

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**MODELAÇÃO DO ATRASO NA NOTIFICAÇÃO DOS  
ÓBITOS NO SISTEMA DE VIGILÂNCIA DIÁRIA DA  
MORTALIDADE**

**Liliana do Carmo Cerdeira Antunes**

TRABALHO DE PROJETO

MESTRADO EM BIOESTATÍSTICA

**2014**

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



## **MODELAÇÃO DO ATRASO NA NOTIFICAÇÃO DOS ÓBITOS NO SISTEMA DE VIGILÂNCIA DIÁRIA DA MORTALIDADE**

**Liliana do Carmo Cerdeira Antunes**

TRABALHO DE PROJETO  
MESTRADO EM BIOESTATÍSTICA

Trabalho de Projeto orientado pela Professora Doutora Marília Antunes e pelo Doutor Baltazar Nunes (INSA) para obtenção do grau de Mestre em Bioestatística

**2014**

## **Agradecimentos**

É com enorme satisfação que expresso o meu mais profundo agradecimento a todos aqueles que tornaram a realização deste trabalho possível.

Gostaria antes de mais de agradecer ao Doutor Baltazar Nunes, coorientador deste trabalho, acima de tudo por acreditar em mim, por todo o apoio, incentivo e disponibilidade e por permitir que me dedicasse a 100% a este trabalho.

À Professora Doutora Marília Antunes, orientadora deste trabalho, pelo apoio, paciência e disponibilidade desmonstrada.

A todos os meus colegas no Departamento de Epidemiologia do Instituto Nacional de Saúde Doutor Ricardo Jorge por todo o apoio, motivação, carinho e bons momentos partilhados.

À minha família, que sempre acreditaram em mim e que tornaram possível todo o caminho percorrido.

Uma última palavra aos meus amigos, que me motivaram nos momentos mais complicados e me fizeram sorrir.

## Resumo

O sistema de vigilância diária da mortalidade (VDM), que monitoriza a mortalidade por “todas as causas” em Portugal, está implementado no Departamento de Epidemiologia (DEP) do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA) e foi desenvolvido em parceria com o Instituto dos Registos e do Notariado (IRN) e com o Instituto de Gestão Financeira e Equipamentos da Justiça (IGFEJ).

O sistema VDM tem como função a monitorização contínua e sistemática da mortalidade observada em Portugal, com o objetivo de detetar e estimar a magnitude do impacto de epidemias, novas doenças, condições climáticas extremas ou outros eventos relacionados com a saúde. Este sistema é visto como uma ferramenta estratégica para a ação em saúde pública. O tempo entre a ocorrência do óbito e a chegada dessa informação ao sistema é crucial para a deteção atempada destes impactos na medida em que, quanto maior é o atraso na notificação dos óbitos ao sistema VDM, maior será o tempo até à deteção destes impactos.

O principal objetivo deste projeto foi o desenvolvimento de uma metodologia que permita estimar diariamente o número de óbitos ocorridos mas ainda não notificados ao VDM, através da modelação do atraso na notificação dos mesmos e, tendo em vista, a diminuição do tempo até à deteção de impactos na mortalidade através da metodologia proposta.

A informação disponível contabiliza um total de 593197 óbitos, correspondendo à informação rececionada de Maio de 2008 a 20 de Janeiro de 2014. A metodologia que se propõe para a concretização dos objetivos assenta na utilização do estimador pontual:

$$\hat{M}_{t,i} = \frac{m_{t,i}}{\hat{p}_i}, \quad i = 1, \dots, 14,$$

onde  $\hat{p}_i$  é a probabilidade estimada de um óbito ser notificado com um atraso igual ou inferior a  $i$  dias,  $m_{t,i}$  o número de óbitos ocorridos a dia  $t$  e notificados com atraso igual ou inferior a  $i$  dias e  $M_{t,i}$  o número total de óbitos ocorridos a dia  $t$  com a informação rececionada até  $i$  dias depois. A estimação das probabilidades associadas aos vários dias de atraso foi obtida através de modelos de regressão logística, tendo como covariáveis: o dia de semana de morte, ser feriado no dia da morte, ser feriado no dia da previsão e ser feriado nos  $k$  dias anteriores à previsão, com  $k = 1, \dots, 4$  e para  $i < 5$ ,  $k = 1, \dots, i-1$ . Todas as variáveis mostraram ser significativas para explicar pelo menos um dos atrasos.

O desempenho da metodologia foi avaliado através de validação cruzada em duas fases: a primeira ajustando os modelos aos dados até 2011 e implementando a metodo-

logia para 2012 e a segunda, ajustando os modelos aos dados até 2012 e implementando a metodologia para 2013. Esta escolha deve-se à ocorrência de dois eventos relacionados com excesso de mortalidade nestes anos, epidemia de gripe em 2012 e onda de calor em 2013.

A metodologia foi capaz de produzir estimativas com qualidade, ou seja, com um erro relativo mediano inferior a 10%, quando se teve acesso a pelo menos um dia de informação. Na prática, corresponde aos momentos de previsão em que pelo menos um dos dias anteriores ao dia da previsão foi um dia útil. A metodologia também permitiu diminuir o tempo até à deteção do início do excesso de mortalidade durante os dois eventos em estudo comparativamente com o sistema atual. O início de excesso de mortalidade durante a epidemia de gripe em 2012 e durante a onda de calor em 2013 foi detetado, através da metodologia, 7 e 6 dias mais cedo, respetivamente.

É absolutamente essencial durante períodos de eventos relacionados com a saúde contar com informação atualizada sobre a mortalidade. A metodologia proposta mostrou ser capaz de estimar diariamente a mortalidade em Portugal, tendo acesso a pelo menos um dia de informação, contribuindo para a redução do tempo até à deteção de excesso de mortalidade. Desta forma, a deteção de impactes destes eventos em tempo real permite que a tomada de ações de mitigação seja mais atempada.

**Palavras-chave:** *Nowcasting*, Sistema de Vigilância diária da Mortalidade, Regressão Logística.

## Abstract

The daily mortality surveillance system (VDM) which monitors the observed “all cause” mortality in Portugal is centered in the Epidemiology Department of the National Health Institute Dr. Ricardo Jorge (INSA). This system was developed jointly with the Institute of Registries and Notaries (IRN) and the Institute of Financial Management and Equipment of Justice (IGFEJ).

The VDM system aims to detect and estimate the magnitude of deaths caused by new diseases epidemics, extreme weather conditions or other health-related events. This system is considered as an important tool for public health action.

The time lag between the onset of health-related event (excess deaths) and its detection by the VDM system is crucial for early detection of impacts on mortality, to the extent that, the greater the reporting delay of deaths to the VDM system, the greater the time until its detection.

The main aim of this project is to develop a method to estimate (nowcast), on a daily basis, the number of deaths occurred but not yet reported to the VDM system, by adjusting for reporting delay. The proposed method should therefore be able to reduce the time until the detection of the onset-related event by the VDM system.

The developed method updates the expected number of deaths on day  $t$ , being the total number of deaths occurred on day  $t$  estimated at day  $t + i$  given by:

$$\hat{M}_{t,i} = \frac{m_{t,i}}{\hat{\rho}_i}, \quad i = 1, \dots, 14, \quad ,$$

where  $\hat{\rho}_i$  is the estimated probability that a death occurred at day  $t$  is reported until day  $t + i$  and  $m_{t,i}$  is the number of deaths on day  $t$  reported to the system until day  $t + i$ . Logistic regression models were used to estimate probabilities associated with delays using the binary variables as covariates: weekday of death, occurrence of a public holiday at day of death, on prediction day and  $k$  days prior to prediction day,  $k = 1, \dots, 4$ ; for  $i < 5$ ,  $k = 1, \dots, i - 1$ . All variables have shown to be statistically significant to explain at least one of the delays.

The performance of the developed method was evaluated in two phases. In the first one by adjusting models to data from 2008 to 2011 and implementing the developed method for 2012. In the second one by adjusting models to data by 2012 and implementing the proposed method for 2013. The choice is due to the occurrence of two health-related events, influenza epidemic in 2012 and the heat wave in 2013. Available information accounts for a total of 593197 deaths.

This method was able to produce estimates with a median relative error lower than 10%, if it had access to at least one day of information. In practice, it corresponds to prediction moments when at least one day contributing to the estimate was a working day. This method also allowed to reduce the time between the onset of health-related event and its detection by the VDM system. The first day of excess mortality during the influenza epidemic in 2012 and during the heat wave in 2013 was detected 7 and 6 days earlier, respectively, using the method proposed.

Therefore, the proposed method was able to estimate the number of deaths in Portugal, having access at least to one day of information, and reduce substantially the time between the onset of health-related event and its detection by the VDM system. It is therefore considered to be absolutely essential during periods of health-related events rely on updated mortality information. Thus, the detection in real time of these impacts on mortality allows mitigation actions to be taken timely.

**Key-words:** *Nowcasting*, Daily Mortality Surveillance System, Logistic Regression.

---

## Conteúdo

---

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Enquadramento e Estado da arte . . . . .	2
1.2	O Sistema de Vigilância Diária da Mortalidade em Portugal . . . . .	4
1.3	Objetivos . . . . .	5
1.4	<i>Nowcasting</i> em sistemas de vigilância da mortalidade . . . . .	6
1.5	Estrutura do Trabalho . . . . .	7
<b>2</b>	<b>Descrição dos dados</b>	<b>8</b>
2.1	Base Dados . . . . .	8
2.2	Variáveis . . . . .	9
2.2.1	Escolha das variáveis a incluir na análise . . . . .	9
2.2.2	Definições . . . . .	11
2.2.3	Operacionalização das variáveis . . . . .	14
<b>3</b>	<b>Metodologia</b>	<b>16</b>
3.1	Estimação das probabilidades dos atrasos . . . . .	17
3.2	Operacionalização da Metodologia . . . . .	18
3.3	Avaliação do desempenho da Metodologia . . . . .	19
3.3.1	Erro das previsões . . . . .	20
3.3.2	Performance . . . . .	21
<b>4</b>	<b>Modelo de Regressão Logística múltiplo</b>	<b>23</b>
4.1	Formulação do modelo . . . . .	24



4.1.1	Identificação das variáveis potencialmente explicativas dos atrasos	24
4.2	Ajustamento do modelo	24
4.3	Quasi-verossimilhança	25
4.4	Seleção dos Modelos	27
4.5	Validação dos Modelos	28
4.5.1	Qualidade do ajustamento	28
4.5.2	Análise de Resíduos	29
<b>5</b>	<b>Resultados</b>	<b>31</b>
5.1	Análise preliminar dos dados	31
5.2	Operacionalização e avaliação da metodologia para 2012	37
5.2.1	Estimação das probabilidades dos atrasos	37
5.2.2	Implementação da Metodologia	45
5.2.3	Avaliação da Metodologia	55
5.3	Operacionalização e avaliação da metodologia para 2013	69
5.3.1	Estimação das probabilidades dos atrasos	69
5.3.2	Implementação da Metodologia	78
5.3.3	Avaliação da Metodologia	87
<b>6</b>	<b>Discussão, conclusão e desenvolvimentos futuros</b>	<b>99</b>
6.1	Discussão	99
6.1.1	Limitações	103
6.2	Conclusão	105
6.3	Desenvolvimentos futuros	106
6.4	Considerações finais	108
	<b>Bibliografia</b>	<b>109</b>
	<b>Anexo</b>	<b>111</b>
<b>A</b>	<b>Restantes Modelos de previsão</b>	<b>112</b>

---

## Lista de Tabelas

---

2.1	Variáveis incluídas na Base dados original . . . . .	8
2.2	Períodos de calor em excesso . . . . .	12
2.3	Períodos de epidemia de gripe . . . . .	13
2.4	Variáveis incluídas na Base dados de teste . . . . .	14
3.1	Previsões para 18 de Junho de 2012 . . . . .	19
3.2	Tabela de classificação dos resultados . . . . .	21
5.1	Medidas descritivas do atraso na notificação . . . . .	32
5.2	Medidas descritivas do atraso na notificação até 2011 . . . . .	37
5.3	Sumário do modelo de previsão 1 . . . . .	38
5.4	Sumário do modelo de previsão 2 . . . . .	39
5.5	Sumário do modelo de previsão 3 . . . . .	41
5.6	Sumário do modelo de previsão 4 . . . . .	42
5.7	Sumário do modelo de previsão 5 . . . . .	44
5.8	Número de previsões efetuadas para cada cenário de receção da in- formação e frequência relativa (%), para cada previsão $t + i$ , com $i = 1, \dots, 5$ , no ano de 2012 . . . . .	55
5.9	Erro relativo mediano (%) das previsões para cada cenário de receção de informação para 2012 . . . . .	56
5.10	Sensibilidade e Especificidade das previsões em $t + i$ , com $i = 1, \dots, 5$ , por cada cenário de receção de informação, para 2012 . . . . .	58

5.11	Valor Preditivo Positivo (VPP) e Valor Preditivo Negativo (VPN) por número de dias sem informação e por dia de previsão $t + i$ , com $i = 1, \dots, 5$	60
5.12	Número de notificações (N) e número de óbitos previstos para o início e fim do período de excesso de mortalidade	63
5.13	Medidas descritivas do atraso na notificação até 2012	69
5.14	Sumário do modelo de previsão 1	70
5.15	Sumário do modelo de previsão 2	72
5.16	Sumário do modelo de previsão 3	73
5.17	Sumário do modelo de previsão 4	75
5.18	Sumário do modelo de previsão 5	76
5.19	Número de previsões efetuadas para cada cenário de receção da informação, para cada previsão $t + i$ , com $i = 1, \dots, 5$ , no ano de 2013	87
5.20	Erro relativo mediano (%) das previsões para cada cenário de receção de informação para 2013	88
5.21	Sensibilidade e Especificidade das previsões em $t + i$ , com $i = 1, \dots, 5$ , por cada cenário de receção de informação, para 2013	90
5.22	Valor Preditivo Positivo (VPP) e Valor Preditivo Negativo (VPN) por número de dias sem informação e por dia de previsão $t + i$ , com $i = 1, \dots, 5$	91
5.23	Número de notificações (N) e número de óbitos previstos para o início e fim do período de excesso de mortalidade	93
A.1	Sumário do modelo de previsão 6 - Teste 2012	112
A.2	Sumário do modelo de previsão 7 - Teste 2012	113
A.3	Sumário do modelo de previsão 8 - Teste 2012	113
A.4	Sumário do modelo de previsão 9 - Teste 2012	114
A.5	Sumário do modelo de previsão 10 - Teste 2012	114
A.6	Sumário do modelo de previsão 11 - Teste 2012	115
A.7	Sumário do modelo de previsão 12 - Teste 2012	115
A.8	Sumário do modelo de previsão 13 - Teste 2012	116
A.9	Sumário do modelo de previsão 14 - Teste 2012	116
A.10	Sumário do modelo de previsão 6 - Teste 2013	117
A.11	Sumário do modelo de previsão 7 - Teste 2013	117
A.12	Sumário do modelo de previsão 8 - Teste 2013	118
A.13	Sumário do modelo de previsão 9 - Teste 2013	118
A.14	Sumário do modelo de previsão 10 - Teste 2013	119
A.15	Sumário do modelo de previsão 11 - Teste 2013	119

## *LISTA DE TABELAS*

---

A.16 Sumário do modelo de previsão 12 - Teste 2013 . . . . .	120
A.17 Sumário do modelo de previsão 13 - Teste 2013 . . . . .	120
A.18 Sumário do modelo de previsão 14 - Teste 2013 . . . . .	121

---

## Lista de Figuras

---

5.1	Frequência relativa do atraso na notificação por ano . . . . .	34
5.2	Frequência relativa do atraso na notificação por dia de semana de ocorrência do óbito . . . . .	35
5.3	Frequência relativa do atraso na notificação em função de ser feriado no dia de ocorrência do óbito . . . . .	35
5.4	Frequência relativa do atraso na notificação - óbitos ocorridos durante um evento relacionado com a saúde . . . . .	36
5.5	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 1 . .	39
5.6	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 2 . .	40
5.7	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 3 . .	42
5.8	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 4 . .	43
5.9	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 5 . .	45
5.10	Previsões em $t + 1$ para 2012 . . . . .	48
5.11	Previsões em $t + 1$ para 2012, excluindo as previsões efetuadas ao domingo e à segunda-feira . . . . .	49
5.12	Previsões em $t + 2$ para 2012 . . . . .	50
5.13	Previsões em $t + 2$ para 2012, excluindo as previsões efetuadas à segunda-feira . . . . .	51
5.14	Previsões em $t + 3$ para 2012 . . . . .	52
5.15	Previsões em $t + 4$ para 2012 . . . . .	53
5.16	Previsões em $t + 5$ para 2012 . . . . .	54
5.17	Erro das previsões por número de dias sem informação . . . . .	56

5.18	Previsões e notificações a $t+1$ durante a Epidemia de Gripe em 2012 . . .	64
5.19	Previsões e notificações a $t+2$ durante a Epidemia de Gripe em 2012 . . .	65
5.20	Previsões e notificações a $t+3$ durante a Epidemia de Gripe em 2012 . . .	66
5.21	Previsões e notificações a $t+4$ durante a Epidemia de Gripe em 2012 . . .	67
5.22	Previsões e notificações a $t+5$ durante a Epidemia de Gripe em 2012 . . .	68
5.23	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 1 . . .	71
5.24	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 2 . . .	73
5.25	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 3 . . .	74
5.26	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 4 . . .	76
5.27	Resíduos da deviance Padronizados vs Índice - Modelo de previsão 5 . . .	77
5.28	Previsões em $t + 1$ para 2013 . . . . .	80
5.29	Previsões em $t + 1$ para 2013, excluindo as previsões efetuadas ao do- mingo e à segunda-feira . . . . .	81
5.30	Previsões em $t + 2$ para 2013 . . . . .	82
5.31	Previsões em $t + 2$ para 2013, excluindo as previsões efetuadas à segunda- feira . . . . .	83
5.32	Previsões em $t + 3$ para 2013 . . . . .	84
5.33	Previsões em $t + 4$ para 2013 . . . . .	85
5.34	Previsões em $t + 5$ para 2013 . . . . .	86
5.35	Erro das previsões por número de dias sem informação . . . . .	89
5.36	Previsões e notificações a $t + 1$ durante a Onda de Calor em 2013 . . .	94
5.37	Previsões e notificações a $t + 2$ durante a Onda de Calor em 2013 . . .	95
5.38	Previsões e notificações a $t + 3$ durante a Onda de Calor em 2013 . . .	96
5.39	Previsões e notificações a $t + 4$ durante a Onda de Calor em 2013 . . .	97
5.40	Previsões e notificações a $t + 5$ durante a Onda de Calor em 2013 . . .	98

# CAPÍTULO 1

---

## Introdução

---

### 1.1 Enquadramento e Estado da arte

O Departamento de Epidemiologia (DEP) do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA) ocupa-se da realização de estudos epidemiológicos sobre o estado de saúde e doença da população Portuguesa e contribui para a concretização da função de “Observatório Nacional de Saúde” do INSA. É neste contexto que são desenvolvidas atividades de vigilância epidemiológica, como a Vigilância Diária da Mortalidade (VDM), que monitoriza a mortalidade por “todas as causas” em Portugal. Este sistema de vigilância foi desenvolvido em parceria com o Instituto dos Registos e do Notariado (IRN) e com o Instituto de Gestão Financeira e Equipamentos da Justiça (IGFEJ).

Genericamente, um sistema de vigilância em saúde pública é definido pela recolha, análise, interpretação e disseminação, sistemática e contínua, da informação relativa a um evento relacionado com a saúde [1]. A disseminação da informação proveniente destes sistemas pode ser usada para a ação imediata em saúde pública, formulação de hipóteses de investigação ou para o planeamento e avaliação de programas [1].

Deste modo, o sistema VDM tem como função a monitorização contínua e sistemática da mortalidade observada em Portugal, com o objetivo de detetar e estimar a magnitude do impacto de epidemias, novas doenças, condições climáticas extremas ou outros eventos relacionados com a saúde.

A mortalidade é um indicador muito poderoso em saúde pública, pois representa o

desfecho mais sério que pode resultar de um evento relacionado com a saúde [2]. Sendo a mortalidade por todas as causas um indicador uniforme e provavelmente o único indicador comparável entre os vários países da Europa [2], está em desenvolvimento, desde 2008, um projeto que visa implementar um sistema de vigilância da mortalidade que funcione a nível Europeu (EuroMOMO), no qual Portugal está envolvido.

Uma das características mais importantes nos sistemas de vigilância em saúde pública é denominada por *Timeliness*. Genericamente, é definido pelo tempo associado à passagem da informação entre as várias etapas de um sistema de vigilância, que podem ir desde a ocorrência do evento em observação pelo sistema até à tomada de ação [1]. Para o sistema VDM, a característica *timeliness* pode ser avaliada pelo tempo entre a ocorrência de impactes na mortalidade e a deteção dos mesmos pelo sistema, já que o objetivo deste sistema é o de detetar estes impactes. Dado isto, quanto maior é o atraso na notificação dos óbitos ao sistema VDM, que é medido pelo tempo entre a ocorrência do óbito e a chegada dessa informação ao sistema, maior será o tempo até à deteção destes impactes, sendo então crucial para determinar o *timeliness* do sistema VDM.

O referido atraso na notificação dos óbitos é um problema comum na vigilância diária da mortalidade por toda a Europa, sendo deste modo de grande importância o desenvolvimento de métodos que permitam estimar os óbitos ocorridos mas ainda não notificados ao sistema de vigilância[3]. Estes métodos são denominados por *nowcasting* que é, por definição, o método de prever o presente, o futuro próximo ou o passado recente, com informação incompleta proveniente de sistemas de recolha contínua e sistemática de informação [4].

Em Portugal, identificar as principais causas do atraso na notificação passa essencialmente por conhecer o processo de registo de um óbito, já que a notificação dos óbitos só se processa após o registo dos mesmos. Dado isto, é importante referir que o registo deve ser efetuado 48 horas após o falecimento ou após o cadáver ter sido encontrado e/ou autopsiado [5]. A circunstância da ocorrência do óbito é importante na medida em que cadáveres encontrados tardiamente e/ou submetidos a autópsia são registados com maior atraso. Do mesmo modo, as mortes ocorridas fora do país também podem ser registadas com maior atraso já que apenas alguns consulados Portugueses têm implementado o Sistema Integrado do Registo e Identificação Civil (SIRIC). Outro potencial motivo de atraso prende-se com o facto de o registo apenas poder ser efetuado em dias úteis, com exceção de algumas conservatórias que permitem registos ao sábado, e por isso, o número de registos efetuados ao fim-de-semana e em feriados é residual ou até nulo.



É absolutamente essencial durante períodos de ocorrência de eventos relacionados com excesso de mortalidade contar com informação atualizada sobre a mortalidade. Desta forma, impactes destes eventos são detetados em tempo real, tornando possível o acompanhamento da tendência epidemiológica dos mesmos e a tomada de ações de mitigação atempada.

### **1.2 O Sistema de Vigilância Diária da Mortalidade em Portugal**

O sistema VDM começou com a necessidade de se avaliar rapidamente o impacto da onda de calor de 2003 na mortalidade em Portugal e está ativo desde então. A tomada de ações de mitigação em saúde pública tende a esperar pela confirmação de impactes na população e assim que essa evidência era obtida, era demasiado tarde para agir. Dado esta situação, tornou-se indispensável criar uma ferramenta que pudesse providenciar evidência de impactes em tempo útil [6].

O primeiro sistema em 2003, um sistema de teste, contava apenas com a informação proveniente de 31 conservatórias do registo civil, contabilizando apenas 40% da mortalidade em Portugal. A primeira versão do VDM foi lançada em 2004 contando com informação proveniente de 67 conservatórias: as 31 conservatórias iniciais juntamente com uma amostra aleatória de 36 conservatórias. Este sistema esteve operacional de 2004 a 2007. A informação trocada consistia apenas no número de óbitos registados por conservatória (total e apenas maiores que 75 anos). A informação era enviada para o DEP através de chamadas telefónicas e fax, requerendo 6 a 7 pessoas diariamente para recolherem estes dados. Apenas as conservatórias que estavam informatizadas usavam como meio de troca de informação o e-mail. Esta informação era trocada apenas durante o período de verão e era enviado um boletim todos os dias úteis da semana para os seguintes parceiros: Direção Geral de Saúde (DGS), Autoridade Nacional de Proteção Civil (ANPC) e para o Ministério da Saúde. Em 2005 foi decidido prolongar a recolha desta informação para todo o ano.

O sistema atual, totalmente automatizado, está ativo desde 2007. Foi possível devido a uma mudança técnica no Ministério da Justiça, através de um sistema chamado SIRIC (Sistema Integrado de Registos de Identificação Civil) que centraliza a informação proveniente de todas as conservatórias.

O SIRIC conta com a informação do Instituto dos Registos e do Notariado e foi implementado pelo Instituto das Tecnologias de Informação na Justiça (ITIJ), atual-

mente o Instituto de Gestão Financeira e Equipamentos da Justiça. Em meados de 2006, o IRN pediu ao DEP que se testasse uma versão automatizada do sistema VDM usando a informação recolhida pelo SIRIC, apesar de nessa altura apenas metade das conservatórias existentes em Portugal reportarem ao sistema (as conservatórias informatizadas). Os dois sistemas funcionaram em simultâneo de Setembro de 2006 a Maio de 2007 mas em Junho de 2007 já todas as conservatórias reportavam ao SIRIC. Atualmente, a troca de informação é efetuada através de apenas um e-mail que contém a informação que chega ao SIRIC no dia anterior. Apenas alguma informação associada aos óbitos é enviada: a data de morte, data de nascimento, sexo e o código geográfico do local de registo do óbito. O e-mail é enviado pelo IGFEJ para o DEP diariamente, incluindo feriados e fins-de-semana. O Boletim da vigilância diária da mortalidade é enviado todos os dias úteis através de e-mail para a DGS, ANPC e Ministério da Saúde, onde são apresentados gráficos da série temporal da mortalidade para Portugal, Portugal Continental e por região (NUTS II de 2002). Também é divulgado o número de óbitos notificados ao sistema até ao dia da emissão do boletim para os últimos 14 dias.

### 1.3 Objetivos

O principal objetivo deste projeto é o desenvolvimento de uma metodologia que permita estimar diariamente o número de óbitos ocorridos mas ainda não notificados na totalidade ao sistema de vigilância diária da mortalidade em Portugal, através da modelação do atraso na notificação dos mesmos. O objetivo secundário é a diminuição do tempo até à deteção de impactes na mortalidade através da metodologia proposta. Para concretização dos objetivos descritos em cima, foram realizadas as seguintes tarefas:

1. Breve revisão dos métodos utilizados para o *nowcasting* em sistemas de vigilância, com principal interesse em sistemas de vigilância da Mortalidade;
2. Desenvolvimento da metodologia estatística adequada à solução do problema e que cumpra os objetivos do projeto.
3. Análise exploratória dos dados da mortalidade;
4. Descrição da implementação da metodologia;
5. Avaliação do desempenho da metodologia.

## 1.4 Nowcasting em sistemas de vigilância da mortalidade

A aplicação destes métodos na área da vigilância da saúde passa, essencialmente, pela deteção e acompanhamento em tempo real de impactes provocados por eventos relacionados com a saúde.

Alguns exemplos destas aplicações são os trabalhos desenvolvidos para o *nowcasting* das hospitalizações durante a pandemia de gripe A/H1N1 nos países baixos em 2009 [7], para o *nowcasting* da taxa de incidência semanal de síndrome gripal e da probabilidade da atividade gripal estar em estado epidémico em Portugal [4], e para o *nowcasting* da mortalidade durante a onda de calor em Inglaterra e no país de Gales em Junho de 2011 [8].

Existem poucos estudos relativamente à aplicação destes métodos em sistemas de vigilância da mortalidade. Pouca atenção tem sido dada ao desenvolvimento destas metodologias no âmbito destes sistemas e mais desenvolvimento nesta área seria de grande benefício [3]. Um estudo conduzido pelo projeto EuroMOMO em 2008, com o objetivo de se efetuar um levantamento sobre os sistemas de vigilância da mortalidade pela Europa, refere que apenas 11 dos 32 países haviam analisado o atraso na notificação dos óbitos [9]. Não é referido se estes países corrigem efetivamente a estimativa da mortalidade para os atrasos na notificação.

O sistema de vigilância da mortalidade da Escócia, automatizado desde 2009, utiliza um fator de correção baseado na função de distribuição empírica dos atrasos, agrupada em quatro categorias: dias úteis, sábados, domingos e feriados [3]. Os autores apenas apresentam os cálculos para a correção da mortalidade por sexo e grupo etário. O número total de mortes ocorridas a dia  $t - i$  ( $0 < i < 13$ ), para o grupo etário  $a$  e sexo  $s$ , estimado a dia  $t$ , é dado por:

$$C_{isat} = \frac{R_{sati}}{D_{d(t-i),g(a,s),w(i)}} \quad (1.1)$$

onde  $R_{sati}$  é o número de mortes ocorridas a dia  $t - i$  que foram notificadas até dia  $t$ ,  $D_{d(t-i),g(a,s),w(i)}$  a proporção esperada de mortes notificadas em  $w(i)$  dias úteis após dia  $t - i$ , para o dia de semana apropriado  $d(t - i)$  (dia útil, sábado, domingo ou feriado). Desta forma, assumem que o atraso na notificação é homogéneo para os óbitos ocorridos em dias úteis. Os autores referem também que o método utilizado poderia beneficiar se as correções para os atrasos fossem modeladas por distribuições probabilísticas.

O projeto EuroMOMO também corrige os óbitos observados para o atraso na no-

tificação. O método utilizado pressupõe que a proporção de mortes registadas durante um certo período é proporcional ao número de dias que as administrações estão abertas. Para cada semana, as proporções de mortes já registadas à data de agregação são calculadas e modeladas em função do número de dias que as administrações estiveram fechadas e uma tendência, através de um modelo linear generalizado (distribuição binomial)[10]. Posteriormente, o total da mortalidade é modelado usando a proporção estimada referida acima e o número de mortes já notificadas, através de um modelo linear generalizado (distribuição Poisson). O número de mortes corrigidas para o atraso é então prevista para um determinado período. O método terá um bom desempenho se a transmissão da informação for regular e “suave”, mesmo que o atraso seja longo. Se houver falta de informação e transmissão de informação irregular, é mais provável que o sistema produza uma mortalidade média em vez de variações esperadas. Não são apresentados detalhes mais técnicos.

É importante referir que este último método é aplicado a dados semanais e não a dados diários, como é o objeto deste projeto.

## 1.5 Estrutura do Trabalho

O presente trabalho é constituído por 7 capítulos e 1 anexo.

No atual capítulo é feita uma breve introdução ao tema, são apresentados os objetivos do projeto e os métodos utilizados para o *nowcasting* dos óbitos em sistemas de vigilância da mortalidade nos estudos revistos.

No segundo capítulo é apresentada a base dados original, as variáveis que se consideraram analisar e a operacionalização das mesmas.

A metodologia que se propõe para o *Nowcasting* dos óbitos no sistema de vigilância diária da mortalidade em Portugal e os métodos utilizados para a avaliação da mesma são apresentados no capítulo terceiro.

Os métodos utilizados para a modelação dos atrasos e para a avaliação dos mesmos são descritos no capítulo quarto e, no quinto capítulo são apresentados os resultados obtidos.

A discussão, conclusões e desenvolvimentos futuros para o projeto são então apresentados no capítulo sexto.

Por fim, o Anexo contém os sumários dos modelos que não foram apresentados juntamente com os resultados.

## CAPÍTULO 2

---

### Descrição dos dados

---

#### 2.1 Base Dados

A quantidade de informação que é enviada relativamente aos óbitos é reduzida. Apenas se tem acesso à data de ocorrência e registo do óbito e a alguma informação demográfica, como a data de nascimento, sexo e local do registo. Alguma desta informação é posteriormente transformada pelo sistema VDM, criando nova informação como a idade e o local de registo por região NUTS III do óbito. A operacionalização da informação que chega diariamente ao sistema VDM está apresentada na tabela 2.1.

Tabela 2.1: Variáveis incluídas na Base dados original

Descrição	Variável	Tipo da variável
Número de identificação do óbito	idObito	Numérico
Data de ocorrência do óbito	dataOb	Data
Data de registo do óbito	data	Data
Código conservatória	codConservatoria	Numérico
Descrição da conservatória	descConservatoria	Texto
Código geográfico da conservatória	codGeo	Numérico
Código NUTS III da conservatória	codNut3	Numérico
Data de nascimento	data2	Data
Idade	idade	Numérico
Sexo	sexo	Texto

A data de receção da informação não é incluída na base dados original (BD) e visto que é com base nesta data que se irá calcular o atraso da notificação dos óbitos, é de extrema importância que esta seja incluída na base dados com que se irá trabalhar.

O ficheiro enviado diariamente ao DEP consiste em registos individuais de óbitos que chegaram à base dados do SIRIC no dia anterior. A cada ficheiro está associada a data de referência que diz respeito à data de agregação da informação. A data de referência vem indicada no e-mail enviado diariamente, que tem como anexo o ficheiro dos óbitos, e que, posteriormente, é atribuída manualmente como nome de cada ficheiro. É importante salientar que a informação que é enviada em cada ficheiro para uma determinada data de referência não diz respeito apenas aos óbitos ocorridos e/ou registados nessa data, aliás, os ficheiros podem conter óbitos e/ou registos mais antigos. O ficheiro dos óbitos é sempre receccionado pelo DEP no dia seguinte à data de referência.

Até dia 3 de Janeiro de 2010 não se tinha acesso à data de referência dos ficheiros informaticamente, e por isso, recorreu-se à conta de e-mail do projeto para recolher a informação até essa data. Duplicou-se a base de dados original e reintroduziram-se todos os registos com a respetiva data de referência. O ficheiro mais antigo dizia respeito aos óbitos que integraram o SIRIC dia 7 de Maio de 2008 e, portanto, apenas se tem informação a partir desta data. A informação foi agregada a dia 20 de Janeiro de 2014. Esta base dados será referida como Base dados de teste (BD teste).

Durante este processo detetaram-se ficheiros referentes a longos períodos e não referentes a um único dia, como o habitual. Estes ficheiros vieram atualizar a informação de 2007 a 2009 que na altura não foi enviada na totalidade. A informação contida nestes ficheiros não foi inserida na BD teste, pois o cálculo do atraso na notificação destes óbitos não seria representativo do atraso real. Apesar de a mortalidade observada para o período de 7 de Maio de 2008 a 3 de Fevereiro de 2009 estar incompleta, é possível calcular o atraso na notificação dos óbitos que chegaram nos ficheiros diários e, que desta forma, seguiram o procedimento atual.

## **2.2 Variáveis**

### **2.2.1 Escolha das variáveis a incluir na análise**

Nesta fase é usual recorrer-se à análise preliminar dos dados e a bibliografia existente para a identificação de variáveis que possam explicar a variável de interesse.

Como foi apresentado na introdução, pouca atenção foi dada ao desenvolvimento

destas metodologias e, por isso, a informação que pode ser retirada de estudos anteriores é reduzida. As variáveis que se consideraram como potenciais variáveis explicativas de cada atraso na notificação, tiveram por base fatores que influenciam o tempo até ao registo, dado que a notificação ao sistema VDM só se realiza após o registo. Como foi referido, um óbito deve ser registado 48 horas após o falecimento ou após o cadáver ter sido encontrado e/ou autopsiado. Como se sabe, os registos são efetuados nas conservatórias do registo civil, as quais (na sua maioria) encerram ao fim-de-semana e aos feriados. É importante referir que, apesar das conservatórias estarem encerradas durante estes dias, é possível, ao contrário do esperado, fazer o registo de um óbito. O número de registos efetuados nestes dias é residual.

A ocorrência do óbito ao sábado, domingo ou a um feriado, são fatores referidos nos estudos revistos como variáveis que influenciam o atraso na notificação. Estas variáveis foram consideradas como potenciais variáveis explicativas dos atrasos, visto que também em Portugal os registos efetuados nestes dias são residuais ou até nulos.

Também se considerou que ser feriado nos dias anteriores à data de previsão, ou seja, nos dias em que é agregada a informação no SIRIC, poderia influenciar o atraso na notificação. Períodos que incluem vários dias em que os registos efetuados são reduzidos e até nulos, poderão levar a um maior atraso na notificação da informação referente a esse período. Estes podem ocorrer quando feriados antecedem ou sucedem fins-de-semana, podendo levar a três ou até quatro dias sem informação, se gerarem tolerâncias de ponto. Estes acontecimentos são frequentes em períodos festivos, como o Natal, a passagem de ano, o Carnaval e a Páscoa.

Sob o pressuposto de que ser feriado no dia seguinte ao de agregação da informação pode ser responsável por um maior número de registos (dada a impossibilidade de o fazer no dia seguinte), a influência desta variável também foi analisada (ser feriado no dia seguinte à data de agregação da informação).

Também se investigou se o dia da semana de ocorrência do óbito poderia influenciar o atraso na notificação. Apesar de a definição de atraso na notificação ser distinta neste trabalho, em Portugal já se tinha verificado que o número de dias de atraso na notificação era influenciado pelo dia da semana de ocorrência do óbito [6].

Períodos de maior mortalidade, levam naturalmente a um maior número de registos, e podem alterar o funcionamento das conservatórias. Consequentemente podem alterar o atraso na notificação. Foram então incluídas variáveis que indicam se o dia da morte ocorreu durante um período de epidemia de gripe ou durante uma onda de calor para se poder analisar o efeito destes fatores.

Em suma, considerou-se que o atraso na notificação poderá depender apenas do dia da semana de ocorrência do óbito, se esse dia, ou os dias anteriores e seguintes, foram um feriado nacional, e se o óbito ocorreu durante um evento relacionado com a saúde.

### 2.2.2 Definições

Nesta secção são apresentados critérios para a definição de um período de evento relacionado com a saúde, nomeadamente um período de epidemia de gripe e de onda de calor. Estas definições são as utilizados pelos sistemas de vigilância implementados no DEP para a monitorização do impacte destes eventos. Também é apresentada a definição utilizada para a determinação de um período com excesso de mortalidade.

#### Onda de calor

É conduzido no DEP, desde 1999, nos meses de Maio a Outubro, um sistema de vigilância e monitorização das ondas de calor com potenciais efeitos na saúde humana, o sistema ICARO (*Importância do Calor - Repercussão nos Óbitos*)[11]. Este sistema foi implemento em parceria com o Instituto de Meteorologia (IM), atualmente o Instituto Português do Mar e da Atmosfera (IPMA), e conta com a participação da Direção Geral de Saúde (DGS) e da Autoridade Nacional de Proteção Civil (ANPC). Desde 2004 que faz parte integrante do Plano de Contingência para temperaturas extremas adversas – Módulo calor [11]. Este sistema conta com a informação diária das temperaturas observadas e previstas cedidas pelo IPMA. Um período de calor em excesso com possível repercussão nos óbitos é definido pelo conjunto de dias consecutivos em que se observam pelo menos um dos três critérios abaixo:

- Índice ICARO diário acima de 1;
- Um distrito com temperatura máxima igual ou superior a 36°C;
- Dois ou mais distritos com temperatura máxima igual ou superior a 35°C.

Os períodos de calor em excesso utilizados neste trabalho são apresentados na tabela 2.2.



Tabela 2.2: Períodos de calor em excesso

2007	2008	2009	2010
04/06	20/06 a 22/06	28/05 a 01/06	31/05 a 01/06
04/07 a 13/07	25/06 a 30/06	12/06 a 13/06	23/06
25/07 a 04/08	14/07 a 19/07	17/06 a 22/07	29/06 a 01/07
06/08	21/07 a 22/07	18/07 a 21/07	03/07 a 11/07
08/08 a 09/08	02/08 a 03/08	25/07 a 26/07	17/07 a 19/07
17/08	05/08	28/07 a 30/07	24/07 a 01/08
02/09 a 06/09	09/08	04/08	03/08 a 15/08
	13/08	10/08 a 19/08	20/08 a 22/08
		21/08 a 23/08	25/08 a 26/08
		27/08 a 31/08	28/08 a 31/08
		07/09 a 09/09	03/09 a 04/09
			12/09 a 14/09
2011	2012	2013	
24/06 a 27/06	14/05 a 17/05	23/06 a 01/07	
30/06 a 01/07	31/05	03/07 a 11/07	
15/07	24/06 a 28/06	17/07	
20/07	16/07 a 26/07	30/07 a 01/08	
24/07 a 29/07	30/07	04/08	
09/08 a 12/08	03/08	09/08 a 26/08	
15/08 a 16/08	07/08 a 12/08	31/08 a 04/09	
19/08 a 22/08	17/08 a 23/08	20/09 a 22/09	
07/09 a 09/09	31/08 a 07/09		
13/09 a 14/09	12/09 a 16/09		

### Atividade Gripal

A definição de um período epidêmico de gripe depende da informação proveniente da rede Médicos Sentinela (MS), uma rede constituída por médicos de medicina geral e familiar, e de outras redes de vigilância laboratorial. A informação proveniente destas redes consiste em estimativas semanais das taxas de incidência de síndrome gripal (SG) e do número de casos de SG positivos para o vírus da gripe [12]. O período epidêmico é definido como um conjunto de semanas consecutivas com estimativas da incidência de SG acima do limite superior de confiança a 95% para a linha base, durante as quais foram reportados casos não esporádicos de gripe com confirmação laboratorial [12].

Os períodos de epidemia de gripe utilizados neste trabalho são apresentados na tabela:

Tabela 2.3: Períodos de epidemia de gripe

Período
15/01/2007 a 18/02/2007
14/01/2008 a 10/02/2008
01/12/2008 a 01/02/2009
26/10/2009 a 20/12/2009
13/12/2010 a 06/02/2011
23/01/2012 a 25/03/2012

### Excesso de mortalidade

O objetivo da monitorização da mortalidade é a deteção e confirmação de excesso de mortalidade, que é função da diferença entre a mortalidade observada e a mortalidade esperada, na ausência de eventos associados a excesso de mortalidade. No sistema VDM, para o cálculo da mortalidade esperada, são ajustados modelos de regressão cíclica à série temporal da mortalidade, usando a biblioteca de **R**, Flubase [6, 13]. Os modelos são ajustados à série temporal excluindo os períodos onde se observaram eventos, para se poder detetar e confirmar períodos com excesso de mortalidade. A previsão da mortalidade obtida através do modelo é considerada a mortalidade esperada ou a linha de base para a mortalidade.

Neste trabalho, um período com excesso de mortalidade vai ser determinado seguindo duas das “Westguard rules” como são popularmente intituladas. Estas regras, criadas por James Westguard, fazem parte de um procedimento de múltiplas regras utilizado para o controle interno de qualidade de um processo em fase analítica. O objetivo é determinar se um processo está em controlo e estas regras definem critérios para a decisão. Existem cinco regras que formam o procedimento de múltiplas regras, mas estas podem e devem ser escolhidas e combinadas da forma mais adequada ao processo que se quer analisar [14]. Os limites são escritos em função da média e do desvio-padrão, sendo do tipo  $(\bar{x} \pm 2s)$ , ou do tipo  $(\bar{x} \pm 3s)$ , dependendo da regra utilizada.

A regra  $1_{2s}$  é uma regra de alerta, que é despoletada quando uma observação está acima do limite de controlo superior ou abaixo do limite de controlo inferior, definidos pelo dobro do desvio-padrão.

A regra  $2_{2s}$  indica que se deve determinar um processo fora de controlo, quando duas observações consecutivas estão acima do limite de controlo superior ou abaixo do limite de controlo inferior, definidos também pelo dobro do desvio-padrão.

A  $1_{3s}$  indica que se deve considerar o processo fora de controlo quando uma observação está acima ou abaixo dos limites superior ou inferior respectivamente do triplo

do desvio-padrão.

Neste trabalho, os limites de controlo foram definidos pelos limites dos intervalos de confiança a 95% e 99% para a linha de base da mortalidade, correspondendo aos limites de controle 2s e 3s, respetivamente. Apenas se aplicou a regra referente ao limite superior, visto que apenas se está interessado em detetar períodos de excesso de mortalidade.

A determinação de um período de excesso de mortalidade foi então definida como:

- Dois dias consecutivos com valores da mortalidade entre os limites superiores de confiança a 95% e 99% para a linha de base da mortalidade e termina após dois dias consecutivos com valores da mortalidade abaixo deste mesmo limite;
- Um dia com valores da mortalidade acima do limite superior de confiança a 99% para a linha de base da mortalidade e termina após dois dias consecutivos para valores da mortalidade abaixo do limite superior de confiança a 95%;

### 2.2.3 Operacionalização das variáveis

As variáveis incluídas da Base dados de teste e a operacionalização das mesmas encontram-se na tabela 2.2.1.

Tabela 2.4: Variáveis incluídas na Base dados de teste

Descrição	Variável ou variáveis	Tipo da variável
Data de referência do ficheiro	dataFicheiro	Data
Data de receção do ficheiro	dataNot	Data
Número de dias de atraso na notificação	atraso	Numérico
Dia da semana da morte	diamorte	Numérico
Feriado no dia da morte	feriadamorte	Binário
Feriado nos $k$ dias da previsão, $k = 1, \dots, 15$	f1,...,f15	Binário
Epidemia de gripe no dia da morte	gripemorte	Binário
Calor em excesso no dia da morte	calormorte	Binário

O atraso na notificação dos óbitos foi calculado como a diferença entre a data de ocorrência do óbito e a data de receção do ficheiro ao DEP. A data de receção foi definida como o dia seguinte à data de referência do ficheiro. Ao calcular os atrasos para o período de estudo, detetaram-se atrasos negativos e nulos, que correspondem, certamente, a erros de introdução da data de referência na Base dados original. Estes registos foram eliminados.

O dia de semana da ocorrência de morte foi gerado automaticamente através da função *as.POSIXlt* da biblioteca de base do **R**.

Todos os feriados foram mapeados desde 2007 até 2013. O mesmo foi feito para os períodos de epidemia de gripe e calor em excesso. O ano escolhido para o início do registo destas variáveis foi baseado na intersecção da data de ocorrência de óbito mais antiga e a data a partir da qual se tem conhecimento dos períodos em que ocorreram estes eventos. A data mais antiga de morte foi dia 1 de Janeiro de 2006, mas apenas se tem conhecimento do impacto destes eventos desde 2007, data a partir da qual o sistema VDM se considera automatizado. Desta forma, apenas foram considerados os óbitos ocorridos a partir do dia 1 de Janeiro de 2007.

## CAPÍTULO 3

---

### Metodologia

---

Como foi apresentado no primeiro capítulo, o principal objetivo deste projeto é o desenvolvimento de uma metodologia que permita estimar diariamente o número de óbitos ocorridos mas ainda não notificados ao sistema VDM, através da modelação do atraso na notificação dos mesmos. A metodologia que se propõe estima diariamente o número total de óbitos ocorridos com base no número de óbitos notificados ao sistema. Considerando  $m_{t,i}$  o número de óbitos ocorridos a dia  $t$  e notificados com atraso igual ou inferior a  $i$  dias,  $M_t$  o número total de óbitos ocorridos a dia  $t$  e  $p_i$  a probabilidade de um óbito ser notificado com atraso igual ou inferior a  $i$  dias, tem-se:

$$m_{t,i} \sim \text{Binomial}(M_t, p_i) \quad (3.1)$$

$$E[m_{t,i}] = M_t \times p_i$$

Manipulando a equação acima, e considerando a informação recebida até dia  $t + i$ , o número de óbitos ocorridos a dia  $t$ ,  $M_{t,i}$ , pode ser estimado por:

$$\hat{M}_{t,i} = \frac{m_{t,i}}{\hat{p}_i}, \quad i = 1, \dots, \quad (3.2)$$

onde  $\hat{p}_i$  é a probabilidade estimada de um óbito ser notificado com um atraso igual ou inferior a  $i$  dias. A estimação de cada  $p_i$  será descrita na secção 3.1.

Visto que, se decorridos  $i$  dias após a data  $t$ , ainda não tiver sido recebida qualquer

notificação de óbitos ocorridos a  $t$ ,  $m_{t,i}$  toma o valor zero, pelo que a estimativa do número de óbitos é dada por:

$$\hat{M}_{t,i} = \frac{m_{t,i}}{\hat{\rho}_i} = \frac{0}{\hat{\rho}_i} = 0 \quad i = 1, \dots$$

É importante recordar que esta situação, em que o número de notificações para uma data  $t$  é nulo passados  $i$  dias, só ocorre quando a morte se deu durante o fim-de-semana ou feriado nacional, dado que as conservatórias estão encerradas. Apesar de só se dar nas condições enunciadas, nem sempre as notificações são nulas, sendo possível chegarem notificações de óbitos registados durante o fim-de-semana ou em feriados nacionais e que, usualmente, são enviadas para o DEP no dia seguinte. Também é importante recordar que os registos efetuados nestes dias são residuais.

Para contornar a limitação do estimador proposto, nos dias em que tal ocorrer, irá assumir-se  $m_{t,i} = 1$ . Desta forma, o estimador do número de óbitos ocorridos a dia  $t$ , dado que  $m_{t,i}$  é zero, é dado por:

$$\hat{M}_{t,i} = \frac{1}{\hat{\rho}_i} - 1 \quad i = 1, \dots \quad (3.3)$$

A escolha do numerador prende-se com o facto de este ser o valor mais próximo de zero, o valor real de notificações para o dia  $t$ , que  $m_{t,i}$  pode tomar. Dado que se assume que se recebeu uma notificação relativa à data  $t$ , após a estimativa do número de óbitos ocorridos, deve retirar-se este valor, dado que esta notificação foi adicionada de forma artificial.

Para a implementação da metodologia proposta vão seguir-se os seguintes passos:

1. Estimação das probabilidades associadas aos atrasos;
2. Operacionalização da metodologia;
3. Avaliação do desempenho da metodologia;

### 3.1 Estimação das probabilidades dos atrasos

Para estimar a probabilidade de um óbito ser notificado com um atraso igual ou inferior a  $i$  dias, é necessário recolher informação sobre o número de óbitos que são

notificados com atraso igual ou inferior a  $i$ . Esta informação pode ser obtida a partir do número de dias de atraso observado na notificação de cada óbito, passando a definir-se a variável  $A_j$  como o número de dias de atraso na notificação de cada óbito  $j$ .

Considerando  $A_{ji}$  a variável que indica se um óbito  $j$  chegou com atraso igual ou inferior a  $i$  dias,

$$A_{ji} = \begin{cases} 1, & \text{se } A_j \leq i \\ 0, & \text{c.c.} \end{cases},$$

$p_{ji}$  é definido como

$$p_{ji} = P[A_{ji} = 1].$$

Desta forma, para cada óbito  $j$ , definem-se as variáveis  $A_{ji}$ , com  $i = 1, \dots$ , à custa da variável  $A_j$ , que indicam se o atraso na notificação do mesmo foi igual ou inferior a  $i$  dias, passando a ter-se:

$$A_{ji} \sim \text{Bernoulli}(p_{ji}) \quad \text{ou} \quad A_{ji} \sim \text{Binomial}(1, p_{ji}).$$

A estimação das probabilidades associadas aos atrasos será conseguida através de modelos de regressão, em particular, modelos de regressão logística. Os modelos de regressão têm como principal objetivo estudar a relação entre variáveis ou, mais particularmente, analisar a influência que uma ou mais variáveis têm sobre uma variável de interesse. No caso do modelo de regressão logística, a variável de interesse segue uma distribuição binomial e, por isso, é o mais indicado para cada variável  $A_{ji}$ , visto que, como foi definido em cima, também estas seguem uma distribuição Binomial. O recurso a este método assenta no objetivo de analisar a influência das variáveis que caracterizam o óbito  $j$ , consideradas na secção 2.2.1, no atraso da notificação dos óbitos e poder estimar a probabilidade de um óbito chegar com atraso igual ou inferior a  $i$  dias, dada a influência destas variáveis. Uma descrição mais técnica do modelo de regressão logística está disponível no capítulo 4.

## 3.2 Operacionalização da Metodologia

A metodologia foi operacionalizada em **R** [15]. Foi desenvolvida uma função que para a data  $t$ , o dia para o qual se quer estimar o número de óbitos ocorridos, devolve :

- a data em que se está a fazer a previsão,  $t + i$ ;
- o número de óbitos notificados até ao dia em que se está a fazer a previsão,  $m_{t,i}$ ;

- o número de óbitos estimado para cada o dia  $t$  através da metodologia adotada,  $\hat{M}_{t,i}$ ;
- o número de óbitos notificados até ao dia corrente.

Para tal, a função recebe apenas o dia para o qual se quer estimar o número de óbitos, ou seja, a data  $t$ . Apenas foram ajustados 14 modelos, passando a definir-se  $i = 1, \dots, 14$ , dado que apenas são apresentados no boletim da vigilância diária da mortalidade os valores da mortalidade para os últimos 14 dias. A tabela 3.1 apresenta o output da função implementada, para o dia 18 de Junho de 2012.

Tabela 3.1: Previsões para 18 de Junho de 2012

Previsão	Data de Previsão	Notificados	$p_i$	Nowcast
$t + 1$	2012-06-19	43	0.1815	236.97
$t + 2$	2012-06-20	183	0.7650	239.21
$t + 3$	2012-06-21	214	0.9330	229.36
$t + 4$	2012-06-22	224	0.9699	230.93
$t + 5$	2012-06-23	229	0.9771	234.36
$t + 6$	2012-06-24	229	0.9739	235.12
$t + 7$	2012-06-25	229	0.9715	235.72
$t + 8$	2012-06-26	231	0.9761	236.65
$t + 9$	2012-06-27	231	0.9796	235.79
$t + 10$	2012-06-28	231	0.9818	235.27
$t + 11$	2012-06-29	231	0.9829	235.00
$t + 12$	2012-06-30	231	0.9839	234.78
$t + 13$	2012-07-01	231	0.9838	234.80
$t + 14$	2012-07-02	231	0.9837	234.80
Total	2014-02-13	233	-	-

### 3.3 Avaliação do desempenho da Metodologia

O primeiro passo começa com a divisão da Base dados de teste (BD teste) em duas partes: a Base Dados A (BD A) e a Base Dados B (BD B). A primeira será utilizada para a determinação e validação dos modelos utilizados para estimar as probabilidades associadas aos atrasos, enquanto que a segunda será utilizada para avaliar o desempenho da Metodologia. Esta técnica, denominada de validação cruzada, permite avaliar



a capacidade de generalização do modelo, isto é, permite avaliar o desempenho dos modelos para um novo conjunto de dados. Esta capacidade é uma característica vital, visto que se pretende que a metodologia desenvolvida seja implementada futuramente no sistema VDM e, portanto, será aplicada diariamente a novos dados da mortalidade. A partição da BD teste irá ser efetuada por conveniência, não sendo alterada a ordem temporal. Dado que em 2012 e 2013 ocorreram eventos com excesso de mortalidade, epidemia de gripe e onda de calor, respectivamente, optou-se por avaliar a metodologia em duas fases. Para tal, foi necessário:

1. Ajustar os modelos com os ficheiros recebidos até dia 31 de Dezembro de 2011 e avaliar o desempenho da metodologia para o ano de 2012.
2. Ajustar os modelos com os ficheiros recebidos até dia 31 de Dezembro de 2012 e avaliar o desempenho da metodologia para o ano de 2013.

Para o primeiro teste, a BD A será composta pelos ficheiros recebidos entre dia 7 de Maio de 2008 e dia 31 de Dezembro de 2011 e a BD B pelos ficheiros recebidos durante o ano de 2012.

Para o segundo teste, a BD A será composta pelos ficheiros recebidos entre dia 7 de Maio de 2008 e dia 31 de Dezembro de 2012 e a BD B pelos ficheiros recebidos durante o ano de 2013.

Como se mostrou no início do presente capítulo, a estimativa do número total de óbitos ocorridos na data  $t$  é escrita em função do número de óbitos notificados até  $t + i$  e, dado que são utilizados dois estimadores, devido à dependência da metodologia em notificações não nulas, é importante avaliar o desempenho de ambos. Também é importante avaliar se as estimativas do número total de óbitos será influenciada pelo número de dias que foi rececionada informação e, por isso, a avaliação do desempenho da metodologia será efetuada em função do número de dias em que não houve receção de informação.

Neste trabalho foi considerado um nível de significância  $\alpha$  de 10%.

#### 3.3.1 Erro das previsões

O erro associado a cada estimativa  $\hat{M}_{t,i}$ , vai ser medido pelo erro relativo percentual, dado por:

$$ER_{t,i}\% = \frac{|M_t - M_{t,i}|}{M_t} \times 100\% \quad (3.4)$$

O erro associado a cada dia de previsão  $i$ , vai ser dado pela mediana do erro relativo percentual das previsões feitas  $i$  dias após a ocorrência de morte.

### 3.3.2 Performance

#### Sensibilidade e Especificidade

A sensibilidade e a especificidade são medidas de performance de classificação e neste estudo vão ser utilizadas para avaliar o desempenho da metodologia desenvolvida relativamente à deteção de excesso de mortalidade. A sensibilidade é definida como a probabilidade de se prever um dia com excesso de mortalidade em dias em que houve um verdadeiro excesso de mortalidade e a especificidade é definida como a probabilidade de se prever um dia sem excesso de mortalidade em dias em que não houve verdadeiramente excesso de mortalidade. Para se obterem estas medidas deve ser construída uma tabela de contingência como a tabela 3.2, e as medidas são pelas expressões dadas em 3.5.

Tabela 3.2: Tabela de classificação dos resultados

		Metodologia	
		excesso	sem excesso
Mortalidade ocorrida	excesso	a	b
	sem excesso	c	d

$$Sensibilidade = \frac{a}{a + b} \quad Especificidade = \frac{d}{c + d} \quad (3.5)$$

Para classificar tanto a mortalidade observada como a mortalidade estimada pela metodologia com e sem excesso de mortalidade seguiu-se o procedimento definido na secção 2.2.2.

#### Valor Preditivo Positivo e Valor Preditivo Negativo

O Valor preditivo positivo (VPP) é definido como a probabilidade de detetar um verdadeiro excesso e é dado como a proporção de verdadeiros dias de excesso de mortalidade entre os dias detetados com excesso pela metodologia proposta. O Valor preditivo negativo (VPN) é definido como a probabilidade de detetar um verdadeiro dia sem excesso de mortalidade e é dado como a proporção de verdadeiros dias sem excesso de

mortalidade entre os dias detetados sem excesso pela metodologia proposta. A sua importância reside em quantificar o desempenho da metodologia relativamente à credibilidade do alerta de excesso de mortalidade. As medidas são calculadas usando a tabela 3.2 e são dadas por:

$$VPP = \frac{a}{a + c} \quad VPN = \frac{d}{b + d}$$

#### **Timeliness na deteção de um período de excesso de mortalidade**

O *timeliness* do sistema VDM será medido pelo *timeliness* na deteção de um período de excesso de mortalidade, visto que o objetivo do sistema é o de detetar e estimar a magnitude do impacto de epidemias, novas doenças, condições climáticas extremas ou outros eventos relacionados com a saúde na mortalidade. Este é medido pela rapidez com que um excesso é detetado e pode ser expresso em unidades de tempo salvo ou como a proporção de tempo salvo [16], dados por:

$$Timeliness = \frac{n - n^*}{n} \quad \text{ou} \quad Timeliness = n - n^*,$$

onde  $n^*$  e  $n$  é o número de dias desde o início de excesso de mortalidade até à sua deteção através da metodologia e através do sistema VDM atual, respetivamente.

## CAPÍTULO 4

---

### Modelo de Regressão Logística múltiplo

---

Como foi referido no capítulo 3, o principal objetivo da aplicação de modelos de regressão é o de estudar a relação entre variáveis ou, mais particularmente, analisar a influência que uma ou mais variáveis têm sobre uma variável de interesse. Às primeiras dá-se o nome de variáveis explicativas e à variável de interesse, dá-se o nome de variável resposta. O modelo de regressão logística integra a classe dos Modelos Lineares Generalizados (MLG), que sintetiza vários modelos de estrutura de regressão linear que têm em comum o facto da variável resposta seguir uma distribuição que pertence à família exponencial. Cada uma das nossas variáveis de interesse, como vimos no capítulo 3, seguem uma distribuição Binomial, que por sua vez, pertence à família exponencial, dado que pode ser escrita da forma:

$$f(y|\theta, \phi) = \exp \left\{ \frac{\omega}{\phi} (y\theta - b(\theta)) + c(y, \phi, \omega) \right\}, \quad (4.1)$$

onde  $\theta$  é um parâmetro de localização,  $\phi$  um parâmetro de dispersão e  $\omega$  uma constante conhecida.

Para a construção de cada modelo de regressão devem seguir-se os seguintes passos:

1. Formulação dos modelos
2. Ajustamento dos modelos
3. Seleção dos modelos

## 4. Validação dos modelos

### 4.1 Formulação do modelo

#### 4.1.1 Identificação das variáveis potencialmente explicativas dos atrasos

Nesta fase é usual recorrer-se à análise preliminar dos dados e a bibliografia existente para a identificação de potenciais variáveis que possam explicar a variável de interesse. A identificação destas variáveis está descrita na 2.2.1

### 4.2 Ajustamento do modelo

No modelo de regressão logística, a variável resposta  $Y$ , é uma variável binária que pode tomar valores 0 e 1, correspondendo respetivamente à ausência e ocorrência do evento de interesse.

Quando várias observações partilham o mesmo vetor de covariáveis, o que acontece frequentemente quando todas as covariáveis são categóricas, considera-se como variável resposta,  $Y$ , a frequência relativa de sucessos do grupo, isto é,  $Y_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji}$ , onde  $n_j$  é número de indivíduos no  $j$ -ésimo grupo e

$$Y_{ji} = \begin{cases} 1 & \text{se a } i\text{-ésima observação do grupo } j \text{ é sucesso} \\ 0 & \text{c.c.} \end{cases}.$$

Como as frequências absolutas seguem uma distribuição  $Binomial(n_j, \pi_j)$ , as frequências relativas vão seguir uma distribuição  $B(n_j, \pi_j)/n_j$  [17], isto é,

$$f(y_j | \pi_j) = \binom{n_j}{n_j y_j} \pi_j^{n_j y_j} (1 - \pi_j)^{n_j - n_j y_j}, \quad y = 0, \frac{1}{n_j}, \dots, 1 \quad (4.2)$$

Prova-se que a distribuição pertence à família exponencial, dado que a sua função massa de probabilidade se pode escrever da forma definida em 4.1, se tivermos:

$$\theta = \ln\left(\frac{\pi_j}{1 - \pi_j}\right),$$

$$b(\theta) = \ln(1 + e^\theta), \quad c(y, \phi) = \ln\left(\frac{n_j}{n_j y_j}\right),$$

$$b'(\theta) = \frac{e^\theta}{1 + e^\theta} = \pi_j, \quad b''(\theta) = V(\mu_j) = \frac{e^\theta}{(1 + e^\theta)^2} = \pi_j(1 - \pi_j),$$

$$\phi = 1, \quad \omega = n_j.$$

A cada grupo  $j$  está associado um vetor de especificação  $\mathbf{z}_j = (1, x_{j1}, \dots, x_{jp})^T$ , com  $j = 1, \dots, m$ , sendo  $m$  o número de grupos, ou seja,  $\sum_{j=1}^m n_j = n$ . Quando existem variáveis qualitativas com  $q$  categorias ( $q > 2$ ), devem ser criadas  $q-1$  variáveis *dummy*, para substituir a variável original no vetor  $\mathbf{z}$ .

Seja  $E[Y_j] = \pi_j$  com função de ligação canónica  $\log\left(\frac{\pi_j}{1-\pi_j}\right) = \text{logit}(\pi_j)$ . O modelo de regressão logística é dado por

$$\text{logit}(\pi_j) = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_k x_{jp}, \quad (4.3)$$

onde a probabilidade de sucesso,  $P(Y_j = 1) = \pi_j$ , é dada por

$$\pi_j = \frac{\exp(\mathbf{z}_j^T \boldsymbol{\beta})}{1 + \exp(\mathbf{z}_j^T \boldsymbol{\beta})}. \quad (4.4)$$

Ajustar o modelo requer que se obtenham estimativas para o vetor  $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \dots, \beta_k)$  e para tal é utilizado o método da máxima verosimilhança. Estas estimativas são obtidas através da resolução do sistema de equações de verosimilhança,  $\frac{\partial l(\boldsymbol{\beta})}{\partial l(\boldsymbol{\beta})_k} = 0$ , com  $k = 0, \dots, p$ , em que  $l(\boldsymbol{\beta})$  representa o logaritmo da verosimilhança de  $\boldsymbol{\beta}$ . Estas equações não têm solução analítica, sendo necessário recorrer a métodos iterativos para a sua resolução. As soluções de máxima verosimilhança para  $\boldsymbol{\beta}$  podem ser determinadas de forma automática através da utilização da função *glm* da biblioteca base do **R**, que utiliza o algoritmo iterativo de mínimos quadrados ponderados para resolver o sistema de equações.

### 4.3 Quasi-verosimilhança

Um fenómeno que acontece frequentemente é as respostas apresentarem uma variância superior à explicada pelo modelo binomial. Este fenómeno é denominado por sobre-dispersão e pode dever-se ao facto de existir heterogeneidade entre as observações que

não é explicada pelas covariáveis, ou por existir correlação entre as observações. Nestes casos, uma das soluções passa por introduzir um parâmetro de sobredispersão,  $\phi > 1$ , desconhecido, admitindo que se tem  $var(Y_j) = \phi V(\mu_j)$ . Sabemos que para o modelo Binomial isto é verdade se se fizer  $\phi = 1$ , mas o modelo já não pode ser escrito na forma da família exponencial. O modelo passa a estar definido apenas através do valor médio e da variância:

$$E(Y_j) = \mu_j = \pi_j, \quad var(Y_j) = \phi V(\mu_j) = \phi \frac{\pi_j(1 - \pi_j)}{n_j},$$

não sendo possível o recurso à verosimilhança para fazer inferência. A solução passa então pelo conceito de *quasi-verosimilhança*.

A variável

$$U = U(\mu_j, Y_j) = \frac{Y_j - \mu_j}{\phi V(\mu_j)},$$

com

$$E(U) = 0, \quad var(U) = \frac{1}{\phi V(\mu_j)}, \quad -E \left[ \frac{\partial U}{\partial \mu_j} \right] = var(U),$$

comporta-se como uma função score, sendo a derivada de log-verosimilhança, e como tal, espera-se que o integral de  $U$  se comporte como uma função log-verosimilhança. A função de *quasi-verosimilhança* (ou *quasi-log-verosimilhança*) é definida por

$$Q_j(\mu_j, y_j) = \int_{y_j}^{\mu_j} \mu_j(t, y) dt = \int_{y_j}^{\mu_j} \frac{y_j - t}{\phi V(t)} dt \quad (4.5)$$

No caso em que temos  $n$  observações de variáveis aleatórias independentes a função *quasi-verosimilhança*, é o somatório de cada contribuição individual:

$$Q_j(\mu_j, y_j) = \sum_{j=1}^n Q_j(\mu_j, y_j) \quad (4.6)$$

As estimativas de  $\beta$  são obtidas ao igualar a função  $s^*(\beta) = \frac{\partial Q}{\partial \beta}$ , que não depende de  $\phi$ , a zero. A esta função dá-se o nome de quasi-score ou função de estimação generalizada. Utilizando este método, a estimação dos parâmetros não é afetada pelo valor de  $\phi$ , e desta forma, os parâmetros estimados coincidem com os parâmetros estimados pelo modelo logístico, mas a inferência é ajustada tendo em conta a sobredispersão.

As propriedades assintóticas dos estimadores de quasi-verosimilhança  $\hat{\beta}^*$ , podem ser obtidas sob condições de regularidade semelhantes às necessárias para os estimadores

de máxima verosimilhança. Então:

$$\hat{\beta}^* \sim N(\beta, (I^*)^{-1}(\hat{\beta}^*)V(\hat{\beta}^*)(I^*)^{-1}(\hat{\beta}^*)),$$

onde

$$I^*(\beta) = E\left(-\frac{\partial s^*(\beta)}{\partial \beta^T}\right)$$

e

$$V(\beta) = \text{cov}(s^*(\beta)).$$

O método de quasi-verosimilhança permite a obtenção de estimadores consistentes e assintoticamente normais para  $\beta$ , com apenas perda de eficiência.

O parâmetro  $\phi$  é estimado através de

$$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^m \frac{(y_j - n_j \hat{\pi}_j)^2}{n_j \hat{\pi}_j (1 - \hat{\pi}_j)}, \quad (4.7)$$

que segue uma distribuição aproximada de um  $\chi^2$  com  $(n-p)$  graus de liberdade.

## 4.4 Seleção dos Modelos

Após a identificação de potenciais variáveis explicativas e de uma primeira análise das variáveis, a etapa seguinte passa pela determinação do modelo, isto é, determinar as covariáveis e respetivos coeficientes a introduzir no modelo final. Quando existe um elevado número de covariáveis, é importante que o modelo final seja o mais parcimonioso possível, incluindo o menor número possível de covariáveis sem por em causa a capacidade de ajustamento e de predição do modelo e de forma a permitir uma boa interpretação do mesmo. Podemos então definir o melhor modelo como sendo aquele que melhor combina os princípios de ajustamento, parcimónia e interpretação. Neste trabalho, dado o número reduzido de covariáveis, as únicas que não serão incluídas nos modelos finais, serão aquelas que nunca se mostraram significativas em nenhum dos modelos.

### Teste de Wald

Para testar a significância dos coeficientes associados a cada covariável, será a utilizada a estatística de *Wald*, que vem implementada na função *glm* do **R**. A estatística



de *Wald* é uma das estatísticas mais utilizadas para inferir sobre  $\beta$  e é baseada na normalidade assintótica do estimador de *quasi-verosimilhança*,  $\hat{\beta}^*$ ,

$$\hat{\beta}^* \sim N(\beta, V(\beta)).$$

Para testar a hipótese

$$H_0 : \beta_k = 0 \quad \text{vs} \quad H_1 : \beta_k \neq 0,$$

sob  $H_0$ , a estatística de *Wald* é dada por

$$\frac{\hat{\beta}_k}{\sqrt{V(\hat{\beta})}} \sim N(0, 1). \quad (4.8)$$

A hipótese nula é rejeitada em favor da hipótese alternativa, quando o valor-p associado ao valor da estatística de teste é inferior a um determinado nível de significância  $\alpha$ , ou seja, quando a probabilidade de se observar um valor para a estatística de teste tão ou mais extremo do que o observado, for inferior a um determinado nível de significância  $\alpha$ .

## 4.5 Validação dos Modelos

### 4.5.1 Qualidade do ajustamento

#### Teste de Hosmer-Lemeshow

Para os modelos de regressão logística, um dos testes mais utilizados para avaliar a qualidade do ajustamento de um modelo aos dados é o teste de Hosmer-Lemeshow [18]. Este teste baseia-se no agrupamento das probabilidades estimadas pelo modelo através de dois métodos:

1. Baseado nos percentis das probabilidades estimadas
2. Baseado em valores fixos das probabilidades estimadas

No primeiro método, definido  $g$  como o número de percentis escolhido, obtém-se sempre para o primeiro grupo as  $n_1 = \frac{n}{g}$  observações com as probabilidades estimadas mais pequenas. O último grupo irá conter as  $n_g = \frac{n}{g}$  observações com as probabilidades

estimadas mais elevadas. No segundo método, criam-se pontes de corte calculados por  $\frac{k}{g}, k = 1, \dots, g - 1$ . A título de exemplo, se  $g = 10$ , o primeiro ponte de corte é 0.1 e o primeiro grupo é definido pelas observações com valores das probabilidades estimadas iguais ou inferiores a 0.1 e o ultimo grupo pelas observações com probabilidade estimada superior a 0.9. O primeiro método mostrou ser melhor do que o segundo em termos de aderência à distribuição e especialmente quando as probabilidades estimadas são pequenas ( $< 0.2$ ). São usualmente usados 10 grupos ( $g=10$ ) e é o que normalmente está implementado nos softwares estatísticos [18]. A estatística de teste é dada por:

$$\hat{C} = \sum_{k=1}^g \left[ \frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right] \quad (4.9)$$

sendo:

$$o_{0k} = \sum_{j=1}^{c_k} (m_j - y_j)$$

$$\hat{e}_{1k} = \sum_{j=1}^{c_k} m_j \hat{\pi}_j$$

$$\hat{e}_{0k} = \sum_{j=1}^{c_k} m_j (1 - \hat{\pi}_j)$$

$$o_{1k} = \sum_{j=1}^{c_k} y_j$$

e  $c_k$  o número de padrões de covariáveis no grupo  $k$ .

Para testar a hipótese

$$H_0 : \text{O modelo está bem ajustado} \quad \text{vs} \quad H_1 : \text{O modelo não está bem ajustado},$$

$C_k$  segue uma distribuição qui-quadrado com  $(g - 2)$  graus de liberdade, sob a hipótese nula. Esta é rejeitada em favor da hipótese alternativa quando o valor-p associado ao valor da estatística de teste é inferior a um determinado nível de significância  $\alpha$ .

#### 4.5.2 Análise de Resíduos

A análise de resíduos é útil, não só para uma avaliação local da qualidade de ajustamento de um modelo como também para ajudar a identificar observações mal ajustadas,

isto é, que não são bem explicadas pelo modelo [17]. Os resíduos representam a diferença entre o valor observado e o valor estimado pelo modelo, sendo esta análise essencialmente gráfica. Os resíduos da *Deviance* (*Deviance residuals*) são os mais utilizados quando se trabalha com modelos de regressão logística e são dados por

$$r_j^d = \text{sign}(y_j - \hat{y}_j) \sqrt{d_j},$$

em que  $\hat{y}_j$  são os valores ajustados pelo modelo e  $d_j$  representa a contribuição da  $j$ -ésima observação para a função da *deviance*, dada por

$$d_j = 2m_j \left[ \ln \left( \frac{y_j}{\hat{y}_j} \right) + (1 - y_j) \ln \left( \frac{1 - y_j}{1 - \hat{y}_j} \right) \right],$$

onde  $m_j$  é o número de grupos.

Os resíduos da *deviance* padronizados são dados por:

$$r_j^{*D} = \frac{r_j^D}{\sqrt{\phi(1 - h_{jj})}}, \quad (4.10)$$

em que  $h_{jj}$  representa a diagonal da matriz de projeção (*hat*), que no caso do modelo logístico é dada por

$$\mathbf{H} = \mathbf{V}^{1/2} \mathbf{Z} (\mathbf{Z}^T \mathbf{V} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{1/2}, \quad (4.11)$$

onde  $\mathbf{Z}$  representa a matriz ( $m \times p$ ) que contém os valores das  $p$  covariáveis para os  $m$  grupos e  $\mathbf{V}$  a matriz diagonal com elemento geral dado por  $v_j = \hat{\pi}_j(1 - \hat{\pi}_j)$ .

Os resíduos padronizados devem ser representados graficamente contra os valores ajustados pelo modelo, ou contra cada índice  $j$ . Espera-se que 95% dos valores dos resíduos estejam compreendidos entre -2 e 2.

## CAPÍTULO 5

---

### Resultados

---

Neste capítulo serão apresentados os resultados da análise preliminar dos dados e em seguida os resultados obtidos para cada ano considerado para a validação da metodologia, 2012 e 2013. Nos resultados para cada ano, primeiro são apresentados os sumários dos modelos considerados para a estimação das probabilidades associadas aos atrasos e a respetiva análise. Apenas são apresentados os modelos que correspondem aos cinco primeiros dias de previsão dado que estes são os mais pertinentes para a deteção atempada de impactes na mortalidade. Os sumários dos restantes modelos são apresentados no anexo A. Em seguida são apresentadas as previsões produzidas através da metodologia e os resultados da sua avaliação.

### 5.1 Análise preliminar dos dados

Esta primeira fase da análise tem como objetivo iniciar o contacto com os dados e começar a analisar a relação do atraso na notificação com os fatores que se consideraram ter influência no mesmo.

A base dados de teste contabiliza um total de 593197 óbitos, referentes aos ficheiros recebidos entre 7 de Maio de 2008 e 20 de Janeiro de 2014. O maior atraso observado foi de 2543 dias. Em dois dias obtém-se metade da informação de um dado dia mas para ter acesso a 90% da informação, são necessários 5 dias. A média dos atrasos é de 6 dias, muito superior à mediana, o que nos indica assimetria positiva na distribuição dos

Tabela 5.1: Medidas descritivas do atraso na notificação

mínimo	máximo	média	desvio-padrão	variância
1	2543	6.14	55.48	3078.43

Percentis						
50%	75%	80%	90%	95%	98%	99%
2	3	4	5	6	11	41

atrasos. Para melhor visualização, as observações cujo atraso na notificação foi superior a 15 dias, foram censuradas nas representações gráficas. A informação considerada compreende um total de 583659 observações.

A figura 5.1, que apresenta o atraso na notificação por mês e por ano, sugere que existe variação ao longo dos meses nos vários anos, mas não parece sugerir uma variação sistemática.

Relativamente ao dia da semana de ocorrência do óbito, a figura 5.2, aponta para a existência de diferenças relativamente ao número de dias de atraso na notificação dos mesmos. O atraso de um dia na notificação é mais frequente quando o óbito ocorre de segunda-feira a sexta-feira, observando-se com maior frequência à sexta-feira. Para melhor compreensão do atraso de um dia na notificação, se um óbito for registado no próprio dia de morte e essa informação chegar à base dados do SIRIC nesse dia, a notificação chega ao sistema VDM no dia seguinte, chegando com atraso de um dia.

Óbitos ocorridos ao fim-de-semana, relativamente aos outros dias de semana de ocorrência do óbito, são notificados com menor frequência com atraso de um dia. Este resultado já seria de esperar visto que os óbitos registados no próprio dia de morte, sendo este sábado ou domingo, são residuais. Os óbitos ocorridos ao domingo são notificados frequentemente com atraso de dois dias, e ocorridos ao sábado, com atraso de três dias. Na prática, os óbitos ocorridos ao sábado são frequentemente registados na segunda-feira seguinte, a informação chega ao SIRIC nesse dia, e estes óbitos chegam ao sistema VDM na terça-feira, com atraso de três dias. Os óbitos ocorridos ao domingo, seguem o mesmo processo que os óbitos ocorridos ao sábado e, chegam ao sistema VDM na terça-feira, com atraso de dois dias. Já os óbitos ocorridos à sexta-feira são notificados com maior frequência quatro dias depois. Analisando este resultado no sentido prático, estes óbitos acabam por seguir o mesmo processo que os óbitos ocorridos ao fim-de-semana, sendo registados na segunda-feira seguinte, e são notificados ao sistema VDM

na terça-feira, chegando então com atraso de quatro dias.

Nos dias em que o dia de ocorrência do óbito foi um feriado nacional, informação apresentada na figura 5.3, a notificação destes ocorre com menos frequência com atraso de um dia, o que seria expectável.

Relativamente aos óbitos ocorridos durante um evento relacionado com a saúde, epidemia de gripe ou calor em excesso, a figura 5.4 sugere não existir diferenças na notificação destes comparativamente com os dias em que tal não acontece.

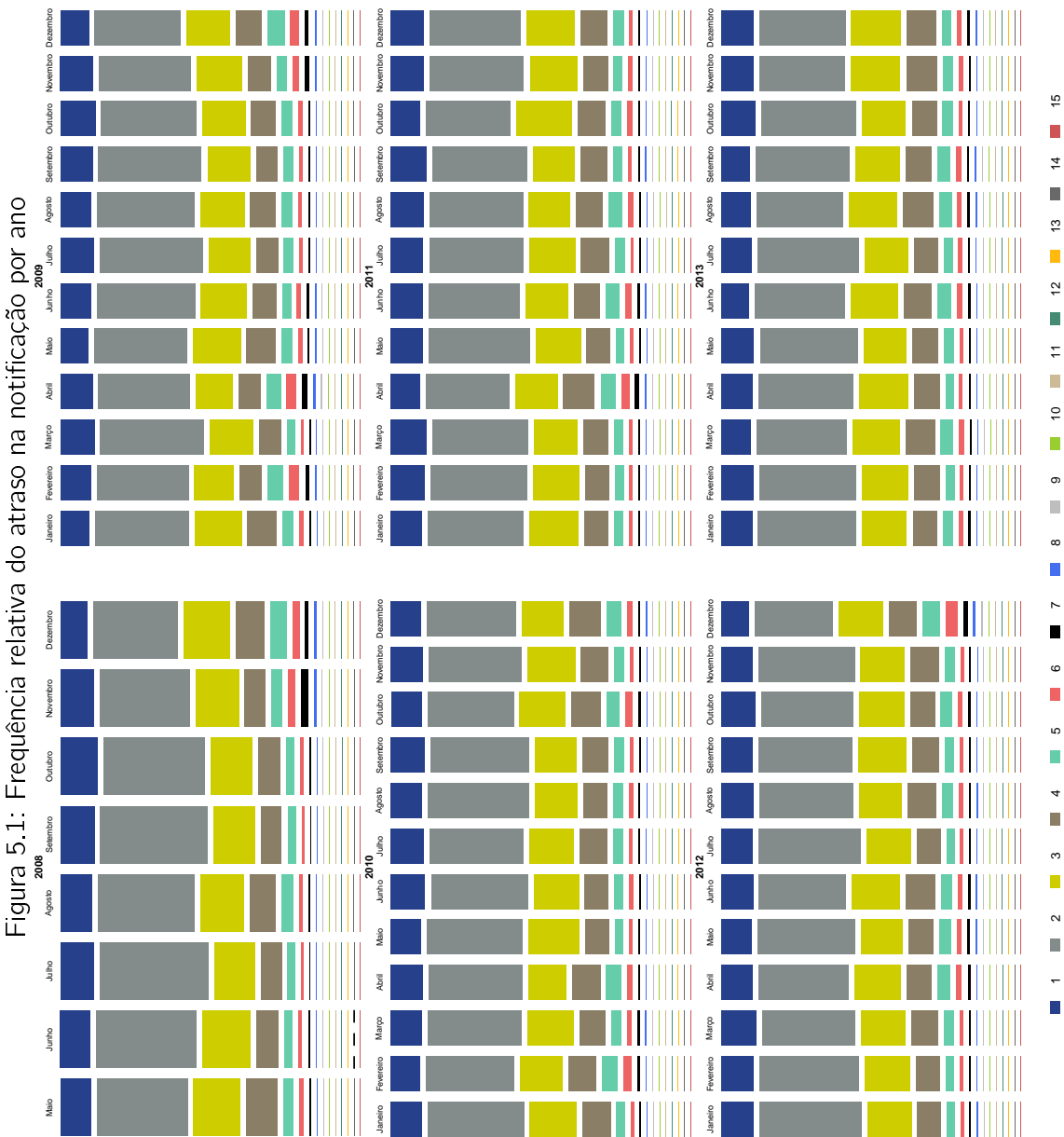


Figura 5.2: Frequência relativa do atraso na notificação por dia de semana de ocorrência do óbito

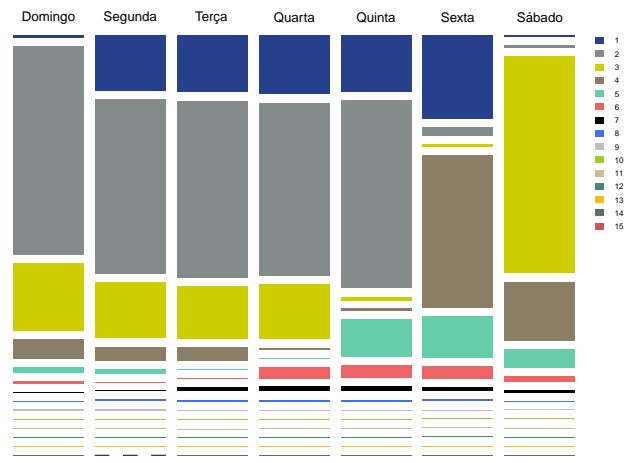


Figura 5.3: Frequência relativa do atraso na notificação em função de ser feriado no dia de ocorrência do óbito

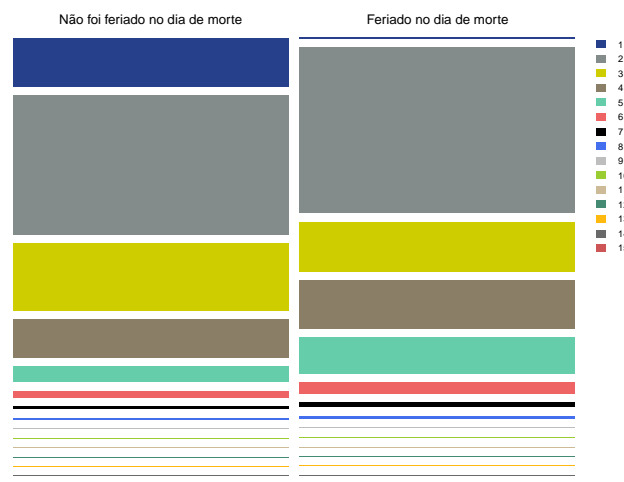
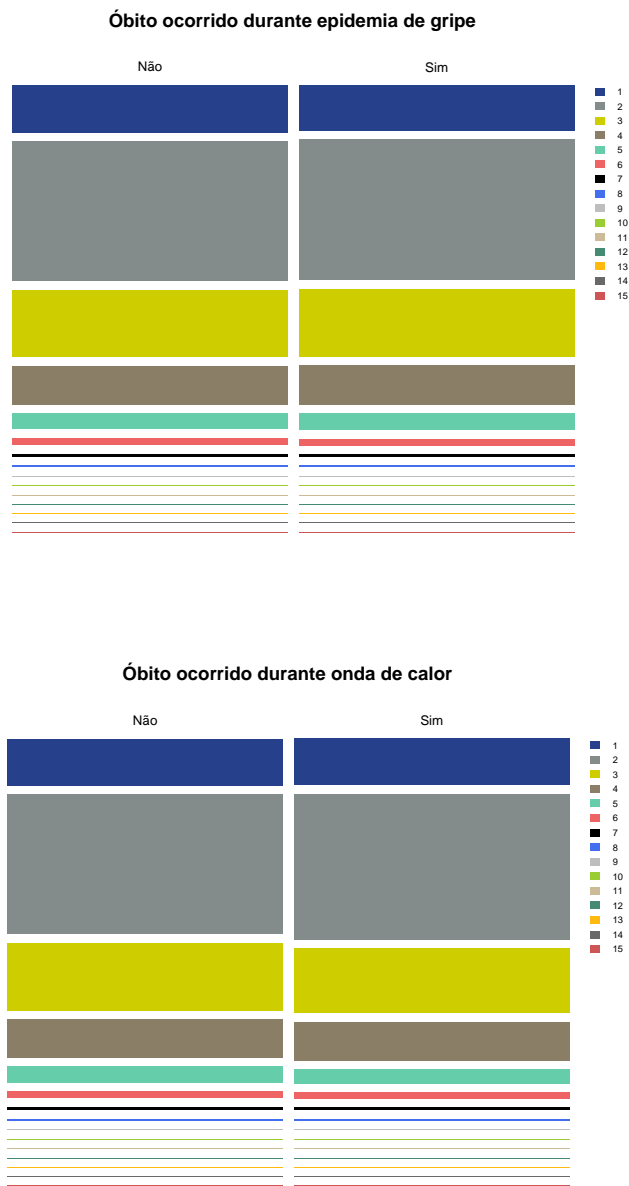




Figura 5.4: Frequência relativa do atraso na notificação - óbitos ocorridos durante um evento relacionado com a saúde



## 5.2 Operacionalização e avaliação da metodologia para 2012

Para avaliar o desempenho da metodologia para o ano de 2012, os modelos foram ajustados considerando os dados recebidos até ao final de 2011, compreendendo um total de 368500 observações. Os valores obtidos para as medidas descritivas dos atrasos não parecem diferir substancialmente dos valores obtidos com a informação completa. Os valores que sofreram alterações foram aqueles que dependem dos atrasos mais longos visto que se perdeu acesso a informação mais recente e, consequentemente, a possíveis notificações com atrasos mais longos.

Tabela 5.2: Medidas descritivas do atraso na notificação até 2011

mínimo	máximo	média	desvio-padrão	variância
1	1811	5.25	38.70	1497.93

Percentis						
50%	75%	80%	90%	95%	98%	99%
2	3	4	5	6	11	34

### 5.2.1 Estimação das probabilidades dos atrasos

Cada modelo, que terá como objetivo estimar a probabilidade de o atraso ser igual ou inferior a  $i$  dias, para simplicidade, será referido como modelo de previsão  $i$ . O óbito ter ocorrido durante um evento relacionado com a saúde, onda de calor ou epidemia de gripe, não se mostrou significativo para a explicação de cada atraso. Os modelos finais foram ajustados excluindo estas variáveis, o que já era sugerido na análise preliminar dos dados

As covariáveis comuns a todos os modelos foram:

- dia de semana de ocorrência do óbito
- ser feriado no dia de ocorrência do óbito
- ser feriado no dia da previsão
- ser feriado nos últimos  $k$  dias anteriores à previsão  $i$ ,  $k = 1, \dots, 4$ . Para  $i < 5$ ,  $k = 1, \dots, i - 1$

**Modelo de previsão 1**

Tabela 5.3: Sumário do modelo de previsão 1

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	-5.1232	0.2075	-24.68	0.0000
Segunda <sup>a</sup>	3.6336	0.2113	17.19	0.0000
Terça <sup>a</sup>	3.7056	0.2112	17.54	0.0000
Quarta <sup>a</sup>	3.7052	0.2112	17.54	0.0000
Quinta <sup>a</sup>	3.6849	0.2114	17.43	0.0000
Sexta <sup>a</sup>	4.1467	0.2105	19.70	0.0000
Sábado <sup>a</sup>	0.0278	0.2927	0.10	0.9249
Feriado morte	-3.9591	0.4880	-8.11	0.0000
Feriado dia da previsão	0.2796	0.0941	2.97	0.0056
$\phi$ : 20.34337				
Null deviance: 39169.84 on 19 degrees of freedom				
Residual deviance: 164.15 on 11 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 1 é apresentado na tabela 5.3.

Todas as variáveis mostraram ser significativas com exceção do sábado como dia de semana de ocorrência do óbito. Este resultado indica-nos que o sábado não difere significativamente do domingo, como dia de ocorrência de morte, no atraso de um dia na notificação dos óbitos. A ocorrência do óbito de segunda a sexta aumenta a probabilidade de um óbito ser notificado no dia seguinte à data de morte comparativamente com a ocorrência do óbito ao fim-de-semana. O efeito do dia de ocorrência do óbito de segunda a quinta-feira não diferem substancialmente (variam entre 3.6 e 3.7), indicando que o efeito que exercem na probabilidade da notificação no dia seguinte à morte é semelhante. O efeito associado à sexta-feira como dia de morte é mais acentuado que os restantes dias úteis, sendo este o que mais aumenta a probabilidade de um óbito ser notificado no dia seguinte ao dia da morte. Por outro lado, ser feriado no dia da morte diminui a probabilidade da notificação no dia seguinte à ocorrência do óbito, e ser feriado no dia seguinte ao dia da morte, aumenta esta probabilidade.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 0.1791$ , com 8 graus de liberdade e um valor-p associado de 1. O teste indica um bom ajustamento.

Os resíduos, apresentados na figura 5.5, não apresentam valores extremos, indicando a ausência de observações mal ajustadas pelo modelo.

Figura 5.5: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 1

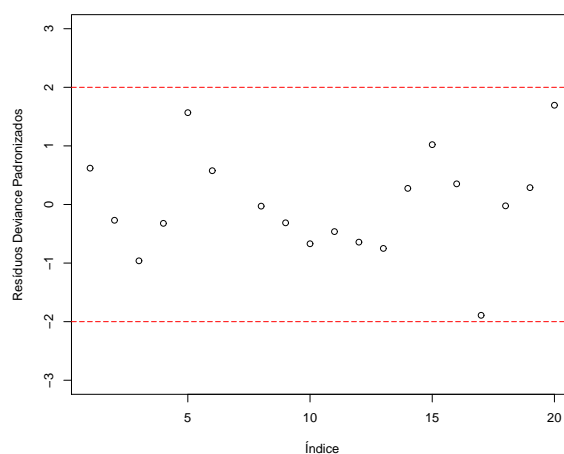
**Modelo de previsão 2**

Tabela 5.4: Sumário do modelo de previsão 2

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	0.7790	0.0795	9.80	0.0000
Segunda <sup>a</sup>	0.4014	0.1160	3.46	0.0028
Terça <sup>a</sup>	0.4433	0.1168	3.80	0.0013
Quarta <sup>a</sup>	0.4404	0.1171	3.76	0.0014
Quinta <sup>a</sup>	0.5886	0.1200	4.90	0.0001
Sexta <sup>a</sup>	-1.6234	0.1129	-14.38	0.0000
Sábado <sup>a</sup>	-3.9360	0.2013	-19.55	0.0000
Feriado morte	-0.5483	0.1828	-3.00	0.0077
Feriado dia anterior à previsão	-2.2994	0.1905	-12.07	0.0000
Feriado no dia da previsão	-0.0073	0.2006	-0.04	0.9714
$\phi$ : 69.80373				
Null deviance: 131582.6 on 27 degrees of freedom				
Residual deviance: 1442.8 on 18 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 2 e a representação gráfica dos resíduos estão apresentados na tabela 5.4 e na figura 5.6, respetivamente.

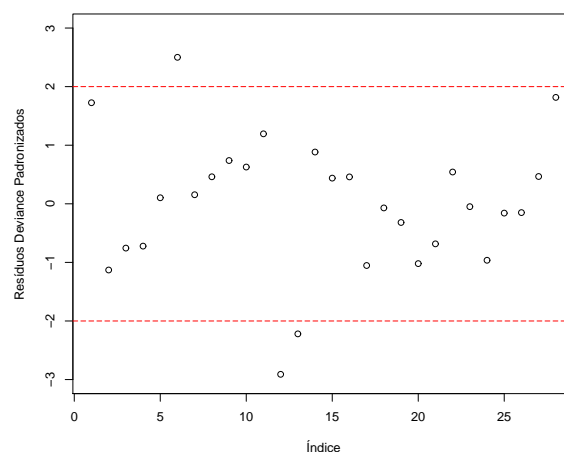
Todas as variáveis mostraram ser significativas, com exceção de ser feriado no dia da previsão, que representa o dia seguinte à data de agregação da informação. É de notar

que os valores dos coeficientes para o dia de semana de ocorrência do óbito diferem substancialmente dos valores para os mesmos no modelo de previsão 1. O dia de semana de ocorrência do óbito de segunda a quinta aumenta a probabilidade de um óbito ser notificado dentro de dois dias após a morte comparativamente com o domingo, mas esta já não é tão acentuada. Indica portanto que a probabilidade da notificação dos óbitos ocorridos ao domingo aumenta no atraso de dois dias na notificação, relativamente ao atraso de um dia. Já a ocorrência do óbito à sexta e ao sábado diminuem a probabilidade do óbito ser notificado em dois dias em comparação com o domingo. Ser feriado no dia anterior à previsão, ou seja, no dia em que é agregada a informação, e ser feriado no dia da morte, diminui a probabilidade de um óbito ser notificado dentro de dois dias.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 0.9647$ , com 8 graus de liberdade e um valor-p associado de 0.9985, indicando um bom ajustamento.

Os resíduos não apresentam observações com valores extremos, indicando a inexistência de observações atípicas.

Figura 5.6: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 2



**Modelo de previsão 3**

Tabela 5.5: Sumário do modelo de previsão 3

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	2.2624	0.1405	16.10	0.0000
Segunda <sup>a</sup>	0.3716	0.2089	1.78	0.0870
Terça <sup>a</sup>	0.3916	0.2086	1.88	0.0717
Quarta <sup>a</sup>	0.3823	0.2083	1.84	0.0780
Quinta <sup>a</sup>	-0.8393	0.1757	-4.78	0.0001
Sexta <sup>a</sup>	-3.0077	0.1682	-17.88	0.0000
Sábado <sup>a</sup>	-1.3395	0.1687	-7.94	0.0000
Feriado morte	-0.6741	0.2472	-2.73	0.0113
Feriado no dia anterior à previsão	-1.5699	0.2176	-7.21	0.0000
Feriado dois dias antes da previsão	-1.5208	0.2150	-7.07	0.0000
Feriado no dia da previsão	-0.2281	0.2583	-0.88	0.3851
$\phi$ : 95.83889				
Null deviance: 90894.1 on 36 degrees of freedom				
Residual deviance: 2938.7 on 26 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 3 encontra-se apresentado na tabela 5.5.

Todas as variáveis mostraram ser significativas, com exceção de ser feriado no dia de previsão. Os óbitos ocorridos de segunda a quarta, comparativamente com os óbitos ocorridos ao domingo, aumentam a probabilidade de um óbito ser notificado com atraso igual ou inferior a três dias, ao contrário dos óbitos ocorridos de quinta a sábado, que diminuem esta probabilidade. Ser feriado no dia da morte e nos dois dias seguintes, também diminui a probabilidade de um óbito ser notificado num prazo de três dias.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 1.8337$ , com 8 graus de liberdade e um valor-p associado de 0.9857. O teste indica que o modelo apresenta um bom ajustamento.

Os resíduos, apresentados na figura 5.7, não apresentam valores extremos, indicando um bom ajustamento local.

Figura 5.7: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 3

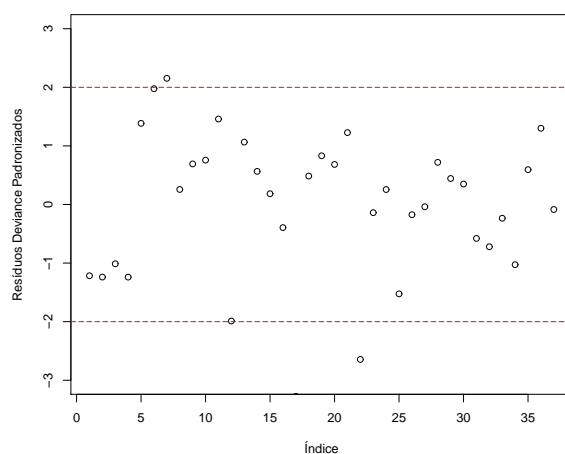
**Modelo de previsão 4**

Tabela 5.6: Sumário do modelo de previsão 4

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.2080	0.1858	17.27	0.0000
Segunda <sup>a</sup>	0.2671	0.2662	1.00	0.3228
Terça <sup>a</sup>	0.2576	0.2692	0.96	0.3455
Quarta <sup>a</sup>	-0.5103	0.2349	-2.17	0.0369
Quinta <sup>a</sup>	-1.7074	0.2090	-8.17	0.0000
Sexta <sup>a</sup>	-1.8338	0.2078	-8.82	0.0000
Sábado <sup>a</sup>	-0.9643	0.2231	-4.32	0.0001
Feriado morte	-0.5487	0.2382	-2.30	0.0275
Feriado dia anterior à previsão	-1.3143	0.2298	-5.72	0.0000
Feriado dois dias antes da previsão	-1.1264	0.2492	-4.52	0.0001
Feriado três dia antes da previsão	-1.5425	0.2191	-7.04	0.0000
Feriado dia da previsão	-0.5549	0.2480	-2.24	0.0320

 $\phi$ : 80.27634

Null deviance: 29838.1 on 45 degrees of freedom

Residual deviance: 3016.5 on 34 degrees of freedom

<sup>a</sup> Referência: Domingo

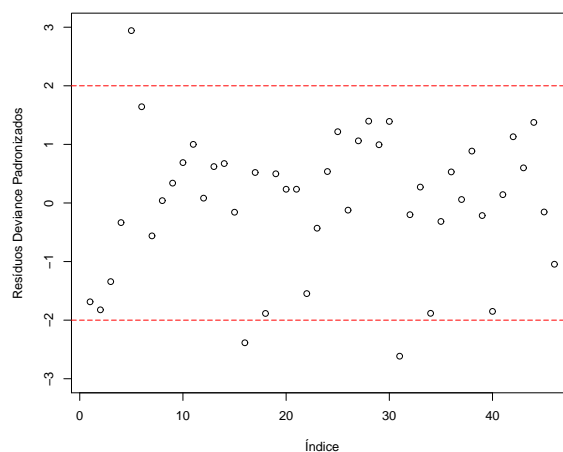
O sumário do modelo de previsão 4 e os respectivos resíduos são apresentados na tabela 5.6 e na figura 5.8.

Todas as variáveis mostraram ser significativas, com exceção dos dias de semana segunda-feira e terça-feira para o dia de ocorrência do óbito. O efeito da ocorrência dos óbitos nestes dias de semana não difere significativamente do efeito da ocorrência do óbito ao domingo no atraso de quatro dias na notificação dos óbitos. Os restantes dias de semana de ocorrência do óbito, relativamente ao domingo, diminuem a probabilidade da notificação do óbito dentro de quatro dias. Ser feriado no dia da previsão, ou seja, no dia em que é agregada a informação, volta a ser importante para explicar este atraso, mas diminuindo a probabilidade da notificação. Ser feriado no dia da morte e nos três dias seguintes também diminui a probabilidade de um óbito ser notificado dentro de quatro dias.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 4.166$ , com 8 graus de liberdade e um valor-p associado de 0.8418. O teste indica um bom ajustamento.

Os resíduos não apresentam valores extremos, indicando um bom ajustamento local do modelo.

Figura 5.8: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 4





**Modelo de previsão 5**

Tabela 5.7: Sumário do modelo de previsão 5

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.5965	0.1492	24.10	0.0000
Segunda <sup>a</sup>	0.1576	0.2115	0.75	0.4604
Terça <sup>a</sup>	-0.1936	0.1997	-0.97	0.3378
Quarta <sup>a</sup>	-0.9406	0.1772	-5.31	0.0000
Quinta <sup>a</sup>	-1.1016	0.1744	-6.32	0.0000
Sexta <sup>a</sup>	-1.1138	0.1748	-6.37	0.0000
Sábado <sup>a</sup>	-0.4897	0.1906	-2.57	0.0138
Feriado morte	-0.1709	0.2244	-0.76	0.4507
Feriado dia anterior previsão	-1.0201	0.1860	-5.49	0.0000
Feriado dois dias antes previsão	-0.8329	0.2043	-4.08	0.0002
Feriado três dias antes previsão	-1.0858	0.1871	-5.80	0.0000
Feriado quatro dias antes previsão	-1.1066	0.1749	-6.33	0.0000
Feriado no dia da previsão	-0.2259	0.2295	-0.98	0.3305
$\phi$ : 36.19589				
Null deviance: 9222.6 on 54 degrees of freedom				
Residual deviance: 1552.4 on 42 degrees of freedom				
<sup>a</sup> Referência:Domingo				

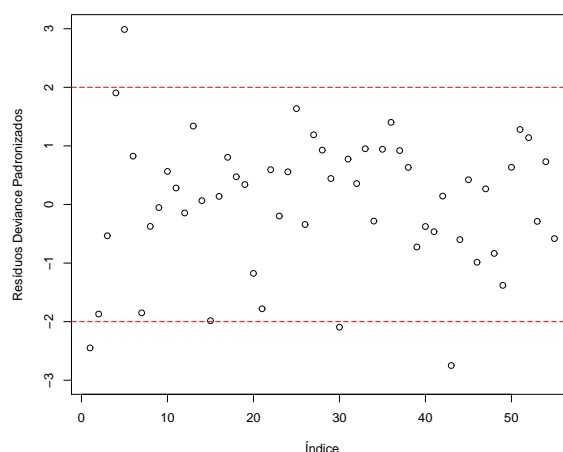
O sumário e os resíduos do modelo de previsão 5 estão apresentados na tabela 5.7 e na figura 5.9, respectivamente.

As variáveis que não mostraram ser significativas foram os dias de semana segunda-feira e terça-feira para o dia de ocorrência do óbito, ser feriado no dia de morte e ser feriado no dia de previsão. Ser feriado nos quatro dias anteriores ao dia de previsão diminui a probabilidade de um óbito ser notificado até cinco dias após a morte, tal como o dia de ocorrência da morte de quarta a sábado, em comparação com o domingo.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 9.0365$ , com 8 graus de liberdade e um valor-p associado de 0.3392. A hipótese nula não é rejeitada, indicando que o modelo está bem ajustado.

Os resíduos não apresentam valores extremos, indicando a inexistência de observações mal ajustadas pelo modelo.

Figura 5.9: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 5



### 5.2.2 Implementação da Metodologia

Nesta secção serão apresentados os resultados obtidos através da metodologia proposta para 2012, considerando as probabilidades estimadas através dos modelos apresentados na secção anterior. Como foi referido, estes modelos foram determinados considerando apenas a informação receccionada até 2011, e serão utilizadas para prever a mortalidade com o dados receccionados em 2012. O objetivo é testar a metodologia nas condições em que deverá ser utilizada, ou seja, tendo em conta as notificações dos óbitos que chegam diariamente e estão disponíveis em cada momento de previsão. As figuras identificadas de 5.10 a 5.16, apresentam as previsões para 2012 para cada dia de previsão, de  $t + 1$  a  $t + 5$ , a mortalidade observada, a linha de base para a mortalidade e os respetivos intervalos de confiança a 95% e a 99%.

As previsões em  $t + 1$ , calculadas considerando o modelo de previsão 1 e apresentadas na figura 5.10, apresentam diversos valores muito abaixo e acima da mortalidade observada. Entende-se que estas correspondem às estimativas obtidas sem informação, ou seja, estimativas obtidas com informação que é agregada ao fim-de-semana e em feriados nacionais. Portanto, dizem respeito às estimativas calculadas quando o número de notificações é nulo e, como tal, a estimativa é calculada assumindo que se recebeu uma notificação, como foi descrito no capítulo 3. Ao retirar apenas as estimativas efetuadas ao domingo e à segunda, informação apresentada na figura 5.11, confirma-se que as observações discrepantes correspondem essencialmente às previsões efetuadas sem

informação e que as previsões que não são efetuadas nestes dias estão muito próximas da mortalidade observada.

As previsões em  $t + 2$ , apresentadas na figura 5.12, são baseadas na informação recebida durante os dois dias seguintes à data de morte e considerando o modelo de previsão 2. Apresentam também bastantes valores muito abaixo da mortalidade observada. Entende-se que estas correspondem às estimativas obtidas sem informação, que para este momento de previsão dizem respeito a estimativas obtidas com informação que é agregada ao fim-de-semana e em feriados nacionais ou tolerâncias de ponto à sexta ou à segunda. Portanto, tal como em  $t + 1$ , dizem respeito às estimativas calculadas quando o número de notificações é nulo. Ao retirar apenas as previsões efetuadas à segunda-feira, ou seja, previsões obtidas com informação agregada ao sábado e domingo, confirma-se que as observações discrepantes correspondem essencialmente às previsões efetuadas sem informação, como se pode verificar na figura 5.13. Também se verifica, mais uma vez, que as previsões efetuadas nos restantes dias estão muito próximas da mortalidade observada.

As previsões em  $t + 3$  estão apresentadas no gráfico 5.14 e são baseadas na informação recebida durante os três dias seguintes à data de morte e no modelo de previsão 3. Observam-se algumas estimativas muito abaixo da mortalidade observada, que analisando mais detalhadamente, correspondem aos dias em que existiram feriados antes ou depois de fins-de-semana, levando a três dias sem informação. Os dias detetados foram: 6 de Abril, 5 de Outubro e os dias 22, 23, 29 e 30 de Dezembro. Uma situação extraordinária observou-se no dia 19 de Maio, que conta com informação agregada de sábado a segunda, mas apenas foram recebidas 19 notificações. É de notar que a mediana das notificações em  $t + 3$ , das mortes ocorridas ao sábado em 2012, foi de 204 óbitos.

As previsões em  $t + 4$  e  $t + 5$  estão apresentadas nos gráficos 5.15 e 5.16, respetivamente. As estimativas que se observam muito abaixo da mortalidade observada dizem respeito a algumas das observações discrepantes detetadas em  $t + 3$ ,  $t + 2$  e  $t + 1$ . Como tal, estas correspondem a quatro dias seguidos sem informação ou com informação irregular.

Através da implementação da metodologia para o ano de 2012 verifica-se que as estimativas efetuadas sem informação ou com informação irregular estão longe da mor-

talidade observada para o dia correspondente, mas as restantes estimativas mostram-se muito próximas da mortalidade observada.

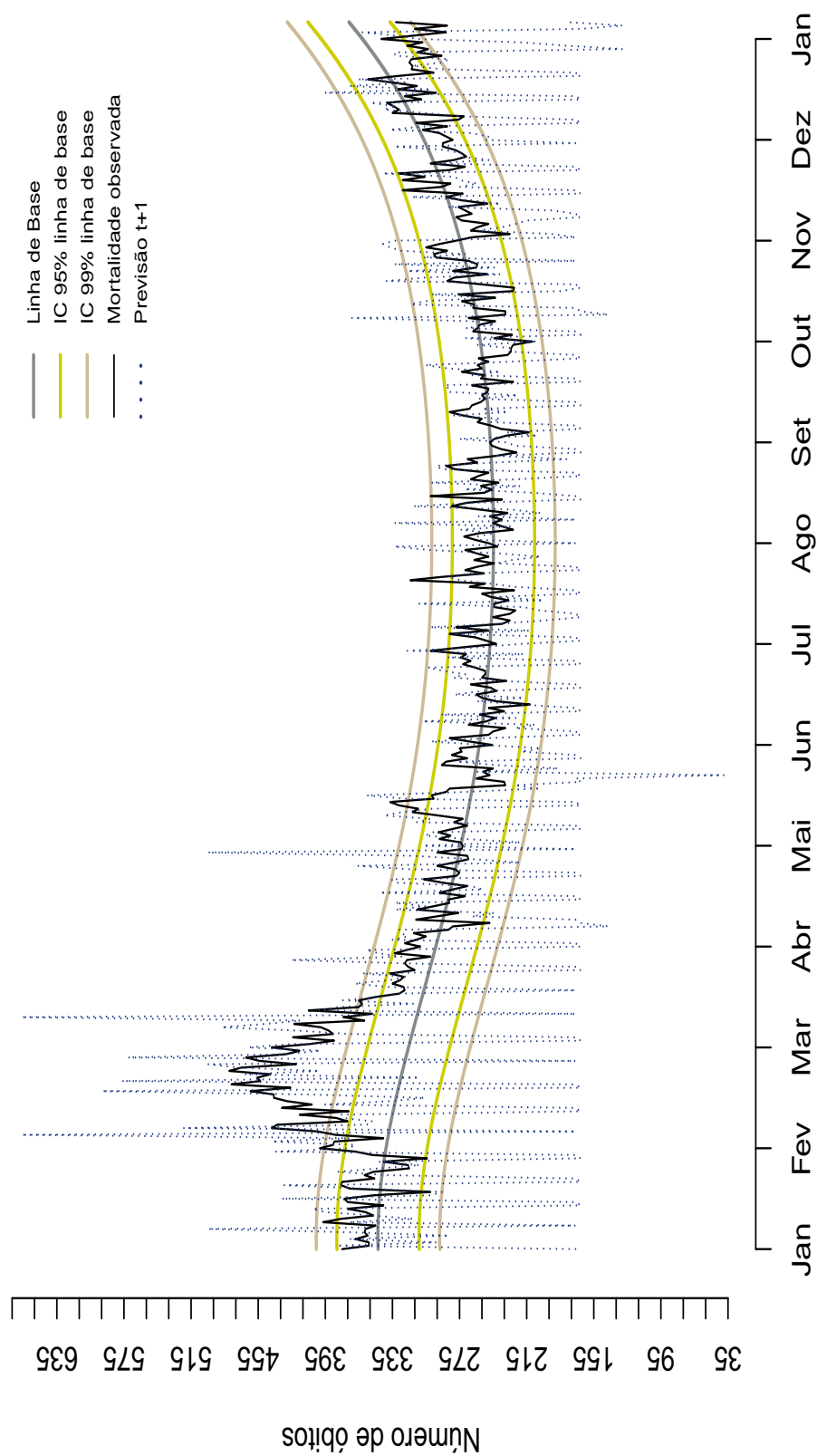
Figura 5.10: Previsões em  $t + 1$  para 2012

Figura 5.11: Previsões em  $t + 1$  para 2012, excluindo as previsões efetuadas ao domingo e à segunda-feira

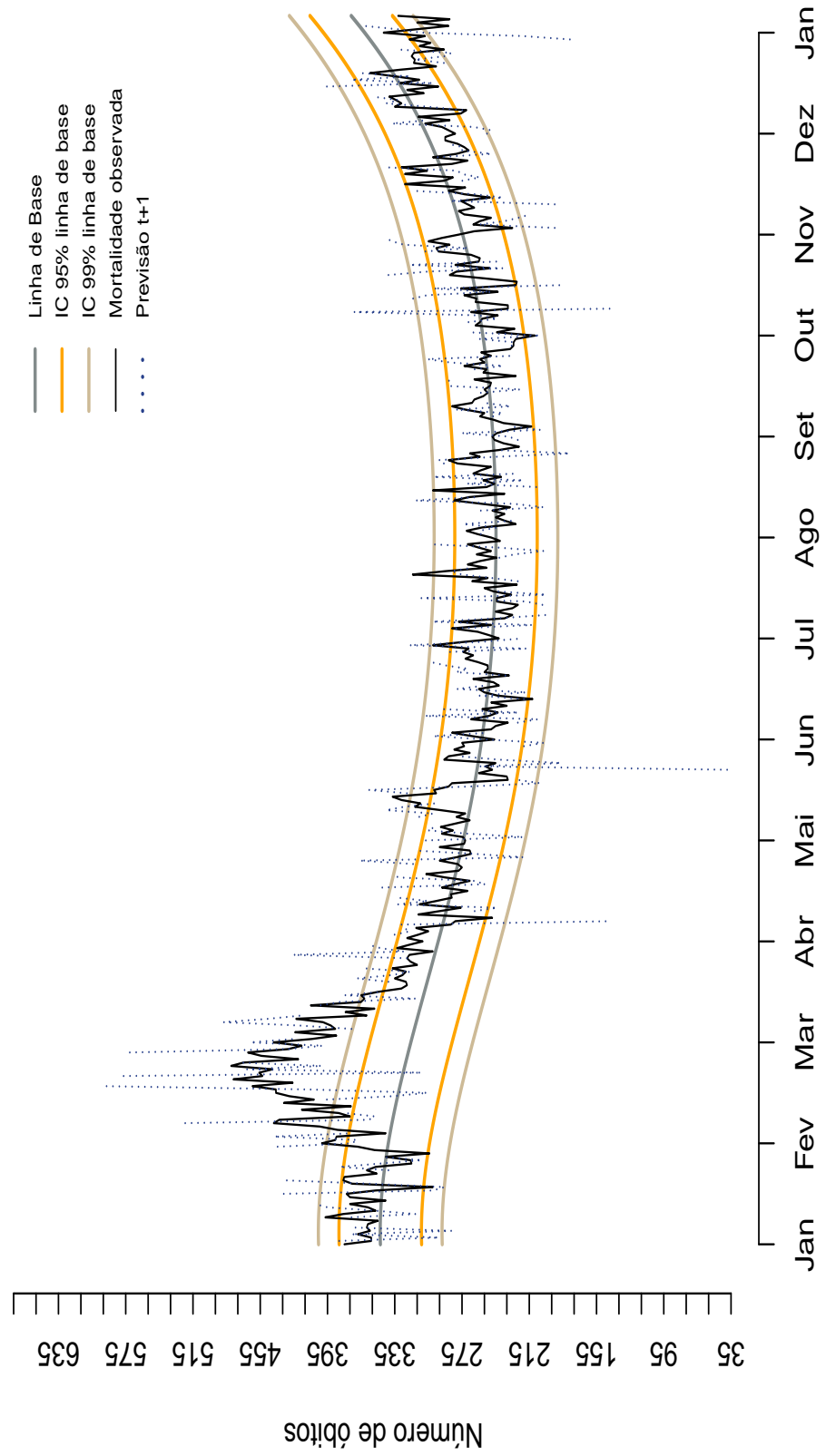


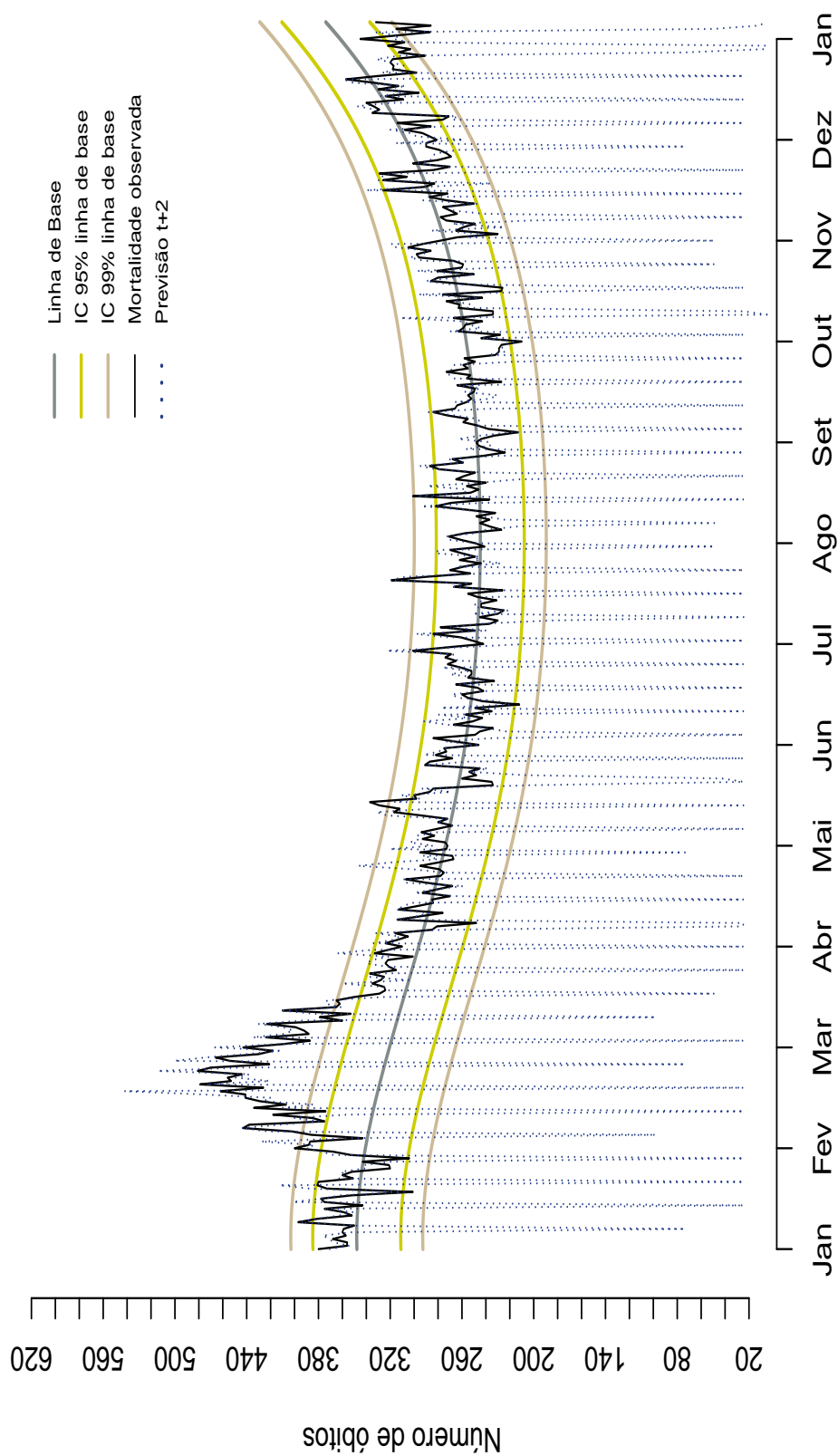
Figura 5.12: Previsões em  $t + 2$  para 2012

Figura 5.13: Previsões em  $t + 2$  para 2012, excluindo as previsões efetuadas à segunda-feira

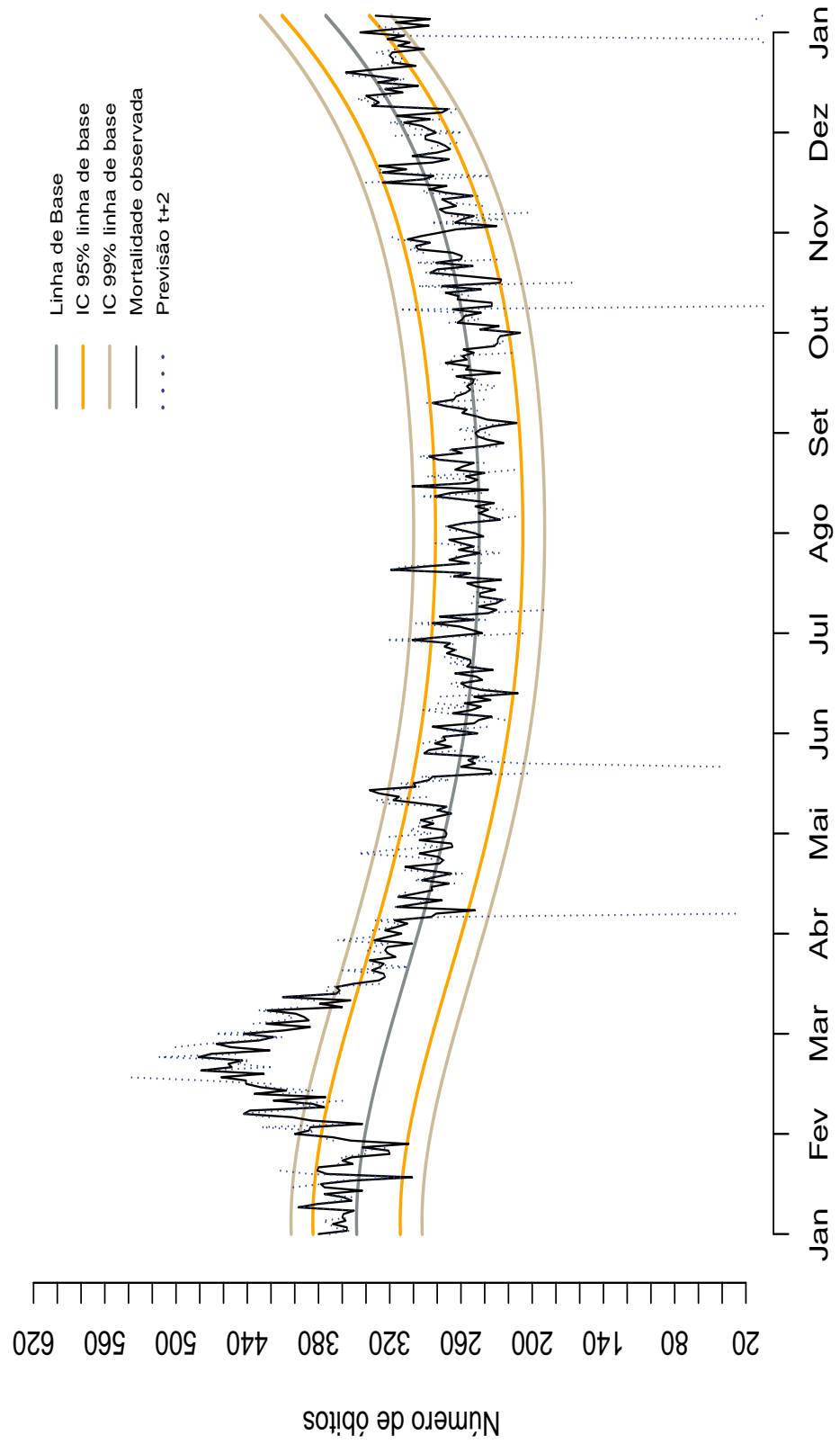




Figura 5.14: Previsões em  $t + 3$  para 2012

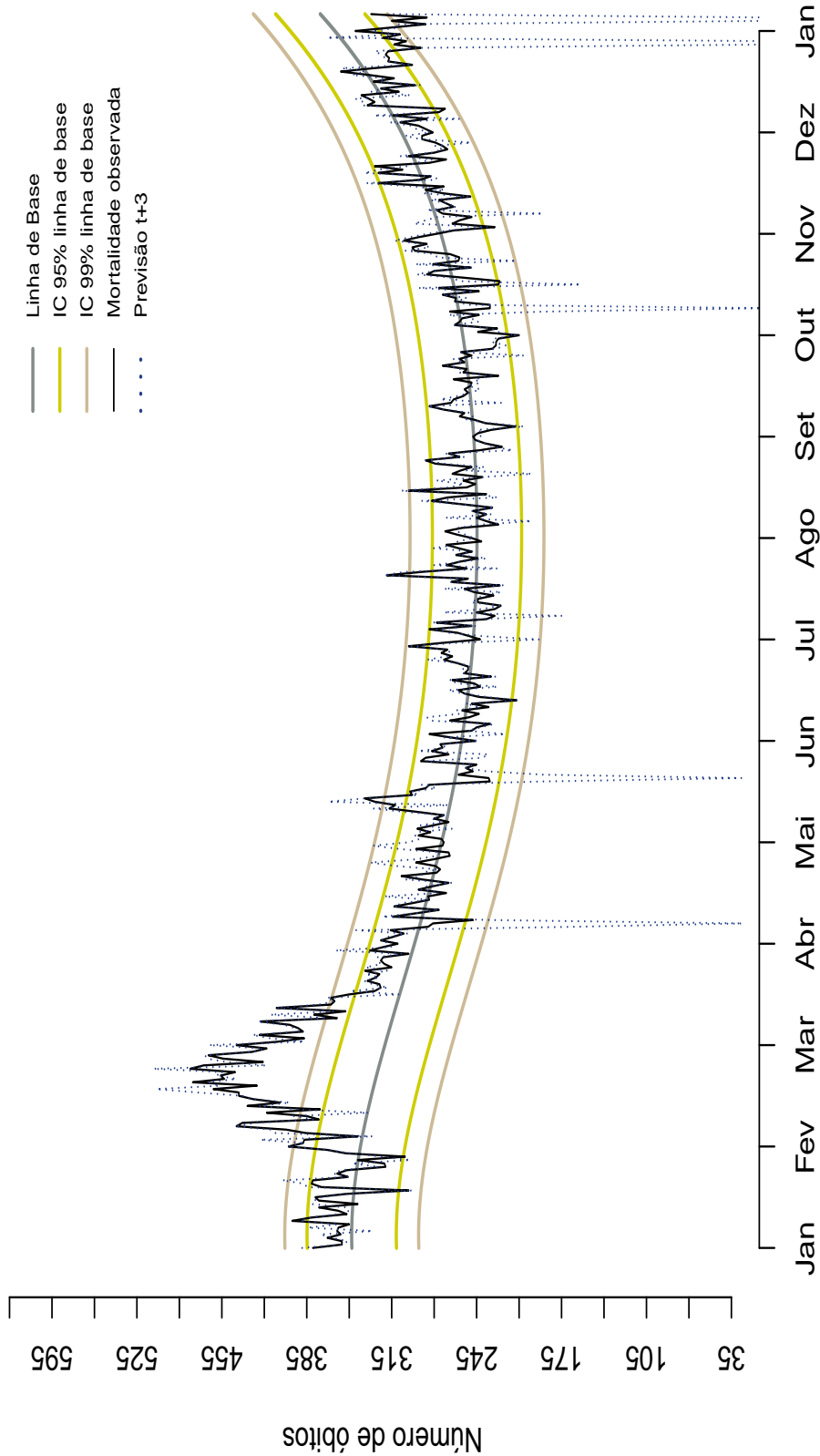


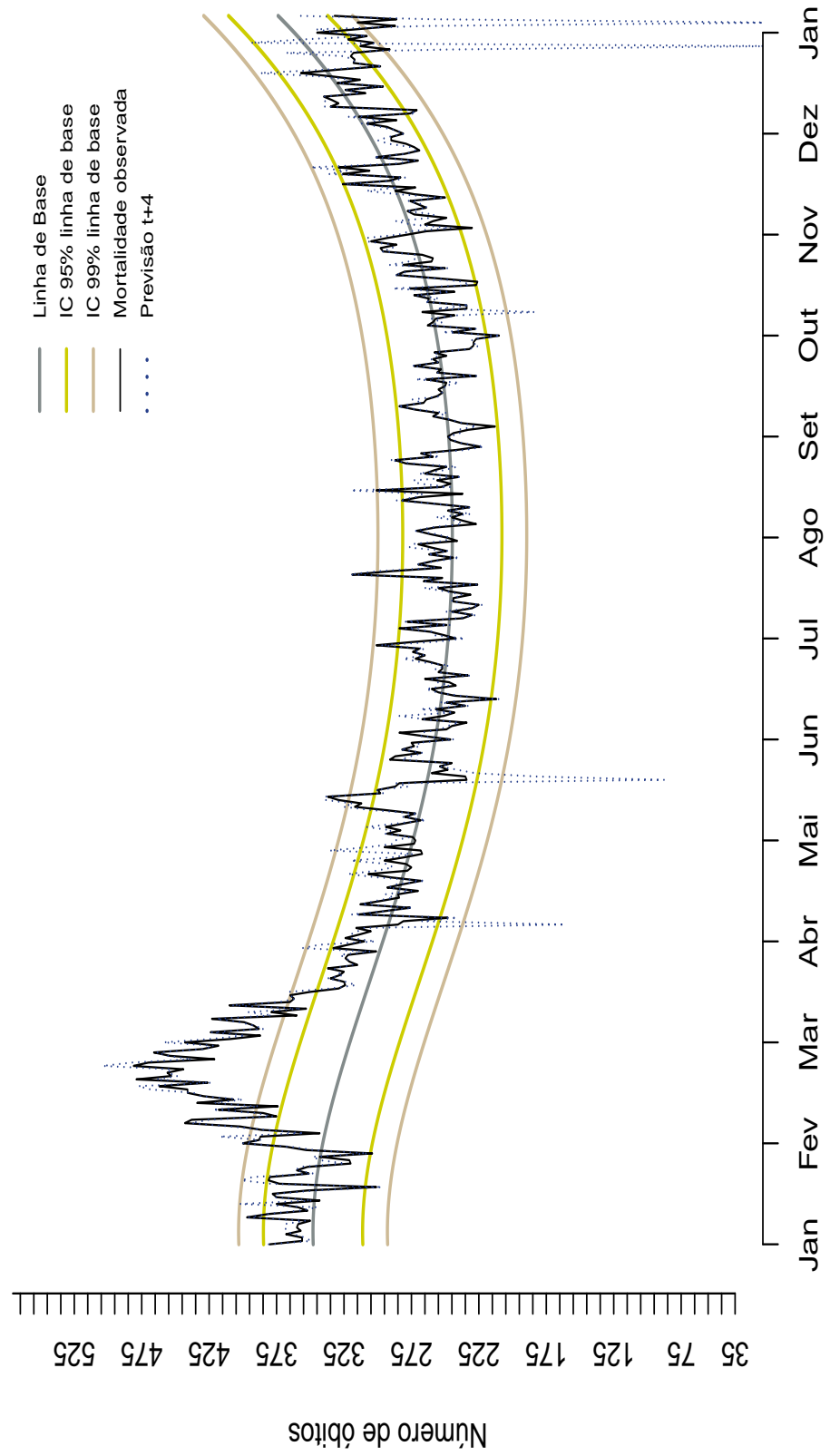
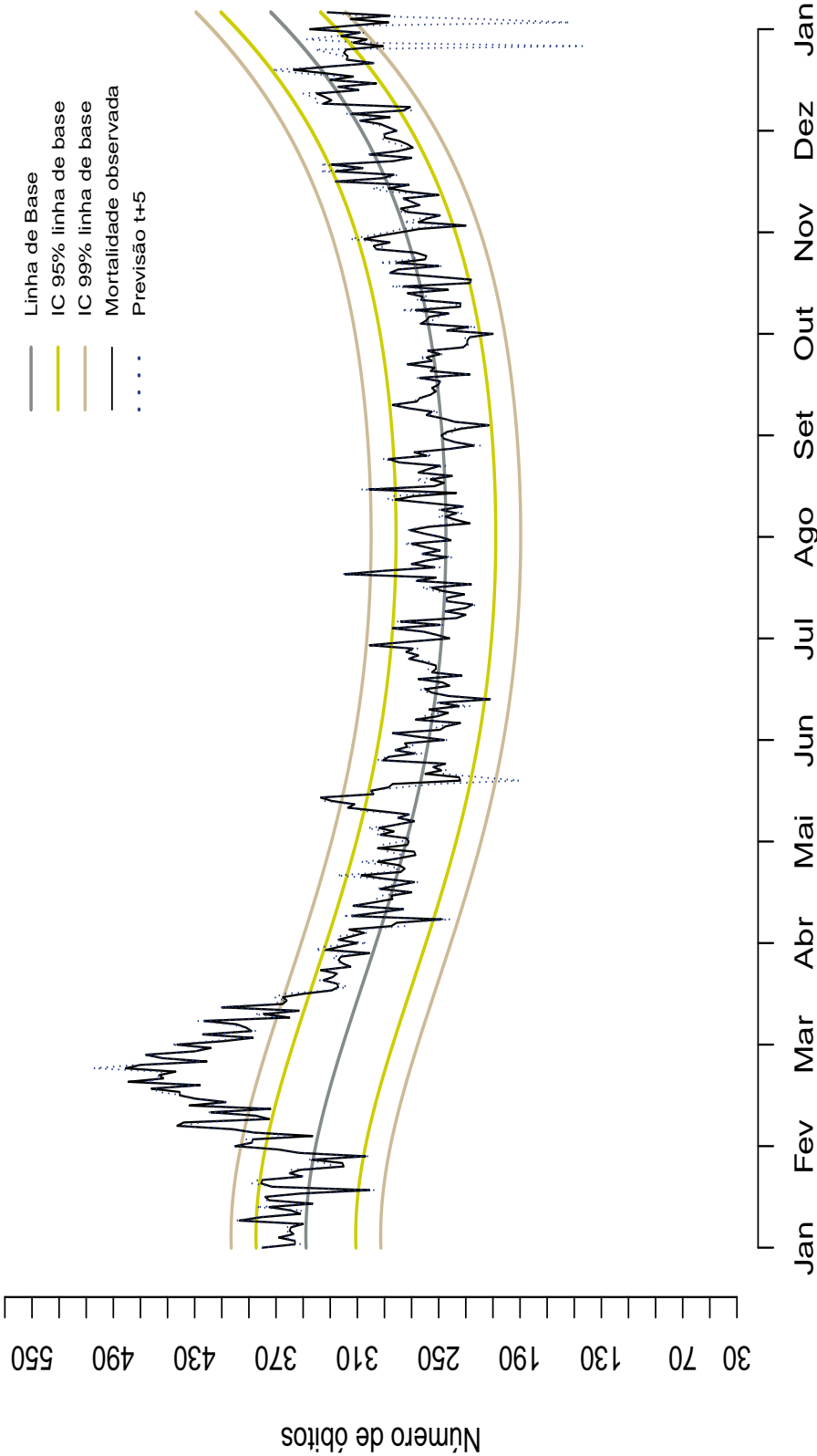
Figura 5.15: Previsões em  $t + 4$  para 2012

Figura 5.16: Previsões emt + 5 para 2012



### 5.2.3 Avaliação da Metodologia

Para poder avaliar o desempenho da metodologia, as medidas foram calculadas para os vários cenários relativos ao número de dias em que não houve receção de informação, como foi descrito no capítulo 3. Por isso, é importante analisar a frequência de cada cenário para este ano 2012. Esta informação está disponível na tabela 5.8 que apresenta a frequência de cada cenário, ou seja, o número de previsões efetuadas em cada um dos cenários relativos à receção da informação.

Tabela 5.8: Número de previsões efetuadas para cada cenário de receção da informação e frequência relativa (%), para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , no ano de 2012

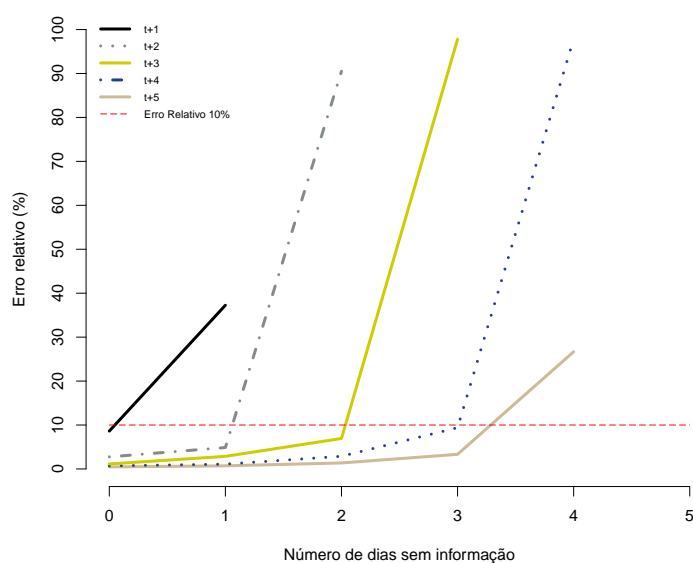
	Informação completa	Número de dias sem informação				
		1 dia	2 dias	3 dias	4 dias	5 dias
$t + 1$	251 (68.6)	115 (31.4)	-	-	-	-
$t + 2$	194 (53.0)	115 (31.4)	57 (15.6)	-	-	-
$t + 3$	140 (38.3)	114 (31.2)	107 (29.2)	5 (1.4)	-	-
$t + 4$	90 (24.6)	108 (29.5)	156 (42.6)	11 (3.0)	1 (0.3)	-
$t + 5$	44 (12.0)	98 (26.8)	201 (54.9)	21 (5.7)	2 (0.6)	0

Os valores apresentados na diagonal da tabela 5.8 representam as previsões efetuadas sem qualquer informação. Durante o ano 2012, não existiu nenhum período em que não se tenha recebido informação durante 5 dias. Nas previsões com 1 dia de atraso, 31.4% das estimativas foram efetuadas sem informação e 68.6% com informação completa que, para este dia de previsão, corresponde a um dia de informação. Nas previsões até 2 dias de atraso, 15.6% das estimativas foram efetuadas sem informação mas, a partir deste dia de previsão, as estimativas efetuadas sem qualquer informação ocorrem com uma frequência relativa igual ou inferior a 1.4%. Até 3 dias de atraso, o mais frequente foi a produção de estimativas com informação completa, e em seguida com pelo menos um dia de informação. Após 4 dias de atraso, o mais frequente é a produção de estimativas considerando dois dias sem informação, em seguida considerando 1 dia sem informação e por fim com informação completa. 39.3% das previsões foram efetuadas com informação completa, 30.1% com um dia sem informação e 35.6% com dois dias sem informação.

Tabela 5.9: Erro relativo mediano (%) das previsões para cada cenário de receção de informação para 2012

Erro Relativo mediano (%)						
Previsão	Informação Completa	Número de dias sem informação				
		1 dia	2 dias	3 dias	4 dias	5 dias
$t + 1$	8.62	36.25	—	—	—	—
$t + 2$	2.75	4.88	90.50	—	—	—
$t + 3$	1.37	2.86	6.93	97.80	—	—
$t + 4$	0.70	1.09	2.89	9.38	97.87	—
$t + 5$	0.46	0.71	1.36	3.33	26.68	0

Figura 5.17: Erro das previsões por número de dias sem informação



### Erro relativo das previsões

A tabela 5.9 e a figura 5.17 apresentam o erro relativo mediano percentual para cada previsão efetuada em  $t + i$ , com  $i = 1, \dots, 5$ , e para cada cenário de receção de informação.

Analisando a tabela e o gráfico referidos acima, o erro relativo das previsões variou

em função do número de dias que se recebeu informação.

O erro relativo mediano das previsões a  $t + 1$  foi inferior a 10% se houve receção de informação, mas sem informação, o erro das previsões passou para 36.25%. Portanto, previsões efetuadas em  $t + 1$ , de terça-feira a sábado, excluindo os dias em que o dia da morte é um feriado nacional, apresentam um erro relativo mediano percentual inferior a 10%.

Verificou-se o mesmo resultado para os restantes momentos de previsão, ou seja, quando houve pelo menos um dia em que foi rececionada informação, as estimativas obtidas apresentaram um erro relativo mediano percentual inferior a 10%, com exceção as previsões em  $t + 5$ . Em  $t + 5$ , o erro associado às previsões efetuadas com apenas um dia de informação apresentou um erro relativo mediano percentual de 26.68%. Este resultado é claro analisando a figura 5.17, pois apenas as estimativas efetuadas sem informação apresentam erros superiores a 10%, visto que são as únicas que estão acima da linha que indica um erro de 10%, com exceção da previsão em  $t + 5$ .

Quando não existiu qualquer informação, as estimativas apresentaram erros muito superiores a 10%.

**Sensibilidade e Especificidade**Tabela 5.10: Sensibilidade e Especificidade das previsões em  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de recepção de informação, para 2012

Previsão	Número de dias sem informação	Sensibilidade	Especificidade
$t + 1$	Informação completa	0.71	0.83
	1 dia	0.20	0.94
$t + 2$	Informação completa	1.00	0.97
	1 dia	0.67	0.95
	2 dias	0	1.00
$t + 3$	Informação completa	1.00	1.00
	1 dia	1.00	0.97
	2 dias	0.79	0.96
	3 dias	0	1.00
$t + 4$	Informação completa	1.00	1.00
	1 dia	1.00	1.00
	2 dias	1.00	0.95
	3 dias	0	1.00
	4 dias	—	1.00
$t + 5$	Informação completa	1.00	1.00
	1 dia	1.00	1.00
	2 dias	1.00	0.98
	3 dias	0	1.00
	4 dias	0	1.00
	5 dias	—	—

A tabela 5.10 apresenta os valores da sensibilidade e especificidade para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de recepção de informação. Em 2012 foram observados 52 dias com excesso de mortalidade, considerando a definição de excesso em 2.2.2, correspondendo a 4 períodos de excesso de mortalidade.

Analisando a referida tabela, a sensibilidade e a especificidade também variaram em função do número de dias que se recebeu informação, tal como o erro relativo mediano das previsões. A probabilidade de um excesso de mortalidade ser detetado no dia seguinte foi de 0.71 se se tiver recebido informação. O que nos indica que a probabilidade de ocorrerem falsos negativos no dia seguinte, ou seja, de um excesso

não ser detetado, se se tiver recebido informação, é de 0.29. Se não se tiver recebido informação, o VPP passa para 0.20.

Após dois dias, a probabilidade de se detetar um excesso de mortalidade foi de 1 se se recebeu informação nos dois dias seguintes à data que se quer estimar. Com apenas um dia de informação, esta passa a ser de 0.67.

A partir de  $t + 2$ , sem receção de informação, a sensibilidade é nula, ou seja, a metodologia não tem capacidade para detetar o excesso de mortalidade.

A probabilidade de se detetar um dia sem excesso de mortalidade é sempre superior a 0.83, seja qual for a quantidade de dias que se recebeu informação.

### **Valor Preditivo Positivo e Negativo**

A tabela 5.11 apresenta o Valor Preditivo Positivo (VPP) e Negativo (VPN) para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de receção de informação.

Os Valores Preditivo Positivo e Negativo também mostraram variar em função do número de dias que se recebeu informação. A probabilidade de um excesso de mortalidade detetado pela metodologia em  $t + 1$  ser verdadeiro é de 0.43, se se tiver recebido informação. Se não se tiver recebido informação, esta passa a ser de 0.33.

Após dois dias, a probabilidade de ser detetado um verdadeiro excesso de mortalidade é de 0.86 se se recebeu informação nos dias seguintes à data de ocorrência. Com apenas um dia de informação, esta passa a ser de 0.66. Sem informação, a metodologia não detetou nenhum excesso em  $t + 2$ .

Após três dias, se se recebeu informação todos os dias, a probabilidade de ser detetado um verdadeiro excesso é de 1, com apenas dois dias de informação, esta passa a ser de 0.86.

A probabilidade de a metodologia detetar um verdadeiro dia sem excesso de mortalidade foi sempre superior a 0.87 para qualquer dia de previsão, mesmo sem se ter tido acesso a informação.



Tabela 5.11: Valor Preditivo Positivo (VPP) e Valor Preditivo Negativo (VPN) por número de dias sem informação e por dia de previsão  $t + i$ , com  $i = 1, \dots, 5$ 

Previsão	Número de dias sem informação	VPP	VPN
$t + 1$	Informação completa	0.43	0.96
	1	0.33	0.89
$t + 2$	Informação completa	0.86	1.00
	1 dia	0.66	0.95
	2 dias	0	0.88
$t + 3$	Informação completa	1.00	1.00
	1 dia	0.85	1.00
	2 dias	0.65	0.97
	3 dias	—	1.00
$t + 4$	Informação completa	1.00	1.00
	1 dia	1.00	1.00
	2 dias	0.80	1.00
	3 dias	—	1.00
	4 dias	—	1.00
$t + 5$	Informação completa	1.00	1.00
	1 dia	1.00	1.00
	2 dias	0.91	1.00
	3 dias	—	1.00
	4 dias	—	1.00
	5 dias	—	—

**Timeliness na deteção de um período de excesso de mortalidade**

A tabela 5.12 apresenta os óbitos notificados e estimados pela metodologia, em cada momento de previsão (representados pelo dia em que seria efetuada a previsão e, da mesma forma, que seriam rececionadas as notificações e o respetivo dia de semana), para o primeiro dia de excesso e para os dias adjacentes. O mesmo é apresentado para o último dia de excesso e os dias adjacentes. A referida tabela tem como objetivo analisar o *timeliness* na deteção de excesso de mortalidade e da mesma forma, analisar o tempo até à deteção do final do excesso de mortalidade.

As figuras identificadas de 5.18 a 5.22 apresentam as previsões e as notificações em cada momento  $t + i$ , com  $i = 1, \dots, 5$ , para o período de epidemia de gripe em

2012, a linha de base para a mortalidade e os respectivos limites superiores dos intervalos de confiança a 95% e a 99%. Uma observação importante é retirada das figuras 5.10 a 5.15, que nos confirmam mais uma vez que os valores discrepantes estimados pela metodologia correspondem aos dias em que as notificações foram nulas ou irregulares.

O primeiro excesso de mortalidade durante a epidemia de gripe em 2012 ocorreu no dia 4 de Fevereiro de 2012. O valor observado da mortalidade para este dia foi de 386 óbitos, superior ao valor do limite superior do intervalo de confiança a 95%, mas abaixo do limite inferior do intervalo de confiança (LS IC) a 99%. Como foi referido, um período de excesso de mortalidade inicia-se com dois dias consecutivos da mortalidade acima do valor do LS IC a 95% (regra  $2_{2s}$ ), ou um dia acima do valor do LS IC a 99% (regra  $1_{3s}$ ).

Como se pode verificar na tabela 5.12, através do sistema VDM atual (sem correção para os atrasos), o excesso durante este período só é detetado relativamente ao dia 6 de Fevereiro, três dias depois, dia 9 de Fevereiro. Na prática, este seria o primeiro excesso detetado. Analisando os gráficos 5.18, 5.19 e 5.20, pode confirmar-se que é em  $t + 3$  que é detetada a primeira observação acima de pelo menos um dos limites superiores dos intervalos de confiança a 95% e a 99%.

A deteção do início do período de excesso de mortalidade, dia 4 de Fevereiro, depende da observação do dia seguinte visto que o valor da mortalidade está abaixo do LS IC 99%. Portanto, através da regra  $2_{2s}$ , são necessários dois dias consecutivos de mortalidade observada acima do LS IC 95% para a mortalidade esperada. Apesar do excesso para o dia 5 de Fevereiro (dia seguinte ao início da mortalidade) ser detetado também dia 9 de Fevereiro, o excesso do dia anterior (início da mortalidade), só é detetado 10 dias depois, dia 14 de Fevereiro. Portanto, o início do período de excesso de mortalidade, que ocorre dia 4 de Fevereiro, só é detetado 10 dias depois através do sistema VDM atual. Analisando a figura 5.21, pode confirmar-se que a primeira observação acima do LS IC a 95% correspondente à segunda observação necessária para o início do excesso de mortalidade, surge apenas em  $t + 4$ .

Através da metodologia, o primeiro excesso detetado, ignorando as previsões efetuadas com informação agregada ao fim-de-semana, corresponde ao início de excesso de mortalidade. Simultaneamente também deteta o excesso observado para o dia seguinte. O início do período de excesso de mortalidade, dia 4 de Fevereiro, é detetado três dias

## CAPÍTULO 5. RESULTADOS

---

depois, dia 7 de Fevereiro, com 410 óbitos (ER=6.22%). O excesso de mortalidade relativo a dia 5 de Fevereiro é detetado dois dias depois, também dia 7 de Fevereiro, com 421 óbitos (ER=4.98%).

O tempo salvo até à deteção do início do período de excesso foi de 7 dias, definindo o *timeliness* em 7 dias ou em 0.7. Até à deteção do primeiro excesso detetado pelas notificações, o tempo salvo foi de 2 dias.

O ultimo dia de excesso de mortalidade é observado a 16 de Março e é detetado pela metodologia no dia seguinte. O primeiro dia sem excesso de mortalidade é observado a 17 de Março mas apenas é detetado quatro dias depois através da metodologia. Na prática, é a dia 20 de Março que a metodologia deteta o final do excesso, para dia 19 de Março.

## CAPÍTULO 5. RESULTADOS

Tabela 5.12: Número de notificações (N) e número de óbitos previstos para o início e fim do período de excesso de mortalidade

Início do excesso de mortalidade								
	3 Feb 2012 (Sexta)		<b>4 Feb 2012 (Sábado)</b>		5 Feb 2012 (Domingo)		6 Feb 2012 (Segunda)	
Dia	Previsão	N	Previsão	N	Previsão	N	Previsão	N
4 Feb (sáb)	337	101	—	—	—	—	—	—
5 Feb (dom)	352	106	666	4	—	—	—	—
6 Feb (seg)	329	106	98	4	171	0	—	—
7 Feb (ter)	354	283	<b>410</b>	294	<b>421</b>	289	521	95
8 Feb (qua)	344	329	383	347	414	375	444	340
9 Feb (qui)	341	331	384	368	408	<b>393</b>	443	<b>414</b>
10 Feb (sex)	341	331	383	371	405	<b>395</b>	438	425
11 Feb (sab)	343	334	382	372	404	395	438	428
12 Feb (dom)	343	334	382	372	405	395	439	428
13 Feb (seg)	343	334	382	373	404	395	440	428
14 Feb (ter)	343	334	385	<b>377</b>	406	398	440	429
LS IC 95% ; 99%	375 ; 394		374 ; 394		375 ; 393		374 ; 393	
Total de óbitos	343		386		402		443	

Timeliness para o início do excesso=0.7

Fim do excesso de mortalidade								
	16 Mar 2012 (Sábado)		<b>17 Mar 2012 (Domingo)</b>		18 Mar 2012 (Segunda)		19 Mar 2012 (Terça)	
Dia (dia da semana)	Previsão	N	Previsão	N	Previsão	N	Previsão	N
17 Mar (dom)	347	94	—	—	—	—	—	—
18 Mar (seg)	326	98	333	2	—	—	—	—
19 Mar (ter)	307	99	49	2	171	1	—	—
20 Mar (qua)	365	292	347	249	<b>323</b>	222	<b>340</b>	62
21 Mar (qui)	364	336	<b>335</b>	303	317	288	325	249
22 Mar (sex)	357	343	332	318	316	304	330	308
LS IC 95% ; 99%	345 ; 364		345 ; 363		344 ; 363		343 ; 364	
Total de óbitos	350		329		324		337	

Figura 5.18: Previsões e notificações a  $t+1$  durante a Epidemia de Gripe em 2012

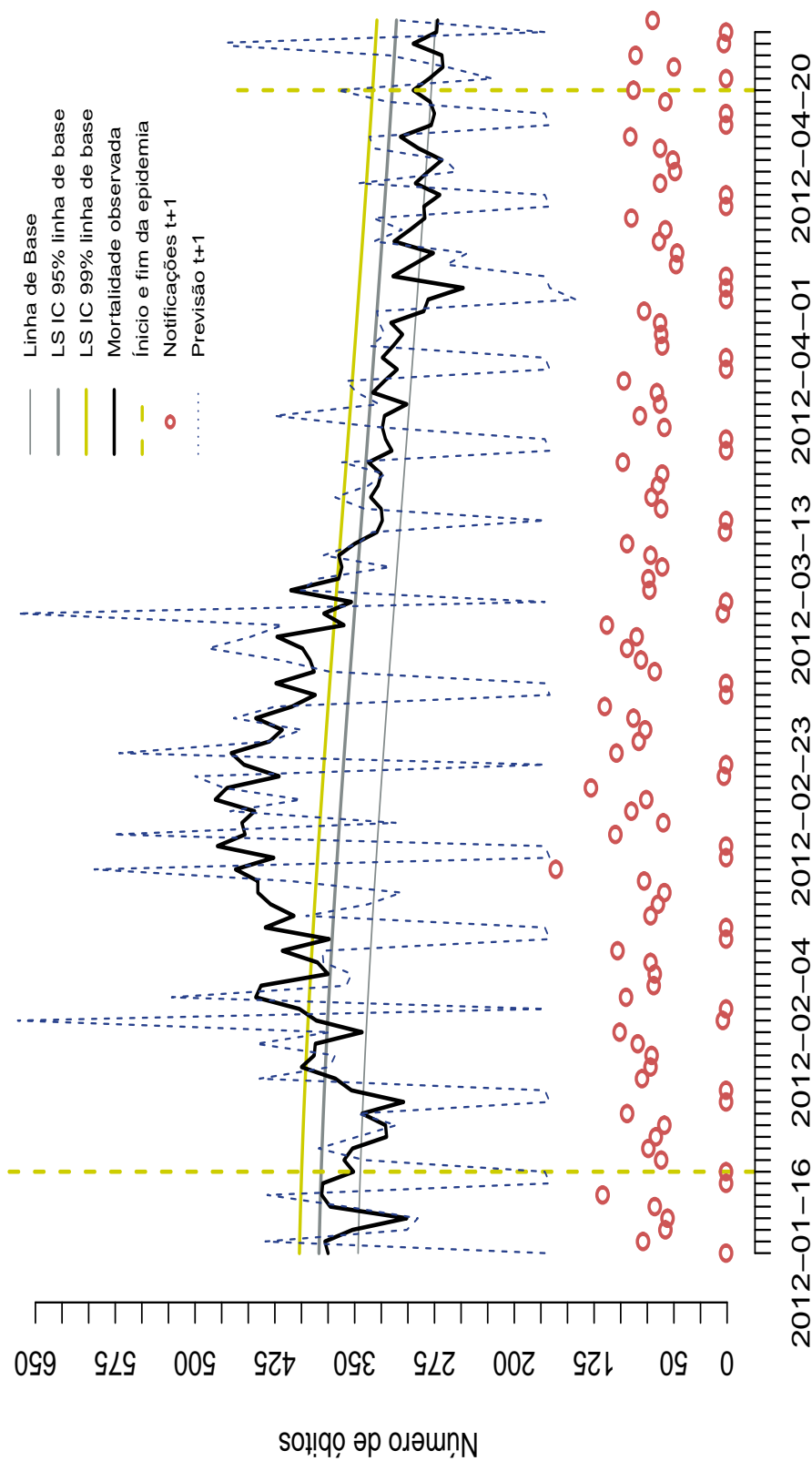


Figura 5.19: Previsões e notificações a t+2 durante a Epidemia de Gripe em 2012

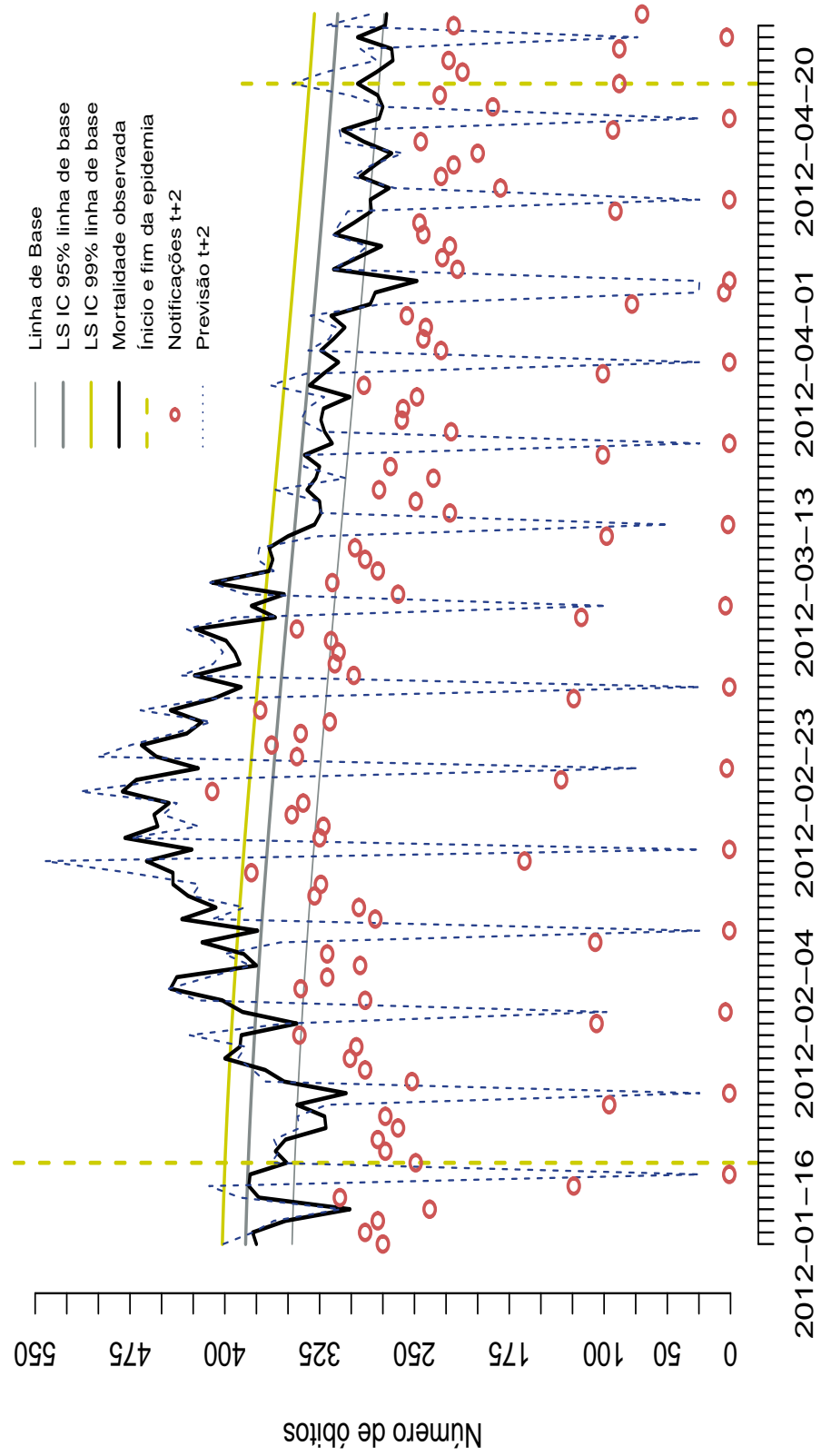


Figura 5.20: Previsões e notificações a t+3 durante a Epidemia de Gripe em 2012

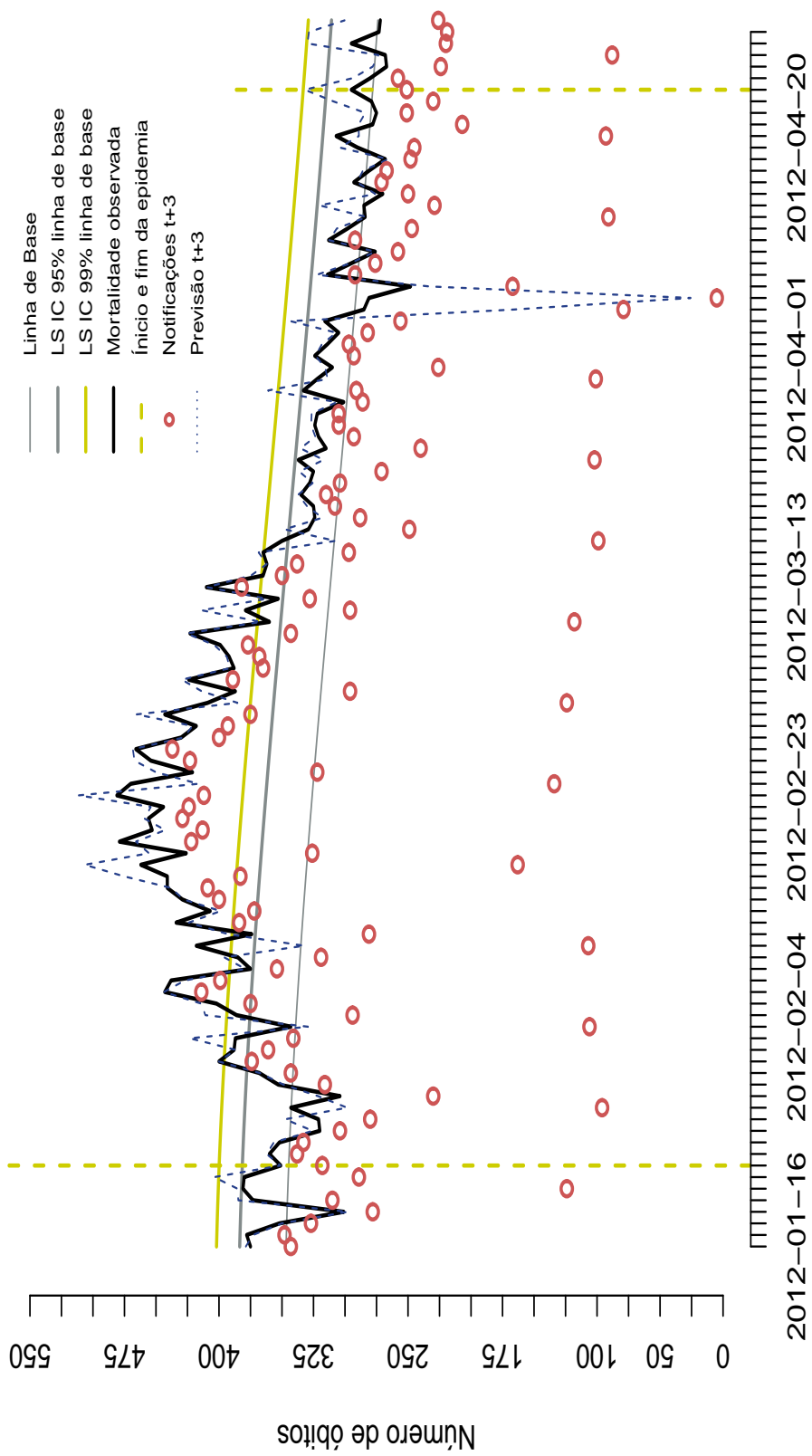


Figura 5.21: Previsões e notificações a t+4 durante a Epidemia de Gripe em 2012

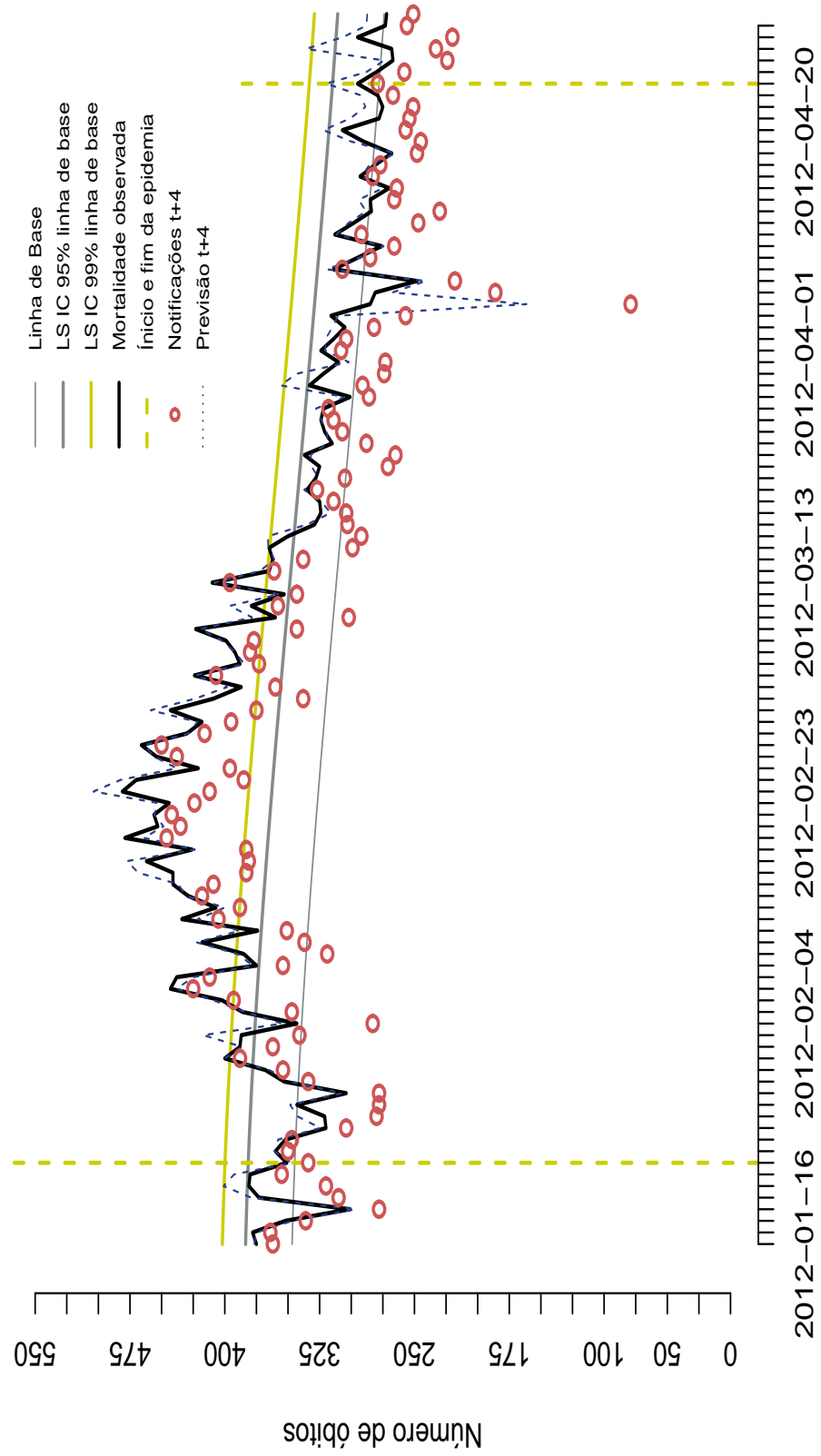
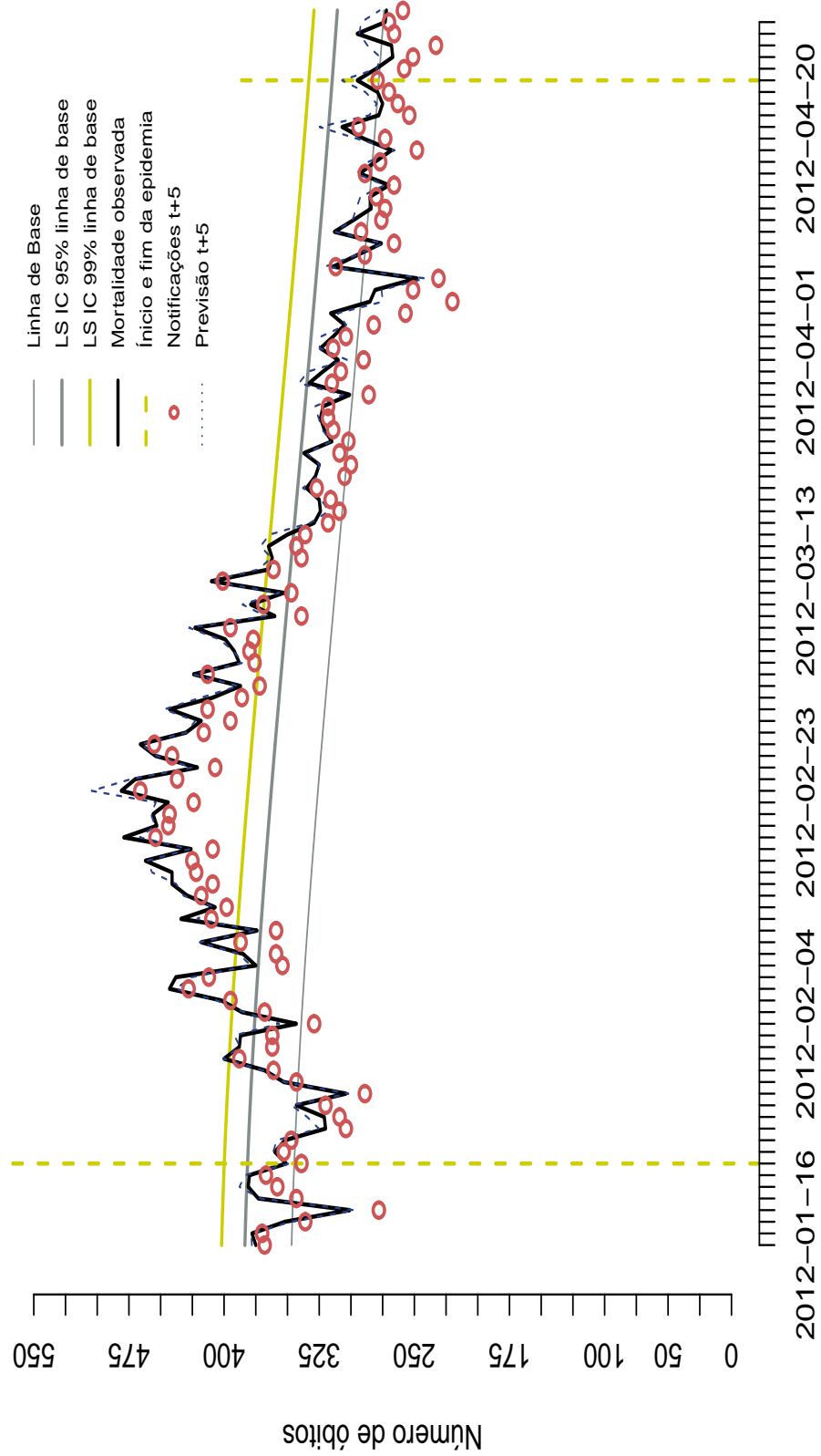




Figura 5.22: Previsões e notificações a t+5 durante a Epidemia de Gripe em 2012



### 5.3 Operacionalização e avaliação da metodologia para 2013

Para avaliar o desempenho da metodologia para o ano de 2013, os modelos foram estimados com base nos dados recebidos até 2012, compreendendo um total de 479669 observações. Os valores obtidos para as medidas descritivas dos atrasos não parecem diferir substancialmente dos valores obtidos com a informação completa, ou apenas com a informação recepcionada até 2011, apresentada para a operacionalização e avaliação da metodologia para 2012. Os valores que sofreram alterações foram aqueles dependentes dos atrasos mais longos, tal como nos dados recepcionados até 2011, visto que se perdeu acesso a informação mais recente.

Tabela 5.13: Medidas descritivas do atraso na notificação até 2012

mínimo	máximo	média	desvio-padrão	variância
1	2145	5,648	46.66	2176.86

Percentis						
50%	75%	80%	90%	95%	98%	99%
2	3	4	5	6	10	36

#### 5.3.1 Estimação das probabilidades dos atrasos

Nesta etapa serão apresentados os sumários dos modelos para estimar cada probabilidade acumulada incluindo os dados de 2012, e os resultados da avaliação dos mesmos. O óbito ter ocorrido durante um evento relacionado com a saúde, onda de calor ou epidemia de gripe, não mostrou ser significativo para explicar cada atraso. Os modelos finais foram ajustados excluindo estas variáveis.

As covariáveis comuns a todos os modelos foram:

- dia de semana de ocorrência do óbito
- ser feriado no dia de ocorrência do óbito
- ser feriado no dia da previsão
- ser feriado nos últimos  $k$  dias anteriores à previsão  $i$ ,  $k = 1, \dots, 4$ . Para  $i < 5$ ,  $k = 1, \dots, i - 1$

**Modelo de previsão 1**

Tabela 5.14: Sumário do modelo de previsão 1

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	-5.3990	0.2621	-20.60	0.0000
Segunda <sup>a</sup>	3.8985	0.2660	14.66	0.0000
Terça <sup>a</sup>	3.9528	0.2659	14.86	0.0000
Quarta <sup>a</sup>	3.9641	0.2659	14.91	0.0000
Quinta <sup>a</sup>	3.9503	0.2660	14.85	0.0000
Sexta <sup>a</sup>	4.4203	0.2652	16.67	0.0000
Sábado <sup>a</sup>	0.1930	0.3562	0.54	0.5987
Feriado morte	-4.1686	0.6169	-6.76	0.0000
Feriado no dia seguinte ao óbito	0.3217	0.1066	3.02	0.0117
$\phi$ : 21.1908				
Null deviance: 51639.2 on 19 degrees of freedom				
Residual deviance: 149.1 on 11 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 1 e os respectivos resíduos estão apresentados na tabela 5.14 e na figura 5.23, respetivamente.

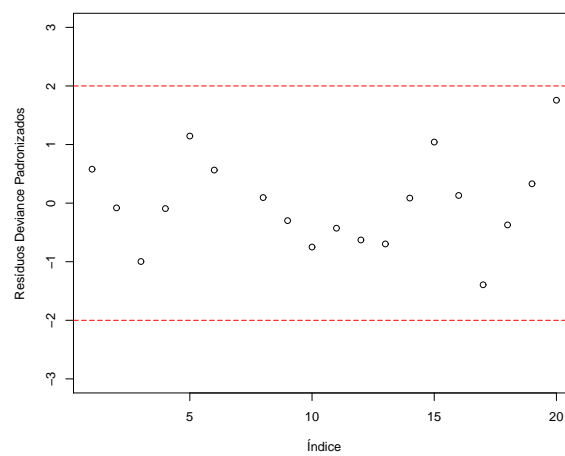
Todas as variáveis mostraram ser significativas, com exceção do sábado como dia de semana de ocorrência do óbito. O mesmo resultado foi observado para o modelo de previsão 1 considerando apenas os dados até 2011 indicando, mais um vez, que o dia de ocorrência do óbito ao sábado não difere significativamente do domingo. A ocorrência do óbito de segunda a sexta, tal como no teste para 2012, aumenta a probabilidade de um óbito ser notificado no dia seguinte à data de morte comparativamente com a ocorrência do óbito ao fim-de-semana. A sexta como dia de morte é o fator que mais aumenta a probabilidade de um óbito ser notificado no dia seguinte, relativamente ao domingo, para este atraso. O mesmo resultado é observado para a variável ser feriado no dia da morte e ser feriado no dia de previsão, o primeiro diminuindo a probabilidade do óbito ser notificado, e o segundo aumentando esta probabilidade. A única diferença observada entre o presente teste e o teste para 2012, reside nos valores dos coeficientes, mas estes não variam substancialmente (superior a 10 %), com exceção do sábado como dia de morte. O valor do coeficiente associado no teste anterior é de 0.0278 e para este teste é de 0.1930, apresentando uma variação <sup>1</sup> de 85.60%.

<sup>1</sup>A variação foi medida pela diferença relativa percentual dada por:  $|\beta_1(2013) - \beta_1(2012)| / \beta_1(2012) * 100$ , onde  $\beta_1(2012)$  representa o valor do coeficiente no modelo 1 no teste para 2012 e  $\beta_1(2013)$  o valor do coeficiente no modelo 1 no teste para 2013.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 0.2038$ , com 8 graus de liberdade e um valor-p associado de 1. O teste indica um bom ajustamento.

Os resíduos não apresentam nenhum valor extremo, indicando a inexistência de observações atípicas.

Figura 5.23: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 1



**Modelo de previsão 2**

Tabela 5.15: Sumário do modelo de previsão 2

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	0.7771	0.0750	10.36	0.0000
Segunda <sup>a</sup>	0.3967	0.1097	3.61	0.0020
Terça <sup>a</sup>	0.4435	0.1104	4.02	0.0008
Quarta <sup>a</sup>	0.4561	0.1110	4.11	0.0007
Quinta <sup>a</sup>	0.6396	0.1143	5.60	0.0000
Sexta <sup>a</sup>	-1.6376	0.1068	-15.33	0.0000
Sábado <sup>a</sup>	-4.1716	0.2095	-19.91	0.0000
Feriado morte	-0.5797	0.1784	-3.25	0.0044
Feriado dia anterior à previsão	-2.3416	0.1868	-12.53	0.0000
Feriado dia da previsão	0.0172	0.1967	0.09	0.9312
$\phi$ : 80.83887				
Null deviance: 175992.2 on 27 degrees of freedom				
Residual deviance: 1754.8 on 18 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 2 está apresentado na tabela 5.15.

Todas as variáveis mostraram ser significativas, com exceção de ser feriado no dia da previsão, observado também para o modelo de previsão 2 no teste para 2012. O dia de semana de ocorrência do óbito de segunda a quinta aumenta a probabilidade de um óbito ser notificado dentro de dois dias após a morte, comparativamente com o domingo, tal como no teste para 2012. A ocorrência do óbito à sexta e ao sábado diminuem a probabilidade do óbito ser notificado em dois dias em comparação com o domingo, tal como ser feriado no dia anterior à previsão, ou seja, no dia em que é agregada a informação, e ser feriado no dia da morte. O coeficiente que mostrou variar substancialmente, relativamente ao teste para 2012, está associado à variável ser feriado no dia de previsão, que no teste para 2012 apresenta um valor de 0.0172 e para este teste, apresenta um valor de -0.0073.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 0.7294$ , com 8 graus de liberdade e um valor-p associado de 0.9994. O teste indica um bom ajustamento.

Os resíduos, apresentados na figura 5.24, não apresentam observações com valores extremos, indicando um bom ajustamento local pelo modelo.

Figura 5.24: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 2

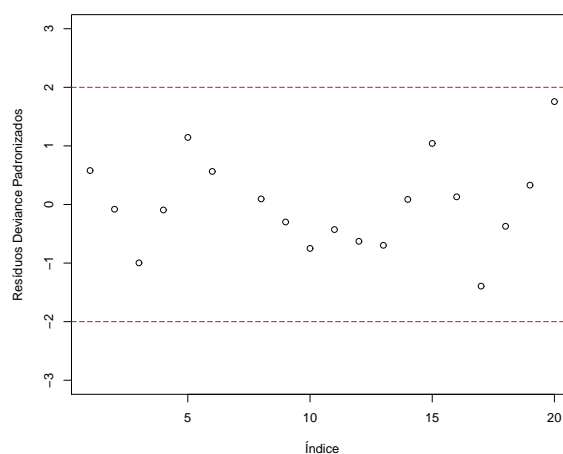
**Modelo de previsão 3**

Tabela 5.16: Sumário do modelo de previsão 3

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	2.2630	0.1355	16.70	0.0000
Segunda <sup>a</sup>	0.3674	0.2011	1.83	0.0791
Terça <sup>a</sup>	0.3989	0.2012	1.98	0.0581
Quarta <sup>a</sup>	0.4250	0.2028	2.10	0.0460
Quinta <sup>a</sup>	-0.8087	0.1697	-4.77	0.0001
Sexta <sup>a</sup>	-3.0460	0.1622	-18.78	0.0000
Sábado <sup>a</sup>	-1.3220	0.1628	-8.12	0.0000
Feriado morte	-0.6669	0.2449	-2.72	0.0114
Feriado dia anterior à previsão	-1.5965	0.2147	-7.44	0.0000
Feriado dois dias antes da previsão	-1.5541	0.2114	-7.35	0.0000
Feriado dia da previsão	-0.2005	0.2563	-0.78	0.4411
$\phi$ : 114.3595				
Null deviance: 120071.8 on 36 degrees of freedom				
Residual deviance: 3580.8 on 26 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário e os resíduos do modelo de previsão 3 encontram-se apresentados na tabela 5.16 e na figura 5.25, respectivamente.

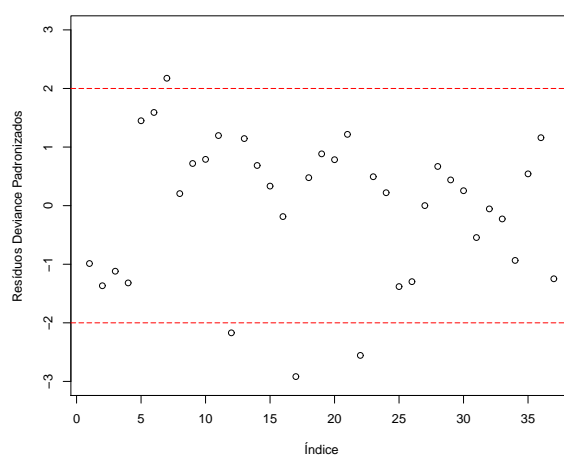
A variável que não se mostrou significativa para explicar o atraso igual ou inferior a

três dias foi a variável que indica se foi feriado no dia da previsão. Os óbitos ocorridos de segunda a quarta, comparativamente com os óbitos ocorridos ao domingo, aumentam a probabilidade de um óbito ser notificado com atraso igual ou inferior a três dias, ao contrário dos óbitos ocorridos de quinta a sábado, que diminuem esta probabilidade. Ser feriado no dia da morte e nos dois dias seguintes, também diminui a probabilidade de um óbito ser notificado num prazo de três dias. Os resultados enunciados foram os mesmos observados para o teste de 2012. Os valores dos coeficientes deste modelo não diferem substancialmente (superior a 10%) dos valores observados para o modelo correspondente no teste para 2012.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 1.6925$ , com 8 graus de liberdade e um valor-p associado de 0.989. O teste indica um bom ajustamento.

Os resíduos não apresentam valores extremos, indicando a inexistência de observações atípicas.

Figura 5.25: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 3



**Modelo de previsão 4**

Tabela 5.17: Sumário do modelo de previsão 4

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.1986	0.1743	18.35	0.0000
Segunda <sup>a</sup>	0.2469	0.2496	0.99	0.3294
Terça <sup>a</sup>	0.2757	0.2529	1.09	0.2833
Quarta <sup>a</sup>	-0.4861	0.2210	-2.20	0.0347
Quinta <sup>a</sup>	-1.6915	0.1963	-8.62	0.0000
Sexta <sup>a</sup>	-1.8117	0.1954	-9.27	0.0000
Sábado <sup>a</sup>	-0.9450	0.2097	-4.51	0.0001
Feriado morte	-0.5139	0.2313	-2.22	0.0331
Feriado dia anterior à previsão	-1.3410	0.2201	-6.09	0.0000
Feriado dois dias antes da previsão	-1.1603	0.2374	-4.89	0.0000
Feriado três dias antes da previsão	-1.5039	0.2110	-7.13	0.0000
Feriado dia da previsão	-0.4060	0.2425	-1.67	0.1032
$\phi$ : 91.32259				
Null deviance: Null deviance: 36945.0 on 45 degrees of freedom				
Residual deviance: Residual deviance: 3511.7 on 34 degrees of freedom				
<sup>a</sup> Referência: Domingo				

O sumário do modelo de previsão 4 está apresentado na tabela 5.17.

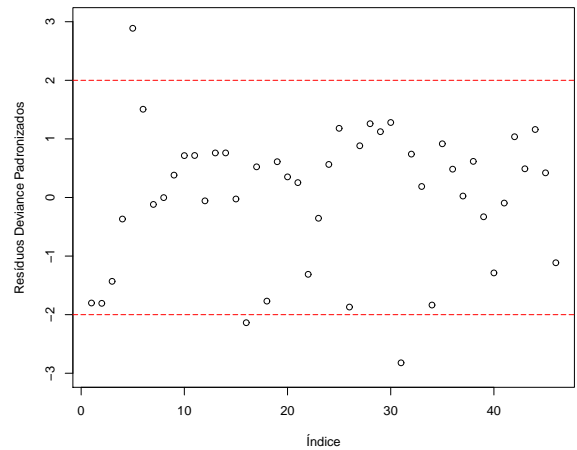
As variáveis que não se mostraram significativas para explicar este atraso foram a segunda e terça-feira de ocorrência do óbito e ser feriado no dia de previsão. Para este modelo de previsão, comparativamente com o observado no teste para 2012, ser feriado no dia da previsão deixa de ser significativo. O efeito da ocorrência dos óbitos segunda e terça não difere significativamente do domingo no atraso de quatro dias na notificação dos óbitos. Os restantes dias de semana de ocorrência do óbito, relativamente ao domingo, ser feriado no dia da morte e nos três dias seguintes, diminuem esta probabilidade.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 1.9015$ , com 8 graus de liberdade e um valor-p associado de 0.9839. O teste indica um bom ajustamento.

Os resíduos para este modelo, apresentados na figurs 5.26, não apresentam nenhum valor extremo, indicando a inexistência de observações mal ajustadas.



Figura 5.26: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 4



**Modelo de previsão 5**

Tabela 5.18: Sumário do modelo de previsão 5

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.5864	0.1384	25.92	0.0000
Segunda <sup>a</sup>	0.1617	0.1966	0.82	0.4156
Terça <sup>a</sup>	-0.1901	0.1850	-1.03	0.3098
Quarta <sup>a</sup>	-0.9299	0.1647	-5.65	0.0000
Quinta <sup>a</sup>	-1.0693	0.1626	-6.58	0.0000
Sexta <sup>a</sup>	-1.0862	0.1627	-6.67	0.0000
Sábado <sup>a</sup>	-0.4833	0.1771	-2.73	0.0092
Feriado morte	-0.1566	0.2163	-0.72	0.4731
Feriado dia anterior à previsão	-0.9896	0.1758	-5.63	0.0000
Feriado dois dias antes da previsão	-0.8906	0.1886	-4.72	0.0000
Feriado três dias antes da previsão	-1.0319	0.1799	-5.74	0.0000
Feriado quatro dias antes da previsão	-1.0594	0.1690	-6.27	0.0000
Feriado dia da previsão	-0.0671	0.2226	-0.30	0.7647
$\phi$ : 40.12825				
Null deviance: 10934.6 on 54 degrees of freedom				
Residual deviance: 1736.2 on 42 degrees of freedom				
<sup>a</sup> Referência: Domingo				

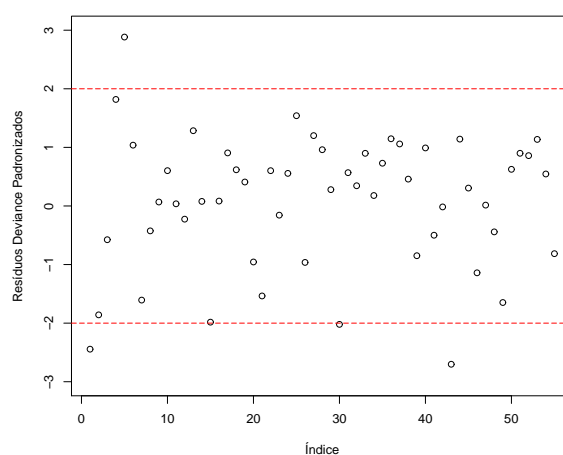
O sumário do modelo de previsão 5 é apresentado na tabela 5.18.  
As variáveis que não se mostraram significativas foram a segunda e terça-feira de

ocorrência do óbito, ser feriado no da morte e no dia de previsão. Os resultados obtidos são muito semelhantes aos resultados obtidos no teste para 2012. Ser feriado nos quatro dias anteriores ao dia de previsão diminui a probabilidade de um óbito ser notificado até cinco dias após a morte, tal como o dia de ocorrência da morte de quarta a sábado, em comparação com o domingo. O coeficiente que se alterou substancialmente, em relação ao teste para 2012, está associado à variável ser feriado no dia da previsão.

O teste de Hosmer-Lemeshow apresentou uma estatística de teste  $\chi^2 = 0.6128$ , com 8 graus de liberdade e um valor-p associado de 0.9997. O teste indica que o modelo está bem ajustado.

Os resíduos, apresentados na figura 5.27, não apresentam nenhum valor extremo.

Figura 5.27: Resíduos da deviance Padronizados vs Índice - Modelo de previsão 5



### 5.3.2 Implementação da Metodologia

Nesta secção serão apresentados os resultados obtidos através da metodologia proposta para 2013, considerando as probabilidades estimadas através dos modelos apresentados na secção anterior, tal como na implementação da metodologia para 2012. Para recordar, estas foram calculadas considerando apenas a informação receccionada até 2012, e serão utilizadas para prever a mortalidade com o dados receccionados durante o ano de 2013. O objetivo é testar a metodologia nas condições em que deverá ser utilizada, ou seja, tendo em conta as notificações dos óbitos que chegam diariamente e estão disponíveis em cada momento de previsão. As figuras identificadas de 5.28 a 5.34, apresentam as previsões para 2013 para cada dia de previsão, de  $t + 1$  a  $t + 5$ , a mortalidade observada, a linha de base para a mortalidade e os respetivos intervalos de confiança a 95% e a 99%.

As previsões em  $t + 1$ , apresentadas na figura 5.28, apresentam diversos valores abaixo e acima da mortalidade observada. Entende-se, como em 2012, que estas correspondem às estimativas obtidas sem informação, ou seja, quando o número de notificações é nulo. Ao retirar apenas as estimativas efetuadas ao domingo e à segunda, informação apresentada na figura 5.29, confirma-se que as observações discrepantes correspondem essencialmente às previsões efetuadas sem informação, tal como em 2012.

As previsões em  $t + 2$ , apresentadas na figura 5.30, são baseadas na informação recebida durante os dois dias seguintes à data de morte. Apresentam também, diversos valores abaixo da mortalidade observada. Entende-se, como em 2012, que estas correspondem às estimativas obtidas sem informação. Ao retirar apenas as previsões efetuadas à segunda-feira, ou seja, previsões obtidas com informação agregada ao sábado e domingo, confirma-se que estas observações correspondem essencialmente às previsões efetuadas sem informação, como se pode verificar na figura 5.31.

As previsões em  $t + 3$  estão apresentadas na figura 5.14 e são baseadas na informação recebida durante os três dias seguintes à data de morte. Observam-se algumas estimativas bastante abaixo da mortalidade observada, que analisados mais detalhadamente, correspondem aos dias em que existiram feriados antes ou depois de fins-de-semana, levando a três dias sem informação. Os dias detetados foram: 29 de Março e 8 de Junho. Uma situação extraordinária observou-se no dia 28 de Setembro, que conta com informação agregada de sábado a segunda, mas apenas foram recebidas 65 notificações. É de notar que a mediana das notificações em  $t + 3$ , das mortes ocorridas ao sábado

em 2013, foi de 215 óbitos.

As previsões em  $t + 4$  estão apresentadas na figura 5.33. As observações muito abaixo da mortalidade esperada verificam ser as mesmas observações detetadas em  $t + 3$ . Como tal, estas dizem respeito a estimativas elaboradas sem informação ou com informação irregular.

As previsões em  $t + 5$  estão apresentadas na figura 5.34, e não se apresentam discordantes com a mortalidade observada.

Figura 5.28: Previsões em  $t + 1$  para 2013

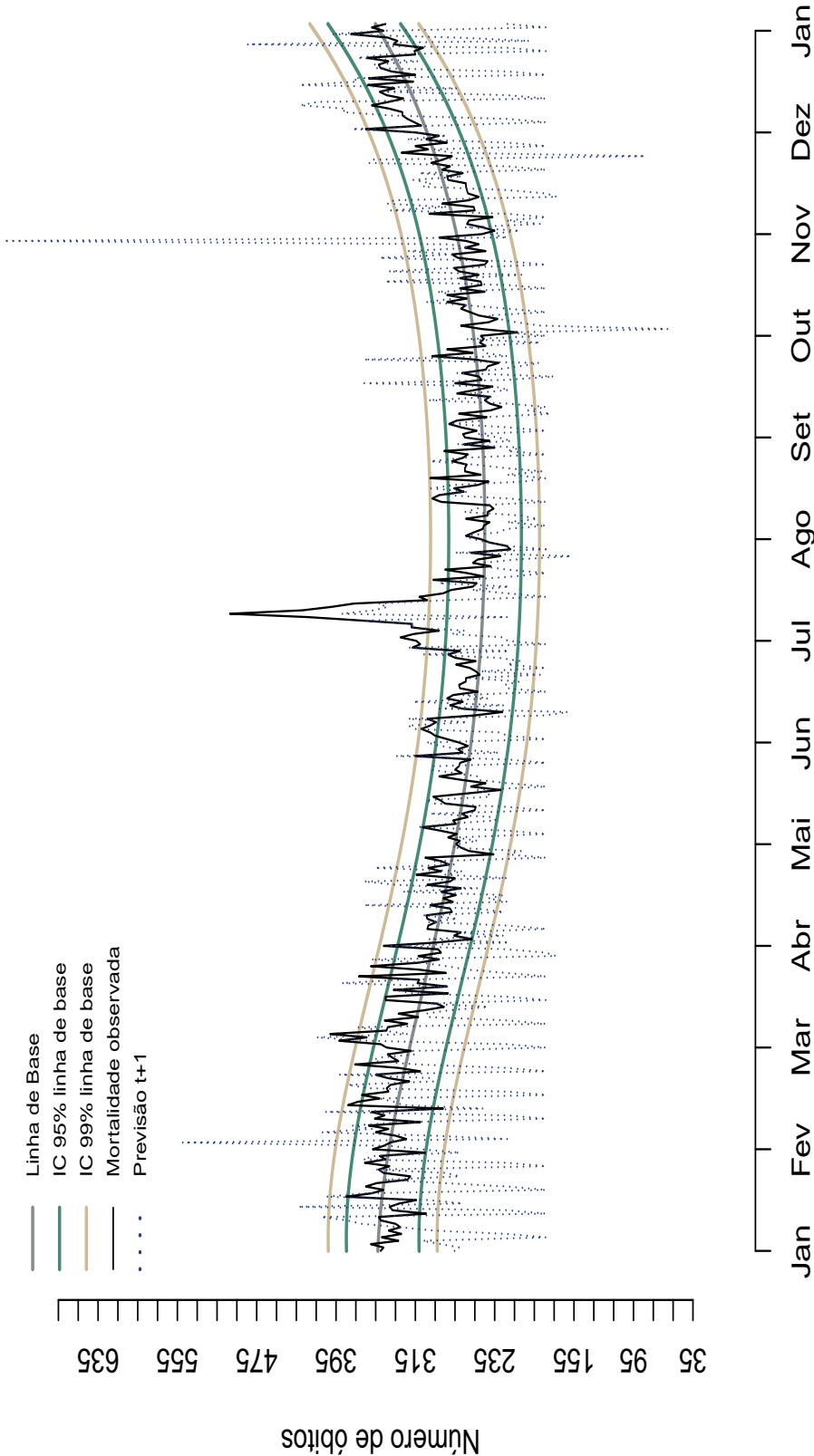


Figura 5.29: Previsões em  $t + 1$  para 2013, excluindo as previsões efetuadas ao domingo e à segunda-feira

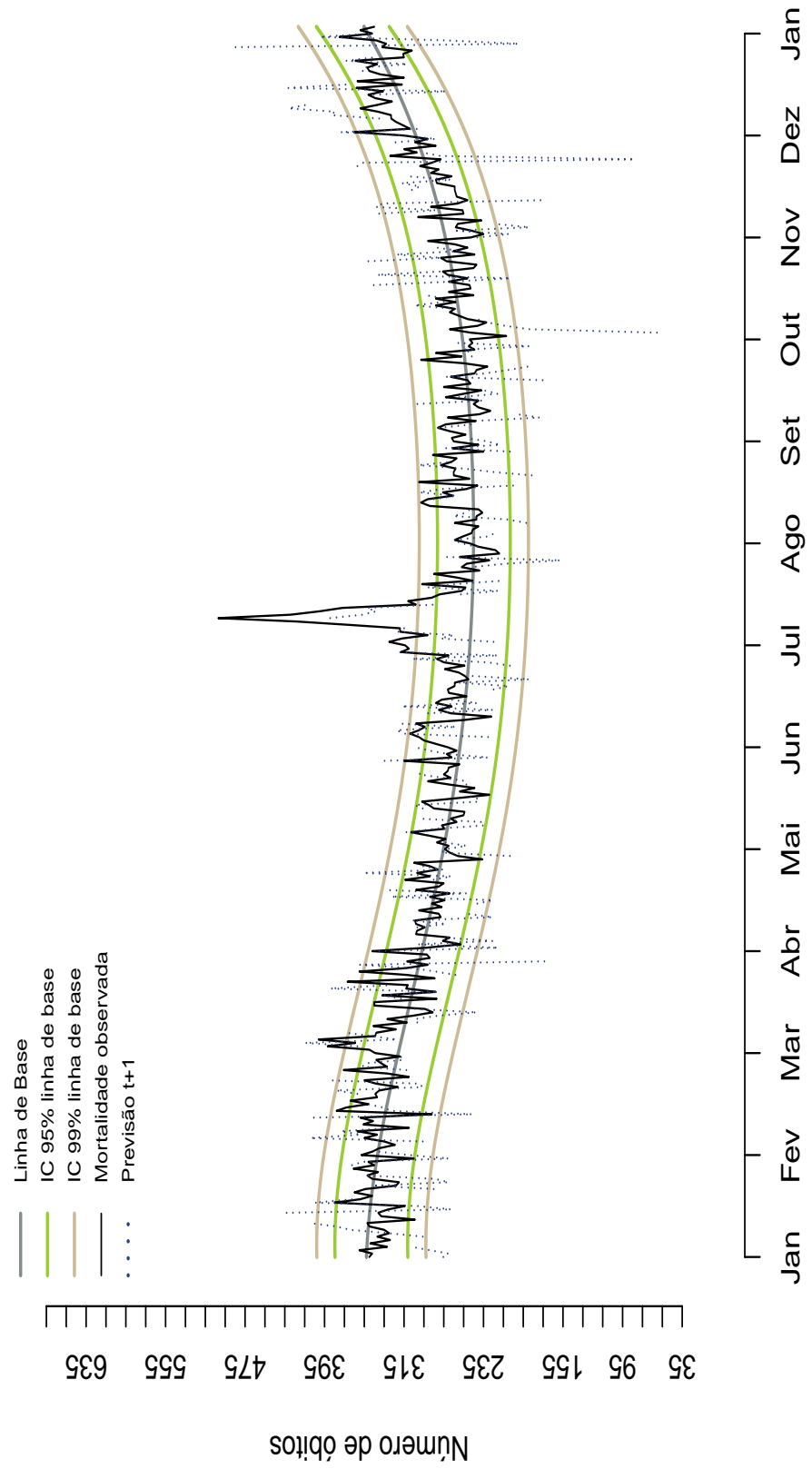


Figura 5.30: Previsões em  $t + 2$  para 2013

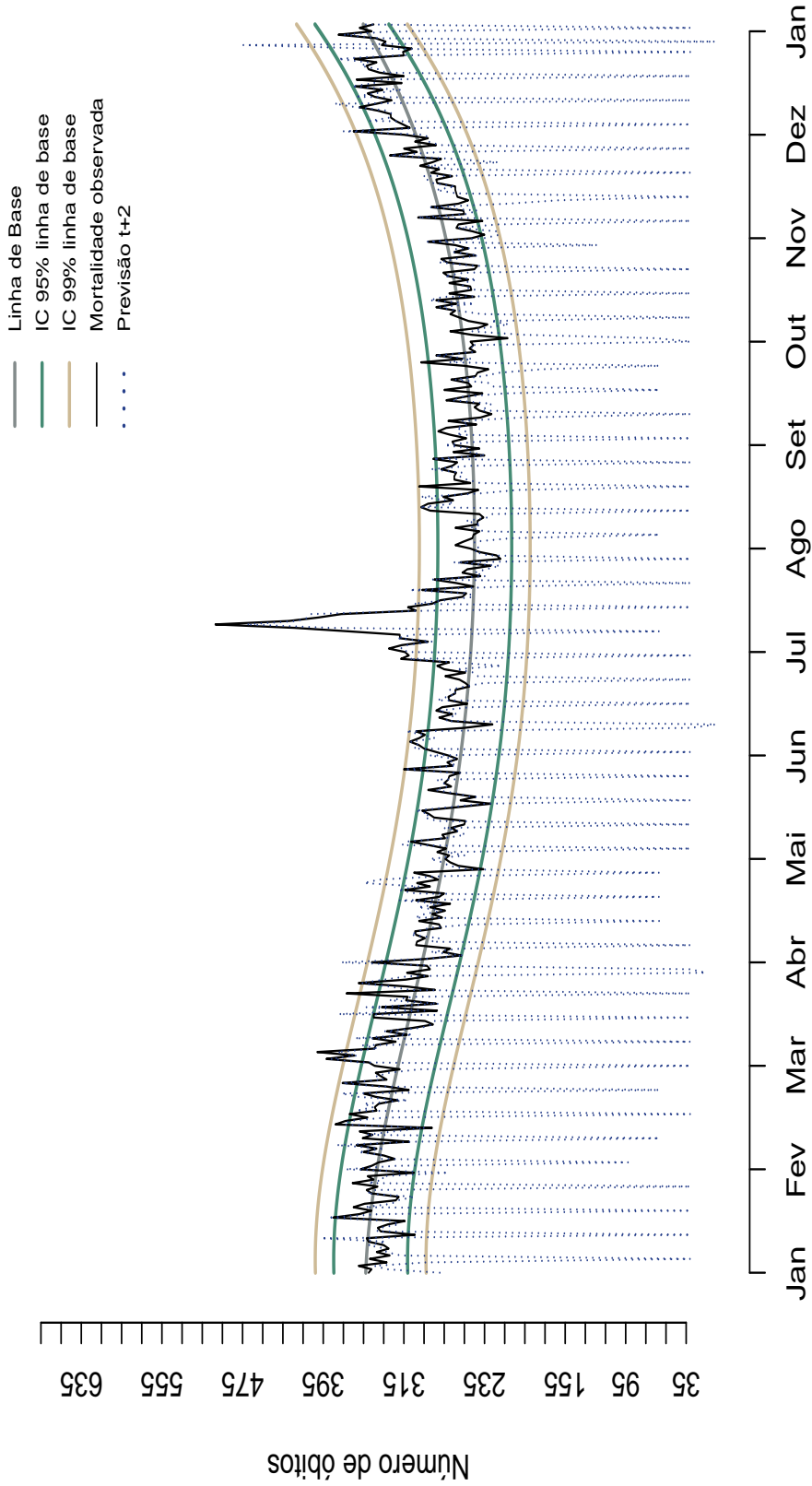


Figura 5.31: Previsões em  $t + 2$  para 2013, excluindo as previsões efetuadas à segunda-feira

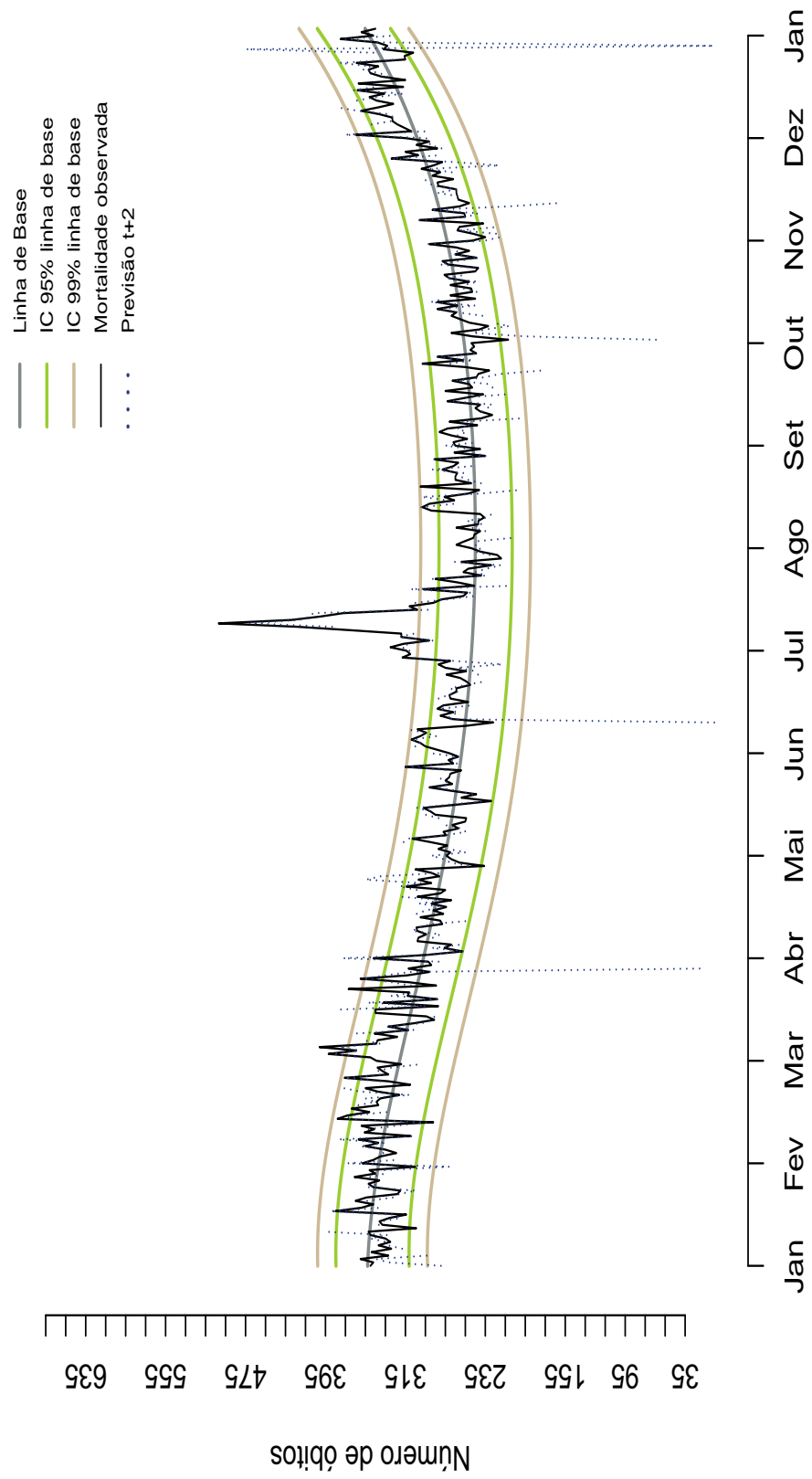




Figura 5.32: Previsões em  $t + 3$  para 2013

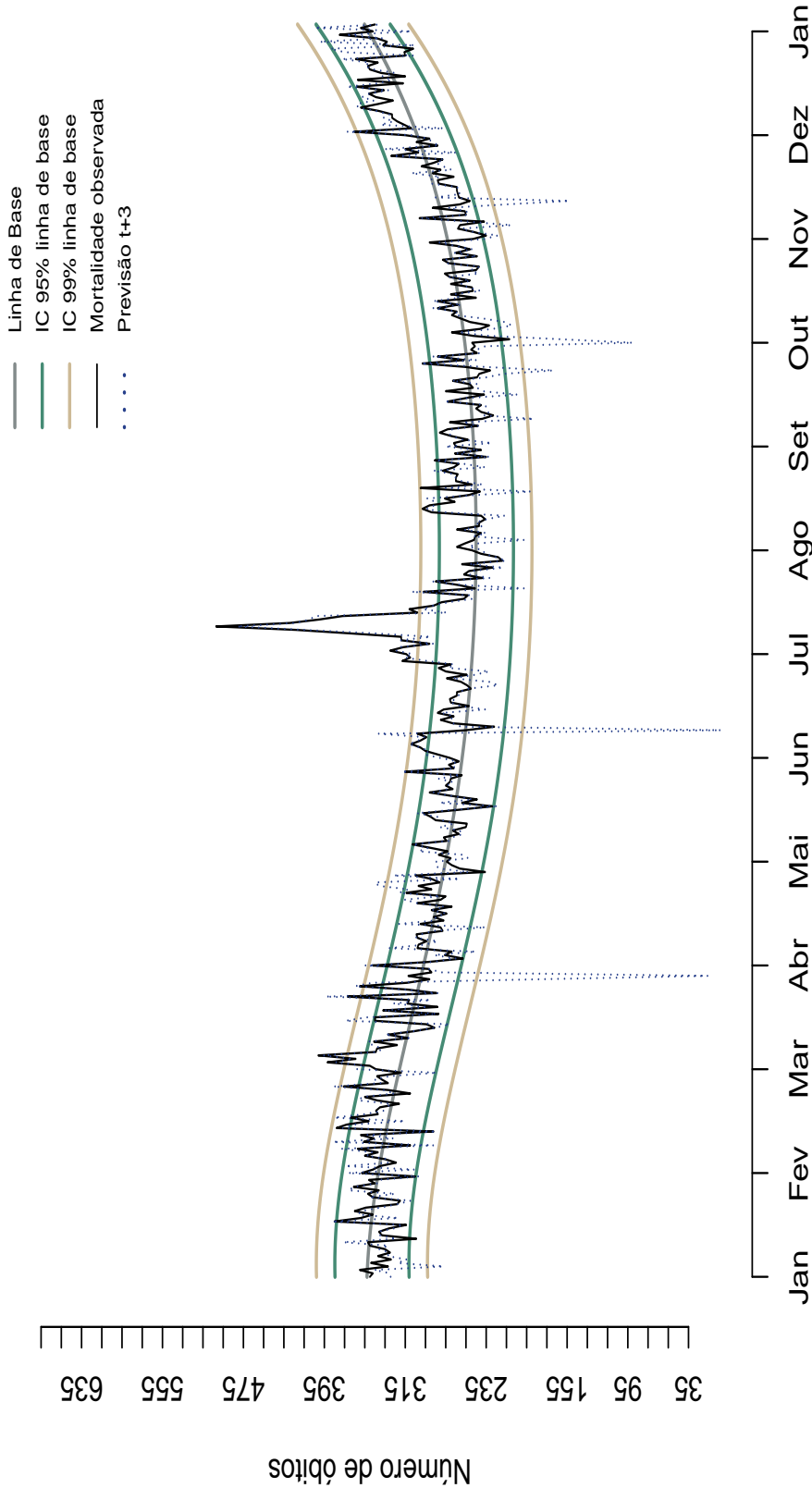


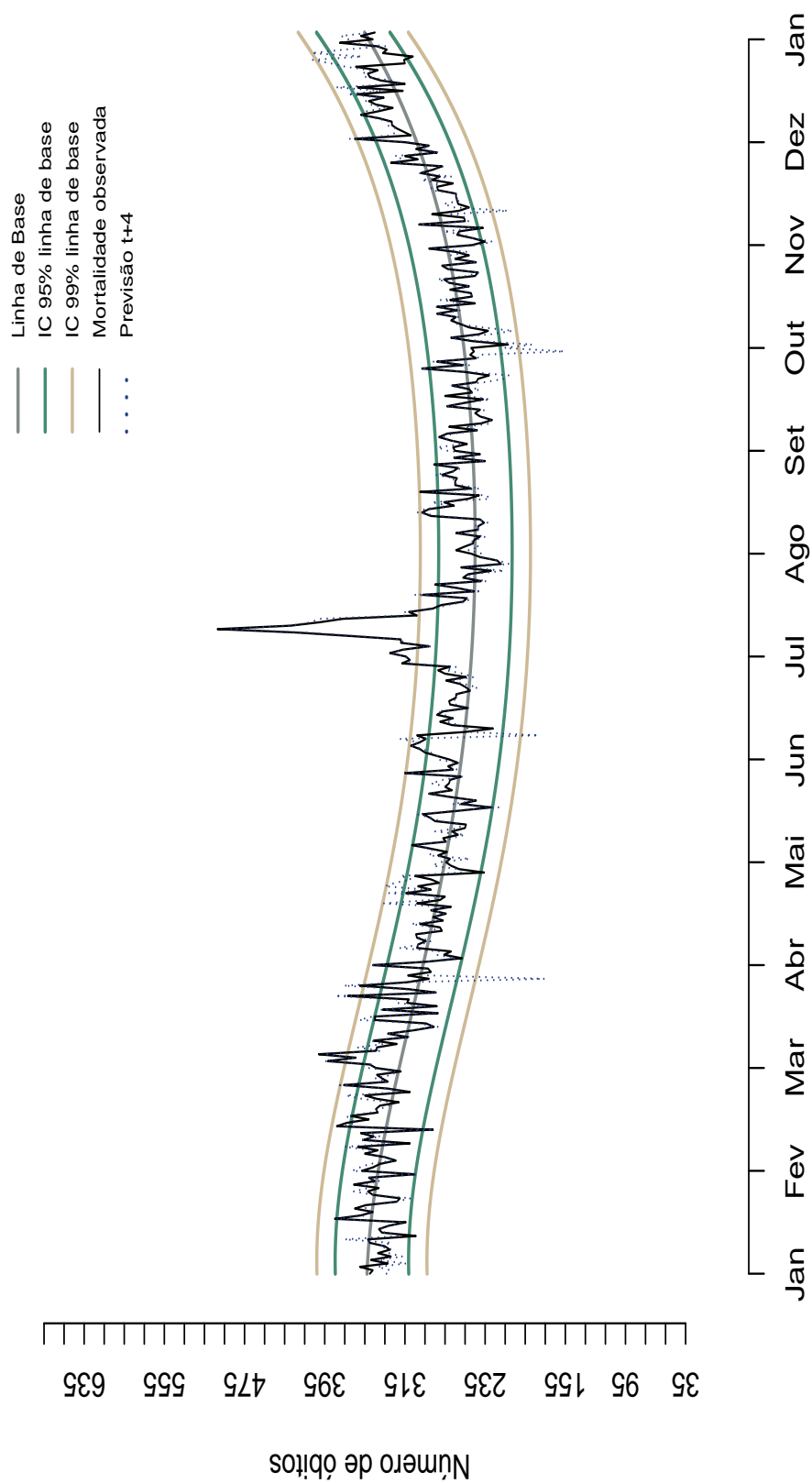
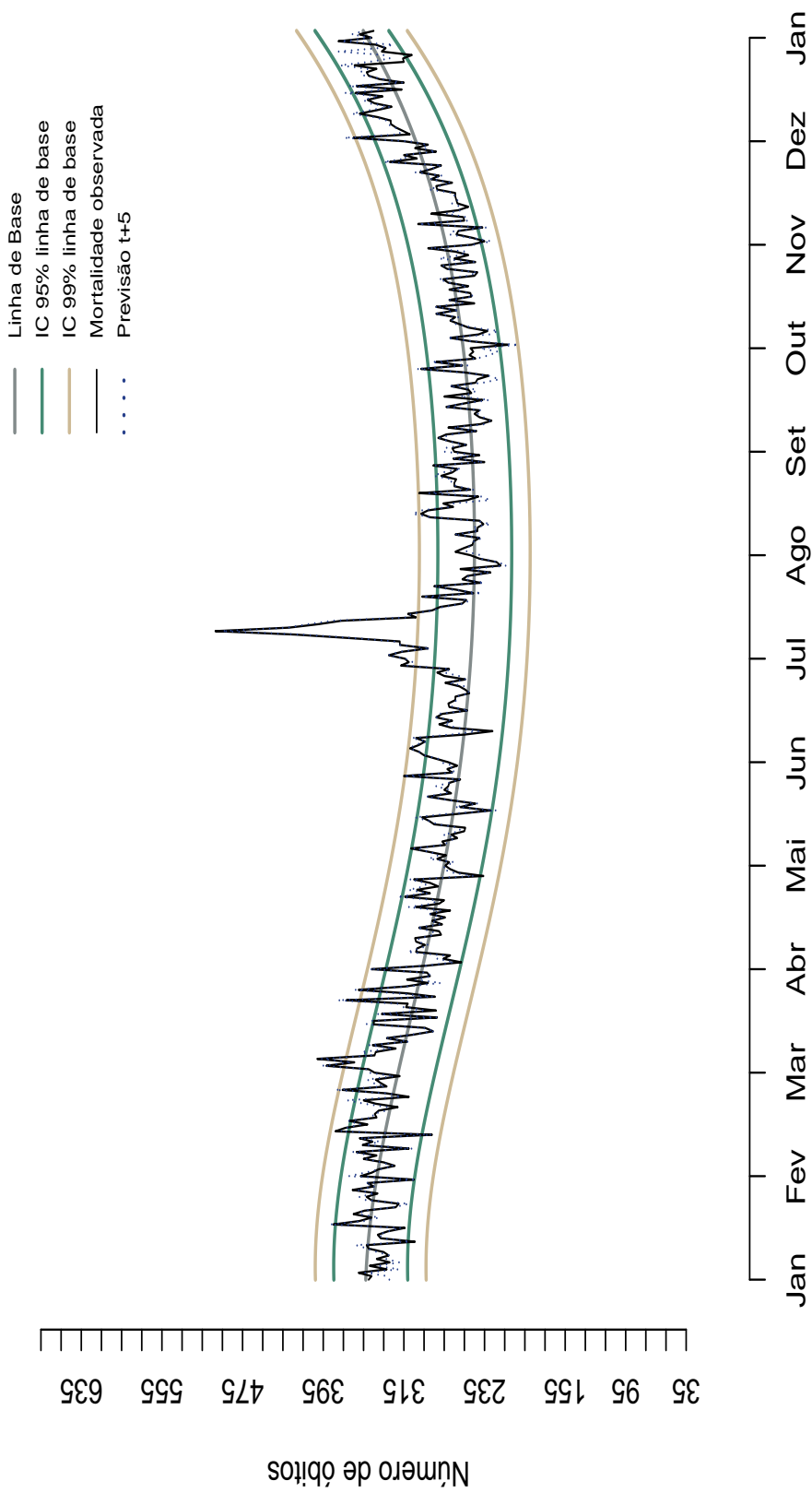
Figura 5.33: Previsões em  $t + 4$  para 2013

Figura 5.34: Previsões em  $t + 5$  para 2013



### 5.3.3 Avaliação da Metodologia

Para poder avaliar o desempenho da metodologia para o ano de 2013, tal como para o ano de 2012, as medidas foram calculadas para os vários cenários relativos ao número de dias em que não houve receção de informação. Por isso, é importante analisar a frequência de cada cenário para 2013. Esta informação está disponível na tabela 5.19 que apresenta o número de previsões efetuadas em cada um dos cenários relativos à receção da informação.

Tabela 5.19: Número de previsões efetuadas para cada cenário de receção da informação, para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , no ano de 2013

	Informação completa	Número de dias sem informação				
		1 dia	2 dias	3 dias	4 dias	5 dias
$t + 1$	251 (69.1)	112 (30.9)	—	—	—	—
$t + 2$	195 (53.7)	113 (31.1)	55 (15.2)	—	—	—
$t + 3$	142 (39.1)	110 (30.3)	109 (30.0)	2 (0.6)	—	—
$t + 4$	92 (25.3)	105 (28.9)	158 (43.5)	8 (2.2)	0	—
$t + 5$	45 (12.4)	98 (26.9)	203 (55.9)	16 (4.4)	1 (0.3)	0

Os valores apresentados na diagonal da tabela referente ao número de dias sem informação representam as previsões efetuadas sem qualquer informação. Durante o ano 2013, não houve nenhum período em que não se tenha recebido informação durante 5 dias. Nas previsões a  $t + 1$ , 30.9% das estimativas foram efetuadas sem informação e 69.1% com informação completa. Nas previsões a  $t + 2$ , 15.2% das estimativas foram efetuadas sem informação. Após  $t + 2$  as estimativas efetuadas sem qualquer informação são inferiores a 0.6%. Até três dias de atraso, o mais frequente é a produção de estimativas sem falta de informação. Nos atrasos até 4 e 5 dias o mais frequente é sempre a produção de estimativas com pelos menos um dia de informação.

#### Erro Relativo das previsões

A tabela 5.9 e a figura 5.17 apresentam o erro relativo mediano percentual para cada previsão efetuada em  $t + i$ , com  $i = 1, \dots, 5$ , e para cada cenário de receção de informação.

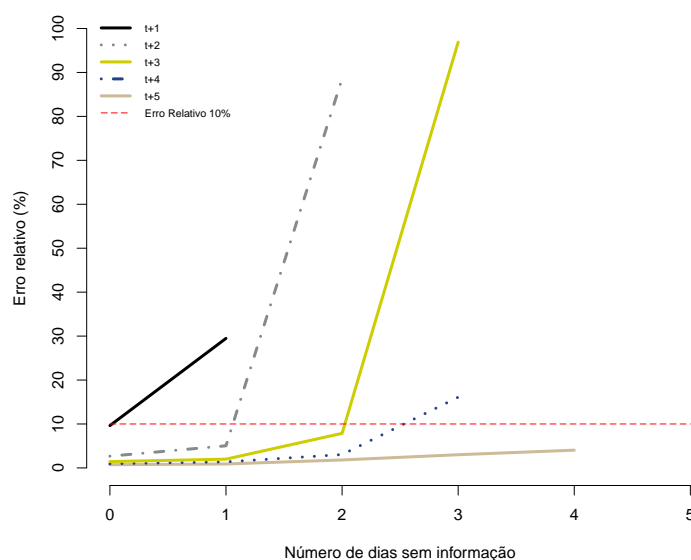
Analisando a tabela e o gráfico referidos acima, o erro relativo das previsões variou em função do número de dias que se recebeu informação. O erro relativo mediano das previsões a  $t + 1$  foi inferior a 10% se houve receção de informação, mas sem informação, o erro das previsões passou para 29.47%. Verificou-se o mesmo resultado

para os restantes momentos de previsão, ou seja, quando houve pelo menos um dia em que foi rececionada informação, as estimativas obtidas apresentaram um erro relativo mediano inferior a 10%. A única exceção foram as previsões em  $t + 4$ , que com apenas um dia de informação, apresentaram um erro relativo mediano de 16.09%. Sem qualquer informação, as estimativas apresentaram valores para o erro relativo mediano bem superiores a 10%. Este resultado é visualizado de imediato através do gráfico 5.35, pois apenas as estimativas efetuadas sem informação apresentam erros superiores a 10%, visto que são as únicas que estão acima da linha que indica um erro de 10%, com exceção da previsão em  $t + 4$ .

Tabela 5.20: Erro relativo mediano (%) das previsões para cada cenário de receção de informação para 2013

Erro Relativo mediano (%)						
Previsão	Informação completa	Número de dias sem informação				
		1 dia	2 dias	3 dias	4 dias	5 dias
$t + 1$	9.63	29.47	—	—	—	—
$t + 2$	2.67	5.03	88.59	—	—	—
$t + 3$	1.44	1.99	7.86	96.89	—	—
$t + 4$	0.94	1.37	3.04	16.09	0	—
$t + 5$	0.73	0.89	1.81	3.00	4.04	0

Figura 5.35: Erro das previsões por número de dias sem informação



### Sensibilidade e Especificidade

A tabela 5.21 apresenta os valores da sensibilidade e especificidade para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de receção de informação para 2013. Em 2013 foram observados 35 dias com excesso de mortalidade, considerando a definição de excesso em 2.2.2, correspondendo a 8 períodos de excesso de mortalidade.

A sensibilidade e a especificidade variaram em função do número de dias que se recebeu informação, tal como em 2012. A probabilidade de um excesso de mortalidade ser detetado no dia seguinte foi de 0.56 se se tiver recebido informação. Se não se tiver recebido informação, esta passa para 0.1. O que nos indica que a probabilidade de ocorrerem falsos negativos no dia seguinte, ou seja, de um excesso não ser detetado se não existir informação, é de 0.9.

Após dois dias, a probabilidade de se detetar um excesso de mortalidade foi de 0.81 se se recebeu informação nos dois dias seguintes à data que se quer estimar. Com apenas um dia de informação, esta passa a ser de 0.67.

A partir de  $t + 2$ , sem receção de informação, a sensibilidade é nula, ou seja, a metodologia não tem capacidade para detetar excesso de mortalidade.

A probabilidade de se detetar um dia sem excesso de mortalidade é sempre superior a 0.83, apesar da quantidade de dias que se recebeu informação.

Tabela 5.21: Sensibilidade e Especificidade das previsões em  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de receção de informação, para 2013

Previsão	Número de dias sem informação	Sensibilidade	Especificidade
$t + 1$	Informação completa	0.56	0.87
	1 dia	0.10	0.94
$t + 2$	Informação completa	0.88	0.98
	1 dia	0.81	0.93
	2 dias	0	1.00
$t + 3$	Informação completa	0.94	0.98
	1 dia	0.80	0.93
	2 dias	0.78	0.97
	3 dias	0	1.00
$t + 4$	Informação completa	0.89	0.99
	1 dia	0.88	0.99
	2 dias	0.83	0.99
	3 dias	—	0.95
	4 dias	—	—
$t + 5$	Informação completa	1.00	1.00
	1 dia	1.00	1.00
	2 dias	1.00	1.00
	3 dias	0	1.00
	4 dias	0	1.00
	5 dias	—	—

### Valor preditivo Positivo e Negativo

A tabela 5.22 apresenta o Valor Preditivo Positivo (VPP) e Negativo (VPN) para cada previsão  $t + i$ , com  $i = 1, \dots, 5$ , por cada cenário de receção de informação.

O Valor Preditivo Positivo e Negativo também mostrou variar em função do número de dias que se recebeu informação. A probabilidade de um excesso de mortalidade detetado pela metodologia no dia seguinte ser verdadeiro, é de 0.43 se se tiver recebido informação. Se não se tiver recebido informação, esta passa a ser de 0.30.

Após dois dias, a probabilidade ser detetado um verdadeiro excesso de mortalidade é de 0.83 se se recebeu informação nos dias seguintes à data de ocorrência. Com apenas um dia de informação, esta passa a ser de 0.62.

Tabela 5.22: Valor Preditivo Positivo (VPP) e Valor Preditivo Negativo (VPN) por número de dias sem informação e por dia de previsão  $t + i$ , com  $i = 1, \dots, 5$ 

Previsão	(N)	VPP	VPN
$t + 1$	Informação completa	0.43	0.95
	1 dia	0.30	0.91
$t + 2$	Informação completa	0.83	0.98
	1 dia	0.62	0.99
	2 dias	0	0.99
$t + 3$	Informação completa	0.88	0.99
	1 dia	0.80	0.98
	2 dias	0.70	0.98
	3 dias	—	1.00
$t + 4$	Informação completa	0.93	0.99
	1 dia	0.89	0.98
	2 dias	0.60	1.99
	3 dias	—	1.00
	4 dias	—	
$t + 5$	Informação completa	0.93	1.00
	1 dia	0.93	1.00
	2 dias	0.97	1.00
	3 dias	0	1.00
	4 dias	—	1.00
	5 dias	—	

Após três dias, se se recebeu informação todos os dias, a probabilidade de ser detetado um verdadeiro excesso foi de 0.88, com apenas dois dias de informação, esta passa a 0.80.

A probabilidade de a metodologia detetar um verdadeiro dia sem excesso de mortalidade foi sempre superior a 0.87 para qualquer dia de previsão e mesmo sem se ter recebido informação.

### Timeliness na deteção de um período de excesso de mortalidade

A tabela 5.23 apresenta os óbitos notificados e estimados pela metodologia, em cada momento de previsão (representados pelo dia em que seria efetuada a previsão e, da



mesma forma, que seriam receccionadas as notificações, e o respetivo dia de semana), para o primeiro dia de excesso e para os dias adjacentes. O mesmo é apresentado para o último dia de excesso. A referida tabela, tal como em 2012, tem como objetivo analisar o *timeliness* na deteção de excesso de mortalidade e, da mesma forma, analisar o tempo até à deteção do final do excesso de mortalidade. Uma conclusão importante é retirada das figuras 5.28 a 5.32, que nos confirmam, mais uma vez, tal como em 2012, que os valores discrepantes estimados pela metodologia correspondem aos dias em que as notificações foram nulas ou irregulares.

O primeiro excesso de mortalidade durante a onda de calor de 2013 ocorreu no dia 28 de Junho de 2013. O valor observado da mortalidade para este dia foi de 318 óbitos, superior ao valor do limite superior do intervalo de confiança a 99%. Como foi referido, um período de excesso de mortalidade inicia-se com dois dias consecutivos da mortalidade acima do valor do LS IC a 95% (regra  $2_{2s}$ ), ou um dia acima do valor do LS IC a 99% (regra 13s).

Através do sistema VDM atual, o excesso para o início do período, dia 28 de Junho, é detetado dia 4 de Julho através da regra  $2_{2s}$ . Associando os 288 óbitos notificados a 3 de julho para dia 28 de Junho, e os 293 óbitos notificados dia 4 de Julho para o dia seguinte. Como se pode verificar na tabela 5.23 e na figura 5.40, através das notificações, o primeiro excesso durante este período é detetado para o dia 28 de Junho, 6 dias depois, quando se observam as duas primeiras notificações acima do LS IC 95%.

Através da metodologia, o início do período de excesso de mortalidade, dia 28 de Junho, é detetado no dia seguinte, com uma previsão de 322 óbitos (ER=1.26%).

O tempo salvo até à deteção do início do período de excesso foi de 5 dias, definindo o *timeliness* até à deteção de excesso de mortalidade em 5 dias ou em 0.83.

O ultimo dia de excesso de mortalidade é observado a 14 de Julho e é detetado pela metodologia dois dias depois. O primeiro dia sem excesso de mortalidade é observado a 15 de Julho e é detetado no dia seguinte pela metodologia.

## CAPÍTULO 5. RESULTADOS

Tabela 5.23: Número de notificações (N) e número de óbitos previstos para o início e fim do período de excesso de mortalidade

Início do excesso de mortalidade								
	27 Jun 2013 (Quinta)		<b>28 Jun 2013 (Sexta)</b>		29 Jun 2013 (Sábado)		30 Jun 2013 (Domingo)	
Dia	Previsão	N	Previsão	N	Previsão	N		
28 Jun (sex)	221	42	—	—	—	—	—	—
29 Jun (sab)	266	214	<b>322</b>	88	—	—	—	—
30 Jun (dom)	266	216	309	92	183	0	—	—
1 Jul (seg)	264	216	293	92	30	0	222	0
2 Jul (ter)	270	249	312	250	<b>316</b>	227	311	213
3 Jul (qua)	271	260	311	<b>288</b>	309	280	308	279
4 Jul (qui)	272	265	312	300	307	<b>293</b>	308	279
5 Jul (sex)	272	266	316	307	309	300	315	303
LS IC 95% ; 99%	284 ; 303		284 ; 303		284 ; 302		284 ; 303	
Total de óbitos	270		318		310		314	
Timeliness para o início do excesso=0.83								
Fim do excesso de mortalidade								
	14 Jul 2013 (dom)		<b>15 Jul 2013 (seg)</b>		16 Jul 2013 (ter)			
Dia	Previsão	Notificados	Previsão	Notificados	Previsão	Notificados		
15 Jul (seg)	222	0	—	—	—	—	—	—
16 Jul (ter)	305	209	<b>257</b>	47	—	—	—	—
17 Jul (qua)	288	261	273	209	<b>220</b>	42		
18 Jul (qui)	284	273	272	254	246	190		
19 Jul (sex)	286	278	274	265	248	232		
20 Jul (sab)	287	280	276	270	253	246		
LS IC 95% ; 99%	282 ; 301		282 ; 300		282 ; 300			
Total de óbitos	287		279		257			

Figura 5.36: Previsões e notificações a  $t + 1$  durante a Onda de Calor em 2013

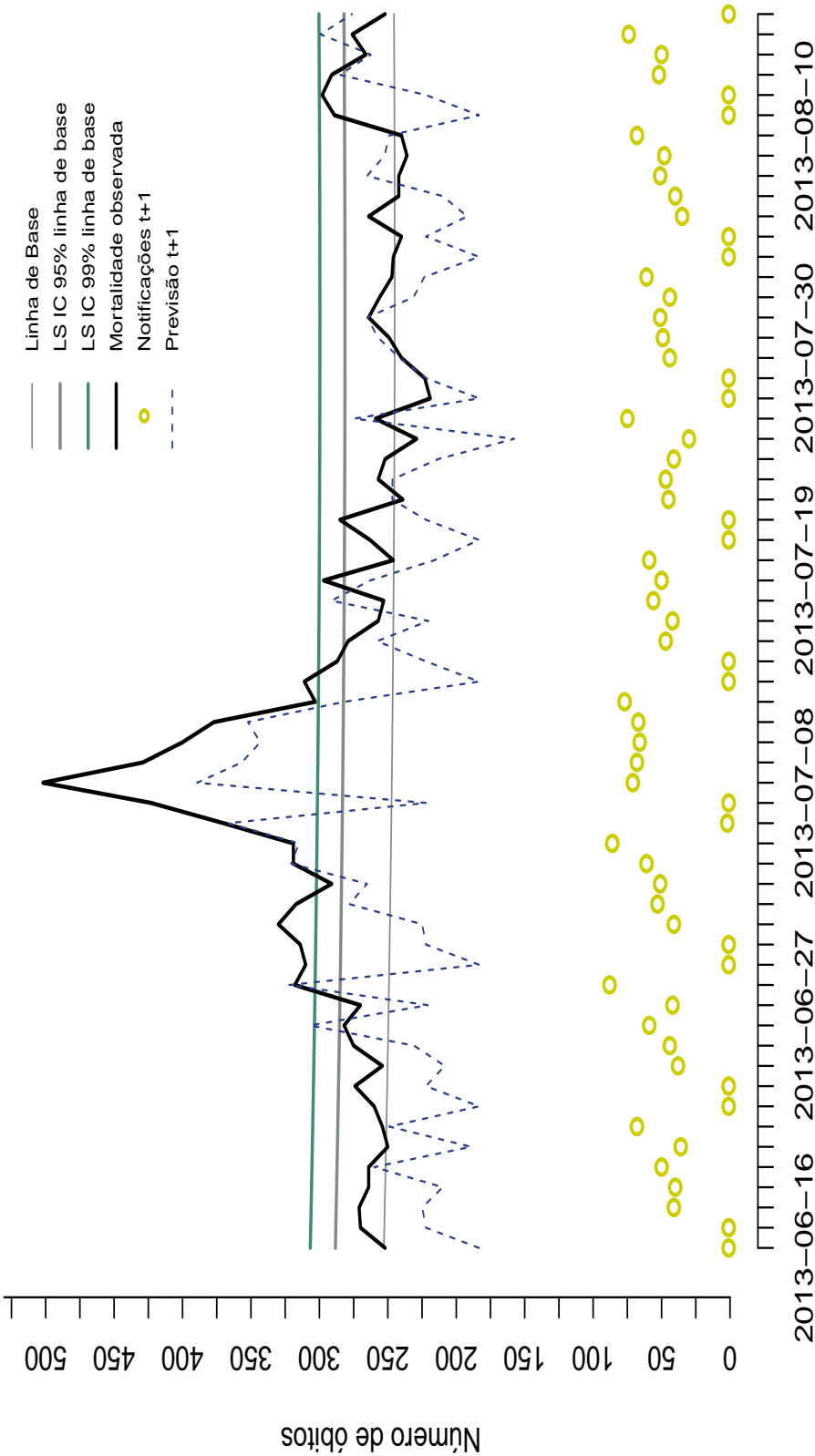


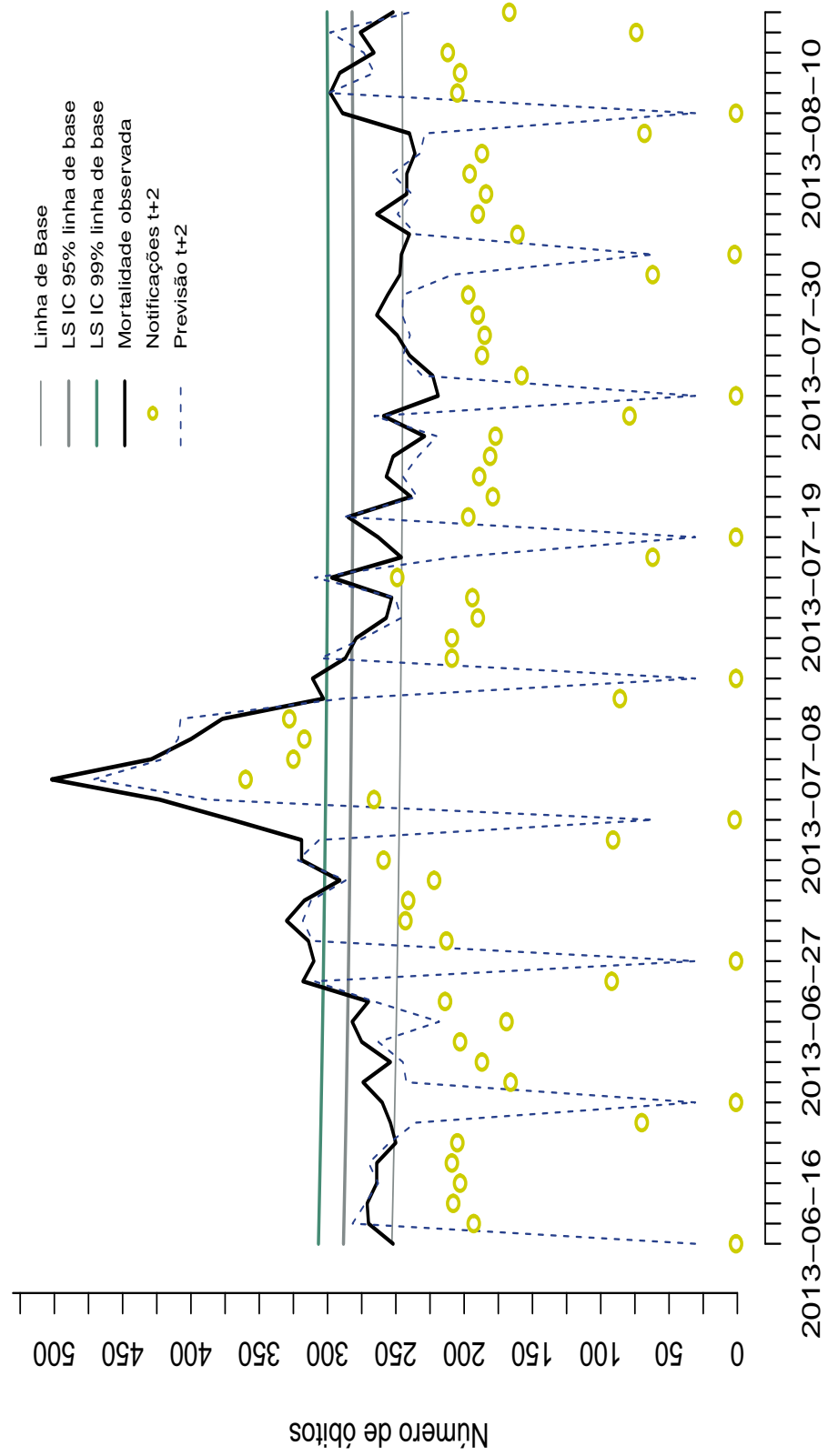
Figura 5.37: Previsões e notificações a  $t + 2$  durante a Onda de Calor em 2013

Figura 5.38: Previsões e notificações a  $t + 3$  durante a Onda de Calor em 2013

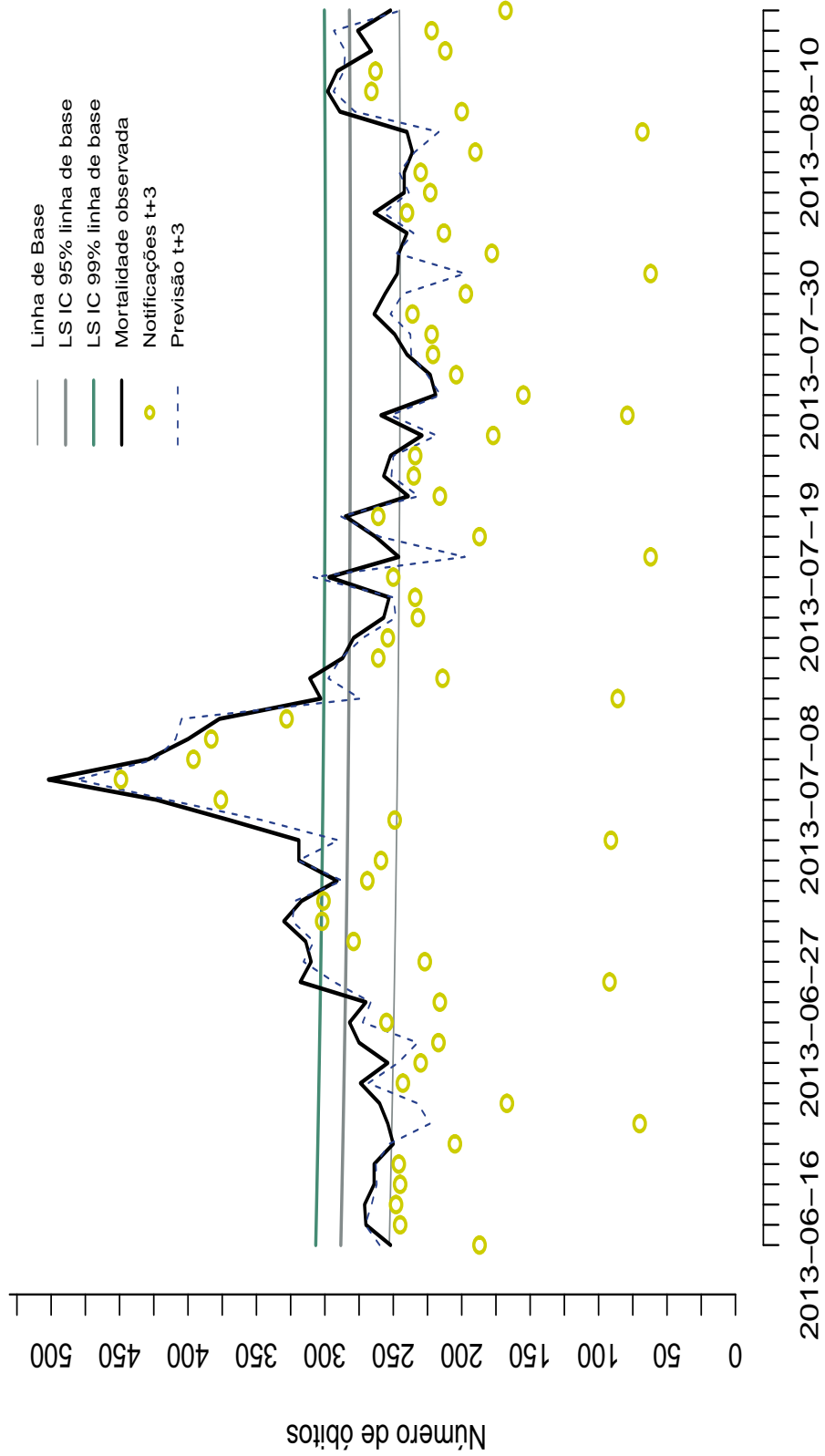


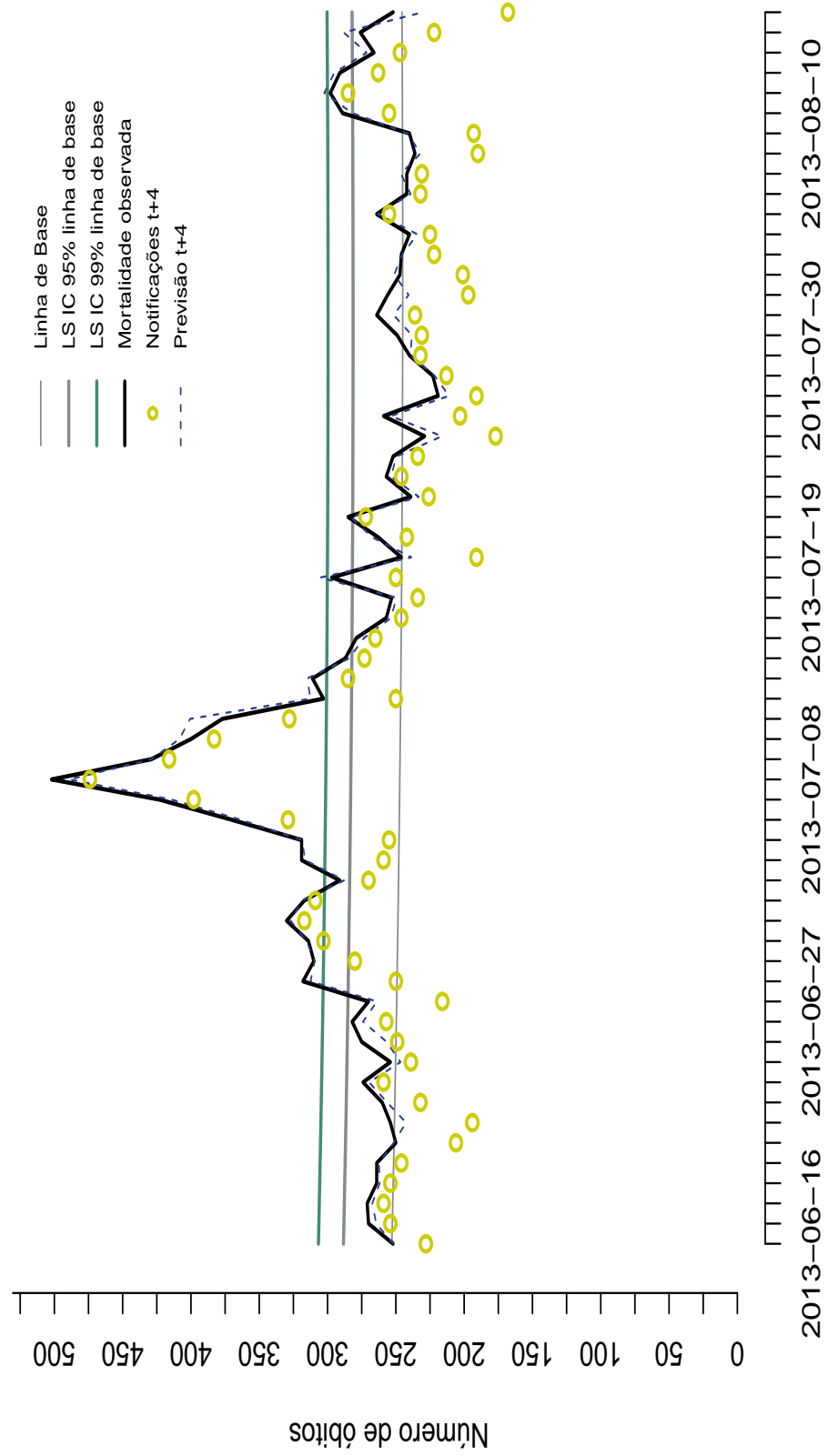
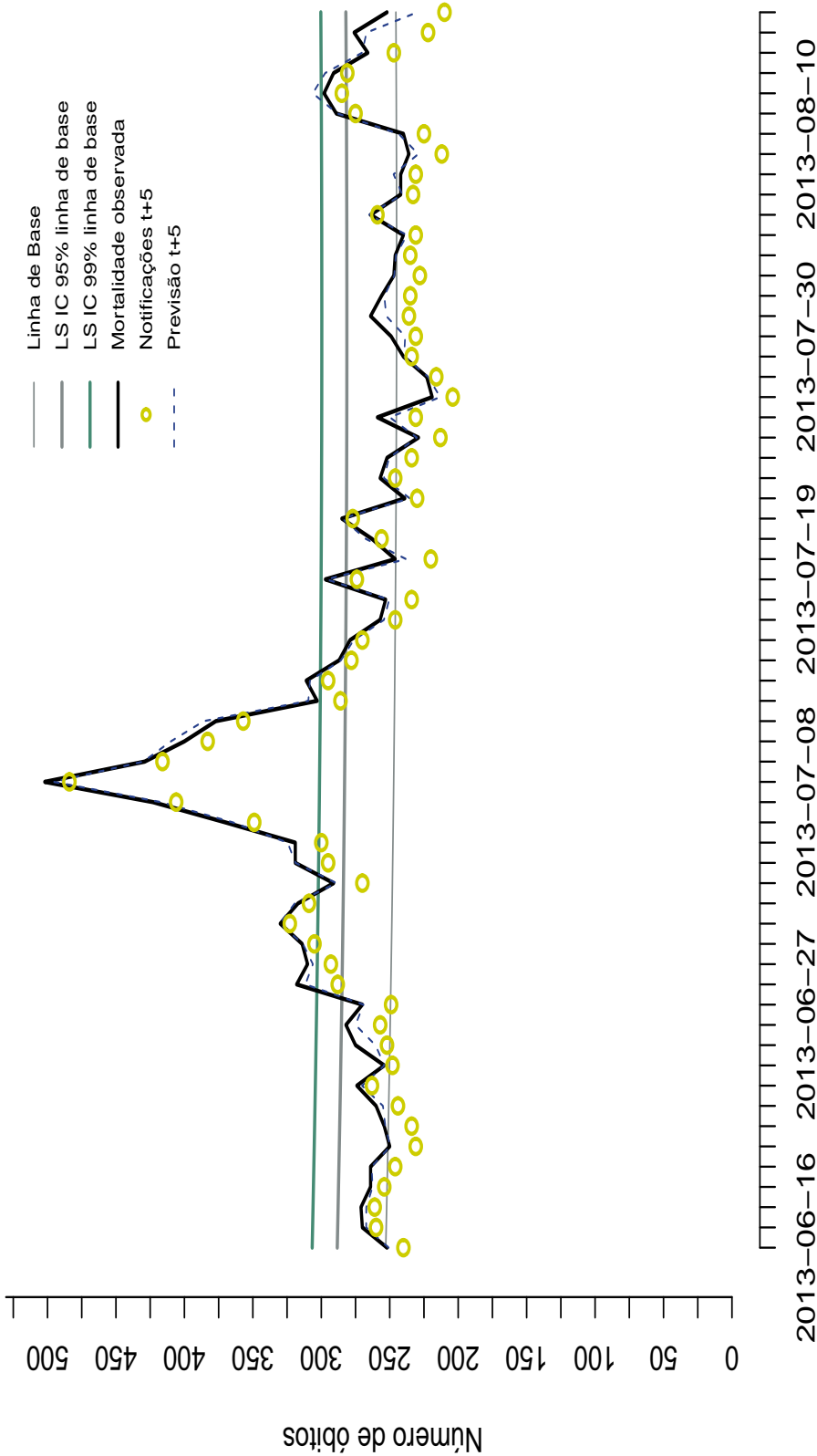
Figura 5.39: Previsões e notificações a  $t + 4$  durante a Onda de Calor em 2013

Figura 5.40: Previsões e notificações a  $t + 5$  durante a Onda de Calor em 2013



---

### Discussão, conclusão e desenvolvimentos futuros

---

#### 6.1 Discussão

De acordo com os resultados obtidos no capítulo anterior foi possível através da metodologia proposta obterem-se estimativas de qualidade de uma forma geral e diminuir o tempo até à detecção de impactes de eventos associados a excessos de mortalidade comparativamente com sistema VDM atual.

A análise realizada às previsões obtidas permitiu constatar a sua qualidade, ou seja, com um erro relativo mediano inferior a 10%, aumenta com o número de dias que se recebeu informação e, também, que esta produzirá estimativas com qualidade se pelo menos um dos dias anteriores à previsão for um dia útil.

As previsões mais afetadas com falta de informação são as previsões no dia seguinte e após dois dias da data que se quer estimar, correspondendo aproximadamente a 31% e a 15% respetivamente do total das previsões efetuadas nestes momentos de previsão durante um ano. Por outro lado, previsões efetuadas sem qualquer informação são muito pouco frequentes ( $\leq 1\%$ ). Portanto, o cenário mais frequente na prática é a produção de estimativas com pelo menos um dia de informação, correspondendo a 69% e a 85% das previsões efetuadas em  $t + 1$  e  $t + 2$ , e a 99% dos restantes momentos de previsão durante um ano, levando a estimativas com qualidade para a maioria das previsões.

Para as previsões sem informação, recorreu-se a um outro estimador que assumia



pelo menos um óbito notificado (equação 3.3). Dado que as estimativas com um erro relativo superior a 10% correspondem essencialmente às estimativas efetuadas sem informação e, consequentemente, através deste estimador, esta alternativa para os dias com notificações nulas não é viável visto que não mostrou produzir bons resultados.

O estudo desenvolvido no âmbito do sistema de vigilância da mortalidade na Escócia [3], que utiliza o mesmo estimador para o número de óbitos em dados diários, não refere qualquer problema relativamente ao número nulo de notificações. Dado que também neste país não é possível fazer-se o registo de um óbito ao fim-de-semana, o método utilizado também deveria sofrer da mesma limitação. Na verdade, a razão para tal deve-se ao facto de os autores terem optado por produzir previsões nestas circunstâncias, aliás, apenas são efetuadas previsões considerando os dias úteis anteriores à previsão. O método utilizado neste estudo atualiza a informação já notificada ao sistema, acabando por não sofrer por falta de informação e, consequentemente por não prever diariamente o número de óbitos ocorridos no país.

Através dos resultados obtidos também foi possível constatar que a metodologia permitiu a diminuição do tempo até à deteção do início do excesso de mortalidade durante os dois eventos relacionados com a saúde ocorridos em 2012 e 2013.

Em 2012, o início do excesso de mortalidade associada à epidemia de gripe, ocorrida a 4 de Fevereiro, foi detetada três dias depois através da metodologia e através do sistema atual (sem correção para os atrasos), seria detetado 10 dias depois. Como tal, a metodologia permitiu a deteção do início de excesso de mortalidade 7 dias mais cedo.

Em 2013, o início do excesso de mortalidade associada à onda de calor, ocorrido a 28 de Junho, foi detetado no dia seguinte através da metodologia. Através do sistema atual, o início da mortalidade associada a este evento só seria detetado dia 4 de Julho. A metodologia permitiu, então, detetar o início do excesso de mortalidade 6 dias mais cedo.

Relativamente aos fatores que se mostraram associados a cada atraso, o efeito dos mesmos variou nos dois testes, o teste para 2012 e o teste para 2013, apesar de não ter variado substancialmente. A razão para tal pode residir na variabilidade do atraso na notificação ao longo dos anos. Apesar de na análise preliminar dos dados esta não sugerir uma variação sistemática, pode, de qualquer forma, alterar os valores dos coeficientes ao incluir-se mais um ano nos dados que são utilizados para a estimação dos mesmos.

Os efeitos que mais se alteram comparando os dois testes foram os efeitos associados aos feriados. Uma interpretação pode estar associada ao facto de o dia da semana em

que os feriados ocorrem difere de ano para ano e de existirem feriados móveis. Desta forma, o dia da semana em que ocorrem relativamente à data de morte pode influenciar atrasos distintos em diversos anos, alterando assim o efeito dos feriados de ano para ano.

Os fatores considerados para a explicação de cada atraso mostraram estar associados a pelo menos um dos atrasos nestes primeiros cinco dias de previsão.

Ser feriado no dia da morte e ser feriado nos dias anteriores à previsão mostraram ser importantes para explicar todos os atrasos e diminuir a probabilidade de um óbito ser notificado. Este resultado era expectável visto que ser feriado implica o encerramento das conservatórias e, consequentemente, impossibilitam o registo de um óbito.

Ser feriado no dia da previsão apenas mostrou ser significativo para os óbitos que são notificados no dia seguinte à data de morte, aumentando a probabilidade desta notificação. Este resultado era expectável, dado que ser feriado leva ao encerramento das conservatórias e, consequentemente, impossibilita o registo no dia seguinte ao óbito, levando a um maior número de óbitos que são registados no dia da morte. A mesma interpretação é dada ao efeito da sexta-feira como dia de ocorrência de morte, também para este atraso, que aumenta a probabilidade de um óbito ser notificado no dia seguinte à morte. Uma vez que, se este não for registado no dia da morte, sexta-feira, só poderá ser registado na segunda-feira seguinte, três dias depois. É importante referir que o funeral do óbito só se poderá realizar após o registo do mesmo, o que também justifica o efeito associado a estes dois dias. Se um óbito ocorrido a uma sexta-feira for registado no próprio dia, o funeral poderá realizar-se durante o fim-de-semana, caso contrário, apenas se poderá realizar a partir da segunda-feira seguinte. O mesmo raciocínio é feito para o efeito de ser feriado no dia seguinte à morte, uma vez que, se o óbito for registado no dia de morte, o funeral poderá realizar-se no dia seguinte, um feriado nacional, ou o mais cedo possível.

O dia da semana de morte também mostrou ser importante para explicar cada atraso. É no atraso de um dia na notificação que o efeito dos dias úteis (segunda a sexta-feira) mostra ser mais acentuado, comparativamente com o fim-de-semana. Quanto maior é o atraso na notificação, menor é o efeito dos dias úteis na notificação e, eventualmente, este fator deixa de ser importante para explicar cada atraso. Aliás, após  $t + 7$ , inclusive, todas as variáveis perdem significância, indicando que os óbitos com atrasos superiores a 7 dias não diferem relativamente ao dia da semana de ocorrência nem aos dias de feriado. Este resultado também pode indicar que estes são os que ocorrem no estrangeiro, ou os que são encontrados tardiamente e sujeitos a autópsia, visto que não dependem dos

## *CAPÍTULO 6. DISCUSSÃO, CONCLUSÃO E DESENVOLVIMENTOS FUTUROS*

---

memos fatores que foram considerados influenciar o processo de registo e, por isso, devem depender da circunstância de ocorrência do óbito. Como foi referido, os dois fatores gerais que influenciam a notificação são os fatores associados ao registo e os fatores associados à circunstância de ocorrência do óbito.

A probabilidade de detetar um excesso de mortalidade aumenta com o número de dias que se recebe informação, como seria de esperar. Os piores cenários para a deteção de excesso de mortalidade são os momentos de previsão sem informação em ambos os anos. Os primeiros momentos de previsão em que é mais provável a deteção de um excesso de mortalidade são em  $t + 2$  com informação completa ou com um dia de informação. Após  $t + 2$ , desde que exista pelo menos um dia de informação, a probabilidade de se detetar um excesso de mortalidade é elevada.

A especificidade em ambos os teste é sempre superior a 0.83 independentemente do número de dias sem informação. O problema da metodologia reside nos falsos negativos que poderá estar relacionada com a impossibilidade de se efetuarem previsões sem informação e da qualidade das previsões diminuir quando o número de dias com informação diminui. A especificidade também é influenciada por esta limitação, que leva a um maior número de previsões muito abaixo da mortalidade observada, estando sempre em conformidade com a existência de dias sem excesso de mortalidade.

A probabilidade de um excesso de mortalidade detetado pela metodologia ser real também aumenta com o número de dias que se recebe informação, como seria de esperar. Os piores cenários para o Valor Preditivo Positivo são em  $t + 1$ , com e sem informação, e em  $t + 2$  sem informação. Na prática, as primeiras previsões de excesso de mortalidade mais fidedignas são em  $t + 2$  com informação, ou em  $t + 3$ , com dois dias de informação ou com informação completa. A probabilidade de se detetar um verdadeiro dia sem excesso de mortalidade através da metodologia foi sempre elevada. O facto do Valor Preditivo Negativo ser sempre elevado, mesmo sem informação, é possivelmente fruto da limitação da metodologia, que sem informação tende a subestimar o número de óbitos ocorridos.

Dado que o estimador proposto para os dias sem que não é rececionada informação não se mostrou viável, deve ser discutido um método alternativo para estes dias. Em primeiro lugar deve discutir-se os métodos revistos para o *Nowcasting* dos óbitos noutros sistemas de vigilância da mortalidade. O projeto EuroMOMO utiliza uma metodologia que obtém a estimativa do número de óbitos ocorridos a dia  $t$  através da modelação

dos atrasos e, posteriormente, da modelação do número de óbitos, em alternativa à estimação pontual utilizada pela metodologia adotada. A modelação do número de óbitos, definindo que número de óbitos em cada dia segue uma distribuição Poisson e, considerando como covariáveis: a probabilidade de um óbito ser notificado com atraso igual ou inferior a  $i$  dias ( $p_i$ ) e o número de óbitos já notificados até dia  $i$  ( $n_{t,i}$ )

$$\log(M_{t,i}) = \beta_0 + \beta_1 \times p_i + \beta_2 \times n_{t,i}.$$

Este método permite fazer estimativas do número de óbitos, mesmo quando não se recebe informação, mas quando tal acontece, as estimativas acabam por ser um valor médio da mortalidade em vez de uma verdadeira variação da mortalidade, pois quando  $n_{t,i} = 0$  e  $p_i \approx 0$ , o que acontece ao fim-de-semana e aos feriados, a equação é dada por

$$\log(M_{t,i}) = \beta_0.$$

Neste caso, as previsões elaboradas sem informação não permitem a estimativa de verdadeiras variações no valor da mortalidade, e conseqüentemente, não permitem a deteção de excesso de mortalidade. Dado isto, este método não resolve as limitações da metodologia adotada, pois também aqui as estimativas com qualidade dependem da chegada de informação e de uma notificação regular.

### **6.1.1 Limitações**

Uma limitação direta da metodologia é não produzir estimativas sem receção de informação, ou seja, quando

$$n_{t,i} = 0, \quad M_{t,i} = \frac{n_{t,i}}{\hat{p}_i} = \frac{0}{\hat{p}_i} = 0,$$

e por isso, não permite fazer uma estimativa do número de óbitos ocorridos. Para contornar esta limitação, substituiu-se o valor de  $n_{t,i}$  quando este era nulo, por 1, ou seja,

$$n_{t,i} = 1, \quad M_{t,i} = \frac{n_{t,i}}{\hat{p}_i} - 1 = \frac{1}{\hat{p}_i} - 1.$$

O segundo estimador não melhorou a qualidade das estimativas, visto que as observações discrepantes detetadas na implementação da metodologia e o erros relativos superiores a 10% diziam respeito especificamente aos dias em que este era utilizado.

A segunda limitação diz respeito à receção da informação. A metodologia não pro-

duzirá estimativas com qualidade se a receção da informação for irregular. Como se viu, as estimativas discrepantes na implementação da metodologia para os dois anos diziam respeito não só às condições enunciadas em cima, mas também à receção irregular de informação. Esta condição ocorre quando a quantidade de informação que é recebida está muito abaixo ou muito acima do esperado para um dia com determinadas condições (as variáveis que foram consideradas para explicar cada atraso). A metodologia assenta no pressuposto de que cada atraso segue uma distribuição binomial e, por isso, a probabilidade de cada atraso é estacionária para as mesmas condições. A consequência deste pressuposto faz com que as notificações que chegam nas mesmas condições (com os mesmos valores para as covariáveis) tenham a mesma probabilidade de sucesso, ou seja, vão representar a mesma proporção relativamente ao total de óbitos para aquele dia. Portanto, quando chegam notificações abaixo do esperado para determinadas condições, a metodologia leva a que sejam consideradas como baixa mortalidade para aquele dia. O mesmo raciocínio é feito para as notificações muito acima do esperado para determinadas condições.

Outra limitação é a heterogeneidade nos dados que não foi explicada pelos modelos considerados, levando a valores dos parâmetros de dispersão associados a cada modelo de previsão muito elevados ( $\phi \gg 1$ ).

A causa pode assentar em falta de informação, ou seja, não terem sido consideradas covariáveis importantes para explicar os atrasos e desta forma, o modelo seleccionado não consegue explicar a heterogeneidade das observações através das covariáveis escolhidas.

Como foi referido, os dois fatores gerais que influenciam a notificação são os fatores associados ao registo e os fatores associados à circunstância de ocorrência do óbito. Neste estudo apenas foram incluídos os fatores associados ao registo, visto que para a previsão são os únicos fatores a que se tem acesso. Esta também pode ser uma fonte de variabilidade nos dados que não consegue ser explicada pelos modelos.

A variação não explicada também poderá estar na influência da idade, sexo ou região do indivíduo. Uma forma de controlar a heterogeneidade provocada por estas variáveis, seria construir-se modelos introduzindo estas variáveis, e obterem-se estimativas da probabilidade de cada atraso por região, sexo e grupo etário. Calcular-se-ia, através da metodologia adotada, o número de óbitos por estrato e, posteriormente, o número total de óbitos nacional. Este seria dado pela soma das estimativas calculadas por estrato. No entanto, esta abordagem pode levar a uma sobrestimação do número total de óbitos e será mais afetada pela falta de notificação. Na prática, chegar informação não implica

que chegue informação de óbitos ocorridos em todas as regiões e, consequentemente, o número de notificações para determinada região poderá nula, tal como a respetiva estimativa dos óbitos, faltando informação para a estimativa do número total de óbitos nacional.

Outra fonte de variação poderá estar na impossibilidade de retirar da Base dados os óbitos ocorridos fora do País. O sistema VDM não permite fazer distinção relativamente a estes óbitos e, por isso, estes foram incluídos na análise efetuada. Como foi descrito, as variáveis incluídas dizem respeito essencialmente a fatores que poderiam influenciar o funcionamento das conservatórias em Portugal. É desconhecido se o procedimento de registo de um óbito ocorrido fora do país e se a agregação desta informação é influenciada pelos mesmos fatores que as conservatórias em Portugal e, por isso, pode ser um fator que introduz variação não explicada nos modelos. No entanto, estes óbitos devem ser residuais comparativamente com os óbitos ocorridos em Portugal.

É de referir que apenas foram incluídos os feriados nacionais, no entanto, um feriado municipal também poderá influenciar o atraso na notificação. Esta pode ser outra fonte de variação que não conseguiu ser explicada pelos modelos atuais.

A ultima limitação deve-se ao facto de as definições utilizadas para a ocorrência de um evento relacionado com a saúde incluírem não só os períodos com excesso de mortalidade mas também todos os alarmes dados pelos respetivos sistemas de vigilância. A influência deste fator, na teoria, dizia respeito ao excesso de mortalidade e à influência que um maior número de registos teria no procedimento regular das conservatórias. Dado isto, o efeito destes fatores pode ter enviesado os resultados, levando à não significância dos mesmos. Na prática, poderiam ser incluídos os períodos em que ocorreu excesso de mortalidade para estimar cada um dos modelos, mas não se teria acesso a essa informação diariamente para a previsão, dado que é desconhecido. Por isso, a informação mais fidedigna que se teria diariamente na prática, diria respeito aos alarmes dados pelos sistemas de vigilância implementados no DEP.

## **6.2 Conclusão**

No que diz respeito aos objetivos deste projeto, a metodologia proposta mostrou ser capaz de concretizar parcialmente o objetivo principal e mostrou ser capaz de concretizar o objetivo secundário.

O objetivo principal, que diz respeito à estimação diária do número de óbitos ocor-

ridos através da modelação dos atrasos, foi atingindo parcialmente. A metodologia proposta mostrou ser capaz de estimar o número de óbitos com qualidade, ou seja, com um erro relativo inferior a 10%, mas não mostrou ser capaz de o fazer diariamente. Os resultados obtidos mostraram que a metodologia depende de pelo menos um dia útil anterior ao dia de previsão para poder produzir estimativas com qualidade.

Durante um ano, o mais frequente é a produção de estimativas sem falta de informação nos momentos de  $t + 1$  a  $t + 3$  e, a partir de  $t + 3$ , são raros os dias em que estimativas são produzidas sem informação.

A produção de estimativas com qualidade também depende da receção de informação de uma forma regular. Nos dois anos considerados para a avaliação da metodologia, este cenário ocorreu duas vezes, uma em cada ano estudado.

Em suma, a metodologia produz estimativas com qualidade quando pelo menos um dos dias considerados para as mesmas foi um dia útil e a transmissão da informação foi regular. Uma vez estas são as condições mais frequentes durante um ano, pode considerar-se que de uma forma geral, a metodologia terá um bom desempenho.

O objetivo secundário, que diz respeito à diminuição do tempo até à deteção de excesso da mortalidade através da metodologia, foi atingido. Os resultados mostraram que para os dois eventos relacionados com a saúde, epidemia de gripe em 2012 e onda de calor em 2013, o início do excesso de mortalidade foi detetado mais cedo através da metodologia.

Desta forma, o objetivo do sistema VDM, que consiste na deteção e estimação da magnitude do impacte destes eventos, pode ser realizado em menos tempo e, por isso, o *timeliness* do sistema VDM foi melhorado. Visto que este é uma das características mais importantes para este sistema, a metodologia mostra ser útil e cumpre o objetivo mais pertinente do projeto.

### 6.3 Desenvolvimentos futuros

O primeiro desenvolvimento futuro passa pelo desenvolvimento de um método alternativo a ser utilizado nos dias em que não se recebe informação.

Este método pode passar por assumir que para estes dias, sem informação ou informação irregular, a mortalidade estará mais próxima da estimativa elaborada pela metodologia para os dias anteriores. Na prática passa por, nos dias em que a metodolo-

gia depende apenas de informação nula ou irregular, modelar a mortalidade em função dos óbitos registados e previstos para os dias anteriores. Esta solução também não poderá detetar uma verdadeira variação na mortalidade, mas melhorará o aspeto da metodologia atual na sua implementação futura no sistema.

O segundo desenvolvimento passa pela estimação estratificada do número de óbitos ocorridos em cada dia, através da introdução de outras variáveis associadas aos óbitos nos modelos atuais. Este passo tem como objetivo diminuir a heterogeneidade não explicada pelos modelos atuais e até poder apresentar-se e analisar-se a informação da mortalidade por cada estrato, o que seria um desenvolvimento interessante. Nesta perspetiva, o número total de óbitos ocorridos em Portugal para um determinado dia seria dado pela soma das estimativas por estrato. Se o método que se propõe desenvolver, para a correção dos dias sem qualquer informação, mostrar bons resultados, esta abordagem não deverá sofrer com maior falta de informação. Poderá, no entanto, levar a uma sobrestimação do número total de óbitos.

O terceiro desenvolvimento futuro será a implementação da metodologia no sistema VDM, que passa pela inclusão das estimativas obtidas diariamente no boletim da Vigilância Diária da Mortalidade. Apesar de nesta fase a metodologia não produzir estimativas diariamente, permite a deteção precoce de excesso de mortalidade, em comparação com o sistema atual e, como tal, mostrou-se útil. Para uma apresentação adequada da informação, as estimativas devem ser calculadas para Portugal, Portugal Continental e Regiões NUTS II, já que o Boletim apresenta a informação para estes estratos. A implementação da mesma não será computacionalmente pesada visto que apenas são necessários os parâmetros estimados para cada modelo de previsão. Os modelos podem e devem ser atualizados de ano para ano, de forma a atualizar os efeitos estimados para novas variações na notificação.

Por fim, seria interessante incluir na implementação da metodologia no sistema VDM, uma medida de fiabilidade da previsão que se está a obter. Esta poderá ser dada pelo erro relativo mediano percentual para as estimativas efetuadas em determinadas condições (por exemplo, em função do número de dias que não foi rececionada informação). Num contexto de excesso de mortalidade poderia incluir-se também o Valor Preditivo Positivo para as previsões efetuadas em determinadas condições. Desta forma, quando um excesso é detetado pela metodologia, seria possível dizer-se a probabilidade deste ser verdadeiro. Na prática, este seria o resultado mais pertinente deste



trabalho.

## **6.4 Considerações finais**

Gostava de referir que este trabalho será apresentado, em forma de poster, na *International Biometric Conference*, que se irá realizar em Florença, Itália em Julho de 2014.

---

## Bibliografia

---

- [1] Centers for Disease Control and Prevention. *Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group*, MMWR, 2001; 18(15):pii=20451. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20451>
- [2] Mazick A, Participants of a workshop on mortality monitoring in Europe. *Monitoring excess mortality for public health action: potential for a future European network*. Euro Surveillance. 2007; 12(1):pii=3107. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=3107>
- [3] Wagner A P, McKenzie E, Robertson C, McMenamin J, Reynolds A, Murdoch H. *Automated mortality monitoring in Scotland from 2009*. Euro Surveill. 2013; 18(15):pii=20451. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20451>
- [4] Nunes B, Natário I, Carvalho L M. *Nowcasting influenza epidemics using non-homogeneous hidden Markov models*. Statist. Med. 2013; 32 2643–2660. Available online: <http://onlinelibrary.wiley.com/doi/10.1002/sim.5670/pdf>
- [5] Instituto dos Registos e Notariado, Ministério da Justiça. Acedido a 14 Mar 2014. Disponível em: <http://www.irn.mj.pt/sections/inicio>
- [6] Nogueira P J, Machado A, Rodrigues E, Nunes B, Sousa L, Jacinto M, Ferreira A, Falcão J M, Ferrinho P. *The new automated daily mortality surveillance system in Portugal*. Euro Surveill; 2010; 15(13):pii=19529. Disponível em: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19529>

## BIBLIOGRAFIA

---

- [7] Donker T, Boven M, Ballegooijen M W, Klooster M T, Wielders C C, Wallinga J. *Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands*, J Epidemiol Community Health 2012;66:866e868. doi:10.1136/jech-2011-200962.
- [8] Green K H, Andrews J N, Bickler G, Pebody G R. *Rapid estimation of excess mortality: nowcasting during the heatwave alert in England and Wales in June 2011*, J Epidemiol Community Health 2012;66:866e868. doi:10.1136/jech-2011-200962.
- [9] Kanieff M, Rago G, Minelli G, Lamagni T, Sadicova O, Šelb J, Vantarakis A, Conti S. *The potential for a concerted system for the rapid monitoring of excess mortality throughout Europe*. Euro Surveill; 2010; 15(43):pii=19697. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19697>
- [10] EuroMOMO. *European monitoring of excess mortality for public health action*, Acedido a 14 Mar 2014. <http://www.euromomo.eu/index.html>
- [11] Departamento de Epidemiologia, Instituto Nacional de Saúde Dr. Ricardo Jorge. Acedido a 17 Mar 2014. Disponível em: <http://www.insa.pt/sites/INSA/Portugues/AreasCientificas/Epidemiologia/Unidades/UnInstrObser/Paginas/ICARO.aspx>
- [12] Instituto Nacional de Saúde Dr. Ricardo Jorge. Boletim de Vigilância Epidemiológica da Gripe, Época 2013/2014, Semana 16 - de 14/04/2014 a 20/04/2014. Acedido a 29 ABR 2014. Disponível em: [http://www.insa.pt/sites/INSA/Portugues/Documents/2014\\_S16.pdf](http://www.insa.pt/sites/INSA/Portugues/Documents/2014_S16.pdf)
- [13] Nunes B, Natario I, Carvalho L. flubase: *Baseline of mortality free of influenza epidemics*. 2009: R package version 1.0.
- [14] James Westguard, Westgard QC. Acedido a 10 FEV 2014. Disponível em: <http://www.westguard.com/westguard-rules/>
- [15] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0. Acedido 10 JAN 2014. Disponível em: <http://www.R-project.org>
- [16] Lombardo J S, Buckeridge D L. *Disease Surveillance: A Public Health Informatics Approach*, Wiley, 2007.
- [17] Turkman M, Silva G. Modelos lineares Generalizados - da teoria à prática, Lisboa, 2000.

- [18] Hosmer D, Lemeshow S, Sturdivant R. *Applied Logistic Regression*, Third Edition, Wiley, 2013.

## ANEXO A

### Restantes Modelos de previsão

Tabela A.1: Sumário do modelo de previsão 6 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.7336	0.1321	28.25	0.0000
Segunda	-0.1115	0.1783	-0.63	0.5349
Terça	-0.4366	0.1691	-2.58	0.0134
Quarta	-0.5303	0.1669	-3.18	0.0028
Quinta	-0.5999	0.1649	-3.64	0.0007
Sexta	-0.5380	0.1668	-3.23	0.0024
Sábado	-0.2801	0.1737	-1.61	0.1144
Feriado morte	-0.1460	0.2299	-0.64	0.5288
Feriado dia anterior à previsão	-0.2936	0.2199	-1.34	0.1890
Feriado dois dias antes da previsão	-0.7711	0.1902	-4.05	0.0002
Feriado três dias antes da previsão	-0.5075	0.2012	-2.52	0.0155
Feriado quatro dias antes da previsão	-0.8178	0.1840	-4.44	0.0001
Feriado dia da previsão	-0.0298	0.2425	-0.12	0.9028
$\phi$ : 22.69144				
Null deviance: 2122.76 on 54 degrees of freedom				
Residual deviance: 800.57 on 42 degrees of freedom				
<sup>a</sup> Referência: Domingo				

ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.2: Sumário do modelo de previsão 7 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.6377	0.0915	39.74	0.0000
Segunda	-0.1099	0.1243	-0.88	0.3811
Terça	-0.0367	0.1275	-0.29	0.7749
Quarta	-0.0171	0.1280	-0.13	0.8946
Quinta	-0.0404	0.1270	-0.32	0.7517
Sexta	-0.1640	0.1239	-1.32	0.1919
Sábado	-0.0304	0.1271	-0.24	0.8121
Feriado morte	-0.0482	0.1963	-0.25	0.8070
Feriado dia anterior à previsão	-0.1550	0.1857	-0.83	0.4081
Feriado dois dias antes da previsão	-0.1759	0.1845	-0.95	0.3454
Feriado três dias antes da previsão	-0.1797	0.1803	-1.00	0.3241
Feriado quatro dias antes da previsão	-0.2846	0.1719	-1.66	0.1045
Feriado previsão	0.0444	0.2053	0.22	0.8297
$\phi$ : 10.90872				
Null deviance: 358.77 on 59 degrees of freedom				
Residual deviance: 265.50 on 47 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.3: Sumário do modelo de previsão 8 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.6161	0.0960	37.68	0.0000
Segunda	0.0994	0.1361	0.73	0.4687
Terça	0.2146	0.1412	1.52	0.1354
Quarta	0.2995	0.1444	2.07	0.0436
Quinta	0.1701	0.1396	1.22	0.2290
Sexta	0.0375	0.1360	0.28	0.7838
Sábado	-0.0262	0.1338	-0.20	0.8456
Feriado morte	-0.0041	0.2286	-0.02	0.9857
Feriado dia anterior à previsão	-0.0981	0.2234	-0.44	0.6625
Feriado dois dias antes da previsão	-0.1276	0.2142	-0.60	0.5543
Feriado dois dias antes da previsão	0.1198	0.2308	0.52	0.6062
Feriado três dias antes da previsão	-0.0679	0.2123	-0.32	0.7507
Feriado quatro dias antes da previsão	-0.1078	0.2204	-0.49	0.6269
$\phi$ : 10.90872				
Null deviance: 267.80 on 59 degrees of freedom				
Residual deviance: 156.58 on 47 degrees of freedom				
<sup>a</sup> Referência: Domingo				

## ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.4: Sumário do modelo de previsão 9 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.7443	0.1035	36.16	0.0000
Segunda	0.1310	0.1486	0.88	0.3825
Terça	0.2396	0.1533	1.56	0.1249
Quarta	0.2910	0.1553	1.87	0.0672
Quinta	0.1972	0.1524	1.29	0.2018
Sexta	-0.0810	0.1438	-0.56	0.5756
Sábado	-0.1574	0.1406	-1.12	0.2685
Feriado morte	-0.0003	0.2451	-0.00	0.9991
Feriado dia anterior à previsão	-0.2166	0.2290	-0.95	0.3490
Feriado dois dias antes da previsão	-0.1218	0.2368	-0.51	0.6096
Feriado três dias antes da previsão	-0.1219	0.2253	-0.54	0.5909
Feriado quatro dias antes da previsão	0.1779	0.2503	0.71	0.4808
Feriado no dia da previsão	0.0312	0.2449	0.13	0.8992
$\phi$ : 12.36128				
Null deviance: 352.75 on 59 degrees of freedom				
Residual deviance: 145.95 on 47 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.5: Sumário do modelo de previsão 10 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.8529	0.1043	36.95	0.0000
Segunda	0.1278	0.1506	0.85	0.4004
Terça	0.2426	0.1554	1.56	0.1253
Quarta	0.2771	0.1565	1.77	0.0831
Quinta	0.1049	0.1507	0.70	0.4896
Sexta	-0.1912	0.1421	-1.35	0.1850
Sábado	-0.1571	0.1423	-1.10	0.2754
Feriado morte	-0.0111	0.2420	-0.05	0.9637
Feriado dia anterior à previsão	-0.0421	0.2419	-0.17	0.8624
Feriado dois dias antes da previsão	-0.2171	0.2287	-0.95	0.3473
Feriado três dias antes da previsão	-0.0805	0.2385	-0.34	0.7374
Feriado quatro dias antes da previsão	-0.1039	0.2235	-0.46	0.6442
Feriado no dia da previsão	0.1854	0.2580	0.72	0.4759
$\phi$ : 11.47607				
Null deviance: 355.78 on 59 degrees of freedom				
Residual deviance: 134.31 on 47 degrees of freedom				
<sup>a</sup> Referência: Domingo				

ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.6: Sumário do modelo de previsão 11 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.9225	0.0973	40.31	0.0000
Segunda	0.1245	0.1394	0.89	0.3766
Terça	0.2638	0.1449	1.82	0.0756
Quarta	0.2058	0.1423	1.45	0.1553
Quinta	0.0239	0.1369	0.17	0.8622
Sexta	-0.1593	0.1320	-1.21	0.2340
Sábado	-0.1381	0.1321	-1.05	0.3015
Feriado morte	-0.0006	0.2233	-0.00	0.9979
Feriado dia anterior à morte	0.1106	0.2334	0.47	0.6379
Feriado dois dias antes da previsão	-0.0090	0.2257	-0.04	0.9682
Feriado três dias antes da previsão	-0.1930	0.2102	-0.92	0.3638
Feriado quatro dias antes da previsão	-0.0541	0.2212	-0.24	0.8078
Feriado no dia da previsão	0.2175	0.2382	0.91	0.3661
$\phi$ : 9.111915				
Null deviance: 273.02 on 56 degrees of freedom				
Residual deviance: 113.03 on 44 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.7: Sumário do modelo de previsão 12 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.9734	0.1125	35.32	0.0000
Segunda	0.1391	0.1610	0.86	0.3930
Terça	0.2121	0.1646	1.29	0.2053
Quarta	0.1460	0.1615	0.90	0.3716
Quinta	0.0565	0.1584	0.36	0.7231
Sexta	-0.1349	0.1523	-0.89	0.3811
Sábado	-0.1011	0.1534	-0.66	0.5139
Feriado morte	-0.0022	0.2529	-0.01	0.9932
Feriado no dia anterior à previsão	0.1471	0.2687	0.55	0.5872
Feriado dois dias antes da previsão	0.1049	0.2682	0.39	0.6977
Feriado três dias antes da previsão	0.0400	0.2633	0.15	0.8802
Feriado quatro dias antes da previsão	-0.1500	0.2451	-0.61	0.5441
Feriado no dia da previsão	0.1652	0.2670	0.62	0.5396
$\phi$ : 11.41681				
Null deviance: 223.12 on 51 degrees of freedom				
Residual deviance: 120.77 on 39 degrees of freedom				
<sup>a</sup> Referência: Domingo				



## ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.8: Sumário do modelo de previsão 13 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.9975	0.1305	30.62	0.0000
Segunda	0.1101	0.1862	0.59	0.5576
Terça	0.1796	0.1903	0.94	0.3512
Quarta	0.1816	0.1902	0.95	0.3455
Quinta	0.0896	0.1859	0.48	0.6326
Sexta	-0.0888	0.1796	-0.49	0.6237
Sábado	-0.0849	0.1793	-0.47	0.6387
Feriado morte	0.0245	0.3000	0.08	0.9352
Feriado dia anterior à previsão	0.1222	0.3086	0.40	0.6942
Feriado dois dias antes da previsão	0.1707	0.3176	0.54	0.5939
Feriado três dias antes da previsão	0.1332	0.3157	0.42	0.6753
Feriado quatro dias antes da previsão	0.0686	0.3111	0.22	0.8266
Feriado no dia da previsão	0.0754	0.3035	0.25	0.8051
$\phi$ : 15.01242				
Null deviance: 209.60 on 51 degrees of freedom				
Residual deviance: 131.62 on 39 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.9: Sumário do modelo de previsão 14 - Teste 2012

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	4.0012	0.1324	30.23	0.0000
Segunda	0.1043	0.1897	0.55	0.5856
Terça	0.2362	0.1972	1.20	0.2382
Quarta	0.2376	0.1966	1.21	0.2340
Quinta	0.1426	0.1927	0.74	0.4637
Sexta	-0.0341	0.1856	-0.18	0.8553
Sábado	-0.0440	0.1843	-0.24	0.8126
Feriado morte	0.0602	0.3179	0.19	0.8509
Feriado dia anterior à previsão	0.0235	0.3105	0.08	0.9400
Feriado dois dias antes da previsão	0.1113	0.3181	0.35	0.7283
Feriado três dias antes da previsão	0.1739	0.3284	0.53	0.5995
Feriado quatro dias antes da previsão	0.1654	0.3302	0.50	0.6191
Feriado dia da previsão	0.0252	0.3195	0.08	0.9374
$\phi$ : 15.5526				
Null deviance: 200.77 on 51 degrees of freedom				
Residual deviance: 120.20 on 39 degrees of freedom				
<sup>a</sup> Referência: Domingo				

ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.10: Sumário do modelo de previsão 6 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.7351	0.1165	32.07	0.0000
Segunda	-0.1197	0.1573	-0.76	0.4511
Terça	-0.4455	0.1491	-2.99	0.0047
Quarta	-0.5315	0.1475	-3.60	0.0008
Quinta	-0.5801	0.1460	-3.97	0.0003
Sexta	-0.5378	0.1471	-3.66	0.0007
Sábado	-0.2691	0.1539	-1.75	0.0878
Feriado morte	-0.1248	0.2104	-0.59	0.5565
Feriado dia anterior à previsão	-0.2629	0.1960	-1.34	0.1871
Feriado dois dias antes da previsão	-0.7173	0.1721	-4.17	0.0002
Feriado três dias antes da previsão	-0.5107	0.1807	-2.83	0.0072
Feriado quatro dias antes da previsão	-0.7553	0.1698	-4.45	0.0001
Feriado dia da previsão	0.0440	0.2195	0.20	0.8420
$\phi$ : 22.66029				
Null deviance: 2381.15 on 54 degrees of freedom				
Residual deviance: 845.72 on 42 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.11: Sumário do modelo de previsão 7 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.6519	0.0656	55.64	0.0000
Segunda	-0.1192	0.0892	-1.34	0.1879
Terça	-0.0475	0.0914	-0.52	0.6053
Quarta	-0.0496	0.0914	-0.54	0.5900
Quinta	-0.0635	0.0908	-0.70	0.4877
Sexta	-0.1712	0.0888	-1.93	0.0598
Sábado	-0.0179	0.0917	-0.19	0.8464
Feriado morte	-0.0779	0.1431	-0.54	0.5888
Feriado dia anterior à previsão	-0.1433	0.1344	-1.07	0.2915
Feriado dois dias antes da previsão	-0.1296	0.1348	-0.96	0.3414
Feriado três dias antes da previsão	-0.1799	0.1310	-1.37	0.1762
Feriado quatro dias antes da previsão	-0.2929	0.1257	-2.33	0.0242
Feriado no dia da previsão	0.1457	0.1526	0.95	0.3446
$\phi$ : 7.196596				
Null deviance: 376.16 on 59 degrees of freedom				
Residual deviance: 257.95 on 47 degrees of freedom				
<sup>a</sup> Referência: Domingo				

## ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.12: Sumário do modelo de previsão 8 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.6308	0.0563	64.51	0.0000
Segunda	0.0803	0.0798	1.01	0.3191
Terça	0.1952	0.0827	2.36	0.0223
Quarta	0.2352	0.0836	2.81	0.0071
Quinta	0.1539	0.0817	1.88	0.0656
Sexta	0.0446	0.0800	0.56	0.5793
Sábado	-0.0149	0.0788	-0.19	0.8510
Feriado morte	-0.0338	0.1362	-0.25	0.8052
Feriado dia anterior à previsão	-0.0176	0.1352	-0.13	0.8972
Feriado dois dias antes da previsão	-0.1075	0.1273	-0.84	0.4024
Feriado três dias antes da previsão	0.1353	0.1369	0.99	0.3280
Feriado quatro dias antes da previsão	-0.0701	0.1265	-0.55	0.5820
Feriado no dia da previsão	-0.0135	0.1339	-0.10	0.9199
$\phi$ : 5.254157				
Null deviance: 256.90 on 60 degrees of freedom				
Residual deviance: 160.74 on 48 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.13: Sumário do modelo de previsão 9 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.7602	0.0588	63.90	0.0000
Segunda	0.1086	0.0841	1.29	0.2031
Terça	0.2031	0.0865	2.35	0.0230
Quarta	0.2326	0.0871	2.67	0.0103
Quinta	0.1754	0.0861	2.04	0.0472
Sexta	-0.0781	0.0816	-0.96	0.3433
Sábado	-0.1478	0.0801	-1.84	0.0712
Feriado morte	-0.0342	0.1405	-0.24	0.8086
Feriado dia anterior à previsão	-0.1360	0.1342	-1.01	0.3163
Feriado dois dias antes da previsão	-0.0280	0.1387	-0.20	0.8407
Feriado três dias antes da previsão	-0.0982	0.1298	-0.76	0.4529
Feriado quatro dias antes da previsão	0.1894	0.1432	1.32	0.1923
Feriado no dia da previsão	0.1043	0.1424	0.73	0.4671
$\phi$ : 5.073713				
Null deviance: 349.78 on 60 degrees of freedom				
Residual deviance: 149.08 on 48 degrees of freedom				
<sup>a</sup> Referência: Domingo				

ANEXO A. RESTANTES MODELOS DE PREVISÃO

Tabela A.14: Sumário do modelo de previsão 10 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.8750	0.0603	64.31	0.0000
Segunda	0.0943	0.0864	1.09	0.2802
Terça	0.1870	0.0886	2.11	0.0400
Quarta	0.2335	0.0895	2.61	0.0121
Quinta	0.0762	0.0864	0.88	0.3821
Sexta	-0.1932	0.0819	-2.36	0.0225
Sábado	-0.1529	0.0823	-1.86	0.0693
Feriado morte	-0.0545	0.1400	-0.39	0.6991
Feriado dia anterior à previsão	0.0124	0.1422	0.09	0.9307
Feriado dois dias antes da previsão	-0.1307	0.1364	-0.96	0.3426
Feriado três dias antes da previsão	0.0176	0.1418	0.12	0.9016
Feriado quatro dias antes da previsão	-0.0774	0.1304	-0.59	0.5557
Feriado no dia da previsão	0.1950	0.1479	1.32	0.1936
$\phi$ : 4.832753				
Null deviance: 360.90 on 60 degrees of freedom				
Residual deviance: 149.77 on 48 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.15: Sumário do modelo de previsão 11 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.9433	0.0425	92.72	0.0000
Segunda	0.0955	0.0606	1.58	0.1219
Terça	0.2103	0.0625	3.36	0.0016
Quarta	0.1659	0.0617	2.69	0.0100
Quinta	-0.0030	0.0595	-0.05	0.9605
Sexta	-0.1547	0.0578	-2.68	0.0103
Sábado	-0.1306	0.0579	-2.26	0.0289
Feriado morte	-0.0520	0.0976	-0.53	0.5965
Feriado no dia anterior à morte	0.0987	0.1014	0.97	0.3355
Feriado dois dias antes da morte	0.0494	0.1005	0.49	0.6253
Feriado três dias antes da morte	-0.1181	0.0947	-1.25	0.2188
Feriado quatro dias antes da morte	0.0336	0.0991	0.34	0.7360
Feriado no dia da previsão	0.2655	0.1056	2.51	0.0156
$\phi$ : 2.203983				
Null deviance: 276.22 on 57 degrees of freedom				
Residual deviance: 117.37 on 45 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.16: Sumário do modelo de previsão 12 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	3.9904	0.0493	80.96	0.0000
Segunda	0.1095	0.0701	1.56	0.1262
Terça	0.1645	0.0712	2.31	0.0262
Quarta	0.1110	0.0703	1.58	0.1221
Quinta	0.0326	0.0691	0.47	0.6395
Sexta	-0.1238	0.0669	-1.85	0.0717
Sábado	-0.0919	0.0675	-1.36	0.1809
Feriado morte	-0.0374	0.1109	-0.34	0.7380
Feriado dia anterior à morte	0.1866	0.1198	1.56	0.1272
Feriado dois dias antes da morte	0.0995	0.1171	0.85	0.4003
Feriado três dias antes da morte	0.0936	0.1177	0.80	0.4313
Feriado quatro dias antes da morte	-0.0787	0.1102	-0.71	0.4795
Feriado dia da previsão	0.1948	0.1189	1.64	0.1092
$\phi$ : 2.788059				
Null deviance: 228.28 on 52 degrees of freedom				
Residual deviance: 128.91 on 40 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.17: Sumário do modelo de previsão 13 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	4.0239	0.0470	85.57	0.0000
Segunda	0.0737	0.0666	1.11	0.2748
Terça	0.1238	0.0676	1.83	0.0749
Quarta	0.1394	0.0679	2.05	0.0469
Quinta	0.0594	0.0666	0.89	0.3779
Sexta	-0.0935	0.0647	-1.45	0.1564
Sábado	-0.0844	0.0647	-1.30	0.1998
Feriado morte	-0.0165	0.1075	-0.15	0.8785
Feriado no dia anterior à morte	0.1395	0.1126	1.24	0.2229
Feriado dois dias antes da morte	0.2205	0.1165	1.89	0.0659
Feriado três dias antes da morte	0.1158	0.1125	1.03	0.3096
Feriado quatro dias antes da morte	0.1159	0.1133	1.02	0.3130
Feriado no dia da previsão	0.1160	0.1109	1.05	0.3021
$\phi$ : 2.462301				
Null deviance: 181.24 on 51 degrees of freedom				
Residual deviance: 102.76 on 39 degrees of freedom				
<sup>a</sup> Referência: Domingo				

Tabela A.18: Sumário do modelo de previsão 14 - Teste 2013

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	4.0301	0.0434	92.79	0.0000
Segunda	0.0658	0.0618	1.07	0.2931
Terça	0.1724	0.0637	2.70	0.0101
Quarta	0.1949	0.0640	3.05	0.0041
Quinta	0.1089	0.0628	1.73	0.0907
Sexta	-0.0420	0.0608	-0.69	0.4946
Sábado	-0.0433	0.0606	-0.71	0.4798
Feriado morte	0.0157	0.1036	0.15	0.8800
Feriado dia anterior à morte	0.0583	0.1035	0.56	0.5763
Feriado dois dias antes da morte	0.1242	0.1056	1.18	0.2465
Feriado três dias antes da morte	0.2140	0.1094	1.96	0.0577
Feriado quatro dias antes da morte	0.1469	0.1066	1.38	0.1758
Feriado no dia da previsão	0.0733	0.1055	0.69	0.4915
$\phi$ : 2.411805				
Null deviance: 164.824 on 51 degrees of freedom				
Residual deviance: 89.074 on 39 degrees of freedom				
<sup>a</sup> Referência: Domingo				