qualquer;

$D'(a+)$ = óbitos registrados e afetados por erros na idade $a+$;

r = taxa de crescimento natural;

k = fator de correção para a cobertura dos óbitos.

Para estimar a cobertura de óbitos para os espaços geográficos dos Estados do Semiárido brasileiro fez-se uso da técnica de Brass (1975) para os anos censitários: 1991, 2000 e 2010.

1. Para os anos intercensitários entre 1991, 2000 e 2010 recorreu-se ao seguinte procedimento. Foram obtidos os óbitos observados entre os anos de 1991 a 2010 para os espaços geográficos dos Estados do Semiárido brasileiro, excluindo Minas Gerais;
2. Com as coberturas de óbitos estimadas pela RIPSAs referentes aos anos de 1991, 2000 e 2010 para os Estados do Semiárido brasileiro calculou-se o fator de correção para cada ano, cuja equação é dada por:

$$FC_i = \frac{1}{Cobertura_i} \cdot 100, \quad i = 1991, 2000 \text{ e } 2010$$

3. Após os cálculos dos fatores de correção, foram obtidos os óbitos corrigidos da seguinte forma:

$$Obito_cor_i = Obito_obs_i \cdot FC_i, \quad i = 1991, 2000 \text{ e } 2010$$

4. Com os óbitos corrigidos para os anos e Estados analisados obteve-se a diferença, da seguinte maneira:

$$D_i = Obito_obs_i - Obito_cor_i, \quad i = 1991, 2000 \text{ e } 2010$$

5. Logo após, calculou-se a média entre as diferenças dos Óbitos, criando duas médias:

$$\bar{D}_1 = \frac{D_{1991} + D_{2000}}{2}$$

$$\bar{D}_2 = \frac{D_{2000} + D_{2010}}{2}$$

6. O próximo passo consistiu em estimar os óbitos esperados para os Estados do Semiárido entre os anos de 1991 a 2010 utilizando as diferenças médias em dois períodos de tempo: \bar{D}_1 foi utilizado no período dos anos (1992-1999) e \bar{D}_2 nos anos (2001-2009), com o seguinte cálculo:

$$Obito_esp_{1992-1999} = Obito_obs_{1992-1999} + \bar{D}_1$$

$$Obito_esp_{2001-2011} = Obito_obs_{2001-2009} + \bar{D}_2$$

7. Finalmente, calculou-se a Cobertura de Óbitos para os Estados do Semiárido entre os anos de 1991 a 2010, exceto para os anos censitários

$$Cobertura_SA_i = \frac{Obito_obs_i}{Obito_esp_i} \cdot 100 \quad i = 1991, \dots, 2010$$

Nota-se que os anos censitários de 1991, 2000 e 2010 foram usados como referência entre toda a série das Coberturas de Óbitos.

3.7 Critérios de Seleção dos Modelos

A escolha do modelo mais apropriado, do ponto de vista estatístico, é um tópico extremamente importante na análise de dados. Busca-se aqui o modelo que apresente resultados plausíveis com a realidade, criando cenários prospectivos sobre a cobertura dos óbitos até atingirem o grau de completitude. Os critérios estabelecidos são:

I - Escolheu-se como ano de cobertura ótimo o primeiro ano que teve a estimativa maior ou igual a 99% ou quando o ponto máximo de inflexão da curva do modelo foi alcançado;

II - O modelo de previsão adequado para o Nordeste e seus Estados será aquele que apresentar os menores erros de estimação, entre os 3 tipos de medidas de precisão (EAM, EPM e EQM) descritos nesse trabalho;

III - Não será admitido nenhum ano de completitude das coberturas antes de 2020 para o Nordeste e seus Estados, ou seja, a partir do ano de 2020 as coberturas dos óbitos passarão a atingir sua universalização. O ano de 2020 foi a previsão mais baixa entre os 3 modelos previsto para o Nordeste, de acordo com a Tabela 4.5;

IV- Caso as projeções para os modelos sejam muito díspares será adotada a média entre elas;

V - Os critérios I a IV serão adotados para as estimativas do Semiárido brasileiro, visto que, o ano de cobertura ótimo para essa região somente atingirá a universalização depois do Estado como um todo.

Capítulo 4

Resultados e Discussões

4.1 Cobertura dos Óbitos para os Estados do Nordeste brasileiro

Nessa seção são apresentadas as estimativas das coberturas de óbitos. A Tabela 4.1 apresenta a série temporal anual de 1991 a 2011 da Cobertura de Óbitos para o Brasil, Nordeste e os Estados do Nordeste. Estas estimativas foram produzidas pela RIPSa e estão disponibilizados no site do Ministério da Saúde (2011). Observa-se claramente uma tendência crescente desses valores ao longo do tempo, sendo que a partir do ano 2000 essa evolução ocorreu de forma mais rápida.

No início da série em 1991, as coberturas para os Estados do Maranhão, Piauí, Ceará e Rio Grande do Norte estavam abaixo de 50%, tendo atingindo o Nordeste uma cobertura de 51,4%. O Estado com a cobertura mais baixa, o Piauí se distanciou em 52,5% com relação a cobertura para o Brasil. Um dos fatores que esclarece a qualidade deficiente nos dados de óbitos, se refere a redução dos gastos na saúde pelo governo brasileiro durante esse período (POLIGNANO, 2004). Essa falta de investimento acabou gerando dificuldades na captação das estatísticas de mortalidade do Brasil e suas regiões, dificultando a melhoria da qualidade dos dados de óbitos.

Observa-se também na Tabela 4.1 que no período de 1991 a 1999 os Estados do PI e MA obtiveram as menores estimativas de coberturas entre os Estados do Nordeste, indicando assim possíveis fontes de erro na qualidade desses óbitos. Segundo Campelo (1996), o Estado do Piauí no primeiro período da série (1991-1999) enfrentou uma situação irregular quanto a veracidade das suas informações desde a década de 70.

Tabela 4.1: Cobertura dos Óbitos para o Brasil, Nordeste e Estados do Nordeste, de 1991 a 2011.

Anos	BR	NE	MA	PI	CE	RN	PB	PE	AL	SE	BA
1991	77,6	51,4	31,4	25,1	41,1	44,9	53,5	70,0	57,7	64,2	56,9
1992	78,6	51,9	31,2	32,6	41,0	46,2	51,0	68,7	56,7	73,9	57,8
1993	82,9	55,4	31,7	36,3	49,5	53,5	58,5	72,8	57,8	67,9	58,9
1994	83,4	55,2	28,2	35,9	52,7	52,5	56,5	70,6	55,7	68,7	61,3
1995	83,6	55,4	28,5	34,1	51,5	58,7	57,9	68,0	58,1	80,2	60,8
1996	84,6	55,6	27,3	30,7	53,6	57,3	56,1	70,2	55,5	74,0	62,9
1997	83,7	56,8	30,6	35,6	57,2	56,8	55,9	72,6	57,8	70,7	61,0
1998	85,8	60,8	34,6	39,0	58,7	59,9	58,0	77,0	67,5	78,4	65,3
1999	85,9	61,9	32,7	40,7	64,5	61,2	57,0	77,6	58,8	80,3	68,6
2000	91,0	79,6	55,2	73,1	80,6	77,7	82,1	91,7	87,1	88,2	77,7
2001	91,7	82,0	64,0	79,5	83,1	80,4	83,9	92,2	87,8	89,0	78,8
2002	92,5	83,9	69,1	82,7	84,7	81,0	85,8	92,7	89,0	89,7	81,1
2003	92,9	84,9	72,3	84,0	86,4	81,8	87,3	92,8	89,3	90,1	81,6
2004	93,1	85,1	73,5	84,7	86,5	81,5	88,2	92,7	89,2	90,3	81,9
2005	93,2	85,4	74,4	84,8	86,4	80,6	88,2	92,9	88,7	89,9	83,0
2006	93,3	85,5	74,2	85,5	86,1	80,4	88,4	93,0	89,7	89,8	83,2
2007	93,6	86,1	74,4	86,8	86,3	81,8	89,4	93,2	90,2	90,0	84,0
2008	94,0	87,2	76,5	87,3	87,9	84,8	89,8	93,3	91,5	91,8	84,9
2009	94,3	88,1	78,5	87,4	88,7	86,2	90,7	93,6	91,7	92,4	86,2
2010	94,2	88,9	78,8	88,2	90,0	87,9	91,1	93,6	92,3	92,5	87,4
2011	94,2	88,8	79,1	88,1	89,8	87,6	91,2	93,5	92,2	92,7	87,4
Mínimo	77,6	51,4	27,3	25,1	41,0	44,9	51,0	68,0	55,5	64,2	56,9
Máximo	94,3	88,9	79,1	88,2	90,0	87,9	91,2	93,6	92,3	92,7	87,4
Amp 1991-1999	8,3	10,5	7,3	15,6	23,5	16,3	7,5	9,6	12,0	16,1	11,7
Amp 1999-2000	5,1	17,7	22,5	32,4	16,1	16,5	25,1	14,1	28,3	7,9	9,1
Amp 2000-2011	3,3	9,3	23,9	15,1	9,4	10,2	9,1	1,9	5,2	4,2	9,7

Fonte: Ministério da Saúde – DATASUS, 2011.

Por conseguinte ainda nesse mesmo período, apenas 63% da população estava coberta e 44,8% dos municípios do PI obtinham informações regulares das estatísticas de óbito. O Maranhão nesse período apresentou um alto nível de subregistro de mortalidade, cujo erro na declaração das idades e o alto índice de causas mal definidas na DO, influenciaram na sua qualidade (QUEIROZ, 2012).

Uma olhada mais detalhada para a evolução das coberturas revela que o ano 2000 pode ser considerado como um marco no tempo, que desencadeou um ritmo de aumento sem precedentes na história. Ao considerar o período de 1999 a 2000, o Brasil acusou menor aumento na comparação da evolução ano a ano de toda a série atingindo 5,1%. Para o Nordeste este aumento foi de 17,7%. Neste mesmo período o Estado com menor evolução foi SE com 7,9% e o maior foi PI com 32,4%

Paes (1999) destaca que os Estados do Nordeste, pelo menos desde 1980, apresentavam importantes inconsistências na captura dos registros de óbitos. Uma das hipóteses colocadas era a pouca intervenção do sistema na captação dos dados pelo MS e pelo Registro Civil, no sentido de melhorar e aumentar os registros de óbitos. Polignano (2004) afirma que no Brasil durante o período de 1991 a 1999, houve uma enorme crise política e econômica que afetou diretamente os investimentos na saúde e atenção básica. Uma das hipóteses levantadas para explicar essa deficiência na captação dos óbitos, se refere ao precário treinamento dos profissionais da saúde quanto a coleta, manipulação e tratamento dos dados e a não padronização dessas tarefas o que propiciou a má qualidade e a não fidedignidade das informações (JORGE, 2010).

No período 1999-2000 o Estado do Nordeste com a maior amplitude foi o Piauí (32,4%) e a menor Alagoas (7,9%). O fato do PI ter a menor cobertura em 1991 possibilitou um avanço mais rápido que os Estados com níveis mais elevados, ou seja, é mais fácil provocar impactos em regiões com níveis de cobertura mais baixos que naqueles com níveis mais elevados.

Entre os fatores que contribuíram para essa melhora no ano 2000 foi o desenvolvimento tecnológico da informática que possibilitou um considerável salto de qualidade nos registros, coletas e processamentos dos dados. Jorge (2007) argumenta que houve uma melhora significativa na qualidade dos registros de óbitos do SIM a partir do ano 2000, e que tal fato se remete a um grande volume de registros da DO devido ao adcionamento dos hospitais na coletas de dados, antes coletados apenas em cartórios, existindo, assim uma mudança metodológica na forma de captação dos registros da DO no Brasil.

No período de 2000 a 2011 todas coberturas da série ficaram acima de 70%, exceto o MA no período de 2000 a 2002. Em 2011, vários Estados atingiram um patamar acima de 90% (PB, PE, AL e o SE). A maior diferença anual encontrada ocorreu no Estado do Maranhão (23,9%) e a menor em Pernambuco (1,9%). No mesmo intervalo, o Brasil teve uma diferença de 3,3% e o Nordeste de 9,3%. É bem razoável esperar que o ritmo de aumento diminua com o tempo, particularmente quando as coberturas atingirem níveis elevados, ou seja, quanto mais próximo de 100% menor é o aumento. Sendo assim, pode-se dizer que o ritmo de crescimento no Nordeste foi mais veloz que no restante do País.

4.2 Análise Residual

Neste trabalho foram calculadas medidas de precisão para avaliar os erros entre as estimativas geradas pelos modelos. Aplicando os critérios de seleção do Erro Quadrático Médio (EQM), Erro Absoluto Médio (EAM) e Erro Percentual Médio (EPM), foi possível analisar qual dos modelos teve o menor desvio em relação ao padrão do Estado. A Tabela 4.2 apresenta as medidas de precisão anual para as estimativas das coberturas completas de órbitos para os Estados do Nordeste.

Tabela 4.2: Medidas de Precisão para estimativas das coberturas completas de órbitos para o Nordeste e os Estados do Nordeste.

Estados	EAM			EPM			EQM		
	Logístico	Gompertz	Holt	Logístico	Gompertz	Holt	Logístico	Gompertz	Holt
Maranhão	6,02	6,51	3,03	-1,47	-2,32	0,97	52,28	58,44	29,37
Piauí	6,98	7,62	4,74	-0,93	-1,85	-4,45	70,01	80,20	56,85
Ceará	3,59	3,82	2,67	-0,17	-0,27	0,67	17,82	19,91	16,49
Rio Grande do Norte	2,98	3,05	2,22	-0,24	-0,27	1,19	14,38	15,01	15,61
Paraíba	4,84	4,97	2,88	-0,68	-0,75	1,61	36,89	38,24	37,70
Pernambuco	3,57	3,63	2,38	-0,24	-0,26	0,45	17,53	18,01	13,14
Alagoas	5,18	5,32	3,76	-0,74	-0,81	1,30	41,97	43,48	46,04
Sergipe	2,55	2,57	3,59	-0,17	-0,18	-2,15	10,38	11,01	29,03
Bahia	2,18	2,23	1,27	-0,12	-0,14	0,63	7,38	7,69	4,84
Nordeste	3,90	4,05	1,67	-0,35	-0,42	1,11	21,72	22,92	15,16

*EAM - Erro Absoluto Médio; EPM - Erro Percentual Médio; EQM - Erro Quadrático Médio.

Observando o EAM percebe-se que o modelo de Holt forneceu desvios menores para todas as previsões dos Estados e NE exceto para SE (desvio de 2,55). Comparando o EAM do modelo Logístico para o NE houve um aumento de 2,23 desvios com relação ao

de Holt. Fazendo esta mesma comparação o modelo de Gompertz teve o pior desempenho entre os modelos, cujos EAM foram bastantes altos com um aumento de 2,38 com relação ao de Holt.

Outra medida de precisão utilizada foi o EPM que fornece uma indicação do tamanho médio do erro, expresso como uma porcentagem do valor observado, independentemente do erro ser positivo ou negativo. Ao analisar o EPM percebe-se que o modelo Logístico obteve os menores valores para o NE e os Estados, cuja variação das estimativas ficaram entre -1,5 a 0.

O EPM de Gompertz foi o segundo a apresentar estimativas razoáveis, cujos valores ficaram entre -2,5 a 0. O modelo de Holt teve o pior desempenho com altos valores de EPM, variando entre -4,5 a 1,7. Lopes (2002) afirma que o valor do EPM deve ser próximo de zero, caso contrário, diz-se que o modelo apresenta viés, ou seja, as estimativas estão desviando-se para cima ou para baixo da linha de tendência. Nota-se essas mudanças de comportamento no modelo de Holt, cujo EPM variou bastante em torno de zero.

O EQM para as estimativas das coberturas para os Estados do Nordeste foram bastantes distintos, variando entre 4 a 80,5. Analisando o EQM percebe-se que o modelo de Holt teve o melhor desempenho, com as menores estimativas para os Estados do Nordeste, seguido pelo modelo Logístico e de Gompertz. Avaliando por Estados, a BA obteve as menores estimativas para os três modelos, com valores entre 4 a 8 e o PI as maiores estimativas do EQM, variando entre 56 a 80,5.

Segundo Keyfitz (1981), o EQM é a medida de precisão mais útil para avaliar a qualidade das estimativas, pois verifica o menor desvio entre um ponto “base” e a média de cada pontos da série temporal usada. Em geral, nota-se que o modelo de Holt obteve o melhor desempenho no conjunto das 3 medidas de precisão, apresentando os menores erros. Analisando o desempenho das modelagens por Estado, escolheu-se o melhor modelo aquele que apresentou os menores erros entre o conjunto das medidas de precisão, como abordado na seção 3.7

4.3 Medidas de Diagnósticos

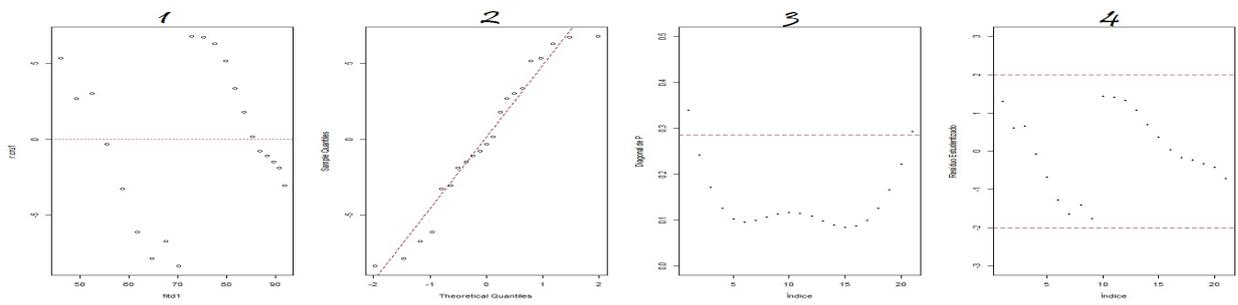
As medidas de diagnósticos são ferramentas utilizadas para análise de resíduos, detecção de valores extremos (outliers), de pontos influentes (alavanca) e para avaliação de colinearidade. Além dessas medidas realizou-se aqui testes baseados em hipóteses estatísticas, para verificar a adequabilidade dos ajustes dos modelos Logístico e Gompertz como pode ser visto na Tabela 4.3.

Tabela 4.3: Estimativas do valor-p residual dos testes Shapiro-Wilk, Breusch-Pagan e Durbin-Watson para o Nordeste e Estados do Nordeste

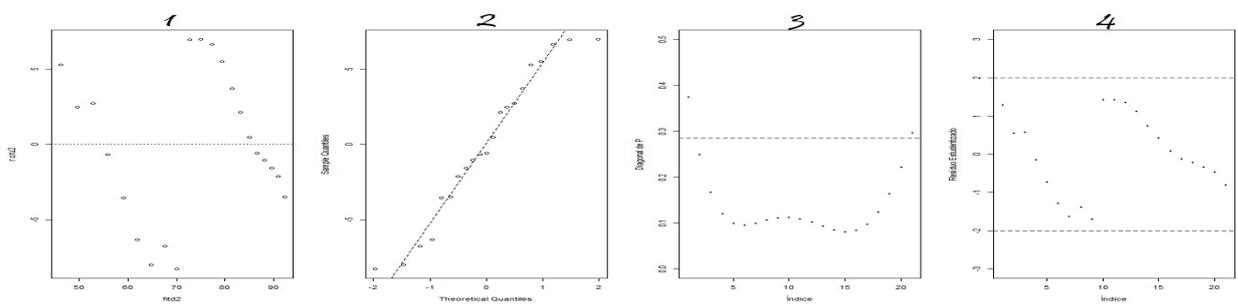
Estados	Logístico			Gompertz		
	Shapiro	Breusch-Pagan	Durbin-Watson	Shapiro	Breusch-Pagan	Durbin-Watson
Maranhão	0,5178	0,9952	0,0000	0,3663	0,9973	0,0000
Piauí	0,1871	1,0000	0,0000	0,2012	1,0000	0,0000
Ceará	0,5734	0,9999	0,0000	0,5174	0,9999	0,0000
Rio Grande do Norte	0,6215	0,9996	0,0020	0,6377	0,9997	0,0000
Paraíba	0,2502	1,0000	0,0000	0,2942	1,0000	0,0020
Pernambuco	0,4962	1,0000	0,0000	0,5066	1,0000	0,0000
Alagoas	0,7214	1,0000	0,0020	0,7686	1,0000	0,0020
Sergipe	0,4796	1,0000	0,2240	0,5167	1,0000	0,2120
Bahia	0,7578	0,9998	0,0000	0,7345	0,9998	0,0000
Nordeste	0,3049	1,0000	0,0000	0,2846	1,0000	0,0000

Para o diagnóstico de heteroscedasticidade, tentou-se encontrar alguma tendência no gráfico. Por isso, se os pontos estão aleatoriamente distribuídos em torno do 0, sem nenhum comportamento ou tendência, tem-se indícios de que a variância dos resíduos é homoscedástica. Nas Figuras 4.1 a 4.10 apresenta-se as análises gráficas dos resíduos para os Modelos Logístico e Gompertz. O Gráfico 1 das figuras representam os resíduos padronizados versus os valores ajustados, onde verificou-se a presença de aleatoriedade nas extremidades dos gráficos, indicando, assim, a homogeneidade das variâncias que pode ser confirmado pelas estimativas da estatística do teste de Breusch-Pagan apresentado na Tabela 4.3.

Figura 4.1: Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Nordeste

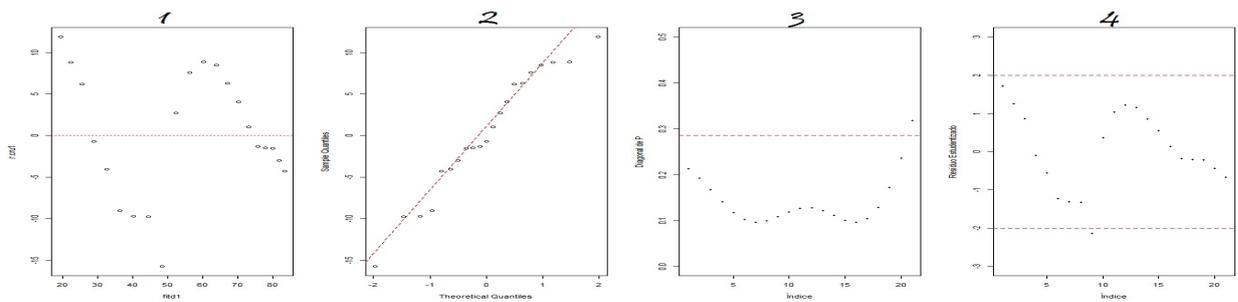


(a) Logístico

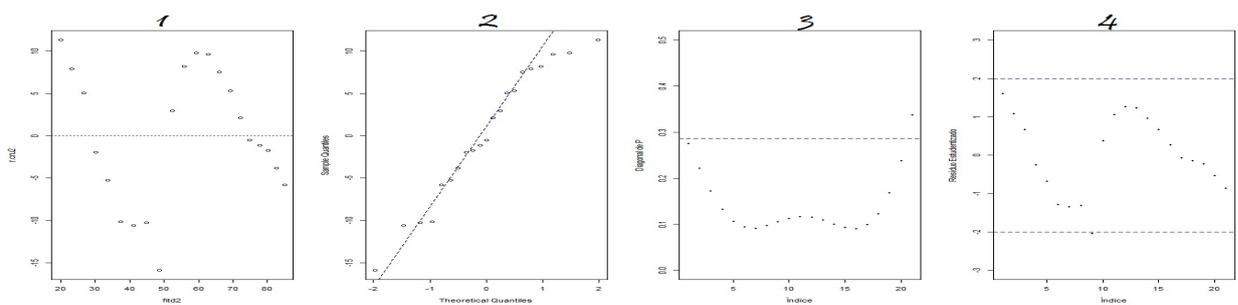


(b) Gompertz

Figura 4.2: Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Maranhão

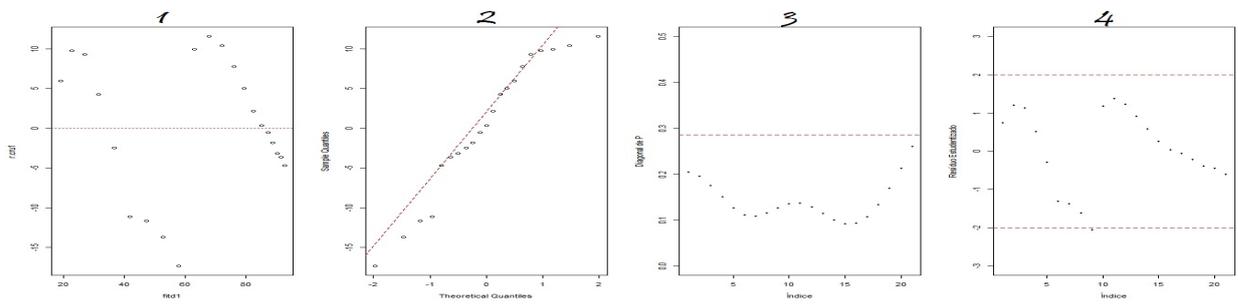


(a) Logístico

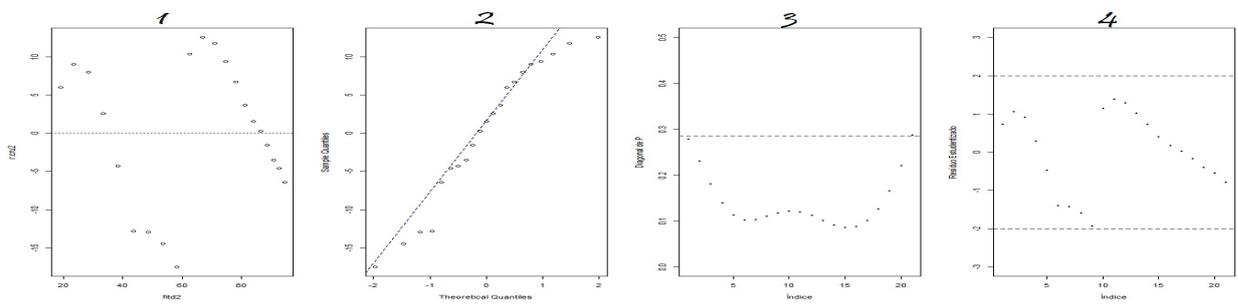


(b) Gompertz

Figura 4.3: Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Piauí

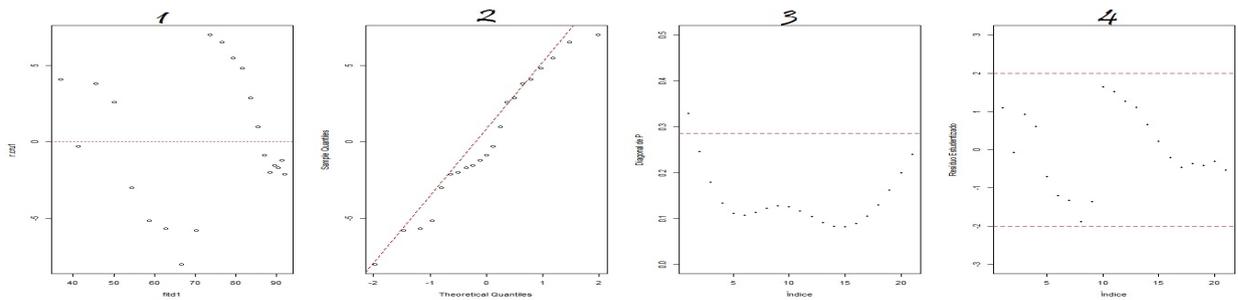


(a) Logístico

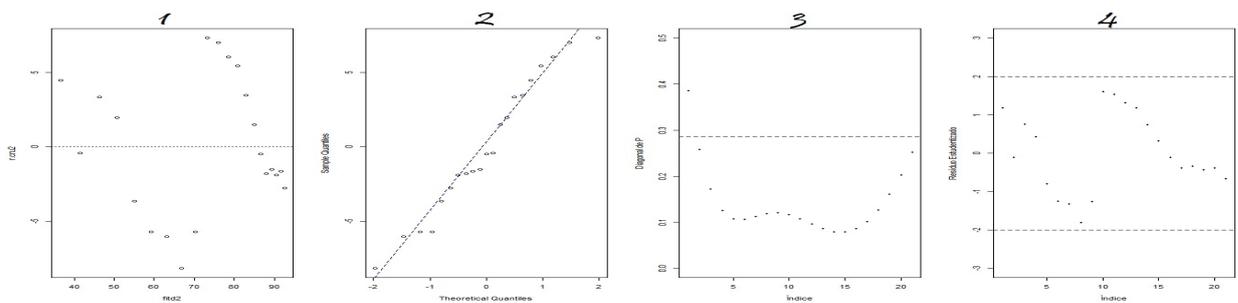


(b) Gompertz

Figura 4.4: Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Ceará



(a) Logístico



(b) Gompertz

Figura 4.5: Análise gráfica dos resíduos do modelo Logístico e Gompertz para o Rio Grande do Norte

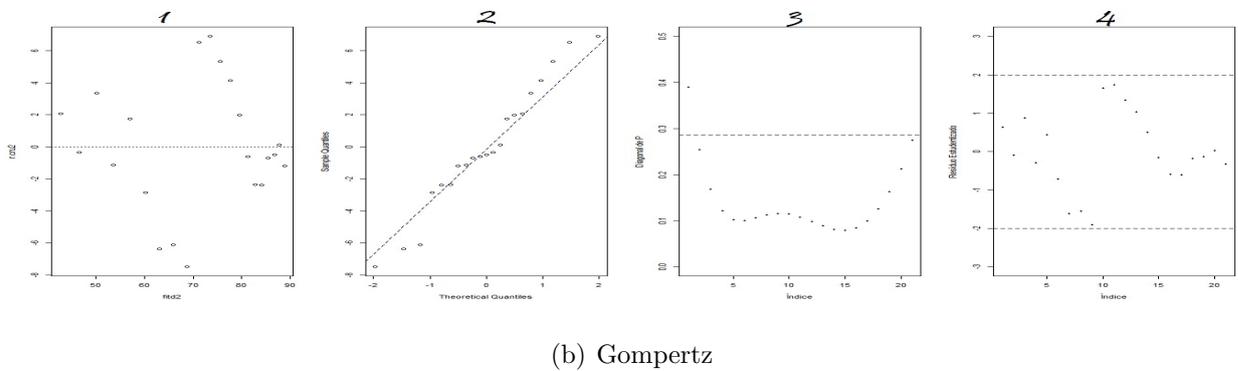
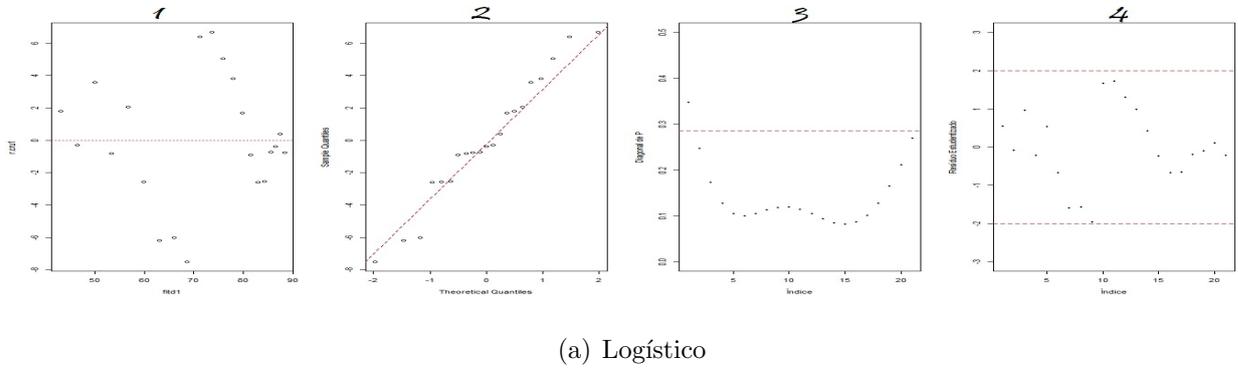


Figura 4.6: Análise gráfica dos resíduos do modelo Logístico e Gompertz para a Paraíba

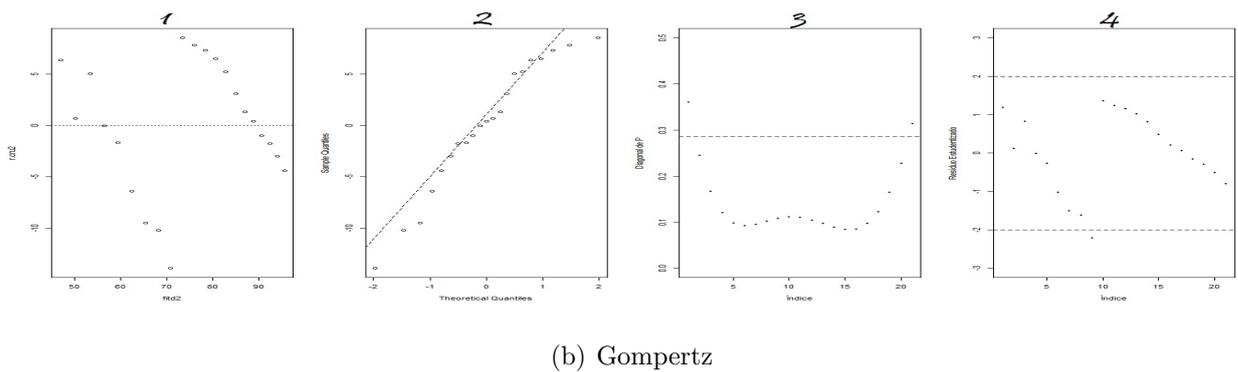
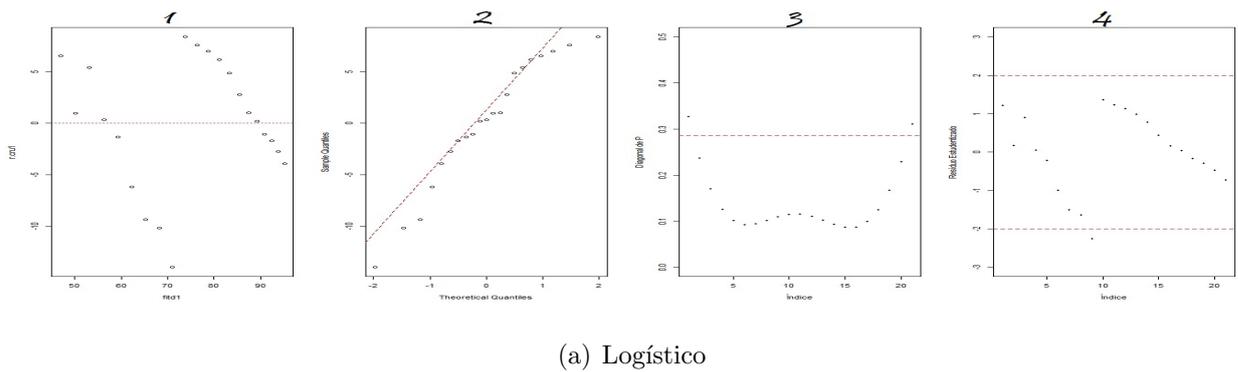
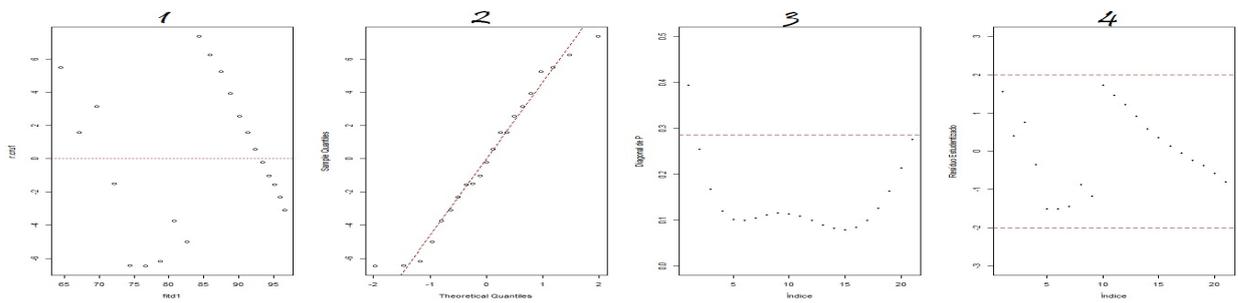
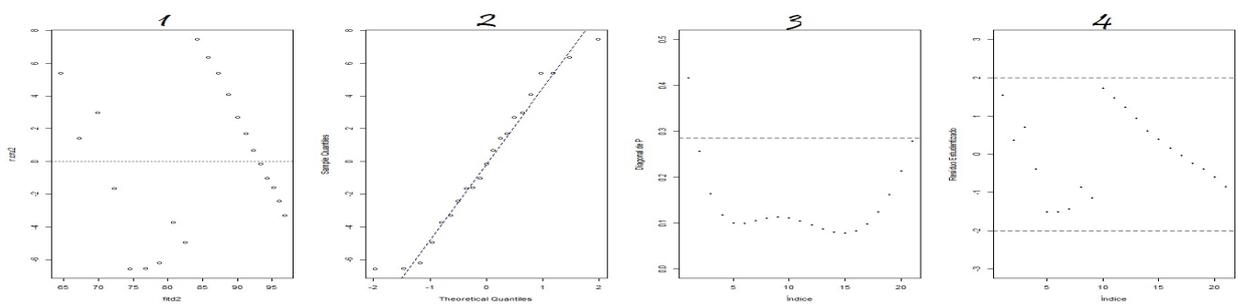


Figura 4.7: Análise gráfica dos resíduos do modelo Logístico e Gompertz para Pernambuco

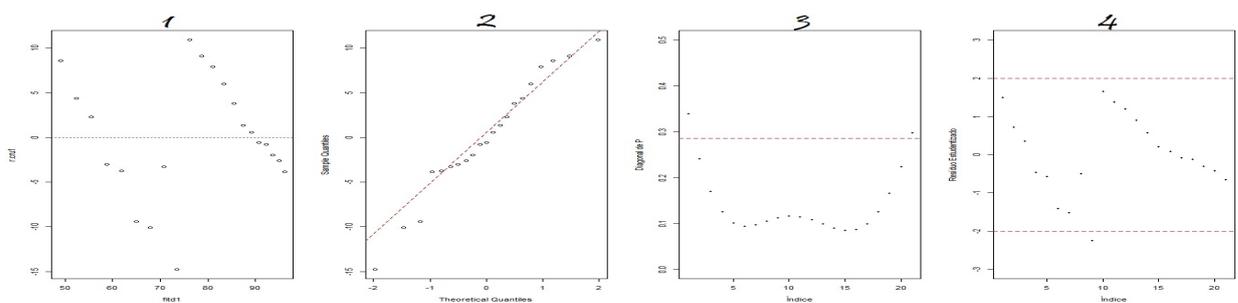


(a) Logístico

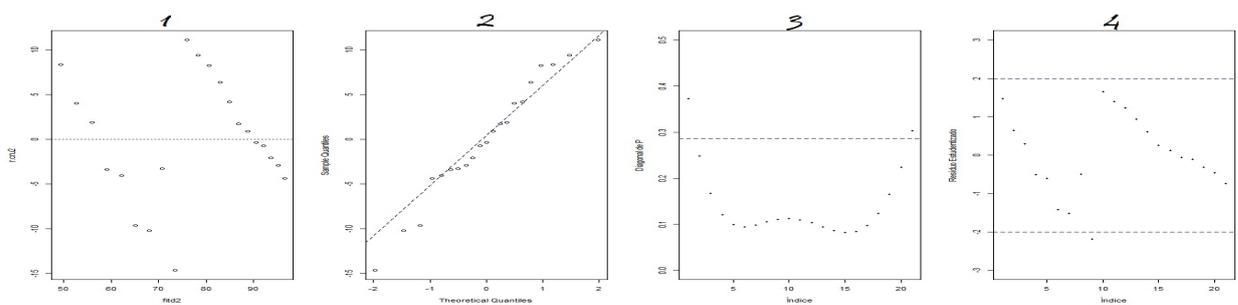


(b) Gompertz

Figura 4.8: Análise gráfica dos resíduos do modelo Logístico e Gompertz para Alagoas

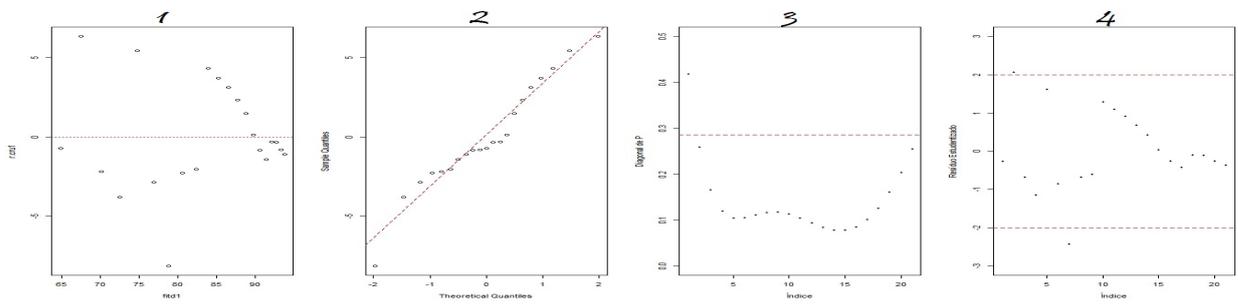


(a) Logístico

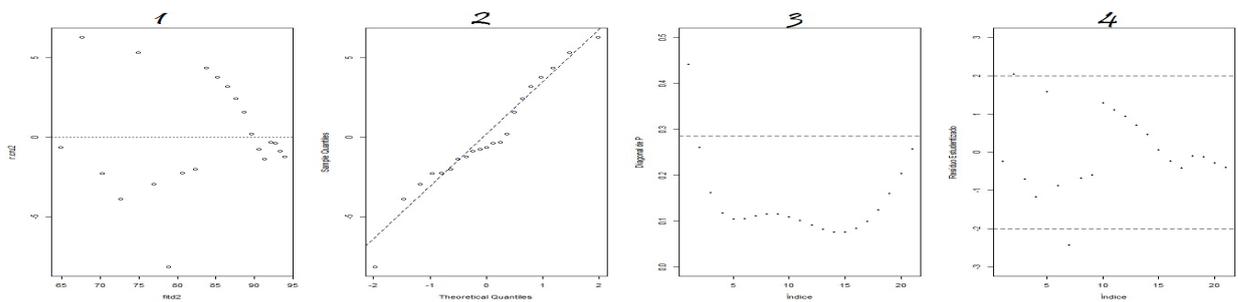


(b) Gompertz

Figura 4.9: Análise gráfica dos resíduos do modelo Logístico e Gompertz para Sergipe

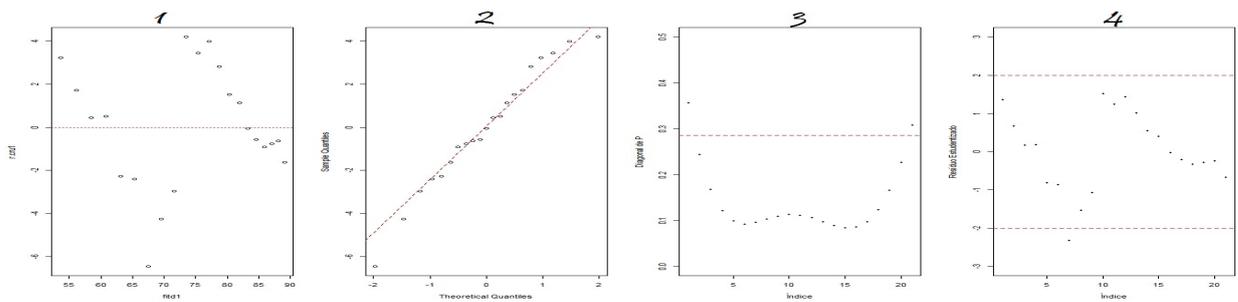


(a) Logístico

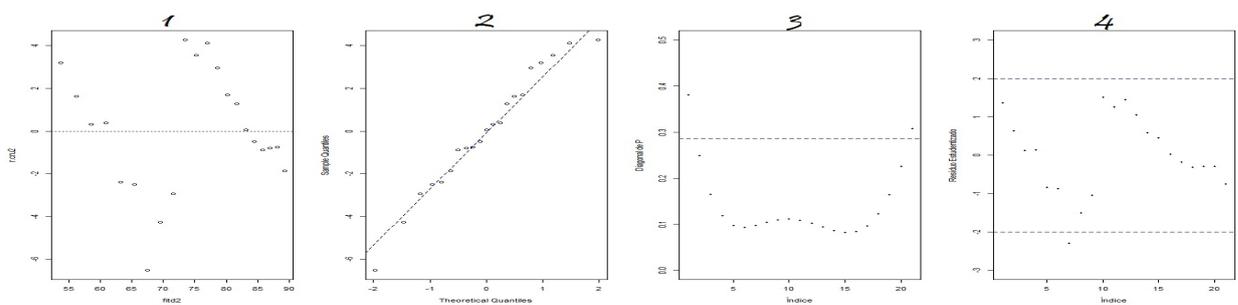


(b) Gompertz

Figura 4.10: Análise gráfica dos resíduos do modelo Logístico e Gompertz para a Bahia



(a) Logístico



(b) Gompertz

Verificando a pressuposição de normalidade (Gráfico 2) pelo "Q-Q Plot" dos resíduos studentizados para o Modelo Logístico nota-se que para o Nordeste e os Estados MA, CE, PE, AL, SE, BA os valores se aproximaram da reta do ajuste, mostrando assim a normalidade dos resíduos. Já para os Estados PI, RN e PB os valores se afastaram da reta do ajuste, por causa da presença de pontos influentes e aberrantes nos extremos. Para o modelo de Gompertz, o Nordeste e os Estados PI, RN e PB apresentaram alguns pontos influentes e aberrantes na hora da estimação dos resíduos, e os demais Estados as estimativas giraram em torno da reta do ajuste, mas com leves desvios.

Além da análise gráfica calculou-se o Teste de Shapiro-Wilk para verificar se os resíduos segue uma distribuição normal. Na Tabela 4.3 nota-se que as estimativas dos p-valores foram maiores que 5% confirmando, assim, a normalidade dos resíduos. Uma análise gráfica para verificar a hipótese de independência dos resíduos pode ser feita por meio do gráfico dos resíduos versus a ordem da coleta dos dados (Gráficos 3 e 4).

Se ao avaliar esses gráficos das Figuras 4.1 a 4.10 e percebendo a existência de uma tendência nos pontos, logo há indícios de dependência dos resíduos. Observando os resíduos do Gráfico 4 conclui-se que não houve dependência entre eles, mostrando um comportamento aleatório. Detectou-se a presença de pontos outliers em quase todos os Estados. Na Tabela 4.3 está o teste de Durbin-Watson que confirma a análise gráfica, os quais mostram que os resíduos são independentes. Apresenta-se na Tabela 4.4 dois testes estatísticos para análise de séries temporais aplicado ao modelo de alisamento exponencial de Holt, são eles o teste Dickey-Fuller e Wilcoxon.

Tabela 4.4: Estimativas do valor-p dos testes Dickey-Fuller e Wilcoxon para o Nordeste e Estados do Nordeste

Estados	P-valores	
	Dickey-Fuller	Wilcoxon
Maranhão	0,5085	0,00006403
Piauí	0,6423	0,00009537
Ceará	0,9387	0,00006403
Rio Grande do Norte	0,6431	0,00006395
Paraíba	0,5971	0,00006403
Pernambuco	0,7265	0,00006395
Alagoas	0,7721	0,00006403
Sergipe	0,9237	0,00009537
Bahia	0,4650	0,00006403
Nordeste	0,6530	0,00006403

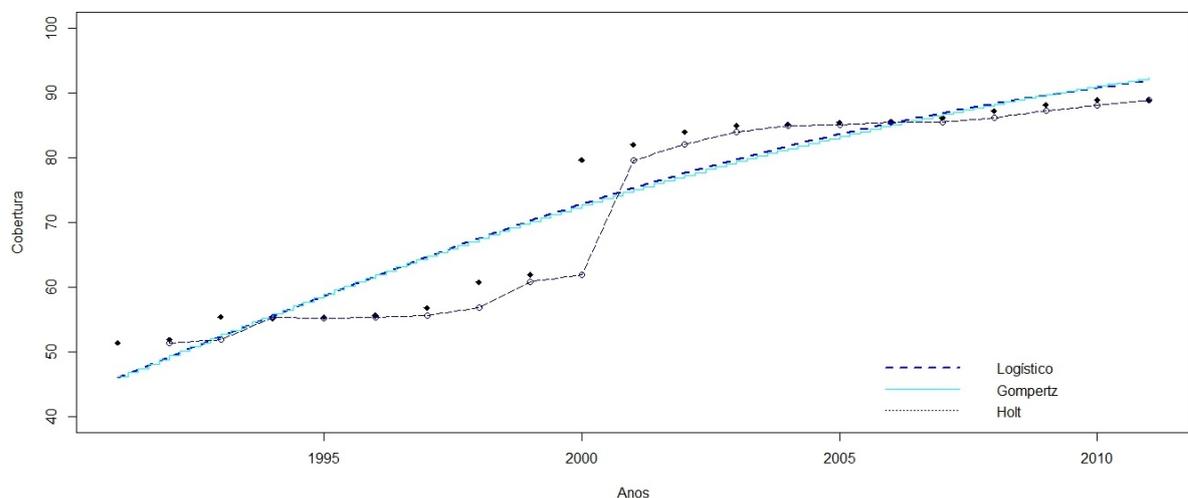
De acordo com as estimativas do p-valor do Teste de Dickey-Fuller as séries temporais abordadas são estacionárias ao longo do tempo com um nível de significância de 5%. O Teste de Wilcoxon afirmou a presença de tendência em todas as séries. Gráficamente notou-se a presença de uma tendência crescente e um comportamento quase estacionário nas séries do Nordeste e Estados do Nordeste.

4.4 Projeções das Coberturas dos Óbitos do Nordeste Brasileiro

Utilizando as técnicas de modelagem em estudo foi possível prever o tempo t em que as coberturas de óbitos para o Nordeste e seus nove Estados iriam atingir a cobertura completa, ou seja, 100%. A Figura 4.11 mostra a série temporal da cobertura de óbito para o Nordeste e as curvas de ajuste para cada modelo adotado. Analisando a adequabilidade dos modelos, é cabível enxergar uma melhor suavização para o modelo de Holt, visto que o Logístico e o de Gompertz não possuem diferenças com relação à plotagem da curva.

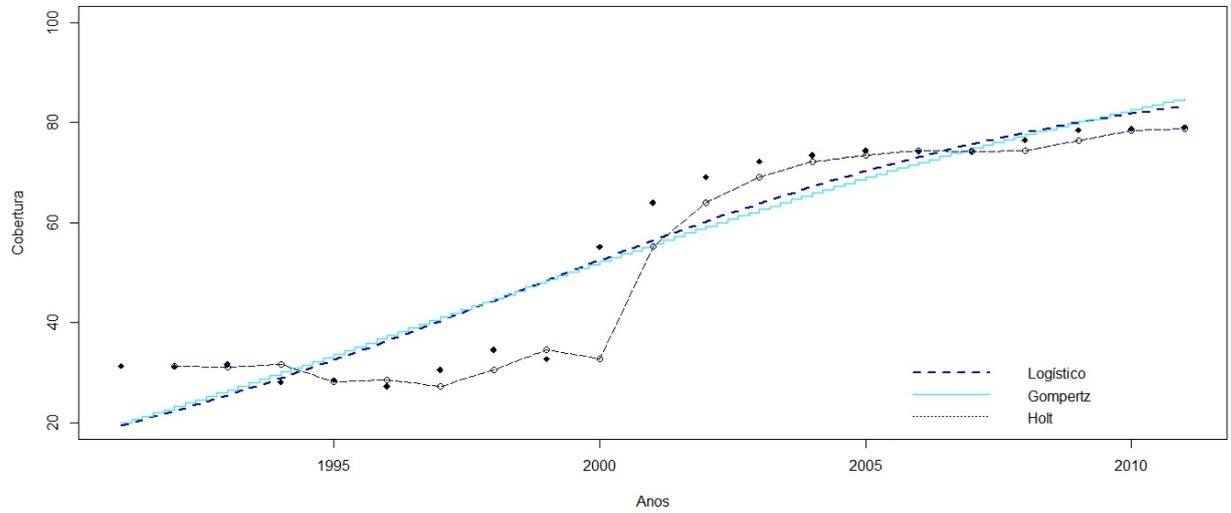
O comportamento das estimativas de coberturas de óbitos fornecidos pelo DATASUS pode ser verificado para cada um dos Estados nas Figuras 4.12, 4.13 e 4.14. Ao detalhar o comportamento das coberturas, verifica-se uma diferença de padrões entre os Estados do RN, PE e SE comparado ao Nordeste, devido a existência de oscilações durante o período de 1991 a 1999.

Figura 4.11: Projeção das coberturas de óbitos estimadas para o Nordeste Brasileiro.

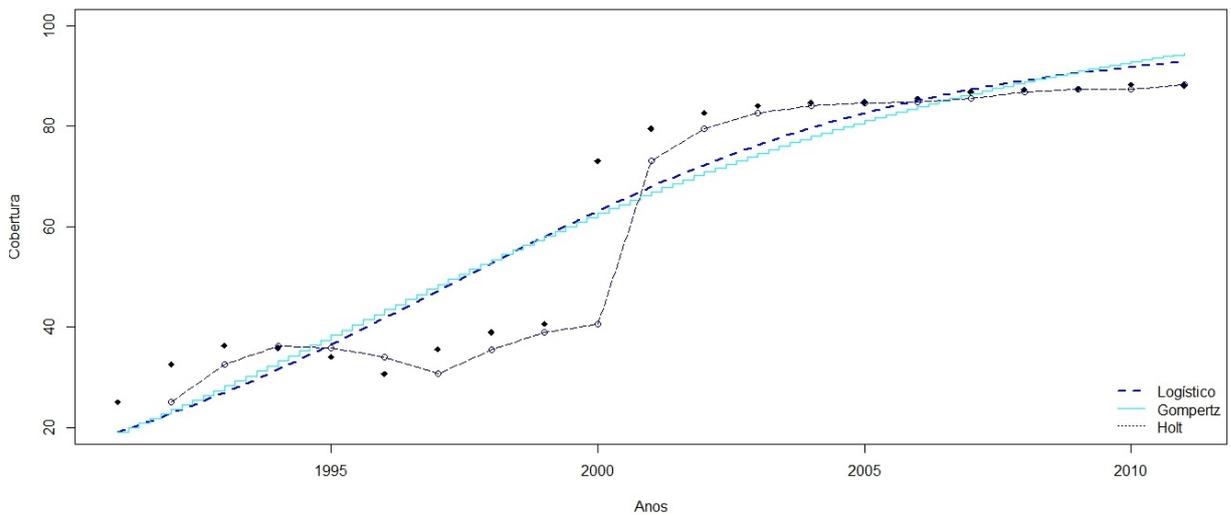


Fonte de dados básicos: Ministério da Saúde – DATASUS, 2011.

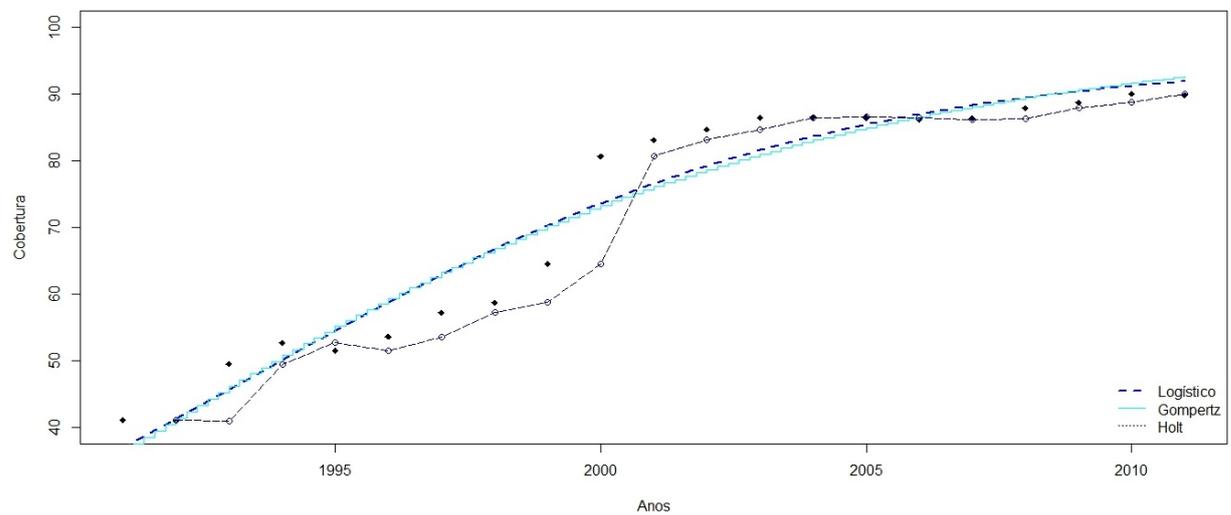
Figura 4.12: Projeção das coberturas de óbitos estimadas para os Estados do Maranhão, Piauí e Ceará.



(a) Maranhão

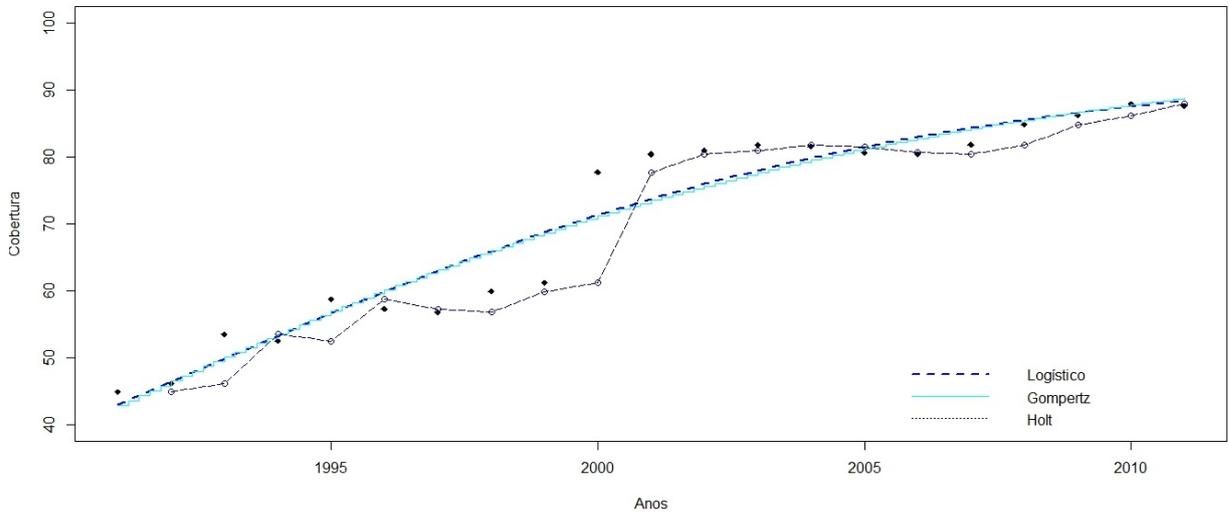


(b) Piauí

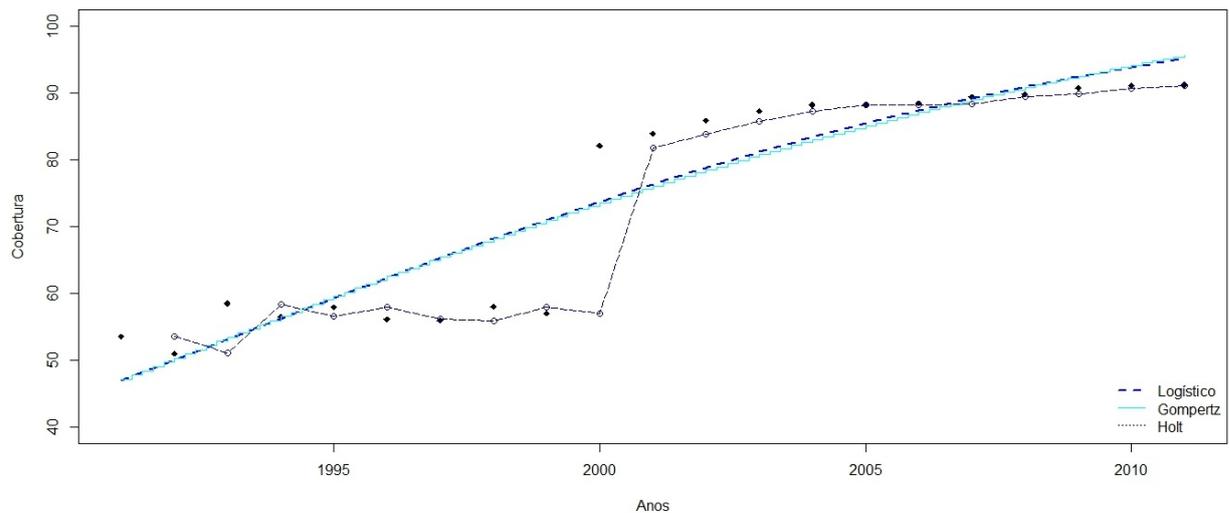


(c) Ceará

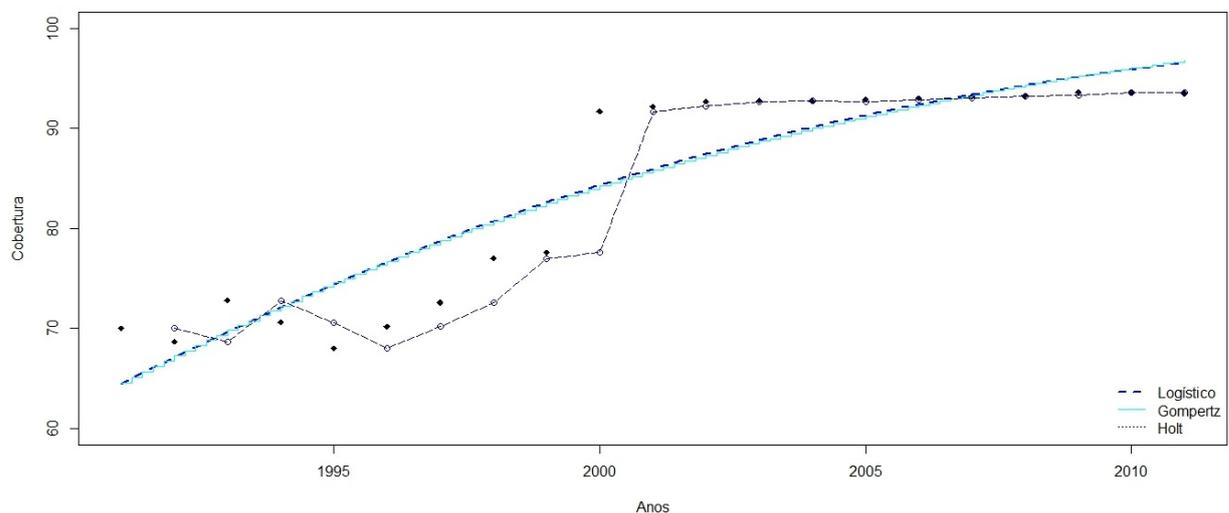
Figura 4.13: Projeção das coberturas de óbitos estimadas para os Estados do Rio Grande do Norte, Paraíba e Pernambuco.



(a) Rio Grande do Norte

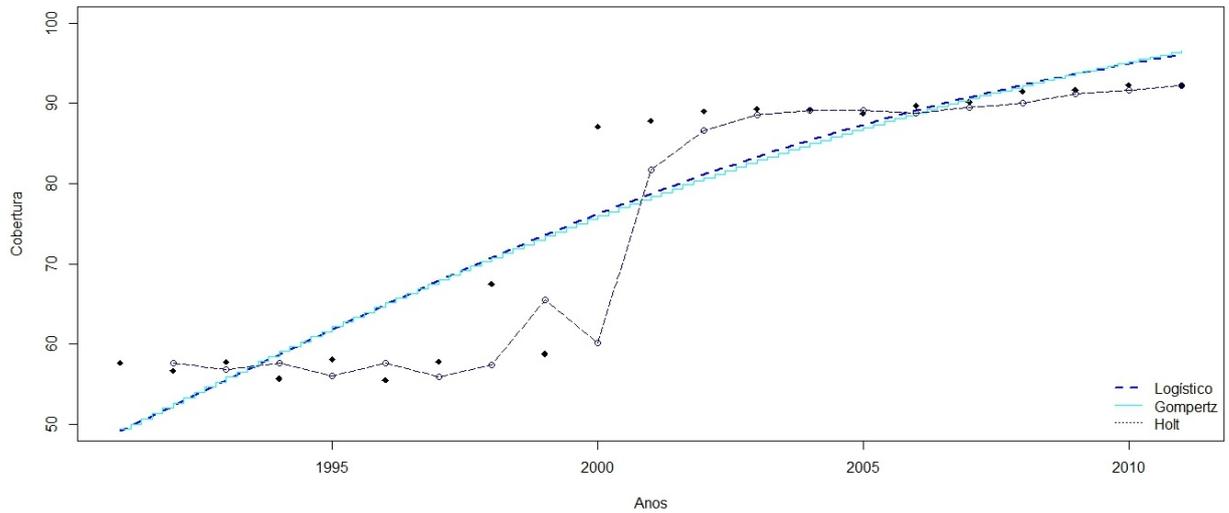


(b) Paraíba

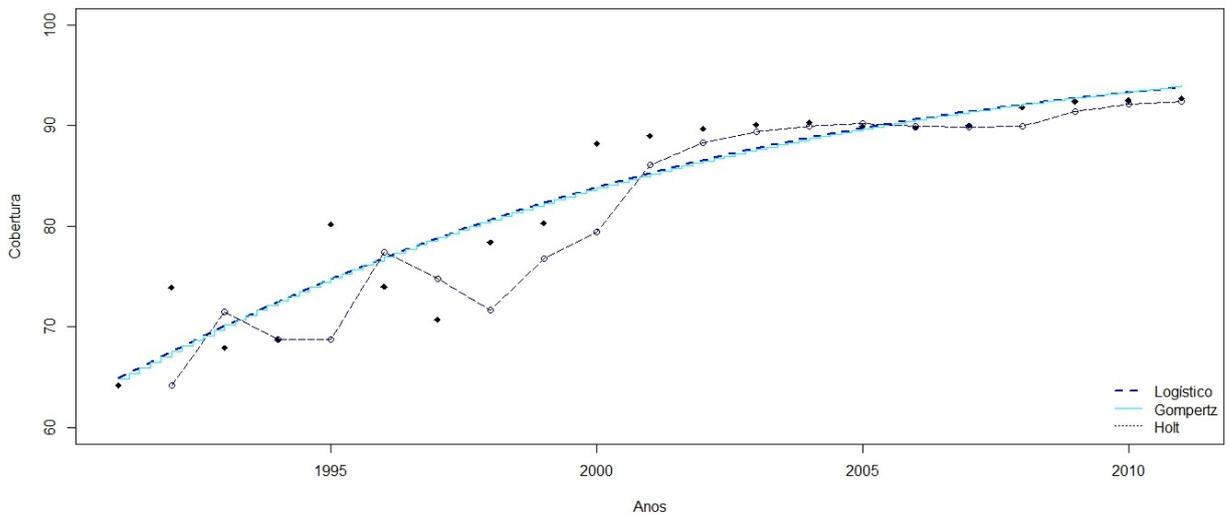


(c) Pernambuco

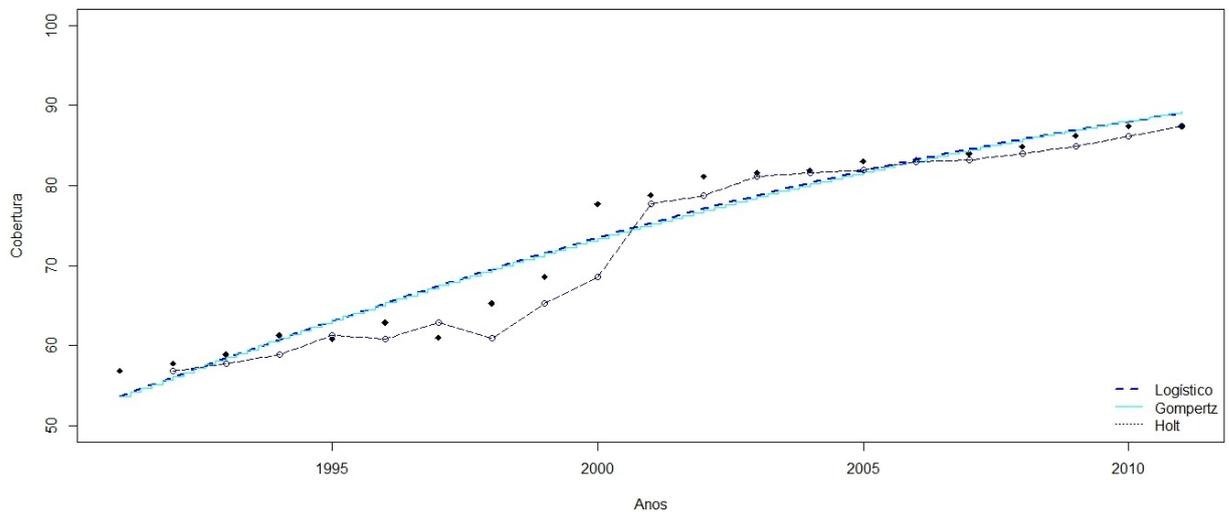
Figura 4.14: Projeção das coberturas de óbitos estimadas para os Estados do Alagoas, Sergipe e Bahia.



(a) Alagoas



(b) Sergipe



(c) Bahia

Essas flutuações continuaram até a chegada do ano 2000, para em seguida mostrarem uma redução mais acentuada, com um comportamento quase constante. Nas figuras estão plotados os valores observados e estimados pelos três métodos: Logístico, Gompertz e Holt. As estimativas das coberturas apresentaram um comportamento de aumento sistemático com uma quebra de ritmo entre 1999 e 2000. Conforme já observado são evidentes dois momentos de tendência, antes e depois desta quebra. No momento anterior o ritmo de aumento foi mais elevado que o segundo para todos os Estados e conseqüentemente o Nordeste, como um todo.

Enquanto o modelo Logístico e o de Gompertz se assemelharam, por sua vez, se distanciaram do modelo de Holt. Enquanto os dois primeiros modelaram razoavelmente o segundo momento, após o ano 2000, não ocorreu o mesmo para o primeiro momento. Já o modelo de Holt acompanhou mais de perto os valores originais da série, antes e depois de 2000. A Tabela 4.5 apresenta as estimativas das Coberturas Completas dos óbitos para o Nordeste e Estados do Nordeste. Onde verifica-se uma diferença de 10 anos entre a cobertura média dos Estados e a região Nordeste quando utilizado o modelo Logístico.

Tabela 4.5: Estimativas das Coberturas Completas de óbitos para o Nordeste e Estados do Nordeste.

Estados	Modelos		
	Logístico	Gompertz	Holt
Maranhão	2030*	2022	2028
Piauí	2032*	2017	2016
Ceará	2024*	2033	2023
Rio Grande do Norte	2025*	2040*	2021
Paraíba	2017	2016	2020
Pernambuco	2020	2018	2023
Alagoas	2019	2017	2018
Sergipe	2026*	2049	2022
Bahia	2030	2028	2023
Nordeste	2035	2021	2020
Média	2026	2027	2021

Fonte de dados básicos: Ministério da Saúde - DATASUS

* Ponto de inflexão da curva abaixo de 100%.

Isto significa maior dispersão das estimativas. Quanto ao modelo de Gompertz essa diferença foi de 7 anos, e pelo modelo de Holt foi inexistente por causa da sua adequabilidade aos valores observados. Evidentemente que estes resultados apresentados na Tabela 4.5 devem ser encarados com cautela, já que é preciso levar em conta os erros inerentes a toda e qualquer previsão. As previsões dos anos pelo modelo de Holt alcançaram o grau de cobertura ótimo com o menor tempo entre os três modelos usados para todos os Estados e o Nordeste, cujas estimativas captaram com mais proximidade as tendências e ciclos apresentados pela série anual.

Comparando os ajustes dos modelos adotados, nota-se uma diferença entre os modelos Logístico e de Gompertz, em que foram encontrados pontos de inflexões diferentes, ou seja, as modelagens tiveram um padrão de crescimento distintos. Observa-se que os Estados do Alagoas e Paraíba seriam os primeiros a atingir o maior valor de cobertura, independente do modelo adotado. Os Estados do Rio Grande do Norte, Sergipe e Bahia seriam os últimos a atingirem uma ótima qualidade de registro das estatísticas de óbitos. Apesar de serem utilizados apenas 21 pontos de informações nas estimativas, os métodos modelaram razoavelmente bem os dados observados, fornecendo valores plausíveis. No entanto, as modelagens apresentaram algumas divergências de comportamento. De acordo com os critérios estabelecidos na seção 3.7 para escolha do melhor modelo, são apresentadas as previsões finais na Tabela 4.6.

Tabela 4.6: Estimativas e Modelos Adotados para o Nordeste e Estados do Nordeste.

Estado	Ano da Previsão	Modelo
Maranhão	2028	Holt
Piauí	2032	Logístico
Ceará	2023	Holt
Rio Grande do Norte	2025	Logístico
Paraíba	2020	Holt
Pernambuco	2023	Holt
Alagoas	2020*	Holt
Sergipe	2022	Holt
Bahia	2023	Holt
Nordeste	2020	Holt

*Adotou-se a previsão do Nordeste, conforme o critério III na seção 3.7

Para efeito de decisão foi adotado o método Logístico para a previsão dos Estados do PI e RN, os quais apresentaram um comportamento bastante sensível à velocidade temporal dos dados. Em virtude desses Estados apresentarem um acelerado aumento da cobertura dos óbitos após o ano 2000, a modelagem de Holt captou mais o segundo momento da série. Como consequência, esses Estados atingiriam mais rapidamente a universalização dos registros de óbitos, o que é pouco provável. Já para os demais Estados do Nordeste foi adotado o modelo de Holt.

4.5 Cobertura e projeções dos Óbitos do espaço geográfico dos Estados do Semiárido brasileiro

Nessa seção é abordada a série temporal da Cobertura dos Óbitos para o espaço geográfico dos Estados do Semiárido Brasileiro, de 1991 a 2011 e suas medidas de precisão. Apresenta-se também as coberturas de óbitos estimada pelo método de Brass para os Estados do Semiárido brasileiro para os anos censitários de 1991, 2000 e 2010. Nota-se na Tabela 4.7 que a cobertura na região saltou de 71% em 1991 para 89% em 2010, sendo que entre 2000 e 2010 ocorreu o maior aumento (10,5%).

Tabela 4.7: Cobertura de óbitos do espaço geográfico dos Estados do Semiárido Brasileiro entre 1991, 2000 e 2010.

Estados	Anos		
	1991	2000	2010
Piauí	56,2	77,1	91,1
Ceará	68,3	78,5	89,2
Rio Grande do Norte	73,6	76,8	92,9
Paraíba	74,3	80,0	92,1
Pernambuco	81,1	87,3	94,2
Alagoas	78,6	79,8	89,2
Sergipe	74,2	80,1	91,3
Bahia	69,0	76,4	84,5
Semiárido	71,0	78,6	89,1

Fonte dos dados básicos: Ministério da Saúde/SIM, 2015; IBGE/ Censo Demográfico 1991, 2000 e 2010.

Em 1991, o PI se destacou bastante entre os estados, onde apresentou estimativas muito inferiores. No entanto, nas décadas seguintes o estado foi o que mais avançou com um salto de 34,8% de 1991 a 2010, assim ficando dentro da estimativa média da região. Em relação ao nível de cobertura, PE foi o estado que se destacou positivamente em todos os anos, obtendo as estimativas mais elevadas. Na Tabela 4.8 apresenta-se as Coberturas dos Óbitos para o espaço geográfico dos Estados do Semiárido brasileiro, de 1991 a 2010.

Na criação das séries temporais para cada estado utilizou-se valores da Cobertura dos Óbitos da RIPSA entre os anos de 1991, 2000 e 2010, para estimar as coberturas de óbitos para o Semiárido, cuja metodologia se encontra detalhada na seção 3.6. Ao observar os dados estimados para cada Estado do Semiárido nota-se uma semelhança com as séries da RIPSA, apresentando uma tendência crescente com o passar dos anos, além de distinguir três períodos de tempo, onde se mostra mudanças repentinas nos valores de cobertura.

Observa-se na Tabela 4.8 que no período de 1991 a 1999 o Estado do PI obteve a menor estimativa de cobertura e PE a maior estimativa entre os estados do Semiárido. Entre os anos 1999 e 2000, as coberturas obtiveram um considerado aumento decorrente da melhoria nos serviços de saúde pública. A partir do ano 2000 a 2010, os dados assumem um comportamento suave em relação a década passada, com uma diferença de 8,8% variando entre 84,5% e 93,3%.

Analisando as coberturas estimadas para o Semiárido, no período de 1991 a 1999, nota-se que na grande maioria das unidades da federação as coberturas estimadas foram as mais baixas da série. A análise para as várias unidades da federação mostra que as maiores discrepâncias na virada do século, foi para o estado do PI. Olhando detalhadamente, nota-se uma variação nas estimativas anuais para cada série dos Estados do Semiárido brasileiro, principalmente entre os anos de 1991 a 2000. Essa variação é reflexo da ineficiência na captação dos dados de registros dos óbitos do Estado como um todo.

Tabela 4.8: Cobertura dos Óbitos para o espaço geográfico dos Estados do Semiárido brasileiro, de 1991 a 2010.

Anos	PI	CE	RN	PB	PE	AL	SE	BA
1991	56,2	68,3	73,7	74,3	81,2	78,6	74,2	69,0
1992	64,1	68,7	73,5	75,5	83,5	78,4	74,8	71,1
1993	65,4	73,1	76,9	78,2	84,9	78,5	76,4	71,5
1994	63,7	74,4	76,0	76,5	84,3	78,9	73,7	72,2
1995	61,3	72,8	76,4	77,3	83,1	77,5	78,3	72,1
1996	60,9	73,8	75,5	76,3	83,9	77,2	72,8	72,1
1997	59,8	75,6	74,9	77,1	83,9	79,2	71,5	71,0
1998	61,3	76,5	76,3	78,1	84,6	82,2	75,7	71,8
1999	67,6	78,3	76,7	74,9	84,7	80,9	74,7	74,3
2000	77,1	78,6	76,8	80,0	87,4	79,9	80,1	76,4
2001	84,5	83,1	84,1	85,3	90,4	84,0	85,0	79,1
2002	84,4	83,9	85,0	86,0	90,7	84,6	85,1	79,5
2003	85,5	84,6	84,4	86,0	90,9	84,4	85,5	80,0
2004	85,3	84,7	85,1	86,6	91,0	84,9	85,5	80,4
2005	85,9	84,2	84,3	85,9	90,7	84,5	86,4	80,1
2006	86,1	84,9	83,9	85,7	90,7	84,6	85,9	81,0
2007	86,3	84,7	84,6	86,3	91,0	85,3	85,7	80,7
2008	87,1	85,7	85,6	87,2	91,3	85,0	86,6	81,5
2009	87,0	86,0	86,2	87,3	91,3	86,0	86,7	81,9
2010	91,1	89,3	92,9	92,2	94,3	89,3	91,3	84,5
Mínimo	56,2	68,3	73,5	74,3	81,2	77,2	71,5	69,0
Máximo	91,1	89,3	92,9	92,2	94,3	89,3	91,3	84,5
Amp 1991 -1999	11,4	10,0	3,4	3,9	3,7	5,0	6,8	5,3
Amp 1999 - 2000	9,5	0,3	0,1	5,1	2,7	-1,0	5,4	2,1
Amp 2000 - 2010	14,0	10,7	16,1	12,2	6,9	9,4	11,2	8,1

No entanto, para prever os anos de cobertura ótima adotou a mesma medida proposta na seção 4.4, onde adotou-se os tipos de modelagens preditivas para cada Estado, devido a complexidade desse estudo. Na Tabela 4.9 estão apresentadas as estimativas dos anos de cobertura ótimas para os Estados do Semiárido. A previsão do ano para o estado do PE alcançou o grau de cobertura ótimo com o menor tempo entre os Estados do Semiárido.

Observa-se que o modelo de alisamento exponencial de Holt se destacou melhor para todos os Estados do Semiárido, mas em contrapartida o modelo Logístico apresentou estimativas mais plausíveis, baseadas pelos critérios de seleção estabelecidos na seção 3.7. Na Tabela 4.10 estão apresentadas as medidas de precisão dos modelos Logístico e de Holt para os Estados do Semiárido brasileiro.

Tabela 4.9: Previsão para as estimativas das coberturas completas de órbitos para os Estados do Semiárido brasileiro.

Estado	Anos	
	Logístico	Holt
Piauí	2035	2019
Ceará	2030	2023
Rio Grande do Norte	2036	2020
Paraíba	2028	2018
Pernambuco	2025	2017
Alagoas	2027	2028
Sergipe	2033	2026
Bahia	2034	2021

Observando o EAM percebe-se que o modelo de Holt forneceu desvios menores comparados ao Logístico, para todas as previsões dos Estados. Ao analisar o EPM percebe-se que o modelo Logístico obteve os menores valores para os Estados (CE e PE), cuja variação das estimativas ficaram entre -1 a 1,4. O modelo de Holt teve o melhor desempenho com baixos valores de EPM, variando entre -0,4 a 0,8. Analisando o EQM percebe-se que o modelo de Holt teve o melhor desempenho para todos os Estados exceto PB e SE, que apresentaram as menores estimativas pelo ajuste do modelo Logístico. No entanto, de acordo com os critérios estabelecidos na seção 3.7 para escolha do melhor modelo, então são apresentados os resultados na Tabela 4.11 para o Semiárido brasileiro.

Tabela 4.10: Medidas de precisão para as estimativas das coberturas completas de órbitos para os Estados do Semiárido brasileiro.

Estado	EAM		EPM		EQM	
	Logístico	Holt	Logístico	Holt	Logístico	Holt
Piauí	3,93	2,13	-0,44	-0,25	10,84	9,05
Ceará	1,21	1,05	-0,20	0,47	2,75	2,48
Rio Grande do Norte	1,79	1,34	1,27	0,77	7,83	5,65
Paraíba	2,56	1,57	1,31	-0,36	2,09	4,50
Pernambuco	1,18	0,92	-0,21	-0,24	5,79	3,67
Alagoas	1,59	1,04	-0,94	0,49	4,57	2,38
Sergipe	2,65	1,88	-1,00	0,29	3,26	6,84
Bahia	1,06	0,80	-0,30	0,10	4,76	1,22

Para efeito de decisão será adotado aqui o método Logístico para a previsão dos Estados do PI, PB e PE os quais apresentam comportamento bastante sensível a velocidade temporal dos dados. Já para os demais Estados do Semiárido brasileiro será adotado o modelo de Holt. No entanto, entre os modelos selecionados o de Holt parece ser o mais indicado para prever o ano de completitude das coberturas dos órbitos, apresentando as menores estimativas. Nota-se que as previsões para os Estados do Semiárido ficaram acima das previsões para os Estados como um todo, conforme o suposto inicial.

Tabela 4.11: Estimativas e Modelos Adotados para os espaços geográficos dos Estados do Semiárido brasileiro.

Estado	Anos	Modelo
Piauí	2035	Logístico
Ceará	2025	Holt
Rio Grande do Norte	2020*	Holt
Paraíba	2028	Logístico
Pernambuco	2025	Logístico
Alagoas	2028	Holt
Sergipe	2026	Holt
Bahia	2025	Holt

*Adotou-se a previsão do Nordeste, conforme o critério III na seção 3.7

Capítulo 5

Conclusão

Estudos relacionados a métodos de projeção da cobertura dos óbito são importantes ferramentas de monitoramento da qualidade dos dados, principalmente para a região do Nordeste brasileiro, onde estudos ligados a esta natureza são escassos e há uma necessidade crescente de fortalecer o debate sobre os níveis de completitude dos registros de óbitos. Com isso, foi realizado um estudo para os Estados do Nordeste entre os anos de 1991 a 2011 buscando descrever através de modelos matemáticos a universalização da cobertura de óbito.

O emprego dessas modelagens possibilitou investigar e identificar a existência de padrões e níveis entre a cobertura dos óbito dos Estados, verificando através de estimativas qual modelo melhor se adequou aos dados observados. Entre as modelagens, Holt apresentou melhor adaptação ao padrão dos dados das séries para alguns Estados, descrevendo as previsões em menores tempos e erros de estimação para o estudo. E para os demais Estados o modelo Logístico se adequou melhor.

As estimativas mostram que os Estados do PI, AL e SE seriam os primeiros a atingirem a universalização, quanto a qualidade dos dados. Como observa-se em todo estudo, este também não está livre de algumas limitações. Logo, para se verificar a adequabilidade dos modelos, faz se necessário o cumprimento de certos pressupostos, um deles se refere a quantidade de informação da série temporal que nesse trabalho consiste em apenas 21 pontos, impedindo assim o pleno desenvolvimento dos modelos.

Com relação aos modelos Logístico e Gompertz, eles são matematicamente iguais onde na literatura são usados como modelos diferentes para se comparar suas capacidades preditivas. Portanto, ambos foram utilizados aqui como exercício para prever a cobertura

de óbito. Em seguida, após identificar o melhor método, elaborou-se a série temporal para o espaço geográfico dos Estados do Semiárido Brasileiro entre 1991 a 2010, tendo como base a cobertura de óbitos estimada pelo método de Brass entre os anos censitários de 1991, 2000 e 2010. As previsões mostraram que os Estados do CE, RN, PE e BA seriam os primeiros a atingirem um grau de completitude elevado, em menor quantidade temporal.

Os resultados encontrados neste trabalho buscaram através de modelos matemáticos prever o ano de cobertura ótimo para o Nordeste, e elaborar uma série para o espaço geográfico dos Estados do Semiárido. As séries das coberturas de óbitos para os Estados do Semiárido possuem estimativas mais próximas da realidade, quanto aos valores de 1991 a 2000, pois o aumento não foi tão brusco como na séries dos Estados como um todo.

Entre os modelos selecionados (Logístico, Gompertz e Holt), esse último é sugerido como o mais indicado para prever o ano de completitude das coberturas dos óbitos, visto que sua curva de ajuste se adequou melhor aos dados e também apresentou as menores medidas de precisão para a maioria dos Estados.

Algumas extensões desse trabalho são propostas à medida que os dados são disponibilizados pelos órgãos produtores dessas informações, sugere-se proceder a atualização do trabalho usando a mesma metodologia, ou alguma adaptação dos procedimentos adotados aqui. Usar outros tipos de modelagens para medir a qualidade dos dados de óbitos.

É preciso reconhecer, no entanto, que como qualquer cenário que se trace, esse aqui reflete uma possibilidade considerada plausível e que somente o futuro poderá confirmar ou não estes cenários. De todo os modos, trata-se esse estudo de uma contribuição esperada para auxiliar no planejamento e avaliação dos gestores com o desempenho das estatísticas vitais, mais especificamente dos óbitos.

Referências Bibliográficas

- [1] ALBANO, G.; GIORNO, V. A. *stochastic model in tumor growth*. J. Theore. Biol., London, v.242, n.2, p.329-336, 2006.
- [2] AUGUSTEYN, R. C. *Growth of the human eye lens*. Mol. Vision, Clifton, n.13, p.252-257, 2007.
- [3] BRASS, W. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Occasional Publication. International Program of Laboratories for Population Statistics, Chapel Hill, 1975.
- [4] BEZERRA, J. *Ecologia de Populações: A curva logística e crescimento populacional*. NEPAM -UNICAMP, p. 1-18, 2008.
- [5] COURBAJE, Y.; FARGUES P. *A method for deriving mortality estimates from incomplete vital statistics*. Popul Stud. 1979;33(1):165-80.
- [6] DATASUS - Departamento de Informática do SUS *Disponível em:* < <http://www2.datasus.gov.br/> >. Acesso em 15 de abril de 2015.
- [7] EHLERS, R. S. *Análise de séries temporais*. [s.l: s.n.]. v. 5
- [8] GOMPERTZ, B. *On the nature of the function expressive of the law of human mortality, and on a new method of determining the value of life contingencies*. Philos. Trans. R. Soc. London, London, v.115, n.1825, p.513-585, 1825.
- [9] HELENA, M.; MELLO, P. D. E.; LAURENTI, R. *Avaliação dos sistemas de informação em saúde no brasil*. v. 18, n. 1, p. 7-18, 2010.
- [10] HOFFMAN, R.; VIEIRA, S. *Análise de regressão: uma introdução à econometria*. São Paulo, HUNITEC, 379p, 1998.

- [11] IBGE - Sistema IBGE de Comunicação Social - Indicadores sociodemográficos e de saúde no Brasil 2009.2012 *Disponível em:* < <http://tabnet.datasus.gov.br/cgi/idb2012/matriz.htm> >. Acesso em 15 de abril de 2015.
- [12] JORGE, M. H. P. D. M.; GOTLIEB, S. L. D. *O Sistema de Informação de Atenção Básica como fonte de dados para os Sistemas de Informações sobre Mortalidade e sobre Nascidos Vivos*. Informe Epidemiológico do Sus, v. 10, n. 1, p. 7–18, 2001.
- [13] JORGE, M. H. P. M.; GOTLIEB, S. L. D.; LAURENTI, R. *O sistema de informações sobre mortalidade: problemas e propostas para o seu enfrentamento II - Mortes por causas externas*. Revista Brasileira de Epidemiologia. 2002;5(2):212-23.
- [14] JORGE, M. H. P. D. M.; LAURENTI, R.; GOTLIEB, S. L. D. *Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC*. Ciência & Saúde Coletiva, v. 12, n. 3, p. 643–654, 2007.
- [15] KEYFITZ, N. *The limits of populations forecasting* Population and Development Review 8(4): 579-593, 1981.
- [16] LO, C. F. *Stochastic Gompertz model of tumour cell growth*. J. Theor. Biol., London, v.248, n.2, p.317-321, 2007.
- [17] LOPES, R. D. *Previsão De Autopeças: Estudo De Caso Em Uma Concessionária De Veículos*. [Dissertação de Mestrado], Florianópolis, 2002.
- [18] MARINHO, P. R. D.; PAES, N. A. *A qualidade dos eventos vitais do Nordeste, relacionamentos com as condições de vida e perspectivas* [Relatório de Pesquisa], João Pessoa-PB, 2010.
- [19] MARTINE, G.; LIMA, R. *Projeções Populacionais: usos, abusos e não-usos*. Iplan, Ipea, Brasília, 1989.
- [20] MORETTIN, P. A.; TOLOI, C. M.C. *Análise de séries temporais*. 2 ed. São Paulo: Blucher, 2006. 564p.
- [21] PAES, N. A; ALBUQUERQUE, M.E.E. *Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras*. Revista de Saúde Pública.1999; 33(1):33-43.

- [22] PAES, N. A. *Avaliação da cobertura dos registros de óbitos dos Estados brasileiros em 2000*. Revista de Saúde Pública, v. 39, n. 6, p. 882–890, 2005.
- [23] PAES, N. A. *Qualidade das estatísticas de óbitos por causas desconhecidas dos Estados brasileiros*. Revista de Saúde Pública, v. 41, n. 3, p. 436–445, 2007.
- [24] PAES, N. A.; REGIS, M.; SILVA, M. *Cenários regionais prospectivos das estatísticas de nascimentos do Nordeste brasileiro.*, Associação Brasileira de Estudos Populacionais, Caxambu-MG, 2010.
- [25] PRESTON, S.; COALE, A.J.; TRUSSELL, J.; WEINSTEIN, M. *Estimating the completeness of reporting of adult deaths in populations that are approximately stable*. Popul Index 1980;46(2):179-202.
- [26] POLIGNANO, M. V. *História das políticas de saúde no Brasil - uma pequena revisão*. Chemistry & . . . , p. 1–35, 2004.
- [27] PROVENZA, M. M.; FABIANO, J.; SILVA, L. D. C. *Análise e Previsão de Séries Temporais do Homicídio Doloso no Rio De Janeiro*. Revista Cadernos de Estudos Sociais e Políticos, v. 4, n. 7, p. 63–85, 2015.
- [28] PORTAL ASA Brasil. *Disponível em: < <http://www.asabrasil.org.br/portal/Default.asp>>*. Acesso em 15 de abril de 2015.
- [29] QUEIROZ, B. L. *Estimativas do Grau de Cobertura e da Esperança de Vida para as Unidades da Federação no Brasil entre 2000 e 2010*. p. 1–20, 2012.
- [30] SEADE, F. *População e reestruturação produtiva de novos elementos para projeções demográficas*. p. 3–8, 1995.
- [31] SILVA, D. A. M.; OLIVEIRA, P. A. *Comparação entre métodos de previsão de demanda aplicados a fios cirúrgicos utilizados em um hospital escola do município de Botucatu*. Revista Tekhne e Logos, Botucatu, SP, Vol. 3, n. 3, Novembro, 2012.
- [32] SOUZA, V. J. DE. *Curvas de Crescimento de Gompertz para o Acompanhamento de Crianças de Alto Risco*. p. 39–58, 2010.

- [33] VIRIATO, C. *Mortalidade por causas externas no município de Teresina (PI) 1971 a 1990*. Campinas, São Paulo, Universidade Estadual de Campinas, Faculdade de Ciências Médica, 1996.

Apêndice A

Rotina no R - Modelo Logístico

```
#-----#  
# Rotina no R utilizada para a obtenção das estimativas do Modelo Logístico #  
#-----#  
  
rm(list=ls(all=TRUE))  
  
DADOS<-read.table("C:/Users/ALISSON/Desktop/monografia/script_r/serie_ripsa.txt", header=T)  
  
#medidas descritivas  
  
descritiva<-summary(DADOS)  
  
y<-DADOS$NE # Nordeste, y=($MA,$PI,$CE,$RN,$PB,$PE,$AL,$SE,$BA)  
  
x<-DADOS$Anos  
  
#----- Modelo Logístico -----#  
  
alpha0 <- max(y) + 1  
z0 <- log((alpha0/y) - 1)  
reg <- lm(z0~x)  
beta0 <- coef(reg)[1]  
gamma0 <- -coef(reg)[2]  
  
log.der <- deriv3(~ a/(1+exp(b - c*x)), c("a","b","c"), function(x,a,b,c){})
```

```

reg.nls <- nls(y ~ log.der(x,a,b,c), start=list(a=alpha0, b=beta0, c=gamma0))

ajuste1 <- summary(reg.nls)

theta.log <- ajuste1$coefficients[,1]

sd <- ajuste1$parameters[,2]

## Diagnósticos

n <- length(y)

p <- length(theta.log)

sigma <- ajuste1$sigma

sigma2 <- (sigma)^2

der.log <- log.der(x,theta.log[1],theta.log[2],theta.log[3])

# procedimentos para obter - testes estatísticos na análise de resíduos

residuosnI1<-ajuste1$resid

shapiro.test(residuosnI1) # teste normalidade

r.cru1 <- residuals(ajuste1)

var(r.cru1)

r.pad1 <- residuals(ajuste1, type="pearson")

var(r.pad1)

grad.log <- attr(der.log,"gradient") # gradiente

C1 <- solve(t(grad.log)%*%grad.log)

sd.theta <- sqrt(diag(C1)*sigma2)

H.log <- attr(der.log, "hessian") # hessiana

FtF <- t(grad.log)%*%grad.log

#extraíndo a matriz gradiente avaliada nas estimativas dos parâmetros

m1c <- lm(y~-1+grad.log)

require(lmtest)

require(car)

gqtest(m1c) # teste para homogeneidade de variâncias

```

```
durbinWatsonTest(m1c)# teste para independência dos resíduos

# Predição
for(i in 2011:2100){
  newph<- data.frame(x=i)
  print(c(trunc(i+1,digits=0),predict(reg.nls,newdata=newph)))
}

# valores preditos para os valores observados
fitd1<-c(der.log[1],der.log[2],der.log[3],der.log[4],der.log[5],der.log[6],der.log[7],der.log[8],
der.log[9],der.log[10],der.log[11],der.log[12],der.log[13],der.log[14],der.log[15],der.log[16],
der.log[17],der.log[18],der.log[19],der.log[20],der.log[21])
valores1<-data.frame(y,fitd1)

# erro dos valores estimados
Epm_1=((sum((y-fitd1)/y))/21)*100
EQM_1=sum((y-fitd1)^2)/21
Eam_1= (sum(abs(y-fitd1)))/21
```

Apêndice B

Rotina no R - Modelo de Gompertz

```
#-----#  
# Rotina no R utilizada para a obtenção das estimativas do Modelo de Gompertz#  
#-----#  
  
rm(list=ls(all=TRUE))  
  
DADOS<-read.table("C:/Users/ALISSON/Desktop/monografia/script_r/serie_ripsa.txt", header=T)  
  
#medidas descritivas  
  
descritiva<-summary(DADOS)  
  
y<-DADOS$NE # Nordeste, y=($MA,$PI,$CE,$RN,$PB,$PE,$AL,$SE,$BA)  
  
x<-DADOS$Anos  
  
#----- Modelo de Gompertz -----#  
  
alpha1 <- max(y) + 1  
  
z1 <- log(-log(y/alpha0))  
  
fit <- lm(z0~x)  
  
beta1 <- coef(fit)[1]  
  
gamma1 <- - coef(fit)[2]  
  
gomp.der <- deriv3(~ a*exp(-exp(b - c*x)), c("a","b","c"), function(x,a,b,c){})
```

```

fit.nls <- nls(y ~ gomp.der(x,a,b,c), start=list(a=alpha1, b=beta1, c=gamma1))

ajuste2 <- summary(fit.nls)

theta.g <- ajuste2$coefficients[,1]

sd <- ajuste2$parameters[,2]

## Diagnósticos

n <- length(y)

p <- length(theta.g)

sigma <- ajuste2$sigma

sigma2 <- (sigma)^2

der.gomp <- gomp.der(x,theta.g[1],theta.g[2],theta.g[3])

# procedimentos para obter - testes estatísticos na análise de resíduos

residuosnI2<-ajuste2$resid

residuosnI2

shapiro.test(residuosnI2) # teste normalidade

r.cru2 <- residuals(ajuste2)

var(r.cru2)

r.pad2 <- residuals(ajuste2, type="pearson")

var(r.pad2)

grad.gomp <- attr(der.gomp,"gradient") # gradiente

C2<- solve(t(grad.gomp)%*%grad.gomp)

sd.theta <- sqrt(diag(C2)*sigma2)

#extraíndo a matriz gradiente avaliada nas estimativas dos parâmetros

m2c <- lm(y~-1+grad.gomp)

require(lmtest)

require(car)

gqtest(m2c) # teste para homogeneidade de variâncias

```

```
durbinWatsonTest(m2c)# teste para independência dos resíduos

# Predição
for(i in 2011:2100){
  newph<- data.frame(x=i)
  print(c(trunc(i+1,digits=0),predict(fit.nls,newdata=newph)))
}

# valores preditos para os valores observados
fitd2<-c(der.gomp[1],der.gomp[2],der.gomp[3],der.gomp[4],der.gomp[5],der.gomp[6],der.gomp[7],der.gomp[8],
der.gomp[9],der.gomp[10],der.gomp[11],der.gomp[12],der.gomp[13],der.gomp[14],der.gomp[15],der.gomp[16],
der.gomp[17],der.gomp[18],der.gomp[19],der.gomp[20],der.gomp[21])
valores2<-data.frame(y,fitd2)

# erro dos valores estimados
Epm_2=((sum((y-fitd2)/y))/21)*100
EQM_2=sum((y-fitd2)^2)/21
Eam_2= (sum(abs(y-fitd2)))/21
```

Apêndice C

Rotina no R - Algoritmo de Holt

```
#-----#  
# Rotina no R utilizada para a obtenção das estimativas do Algoritmo de Holt #  
#-----#  
  
rm(list=ls(all=TRUE))  
  
DADOS<-read.table("C:/Users/ALISSON/Desktop/monografia/script_r/serie_ripsa.txt", header=T)  
  
#medidas descritivas  
  
descritiva<-summary(DADOS)  
  
y<-DADOS$NE # Nordeste, y=($MA,$PI,$CE,$RN,$PB,$PE,$AL,$SE,$BA)  
  
x<-DADOS$Anos  
  
#----- Algoritmo de Holt -----#  
  
# Pacotes  
  
library(forecast)  
  
library(tseries)  
  
library(BSDA)  
  
serie=ts(y,start=c(1991), frequency = 1)
```

```

# Pressupostos

ndiffs(serie) # Teste paramétrico das diferenças

adf.test(serie) # Teste de Dickey-Fuller (estacionariedade)

SIGN.test(serie) #teste do sinal(tendência)

# modelo do algoritmo exponencial simples de HOLT

m=HoltWinters(serie, gamma=FALSE)

# coeficientes estimados do modelo

alpha=m$alpha # nivel

beta=m$beta # tendencia

# Erro dos valores estimados

fitd3=m$fitted[,1]

y<-y[3:21]

Epm_3=((sum((y-fitd3)/y))/21)*100

EQM_3=sum((y-fitd3)^2)/21

Eam_3= (sum(abs(y-fitd3)))/21

# previsão

p<- predict(m, 24, prediction.interval = TRUE)

#----- Gráficos (dados com curvas ajustadas para os Modelos -----#

plot(y~x,pch=18, ylab="Cobertura", xlab="Anos", ylim=c(40,100)) #NORDESTE, CE, RN, PB

plot(y~x,pch=18, ylab="Cobertura", xlab="Anos", , ylim=c(20,100)) #MA, PI

plot(y~x,pch=18, ylab="Cobertura", xlab="Anos", ylim=c(60,100)) #PE, SE

plot(y~x,pch=18, ylab="Cobertura", xlab="Anos", ylim=c(50,100)) #AL ,BA

```

```
with(list(theta=theta.log), curve(theta[1]/(1 + exp(theta[2] - (theta[3]*x))),
col="blue",add=TRUE,lty=2,type="l",lwd=2))
with(list(theta=theta.g), curve(theta[1]*exp(-exp(theta[2] - theta[3]*x)),
col="cyan",add=TRUE,lty=1,type="s",lwd=1))
with(list(theta=alpha),lines(fitted(m)[,1],lty=5, type= "o",lwd=1.5,col="blue4"))
legend(x="bottomright",c("Logístico", "Gompertz","Holt"), col = c("blue","cyan","blue4"), lty = c(2, 1, 3), lwd=c(2,
```

Apêndice D

Rotina no R - Cobertura dos Estados do Semiárido

```
#-----#
# Rotina no R utilizada para a obtenção das estimativas da Cobertura dos Estados do Semiárido #
#-----#

rm(list=ls(all=TRUE))

DADOS<-read.table("C:/Users/ALISSON/Desktop/monografia/serie_semiarido/serie_obitos_obs_semiarido.txt", header=T)

Cobertura<-read.table("C:/Users/ALISSON/Desktop/monografia/serie_semiarido/serie_cob_obs_est.txt", header=T)

# Agrupando os óbitos de 1991,2000 e 2010

Obito_geral<-rbind(DADOS[1,-1],DADOS[10,-1],DADOS[20,-1])

# Fator de Correção dos Óbitos

fc_geral<-(1/Cobertura[,-1])*100

# Corrigindo os Óbitos Observados com o Fator de Correção

Obito_cor<-ceiling(Obito_geral*fc_geral)

# Diferença entre os Óbitos Corrigidos e os Óbitos Observados
```

```
Dif_geral<-Obito_cor-Obito_geral

# média das diferenças dos óbitos

Media_dif1<-floor(colMeans(Dif_geral[1:2,1:8])) #1992 a 1999

Media_dif2<-floor(colMeans(Dif_geral[2:3,1:8])) #2001 a 2011

# Calculando os Óbitos Esperados, somando os obitos_obs com as medias_dif

Obito_esp1<-DADOS[1,-1]+Media_dif1
Obito_esp2<-DADOS[2,-1]+Media_dif1
Obito_esp3<-DADOS[3,-1]+Media_dif1
Obito_esp4<-DADOS[4,-1]+Media_dif1
Obito_esp5<-DADOS[5,-1]+Media_dif1
Obito_esp6<-DADOS[6,-1]+Media_dif1
Obito_esp7<-DADOS[7,-1]+Media_dif1
Obito_esp8<-DADOS[8,-1]+Media_dif1
Obito_esp9<-DADOS[9,-1]+Media_dif1

Obito_esp10<-DADOS[10,-1]+Media_dif2
Obito_esp11<-DADOS[11,-1]+Media_dif2
Obito_esp12<-DADOS[12,-1]+Media_dif2
Obito_esp13<-DADOS[13,-1]+Media_dif2
Obito_esp14<-DADOS[14,-1]+Media_dif2
Obito_esp15<-DADOS[15,-1]+Media_dif2
Obito_esp16<-DADOS[16,-1]+Media_dif2
Obito_esp17<-DADOS[17,-1]+Media_dif2
Obito_esp18<-DADOS[18,-1]+Media_dif2
Obito_esp19<-DADOS[19,-1]+Media_dif2
Obito_esp20<-DADOS[20,-1]+Media_dif2
Obito_esp21<-DADOS[21,-1]+Media_dif2

Obito_esp1<-rbind(Obito_esp2,Obito_esp3,Obito_esp4,Obito_esp5,
```

```
Obito_esp6,Obito_esp7,Obito_esp8,Obito_esp9) # De 1992 a 1999

Obito_esp2<-rbind(Obito_esp11,Obito_esp12,Obito_esp13,Obito_esp14,Obito_esp15,Obito_esp16,Obito_esp17,
Obito_esp18,Obito_esp19) # De 2001 a 2009 e 2011

# Cobertura para os Estados do Semiárido

Cob_semi1<-round(((DADOS[2:9,-1]/Obito_esp1)*100),digits=1)

Cob_semi2<-round(((DADOS[11:19,-1]/Obito_esp2)*100),digits=1)

Cob_semi<-rbind(Cobertura[1,-1],Cob_semi1,Cobertura[2,-1],Cob_semi2,Cobertura[3,-1])
```