

PREVISÃO DO TEMPO DE PERMANÊNCIA EM DOIS LARES DO DISTRITO DE SANTARÉM

ESTIMATION OF THE PERMANENCE TIME IN TWO RESIDENTIAL HOMES IN THE SANTARÉM DISTRICT

Pedro Sobreiro¹; Domingos Martinho²; António Pratas²; Marco Tereso²

*¹Instituto Politécnico de Santarém; ²ISLA Santarém
sobreiro@esdrm.ipsantarem.pt; domingos.martinho@islasantarem.pt;
antonio.pratas@islasantarem.pt; marco.tereso@islasantarem.pt*

Resumo

A previsão do tempo de permanência por parte dos utentes de lares, baseada na análise dos dados existentes, constitui um indicador da maior importância para os gestores deste tipo de equipamentos ajudando-os a planear os recursos necessários. Neste estudo desenvolveu-se uma análise para determinar o tempo de permanência de um idoso num lar e quais são as variáveis que influenciam esse tempo de permanência. Os dados foram analisados tendo por base o modelo de regressão de Cox onde se verificou que as variáveis faturação acumulada, Idade e distância da residência ao lar têm impacto no tempo de permanência. Na aplicação do teste log-rank encontramos diferenças na sobrevivência nas variáveis faturação acumulada e idade. Os resultados obtidos podem ser úteis para avaliação do tempo que decorre até ao abandono e dessa forma permitir planear os recursos necessários para acompanhar os utentes dos lares.

Palavras-chave: abandono, lares, tempo de permanência.

Abstract

Predicting the residence time of residential homes allows identifying if an event can occur, according to existing data, in order to plan the necessary resources. In this study, we developed a survival analysis to determine the length of time an elderly person lives in a nursing home and what are the variables that influence the length of stay. The dataset was analyzed using the Cox regression model and was found that variables accumulated billing; age and distance have impact on the length of stay. In the log-rank test we found differences in the survival in the variables accumulated billing and age. The results can be useful for evaluating the time until abandonment and allow planning the necessary resources to follow the users of residential homes.

Keywords: dropout, residential homes, length of stay.

1. INTRODUÇÃO

A previsão do tempo de permanência em lares permite avaliar os recursos necessários para acompanhar os idosos que os frequentam. Esta informação sobre o tempo de estadia nos lares é fundamental para prever os recursos financeiros necessários (Forder & Fernandez, 2011). A forma mais comum de se abandonar este tipo de instituições é devido à morte, sendo por isso importante estimar a duração da estadia após a entrada (Lee, Chau, Hui, Chan, & Woo, 2009). Apesar da importância de saber o tempo de estadia num lar, tem sido desenvolvida pouca investigação que nos permita ter uma perspetiva exata de quanto tempo as pessoas vivem nos lares antes de os abandonarem, na maioria das vezes, devido à morte (Forder & Fernandez, 2011).

Os métodos de análise baseados na aplicação de métodos estatísticos têm aumentado a sua popularidade em várias áreas, desde a medicina até à economia (Singer & Willett, 1993). De entre esses métodos a análise das probabilidades de sobrevivência tem sido uma das formas mais utilizadas para realizar a previsão de estadia em lares, sendo apresentada como uma curva de sobrevivência (Bland & Altman, 1998) que permite identificar os eventos de uma forma cronológica e avaliar simultaneamente quanto tempo decorre até que um evento ocorre (Williams, 2008).

A análise de probabilidades de sobrevivência consiste na aplicação de um conjunto de métodos estatísticos que permitem modelar o tempo até à ocorrência de um evento. O tempo de abandono é representado por T , que consiste numa variável aleatória que indica o período de tempo do evento ocorrer para um indivíduo selecionado aleatoriamente da população, representando a probabilidade de um evento ocorrer em cada período de tempo que não ocorreu já anteriormente, conhecido como uma função cronológica de risco (Singer & Willett, 1993). A função de sobrevivência é a probabilidade de um indivíduo sobreviver depois do tempo t , $S(t) = P(T > t)$, $t \geq 0$, com as características $S(0) = 1, S(\infty) = 0$. A distribuição da função é representada por F , definida como $F(t) = P(T \leq 0)$, para $t \geq 0$. A função de densidade da distribuição é representada com f onde:

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{P[t \leq T < t + dt]}{dt}$$

$f(t)dt$ representa a probabilidade de um evento ocorrer no momento t .

A necessidade de representar a evolução da probabilidade de ocorrer um evento ao longo do tempo, levou à definição da função de risco, representada como:

$$\lambda(t) = \lim_{dt \rightarrow 0^+} \frac{P[t \leq T < t + dt | T \geq t]}{dt}$$

$\lambda(t)$ representa a probabilidade de um evento no momento t .

O risco de sobrevivência h_j é definido como uma probabilidade condicional tal que um indivíduo experienciar um evento aleatoriamente num período de tempo j , considerando que não se deparou com o evento antes do momento j :

$$h_j = Pr[T = j | T \geq j]$$

A curva de sobrevivência baseada no risco de sobrevivência mostra como é que as taxas de abandono evoluem ao longo do tempo e como a tendência pode ser influenciada por outras variáveis.

As variáveis a utilizar neste tipo de análise podem ser quaisquer outros dados, utilizados como variáveis independentes, que podem influenciar o tempo de sobrevivência (Williams, 2008). Num estudo conduzido por Williams (2008) são consideradas variáveis independentes tais com: a idade, o sexo, a falta de apetite, os distúrbios no sono, a zona de residência, o historial de doenças crónicas (respiratórias, diabetes, etc.) e os medicamentos tomados. Este autor concluiu que essas variáveis são preditores no cálculo da probabilidade de mortalidade ocorrer. Mark, Alistair, Lindsay e Morris (2001) concluíram que as doenças crónicas, a falta de apetite e a medicação de cada um dos idosos são fatores que permitem determinar a probabilidade de ocorrer o óbito. Forder e Fernandez (2011) concluíram que a idade e o sexo são fortes preditores do tempo de estadia no lar. Ao contrário de Forder e Fernandez (2011), um estudo desenvolvido por Shah, Carey, Harris, DeWilde e Cook (2013) conclui que a idade e as características diagnósticas, nomeadamente as doenças, são fatores de fraca predição do risco de morte em idosos residentes em lares. Um estudo realizado na Suécia envolvendo 66 lares identifica o sexo, o número de doenças, os sintomas de depressão e as condições de nutrição como preditores significativos da sobrevivência (Lai, Ho, Chin, & Kwong, 2018). Dantas et al. (1991) utilizaram outras variáveis como as capacidades de independência, a continência e situações de doença crónica, como por exemplo a diabetes, para determinar o tempo de sobrevivência em lares.

Constata-se assim que existem muitas variáveis que podem ser utilizadas para determinar o tempo de permanência em lares. A idade, o sexo, a falta de apetite (condicionante das condições de nutrição), distúrbios no sono, histórico de doenças crónicas, os medicamentos tomados (associados aos diversos tipos de patologias), a zona de residência, são as variáveis que surgem com maior frequência como preditores na mortalidade.

Verifica-se ainda que, embora exista algum consenso sobre as variáveis que podem ser utilizadas, nem todos essas variáveis parecem ter a mesma influência como fator de predição.

2. METODOLOGIA

O objetivo deste estudo consiste na aplicação de técnicas de análise de sobrevivência para determinar o tempo de permanência de um idoso no lar e determinar as variáveis que têm impacto na predição.

O desenvolvimento deste estudo teve por base os dados disponibilizados pelos responsáveis de duas instituições do distrito de Santarém, contendo a informação de 601 pessoas distribuída por dois ficheiros, o primeiro respeitante aos dados dos utentes e o segundo referente à faturação realizada. Para garantir a privacidade dos dados

eliminou-se a morada, o nome e os contactos dos utentes. Procedeu-se ao cruzamento entre os dados dos utentes e da faturação tendo-se para o efeito assumido que os utentes sem registo de conta corrente não podiam ser considerados pelo que foram eliminados.

Após o pré-tratamento obtiveram-se 200 registos válidos incluindo as seguintes variáveis: sexo; data de nascimento; faturação acumulada (fatAcum); tempo de estadia; data de entrada; data de saída; idade; distancia da residência ao lar. A faturação acumulada foi calculada e inserida na variável fatAcum. O tempo de estadia foi calculado a partir da data de inicio e data fim e foi representado em número de meses. Para determinar a data de inicio utilizou-se a data da primeira fatura emitida ao utente e para data de fim a data da última fatura, considerando a última fatura a data a partir da qual ocorreram 60 dias sem emissão de faturas. Este procedimento foi adotado uma vez que nos registos de conta corrente não existia, de forma explicita, a informação referente à data de entrada e à data de saída. A distância da residência do utente à instituição foi calculada através de uma rotina que utiliza uma API para o google maps enviando o código postal da instituição e do utente e tratando o resultado obtido em XML. A idade do utente foi calculada a partir da data de nascimento. No final do tratamento dos dados obtiveram-se as seguintes variáveis: data inicio, data de fim, sexo, idade, estadia, distância e faturação acumulada (fatAcum).

O processamento dos dados foi realizado em Python (Continuum Analytics, 2016), usando Pandas (McKinney & others, 2010) e NumPy (Walt, Colbert, & Varoquaux, 2011). Nesta pesquisa, seguiu-se a metodologia de mineração de dados CRISP-DM (Processo Padrão da Indústria Cross para Mineração de Dados) (Wirth & Hipp, 2000), que consiste em seis etapas: (1) entender o domínio e desenvolver as metas do estudo, (2) compreensão de dados, (3) processamento de dados, (4) desenvolvimento de modelos, (5) avaliação e (6) implantação de modelos.

Para determinar a curva de sobrevivência a partir do momento inicial (data de início), momento final (data de fim) e situação da análise de sobrevivência usou-se o estimador Kaplan-Meier que fornece um método para estimar a curva de sobrevivência. Considerou-se que os coeficientes obtidos na regressão de Cox indicam o risco associado à respetiva variável (Grambsch & Therneau, 1994). O teste log rank permitiu realizar uma comparação estatística de dois grupos e utilizou-se o modelo de riscos proporcionais de Cox que permite incluir covariáveis adicionais (Bewick, Cheek, & Ball, 2004). O estimador de Kaplan-Meier, usado para obter informações sobre o tempo do evento de dropout, é o método mais amplamente usado para estimar as funções de sobrevivência (Efron, 1988). Simplificando a análise, as probabilidades de sobrevivência são apresentadas como uma curva de sobrevivência. A curva de sobrevivência é uma

representação das probabilidades de sobrevivência correspondentes a um tempo em que os eventos são observados (Bland & Altman, 1998). A análise de sobrevivência foi desenvolvida usando as linhas de vida do pacote (Cameron Davidson-Pilon et al., 2017).

3. RESULTADOS

Os resultados obtidos apresentam uma análise de sobrevivência considerando todos os utentes, tendo-se ainda avaliado a sobrevivência considerando o sexo, a idade, a distância ao lar e o valor da faturação acumulada (fatAcum).

3.1 Distribuição por sexo

Conforme se pode observar na figura 1, a amostra, composta por 200 registos, divide-se em 110 utentes do sexo feminino (55%) e 90 do sexo masculino (45%).

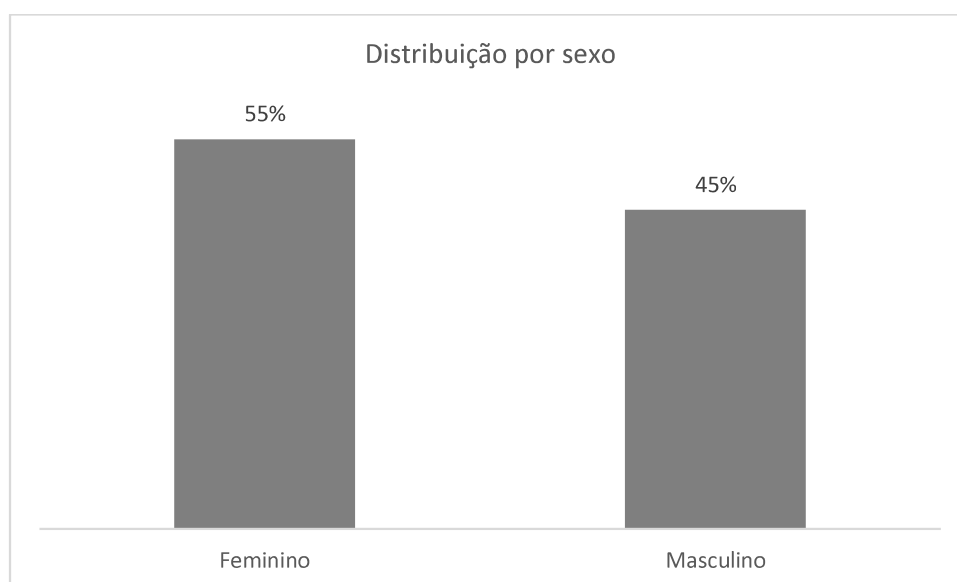


Figura 7. Distribuição dos utentes por sexo

3.2 Área de residência dos utentes

Os utentes residem principalmente em duas zonas geográficas: na zona centro e na zona de Lisboa, como se pode observar na figura 2.



Figura 8. Representação geográfica dos utentes

3.3 Curva de sobrevivência

A curva de sobrevivência foi calculada a partir da tabela de sobrevivência (*Quadro 2*), onde a coluna *Removed* contem o número de observações removidas durante o período de tempo (mês), quer por abandono, o valor indicado na coluna *Observed*, ou por censura (*Censored*), representando o total. A coluna *Entrance* representa as observações que entraram no estudo, ou seja todas observações analisadas começaram no momento 0 com 200 utentes e 0 em todos os outros momentos. A coluna *At_risk* contem as pessoas que ainda não abandonaram num determinado momento. O valor é igual no momento 0 em *Entrance* e *At_risk*. Nos momentos restantes o valor é igual à diferença entre o valor *At_risk* anterior e o valor *Removed* mais o *Entrance*. Se considerarmos o evento no momento 2 o valor *At_risk* é igual ao valor anterior *At_risk* menos o *Removed* mais o *Entrance*, i.e. $200 - 19 + 0 = 181$. Nos resultados obtidos nos primeiros 12 meses podemos comprovar que metade dos utentes abandona o lar.

Apesar de não termos dados censurados, o que representaria uma falta de informação, não se verifica o evento de abandono. Podem existir três tipos de censura: (1) *right censoring* caso de um evento não ocorrer durante o estudo, mas depois de terminar; (2) *left censoring*, se um evento ocorreu antes de se iniciar o estudo e (3) *interval censoring*, quando existe um momento em que ocorre o evento durante um intervalo de tempo, mas não se consegue determinar exatamente o momento em que ocorreu.

Event_at	Removed	Observed	Censored	Entrance	At_risk
0	0	0	0	200	200
1	19	19	0	0	200
2	8	8	0	0	181
3	13	13	0	0	173
4	7	7	0	0	160
5	10	10	0	0	153
6	6	6	0	0	143
7	11	11	0	0	137
8	6	6	0	0	126
9	16	15	1	0	120
10	1	1	0	0	104
11	3	3	0	0	103
12	8	8	0	0	100

Quadro 2. Eventos nos primeiros 12 meses

Depois da tabela de sobrevivência são calculadas as probabilidades de sobrevivência, onde a probabilidade de um indivíduo sobreviver até ao período i é:

$$p_i = \frac{r_i - d_i}{r_i}$$

Onde r_i é número de indivíduos que sobreviveram no início do período, d_i número de indivíduos que morreram durante o período de tempo. As probabilidades de sobrevivência estão representadas no *Quadro 3*, onde também foi adicionada a estimativa do tempo de sobrevivência considerando o mês onde se encontra (tEstimado).

Event_at	P_i	tEstimado
0	1.000	12
1	0.905	12
2	0.865	11
3	0.800	10
4	0.765	10
5	0.715	10
6	0.685	9
7	0.630	9
8	0.600	8
9	0.525	10
10	0.519	9
11	0.504	9
12	0.464	9

Quadro 3. Probabilidade de sobrevivência primeiros 12 meses

A probabilidade de sobrevivência calculada para cada mês é representada na curva de sobrevivência, apresentada na figura 3.

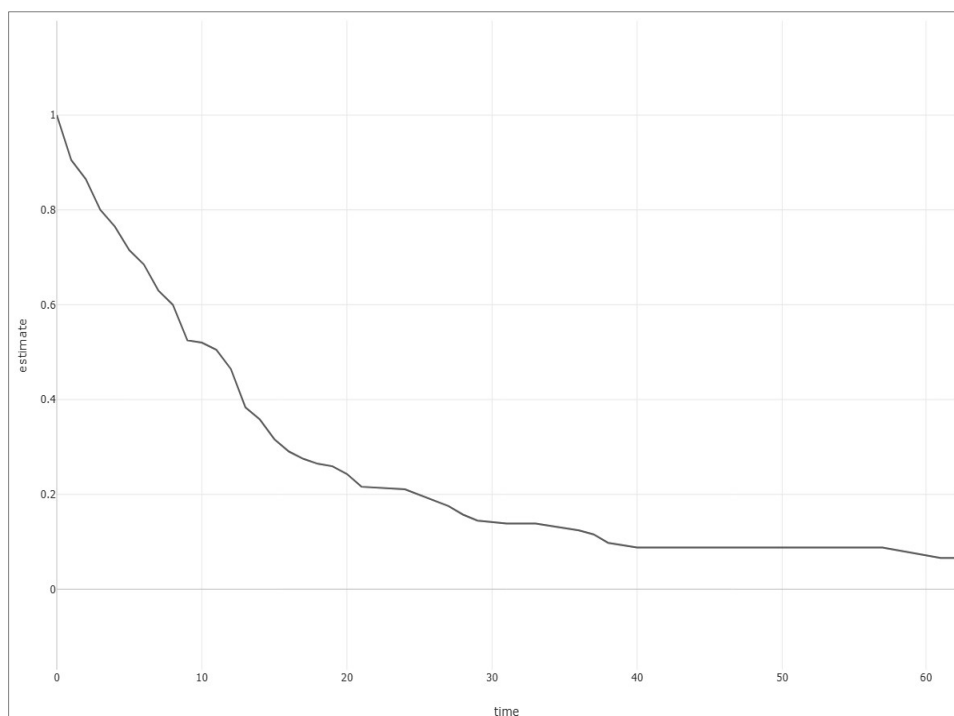


Figura 9. Curva de sobrevivência.

Os fatores que podem ter impacto na sobrevivência foram analisados através de uma regressão de Cox, onde um valor positivo apresenta um prognóstico pior na sobrevivência e um negativo um prognóstico melhor. O estudo permitiu determinar qual o impacto de cada uma das variáveis na permanência dos utentes dos lares, tendo-se analisado as variáveis: sexo, idade, faturação acumulada (facAcum) e distancia do lar à residência.

Os resultados obtidos na aplicação da regressão de Cox indicam o número de utentes que abandonaram (n_event) e o valor de p (likelihood_ratio_test_p_value) inferior a 0.05, permitindo concluir quais as variáveis com impacto no abandono (*Quadro 4*).

N	n_event	likelihood_ratio_test	likelihood_ratio_test_p_value
200	175	64.357291771449	3.513855872939e-13

Quadro 4: Estimativas dos parâmetros

A avaliação do contributo de cada uma das variáveis para o abandono é representada no *Quadro 5*. A curva de sobrevivência por sexo está de acordo com a regressão de Cox, onde o sexo masculino apresenta uma sobrevivência maior, mas um $p > 0.05$ não nos permite afirmar que tem impacto no evento de abandono.

As variáveis faturação acumulada (fatAcum), idade e distância apresentam um impacto, com significado estatístico, no evento de abandono ($p < 0.05$), conforme se pode verificar no *Quadro 5*, onde fatAcum, sexo e idade apresentam valores negativos no coeficiente, apesar de apenas fatAcum e idade terem valores significativos com ($p < 0.01$). A distancia da residência ao lar apresenta um valor positivo significativo ($p < 0.01$).

Term	Estimate	p_value
sexoMasculino	-0.08892	0.56536
fatAcum	-0.00006	0.00000
Idade	-0.02296	0.00013
Distancia	+0.00503	0.03834

Quadro 5: Estimativa dos parâmetros

Conforme referido anteriormente, utilizou-se o modelo de regressão de Cox para analisar as variáveis utilizadas, e como é que cada uma delas contribuiu para o evento de abandono tendo-se para o efeito construído um modelo para previsão da sobrevivência.

3.4 Sobrevivência por sexo, idade, distância e faturação acumulada

A tabela de sobrevivência, respetiva probabilidade de sobrevivência e o tempo estimado para o sexo feminino e masculino são apresentadas no *Quadro 6* e *Quadro 7*, respetivamente.

Event_at	Removed	Observed	Censored	Entrance	At_risk	P _i	tEstimado
0	0	0	0	110	110	1.000	12
1	8	8	0	0	110	0.927	12
2	7	7	0	0	102	0.863	11
3	9	9	0	0	95	0.781	10
4	3	3	0	0	86	0.754	10
5	8	8	0	0	83	0.681	10
6	3	3	0	0	75	0.654	9
7	7	7	0	0	72	0.590	9
8	2	2	0	0	65	0.572	8
9	11	10	1	0	63	0.481	10
10	0	0	0	0	63	0.481	9
11	3	3	0	0	52	0.454	9
12	4	4	0	0	49	0.416	9

Quadro 6. Eventos nos primeiros 12 meses por sexo feminino

Event_at	Removed	Observed	Censored	Entrance	At_risk	P _i	tEstimado
0	0	0	0	90	90	1.000	13
1	11	11	0	0	90	0.877	12
2	1	1	0	0	79	0.866	11
3	4	4	0	0	78	0.822	11
4	4	4	0	0	74	0.777	11
5	2	2	0	0	70	0.755	10
6	3	3	0	0	68	0.722	10
7	4	4	0	0	65	0.677	10
8	4	4	0	0	61	0.633	9
9	5	5	0	0	57	0.577	9
10	1	1	0	0	52	0.566	9
11	0	0	0	0	51	0.522	9
12	4	4	0	0	51	0.522	9

Quadro 7. Eventos nos primeiros 12 meses por sexo masculino

Na figura 4 representa-se a curva de abandono ao longo do tempo por sexo. Conforme se pode observar as duas curvas mantêm-se quase a par. Apesar disso pode afirmar-se que se verifica uma maior taxa de mortalidade relacionada com o sexo feminino.

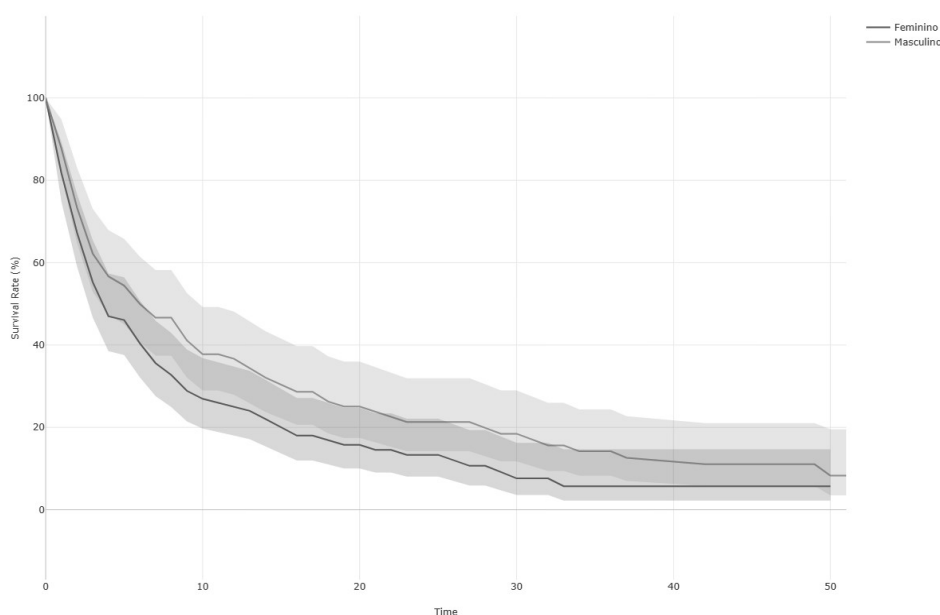


Figura 10. Curva de sobrevivência por sexo

A comparação das curvas de sobrevivência é realizada com o teste log-rank, onde realizamos o teste da hipótese nula assumindo que não existiam diferenças entre as curvas de sobrevivência. Obteve-se um valor do $\chi^2=2.1762$ com um $p=0.14$, que confirma que não se verificam diferenças na sobrevivência considerando o sexo.

Os resultados obtidos para a variável sexo foram também desenvolvidos para a idade, onde para simplificar a definição dos grupos se criaram cinco escalões a partir do cálculo

dos decis (20%,40%,60% e 80%). Na *Figura 11* pode observar-se a representação da sobrevivência por idade.

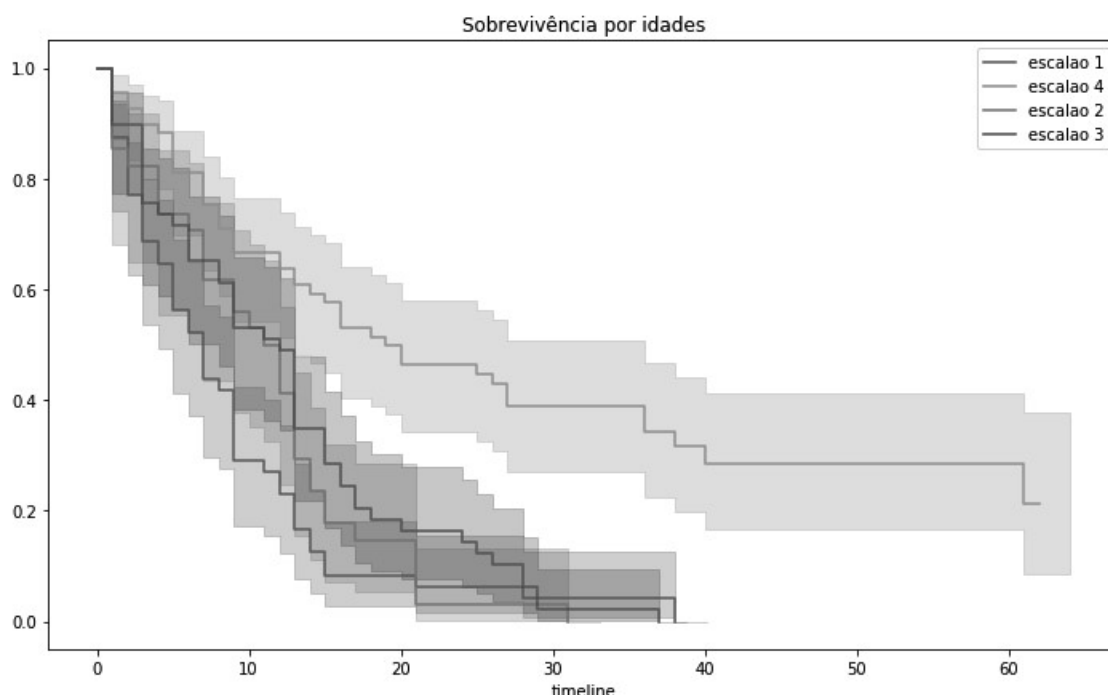


Figura 11. Curva de sobrevivência por idade

A comparação das curvas de sobrevivência é realizada com o teste log-rank, onde realizamos o teste da hipótese nula que assumindo que não existiam diferenças entre as curvas de sobrevivência. Obteve-se um valor do $\chi^2=45.648$ com um $p=0.000$, rejeitando a hipótese nula que assumia não existirem diferenças na sobrevivência considerando os escalões da idade. Estes valores obtidos permitiram aplicar um teste pairwise log-rank para percebermos onde se encontravam as diferenças. Foram encontradas diferenças entre os escalões 1-4, 2-4, 3-4 com $p<0.05$. Nos escalões 1-2, 1-3, 2-3 não existiram diferenças.

Na análise da sobrevivência por distância, criaram-se cinco escalões a partir do cálculo dos decis (20%,40%,60% e 80%). Na *Figura 12* apresenta-se a representação da sobrevivência por distância. Obteve-se um valor do $\chi^2=1.6158$ com um $p=0.4458$, não se rejeitando a hipótese nula que assumia não existirem diferenças na sobrevivência considerando os escalões da idade.

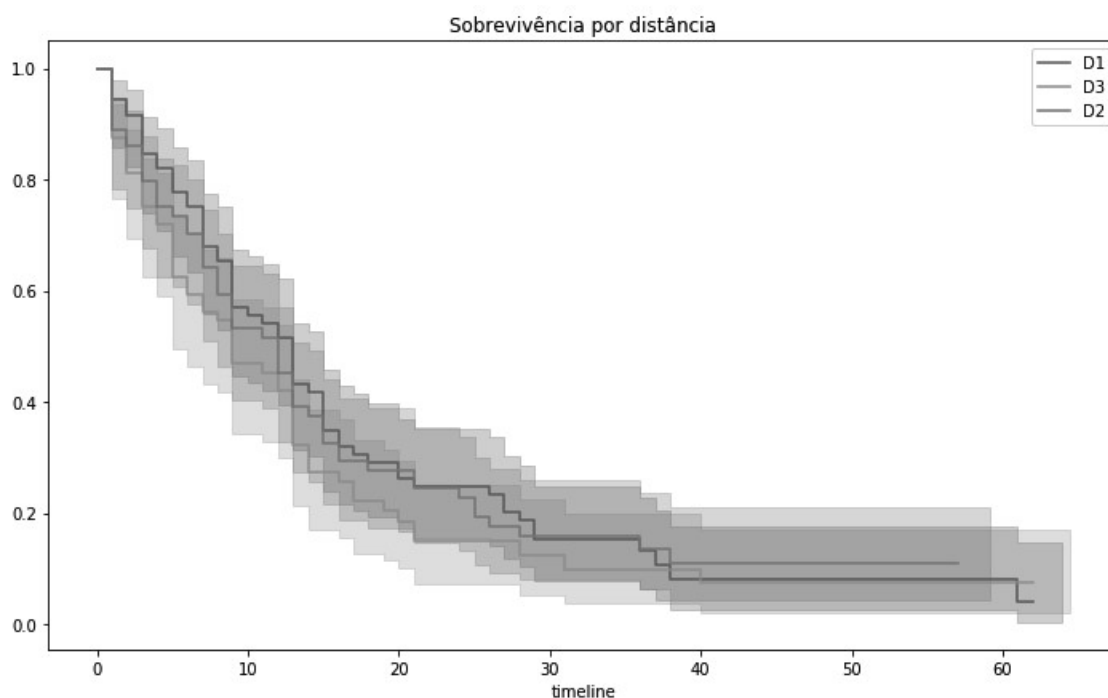


Figura 12. Curva de sobrevivência por distâncias

Na análise da sobrevivência por valor gasto, criamos três escalões a partir do cálculo dos decis (33% e 66%). Na *Figura 13* apresenta-se a representação da sobrevivência por distância. Obtivemos um valor do $\chi^2=59.508$ com um $p=0.0000$, rejeitando a hipótese nula que assumia não existirem diferenças na sobrevivência considerando os escalões valor gasto. Estes valores obtidos permitiram aplicar um teste pairwise log-rank para se perceber onde se encontravam as diferenças. Foram encontradas diferenças entre os escalões 1-2, 1-3 e 2-3 com $p<0.05$.

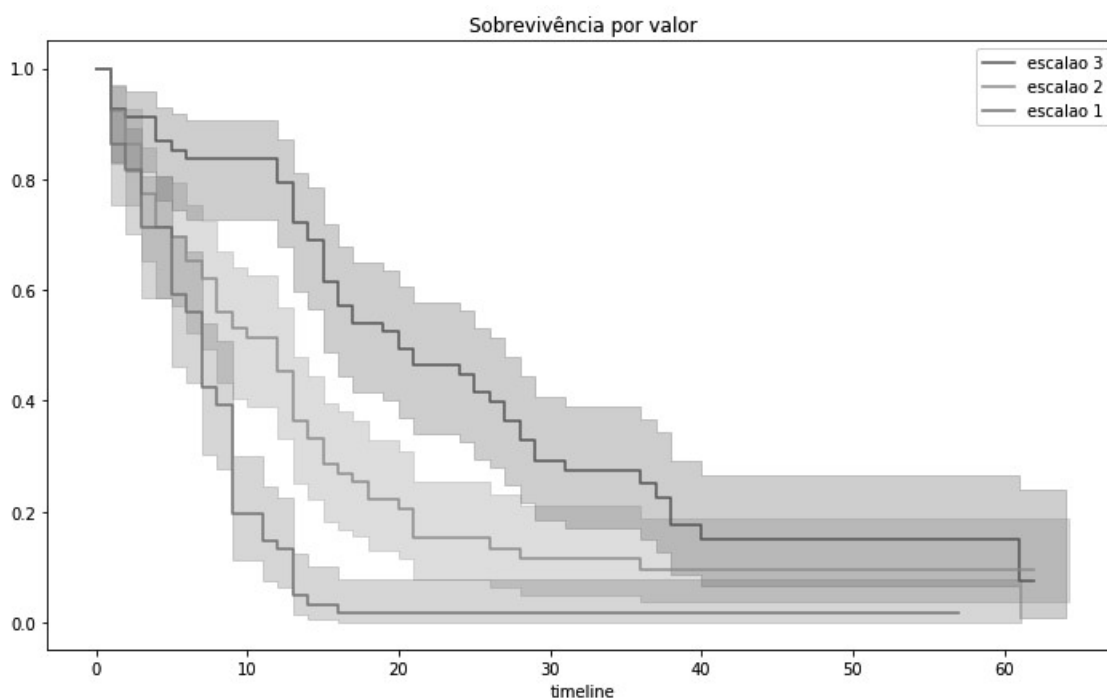


Figura 13. Curva de sobrevivência por valor gasto

4. DISCUSSÃO DOS RESULTADOS

Os resultados obtidos são diferentes dos verificados no estudo de Lee et al. (2009) que apresentam uma sobrevivência de 50% até aos primeiros dois anos. Forder e Fernandez (2011) referem 462 dias para metade dos residentes de um lar falecerem e Dale Mark et al. (2001) indicam o tempo médio de sobrevivência de 5.9 anos e Lai et al. (2018) indicam 22 meses.

Ao contrário dos estudos referidos anteriormente, os resultados obtidos neste estudo mostram que se verificam 50% de óbitos nos primeiros 12 meses de permanência. Contudo, isto pode dever-se às características dos lares analisados. Alguns lares apresentam taxas de mortalidade mais altas do que outros, apesar de poder ser relacionado com as características dos residentes (Dale Mark et al., 2001).

A maior taxa de mortalidade associado ao sexo feminino que se verifica neste estudo, não se verifica em outros estudos, onde sexo masculino apresenta uma menor sobrevivência (Dontas et al., 1991).

Os resultados mostram que existem três variáveis com impacto na sobrevivência, faturação acumulada (fatAcum), Idade e distancia. Na comparação da sobrevivência (logrank) de acordo com as variáveis fatAcum, idade, distância e sexo, não se obtiveram diferenças para as variáveis distância e sexo. Foram encontradas diferenças para as variáveis fatAcum e idade.

5. LIMITAÇÕES E TRABALHO FUTURO

Neste estudo apresentam-se resultados preliminares para analisar a sobrevivência de clientes de lares na região centro do país tendo por base uma amostra com uma dimensão de 200 elementos. A dimensão da amostra coloca alguns constrangimentos em relação à generalização dos resultados pelo que uma das vertentes do trabalho futuro consiste na ampliação do estudo a um universo alargado de lares, incorporando novas variáveis, de modo a que os resultados que vierem a ser obtidos possam ser generalizáveis.

Por outro lado, considera-se do maior interesse desenvolver uma investigação utilizando a aprendizagem baseada em computador de modo a prever o tempo de estadia no lar, o que constituirá uma nova abordagem uma vez que apesar de existirem alguns estudos, como Dale Mark et al. (2001) que utilizam o *logistic regression* para prever o tempo de vida, nenhum deles recorre à aprendizagem baseada em computador. Simultaneamente verifica-se o aparecimento de outras técnicas para a análise da sobrevivência, nomeadamente baseadas em redes neuronais (e.g. Baesens, Gestel, Stepanova, Poel, & Vanthienen (2005)), cuja aplicação seria interessante no contexto da investigação realizada.

REFERÊNCIAS

- Baesens, B., Gestel, T. V., Stepanova, M., Poel, D. V. den, & Vanthienen, J. (2005). Neural network survival analysis for personal loan data. *Journal of the Operational Research Society*, 56(9), 1089–1098. doi:10.1057/palgrave.jors.2601990
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: Survival analysis. *Critical Care*, 8(5), 389–394. doi:10.1186/cc2955
- Bland, J. M., & Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). *BMJ (Clinical Research Ed.)*, 317(7172), 1572.
- Cameron Davidson-Pilon, Jonas Kalderstam, Ben Kuhn, Andrew Fiore-Gartland, Luis Moneda, Alex Parij, ... André F. Rendeiro. (2017). *CamDavidsonPilon/lifelines: v0.13*. Zenodo. doi:10.5281/zenodo.1127755
- Continuum Analytics. (2016). *Anaconda Software Distribution*. Obtido de <https://continuum.io>
- Dale Mark C., Burns Alistair, Panter Lindsay, & Morris Julie. (2001). Factors affecting survival of elderly nursing home residents. *International Journal of Geriatric Psychiatry*, 16(1), 70–76. doi:10.1002/1099-1166(200101)16:1<70::AID-GPS277>3.0.CO;2-6

- Dontas, A. S., Tzonou, A., Kasviki-Charvati, P., Georgiades, G. L., Christakis, G., & Trichopoulos, D. (1991). Survival in a residential home: an eleven-year longitudinal study. *Journal of the American Geriatrics Society*, 39(7), 641–649.
- Efron, B. (1988). Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association*, 83(402), 414–425. doi:10.1080/01621459.1988.10478612
- Forder, J., & Fernandez, J.-L. (2011). Length of stay in care homes, 28.
- Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3), 515–526. doi:10.1093/biomet/81.3.515
- Lai, C. K. Y., Ho, L. Y. W., Chin, K. C. W., & Kwong, E. W. Y. (2018). Survival prediction among nursing home residents: A longitudinal study. *Geriatrics & Gerontology International*, 18(3), 428–433. doi:10.1111/ggi.13197
- Lee, J. S. W., Chau, P. P. H., Hui, E., Chan, F., & Woo, J. (2009). Survival prediction in nursing home residents using the Minimum Data Set subscales: ADL Self-Performance Hierarchy, Cognitive Performance and the Changes in Health, End-stage disease and Symptoms and Signs scales. *European Journal of Public Health*, 19(3), 308–312. doi:10.1093/eurpub/ckp006
- McKinney, W., & others. (2010). Data structures for statistical computing in python. Em *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56). SciPy Austin, TX. Obtido de <https://pdfs.semanticscholar.org/f6da/c1c52d3b07c993fe52513b8964f86e8fe381.pdf>
- Shah, S. M., Carey, I. M., Harris, T., DeWilde, S., & Cook, D. G. (2013). Mortality in older care home residents in England and Wales. *Age and Ageing*, 42(2), 209–215. doi:10.1093/ageing/afs174
- Singer, J. D., & Willett, J. B. (1993). It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics*, 18(2), 155–195. doi:10.3102/10769986018002155
- Walt, S. van der, Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13(2), 22–30. doi:10.1109/MCSE.2011.37
- Williams, C. S. (2008). Surviving Survival Analysis – An Applied Introduction. Institute for Advanced Analytics. Obtido de <https://analytics.ncsu.edu/sesug/2008/ST-147.pdf>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Em *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (pp. 29–39).

PERFIL ACADÉMICO E PROFISSIONAL DOS AUTORES

Pedro Sobreiro – Professor Adjunto na Escola Superior de Desporto de Rio Maior (Instituto Politécnico de Santarém) onde leciona Gestão de Sistemas de Informação e Estatística no curso de Gestão das Organizações Desportivas. Responsável pela investigação em Gestão do Desporto. Professor Convidado no ISLA Santarém onde colabora na realização de estudos e projetos de investigação. Realizou o Doutoramento em Gestão por Processos nas Organizações Desportivas, possui o mestrado e o título de especialista em Informática.

Domingos Martinho – Professor Coordenador e Diretor do ISLA Santarém. Mestre em informática, doutorado com especialização em tecnologias da informação e comunicação na educação pela Universidade de Lisboa. Professor Especialista em Ciências Informáticas. Desenvolve investigação nas áreas da aplicação das metodologias e tecnologias de e-learning na educação, aprendizagem baseada em computador e aplicação de business intelligence. Membro do CEPESE e da UI&D do ISLA Santarém onde coordena o núcleo de Informação e Tecnologia.

António Pratas – Professor Assistente no Isla Santarém. Licenciado em Informática de Gestão pelo ISLA Santarém. Professor Especialista em Ciências Informáticas. Membro do Núcleo de Informação e Tecnologia da UI&D do ISLA Santarém. Membro do Conselho Pedagógico e do Conselho Técnico-Científico do Isla Santarém.

Marco Tereso – Professor Assistente no ISLA de Santarém, onde leciona unidades curriculares de programação informática. Licenciado em Engenharia Informática, ramo de Tecnologias de Informação e Multimédia pela ESTGOH-IPC. Mestre em Informática e Sistemas, ramo de Desenvolvimento de Software pelo ISEC-IPC. Colabora na investigação e desenvolvimento de projetos da UI&D do ISLA de Santarém.

Endereço postal:

ISLA Santarém

Morada: Largo Cândido dos Reis, 2000-241 Santarém

Cidade (País): Portugal