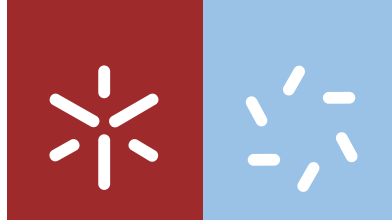


**Universidade do Minho**  
Escola de Ciências

Adriana Maria da Silva Vieira

**Modelação de Admissões e Internamentos  
na Urgência do Hospital de Braga**



**Universidade do Minho**  
Escola de Ciências

Adriana Maria da Silva Vieira

## **Modelação de Admissões e Internamentos na Urgência do Hospital de Braga**

Dissertação de Mestrado  
Mestrado em Estatística

Trabalho realizado sob orientação da  
**Professora Doutora Inês Sousa**  
e da  
**Doutora Sónia Dória**

julho de 2016

Nome: Adriana Maria da Silva Vieira

Endereço eletrónico: adrimsvieira@gmail.com

Título da dissertação: Modelação de Admissões e Internamentos na Urgência do Hospital de Braga

Orientadores: Professora Doutora Inês Sousa e Doutora Sónia Dória

Ano de conclusão: 2016

Mestrado em Estatística

**É autorizada a reprodução integral desta dissertação apenas para efeitos de investigação, mediante declaração escrita do interessado, que a tal se compromete.**

Universidade do Minho, \_\_\_\_\_ de \_\_\_\_\_ de 2016

A autora: \_\_\_\_\_

# AGRADECIMENTOS

À Professora Doutora Inês Sousa e à Doutora Sónia Dória, pela sua dedicação, apoio e partilha de conhecimentos. Agradeço essencialmente todos os conselhos, sugestões e empenho, mas também a amizade.

Ao Rui, pela simpatia e disponibilidade com que sempre se mostrou.

Aos meus pais, Fernanda e Jacinto e ao meu irmão Sérgio, pela ajuda e compreensão demonstradas em todos os momentos e por sempre terem acreditado em mim. Uma palavra de gratidão por todo amor e apoio incondicionais, bem como por todos os sacrifícios, que me permitem estar aqui.

Ao Davide, pela inabalável confiança que deposita em mim e nos meus sonhos, por nunca me ter deixado desistir e pela necessidade de me ver constantemente a sorrir. Agradeço o carinho, a dedicação e a infinita paciência.

Aos restantes professores e a todos os meus amigos e familiares, que de alguma forma contribuíram para o meu crescimento científico e pessoal, oferecendo a sua ajuda em todas as circunstâncias.

A todos o meu muito obrigado!



## RESUMO

Este estudo teve como objetivos, através da modelação de padrões e tendências, desenvolver modelos que permitissem prever, tanto o número de pacientes admitidos diariamente nos três serviços de urgência (geral, obstétrico e pediátrico), como o número de pacientes internados via urgência, no Hospital de Braga, em Portugal. Para tal foram usadas técnicas de regressão linear e de Poisson, respetivamente, bem como análise de séries temporais.

Recorrendo ao uso de variáveis sazonais, meteorológicas, ambientais e epidemiológicas, tentou encontrar-se equações matemáticas que permitissem a previsão futura do número de admissões em ambos os serviços de urgência e internamento.

O uso de variáveis que não as sazonais revelou-se de pouco valor preditivo, apesar de existirem variáveis estatisticamente significativas. Além disso, para o hospital é mais simples utilizar apenas dados internos ou do conhecimento geral. Assim, foi decidida a utilização de apenas variáveis sazonais, sendo estas o dia da semana, o mês do ano e os feriados. Adicionalmente utilizou-se também a variável  $t$  que representa o tempo, em dias, decorrido desde 1 de janeiro de 2012.

No caso das admissões no serviço de urgência, estas foram modeladas recorrendo a técnicas de regressão linear múltipla, tendo-se verificado correlação temporal nos resíduos de cada modelo, sendo o número de admissões de um determinado dia correlacionadas com as dos 7 dias anteriores no caso da urgência geral, e com as dos 8 dias anteriores no caso de ambas urgências, obstétrica e pediátrica.

Por outro lado, no caso das admissões no serviço de internamento via urgência, a modelação foi feita com recurso à regressão de Poisson, não se verificando a existência de qualquer tipo de correlação temporal. Neste caso, também não se registou qualquer evidência de que os feriados estejam relacionados com o número de internamentos.

Toda a modelação foi feita com recurso ao *software R* versão 3.2.2., (2015-08-14).

**Palavras-chave:** previsão, hospital, serviço de urgência, serviço de internamento, regressão linear, modelos lineares generalizados, séries temporais.



# ABSTRACT

This study aims, through patterns and trend modelling, to develop models that allow to predict the number of patients that visit, on a daily basis, the three emergency departments (general, obstetric and pediatric), as well as, the number of hospitalized patients, at Braga's Hospital, Portugal. For such, were used, respectively, linear and Poisson regression, as well as, time series analysis.

Using calendar, meteorological, environmental and epidemiological variables, we tried to get mathematical equations that allow to forecast the future number of admissions, both in emergency and hospitalization services. The use of other variables than the calendar ones proved to have little predictive value, despite existing statistically significant variables. Furthermore, is simpler for the hospital use only internal data or general knowledge. Thus, it was decided to use only calendar variables, being these the day of the week, the month of the year and holidays. Additionally, we used the variable  $t$ , representing the elapsed time, in days, since January 1, 2012.

In the case of emergency department admissions, these were modeled using multiple linear regression techniques, having been verified temporal correlation in the residuals of each model, being the number of admissions for a given day correlated with the previous 7 days in case of general emergency, and with the previous 8 days for both emergency, obstetric and pediatric. On the other hand, in the hospitalization admission case, modeling was done resorting to the Poisson regression, not having been verified the existence of any type of temporal correlation. In this case, it was also not registered any evidence that the holidays are related to the number of admissions.

All modeling was performed with R software, version 3.2.2., (2015-08-14).

**Keywords:** forecast, hospital, emergency department, internment department, linear regression, generalized linear models, time series.





# CONTEÚDO

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Hospital de Braga</b>	<b>7</b>
<b>3</b>	<b>Metodologias</b>	<b>10</b>
3.1	Modelos Lineares . . . . .	10
3.1.1	Regressão Linear . . . . .	10
3.1.1.1	Pressupostos do Modelo de Regressão . . . . .	11
3.1.1.2	Estimação dos Parâmetros de um Modelo de Regressão . . . . .	12
3.1.1.3	Inferências no Modelo de Regressão . . . . .	13
3.1.1.4	Qualidade do Modelo . . . . .	14
3.1.1.5	Seleção de Variáveis . . . . .	15
3.1.1.6	Análise de Resíduos e Multicolinearidade . . . . .	15
3.1.2	Modelos Lineares Generalizados . . . . .	17
3.1.2.1	A Família Exponencial . . . . .	18
3.1.2.2	Valor Médio e Variância . . . . .	18
3.1.2.3	Componentes de um MLG . . . . .	19
3.1.2.4	Estimação dos parâmetros de um MLG . . . . .	20
3.1.2.5	Inferências nos MLG's . . . . .	22
3.1.2.6	Qualidade do Modelo . . . . .	22
3.1.2.7	Análise de Resíduos . . . . .	23
3.1.2.8	Modelo de Poisson . . . . .	23
3.1.2.9	Modelo de Binomial Negativa . . . . .	25
3.2	Modelos de Séries Temporais . . . . .	26
3.2.1	Processos Estacionários . . . . .	27
3.2.1.1	O Periodograma . . . . .	29

3.2.1.2	Processos autoregressivos de ordem $p$ . . . . .	30
3.2.1.3	Processos de médias móveis de ordem $q$ . . . . .	30
3.2.1.4	Processos mistos autoregressivos e médias móveis . . . . .	30
3.2.1.5	Função de autocovariância e função de autocorrelação . . . . .	31
3.3	Transformações de Box Cox . . . . .	32
3.4	MAPE . . . . .	32
3.5	Teorema do Limite Central . . . . .	33
<b>4</b>	<b>Análise Exploratória</b> . . . . .	<b>34</b>
4.1	Serviço de Urgência . . . . .	34
4.1.1	Urgência Geral . . . . .	36
4.1.2	Urgência Obstétrica . . . . .	44
4.1.3	Urgência Pediátrica . . . . .	52
4.2	Dados Meteorológicos . . . . .	60
4.3	Dados Poluição Atmosférica . . . . .	62
4.4	Dados Gripe . . . . .	66
<b>5</b>	<b>Modelação das Séries Temporais</b> . . . . .	<b>70</b>
5.1	Admissões na Urgência . . . . .	70
5.1.1	Urgência Geral . . . . .	71
5.1.1.1	Análise dos Resíduos . . . . .	77
5.1.1.2	Validação do Modelo e Predição . . . . .	78
5.1.2	Urgência Obstétrica . . . . .	85
5.1.2.1	Análise dos Resíduos . . . . .	90
5.1.2.2	Validação do Modelo e Predição . . . . .	91
5.1.3	Urgência Pediátrica . . . . .	92
5.1.3.1	Análise dos Resíduos . . . . .	96
5.1.3.2	Validação do Modelo . . . . .	97
5.2	Internamentos Via Urgência Geral . . . . .	99

5.2.1	Análise dos Resíduos . . . . .	103
5.2.1.1	Validação do Modelo . . . . .	103
<b>6</b>	<b>Conclusões</b>	<b>105</b>
	<b>Bibliografia</b>	<b>108</b>
	<b>Anexo A - Tabela dos Valores Previstos <i>vs</i> Reais da Urgência Geral</b>	<b>113</b>
	<b>Anexo B - Tabela dos Valores Previstos <i>vs</i> Reais da Urgência Obstétrica</b>	<b>116</b>
	<b>Anexo C - Tabela dos Valores Previstos <i>vs</i> Reais da Urgência Pediátrica</b>	<b>120</b>

## LISTA DE FIGURAS

2.1	Mapa dos Agrupamentos de Centros de Saúde . . . . .	7
4.1	Total de Pacientes Admitidos na Urgência Geral, diariamente . . . . .	41
4.2	Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Sexo . . . . .	42
4.3	Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Local de Residência . . . . .	42
4.4	Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta a Cor da Pulseira . . . . .	43
4.5	Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Destino . . . . .	43
4.6	Total de Pacientes Admitidos na Urgência Geral, com destino ao Serviço de Internamento . . . . .	44
4.7	Total de Pacientes Admitidos na Urgência Obstétrica, diariamente . . . . .	49
4.8	Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o Local de Residência . . . . .	50
4.9	Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o a Cor da Pulseira . . . . .	50
4.10	Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o Destino . . . . .	51
4.11	Total de Pacientes Admitidos na Urgência Obstétrica, com destino ao Serviço de Internamento . . . . .	51
4.12	Total de Pacientes Admitidos na Urgência Pediátrica, diariamente . . . . .	56
4.13	Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Sexo . . . . .	57

4.14	Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Local de Residência . . . . .	58
4.15	Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta a Cor da Pulseira . . . . .	58
4.16	Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Destino . . . . .	59
4.17	Total de Pacientes Admitidos na Urgência Pediátrica, com destino ao Serviço de Internamento . . . . .	59
4.18	Temperaturas máxima (a) e mínima (b) registadas, diariamente . . . . .	62
4.19	Concentrações de NO <sub>2</sub> (a), NO (b), O <sub>3</sub> (c) e P <sub>10</sub> (d) registadas, diariamente . . . . .	65
4.20	Número de consultas devido a gripe registadas, semanalmente . . . . .	69
5.1	Periodograma dos resíduos do Modelo 2 . . . . .	73
5.2	FAC (a) e FACP (b) dos resíduos . . . . .	74
5.3	Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 <i>vs</i> os resíduos (b) e do tempo <i>vs</i> os resíduos (c) . . . . .	77
5.4	Previsão do número de pacientes admitidos na urgência geral, diariamente . . . . .	78
5.5	Probabilidade de acertar no número de médicos, para valores de $p_c$ entre 50 e 90, no dia 01/01/2016 . . . . .	81
5.6	Probabilidade de acertar no número de médicos, para valores de $p_c$ entre 50 e 90, no dia 02/01/2016 . . . . .	82
5.7	Probabilidade de acertar no número de médicos, para valores de $p_c$ entre 50 e 90, no dia 03/01/2016 . . . . .	82
5.8	Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 01/01/2016, para valores de $p_c$ entre 50 e 90 . . . . .	83
5.9	Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 02/01/2016, para valores de $p_c$ entre 50 e 90 . . . . .	84

5.10	Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 03/01/2016, para valores de $p_c$ entre 50 e 90 . . . . .	85
5.11	Periodograma dos resíduos do Modelo 1 . . . . .	86
5.12	FAC (a) e FACP (b) dos resíduos . . . . .	87
5.13	Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 <i>vs</i> os resíduos (b) e do tempo <i>vs</i> os resíduos (c) . . . . .	90
5.14	Previsão do número de pacientes admitidos na urgência obstétrica, diariamente	91
5.15	Periodograma dos resíduos do Modelo 1 . . . . .	93
5.16	FAC (a) e FACP (b) dos resíduos . . . . .	94
5.17	Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 <i>vs</i> os resíduos (b) e do tempo <i>vs</i> os resíduos (c) . . . . .	97
5.18	Previsão do número de pacientes admitidos na urgência pediátrica, diariamente	98
5.19	Periodograma dos resíduos do Modelo 1 . . . . .	100
5.20	FAC (a) e FACP (b) dos resíduos . . . . .	101
5.21	Representação gráfica da transformação, proposta por McCullagh e Nelder, dos valores ajustados do modelo 1 <i>vs</i> os resíduos (a) e do tempo <i>vs</i> os resíduos (b)	103
5.22	Previsão do número de pacientes admitidos no serviço de internamento via urgência geral, diariamente . . . . .	104

## LISTA DE TABELAS

2.1	Local de Residência . . . . .	8
2.2	Sistema de Triagem de Manchester (Matias <i>et al.</i> , 2008) . . . . .	9
3.1	Comparação da FAC e FACP para os processos ARMA(p,q) . . . . .	32
4.1	Descrição da Base de Dados Urgência . . . . .	35
4.2	Tabela de Frequências Absolutas e Relativas do Tipo de Urgência . . . . .	36
4.3	Tabela Sumária da Variável Quantitativa Idade . . . . .	37
4.4	Tabela de Frequências Absolutas e Relativas . . . . .	37
4.5	Tabela Sumária do Número de Admissões, por dia, de acordo com categorização de feriado . . . . .	38
4.6	Tabela Sumária do Número de Admissões, por dia, em cada dia da semana . . . . .	38
4.7	Tabela Sumária do Número de Admissões, por dia, em cada mês . . . . .	39
4.8	Tabela Sumária do Número de Admissões, por dia, em cada estação do ano . . . . .	39
4.9	Tabela Sumária da Variável Quantitativa Idade . . . . .	44
4.10	Tabela de Frequências Absolutas e Relativas . . . . .	45
4.11	Tabela Sumária do Número de Admissões, por dia, em cada dia da semana . . . . .	47
4.12	Tabela Sumária do Número de Admissões, por dia, em cada mês . . . . .	47
4.13	Tabela Sumária do Número de Admissões, por dia, em cada estação do ano . . . . .	48
4.14	Tabela Sumária do Número de Admissões, diárias, de acordo com categorização de feriado . . . . .	48
4.15	Tabela Sumária da Variável Quantitativa Idade . . . . .	53
4.16	Tabela de Frequências Absolutas e Relativas . . . . .	53
4.17	Tabela Sumária do Número de Admissões, por dia, em cada dia da semana . . . . .	54
4.18	Tabela Sumária do Número de Admissões, por dia, em cada mês . . . . .	55
4.19	Tabela Sumária do Número de Admissões por dia em cada estação do ano . . . . .	55



4.20	Tabela Sumária do Número de Admissões, por dia, de acordo com categorização de feriado . . . . .	55
4.21	Descrição da Base de Dados Meteo . . . . .	60
4.22	Tabela Sumária do Valor Médio Diário em ambas as estações . . . . .	61
4.23	Descrição da Base de Dados Ar . . . . .	63
4.24	Tabela Sumária das Variável Quantitativas . . . . .	64
4.25	Classificação do Índice de Qualidade do Ar pela APA . . . . .	64
4.26	Descrição da Base de Dados Gripe . . . . .	67
4.27	Tabela Sumária das Variáveis Quantitativas . . . . .	68
4.28	Tabela sumária do número total de casos de gripe registados semanalmente . .	68
5.1	Estimativas para os coeficientes do autogressivo de ordem 7 . . . . .	75
5.2	Tabela sumária do Modelo 3 . . . . .	76
5.3	Estimativas para os coeficientes do autogressivo de ordem 8 . . . . .	88
5.4	Tabela sumária do Modelo 3 . . . . .	89
5.5	Tabela sumária do Modelo 3 . . . . .	95
5.6	Estimativas para os coeficientes do autogressivo de ordem 8 . . . . .	96
5.7	Tabela sumária do Modelo 1 . . . . .	102

# ABREVIATURAS

**ACES** - Agrupamentos de Centros de Saúde

**AIC** - Critério de Informação de Akaike

**APA** - Agência Portuguesa do Ambiente

**ARS** - Administração Regional de Saúde

**DP** - Desvio Padrão

**FAC** - Função de Autocorrelação

**FACP** - Função de Autocorrelação Parcial

**HB** - Hospital de Braga

**MLG** - Modelos Lineares Generalizados

**MMV** - Método da Máxima Verosimilhança

**SI** - Serviço de Internamento

**SNS** - Serviço Nacional de Saúde

**SSE** - Soma dos Quadrados dos Resíduos

**SST** - Soma dos Quadrados dos Desvios das Observações em Relação à Sua Média

**SU** - Serviço de Urgência

**TLC** - Teorema do Limite Central

**VIF** - *variance inflation factor*



## Introdução

---

Os hospitais públicos são parte essencial do Serviço Nacional de Saúde (SNS), apresentando-se como sistemas complexos, cujas variadas áreas de atuação estão divididas segundo especialidades médicas, contando com o auxílio de equipamentos tecnológicos para a resolução dos problemas de saúde dos pacientes que todos os dias lá chegam.

Uma das principais e mais complexas áreas de um hospital é o seu Serviço de Urgência, cujo principal objetivo passa pela prestação de cuidados de saúde a todos que deles necessitem. Estes cuidados de saúde têm como alvo situações de caráter urgente, isto é, crises agudas de que o paciente padeça, cujo atendimento e tratamento devam ser imediatos.

No entanto, cada vez mais se tem registado um aumento de deslocações às urgências por casos classificados de pouco ou até mesmo não urgentes. Esta situação deve-se essencialmente a um envelhecimento da população cada vez mais evidente e conseqüente aumento da procura de cuidados médicos. Por outro lado, também a falta de acessibilidade a cuidados de saúde primários pode constituir um fator para o aumento de admissões, nos serviços de urgência, de casos não urgentes.

Este aumento registado leva muitas vezes à sobrelotação da urgência de um hospital, levando a tempos de espera excessivos e redução da qualidade dos cuidados de saúde prestados, o que tem como conseqüências a deterioração da resposta clínica e insatisfação dos pacientes.

Tendo em conta dados da Comissão de Reavaliação da rede Nacional de Emergência/Urgência, em 2010, sabe-se que aproximadamente 46% dos casos admitidos nos serviços de urgência a nível nacional não constituem episódios urgentes. Verifica-se, assim que cerca de 6 milhões de admissões anuais nos serviços de urgência de Portugal Continental se ficam

a dever a um uso abusivo e não urgente. Estes valores variam consoante a região do país a considerar, sendo no norte que se registam os mais baixos valores de utilização dos serviços de urgência - cerca de 547 episódios por cada mil habitantes.

Assim sendo, em termos de gestão hospitalar é necessária uma correta alocação dos recursos, humanos e materiais. Num hospital central, que serve uma grande população, a tarefa pode tornar-se dificultada com as chegadas imprevistas aos serviços do mesmo. Por essa razão a previsão é essencial.

Sendo os hospitais financiados por um modelo que impõe restrições ao financiamento, a afetação dos recursos deve garantir um equilíbrio entre o seu custo e a sua efetividade. Ou seja, uma previsão adequada das admissões no serviço de urgência permite um melhor processo de escala das equipas clínicas, o que conseqüentemente garantirá menores tempos de espera no atendimento. Além disso, uma correta previsão do número de admissões na urgência, leva à previsão do número de camas no hospital necessárias, para dar resposta à afluência de casos urgentes.

A incerteza acerca da necessidade de camas disponíveis para dar resposta à afluência de casos urgentes obriga, frequentemente, a cancelar, no próprio dia, atividade cirúrgica que já havia sido programada com antecedência.

O presente trabalho tem então, como principais objetivos a previsão do número de pacientes que chegam diariamente ao Serviço de Urgência, bem como, afim de perceber a necessidade de camas no meio hospitalar, o número de pacientes que segue da urgência para o Serviço de Internamento.

Estudos prévios demonstram que a afluência ao serviço de urgência não é aleatória, esta depende muitas vezes de variáveis meteorológicas (Batal *et al.*, 2001; Attia *et al.*, 1998; Noble *et al.* 1971; Holleman *et al.*, 1996; Diehl *et al.*, 1981; Diaz *et al.*, 2001), sazonais (Batal *et al.*, 2001; Holleman *et al.*, 1996; Diehl *et al.*, 1981), ambientais (Diaz *et al.*, 2001) e mesmo epidemiológicas (Araz *et al.*, 2014; Fuhrmann *et al.*, 2016).

Attia *et al.* (1998) não encontraram diferenças significativas entre condições atmosféricas favoráveis e desfavoráveis nas admissões hospitalares, numa unidade pediátrica.

Já Noble *et al.* (1971) concluíram que com condições meteorológicas favoráveis o número de pacientes a dirigir-se ao serviço de urgência, sem recorrer ao uso de ambulância (possivelmente casos não tão urgentes) aumentou, em três hospitais de Boston.

Num hospital de veteranos, Holleman *et al.* (1996) conseguiram modelar o volume de pacientes admitidos usando variáveis meteorológicas e sazonais em simultâneo e obtiveram como variáveis com valor significativamente explicativo: temperaturas máximas, queda de neve, estação do ano, semana do mês, dia da semana, feriados e dia de entrega do cheque federal.

Batal *et al.* (2001) tentaram também explicar o número de admissões numa clínica em Denver usando variáveis sazonais (dia da semana, mês do ano, estação do ano, feriados e dias antes e depois de um feriado) e variáveis meteorológicas (temperaturas máximas, médias e mínimas, queda de neve e chuva). À semelhança do concluído por Holleman *et al.* (1996) ao nível das variáveis meteorológicas, apenas a temperatura máxima diária e a queda de neve se revelaram estatisticamente significativas.

Tanto Batal *et al.* (2001), como Holleman *et al.* (1996) e Diehl *et al.* (1981) perceberam que, apesar de estatisticamente significativas, as variáveis meteorológicas acrescentam pouco valor preditivo ao modelo, após introduzidas as variáveis sazonais, sendo que por vezes acabam mesmo por fazer com que as variáveis sazonais percam algum do seu valor explicativo. Da mesma forma acabou por concluir-se isso mesmo neste estudo, como se poderá ver mais à frente.

Por sua vez, Diaz *et al.* (2001) concluíram, usando como variáveis explicativas as concentrações de certos poluentes no ar (ozono, dióxido de enxofre, óxidos de azoto, dióxido de azoto e totalidade de partículas suspensas) de Madrid, bem como certas variáveis meteorológicas (temperaturas máxima, média e mínima e humidade relativa), numa modelação de séries temporais, a existente relação entre o número de admissões e temperaturas, humidade relativa e concentrações de ozono. Observaram ainda que, os poluentes mostram em geral um padrão com pico máximo nos meses de inverno, à exceção do ozono que, por necessitar de luz solar para a sua formação, registou um pico nos meses quentes. Quanto ao número de pacientes regista-se afluência máxima nos meses de inverno, relativamente

aos meses de verão onde a afluência ao hospital é menor.

Quanto a variáveis epidemiológicas, Fuhrmann *et al.* (2016) concluíram que o número de admissões num serviço de urgência na Carolina do Norte aumenta com o aumento de casos de gripe e pneumonia.

Na maioria dos estudos os métodos utilizados são a regressão linear e a análise de séries temporais, já que ambas são de simples implementação, são capazes de modelar padrões e tendências e providenciam resultados de fácil interpretação (Jones *et al.*, 2008; Wiler *et al.*, 2011).

Outras conclusões retiradas de relevante interesse: registam-se valores máximos de visitas à urgência às segundas-feiras, decrescentes ao longo da semana, registando valores mais baixos aos fins de semana (Batal *et al.*, 2001; Wargon *et al.*, 2010; Marcilio *et al.*, 2013; Holleman *et al.*, 1996; Diehl *et al.*, 1981; Glass e Friedman, 1977); é preferível a modelação usando apenas variáveis sazonais *versus* variáveis meteorológicas (Marcilio *et al.*, 2013); existe em geral, presente nos dados de admissões nos serviços de urgência uma sazonalidade ao nível da semana (Jones *et al.*, 2008); a variável meteorológica que mais impacto costuma ter no número de admissões é a temperatura máxima (Jones *et al.*, 2008; Diaz *et al.*, 2001); existe pouca variação no número de admissões entre os meses do ano (Marcilio *et al.*, 2013; Champion *et al.*, 2007).

Assim sendo, neste estudo, onde os objetivos são:

- descrever a progressão do número de casos que dão entrada na urgência do Hospital de Braga, de 2012 a 2015, de acordo com o serviço de urgência;
- identificar fatores de risco significativos para explicar o número de casos de urgência no Hospital de Braga, por serviço de urgência;
- caracterizar um modelo de predição para o número de casos de urgência no Hospital de Braga em dias futuros, para cada um dos serviços de urgência;
- descrever a progressão do número de casos que dão entrada no internamento do Hospital de Braga, de 2012 a 2015, via urgência geral;

- identificar fatores de risco significativos para explicar o número de internamentos no Hospital de Braga, via urgência geral;
- caracterizar um modelo de predição para o número de admissões no serviço de internamento no Hospital de Braga em dias futuros, via urgência geral;

utilizar-se-ão variáveis sazonais, meteorológicas, ambientais e epidemiológicas numa tentativa de resposta a cada um dos objetivos.

O número de admissões e internamentos no hospital serão considerados séries temporais, com objetivo de perceber a existência, ou não, de correlação temporal posterior à modelação incluindo variáveis explicativas.

Será analisada uma base de dados, relativa à chegada diária de pacientes ao Serviço de Urgência entre 1 de janeiro de 2012 e 31 de dezembro de 2015. Com objetivo de testar os modelos finais utilizar-se-ão dados relativos ao primeiro trimestre de 2016.

Ter-se-ão como variáveis explicativas: dia da semana, mês do ano, estação do ano, feriados, temperatura máxima e mínima do dia anterior, número de consultas de gripe na semana anterior, concentrações de alguns poluentes.

Os feriados tidos em conta serão: Ano Novo, Carnaval, Sexta-feira Santa, Páscoa, Dia da Liberdade, Dia do Trabalhador, Corpo de Deus (do ano de 2012), Dia de Portugal, Assunção de Nossa Senhora, Dia da Implantação da República (do ano de 2012), Dia de Todos os Santos (do ano de 2012), Dia da restauração da Independência (do ano de 2012), Dia da Imaculada Conceição e Dia de Natal.

Caso as variáveis meteorológicas, de poluição atmosférica e de gripe sejam incluídas no modelo teórico para o número de pacientes no hospital, esta informação será necessária em dias futuros para a previsão do número de admissões. Assim, para fazer predição em tempo real, o hospital necessitará de obter, diariamente, toda uma informação extra, proveniente de fontes externas. Portanto, no caso de se verificar, à semelhança do registado em estudos anteriores, que a inclusão destas variáveis não melhora significativamente a qualidade de ajustamento do modelo, optar-se-á pela não inclusão das mesmas.



Este estudo contará com seis capítulos além da introdução.

No Capítulo 2 tem-se uma descrição do Hospital de Braga, centrando a atenção no Serviço de Urgência do mesmo.

O Capítulo 3 conta com um sumário das metodologias estatísticas utilizadas na elaboração do estudo. Estas metodologias são essencialmente relativas à análise de modelos de regressão e modelos lineares generalizados, bem como à análise de séries temporais, com especial foco nos processos estacionários. Além disso, pode ainda encontrar-se referência às transformações de Box-Cox, ao erro absoluto percentual médio e ainda ao Teorema do Limite Central.

No Capítulo 4 apresentam-se as bases de dados disponíveis para estudo e respetiva análise exploratória.

Conta-se com uma base de dados disponibilizada pelo Hospital de Braga, com informações relativas à entrada dos diferentes pacientes, diariamente; uma base de dados cedida pela Agência Portuguesa do Ambiente, com as medições, horárias, das concentrações de certos poluentes na zona do Minho; uma base de dados da Administração Regional de Saúde do Norte com os valores semanais relativos às consultas registadas, relacionadas com o vírus da gripe e, por fim, uma base de dados, cedida pela *National Oceanic and Atmospheric Administration* do Departamento do Comércio dos Estados Unidos da América, com informações relativas a temperaturas máximas e mínimas, bem como à precipitação em território português. Tem-se, ainda em conta, para a análise exploratória, as variáveis sazonais em estudo: dia da semana, feriado, mês do ano e estação do ano.

Já no Capítulo 5 apresenta-se a modelação dos dados propriamente dita. Com recurso às metodologias explicitadas no Capítulo 3 modelaram-se o número de admissões nos três serviços de urgência, bem como número de internamentos via urgência, encontrando fórmulas matemáticas que permitem a previsão futura destes números. Além disso, encontra-se também a validação de cada um dos modelos, usando dados fornecidos pelo hospital relativos ao primeiro trimestre de 2016.

No Capítulo 6 podem encontrar-se as principais conclusões deste estudo.

### Hospital de Braga

O Hospital de Braga (HB), inaugurado em Maio de 2011, na zona Este da cidade de Braga, permitiu uma grande melhoria nos cuidados de saúde prestados, comparativamente com o Hospital de S. Marcos, com mais de 500 anos, o qual veio substituir.

O HB cobre uma área de cerca de 1.1 milhões de pacientes, sendo prestador de cuidados em 1<sup>a</sup> e 2<sup>a</sup> linha para diferentes áreas geográficas (Rede de Referência Hospitalar de Urgência/Emergência, 2001). Isto significa que de acordo com a área geográfica a que o paciente pertence, o Hospital de Braga é, tendo em atenção o tipo de urgência e/ou serviço a que o paciente se pretende dirigir, o único disponível (1<sup>a</sup> linha), ou então uma opção relativamente a outros hospitais da região (2<sup>a</sup> linha).

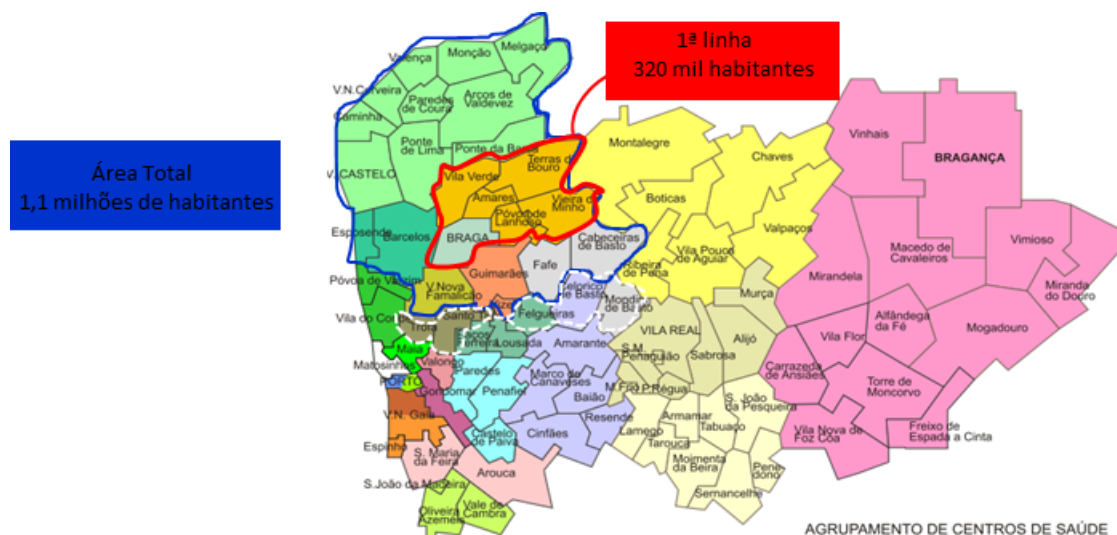


Figura 2.1: Mapa dos Agrupamentos de Centros de Saúde

A Administração Regional de Saúde (ARS) define os agrupamentos de centros de saúde

(ACES) de acordo com as áreas geográficas a que os centros de saúde no SNS pertencem.

A Tabela 2.1 mostra o agrupamento dos concelhos com a respetiva nomenclatura para cada ACES. A Figura 2.1 representa geograficamente os mesmos agrupamentos.

Assim sendo, Hospital de Braga é o responsável de 1<sup>a</sup> linha por cerca de 320 mil habitantes dos ACES Cávado I e Cávado II e o responsável em 2<sup>a</sup> linha pela população dos restantes concelhos dos distritos de Braga e Viana do Castelo, além dos restantes pacientes do território nacional e internacional.

Tabela 2.1: Local de Residência

<b>ACES/Distrito</b>	<b>Concelho/Local de Residência</b>
ACES Cávado I	Braga
ACES Cávado II	Amares, Póvoa de Lanhoso, Terras de Bouro, Vieira do Minho e Vila Verde
ACES Cávado III	Barcelos e Esposende
ACES Alto Ave	Cabeceiras de Basto, Celorico de Basto, Fafe, Guimarães e Vizela
ACES Ave	Vila Nova de Famalicão
Viana do Castelo	Arcos de Valdevez, Caminha, Melgaço, Monção, Paredes de Coura, Ponte da Barca, Ponte de Lima, Valença, Viana do Castelo e Vila Nova de Cerveira
Outros	Restantes concelhos de Portugal
Estrangeiro	Morada fiscal fora de Portugal

A admissão de um paciente no hospital pode faz-ser de duas formas: eletiva ou emergente.

Quando a admissão de um paciente é programada com dias ou até mesmo semanas de antecedência tem-se uma admissão eletiva.

Por outro lado, quando a admissão do paciente é feita através do Serviço de Urgência (SU), de forma não planeada tem-se uma admissão emergente. Este tipo de admissão pode dever-se a fatores epidemiológicos (cronicidade de patologias), meteorológicos (picos de

frio ou calor), económicos (crise económica, dispensa de pagamento de taxas moderadoras, menor recurso a serviços privados) e até sociais (dias da semana, horas do dia).

O Serviço de Urgência do HB está repartido em três outros serviços de forma a suprimir as necessidades de cada paciente:

- Urgência Geral;
- Urgência Obstétrica;
- Urgência Pediátrica.

A triagem nos três serviços de urgência do HB é feita de acordo com o sistema de triagem de Manchester (Matias *et al.*, 2008) que, após o seguimento de um conjunto de questões de ordem clínica, permite atribuir ao paciente uma pulseira colorida de acordo com a gravidade da sua situação. A cor atribuída ao paciente determina o tempo de espera máximo a que o paciente deverá ser submetido até à primeira observação médica. Na Tabela 2.2 estão identificados estes tempos teóricos máximos de espera até à 1<sup>a</sup> observação.

Tabela 2.2: Sistema de Triagem de Manchester (Matias *et al.*, 2008)

Cor da Pulseira	Situação	Tempo até à 1. <sup>a</sup> observação médica
Vermelho	Emergente	0 minutos
Laranja	Muito Urgente	10 minutos
Amarelo	Urgente	60 minutos (1 hora)
Verde	Pouco Urgente	120 minutos (2 horas)
Azul	Não Urgente	240 minutos (4 horas)

Após atendimento médico os pacientes podem tomar vários destinos, incluindo o Serviço de Internamento (SI). Apesar da capacidade máxima do SI variar, este dispõe de 705 camas (médicas e cirúrgicas), que podem ser ocupadas tanto por pacientes provenientes do SU (cerca de 56% da totalidade de internamentos), bem como por pacientes cujo internamento é programado, como é o caso, por exemplo, de pacientes submetidos a cirurgias, exames invasivos e cesarianas eletivas.

## Metodologias

---

### 3.1 Modelos Lineares

#### 3.1.1 Regressão Linear

A regressão linear teve os seus inícios no século XIX e a sua primeira documentação foi feita por Galton, em Inglaterra, num artigo científico que demonstrava a existência de uma relação linear entre o diâmetro dos grãos de ervilhas e o diâmetro dos grãos descendentes (Hanley, 2004).

A técnica da regressão visa, portanto, estudar a relação entre uma variável resposta (ou dependente),  $Y$ , e  $p$  variáveis explicativas (ou independentes),  $X_1, X_2, \dots, X_p$ , com objetivo de explicar determinado fenómeno e fazer previsão, através de uma fórmula matemática,

$$Y = X\beta + \epsilon$$

onde  $Y$  é um vetor de dimensão  $n \times 1$ ,  $X$  representa a matriz de desenho do modelo, com dimensão de  $n \times (p + 1)$ , associada a um vetor  $\beta$ , de dimensão  $(p + 1) \times 1$ , do qual figuram os parâmetros da regressão e  $\epsilon$  é o vetor de erros aleatórios, de dimensão  $n \times 1$ , com distribuição gaussiana. Assim sendo, também  $Y$  tem distribuição normal e, portanto,  $\epsilon$  e  $Y$  são variáveis aleatórias com

- $E[Y]=X\beta$ ;
- $E[\epsilon]=0$ ;

- $\text{Var}[Y]=\text{Var}[\epsilon]=\sigma^2 I_n$ , onde  $I_n$  é a matriz identidade de ordem  $n$ ;

e  $X$  é determinístico, medido sem erro.

Assim sendo, dada uma amostra aleatória,  $(x_{11}, x_{12}, \dots, x_{1p}, y_1), \dots, (x_{n1}, x_{n2}, \dots, x_{np}, y_n)$ , de  $n$  observações independentes, onde  $x_{ij}$  e  $y_i$  representam, respetivamente, os valores das variáveis  $X_j$  e  $Y$  para o indivíduo  $i$ , tem-se:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n$$

em que

- $y_i$  -  $i$ -ésima observação da variável resposta;
- $x_{ij}$  -  $i$ -ésima observação da covariável  $x_j$ ;
- $\beta_0$  - valor esperado de  $Y$  quando as variáveis explicativas são iguais a zero;
- $\beta_j$  - aumento, em média, do valor de  $Y$  com o aumento de uma unidade em  $x_j$ , no caso desta ser quantitativa, mantendo as restantes variáveis constantes; aumento, em média, do valor de  $Y$  de cada categoria de  $x_j$ , em relação à categoria de referência, no caso de  $x_j$  ser categórica, mantendo as restantes variáveis constantes;
- $\epsilon_i$  - erro aleatório associado à resposta do indivíduo  $i$ .

### 3.1.1.1 Pressupostos do Modelo de Regressão

Como pressupostos do modelo de regressão, isto é, condições que têm de se ver cumpridas para o modelo ser válido, tem-se

- $E[\epsilon_i]=0$ ;
- $\text{Var}[\epsilon_i]=\sigma^2, \forall i$ , isto é, a variância é constante (homocedasticidade);
- $\epsilon_i$ 's são independentes;
- $\epsilon_i \sim N(0, \sigma^2)$ ;

- $\text{cov}(y_i|\mathbf{x}_i, y_j|\mathbf{x}_j)=0$ , para  $i \neq j$ , onde  $i, j = 1, \dots, n$ ;
- as variáveis explicativas não devem estar correlacionadas.

É essencial que nenhuma destas condições seja violada, sob consequência do modelo não ser adequado.

Mais à frente serão identificadas algumas técnicas para perceber se estes pressupostos são, ou não, cumpridos.

### 3.1.1.2 Estimação dos Parâmetros de um Modelo de Regressão

O maior problema dos modelos lineares prende-se com a estimação dos coeficientes do modelo, de modo a que os erros sejam os mínimos possíveis. Para tal utiliza-se mais comumente o Método dos Mínimos Quadrados (Chatterjee e Hadi, 2006), que acaba por ser um caso particular do Método da Máxima Verosimilhança para os parâmetros  $\beta$  (Rao *et al.*, 2008).

O vetor dos estimadores de mínimos quadrados é dado então por

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

vetor este que é centrado ( $E[\hat{\beta}] = \beta$ ), tem como variância os elementos da diagonal de  $\sigma^2(X^T X)^{-1}$ , é uma combinação linear de  $Y$  e é aquele com menor variância.

Pode então escrever-se

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y$$

e, portanto, considerando novamente uma amostra aleatória de  $n$  observações independentes tem-se  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}$ , donde se sabe que os valores de  $\hat{y}_i$  representam os valores estimados de  $y_i$  e que  $e_i = y_i - \hat{y}_i$  representam os resíduos.  $e = Y - \hat{Y}$  é o vetor, de dimensão  $n \times 1$ , dos resíduos do modelo.

Encontrado um estimador para  $\beta$ , é necessário encontrar também para  $\sigma^2$ . Assim, tem-se, como estimador centrado para  $\sigma^2$  (Chatterjee e Hadi, 2006),

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1} = \frac{\sum_{i=1}^n e_i^2}{n - p - 1} = \frac{SSE}{n - p - 1}$$

onde SSE representa a soma dos quadrados dos resíduos.

### 3.1.1.3 Inferências no Modelo de Regressão

Por vezes existe interesse em saber se um determinado parâmetro do modelo é igual a algum valor específico. Para tal é possível recorrer-se a um teste de hipóteses, o qual vai testar

$$H_0 : \beta_j = b_j \quad \text{vs.} \quad H_1 : \beta_j \neq b_j,$$

com estatística de teste

$$T = \frac{\hat{\beta}_j - b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-p-1}$$

onde  $C_{jj}$  é o  $j$ -ésimo elemento da diagonal da matriz  $C = (X^T X)^{-1}$  (Chatterjee e Hadi, 2006).

O p-valor associado a este teste é dado por  $2 \times P(t_{n-p-1} > |T|)$ .

**Nota:** Usualmente, o interesse passa por saber se algum dos parâmetros é igual a zero, pois isso informa da significância, ou não, que a variável explicativa correspondente tem no modelo.

Pode também ser necessária a estimação do valor esperado para um determinado  $x_0$ , isto é,  $\hat{E}[Y_0]$ . Tem-se  $\hat{E}[Y_0] = \hat{E}[Y|x = x_0] = x_0^T \hat{\beta}$ , cujo intervalo de confiança a  $(1-\alpha)100\%$  é dado por



$$\left] \hat{E}[Y_0] - t_{1-\frac{\alpha}{2};n-p-1} \sqrt{\hat{\sigma}^2 x_0^T C x_0} ; \hat{E}[Y_0] + t_{1-\frac{\alpha}{2};n-p-1} \sqrt{\hat{\sigma}^2 x_0^T C x_0} \left[.$$

Se, por outro lado, o objetivo passar por obter uma estimativa para  $Y$ , dado um ponto  $x_0$ , tem-se  $\hat{Y}_0 = x_0^T \hat{\beta}$ , com um intervalo de confiança a  $(1 - \alpha)100\%$  dado por

$$\left] \hat{Y}_0 - t_{1-\frac{\alpha}{2};n-p-1} \sqrt{\hat{\sigma}^2 [1 + x_0^T C x_0]} ; \hat{Y}_0 + t_{1-\frac{\alpha}{2};n-p-1} \sqrt{\hat{\sigma}^2 [1 + x_0^T C x_0]} \left[.$$

### 3.1.1.4 Qualidade do Modelo

A qualidade do modelo pode ser obtida de diversas formas. Usualmente recorre-se ao coeficiente de determinação ou mesmo a métodos gráficos. Neste estudo o foco será o coeficiente de determinação (Chatterjee e Hadi, 2006).

O coeficiente de determinação, designado por  $R^2$  ( $0 \leq R^2 \leq 1$ ), não é mais do que a percentagem da variabilidade de  $Y$  que é explicada pelo modelo. Mas atenção, o  $R^2$  aumenta sempre que são colocadas mais variáveis explicativas no modelo, sejam estas, ou não, significativas. Assim sendo, o mais razoável é utilizar o coeficiente de determinação ajustado ( $R_a^2$ ), dado este apenas aumentar se a variável introduzida for de facto relevante para o modelo.

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}} = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$

com  $R^2 = 1 - \frac{SSE}{SST}$ , onde  $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  e  $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ .

Se o objetivo for comparar dois modelos entre si, há várias formas de o fazer.

No caso dos modelos a comparar serem aninhados (as variáveis explicativas de um dos modelos formam um subconjunto das variáveis explicativas do outro modelo), pode testar-se apenas se os parâmetros correspondentes às variáveis explicativas, que não aparecem em ambos os modelos, são simultaneamente nulos.

Por outro lado, se os modelos não forem aninhados o mais adequado será usar  $R_a^2$  ou ainda o critério de informação de Akaike (Rao *et al.*, 2008).

O critério de informação de Akaike (AIC) é uma medida para a seleção de modelos, modelos esses que podem ser ordenados de acordo com o valor deste critério, sendo o modelo tanto melhor quanto menor for o seu AIC, que é dado por

$$AIC = -2\log(L) + 2p,$$

onde L representa a verossimilhança do modelo e p o número de parâmetros.

Ambos, AIC e  $R_a^2$ , podem ser também usados na comparação de modelos aninhados.

**Nota:** No caso dos modelos lineares generalizados pode usar-se, entre outras medidas, o AIC como forma de comparação entre modelos.

### 3.1.1.5 Seleção de Variáveis

Os três métodos mais utilizados na seleção de variáveis explicativas são *Backward*, *Forward* e *Stepwise*.

No entanto, apenas será utilizado o processo *stepwise*, o mais completo dos três, que não passa de uma combinação dos outros dois. É essencialmente um processo *forward*, pois começa-se apenas com uma variável e vai-se adicionando variáveis uma a uma. No entanto, em cada passo, é feita uma análise das variáveis já introduzidas, de maneira a que todas as variáveis se mantenham significativas após introdução de uma nova variável, o que corresponderia a um processo *backward*, onde se começa com todas as variáveis possíveis e se vão removendo uma a uma até todas serem estatisticamente significativas.

**Nota:** Para os modelos lineares generalizados a seleção de variáveis é feita da mesma forma.

### 3.1.1.6 Análise de Resíduos e Multicolinearidade

A análise dos resíduos é um conjunto de técnicas que permite descobrir possíveis falhas em alguns dos pressupostos dos modelos.

Computacionalmente, as medidas usadas na análise de resíduos têm como base os resíduos padronizados  $e_i^s$ , que não são mais do que o quociente entre os resíduos  $e_i$  e o desvio-padrão estimado  $\hat{\sigma}$ .

A normalidade dos erros pode ser testada usando um simples teste de Shapiro-Wilks, ou Kolmogorov-Smirnov sobre os resíduos, sob hipótese nula dos erros terem distribuição normal. Pode também testar-se este pressuposto traçando um Normal QQ-plot. Se os erros seguirem uma distribuição gaussiana, os pontos do gráfico devem estar na sua maioria sobre a reta traçada.

Para verificar a homogeneidade da variância, a média nula e a independência dos erros, basta representar graficamente os resíduos *versus* os valores estimados da variável resposta. Os pontos do gráfico devem distribuir-se aleatoriamente em torno do resíduo nulo, formando uma mancha de largura uniforme. Se os resíduos seguirem um padrão, a condição de independência não é satisfeita. Já se a dispersão dos resíduos aumentar ou diminuir com os valores da variável resposta, o pressuposto da homocedasticidade não se vê cumprido.

Para detetar a multicolinearidade, isto é, correlação entre variáveis explicativas, pode recorrer-se tanto à matriz de correlação das variáveis explicativas (se existem variáveis altamente correlacionadas deve eliminar-se aquela que fizer mais sentido, tendo em atenção os dados), como ao VIF (*variance inflation factor*, em inglês).

Recorde-se que é pressuposto do modelo de regressão não existir correlação entre as variáveis independentes.

Neste caso, optou-se pela utilização do VIF, que é dado por

$$VIF = \frac{1}{1 - R_j^2}$$

onde  $R_j^2$  é o coeficiente de determinação da regressão da variável  $X_j$  em função das restantes variáveis explicativas.

Quando  $VIF \simeq 1$  as variáveis explicativas são independentes. Já se  $VIF > 10$  as

variáveis explicativas estão correlacionadas. Neste caso será usado como ponto de corte um *VIF* superior a 10 (Craney e Surlles, 2002). Há quem considere pontos de corte mais baixos, como 5.

**Nota:** No caso dos modelos lineares generalizados o cálculo do *VIF* é feito considerando a distribuição gaussiana, à semelhança do que acontece nos modelos de regressão linear.

### 3.1.2 Modelos Lineares Generalizados

Com a existência de dados que não eram explicados de maneira linear ou normal, foi necessário o desenvolvimento de novos tipos de modelos, como, por exemplo, os modelos *log-log* para ensaios de diluição (Fisher, 1922), *probit* (Bliss, 1935) e *logit* (Berkson, 1944; Dyke e Patterson, 1952) para proporções, *log-lineares* para dados de contagens (Birch, 1963) ou os modelos de regressão para análise de sobrevivência (Feigl e Zelen, 1965; Zippin e Armitage, 1966; Glasser, 1967). Todos estes modelos apresentam uma estrutura linear e, em todos a variável resposta segue uma distribuição pertencente à família exponencial de distribuições.

Mais tarde, Nelder e Wedderburn, em 1972, introduziram então os modelos lineares generalizados (MLG's), sendo estes uma síntese de todos os outros modelos que tinham sido usados até à data para modelar dados não gaussianos.

Os MLG's são, portanto, não só extensão de todos esses modelos, mas também do próprio modelo linear clássico, que modelam o valor esperado da variável resposta. Têm, no entanto, duas particularidades, a distribuição da variável resposta pertence sempre à família exponencial de distribuições e, a relação entre variável resposta e variáveis independentes é dada por uma qualquer função diferenciável, mantendo-se sempre a estrutura de linearidade.

Obviamente os MLG's são ainda limitados, já que mantêm a estrutura linear, só podem ser aplicados a dados cujas distribuições pertençam à família exponencial e exigem a independência das respostas.

### 3.1.2.1 A Família Exponencial

A variável resposta,  $Y$ , tem distribuição pertencente à família exponencial se a sua função densidade de probabilidade ou função massa de probabilidade se escrever da forma

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

onde  $\theta$  é o parâmetro de localização,  $\phi$  o parâmetro de dispersão e as funções  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções reais conhecidas para cada distribuição.

No caso de  $\phi$  ser conhecido tem-se uma distribuição da família exponencial com parâmetro canônico  $\theta$ . Por outro lado, caso  $\phi$  seja desconhecido a distribuição pode não fazer parte da família exponencial. Admite-se ainda, que a função  $b(\cdot)$  é diferenciável, que o suporte da distribuição não depende dos parâmetros e que em muitas situações,  $a(\phi) = \frac{\phi}{\omega}$ , onde  $\omega$  é uma constante conhecida (Turkman e Silva, 2000).

### 3.1.2.2 Valor Médio e Variância

Sendo  $Y$  uma variável resposta com distribuição pertencente à família exponencial, tem-se

- $E[Y] = \mu = b'(\theta)$ ;
- $\text{Var}(Y) = b''(\theta)a(\phi)$ .

onde  $b'(\theta)$  e  $b''(\theta)$  são a primeira e segunda derivadas de  $b(\theta)$ , respetivamente.

Como se pode ver, a variância de  $Y$  é o produto de duas funções:  $b''(\theta)$ , que apenas depende do parâmetro de localização  $\theta$ ;  $a(\phi)$ , que depende apenas do parâmetro de dispersão  $\phi$ . Ao termo  $b''(\theta)$  dá-se o nome de função de variância ( $V(\mu)$ ), função esta que caracteriza a distribuição.

$$V(\mu) = \frac{d\mu}{d\theta}$$

### 3.1.2.3 Componentes de um MLG

Um modelo linear generalizado assenta sobre três componentes distintas

1. Componente Aleatória: identifica a variável resposta,  $Y$ , a modelar, tratando-se esta de uma variável aleatória, da qual se recolhem  $n$  observações independentes e cuja distribuição pertence à família exponencial;
2. Componente Sistemática: especifica as variáveis explicativas, considerando uma combinação linear das mesmas.

O valor esperado  $\mu_i$  e o preditor linear  $\eta_i = X_i^T \beta$  estão relacionados através da seguinte condição

$$\mu_i = h(\eta_i) = h(X_i^T \beta)$$

onde  $h$  é uma função monótona e diferenciável,  $\beta$  é um vetor de parâmetros,  $X_i$  é o vetor das variáveis explicativas e  $g = h^{-1}$  é a função de ligação que relaciona  $\mu_i$  com o preditor linear, ou seja,

$$g(\mu_i) = \eta_i, \quad i = 1, \dots, n$$

3. Função de Ligação: relaciona as componentes aleatória e sistemática através de uma função, a chamada função de ligação,  $g = h^{-1}$ , função esta monótona e diferenciável.

São de especial interesse os casos em que preditor linear e parâmetro de localização coincidem, isto é,  $\theta_i = \eta_i$ , donde se obtém que  $\theta_i = X_i^T \beta$ . Quando isto acontece a função de ligação é denominada de função de ligação canónica. As funções de ligação canónicas são mais utilizadas no contexto dos MLG's, já que estas garantem a concavidade da função de verosimilhança e, portanto, muitos resultados assintóticos são obtidos mais facilmente.

### 3.1.2.4 Estimação dos parâmetros de um MLG

A estimação dos parâmetros de um modelo linear generalizado é feita usando, tipicamente, o método da máxima verosimilhança (MMV) (Nelder e Wedderburn, 1972). Desta forma, os estimadores obtidos são consistentes, assintoticamente eficientes, com distribuição assintoticamente normal.

O MMV tem como objetivo a maximização da função de verosimilhança, o que é equivalente a maximizar a função *log-verosimilhança*. O logaritmo da função de verosimilhança, em função de  $\beta$ , supondo  $\phi$  conhecido, é dado por

$$\log(L(\beta)) = l(\beta) = \sum_{i=1}^n \left\{ \frac{w_i(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi, w_i) \right\} = \sum_{i=1}^n l_i(\beta)$$

onde  $l_i(\beta) = \frac{w_i(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi, w_i)$  é a contribuição de cada observação da variável resposta para a verosimilhança.

Assim sendo, os estimadores de máxima verosimilhança para  $\beta$  são calculados como solução das equações de verosimilhança,

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta_j}, \quad j = 1, \dots, p$$

equações essas dadas por (Turkman e Silva, 2000)

$$\sum_{i=1}^n \frac{(y_i - \mu_i)z_{ij}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \dots, p$$

onde  $z_{ij} = \frac{\partial \eta_i(\beta_i)}{\partial \beta_j}$ .

Geralmente, estas equações não são lineares, tendo, por isso, de ser resolvidas com recurso a processos iterativos de maximização. Na sua resolução usa-se, em geral, o método de *scores* de Fisher, por ser este o mais simples. Este método utiliza como matriz de covariância, a matriz de informação de Fisher  $I(\beta)$ ,

$$I(\beta) = E \left[ -\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} \right] = E \left[ -\frac{\partial S(\beta)}{\partial \beta} \right]$$

onde  $S(\beta) = \sum_{i=1}^n S_i(\beta) = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta_j}$ ,  $j = 1, \dots, p$ , é a chamada função *score*.

Na forma matricial,  $I(\beta) = X^T W X$ , em que  $W$  é a matriz diagonal de ordem  $n$  e o seu  $i$ -ésimo elemento é dado por

$$W_i = \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{\text{var}(Y_i)}.$$

Assim sendo, o estimador de máxima verosimilhança de  $\beta$  na  $(k+1)$ -ésima iteração é dado por

$$\hat{\beta}^{(k+1)} = (X^T W^{(k)} X)^{-1} X^T W^{(k)} u^{(k)}$$

em que  $u^{(k)} = \eta_i^{(k)} + \left(y_i - \mu_i^{(k)}\right) \frac{\partial \eta_i^{(k)}}{\partial \mu_i^{(k)}}$  (Turkman e Silva, 2000).

Para o cálculo do estimador de  $\beta$  é irrelevante o conhecimento do parâmetro de dispersão.

O parâmetro de dispersão pode ser estimado, também, usando o MMV, no entanto existe um método mais simples, que assenta na distribuição de amostragem, para grandes valores de  $n$ , da estatística de Pearson Generalizada (Turkman e Silva, 2000). Assim sendo, tem-se

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{\omega_i (y_i - \hat{\mu}_i)^2}{V(\mu_i)}$$

sendo este um estimador consistente de  $\phi$ .

**Nota:** Para o caso particular em que a distribuição dos dados é gaussiana tem-se

- $\theta = \mu$ ;
- $\phi = \sigma^2$  e  $\omega = 1 \Rightarrow a(\phi) = \sigma^2$ ;
- $c(y, \phi) = -\frac{1}{2}\left(\frac{y^2}{\phi} + \log(2\pi\phi)\right)$ ;
- $b(\theta) = \frac{\theta^2}{2}$ ;



- $E[Y] = b'(\theta) = \mu = \theta$ ;
- $Var(Y) = \sigma^2$ ;
- a função variância é dada por  $V(\mu) = 1$ ;
- a função de ligação canónica é a função identidade:  $\eta = \mu$ .

sendo os resultados equivalentes à regressão linear da secção anterior.

### 3.1.2.5 Inferências nos MLG's

Existem essencialmente três estatísticas para testar

$$H_0 : C\beta = \delta \quad vs. \quad H_1 : C\beta \neq \delta,$$

onde  $C$  é uma matriz  $q \times p$  com  $q \leq p$ , de característica completa  $q$  e  $\delta$  um vetor de dimensão  $q$ .

As três: Estatística de Wald (utilizada em geral para testar hipóteses nulas sobre componentes individuais), Estatística de Razão de Verosimilhanças (utilizada, em geral, para comparar modelos encaixados) e a Estatística Score (baseada nas propriedades assintóticas da função *score*), são deduzidas das distribuições assintóticas dos estimadores de máxima verosimilhança (Turkman e Silva, 2000).

### 3.1.2.6 Qualidade do Modelo

Com objetivo de se perceber a qualidade de ajustamento do modelo aos dados é possível recorrer-se, além do Critério de Akaike à Análise do Desvio.

Seja  $D(y, \hat{\mu}) = -2\phi(l_M(\hat{\beta}_M) - l_S(\hat{\beta}_S))$  a função desvio onde  $l_S(\hat{\beta}_S)$  é a log-verosimilhança do modelo saturado (modelo em que o número máximo de parâmetros é igual a  $n$ ) e  $l_M(\hat{\beta}_M)$  é a log-verosimilhança do modelo em estudo.

Considerando  $M_1$  e  $M_2$ , modelos encaixados, com  $p_1$  e  $p_2$  parâmetros, respetivamente, tem-se, sob hipótese nula de que o modelo mais simples,  $M_2$ , é melhor

$$\frac{D(M_2) - D(M_1)}{\phi} \sim \chi_{p_1 - p_2}^2$$

onde  $D(M_1)$  é a função desvio do modelo  $M_1$  e  $D(M_2)$  é a função desvio do modelo  $M_2$ ;

### 3.1.2.7 Análise de Resíduos

Para que o modelo seja adequado é necessário que os pressupostos do modelo sejam cumpridos.

Com objetivo de avaliar a evidência de variância não constante e média nula, basta representar graficamente os resíduos *versus* os valores previstos  $\hat{\mu}$ , ou uma transformação destes ( $2\sqrt{\hat{\mu}}$  no caso do modelo de Poisson, segundo McCullagh e Nelder (1989)). Os pontos do gráfico devem estar dispostos em torno de zero, com uma amplitude constante, sem estarem dispostos em nenhum padrão perceptível. Não sendo visível qualquer padrão pode considerar-se também a independência entre observações.

### 3.1.2.8 Modelo de Poisson

No caso da variável resposta  $Y$ , conter dados de contagem é usual utilizar um modelo de regressão de Poisson, isto é,  $Y \sim Poisson(\mu)$ .

Neste modelo

- $\theta = \log(\mu)$ ;
- $\phi = 1$  e  $\omega = 1 \Rightarrow a(\phi) = 1$ ;
- $c(y, \phi) = -\log(y!)$ ;
- $b(\theta) = e^\theta$ ;
- $E[Y] = b'(\theta) = e^\theta = \mu$ ;

- $Var(Y) = b''(\theta)a(\phi) = e^\theta = \mu$ ;
- a função variância é dada por  $V(\mu) = \mu$ ;
- a função de ligação canónica é a função logarítmica:  $\eta = \log(\mu)$ .

e, portanto,

$$f(y; \mu) = \exp \{y \log(\mu) - \mu - \log(y!)\}.$$

A distribuição de Poisson é, também, bastante utilizada na modelação de taxas, surgindo como alternativa à distribuição binomial.

A variável dependente  $Y$ , continua a ser composta por dados de contagens e, portanto,  $Y \sim Poisson(\mu)$ . No entanto, tem-se que  $\mu = \frac{\lambda}{l}$  onde  $\lambda$  representa o número de eventos esperados e  $l$  a quantidade total de indivíduos em cada sub-grupo de interesse.

Como a função de ligação é exatamente a mesma, tem-se

$$\log(\mu) = X^T \beta \Leftrightarrow \log(\lambda) = \log(l) + X^T \beta.$$

Este último dado é conhecido como *offset*, pois não lhe é atribuído qualquer coeficiente.

Quando o termo *offset* está presente, o *software R* estima os preditores lineares e os valores ajustados para  $\lambda_i$  ao invés de estimar  $\mu_i$ , ou seja,

- $\eta_i = \log(l_i) + \beta x_i$ ;
- $\lambda_i = \exp(\eta_i) = \exp(\log(l_i) + \beta x_i)$ .

No manuseamento de dados de contagem, o modelo de regressão de Poisson revela-se muitas vezes ineficaz, já que os dados apresentam muitas vezes sobredispersão.

Em dados de contagem reais verifica-se muitas vezes que  $Var(Y) > E[Y]$ , o que vai contra o que se sabe do modelo de Poisson, onde a média e variância são iguais. Quando isto

acontece está-se perante um caso de sobredispersão no modelo, o que pode levar a grandes flutuações aquando do cálculo dos erros padrão das estimativas. Esta pode ser detetada analisando o valor do desvio (D); quando este é maior do que um está-se, usualmente, na presença de uma situação de sobredispersão.

A sobredispersão pode dever-se a vários fatores, como falta de covariáveis, existência de *outliers* na variável resposta, escolha errada da função de ligação e correlação entre as observações (Hilbe, 2007). A existência de sobredispersão pode ainda dever-se à existência de correlação temporal nos dados (Zeger, 1998).

### 3.1.2.9 Modelo de Binomial Negativa

Uma das alternativas mais usuais aquando da existência de sobredispersão nos dados é a utilização de um modelo binomial negativo.

Assim sendo, dado  $Y \sim BN(\mu, k)$  tem-se para este modelo

- $\theta = \log\left(\frac{\mu}{\mu+k}\right)$ ;
- $\phi = 1$  e  $\omega = 1 \Rightarrow a(\phi) = 1$ ;
- $c(y, k) = \log\left(\frac{\Gamma(k+y)}{\Gamma(k)y!}\right)$ ;
- $b(\theta) = -k \log\left(\frac{k}{k+\mu}\right)$ ;
- $E[Y] = b'(\theta) = \frac{ke^\theta}{1-e^\theta} = \mu$ ;
- $Var(Y) = b''(\theta)a(\phi) = \frac{ke^\theta}{(1-e^\theta)^2} = \mu + \frac{\mu^2}{k}$ ;
- a função de ligação canónica é a função logarítmica:  $\eta = \log(\mu)$ .

e, portanto,

$$f(y; \mu, k) = \exp\left\{y \log\left(\frac{\mu}{\mu+k}\right) - k \log\left(\frac{k}{k+\mu}\right) + \log\left(\frac{\Gamma(k+y)}{\Gamma(k)y!}\right)\right\}.$$

## 3.2 Modelos de Séries Temporais

Uma série temporal pode ser definida como um conjunto de observações de certa variável ( $X$ ),  $x_1, x_2, \dots, x_t$ , ordenadas, com  $t \in \mathbb{N}$  (série temporal discreta) ou  $t \in \mathbb{R}$  (série temporal contínua) correlacionadas no tempo, isto é, uma série temporal pressupõe a existência de uma estrutura de correlação temporal. As observações são, em geral, recolhidas aleatoriamente e, portanto, independentes e identicamente distribuídas (i.i.d.) (Brockwell e Davis, 2002).

Numa análise de séries temporais, os principais objetivos passam por encontrar o melhor modelo que explique o comportamento de determinado fenómeno, ao longo do tempo e, usar esse modelo para fazer previsão, simulações e tomadas de decisão. Para tal recorre-se a funções determinísticas do tempo, ao passado já observado e ainda a outras variáveis aleatórias.

Sendo uma série temporal um conjunto de variáveis aleatórias ordenadas no tempo, pode dizer-se que uma série temporal é um processo estocástico.

Assim sendo, dado um processo estocástico  $X_t$ , tal que  $\forall t$  se tem  $E[X_t^2] < \infty$ , então

- $E[X_t] = \mu_t$ , representa a função valor médio;
- $Var(X_t) = E[(X_t - \mu_t)^2] = \sigma_t^2$ , representa a variância;
- $Cov(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] = \gamma(t_1, t_2)$ , representa a covariância;
- $Corr(X_{t_1}, X_{t_2}) = \frac{\gamma(t_1, t_2)}{\sigma_{t_1} \sigma_{t_2}} = \rho(t_1, t_2)$ , representa a função de correlação.

O comportamento de uma série temporal  $X_t$ , é explicado pelos seus valores passados e pelos valores passados e presentes de um erro estocástico, isto é

$$X_t = f(X_{t-1}, X_{t-2}, \dots, \epsilon_t, \epsilon_t - 1, \dots)$$

onde  $\epsilon_t$  são i.i.d. com distribuição  $N(0, \sigma_\epsilon^2)$ .  $\sigma_\epsilon^2$  designa-se por variância dos erros aleatórios.

### 3.2.1 Processos Estacionários

Muitas vezes uma série temporal não apresenta um comportamento estacionário, isto é, a série sofre alterações significativas à medida que o tempo vai passando. Mas o que significa, matematicamente falando, uma série temporal ser estacionária?

Apesar da definição de série temporal estacionária existir, na prática esta não é usualmente utilizada por ser de difícil cumprimento. Assim usa-se antes, como forma de garantir estacionaridade, a definição de série fracamente estacionária. Pode dizer-se então que uma série é estacionária (fracamente estacionária) se

- $\mu_t = \mu$ , isto é, o valor médio não depende do tempo;
- $\sigma_t^2 = \sigma^2$ , isto é, o valor da variância não depende do tempo;
- $Cov(X_{t_1}, X_{t_2}) = \gamma(|t_2 - t_1|)$ , isto é, a covariância entre as observações, para diferentes pontos no tempo depende apenas do desfasamento  $t_2 - t_1$ .

Um caso de modelo estacionário é o chamado ruído branco, onde

- $E[X_t] = \mu$  (usualmente  $\mu = 0$ );
- $Var(X_t) = \sigma^2$ , independentemente de  $t$ ;
- $Cov(X_{t_1}, X_{t_2}) = 0$ ,  $t_1 \neq t_2$ .

Não estando na presença de uma série estacionária, existem diversas formas de transformar os dados para que estes se tornem estacionários.

Inicialmente deve perceber-se se a não estacionaridade se deve apenas à existência de variância não constante. Assim, e se depois de estabilizada a variância por meio de uma transformação, se continuar a não registar estacionaridade, é altura de recorrer a outros

métodos. As transformações usualmente utilizadas são as transformações de Box-Cox que estão descritas em mais detalhe na secção 3.3.

Pode então retirar-se à série

- tendência: identifica a inclinação da série, ao longo do tempo, e pode ser consequência dos valores observados serem dependentes de uma componente temporal;
- sazonalidade: identifica certos períodos "especiais" no tempo, onde ocorrem aumentos/ decréscimos que se repetem regularmente;
- movimentos oscilatórios: identificam fases alternadas de expansão e depressão que não apresentam qualquer periodicidade.

Alguns dos modelos mais usados para modelar a tendência de uma série temporal são os modelos polinomiais (Chatfield, 2004). Os três mais utilizados são

- $X_t = a + \epsilon_t$ , quando se está na presença de uma tendência simples;
- $X_t = a + bt + \epsilon_t$ , quando se está na presença de uma tendência linear;
- $X_t = a + bt + ct^2 + \epsilon_t$ , quando se está na presença de uma tendência quadrática.

onde  $a$ ,  $b$  e  $c$  são coeficientes a ser estimados.

Um polinómio com grau superior a 2 já não faz sentido, sendo de difícil interpretação.

Com objetivo de modelar a sazonalidade tem de se perceber quais os fatores que podem estar a repetir-se regularmente, podendo estes ser meses, semanas ou datas especiais. Posto isso, basta definir um modelo de regressão usando variáveis *dummy*.

Por exemplo, no caso do fator de sazonalidade serem os meses do ano, tem-se

$$X_t = a + bt + \alpha_1 \text{ Se(Mês=Janeiro)} + \dots + \alpha_{12} \text{ Se(Mês=Dezembro)} + \epsilon_t.$$

Um modelo adequado para ajustar um ciclo é (Brockwell e Davis, 2002)

$$X_t = c + A \cos(\omega t) + B \sin(\omega t) + \epsilon_t$$

onde  $\omega = 2\pi f$  e  $f$  representa a frequência, em ciclos, por unidade de tempo ( $f \geq 0$ ).  $c$ ,  $A$  e  $B$  representam os coeficientes a ser estimados.

Por vezes existe a necessidade de adicionar vários ciclos afim de melhor o modelo. Estes são conhecidos como harmônicos pois, as suas frequências são múltiplas da primeira frequência encontrada,  $\omega$ . Por exemplo, no caso de existirem três ciclos, tem-se

$$X_t = c + A_1 \cos(\omega t) + B_1 \sin(\omega t) + A_2 \cos(2\omega t) + B_2 \sin(2\omega t) + A_3 \cos(3\omega t) + B_3 \sin(3\omega t) + \epsilon_t.$$

### 3.2.1.1 O Periodograma

O periodograma é uma ferramenta, aplicada aos resíduos de um modelo, que permite descobrir se uma série contém ciclos e quais a sua amplitude,  $R$  e frequência,  $f$  ( $0 \leq f \leq \frac{1}{2}$ ) (Brockwell e Davis, 2002).

Um ciclo único de amplitude  $R$  e frequência  $f$  terá um periodograma com um pico no ponto de frequência  $f$ , com altura de cerca de  $\frac{1}{2}nR^2$  e largura de  $\frac{1}{n}$ . O maior pico será ladeado de picos menores, com frequências de  $f \pm \frac{1}{n}$ ,  $f \pm \frac{2}{n}$ ...

Se a série tiver variados ciclos, então, as suas frequências diferem em pelo menos um ciclo. No caso de existir mais do que um ciclo, o periodograma mostrará picos distintos, cada um correspondente a determinado ciclo.

A existência de tendência na série temporal será também visível no periodograma. Se existir um pico na frequência zero, muitas vezes maior do que qualquer outro pico, então a série temporal apresenta uma tendência que tem de ser removida antes de examinar novamente o periodograma, com objetivo de encontrar um ciclo.



### 3.2.1.2 Processos autoregressivos de ordem $p$

$X_t$  é um processo autoregressivo de ordem  $p$ , designado por  $AR(p)$ , se é estacionário e se para cada  $t$

$$X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \epsilon_t.$$

Ou seja, num processo autoregressivo de ordem  $p$ , o valor de  $X_t$  depende tanto das observações passadas e valores passados de outras variáveis ( $\mu$ ) como de um processo de ruído branco,  $\epsilon_t$ .  $\epsilon_t$  é independente de  $X_{t+k}$ , para todo o  $k$ .

$\phi_i, i = 1, \dots, p$  são os coeficientes de regressão estimados.

### 3.2.1.3 Processos de médias móveis de ordem $q$

$X_t$  é um processo de médias móveis de ordem  $q$ , designado por  $MA(q)$ , se é estacionário e se para cada  $t$

$$X_t = \mu + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q}.$$

Ou seja, num processo de médias móveis de ordem  $q$ , o valor de  $X_t$  depende tanto dos valores passados dos erros, como do presente ruído branco,  $\epsilon_t$ , além dos valores passados de outras variáveis ( $\mu$ ).

$\theta_i, i = 1, \dots, q$  são os parâmetros de regressão estimados.

### 3.2.1.4 Processos mistos autoregressivos e médias móveis

Os modelos mistos, designados de  $ARMA(p, q)$ , não são mais do que uma combinação de ambos os processos anteriores,  $AR(p)$  e  $MA(q)$ . Assim sendo,  $X_t$  é definida em função dos valores passados da série e dos valores passados dos erros.  $\phi_i$  e  $\theta_i$  ( $i = 1, \dots, p$ ) são os parâmetros do modelo, que, no caso de ser multivariado, é dado por

$$X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q}.$$

### 3.2.1.5 Função de autocovariância e função de autocorrelação

Num processo estacionário existem duas funções essenciais para que se possa definir um modelo: as funções de autocovariância e autocorrelação.

A função de autocovariância mede a intensidade com que pares de valores, com um espaçamento de amplitude  $k$ , se acompanham e é dada por

$$\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)].$$

Assim sendo, tem-se que a função de autocorrelação (FAC), que mede a correlação entre pares de valores, separados por um intervalo de amplitude  $k$ , é dada por

$$\rho_k = Corr(X_t, X_{t+k}) = \frac{\gamma_k}{\gamma_0}$$

pelo que se tem  $\sigma^2 = \gamma_0$ .

A função de autocorrelação assume sempre valores entre -1 e 1.

Na maioria das vezes, além da correlação, interessa também estudar a correlação parcial entre  $X_t$  e  $X_{t+k}$ , que é dada pela função de autocorrelação parcial, designada de FACP. Esta mede, além da correlação global entre pares de valores separados por um intervalo de amplitude  $k$ , a correlação parcial entre os mesmos quando se fixam variáveis intermédias  $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ .

Usando ambas FAC e FACP pode então chegar-se a uma conclusão quanto ao modelo a ser utilizado na modelação da série temporal, de acordo com o que se pode ver na tabela 3.1.

Tabela 3.1: Comparação da FAC e FACP para os processos ARMA(p,q)

<b>Processo</b>	<b>FAC</b>	<b>FACP</b>
AR(p)	Decresce exponencial e/ou sinusoidalmente para zero	Anula-se a partir de um intervalo maior do que p
MA(q)	Anula-se a partir de um intervalo maior do que q	Decresce exponencial e/ou sinusoidalmente para zero
ARMA(p,q)	Decresce exponencial e/ou sinusoidalmente para zero	Decresce exponencial e/ou sinusoidalmente para zero

### 3.3 Transformações de Box Cox

Box e Cox (1964) propuseram uma transformação paramétrica com objetivo de estabilizar a variância de uma série ao longo do tempo.

$$Y_t = \begin{cases} \frac{(X_t-1)^\lambda}{\lambda}, & \text{se } \lambda \neq 0. \\ \log(X_t), & \text{se } \lambda = 0. \end{cases}$$

Estas transformações podem ser utilizadas no contexto de séries temporais. No entanto, são também utilizadas no contexto de MLG's com objetivo de transformar a variável resposta, de forma a que esta siga uma distribuição gaussiana, sendo assim possível aplicar técnicas de regressão linear.

### 3.4 MAPE

O erro percentual absoluto médio (*mean absolute percentage error* em inglês) é uma medida que expressa o erro cometido, em percentagem, no ajuste de um modelo, isto é, mede a precisão com que o modelo estima. O MAPE é uma medida que não depende

da escala dos dados, pelo que facilita a comparação entre múltiplos modelos (Jones *et al.*, 2008).

Ou seja, caso  $MAPE = 5\%$  então, em média, a estimação feita pelo modelo em estudo, quando comparada com os valores reais observados, estará errada em cerca de 5%.

Tem-se

$$MAPE = 100\% \times \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

onde  $y_i$  representa os valores realmente observados e  $\hat{y}_i$  os valores estimados pelo modelo. Note-se ainda que os valores de  $y_i$  não podem ser zero, pelo que esta medida apenas pode ser utilizada nos casos em que a variável resposta toma valores não nulos.

### 3.5 Teorema do Limite Central

O Teorema do Limite Central (TLC) diz que para um grande número de variáveis aleatórias com idêntico comportamento, a distribuição da soma das variáveis aleatórias aproxima-se de uma distribuição normal (Ross, 2010), isto é, sendo  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com média  $\mu$  e variância  $\sigma^2$ , quando  $n \rightarrow \infty$

$$\frac{(\sum_{i=1}^n X_i) - n\mu}{\sigma\sqrt{n}} \sim N(0, 1) \Leftrightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

### Análise Exploratória

---

#### 4.1 Serviço de Urgência

A base de dados "Urgência" é composta por 749326 indivíduos e 16 variáveis, 2 quantitativas e 5 qualitativas, sendo que as restantes correspondem a data e horas. Na Tabela 4.1 pode ver-se a descrição desta base de dados, que não contém qualquer valor em falta.

Dada a existência de muitos níveis em certas variáveis optou por fazer-se o agrupamento dos mesmos.

Um desses casos foi a variável Concelho (com 271 níveis), onde o agrupamento das classes foi de acordo com os ACES, acabando por ficar-se com a variável Local de Residência já apresentada na Tabela 2.1. O outro caso observou-se na variável Destino (com 14 níveis) que foi agrupada em 5 grupos com características semelhantes. Assim, a classe "Abandono" comporta 4 dos anteriores níveis ("Abandono antes da Triagem", "Abandono Após Observação Médica", "Abandono após Triagem" e "Saída Contra Parecer Médico"), "Exterior" comporta 5 níveis ("Alta Administrativa", "ARS/ Centro de Saúde", "Consulta Externa", "Exterior Não Referenciado" e "Rede Cuidados Continuados"), "Outro Hospital" contém apenas os indivíduos que foram, de facto, transferidos para outro hospital, "Falecido" contém 3 das anteriores categorias ("Falecido com Autópsia", "Falecido sem Autópsia" e "Morgue (Entrou Cadáver)") e "Serviço de Internamento" contém apenas os pacientes transferidos para o internamento do HB.

Relativamente à variável Cor Triagem é de notar que apesar das pulseiras coloridas tomarem apenas 5 cores, de acordo com a triagem de Manchester, esta tem 6 níveis. Isto acontece devido à existência de pacientes que apesar de fazerem o seu registo no SU

abandonam o mesmo antes de efetuada a triagem e, portanto, é necessária a existência de um nível "Não Triado" que não é na realidade nenhuma cor de pulseira.

Note-se finalmente que nem todas as variáveis serão estudadas ao longo do trabalho, como é o caso das horas, já que não são de interesse para o objetivo do mesmo.

Tabela 4.1: Descrição da Base de Dados Urgência

Variável	Descrição da Variável	Tipo de Variável
Episódio	Episódio da urgência atribuído a cada paciente	Quantitativa ( $\in \mathbb{N}$ )
Local	Urgência na qual o paciente é admitido	Qualitativa (3 níveis)
Sexo	Sexo do paciente	Qualitativa (2 níveis)
Idade	Idade do paciente	Quantitativa (anos)
Data.Nascimento	Data de nascimento do paciente	Qualitativa (datas)
Concelho	Concelho de residência do paciente	Qualitativa (271 níveis)
Data.Admissão	Data de admissão do paciente na urgência	Qualitativa (datas)
Hora.Admissão	Hora de admissão do paciente na urgência	Qualitativa (horas)
Data.Altas	Data da alta do paciente	Qualitativa (datas)
Hora.Altas	Hora da alta do paciente	Qualitativa (horas)
Cor.Triagem	Cor da pulseira atribuída ao paciente na triagem	Qualitativa (6 níveis)
Data.Triagem	Data a que o paciente é triado	Qualitativa (datas)
Hora.Triagem	Hora a que o paciente é triado	Qualitativa (horas)
Data.1 <sup>a</sup> .Obs.	Data em que ocorre a 1. <sup>a</sup> observação médica	Qualitativa (datas)
Hora.1 <sup>a</sup> .Obs.	Hora a que ocorre a 1. <sup>a</sup> observação médica	Qualitativa (horas)
Destino	Destino do paciente aquando da alta	Qualitativa (14 níveis)

Ao longo dos quatro anos em estudo observou-se a chegada de 509413 pacientes à urgência geral, 52657 à urgência obstétrica e 187256 à urgência pediátrica. Pode então perceber-se que a maioria das admissões feitas no SU é feita no serviço geral (cerca de

67.98%), enquanto que o serviço pediátrico constitui aproximadamente 25% e o serviço obstétrico aproximadamente 7% do total das admissões. Estes resultados podem ser observados através da leitura da Tabela 4.2.

Tabela 4.2: Tabela de Frequências Absolutas e Relativas do Tipo de Urgência

<b>Tipo de Urgência</b>	<b>Freq. Absoluta</b>	<b>Freq. Relativa * 100%</b>
Geral	509413	67.98%
Obstétrica	52657	7.03%
Pediátrica	187256	24.99%

#### 4.1.1 Urgência Geral

Como já referido, entre 1 de janeiro de 2012 e 31 de dezembro de 2015 foram admitidos na urgência geral do HB cerca de 509413 pacientes, sendo a maioria do sexo feminino (54.32%). Na Tabela 4.4 é visível este resultado, bem como frequências absolutas e relativas para as restantes variáveis qualitativas de interesse.

Pode ver-se que a maior parte dos pacientes têm morada na zona do ACES Cávado I e Cávado II (50.15% e 29.96%, respetivamente), como seria expetável, uma vez ser este o hospital de referência de 1<sup>a</sup> linha para estas duas zonas.

Os cuidados de saúde prestados são, em geral, urgentes e pouco urgentes (60.27% e 29.15 %, respetivamente). Note-se ainda 0.32% dos pacientes que após serem registados não comparecem ao serviço de triagem. Isto deve-se, provavelmente, a tempos de espera excessivos.

Aquando da sua saída do SU os pacientes tomam variados destinos, sendo que 84.70% destes são reencaminhados para o exterior do SU, seja para os centros de saúde, consultas externas no HB ou simplesmente para as suas residências. Da totalidade de pessoas recebidas somente cerca de 48908 (9.6%) seguiram para o SI.

Os pacientes admitidos no SU Geral têm idades compreendidas entre os 19 e os 106 anos, sendo a média das suas idades cerca de 53.60 anos, com um desvio padrão de 20.12,

como descrito na Tabela 4.3.

Tabela 4.3: Tabela Sumária da Variável Quantitativa Idade

Variável	Mín.	1.º Q	Mediana	3.º Q	Máx.	Média	Desvio Padrão
Idade	19.00	37.00	53.00	70.00	106.00	53.60	20.12

Tabela 4.4: Tabela de Frequências Absolutas e Relativas

Variável	Categoria	Freq. Absoluta	Freq. Relativa * 100%
Sexo	F	276695	54.32%
	M	232718	45.68%
Local de Residência	ACES Cávado I	255463	50.15%
	ACES Cávado II	152603	29.96%
	ACES Cávado III	42783	8.40%
	ACES Ave	14664	2.88%
	ACES Alto Ave	19925	3.91%
	Viana do Castelo	11702	2.30%
	Outros	11682	2.29%
	Estrangeiro	591	0.12%
Cor da Pulseira	Azul	3637	0.71%
	Verde	148512	29.15%
	Amarelo	307019	60.27%
	Laranja	45863	9.00%
	Vermelho	2756	0.54%
	Não Triado	1626	0.32%
Destino	Abandono	15413	3.03%
	Exterior	431490	84.70%
	Outro Hospital	12308	2.42%
	Falecido	1294	0.25%
	Internamento	48908	9.60%



Nas tabelas que se seguem encontra-se uma breve análise descritiva do número de admissões diárias na urgência geral, de acordo com o dia da semana, mês, estação do ano e o facto de ser, ou não, feriado.

Sabe-se então que, o número de pacientes admitidos na urgência geral é menor nos dias de feriado que nos restantes dias do ano, como se pode ver na Tabela 4.5, com uma média de 302.80 ( $\pm 39.27$ ) *versus* uma média de 350.10 pacientes ( $\pm 44.55$ ). É possível então que o facto do dia em estudo ser, ou não, feriado, tenha interesse na modelação do número de pacientes admitidos.

Da leitura da Tabela 4.6 pode concluir-se que o número de pacientes que se dirigem a este serviço de urgência é maior à segunda-feira, com uma média de cerca de 399.09 pacientes ( $\pm 35.70$ ) e vai diminuindo ao longo da semana, atingido o seu mínimo ao domingo com uma média de 292.19 pacientes ( $\pm 27.37$ ).

Tabela 4.5: Tabela Sumária do Número de Admissões, por dia, de acordo com categorização de feriado

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Feriado	204.00	278.00	311.50	329.20	406.00	302.80	39.27
Outros	221.00	320.00	352.00	379.00	492.00	350.10	44.55

Tabela 4.6: Tabela Sumária do Número de Admissões, por dia, em cada dia da semana

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Segunda-feira	274.00	374.00	400 .00	422.00	492.00	399.09	35.70
Terça-feira	247.00	338.00	358.00	379.00	460.00	359.56	33.44
Quarta-feira	231.00	336.00	362.00	381.00	429.00	356.57	34.59
Quinta-feira	237.00	337.00	358.00	378.00	472.00	356.29	33.02
Sexta-feira	276.00	339.00	357.00	376.25	489.00	358.37	31.56
Sábado	221.00	299.75	318.50	338.00	432.00	318.56	29.53
Domingo	204.00	274.00	291.00	312.00	367.00	292.19	27.37

Quanto ao número de pacientes admitidos diariamente de acordo com o mês do ano, cujo sumário pode ser visto na Tabela 4.7, os valores mais altos registam-se nos meses de julho (média de  $362.15 \pm 46.73$ ) e agosto (média de  $367.16 \pm 49.28$ ), seguidos dos outros meses de verão, o que vem de encontro ao que se pode observar na Tabela 4.8, donde se pode concluir uma afluência de em média 359.96 pacientes ( $\pm 46.85$ ), diariamente, durante a estação mais quente.

Tabela 4.7: Tabela Sumária do Número de Admissões, por dia, em cada mês

<b>Mês</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Janeiro	226.00	316.75	351.00	382.25	489.00	347.41	46.21
Fevereiro	266.00	327.00	350.00	373.00	423.00	347.42	36.37
Março	248.00	309.75	338.00	370.25	445.00	339.09	40.95
Abril	204.00	306.00	338.00	367.25	429.00	334.27	46.96
Maió	239.00	319.25	346.00	369.00	440.00	343.88	42.50
Junho	262.00	319.50	350.50	379.50	463.00	350.85	43.86
Julho	255.00	334.75	367.50	393.25	481.00	362.15	46.73
Agosto	250.00	333.75	367.50	398.50	492.00	367.16	49.28
Setembro	256.00	319.00	355.50	380.00	465.00	352.55	44.79
Outubro	249.00	327.75	350.50	376.25	457.00	351.77	40.62
Novembro	238.00	310.50	341.50	365.50	446.00	336.73	40.41
Dezembro	237.00	320.50	350.00	378.50	473.00	350.04	50.43

Tabela 4.8: Tabela Sumária do Número de Admissões, por dia, em cada estação do ano

<b>Estação</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Primavera	204.00	312.00	344.00	372.00	463.00	341.08	44.61
Verão	250.00	330.50	361.00	391.00	492.00	359.96	46.85
Outono	238.00	318.00	346.00	372.00	457.00	345.56	41.09
Inverno	226.00	317.00	350.00	377.00	489.00	348.05	45.69

Da análise gráfica das séries temporais seguintes, onde as linhas sólidas com cor mais escura correspondem a um estimador *smooth.spline* (função *smooth.spline* do *R*) não paramétrico - este é equivalente a um estimador de médias móveis pesadas que, visualmente, sumaria os dados - pode observar-se uma grande diferença entre o registado no ano de 2012, principalmente nos seis primeiros meses, e os anos de 2013, 2014 e 2015, o que se deve provavelmente ao processo de transição entre hospitais.

No que ao total de pacientes admitidos diariamente diz respeito, pode perceber-se, observando a Figura 4.1, uma grande afluência nos meses quentes, havendo em geral uma depressão nos meses de março, abril e novembro. O máximo de pacientes atendidos na urgência geral ocorreu em agosto de 2014, com cerca de 500 pacientes, e o mínimo em abril de 2012 com perto de 200 pacientes observados.

Observando ambas as séries temporais, representadas na Figura 4.2, tendo estas em atenção o sexo do paciente, além da já mencionada predominância de pacientes do sexo feminino, pode perceber-se que ambas as séries se comportam de forma bastante semelhante ao longo do tempo, e portanto o sexo do paciente não está diretamente relacionado com o facto deste se dirigir ao serviço de urgência.

De acordo com o local de residência dos pacientes, por leitura das séries temporais representadas na Figura 4.3, é reforçado o facto da maioria dos pacientes admitidos na urgência pertencerem a zonas dos ACES Cávado I e II, tendo ambas as séries temporais distribuições bastante semelhantes, ao longo do tempo. Quanto à série que representa as admissões de indivíduos do ACES Cávado III, pode ver-se que esta é constante ao longo do tempo. Todas as outras séries temporais parecem sobrepor-se, com comportamentos semelhantes.

Tendo em conta a Figura 4.4, onde estão representadas as diferentes séries temporais de acordo com a cor da pulseira atribuída ao paciente aquando da triagem, percebe-se claramente uma predominância de pulseiras verdes e amarelas, o que vem de encontro ao descrito na Tabela 4.4. No entanto, apesar do comportamento da série temporal das pulseiras verdes estar de acordo com o comportamento da série temporal total, no caso da série temporal das pulseiras amarelas pode notar-se uma ligeira diferença. Ao contrário

da primeira, pode ver-se nesta, a existência de mais casos nos meses de inverno. Note-se que as pulseiras são atribuídas de acordo com a gravidade da situação do paciente, o que indica a existência de casos urgentes nos meses de inverno, sendo que o que faz o total de admissões aumentar nos meses quentes são casos pouco urgentes. Isto pode ser corroborado pelo comportamento constante da série temporal das pulseiras laranja ao longo do tempo, isto é, a existência de casos muito urgentes não estará dependente de situações de cariz social, climatérico ou mesmo sazonal.

Da mesma forma, acaba por perceber-se isso mesmo quando observada a Figura 4.5, que diz respeito ao total de pacientes admitidos diariamente no SU geral, de acordo com o destino que tomam após terem alta. A maioria dos casos, e os que seguem a tendência da série temporal correspondente à totalidade de pacientes, são casos que acabam por ter como destino o exterior do hospital, enquanto que a série temporal cujos pacientes seguem rumo ao SI do hospital é bastante mais constante ao longo do tempo, quando comparada com as restantes.

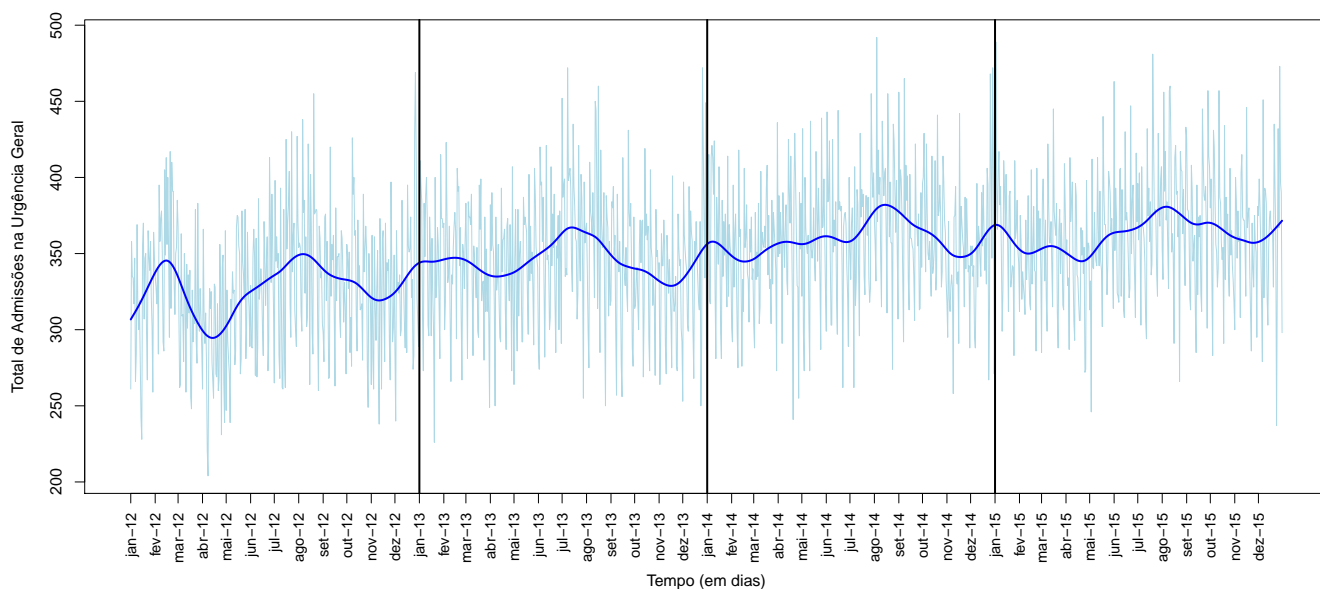


Figura 4.1: Total de Pacientes Admitidos na Urgência Geral, diariamente

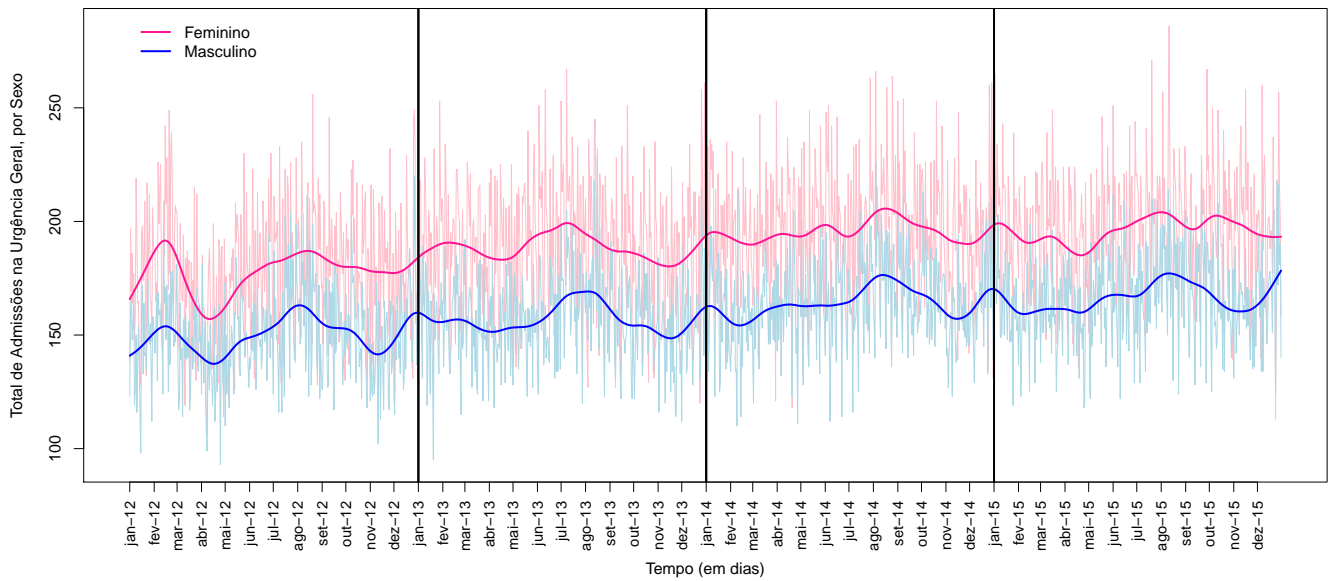


Figura 4.2: Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Sexo

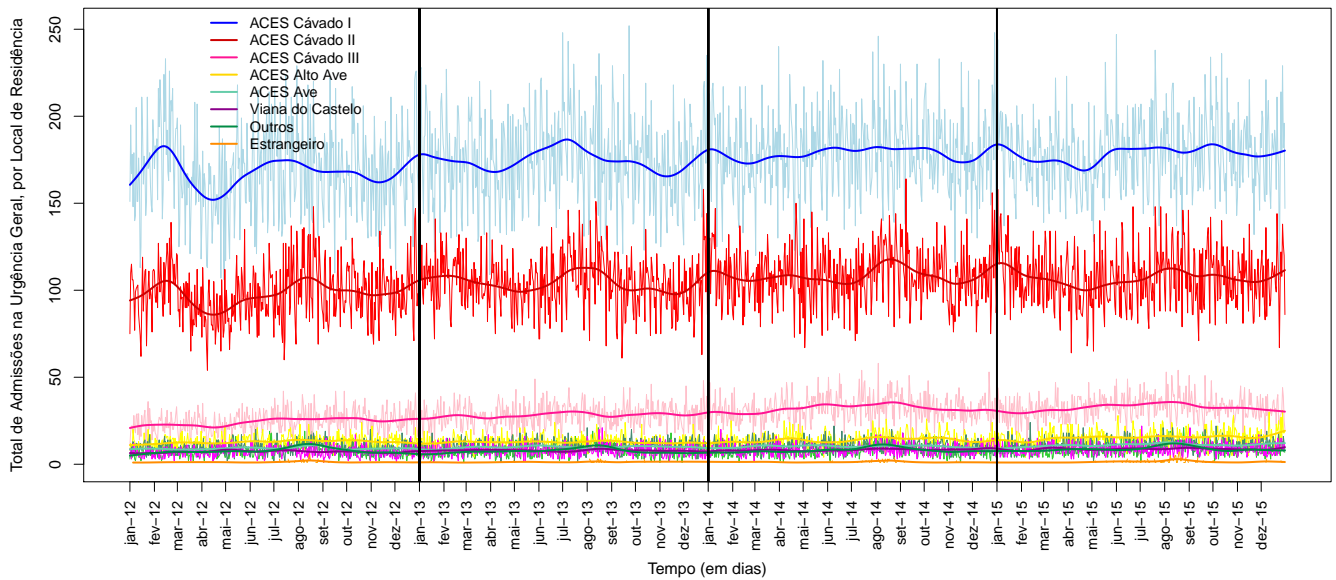


Figura 4.3: Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Local de Residência

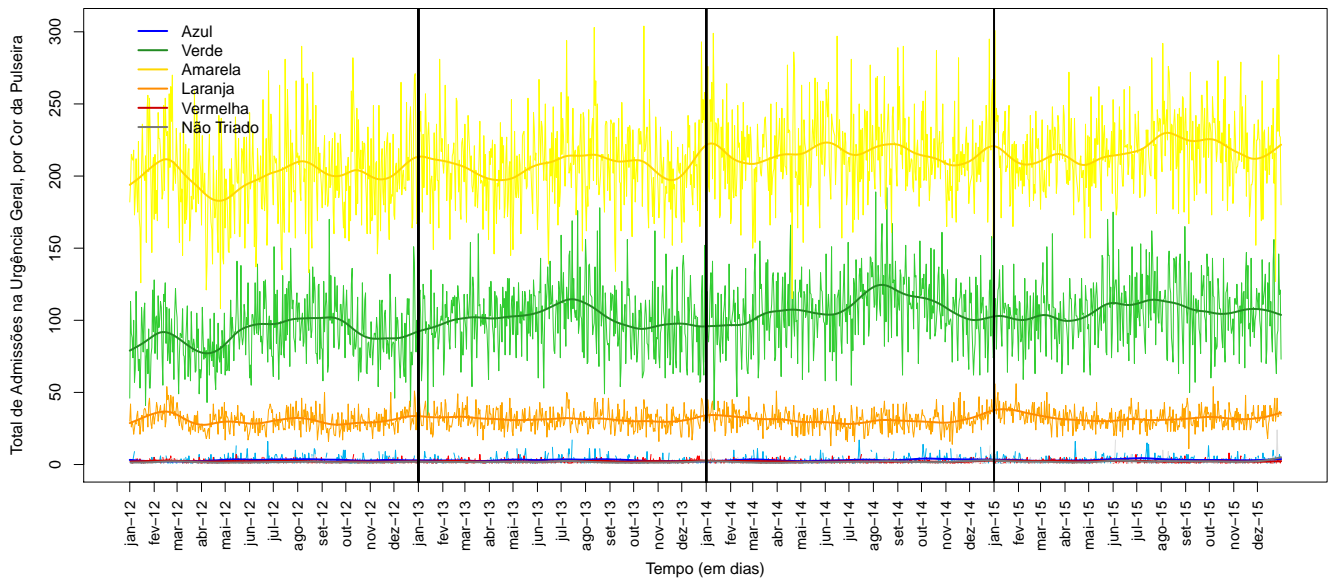


Figura 4.4: Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta a Cor da Pulseira

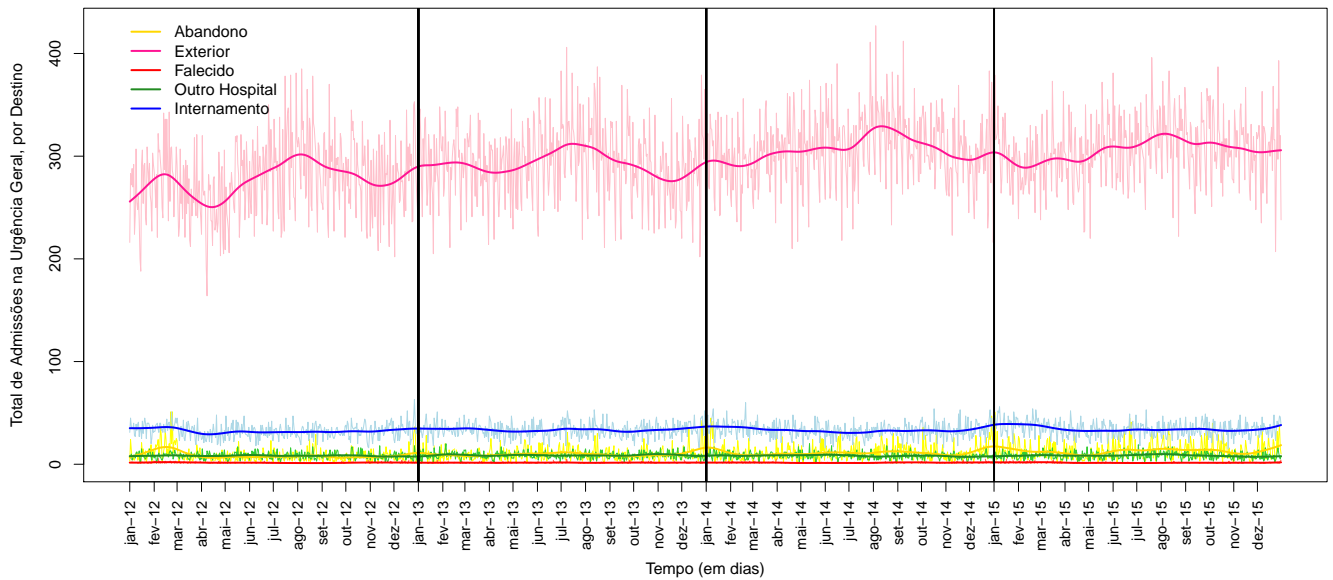


Figura 4.5: Total de Pacientes Admitidos na Urgência Geral, diariamente, tendo em conta o Destino

Tendo em atenção apenas a série temporal do total de pacientes que seguem para o SI, diariamente, representada na Figura 4.6, percebe-se que de facto a maioria ocorre nos meses de inverno, sendo bastante constante nos restantes. Note-se ainda que o máximo de internamentos é cerca de 60 e ocorre em dezembro de 2012, enquanto que o mínimo ocorreu em outubro de 2015, com cerca de 15 internamentos.

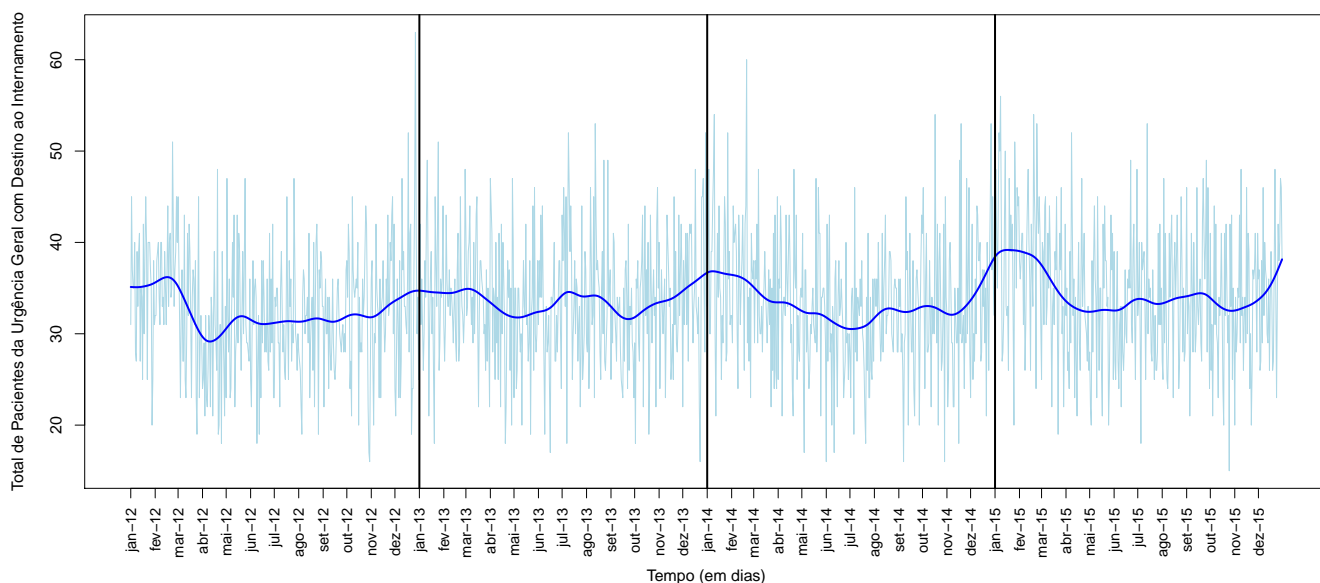


Figura 4.6: Total de Pacientes Admitidos na Urgência Geral, com destino ao Serviço de Internamento

### 4.1.2 Urgência Obstétrica

Sendo todos os pacientes neste serviço do sexo feminino, pode dizer-se que entre 1 de janeiro de 2012 e 31 de dezembro de 2015 foram admitidas na urgência obstétrica do HB cerca de 52657 mulheres.

Tabela 4.9: Tabela Sumária da Variável Quantitativa Idade

Variável	Mín.	1.º Q	Mediana	3.º Q	Máx.	Média	Desvio Padrão
Idade	2.00	27.00	31.00	34.00	92.00	30.70	5.44

Na Tabela 4.10 encontram-se as frequências absolutas e relativas para as variáveis qualitativas, donde se retiraram as conclusões seguintes.

Tabela 4.10: Tabela de Frequências Absolutas e Relativas

Variável	Categoria	Freq. Absoluta	Freq. Relativa * 100%
Local de Residência	ACES Cávado I	27198	51.65%
	ACES Cávado II	12647	24.02%
	ACES Cávado III	9532	18.10%
	ACES Ave	733	1.39%
	ACES Alto Ave	973	1.85%
	Viana do Castelo	598	1.14%
	Outros	950	1.80%
	Estrangeiro	26	0.05%
Cor da Pulseira	Azul	404	0.77%
	Verde	20350	38.65%
	Amarelo	24286	46.12%
	Laranja	7534	14.31%
	Vermelho	6	0.01%
	Não Triado	77	0.15%
Destino	Abandono	252	0.48%
	Exterior	40663	77.22%
	Outro Hospital	13	0.02%
	Falecido	0	0.00%
	Internamento	11729	22.27%

A maioria destas pacientes são de áreas pertencentes aos ACES Cávado I, Cávado II e Cávado III (51.65% , 24.02% e 18.10%, respetivamente), como já seria de esperar, sendo o HB prestador de cuidados obstétricos de 1<sup>a</sup> instância nestes ACES.

Os cuidados de saúde prestados são, em geral, urgentes, pouco urgentes e muito urgentes



(46.12%, 38.65% e 14.31%, respetivamente). A existência de uma maior percentagem de casos muito urgentes deve-se, provavelmente, à chegada de mulheres em trabalho de parto à urgência. Note-se ainda a existência de apenas 6 casos emergentes nos quatro anos em estudo.

Após a alta, aproximadamente 77.22% das pacientes têm como destino o exterior do HB, enquanto que 22.27% segue para o SI. Este aumento de pacientes internadas, quando comparado com a urgência geral, deve-se também ao facto da maioria das pacientes seguirem para dar à luz. Note-se também a não existência de nenhuma paciente falecida neste serviço.

As pacientes admitidas neste serviço têm idades compreendidas ente os 2 e os 92 anos, com uma média de idades de 30.70 anos, com um desvio padrão de 5.44, como descrito na Tabela 4.9.

À semelhança do que foi feito para a urgência geral, encontra-se de seguida uma breve análise descritiva do número de admissões diárias na urgência obstétrica, de acordo com o dia da semana na Tabela 4.11, com o mês na Tabela 4.12, com a estação do ano na Tabela 4.13 e com o facto de ser, ou não, feriado na Tabela 4.14.

O número de pacientes admitidos na urgência obstétrica, é menor nos dias de feriado, à semelhança do que acontece na urgência geral, com uma média de 30.23 pacientes ( $\pm 8.24$ ). Desta vez, no entanto, a diferença não é tão discrepante, já que a média de pacientes admitidos, em cada dia, quando este não é um feriado é de cerca de 36.22 ( $\pm 11.05$ ). No entanto, é ainda provável que se venha a registar uma significância do facto de determinado dia ser, ou não, feriado aquando da modelação do número de pacientes admitidas nesta urgência.

Atendendo ao dia da semana, pode concluir-se que o número de pacientes que se dirigem a este serviço de urgência é maior à segunda-feira, com uma média de cerca de 46 pacientes ( $\pm 9.74$ ), diminuindo ao longo da semana, atingido o seu mínimo ao domingo com uma média de 23.16 pacientes ( $\pm 5.48$ ). É de notar, no entanto, um ligeiro "pico" à sexta-feira (média de 40.71 pacientes com um desvio padrão de 8.83), véspera de fim de semana.

Tabela 4.11: Tabela Sumária do Número de Admissões, por dia, em cada dia da semana

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Segunda-feira	26.00	39.00	45.00	53.00	85.00	46.00	9.74
Terça-feira	18.00	34.00	40.00	47.00	82.00	40.73	9.13
Quarta-feira	15.00	33.00	38.00	43.00	60.00	37.61	7.61
Quinta-feira	17.00	32.00	37.00	43.00	62.00	37.91	8.00
Sexta-feira	20.00	35.00	40.00	46.00	79.00	40.71	8.83
Sábado	11.00	22.00	26.00	29.25	43.00	26.15	5.75
Domingo	12.00	19.00	23.00	27.00	41.00	23.16	5.48

Tabela 4.12: Tabela Sumária do Número de Admissões, por dia, em cada mês

<b>Mês</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Janeiro	15.00	28.00	35.00	43.00	63.00	35.38	10.13
Fevereiro	12.00	27.00	33.00	38.00	58.00	32.63	9.08
Março	13.00	25.75	32.50	39.00	65.00	32.52	9.39
Abril	15.00	27.00	34.00	42.00	57.00	34.65	10.22
Maio	11.00	30.75	38.00	43.25	62.00	36.58	9.79
Junho	18.00	26.00	36.00	44.00	64.00	36.37	11.12
Julho	12.00	28.00	38.00	45.25	66.00	37.81	11.93
Agosto	13.00	29.75	40.50	50.25	62.00	40.20	12.25
Setembro	16.00	30.00	38.50	47.00	85.00	39.18	12.29
Outubro	14.00	29.00	34.50	41.25	64.00	34.91	10.02
Novembro	12.00	26.00	33.50	40.00	65.00	33.67	10.06
Dezembro	17.00	29.00	37.00	45.25	82.00	38.29	12.54

Tendo em atenção o mês do ano, os valores mais altos registam-se nos meses de agosto (média de  $40.20 \pm 12.15$ ) e setembro (média de  $39.18 \pm 12.29$ ), seguidos dos outros meses quentes, o que mais uma vez vem de encontro ao que se concluiu quanto ao serviço de

urgência geral, ou seja, a afluência é maior no verão, com uma média diária de 38.76 pacientes ( $\pm 12.21$ ). Dos meses frios aquele em que se nota um maior número de chegadas à urgência é o mês de dezembro (média de  $38.29 \pm 12.54$ ).

Tabela 4.13: Tabela Sumária do Número de Admissões, por dia, em cada estação do ano

<b>Estação</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Primavera	11.00	27.75	36.00	43.00	64.00	35.73	10.18
Verão	12.00	29.00	39.00	48.00	85.00	38.76	12.21
Outono	12.00	28.00	35.00	42.00	72.00	35.20	10.55
Inverno	12.00	27.00	34.00	41.00	82.00	34.44	10.61

Tabela 4.14: Tabela Sumária do Número de Admissões, diárias, de acordo com categorização de feriado

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Feriado	17.00	24.75	29.00	34.00	59.00	30.23	8.94
Outros	11.00	28.00	36.00	44.00	85.00	36.22	11.05

De seguida encontram-se representações gráficas de algumas séries temporais, com características de interesse, onde as linhas correspondem a um estimador *smooth.spline* não paramétrico (função *smooth.spline* do *R*).

Tendo em atenção a série temporal da totalidade de pacientes admitidas na urgência obstétrica, representada na Figura 4.7, pode ver-se claramente uma maior afluência nos meses mais quentes, existindo sempre uma menor afluência nos meses de fevereiro, março e novembro. O máximo de pacientes atendidas ocorreu em setembro de 2014, com cerca de 85 pacientes a visitarem o serviço de urgência obstétrica.

Na Figura 4.8 encontra-se a representação gráfica do número de pacientes admitidas de acordo com o local de residência. Daqui conclui-se que a maioria das pacientes advém da zona do ACES Cávado I, seguindo-se as pacientes admitidas oriundas dos ACES Cávado II e III, tendo as segundas zonas distribuições semelhantes e mais constantes ao longo do

tempo, do que a primeira, sendo que, relativamente à série que representa o número de admissões de pacientes da zona do ACES Cávado II, se pode ver uma maior afluência nos meses de agosto, setembro e outubro.

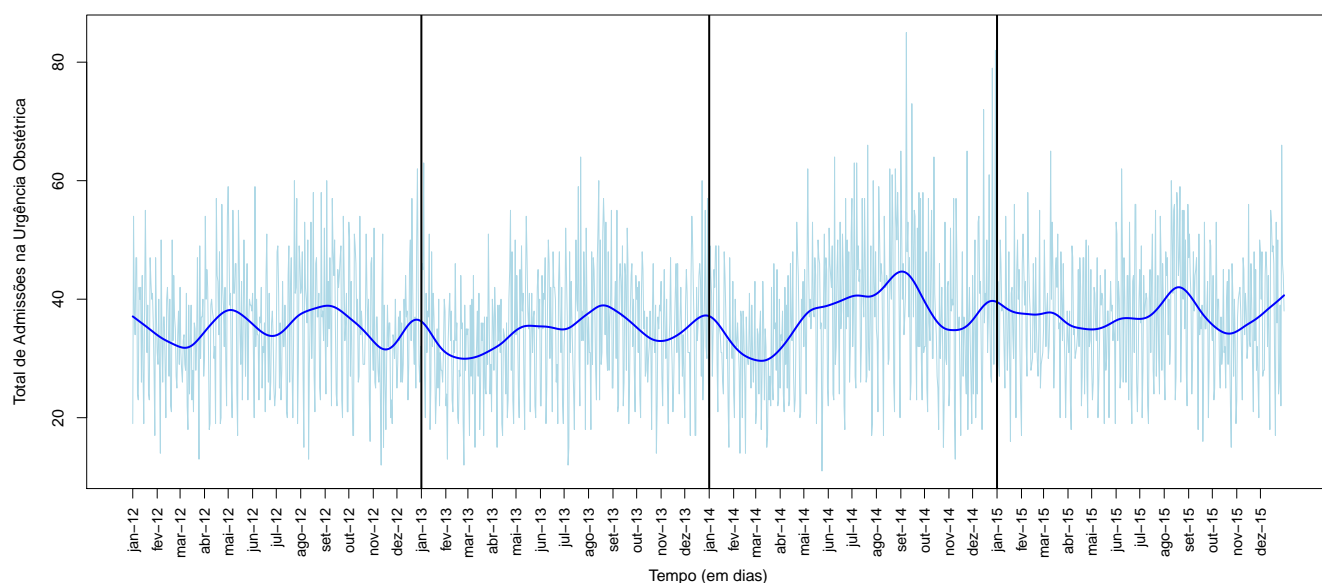


Figura 4.7: Total de Pacientes Admitidos na Urgência Obstétrica, diariamente

Atendendo à cor da pulseira, conclui-se, após observação da Figura 4.9, a existência de mais admissões cujas pulseiras atribuídas são da cor verde, comparativamente à cor amarela, nos meses de agosto e setembro dos anos de 2013, 2014 e 2015. O comportamento da série temporal das pulseiras amarelas é mais constante ao longo do tempo, do que a da série das pulseiras verdes. De notar ainda a existência de vários casos cujas pulseiras atribuídas são azuis nos anos de 2013 e 2014 e também o facto do comportamento da séries das pulseiras laranjas se apresentar constante ao longo do tempo.

Quando tido em conta o destino da paciente após a alta, as conclusões, analisado o gráfico da Figura 4.10, são as mesmas que para a urgência geral. Se por um lado, a série temporal correspondente às pacientes que seguem para o exterior do SU se assemelha à série temporal da totalidade de pacientes admitidas na urgência obstétrica, por outro, a série correspondente ao número de pacientes internadas parece constante ao longo do tempo.

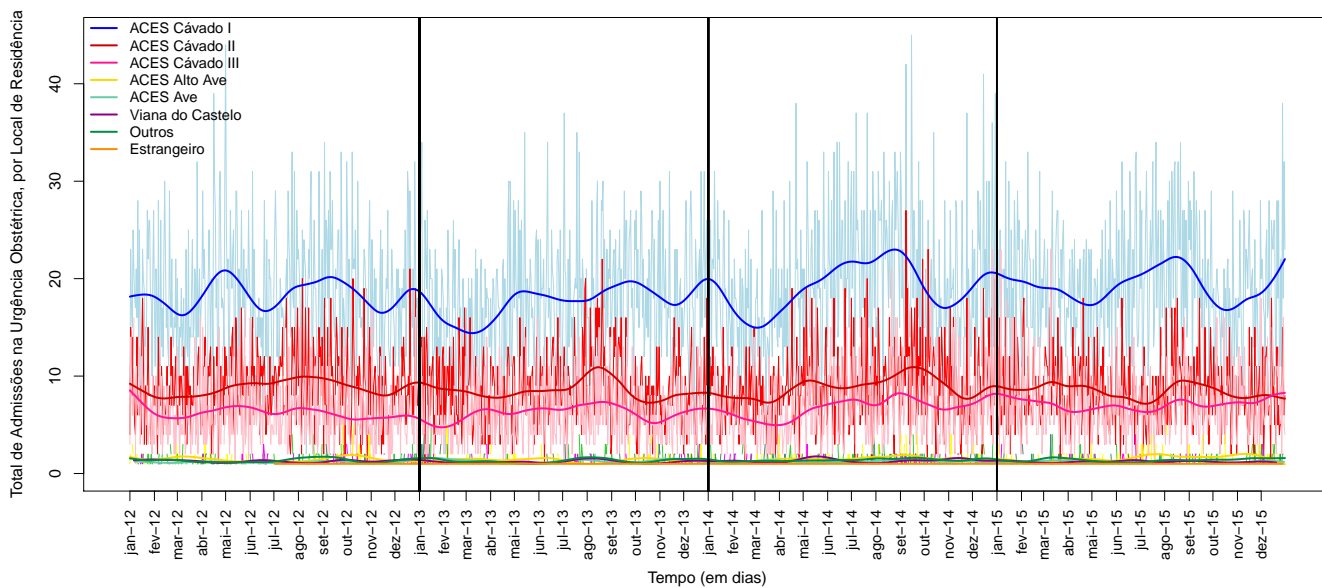


Figura 4.8: Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o Local de Residência

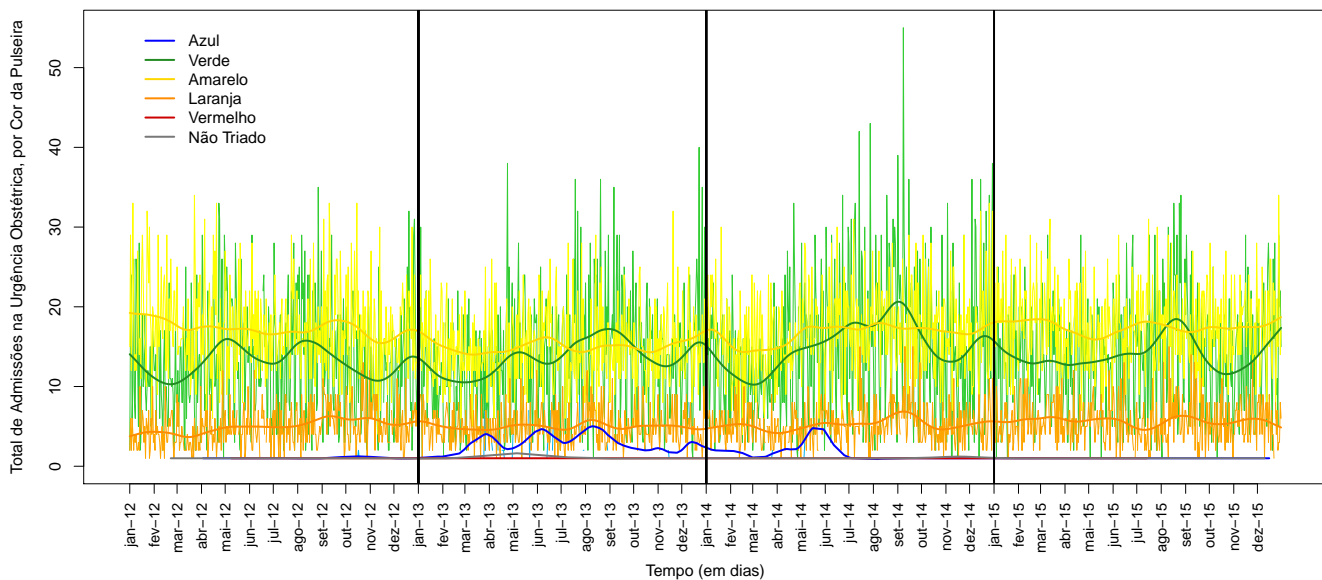


Figura 4.9: Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o a Cor da Pulseira

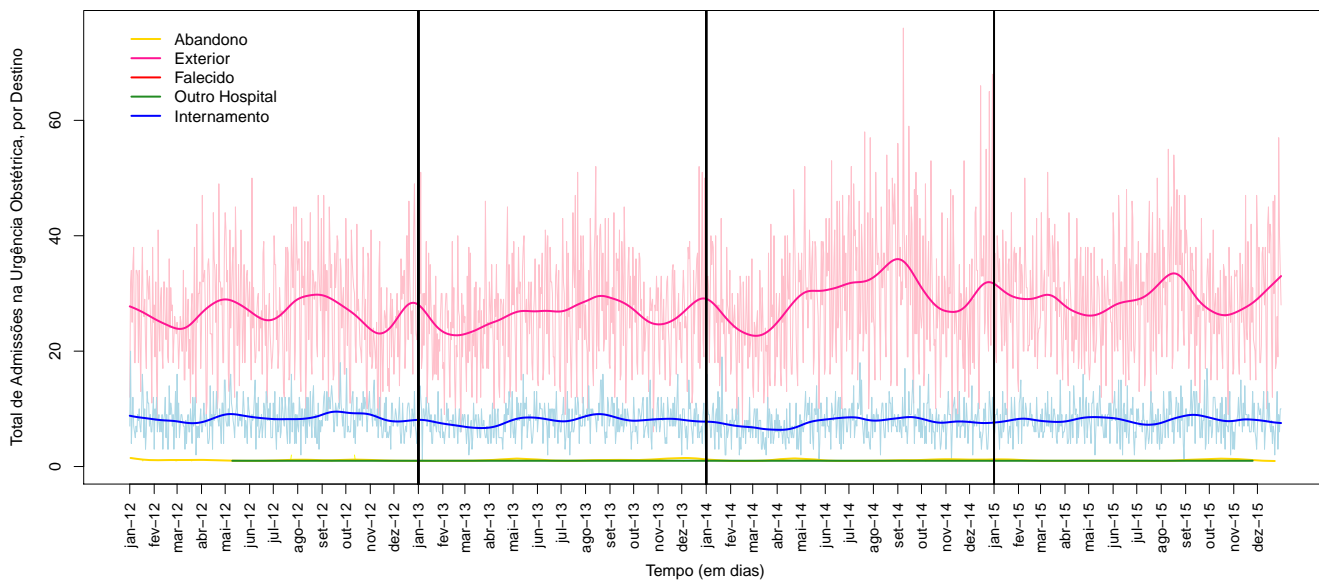


Figura 4.10: Total de Pacientes Admitidos na Urgência Obstétrica, diariamente, tendo em conta o Destino

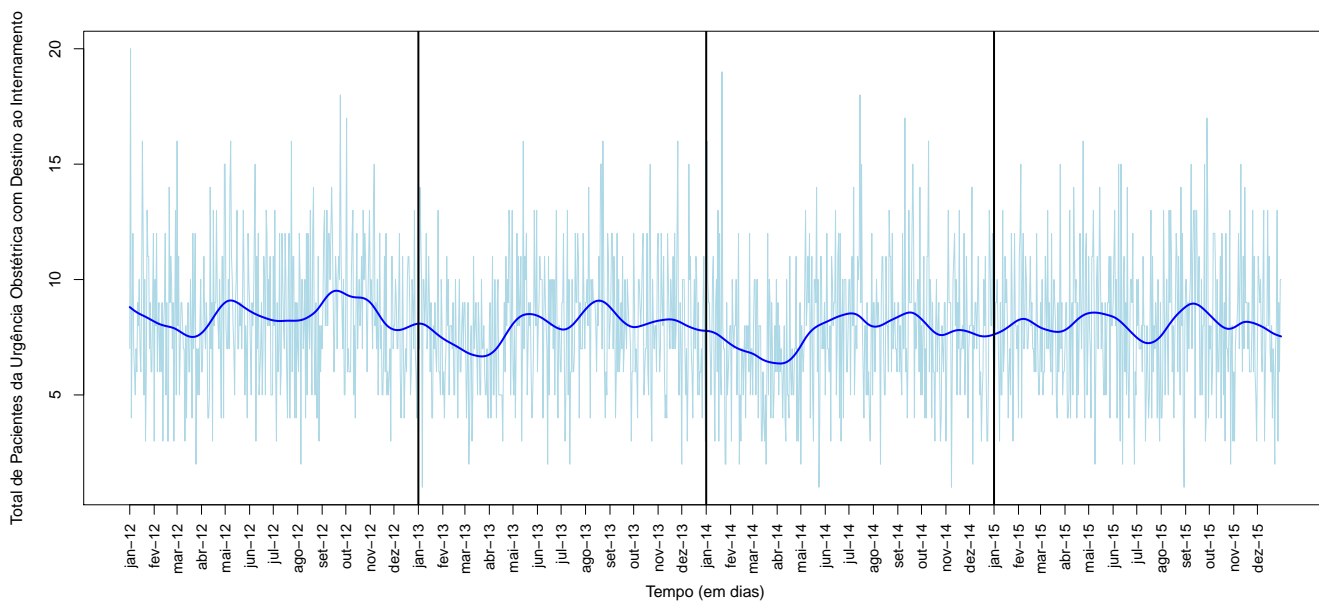


Figura 4.11: Total de Pacientes Admitidos na Urgência Obstétrica, com destino ao Serviço de Internamento

Tendo em atenção apenas a série temporal do total de pacientes internadas, diariamente, representada graficamente na Figura 4.11, observa-se um número mais baixo de internamentos nos meses de fevereiro, março, abril e maio de 2012, 2013 e 2014. Enquanto isso, nos meses de agosto e setembro parece ocorrer um significativo aumento de internamentos. Note-se ainda que o máximo de internamentos é de cerca de 20 e ocorre em janeiro de 2012, enquanto que o mínimo ocorre em maio de 2014 e agosto de 2015, com apenas 1 internamento em cada um desses dias.

### 4.1.3 Urgência Pediátrica

Entre 1 de janeiro de 2012 e 31 de dezembro de 2015 foram admitidas na urgência pediátrica do HB cerca de 187256 crianças, com idades compreendidas entre os 0 e os 18 anos, com uma média de idades de 6.41 ( $\pm$  5.56 anos). Estes resultados são visíveis na Tabela 4.15.

Os pacientes são maioritariamente do sexo masculino, cerca de 53.84%, *versus* os cerca de 46.16% de afluência do sexo feminino. Esta diferença deve-se, provavelmente, ao facto de nascerem mais crianças do sexo masculino do que do sexo feminino.

A maioria dos pacientes advêm de residências na zona dos ACES Cávado I e Cávado II (57.36% e 31.77%, respetivamente), à semelhança do que acontece na urgência geral.

Os cuidados de saúde prestados são, em geral, urgentes e pouco urgentes (46.33% e 41.26%, respetivamente).

Após a alta, a grande maioria dos pacientes (aproximadamente 94.08%) têm como destino o exterior do HB, sendo que somente cerca de 2.93% segue para o SI, percentagem muito semelhante à do abandono do SU. Atente-se ainda para a existência de 6 falecidos durante os 4 anos em estudo.

As conclusões acima foram retiradas tendo em atenção as frequências relativas e absolutas presentes na Tabela 4.16.

Tabela 4.15: Tabela Sumária da Variável Quantitativa Idade

Variável	Mín.	1.º Q	Mediana	3.º Q	Máx.	Média	Desvio Padrão
Idade	0.00	2.00	5.00	11.00	18.00	6.41	5.56

Tabela 4.16: Tabela de Frequências Absolutas e Relativas

Variável	Categoria	Freq. Absoluta	Freq. Relativa * 100%
Sexo	F	86437	46.16%
	M	100819	53.84%
Local de Residência	ACES Cávado I	107417	57.36%
	ACES Cávado II	59496	31.77%
	ACES Cávado III	9970	5.32%
	ACES Ave	2149	1.15%
	ACES Alto Ave	3089	1.65%
	Viana do Castelo	2173	1.16%
	Outros	2767	1.48%
	Estrangeiro	195	0.10%
Cor da Pulseira	Azul	1523	0.81%
	Verde	77265	41.26%
	Amarelo	86760	46.33%
	Laranja	21051	11.24%
	Vermelho	238	0.13%
	Não Triado	419	0.22%
Destino	Abandono	4686	2.50%
	Exterior	176166	94.08%
	Outro Hospital	908	0.49%
	Falecido	6	0.00%
	Internamento	5490	2.93%



Como se pode ver na Tabela 4.17, o número de pacientes admitidos no serviço pediátrico é maior à segunda-feira (média de 144.11 indivíduos com um desvio padrão de 30.15), diminuindo ao longo da semana, atingindo o seu mínimo ao sábado, com uma média de 115.33 pacientes ( $\pm 19.35$ ) e não ao domingo como em ambos os outros serviços de urgência. Ao domingo existe uma ligeira subida na afluência à urgência, em relação ao sábado, com uma média de 121.87 pacientes e desvio padrão de 23.28.

Atentando ao mês do ano, observada a Tabela 4.18 sabe-se que os valores diários mais altos registam-se nos meses de janeiro e fevereiro (média de  $153.03 \pm 26.20$  e  $144.28 \pm 27.52$ , respetivamente) e os mais baixos no mês de agosto (média de  $98.65$  pacientes  $\pm 14.12$ ), o que vem de encontro ao que é visível na Tabela 4.19, em que os dias onde existe maior afluência à urgência se registam na estação invernal (com uma média de 143.74 e um desvio padrão de 27.28) e menor afluência na estação de verão (média de  $104.76 \pm 15.86$ ).

Os resultados da Tabela 4.20 permitem concluir que, o número de pacientes admitidos na urgência pediátrica é também, à semelhança do registado para ambas os outros serviços de urgência, menor em dias de feriado, com uma média, diária, de 116 pacientes ( $\pm 21.87$ ), do que nos restantes dias, onde se verifica uma média de admissões diárias de 128.5 ( $\pm 26.27$ ). Assim sendo, é provável que o facto de determinado dia ser, ou não, feriado se venha a registar significativa aquando da modelação do número de pacientes admitidas nesta urgência.

Tabela 4.17: Tabela Sumária do Número de Admissões, por dia, em cada dia da semana

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Segunda-feira	81.00	122.00	146.00	163.00	252.00	144.11	30.15
Terça-feira	77.00	112.00	130.00	149.00	218.00	131.15	26.58
Quarta-feira	65.00	107.00	128.00	147.00	204.00	128.88	27.21
Quinta-feira	78.00	111.00	131.00	146.00	195.00	129.44	24.18
Sexta-feira	76.00	111.75	126.00	144.00	200.00	126.34	22.44
Sábado	64.00	101.75	115.00	126.00	181.00	115.33	19.35
Domingo	74.00	107.00	120.00	136.00	199.00	121.87	23.28

Tabela 4.18: Tabela Sumária do Número de Admissões, por dia, em cada mês

<b>Mês</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Janeiro	92.00	136.75	152.00	172.00	227.00	153.03	26.20
Fevereiro	88.00	126.00	139.00	159.00	252.00	144.28	27.52
Março	79.00	109.00	129.00	142.25	192.00	127.85	24.94
Abril	69.00	102.75	119.00	131.25	170.00	118.34	21.43
Maiο	78.00	119.00	131.00	145.00	166.00	130.62	17.31
Junho	86.00	112.00	122.50	134.00	181.00	124.30	18.19
Julho	76.00	102.00	110.50	120.25	148.00	112.02	14.13
Agosto	64.00	88.75	97.50	107.00	136.00	98.65	14.12
Setembro	65.00	90.00	102.00	123.50	180.00	109.23	23.63
Outubro	99.00	126.00	140.00	151.00	180.00	139.21	17.32
Novembro	95.00	123.75	140.00	153.00	199.00	139.88	21.70
Dezembro	80.00	124.00	141.00	159.00	193.00	141.37	22.62

Tabela 4.19: Tabela Sumária do Número de Admissões por dia em cada estação do ano

<b>Estação</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Primavera	69.00	111.00	124.50	140.00	188.00	124.75	20.80
Verão	64.00	94.00	103.50	115.00	149.00	104.76	15.86
Outono	95.00	126.00	139.50	153.00	199.00	139.85	19.82
Inverno	80.00	124.00	141.00	160.00	252.00	143.74	27.28

Tabela 4.20: Tabela Sumária do Número de Admissões, por dia, de acordo com categorização de feriado

<b>Dia</b>	<b>Mín.</b>	<b>1.º Q</b>	<b>Mediana</b>	<b>3.º Q</b>	<b>Máx.</b>	<b>Média</b>	<b>Desvio Padrão</b>
Feriado	78.00	99.75	116.00	130.20	160.00	116.00	21.87
Outros	64.00	110.00	127.00	146.00	252.00	128.5	26.27

Da análise gráfica das séries temporais, onde as linhas correspondem a um estimador *smooth.spline* não paramétrico (função *smooth.spline* do *R*), pode observar-se, à semelhança do que foi visto no caso da urgência geral, uma grande diferença entre o ano de 2012 (principalmente os primeiros 6 meses) e os restantes, que fica a dever-se à recente inauguração das instalações.

Respeitante ao total de pacientes admitidos diariamente observa-se, na Figura 4.12 um pico bastante significativo no início de 2012, onde se atinge o número máximo de crianças admitidas nos 4 anos em estudo (cerca de 250 pacientes no mês de fevereiro). Existe em geral uma depressão nos meses de primavera e verão, sendo neste último atingido o mínimo de pacientes atendidos neste serviço (cerca de 65 pacientes).

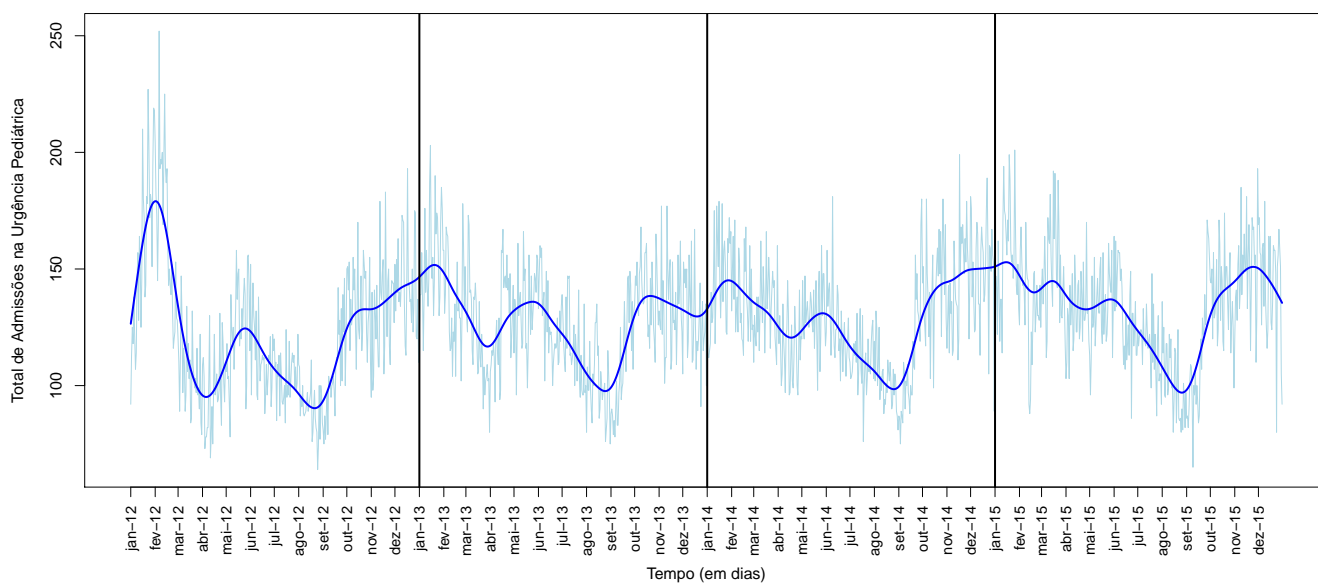


Figura 4.12: Total de Pacientes Admitidos na Urgência Pediátrica, diariamente

Tendo em conta o sexo do paciente tem-se o gráfico da Figura 4.13. Aqui, além da já mencionada predominância de pacientes do sexo masculino, pode perceber-se que ambas as séries se comportam de forma bastante semelhante ao longo do tempo e, portanto, o sexo do paciente não está diretamente relacionado com o facto deste se dirigir à urgência, à semelhança do observado no SU geral.

De acordo com o local de residência, é reforçado, a partir da leitura do gráfico da Figura 4.14, o facto da maioria dos pacientes admitidos na urgência pertencerem a zonas dos ACES Cávado I e II, tendo ambas as séries temporais distribuições bastante semelhantes, ao longo do tempo, apesar da segunda se comportar de maneira um pouco mais suave. Quanto à série que representa as admissões de indivíduos do ACES Cávado III, pode ver-se que esta é constante ao longo do tempo. Todas as outras séries temporais parecem sobrepôr-se, com comportamentos semelhantes.

Na Figura 4.15 pode ver-se o gráfico das séries temporais de cada categoria da variável Cor da Pulseira, onde podem notar-se comportamentos semelhantes para ambas as séries que correspondem a pulseiras verdes e amarelas. Quanto à série temporal correspondente à cor laranja já não existe uma tão constante distribuição ao longo do tempo, como acontecia para os outros dois serviços de urgência. Esta última possibilita concluir que os casos muito urgentes têm lugar, em geral, nos meses frios, devido, possivelmente, há existência de mais problemas do foro respiratório, diminuindo na altura de primavera e verão.

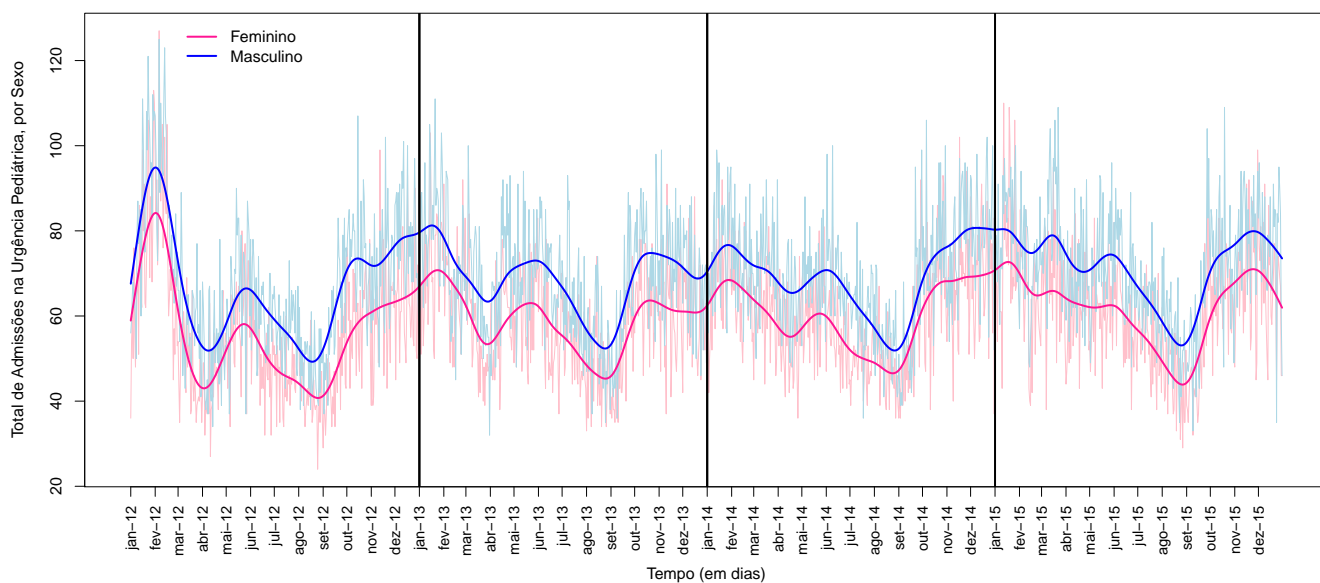


Figura 4.13: Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Sexo

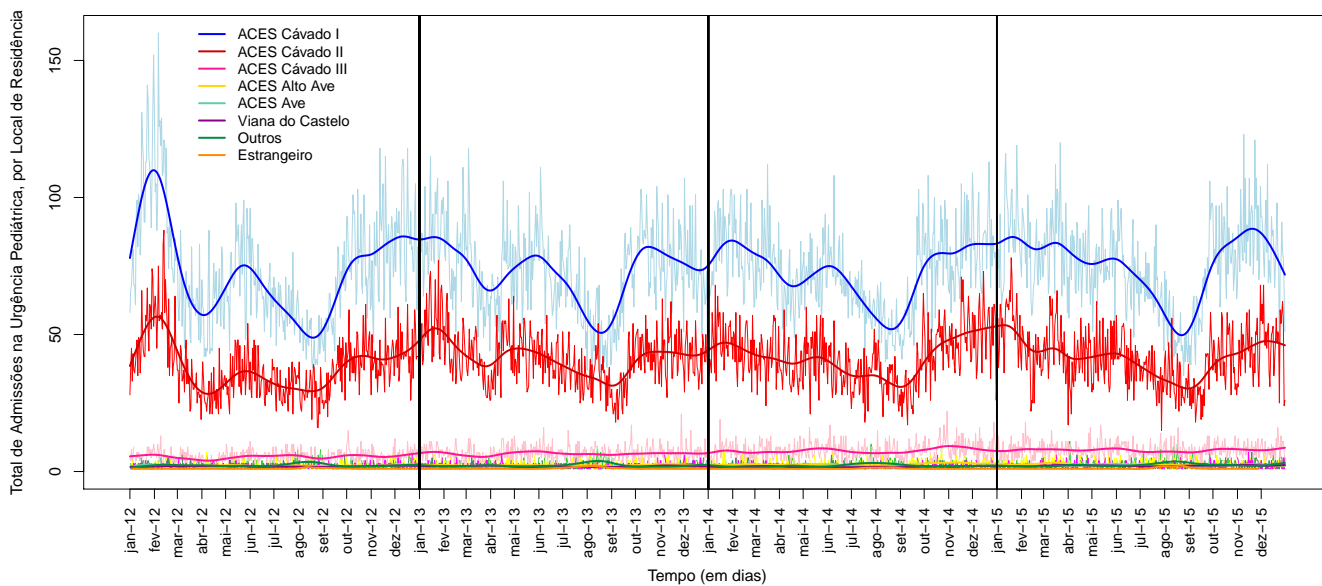


Figura 4.14: Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Local de Residência

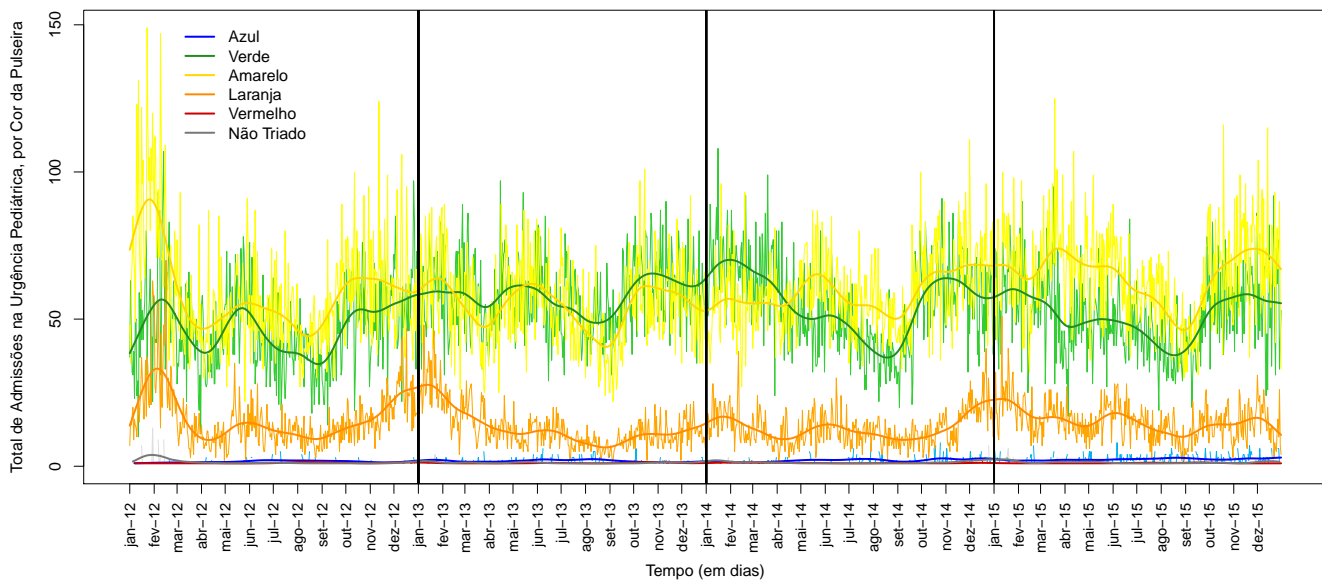


Figura 4.15: Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta a Cor da Pulseira

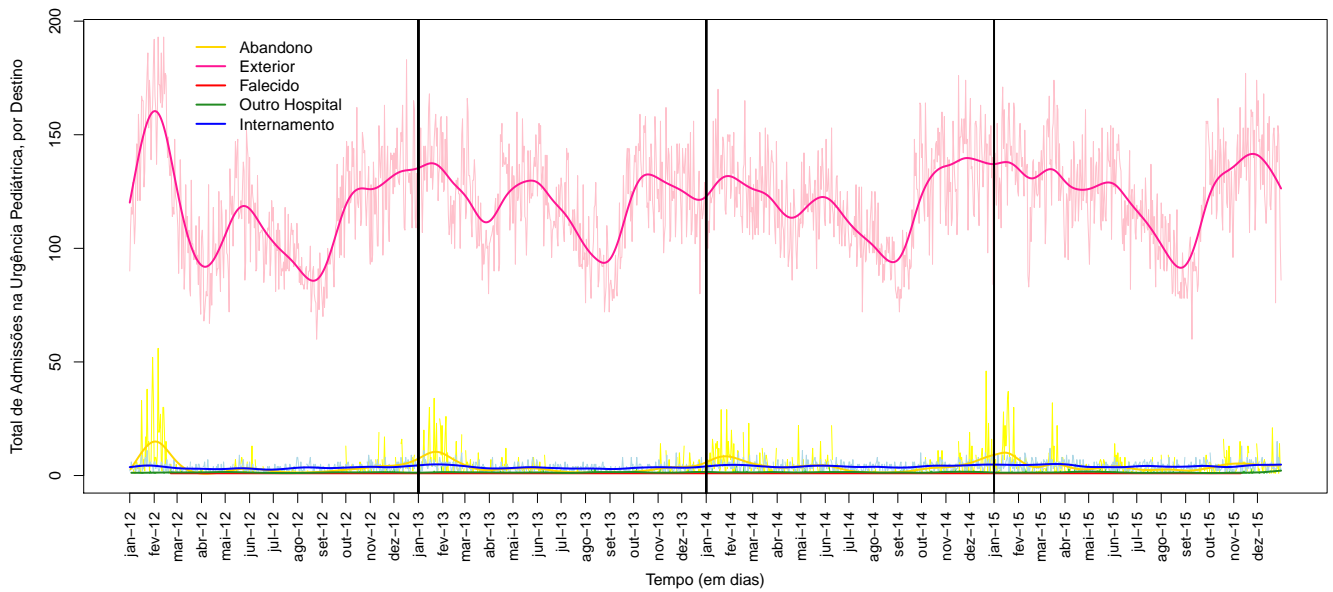


Figura 4.16: Total de Pacientes Admitidos na Urgência Pediátrica, diariamente, tendo em conta o Destino

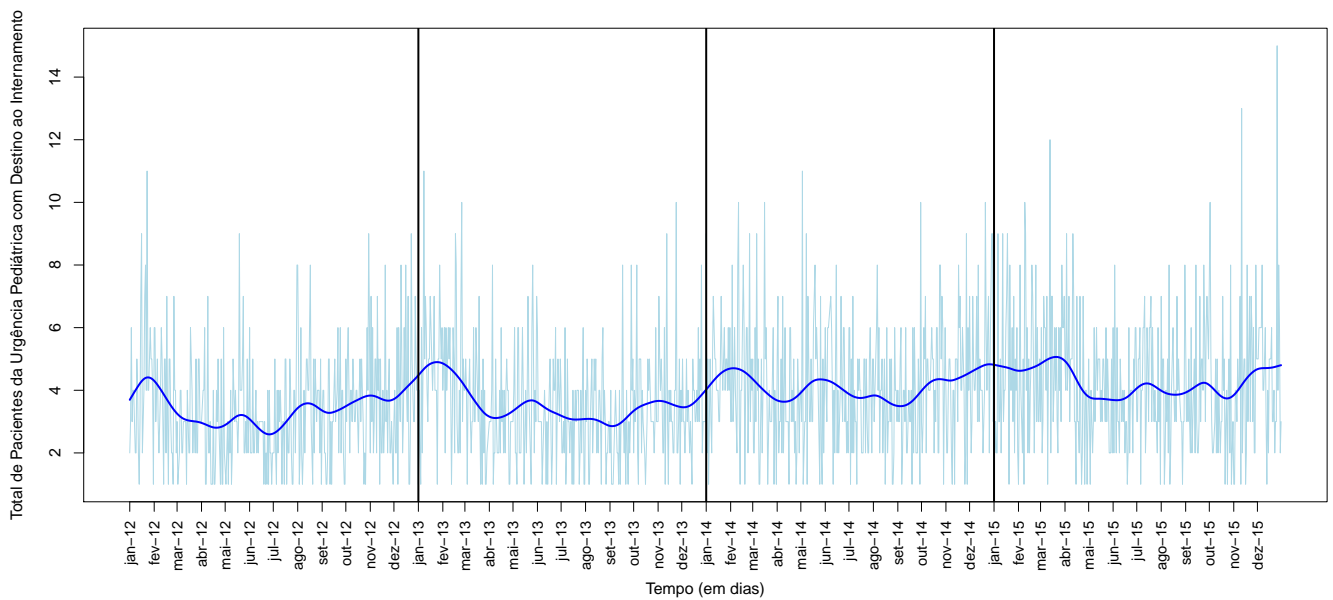


Figura 4.17: Total de Pacientes Admitidos na Urgência Pediátrica, com destino ao Serviço de Internamento

Como já visto, apenas uma minoria das crianças são internadas, face à quantidade cuja alta tem como destino o exterior. Pode ver-se na Figura 4.16 que, ao contrário da irregularidade da série temporal correspondente às crianças cujo destino é o exterior, a série daquelas cujo destino é o SI é aparentemente constante ao longo do tempo.

Observando, no entanto, somente a distribuição do número de crianças admitidas, cujo destino é o SI, ao longo do tempo nota-se que o número de internamentos é maior nos meses de inverno, assim como o número de admissões, o que está correlacionado. O número de internamentos neste serviço varia entre os 0 e os cerca de 15, estes últimos registados em dezembro de 2015. Estes resultados foram concluídos da análise do gráfico da Figura 4.17.

## 4.2 Dados Meteorológicos

A base de dados "Meteo" é composta por 2922 observações e 6 variáveis, 3 quantitativas, 1 qualitativa e 1 data. As observações foram registadas entre 1 de janeiro de 2012 e 31 de dezembro de 2015, em duas estações diferentes, localizadas em Bragança e Coimbra, sendo estas as estações mais próxima da zona de Braga com registo das variáveis de interesse.

Na Tabela 4.21 encontra-se a descrição desta base de dados.

Tabela 4.21: Descrição da Base de Dados Meteo

Variável	Descrição da Variável	Tipo de Variável
Estação	Código da estação onde foram efetuadas as medições	Qualitativa (2 níveis)
Nome da Estação	Estação onde foram efetuadas as medições	Qualitativa (2 níveis)
Data	Data da medição	Qualitativa (datas)
TMAX	Temperatura máxima registada, diariamente	Quantitativa (°C)
TMIN	Temperatura mínima registada, diariamente	Quantitativa (°C)
PRECIP	Precipitação observada, diariamente	Quantitativa (mm)

Calculou-se para cada uma das variáveis, o valor médio diário dos valores registados em ambas as estações de Coimbra e Bragança, obtendo-se os resultados da Tabela 4.22, dos quais se pode concluir

- uma temperatura máxima, diária, entre os 6.15<sup>o</sup>C e os 38.40<sup>o</sup>C, com uma média de 19.47<sup>o</sup>C ( $\pm 7.21$ ) e 234 observações em falta;
- uma temperatura mínima entre os -8.80<sup>o</sup>C e os 24.00<sup>o</sup>C, com uma média de 9.36<sup>o</sup>C ( $\pm 5.07$ ) e 256 observações em falta;
- uma precipitação entre os 0mm e os 65.50mm, com uma média de 7.61mm ( $\pm 8.18$ ) e 1072 observações em falta.

Assim sendo, podem já excluir-se os dados relativos à precipitação pela existência de vários valores em falta.

Tabela 4.22: Tabela Sumária do Valor Médio Diário em ambas as estações

Variável	Mín.	1. <sup>o</sup> Q	Mediana	3. <sup>o</sup> Q	Máx.	Média	D.P.	NA's
TMAX	6.15	13.20	18.80	25.20	38.40	19.47	7.21	234
TMIN	-8.80	5.80	10.00	13.25	24.00	9.36	5.07	256
PRECIP	0.00	1.00	4.30	10.90	65.50	7.61	8.18	1072

Tendo em conta o gráfico da Figura 4.18 (a), pode perceber-se que a existência de valores em falta da temperatura máxima dá-se principalmente no final do ano de 2015. A tendência de ambas as séries representadas na Figura 4.18 está de acordo com o esperado, temperaturas mais altas nos meses de verão e mais baixas nos de inverno, sendo intermédias nos restantes. Assim sendo, é possível que estes fatores influenciem o número de admissões na urgência.



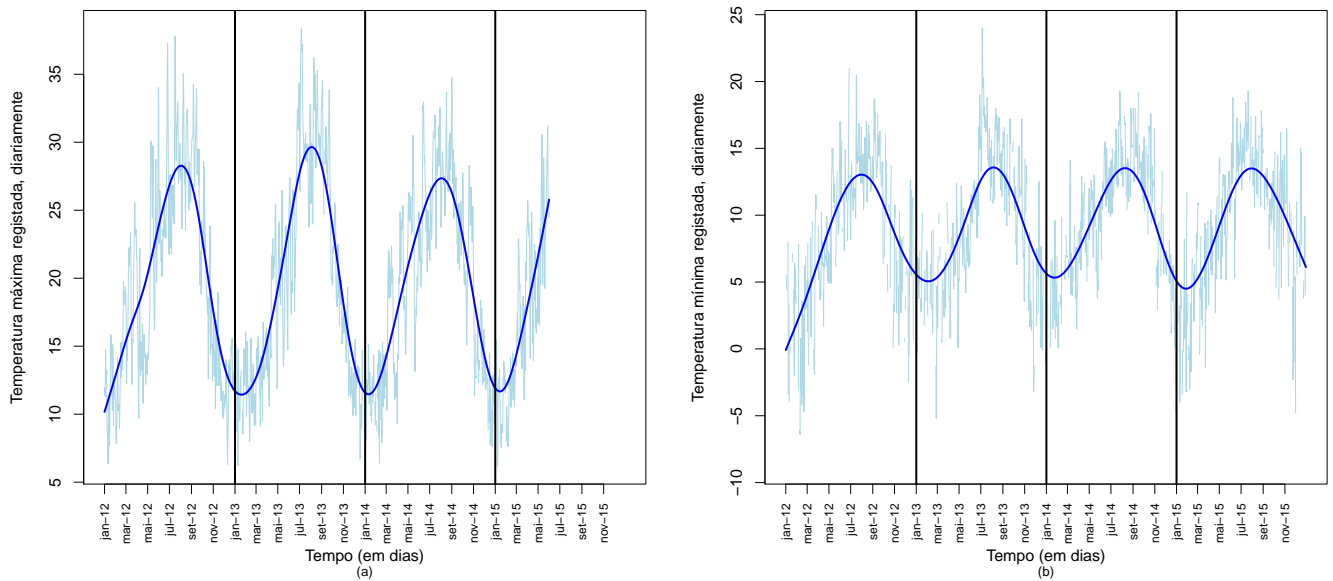


Figura 4.18: Temperaturas máxima (a) e mínima (b) registadas, diariamente

### 4.3 Dados Poluição Atmosférica

A base de dados "Ar" é composta por 78912 observações e 7 variáveis, 4 quantitativas, 1 qualitativa e 2 correspondentes a datas ou horas. As concentrações de cada poluente foram registadas, de hora em hora, entre 1 de janeiro de 2012 e 31 de dezembro de 2014, em três estações diferentes, localizadas em Braga, Guimarães e Viana do Castelo. Note-se que nem todas as estações no país registam os mesmos poluentes, portanto optou-se por estudar as concentrações de poluentes existentes nas zonas de Portugal que possam ter influência sobre o HB. Dado não existirem ainda dados validados para o ano de 2015, assumiram-se dados faltantes para as datas desse ano.

Na Tabela 4.23 encontra-se a descrição da base de dados referida.

Tabela 4.23: Descrição da Base de Dados Ar

Variável	Descrição da Variável	Tipo de Variável
Estação	Estação onde foram efetuadas as medições	Qualitativa (3 níveis)
Data	Data da medição	Qualitativa (datas)
Hora	Hora da medição	Qualitativa (horas)
NO <sub>2</sub>	Concentração de Dióxido de Azoto	Quantitativa ( $\mu\text{g}/\text{m}^3$ )
NO	Concentração de Monóxido de Azoto	Quantitativa ( $\mu\text{g}/\text{m}^3$ )
O <sub>3</sub>	Concentração de Ozono	Quantitativa ( $\mu\text{g}/\text{m}^3$ )
P <sub>10</sub>	Concentração de Partículas < 10 $\mu\text{g}$	Quantitativa ( $\mu\text{g}/\text{m}^3$ )

Em termos de estudo o pretendido seria um valor médio da concentração de cada um dos poluentes, diariamente. Para esse efeito calculou-se, primeiramente, a média diária da concentração de cada poluente, obtendo-se assim um valor para a concentração de cada um dos poluentes, para cada dia, em cada uma das três estações. Calculou-se, por fim, a média, diária, da concentração de cada um dos poluentes nas três estações.

Em termos sumários pode observar-se, a partir da Tabela 4.24, para cada tipo de poluente que

- as concentrações de NO<sub>2</sub> encontram-se entre os  $0.98\mu\text{g}/\text{m}^3$  e os  $64.30\mu\text{g}/\text{m}^3$ , com uma média de  $23.84\mu\text{g}/\text{m}^3$  ( $\pm 10.99$ ). Existem 395 observações em falta;
- as concentrações de NO encontram-se entre os  $1.13\mu\text{g}/\text{m}^3$  e os  $108.90\mu\text{g}/\text{m}^3$ , com uma média de  $16.43\mu\text{g}/\text{m}^3$  ( $\pm 14.73$ ). Existem 395 observações em falta;
- as concentrações de O<sub>3</sub> encontram-se entre os  $4.96\mu\text{g}/\text{m}^3$  e os  $115.10\mu\text{g}/\text{m}^3$ , com uma média de  $58.25\mu\text{g}/\text{m}^3$  ( $\pm 18.66$ ). Existem 436 observações em falta;
- as concentrações de P<sub>10</sub> encontram-se entre os  $3.47\mu\text{g}/\text{m}^3$  e os  $59.61\mu\text{g}/\text{m}^3$ , com uma média de  $15.06\mu\text{g}/\text{m}^3$  ( $\pm 9.23$ ). Existem 395 observações em falta;

donde se conclui, tendo em atenção a classificação do Índice de Qualidade do Ar presente na Tabela 4.25, que o ar da região, nos três anos registados, considerando os valores máximos da concentração de cada poluente (os valores de referência para NO serão semelhantes aos de NO<sub>2</sub>), tem fraca qualidade, já que segundo a APA, a degradação da qualidade do ar estará mais dependente da pior classificação verificada entre os diferentes poluentes e existem valores de P<sub>10</sub> que atingem os 59.61µg/m<sup>3</sup>.

No entanto, considerando os valores médios observados, a qualidade do ar pode ser considerada muito boa.

Tabela 4.24: Tabela Sumária das Variável Quantitativas

Variável	Mín.	1.º Q	Mediana	3.º Q	Máx.	Média	D.P.	NA's
NO <sub>2</sub>	0.98	17.07	23.19	31.04	64.30	23.84	10.99	395
NO	1.13	7.97	11.79	19.93	108.90	16.43	14.73	395
O <sub>3</sub>	4.96	45.54	57.79	70.06	115.10	58.25	18.66	436
P <sub>10</sub>	3.47	8.88	12.02	18.67	59.61	15.06	9.23	395

Tabela 4.25: Classificação do Índice de Qualidade do Ar pela APA

Classificação	NO <sub>2</sub>	O <sub>3</sub>	P <sub>10</sub>
Mau	> 400	> 240	> 120
Fraco	200 - 399	180 - 239	50 - 119
Médio	140 - 199	120 - 179	35 - 49
Bom	100 - 139	60 - 119	20 - 34
Muito Bom	0 - 99	0 - 59	0 - 19

Tendo em conta as representações gráficas de séries temporais, relacionadas com o comportamento das concentrações dos poluentes atmosféricos ao longo do tempo (o ano de 2015 não está representado, visto a não existência de dados para esse ano), pode ver-se

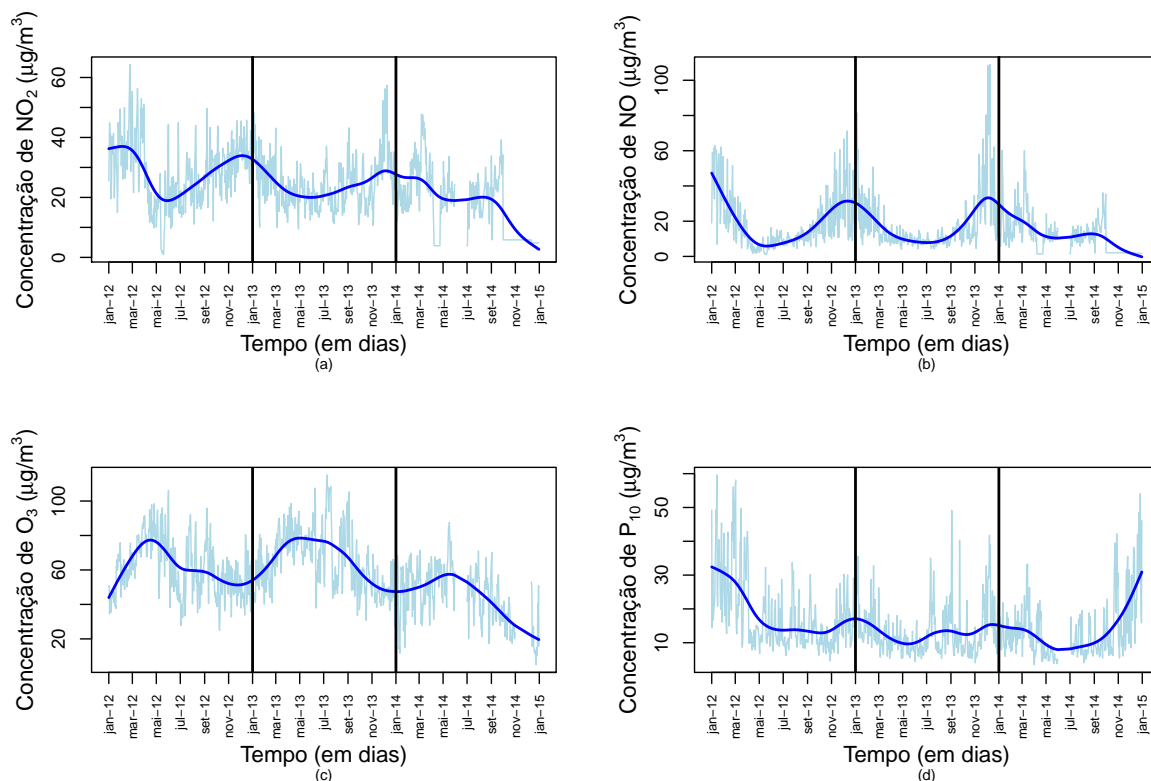


Figura 4.19: Concentrações de  $\text{NO}_2$  (a),  $\text{NO}$  (b),  $\text{O}_3$  (c) e  $\text{P}_{10}$  (d) registadas, diariamente

- tendo em atenção a Figura 4.19 (a) que, a concentração de  $\text{NO}_2$  parece diminuir ao longo do tempo, sendo mais alta nos meses mais frios dos anos de 2012 e 2013, atingindo valores próximos de  $0\mu\text{g}/\text{m}^3$  no fim de 2014. O seu valor máximo foi atingido em fevereiro de 2012, cerca de  $65\mu\text{g}/\text{m}^3$ , o que permite concluir que caso fosse este o único poluente a ter em atenção, a qualidade do ar seria muito boa;
- na Figura 4.19 (b) que, a concentração de  $\text{NO}$  é mais alta nos meses mais frios dos anos de 2012 e 2013, atingindo valores próximos de  $0\mu\text{g}/\text{m}^3$  no fim de 2014, tal como a anterior, o que permite concluir que caso fosse este o único poluente a ter em atenção, a qualidade do ar seria muito boa. No entanto, em dezembro de 2013 foi atingido o valor máximo de cerca de  $100\mu\text{g}/\text{m}^3$ , e, portanto, nesta altura a qualidade do ar seria classificada como boa;

- atendendo à Figura 4.19 (c) que, a concentração de  $O_3$  é mais alta nos meses de primavera, atingindo os seus valores mais baixos (cerca de  $10\mu g/m^3$ ) no fim de 2014. O seu valor máximo foi atingido em julho de 2013, cerca de  $110\mu g/m^3$ , o que permite concluir que caso fosse este o único poluente a ter em atenção, a qualidade do ar seria boa ou muito boa;
- dada a Figura 4.19 (d) que, a concentração de  $P_{10}$  é mais alta no início de 2012 e fim de 2014 (cerca de  $35\mu g/m^3$ ), apresentando em geral valores mais elevados nos meses frios, mas sem grandes oscilações relativamente aos restantes meses. Assim sendo, caso fosse este o único poluente em estudo, a qualidade do ar seria boa ou média. No entanto, em fevereiro de 2012 foi atingido o valor máximo de cerca de  $60\mu g/m^3$ , e, portanto, nesta altura a qualidade do ar seria classificada como fraca.

## 4.4 Dados Gripe

A base de dados "Gripe" é composta por 209 observações e 7 variáveis, 6 quantitativa e 1 data, cuja descrição se encontra na Tabela 4.26. O número total de consultas relacionadas com casos de gripe foram registadas, semanalmente, nas várias zonas de interesse, entre a última semana de 2011 e a última de 2015.

Tendo em atenção a Tabela 4.27, que sumaria as variáveis quantitativas, sabe-se que a maioria das consultas relacionadas com casos de gripe ocorreram no ACES Alto Ave e na zona de Viana do Castelo, com medianas de 21.50 e 23.50, respetivamente. Neste caso têm-se em conta os valores das medianas pois, os da média são inflacionados pelos máximos registados. De referir ainda a existência de poucas consultas no ACES Cávado II, com uma mediana de 6.5.

Tabela 4.26: Descrição da Base de Dados Gripe

Variável	Descrição da Variável	Tipo de Variável
Semana	Semana do ano	Qualitativa (datas)
ACES Alto Ave	Número de consultas registadas no ACES Alto Ave, em cada semana	Quantitativa ( $\in \mathbb{N}$ )
ACES Cávado I ( $\in \mathbb{N}$ )	Número de consultas registadas no ACES Cávado I, em cada semana	Quantitativa ( $\in \mathbb{N}$ )
ACES Cávado II	Número de consultas registadas no ACES Cávado II, em cada semana	Quantitativa ( $\in \mathbb{N}$ )
ACES Cávado III	Número de consultas registadas no ACES Cávado III, em cada semana	Quantitativa ( $\in \mathbb{N}$ )
ACES Ave	Número de consultas registadas no ACES Ave, em cada semana	Quantitativa ( $\in \mathbb{N}$ )
Alto Minho	Número de consultas registadas na zona de Viana do Castelo, em cada semana	Quantitativa ( $\in \mathbb{N}$ )

Tabela 4.27: Tabela Sumária das Variáveis Quantitativas

Variável	Mín.	1.º Q	Mediana	3.º Q	Máx.	Média	D.P.	NA's
ACES Alto Ave	1.00	8.00	21.50	51.75	532.00	56.65	92.56	11
ACES Cávado I	1.00	7.00	18.00	38.00	371.00	41.37	65.66	16
ACES Cávado II	1.00	3.00	6.50	15.00	145.00	16.84	26.51	35
ACES Cávado III	1.00	4.00	12.50	30.50	419.00	32.51	57.69	25
ACES Ave	1.00	6.00	15.00	34.00	332.00	30.2	48.67	14
Alto Minho	1.00	11.25	23.50	59.25	641.00	57.38	95.03	11

Dado o interesse em ter apenas uma variável relacionada com o número de casos de gripe registados, fez-se uma soma, semanal, de todos os casos de gripe, isto é, para cada semana obteve-se apenas o número total de casos de gripe no território em estudo. Assim

sendo, e com os agora 10 valores em falta, tem-se uma média de 228 consultas semanais, contra uma mediana de 94, fazendo-se notar claramente o papel do número máximo de consultas, como esperado. Estes resultados podem ser consultados na Tabela 4.28.

Tabela 4.28: Tabela sumária do número total de casos de gripe registados semanalmente

Variável	Mín.	1. <sup>o</sup> Q	Mediana	3. <sup>o</sup> Q	Máx.	Média	D.P.	NA's
NConsultas	2.00	38.00	94.00	218.00	2393.00	228.00	371.82	10

Atendendo então à representação gráfica do número total de casos de gripe, registados semanalmente, observável na Figura 4.20, pode perceber-se uma existência mais acentuada de casos de gripe nas semanas iniciais de cada ano, o que é de esperar, dadas as condições climáticas (temperaturas mais baixas). Note-se ainda o número de casos de gripe no início do ano de 2015: cerca de 2500 casos registados, aproximadamente mais 1000 do que nos anos anteriores.

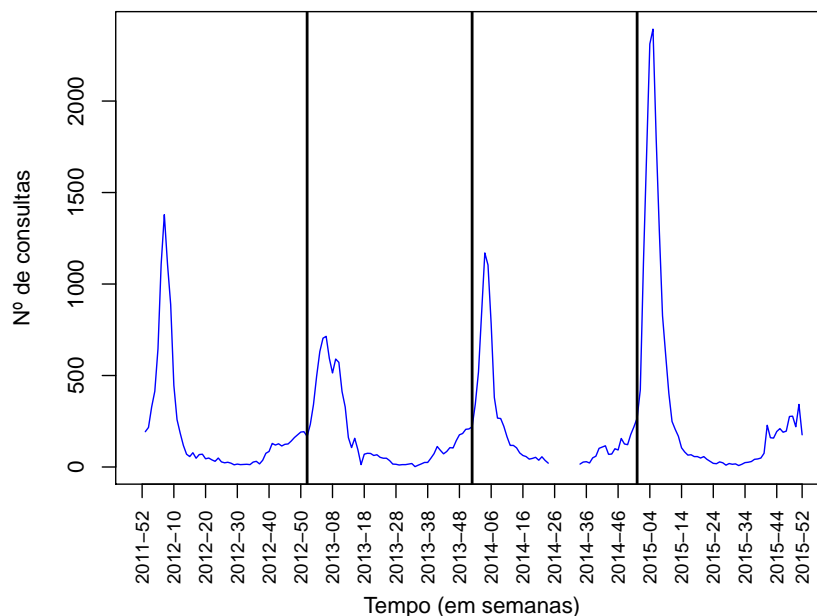


Figura 4.20: Número de consultas devido a gripe registadas, semanalmente

### Modelação das Séries Temporais

---

#### 5.1 Admissões na Urgência

O objetivo aqui passa por modelar o número de pacientes admitidos no SU para que se possa fazer uma previsão futura de maneira a escalar o melhor possível os recursos hospitalares.

Para tal usar-se-ão como variáveis explicativas o dia da semana (categoria de referência é segunda-feira), o mês (categoria de referência é janeiro), a estação (categoria de referência é inverno), o facto de cada dia ser, ou não, feriado (categoria de referência é "outros"), as concentrações atmosféricas de  $\text{NO}_2$ ,  $\text{NO}$ ,  $\text{O}_3$  e  $\text{P}_{10}$ , os registos de temperaturas máximas (TMAX) e mínimas (TMIN) registadas no dia anterior e também o número de consultas devidas a casos de gripe na semana anterior (NConsultas).

Além disso, observados os gráficos das Figuras 4.1, 4.7 e 4.12, parece existir uma certa tendência linear nas séries temporais. Por essa razão será também vista como variável explicativa o tempo (designado por  $t$ ), variável que a cada dia faz corresponder um valor temporal, onde o dia 01/01/2012 é tomado como o dia 1 e o dia 31/12/2015 como o dia 1461.

A modelação será feita para cada um dos serviços de urgência, Geral, Obstétrico e Pediátrico, visto cada um ter características diferentes dos outros. Além disso, a alocação de recursos hospitalares não é a mesma nos três serviços.

Antes de mais é necessário perceber, através do cálculo de VIF, se existe alguma variável no modelo que possa criar problemas de multicolinearidade, isto é, se existe alguma variável



explicativa que correlacionada com as restantes possa por em causa a validade do modelo.

O cálculo de VIF é igual para os três SU's, já que apenas é determinado com o uso do  $R^2$  de cada uma das variáveis explicativas em função das restantes. Assim sendo, como a variável dependente não é tida em conta e as variáveis explicativas são as mesmas para as três urgências, as conclusões retiradas acerca de quais as variáveis não podem figurar nos modelos é igual para os três casos.

Calculado então o VIF acaba por retirar-se a variável referente à estação do ano. Posto isto, todas as restantes variáveis, podem ser utilizadas.

À semelhança, na modelação do número de internamentos via urgência geral (secção 5.2) verifica-se o mesmo, uma vez que o cálculo de VIF é efetuado da mesma forma. Ou seja, também na modelação do número de internamentos via urgência geral não será possível a utilização da variável referente à estação do ano.

### 5.1.1 Urgência Geral

Como se quer modelar o número de admissões na urgência geral do HB, tem-se então como variável resposta ( $Y_t$ ) o número de pacientes que, entre 2012 e 2015, foram admitidos nesta urgência, diariamente. Assim sendo, a variável resposta não passa de uma contagem pelo que o natural a fazer seria a utilização de um modelo linear generalizado com distribuição de Poisson.

Usando o método de regressão *stepwise* começou-se por modelar a variável resposta em função do tempo  $t$ . No entanto, analisando o periodograma deste modelo chegou-se à conclusão que existia alguma tendência não modelada, dada a existência de um pico na frequência zero. Assim foram-se inserindo e removendo variáveis explicativas até que o periodograma não apresentasse um pico inicial, sinal de que a tendência estava modelada. Isto aconteceu aquando da introdução da variável referente ao mês do ano.

Modelada então a tendência pode voltar a fazer-se o periodograma. Este mostra a existência de, aparentemente, três ciclos. Tentou-se então o uso dos ciclos harmónicos.

No entanto, introduzidas, em lugar dos ciclos harmônicos, as variáveis referentes a dias da semana e feriado, ambas estatisticamente significativas, a questão parece ficar resolvida já que o periodograma se observa constante, sem nenhum pico evidente. Sendo assim, optou-se descartar o modelo com ciclos harmônicos, pois além de mais difícil interpretação, estes não resolviam realmente o problema. Ou seja, utilizar como parte fixa do modelo as variáveis Dia, mês, semana e feriado é uma alternativa ao uso de ciclos harmônicos, que se ajusta adequadamente aos dados.

Adicionando as restantes variáveis explicativas concluiu-se a não significância de NO e P<sub>10</sub>. Assim, como variáveis explicativas para o modelo tem-se: t, mês, feriado (ser ou não feriado), dia da semana, NConsultas, TMAX, TMIN, NO<sub>2</sub> e O<sub>3</sub>.

Entretanto notou-se a existência de sobredispersão dos dados (desvio de aproximadamente 2), pelo que se optou por ajustar um modelo cuja distribuição é Binomial Negativa com, exatamente, as mesmas variáveis explicativas.

Observa-se agora que as variáveis NO<sub>2</sub> e TMIN já não são estatisticamente significativas.

O passo seguinte seria então estudar a existência de correlação temporal nos resíduos do modelo.

No entanto, e visto a análise de séries temporais ser, em geral, feita com recurso à distribuição normal, decidiu-se realizar exatamente a mesma modelação agora assumindo como distribuição dos dados a distribuição normal. Isto pode ser feito já que em meio a tantas observações o TLC permite ajustar qualquer distribuição a uma distribuição normal. Na verdade os resultados dos modelos de Poisson, Binomial Negativa e Normal não se mostraram distintos.

Inicialmente fez-se então uma transformação de Box Cox com objetivo de estabilizar média e variância. Assim, tem-se como variável resposta  $Y_t^* = g(Y_t)$ , onde  $g(Y_t) = \frac{(Y_t-1)^\lambda}{\lambda}$ , com  $\lambda = 0.578842$  (funções *BoxCox* e *BoxCox.lambda* do R).

Voltou então a estudar-se o modelo usando regressão *stepwise* começando por considerar-se como variável explicativa apenas t. O periodograma volta a demonstrar a existência de tendência não modelada, com um pico na frequência zero. Esta tendência é modelada com a introdução da variável mês.

O periodograma apresenta agora três ciclos visíveis. O uso de ciclos harmônicos *versus* variáveis sazonais (feriado e dia da semana) revela-se novamente não compensatória, dado o grau de complexidade dos primeiros.

É agora altura de introduzir variáveis atmosféricas, meteorológicas e epidemiológicas. As variáveis NO<sub>2</sub>, P<sub>10</sub> e TMIN não se revelaram estatisticamente significativas, enquanto que NO e O<sub>3</sub> são marginalmente significativas (p-valores de 0.09 e 0.06, respetivamente).

O modelo que tem como variáveis explicativas t, mês, dia da semana, feriado, NO, O<sub>3</sub>, NConsultas e TMAX (Modelo 1) tem como  $R_a^2$  0.661, enquanto que o modelo composto apenas pelas variáveis sazonais e t (Modelo 2) tem um  $R_a^2$  de 0.641.

Posto isto, o natural seria adotar o Modelo 1. No entanto, o Modelo 1 é bem mais complexo do que o Modelo 2 e a percentagem de variabilidade explicada pelo modelo não é muito maior. Além disso, como o Modelo 1 comporta variáveis onde existiam dados faltantes (variáveis meteorológicas, epidemiológicas e atmosféricas), existe uma supressão de dados (512 das 1461 observações estão em falta) na hora do ajuste do modelo, o que pode levar a erros.

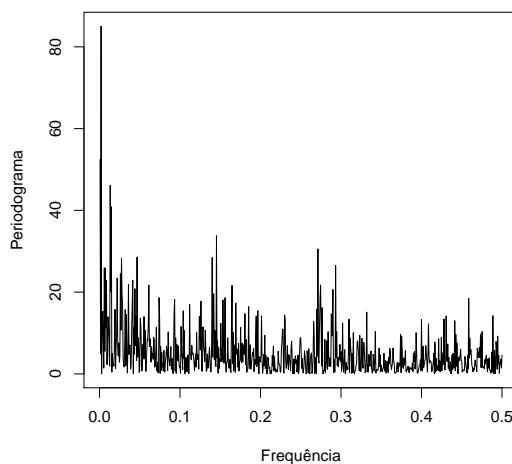


Figura 5.1: Periodograma dos resíduos do Modelo 2

Assim sendo, optou-se por considerar o Modelo 2 como o "melhor" modelo. Tendo

em atenção a Figura 5.1 vê-se um típico periodograma de uma série correlacionada, sem nenhum pico isolado evidente. Assim sendo, é agora altura de construir as representações gráficas das FAC e FACP dos resíduos do modelo, para tentar perceber se existe alguma correlação no modelo.

Observando agora o gráfico da Figura 5.2 vê-se que a FAC decai exponencialmente para zero e a FACP se anula a partir de um intervalo superior a 7. Assim conclui-se a existência de uma estrutura autoregressiva de ordem 7 para os resíduos, pelo que se pode dizer que as observações de cada dia estão correlacionadas com as dos sete dias anteriores, o que tem lógica tendo em conta os dados, ou seja, existe correlação semanal e, portanto, o número de pacientes admitidos num dia  $t$  está dependente do número de pacientes admitidos nos sete dias antecedentes.

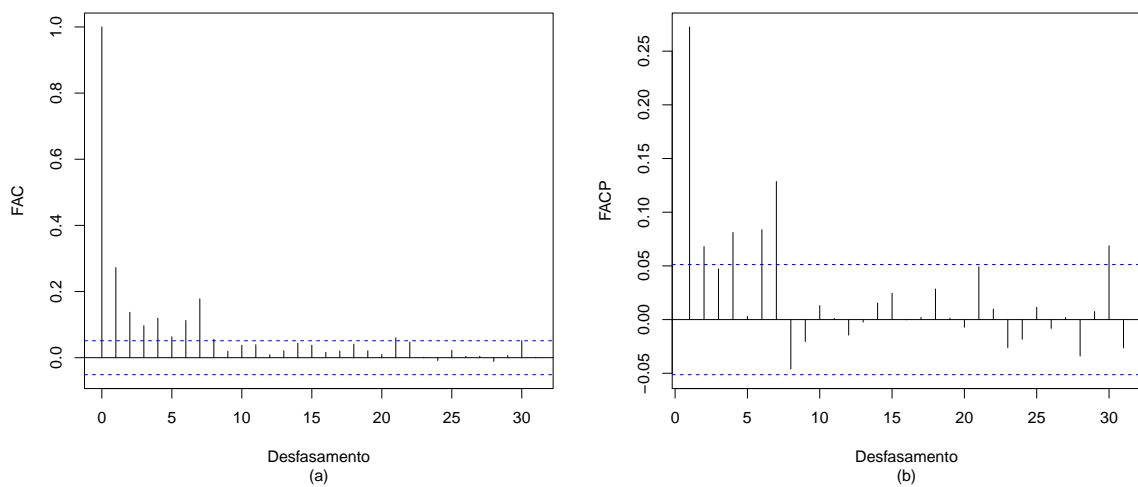


Figura 5.2: FAC (a) e FACP (b) dos resíduos

Assim sendo, tem-se, com um MAPE de 3.39%, o modelo (Modelo 3)

$$Y_t^* = \mu + \epsilon_t \quad (5.1)$$

cuja parte fixa do modelo é dada por

$$\begin{aligned} \mu = & \beta_0 + \beta_1 t + \beta_2 \text{Feriado} + \beta_3 \text{Terca} + \beta_4 \text{Quarta} + \beta_5 \text{Quinta} + \beta_6 \text{Sexta} + \beta_7 \text{Sabado} \\ & + \beta_8 \text{Domingo} + \beta_9 \text{Fevereiro} + \beta_{10} \text{Marco} + \beta_{11} \text{Abril} + \beta_{12} \text{Maio} + \beta_{13} \text{Junho} \\ & + \beta_{14} \text{Julho} + \beta_{15} \text{Agosto} + \beta_{16} \text{Setembro} + \beta_{17} \text{Outubro} + \beta_{18} \text{Novembro} + \beta_{19} \text{Dezembro} \end{aligned}$$

e os erros dados por

$$\begin{aligned} \epsilon_t = & \theta_1(Y_{t-1}^* - \mu) + \theta_2(Y_{t-2}^* - \mu) + \theta_3(Y_{t-3}^* - \mu) + \theta_4(Y_{t-4}^* - \mu) + \theta_5(Y_{t-5}^* - \mu) \\ & + \theta_6(Y_{t-6}^* - \mu) + \theta_7(Y_{t-7}^* - \mu) + \delta_t. \end{aligned}$$

$\delta_t$  representa o ruído branco, com variância  $\hat{\sigma}^2 = 4.695$ .

As restantes estimativas para os parâmetros do modelo podem ser vistas nas Tabelas 5.2 e 5.1.

Tabela 5.1: Estimativas para os coeficientes do autogressivo de ordem 7

<b>Coefficiente</b>	<b>Estimativa</b>	<b>Erro padrão</b>
$\hat{\theta}_1$	0.237	0.026
$\hat{\theta}_2$	0.046	0.027
$\hat{\theta}_3$	0.014	0.027
$\hat{\theta}_4$	0.073	0.027
$\hat{\theta}_5$	-0.024	0.027
$\hat{\theta}_6$	0.052	0.027
$\hat{\theta}_7$	0.1301	0.026

Tabela 5.2: Tabela sumária do Modelo 3

Coefficiente	Estimativa	Erro padrão	z	p-value
$\hat{\beta}_0$	51.898	0.269	193.093	< 2e-16
$\hat{\beta}_1$	0.003	0.0001	18.747	< 2e-16
$\hat{\beta}_2$	-3.579	0.359	-9.981	< 2e-16
$\hat{\beta}_3$	-3.130	0.226	-13.824	< 2e-16
$\hat{\beta}_4$	-3.450	0.226	-15.251	< 2e-16
$\hat{\beta}_5$	-3.451	0.226	-15.251	< 2e-16
$\hat{\beta}_6$	-3.228	0.227	-14.236	< 2e-16
$\hat{\beta}_7$	-6.740	0.226	-29.769	< 2e-16
$\hat{\beta}_8$	-9.044	0.226	-39.973	< 2e-16
$\hat{\beta}_9$	-0.018	0.301	-0.061	<b>0.951</b>
$\hat{\beta}_{10}$	-0.764	0.294	-2.604	0.009
$\hat{\beta}_{11}$	-1.236	0.297	-4.166	3.28e-05
$\hat{\beta}_{12}$	-0.563	0.294	-1.916	0.056
$\hat{\beta}_{13}$	-0.006	0.297	-0.020	<b>0.984</b>
$\hat{\beta}_{14}$	0.574	0.295	1.946	0.052
$\hat{\beta}_{15}$	1.168	0.295	3.958	7.92e-05
$\hat{\beta}_{16}$	-0.299	0.298	-1.002	<b>0.316</b>
$\hat{\beta}_{17}$	-0.491	0.296	-1.658	<b>0.098</b>
$\hat{\beta}_{18}$	-1.718	0.299	-5.738	1.16e-08
$\hat{\beta}_{19}$	-0.586	0.298	-1.966	0.049

Assim sendo, sabe-se que, além de resíduos com estrutura de correlação, a variável  $Y_t^*$

- tende a aumentar, em média, com o aumento do  $t$ ;
- toma, em média, menores valores em dias de feriado, quando comparado com os restantes dias;

- diminui, em média, ao longo da semana, tomando como referência a segunda-feira;
- tende a diminuir, em média, em março, abril, maio, novembro e dezembro, a aumentar em julho e agosto e a manter-se aproximadamente igual, ao registado em janeiro, nos meses de fevereiro, junho, setembro e outubro.

### 5.1.1.1 Análise dos Resíduos

Sendo o modelo final o Modelo 3 tem-se como resíduos do mesmo os valores estimados do ruído branco  $\delta_t$ .

Pelo gráfico da Figura 5.3 (a) pode então ver-se que pontos estão na sua maioria sobre a reta traçada, pelo que os resíduos seguem uma distribuição normal. Tendo ainda em atenção os gráficos da Figura 5.3 (b) e Figura 5.3 (c) pode ver-se que os pontos se distribuem aleatoriamente em torno do resíduo nulo, formando uma mancha de largura uniforme. Assim sendo, vêem-se cumpridos os pressupostos do modelo.

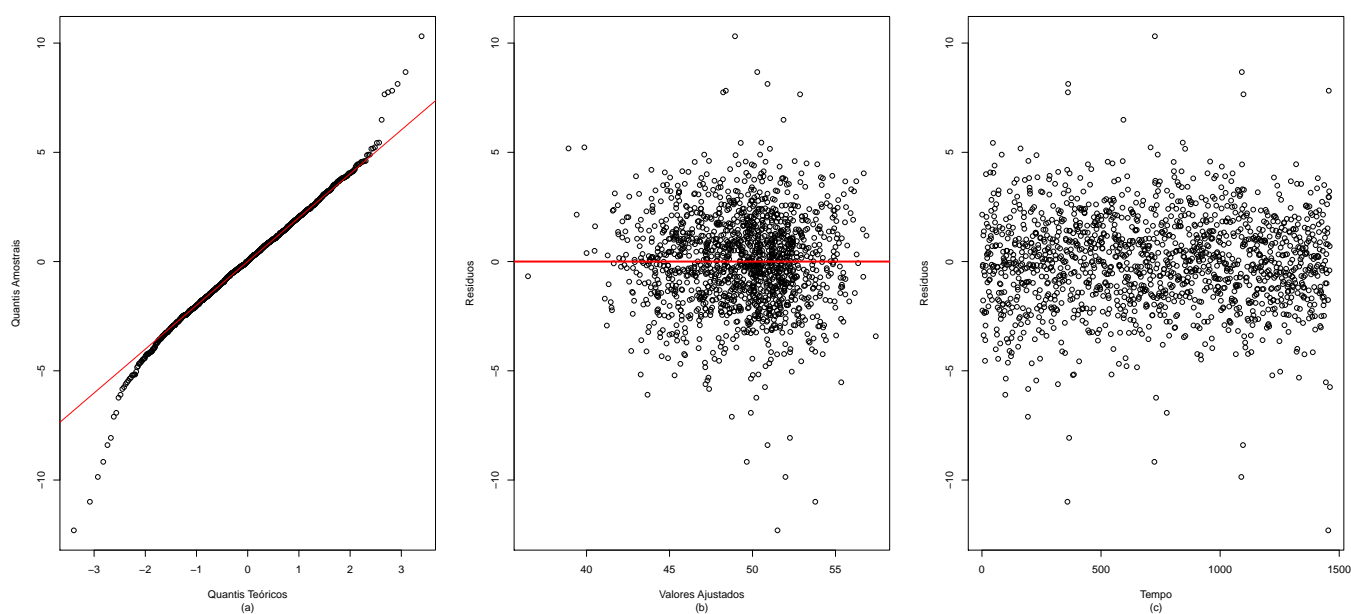


Figura 5.3: Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 *vs* os resíduos (b) e do tempo *vs* os resíduos (c)

### 5.1.1.2 Validação do Modelo e Predição

Com objetivo de testar a adequação do modelo aos dados utilizaram-se os dados do primeiro trimestre de 2016, isto é, dados relativos à admissão de pacientes na urgência geral desde 01/01/2016 a 31/03/2016, o que perfaz uma totalidade de 91 dias.

Para tal fez-se primeiramente previsão de  $Y$  para o ano de 2016, construindo-se os intervalos de confiança a 95%. Para tal consideraram-se como os valores previstos para a variável resposta a soma entre os valores previstos pela parte fixa do modelo e pelo processo autoregressivo. O mesmo foi feito no caso das urgências obstétrica e pediátrica.

Tendo em conta a representação gráfica da Figura 5.4 vê-se a amarelo os valores previstos pelo modelo para o ano de 2016. As linhas a vermelho representam uma suavização dos valores previstos e do intervalo de confiança (a tracejado). A azul tracejado tem-se os valores que realmente foram registados no primeiro trimestre de 2016.

Vê-se ainda a rosa a previsão que o modelo faz para os anos de 2012, 2013, 2014 e 2015, sendo esta bastante semelhante ao registado verdadeiramente (linha azul).

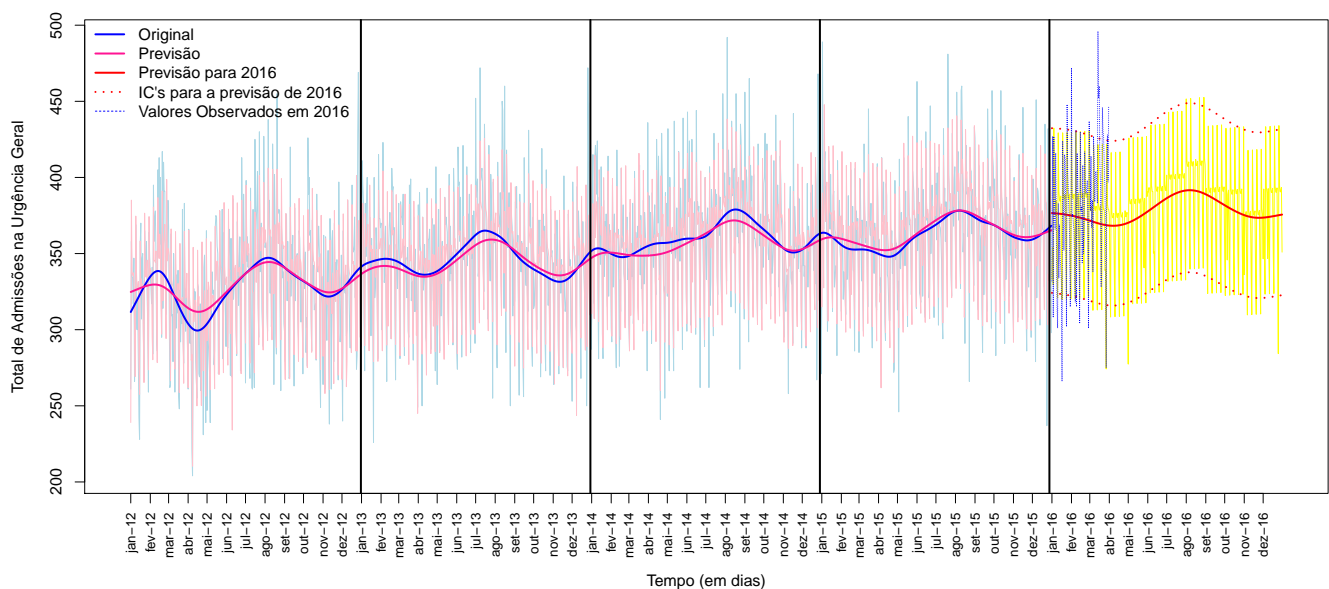


Figura 5.4: Previsão do número de pacientes admitidos na urgência geral, diariamente



Em 80 dos 91 dias a previsão feita pelo modelo cai dentro do intervalo de confiança, sendo que o modelo prevê acertadamente 87.9% dos dias. Em 9 dias apenas o valor observado se encontra acima do limite superior do IC e em 2 dias abaixo do limite inferior. Estes dias são, maioritariamente, pertencentes ao mês de março, onde neste ano de 2016 se registou um número de admissões no SU superior ao registado para o mesmo mês em anos anteriores.

No Anexo A são apresentados os valores previstos pelo modelo, com intervalos de confiança e respetivos valores reais observados. Nas tabelas, a cor vermelha estão identificados os dias/valores em que o valor observado ficou acima do limite superior do intervalo de confiança previsto e a cor azul os dias onde o valor observado ficou abaixo do limite inferior do intervalo de confiança previsto.

Para o hospital a informação sobre o número de pacientes que entram na urgência tem de ser traduzida em tomada de decisão sobre o número de médicos necessários na urgência, diariamente.

Desta forma, interessa fazer também inferência sobre o número de médicos necessários diariamente no serviço de urgência.

O número de médicos necessários terá um arredondamento superior, a número inteiro, da divisão do número de pacientes por um valor de ponto de corte. Este valor de ponto de corte (pc), será teoricamente o número de pacientes que um médico poderá ver diariamente. Por discussão com a direção de produção, o arredondamento deverá ser feito a 0.1. Assim, se tivermos um valor estimado de 357 pacientes para um ponto de corte  $pc=70$ , seriam necessários  $357/70=4.1$  médicos, isto é, 5 médicos.

Será de interesse então saber, segundo estes critérios de arredondamento, qual a probabilidade de o nosso modelo acertar no número de médicos estimado de acordo com aquele dado pelo estimador pontual  $\hat{Y}$ .

Isto é, calcular

$$P\left(\text{acertar no número de médicos como dado por } \hat{Y}\right) = P\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 0.1 < \frac{Y}{pc} < \left\lceil \frac{\hat{Y}}{pc} \right\rceil + 1 + 0.1\right)$$

onde  $\lceil \cdot \rceil$  representa a parte inteira de um número e  $pc$  um ponto de corte escolhido, como definido anteriormente.

Esta probabilidade é calculada como

$$\begin{aligned} & P\left(Y < \left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 1 + 0.1\right) \times pc\right) - P\left(Y < \left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 0.1\right) \times pc\right) = \\ & = P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 1 + 0.1\right) \times pc\right)\right) - P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 0.1\right) \times pc\right)\right) \end{aligned}$$

Note que  $Y^*$  tem distribuição conhecida.

$$Y^* \sim N(X\beta, \Sigma)$$

onde o modelo para  $X\beta$  é dado por Modelo 3 em (5.1) e a estrutura de variância/covariância é aquela utilizada para modelar a série temporal.

Da mesma forma, pode calcular-se

1.  $P\left(\text{necessitar de mais de um médico do que o dado por } \hat{Y}\right)$ ;
2.  $P\left(\text{necessitar de mais de dois médicos do que o dado por } \hat{Y}\right)$ ;
3.  $P\left(\text{necessitar de menos de um médico do que o dado por } \hat{Y}\right)$ ;
4.  $P\left(\text{necessitar de menos de dois médicos do que o dado por } \hat{Y}\right)$ .

Cada uma destas probabilidades pode facilmente ser calculada como anteriormente, isto é,

1.  $P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 2 + 0.1\right) \times pc\right)\right) - P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 1 + 0.1\right) \times pc\right)\right)$ ;
2.  $P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 3 + 0.1\right) \times pc\right)\right) - P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 2 + 0.1\right) \times pc\right)\right)$ ;

3.  $P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil + 0.1\right) \times pc\right)\right) - P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil - 1 + 0.1\right) \times pc\right)\right);$
4.  $P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil - 1 + 0.1\right) \times pc\right)\right) - P\left(Y^* < g\left(\left(\left\lceil \frac{\hat{Y}}{pc} \right\rceil - 2 + 0.1\right) \times pc\right)\right).$

As Figuras 5.5, 5.6, 5.7, 5.8, 5.9 e 5.10 representam estas probabilidades para valores de  $pc$  entre 50 e 90, para os dias 01/01/2016, 02/01/2016 e 03/01/2016.

A probabilidade de acertar no número de médicos no dia 01/01/2016 é maior para pontos de corte próximos de 73 e 90, como se pode ver pela Figura 5.5. No entanto, a probabilidade de, neste dia, ser necessário mais um médico, como se pode ver na Figura 5.8 (a), ronda no máximo os 0.4., para valores de  $pc$  de, aproximadamente 57 e 68; a probabilidade de ser necessário menos um médico é exatamente o complementar, como se vê pela Figura 5.8 (c). Quanto às probabilidade de serem precisos mais dois ou menos dois médicos, pode ver-se nos gráficos das Figuras 5.8 (b) e 5.8 (d), que estas rondam a probabilidade zero.

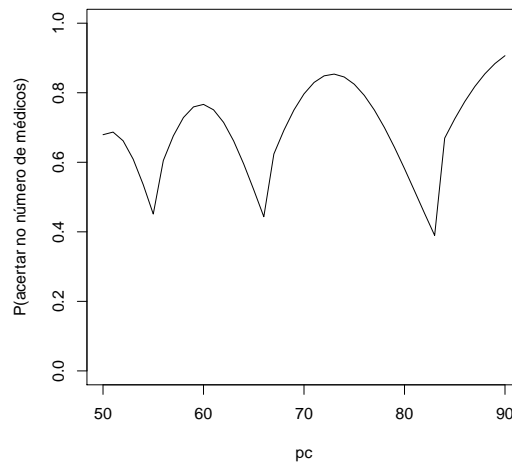


Figura 5.5: Probabilidade de acertar no número de médicos, para valores de  $pc$  entre 50 e 90, no dia 01/01/2016

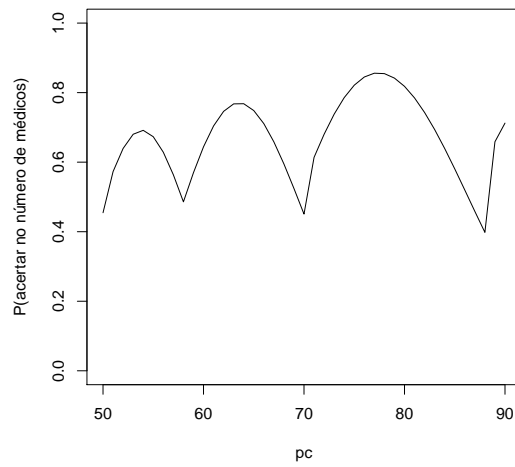


Figura 5.6: Probabilidade de acertar no número de médicos, para valores de pc entre 50 e 90, no dia 02/01/2016

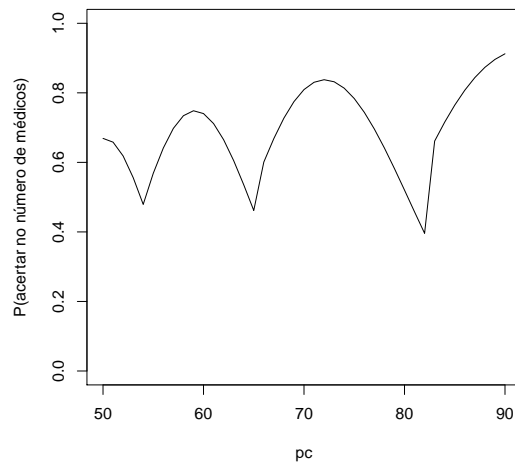


Figura 5.7: Probabilidade de acertar no número de médicos, para valores de pc entre 50 e 90, no dia 03/01/2016

Já no dia 02/01/2016, a probabilidade de acertar no número de médicos, é maior para pontos de corte próximos de 78, como se pode ver pela Figura 5.6. No entanto, a

probabilidade de, neste dia, ser necessário mais um médico, como se pode ver na Figura 5.9 (a), ronda no máximo os 0.4., para valores de  $pc$  de, aproximadamente 60 e 73; a probabilidade de ser necessário menos um médico é exatamente o complementar, como se vê pela Figura 5.9 (c). Quanto às probabilidade de serem precisos mais dois ou menos dois médicos, pode ver-se nos gráficos das Figuras 5.9 (b) e 5.9 (d), que estas rondam a probabilidade zero.

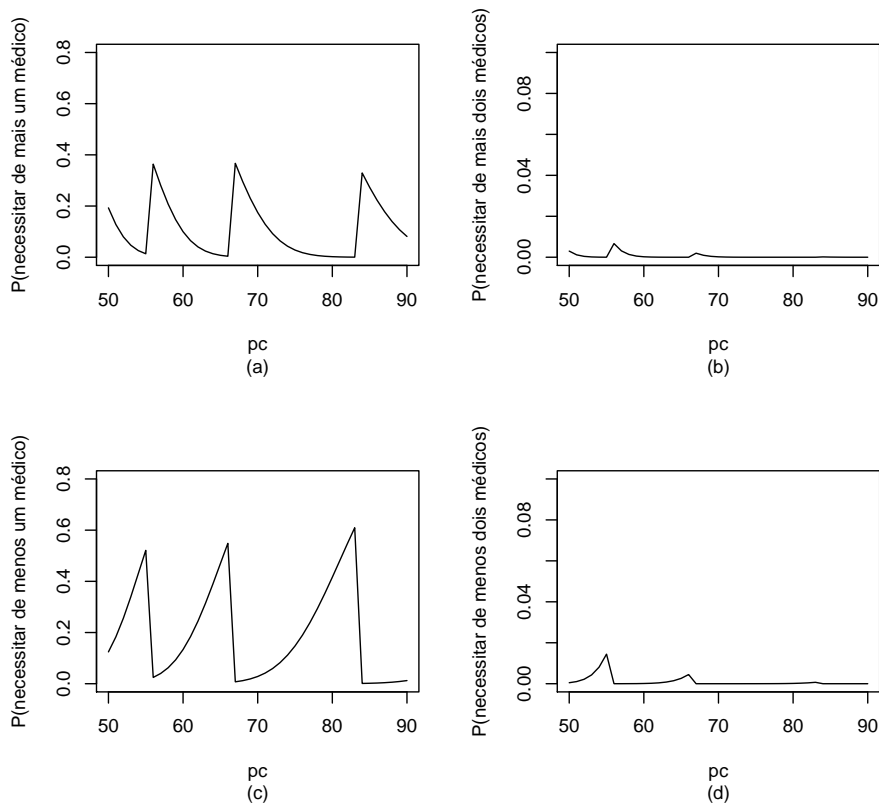


Figura 5.8: Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 01/01/2016, para valores de  $pc$  entre 50 e 90

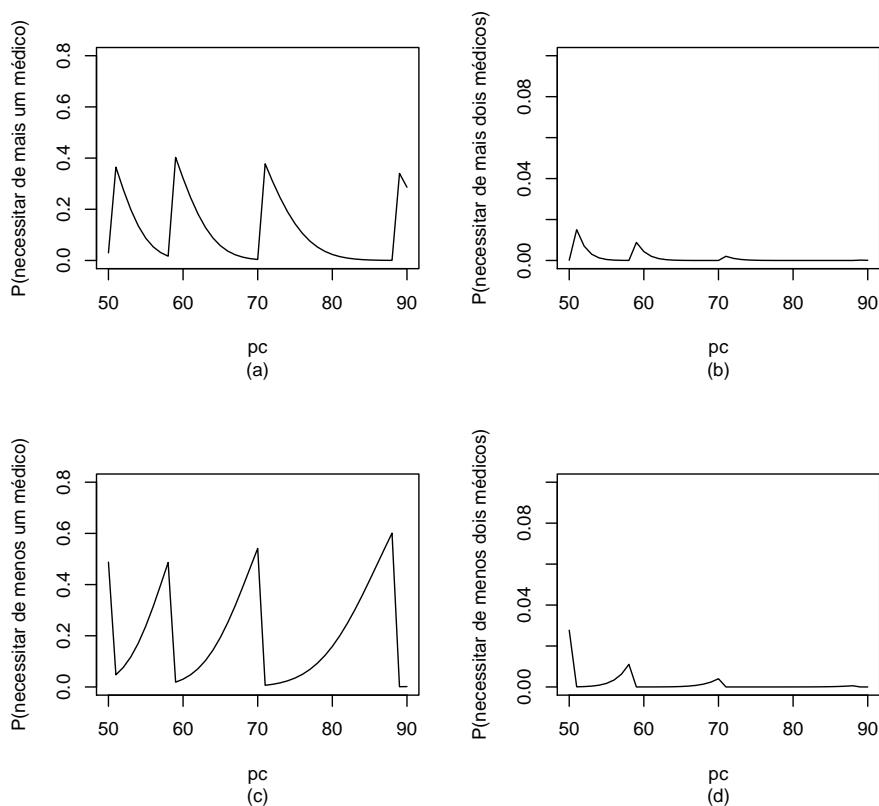


Figura 5.9: Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 02/01/2016, para valores de  $pc$  entre 50 e 90

No 03/01/2016, as cinco probabilidades são semelhantes ao registado para o dia 01/01/2016, como se pode ver nas Figuras 5.7 e 5.10.

No entanto, como é óbvio o hospital não pode mudar o ponto de corte todos os dias. Assim, tendo a direção do hospital decidido um ponto de corte adequado às suas necessidades é possível calcularem as cinco probabilidades em questão. Como se pode ver será provavelmente raro um caso em que a diferença de médicos realmente necessários e o número de médicos previstos seja maior do que um, seja por defeito ou excesso. Definido então um ponto de corte, o hospital tem a possibilidade de olhando para as probabilidades calculadas decidir se será compensatório ou não a escala de menos ou mais um médico.

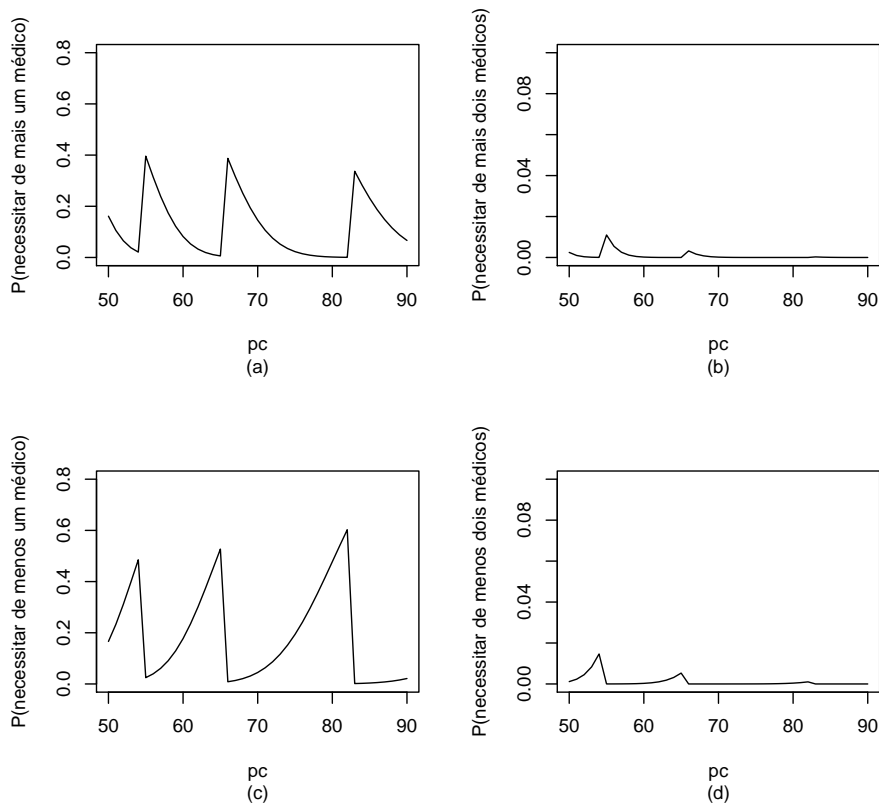


Figura 5.10: Probabilidade de necessitar de mais um médico (a), mais dois médicos (b), menos um médico (c) ou menos dois médicos (d), do que o previsto para o dia 03/01/2016, para valores de  $pc$  entre 50 e 90

### 5.1.2 Urgência Obstétrica

O objetivo nesta secção é, novamente, modelar o número de pacientes admitidos, mas agora no serviço de urgência obstétrico. O mesmo será feito no caso da urgência pediátrica.

Optou-se por apenas ter em conta a distribuição dos dados como gaussiana, visto na secção anterior ter-se chegado à conclusão que os resultados eram semelhantes.

Primeiramente fez-se então uma transformação de Box Cox com objetivo de estabilizar média e variância. Assim, tem-se como variável resposta  $Y_t^* = g(Y_t)$ , onde  $g(Y_t) = \frac{(Y_t - 1)^\lambda}{\lambda}$ , com  $\lambda = 0.144861$ .

Começou por estudar-se o modelo com apenas  $t$  como variável explicativa. O periodograma dos resíduos deste modelo mostra então uma tendência não modelada, que acaba por desaparecer com a introdução da variável respeitante ao mês na parte fixa do modelo. O periodograma demonstra agora a existência de três ciclos.

Foram então utilizados ciclos harmónicos, que à semelhança do ocorrido no caso da modelação de admissões na urgência geral, se demonstraram bastante complexos, não sendo preferíveis à simples introdução das variáveis dia da semana e feriado, que removem todos os ciclos que aparentam existir. Este modelo (Modelo 1) tem um coeficiente de determinação ajustado de 0.5766, o que significa que aproximadamente 57.66% da variabilidade da variável resposta é explicada pelo modelo.

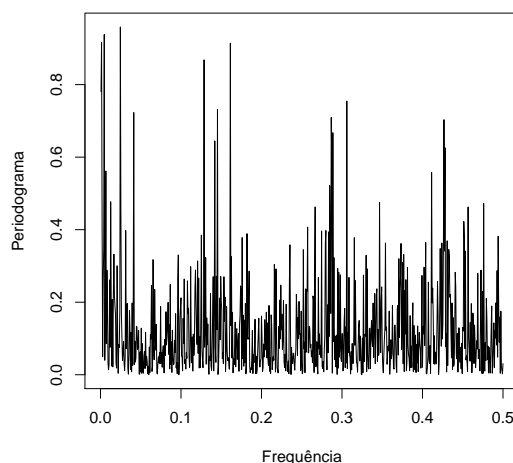


Figura 5.11: Periodograma dos resíduos do Modelo 1

Com a introdução de variáveis atmosféricas, meteorológicas e epidemiológicas verificou-se que apenas a variável  $O_3$  se mostrava estatisticamente significativa. Introduzindo uma variável respeitante às fases da lua (tomando valor 1 quando é lua nova ou lua cheia e 0 nas restantes fases) pode perceber-se que esta também não é significativa para o modelo. Assim, este segundo modelo (Modelo 2) conta com um  $R_a^2$  de aproximadamente 0.5839. De referir ainda que a variável respeitante à concentração atmosférica do ozono contém



436 observações em falta, pelo que se perde muita informação. Assim, apesar do Modelo 2 ser teoricamente "melhor" do que o Modelo 1, será o primeiro modelo que passará a ser utilizado, cujo periodograma pode ser visto na Figura 5.11, periodograma esse indicador de correlação.

Com objetivo de estudar a possível correlação, é necessário perceber o comportamento das funções de autocorrelação e autocorrelação parcial sobre os resíduos.

Tendo então em atenção a representação gráfica das mesmas, Figura 5.12 (a) e Figura 5.12 (b), observa-se que a FAC decresce sinusoidalmente para zero e a FACP anula-se a partir de 8. Assim os resíduos do Modelo 2 contam com uma estrutura autoregressiva de ordem 8, pelo que se tem que o número de pacientes admitidos em determinado dia  $t$  está dependente do número de pacientes admitidos nos 8 dias anteriores.

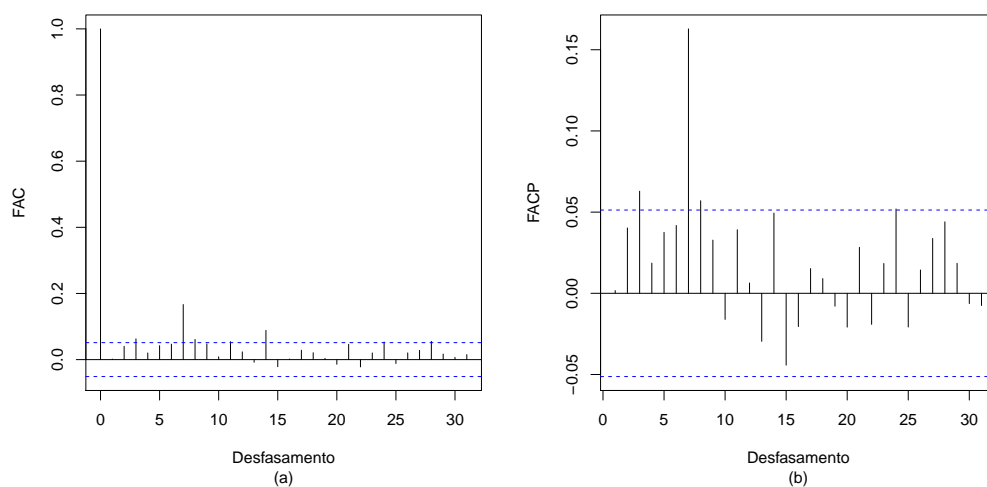


Figura 5.12: FAC (a) e FACP (b) dos resíduos

Assim sendo, tem-se, com um MAPE de 6%, o modelo (Modelo 3)

$$Y_t^* = \mu + \epsilon_t \quad (5.2)$$

cuja parte fixa do modelo é dada por

$$\begin{aligned} \mu = & \beta_0 + \beta_1 t + \beta_2 \text{Feriado} + \beta_3 \text{Terca} + \beta_4 \text{Quarta} + \beta_5 \text{Quinta} + \beta_6 \text{Sexta} + \beta_7 \text{Sabado} \\ & + \beta_8 \text{Domingo} + \beta_9 \text{Fevereiro} + \beta_{10} \text{Marco} + \beta_{11} \text{Abril} + \beta_{12} \text{Maio} + \beta_{13} \text{Junho} \\ & + \beta_{14} \text{Julho} + \beta_{15} \text{Agosto} + \beta_{16} \text{Setembro} + \beta_{17} \text{Outubro} + \beta_{18} \text{Novembro} + \beta_{19} \text{Dezembro} \end{aligned}$$

e os erros dados por

$$\begin{aligned} \epsilon_t = & \theta_1(Y_{t-1}^* - \mu) + \theta_2(Y_{t-2}^* - \mu) + \theta_3(Y_{t-3}^* - \mu) + \theta_4(Y_{t-4}^* - \mu) + \theta_5(Y_{t-5}^* - \mu) \\ & + \theta_6(Y_{t-6}^* - \mu) + \theta_7(Y_{t-7}^* - \mu) + \theta_8(Y_{t-8}^* - \mu) + \delta_t. \end{aligned}$$

$\delta_t$  representa o ruído branco, com variância  $\hat{\sigma}^2 = 0.1162$ .

As estimativas para os parâmetros do modelo podem ser vistas nas Tabelas 5.4 e 5.3.

Tabela 5.3: Estimativas para os coeficientes do autogressivo de ordem 8

<b>Coefficiente</b>	<b>Estimativa</b>	<b>Erro padrão</b>
$\hat{\theta}_1$	-0.021	0.026
$\hat{\theta}_2$	0.028	0.026
$\hat{\theta}_3$	0.054	0.026
$\hat{\theta}_4$	0.007	0.026
$\hat{\theta}_5$	0.029	0.026
$\hat{\theta}_6$	0.041	0.026
$\hat{\theta}_7$	0.164	0.026
$\hat{\theta}_8$	0.058	0.026

Tabela 5.4: Tabela sumária do Modelo 3

Coefficiente	Estimativa	Erro padrão	z	p-valor
$\hat{\beta}_0$	5.002	0.041	123.268	< 2e-16
$\hat{\beta}_1$	1e-04	2.24e-05	4.473	8.33e-06
$\hat{\beta}_2$	-0.287	0.054	-5.299	1.34e-07
$\hat{\beta}_3$	-0.205	0.034	-5.999	2.51e-09
$\hat{\beta}_4$	-0.338	0.034	-9.910	< 2e-16
$\hat{\beta}_5$	-0.325	0.034	-9.509	< 2e-16
$\hat{\beta}_6$	-0.202	0.034	-5.909	4.30e-09
$\hat{\beta}_7$	-0.942	0.034	-27.573	< 2e-16
$\hat{\beta}_8$	-1.135	0.034	-33.220	< 2e-16
$\hat{\beta}_9$	-0.131	0.045	-2.877	0.004
$\hat{\beta}_{10}$	-0.127	0.044	-2.852	0.004
$\hat{\beta}_{11}$	-0.035	0.045	-4.166	<b>0.438</b>
$\hat{\beta}_{12}$	0.056	0.044	1.258	<b>0.209</b>
$\hat{\beta}_{13}$	0.042	0.045	0.930	<b>0.352</b>
$\hat{\beta}_{14}$	0.062	0.045	1.395	<b>0.163</b>
$\hat{\beta}_{15}$	0.199	0.045	4.469	8.46e-06
$\hat{\beta}_{16}$	0.137	0.045	3.045	0.002
$\hat{\beta}_{17}$	-0.063	0.045	-1.418	<b>0.157</b>
$\hat{\beta}_{18}$	-0.106	0.045	-2.347	0.019
$\hat{\beta}_{19}$	0.100	0.045	2.224	0.026

Assim sendo, sabe-se que, além de resíduos com estrutura de correlação, a variável  $Y_t^*$

- tende a aumentar, em média, com o aumento do t;
- toma, em média, menores valores em dias de feriado, quando comparado com os restantes dias;

- diminui, em média, ao longo da semana, tomando como referência a segunda-feira;
- tende a diminuir, em média, em fevereiro, março e novembro, a aumentar em agosto setembro e dezembro e a manter-se aproximadamente igual, ao registrado em janeiro, nos meses de abril, maio, junho, julho e outubro.

### 5.1.2.1 Análise dos Resíduos

Como já referido os resíduos do Modelo 3 serão os valores estimados do ruído branco  $\delta_t$ .

Pelo gráfico da Figura 5.13 (a) pode então ver-se que os pontos estão na sua maioria sobre a reta traçada, pelo que os resíduos seguem uma distribuição normal. Tendo ainda em atenção os gráficos da Figura 5.13 (b) e Figura 5.13 (c) pode ver-se que os pontos se distribuem aleatoriamente em torno do resíduo nulo, formando uma mancha de largura uniforme. Assim sendo, vêem-se cumpridos os pressupostos do modelo.

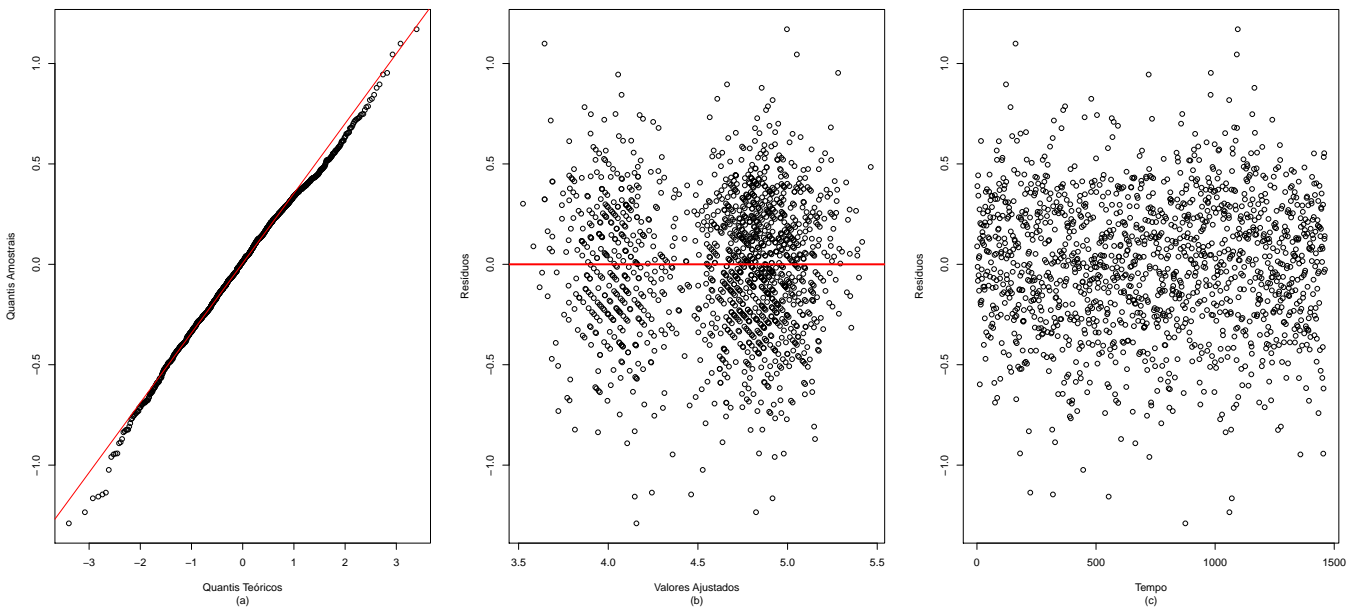


Figura 5.13: Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 *vs* os resíduos (b) e do tempo *vs* os resíduos (c)

### 5.1.2.2 Validação do Modelo e Predição

Com objetivo de testar a adequação do modelo aos dados utilizaram-se os dados referentes à admissão de pacientes na urgência geral desde 01/01/2016 a 31/03/2016, tendo-se assim dados para uma totalidade de 91 dias.

Inicialmente fez-se previsão de  $Y$  para o ano de 2016, construindo-se os intervalos de confiança a 95%.

Tendo em conta a representação gráfica da Figura 5.14 vê-se a amarelo os valores previstos pelo modelo para o ano de 2016. As linhas a vermelho representam uma suavização dos valores previstos e do intervalo de confiança (a tracejado). A azul tracejado tem-se os valores que realmente foram registados no primeiro trimestre de 2016.

Vê-se ainda a rosa a previsão que o modelo faz para os anos de 2012, 2013, 2014 e 2015, sendo esta bastante semelhante ao registado verdadeiramente (linha azul).

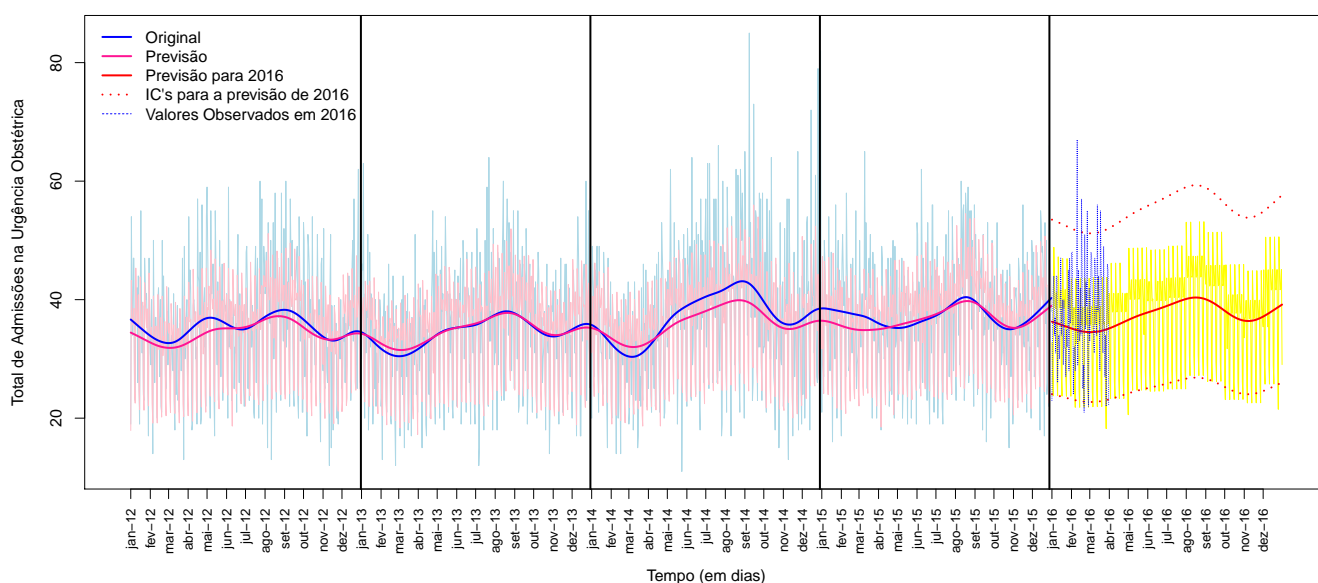


Figura 5.14: Previsão do número de pacientes admitidos na urgência obstétrica, diariamente

Em 78 dos 91 dias a previsão feita pelo modelo cai dentro do intervalo de confiança,

sendo que o modelo prevê acertadamente 85.71% dos dias. Em 11 dias apenas o valor observado se encontra acima do limite superior do IC e em 2 dias abaixo do limite inferior. Estes dias são, maioritariamente, pertencentes ao mês de fevereiro e março.

No Anexo B são apresentados os valores previstos pelo modelo, com intervalos de confiança e respetivos valores reais observados. Nas tabelas, a cor vermelha estão identificados os dias em que o valor observado ficou acima do limite superior do intervalo de confiança previsto e a cor azul os dias/valores onde o valor observado ficou abaixo do limite inferior do intervalo de confiança previsto.

### 5.1.3 Urgência Pediátrica

Como já referido, nesta secção, supõe-se uma distribuição normal para a variável resposta - número de pacientes admitidos, diariamente, na urgência pediátrica.

Inicialmente fez-se uma transformação de Box Cox com objetivo de estabilizar média e variância. Assim, tem-se como variável resposta  $Y_t^* = g(Y_t)$ , onde  $g(Y_t) = \frac{(Y_t-1)^\lambda}{\lambda}$ , com  $\lambda = 0.02747$ .

Começou por estudar-se o modelo com apenas  $t$  como variável explicativa.

O periodograma dos resíduos do modelo mostra um pico na frequência zero e portanto existe uma tendência não modelada, que desaparece com a introdução da variável respeitante ao mês na parte fixa do modelo. O periodograma demonstra, à semelhança do que aconteceu na modelação das admissões de ambas as outras urgências, três picos espaçados, o que indica a existência de três ciclos distintos.

Tentou-se então a introdução de ciclos harmónicos, que, mais uma vez, se demonstraram não significativos relativamente à opção de introduzir feriado e dia da semana na parte fixa do modelo. Este modelo (Modelo 1) tem um coeficiente de determinação ajustado de 0.4837, o que significa que aproximadamente 48.37% da variabilidade da variável  $Y_t^*$  é explicada pelo modelo.

Com a introdução de variáveis atmosféricas, meteorológicas e epidemiológicas verificou-

se que apenas a variável NO se mostrava estatisticamente significativa.  $P_{10}$  revela-se marginalmente significativa (p-valor de 0.077), isto é, a um nível de significância de 10%  $P_{10}$  poderia ser classificada como significativa para o modelo. Este segundo modelo, composto pelas variáveis presentes no Modelo 1, NO e  $P_{10}$  (Modelo 2) tem um  $R_a^2$  de aproximadamente 0.5173. O modelo contém um total de 395 observações faltantes, causada pelo número de NA's das variáveis atmosféricas, pelo que é perdida alguma informação. Assim, apesar do Modelo 2 ser teoricamente "melhor" do que o Modelo 1, será o primeiro modelo que passará a ser utilizado, cujo periodograma pode ser visto na Figura 5.15, periodograma esse indicador de correlação.

Note-se ainda que apesar de parecer existir na frequência zero indicação de tendência isso não se verifica. Não existe nenhum pico distinto para tal. Além disso, também não existe nenhum pico isolado indicador da existência de nenhum ciclo não identificado.

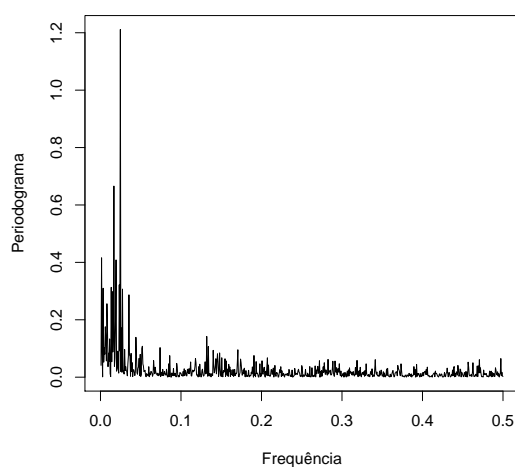


Figura 5.15: Periodograma dos resíduos do Modelo 1

Com objetivo de estudar a possível correlação indicada pelo periodograma, é necessário perceber o comportamento das funções de autocorrelação e autocorrelação parcial sobre os resíduos.

Tendo então em atenção a Figura 5.16 (a) e a Figura 5.16 (b), representações gráficas

da FAC e FACP, respetivamente, observa-se que a FAC decresce sinusoidalmente para zero e a FACP anula-se a partir do desfasamento 8. Assim os resíduos do Modelo 2 contam com uma estrutura autoregressiva de ordem 8, pelo que se tem que o número de pacientes admitidos em determinado dia  $t$  está dependente do número de pacientes admitidos nos 8 dias anteriores, à semelhança do que acontece na urgência obstétrica.

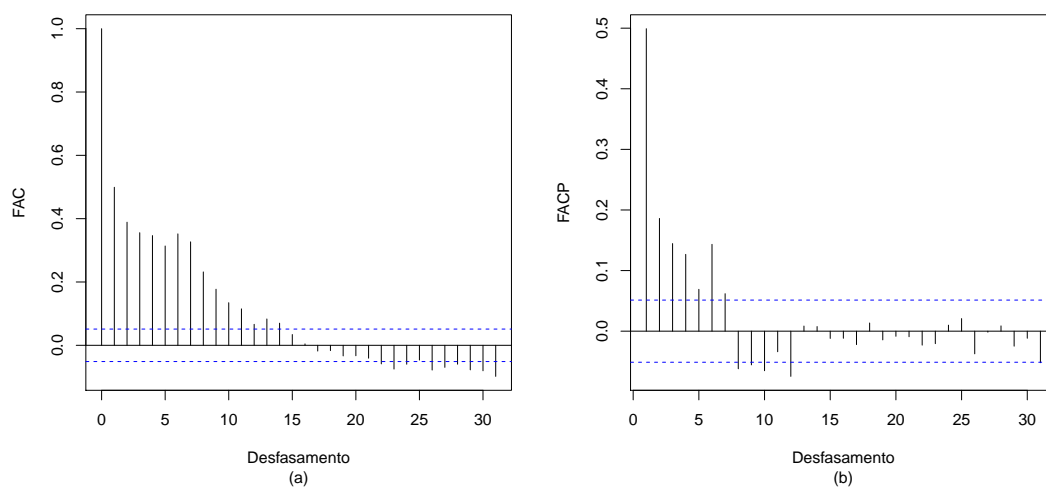


Figura 5.16: FAC (a) e FACP (b) dos resíduos

Assim sendo, tem-se, com um MAPE de 2.12%, o modelo (Modelo 3)

$$Y_t^* = \mu + \epsilon_t \quad (5.3)$$

cuja parte fixa do modelo é dada por

$$\begin{aligned} \mu = & \beta_0 + \beta_1 t + \beta_2 \text{Feriado} + \beta_3 \text{Terca} + \beta_4 \text{Quarta} + \beta_5 \text{Quinta} + \beta_6 \text{Sexta} + \beta_7 \text{Sabado} \\ & + \beta_8 \text{Domingo} + \beta_9 \text{Fevereiro} + \beta_{10} \text{Marco} + \beta_{11} \text{Abril} + \beta_{12} \text{Maio} + \beta_{13} \text{Junho} \\ & + \beta_{14} \text{Julho} + \beta_{15} \text{Agosto} + \beta_{16} \text{Setembro} + \beta_{17} \text{Outubro} + \beta_{18} \text{Novembro} + \beta_{19} \text{Dezembro} \end{aligned}$$

e os erros dados por



$$\begin{aligned} \epsilon_t = & \theta_1(Y_{t-1}^* - \mu) + \theta_2(Y_{t-2}^* - \mu) + \theta_3(Y_{t-3}^* - \mu) + \theta_4(Y_{t-4}^* - \mu) + \theta_5(Y_{t-5}^* - \mu) \\ & + \theta_6(Y_{t-6}^* - \mu) + \theta_7(Y_{t-7}^* - \mu) + \theta_8(Y_{t-8}^* - \mu) + \delta_t. \end{aligned}$$

$\delta_t$  representa o ruído branco, com variância  $\hat{\sigma}^2 = 0.019$ .

As estimativas para os parâmetros do modelo podem ser vistas nas tabelas 5.5 e 5.6.

Tabela 5.5: Tabela sumária do Modelo 3

<b>Coefficiente</b>	<b>Estimativa</b>	<b>Erro padrão</b>	<b>z</b>	<b>p-valor</b>
$\hat{\beta}_0$	5.463	0.020	277.567	< 2e-16
$\hat{\beta}_1$	8.97e-05	1.09e-05	8.267	3.08e-16
$\hat{\beta}_2$	-0.117	0.026	-4.462	8.74e-06
$\hat{\beta}_3$	-0.104	0.017	-6.283	4.38e-10
$\hat{\beta}_4$	-0.129	0.017	-7.791	1.26e-14
$\hat{\beta}_5$	-0.119	0.017	-7.193	1.02e-12
$\hat{\beta}_6$	-0.142	0.017	-8.521	< 2e-16
$\hat{\beta}_7$	-0.245	0.017	-14.745	< 2e-16
$\hat{\beta}_8$	-0.183	0.017	-11.072	< 2e-16
$\hat{\beta}_9$	-0.073	0.022	-3.312	0.001
$\hat{\beta}_{10}$	-0.215	0.022	-9.998	< 2e-16
$\hat{\beta}_{11}$	-0.301	0.022	-13.843	< 2e-16
$\hat{\beta}_{12}$	-0.184	0.022	-8.547	< 2e-16
$\hat{\beta}_{13}$	-0.245	0.022	-11.281	< 2e-16
$\hat{\beta}_{14}$	-0.372	0.022	-17.200	< 2e-16
$\hat{\beta}_{15}$	-0.512	0.022	-23.691	< 2e-16
$\hat{\beta}_{16}$	-0.420	0.022	-19.200	< 2e-16
$\hat{\beta}_{17}$	-0.129	0.022	-5.940	3.56e-09
$\hat{\beta}_{18}$	-0.128	0.022	-5.850	6.08e-09
$\hat{\beta}_{19}$	-0.117	0.022	-5.357	9.82e-08

Tabela 5.6: Estimativas para os coeficientes do autogressivo de ordem 8

<b>Coeficiente</b>	<b>Estimativa</b>	<b>Erro padrão</b>
$\hat{\theta}_1$	0.338	0.026
$\hat{\theta}_2$	0.099	0.028
$\hat{\theta}_3$	0.072	0.028
$\hat{\theta}_4$	0.090	0.028
$\hat{\theta}_5$	0.017	0.028
$\hat{\theta}_6$	0.129	0.028
$\hat{\theta}_7$	0.082	0.028
$\hat{\theta}_8$	-0.060	0.027

Assim sendo, sabe-se que, além de resíduos com estrutura de correlação, a variável  $Y_t^*$

- tende a aumentar, em média, com o aumento do  $t$ ;
- toma, em média, menores valores em dias de feriado, quando comparado com os restantes dias;
- diminui, em média, ao longo da semana, tomando como referência a segunda-feira;
- tende a diminuir, em média, em todos os meses do ano, quando comparados com o registado no mês de janeiro.

### 5.1.3.1 Análise dos Resíduos

Como já se sabe os resíduos do Modelo 3 serão os valores estimados do ruído branco  $\delta_t$ .

Pelo gráfico da figura 5.17 (a) pode então ver-se que pontos estão na sua maioria sobre a reta traçada, pelo que os resíduos seguem uma distribuição normal. Tendo ainda em atenção os gráficos da Figura 5.17 (b) e Figura 5.17 (c), pode ver-se que os pontos se distribuem aleatoriamente em torno do resíduo nulo, formando uma mancha de largura uniforme. Assim sendo, vêem-se cumpridos os pressupostos do modelo.

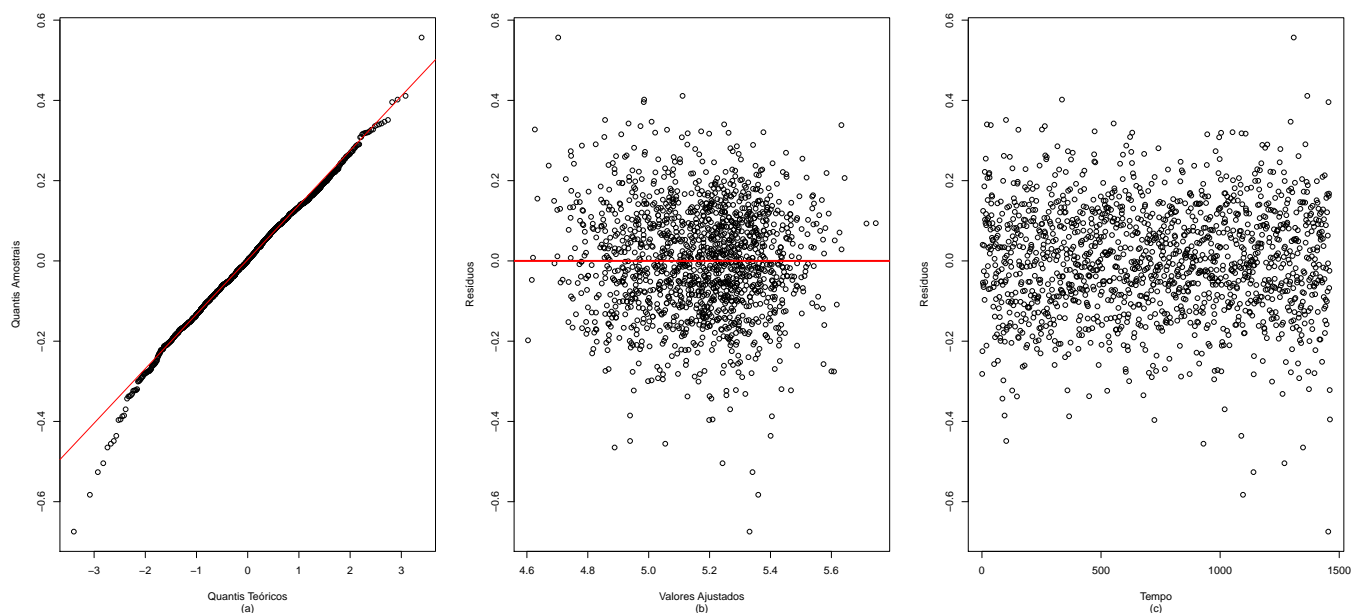


Figura 5.17: Representação gráfica do Normal Q-Q plot (a), dos valores ajustados do modelo 3 *vs* os resíduos (b) e do tempo *vs* os resíduos (c)

### 5.1.3.2 Validação do Modelo

Com objetivo de testar a adequação do modelo aos dados utilizaram-se os dados referentes à admissão de pacientes na urgência geral desde 01/01/2016 a 31/03/2016, tendo-se assim dados para uma totalidade de 91 dias.

Inicialmente fez-se previsão de  $Y$  para o ano de 2016, construindo-se os intervalos de confiança a 95%.

Tendo em conta a representação gráfica da Figura 5.18 vê-se a amarelo os valores previstos pelo modelo para o ano de 2016. As linhas a vermelho representam uma suavização dos valores previstos e do intervalo de confiança (a tracejado). A azul tracejado tem-se os valores que realmente foram registados no primeiro trimestre de 2016.

Vê-se ainda a rosa a previsão que o modelo faz para os anos de 2012, 2013, 2014 e 2015, sendo esta bastante semelhante ao registado verdadeiramente (linha azul).

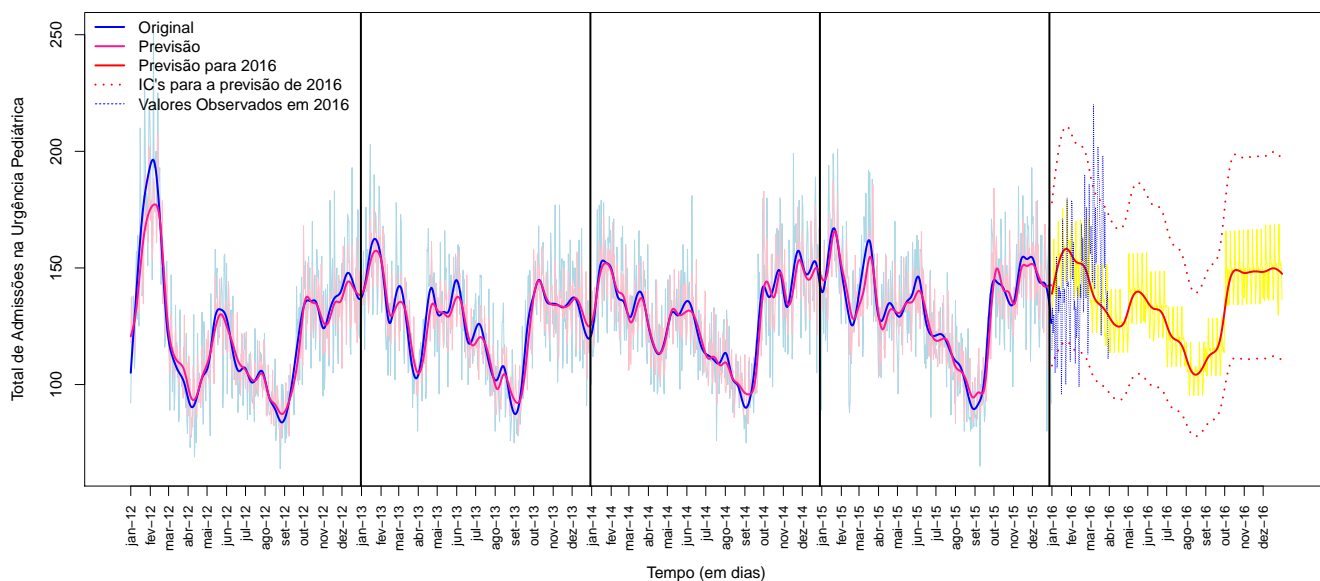


Figura 5.18: Previsão do número de pacientes admitidos na urgência pediátrica, diariamente

Em 80 dos 91 dias a previsão feita pelo modelo cai dentro do intervalo de confiança, sendo que o modelo prevê acertadamente 87.91% dos dias. Em 7 dias apenas o valor observado se encontra acima do limite superior do IC e em 4 dias abaixo do limite inferior. Estes dias são, maioritariamente, pertencentes ao mês de março, à semelhança do que acontece na urgência geral.

No Anexo C são apresentados os valores previstos pelo modelo, com intervalos de confiança e respetivos valores reais observados. Nas tabelas, a cor vermelha estão identificados os dias em que o valor observado ficou acima do limite superior do intervalo de confiança previsto e a cor azul os dias/valores onde o valor observado ficou abaixo do limite inferior do intervalo de confiança previsto.

## 5.2 Internamentos Via Urgência Geral

O objetivo aqui passa por modelar o número de pacientes admitidos no serviço de internamento via serviço de urgência, isto é, modelar o número de pacientes que admitidos na urgência tenham como destino o SI do hospital. Desta forma é possível prever o número de camas necessárias para internamentos via urgência e assim agendar de forma mais acertada os internamentos planeados.

Esta modelação terá como foco apenas os internamentos via urgência geral, já que em ambos os outros serviços não se registam quaisquer problemas de faltas de camas no internamento.

Para a modelação em si usar-se-ão como variáveis explicativas o dia da semana, o mês do ano, a estação do ano, o facto de cada dia ser feriado, ou não, as concentrações atmosféricas de NO<sub>2</sub>, NO, O<sub>3</sub> e P<sub>10</sub>, os registos de temperaturas máximas e mínimas registadas no dia anterior e no próprio dia e também o número de consultas devidas a casos de gripe na semana anterior e na própria semana.

Além disso, observado o gráfico da figura 4.6, parece existir uma ligeira tendência linear na série. Por essa razão será também vista como variável explicativa o tempo, variável que a cada dia faz corresponder um valor temporal, onde o dia 01/01/2012 é tomado como o dia 1 e o dia 31/12/2015 como o dia 1461.

Dado o número de pacientes admitidos no SI via urgência num determinado dia estar dependente do número de pacientes admitidos na urgência nesse mesmo dia, optou-se usar como distribuição para a variável resposta a distribuição de Poisson. No entanto, a modelação terá em conta a taxa de pacientes internados, ao invés de apenas o número de pessoas.

Ou seja  $Y_t \sim Poisson(\mu_t)$  onde  $\mu_t = \frac{\lambda t}{l}$ , em que  $\lambda$  representa o número de pessoas internadas no dia  $t$  e  $l$  o número de pacientes admitidos na urgência nesse mesmo dia  $t$ .

Começou-se então por ajustar o modelo linear generalizado de Poisson com as variável explicativa  $t$  e o termo *offset*, cujo periodograma mostra um pico na frequência zero,

indicando existência de tendência não modelada. Esta tendência viu-se modelada com a introdução da variável mês no modelo. Feito o periodograma pode ver-se a existência de um ciclo. Novamente a introdução de um ciclo harmónico não é compensatória, sendo a introdução da variável dia da semana na parte fixa do modelo suficiente.

A colocação das restantes variáveis no modelo revelaram que apenas  $O_3$  e TMIN são estatisticamente significativas. Ao contrário do que foi concluído dos outros modelos o facto de ser ou não feriado não tem qualquer influência no número de pessoas internadas, o que faz sentido, dado que a gravidade dos casos não está dependente do ser ou não feriado.

Devido à falta de algumas observações nas variáveis  $O_3$  e TMIN o modelo contendo estas variáveis verá 617 observações removidas, o que significa que mais de 40% dos 1461 dias em estudo não serão contabilizados.

Assim sendo, optou-se por ter em conta o modelo com as variáveis explicativa  $t$ , mês e dia da semana, com termo *offset* (Modelo 1) e cujo periodograma pode ser visto na Figura 5.19.

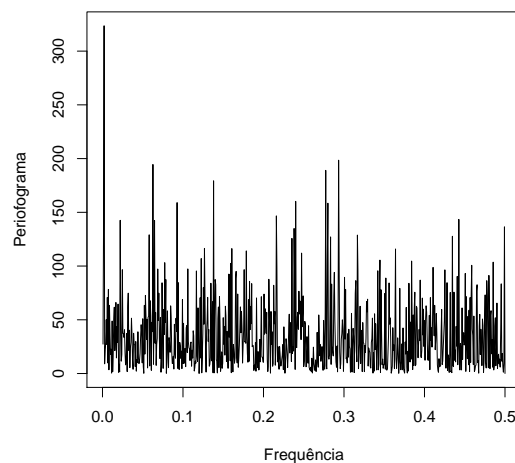


Figura 5.19: Periodograma dos resíduos do Modelo 1

É agora altura de estudar as representações gráficas das funções de correlação e auto-correlação, com objetivo de perceber se os resíduos do modelo têm algum tipo de estrutura

de correlação temporal.

Assim, tendo em atenção os gráficos da Figura 5.20 pode perceber-se que a FAC decai bruscamente para zero e a FACP decai sinusoidalmente para zero, ambas mantendo-se dentro das bandas de confiança. Portanto conclui-se que não existe correlação temporal nos dados.

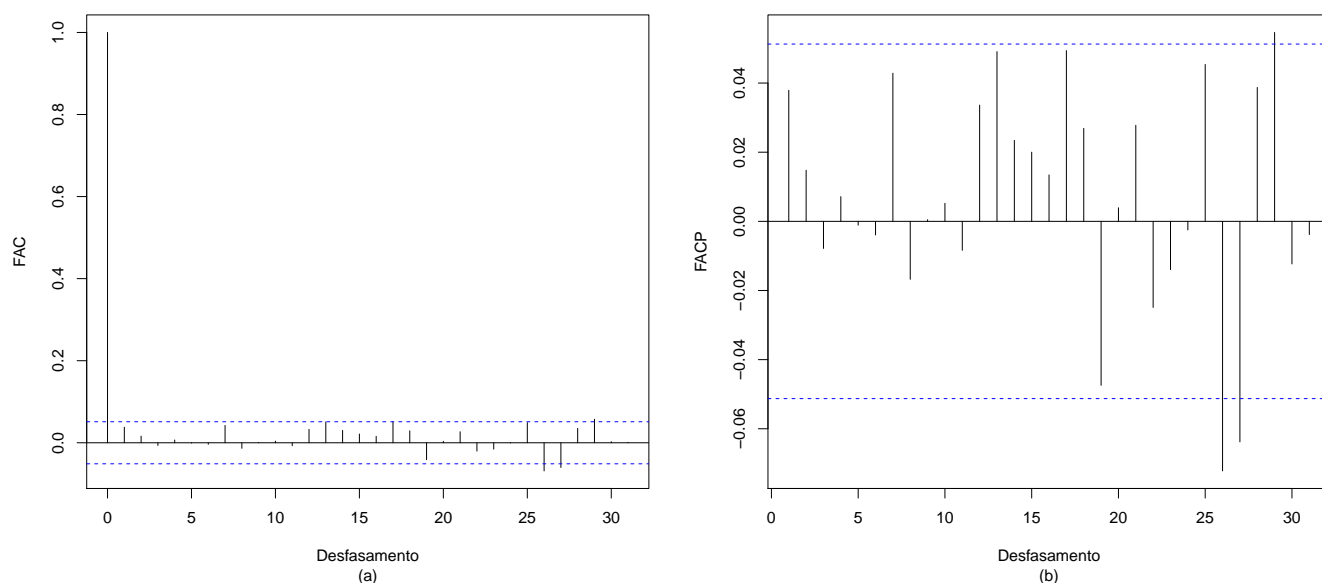


Figura 5.20: FAC (a) e FACP (b) dos resíduos

Desta forma tem-se como modelo para o número de pacientes internados via urgência geral tem-se, com um MAPE de 14.95%, o modelo

$$\begin{aligned} \lambda = & \exp(\log(l) + \beta_0 + \beta_1 t + \beta_2 \text{Terca} + \beta_3 \text{Quarta} + \beta_4 \text{Quinta} + \beta_5 \text{Sexta} + \beta_6 \text{Sabado} \\ & + \beta_7 \text{Domingo} + \beta_8 \text{Fevereiro} + \beta_9 \text{Marco} + \beta_{10} \text{Abril} + \beta_{11} \text{Maio} + \beta_{12} \text{Junho} \\ & + \beta_{13} \text{Julho} + \beta_{14} \text{Agosto} + \beta_{15} \text{Setembro} + \beta_{16} \text{Outubro} + \beta_{17} \text{Novembro} + \beta_{18} \text{Dezembro}) \end{aligned}$$

cujas estimativas para  $\hat{\beta}$  se encontram na tabela 5.7.

Tabela 5.7: Tabela sumária do Modelo 1

<b>Coefficiente</b>	<b>Estimativa</b>	<b>Erro padrão</b>	<b>z</b>	<b>p-valor</b>
$\hat{\beta}_0$	-2.299	0.019	-118.183	< 2e-16
$\hat{\beta}_1$	-4.29e-05	1.11e-05	-3.851	0.0001
$\hat{\beta}_2$	0.071	0.017	4.322	1.55e-05
$\hat{\beta}_3$	0.087	0.016	5.274	1.33e-07
$\hat{\beta}_4$	0.116	0.016	7.120	1.08e-12
$\hat{\beta}_5$	0.110	0.016	6.724	1.77e-11
$\hat{\beta}_6$	0.070	0.017	4.081	4.48e-05
$\hat{\beta}_7$	0.026	0.018	1.475	<b>0.140</b>
$\hat{\beta}_8$	0.003	0.021	0.138	<b>0.890</b>
$\hat{\beta}_9$	-0.050	0.021	-2.344	0.019
$\hat{\beta}_{10}$	-0.084	0.022	-3.840	0.0001
$\hat{\beta}_{11}$	-0.101	0.022	-4.685	2.80e-06
$\hat{\beta}_{12}$	-0.151	0.022	-6.811	9.72e-12
$\hat{\beta}_{13}$	-0.146	0.022	-6.733	1.66e-11
$\hat{\beta}_{14}$	-0.132	0.022	-6.105	1.03e-09
$\hat{\beta}_{15}$	-0.132	0.022	-5.958	2.55e-09
$\hat{\beta}_{16}$	-0.102	0.022	-4.718	2.38e-06
$\hat{\beta}_{17}$	-0.048	0.022	-2.173	0.030
$\hat{\beta}_{18}$	-0.037	0.022	-1.726	<b>0.084</b>

Além do Modelo 1, cujo valor de AIC é de 9353.1 testaram-se também outros modelos, com objetivo de tentar diminuir o erro médio.

Considerou-se a colocação da cor da pulseira da triagem como variável explicativa, obtendo-se um modelo com AIC igual a 27016 e MAPE=14.76% e também a colocação de uma variável explicativa que tinha em consideração a gravidade do paciente (pulseiras azul e verde são não urgentes e amarelo, laranja e vermelho são urgentes), donde se tem um



modelo com AIC igual a 9241.2 e MAPE de 14.95%. Assim sendo, dada a complexidade dos modelos acima descritos *versus* os valores de AIC e MAPE conclui-se que o Modelo 1 é o mais adequado.

## 5.2.1 Análise dos Resíduos

Pelos gráficos da Figura 5.21 (a) e Figura 5.21 (b), pode ver-se que os pontos se distribuem aleatoriamente em torno do resíduo nulo, formando uma mancha de largura uniforme. Assim sendo, vêem-se cumpridos os pressupostos do modelo.

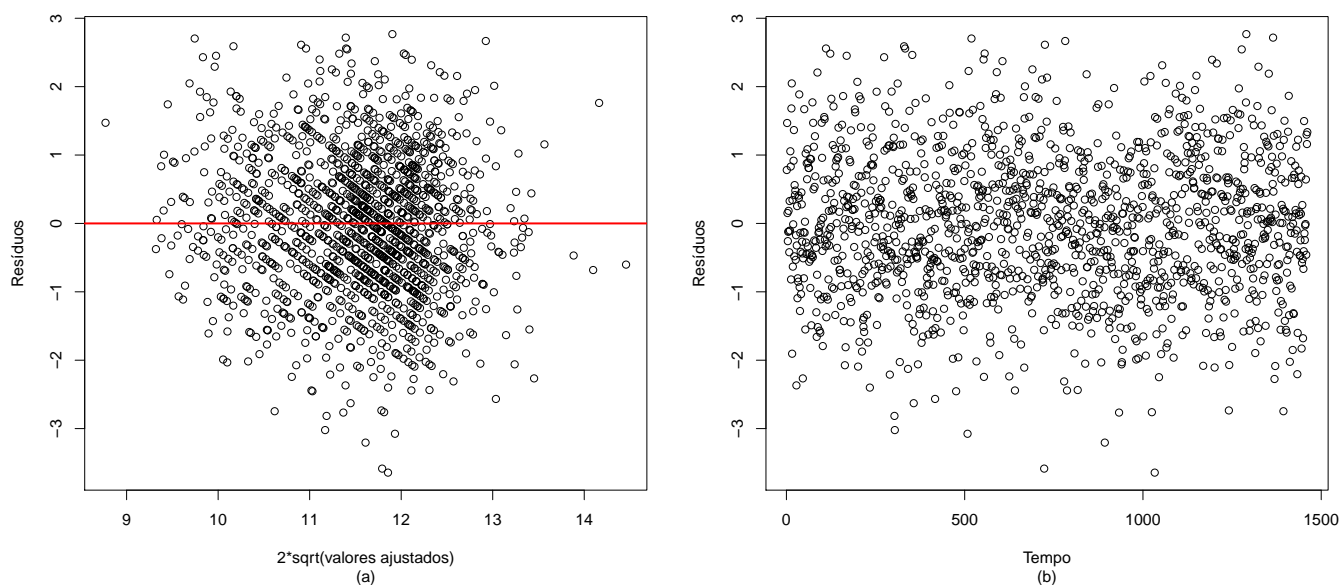


Figura 5.21: Representação gráfica da transformação, proposta por McCullagh e Nelder, dos valores ajustados do modelo 1 *vs* os resíduos (a) e do tempo *vs* os resíduos (b)

### 5.2.1.1 Validação do Modelo

Com objetivo de testar a adequação do modelo aos dados utilizaram-se os dados referentes à admissão de pacientes na urgência geral desde 01/01/2016 a 31/03/2016, tendo-se assim dados para uma totalidade de 91 dias.

Inicialmente fez-se previsão de  $Y$  para o ano de 2016, construindo-se os intervalos de confiança a 95%.

Tendo em conta a representação gráfica da figura 5.18 vê-se a amarelo os valores previstos pelo modelo para o ano de 2016. As linhas a vermelho representam uma suavização dos valores previstos e do intervalo de confiança (a tracejado). A azul tracejado tem-se os valores que realmente foram registados no primeiro trimestre de 2016.

Vê-se ainda a rosa a previsão que o modelo faz para os anos de 2012, 2013, 2014 e 2015, sendo esta bastante semelhante ao registado verdadeiramente (linha azul).

Em 15 dos 91 dias a previsão feita pelo modelo cai dentro do intervalo de confiança, sendo que o modelo prevê acertadamente 16.48% dos dias. Em 29 dias o valor observado se encontra acima do limite superior do IC e em 47 dias abaixo do limite inferior.

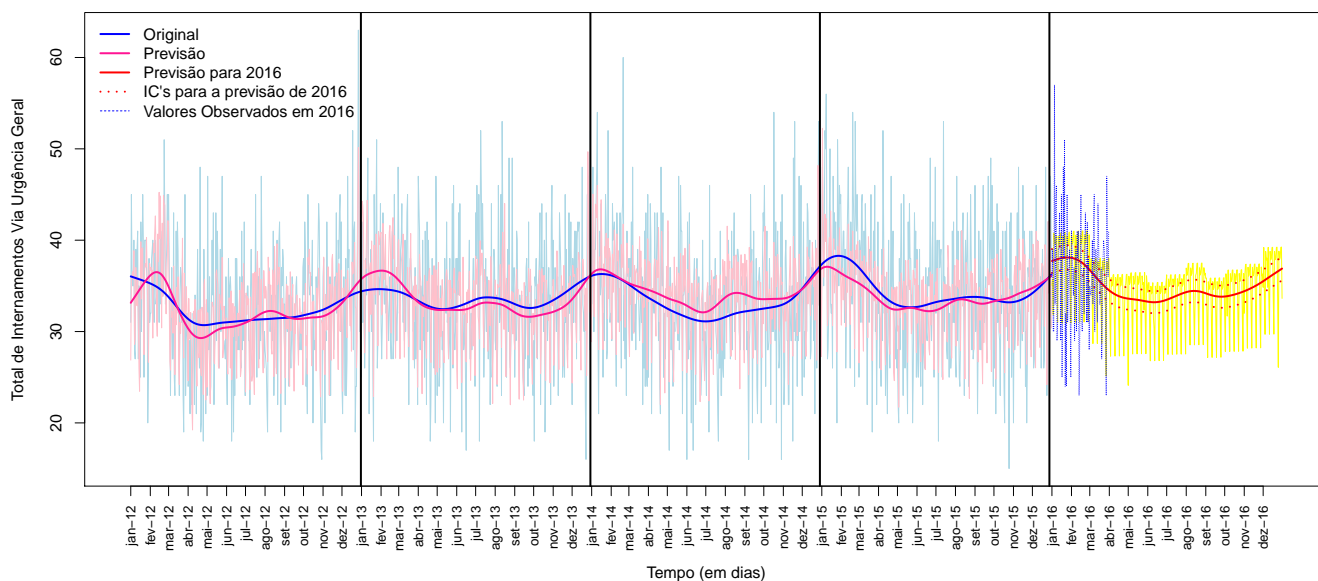


Figura 5.22: Previsão do número de pacientes admitidos no serviço de internamento via urgência geral, diariamente

### Conclusões

---

O número de admissões nos três serviços de urgência do HB têm modelações distintas, apesar de compostos pelas mesmas variáveis explicativas nas suas partes fixas.

Nota-se a existência de correlação temporal ao nível de 7 e 8 dias, o que significa que, em geral, o número de admissões em determinado dia estará correlacionado com o número registado na semana anterior.

Assim sendo, para a modelação do número de admissões nos três serviços de urgência têm-se modelos parcialmente regressivos com resíduos autoregressivos.

A parte fixa de cada um dos três modelos de regressão tem como covariáveis, o dia da semana, o mês do ano, feriados e uma variável  $t$ , que toma o valor 1 para o dia 01/01/2012, aumentando em uma unidade por cada dia que passa. Apesar de algumas variáveis meteorológicas, ambientais e epidemiológicas se revelarem significativas, a sua introdução no modelo não se traduz numa melhoria acentuada do valor preditivo do mesmo, à semelhança do que se registou em outros estudos (Batal *et al.*, 2001; Holleman *et al.*, 1996; Diehl *et al.*, 1981). Assim, optou-se pela sua não inclusão. Desta maneira tem-se modelos mais simples, fáceis de explicar e de aplicar, dado que desta maneira o hospital apenas necessita de dados internos e variáveis sazonais do conhecimento geral.

Para o caso da urgência geral, o modelo final tem resíduos autoregressivos de ordem 7, o que quer dizer que existe correlação semanal dos dados, isto é, o número de admissões de um dia está relacionado com o número de admissões dos sete dias anteriores, com um erro absoluto percentual médio de 3.39%. Este resultado permite, aquando da comparação dos valores previstos com os realmente observados no primeiro trimestre do ano de 2016, ver que o modelo teria previsto, adequadamente, o número de admissões em 87.9% das vezes,

enquanto que teria previsto por defeito 9.9% e por excesso em 2.2%.

Por outro lado, para ambas as urgências obstétrica e pediátrica, os modelos finais têm resíduos autoregressivos de ordem 8, o que significa que o número de admissões de um dia estão relacionado com o número de admissões dos oito dias anteriores, com erros absolutos percentuais médios de 6% e 2.12%, respetivamente, o que, tendo em consideração dados do primeiro trimestre de 2016, teria levado a uma previsão adequada em 85.71% das vezes, por defeito em 12.2% e por excesso em 2.09% na urgência obstétrica e, a uma previsão adequada em 87.91%, por defeito em 7.69% e por excesso em 4.4% na urgência pediátrica.

No entanto, quando o assunto passa a ser a previsão do número de pacientes internados via urgência os modelos já não se comportam da mesma maneira. Neste caso os feriados já não acrescentam qualquer tipo de valor explicativo para o modelo, que tem na sua parte fixa as variáveis t, dia da semana e mês do ano. Os resíduos deste modelo não apresentam qualquer tipo de estrutura de correlação temporal, o que indica que depois de retirada a sazonalidade e tendência não existe nenhuma relação entre o número de pessoas internadas num determinado dia e nos dias anteriores. Este modelo tem um erro absoluto percentual médio de 14.95%, que é bastante alto, quando comparado com os valores obtidos para as admissões nas urgências. Daqui concluiu-se que quando tidos em conta os valores realmente observados para o primeiro trimestre de 2016 a previsão é adequado em apenas 16.48% dos dias, enquanto que ocorre por defeito em 31.87% e por excesso em 51.65%.

À semelhança do que foi visto noutros estudos (Batal *et al.*, 2001; Wargon *et al.*, 2010; Marcilio *et al.*, 2013; Holleman *et al.*, 1996; Diehl *et al.*, 1981; Glass e Friedman, 1977) o número de admissões no SU do HB também tem o seu pico à segunda-feira, decrescendo ao longo da semana, atingindo valores mínimos ao fim de semana. Este, e o facto de se registarem menos visitas ao SU em dias de feriado, levam à mesma conclusão retirada por Batal *et al.* (2001): existem possibilidades de, as pessoas preferirem atrasar a sua visita à urgência a perder o seu tempo de descanso, ou até mesmo de ficarem doentes no decorrer dos excessos cometidos nos feriados e fins de semana.

No que toca ao mês do ano, o comportamento não é o mesmo para os três serviços de urgência.

No caso das urgências geral e obstétrica as admissões são mais frequentes nos meses quentes e não nos meses de inverno, o que contraria o encontrado por Batal *et al.* (2001) e Holleman *et al.* (1996), vindo de encontro ao que Diehl *et al.* (1981) e Glass & Friedman (1977) deduziram.

No caso da urgência pediátrica verifica-se exatamente o contrário.

As admissões nas três urgências do HB foram modeladas com sucesso, tendo sido obtidos bons resultados, já que deram aso a erros de pequena escala. É no entanto necessário continuar a estudar e atualizar cada um dos modelos com dados recentes. Isto é, para prever o número de pacientes que vai chegar, diariamente, a cada uma das urgências do hospital, num determinado mês, é adequado ter em atenção, além dos dados de 2012 a 2015, todos os dados que se conseguirem até à data da estimação, atualizando o modelo com os mesmos, o que possibilitará a estimação de novos parâmetros e uma consequente previsão mais atualizada. Assim, será possível, a longo prazo uma correta alocação dos recursos hospitalares, promovendo tempos de espera mais curtos, melhores cuidados de saúde e consequente satisfação dos pacientes.

Quanto às admissões no SI via urgência é óbvia a existência de padrões que não estão a ser percebidos, variáveis que não estão a ser tidas em conta, ou até mesmo fenómenos para os quais não existe explicação aparente que precisam de ser entendidos, sendo por isso necessário um estudo mais aprofundado e dedicado unicamente a este assunto.

Em termos de trabalhos futuros, seria viável a utilização das horas como variável explicativa, ou até mesmo a condução de um estudo semelhante em termos horários e não diários, com objetivo de tentar prever as admissões nas urgências em cada hora do dia.

De interesse também seria a introdução da região de onde o paciente é oriundo como uma variável explicativa, analisando a sua significância.

## BIBLIOGRAFIA

Araz, O.M., Bentley, D., Muelleman, R.L. (2014). *Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska*. American Journal of Emergency Medicine, 32 (9), 1016-23.

Attia, M.W., Edward, R. (1998). *Effect of weather on the number and the nature of visits to a pediatric ED*. American Journal of Emergency Medicine, 16, 374-5.

Batal, H., Tench, J., McMillan, S., Adams, J., Mehler, P.S. (2001). *Predicting patient visits to an urgent care clinic using calendar variables*. Academic Emergency Medicine, 8, 48-53.

Berkson, J. (1944). *Application of the logistic function to bio-assay*. Journal of the American Statistical Association, 39, 357-365.

Birch, M.W. (1965). *The detection of partial association II: the general case*. Journal of the Royal Statistical Society, Series B, 27, 111-124.

Bliss, C.I. (1935). *The calculation of the dosage-mortality curve*. Annals of Applied Biology, 22, 134- 78.

Box, G.E.P., Cox, D.R. (1964). *An Analysis of Transformations*. Journal of the Royal Statistical Society, Series B, 26 (2), 211-252.

Brockwell, P.J., Davis, R.A. (2002). *Introduction to Time Series and Forecasting, Second Edition*. New York: Springer-Verlag. Second Edition.

- Champion, R., Kinsman, L.D., Lee, G.A., Masman, K.A., May, E.A., Mills, T.M., Taylor, M.D., Thomas, P.R., Williams, R.J. (2007). *Forecasting emergency department presentations*. Australian Health Review: A Publication of the Australian Hospital Association, 31 (1), 83-90.
- Chatfield, C. (1996). *The Analysis on Time Series - An Introduction*. New York: Chapman & Hall. Sixth Edition.
- Chatterjee, S., Hadi, A.S. (2006). *Regression Analysis by Example*. New Jersey: John Wiley & Sons. Fourth Edition.
- Comissão de Reavaliação da Rede Nacional de Emergência/Urgência. (2012). *Reavaliação da Rede Nacional de Emergência/Urgência - Relatório*.
- Craney, T.A., Surles, J.G. (2002). *Model-Dependent Variance Inflation Factor Cutoff Values*. Quality Engineering, 14(3), 391-403.
- Diaz, J., Alberdi, J.C., Pajares, M.S., Lopez, R., Lage, M.B., Otero, A. (2001). *A model for forecasting emergency hospital admissions: effect of environmental variables*. Journal of Environmental Health, 64 (3), 9-15.
- Diehl, A.K., Morris, M.D., Mannis, S.A. (1981). *Use of calendar and weather data to predict walk-in attendance*. Southern Medical Journal, 74, 709-12.
- Diggle, P.J. (2000). *Time Series, A Biostatistical Introduction*. Oxford Statistical Science Series, Oxford University Press.
- Direção Geral de Saúde. (2001). *A Rede de Referência Hospitalar de Urgência/Emergência*.

Dyke, G.V., Patterson, H.D. (1952). *Analysis of factorial arrangements when the data are proportions*. Biometrics, 8, 1-12.

Feigl, P., Zelen, M. (1965). *Estimation of exponential survival probabilities with concomitant information*. Biometrics, 21, 826-838.

Fisher, R.A. (1922). *On the mathematical foundations of theoretical statistics*. The Royal Society, 222, 309-68.

Fuhrmann, C.M., Sugg, M.M., Konrad II, C.E., Walker, A. (2016). *Impact of Extreme Heat Events on Emergency Department Visits in North Carolina (2007–2011)*. Journal of Community Health, 41, 146-156.

Glass, R., Friedmann, D. (1977). *Trends in the demand for emergency room services: the Mount Sinai*. Journal of Emergency Medicine, 15, 397-9.

Glasser, M. (1967). *Exponential survival with covariance*. Journal of the American Statistical Association, 62, 561-568.

Hanley, J.A. (2004). *"Transmuting" Women into Men: Galton's Family Data on Human Stature*. The American Statistician, 58 (3), 237-243.

Hilbe, J.M. (2008). *Negative Binomial Regression*. UK: Cambridge University Press. Second Edition.

Holleman, D.R., Bowling, R.L., Gathy, C. (1996). *Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data*. Journal of General Internal Medicine, 11, 237-9.



Jones, S.S., Thomas, A., Evans, R.S., Welch, S.J., Haug, P.J., Snow, G.L. (2008). *Forecasting Daily Patient Volumes in the Emergency Department*. Academic Emergency Medicine, 15 (2), 159-170.

Marcilio, I., Hajat, S., Gouveia, N. (2013). *Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings*. Academic Emergency Medicine, 20 (8), 769-777.

Matias, C., Oliveira, R., Duarte, R., Bico, P., Mendonça, C., Nuno, L., Almeida, A., Rabaçal, C., Afonso, S. (2008). *Triagem de Manchester nas Síndromes Coronárias Agudas*. Revista Portuguesa de Cardiologia, 27 (2), 205-216.

McCullagh, P., Nelder, J. (1989). *Generalized Linear Models*. London: Chapman & Hall. Second Edition.

Nelder, J., Wedderburn, R. (1972). *Generalized linear models*. Journal of the Royal Statistical Society A, 135, 370–384.

Noble, J.H., La Montagne, M.E., Bellotti, C., Wechsler, H. (1971). *Variations in visits to hospital emergency care facilities: ritualistic and meteorological factors affecting supply and demand*. Medical Care, 9, 415-27.

Rao, C.R., Toutenburg, H., Shalabh, Heumann, C. (2008). *Linear Models and Generalizations - Least Squares and Alternatives*. New York: Springer-Verlag. Third Edition.

Ross, S. (2010). *A first course in probability*. New Jersey: Pearson Education. Eighth Edition.

Turkman, A., Silva, G. (2000). *Modelos Lineares Generalizados - da teoria à prática*. Lis-

boa: Sociedade Portuguesa de Estatística.

Wargon M., Casalino, E., Guidet, B. (2010). *From Model to Forecasting: A Multicenter Study in Emergency Departments*. Society for Academic Emergency Medicine, 17 (9), 970-8.

Wiler, J.L., Griffey, R.T., Olsen, T. (2011), *Review of Modeling Approaches for Emergency Department Patient Flow and Crowding Research*. Academic Emergency Medicine, 18 (12), 1371-1379.

Zeger, S. (1988). *A regression model for time series of counts*. Biometrika, 75, 621-29.

Zippin, C., Armitage, P. (1966). *Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter*. Biometrics, 22, 665- 672.

## Anexo A - Tabela dos Valores Previstos *vs* Reais da Urgência Geral

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-01-01	285.76	333.27	383.83	332
2016-01-02	303.89	353.98	407.26	427
2016-01-03	280.66	329.42	381.43	308
2016-01-04	377.43	432.44	490.57	427
2016-01-05	341.57	394.62	450.86	380
2016-01-06	332.67	385.16	440.87	366
2016-01-07	322.74	374.69	429.88	416
2016-01-08	331.45	384.61	441.07	364
2016-01-09	295.75	346.69	401.00	317
2016-01-10	271.50	320.76	373.44	301
2016-01-11	373.26	429.29	488.58	383
2016-01-12	336.70	390.47	447.55	334
2016-01-13	332.05	385.52	442.32	369
2016-01-14	330.84	384.25	440.99	346
2016-01-15	334.10	387.75	444.73	341
2016-01-16	295.25	346.31	400.75	330
2016-01-17	270.57	319.88	372.62	266
2016-01-18	373.16	429.29	488.69	424
2016-01-19	336.19	390.01	447.14	395
2016-01-20	332.38	385.95	442.85	352
2016-01-21	332.24	385.80	442.69	415
2016-01-22	334.92	388.66	445.72	412
2016-01-23	295.47	346.56	401.05	373
2016-01-24	270.70	320.03	372.79	302

<b>Data</b>	<b>Limite Inferior</b>	<b>Valor Previsto</b>	<b>Limite Superior</b>	<b>Valor Real</b>
2016-01-25	373.40	429.55	488.98	448
2016-01-26	336.35	390.18	447.34	403
2016-01-27	332.66	386.26	443.18	390
2016-01-28	332.66	386.25	443.17	378
2016-01-29	335.28	389.04	446.13	361
2016-01-30	295.72	346.83	401.33	315
2016-01-31	270.91	320.26	373.04	340
2016-02-01	373.43	429.58	489.01	472
2016-02-02	336.37	390.20	447.36	374
2016-02-03	332.69	386.29	443.22	376
2016-02-04	332.71	386.30	443.23	369
2016-02-05	335.32	389.09	446.18	366
2016-02-06	295.74	346.86	401.36	335
2016-02-07	270.93	320.28	373.06	318
2016-02-08	373.68	429.84	489.29	431
2016-02-09	296.18	347.33	401.86	315
2016-02-10	332.92	386.53	443.48	416
2016-02-11	332.94	386.55	443.49	397
2016-02-12	335.56	389.34	446.45	371
2016-02-13	295.96	347.09	401.61	350
2016-02-14	271.14	320.51	373.30	304
2016-02-15	373.91	430.10	489.56	430
2016-02-16	336.83	390.69	447.88	378
2016-02-17	333.15	386.78	443.73	381
2016-02-18	333.17	386.79	443.75	344
2016-02-19	335.78	389.58	446.70	408
2016-02-20	296.17	347.32	401.85	350

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-02-21	271.35	320.73	373.54	367
2016-02-22	374.15	430.35	489.82	417
2016-02-23	337.05	390.93	448.13	394
2016-02-24	333.38	387.02	443.99	389
2016-02-25	333.39	387.03	444.00	391
2016-02-26	336.01	389.82	446.96	398
2016-02-27	296.39	347.55	402.10	368
2016-02-28	271.56	320.95	373.78	301
2016-02-29	374.39	430.60	490.08	437
2016-03-01	328.66	382.00	438.67	416
2016-03-02	325.03	378.12	434.56	378
2016-03-03	325.04	378.14	434.58	414
2016-03-04	327.63	380.90	437.50	367
2016-03-05	288.45	339.05	393.06	385
2016-03-06	263.90	312.73	365.01	338
2016-03-07	365.62	421.29	480.25	427
2016-03-08	328.89	382.23	438.92	413
2016-03-09	325.25	378.36	434.81	384
2016-03-10	325.26	378.37	434.83	403
2016-03-11	327.85	381.13	437.75	404
2016-03-12	288.66	339.28	393.30	395
2016-03-13	264.11	312.95	365.24	383
2016-03-14	365.85	421.54	480.52	496
2016-03-15	329.11	382.47	439.17	452
2016-03-16	325.47	378.59	435.06	460
2016-03-17	325.49	378.61	435.08	410
2016-03-18	328.08	381.37	438.00	396

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-03-19	288.87	339.50	393.54	328
2016-03-20	264.31	313.17	365.48	372
2016-03-21	366.08	421.79	480.78	446
2016-03-22	329.33	382.71	439.42	430
2016-03-23	325.69	378.83	435.31	426
2016-03-24	325.71	378.85	435.33	407
2016-03-25	288.32	338.92	392.92	394
2016-03-26	289.08	339.73	393.78	364
2016-03-27	228.14	274.22	323.80	275
2016-03-28	366.32	422.03	481.04	411
2016-03-29	329.56	382.95	439.68	428
2016-03-30	325.92	379.07	435.56	397
2016-03-31	325.93	379.09	435.58	447

## Anexo B - Tabela dos Valores Previstos *vs* Reais da Urgência Obs-tétrica

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-01-01	21.65	32.74	48.38	23
2016-01-02	18.18	27.79	41.44	35
2016-01-03	15.40	23.77	35.77	42
2016-01-04	33.04	48.83	70.68	44
2016-01-05	28.66	42.70	62.24	34
2016-01-06	25.97	38.91	57.00	37
2016-01-07	25.67	38.50	56.46	31
2016-01-08	27.41	41.18	60.47	44
2016-01-09	17.38	26.83	40.38	44
2016-01-10	15.29	23.79	36.06	26
2016-01-11	31.67	47.22	68.88	33
2016-01-12	27.92	41.92	61.54	34
2016-01-13	25.58	38.61	56.94	30
2016-01-14	25.71	38.79	57.20	38
2016-01-15	27.70	41.63	61.16	47
2016-01-16	17.30	26.74	40.26	34
2016-01-17	15.23	23.72	35.98	38
2016-01-18	31.47	46.96	68.54	40
2016-01-19	27.76	41.71	61.28	29
2016-01-20	25.53	38.55	56.87	40
2016-01-21	25.73	38.83	57.27	32
2016-01-22	27.76	41.71	61.28	32
2016-01-23	17.30	26.73	40.26	27
2016-01-24	15.22	23.71	35.95	35

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-01-25	31.44	46.91	68.49	32
2016-01-26	27.73	41.68	61.23	34
2016-01-27	25.52	38.54	56.86	45
2016-01-28	25.74	38.84	57.29	28
2016-01-29	27.77	41.73	61.31	47
2016-01-30	17.30	26.74	40.27	36
2016-01-31	15.22	23.71	35.96	37
2016-02-01	29.03	43.52	63.79	48
2016-02-02	25.57	38.61	56.96	48
2016-02-03	23.51	35.68	52.86	40
2016-02-04	23.72	35.97	53.27	24
2016-02-05	25.62	38.68	57.05	32
2016-02-06	15.87	24.66	37.31	32
2016-02-07	13.94	21.83	33.27	28
2016-02-08	29.04	43.53	63.80	39
2016-02-09	21.33	32.55	48.48	41
2016-02-10	23.52	35.69	52.88	67
2016-02-11	23.73	35.98	53.29	29
2016-02-12	25.63	38.69	57.07	45
2016-02-13	15.88	24.67	37.32	35
2016-02-14	13.94	21.84	33.28	33
2016-02-15	29.05	43.55	63.83	42
2016-02-16	25.59	38.64	57.00	34
2016-02-17	23.53	35.70	52.90	57
2016-02-18	23.74	36.00	53.31	43
2016-02-19	25.64	38.71	57.10	25
2016-02-20	15.88	24.68	37.34	29



Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-02-21	13.95	21.85	33.29	21
2016-02-22	29.07	43.57	63.85	51
2016-02-23	25.60	38.65	57.02	35
2016-02-24	23.54	35.72	52.92	43
2016-02-25	23.75	36.01	53.33	39
2016-02-26	25.65	38.72	57.12	55
2016-02-27	15.89	24.69	37.35	26
2016-02-28	13.96	21.86	33.31	22
2016-02-29	29.08	43.58	63.88	42
2016-03-01	25.68	38.76	57.17	33
2016-03-02	23.61	35.82	53.06	39
2016-03-03	23.82	36.11	53.47	48
2016-03-04	25.73	38.83	57.27	47
2016-03-05	15.94	24.76	37.46	41
2016-03-06	14.00	21.92	33.40	36
2016-03-07	29.17	43.71	64.04	34
2016-03-08	25.69	38.78	57.20	31
2016-03-09	23.62	35.84	53.08	47
2016-03-10	23.83	36.13	53.49	43
2016-03-11	25.74	38.85	57.29	36
2016-03-12	15.95	24.77	37.48	34
2016-03-13	14.01	21.93	33.42	56
2016-03-14	29.18	43.72	64.07	38
2016-03-15	25.70	38.80	57.22	44
2016-03-16	23.63	35.85	53.11	40
2016-03-17	23.84	36.14	53.52	28
2016-03-18	25.75	38.87	57.32	55

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-03-19	15.96	24.78	37.49	45
2016-03-20	14.02	21.94	33.43	42
2016-03-21	29.19	43.74	64.09	33
2016-03-22	25.71	38.81	57.24	31
2016-03-23	23.65	35.87	53.13	49
2016-03-24	23.85	36.16	53.54	37
2016-03-25	21.48	32.77	48.79	27
2016-03-26	15.97	24.80	37.51	34
2016-03-27	11.50	18.23	28.09	22
2016-03-28	29.20	43.76	64.12	35
2016-03-29	25.72	38.83	57.26	46
2016-03-30	23.66	35.88	53.15	45
2016-03-31	23.86	36.17	53.56	22

## Anexo C - Tabela dos Valores Previstos *vs* Reais da Urgência Pediátrica

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-01-01	101.23	128.32	162.41	115
2016-01-02	105.15	135.01	173.06	133
2016-01-03	108.25	139.67	179.89	122
2016-01-04	125.61	162.45	209.71	130
2016-01-05	114.05	148.20	192.21	123
2016-01-06	106.76	139.10	180.90	105
2016-01-07	111.27	145.77	190.59	134
2016-01-08	113.96	150.24	197.66	155
2016-01-09	103.59	136.91	180.58	107
2016-01-10	108.90	144.10	190.27	110
2016-01-11	128.44	169.99	224.51	142
2016-01-12	117.27	155.53	205.81	131
2016-01-13	114.97	152.66	202.27	142
2016-01-14	117.40	156.06	206.99	137
2016-01-15	115.82	154.11	204.60	140
2016-01-16	105.81	140.96	187.37	96
2016-01-17	111.91	149.10	198.20	113
2016-01-18	131.85	175.51	233.12	171
2016-01-19	120.62	160.74	213.73	150
2016-01-20	118.40	157.85	209.98	135
2016-01-21	119.85	159.82	212.63	141
2016-01-22	117.75	157.06	209.04	126
2016-01-23	107.73	143.82	191.57	100
2016-01-24	113.88	151.99	202.40	121

<b>Data</b>	<b>Limite Inferior</b>	<b>Valor Previsto</b>	<b>Limite Superior</b>	<b>Valor Real</b>
2016-01-25	133.97	178.60	237.56	180
2016-01-26	122.49	163.43	217.55	142
2016-01-27	120.04	160.20	213.31	144
2016-01-28	121.26	161.82	215.46	133
2016-01-29	119.04	158.90	211.61	125
2016-01-30	108.87	145.43	193.82	110
2016-01-31	114.99	153.54	204.55	146
2016-02-01	126.82	169.21	225.25	179
2016-02-02	115.85	154.69	206.08	178
2016-02-03	113.45	151.51	201.88	145
2016-02-04	114.52	152.93	203.76	161
2016-02-05	112.37	150.09	200.01	132
2016-02-06	102.71	137.28	183.08	136
2016-02-07	108.44	144.88	193.13	109
2016-02-08	127.45	170.06	226.40	133
2016-02-09	105.00	140.32	187.10	120
2016-02-10	113.92	152.15	202.74	125
2016-02-11	114.96	153.53	204.57	142
2016-02-12	112.78	150.64	200.74	120
2016-02-13	103.06	137.75	183.70	99
2016-02-14	108.78	145.33	193.73	123
2016-02-15	127.81	170.55	227.05	143
2016-02-16	116.70	155.83	207.60	137
2016-02-17	114.21	152.53	203.24	169
2016-02-18	115.23	153.88	205.04	153
2016-02-19	113.02	150.96	201.17	163
2016-02-20	103.26	138.02	184.06	131

Data	Limite Inferior	Valor Previsto	Limite Superior	Valor Real
2016-02-21	108.98	145.61	194.09	142
2016-02-22	128.04	170.84	227.45	190
2016-02-23	116.89	156.08	207.94	160
2016-02-24	114.38	152.76	203.56	137
2016-02-25	115.40	154.11	205.33	176
2016-02-26	113.18	151.17	201.45	145
2016-02-27	103.40	138.20	184.30	113
2016-02-28	109.12	145.79	194.33	127
2016-02-29	128.19	171.04	227.71	186
2016-03-01	103.28	138.04	184.09	162
2016-03-02	101.05	135.09	180.18	156
2016-03-03	101.95	136.28	181.76	168
2016-03-04	99.98	133.67	178.30	157
2016-03-05	91.30	122.16	163.07	144
2016-03-06	96.37	128.88	171.96	174
2016-03-07	113.27	151.29	201.61	220
2016-03-08	103.37	138.16	184.25	181
2016-03-09	101.13	135.20	180.33	141
2016-03-10	102.03	136.39	181.90	176
2016-03-11	100.05	133.77	178.43	152
2016-03-12	91.37	122.25	163.19	156
2016-03-13	96.44	128.98	172.09	171
2016-03-14	113.35	151.39	201.75	202
2016-03-15	103.44	138.26	184.37	188
2016-03-16	101.21	135.30	180.45	188
2016-03-17	102.10	136.48	182.02	169
2016-03-18	100.12	133.86	178.55	156

<b>Data</b>	<b>Limite Inferior</b>	<b>Valor Previsto</b>	<b>Limite Superior</b>	<b>Valor Real</b>
2016-03-19	91.43	122.33	163.29	121
2016-03-20	96.50	129.06	172.20	150
2016-03-21	113.42	151.49	201.88	198
2016-03-22	103.50	138.34	184.49	198
2016-03-23	101.27	135.38	180.57	177
2016-03-24	102.16	136.56	182.13	150
2016-03-25	90.34	120.88	161.37	174
2016-03-26	91.49	122.41	163.39	141
2016-03-27	87.07	116.54	155.62	138
2016-03-28	113.49	151.58	201.99	132
2016-03-29	103.57	138.43	184.60	140
2016-03-30	101.33	135.46	180.67	111
2016-03-31	102.22	136.64	182.24	120