



**Universidade de Évora - Escola de Ciências e Tecnologia**

**Mestrado em Modelação Estatística e Análise de Dados**

Dissertação

**Modelação Estatística da Afluência ao Serviço de Urgência  
do HESE**

Loide Raquel Lopes Ascenso

Orientador(es) | Gonçalo João Jacinto

Paulo de Jesus Infante dos Santos

Évora 2023

---

---

---

---



**Universidade de Évora - Escola de Ciências e Tecnologia**

**Mestrado em Modelação Estatística e Análise de Dados**

Dissertação

**Modelação Estatística da Afluência ao Serviço de Urgência  
do HESE**

**Loide Raquel Lopes Ascenso**

Orientador(es) | Gonçalo João Jacinto

Paulo de Jesus Infante dos Santos

Évora 2023

---

---

---

---



A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente | Anabela Afonso (Universidade de Évora)

Vogais | Gonçalo João Jacinto (Universidade de Évora) (Orientador)  
Russell Alpizar-Jara (Universidade de Évora) (Arguente)



## **Agradecimentos**

É frequente encontrarmos uma página de agradecimentos em estudos académicos e dissertações. Esta, no entanto, é mais do que uma banalidade protocolar: é um gesto – simples, mas genuíno – de reconhecimento pela preciosa ajuda e pelo apoio de que beneficiei no decurso deste estudo realizado no HESE.

Antes de mais, dedico uma palavra especial aos meus orientadores, Professor Doutor Paulo Infante e Professor Doutor Gonçalo Jacinto, sem cuja colaboração este estudo não seria possível. A disponibilidade, atenção, amizade e valia científica são aritmeticamente inestimáveis.

Uma palavra também muito especial ao Hugo Quintino, durante o período de recolha dos dados e acompanhamento, bem como partilha da sua experiência e de conhecimentos adquiridos por anos de experiência profissional no HESE.

À minha amiga Sandra Morgadinho pelo apoio e partilha do seu conhecimento na área da saúde.

Aos meus pais pelo carinho e incentivos durante este percurso.

A todos estes – e também àqueles cujos nomes não estão citados, mas direta ou indiretamente contribuíram com o seu apoio e auxílio para a realização deste estudo – duas pequenas grandes palavras: muito obrigada.

<b>Índice</b>	
<b>Índice de Tabelas</b>	<b>V</b>
<b>Índice de Figuras</b>	<b>VIII</b>
<b>Lista de abreviaturas</b>	<b>XI</b>
<b>Resumo</b>	<b>XII</b>
<b>Abstract</b>	<b>XIII</b>
<b>1 Introdução .....</b>	<b>1</b>
<b>1.1 Motivação e objetivos do estudo .....</b>	<b>1</b>
<b>1.2 Enquadramento e Motivação .....</b>	<b>3</b>
<b>1.3 Contribuições.....</b>	<b>3</b>
<b>1.4 Organização da tese .....</b>	<b>4</b>
<b>2 Revisão de literatura .....</b>	<b>5</b>
<b>3 O Hospital Espírito Santo – Évora, E.P.E.....</b>	<b>11</b>
<b>4 Metodologia estatística .....</b>	<b>13</b>
<b>4.1 Modelos Lineares Generalizados .....</b>	<b>13</b>
<b>4.1.1. Análise de Resíduos e Multicolinearidade.....</b>	<b>14</b>
<b>4.1.2 O Modelo de Regressão Logística .....</b>	<b>15</b>
<b>4.2 Controlo Estatístico de Qualidade .....</b>	<b>21</b>
<b>4.2.1 Conceito de Variabilidade .....</b>	<b>22</b>
<b>4.2.2 Noção de Controlo Estatístico do Processo .....</b>	<b>23</b>
<b>4.2.2.1 Cartas de controlo .....</b>	<b>23</b>
<b>4.2.3 Outras ferramentas do controlo estatístico de qualidade .....</b>	<b>29</b>
<b>5 Base de dados utilizada e metodologia.....</b>	<b>33</b>
<b>5.1 Caracterização da população atendida no Serviço de Urgência.....</b>	<b>34</b>
<b>5.2 Caracterização da procura pelo Serviço de Urgência - HESE.....</b>	<b>35</b>
<b>5.3 Caracterização da procura do Serviço de Urgência por tipo de utilizador .....</b>	<b>53</b>
<b>5.4 Caracterização da afluência ao Serviço de Urgência 2018 – 2021 .....</b>	<b>56</b>
<b>5.5 Caracterização do hiperutilizador 2018 – 2021 .....</b>	<b>67</b>
<b>5.6 Modelos de regressão logística .....</b>	<b>69</b>

<b>5.6.2 Diagnóstico do modelo .....</b>	<b>74</b>
<b>5.7 Ajustamento do modelo GAM .....</b>	<b>75</b>
<b>5.8 Ajustamento de um Modelo Linear Generalizado para o Hiperutilizador – período 1 de janeiro de 2018 a 10 de Março de 2020. ....</b>	<b>77</b>
<b>5.8.1 Análise Multivariada.....</b>	<b>79</b>
<b>5.9 Controlo dos códigos de Triagem no serviço 1 de Maio de 2020 até 31 de Maio de 2021</b>	<b>87</b>
<b>6 Conclusões .....</b>	<b>97</b>
<b>7 Bibliografia.....</b>	<b>99</b>

## Índice de tabelas

Tabela 1: Exemplos de estudos realizados no âmbito da modelação .....	10
Tabela 2:Resumo da informação relacionada com a Triagem de Manchester .....	12
Tabela 3: Descrição das variáveis da Base de Dados .....	33
Tabela 4: Tabela sumária da variável Idade (2018 e 2019).....	34
Tabela 5: Tabela sumária de valores percentuais e contagens (2018 e 2019).....	36
Tabela 6: Tabela de contagens e percentagens das admissões ao SU pelo tipo de dia em relação ao feriado.....	42
Tabela 7: Tabela de contagens e percentagens das admissões ao SU pelo tipo de dia em relação ao feriado.....	43
Tabela 8: Tabela de contagens e percentagens das admissões ao SU por turno .....	43
Tabela 9: Tabela de contagens e percentagens das admissões ao SU por estação do ano ..	43
Tabela 10: Distribuição da cor das pulseiras por idade - 2018.....	44
Tabela 11: Distribuição da cor das pulseiras por por idade - 2019 .....	44
Tabela 12: Distribuição da idade do utente por cor do protocolo e por género, 2018 .....	45
Tabela 13: Distribuição da idade do utente por cor do protocolo e por género, 2019 .....	45
Tabela 14: Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a idade por cor da triagem .....	46
Tabela 15: Distribuição das temperaturas mínima e máxima por dia da semana.....	48
Tabela 16: Distribuição das temperaturas mínima e máxima (2018 e 2019) .....	48
Tabela 17: Distribuição das temperaturas mínima por cor da pulseira, 2018 .....	49
Tabela 18: Distribuição das temperaturas máxima por cor da pulseira, 2018.....	49
Tabela 19: Distribuição das temperaturas mínima por cor da pulseira, 2019 .....	50
Tabela 20: Distribuição das temperaturas máxima por cor da urgência, 2019.....	50
Tabela 21: Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a temperatura máxima por cor da triagem .....	51
Tabela 22:Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a temperatura mínima por cor da triagem.....	51
Tabela 23: Idade média por tipo de utilizador .....	53
Tabela 24: Percentagem de admissões por cor, por tipo de utilizador .....	54
Tabela 25: Média da proporção de admissões ao SU por tipo de utilizador, por Turno .....	54
Tabela 26: Número de admissões ao SU por estação do ano, por tipo de utilizador .....	55
Tabela 27: Número médio de admissões ao SU por mês, por tipo de utilizador .....	55



Tabela 28: Número médio de admissões ao SU por tipo de dia, por tipo de utilizador .....	56
Tabela 29: Número médio de admissões ao SU por sexo, por tipo de utilizador.....	56
Tabela 30: Estatísticas descritivas da variável Idade por Sexo .....	57
Tabela 31: Motivos de admissão ao SU de caráter urgente.....	59
Tabela 32: Valores percentuais dos tempos de espera de acordo com o código de triagem e tempos de espera correspondentes (vermelho – 0 minutos; laranja – 10 minutos; amarelo – 50 minutos; verde – 120 minutos; azul – 240 minutos) .....	65
Tabela 33: Estatística descritiva da afluência ao SU por ano e por tipo de utilizador .....	68
Tabela 34: Estatística descritiva e valor p para o teste de Mann-Whitney, efetuada aos tempos de espera por tipo de utilizador.....	69
Tabela 35: Variáveis analisadas - classificação.....	69
Tabela 36: Medidas descritivas para as variáveis contínuas (médias, desvios-padrão, máximo e mínimo).....	71
Tabela 37: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas .....	71
Tabela 38: Coeficientes estimados ( $\beta$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95% para os diferentes modelos univariados, 2018-2021.....	72
Tabela 39: Coeficientes estimados ( $\beta$ ) do modelo de regressão logística multivariado para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p (Wald) associados, para o modelo multivariado preliminar, 2018-2021 .....	73
Tabela 40: Frequências absolutas para a variável idade agrupada em classes .....	74
Tabela 41: Frequências absolutas das classes reagrupadas .....	75
Tabela 42: Coeficientes estimados ( $\beta$ ) do modelo GAM para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p associados, estimativas dos termos suavizados, graus de liberdade (edf) e valores do teste F, 2018-2021. ....	76
Tabela 43: Estimativas para os termos suavizados, graus de liberdade (edf), dimensão (k'), valores de proximidade e graus de liberdade (k-index) e valores p. ....	77
Tabela 44: Medidas descritivas para as variáveis contínuas .....	78
Tabela 45: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas .....	78

Tabela 46: Coeficientes estimados ( $\beta$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95%, 2018-2020 .....	79
Tabela 47: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas .....	81
Tabela 48: Medidas descritivas para as variáveis contínuas .....	82
Tabela 49: Coeficientes estimados ( $\beta$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95%, correspondentes ao modelo que usou os dados de 2018-2020.....	82
Tabela 50: Totais para as categorias da variável indicador .....	83
Tabela 51: Coeficientes estimados ( $\beta$ ) do modelo de regressão logística multivariado para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\sigma\beta$ ), valores p (Wald) associados.....	84
Tabela 52: Fases da modelação, classes do modelo e limitações .....	86

## Índice de Figuras

Figura 1: Número de visitas por 100 mil habitantes ao Serviço de Urgência, 2001 e 2011 (Fonte: Relatório grupo de trabalho serviço de urgência, 2019) .....	7
Figura 2: <i>Spline</i> cúbica é uma curva construída a partir de seções de polinómios cúbicos (Wood, 2006, Fig. 3.3, p.122) .....	20
Figura 3: Curvas de suavização obtidas por <i>splines</i> de regressão cúbica usando 3 nós, 4 nós, 5 nós, etc., e 10 nós. Ambos os pontos finais também contam como nós (Zuur, 2009, Fig. 3.9, p. 48) .....	20
Figura 4: Carta de controlo Shewhart.....	24
Figura 5: Metodologia de controlo em dados agrupados – “ <i>Drilling down into into aggregate data</i> ”. (Carey, R. e Stake, L. (2003)) .....	29
Figura 6: Diagrama de Pareto para os defeitos de um tanque .....	30
Figura 7: Exemplos de Diagramas de Pareto.....	31
Figura 8: Boxplot para a idade dos utentes por sexo, 2018.....	35
Figura 9: Boxplot para a idade dos utentes por sexo, 2019 .....	35
Figura 10: Representação gráfica das admissões por dia da semana, 2018 .....	37
Figura 11: Representação gráfica das admissões por dia da semana, 2019 .....	37
Figura 12: Chegada mensal de utentes ao SU (2018 - 2019) .....	38
Figura 13: Chegada mensal de utentes ao SU por ano .....	39
Figura 14: Chegada por hora de utentes ao SU (2018 - 2019) .....	39
Figura 15: Chegada dos utentes ao SU por hora, 2018, 2019 e média 2018 e 2019 .....	40
Figura 16: Número de admissões de acordo com a cor da urgência .....	41
Figura 17: Boxplot da Idade por cor da triagem, por ano .....	41
Figura 18: Distribuição das cores por género 2018.....	42
Figura 19: Distribuição das cores por género 2019.....	42
Figura 20: Afluência ao SU de acordo com o dia .....	47
Figura 21: Distribuição das temperaturas mínima e máxima em 2018 e 2019 .....	47
Figura 22: Afluência por hora e por cor .....	52
Figura 23: Afluência por cor e por turno .....	52
Figura 24: Afluência por utilizador .....	53
Figura 25: Boxplot da variável idade por sexo.....	57
Figura 26: Percentagem de utilizadores por sexo.....	57
Figura 27: Motivo de admissão ao SU .....	58

Figura 28: Motivos de admissão ao SU de caráter pouco ou nada urgente.....	58
Figura 29: Afluência o SU por ano.....	60
Figura 30: Afluência ao SU por ano nos primeiros cinco meses do ano.....	60
Figura 31: Boxplot da idade por ano .....	61
Figura 32: Afluência o SU por Turno.....	61
Figura 33: Afluência ao SU por hora .....	61
Figura 34: Afluência ao SU por dia da semana .....	62
Figura 35: Classificação por tipo de dia .....	62
Figura 36: Afluência ao SU por tipo de dia e por código de urgência .....	63
Figura 37: Histograma dos tempos de espera.....	64
Figura 38: Tempos de espera por intervalos.....	64
Figura 39: % de episódios de acordo com a classificação do Protocolo de Manchester.....	65
Figura 40: Concelhos com % acima de 1% .....	66
Figura 41: Afluência ao SU conforme pertence ou não ao distrito de Évora.....	66
Figura 42: Afluência de acordo com a zona geográfica .....	67
Figura 43: Afluência ao SU por tipo de utilizador .....	68
Figura 44: Linearidade com o logit: i) idade; ii) tempo de espera; iii) indicador.....	83
Figura 45: Carta U para códigos vermelhos por período horário (Rational Ordering) .....	87
Figura 46: Carta U para códigos vermelhos por período horário (Rational subgrouping)..	88
Figura 47: Carta U para códigos laranjas por período horário (Rational Ordering).....	88
Figura 48: Carta U para códigos laranjas por período horário (Rational sugrouping) .....	89
Figura 49: Carta U para códigos amarelos por período horário (Rational Ordering).....	89
Figura 50: Carta U para códigos amarelos por período horário (Rational subgrouping) ...	90
Figura 51: Carta U para códigos verdes por período horário (Rational Ordering).....	90
Figura 52: Carta U para códigos verdes por período horário (Rational subgrouping) .....	91
Figura 53: Carta U para códigos laranjas por dia da semana (Rational Ordering).....	92
Figura 54: Cartas U para códigos laranjas para cada dia da semana.....	92
Figura 55: Carta U para códigos amarelos por dia da semana (Rational Ordering).....	93
Figura 56: Cartas U para códigos amarelos para cada dia da semana.....	93
Figura 57: Carta U para códigos verdes por dia da semana (Rational Ordering).....	94
Figura 58: Cartas U para códigos verdes para cada dia da semana.....	94
Figura 59: Boxplot para os tempos de espera por cor do código de triagem .....	95
Figura 60: Carta P para a proporção de tempos de espera acima do especificado .....	96
Figura A61.1: <i>Receiver Operating Characteristic</i> (ROC) do modelo ajustado .....	108

## **Lista de abreviaturas**

**ADW** Sistema de Informação e Gestão do Hospital

**AIQ** Amplitude interquartil

**GAM** Modelos Aditivos Generalizados

**GLM** Modelos Lineares Generalizados

**HESE** Hospital do Espírito Santo de Évora

**HU** Hiperutilizador

**LC** Linha Central

**LIC** Limite Inferior de Controlo

**LSC** Limite Superior de Controlo

**K-S** Teste de Kolmogorov-Smirnov

**RORS** Rational Ordering ou Rational Subgrouping

**SU** Serviço de Urgência



## Resumo

Os hospitais públicos apresentam-se como sistemas complexos nos quais as suas áreas de intervenção se dividem de acordo com as especialidades médicas. Estas contam com o apoio de equipamentos tecnológicos para a resolução de problemas de saúde identificados nos pacientes que diariamente se apresentam neste serviço.

O serviço de urgência garante a oferta contínua e diferenciada de cuidados de saúde a toda a população, durante todos os dias do ano. Pelo seu caráter de fácil acesso, a população opta frequentemente por resolver os seus problemas de saúde, independentemente da sua urgência, nos serviços de urgência resultando num excesso de procura para o qual, por vezes, o serviço de urgência não foi inicialmente dimensionado.

A grande afluência de utentes ao serviço de urgência dos hospitais constitui um problema que afeta não apenas responsáveis pelo serviço como os próprios utentes, pelos problemas de gestão, escassez dos recursos disponíveis, humanos e materiais, mas principalmente no aumento dos tempos de espera dos pacientes podendo comprometer a qualidade do atendimento e segurança do utente.

Este projeto visa o estudo desta problemática que abrange as chegadas totais de utentes ao serviço de urgência, modelação da afluência ao serviço de urgência e/ou comparação de modelos que possam auxiliar a equipa de gestão do serviço a otimizar o serviço. Pretende-se com este estudo criar/conhecer o perfil do utilizador, no que diz respeito à procura do serviço de urgência do HESE fazendo uma caracterização do serviço, relacionando chegada de utentes e o fluxo dentro do serviço com o perfil do paciente e as suas características. Serão utilizadas ferramentas estatísticas adequadas para analisar os dados fornecidos pelo HESE de forma a responder aos objetivos do estudo. Aplicamos a carta de controlo u para a monitorização do número de casos por cor, tendo em conta as unidades de tempo, dia do ano e dia da semana. Utilizamos os modelos lineares generalizados (GLM's), em particular o modelo de regressão logística, considerando a variável resposta o tipo de utilizador (utilizador normal e hiperutilizador).

**Palavras-chave:** análise estatística de dados; cartas de controlo; modelos lineares generalizados; perfil do utente do serviço de urgência de um hospital; hiperutilizadores

## **Statistical Modeling of Affluence to the HESE Emergency Service**

Public hospitals and present themselves as complex systems in which their areas of intervention are divided according to medical specialties. These have the support of technological equipment for the resolution of health problems identified in patients who present themselves to this service.

The emergency service guaranteed a continuous and differentiated offer of healthcare for the entire population, every day of the year. Due to its easy access character, a population often chooses to solve their health problems, regardless of their urgency, in the emergency services available in an excess of demand for which, sometimes, the emergency service was not dimensioned.

The large influx of users to the institutions' emergency services is a problem that affects not only those responsible for the service, but also the users themselves, due to management problems, scarcity of available human and material resources, but mainly due to the increase in waiting times of patients, which could compromise the quality of care and user safety.

This project aims to study this issue that covers the possibilities of users to the emergency room, modeling the inflow to the emergency room and / or comparison of models that can help the service management team to optimize the service. The aim of this study is to create / know the user profile, with regard to the demand for the HESE emergency service, characterizing the service, relating the arrival of users and the flow within the service with the patient's profile and their features. Tools will be studied to analyze the data provided by HESE in order to respond to the study objectives. We apply the u control chart to monitor the number of cases by color, taking into account the time units, day of the year and day of the week. We use generalized linear models (GLM's), in particular the logistic regression model, considering the response variable as the type of user (normal user and hyperuser).

**Keywords:** Statistical analysis of data; control charts; generalized linear models; profile of emergency department users in a hospital; hyperusers.



# 1 Introdução

Ao longo deste capítulo será apresentada toda a envolvimento deste projeto, no estudo da afluência de utentes ao serviço de urgência do HESE. Começamos por apresentar o enquadramento e motivação do estudo do problema, os objetivos a que se propõem, a metodologia e a estrutura que terá a presente dissertação.

## 1.1 Motivação e objetivos do estudo

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. Cuidados estes de caráter urgente, ou seja, crises agudas que ocorrem num indivíduo e cujo tratamento deva ser imediato.

No entanto, constata-se um aumento na procura pelo Serviço de Urgência como um fenómeno a nível mundial (Carret *et al.*, 2007), aumento esse observado em 14 dos 22 países membros da OCDE, com um incremento de aproximadamente 7% no volume anual de atendimentos (Berchet, 2015). São vários os fatores que têm sido utilizados para justificar as causas que levam a este aumento da procura pelo SU, como por exemplo envelhecimento da população, o aumento do número de pessoas com múltiplas condições crónicas e as mudanças comportamentais associadas à forma como as pessoas optam por ter acesso aos Serviços de Saúde (Coster *et al.*, 2017).

Ainda que um Serviço de Urgência tenha como principal intuito a prestação imediata mediante uma situação urgente, tem-se verificado um aumento de deslocações às urgências de um hospital por casos pouco urgentes ou até mesmo não urgentes.

O uso inapropriado do SU é definido como qualquer visita por alguém com uma condição pouco ou nada urgente que poderia ser tratada de forma mais eficaz e a um custo reduzido, em regime ambulatorio, não urgente.

A procura por serviço de urgência nos hospitais é considerada uma problema internacional que afeta desde os profissionais de saúde como também os próprios utentes. Esta superlotação dos serviços está associada a diversos fatores, entre eles o reduzido acesso a outros serviços de emergência médica ou de cuidados primários e tem provocado atrasos

nos cuidados a doentes urgentes e aumento da mortalidade (Hoot *et al.*, 2007). A grande variedade de fatores intervenientes nesta afluência (variáveis ambientais, eventos locais, número de servidores necessários no serviço, constituição das equipas, etc.) leva a uma difícil tarefa por parte das entidades responsáveis e profissionais de saúde, decorrendo desta incorreta utilização consequências negativas. Entre elas estão a diminuição da qualidade dos serviços prestados aos doentes que carecem efetivamente de cuidados de urgência, desvios de recursos humanos e a menor especialização dos mesmos em Medicina de Urgência. Outra consequência poderá estar relacionada com o atendimento inadequado a doentes que não necessitam do tipo de atendimento prestado em Serviço de Urgência, mas que obtiveram esse benefício. Estes fatores contribuem para uma clara redução do Serviço de Saúde.

Identificar um utilizador frequente dos Serviços de Urgência do Hospital Espírito Santo - Évora, percebendo os fatores que o levam pela procura abusiva a este tipo de serviços, caracterizando o perfil de um hiperutilizador dos SU, constitui a principal motivação e interesse pelo tema. Tendo uma descrição e caracterização deste tipo de pacientes, poderá ser possível elaborar estratégias para mitigar a reincidência e a utilização frequente dos SU, assim como uma gestão ótima dos recursos existentes.

Apresenta-se, neste contexto, o objetivo geral do estudo: Modelação da afluência ao Serviço de Urgência (SU) do Hospital Espírito Santo - Évora, E.P.E. (HESE).

Este projeto visa o estudo desta problemática que abrange as chegadas totais de utentes ao serviço de urgência, modelação da afluência ao serviço de urgência e/ou comparação de modelos que possam auxiliar a equipa de gestão do serviço a otimizar o serviço. Pretende-se com este estudo criar/conhecer o perfil do utilizador, no que diz respeito à procura do serviço de urgência do HESE fazendo uma caracterização do serviço, relacionando chegada de utentes e o fluxo dentro do serviço com o perfil do paciente e as suas características.

Inicialmente será feita uma análise descritiva exploratória dos dados, com o fim de conhecer a população em estudo, bem como as características sociodemográficas dos elementos do estudo, utilizando dados facultados pelo HESE. Iremos aplicar técnicas estatísticas adequadas para atingir os objetivos propostos. Usamos testes de hipóteses paramétricos e não paramétricos, metodologias do controlo estatístico da qualidade e modelos de regressão lineares generalizados, GLM's, que poderão permitir identificar quais as características que influenciam uma utilização abusiva do serviço de urgência e quantificar o impacto de cada uma dessas características na probabilidade de um utente apresentar um comportamento abusivo.

## **1.2 Enquadramento e Motivação**

O desenho do estudo baseia-se na recolha de dados provenientes do ADW (sistema de informação e gestão do hospital).

Este estudo permitiu identificar os utilizadores frequentes e hiperutilizadores do SU, utilizando uma base de dados que contém dados demográficos e relacionados com a visita ao serviço.

Inicialmente realizou-se uma análise a dados referentes aos anos de 2018 e 2019 e posteriormente foram incluídos dados relativos a 2020 e os primeiros 5 meses de 2021. O motivo prende-se com a comparação da afluência ao SU em anos que antecedem a pandemia com o período em que a mesma se instalou, trazendo com ela, não só os períodos de confinamento para a população, mas também um comportamento em termos de afluência a um Serviço de Urgências, condicionado à situação pandémica.

## **1.3 Contribuições**

As principais contribuições deste estudo são:

- Analisar a afluência de utentes ao Serviço de Urgência do Hospital Espírito Santo – Évora;
- Implementar e analisar cartas de controlo para a Triagem no Serviço de Urgência de acordo com o Protocolo de Manchester, para alguns códigos de triagem nos diferentes períodos;
- Modelar a afluência de utilizadores e hiperutilizadores ao serviço de urgência, e perceber que fator ou fatores pesam nesta afluência.

## 1.4 Organização da tese

Este projeto tem como orientação o desenvolvimento do processo de estudo desde a recolha e revisão de literatura até finalização do mesmo, expondo os procedimentos utilizados e justificando a sua escolha.

Para além do Capítulo 1 de Introdução, o relatório conta com mais cinco capítulos. A revisão de literatura é feita no Capítulo 2, na qual constam artigos e referências bibliográficas relevantes para o estudo.

No Capítulo 3 descreve-se o Hospital Espírito Santo – Évora, centrando a atenção no seu Serviço de Urgência.

O Capítulo 4 conta com as metodologias estatísticas utilizadas, nomeadamente, ferramentas do Controlo Estatístico de Qualidade e Modelação Estatística.

No Capítulo 5 faz-se uma apresentação da base de dados utilizada e análise exploratória dos dados, fazendo uma caracterização da população que chega ao SU, caracterização da chegada mensal, semanal e diária. Contamos com uma base de dados disponibilizada pelo HESE, com informações relativas à chegada diária de pacientes ao serviço de urgência entre 1 de janeiro de 2018 e 31 de maio de 2021. Serão analisadas variáveis tais como dias da semana, mês do ano, estação do ano, feriados, temperatura mínima e máxima por dia, concelho de residência do utente, sexo, motivo da admissão, hora de entrada no SU e tempo de espera.

Com recurso às metodologias explicitadas no Capítulo 4 modela-se o número de admissões no serviço de urgência do HESE, tendo em conta códigos verdes, laranjas, amarelos e vermelhos, usando dados fornecidos pelo hospital relativos aos anos de recolha. Os códigos verdes, laranjas e amarelos foram aqueles que registaram maior afluência no período de estudo. Os códigos vermelhos são estudados pelo caráter de urgência que representa. Faz-se ainda uma análise tendo por base a classificação de um utente como utilizador e hiperutilizador<sup>1</sup>.

No Capítulo 6 estão presentes as principais conclusões deste estudo.

---

<sup>1</sup> Um indivíduo que se desloca às urgências 4 ou mais vezes por ano é a definição mais consensual (LaCalle E, Rabin E., 2010, Doupe MB, et al., 2012, Oliveira A. 2008).

## 2 Revisão de literatura

Berchet (2015) define urgência como aquela que necessita de cuidados médicos não podendo ser satisfeita através da administração de outro tipo de cuidados, como por exemplo os cuidados primários. Caracteriza-a ainda como aquela em que o estado de saúde do indivíduo é tão grave ou não sobrevive no momento de admissão ao serviço, sendo imediatamente encaminhado para o internamento ou transferido para uma unidade de saúde cuja prestação de cuidados de saúde seja mais diferenciada.

Emergência e urgência médica são entendidas como a situação clínica de instalação súbita na qual, respetivamente, se verifica ou há risco de compromisso ou falência de uma ou mais funções vitais (Despacho nº18459 de 2006). Estes dois conceitos podem ser distinguidos, pela Comissão de Reavaliação da Rede Nacional de Urgência/Emergência (CRRNEU), definindo como urgência uma situação clínica “que exige intervenção de avaliação e/ou correção em curto espaço de tempo (curativa ou paliativa)”, e por emergência uma situação que “exige um risco de perda de vida ou de função orgânica, necessitando de intervenção em curto espaço de tempo” (Relatório CRRNEU, fevereiro 2012).

Entre 2008 e 2012, a afluência às urgências aumentou de forma geral em todos os países a uma taxa anual de 4,3% (Durand *et al.*, 2011), prevendo-se que este número aumente nos próximos anos (Pine *et al.*, 2009), dadas as diversas transformações sociais.

São vários os fatores que têm sido utilizados para justificar as causas que levam a este aumento da procura pelo SU, como por exemplo envelhecimento da população, o aumento do número de pessoas com múltiplas condições crónicas e as mudanças comportamentais associadas à forma como as pessoas optam por ter acesso aos Serviços de Saúde (Coster *et al.*, 2017).

O SU garante a oferta contínua e diferenciada de cuidados de saúde a toda a população, durante todos os dias do ano. Pelo seu carácter de fácil acesso e outras variáveis tais como falta de meios para um atendimento, a população opta frequentemente por resolver os seus problemas de saúde, independentemente da sua urgência, nos serviços de urgência resultando num excesso de procura para o qual, por vezes, o SU não foi inicialmente dimensionado. Entre os fatores que influenciam a elevada procura do SU, a literatura tem indicado que essa procura se deve, também, ao baixo desempenho do sistema de saúde e da rede hospitalar em particular, conduzindo à baixa qualidade dos serviços em termos de

assistência (Committee on the Future of Emergency Care in the United States Health System, 2007).

A grande afluência de utentes ao SU dos hospitais constitui um problema que afeta não apenas responsáveis pelo serviço como os próprios utentes, pelos problemas de gestão, escassez dos recursos disponíveis, humanos e materiais. E, também, atrasos no fluxo de pacientes podendo comprometer a qualidade do atendimento e segurança do utente.

Este aumento está também relacionado com a utilização indevida por parte dos pacientes que se encontram em situação clínica de atenção médica e recorrem ao SU (Lacalle e Rabin, 2010). Entende-se como utilização inadequada do SU como o atendimento realizado por um indivíduo com uma condição pouco ou nada urgente e que poderia ser tratada de forma eficiente e com menor custo em regime ambulatorio (Christ *et al.*, 2010).

O grande objetivo dos SU reside na diminuição da morbimortalidade e, também, as sequelas incapacitantes, sendo caracterizado como uma componente importante na assistência para a saúde (Azevedo *et al.*, 2010).

Estudos mostram que muitos utilizadores evidenciam uma procura privilegiada e persistente pelos SU, procura essa com carácter repetitivo. Sendo um grupo considerado pequeno, quando comparado com o total de utilizadores, estes chamados hiperutilizadores absorvem uma percentagem elevada da oferta nestes serviços (Lacalle e Rabin, 2010). As consequências desta procura são fator determinante para a superlotação e aumento do custo dos SU (Lacalle *et al.*, 2013).

Este aumento inadequado registado leva muitas vezes à sobrelotação da urgência de um hospital, levando a tempos de espera excessivos e à consequente redução da qualidade dos cuidados de saúde prestados. Pelo consumo de recursos de emergência tais como, profissionais de saúde e equipamentos médicos, desviando-os de situações de natureza mais grave, não só aumenta a carga de trabalho profissional como os afeta negativamente, diminui a satisfação do utente e reduz a qualidade do atendimento com tempos de espera e diagnósticos ou tratamentos tardios (Rocovich e Patel, 2012; Durand *et al.*, 2012).

À parte dos fatores negativos já identificados, surgem os custos associados à alta prevalência de episódios de serviço de urgência inadequados. Em Inglaterra, por exemplo, o custo de visitas inapropriadas foi estimado em cerca de £ 100 milhões entre os anos de 2011 e 2012 (McHale *et al.*, 2013). Em Portugal não existem dados publicados, no entanto sabe-se que é o país com maior número, por 100 mil habitantes de episódios de Serviço de Urgência, de acordo com a OCDE (Fig 1).

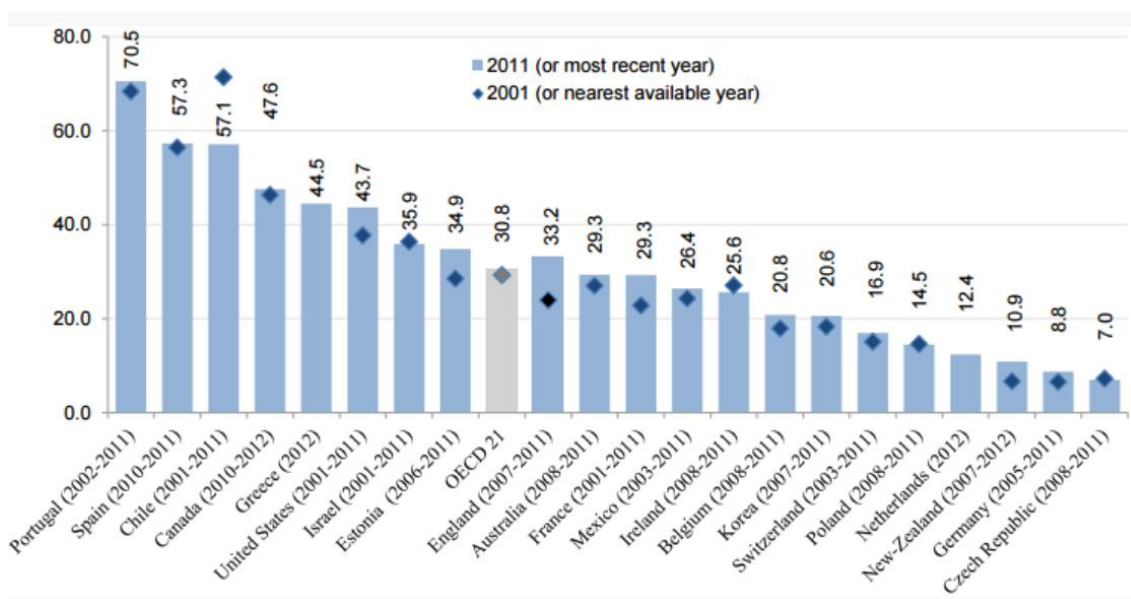


Figura 1: Número de visitas por 100 mil habitantes ao Serviço de Urgência, 2001 e 2011 (Fonte: Relatório grupo de trabalho serviço de urgência, 2019)

Sabe-se, pela Comissão de Reavaliação da rede Nacional de Emergência/Urgência, em 2010, que aproximadamente 46% dos casos admitidos nos SU a nível nacional não constituem episódios urgentes, o que corresponde a cerca de 6 milhões de admissões nos SU em Portugal Continental não urgentes e portanto abusivas. Estes valores variam conforme a região do país, sendo no Alentejo e em particular no Hospital Espírito Santo - Évora, E. P. E., nos anos de 2018 e 2019, foram registados cerca de 40% de admissões não urgentes, num total de 146873 admissões<sup>2</sup>.

Em estudos já realizados está desmonstrado que a afluência ao serviço de urgência não é aleatória, e que depende muitas vezes de condições 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).

Em Boston, Noble *et al.* (1971) concluíram que em condições meteorológicas favoráveis o número de pacientes a dirigir-se ao serviço de urgência, dispensando o uso de ambulância aumentou. Batal *et al.* (2001), numa tentativa de explicar o número de admissões numa clínica em Denver, utilizaram como variáveis o dia da semana, mês do ano, estação do ano, feriados e dias antecedentes e precedentes a feriados, variáveis meteorológicas (temperaturas máximas, temperaturas mínimas, queda de neve e chuva). Destas, apenas a

<sup>2</sup> Fonte: <https://www.sns.gov.pt/monitorizacao-do-sns/servicos-de-urgencia/caracterizacao-urgencias/> ARS Alentejo (ARS) + H. Espírito Santo - Évora, E.P.E. (abrev\_n4\_cs\_aces\_hosp)

temperatura máxima diária e a queda de neve se revelaram estatisticamente significativas. Diaz *et al.* (2001) concluíram que o número de pacientes registados valores máximos nos meses de inverno relativamente aos meses de verão, onde a afluência ao hospital é inferior.

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

Registam-se valores máximos de admissões às urgências às segundas-feiras, e que decrescem com o passar da semana, sendo o fim de semana o que regista menores valores (Batal *et al.*, 2001; Wargon *et al.*, 2010; Marcilio *et al.*, 2013). Para Jones *et al.*, 2008; Diaz *et al.*, 2001, a variável meteorológica com maior impacto no número de admissões ao serviço de urgência é a temperatura máxima. Marcilio *et al.*, 2013; Champion *et al.*, 2007 concluem que existe pouca variação no número de admissões entre os meses do ano.

Identificar um utilizador frequente dos Serviços de Urgência do Hospital Espírito Santo - Évora, percebendo os fatores que o levam pela procura abusiva a este tipo de serviços, caracterizando o perfil de um hiperutilizador dos SU, constitui a principal motivação e interesse pelo tema. Tendo uma descrição e caracterização deste tipo de pacientes, poderá ser possível elaborar estratégias para mitigar a reincidência e a utilização frequente dos SU, assim como uma gestão ótima dos recursos existentes.

Na literatura existem ainda poucos estudos que exploram a previsão do número de admissões diário de utentes ao SU, Friede *et al.*, (2009), Schwartz *et al.*, (2004). Rotstein *et al.*, (1997), desenvolveram um modelo linear generalizado incluindo variáveis para o dia da semana, mês e feriados para prever o número de atendimentos diários num SU em Israel. Os autores descrevem que o erro médio de previsão correspondeu a 10% da média das visitas diárias e recomendaram que a previsão do número de atendimentos no SU fosse feita com suporte no estudo de séries temporais.

Batal *et al.*, (2001) usaram regressão linear múltipla para prever o número de admissões num SU nos Estados Unidos. O modelo obtido para a previsão incluiu variáveis para perceber o efeito do dia da semana, mês, feriados, períodos de inverno e média móvel do volume de utentes nos 7 dias anteriores. Também foram estudadas variáveis relacionadas com o clima (temperatura máxima, temperatura mínima e temperatura média e precipitação de água (chuva, neve, granizo). Os autores concluíram que o dia da semana foi o fator de maior impacto na previsão, com maior número de admissões ao SU às segundas-feiras e tendência decrescente ao longo da semana. As condições meteorológicas pouco acrescentaram no modelo.



Para avaliar a qualidade da previsão, os autores aplicaram no próprio SU os valores previstos com o estudo, adequando a dimensão da equipa de serviço de acordo com o número previsto de pacientes. Verificou-se uma melhoria significativa na organização da equipa de profissionais de saúde no SU, associada à melhoria dos indicadores do grau de satisfação por parte de utentes.

A estrutura dos dados em estudos de modelação da admissão de utentes a um SU, varia. Existem estudos com um espaço temporal de recolha de 6 anos (Jones, 2007) e outros que apenas têm como base 1 ano de recolha (McCarthy *et al.*, 2008). Contudo sabe-se que esta dimensão, no que diz respeito ao número de observações, tem impacto na validação do modelo, podendo originar *overfitting*<sup>3</sup> ou *underfitting*<sup>4</sup>.

Ao contrário da maioria dos autores de estudos neste âmbito, são feitas abordagens ao nível da modelação de dados provenientes de hospitais distintos para a criação de modelos mais gerais e com aplicabilidade em qualquer hospital (Boyle e Jessup *et al.*, 2011; Boyle e Wallis *et al.*, 2008). Se por um lado um modelo geral seria o ideal para o SNS, de modo a reunir as condições para qualquer hospital deste sistema, por outro lado quanto mais específico puder ser o modelo, relativamente à realidade do hospital, melhor desempenho terá. Por ser uma previsão na qual não se podem incluir características da entidade de saúde nem dos utentes, o ponto de partida reside no número de admissões num SU.

Outros estudos mais específicos, como a previsão de ocupação de camas são levados como fatores populacionais ou hospitalares. São muitos os estudos que consideram os fatores meteorológicos determinantes no número de admissões a um SU, mas a maior parte conclui que estas mesmas variáveis acabam por piorar a performance dos modelos. Por outro lado, as variáveis relacionadas com datas (calendário) mostraram um forte poder preditivo nesta temática (Batal *et al.*, 2001; McCarthy *et al.*, 2008; Calegari *et al.*, 2016).

Entre as várias metodologias estudadas (modelos lineares clássicos ou modelos temporais) sabemos que existem pontos fortes e limitações. Se por um lado investigadores obtiveram bons resultados com a utilização de modelos lineares (Boyle e Wallis *et al.*, 2008; Jones *et al.*, 2008; McCarthy *et al.*, 2008; Marcilio *et al.*, 2013), outros conseguiram-no através de séries temporais, tanto pela suavização exponencial (Champion *et al.*, 2007; Boyle *et al.*, 2011) ou modelos ARIMA (Díaz-Hierro *et al.*, 2012; Calegari *et al.*, 2016).

---

<sup>3</sup> *Overfitting* é a palavra inglesa para designar o sobre-ajuste do modelo aos dados, ou seja, um ajuste excessivo na validação que se reflete num baixo desempenho com novos dados.

<sup>4</sup> *Underfitting*, ao contrário de *overfitting*, designa o sub-ajuste do modelo aos dados que origina um erro elevado na validação e no teste.

Na Tabela 1 encontra-se um resumo de alguns estudos realizados no âmbito da modelação de dados desta natureza.

Artigo	Objetivo	Metodologias	Variáveis	Conclusões
Batal <i>et al.</i> , 2001	Admissões Diárias	Regressão linear	Calendário, Meteorológicas	Variáveis de calendário são significativas e variáveis meteorológicas são pouco relevantes
Boyle e Wallis <i>et al.</i> , 2008	Admissões mensais	Regressões		O melhor método é aquele que tem menor erro absoluto médio no teste.
McCarthy <i>et al.</i> , 2008	Admissões horárias	Regressão de Poisson	Calendário, meteorológicas	Variáveis meteorológicas não têm efeito significativo

Tabela 1: Exemplos de estudos realizados no âmbito da modelação

No presente estudo trabalham-se variáveis relacionadas ao volume de admissões do SU no HESE, mas também relacionadas com o perfil dos utentes e características nas admissões.

### **3 O Hospital Espírito Santo – Évora, E.P.E.**

O Hospital do Espírito Santo de Évora, E.P.E. (HESE) é a maior e principal unidade hospital da região do Alentejo, com uma atuação que, para além de receber os utentes do Alentejo Central (cerca de 152.865 habitantes), trata também doentes provenientes do Alto Alentejo, Baixo Alentejo e Alentejo Litoral (345.939 habitantes), perfazendo um total de aproximadamente 498.804 habitantes<sup>5</sup>.

O HESE dispõe de um Serviço de Urgência Polivalente, único em toda a região do Alentejo, prestando cuidados de saúde diferenciados em tempo útil, garantindo padrões de qualidade elevados do desempenho técnico-científico e destaca-se como referência no Alentejo para as mais diferenciadas áreas.

A admissão de um paciente no hospital pode fazer-se de forma eletiva ou emergente, sendo a primeira quando a admissão do paciente é programada com um período temporal de antecedência que pode ser de dias ou semanas. A admissão emergente, é aquela que não carece de qualquer planeamento e por isso feita através do Serviço de Urgência (SU).

Neste tipo de admissão os motivos prendem-se, sobretudo, a fatores epidemiológicos (cronicidade de patologias), meteorológicos (picos de frio ou de calor), fatores económicos (crise económica, isenção de taxas moderadoras, recurso a serviços privados reduzidos) e também fatores sociais (dias festivos, dias da semana, hora do dia).

O Serviço de Urgência Geral integra o Departamento de Urgência/Emergência do HESE EPE, e é constituído por um Serviço de Urgência Geral de Adultos, multidisciplinar e multiprofissional e integra a Unidade Médico – Cirúrgica (UMC) e a Viatura Médica de Emergência e Reanimação (VMER). É um serviço que disponibiliza um atendimento permanente em várias valências, tendo como missão a observação, o tratamento e encaminhamento de doentes emergentes, muito urgentes e urgentes, definidos segundo o Sistema de Triagem – Metodologia de Manchester.

A triagem nos serviços de urgência do HESE é feita tendo por base um conjunto de questões de ordem clínica, que dão seguimento à atribuição da cor da pulseira de acordo com

---

<sup>5</sup> <http://www.hevora.min-saude.pt/2019/12/27/o-hospital/>

a gravidade da situação. A cor atribuída ao paciente determina o tempo de espera máximo a que o paciente deverá ser submetido até à primeira observação médica.

O Sistema de Triagem de Manchester é um método de triagem inicial dos utentes que recorrem ao Serviço de Urgência, e que está implementado em Portugal. Este permite uma triagem de prioridades através da identificação de critérios de gravidade – os discriminadores – e classifica o paciente numa de 5 categorias identificadas por um número, nome, cor e tempos alvo de observação inicial.

Na Tabela 2 estão identificados os tempos estimados máximos de acordo com a gravidade da situação.

Cor da Pulseira	Situação	Tempo até à primeira observação médica
Vermelho	Emergente	0 minutos
Laranja	Muito urgente	10 minutos
Amarelo	Urgente	50 minutos
Verde	Pouco Urgente	120 minutos (2 horas)
Azul	Não Urgente	240 minutos (4 horas)
Branco <sup>6</sup>	Não urgente	-

Tabela 2:Resumo da informação relacionada com a Triagem de Manchester

A implementação do Protocolo de Manchester permite que os atendimentos sejam feitos com mais eficiência, já que o tempo pode representar a diferença entre salvar uma vida e perder um paciente. Além disso, como o sistema prevê uma organização adequada em termos de prioridades, é possível fazer um encaminhamento de pacientes para outras unidades médicas, de forma que a procura possa ser distribuída por unidades médicas próximas, se for o caso. Além disso, com a diminuição do risco clínico, evita-se o uso de recursos de urgência desnecessários.

Com a implementação do Protocolo de Manchester, os pacientes também são beneficiados, uma vez que submetidos a esta metodologia de classificação de risco, estão assegurados que não correrão risco de vida, e além disso, pode dar uma previsão média do tempo que leva até ao atendimento. Além disso, a utilização do Protocolo Manchester é a garantia de oferta de um serviço homogêneo, tendo em vista que, independentemente do

---

<sup>6</sup> Em Portugal, com assentimento do Grupo de Triagem de Manchester, foi introduzida a categoria de urgência “Branca”, para designar as situações que não estão relacionadas com uma queixa clínica e que utilizam os serviços de urgência como forma de consulta de rotina.

horário, do dia da semana ou do profissional de saúde, a instituição de saúde terá o mesmo padrão no atendimento.

## 4 Metodologia estatística

Na presente investigação foram ajustados modelos de regressão logística cuja variável resposta toma os valores de 1 se é um hiperutilizador e de 0 se não o é. Foram também utilizadas cartas de controlo, em particular a carta u, por ser mais adequada para a proporção de casos numa determinada unidade de tempo. Na análise exploratória dos dados, para além das ferramentas da estatística descritiva, foram utilizados testes não paramétricos dada a natureza e distribuição das variáveis, entre eles o teste de Games-Howell por ser o mais robusto a desvios da normalidade e quando o pressuposto da homogeneidade das variâncias é violado.

Para a análise dos dados foram utilizados os softwares RStudio versão 4.1.2 (2021-11-01) e SPSS versão 28.

### 4.1 Modelos Lineares Generalizados

Entende-se regressão como modelos estatísticos para modelar a relação funcional entre variáveis (Montgomery, Peck e Vining, 2006). Em particular, interessa referir modelos que pertencem a um grande grupo de modelos designado por Modelos Lineares Generalizados.

Por este motivo, foi necessário o desenvolvimento de outros 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, 1965) 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.

Em 1972, Nelder e Wedderburn, introduziram então os modelos lineares generalizados (MLG's ou GLM's<sup>7</sup>), sendo estes uma síntese de todos os outros modelos que tinham sido usados até à data para modelar dados não gaussianos.

Os GLM's são, portanto, não apenas a extensão de todos os outros 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 a variável resposta e variáveis independentes é dada por uma qualquer função diferenciável, mantendo-se a estrutura de linearidade. São, pois, casos particulares dos modelos lineares generalizados os seguintes modelos (Turkman, 2000): modelo de regressão linear clássico; modelos de análise de variância e covariância; modelo de regressão logística; modelo de regressão de Poisson; modelos log-lineares para tabelas de contingência multidimensionais; modelo *probit* para estudos de proporções, etc.

Nos GLM's a comparação de modelos é feita com recurso ao AIC e a seleção de variáveis é feita de forma análoga aos modelos lineares. Este critério (AIC) consiste numa medida para a seleção de modelos, modelos esses que podem ser ordenados em conformidade com o valor deste critério, sendo o modelo tanto melhor quanto menor for o seu AIC.

O AIC é dado por

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

Onde  $L$  representa a verosimilhança do modelo e  $p$  o número de parâmetros.

#### **4.1.1. Análise de Resíduos e Multicolinearidade**

Na análise de resíduos pretende-se descobrir possíveis falhas em alguns dos pressupostos dos modelos. As medidas usadas na análise de resíduos têm como base os resíduos padronizados  $e_i^s$ , que são dados pelo quociente entre os resíduos  $e_i$  e o desvio-padrão estimado  $\hat{\sigma}$ .

---

<sup>7</sup> Do inglês, Generalized Linear Models

A normalidade dos resíduos é testada mediante um teste de Shapiro-Wilks ou Kolmogorov-Smirnov, sob a hipótese nula dos erros terem distribuição normal.

Para verificar a homogeneidade da variância, a média nula e a independência dos erros, basta representar graficamente os resíduos e valores estimados da variável resposta.

Na análise da multicolinearidade, isto é, correlação entre as variáveis explicativas pode recorrer-se à matriz de correlações e deve eliminar-se aquela ou aquelas que apresentam valores correlacionais mais elevados, tendo em atenção aos dados. Pode também recorrer-se ao VIF (*variance inflation factor*), 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  $VIF \approx 1$  as variáveis explicativas são independentes. Se  $VIF > 10$  as variáveis explicativas estão correlacionadas. Neste caso será usado como ponto de corte um  $VIF$  superior a 5 ou 10 (Craney e Surles, 2002).

#### 4.1.2 O Modelo de Regressão Logística

O modelo de regressão logística é uma classe particular dos modelos GLM's e aplica-se quando a variável resposta é de natureza nominal ou ordinal, e esta é função das variáveis explicativas, as quais podem ser nominais, ordinais ou contínuas.

A função *logit* é designada por função de ligação nos modelos lineares generalizados e permitem linearizar a variável resposta, podendo esta ser modelada em função de um modelo linear. A função *logit* é dada por

$$\text{logit}(\hat{\pi}) = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right).$$

O rácio  $\frac{\hat{\pi}}{1-\hat{\pi}}$  designa-se por *chances* ou *odds* e traduz as chances de se observar o sucesso ( $Y = 1$ ) relativamente ao insucesso ( $Y = 0$ ). As chances são calculadas da seguinte forma:

$$P(Y = 1) = \frac{\hat{\pi}}{1-\hat{\pi}} = \frac{\#\{Y=1\}/n}{\#\{Y=0\}/n} = \frac{\#\{Y=1\}}{\#\{Y=0\}},$$

e representa-se na forma  $\#\{Y = 1\} : \#\{Y = 0\}$ , ou seja, o número de indivíduos que detém a característica de interesse comparativamente ao número de sujeitos que não possui essa característica. Assim, o modelo da regressão logística, para mais de uma variável independente  $X_i (i = 1, \dots, p)$  é:

$$\text{logit}(\hat{\pi}_j) = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_p X_{pj}.$$

Neste modelo, o coeficiente  $\beta_0$  é o valor do  $\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right)$  quando todos os  $X_i = 0 (i = 1, \dots, p)$  e  $\beta_1, \beta_2, \dots, \beta_p$  são os coeficientes *logit*. Usado quando a variável resposta tem apenas 2 categorias (0 e 1). A função usada na regressão logística para estimar a probabilidade de uma determinada realização  $j (j = 1, \dots, n)$  da variável dependente ser o “sucesso”,  $P[Y_i = 1] = \hat{\pi}_j$ , é a função logística cuja forma geral,

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

Onde, para o modelo de regressão logística multivariado tem-se

$$\hat{\pi}_j = \frac{e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_p X_{pj}}}{1 + e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_p X_{pj}}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1j} + \dots + \beta_p X_{pj})}},$$

onde  $\hat{\pi}$  é o vetor das probabilidades estimadas,  $X$  é a matriz das variáveis independentes e  $\beta$  é o vetor dos coeficientes de regressão logística. O ajustamento consiste em linearizar esta função com a transformação *logit*, (Hosmer e Lemeshow, 2013).

Para a interpretação dos valores de  $\beta$  recorre-se à exponencial destes coeficientes, dada a sua dificuldade de interpretação:

$$\text{Exp}(\beta_i) = \frac{\frac{P[Y=1|X_i=x+1]}{1-(P[Y=1|X_i=x+1])}}{\frac{P[Y=1|X_i=x]}{1-(P[Y=1|X_i=x])}}.$$

A  $\text{Exp}(\beta_i)$  representa o rácio das *chances (Odds ratio)* do “sucesso” *versus* “insucesso” quando a variável independente  $X_i$  permanece constante. Este valor consiste numa aproximação da razão das *chances* do “sucesso” *versus* “insucesso” por unidade da variável independente, ou seja, quando  $X_i$  aumenta uma unidade, as *chances* de obter “sucesso” variam  $\beta_i$  unidades.



O modelo de regressão logística descrito obedece aos seguintes pressupostos:

- Linearidade e aditividade: a escala  $\text{logit}(\hat{\pi})$  é aditiva e linear (mas  $\pi$  não);
- Proporcionalidade: a contribuição de cada  $X_i (i = 1, \dots, p)$  é proporcional ao seu valor com um fator  $\beta_i$ ;
- Efeito constante: a contribuição de uma variável independente é constante, e independente da contribuição das outras variáveis independentes;
- Os erros são independentes e apresentam distribuição binomial;
- Os preditores não são multicolineares.

A validação dos pressupostos do modelo pode fazer-se graficamente através da análise de resíduos. Considerando que as variáveis estão centradas e padronizadas, tem-se  $R = (X^T X)^{-1}$  em que os elementos da diagonal dessa matriz chamados fatores de inflação de variância (VIF) e representam o incremento da variância devido à presença de multicolinearidade (Montgomery, Peck, & Vining, 2021). O VIF pode ser dado pela seguinte equação:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, p,$$

Onde  $p$  é o número das variáveis independentes e  $R_j^2$  é o coeficiente correlação múltipla, que resulta da regressão de  $X_j$  das  $p - 1$  variáveis independentes.

No que diz respeito à multicolinearidade o seu diagnóstico pode ser obtido calculando a tolerância a partir do  $R^2 (T = 1 - R^2)$  obtido pela regressão múltipla entre cada uma das variáveis independentes e as restantes variáveis independentes do modelo.

A matriz de variâncias e covariâncias para as estimativas dos coeficientes do modelo é dada por:

$$V(\hat{\beta}_j) = \sigma^2 (X^T X)^{-1} = \sigma^2 VIF = \sigma^2 (1 - R_j^2)^{-1}.$$

Se  $R_j^2$  for muito próximo de 1, existe uma correlação elevada entre a variável  $X_j$  e as restantes variáveis, e assim  $1 - R_j^2$  estará próximo de zero, o que se traduz num VIF muito elevado, apontando para a colinearidade dessa covariável.

Os parâmetros da regressão logística são estimados pelo Método da Máxima Verosimilhança. A verosimilhança de uma amostra aleatória de  $n$  observações de variáveis independentes de média  $\pi_i, i = 1, 2, \dots, n$  é dada por

$$L(\mathbf{y}, \boldsymbol{\beta}) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}.$$

Onde  $\pi_i = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}$ , é a probabilidade de ocorrência de  $y_i = 1$  com os valores amostrais das variáveis explicativas  $x_1, \dots, x_p$ , para o  $i$ -ésimo indivíduo.

De uma transformação logarítmica na função anterior obtém-se,

$$l(\mathbf{y}; \boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln(\pi_i) + (n - \sum_{i=1}^n y_i) \ln(1 - \pi_i).$$

De seguida calcula-se a função *score* derivando a função  $l(\mathbf{y}; \boldsymbol{\beta})$  em ordem a  $\boldsymbol{\beta}$  obtendo-se

$$U(\boldsymbol{\beta}) = \frac{\partial l(\mathbf{y}; \boldsymbol{\beta})}{\partial(\boldsymbol{\beta})} = \mathbf{x}^T (\mathbf{y} - \boldsymbol{\pi}).$$

A Hessiana é a matriz das segundas derivadas e é dada por

$$H(\boldsymbol{\beta}) = \frac{\partial^2 l(\boldsymbol{\beta})}{\partial(\boldsymbol{\beta}) \partial(\boldsymbol{\beta})^T} = -\mathbf{x}^T \mathbf{W} \mathbf{x},$$

sendo  $\mathbf{W}$  uma matriz diagonal,  $(n \times n)$ , cujos elementos da diagonal principal são obtidos através dos produtos  $\pi_i(1 - \pi_i)$  e as estimativas de  $\boldsymbol{\beta}$  através de métodos numéricos, como por exemplo o método de Newton Raphon.

Para ajustar os modelos seguimos a mesma metodologia recomendada por Hosmer e Lemeshow: para o modelo inicial foram selecionadas todas as variáveis que se revelaram significativas na fase univariada (valor  $p < 0.20$ ). A partir deste modelo foram eliminadas sucessivamente, e por ordem decrescente dos valores  $p$  de Wald, todas as variáveis não significativas (valor  $p > 0.10$ ). Posteriormente verificamos se alguma(s) das variáveis que não foram incluídas no modelo inicial se mostra(m) agora significativas em presença das que estão no modelo, caso em que são adicionadas ao modelo. De seguida verifica-se a adequabilidade do número de categorias de cada variável, e testa-se o pressuposto de linearidade para as variáveis contínuas, com o *logit*. Foram testadas via teste de razão de verosimilhança as interações entre variáveis que faziam sentido no contexto do estudo (valor  $p < 0.01$ ) e foi feita a análise de resíduos para a pesquisa de observações influentes ou outliers.

A qualidade do ajustamento dos modelos de regressão logística foi verificada pelo teste de Hosmer-Lameshow e teste Cessie-van Houwelingen. A capacidade discriminativa dos modelos foi avaliada pela área sob a curva ROC (Area under Curve, AUC), sendo o modelo final validado usando a técnica de validação por bootstrap.

#### 4.1.3 Modelos Aditivos Generalizados

Os modelos aditivos generalizados (Hastie e Tibshirani, 1986, 1990) são uma extensão dos modelos lineares generalizados com um preditor linear que envolve uma soma de funções *suaves* de covariáveis. Um modelo aditivo generalizado standard é da forma (Wood, 2017, p.249)

$$g(\mu_i) = A_{i\gamma} + \sum f_j(x_{ji}), \quad y_i \sim EF(\mu_i, \phi),$$

onde  $A_i$  é a  $i$ -ésima linha de uma matriz do modelo paramétrico, com os parâmetros específicos utilizados na matriz do modelo representados por  $\gamma$ ,  $f_j$  é uma função não paramétrica, isto é, cuja forma não é especificada, estimada através de curvas de alisamento da covariável  $x_j$ , e  $EF(\mu_i, \phi)$  a distribuição da família exponencial com média  $\mu_i$  e parâmetro de escala  $\phi$ . A variável resposta,  $y_i$ , é modelada admitindo a independência dos diferentes valores de  $i$ , dado o valor de  $\mu_i$ . Uma generalização pode ser dada por (Wood, 2017, p.249)

$$g(\mu_i) = A_{i\gamma} + \sum_j L_{ij} f_j(x_{ji}), \quad y_i \sim EF(\mu_i, \phi),$$

onde  $L_{ij}$  é uma função linear limitada de  $f_j$ . A família de curvas de alisamento (*splines*) inclui várias formas de estimar a função não paramétrica  $f_j$ , entre elas, *splines* cúbicos, B-*splines*, *splines* naturais, *splines* finas e *splines* de suavização.

##### 4.1.3.1 Splines cúbicos de alisamento

Uma *spline* cúbica é uma curva construída tendo por base seções de polinómios cúbicos unidos por nós, de tal forma que a curva seja contínua. Este método é feito através da 1ª e da 2ª derivadas da função. Um exemplo de uma *spline* cúbica de alisamento pode ver-se na Figura 2.

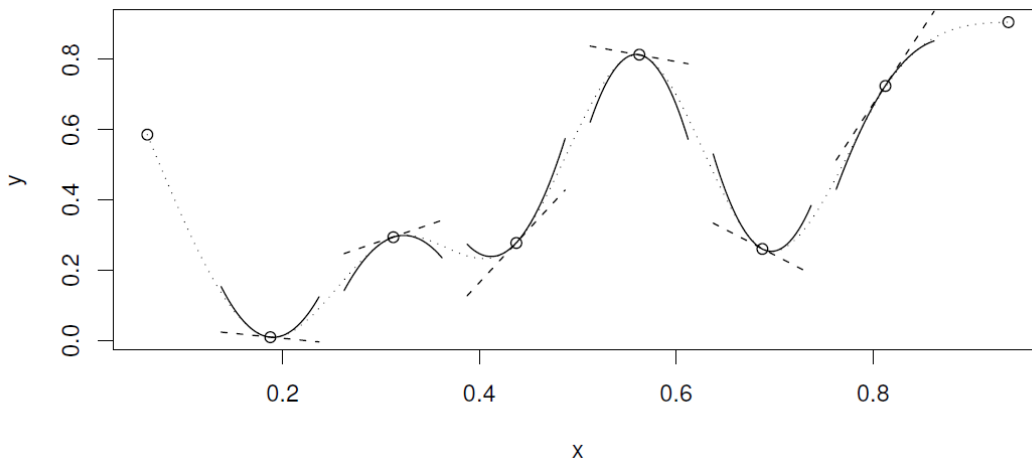


Figura 2: *Spline* cúbica é uma curva construída a partir de seções de polinômios cúbicos (Wood, 2006, Fig. 3.3, p.122)

A curva representada é composta por 7 seções unidas (°) em que os dois pontos finais são chamados nós da *spline*. Cada seção tem diferentes coeficientes, as linhas a tracejado representam os gradientes da *spline* nos nós e as curvas contínuas correspondem à 1ª e 2ª derivadas nos nós, e quanto maior o número de nós utilizados, menor será o alisamento da curva, Figura 3.

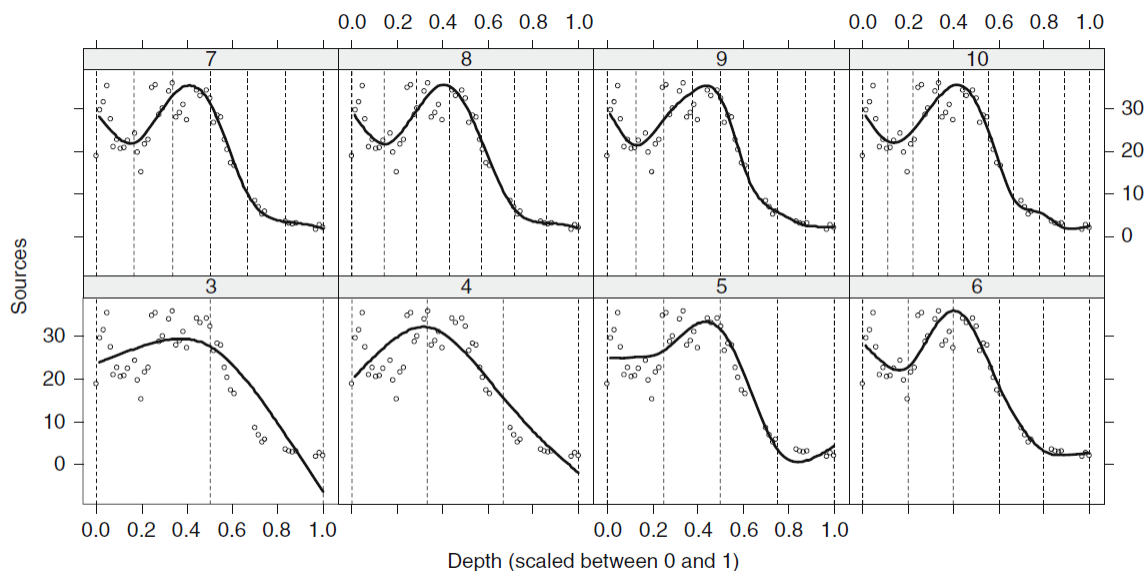


Figura 3: Curvas de suavização obtidas por *splines* de regressão cúbica usando 3 nós, 4 nós, 5 nós, etc., e 10 nós. Ambos os pontos finais também contam como nós (Zuur, 2009, Fig. 3.9, p. 48)

Em Keele (2008), pode ver-se uma recomendação de utilizar 3 nós para amostras com  $n < 30$  e 5 nós para  $n > 100$ . A abordagem dos modelos aditivos generalizados, baseada em curvas de alisamento, permite avaliar a relação entre as variáveis explicativas e a resposta sem necessidade de definir previamente a forma desta relação, no entanto os coeficientes correspondentes a essas curvas nos modelos de regressão não são interpretáveis e a estimação dos riscos envolve procedimentos computacionais intensos.

## 4.2 Controlo Estatístico de Qualidade

O principal objetivo do Controlo Estatístico da Qualidade é atingir uma garantia de Qualidade em organizações, nos produtos e/ou serviços, através da utilização de técnicas estatísticas convenientes (Gomes et al., 2010).

A qualidade de um produto pode ser descrita e avaliada de várias maneiras (Montgomery, 2019). Em geral, distinguem-se as principais dimensões da qualidade. Garvin (1987) propõe oito dimensões da qualidade:

1. **Desempenho:** o produto vai corresponder ao que se espera?
2. **Confiança:** com que frequência o produto vai falhar?
3. **Durabilidade:** quanto tempo vai durar o produto?
4. **Manutenção:** qual a facilidade de reparação do produto?
5. **Estética:** qual vai ser o aspeto do produto?
6. **Características:** qual a função do produto?
7. **Qualidade percebida (aparente):** qual a reputação da entidade/empresa responsável ou do seu produto?
8. **Conformidade com especificações:** o produto é exatamente igual ao produto concebido pelo desenhador?

Estas oito dimensões são habitualmente adequadas para a descrição da qualidade na maioria das situações industriais e de negócios. No entanto, nas instituições de serviços e negócios de transações (tais como instituições bancárias e financeiras, serviços de saúde e serviço ao cliente) podemos acrescentar as três seguintes dimensões, (Montgomery, 2019):

1. **Sensibilidade:** Quanto tempo foi necessário para que o fornecedor do serviço respondesse ao seu pedido de serviço? Quão disposto a ajudar se mostrou o fornecedor do serviço? Quão prontamente foi o seu pedido foi atendido?

2. **Profissionalismo:** Esta dimensão consiste no conhecimento e competências do fornecedor do serviço, e está relacionada com a competência da organização em fornecer os serviços pedidos.
3. **Atenção:** Em geral, os clientes esperam cuidado e atenção personalizada por parte dos fornecedores de serviços. Um cliente pretende que as suas necessidades e preocupações sejam importantes para quem fornece o serviço e que as mesmas sejam alvo de atenção.

Como se pode então definir a qualidade? Uma definição de qualidade pode ser dada com base na utilização que cada um de nós dá a determinado produto. Pelo senso comum, um produto tem qualidade quando satisfaz uma ou mais características que nele se pretendem encontrar. Alguns autores definem qualidade como atender às especificações (do consumidor).

O conceito de qualidade é bastante vasto e abrangente e pode ser visto de duas maneiras:

- **Qualidade do Projeto do Produto:** proporciona uma medida de adequação do projeto às necessidades da adequação ao uso.
- **Qualidade Conforme:** até onde é que o produto satisfaz os requisitos do projeto; uma variabilidade reduzida nos produtos traduz uma melhor Qualidade e uma consequente redução de desperdícios, aumentando a produtividade (Deming, 1986).

#### **4.2.1 Conceito de Variabilidade**

A qualidade de um produto é avaliada em função da variabilidade desse mesmo produto, relativamente a um valor desejado de uma característica de qualidade. Em Montgomery (2019) pode ler-se: “Qualidade é inversamente proporcional à variabilidade.” Para a determinação da qualidade de um produto e para que o seu processo seja possível de controlar é necessário saber distinguir as causas de variação: causas de variação aleatória e causas de variação assinaláveis.

A variabilidade natural, devida à aleatoriedade inerente à natureza, torna praticamente impossível a produção de dois produtos absolutamente iguais. Se a variabilidade for pequena, sem implicar consequências no consumidor, considera-se uma variabilidade aceitável (causas de variabilidade aleatória). No caso em que as consequências da variabilidade são visíveis e afetam o consumidor considera-se inaceitável (causas de variabilidade assinaláveis). Como referimos anteriormente, as causas de variabilidade

podem ser controláveis, tais como a qualidade da matéria-prima, ajustes de máquinas, procedimentos utilizados, cuidado dos operadores, entre outros, ou incontroláveis devido a causas aleatórias.

A Avaliação na Saúde envolve a seleção de critérios para julgar e comparar a adequabilidade, benefícios, efeitos adversos e custos de tecnologias, serviços ou programas de saúde; esses critérios constituem os indicadores de qualidade na saúde (Donabedian, 1980b; Reis, 1995)

#### **4.2.2 Noção de Controlo Estatístico do Processo**

Pode dizer-se que o controlo estatístico do processo nasceu com o trabalho pioneiro de Walter Shewhart por volta de 1930, no estudo da aleatoriedade dos processos industriais. Desenvolveu uma ferramenta estatística que permite determinar se a variabilidade de um processo é aleatória ou devida a causas de variação assinaláveis, a carta de controlo.

Pode dizer-se que o controlo estatístico do processo consiste, basicamente, na interpretação das cartas de controlo de processos e na utilização de técnicas para a identificação de causas de variação e oportunidades de melhoria da qualidade. Estas técnicas visam auxiliar a obtenção dos padrões especificados da qualidade e reduzir a variabilidade inerente a estes padrões especificados. O controlo estatístico do processo permite a manutenção contínua do processo, possibilitando uma ação imediata assim que o problema for detetado, enquadrando-se na filosofia que preconiza a construção da qualidade dentro do processo e prevenção de problemas. Este justifica-se pela existência e necessidade da avaliação da variabilidade com o propósito da melhoria do nível da qualidade.

##### **4.2.2.1 Cartas de controlo**

As cartas de controlo são as técnicas do controlo estatístico do processo mais conhecidas e mais utilizadas. Trata-se de comparações gráficas do estado de um processo (medido por uma característica de qualidade, como a média de uma medida, número de defeitos, entre outros). A interpretação dos valores observados num processo deve ser a mais objetiva possível e conter o máximo de informação detalhada, com a medida de localização central (média e mediana) e a dispersão (amplitude e desvio padrão).

Antes da recolha das amostras é importante que se conheça a característica da qualidade que vai ser estudada. Esta pode ser classificada de duas formas: variáveis ou

atributos. A primeira aplica-se aos casos em que a característica da qualidade é mensurável, isto é, se puder ser medida e expressa por um número. Neste caso utilizam-se as cartas para a média (carta  $\bar{X}$ ), para a amplitude (carta R) e para o desvio padrão (carta S), baseadas na função de probabilidade Normal  $(\mu, \sigma^2)$ . Quando a característica da qualidade não puder ser medível, isto é, é classificada apenas como defeituosa ou não defeituosa, ou podermos contar o número de defeitos por unidade produzida, utilizamos as cartas de controlo para atributos. Temos as Cartas p, para a proporção de unidades defeituosas, ou seja, não conforme com as especificações, as Cartas c, para o número de não conformes associadas a modelos Binomial e de Poisson, respetivamente e as Cartas u, para o número de defeitos por unidade. Modificações sugeridas a este tipo de cartas podem ser encontradas, por exemplo, em Borrer et al. (1998), McCool e Joyner-Motley (1998), Morais e Natário (1998), Reynolds e Stoumbos (1999), Fang (2003), Khoo (2004), Wu e Luo (2004), Morais e Pacheco (2006), Chan et al. (2007), Wu e Jiau (2007) e Wu e Wang (2007).

Na Figura 2 pode observar-se a tradicional carta de controlo tipo Shewhart. A carta contém três linhas paralelas onde a linha central representa o valor médio da característica da qualidade e as duas linhas simetricamente colocadas acima e abaixo da linha central o limite superior de controlo e limite inferior de controlo, respetivamente. Dada uma estatística E, seja  $\mu_E$  a sua média e  $\sigma_E$  o desvio padrão. Então a carta fica definida com a linha central (LC) igual a  $\mu_E$ , limite superior de controlo (LSC) igual a  $\mu_E + L\sigma_E$  e limite inferior de controlo (LIC) igual a  $\mu_E - L\sigma_E$ , onde L representa a distância entre os limites de controlo à linha central. Por fim marca-se os valores referentes à estatística E. Considera-se que um processo está estável se nenhum valor estiver fora dos limites de controlo.

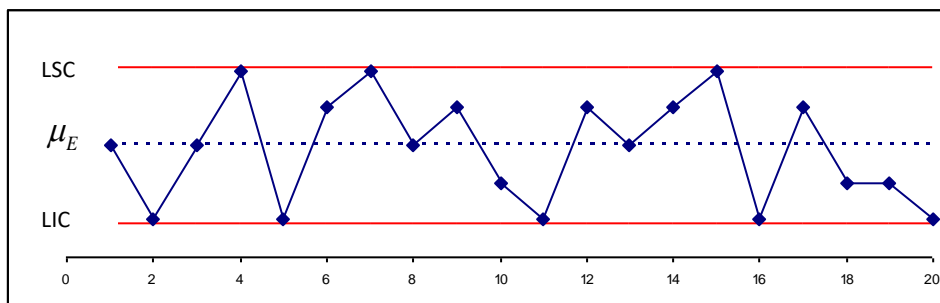


Figura 4: Carta de controlo Shewhart



No presente estudo, dadas as características do processo<sup>8</sup>, optou-se por escolher a carta u para o estudo das variáveis. Independentemente do tipo de carta de controlo que se utiliza existem alguns pressupostos que precisam ser satisfeitos de modo a permitir a sua utilização, e, portanto, que os seus resultados sejam válidos.

- As observações sejam independentes e identicamente distribuídas (pressuposto IID), ou que as amostras sejam retiradas de forma aleatória e que o seu processo esteja estatisticamente sob controlo;
- As observações (ou as médias das amostras das observações) sigam alguma distribuição de probabilidade específica, tais como a Normal, Binomial ou Poisson.

Caso não sejam satisfeitos os pressupostos, todo o valor das cartas de controlo é posto em causa. Em Montgomery (2019) pode ver-se algumas formas de lidar com as situações descritas nos pontos anteriores, sugerindo o ajuste do processo um modelo de séries temporais ARIMA (Média Móvel Autoregressiva Integrada), e aplicar as cartas de controlo aos usuais resíduos obtidos. As cartas de controlo baseadas num modelo de séries temporais são indicadas para o caso em que as observações são correlacionadas como é o caso de processos químicos onde várias medidas num processo ou na característica de um produto são geralmente correlacionadas. A utilização de um modelo apropriado de séries temporais neste tipo de processos permite remover a autocorrelação dessas mesmas observações.

#### **4.2.2.2 Cartas de controlo para variáveis qualitativas**

Muitas características da qualidade não podem ser representadas numericamente de forma apropriada. Nessas situações representamos cada ocorrência como conforme ou não conforme considerando as especificações para determinada característica da qualidade. A terminologia “defeituoso” ou “não defeituoso” é, em geral, utilizada para classificar um produto. No entanto, é comum utilizar a terminologia “conforme” ou “não conforme”. As características de qualidade deste tipo são chamadas de atributos.

Frequentemente o número de unidades amostrais, que constituem os subgrupos (conjunto de unidades selecionadas num determinado momento), é variável, e nestas

---

<sup>8</sup> As características presentes ao estudo serão apresentadas num capítulo posterior.

circunstâncias interessa controlar a taxa de defeitos por unidade. A carta a utilizar é a carta u uma vez que a amostra tem a sua dimensão variável ao longo do tempo.

A carta u faz a monitorização do número de não conformidades por unidade produzida. Semelhante à carta c mas com a particularidade de que o número de não conformidades é expresso em relação a cada unidade. A carta u é útil quando a amostra contém mais do que uma unidade e quando a dimensão da amostra varia.

Admitindo que cada ponto da carta de controlo foi obtido com base em  $n$  unidades de inspeção, temos que  $Y$  (o número de defeitos nas  $n$  unidades) tem uma distribuição de Poisson com parâmetros  $nc$  - soma de variáveis aleatórias Poisson com parâmetro  $c$  é uma Poisson com parâmetro  $nc$ .  $Y$  tem distribuição definida por

$$P(Y = k) = \frac{(nc)^k e^{-nc}}{k!}, k = 0, 1, \dots, +\infty, c > 0.$$

Existindo a possibilidade de considerar um número médio de defeitos por unidade, e sendo  $Y$  o número de defeitos presentes nas  $n$  unidades, então o número médio de defeitos por unidade é:

$$U = \frac{Y}{n}, Y \sim P(nc).$$

Consequentemente os parâmetros da variável aleatória  $U$  são:

$$\mu = E[U] = E\left[\frac{Y}{n}\right] = \frac{1}{n} E[Y] = n \frac{c}{n} = c$$

e

$$\sigma^2 = VAR[U] = VAR\left(\frac{Y}{n}\right) = \frac{1}{n^2} VAR[Y] = n \frac{c}{n^2} = \frac{c}{n}$$

O estimador do parâmetro  $c$  é dado por

$\hat{c} = \bar{u} = \frac{\sum_{i=1}^m u_i}{m}$  (média do número médio de defeitos por unidade em  $m$  amostras de tamanho  $m$ ).

O valor da variável  $u$  num subgrupo que contenha  $n_i$  unidades amostrais onde sejam encontrados  $c$  defeitos, tem como limites de controlo, para este tipo de carta, definidos da seguinte forma:

$$u = \frac{c}{n_i}$$

$$LIC = \bar{u} - 3 \sqrt{\frac{\bar{u}}{n_i}}$$

$$LC = \bar{u} = \frac{c_1 + c_2 + \dots + c_k}{n_1 + n_2 + \dots + n_k}$$

$$LSC = \bar{u} + 3 \sqrt{\frac{\bar{u}}{n_i}}$$

onde,

- $n_i$  representa o número de unidades na amostra  $i, i = 1, \dots, k$ ;
- $c_1, \dots, c_k$  representa o número de defeitos;
- $n_1, \dots, n_k$  representa o tamanho dos  $k$  subgrupos.

A carta  $u$  é recomendada quando o produto é composto por várias partes e características da qualidade, mas também quando a dimensão é variável.

Os requisitos para que o número de não conformidades obedeça a uma Poisson são:

- A frequência média de não conformidades deve ser proporcional à quantidade considerada;
- As não conformidades devem ocorrer de forma independente.

Se a dimensão for variável, também será variável o valor dos limites de controlo na carta. Esta abordagem envolve a construção de uma carta de controlo com base no número médio de não conformidades por unidade de inspeção e as regras de deteção de pontos fora de controlo são as habituais utilizadas nas cartas de controlo.

#### 4.2.2.3 Racional Ordering e Racional Subgrouping

Quando nos deparamos com um processo no qual pode não estar muito evidente as causas de variação relativamente a uma unidade, deve-se proceder à construção de uma carta na qual estão todas as unidades e tentar determinar quando é que uma unidade representa uma causa assinalável quando analisada como uma parte do conjunto total considerado.

A eficácia das cartas de controlo num processo depende muito do tipo de amostragem. Esta pode ser feita sob dois contextos: ordenação racional e subamostragem ou subgrupo racional (*Racional Ordering* e *Racional Subgrouping*). A ordenação racional envolve a melhor seleção do melhor *o quê, como e quando* serão feitas as medições. A frequência de amostragem é racional se for suficientemente frequente para conseguir obter uma monitorização das mudanças do processo. Neste sentido os planos de amostragem devem levar-nos a análises que forneçam uma visão mais profunda e não apenas a apresentação de números.

*Rational Ordering* neste contexto significa que a amostra para todas as unidades presentes é feita de forma sequencial, utilizando uma ordenação de datas, por exemplo, dias, semanas, meses. Com esta metodologia é possível, por exemplo, estratificar uma amostra em subgrupos de acordo com as características do processo e do interesse do estudo.

*Rational Subgrouping*, por outro lado, consiste numa metodologia que, em vez de organizar os dados de forma sequencial, como é feito na *Rational Ordering*, numa série temporal, estes são tratados para cada unidade permitindo fazer uma comparação transversal das várias unidades, utilizando apenas uma unidade amostral para cada subconjunto. O controlo por subgrupo racional traduz-se numa boa alternativa relativamente à ordenação racional quando há poucos dados para cada unidade ou quando as unidades são tantas, que uma carta de controlo construída pelo método de ordenação racional, poderia ficar pouco interpretável. Cada ponto da carta representa a taxa para um elemento individual com base no seu comportamento ao longo do tempo, um ano por exemplo.

No controlo feito por subgrupo racional é habitualmente utilizada uma carta p para atributos ou mesmo as cartas X e S para a variação numa amostra. No presente estudo a carta utilizada será a carta u para atributos.

Depois de estratificadas todas as unidades na mesma carta, por uma das metodologias atrás discutidas, podem surgir dúvidas relativamente às causas assinaláveis detetadas. Uma ou mais unidades podem constituir causas assinaláveis, e nesse caso perceber se essas causas ocorrem dentro da unidade e comparar com as outras unidades. Neste sentido, construir uma carta de controlo para cada unidade poderá ser útil na medida em que permite detetar de forma mais rápida o que pode estar a provocar a instabilidade no processo, quando existe. No caso de se encontrar apenas uma causa de variação comum, depois de estratificados os dados, o processo é analisado sob o ponto de vista de cada variável e agrupar de acordo com a metodologia *Rational Ordering* ou *Rational Subgrouping*. Se uma causa assinalável for detetada, procede-se de acordo com a metodologia inicial de construção de uma carta,

tentando apurar as causas de instabilidade do processo. No caso de o processo estar estável, mas inaceitável, pode ser adaptado o esquema da Figura 5, *PDSA cycle*<sup>9</sup>, de forma a implementar mudanças e estimar a eficiência de possíveis intervenções.

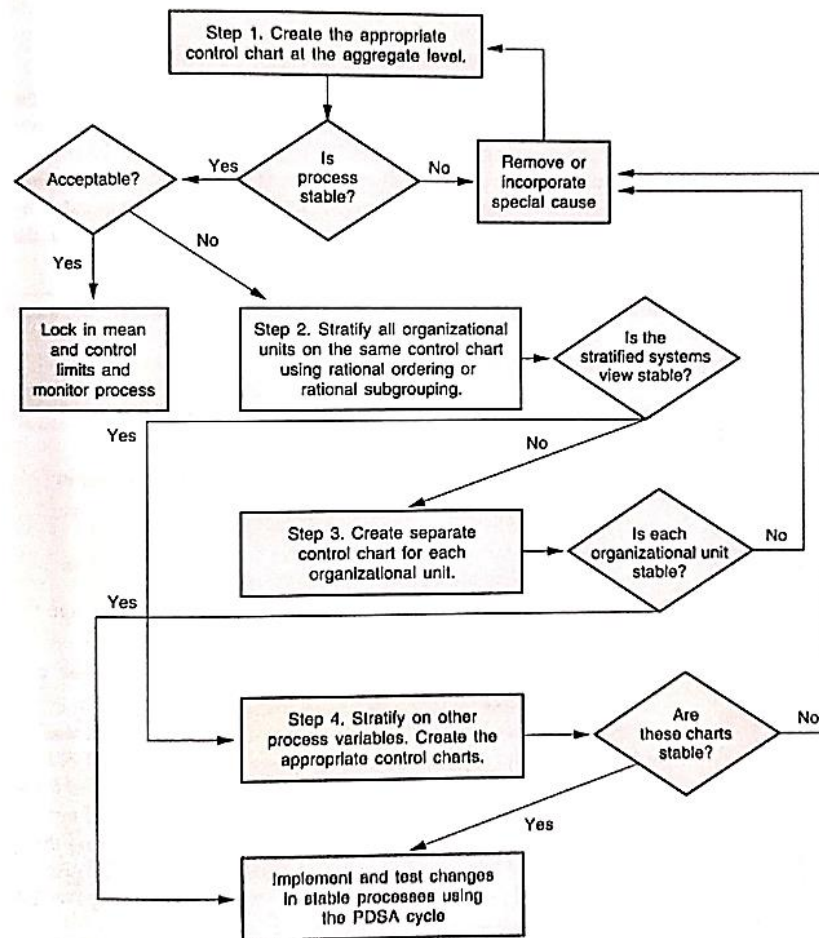


Figura 5: Metodologia de controlo em dados agrupados – “Drilling down into into aggregate data”. (Carey, R. e Stake, L. (2003))

#### 4.2.3 Outras ferramentas do controlo estatístico de qualidade

As cartas de controlo e o estudo da capacidade do processo apresentam melhores resultados quando utilizados juntamente com um conjunto de ferramentas integradas na melhoria da qualidade. Essas ferramentas permitem identificar oportunidades de melhoria e assim reduzir a variabilidade. Podem enunciar-se, entre outras, as seguintes ferramentas (Montgomery, 2019):

<sup>9</sup> plan-do-study-act cycle

- Histograma de frequências: a sua forma mostra a natureza da distribuição das observações e analisar a variabilidade. Além disso os limites de especificação podem ser aqui utilizados na determinação da capacidade do processo.
- Diagrama de Pareto: É um diagrama de barras de atributos organizado por categorias onde está representada a frequência do número de defeitos pelo tipo de defeito, de modo que a identificação dos defeitos mais comuns seja mais rápida. Tal análise deve ser feita tendo sempre em conta os custos de cada defeito (um defeito pouco frequente pode ter efeitos financeiros consideráveis requerendo, por isso, uma atenção prioritária – “poucos essenciais muitos triviais”). Para ilustrar o Diagrama de Pareto, segue-se o seguinte exemplo<sup>10</sup>.

Considere-se uma amostra de 166 defeitos de um tanque, organizados num Diagrama de Barras onde estão representados o número total de ocorrências (do defeito) contra os vários tipos de defeito. A Figura 6 é uma representação deste gráfico.

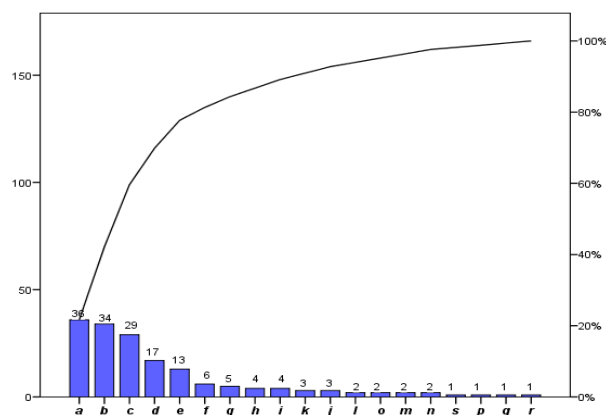


Figura 6: Diagrama de Pareto para os defeitos de um tanque

Fonte: Montgomery, 2012

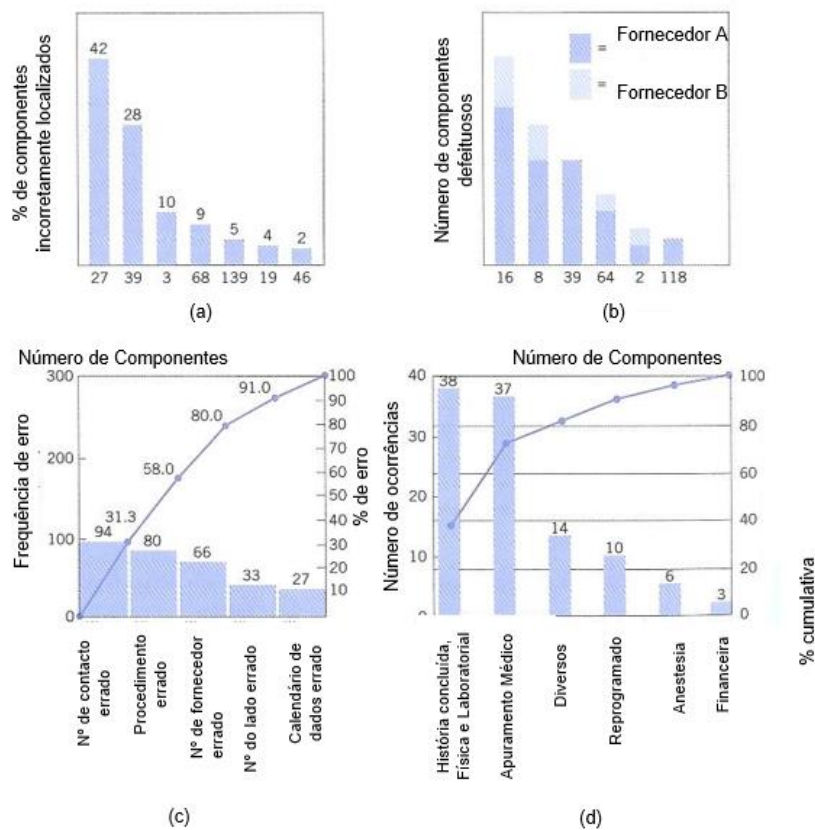
Através desta representação gráfica é fácil identificar o tipo de defeitos que ocorre com mais frequência. Na Figura 6 é fácil ver que os defeitos *a*, *b* e *c* são os defeitos mais frequentes, pressupondo que as respetivas causas do seu aparecimento devem ser identificadas para uma posterior exclusão. Esta ilustração remete para a regra 80:20, ou seja, 80% de determinado resultado é explicado por 20% dos fatores.

Note-se que, através do Diagrama de Pareto, não são imediatamente identificados todos os defeitos, mas apenas uma parte correspondente aqueles que ocorrem mais frequentemente. Considere-se como exemplo, ainda na Figura 6, o defeito *o* cuja frequência,

<sup>10</sup> Montgomery, 2001, pp.178-181.

2 em 166 defeitos, representa 1,2% dos defeitos. No entanto este defeito<sup>11</sup> pode ocorrer como consequência de determinados problemas técnicos, podendo um defeito, nestas condições, pertencer a uma categoria de defeitos mais graves. Quando no conjunto de defeitos existe uma mistura daqueles de onde possam advir consequências assinaláveis e de outros de importâncias menor, um dos seguintes métodos deve ser utilizado:

1. Utilizar um sistema de ponderação<sup>12</sup> para modificar a frequência das contagens.
2. Acompanhar a análise do Diagrama de frequências de Pareto com um Diagrama de custo ou exposição de Pareto.



Fonte: Montgomery, 2001

Figura 7: Exemplos de Diagramas de Pareto

<sup>11</sup> O contexto deste exemplo pode ver-se em Montgomery 2001, pp. 178 – 181.

<sup>12</sup> O processo de sistema de ponderação não será aqui abordado. Pode ver-se em Montgomery 2001, Capítulo 6.

Existem muitas variações da representação básica do Diagrama de Pareto. A Figura 7 (a) é uma representação do Diagrama de Pareto aplicado a um processo de montagem eletrónica utilizando um processo de montagem de componentes por superfície. O eixo vertical indica a percentagem de componentes incorretamente localizados, e o eixo horizontal o número total de componentes, um código que deteta o dispositivo sobre a placa de circuito impresso. Note-se que as localizações 27 e 39 acumulam 70% dos erros. Este facto pode ser uma consequência do *tipo* ou *tamanho* dos componentes nestas localizações, ou o *posicionamento* destas localizações relativamente à placa do desenho. Figura 7 (b) é uma outra representação do Diagrama de Pareto aplicado à indústria eletrónica. Sobre o eixo vertical está representado o número de componentes defeituosas, e no eixo horizontal o número total de componentes. Cada barra vertical de frequências foi dividida em duas partes representando os dois tipos de fornecedores, e tem-se assim um Diagrama de Pareto Empilhado. Esta análise indica claramente que o fornecedor A proporciona uma quota desproporcional de componentes defeituosos.

O Diagrama de Pareto é utilizado por equipas responsáveis pela melhoria da qualidade cujos procedimentos são organizados como na Figura 7 (c). A equipa investiga quebras nas vendas com o intuito de reduzir o número dessas mesmas quebras. Este tipo de Diagrama de Pareto é formado por duas escalas. Uma para o erro de frequência atual e outra para a percentagem de erros. A Figura 7 (d) é uma representação de um Diagrama de Pareto construído por uma equipa responsável pela melhoria da qualidade num hospital, de forma refletir as causas do cancelamento do programa de cirurgia ambulatorio.

De uma forma geral, o Diagrama de Pareto é uma das ferramentas do controlo de qualidade mais utilizadas. Com um conhecimento e prática amplos desta ferramenta, o Diagrama de Pareto deixa de ter limitações nesta área, passando a ser indispensável.



## 5 Base de dados utilizada e metodologia

Recolhemos dados referentes aos anos de 2018 (n=73977) e 2019 (n=74143) integrantes numa base de dados, composta por 11 variáveis de interesse. Inicialmente são apenas considerados os dados 2018-2019 para analisar a afluência ao SU e posteriormente, numa segunda fase, juntamos os dados de 2020 e 2021 que foram recolhidos mais tarde, não apenas para comparar a afluência ao SU mas porque houve a necessidade de recolher outras variáveis de interesse do ponto de vista clínico.

A lista de variáveis encontra-se na Tabela 3.

Variável	Descrição da variável	Tipo de variável
Dias do ano	Designação do dia por mês	Data
Mês	Mês do ano	Qualitativa nominal
Cor do Protocolo	Cor do Protocolo do episódio	Qualitativa nominal
Idade	Idade do paciente	Quantitativa contínua
Sexo	Sexo do paciente	Qualitativa nominal
Situação de feriado	Designação se o dia corresponde a véspera de feriado, feriado, pós feriados ou nenhum dos anteriores	Qualitativa nominal
Turno	Turno em funcionamento à hora do episódio	Qualitativa ordinal
Estação do ano	Estação do ano em que ocorreu o episódio	Qualitativa nominal
Dia da semana	Designação do dia da semana	Qualitativa nominal
Hora do dia	Hora do dia em que ocorreu o episódio	Qualitativa ordinal
Temperatura	Temperaturas mínima diária	Quantitativa contínua
Temperatura	Temperaturas máxima diária	Quantitativa contínua
Concelho	Concelho de residência do utente	Qualitativa nominal
Tempo de espera	Tempo de espera até à 1ª intervenção	Quantitativa contínua
Dia em relação a feriado	Dias não úteis (feriados ou fins de semana)	Qualitativa nominal

Tabela 3: Descrição das variáveis da Base de Dados

Os feriados tidos em conta serão: Ano Novo, Carnaval, Sexta-feira Santa, Páscoa, Dia da Liberdade, Dia do Trabalhador, Corpo de Deus (anos 2018, 2019, 2020, 2021), Dia de

Portugal, Feriado Municipal, Assunção de Nossa Senhora, Dia da Implantação da República, dia de todos os Santos, Dia da Restauração da Independência, Dia da Imaculada Conceição e Dia de Natal. Neste Capítulo apresenta-se a modelação estatística dos dados, os modelos obtidos e validação. Ao longo dos anos de 2018, 2019, 2020 e 2021 (cinco primeiros meses) observámos a chegada de 222045 episódios ao SU.

## 5.1 Caracterização da população atendida no Serviço de Urgência

Nesta secção caracteriza-se a população admitida ao serviço de urgência bem como algumas características do sistema relacionando-o com as admissões e tratamento de utentes. Inicialmente estão envolvidos 148120 episódios, que totalizam os anos de recolha, sendo 73977 de 2018 e 74143 de 2019, que constituem os elementos de análise da primeira fase do estudo.

Idade	2018	2019
Nº de episódios	73977	74143
Média	42,23	42,34
Desvio-padrão	29,53	29,38
Mínimo	0	0
Máximo	107	119
Percentis		
	25	15
	50	41
	75	69

Tabela 4: Tabela sumária da variável Idade (2018 e 2019)

Comparando aos anos de 2018 e 2019 verifica-se que a idade média dos utentes é muito semelhante, isto é, de cerca de 42 anos ( $M=42.23$ ;  $DP=29.53$  para 2018;  $M=42.34$ ;  $DP=29.38$  para 2019), Tabela 4, as diferenças não são significativas ( $U = 2734290983.50$ ,  $W=5470478283.50$ ,  $p>0.10$ ).

Como forma de completar os dados anteriores seguem-se as Figuras 8 e 9 nas quais se pode ver o Boxplot das idades por sexo, em ambos os anos de pesquisa.

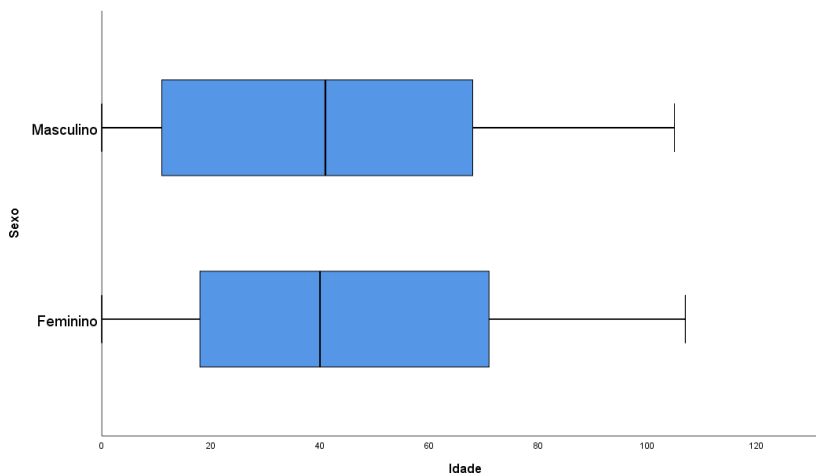


Figura 8: Boxplot para a idade dos utentes por sexo, 2018

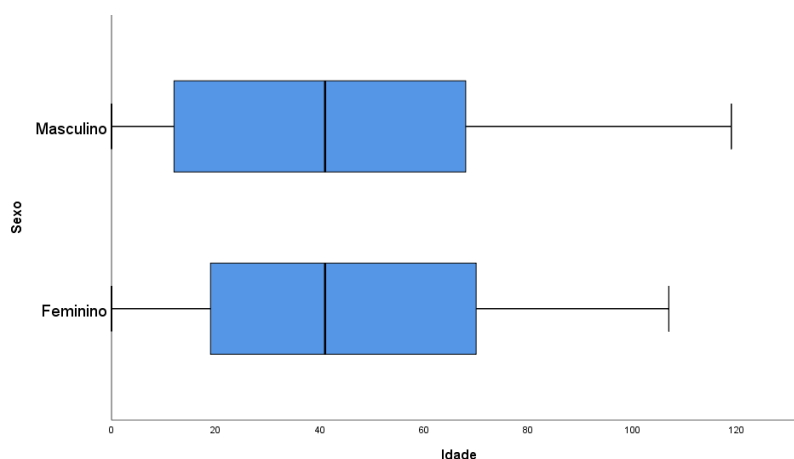


Figura 9: Boxplot para a idade dos utentes por sexo, 2019

## 5.2 Caraterização da procura pelo Serviço de Urgência - HESE

Na Tabela 5 encontra-se a distribuição de chegadas dos utentes nos anos 2018 e 2019, por Estação do ano, cor do protocolo, género, mês, data (véspera de feriado, feriado e dia após o feriado) e o turno correspondente à admissão realizada. Relativamente à cor do protocolo, pode ver-se o registo de admissões correspondente à categoria Não Aplicável e que por não constituir um código de triagem específico, foi apenas contemplada na análise exploratória inicial.

Variável	2018		2019		
	N	%	N	%	
Estação	Inverno	16537	22,40%	16727	22,60%
	Primavera	18668	25,20%	18521	25,00%
	Verão	18237	24,70%	17785	24,00%
	Outono	20535	27,80%	21110	28,50%
Cor do protocolo	Branco	2180	2,90%	3064	4,10%
	Vermelho	1271	1,70%	350	0,50%
	Laranja	26350	35,60%	11719	15,80%
	Amarelo	32336	43,70%	32669	44,10%
	Verde	10996	14,90%	24641	33,20%
	Azul	304	0,40%	1159	1,60%
	Não Aplicável	539	0,70%	540	0,70%
Gênero	Feminino	39407	53,30%	39584	53,40%
	Masculino	34570	46,70%	34559	46,60%
Mês	janeiro	6614	8,90%	6532	8,80%
	fevereiro	6056	8,20%	6237	8,40%
	março	6099	8,20%	6163	8,30%
	abril	5933	8,00%	5869	7,90%
	maio	6395	8,60%	6532	8,80%
	junho	6227	8,40%	5866	7,90%
	julho	6086	8,20%	6062	8,20%
	agosto	6231	8,40%	5854	7,90%
	setembro	5740	7,80%	6029	8,10%
	outubro	6282	8,50%	6603	8,90%
	novembro	6056	8,20%	5980	8,10%
	dezembro	6258	8,50%	6416	8,70%
Data festiva	Véspera de feriado	2849	3,90%	2528	3,40%
	Feriado	2873	3,90%	2804	3,80%
	Pós-feriado	3217	4,30%	3250	4,40%
	Não se aplica	65038	87,90%	65561	88,40%
Turno	00:00-07:59	7975	10,80%	7838	10,60%
	08:00-15:59	36175	48,90%	36513	49,20%
	16:00-23:59	29827	40,30%	29792	40,20%

Tabela 5: Tabela sumária de valores percentuais e contagens (2018 e 2019)

De acordo com os dados da Tabela 5 verifica-se que em ambos os anos houve um registo maior de casos no outono do que nas restantes estações do ano, e ligeiramente menos no inverno. Na triagem predominam códigos amarelos, laranjas e verdes e há mais mulheres que homens. A distribuição das admissões tendo em conta o mês não é significativamente diferente e o turno com menos afluência corresponde ao período horário compreendido entre as 0h-07:59. O turno 08:00-15:59 é o turno com maior afluência ao SU. O teste do Qui-quadrado de Pearson permite admitir que existe associação da afluência ao SU entre o ano e o mês de admissão ( $\chi^2_{(11)} = 45,39, p < 0.001$ ). No que diz respeito ao dia da semana, este é independente do ano, a incidência de utentes do SU por dia da semana, não está associada ao ano em que ocorreu ( $\chi^2_{(6)} = 9.60, p = 0.142$ ).

A análise pode ainda ser realizada tendo em conta o dia da semana, Figura 10 e Figura 11. Verifica-se um crescimento de domingo para 2ª feira, sendo este o dia com maior percentagem de admissões ao SU, tanto em 2018 como em 2019. Entre 2ª feira e Sábado, verifica-se um decréscimo na afluência de utentes, sendo o Sábado o dia com menor valor de admissões. Entre Sábado e Domingo, volta a subir o número de admissões, em 2018 e 2019, não sendo este aumento de forma abrupta, como se regista de Domingo para 2ª feira. Em 2019 verifica-se uma diminuição abrupta da afluência voltando esta a subir à 5ª feira.

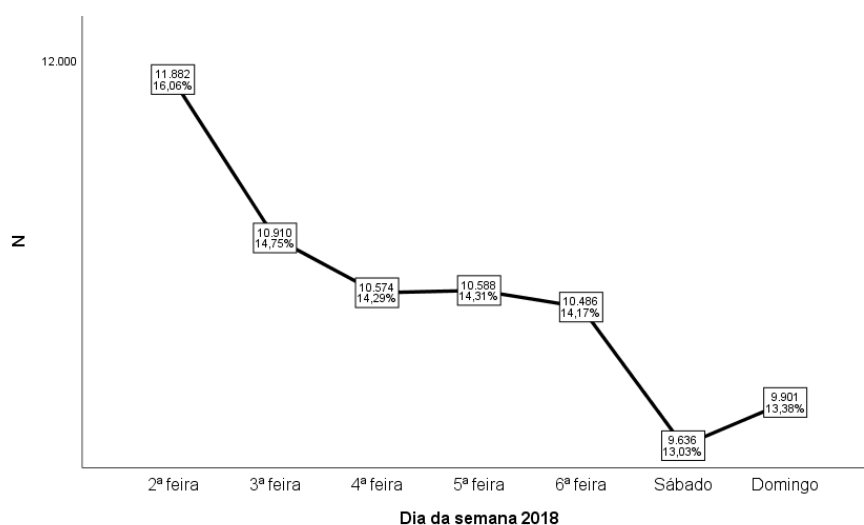


Figura 10: Representação gráfica das admissões por dia da semana, 2018

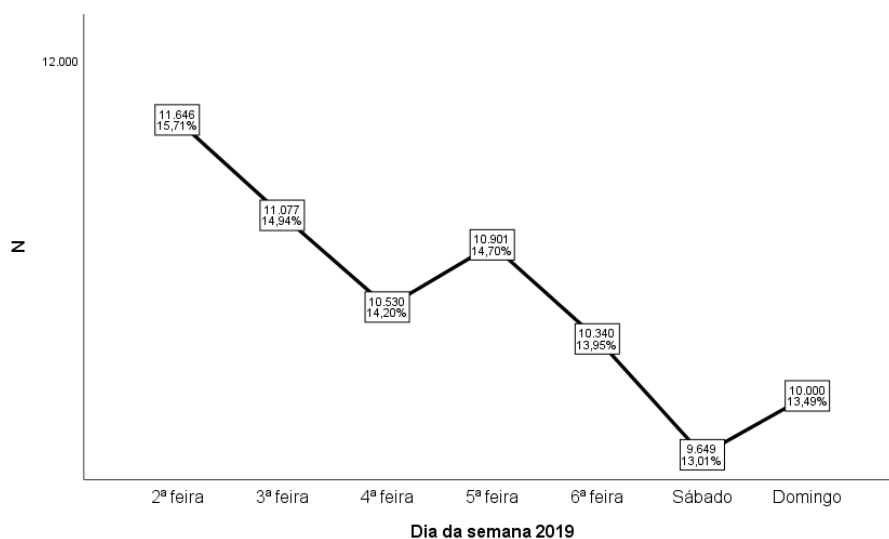


Figura 11: Representação gráfica das admissões por dia da semana, 2019

Janeiro é o mês com maior percentagem (8.94%) de admissões ao SU, em 2018, Figura 12. Se analisarmos os resultados para 2019 verificamos que é o mês de Outubro (8.91%) o que regista maior valor percentual de admissões, seguindo-se o mês de Janeiro (8,81%).

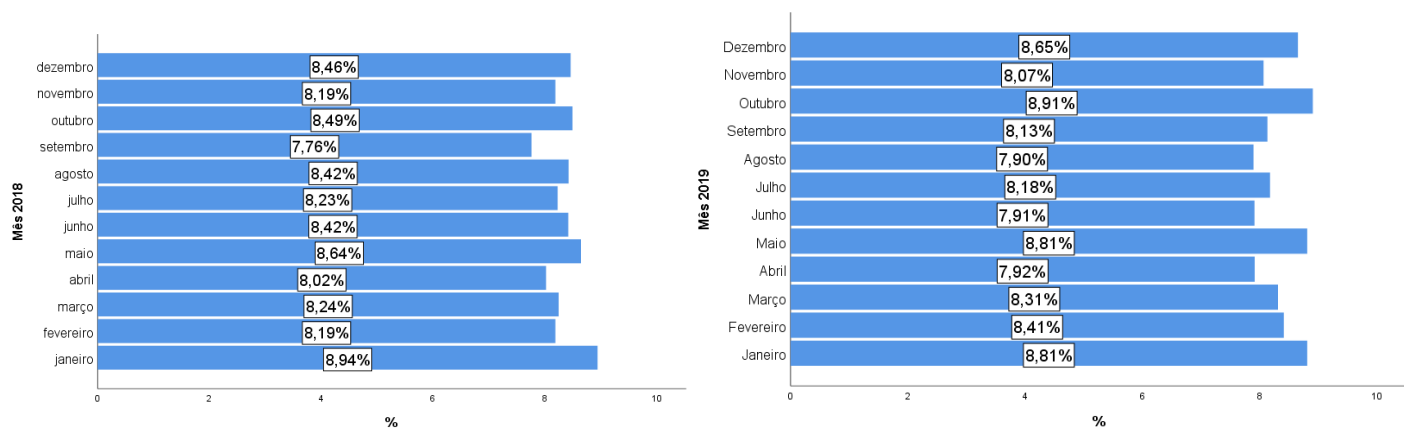


Figura 12: Chegada mensal de utentes ao SU (2018 - 2019)

Para completar a análise por ano, do número de chegadas ao SU, a Figura 13 representa a afluência ao SU, por ano, no qual é fácil de perceber a variação que existe com o avançar dos meses. De Janeiro para Abril regista-se um decréscimo bastante acentuado no número de admissões, que volta a subir Maio. Os picos são portanto Janeiro, Maio, Outubro e embora ligeiramente menor, o mês de Dezembro. Abril e Setembro constituem os meses com menor afluência ao SU. Verifica-se ainda que nos meses de Julho, Agosto e Setembro 2018 e 2019 registam variações opostas.

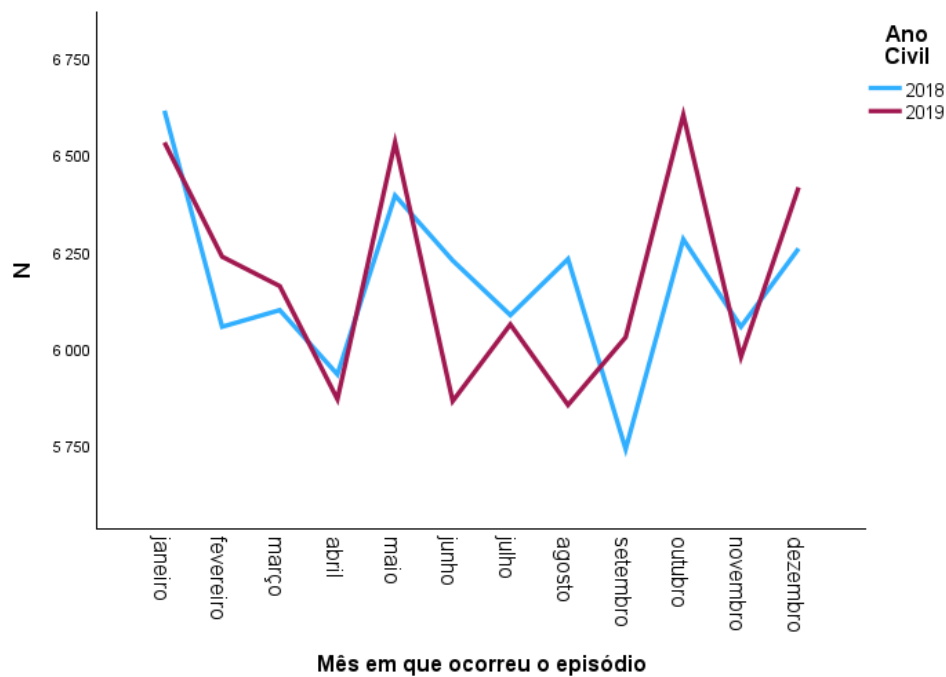


Figura 13: Chegada mensal de utentes ao SU por ano

Em 2018, mais de 50% das admissões ao SU são registadas entre as 9h e as 22h, 82,74%, Figura 14, sendo os restantes 17,26% para utentes que deram entrada entre as 23h e as 8h. No que diz respeito ao ano de 2019, a variação é análoga com a diferença que a maior afluência ocorreu entre as 9h e as 21h (79,03%) e apenas 20,07% entre as 21h e as 8h.

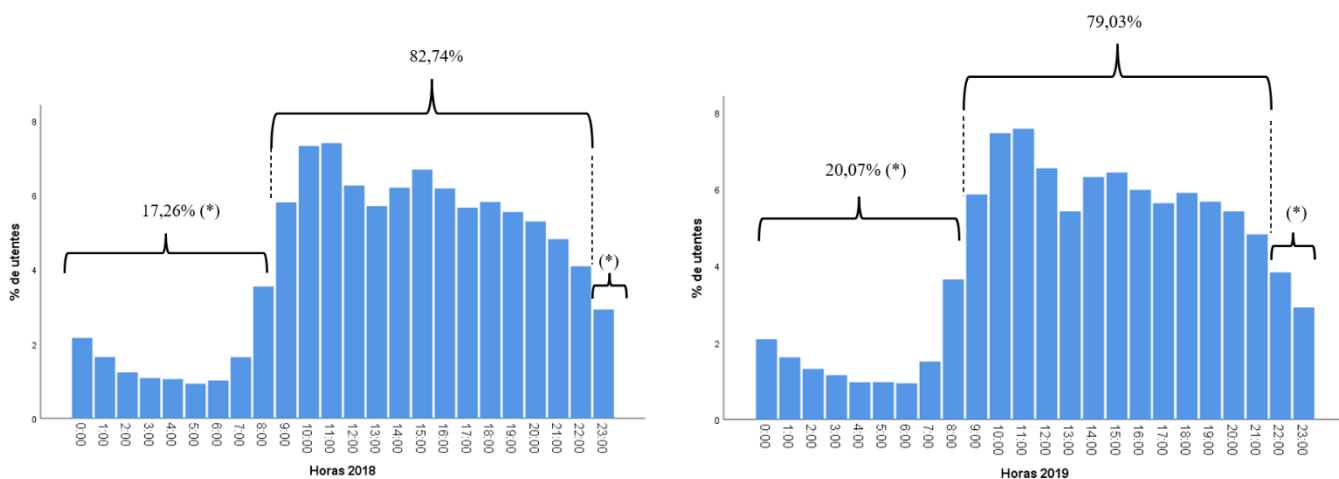


Figura 14: Chegada por hora de utentes ao SU (2018 - 2019)

Nas representações gráficas da Figura 14 está claro que o pico do número de admissões ao SU é por volta das 10h e as 11h da manhã, voltando a ocorrer às 15h sendo os períodos mais “mortos” de madrugada, entre 1h e 7h da manhã, a partir do qual o aumento é bastante acentuado.

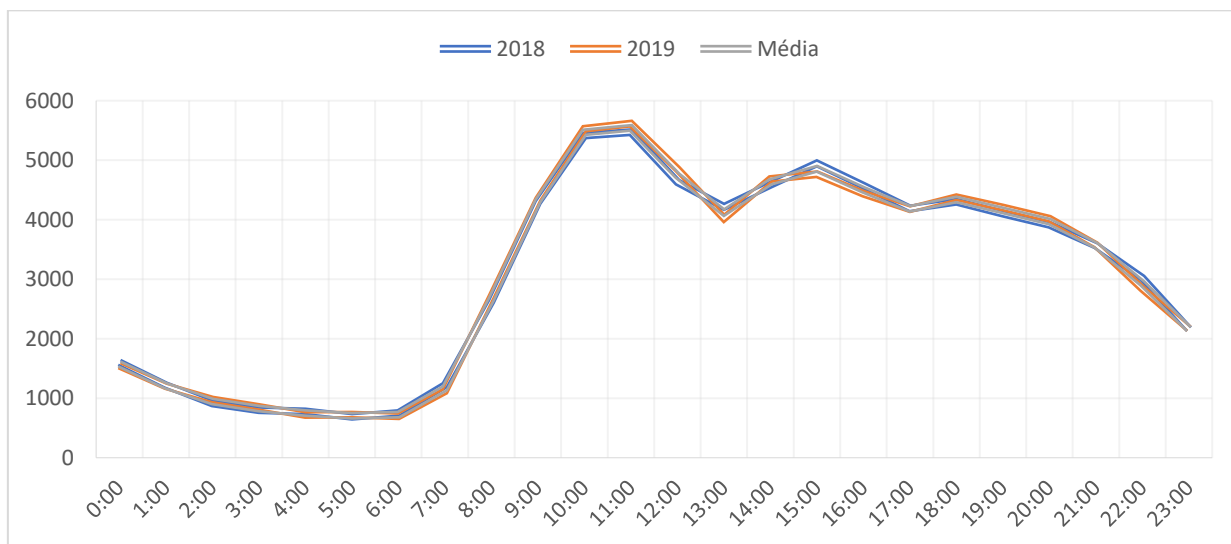


Figura 15: Chegada dos utentes ao SU por hora, 2018, 2019 e média 2018 e 2019

Com a leitura do gráfico da Figura 16, verifica-se que em 2018 a cor identificada por amarelo (43.71%) teve maior registo no número de admissões seguindo-se a cor laranja (35.62%). Para as situações de Azul, os valores são residuais não chegando a 1%. Os códigos brancos e vermelhos registam igualmente valores muito baixos (2.95% e 1.72%, respetivamente).

Para 2019 a cor Amarela é igualmente a que teve maior registo (44.06%) seguindo-se a pulseira Verde (33.23%). Neste ano de pesquisa as cores com menores valores percentuais são o Vermelho (0.47%) e o Azul (1.56%), também com valores residuais. Situação ligeiramente diferente do ano de 2018, comparativamente com as cores de triagem Laranja e Vermelho. Em 2018 houve um maior número de códigos laranjas comparando com 2019, verificando-se o inverso no caso de códigos verdes. Estes resultados mostram que houve um aumento na afluência de não urgentes em 2019.



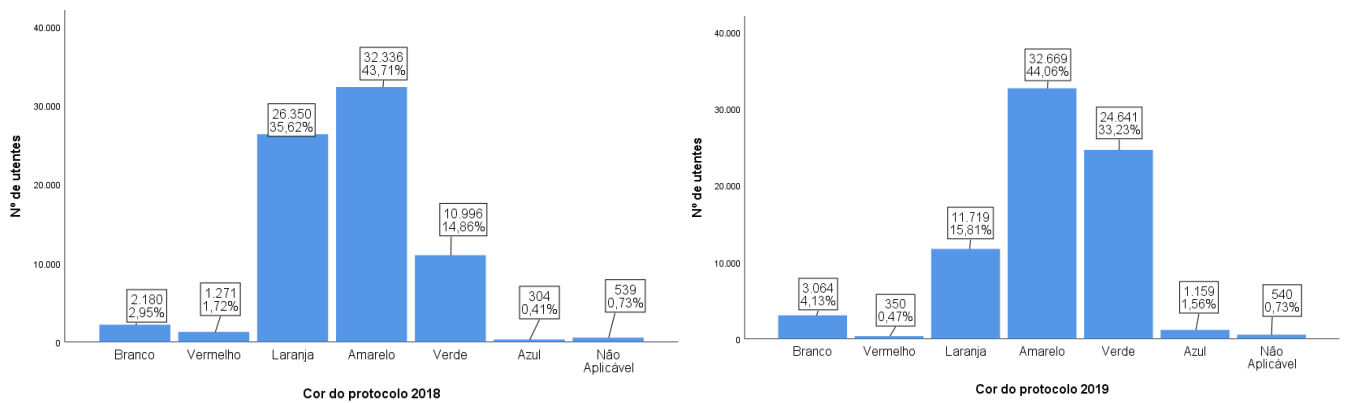


Figura 16: Número de admissões de acordo com a cor da urgência

O Boxplot das idades, Figura 17, por cor do protocolo mostram algumas diferenças no que diz respeito aos dois anos de recolha. Em 2018, para a pulseira com a cor Laranja verifica-se através dos gráficos de box-plot que a afluência ao SU foi essencialmente por jovens/adultos e em 2019 por indivíduos mais velhos. A situação reverte-se quando comparamos a cor verde, ou seja, mediana e quartis muito inferiores em 2019 comparativamente com a mesma cor em 2018.

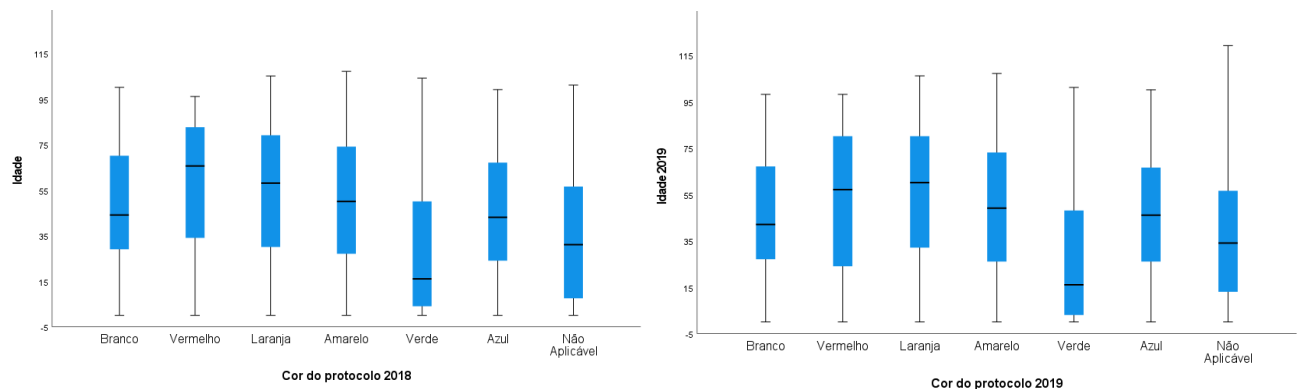


Figura 17: Boxplot da Idade por cor da triagem, por ano

O número de admissões discriminando a cor das pulseiras, tendo em conta o género do utente, pode ser feita através das representações gráficas, Figura 18 e Figura 19. São as mulheres a registar maior percentagem de admissões às urgências, e isto verifica-se para todos os tipos de urgência (cor da pulseira), tanto em 2018 como em 2019. Verifica-se que existe associação entre o género do utilizador e a cor da pulseira ( $\chi^2_{(6)} = 210.52, p < 0.001$ ) para 2018 e também para 2019 ( $\chi^2_{(6)} = 228.77, p < 0.001$ ).

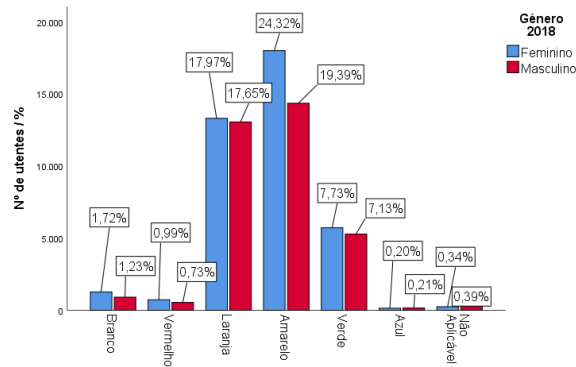


Figura 18: Distribuição das cores por género 2018

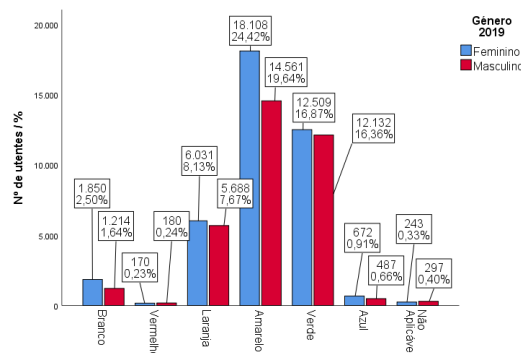


Figura 19: Distribuição das cores por género 2019

A análise das admissões pelo tipo de dia em relação ao dia feriado, mostra que a maior afluência se dá no dia seguinte ao feriado, comparando apenas (véspera, feriado e dia após o feriado). Este facto ocorre em ambos os anos 2018 e 2019, Tabela 6.

	2018		2019		Total	
	n	%	n	%	n	%
Véspera de feriado	2849	3,90%	2528	3,40%	5377	3,63%
Feriado	2873	3,90%	2804	3,80%	5677	3,83%
Pós-feriado	3217	4,30%	3250	4,40%	6467	4,37%
Não se aplica	65038	87,90%	65561	88,40%	130599	88,17%

Tabela 6: Tabela de contagens e percentagens das admissões ao SU pelo tipo de dia em relação ao feriado

Entre 2018 e 2019 verifica-se um decréscimo residual nas admissões nos dias de feriado de 3.9% para 3.8%. O teste do Qui-quadrado foi utilizado para averiguar a relação entre o tipo de dia em relação ao dia feriado e a cor da pulseira, e conclui-se que existem diferenças na afluência discriminada a cor da pulseira e a situação do dia em relação ao dia

de feriado ( $\chi^2_{(18)} = 45.88, p < 0.001$ ) para 2018 e também para 2019 ( $\chi^2_{(18)} = 29.99, p < 0.001$ ).

	2018		2019		Total	
	n	%	n	%	n	%
Não Feriado	71104	96,1	71339	96,2	142443	96,2%
Feriado	2873	3,9	2804	3,8	5677	3,8%

Tabela 7: Tabela de contagens e percentagens das admissões ao SU pelo tipo de dia em relação ao feriado

A distribuição da afluência ao SU, por turno e por ano não apresenta grandes diferenças e o comportamento dos valores é bastante semelhante, no sentido em que o turno 00:00-07:59 tem menor percentagem de admissões, o turno com maior registo é entre 08:00-15:59, seguindo-se o turno 16:00-23:59, Tabela 8.

	2018		2019		Total	
	n	%	n	%	n	%
00:00-07:59	7975	10,78	7838	10,57	15813	10,68%
08:00-15:59	36175	48,90	36513	49,25	72688	49,07%
16:00-23:59	29827	40,32	29792	40,18	59619	40,25%

Tabela 8: Tabela de contagens e percentagens das admissões ao SU por turno

A estação do ano, Tabela 9, com maior volume de admissões é o Outono e isto verifica-se para 2018 e para 2019, seguindo-se a Primavera, depois o Verão e por fim o Inverno. Este comportamento é análogo em ambos os períodos de recolha e poderá estar relacionado com as alterações das temperaturas (podemos ver mais à frente) bem como a qualidade do ar nestas estações.

	2018		2019		Total	
	n	%	n	%	n	%
Inverno	16537	22,4	16727	22,6	33264	22,5%
Primavera	18668	25,2	18521	25	37189	25,1%
Verão	18237	24,7	17785	24	36022	24,3%
Outono	20535	27,8	21110	28,5	41645	28,1%

Tabela 9: Tabela de contagens e percentagens das admissões ao SU por estação do ano

A estação do ano e turno no qual foi feita a admissão ao SU apresenta diferenças significativas, isto é, verifica-se a relação de dependência entre este par de variáveis uma vez que o teste do Qui-quadrado obteve um valor  $p$  de 0.00 ( $\chi^2_{(6)} = 68.67, p < 0.001$ ) para 2018 e da mesma forma para 2019 ( $\chi^2_{(6)} = 92.49, p < 0.001$ ).

No que diz respeito à idade, os valores médios mais elevados registam-se para códigos vermelhos, laranjas e amarelos (M=56, DP=30; M=54, DP=29; M=50, DP=27, respetivamente. São os códigos verdes que registam a idade média mais baixa (M=28, DP=28), Tabela 10. Os dados das Tabelas 10 e 11 mostram uma variação bastante diferente de 2018 para 2019, no que diz respeito à idade dos utentes de acordo com a cor do episódio.

		Cor do protocolo 2018					
		Branco	Vermelho	Laranja	Amarelo	Verde	Azul
Idade 2018	Média	47	56	54	50	28	45
	Mediana	44	66	58	50	16	43
	Desvio padrão	25	30	29	27	28	25
	Percentil 25	29	34	30	27	4	24
	Percentil 75	70	83	79	74	50	67

Tabela 10: Distribuição da cor das pulseiras por idade - 2018

As idades médias mais elevadas, em 2019, ocorreram para episódios laranjas, vermelhos e amarelos (M=55, DP=28; M=51, DP=31; M=49, DP=27, respetivamente).

		Cor do protocolo 2019					
		Branco	Vermelho	Laranja	Amarelo	Verde	Azul
Idade 2019	Média	45	51	55	49	27	46
	Mediana	42	57	60	49	16	46
	Desvio padrão	25	31	28	27	28	24
	Percentil 25	27	24	32	26	3	26
	Percentil 75	67	80	80	73	48	67

Tabela 11: Distribuição da cor das pulseiras por por idade - 2019

A análise da idade dos utentes, 2018 e 2019, de acordo com a cor do protocolo por sexo do utente, pode ser analisada com os dados da Tabela 12 e Tabela 13.

Em 2018 as idades médias mais elevadas registadas pelas mulheres decorrem de códigos azuis (M=63; DP=28) e nos homens para códigos verdes (M=54; DP=29).

			Idade 2018					
			Média	Mediana	Desvio padrão	Percentil 25	Percentil 75	
Género 2018	Feminino	Cor do protocolo 2018	Branco	46	38	23	30	66
			Vermelho	44	41	25	24	66
			Laranja	30	19	28	4	53
			Amarelo	50	50	27	28	75
			Verde	53	55	28	30	81
			Azul	63	74	28	44	85
	Masculino	Cor do protocolo 2018	Branco	50	55	27	25	72
			Vermelho	47	50	26	23	69
			Laranja	26	14	27	3	47
			Amarelo	49	50	27	26	72
			Verde	54	61	29	32	78
			Azul	50	55	31	23	79

Tabela 12: Distribuição da idade do utente por cor do protocolo e por género, 2018

			Idade 2019					
			Média	Mediana	Desvio padrão	Percentil 25	Percentil 75	
Género 2019	Feminino	Cor do protocolo 2019	Branco	45	39	23	29	65
			Vermelho	55	66	31	30	81
			Laranja	55	57	28	31	81
			Amarelo	50	49	27	27	74
			Verde	29	19	28	4	51
			Azul	46	45	24	26	68
	Masculino	Cor do protocolo 2019	Branco	45	48	28	18	70
			Vermelho	47	54	31	17	75
			Laranja	55	61	28	37	79
			Amarelo	48	50	27	25	72
			Verde	25	13	27	3	45
			Azul	46	47	25	25	64

Tabela 13: Distribuição da idade do utente por cor do protocolo e por género, 2019

Para o ano 2019 a idade média mais elevada nas mulheres é registada para códigos vermelhos e laranjas (M=55; DP=31, M=55; DP=28, respetivamente) e para os homens em códigos laranjas (M=55; DP=28).

Numa análise global à idade de acordo com a cor de triagem, e uma vez que as condições de aplicabilidade de uma ANOVA paramétrica não foram verificadas (homogeneidade e normalidade) aplicou-se o teste Games-Howell e dos diferentes valores p verificou-se a existência de diferenças estatisticamente significativas entre pelo menos dois grupos, e ao nível de significância de 5%, Tabela 14.

Grupos (i-j)	Diferença média (i-j)	valor p
Branco - Vermelho	0.28	0.999
Branco - Laranja	-9.73	<0.001
Branco - Amarelo	3.27	<0.001
Branco - Verde	-10.60	<0.001
Branco - Azul	2.14	0.056
Vermelho - Laranja	-10.00	<0.001
Vermelho - Amarelo	2.99	<0.001
Vermelho - Verde	-10.90	<0.001
Vermelho - Azul	1.86	0.368
Laranja - Amarelo	13.00	<0.001
Laranja - Verde	-0.90	<0.001
Laranja - Azul	11.90	<0.001
Amarelo - Verde	-13.90	<0.001
Amarelo - Azul	-1.13	0.565
Verde - Azul	12.80	<0.001

Tabela 14: Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a idade por cor da triagem

Existem diferenças significativas, na idade, entre códigos brancos e códigos laranjas e a estimativa para essa diferença é de 9.73 anos a menos para códigos brancos quando comparados com códigos laranjas. No caso da comparação entre códigos brancos e amarelos, estima-se que a diferença seja de 3.28 anos a mais para utilizadores com códigos brancos em relação a utilizadores com códigos amarelos. Para os restantes códigos a análise é feita de forma análoga e podem ver-se diferenças substanciais entre, por exemplo, códigos laranjas e códigos amarelos, e entre códigos amarelos e códigos azuis.

Foi estudado o padrão de afluência por dia do ano, ao longo de 2018 e 2019 tendo-se verificado dias/meses de maior volume de admissões. Meses como Fevereiro, Junho e Outubro registam um número mais elevado na procura pelo serviço e meses mais quentes (Julho e Agosto) uma menor procura, Figura 20.

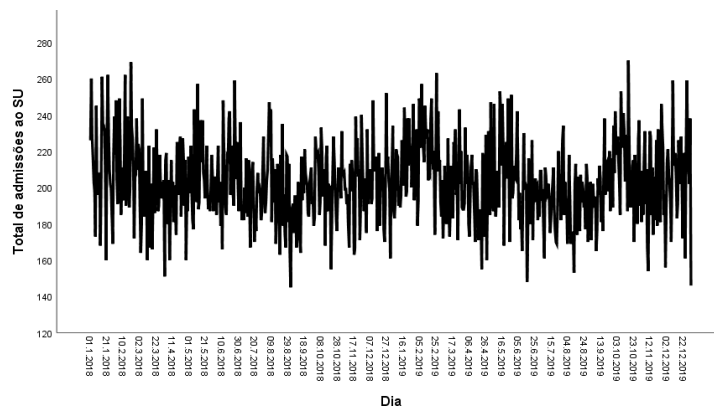


Figura 20: Afluência ao SU de acordo com o dia

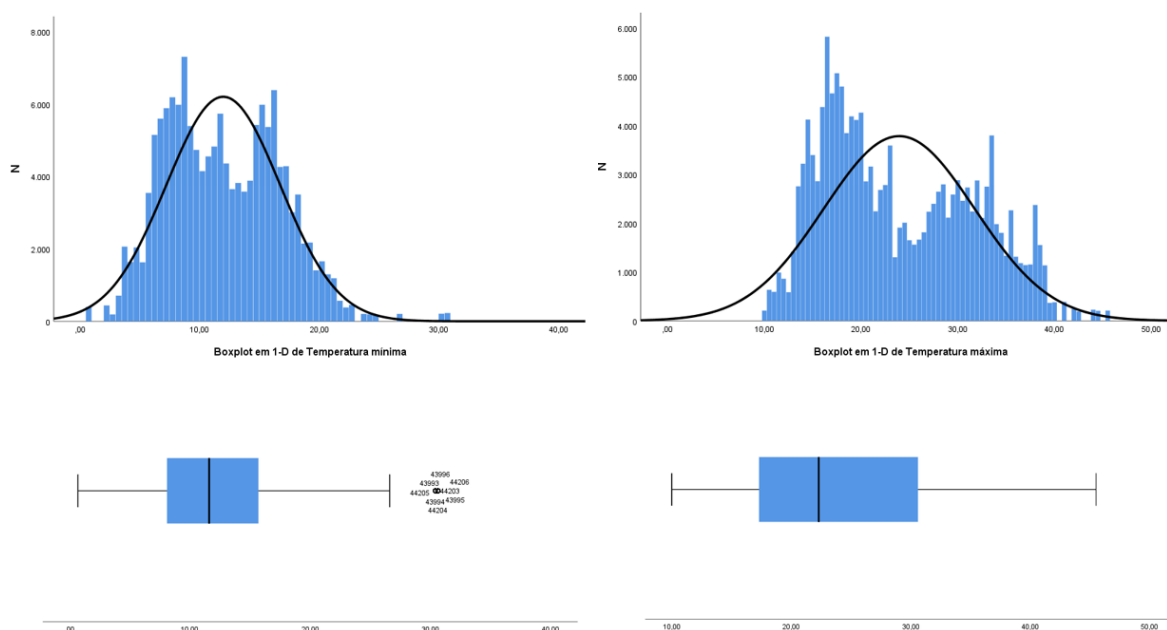


Figura 21: Distribuição das temperaturas mínima e máxima em 2018 e 2019

O Boxplot das temperaturas mínimas registadas apresenta outliers moderados, e correspondem a valores na ordem dos 30°C, mínimos para aqueles dias. O valor do teste de Kolmogorov-Smirnov não valida a normalidade dos dados (valor  $p < 0.001$ ) em qualquer uma das situações.

Em 2018 a temperatura mínima média registada foi à 2ª feira ( $M=11,56$ ;  $DP=5,36$ ) sendo em 2019 à 3ª feira ( $M=11,83$ ;  $DP=4,17$ ). As temperaturas máximas para 2018 registaram um valor médio mais elevado ao Sábado ( $M=23,73$ ;  $DP=8,74$ ) e em 2019 também ao Sábado ( $M=25,27$ ;  $DP=7,32$ ), Tabela 15.

		Dia da semana						
		2ª feira	3ª feira	4ª feira	5ª feira	6ª feira	Sábado	Domingo
Temperatura mínima 2018	Média	11,56	11,85	11,85	11,87	11,81	12,02	12,10
	Mediana	10,20	11,90	12,30	11,70	11,00	11,90	11,10
	Desvio padrão	5,36	5,17	4,99	4,85	4,86	5,60	5,64
	Percentil 25	7,10	7,50	7,70	7,80	8,00	7,10	7,30
	Percentil 75	16,20	16,30	15,60	15,90	15,10	15,80	16,40
Temperatura máxima 2018	Média	23,29	23,18	23,50	23,46	23,61	23,73	23,43
	Mediana	20,10	21,10	20,60	22,30	21,90	20,40	20,70
	Desvio padrão	8,24	7,87	7,74	7,61	8,29	8,74	8,77
	Percentil 25	16,20	16,70	16,90	17,00	16,20	16,70	16,30
	Percentil 75	30,40	29,50	29,70	29,80	31,00	32,30	32,30
Temperatura mínima 2019	Média	12,03	11,83	11,90	12,04	12,13	12,25	12,15
	Mediana	11,70	11,00	11,30	11,50	11,60	11,50	11,90
	Desvio padrão	4,04	4,17	4,09	4,54	4,56	4,22	4,18
	Percentil 25	8,90	8,40	8,90	8,20	8,20	9,40	9,10
	Percentil 75	15,40	15,80	15,70	15,30	15,50	16,00	15,40
Temperatura máxima 2019	Média	24,64	24,55	23,92	23,86	24,16	25,27	24,62
	Mediana	23,20	23,50	21,90	21,20	22,80	25,60	23,80
	Desvio padrão	7,06	7,12	7,49	7,85	7,48	7,32	7,24
	Percentil 25	19,10	19,00	18,00	16,80	18,20	18,90	17,80
	Percentil 75	31,50	30,80	30,00	31,30	29,80	30,20	30,90

Tabela 15: Distribuição das temperaturas mínima e máxima por dia da semana

A Tabela 16 apresenta um resumo dos dados relativos às temperaturas mínimas e máximas nos anos de 2018 e 2019. 2018 registou temperaturas mínimas entre 0,70°C e 30,60°C. Já para o ano de 2019 as mínimas oscilaram entre 3,10°C e 21,60°C. As temperaturas máximas em 2018 atingiram os 45,50°C variando entre 10°C e este valor. Em 2019 as máximas variaram entre os 11,70°C e os 41,20°C, tendo sido registadas no ano 2019 temperaturas mais elevadas (M=24,42; 7,38).

	Temperatura mínima 2018	Temperatura máxima 2018	Temperatura mínima 2019	Temperatura máxima 2019
Média	11,85	23,45	12,04	24,42
Desvio padrão	5,22	8,18	4,26	7,38
Mínimo	0,70	10,00	3,10	11,70
Máximo	30,60	45,50	21,60	41,20
Percentis	25	7,50	16,50	18,30
	50	11,60	21,10	23,00
	75	15,90	30,30	15,40

Tabela 16: Distribuição das temperaturas mínima e máxima (2018 e 2019)



As distribuições das temperaturas mínimas por cor da pulseira encontram-se na Tabela 17. A cor vermelha apresenta um valor médio superior ( $M=12,04$ ;  $DP=4,98$ ) e um valor médio mínimo inferior correspondente à cor Azul ( $M=11,46$ ;  $DP=5,03$ )

Para as temperaturas máximas, também neste ano, a cor da pulseira que registou maior valor médio foi igualmente o vermelho ( $M=23,75$ ;  $DP=7,91$ ) e a temperatura média máxima inferior para códigos azuis ( $M=22,81$ ;  $DP=8,21$ ), Tabela 18.

		Temperatura mínima 2018				
		Média	Mediana	Desvio padrão	Percentil 25	Percentil 75
Cor do protocolo 2018	Branco	11,96	11,65	5,28	7,55	15,90
	Vermelho	12,04	11,70	4,98	7,60	16,10
	Laranja	11,87	11,60	5,19	7,50	15,90
	Amarelo	11,92	11,60	5,25	7,60	15,90
	Verde	11,60	10,90	5,19	7,40	15,60
	Azul	11,46	11,00	5,03	7,20	15,65

Tabela 17: Distribuição das temperaturas mínima por cor da pulseira, 2018

		Temperatura máxima 2018				
		Média	Mediana	Desvio padrão	Percentil 25	Percentil 75
Cor do protocolo 2018	Branco	23,59	21,10	8,23	16,60	30,80
	Vermelho	23,75	21,60	7,91	17,00	30,50
	Laranja	23,46	21,10	8,16	16,50	30,30
	Amarelo	23,58	21,10	8,21	16,60	30,50
	Verde	23,02	20,40	8,13	16,45	29,50
	Azul	22,81	19,95	8,21	16,30	29,90

Tabela 18: Distribuição das temperaturas máxima por cor da pulseira, 2018

Em 2019, a temperatura média mínima mais alta foi de 12,71 ( $M=12,71$ ;  $DP=4,12$ ) correspondente à cor Branca e a média mínima mais baixa correspondente à cor Laranja ( $M=11,90$ ;  $DP=4,31$ ), Tabela 19.

A variação das temperaturas máximas para este ano regista um valor médio mais elevado correspondente à cor Branca ( $M=25,40$ ;  $DP=7,47$ ) e um valor médio mais baixo para códigos vermelhos ( $M=24,15$ ;  $DP=7,10$ ), Tabela 20.

Temperatura mínima 2019						
		Média	Mediana	Desvio padrão	Percentil 25	Percentil 75
Cor do protocolo 2019	Branco	12,71	12,70	4,12	9,40	15,90
	Vermelho	11,96	11,60	4,06	8,80	15,10
	Laranja	11,90	11,40	4,31	8,50	15,40
	Amarelo	12,11	11,60	4,27	8,70	15,50
	Verde	11,93	11,50	4,25	8,50	15,40
	Azul	12,00	11,50	4,16	8,60	15,40

Tabela 19: Distribuição das temperaturas mínima por cor da pulseira, 2019

Temperatura máxima 2019						
		Média	Mediana	Desvio padrão	Percentil 25	Percentil 75
Cor do protocolo 2019	Branco	25,40	24,80	7,47	18,90	31,50
	Vermelho	24,15	22,90	7,10	18,30	30,60
	Laranja	24,19	22,60	7,41	17,90	30,60
	Amarelo	24,56	23,10	7,41	18,40	30,90
	Verde	24,23	22,80	7,31	18,20	30,50
	Azul	24,66	23,20	7,39	18,40	31,00

Tabela 20: Distribuição das temperaturas máxima por cor da urgência, 2019

Para a realização do teste de comparação múltiplas, tendo em conta a cor da triagem e as temperaturas (mínima e máxima), e uma vez que as condições de aplicabilidade não foram validadas para uma abordagem paramétrica, realizou-se o teste de Games-Howell por ser a opção mais robusta nestas situações. Para este teste, e à semelhança ao teste realizado para a variável idade, optou-se por retirar do conjunto de dados a categoria “Não Aplicável”, uma vez que se pretende apenas fazer a análise por cor do protocolo. Para a temperatura máxima existem diferenças estatisticamente significativas, com um nível de significância de 5%, para pelo menos um par de grupos de códigos de triagem apresentados na Tabela 21.

Grupos (i-j)	Diferença média (i-j)	valor p
Branco – Vermelho	-0.82	0.003
Branco – Laranja	-0.96	<0.001
Branco – Amarelo	-0.58	<0.001
Branco – Verde	-0.79	<0.001
Branco – Azul	-0.38	0.551
Vermelho – Laranja	-0.15	0.977
Vermelho – Amarelo	0.24	0.829
Vermelho – Verde	0.02	1.000
Vermelho – Azul	0.44	0.610
Laranja – Amarelo	0.38	<0.001
Laranja – Verde	0.17	0.038
Laranja – Azul	0.58	0.047
Amarelo – Verde	-0.21	<0.001
Amarelo – Azul	0.20	0.919
Verde – Azul	0.42	0.316

Tabela 21: Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a temperatura máxima por cor da triagem

As diferenças apresentadas são significativas, no entanto são diferenças mínimas. Indivíduos que se dirigiram ao SU com códigos brancos foram em dias cuja temperatura máxima apresenta uma estimativa de 0.82 unidades inferior a indivíduos que se dirigiram ao SU com códigos vermelhos. Entradas no Serviço de Urgência com códigos laranjas correspondem a dias em que a estimativa da temperatura máxima é cerca de 0.38 unidades superior a registros com códigos amarelos. De forma análoga, fez-se o teste para as temperaturas mínimas e os resultados estão apresentados na Tabela 22.

Grupos (i-j)	Diferença média (i-j)	valor p
Branco - Vermelho	-0.38	0.057
Branco - Laranja	-0.52	<0.001
Branco - Amarelo	-0.38	<0.001
Branco - Verde	-0.57	<0.001
Branco - Azul	-0.51	<0.001
Vermelho - Laranja	-0.14	0.870
Vermelho - Amarelo	0.00	1.000
Vermelho - Verde	-0.19	0.613
Vermelho - Azul	-0.13	0.965
Laranja - Amarelo	0.14	<0.001
Laranja - Verde	-0.05	0.629
Laranja - Azul	0.00	1.000
Amarelo - Verde	-0.19	<0.001
Amarelo - Azul	-0.13	0.859
Verde - Azul	0.06	0.996

Tabela 22: Comparações Múltiplas do teste de Games-Howell, diferenças médias e valores p para as diferenças significativas, para a temperatura mínima por cor da triagem

Ao nível de significância de 5% conclui-se que indivíduos que foram admitidos ao SU com códigos brancos, fizeram-no em dias cuja estimativa da temperatura mínima foi cerca de 0.52 unidades inferior a indivíduos que se dirigiram ao serviço com códigos

laranjas. Para códigos laranjas, as entradas no SU ocorreram em dias cuja temperatura mínima tem uma estimativa de 0.14 unidades superior a códigos amarelos.

Foi estudado o padrão de utilização do SU em relação à hora de admissão de acordo com a cor da urgência. Nos picos de maior afluência, e à semelhança do que ocorre ao longo das 24 horas, a cor Amarela é a predominante seguindo-se a cor Laranja e logo depois códigos verdes, Figura 22.

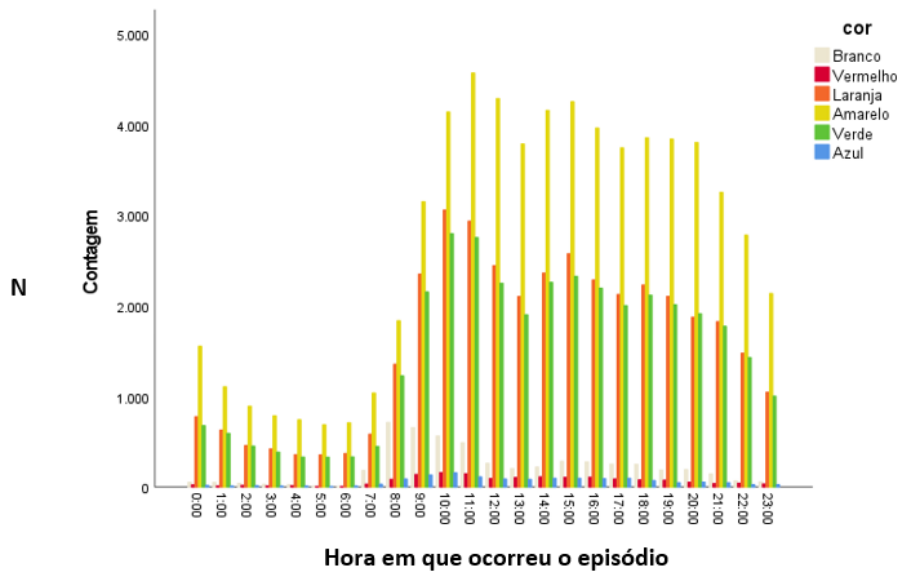


Figura 22: Afluência por hora e por cor

Analisando a prioridade atribuída aos utilizadores do SU, observamos a predominância da cor Amarela nos três turnos, Figura 23. A distribuição da afluência apresentou diferenças significativas comparativamente aos três turnos onde ocorreu o episódio ( $\chi^2_{(10)} = 873.57, p < 0.001$ )

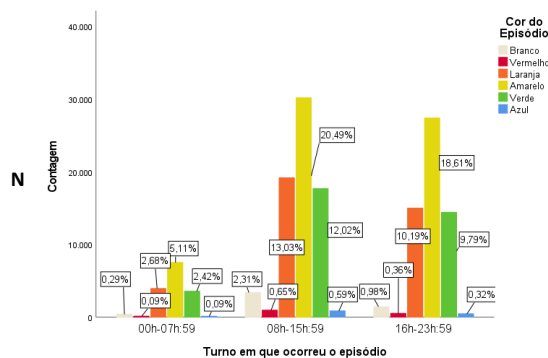


Figura 23: Afluência por cor e por turno

### 5.3 Caracterização da procura do Serviço de Urgência por tipo de utilizador

Muitas pesquisas acerca dos utilizadores das urgências incidem sob o subgrupo dos hiperutilizadores (Hunt et al. 2006). A definição de hiperutilizador varia de acordo com os estudos, mas, em geral, 4 ou mais visitas ao SU por ano é a mais consensual (LaCalle e Rabin, 2010, Doupe *et al.*, 2012, Oliveira, 2008).

Em 2018 e 2019 os hiperutilizadores representam 13,81% dos utentes que visitaram o SU do HESE perfazendo um total de 8714 utentes, Figura 24.

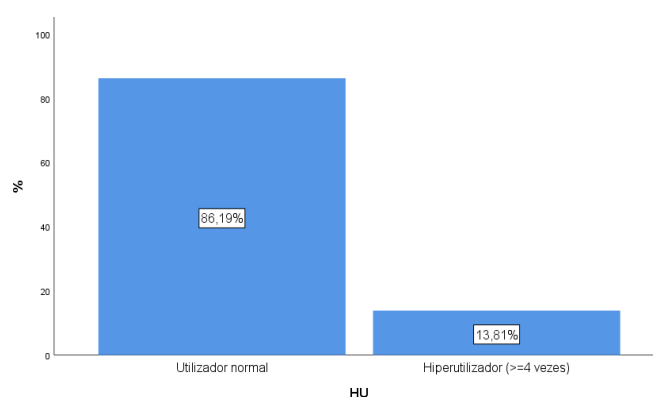


Figura 24: Afluência por utilizador

A idade média de um utilizador normal não ultrapassa os 44,23 anos (M=44,23; DP=26,26), e a idade média de um utente considerado hiperutilizador é ligeiramente inferior (M=41,84; DP=20,99), Tabela 23.

	Utilizador normal	Hiperutilizador
<b>Idade</b>		
Média	44,23	41,84
Mediana	44,00	40,00
Desvio padrão	26,16	20,99
Percentil 25	23,00	26,00
Percentil 75	66,00	56,00

Tabela 23: Idade média por tipo de utilizador

Na análise do volume percentual de idas ao SU, por cor do episódio, e por tipo de utilizador, aumenta quando se analisam os códigos laranjas, vermelho e verdes, Tabela 24. Nos utentes considerados utilizadores normais do Serviço de Urgência a percentagem é mais elevada nos códigos Brancos e idêntica em códigos vermelhos, amarelos e verdes.

Cor do protocolo	Hiperutilizador	Utilizador normal
Branco	14,95%	85,05%
Vermelho	15,77%	84,23%
Laranja	16,73%	83,27%
Amarelo	15,25%	84,75%
Verde	15,59%	84,41%

Tabela 24: Percentagem de admissões por cor, por tipo de utilizador

A variação média da afluência por tipo de utilizador nos diferentes turnos de serviço no Hospital é evidente, Tabela 25, quando se comparam utilizadores normais e hiperutilizadores. Os turnos 08:00-15:59 e 16:00-23:59 têm uma afluência média muito superior ao turno 00:00-07:59, tanto em utilizadores normais do SU como para os hiperutilizadores. Os três turnos apresentam diferenças estatisticamente significativas no que diz respeito à afluência por tipo de utilizador ( $U_{00:00-07:59} = 161172410.5$ ,  $U_{08:00-15:59} = 54507467.5$ ,  $U_{16:00-23:59} = 67245398.5$ ,  $p < 0,001$ ).

	Utilizador normal Média	Hiperutilizador Média
Turno 00:00-07:59	0,16	0,66
Turno 08:00-15:59	0,73	2,96
Turno 16:00-23:59	0,59	2,49

Tabela 25: Média da proporção de admissões ao SU por tipo de utilizador, por Turno

As estações do ano com maior afluência média são a Primavera e o Outono, nos utilizadores normais (M=0,37; M=0,39) e hiperutilizadores (M=1,57; M=1,60), Tabela 26. O teste às diferenças para cada estação do ano, apresentou valores p inferiores a 0,05, pelo que existem diferenças na afluência por mês, considerando o tipo de utilizador do SU.

	Utilizador normal	Hiperutilizador
	Média	Média
Inverno	0,36	1,48
Primavera	0,37	1,57
Verão	0,36	1,46
Outono	0,39	1,60

Tabela 26: Número de admissões ao SU por estação do ano, por tipo de utilizador

A média diária do número de admissões ao SU, por mês e por tipo de utilizador varia, conforme Tabela 27. Um utilizador normal do SU apresenta valores médios que não ultrapassam 0,14 enquanto um utente hiperutilizador tem um registo médio que varia entre 0,48 e 0,54 idas. O mês de Janeiro tem média diária de 0,14, constituindo o mês com maior afluência para um utente normal do SU. Um hiperutilizador regista o maior valor médio diário no mês de Maio.

	Utilizador normal	Hiperutilizador
	Média	Média
Janeiro	0,14	0,53
Fevereiro	0,13	0,49
Março	0,12	0,50
Abril	0,12	0,50
Maio	0,13	0,54
Junho	0,12	0,51
Julho	0,12	0,50
Agosto	0,12	0,50
Setembro	0,12	0,48
Outubro	0,13	0,54
Novembro	0,12	0,50
Dezembro	0,12	0,54

Tabela 27: Número médio de admissões ao SU por mês, por tipo de utilizador

O dia em relação a um dia de feriado foi também analisado, conforme o episódio tenha decorrido de um utilizador ou de um hiperutilizador, Tabela 28. A afluência média é superior nos hiperutilizadores, mas a análise dentro de cada tipo de utilizador não oscila significativamente nos dias antes de feriado, feriado e após feriado. Os dias do ano que não

correspondem a esta categorização do dia em relação ao dia de feriado registaram maiores valores médios em ambos os tipos de utilizador.

	Utilizador normal Média	Hiperutilizador Média
Dia antes do Feriado	0,05	0,22
Feriado	0,06	0,23
Dia após o dia de feriado	0,06	0,27
Dia normal	1,31	5,40

Tabela 28: Número médio de admissões ao SU por tipo de dia, por tipo de utilizador

Também o sexo do utente foi analisado na perspetiva utilizador normal e hiperutilizador, e são as mulheres a registar maiores valores médios na afluência. Uma mulher considerada utilizadora normal do SU recorre, em média, 0,8 vezes a este serviço.

Uma mulher hiperutilizadora apresenta uma média de 3,28, Tabela 29.

	Utilizador normal Média	Hiperutilizador Média
Sexo Feminino	0,80	3,28
Sexo Masculino	0,68	2,83

Tabela 29: Número médio de admissões ao SU por sexo, por tipo de utilizador

#### 5.4 Caracterização da afluência ao Serviço de Urgência 2018 – 2021

Aos dados tratados nas secções anteriores foram adicionados dados relativamente aos anos 2020 e os primeiros cinco meses de 2021, com a finalidade de estudar a evolução da afluência ao SU não apenas numa situação atípica que traz consequências diretas na afluência a um Hospital, mas também de forma a analisar um período temporal maior. Foram também incluídas as variáveis concelho de residência, motivo do episódio e tempo de espera. Estas variáveis não só contribuem para uma melhor noção do tipo de utilizador, mas também para um breve estudo dos tempos de espera tomando a Triagem de acordo com as especificações técnicas presentes no Protocolo de Manchester. Inicialmente começamos por fazer uma análise descritiva do novo conjunto de dados. Pouco mais de 50% dos utilizadores



do SU, neste período são do sexo feminino, com idades médias de 45 anos e 43 anos, para o sexo feminino e sexo masculino, respetivamente, Figuras 25 e 26.

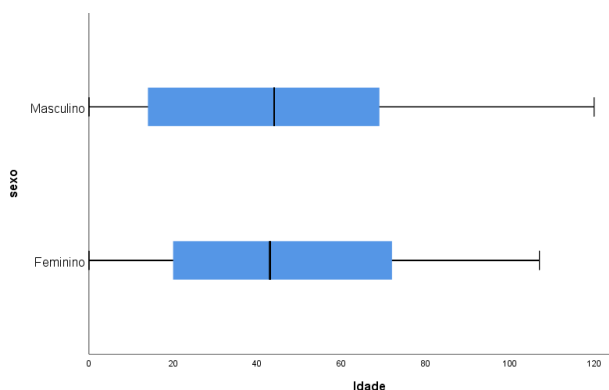


Figura 25: Boxplot da variável idade por sexo

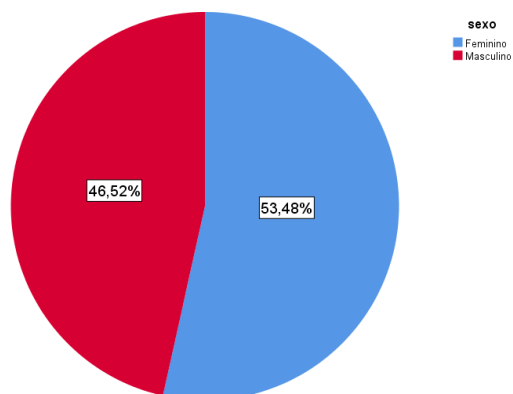


Figura 26: Percentagem de utilizadores por sexo

		Idade						
		Média	Mediana	Máximo	Mínimo	Desvio padrão	Percentil 25	Percentil 75
Sexo	Feminino	45	43	107	0	29	20	72
	Masculino	43	44	120	0	29	14	69

Tabela 30: Estatísticas descritivas da variável Idade por Sexo

Quanto ao motivo de entrada no SU, 11,89% tiveram origem em problemas nos membros seguindo-se 10,28% com problemas de indisposição no adulto, Figura 27.

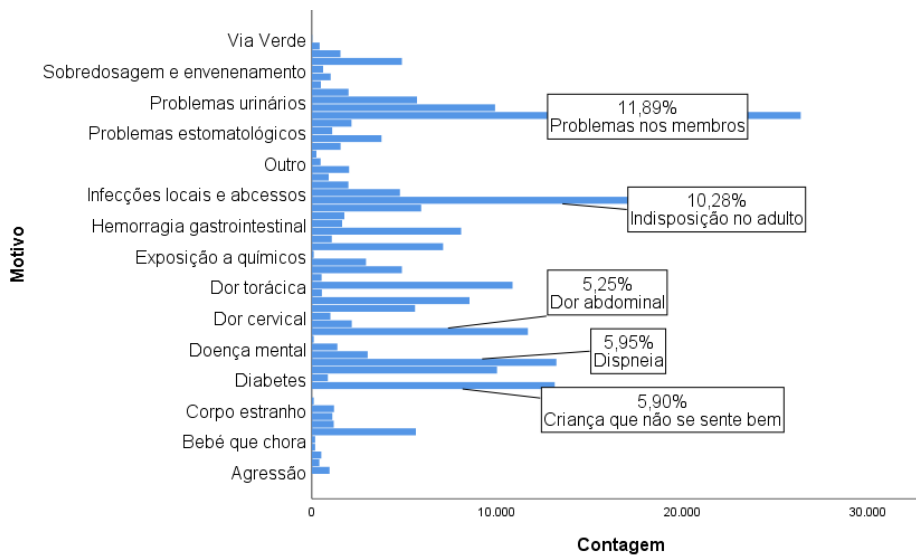


Figura 27: Motivo de admissão ao SU

Os motivos que levam um utente a um Serviço de Urgência devem estar de acordo com o código atribuído no momento da Triagem. De seguida estão representados os valores mais significativos, por tipo de urgência, considerando episódios pouco urgentes ou nada urgentes as cores verde, azul e genérico, Figura 28, e de caráter urgente os episódios amarelos, laranjas e vermelhos, Tabela 31.

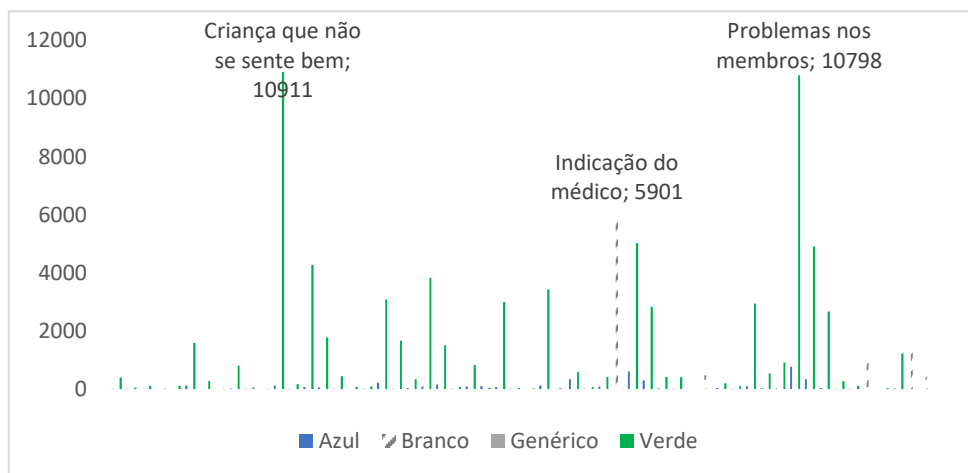


Figura 28: Motivos de admissão ao SU de caráter pouco ou nada urgente

<b>Motivo</b>	<b>Amarelo</b>	<b>Laranja</b>	<b>Vermelho</b>
Agressão	525	26	2
Cefaleia	3278	636	0
Comportamento estranho	611	293	1
Convulsões	758	232	104
Corpo estranho	328	29	5
Criança que não se sente bem	1923	141	13
Diabetes	367	220	107
Diarreia e/ou vômitos	5490	163	1
Dispneia	3942	7306	120
Dispneia na criança	200	2263	106
Doença mental	1099	206	0
Dor abdominal	7505	859	0
Dor abdominal em criança	441	35	0
Dor cervical	591	42	0
Dor de garganta	1623	36	4
Dor lombar	5100	1744	0
Dor testicular	242	214	0
Dor torácica	5451	4457	11
Embriaguez aparente	300	79	0
Erupção cutânea	1398	406	0
Estado de inconsciência / síncope	1861	870	158
Exposição a químicos	30	45	8
Feridas	3271	267	1
Grande traumatismo	538	521	20
Gravidez	3632	3490	8
Hemorragia gastrointestinal	1310	244	2
Hemorragia vaginal	1106	146	0
Indisposição no adulto	11320	5732	126
Infecções locais e abscessos	1604	14	0
Lesão toraco-abdominal	1465	71	0
Mordeduras e picadas	381	117	0
Palpitações	755	691	0
Problemas de ouvidos	698	14	0
Problemas estomatológicos	510	19	0
Problemas faciais	1135	67	1
Problemas nos membros	14094	724	2
Problemas oftalmológicos	3663	805	178
Problemas urinários	2752	226	0
Queda	1419	279	3
Queimaduras profundas e superficiais	232	147	2
Sobredosagem e envenenamento	98	475	4
T.C.E. - Trauma crânio-encefálico	3321	291	6

Tabela 31: Motivos de admissão ao SU de caráter urgente

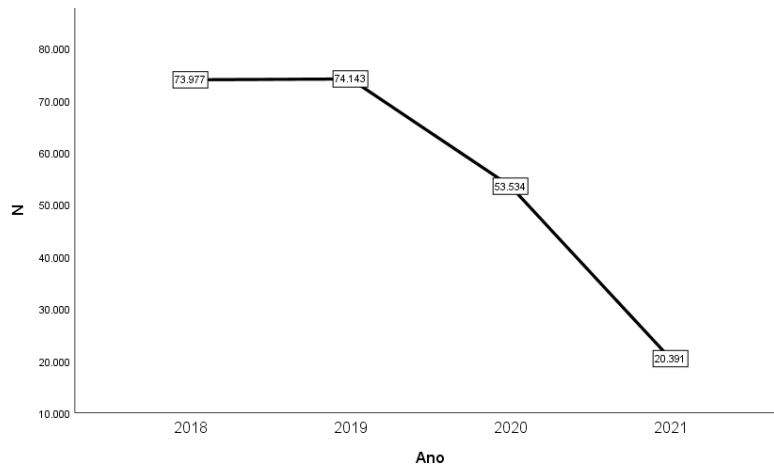


Figura 29: Afluência o SU por ano

O gráfico da Figura 29 reflete a variação na afluência ao Serviço de Urgência, durante os 3 anos e 5 meses em estudo, e é fácil perceber a variação nos anos de pandemia, (comparando com 2020 pois 2021 contemplou apenas 5 meses) comparativamente com 2018 e 2019. 2020 teve um decréscimo significativo a partir de fevereiro, mês em que foi declarada a situação pandêmica no país. Comparativamente com o mesmo período em 2021, a afluência varia no sentido oposto e o volume de entrada no serviço foi significativamente inferior. O aumento em 2021 deu-se à medida que se aproximou o levantamento das restrições no país, ou seja, maio deste ano, Figura 30.

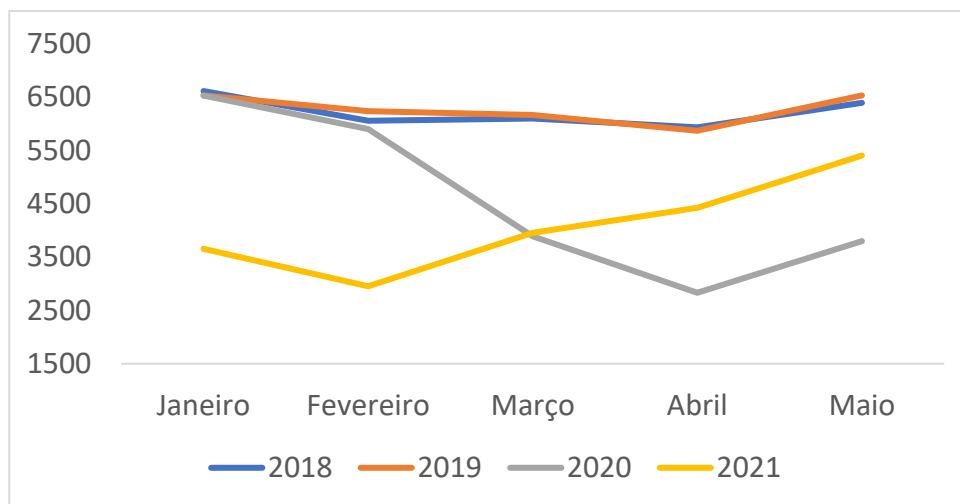


Figura 30: Afluência ao SU por ano nos primeiros cinco meses do ano

O receio do recurso às urgências do hospital, e as restrições impostas levou à diminuição da afluência, o que vem reforçar o conceito de falsas urgências, já identificadas na análise relativa aos anos 2018 e 2019, pré pandemia.

A idade mediana dos utentes manteve-se igual nos anos de 2018 e 2019 aumentando para 47 anos no ano de 2020. Entre janeiro e maio de 2021 teve um aumento para os 50 anos, o que mostra que durante o período crítico da pandemia, indivíduos mais velhos foram mais vezes ao SU, Figura 31.

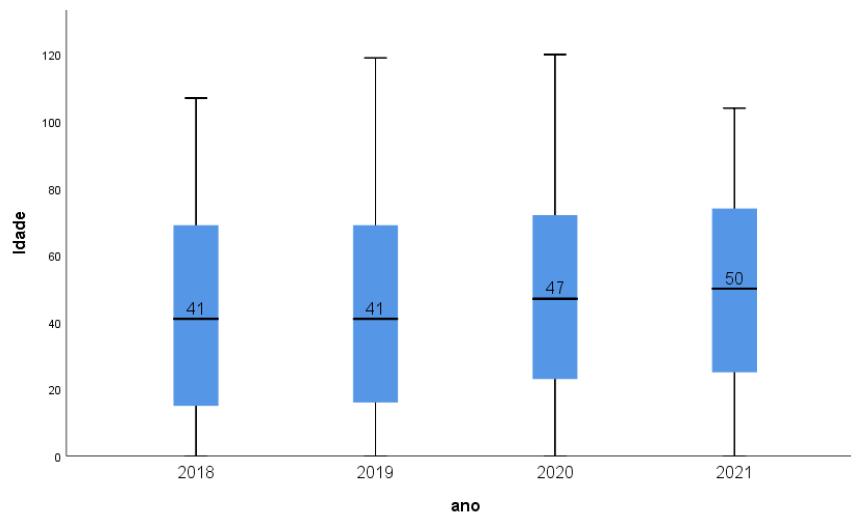


Figura 31: Boxplot da idade por ano

Globalmente o pico horário surge por volta das 10h:00 da manhã, voltando a subir às 15h:00, Figura 33, sendo o período da noite aquele em que ocorrem menos episódios, e por isso o turno das 00h00-08h00 apenas com 10,5% dos episódios registados, Figura 32.

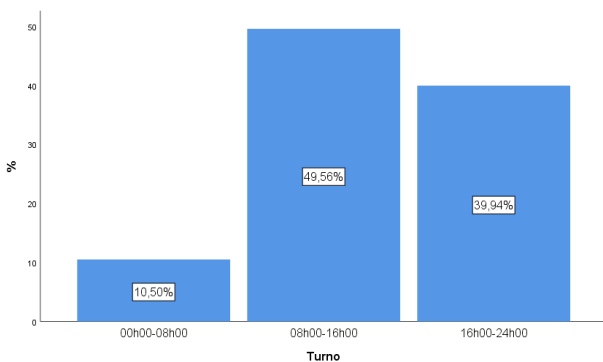


Figura 32: Afluência o SU por Turno

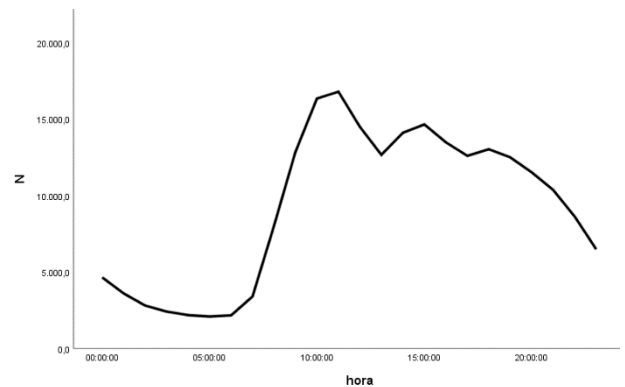


Figura 33: Afluência ao SU por hora

Cerca de 15,79% dos utilizadores recorre ao serviço de urgência à 2ªfeira e o fim de semana é o período com menor registo de episódios representando apenas 26,19% dos episódios, Figura 34.

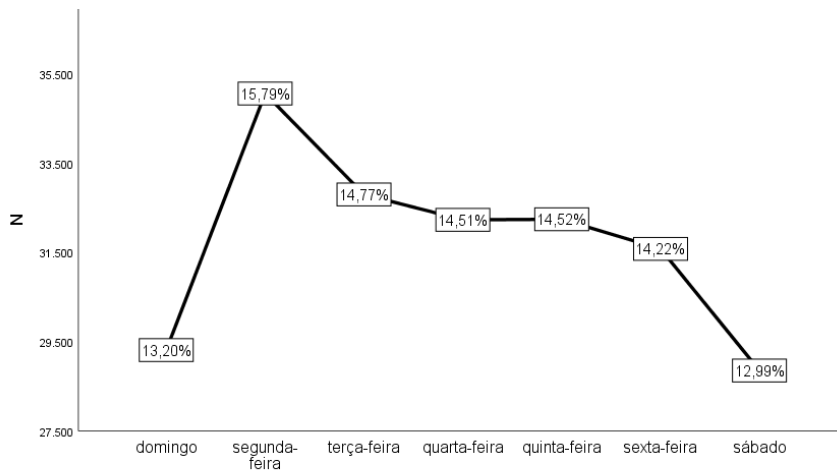


Figura 34: Afluência ao SU por dia da semana

Numa perspetiva de análise ao tipo de dia da semana, classificámos esta variável considerando dias úteis e dias de fins de semana ou feriados, Figura 35. Os dias não úteis (fins de semana ou feriados) representam cerca de 28% da afluência.

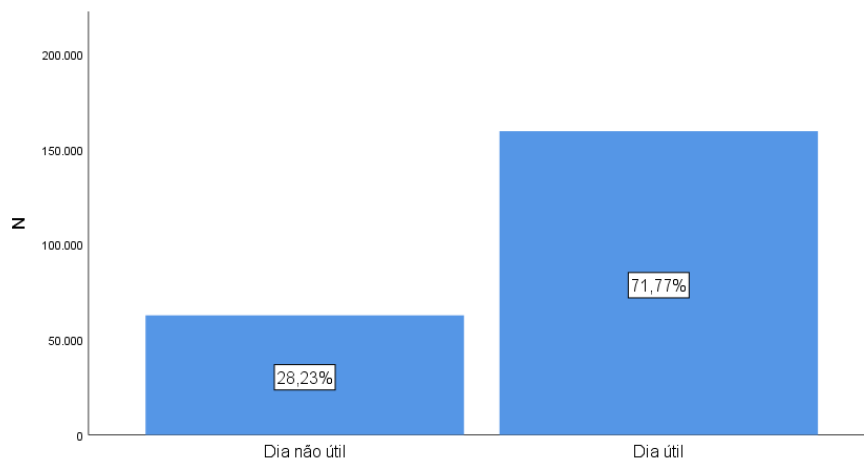


Figura 35: Classificação por tipo de dia

Nos dias úteis são os códigos amarelos a registar maior representatividade com cerca de 43,9% de episódios e para os dias não úteis é também o amarelo a cor com maior abundância (44,6%), Figura 36.

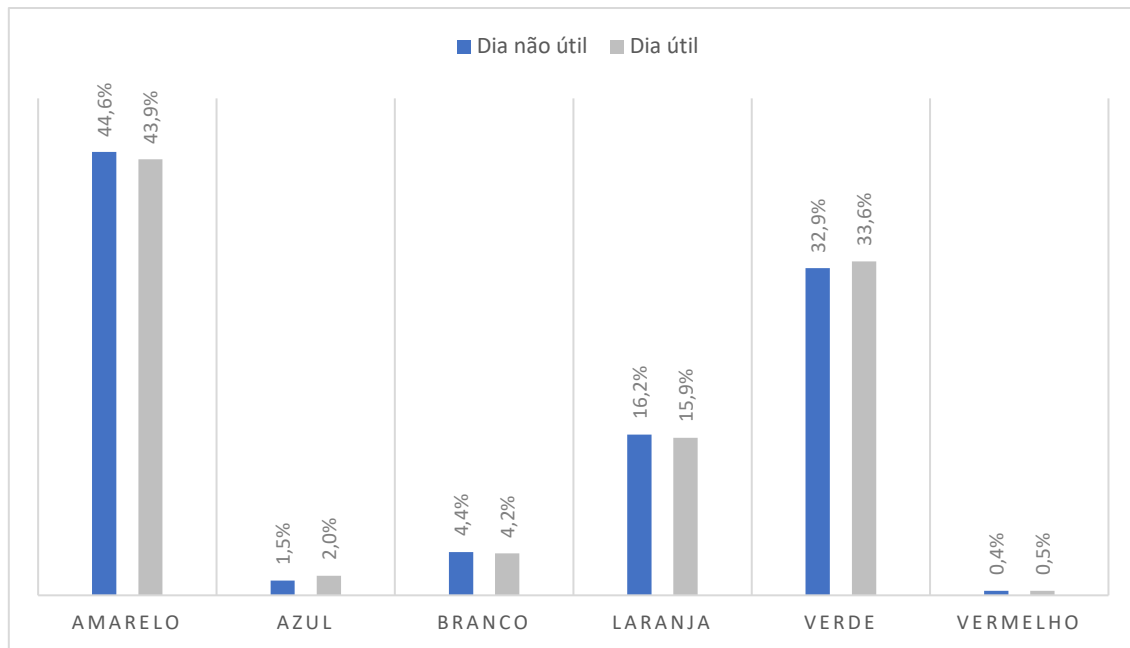


Figura 36: Afluência ao SU por tipo de dia e por código de urgência

O teste do Qui-quadrado de Pearson ( $\chi^2_{(10)} = 72.804, p < 0.001$ ) permite concluir que existe evidência estatística para supor que não existe relação entre o código do protocolo e o tipo de dia.

A variável tempo de espera, que inicialmente estava em horas e minutos, foi convertida para horas, e o tempo mediano de espera até à primeira observação foi de 0,6 horas (36 minutos). 25% dos registos têm um tempo máximo de espera de 0,3 horas (18 minutos), a 50% dos registos teve um tempo máximo de espera de 0,6 horas (36 minutos), correspondentes ao 2º Quartil e o 3º Quartil foi igual 1,3 horas (78 minutos).

O histograma de frequência apresenta uma dispersão que resulta de tempos de espera atípicos (muito elevados) quando comparados com a esmagadora maioria de tempos ocorridos, que se variaram entre as 0 horas e as 5 horas, aproximadamente, Figuras 37 e 38.

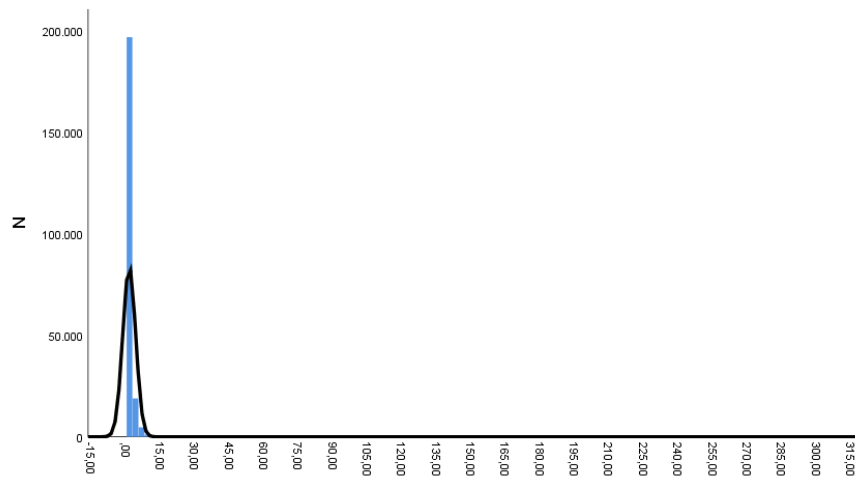


Figura 37: Histograma dos tempos de espera

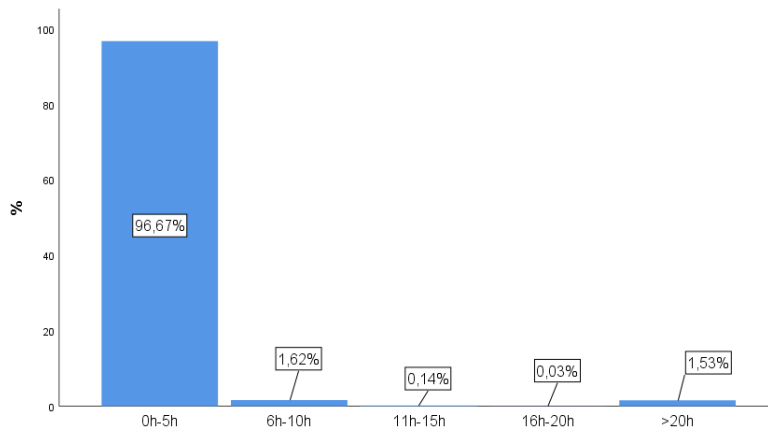


Figura 38: Tempos de espera por intervalos

Para o presente estudo foram considerados apenas tempos de espera inferiores a 12 horas, uma vez que situações em que o tempo excede este limite, poderá ser falha de sistema e/ou erros. De acordo com a classificação do Protocolo de Manchester, representámos os tempos de espera, Figura 39. Mais de 50% dos episódios, tiveram tempo de atendimento correspondente a Urgente, isto é, que durou no máximo 1 hora.



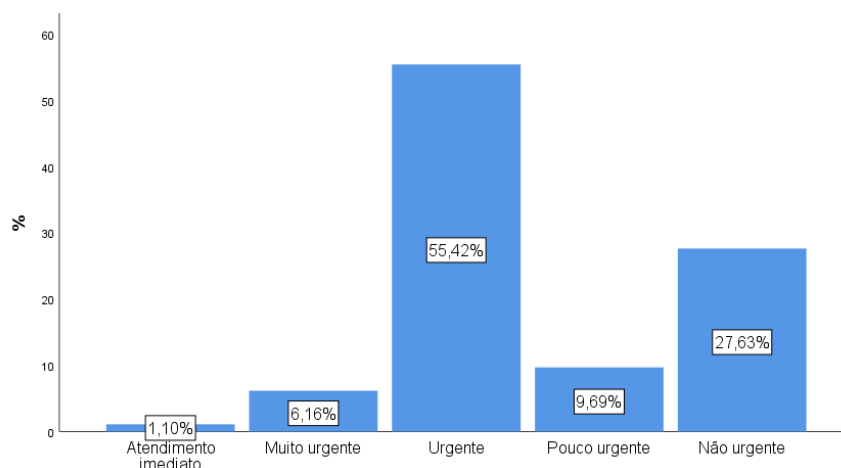


Figura 39: % de episódios de acordo com a classificação do Protocolo de Manchester

Esta distribuição permite apenas perceber como funcionou o atendimento em termos de tempo. Interessa também conhecer os tempos de espera de acordo a cor do episódio que realmente foi registado, Tabela 32.

	Atendimento imediato	%	Muito urgente	%	Urgente	%	Pouco urgente	%	Não urgente	%
Verde	298	0,25	3522	0,26	37594	0,31	8817	0,41	23292	0,38
Laranja	100	0,08	4617	0,34	24026	0,20	454	0,02	5893	0,10
Amarelo	315	0,26	4449	0,33	53859	0,44	11206	0,52	27155	0,45
Vermelho	9	0,01	198	0,01	510	0,00	8	0,00	275	0,00
Azul	75	0,06	102	0,01	1527	0,01	549	0,03	1795	0,03
Genérico	0	0,00	0	0,00	2	0,00	0	0,00	0	0,00
Branco	399	0,33	711	0,05	5330	0,04	449	0,02	2499	0,04

Tabela 32: Valores percentuais dos tempos de espera de acordo com o código de triagem e tempos de espera correspondentes (vermelho – 0 minutos; laranja – 10 minutos; amarelo – 50 minutos; verde – 120 minutos; azul – 240 minutos)

Nos episódios classificados como atendimento imediato, registaram-se mais de 50% com códigos brancos ou verdes. Códigos Laranjas e Amarelos tiveram um tempo de atendimento correspondente à categoria de Muito Urgente, constituindo mais de 50% dos episódios nesta categoria. Cerca de 33% dos registos cujo tempo de espera pertence a situações Urgentes, teve cor amarela havendo e cerca de 34% de utentes cuja urgência devia ter tido um atendimento num tempo mais curto por se tratar de um código Laranja.

Foram registados 271 concelhos de origem dos utentes, entre os quais os que tiveram um valor percentual acima de 1%, em termos de afluência, estão representados na Figura 40.

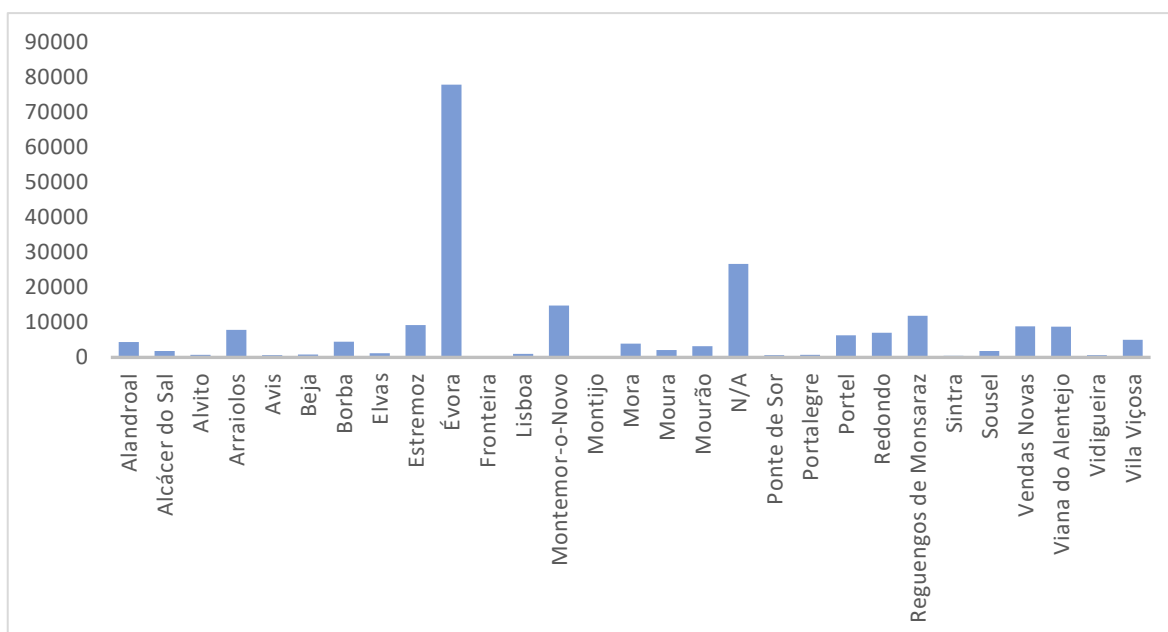


Figura 40: Concelhos com % acima de 1%

Évora registou a maior percentagem de utentes que recorrem ao Serviço de Urgência, como seria de esperar, constituindo cerca de 35,1% dos episódios, assim como concelhos que estão na proximidade, embora se tenham verificado concelhos como Lisboa, Sintra, entre outros, que não pertencem aos 14 municípios que constituem o distrito. Agruparam-se os concelhos de residência dos utentes conforme pertencem a Évora ou não pertencem, e pouco mais de 1/5 (21,7%) são concelhos que não fazem parte do distrito, Figura 41.

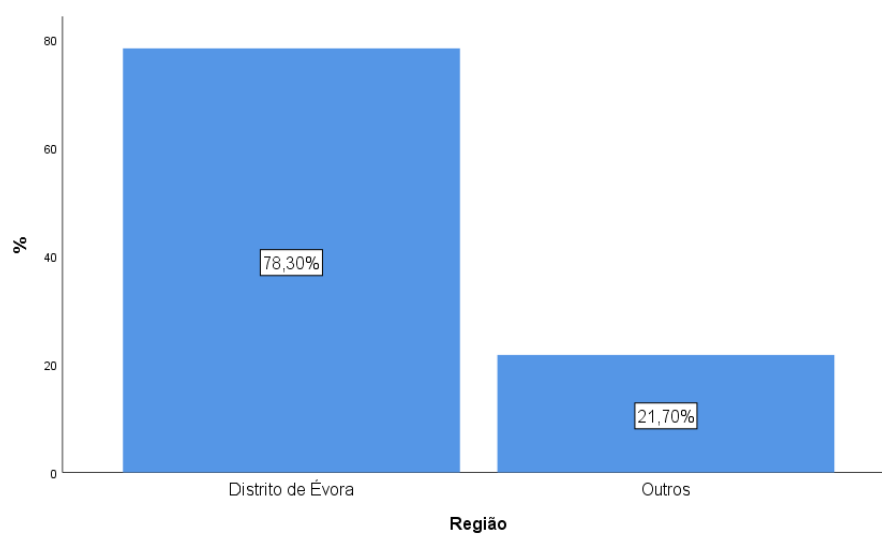


Figura 41: Afluência ao SU conforme pertence ou não ao distrito de Évora

No que diz respeito à afluência de acordo com a triagem feita à chegada, existem diferenças significativas conforme sejam utentes do distrito de Évora ou que residem fora dele ( $\chi^2_{(5)} = 186.739, p < 0.001$ ), o que pode ser justificado pela proximidade em relação ao Hospital.

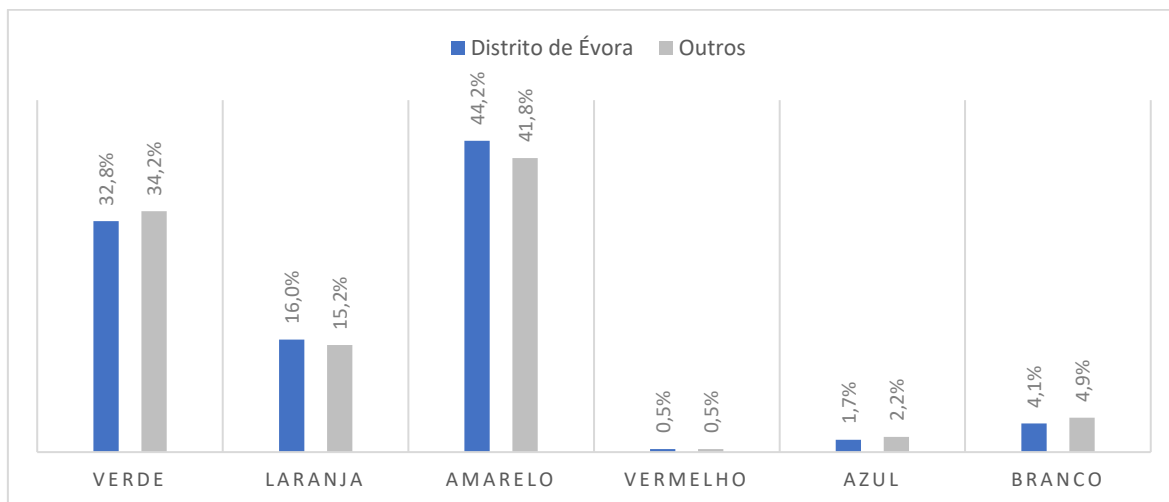


Figura 42: Afluência de acordo com a zona geográfica

A triagem aos utentes de regiões fora do distrito de Évora são, maioritariamente, categorizados por códigos amarelo, laranja ou verde, estando os vermelhos numa percentagem residual, Figura 42.

## 5.5 Caracterização do hiperutilizador 2018 – 2021

Na secção 5.3 foi feita a análise do utente conforme considerado um utilizador normal (menos de 4 idas ao SU por ano), ou hipertulizador (4 ou mais idas ao SU por ano), nos anos 2018 e 2019, e de acordo com algumas das variáveis disponíveis. A análise que se segue contempla dados até maio de 2021, e com as variáveis estudadas na secção 5.4, fazendo a comparação do tipo de utilizador que recorre ao Serviço de Urgência. Foram excluídos 21073 registos por não estarem associados a qualquer número de processo/registo o que não permitiria distinguir entre um utilizador normal do serviço e um hiperutilizador, que constitui o objeto de estudo desta secção. Assim, dos 82693 utilizadores diferentes, que em conjunto deram origem a 200972 entradas no Serviço de Urgência, 18,18% são hiperutilizadores, Figura 43, com idade média igual a 44,43 anos.

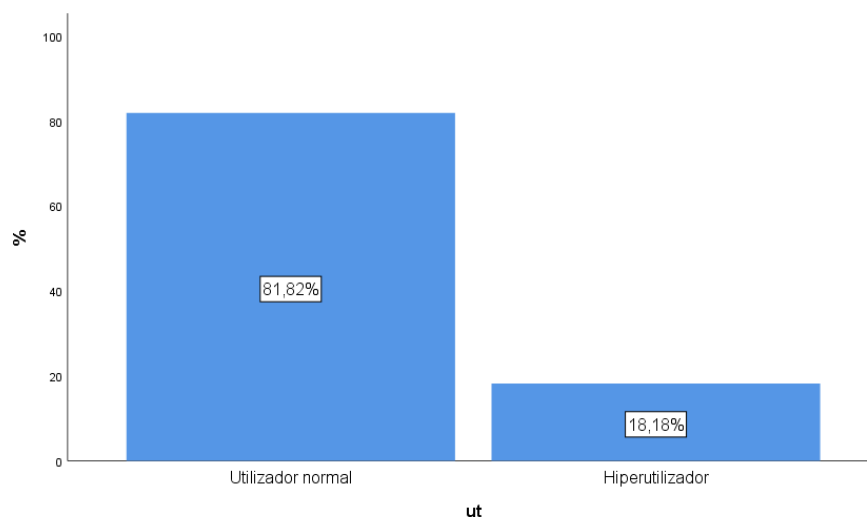


Figura 43: Afluência ao SU por tipo de utilizador

O valor para a mediana do número de vezes que um hiperutilizador recorreu ao SU, é superior durante os três primeiros anos de recolha, e igual 2 vezes em 2018 e 2019 (AIQ=3; AIQ=2, respetivamente). Quando analisamos 2020 verificamos que o valor da mediana reduziu para 1 vez (AIQ=2), Tabela 33.

	Utilizador normal			Hiperutilizador		
	Mediana	Percentil 25	Percentil 75	Mediana	Percentil 25	Percentil 75
2018	0	0	1	2	0	3
2019	0	0	1	2	1	3
2020	0	0	1	1	0	2
2021	0	0	0	0	0	1

Tabela 33: Estatística descritiva da afluência ao SU por ano e por tipo de utilizador

Em 2021 registou a esmagadora maioria de episódios que atingem cerca de 20 idas ao SU num utilizador, que corresponde a um indivíduo do sexo feminino, com 60 anos, do distrito de Évora, e que em 4 anos perfaz um total de 237 idas. As idas desta utente registaram, na maioria das vezes, códigos amarelos ou verdes.

No que diz respeito ao tempo de espera até à 1ª consulta, a mediana do tempo de espera de um utilizador normal foi de 0,68 horas (41 minutos, AIQ=1.02 horas) e de um hiperutilizador igual a 0,95 horas (57 minutos, AIQ=0,83 horas), e há evidência para acreditar que existem diferenças estatisticamente significativas nos tempos de espera, por tipo de utilizador, Tabela 34.

		Utilizador normal	Hiperutilizador	valor p
Tempo de espera em horas	Média	1,17	1,18	<0.001
	Mediana	0,68	0,95	
	Desvio padrão	2,54	1,19	
	Percentil 25	0,35	0,63	
	Percentil 75	1,37	1,46	

Tabela 34: Estatística descritiva e valor p para o teste de Mann-Whitney, efetuada aos tempos de espera por tipo de utilizador

## 5.6 Modelos de regressão logística

Pretendemos conhecer o perfil de um hiperutilizador do Serviço de Urgência do HESE, perceber as suas motivações em termos de idas ao serviço e avaliar a sua associação com algumas variáveis já estudadas. A variável resposta apresenta duas categorias que correspondem aos dois tipos de utilizador considerados: 0 – Utilizador normal do serviço com 67663 registos, e 1 – Hiperutilizador do serviço com 15030 registos.

Recorremos à regressão logística para modelar a probabilidade de um utilizador ser um hipertutilizador em função dos fatores identificados na análise univariada.

As variáveis utilizadas encontram-se na Tabela 35.

Covariável	Tipo de variável
Sexo	Catórica (0 – Feminino; 1 – Masculino)
Idade	Contínua
Tempo de espera em horas	Contínua
Indicador	Contínua
% de admissões Turno 2	Catórica (0 – Não; 1 – Sim)
% de admissões Turno 3	Catórica (0 – Não; 1 – Sim)
% de admissões em dias úteis	Contínua
Concelho	Catórica (0 – Évora; 1 – Fora de Évora a menos de 50km; 2 – Fora de Évora a mais de 50km)
Indicação médica	Catórica (0 – Não; 1 – Sim)

Tabela 35: Variáveis analisadas - classificação

A variável Indicador foi obtida atribuindo um peso aos códigos por tipo de urgência, de acordo com o tempo de espera em minutos associado a cada código, sendo 0 para códigos vermelhos, 10 para códigos laranjas, 50 para códigos amarelos, 120 para códigos verdes e 240 para códigos azuis, e posteriormente calculando a média por utente. Com este procedimento criámos uma variável que serve de indicador da urgência média por episódio por cada utilizador. Para a percentagem de admissões por turno, consideraram-se os três

turnos de trabalho no HESE, e para cada um foi determinada a percentagem de episódios que ocorreu naqueles horários. Posteriormente e considerando a percentagem de admissões nos turnos 2 e 3, categorizamos estas variáveis conforme a percentagem até 50% ou mais de 50%, passando assim a uma variável com apenas duas categorias.

No que diz respeito ao Concelho, consideram-se residentes de Évora, a menos de 50 km de Évora e mais de 50 km de Évora, resultando assim na seguinte categorização 0 – Évora; 1 – Fora de Évora a mais de 50km; 2 – Fora de Évora a menos de 50km.

Entre os 57 motivos de admissão ao SU, registados pelos utentes no período em estudo, agruparam-se em categorias separadas os motivos que somam aproximadamente 80% das admissões totais e os restantes 20% na categorias Outros (Princípio de Pareto 80:20), e que resultou nas seguintes variáveis: Patologias; Agressão; Dores e outros problemas de razão não patológica; Indicação médica; Problemas associados a crianças; Outros motivos. Cada um dos motivos foi transformado numa variável com apenas duas categorias, conforme o utilizador foi ao Serviço de Urgência por dado motivo ou não (0 – Não; 1 – Sim).

Para a variável tempo de espera até à primeira consulta, em horas, não foram considerados para o modelo tempos médios de espera superiores a 12 horas, uma vez que correspondem a valores que não são reais, e que poderão estar no sistema de gestão por erro ou outro motivo (informação obtida por parte do Serviço de Gestão do Hospital). A proveniência do utente foi considerada cada vez que o mesmo se deslocou ao serviço de urgência e teve em conta o centro hospitalar (CH), centro de saúde (CS), INEM (INEM), iniciativa própria (INP), SNS24 e outro. De forma análoga ao motivo de ida ao Serviço de Urgência, estas variáveis foram categorizadas em 0 – Não e 1 – Sim.

Na Tabela 36 estão alguns valores descritivos (médias, medianas, desvios-padrão, máximos e mínimos) para as variáveis contínuas que foram incluídas para os modelos iniciais, e na Tabela 37 as frequências absolutas e relativas (%) para as variáveis categóricas e respetivas categorias, que entraram no modelo.

Covariável	Média	Mediana	Desvio padrão	Máximo	Mínimo
Idade	46.61	47.00	27.44	107.00	0.00
Indicador	66.15	50.00	40.91	240.00	0.00
Tempo de espera (em horas)	1.130	0.74	1.20	11.88	0.01
% de admissões em dias úteis (pdu)	0.72	1.00	0.37	1.00	0.00

Tabela 36: Medidas descritivas para as variáveis contínuas (médias, desvios-padrão, máximo e mínimo)

Covariável	Totais por categorias
Sexo	
Feminino	43870 (53.45%)
Masculino	38203 (46.55%)
Concelho	
Évora	34106 (41.56%)
Fora do concelho a mais de 50 km	24447 (29.79%)
Fora do concelho a menos de 50 km	23520 (28.66%)
Indicação Médica	41 (0.05%)
Sim	82032 (99.95%)
Não	
Admissões no Turno 2	35558 (43.32%)
Sim	46515 (56.68%)
Não	
Admissões no Turno 3	25795 (31.43%)
Sim	56278 (68.57%)
Não	

Tabela 37: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas

Para iniciar a modelação do modelo de regressão logística, realiza-se em primeiro lugar uma análise univariada para avaliar a significância de forma individual das variáveis explicativas. Os modelos univariados com cada uma destas variáveis são dados na seguinte tabela:

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p	OR	IC <sub>95%</sub> OR
Constante	-1.497	0.009	<0.001		
Sexo					
Feminino (referência)					
Masculino	-0.140	0.018	<0.001	0.869	(0.839; 0.901)
Idade	-0.004	0.0003	<0.001	0.996	(0.995; 0.997)
Indicador	0.0002	0.0002	0.323	1.000	(0.999; 1.001)
Tempo de espera (em horas)	0.014	0.007	0.065	1.014	(0.999; 1.029)
Turno 2					
Não (referência)					
Sim	-0.171	0.0183	<0.001	0.842	(0.813; 0.874)
Turno 3					
Não (referência)					
Sim	-0.324	0.020	<0.001	0.723	(0.695; 0.752)
% de admissões em dias úteis (pdu)	-0.069	0.024	0.005	0.934	(0.890; 0.979)
Concelho					
Évora (referência)					
Fora do concelho a menos de 50 km	-0.046	0.022	0.036	0.955	(0.915; 0.997)
Fora do concelho a mais de 50 km	-0.020	0.022	0.356	0.980	(0.939; 1.023)
Admissão por indicação do Médico					
Não (referência)					
Sim	2.266	0.336	<0.001	9.640	(5.095; 19.236)

Tabela 38: Coeficientes estimados ( $\hat{\beta}$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\hat{\beta}}$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95% para os diferentes modelos univariados, 2018-2021.

Com base na Tabela 38, iremos incluir todas as variáveis no modelo multivariado preliminar. A única variável que poderia não entrar nesta primeira fase seria a variável indicador, mas como achamos que clinicamente poderá ser uma variável relevante não a excluimos.

### 5.6.1 Análise Multivariada

Começamos por construir um modelo inicial de efeitos principais usando variáveis explicativas que mostram evidência de serem relevantes quando usadas como preditores únicos, isto é, as variáveis com valor  $p < 0,20$  identificadas acima, e para a variável Concelho juntamos as categorias Fora de Évora a menos de 50km e Fora de Évora a mais de 50km. A variável Concelho passa a ter duas categorias que contemplam utentes de Évora ( $n=67064$ ) e fora de Évora ( $n=15009$ )



Procede-se então à remoção das covariáveis não significativas, por ordem decrescente das suas significâncias e até que no modelo fiquem apenas variáveis significativas a um nível de 10%.

Depois de obtido o modelo com todas as variáveis significativas, analisam-se se existe alguma interação significativa, obtendo-se o seguinte modelo multivariado:

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p
Constante	-0.9212	0.0042	<0.001
Sexo			
Feminino (referência)			
Masculino	-0.214	0.025	<0.001
Idade	-0.001	.0001	0.028
Tempo de espera (em horas)	-0.101	0.020	<0.001
Turno 2			
Não (referência)			
Sim	-0.479	0.022	<0.001
Turno 3			
Não (referência)			
Sim	-0.253	0.038	<0.001
% de admissões em dias úteis (pdu)	-0.055	0.025	0.026
Concelho			
Évora (referência)			
Fora do concelho	0.141	0.036	<0.001
Admissão por indicação do Médico			
Não (referência)			
Sim	11.347	3.359	0.001
Sexo(Homem):tempoespera	0.046	0.015	0.002
idade:tempoespera	0.002	0.0003	<0.001
idade:Pt3	-0.009	0.001	<0.001
idade:concelhoCAT	-0.004	0.001	<0.001
tempoespera:Medico	-1.507	0.515	0.003

Tabela 39: Coeficientes estimados ( $\hat{\beta}$ ) do modelo de regressão logística multivariado para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\hat{\beta}}$ ), valores p (Wald) associados, para o modelo multivariado preliminar, 2018-2021

## 5.6.2 Diagnóstico do modelo

As variáveis mostraram-se significativas no entanto o valor do  $R^2$  de Nagelkerke é de 2.55% é baixo e o teste de Hosmer&Lemeshow mostra que o modelo encontra-se mal ajustado (valor  $p < 0.001$ ), assim como o teste de Cessie-van Houwelingen (valor  $p < 0.001$ ) e com uma capacidade discriminativa baixa ( $AUC=0.599$ , Figura A1.1 em anexo).

O método dos quartis e lowess foram utilizados para averiguar a linearidade no logit e permitiu concluir que podemos assumir que a variável idade é linear, pelo que está na forma funcional correta não sendo necessário neste caso uma transformação. No entanto as variáveis tempo de espera e percentagem de admissões em dias úteis verificaram não ser lineares no logit, (Figuras A1.2, A1.3 e A1.4, em anexo).

Para contornar este pressuposto do modelo, optou-se por categorizar as variáveis contínuas, obtendo-se duas categorias para a variável percentagem de admissões em dias úteis, em que 0 – Não foi ( $n=12823$ ) e 1 – Foi ( $n=69250$ ), isto é, a variável indica se determinado utilizador foi ao Serviço de Urgência num dia útil ou não. O tempo de espera foi agrupado conforme o utilizador tenha ultrapassado, pelo menos uma vez, o tempo de espera na urgência ( $n=36822$ ), ou não tenha ( $n=45251$ ). O modelo apresentou algumas melhorias (o AIC passou de 76820.24 para 72198.61, deviance igual a 4623.6, valor  $p < 0.001$ , e o  $R^2 = 11.34\%$ ), mas ainda com uma capacidade discriminativa baixa uma vez que  $AUC=0.667$ , como se pode ver na curva ROC (Figura A1.5, em anexo), no entanto o teste Cessie-van Houwelingen com valor  $p=0.598$ .

Optou-se por agrupar a variável idade em classes, com amplitude conforme indicação do Hospital, uma vez que foi o método utilizado pelo serviço de gestão, para agrupar as faixas etárias dos utentes, Tabela 40.

Classe	<i>n</i>
[0,20)	17635
[20,45)	20938
[45,65)	17849
[65,75)	8779
[75,84)	8803
[84,107]	8069

Tabela 40: Frequências absolutas para a variável idade agrupada em classes

Depois de retiradas as interações que não se mostraram significativas, obteve-se  $R^2$  de Nagelkerke igual a 13,76% e o teste de razão de verossimilhanças com uma *deviance* de 1321.4, valor  $p < 0.001$ . A capacidade discriminativa teve um aumento residual ( $AUC = 0.698$ , Figura A1.6 em anexo), mas os testes de Hosmer&Lemeshow e Cessie-van Houwelingen mostram um mau ajustamento (valor  $p < 0.001$  em ambos).

A classe [75, 84) foi agrupada com a classe imediatamente acima, obtendo-se os seguintes valores, Tabela 41:

Classe	<i>n</i>
[0,20)	17635
[20,45)	20938
[45,65)	17849
[65,75)	8779
[75,107]	16872

Tabela 41: Frequências absolutas das classes reagrupadas

Uma vez que se impuseram estas limitações no ajustamento aos dados, e devido à complexidade das transformações das variáveis contínuas, sugeridas pelo método dos Polinómios fraccionários no modelo de regressão logístico, optámos por uma abordagem segundo ajustamento de um modelo generalizado aditivo (GAM), com todas as variáveis inicialmente testadas no modelo GLM, que se apresenta na secção seguinte.

## 5.7 Ajustamento do modelo GAM

O ajustamento de modelos GAMs é uma alternativa para a modelação de relações não lineares, e que não apresentam uma forma definida.

Na Tabela 39 apresentam-se os resultados para o ajustamento do modelo GAM, utilizando o mesmo modelo logístico apenas com as transformações suavizadas nas variáveis contínuas, usando 8 nós. O modelo obtido, com as variáveis significativas, depois de retirar a variável *concelho* uma vez que não foi significativa, é o seguinte:

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p
Constante	-0.764	0.060	<0.001
Sexo			
Feminino (referência)			
Masculino	-0.128	0.0267	<0.001
Idade	-0.005	0.001	<0.001
Turno 2			
Não (referência)			
Sim	-0.109	0.032	<0.001
Turno 3			
Não (referência)			
Sim	-0.294	0.036	<0.001
Indicação do médico			
Não (referência)			
Sim	4.138	0.546	<0.001
Covariável	edf	Ref.df	Valor p
s(indicador)	6.849	6.991	<0.001
s(tempoespera)	6.501	6.850	<0.001
s(pdu)	6.993	7.000	<0.001

Tabela 42: Coeficientes estimados ( $\hat{\beta}$ ) do modelo GAM para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\hat{\beta}}$ ), valores p associados, estimativas dos termos suavizados, graus de liberdade (edf) e valores do teste F, 2018-2021.

No ajustamento do modelo GAM as variáveis contínuas são modeladas como uma função de suavização (*splines*) e as restantes de forma semelhante à utilizada num GLM. O desvio explicado, é de 49,4% e a variância dos resíduos igual a 1 (Figura A2.1, em anexo). Na tabela 42 estão representadas as estimativas para os termos suavizados, e estas incluem os graus de liberdade (edf) e os valores do teste F, cuja finalidade é determinar se a função suavizada reduz significativamente o desvio do modelo.

Todos os coeficientes das variáveis explicativas mostraram-se significativos e os valores p para os termos suavizados são igualmente significativos, com exceção do tempo de espera (valor p<0.001).

Nas Figuras A2.2, A2.3 e A2.4 (em anexo) pode ver-se o efeito não linear das variáveis contínuas para 8 graus de liberdade.

Na Tabela 43 apresentam-se os resultados para a adequabilidade dos termos suavizados.

Covariável	k'	edf (df)	k-index	Valor p (k=8)
s(indicador)	7.00	6.849	0.61	<0.001
s(tempoespera)	7.00	6.501	0.96	0.14
s(pdu)	7.00	6.993	0.77	<0.001

Tabela 43: Estimativas para os termos suavizados, graus de liberdade (edf), dimensão ( $k'$ ), valores de proximidade e graus de liberdade (k-index) e valores p.

Os valores apresentados permitem comparar os graus de liberdade (edf) tendo como dimensão máxima ( $k'$ ). Os valores p obtidos (índice  $k < 1$ ) mostram que  $k$  é muito baixo e existe uma grande proximidade entre os graus de liberdade e os valores de  $k'$ .

Comparando o modelo logístico, obtido com a variável Idade agrupada em categorias, conforme também já descrito, obteve-se um melhor ajustamento para o modelo GAM, (AIC= 39528.33, deviance=-31396, valor  $p < 0.001$ ), assim como o coeficiente de determinação  $R^2 = 54.00\%$ , o que se mostrou mais elevado quando comparado com os resultados obtidos no ajustamento via GLM.

### **5.8 Ajustamento de um Modelo Linear Generalizado para o Hiperutilizador – período 1 de janeiro de 2018 a 10 de Março de 2020.**

Na secção anterior foram usados dados referentes ao período completo de recolha, isto é, dados de 1 de janeiro de 2018 até 31 de maio de 2021. Nesta secção apresentamos o ajustamento de um modelo de regressão logística, que contempla dados entre 1 de janeiro de 2018 e 10 de Março de 2020. O período de recolha foi limitado a este intervalo uma vez que qualquer utente, a partir desta data, teve limitações nas idas ao Serviço de Urgência, impostas pela declaração de pandemia. Um indivíduo que até aqui se dirigia ao SU pelos mais diversos motivos, a partir desta data ficou impossibilitado de o fazer com a frequência habitual, e em termos explicativos e preditivos, tem interesse ajustar um modelo que melhor se aproxime da realidade, não contemplando comportamentos atípicos. Obtivemos um total de 66413 indivíduos categorizados em utilizador normal (56502) e hiperutilizador (9911).

As variáveis utilizadas a entrar na análise univariada são as mesmas que utilizamos no ajustamento com os dados do período de recolha total, assim como as suas categorias no

que diz respeito às variáveis categóricas. A variável tempo de espera foi igualmente tomada apenas para tempos até 12 horas.

Para as variáveis contínuas obtiveram-se as seguintes medidas descritivas, Tabela 44:

Covariável	Média	Mediana	Desvio padrão	Máximo	Mínimo
Idade	45.62	46.00	27.73	107.00	0.00
Indicador	66.56	50.00	40.88	240.00	0.00
Tempo de espera (em horas)	1.23	0.80	1.30	11.88	0.01
% de admissões em dias úteis (pdu)	0.71	1.00	0.38	1.00	0.00

Tabela 44: Medidas descritivas para as variáveis contínuas

Na Tabela 45 apresentam-se as frequências absolutas e relativas (%) para as categorias de cada variável categórica, que entram no modelo.

Covariável	Totais por categorias
Sexo	
Feminino	35648 (53.67%)
Masculino	30765 (46.32%)
Concelho	
Évora	25108 (37.81%)
Fora do concelho a mais de 50 km	21317 (32.10%)
Fora do concelho a menos de 50 km	19988 (30.09%)
Indicação Médica	
Sim	3311 (4.99%)
Não	63102(95.01%)
Admissões no Turno 2	37678 (56.73%)
Sim	28735 (43.27%)
Não	
Admissões no Turno 3	44937 (67.66%)
Sim	21476 (32.34%)
Não	

Tabela 45: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas

O teste do qui-quadrado de Pearson foi utilizado para testar a associação entre o tipo de utilizador e as variáveis categóricas que o caracterizam e à afluência. O nível de significância foi fixado em 5%, obtendo-se valor  $p < 0.05$  para todas as variáveis, com

exceção da variável concelho. O modelo univariado com cada uma destas variáveis é apresentado na seguinte tabela:

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	Valor p	OR	$IC_{95\%}OR$
Constante	-1.740	0.011	<0.001		
Sexo					
Feminino (referência)					
Masculino	-0.090	0.022	<0.001	0.914	(0.876; 0.954)
Idade	-0.007	0.0004	<0.001	0.993	(0.992; 0.994)
Indicador	0.001	0.0003	<0.001	1.001	(1.0007; 1.0017)
Tempo de espera (em horas)	-0.007	0.008	0.392	0.993	(0.976; 1.009)
Concelho					
Évora					
Fora do concelho a mais de 50 km	-0.045	0.026	0.087	0.956	(0.908; 1.006)
Fora do concelho a menos de 50 km	0.009	0.026	0.708	1.009	(0.959; 1.064)
Turno 2					
Não (referência)					
Sim	-0.184	0.022	<0.001	0.831	(0.796; 0.868)
Turno 3					
Não (referência)					
Sim	-0.303	0.024	<0.001	0.739	(0.704; 0.775)
% de admissões em dias úteis (pdu)	-0.032	0.028	0.253	0.968	(3.817; 4.420)
Admissão por indicação do Médico					
Não (referência)					
Sim	1.412	0.037	<0.001	4.108	(3.817; 4.420)

Tabela 46: Coeficientes estimados ( $\hat{\beta}$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\beta}$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95%, 2018-2020

Para um valor  $p < 0,20$  consideram-se significativas para o modelo todas as variáveis, com exceção do tempo de espera e a percentagem de admissões em dias úteis.

### 5.8.1 Análise Multivariada

Utilizando a metodologia já descrita no ajustamento do modelo de regressão logística para todos os dados, obtém-se o modelo multivariado com todas as variáveis significativas, incluindo interações significativas. A variável indicador é a primeira a ser removida uma vez que apresenta o maior valor (valor  $p=0.836$ ) e nesta fase todas se revelaram significativas, no entanto o modelo não apresenta melhorias significativas (deviance= 0.044, valor  $p=$

0.834). O ajustamento com a variável tempo de espera, que inicialmente não era significativa, agora apresenta um valor p que permite incluí-la no modelo (valor  $p=0.018$ ), o valor do AIC passou de 53836.25 para 53832.76 (deviance= 5.4934, valor  $p= 0.019$ ), mas ambos com uma capacidade discriminativa baixa, (Figuras A3.1 e A3.2, respetivamente). Na linearidade com o logit a variável idade está na sua forma funcional (Figura A3.4, em anexo) pelo que entra como uma variável contínua sem necessidade de alterações. No caso do tempo de espera em horas, o seu comportamento afasta-se da linearidade e por isso houve a necessidade de a tornar categórica. Este procedimento foi análogo ao já realizado no ajustamento para o período 2018-2021, isto é, consideram-se indivíduos que ultrapassaram pelo menos uma vez o tempo de espera de acordo com o Protocolo de Manchester ou não. A variável tempo de espera categorizada com duas categorias sendo 0 – indivíduos que não ultrapassaram o tempo de espera ( $n= 35243$ , 53.07%) e 1 – indivíduos que ultrapassaram pelo menos uma vez o tempo de espera ( $n= 31170$ , 46.93%). O mesmo acontece para a variável indicador (Figura A3.5, em anexo), e por isso esta variável foi categorizada considerando os seguintes intervalos  $[0,10)$  com 1806 observações,  $[10, 50)$  com 13664 observações e  $[50, 240]$  com 50943 observações.

No ajustamento todas as variáveis foram significativas e este ajustamento apresentou melhorias (AIC= 52884, deviance=931.17, valor  $p<0.001$ ), no entanto a capacidade discriminativa manteve-se baixa (Figura A3.6, em anexo).

Na procura de interações significativas, e depois de as incluir no modelo multivariado, retirando por ordem decrescente as de maior valor p, obteve-se então um modelo que apresentou melhorias (deviance=20101, valor  $p<0.001$ ) quando comparado com o ajustamento sem interações, e com uma capacidade discriminativa superior (AUC=0.695, Figura A3.7 em anexo). Não se obteve um bom ajustamento com o teste de Cessie-Van Houwelingen (valor  $p<0.001$ ) nem com Hosmer & Lemeshow (valor  $p<0.001$ ) pelo que tentámos um ajustamento com a variável idade agrupada com as categorias utilizadas na secção em que foi analisado o período 2018-2021. Todas as variáveis foram significativas, incluindo as categorias para a variável idade, o modelo continuou com uma capacidade discriminativa baixa (AUC=0.684), e o teste de Cessie-Van Houwelingen apresentou um valor que permite validar o ajustamento (valor  $p= 0.0566$ ). Neste modelo obtivemos cinco interações significativas e estando as variáveis categorizadas, a interpretação pode tornar-se confusa e até mesmo perder a sua relevância. A variável idade apresentou algumas categorias não significativas nas interações com a variável percentagem de admissões no turno 2, pelo



que começamos por testar o impacto de a retirar do modelo. Não se conseguiu obter um ajustamento uma vez que valor  $p=0.0019$  para o teste de Cessie-Van Houwelingen. Optou-se por deixar de parte esta interação, e procedeu-se de forma análoga para retirar a interação idade com a variável percentagem de admissões no turno 3 que resultou num modelo que não se ajusta (valor  $p<0.001$ ) e a capacidade discriminativa não se altera significativamente ( $AUC=0.683$ ). O mesmo acontece quando a interação entre a idade e o concelho é retirada ( $AUC=0.682$ ; valor  $p<0.001$ ). Retirou-se a interação entre a idade e o tempo de espera, para ver o impacto em termos de qualidade do ajustamento, e de seguida a interação entre o médico e Pt3 mas incluindo novamente a idade e o tempo de espera, uma vez que foi a que apresentou valores  $p$  mais baixos e, além disso, é uma interação que no contexto do estudo poderá ter relevância na sua interpretação. Em nenhuma das tentativas se conseguiu contornar o problema do ajustamento, tendo o teste de Cessie-Van Houwelingen (valor  $p<0.0001$ ) nem com o Hosmer & Lemeshow (valor  $p<0.0001$ ).

Pela dificuldade em conseguir um bom ajustamento optámos por retirar dos dados todos os registos com idades inferiores a 18 anos. Esta opção veio na perspetiva de um hiperutilizador poder ser um indivíduo adulto e não uma criança, que depende dos pais ou outros, para se deslocar ao Serviço de Urgência. Retiramos também todas as entradas no SU com códigos vermelhos uma vez que um potencial hiperutilizador não recorrerá ao serviço por motivos de urgência imediata.

A metodologia utilizada foi análoga, e obtiveram-se as seguintes medidas descritivas, para as variáveis categóricas, Tabela 47:

Covariável	Totais por categorias
Sexo	
Feminino	28643 (55.50%)
Masculino	22968 (44.50%)
Concelho	
Évora	20108 (38.96%)
Fora do concelho a mais de 50 km	16678 (32.31%)
Fora do concelho a menos de 50 km	14825 (28.72%)
Indicação Médica	
Sim	2465 (4.78%)
Não	49146 (95.22%)
Admissões no Turno 2	
Sim	28315 (54.86%)
Não	23296 (45.14%)
Admissões no Turno 3	
Sim	35954 (69.66%)
Não	15657 (30.34%)

Tabela 47: Totais por categoria e respetivas frequências relativas (%) para as variáveis categóricas

No que diz respeito às variáveis contínuas obtiveram-se os seguintes valores:

Covariável	Média	Mediana	Desvio padrão	Máximo	Mínimo
Idade	56.45	57.00	21.29	106.00	19
Indicador	59.29	50.00	39.49	240.00	0.00
Tempo de espera (em horas)	1.32	0.84	1.40	11.88	0.01
% de admissões em dias úteis (pdu)	0.71	1.00	0.39	1.00	0.00

Tabela 48: Medidas descritivas para as variáveis contínuas

### 5.8.2 Análise Univariada – Período 2018 a 10 de Março de 2020

Os coeficientes no modelo univariado encontram-se na Tabela 49.

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p	OR	IC <sub>95%</sub> OR
Constante	-1.938	0.013	<0.001		
Sexo					
Feminino (referência)					
Masculino	-0.176	0.027	<0.001	0.839	(0.796; 0.884)
Idade	0.008	0.001	<0.001	1.008	(1.007; 1.009)
Indicador	-0.003	0.000	<0.001	0.997	(0.997; 0.998)
Tempo de espera (em horas)	0.024	0.009	0.008	1.025	(1.006; 1.043)
Turno 2					
Não (referência)					
Sim	-0.098	0.026	<0.001	0.907	(0.860; 0.955)
Turno 3					
Não (referência)					
Sim	-0.515	0.032	<0.001	0.597	(0.561; 0.635)
% de admissões em dias úteis (pdu)	0.007	0.034	0.839	1.007	(0.942; 1.077)
Concelho					
Évora (referência)					
Fora do concelho a menos de 50 km	-0.102	0.031	0.001	0.903	(0.849; 0.960)
Fora do concelho a mais de 50 km	-0.148	0.033	<0.001	0.862	(0.808; 0.919)
Admissão por indicação do Médico					
Não (referência)					
Sim	1.433	0.044	<0.001	4.192	(3.841; 4.57)

Tabela 49: Coeficientes estimados ( $\hat{\beta}$ ) do modelo de regressão logística para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\hat{\beta}}$ ), valores p (Wald) associados, razão de chances (OR), e respetivos intervalos de confiança a 95%, correspondentes ao modelo que usou os dados de 2018-2020.

No modelo univariado a variável pdu (percentagem de admissões em dias úteis) não é significativa. Ajustou-se o modelo multivariado, com todas as variáveis significativas ( $p < 0.2$ ) e testou-se o impacto da variável pdu em conjunto com as restantes, e esta continua a não ser significativa. Para o modelo multivariado obteve-se  $R^2 = 0.0608$  e um  $AIC=37402$ , (valor  $p < 0.001$ ), o que evidencia um melhor modelo quando comparado com os valores obtidos antes de separar os dados conforme descrito. Na linearidade com o logit para as variáveis contínuas, concluiu-se que a idade e o tempo de espera se mantêm na sua forma funcional, Figura 44. A variável indicador foi reagrupada em três categorias, sendo as duas primeiras criadas de acordo com o comportamento da variável na linearidade com o logit (Figura 44, iii).

Categorias	n
[0, 10)	1629
[10, 50)	12716
[50, 240]	37266

Tabela 50: Totais para as categorias da variável indicador

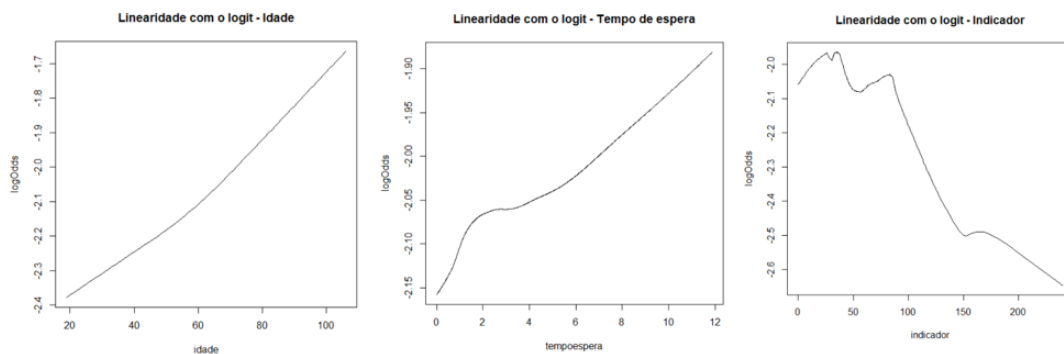


Figura 44: Linearidade com o logit: i) idade; ii) tempo de espera; iii) indicador

Ajustado o modelo com a variável indicador categorizada, obteve-se  $AIC= 36143.13$  (deviance= 1261.3, valor  $p < 0.001$ ) e  $R^2 = 0.105$ , sendo por isto um melhor ajustamento em relação ao modelo com a variável na sua forma original.

Depois de testadas todas interações, foram incluídas no modelo todas as que se revelaram significativas. Para o modelo final obteve-se  $AIC=36178.99$  e  $R^2 = 0.105$ , e os seguintes valores (Tabela 51):

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p
Constante	-3.210	0.174	<0.001
Sexo			
Feminino (referência)			
Masculino	-1,011	0.089	<0.001
Idade	-0.003	0.001	0.006
Indicador_cat			
[0, 10) (referência)			
[10,50)	2.381	0.165	<0.001
[50,240]	1.485	0.165	<0.001
Tempo de espera (em horas)	0.092	0.014	<0.001
Turno 2			
Não (referência)			
Sim	-0.793	0.086	<0.001
Turno 3			
Não (referência)			
Sim	-0.715	0.049	<0.001
Concelho			
Évora (referência)			
Fora do concelho a menos de 50 km	-0.109	0.034	<0.001
Fora do concelho a mais de 50 km	-0.141	0.086	<0.001
Admissão por indicação do Médico			
Não (referência)			
Sim	1.360	0.051	<0.001
Sexo(Homem)*idade	0.014	0.001	<0.001
Sexo(Homem)*tempoHoras	0.059	0.019	0.002
Sexo(Homem)*pt2CAT(Sim)	0.007	0.001	<0.001
idade:pt2CAT(Sim)	0.007	0.001	<0.001
tempoHoras:pt3CAT(Sim)	-0.057	0.021	0.008
pt3CAT(Sim):Medico(Sim)	0.507	0.124	<0.001

Tabela 51: Coeficientes estimados ( $\hat{\beta}$ ) do modelo de regressão logística multivariado para um indivíduo ser um hiperutilizador do SU, respetivos desvios-padrão ( $\hat{\sigma}_{\hat{\beta}}$ ), valores p (Wald) associados

### 5.8.2 Diagnóstico do modelo

No diagnóstico do modelo, começamos por realizar o teste de bondade de ajustamento de Hosmer & Lemeshow para o qual obtivemos, com 8 graus de liberdade, um valor p de 0.1272. O coeficiente de determinação,  $R^2$  de Nagelkerke apresenta um valor de 0.106 e uma capacidade discriminativa de 0.697 (AUC=0.697, Figura A4.1 em anexo), pelo que a capacidade discriminativa é aceitável. A Sensibilidade e Especificidade do modelo permitem avaliar se o modelo é eficiente na classificação dos indivíduos. O modelo apresenta uma Sensibilidade de 73.35% e uma Especificidade de 54.07%. O valor da distância de Cook (Figura A4.3 em anexo) para os diferentes indivíduos não apresenta observações influentes

a considerar, assim como os resíduos deviance (Figura A4.2 em anexo). Os gráficos dos Dfbetas (Figura A4.5 em anexo), permitem concluir que também que não há nenhuma observação a destacar na estimação de cada um dos coeficientes de regressão.

### 5.8.3 Interpretação dos coeficientes

Na Tabela 51 apresentamos os valores dos coeficientes do modelo multivariado ajustado para um indivíduo ser um hiperutilizador do serviço de urgência. Obtidas as estimativas do modelo e a sua validação, pode concluir-se o seguinte:

- Um indivíduo cujo indicador do código de triagem esteja entre 10 e 50, ou seja, códigos laranjas ou amarelos, tem cerca de 11 vezes mais chances de ser um hiperutilizador enquanto um utilizador cujo indicador se encontra entre os 50 e os 240, têm cerca de 4 vezes mais chances de serem hiperutilizadores do serviço. Indicadores que incluem códigos amarelos, verdes ou azuis, têm 4 vezes mais chances de ser um hiperutilizador;
- As possibilidades de um indivíduo ser um hiperutilizador do Serviço de Urgência descem em cerca de 10% se residir fora do concelho de Évora;
- Com o aumento da idade, um indivíduo do sexo masculino tem mais hipóteses de ser um hiperutilizador do Serviço de Urgência;
- Indivíduos do sexo masculino, mais velhos, e que recorrem ao Serviço de Urgência no turno 2, com maiores tempos de espera, têm mais chances de ser um hiperutilizador do Serviço de Urgência;
- Utilizadores que recorrem ao Serviço de Urgência no turno 3, com tempos de espera superiores e por indicação do médico, têm mais de três vezes mais chances de ser um hiperutilizador do que um utilizador que não recorre ao turno 3 nem por indicação do médico e com um tempo de espera inferior;

As diferentes fases da modelação, nomeadamente as dificuldades que foram surgindo, são resultado de um conjunto de dados cujo comportamento é manifestamente diferente tendo em conta as circunstâncias vividas a partir de 10 de Março de 2020. Na Tabela 52 pode ver-se um resumo das principais fases da modelação, todas descritas nas secções anteriores, em consequência das alterações que foram sendo feitas, bem como algumas das limitações que justificaram o não ajustamento do modelo.

<b>Modelo</b>	<b>Classe</b>	<b>Limitações</b>
Dados correspondentes ao período 01/01/2018-31/05/2021	GLM	Não se obtém um bom ajustamento; valor p do teste de Hosmer & Lemeshow muito baixo
Dados correspondentes ao período 01/01/2018-31/05/2021	GAM	Os coeficientes correspondentes às curvas de alisamento não são interpretáveis.
Dados correspondentes ao período 01/01/2018-10/03/2020	GLM	Interações de difícil interpretação; quando retiradas do modelo não se obtém ajustamento.
Dados correspondentes ao período 01/01/2018-10/03/2020 com idades superiores a 18 anos e sem códigos vermelhos	GLM	Obtém-se um bom ajustamento pelo teste de Hosmer&Lemeshow; quando se retiram as interações que não têm interpretação em contexto real, o valor p baixa ligeiramente.

Tabela 52: Fases da modelação, classes do modelo e limitações

## 5.9 Controlo dos códigos de Triagem no serviço 1 de Maio de 2020 até 31 de Maio de 2021

Esta secção é dedicada ao estudo dos códigos de Triagem no serviço de 1 de maio de 2020 até 31 de Maio de 2021. Optou-se por fazer a análise apenas para este conjunto de dados uma vez que é o mais recente dentro do conjunto total.

Começamos por apresentar cartas u para códigos verdes, laranjas e amarelos, por turno, considerando os três turnos, com recurso à metodologia ROSR, Figura 45.

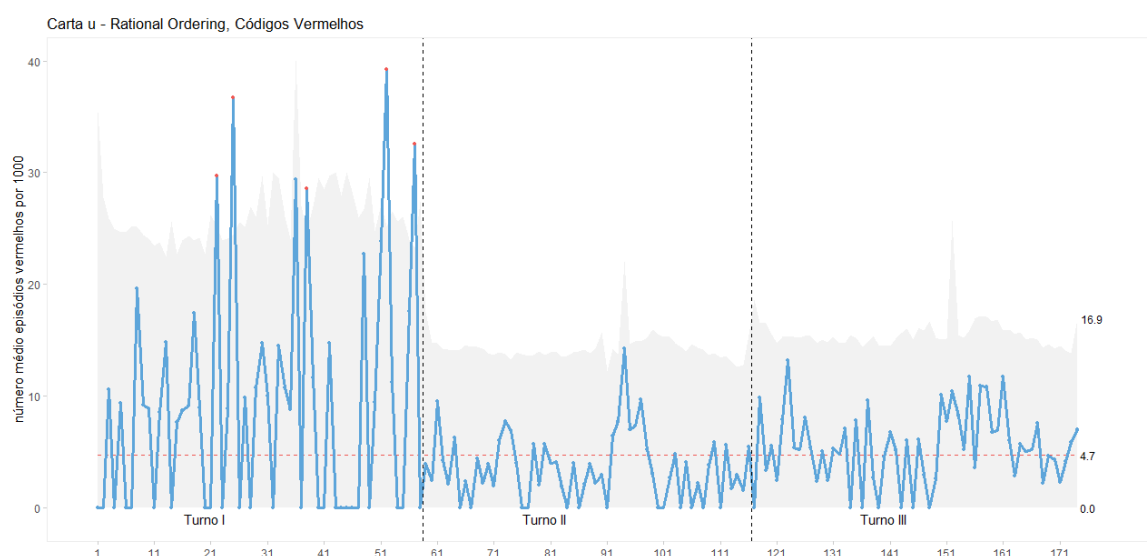


Figura 45: Carta U para códigos vermelhos por período horário (Rational Ordering)

Da observação da carta, Figura 45, podemos concluir que o turno I é uma causa assinalável relativamente aos outros dois turnos, com um número de códigos vermelhos superior

Quando se estratifica por turnos, a linha central varia entre turnos, como se pode ver na Figura 46, sendo esta inferior nos turnos II e III, comparando com o turno I que apresenta um ponto acima do LSC.

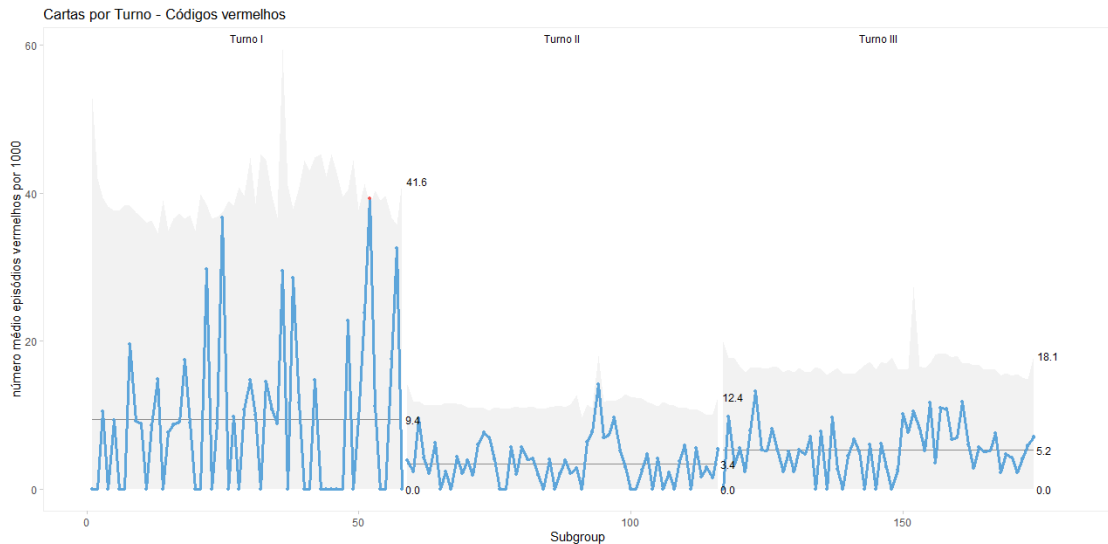


Figura 46: Carta U para códigos vermelhos por período horário (Rational subgrouping)

Relativamente aos códigos laranjas, Figura 47, pode observar-se que o turno I tem praticamente todos os valores acima da linha central, sendo uma clara causa assinalável. No turno II pode observar-se um período com vários valores abaixo do LIC, para além dos três pontos abaixo do LIC. Já no turno III pode observar-se um período com pontos acima do LSC, Figura 47. Podemos observar a carta do turno I sob controlo e com uma linha central bastante superior às cartas dos turnos II e III (que apresentam um período com pontos acima do LSC), Figura 48.

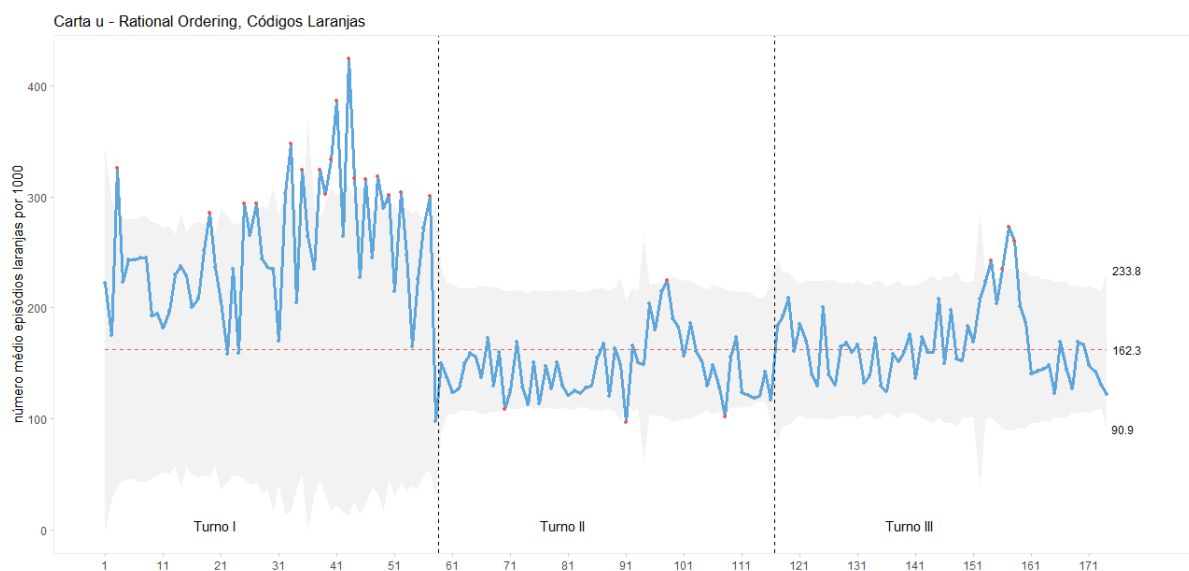


Figura 47: Carta U para códigos laranjas por período horário (Rational Ordering)



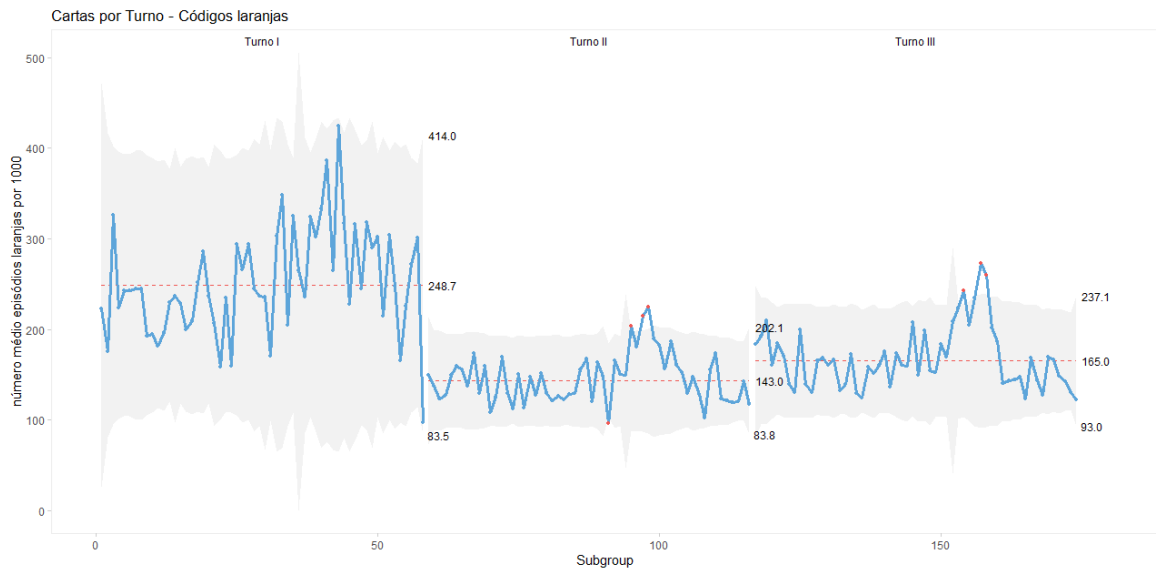


Figura 48: Carta U para códigos laranjas por período horário (Rational sugrouping)

Entre os códigos amarelos nota-se pouco a influência do fator turno. E quando se separam as cartas por turno, pode observar-se proximidade entre as linhas centrais, Figuras 49 e 50.

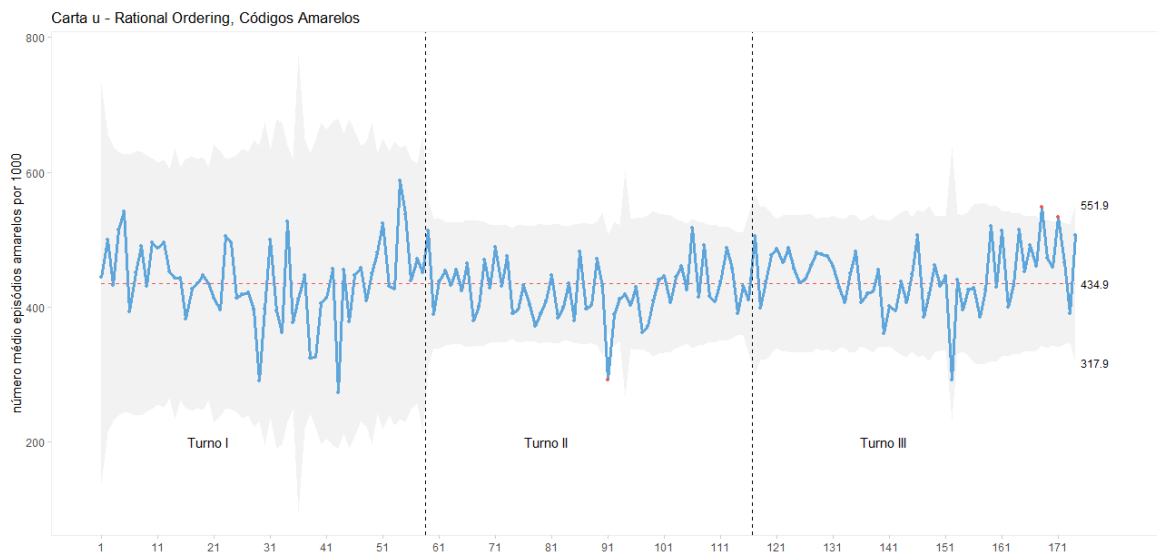


Figura 49: Carta U para códigos amarelos por período horário (Rational Ordering)

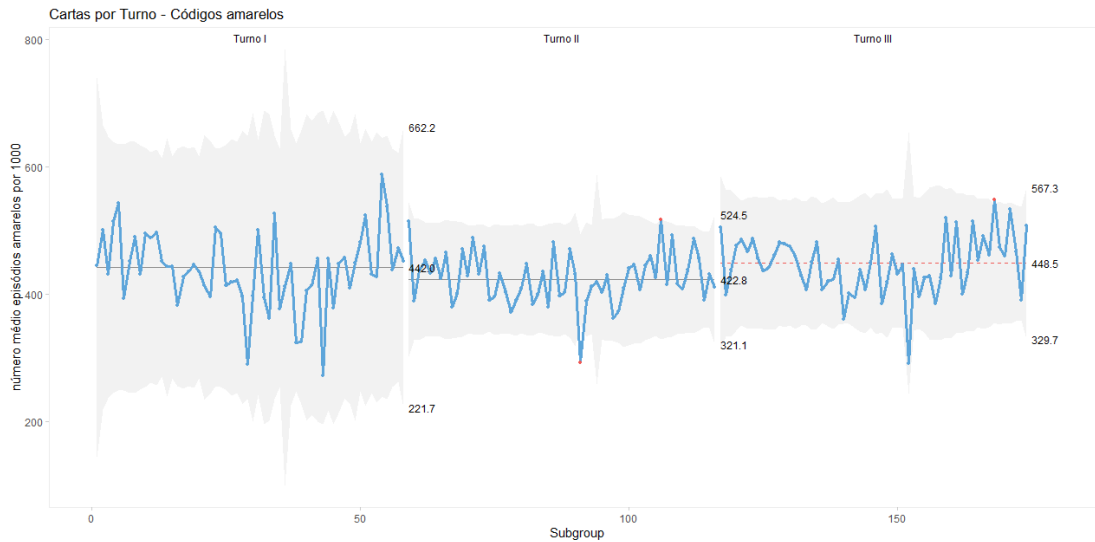


Figura 50: Carta U para códigos amarelos por período horário (Rational subgrouping)

O número médio de episódios verdes apresenta um comportamento oposto ao dos episódios vermelhos e laranjas, Figura 51. O turno I apresenta quase todos os pontos abaixo da linha central, enquanto nos outros dois turnos surgem vários pontos acima do LSC. Quando se separa o processo com uma carta para cada turno, Figura 52, conclui-se que no turno I o processo está estável, com um número médio de códigos verdes próximo dos 228 por 1000, enquanto que nos restantes turnos a média é bastante superior e próxima dos 311 episódios por 1000. Contudo, em ambos os turnos surgem alguns pontos fora dos limites de controlo que deverão requerer investigação por parte do Hospital. Em particular, é evidente o aumento do número de códigos verdes no turno II no segundo e no quarto trimestre, Figura 52.

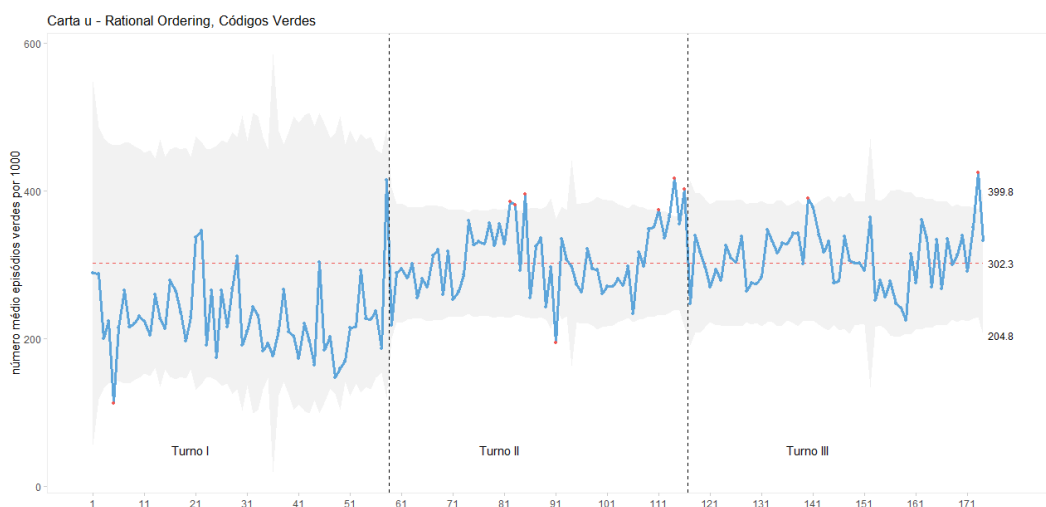


Figura 51: Carta U para códigos verdes por período horário (Rational Ordering)

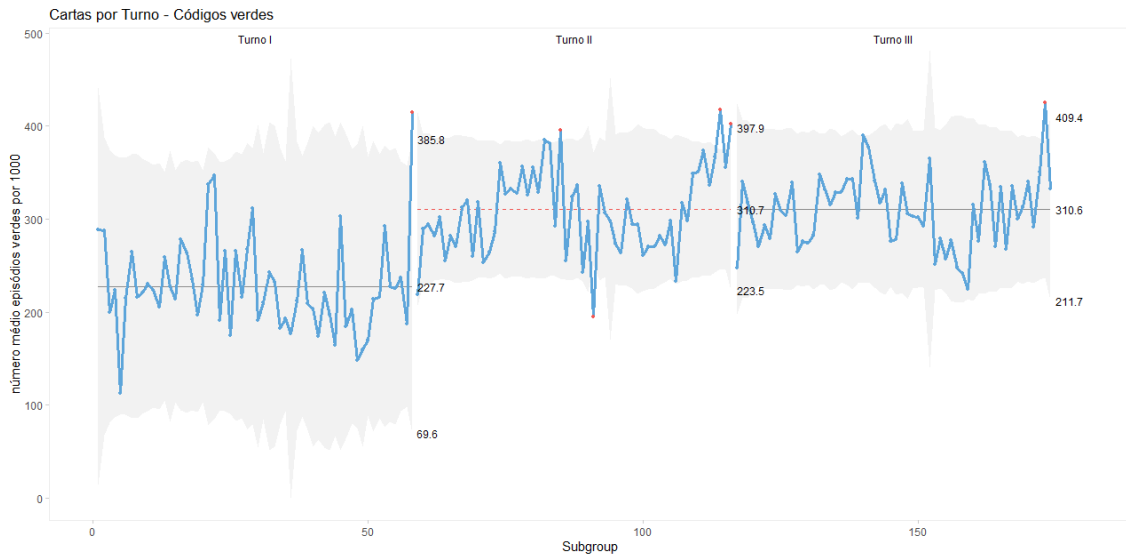


Figura 52: Carta U para códigos verdes por período horário (Rational subgrouping)

Aplicando a metodologia RORS ao dia da semana, as Figuras 53 a 58 representam o número de códigos por 1000 episódios. Uma vez que não se registaram códigos vermelhos em todos os dias, e aqueles que se registaram foram muito poucos, originando muitos zeros no conjunto de dados, não se justifica a construção de uma carta para monitorização de códigos vermelhos. Para códigos laranjas, podemos observar que com exceção da segunda-feira, todos os dias da semana têm pontos acima do LSC, e quando separamos o processo por dia da semana, o domingo é o dia com maior número médio de admissões (próximo de 172 por 1000), Figura 54.

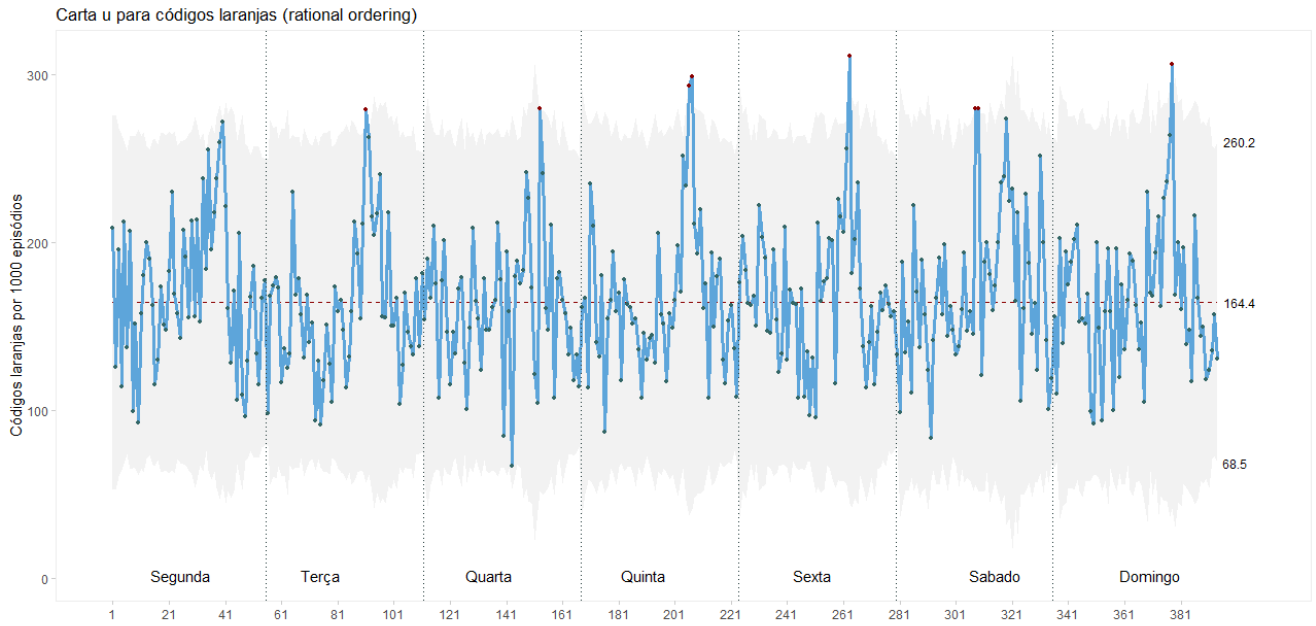


Figura 53: Carta U para códigos laranjas por dia da semana (Rational Ordering)

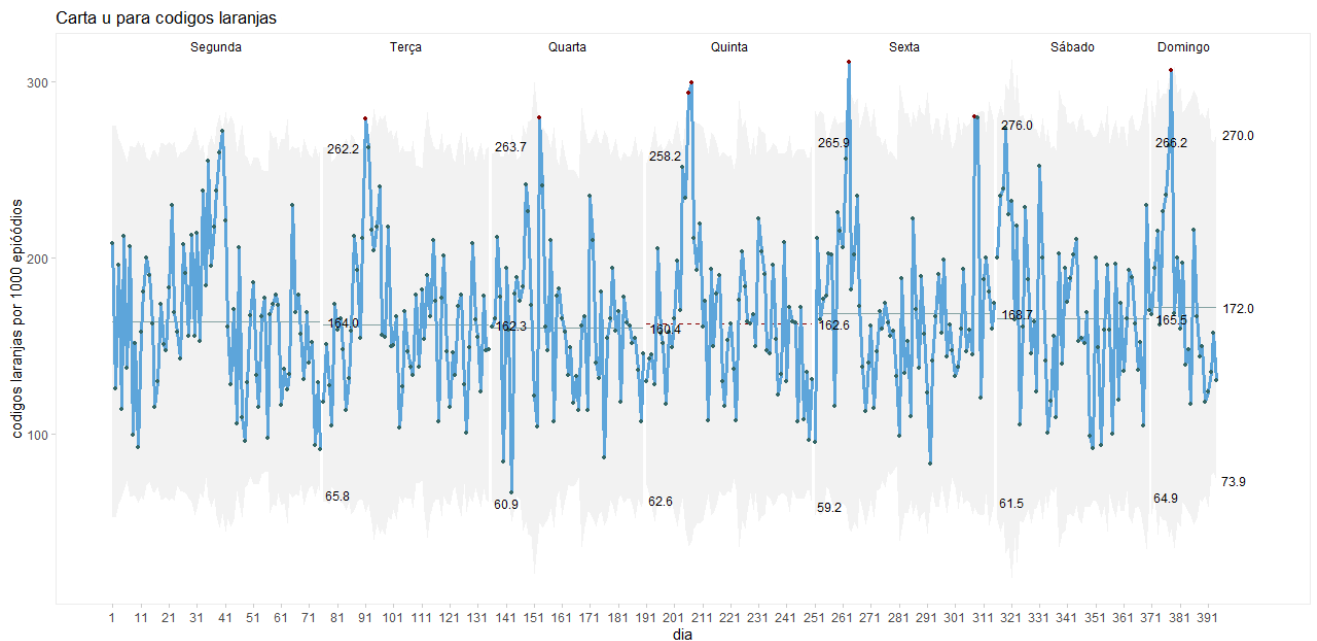


Figura 54: Cartas U para códigos laranjas para cada dia da semana

Para códigos amarelos, Figura 55, o processo encontra-se estável e quando separamos o processo por dia da semana, Figura 56, verifica-se que o número médio de episódios reduz à quinta-feira para 430.3 por 1000.

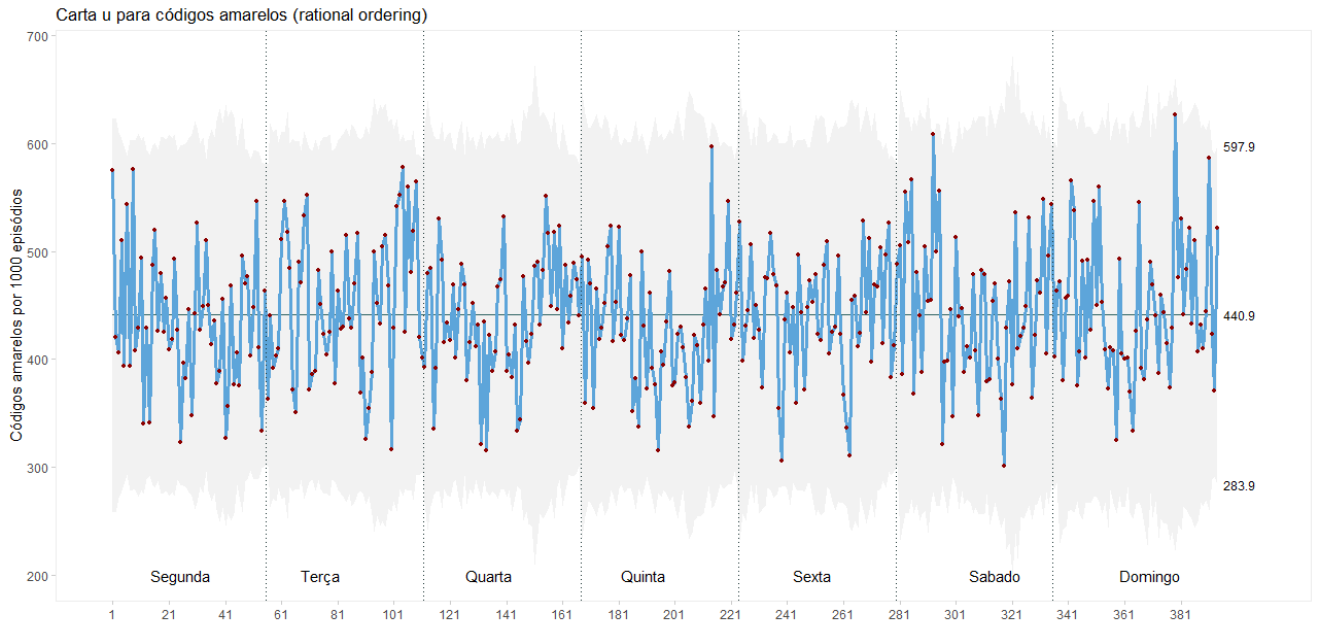


Figura 55: Carta U para códigos amarelos por dia da semana (Rational Ordering)

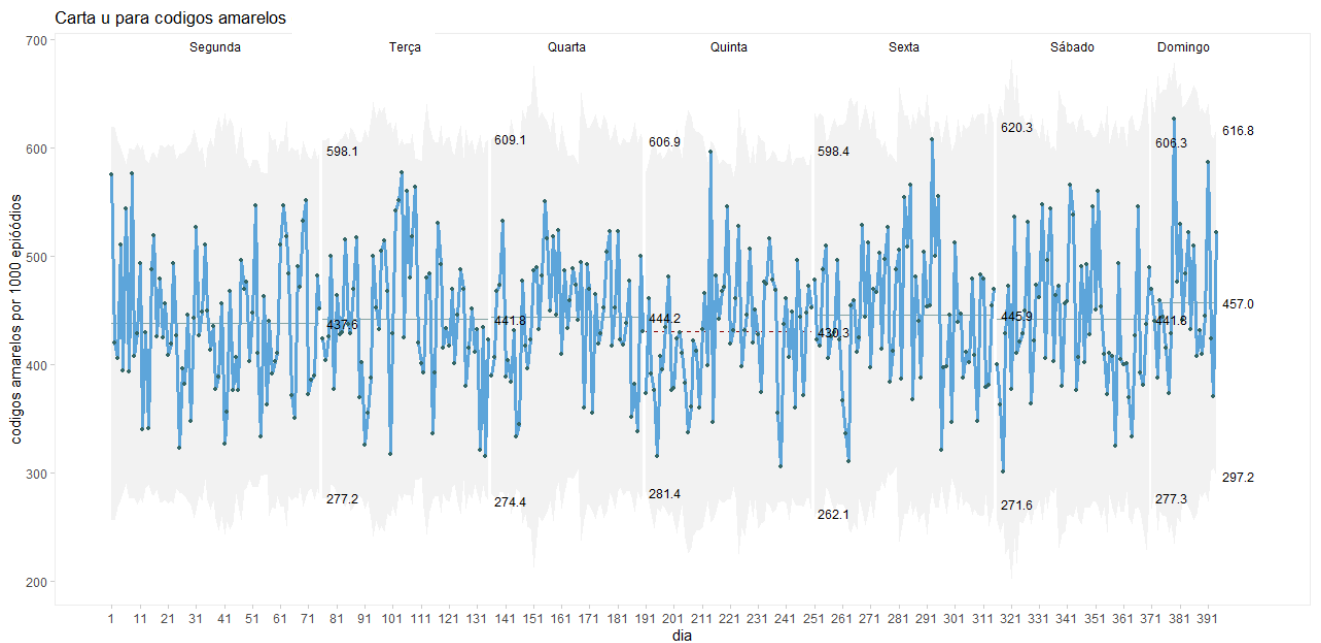


Figura 56: Cartas U para códigos amarelos para cada dia da semana

Códigos verdes, Figura 57, apresentam alguns pontos acima do LSC para a segunda, quarta e sexta e a quinta-feira sofre uma alteração oposta àquela que foi verificada em códigos amarelos uma vez que códigos verdes têm um aumento do número de episódios com a linha central igual a 322.2 por 1000.

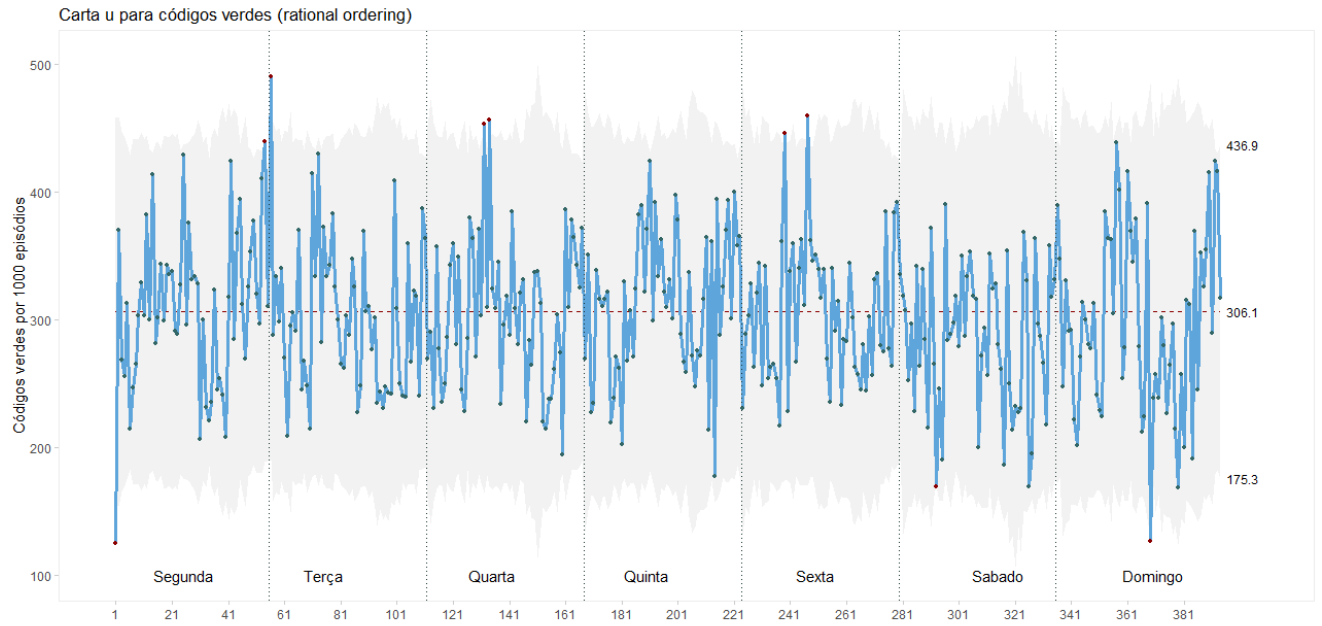


Figura 57: Carta U para códigos verdes por dia da semana (Rational Ordering)

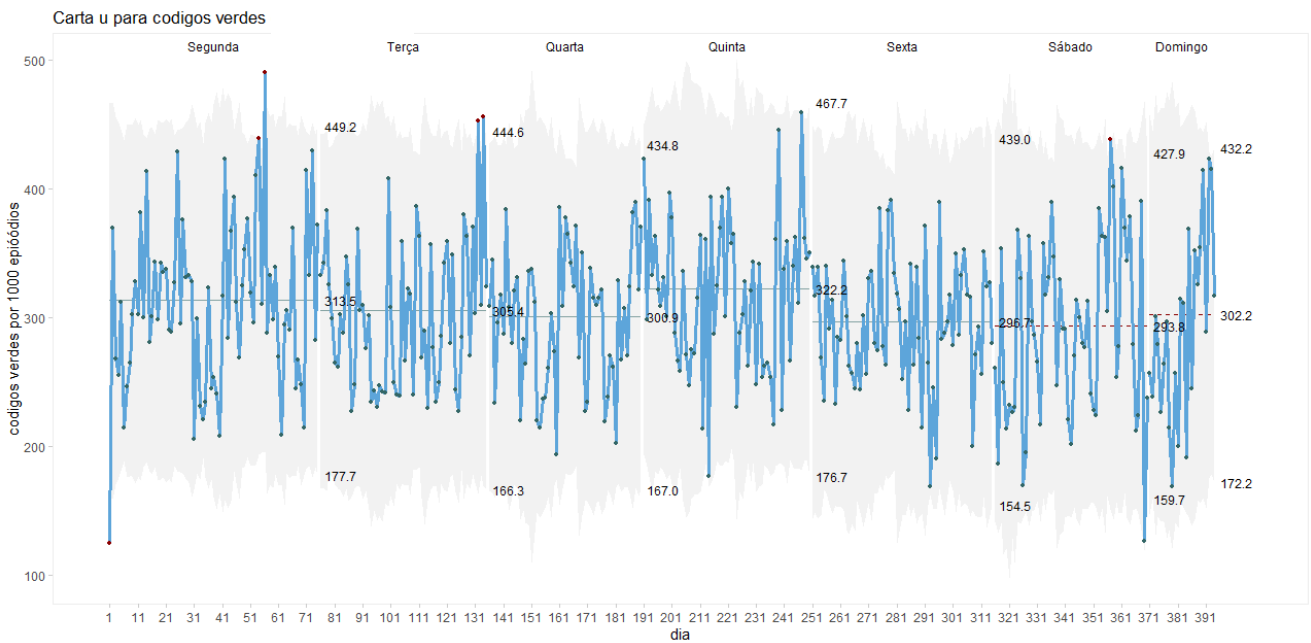


Figura 58: Cartas U para códigos verdes para cada dia da semana

### 5.9.1 Análise dos tempos de espera

Começa-se por representar os tempos de espera, de acordo com a cor da triagem, considerando valores até 12 horas. A decisão de eliminar observações cujo tempo de espera fosse superior a 12 horas, veio da parte do Departamento de Gestão do Hospital, uma vez que acima deste valor é considerado erro no registo. Foram consideradas as cores de triagem Verde, Amarelo, Laranja e Vermelho.

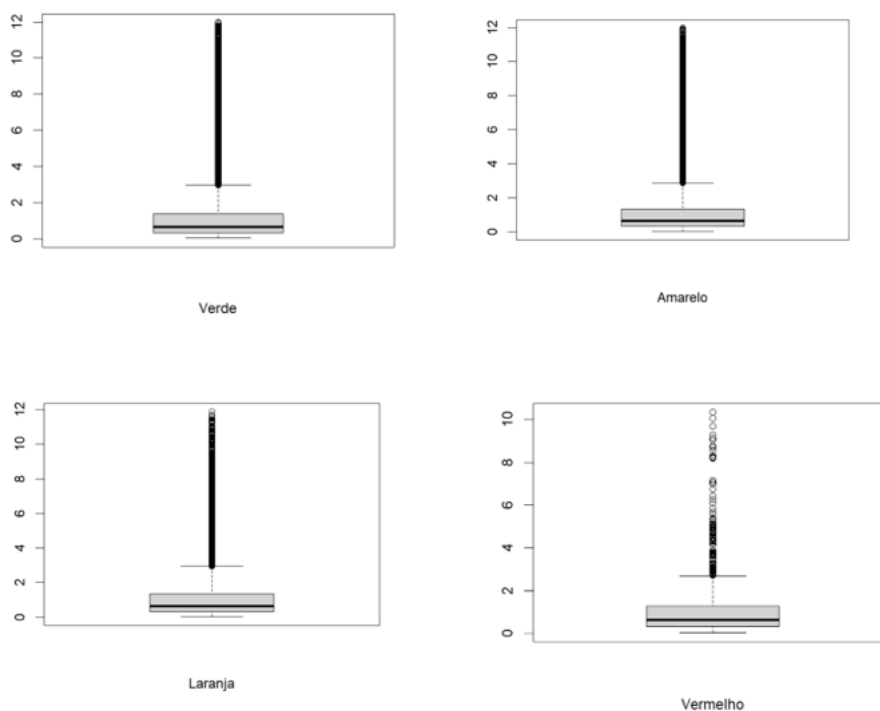


Figura 59: Boxplot para os tempos de espera por cor do código de triagem

Na Figura 59 pode ver-se a existência de muitos outliers, e estes mostram que em qualquer tipo de triagem, as especificações do Protocolo de Manchester não estão a ser cumpridas. Esta situação merece uma análise cuidada e nesse sentido construiu-se uma carta p, para a proporção de tempos de espera acima dos estabelecidos nas especificações do Protocolo de Manchester, Figura 60

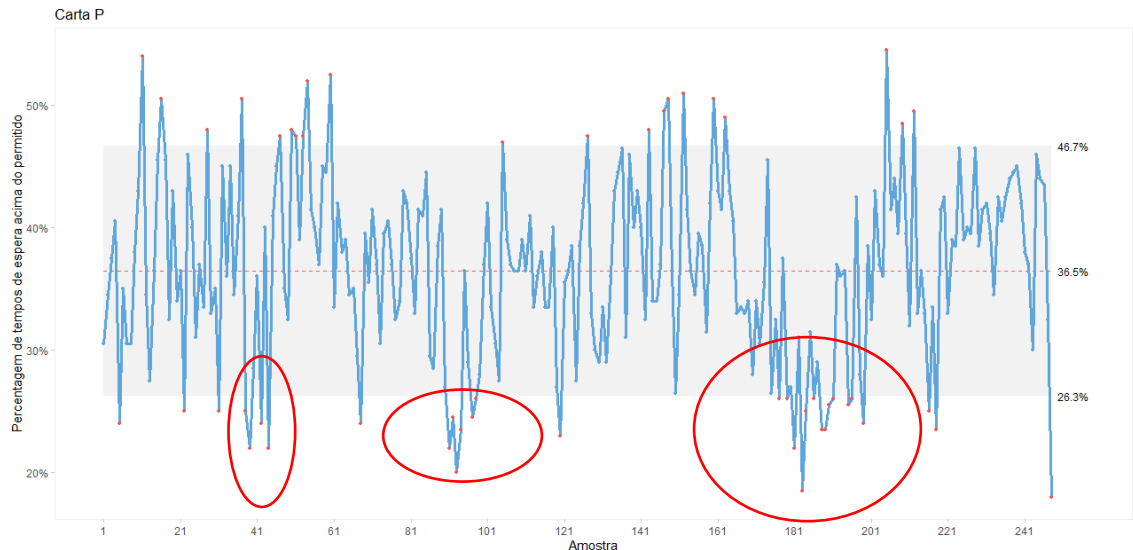


Figura 60: Carta P para a proporção de tempos de espera acima do especificado

Através da análise da carta para a proporção dos tempos de espera, registam-se alguns períodos que correspondem a uma melhoria do processo, vários pontos abaixo do LIC, e que se registam em três períodos distintos do ano. O primeiro período que representa uma melhoria do processo corresponde ao início do mês de julho de 2020. O processo volta a apresentar melhoria que corresponde ao mês de setembro de 2020, sensivelmente entre os dias 12 e 21. O período compreendido entre 177 e 200, corresponde aos meses de fevereiro e março de 2021, e mostra uma clara evidência de melhoria no processo para a proporção dos tempos de espera (todos os pontos abaixo do LIC). No que diz respeito ao período compreendido entre fevereiro e março de 2021, a melhoria deve-se á menor afluência ao SU uma vez que em janeiro se atingiu um pico trágico da COVID, que atingiu o máximo em janeiro de 2021. Relativamente aos outros dois períodos registados, e depois de se ter investigado se poderia estar relacionado com a cor da triagem, com o turno com maior ou menor volume de admissões e mesmo com o motivo de admissão, não se conseguiu apurar um motivo que justifique estes dois períodos de melhoria do processo. O HESE poderá com esta informação e através de recursos disponíveis, tentar perceber o que pode ter ocorrido que contribuiu para esta mudança na proporção dos tempos de espera acima do especificado de acordo com a cor dos códigos de triagem.



## 6 Conclusões

Um dos objetivos deste estudo era a caracterização do utilizador frequente do SU do Hospital Espírito Santo, EPE de Évora, através da criação de um perfil destes utilizadores e desta forma auxiliar na gestão de recursos e melhor responder às necessidades destes utilizadores. Conhecer o comportamento de um utilizador é essencial na medida em que se pretende identificar algum padrão de utilização que possa ser útil para elaborar modelos preditivos e explicativos dessa utilização, e com eles facilitar a distribuição dos recursos materiais e humanos, reduzindo custos e aumentando a qualidade do serviço prestado. Inicialmente dispúnhamos de dados relativos a um período compreendido entre 2018 e 2019 e posteriormente, a estes, foram adicionados dados até maio de 2021, constituindo assim um conjunto de registos que envolveram períodos distintos em termos de funcionamento de um Serviço de Urgência, causado pela Pandemia COVID-19. As diferentes fases da modelação, nomeadamente as dificuldades que foram surgindo, são resultado de um conjunto de dados cujo comportamento é manifestamente diferente tendo em conta as circunstâncias vividas a partir de 10 de Março de 2020. Até se chegar ao modelo final, registaram-se as limitações que justificaram o não ajustamento do modelo. Estas limitações derivam essencialmente da grande heterogeneidade dos dados uma vez que o período temporal 2018 – 2021 era constituído por comportamento em termos de afluência muito diferentes. Além disto, foi importante perceber que códigos vermelhos e idades inferiores a 18 anos poderiam não constituir um potencial utilizador do SU. Assim sendo, com base nos resultados obtidos um hiperutilizador é um indivíduo cujo indicador do código de triagem esteja entre 10 e 50 (códigos laranjas ou amarelos) tem cerca de 11 vezes mais chances de ser um hiperutilizador enquanto um utilizador cujo indicador se encontra entre os 50 e os 240. Indicadores que incluem códigos amarelos, verdes ou azuis, têm 4 vezes mais chances de ser um hiperutilizador. Concluiu-se também que a probabilidade de um indivíduo ser um hiperutilizador do Serviço de Urgência desce em cerca de 10% se residir fora do concelho de Évora e que com o aumento da idade, um indivíduo do sexo masculino tem mais hipóteses de ser um hiperutilizador do Serviço de Urgência. O modelo permitiu ainda concluir que indivíduos do sexo masculino, mais velhos, e que recorrem ao Serviço de Urgência no turno 2 (ou seja o período entre as 16h-00h), com maiores tempos de espera, têm mais chances de ser um hiperutilizador do Serviço de Urgência e que utilizadores que recorrem ao Serviço de Urgência no turno 3 (00h-08h), com tempos de espera superiores e por indicação do médico, têm mais de três vezes mais chances de ser um hiperutilizador do que um utilizador

que não recorre ao turno 3 (00h-08h) nem por indicação do médico e com um tempo de espera inferior.

Poderá ser necessária uma consciencialização da população para a utilização indevida do SU, no sentido em que existem alternativas a este serviço assim como criar programas educacionais integrados nas consultas de acompanhamento e avaliar o seu impacto junto das admissões consideradas não urgentes.

A modelação da afluência ao SU contou também com a monitorização da mesma, com recurso a cartas de controlo por código de triagem, contemplando o período entre maio de 2020 e maio de 2021. Utilizando a metodologia do Controlo Estatístico do Processo (SPC) no SU, pode concluir-se que, em termos globais, nomeadamente com a implementação de cartas de controlo adequadas e uma correta aplicação das técnicas estatísticas, o HESE poderá ganhar maior estabilidade na afluência ao SU no que diz respeito ao número de episódios e tempos de espera, melhorando a qualidade do seu serviço.

Verificou-se através das cartas propostas que, é possível ter um conhecimento amplo da afluência por cor do código da triagem tendo em conta o dia da semana e o turno em que foi registado o episódio. Foi também possível, através das cartas, verificar que a variação da afluência de acordo com o código de triagem sofre alterações e que estas advêm de diversos fatores. Para códigos laranjas, Domingo é o dia com maior número de admissões e para códigos amarelos a afluência reduz à 5ª feira. Códigos verdes têm uma variação oposta a códigos amarelos, à 5ª feira, uma vez que registam um aumento enquanto amarelos diminuem.

Para a proporção dos tempos de espera acima dos especificados de acordo com a cor do código de triagem, registaram-se três períodos que representam melhoria do processo, sendo um deles (entre fevereiro e março de 2021) justificado pelo anterior pico máximo em janeiro de 2021, e que originou uma quebra significativa na afluência ao SU, contribuindo diretamente para que os tempos de espera fossem mais facilmente cumpridos. Relativamente aos dois períodos não identificados, o hospital poderá analisar recorrendo aos meios disponíveis para que, com a informação dada pela carta, consiga concluir acerca do que pode ter contribuído para originar essa melhoria nos tempos de espera e com isso otimizar o funcionamento do serviço.

Como consequência deste estudo, sugere-se a aplicação dos métodos aqui descritos para a análise da afluência no seu normal funcionamento possibilitando uma análise desta no sentido de traçar estratégias com o objetivo de reduzir a afluência melhorando assim o funcionamento do SU.

## 7 Bibliografia

ARS Alentejo (ARS) + H. Espírito Santo - Évora, E.P.E. (abrev\_n4\_cs\_aces\_hosp)

Attia, M. W., & Edward, R. (1998). Effect of weather on the number and the nature of visits to a pediatric ED. *The American journal of emergency medicine*, 16(4), 374–375. [https://doi.org/10.1016/s0735-6757\(98\)90131-0](https://doi.org/10.1016/s0735-6757(98)90131-0)

Azevedo, A.L.M., & Costa, A.M. The narrow entrance door of Brazil's National Health System (SUS): an evaluation of accessibility in the Family Health Strategy. *Interface - Comunic., Saúde, Educ.*, v.14, n.35, p.797-810, out./dez. 2010.

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: official journal of the Society for Academic Emergency Medicine*, 8(1), 48–53. <https://doi.org/10.1111/j.1553-2712.2001.tb00550.x>

Berchet, C. (2015). Emergency Care Services: Trends, Drivers and Interventions to Manage the Demand. *OECD Health Working Papers*, 83. DOI:10.1787/5jrts344crns-en

Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 39(227), 357–365. <https://doi.org/10.2307/2280041>

Birch, M. W. (1964). The Detection of Partial Association, I: The  $2 \times 2$  Case. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 313–324. <http://www.jstor.org/stable/2984428>

Bliss, C. I. (1935). The Calculation of the Dosage-Mortality Curve. *Annals of Applied Biology*, 22, 134-167. <https://doi.org/10.1111/j.1744-7348.1935.tb07713.x>

Borrer, C.M., Champ, C.W., & Rigdon, S.E. (1998). Poisson EWMA Control Charts. *Journal of Quality Technology*, 30(4), 352-361, DOI: [10.1080/00224065.1998.11979871](https://doi.org/10.1080/00224065.1998.11979871)

Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P., & Fitzgerald, G. (2012). Predicting emergency department admissions. *Emergency medicine journal : EMJ*, 29(5), 358–365. <https://doi.org/10.1136/emj.2010.103531>

Boyle, J., Wallis, M., Jessup, M., Crilly, J., Lind, J., Miller, P., & Fitzgerald, G. (2008). Regression forecasting of patient admission data. *Annual International Conference of*

*the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2008, 3819–3822.*  
<https://doi.org/10.1109/IEMBS.2008.4650041>

Calegari, R., Fogliatto, F.S., Lucini, F.R., Neyeloff, J., Kuchenbecker, R.S., & Schaan, B.D. (2016). Forecasting Daily Volume and Acuity of Patients in the Emergency Department. *Computational and Mathematics methods in Medicine, 2016*. Disponível em | <https://doi.org/10.1155/2016/3863268>

Carey, R., & Stake, L. (2003). *Improving Healthcare with Control Charts: Basic and Advanced SPC Methods and Case Studies*. Milwaukee: ASQ quality Press.

Carret, M. L., Fassa, A. G., & Kawachi, I. (2007). Demand for emergency health service: factors associated with inappropriate use. *BMC health services research, 7*(1), 131. DOI:10.1186/1472-6963-7-131

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. <https://doi.org/10.1071/ah070083>

Chan, L-Y., Ouyang, J., & Lau, H. Y-K. (2007). A Two-Stage Cumulative Quantity Control Chart for Monitoring Poisson Processes. *Journal of Quality Technology, 39*(3) 203-223. DOI: [10.1080/00224065.2007.11917689](https://doi.org/10.1080/00224065.2007.11917689)

Christ, M., Grossmann, F., Winter, D., Bingisser, R., & Platz, E. (2010). Modern triage in the emergency department. *Deutsches Arzteblatt international, 107*(50), 892–898. <https://doi.org/10.3238/arztebl.2010.0892>

Comissão de Reavaliação da Rede Nacional de Emergência/urgência. (2012). Reavaliação da Rede Nacional de Emergência e Urgência - Relatório CRRNEU. Acedido em 15 de Maio de 2021 em: <https://www.dgs.pt/ficheiros-de-upload-2013/cnt-rel-crrneu-pdf.aspx>.

Coster, J.E., Turner, J.K., Bradbury, D., & Cantrell, A. (2017). Why Do People Choose Emergency and Urgent Care Services? A Rapid Review Utilizing a Systematic Literature Search and Narrative Synthesis. *Academic Emergency Medicine, 26* (9).

- Craney, T.A., & Surlles, J.G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, 14(3), 391-403. <https://doi.org/10.1081/QEN-120001878>
- Deming, W. E. (1986). *Out of the Crisis*. MA: Massachusetts Institute of Technology
- Despacho nº18459/2006 de 12 de setembro. Diário da República, nº 176 – II Serie- A. Lisboa: Ministério da Saúde. Disponível em <https://files.dre.pt/2s/2006/09/176000000/1861118612.pdf>
- Dharsi, A. (2006). The Future of Emergency Care in the United States Health System. *Journal of Pediatric Surgery*, 41(11) <https://doi.org/10.1016/j.jpedsurg.2006.07.016>
- Díaz, J., Alberdi, J. C., Pajares, M. S., López, C., López, 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(6), 709–712. <https://doi.org/10.1097/00007611-198106000-00020>
- Donabedian, A., 1980b. *The Criteria and Standards of Quality (Vol.II)*. Ann Arbor: Michigan Health Administration Press.
- Donabedian, A., (1980). *The Criteria and Standards of Quality. Journal of Ambulatory Care Management*, 6, 80. DOI:10.1097/00004479-198305000-00015
- Doupe, M. B., Palatnick, W., Day, S., Chateau, D., Soodeen, R. A., Burchill, C., & Derksen, S. (2012). Frequent users of emergency departments: developing standard definitions and defining prominent risk factors. *Annals of emergency medicine*, 60(1), 24–32. <https://doi.org/10.1016/j.annemergmed.2011.11.036>
- Durand, A. C., Gentile, S., Devictor, B., Palazzolo, S., Vignally, P., Gerbeaux, P., & Sambuc, R. (2011). ED patients: how nonurgent are they? Systematic review of the emergency medicine literature. *The American journal of emergency medicine*, 29(3), 333–345. <https://doi.org/10.1016/j.ajem.2010.01.003>
- Durand, A.C., Palazzolo, S., Tanti-Hardouin, N., Gerbeaux, P., Sambuc, R., & Gentile, S. (2012). Nonurgent patients in emergency departments: rational or irresponsible consumers? Perceptions of professionals and patients. *BMC Research Notes*, 5(1), 525. <https://doi.org/10.1186/1756-0500-5-525>

- Dyke, G. V., & Patterson, H. D. (1952). Analysis of Factorial Arrangements when the Data are Proportions. *Biometrics*, 8(1), 1–12. <https://doi.org/10.2307/3001521>
- Fang, Y. (2003). c-charts, X-charts, and the Katz Family of Distributions. *Journal of Quality Technology*, 35(1), 104-114, DOI: <https://doi.org/10.1080/00224065.2003.11980195>
- Feigl, P., & Zelen, M. (1965). Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics*, 21(4), 826–838. <https://doi.org/10.2307/2528247>
- Fisher R. A. (1922). On the mathematical foundations of theoretical statistics *Philosophical Transactions of the Royal Society of London. Series A*, 222. 309–368. <http://doi.org/10.1098/rsta.1922.0009>
- Friede, K. A., Osborne, M. C., Erickson, D. J., Roesler, J. S., Azam, A., Croston, J. K., McGonigal, M. D., & Ney, A. L. (2009). Predicting trauma admissions: the effect of weather, weekday, and other variables. *Minnesota medicine*, 92(11), 47–49.
- Fuhrmann, C. M., Sugg, M. M., Konrad, C. E., 2nd, & Waller, A. (2016). Impact of Extreme Heat Events on Emergency Department Visits in North Carolina (2007-2011). *Journal of community health*, 41(1), 146–156. <https://doi.org/10.1007/s10900-015-0080-7>
- Garvin, D.A. (1987). Competing in the Eight Dimensions of Quality. *Harvard Business Review*, 87, 101-109.
- Glasser, M. (1967). Exponential Survival with Covariance. *Journal of the American Statistical Association*, 62 (318), 561-568. DOI: [10.1080/01621459.1967.10482929](https://doi.org/10.1080/01621459.1967.10482929)
- Gomes, M.I., Figueiredo, F., & Barão, M.I. *Controlo Estatístico da Qualidade*. Edições SPE, 2ª Ed.
- Grupo de Trabalho Serviços de Urgência. (2019). Relatório Grupo de Trabalho – Serviços de Urgência. Acedido em 15 de Maio de 2021 em: <http://www.aenfermagemasleis.pt/2019/11/29/relatorio-do-grupo-de-trabalho-servicos-de-urgencia-apresenta-propostas-de-melhoria/>.
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science* 1, 297–318.
- Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall.

Holleman, D.R., Bowling, R.L., & Gathy, C. Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. *Journal of General Internal Medicine* 11, 237–239 (1996).

Hoot, N. R., Zhou, C., Jones, I., & Aronsky, D. (2007). Measuring and Forecasting Emergency Department Crowding in Real Time. *Annals of Emergency Medicine*, 49(6), 747–755. doi:10.1016/j.annemergmed.2007.01.017

Hosmer, D.W., Sturdivant. R. X., & Lemeshow, S. L. (2013). *Applied Logistic Regression* (3<sup>rd</sup> ed.). Wiley & Sons, Inc.

Governo da República Portuguesa - Ministério da Saúde (2023). *Monitorização Diária*. Acedido em 15 de Maio de 2021, em: <https://www.sns.gov.pt/monitorizacao-do-sns/servicos-de-urgencia/caracterizacao-urgencias/>

Hunt, K. A., Weber, E. J., Showstack, J. A., Colby, D. C., & Callaham, M. L. (2006). Characteristics of frequent users of emergency departments. *Annals of emergency medicine*, 48(1), 1–8. <https://doi.org/10.1016/j.annemergmed.2005.12.030>

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 : official journal of the Society for Academic Emergency Medicine*, 15(2), 159–170. <https://doi.org/10.1111/j.1553-2712.2007.00032.x>

Jones, S.S. (2007). Real-time demand forecasting in the emergency Department. AMIA Symposium proceedings. AMIA Symposium, 2007, 1-5

Khoo, M. B. (2004). A moving average control chart for monitoring the fraction non-conforming. *Quality and Reliability Engineering International*., 20(6), 617-635.

LaCalle, E. J., Rabin, E. J., & Genes, N. G. (2013). High-frequency users of emergency department care. *The Journal of Emergency Medicine*, 44(6), 1167–1173. <https://doi.org/10.1016/j.jemermed.2012.11.042>

LaCalle, E., & Rabin, E. (2010a). Frequent Users of Emergency Departments: The Myths, the Data, and the Policy Implications. *Annals of Emergency Medicine*, 56 (1), 42-48

<https://doi.org/10.1016/j.annemergmed.2010.01.032>

- Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, 20(8), 769–777. <https://doi.org/10.1111/acem.12182>
- McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., & Kelen, G. D. (2008). The challenge of predicting demand for emergency department services. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, 15(4), 337–346. <https://doi.org/10.1111/j.1553-2712.2008.00083.x>
- McCool, J.I., & Joyner-Motley, T. (1998). Control Charts Applicable When the Fraction Nonconforming is Small. *Journal of Quality Technology*, 30(3), 240–247, DOI: [10.1080/00224065.1998.11979852](https://doi.org/10.1080/00224065.1998.11979852)
- McHale, P., Wood, S., Hughes, K., Bellis, M. A., Demnitz, U., & Wyke, S. (2013). Who uses emergency departments inappropriately and when - a national cross-sectional study using a monitoring data system. *BMC Medicine*, 11(1), 258. <https://doi.org/10.1186/1741-7015-11-258>
- Ministério da Saúde (2023). *Hospital do Espírito Santo*. Acedido em 15 de Maio de 2021, em: <http://www.hevora.min-saude.pt/2019/12/27/o-hospital/>
- Montgomery, D.C. (2019). *Introdução ao Controlo Estatístico da Qualidade* (7<sup>o</sup> ed) Ed. Rio de Janeiro: LTC
- Montgomery, D.C. (2012). *Introduction to Statistical Quality Control* (7th Ed). Wiley.
- Montgomery, D.C., Peck, E.A., & Vining, G.G. (2021). *Introduction to Linear Regression Analysis* (6th Ed.). Hoboken: Wiley-Interscience
- Morais, M. C., & Natário, I. (1998) Improving an upper one side c-chart. *Communications in Statistics – Theory and Methods* 27, p. 353-364.
- Morais, M. C., & Pacheco, A. (2006a). Combined CUSUM-Shewhart schemes for binomial data. *Economic Quality Control*, 21, 43-57.



- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370–384. <https://doi.org/10.2307/2344614>
- Noble, J. H., LaMontagne, 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(5), 415–427. <https://doi.org/10.1097/00005650-197109000-00005>
- Oliveira, A. (2008). Hiperutilizadores e urgência. *Ata Médica Portuguesa*, 21(6), 553-558
- Perea-Milla, E., Pons, S. M., Rivas-Ruiz, F., Gallofre, A., Jurado, E. N., Ales, M. A., Jimenez-Puente, A., Fernandez-Nieto, F., Cerda, J. C., Carrasco, M., Martin, L., Cano, D. L., Gutierrez, G. E., Macías, R. C., & Garcia-Ruiz, J. A. (2007). Estimation of the real population and its impact on the utilisation of healthcare services in Mediterranean resort regions: an ecological study. *BMC health services research*, 7, 13. <https://doi.org/10.1186/1472-6963-7-13>
- Pines, J. M., & Yealy, D. M. (2009). Advancing the science of emergency department crowding: measurement and solutions. *Annals of Emergency Medicine*, 54(4), 511–3. doi:10.1016/j.annemergmed.2009.05.023
- Reis, L.G. (1995). Introdução à Avaliação da Qualidade em Serviços de Saúde. In: Avaliação da Qualidade do Uso da Ocitocina no Trabalho de Parto em Maternidades Públicas do Município do Rio de Janeiro. Rio de Janeiro: ENSP/FIOCRUZ.
- Reynold, M.R., & Stoumbos, Z.G. (2004). Should Observations Be Grouped for Effective Process Monitoring?, *Journal of Quality Technology*, 36(4), 343-366, DOI: [10.1080/00224065.2004.11980283](https://doi.org/10.1080/00224065.2004.11980283)
- Rocovich, C., & Patel, T. (2012). Emergency department visits: Why adults choose the emergency room over a primary care physician visit during regular office hours? *World Journal of Emergency Medicine*, 3(2), 91–7. <https://doi.org/10.5847/wjem.j.1920-8642.2012.02.002>
- Rotstein, Z., Wilf-Miron, R., Lavi, B., Shahar, A., Gabbay, U., & Noy, S. (1997). The dynamics of patient visits to a public hospital ED: a statistical model. *The American*

*journal of emergency medicine*, 15(6), 596–599. [https://doi.org/10.1016/s0735-6757\(97\)90166-2](https://doi.org/10.1016/s0735-6757(97)90166-2)

Schwartz, J., Samet, J. M., & Patz, J. A. (2004). Hospital admissions for heart disease: the effects of temperature and humidity. *Epidemiology (Cambridge, Mass.)*, 15(6), 755–761. <https://doi.org/10.1097/01.ede.0000134875.15919.0f>

Turkman, A., & Silva, G. (2000). *Modelos Lineares Generalizados – da teoria à prática*. Lisboa: Sociedade Portuguesa de Estatística

Wargon, M., Casalino, E. & Guidet, B. (2010). From Model to Forecasting: A Multicenter Study in Emergency Departments. *Academic Emergency Medicine*, 17 (9), 970-978. <https://doi.org/10.1111/j.1553-2712.2010.00847.x>

Wood, S. (2017). *Generalized Additive Models An Introduction With R*, 2nd Edition, 249.

Wu, Z., & Wang, Q. (2007). An NP Control Chart Using Double Inspections. *Journal of Applied Statistics*, 34(7), 843-855, DOI: [10.1080/02664760701523492](https://doi.org/10.1080/02664760701523492)

Wu, Z., & Lou, H. (2004) Optimal design of the adaptive sample size and sampling interval npcontrol chart. *Quality and Reliability Engineering International*, 20(6), 553-570

Zippin, C., & Armitage, P. (1966). Use of Concomitant Variables and Incomplete Survival Information in the Estimation of an Exponential Survival Parameter. *Biometrics*, 22(4), 665–672. <https://doi.org/10.2307/2528067>

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A., A., & Smith, G., M. (2009). *Mixed Effect Models and Extensions in Ecology*, Springer.

# Anexos

## A1. Modelos Lineares Generalizados – Dados 2018 a 2021

Figura A61.1: Receiver Operating Characteristic (ROC) do modelo ajustado

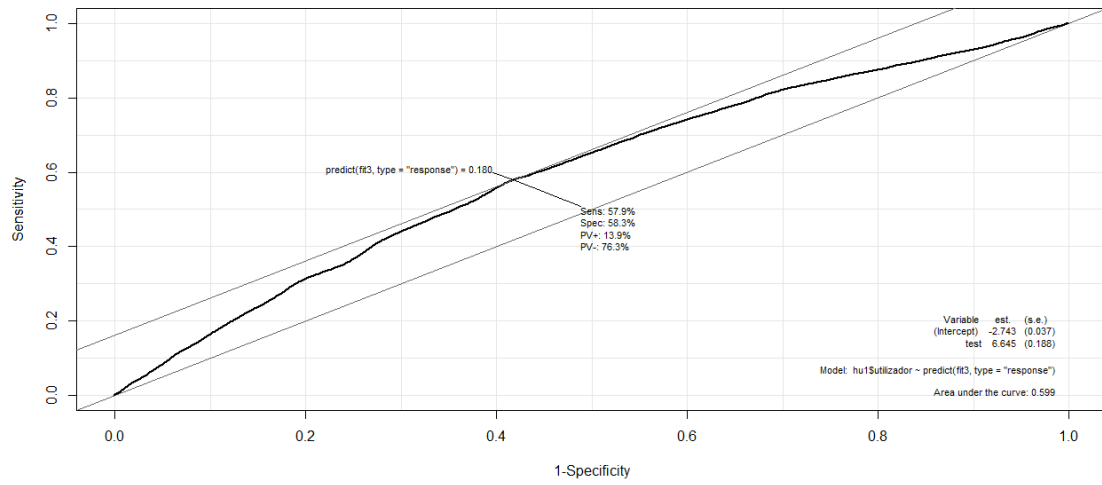


Figura A1.2: Linearidade com o logit - Idade

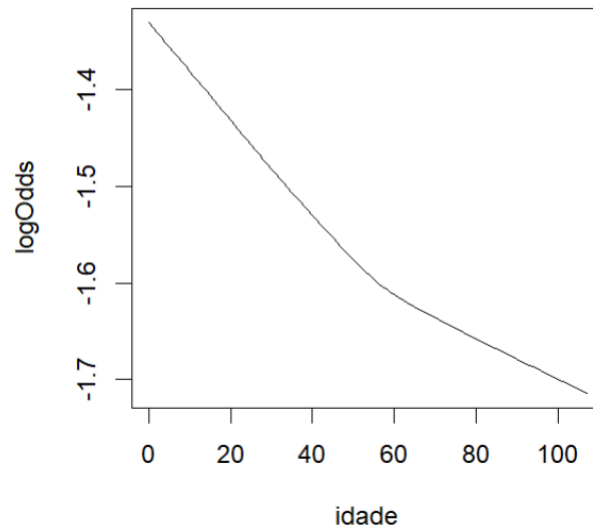


Figura A1.3: Linearidade com o logit – Tempo de espera

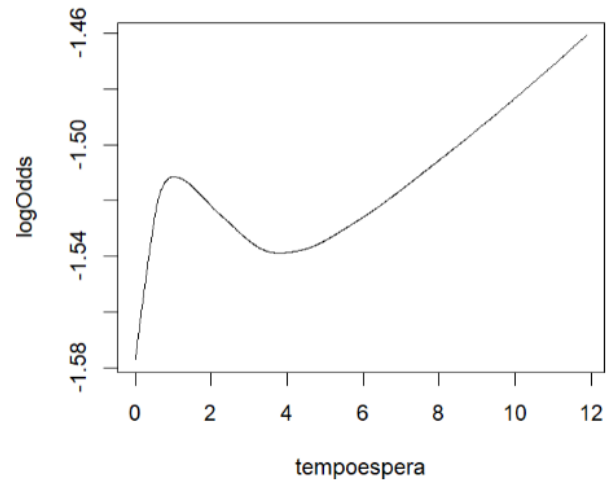


Figura A1.4: Linearidade com o logit – Percentagem de admissões em dias úteis

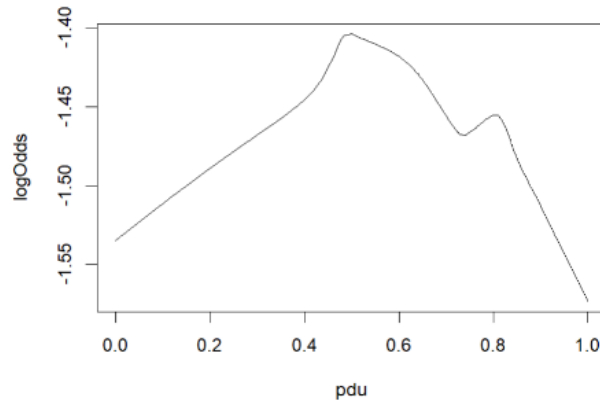


Figura A1.5: Receiver Operating Characteristic (ROC) do modelo ajustado com as transformações às variáveis contínuas

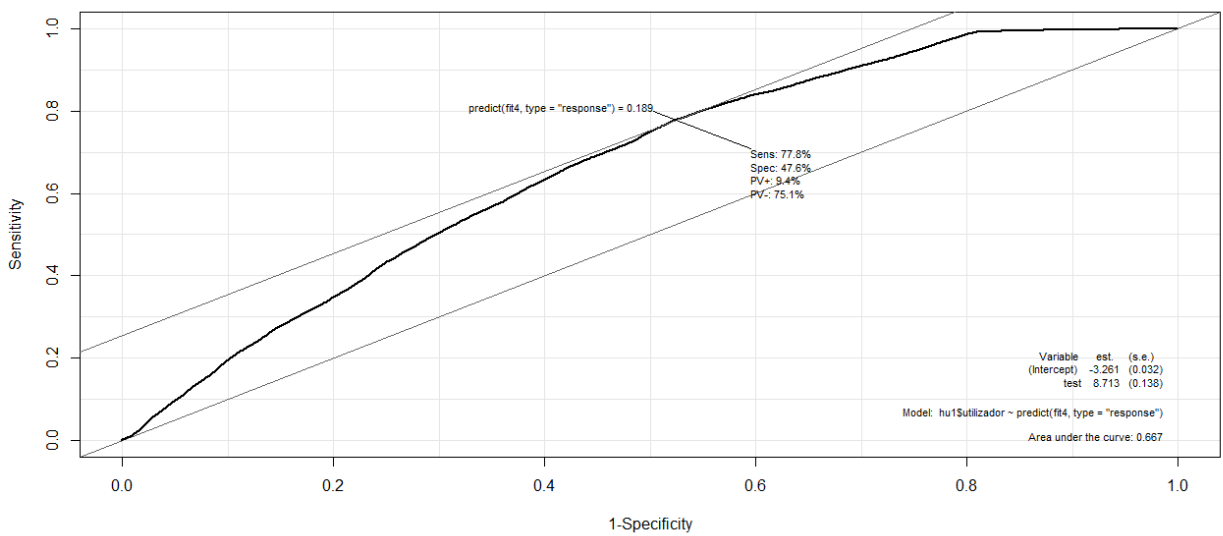


Figura A1.6: Receiver Operating Characteristic (ROC) do modelo ajustado com a variável idade categorizada

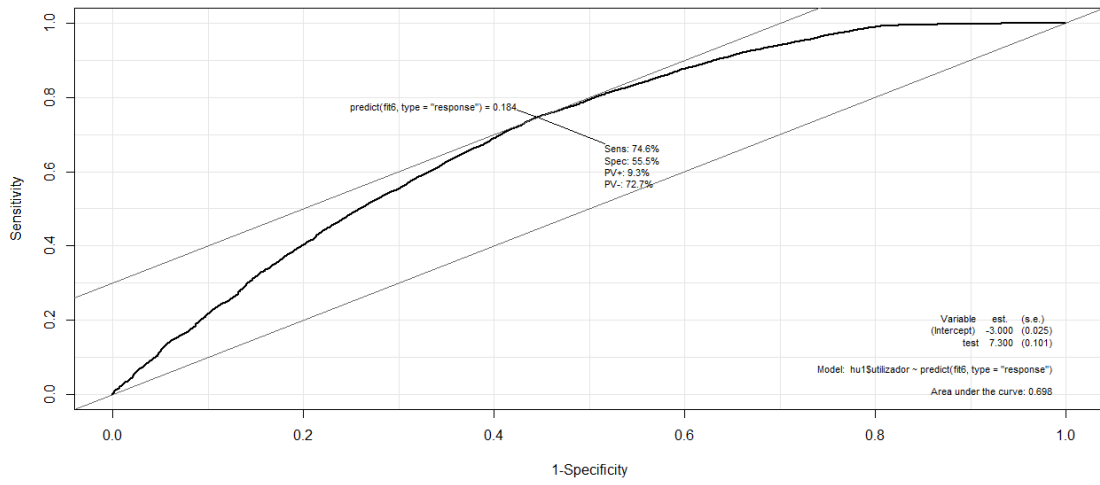
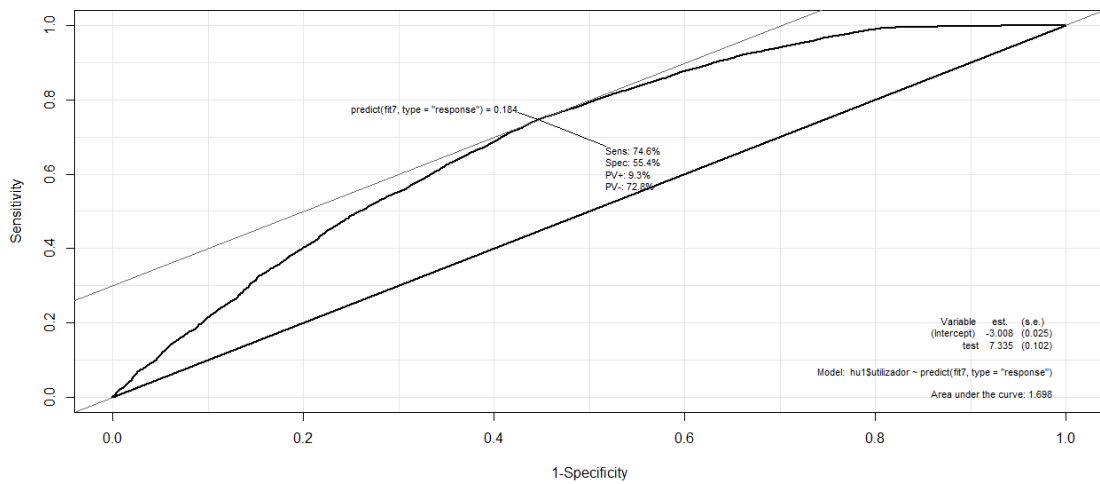


Figura A1.7: Receiver Operating Characteristic (ROC) do modelo ajustado com a junção de categorias na idade



## A2. Modelos aditivos generalizados

Figura A2.1: Ajustamento do GAM, k=8 graus de liberdade

```
Formula:
utilizador ~ sexo + idade + s(indicador, k = 8) + s(tempoespera,
  k = 8) + Pt2 + Pt3 + s(pdu, k = 8) + Medico

Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.7635785  0.0601330 -12.698 < 2e-16 ***
sexo1        -0.1281402  0.0266113  -4.815 1.47e-06 ***
idade        -0.0048920  0.0005313  -9.208 < 2e-16 ***
Pt21         -0.1085285  0.0323187  -3.358 0.000785 ***
Pt31         -0.2943198  0.0357255  -8.238 < 2e-16 ***
Medico1       4.1378011  0.5464344   7.572 3.66e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(indicador)  6.849  6.991  1083 <2e-16 ***
s(tempoespera) 6.501  6.850  1047 <2e-16 ***
s(pdu)         6.993  7.000  6682 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.54  Deviance explained = 49.4%
UBRE = -0.51838  Scale est. = 1          n = 82073
> |
```

Figura A2.2: Efeito não da variável tempo de espera com k=8

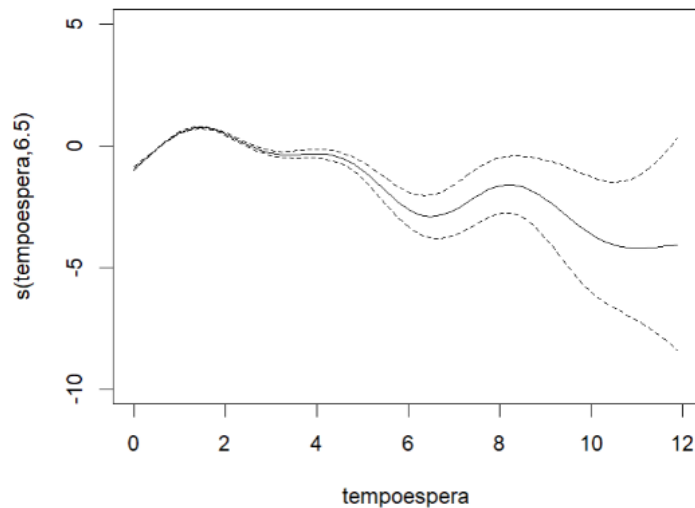


Figura A2.3: Efeito não da variável indicador com k=8

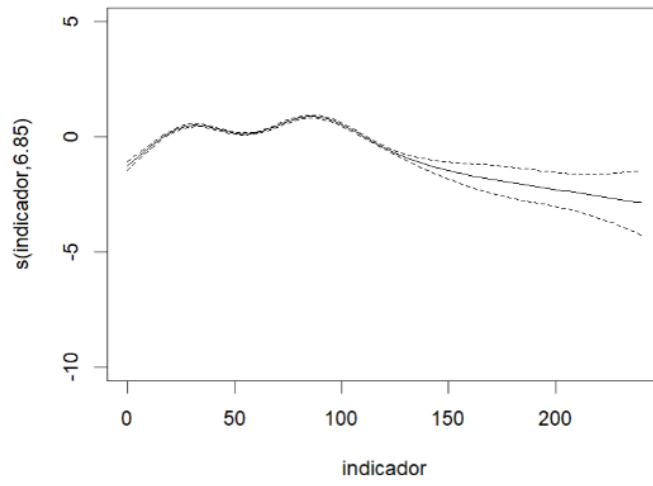
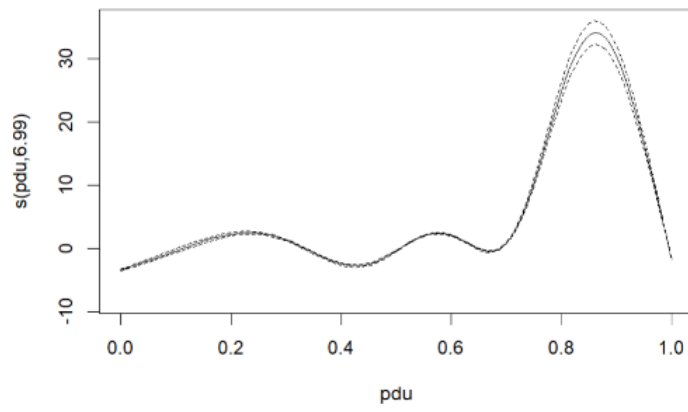


Figura A2.4: Efeito não da variável % de admissões em dias úteis com k=8





### A3 Modelo de regressão logística para dados pré-pandemia (1 de janeiro de 2018 a 10 de Março de 2020)

Figura A3.1: Receiver Operating Characteristic (ROC) do modelo de regressão logístico - pré-pandemia

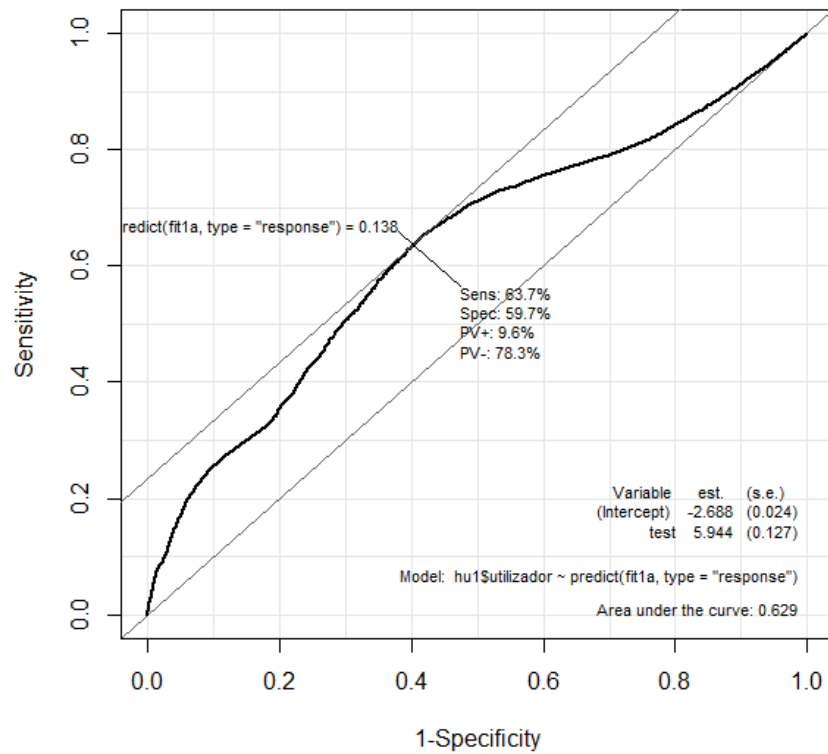


Figura A3.2: Receiver Operating Characteristic (ROC) do modelo de regressão logístico com o tempo de espera

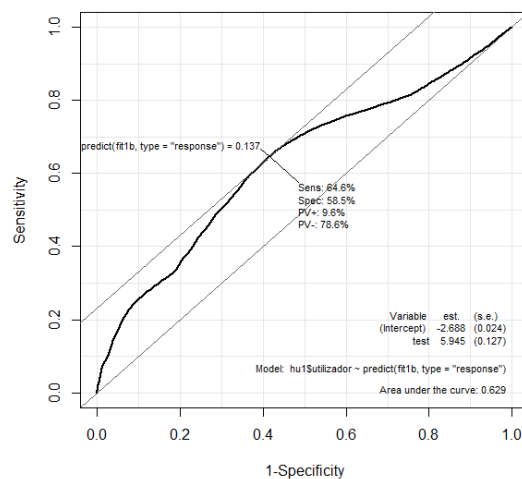


Figura A3.3: Linearidade com o logit - Idade

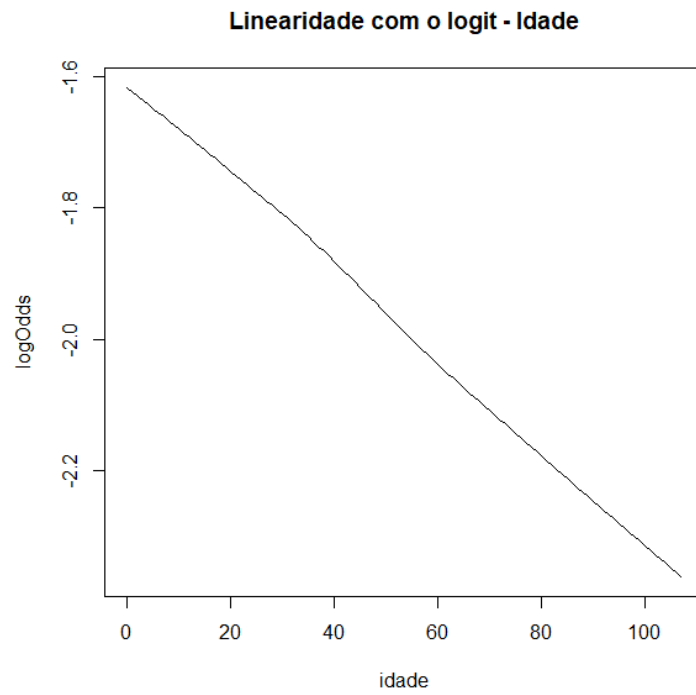


Figura A3.4: Linearidade com o logit – Tempo de espera

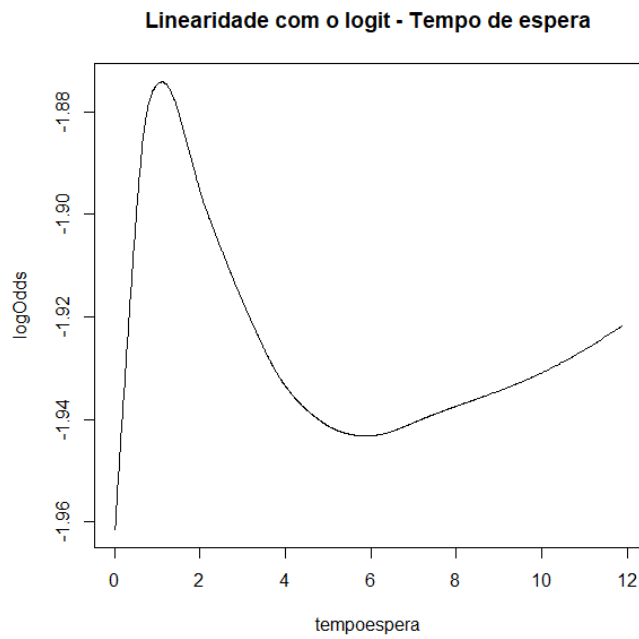


Figura A3.5: Linearidade com o logit – Indicador

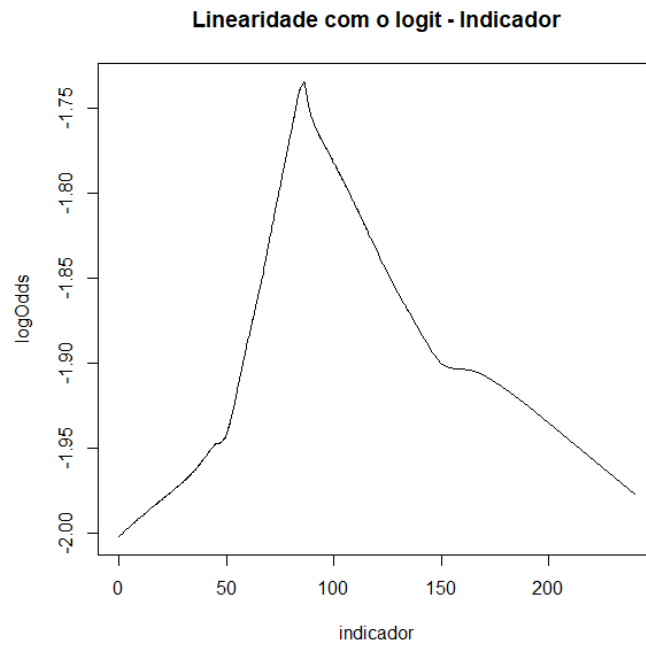


Figura A3.6: Receiver Operating Characteristic (ROC) do modelo de regressão logístico pré-pandemia com o tempo de espera e o indicador categorizados

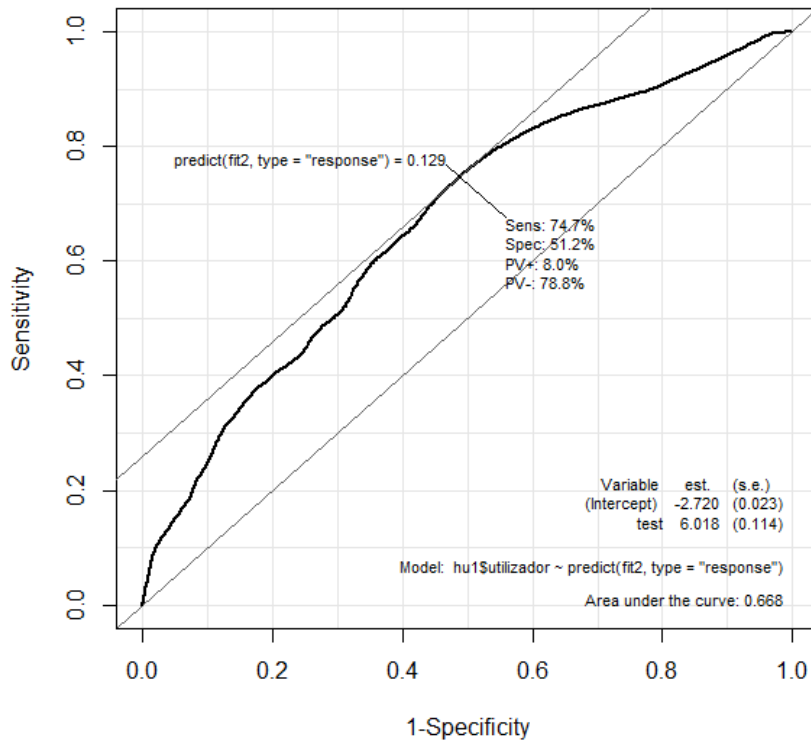
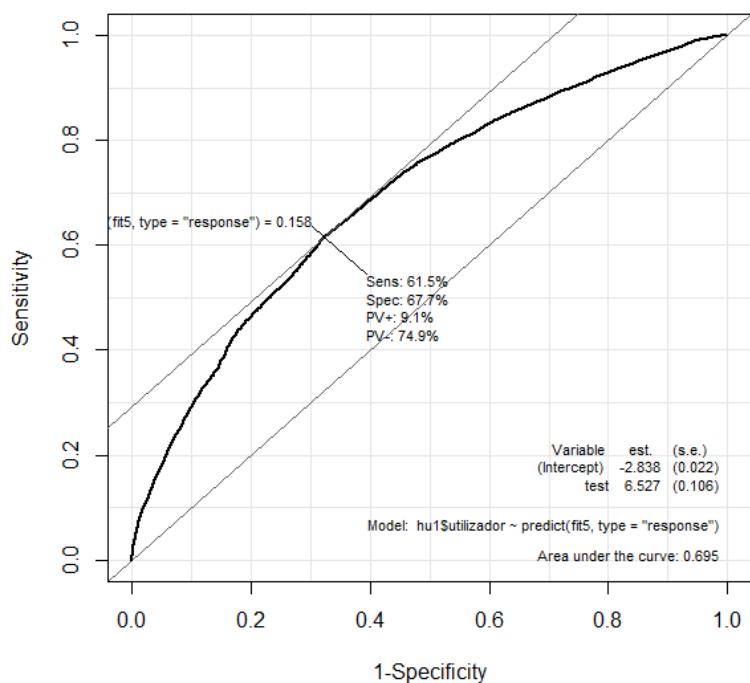


Figura A3.7: Receiver Operating Characteristic (ROC) do modelo de regressão logístico pré-pandemia do modelo com interações



A4: Modelo de regressão logística para dados de 1 de janeiro de 2018 a 10 de março de 2020, sem vermelhos e com idades superiores a 18 anos

Figura A4.1: Receiver Operating Characteristic (ROC) do modelo de regressão logístico pré-pandemia sem vermelhos e idades superiores a 18 anos

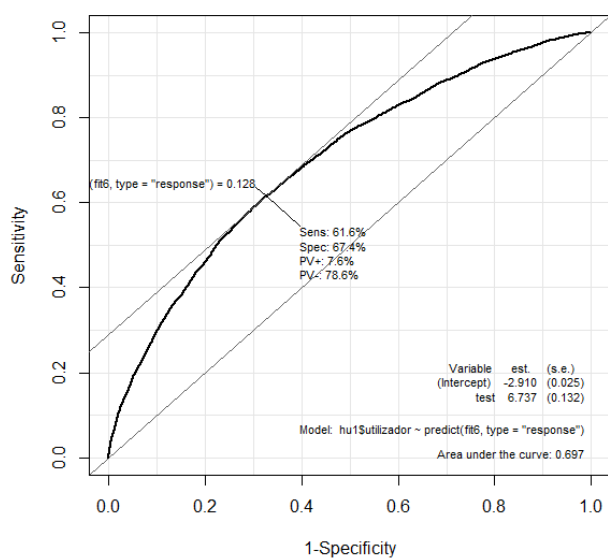


Figura A4.2: Resíduos deviance do modelo final

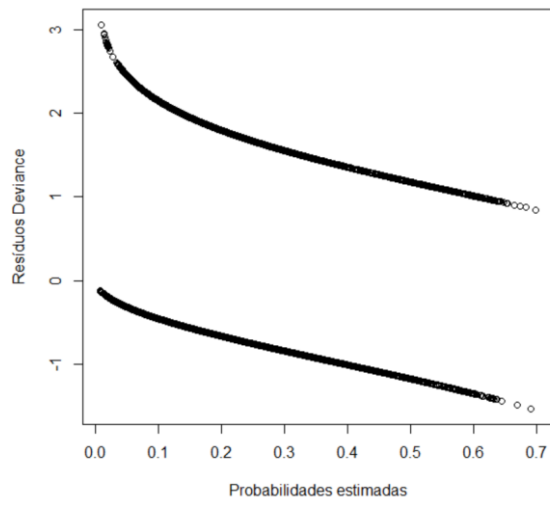


Figura A4.3: Distância de Cook

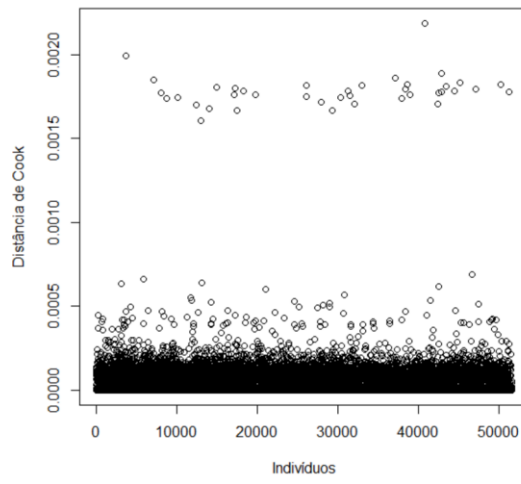


Figura A4.4: Gráfico semi-normal

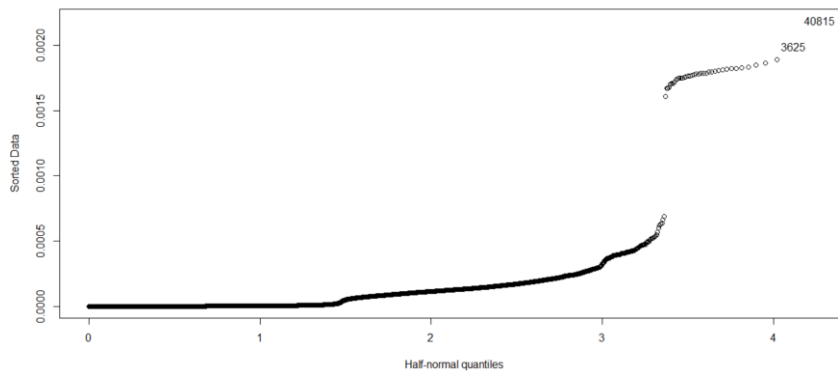


Figura A4.5: Dfbetas para as covariáveis do modelo final

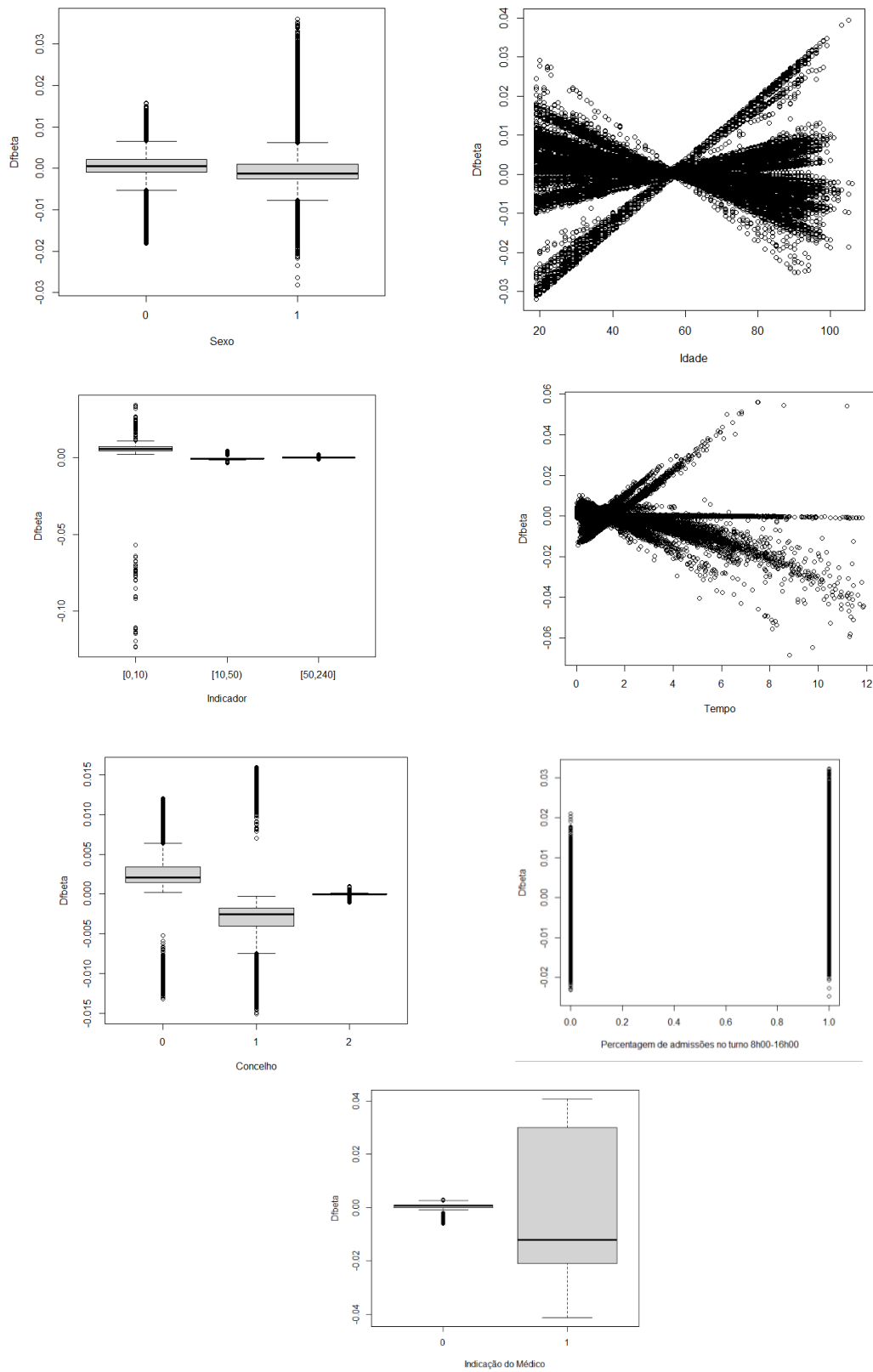


FIGURA A4.6:

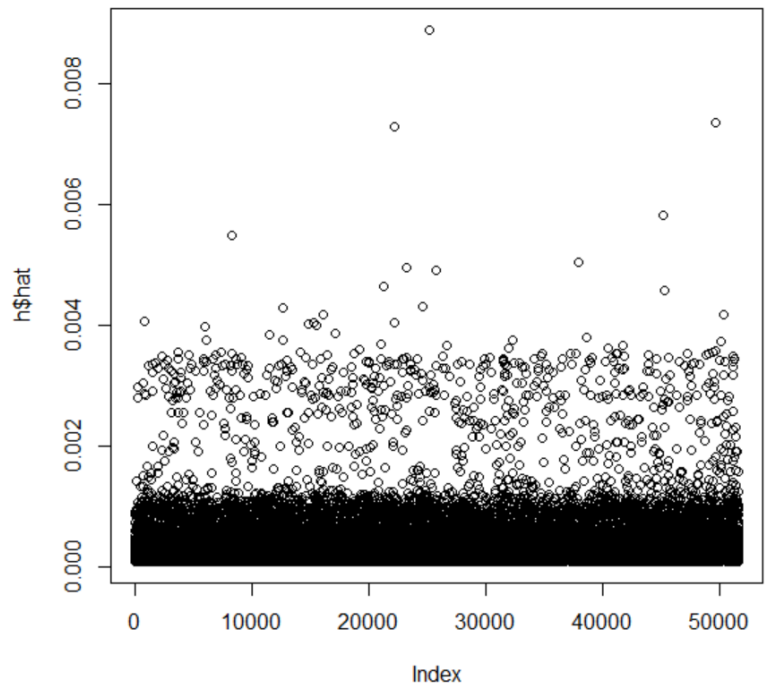
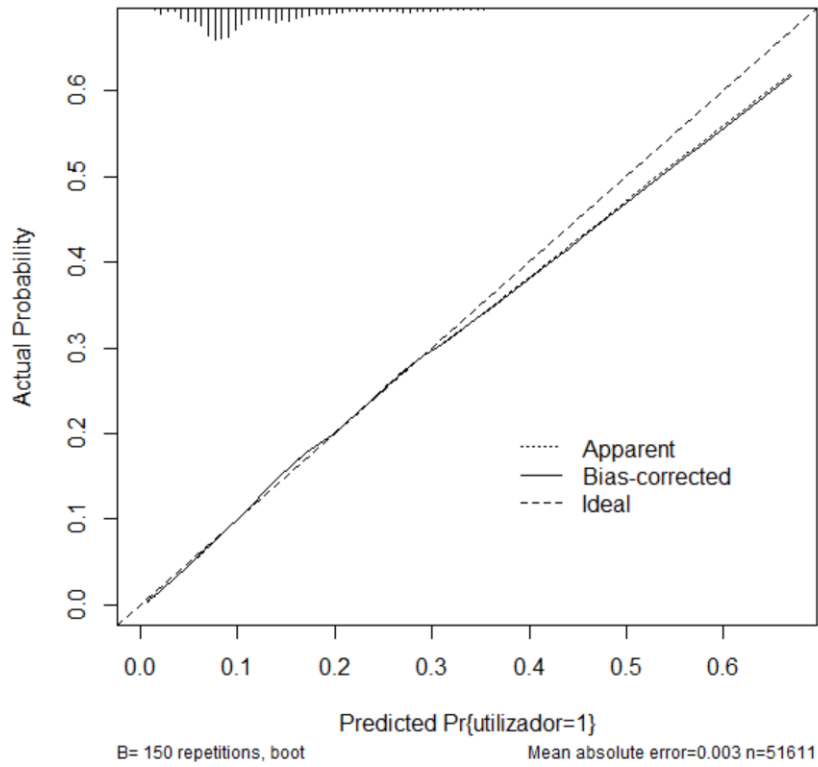


FIGURA A4.7:



## Significância do modelo

Model:

```
utilizador ~ sexo + idade + tempoHoras + IndicadorCat + concelho +  
  pt2CAT + pt3CAT + Medico + sexo * idade + idade * pt2CAT +  
  tempoHoras * pt3CAT + pt3CAT * Medico
```

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		36149	36179			
IndicadorCat	2	37238	37264	1088.88	< 2.2e-16	***
concelho	2	36169	36195	20.07	4.388e-05	***
sexo:idade	1	36252	36280	103.29	< 2.2e-16	***
idade:pt2CAT	1	36170	36198	21.19	4.156e-06	***
tempoHoras:pt3CAT	1	36156	36184	7.12	0.007602	**
pt3CAT:Medico	1	36165	36193	16.18	5.747e-05	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

~ |